STRUCTURE AND EVOLUTION OF HUMAN IMMUNOGLOBULIN C_γ GENES

Thesis by

Jay William Ellison

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1983

(Submitted November 3, 1982)

To my lover and best friend,

Barbara

Acknowledgements

I would like to thank my first research advisor, Eric Davidson, for the rigorous training I received in his lab, and my thesis advisor, Lee Hood, for the opportunity to study human genes in an independent atmosphere. I feel very fortunate in having been able to work under both of these men during my time here.

I also thank the many postdocs and graduate students with whom I've interacted for showing me both how to and how not to do science. I especially appreciate the help (and the wit) of that sagacious Texan, Steve Crews. I'm sure that one of these days he'll be able to go to his right for a ground ball.

I'm very grateful to the people who helped me rush to prepare this thesis—Bernita, Bertha, Ria, and Karyl—while I had my hand in the costodiaphragmatic pleural recess of my cadaver.

Lastly I'd like to thank my parents and close friends—Valerie, Elaine, and especially Barbara—for their encouragement and understanding during the rough times.

Abstract

In order to learn about the evolution of the human immunoglobulin C_γ gene family, the structural features of individual C_γ genes were examined. The complete nucleotide sequences were determined for three members of the gene family—the $C_{\gamma 1}$, $C_{\gamma 2}$, and $C_{\gamma 4}$ genes. A comparison of these sequences with those of the three reported mouse C_γ genes ($C_{\gamma 1}$, $C_{\gamma 2a}$, $C_{\gamma 2b}$) fails to reveal any pairs of corresponding genes in the two species. Moreover, the sequence homology shared by human C_γ genes in both coding and noncoding regions (about 95%) is significantly greater than that seen within the mouse C_γ family (about 70-80%). The presumably neutral mutations accumulated in the noncoding regions of the human genes have been used to estimate that approximately 6-8 million years have elapsed since the divergence of these genes from a common ancestral sequence. This divergence is considerably more recent than inferred for the mouse C_γ genes, and suggests that gene duplication or gene correction events have occurred more recently in humans than in mice.

In contrast to the C_H domain exons and adjacent noncoding regions, the hinge exons of human C_γ genes are quite divergent both in length and sequence. This coding sequence variability is seen to extend into the regions of C_H domains which border the hinge in the polypeptide chain. This divergence is interpreted as being the result of natural selection for particular hinge structures in the IgG subclasses. The implication is that these polypeptide regions are important for immunologic effector functions carried out by IgG molecules.

The arrangement of the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes in human chromosomal DNA has been determined to be 5'-$C_{\gamma 2}$-17 kilobase pairs-$C_{\gamma 4}$-3'. The genetic processes generating hybrid IgG molecules from these two genes are discussed, along with the relationship of these processes to gene duplication and gene correction.

Table of Contents

CHAPTER 1

Introduction

Immunoglobulin (Ig) molecules are composed of two heavy (H) chains and two light (L) chains. Each of these two types of polypeptides consists of an N-terminal variable (V_H or V_L) and a C-terminal constant (C_H or C_L) region. Both V_H and V_L regions, as well as C_L regions, are approximately 100-110 amino acid residues in length, while C_H regions are three to four times larger. This extra size is believed to be the consequence of multiplication of a primordial C_H coding unit, resulting in an ancestral C_H gene encoding four structural domains (C_H domains), each approximately the size of the V and C_L domains (Hill et al., 1966). There are five major types of C_H regions, reflecting the multiplication of the ancestral multidomain C_H gene, followed by divergence among the members of the resulting C_H gene family. These five types of C_H regions—C_μ, C_δ, C_γ, C_ε, and C_α, determine, respectively, the five major classes of antibody molecules—IgM, IgD, IgG, IgE, and IgA.

IgG is the dominant class of antibody in the serum of mammals, comprising about 70% of the total immunoglobulin in human serum. The human IgG pool consists of four closely-related subclasses—IgG1, IgG2, IgG3, and IgG4. The corresponding C_γ regions—$C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\gamma 3}$, and $C_{\gamma 4}$—are encoded by distinct germline genes (Kunkel et al., 1964; Martensson, 1966), which presumably arose through multiplication of an ancestral C_γ gene. More than one type of IgG (reflective of multiple C_γ genes) have also been observed in other mammals, as indicated in Table 1, although the number of subclasses varies for different species (Nisonoff et al., 1975).

In addition to the recognition and binding of foreign antigens by IgG molecules, other important biological activities are exhibited by these immunoglobulins. These activities, some of which are listed in Table 2, involve interactions of the C_γ regions with other biological molecules, in most cases specific cell-surface receptors. Although many of those functions or activities are exhibited by all four IgG subclasses, there are some variations. For example, IgG binding by receptors on several types of

cells appears stronger for the IgG1 and IgG3 subclasses than for IgG2 and IgG4, and IgG4 is the only subclass unable to bind the complement component C1q (for reviews or subclass variations see Spiegelberg, 1974; Unkeless et al., 1981; Winkelhake, 1978).

Such variability among the IgG subclasses is intriguing in view of their extensive structural similarities. The C_γ polypeptide regions are comprised of four protein domains: C_H1 (covalently linked to the N-terminal V_H region), C_H2, C_H3, and a small "hinge" region lying between the C_H1 and C_H2 domains. Protein sequencing studies have shown that the four subclasses of C_γ regions are over 90% homologous in their C_H domains, while hinge regions are much more divergent both in length and sequence (Edelman et al., 1969; Frangione et al., 1980; Pink et al., 1970; Wang et al., 1980). Particularly notable is the different number of cysteine residues in the four types of C_γ hinge, which results in different numbers of disulfide linkages between heavy chains in the four IgG subclasses. Schematic diagrams showing both these linkages and those between heavy and light chains are shown in Figure 1, where we see that γ1 polypeptides have a different point of attachment to light chains than the other types of γ chains.

The structural variability seen among the C_γ hinges could be interpreted to mean that these regions have little functional significance, and thus have diverged rather freely, while the C_H domains have been highly conserved because only certain structures will suffice to carry out the functions of the IgG molecules. Such a high degree of C_H domain conservation is not observed for mouse C_γ regions, where the level of sequence homology between individual C_γ genes is on the order of 70-80% (Miyata et al., 1980; Ollo et al., 1981; Yamawaki-Kataoka et al., 1981). This greater divergence within the mouse C_γ genes suggests another possible interpretation of the sequence relationships among the human C_γ regions: that the human C_γ genes have diverged from a common ancestral C_γ sequence more recently than their mouse

counterparts, and that hinge regions have evolved particularly rapidly since that divergence.

Indeed, some years ago it was proposed that the C_γ gene duplications giving rise to IgG subclasses in various mammals took place after the divergence of all modern mammals from a common ancestor (Pink et al., 1970), which is believed to have taken place about 80 million years ago (Romero-Herrera et al., 1973). Consistent with this idea is the observation of different numbers of IgG subclasses in different mammals (Table 1), which is interpreted as reflecting different frequencies of these duplication events.

The above model of mammalian C_γ gene evolution assumes that members of a given C_γ gene family evolve independently of one another, and that the accumulated differences in the C_γ genes reflect the time elapsed since the duplication event. Recent molecular analyses of mouse C_γ genes suggest that this simple model may not be correct. These studies indicate that, during evolution, genetic information has been exchanged between nonallelic mouse C_γ genes (Miyata et al., 1980; Ollo et al., 1981; Yamawaki-Kataoka et al., 1981). This interpretation implies that C_γ genes do not freely diverge from one another, but rather that sequence homology may be periodically renewed within the gene family. The processes responsible for this sequence homogenization result in the replacement of all or part of the sequence of one gene with the sequence of a homologous nonallelic gene. These processes have collectively been referred to as "gene correction" (Hood et al., 1975).

Two possible scenarios of mammalian C_γ gene evolution are shown in Figure 2, with reference to the genes of humans and mice. One model assumes that the common ancestor of these species had a single C_γ gene, and that the gene duplication events occurred separately in the two species following their divergence from one another. The other model pictures the original C_γ gene duplication as having occurred before mammalian speciation. Following divergence of humans and mice, gene correction

events have taken place (separately) in the two species. Since the human C_γ genes are more similar to one another than are the mouse genes, the pictured scenarios imply that either duplication or correction events have occurred more recently in the human evolutionary line.

The purpose of the research described in this thesis was to examine structural features of human C_γ genes to learn more about the evolution of this small multigene family. Although structural homologies in coding regions are evident from the protein sequence data, nucleotide sequence analysis of these regions would reveal how many codogenically silent nucleotide substitutions have been accumulated, and thus may tell us something about how strongly natural selection has acted at the amino acid level to maintain certain structures. The noncoding flanking and intervening regions are believed to be phenotypically silent, and thus not subjected to selective forces. Examination of these areas would thus reveal the extent of unselected divergence. By using calibrated values of the rate of unselected nucleotide substitution, we may then estimate the time of divergence of any two genes from an identical sequence. According to the discussions in the above paragraph and the models in Figure 2, we could thus estimate the time of occurrence of either (1) the original gene duplication event(s) which created or added to the C_γ gene family; or (2) the most recent gene correction event, which made some or all of the C_γ genes identical in sequence. It may even be possible to deduce the occurrence of gene correction events, if structural remnants, or "footprints" of these events exist.

The mouse C_γ genes have recently been rather well characterized, including the determination of their tandem linkage relationships (Honjo et al., 1979; Tucker et al., 1979; Yamawaki-Kataoka et al., 1980, 1982; Shimizu et al., 1982). Comparisons of human and mouse C_γ genes may reveal important relationships between the gene families in the two species. It should also be interesting to determine the

molecular linkage relationships of the human C_γ genes, in order to confirm or refute the tentative linkage assignments based on protein and genetic analyses (Natvig et al., 1967; Kunkel et al., 1969; Natvig and Kunkel, 1974).

Finally, this work may shed some light on the mode of evolution of the hinge regions of human IgG molecules. Analyses of these regions may yield implications regarding such issues as the types of nucleotide substitutions occurring during evolution, natural selective forces, and immunoglobulin structure and function.

REFERENCES

Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., and Waxdal, M. J. (1969). The covalent structure of an entire γG1 immunoglobulin molecule. Proc. Natl. Acad. Sci. USA **63**, 78-85.

Frangione, B., Rosenwasser, E., Prelli, F., and Franklin, E. C. (1980). Primary structure of human γ3 immunoglobulin deletion mutant: γ3 heavy-chain disease protein Wis. Biochemistry **19**, 4304-4308.

Hill, R. L., Delaney, R., Follow, R. E., and Lebowitz, H. E. (1966). The evolutionary origins of the immunoglobulin. Proc. Natl. Acad. Sci. USA **56**, 1762-1769.

Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N., and Mano, Y. (1979). Cloning and complete nucleotide sequence of mouse immunoglobulin γ1 chain gene. Cell **18**, 559-568.

Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975). The organization, expression, and evolution of antibody genes and other multigene families. Ann. Rev. Genet. **9**, 305-353.

Kunkel, H. G., Allen, J. C., Grey, H. M., Martensson, L., and Grubb, R. (1964). A relationship between the H chain groups of 75 γ-globulin and the Gm system. Nature **203**, 413-414.

Kunkel, H. G., Natvig, J. B., and Joslin, F. G. (1969). A "Lepore" type of hybrid γ globulin. Proc. Natl. Acad. Sci. USA **62**, 144-149.

Martensson, L. (1966). Genes and immunoglobulins. Vox Sang. **11**, 521-545.

Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M., and Honjo, T. (1980). Nucleotide sequence divergence of mouse immunoglobulin γ1 and γ2b chain genes and the hypothesis of intervening sequence-mediated domain transfer. Proc. Natl. Acad. Sci. USA **77**, 2143-2147.

Natvig, J. B., Kunkel, H. G., and Gedde-Dahl, T., Jr. (1967). Genetic studies on the heavy chain subgroups of gamma-G globulin. In Gamma Globulins (Killander, J., ed.), pp. 313-328, Wiley & Sons, New York.

Natvig, J. B. and Kunkel, H. G. (1974). A hybrid IgG4-IgG2 immunoglobulin. J. Immunol. **112**, 1277-1284.

Nisonoff, A., Hopper, J. E., and Spring, S. B. (1975). Immunoglobulins of the rabbit, mouse, guinea pig, and horse. In The Antibody Molecule (Dixon, F. J., Jr., and Kunkel, H. G., eds.), pp. 314-318, Academic Press, New York.

Ollo, R., Auffray, C., Morchamps, C., and Rougeon, F. (1981). Comparison of more immunoglobulin γ2a and γ2b chain genes suggests that exons can be exchanged between genes in a multigenic family. Proc. Natl. Acad. Sci. USA **78**, 2442-2446.

Pink, J. R. L., Buttery, S. H., DeVries, G. M., and Milstein, C. (1970). Human immunoglobulin subclasses: Partial amino acid sequence of the constant region of a γ4 chain. Biochem. J. **117**, 33-47.

Romero-Harrera, A. E., Lehmann, H., Joysey, K. A., and Friday, A. F. (1973). Molecular evolution of myoglobin and the fossil record: A phylogenetic synthesis. Nature **246**, 389-395.

Shimizu, A., Takahashi, N., Yaoita, Y., and Honjo, T. (1982). Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell **28**, 499-506.

Spiegelberg, H. L. (1974). Biological activities of immunoglobulins of different classes and subclasses. Adv. Immunol. **19**, 259-294.

Tucker, P. W., Marcu, K. B., Newell, N., Richards, J., and Blattner, F. R. (1979). Sequence of the cloned gene for the constant region of murine γ2b immunoglobulin heavy chain. Science **206**, 1303-1306.

Unkeless, J. C., Fleit, H., and Mellman, I. S. (1981). Structural aspects and heterogeneity of immunoglobulin F_c receptors. Adv. Immunol. **31**, 247-270.

Wang, A.-C., Tung, E., and Fudenberg, H. H. (1980). The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and functional implications. J. Immunol. **125**, 1048-1054.

Winkelhake, J. L. (1978). Immunoglobulin structure and effector functions. Immunochemistry **15**, 695-714.

Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M., and Honjo, T. (1980). Complete nucleotide sequence of immunoglobulin γ2b chain gene cloned from newborn mouse DNA. Nature **283**, 786-789.

Yamawaki-Kataoka, Y., Miyata, T., and Honjo, T. (1981). The complete nucleotide sequence of mouse immunoglobulin γ2a gene and evolution of heavy chain genes: Further evidence for intervening sequence-mediated domain transfer. Nucleic Acids Res. **9**, 1365-1381.

Table 1. IgG Subclasses in Various Mammals

Species	IgG Subclasses	No. of Functional C_γ Genes
Human	IgG1, IgG2, IgG3, IgG4	4
Mouse	IgG1, IgG2a, IgG2b, IgG3	4
Rat	IgG1, IgG2a, IgG2b	3
Rabbit	IgG (only one type)	1
Guinea Pig	IgG1, IgG2	2
Dog	IgG1, IgG2a, IgG2b, IgG2c	4
Horse	IgGa, IgGb, IgGc, IgG(T), IgGB	5
Sheep	IgG1, IgG2	2

Table 2. Secondary Biological Activities of IgG Antibodies

Transfer across placenta
Complement (C1q) binding
Binding to cystic fibrosis factor
Binding to Staphylococcal A protein
Role as antigen to rheumatoid factors
Cytophilic properties; binding to receptors on: monocytes macrophages neutrophiles lymphocytes

Figure 1. Interchain disulfide bonds in human IgG molecules. The heavy and light polypeptide chains of the human IgG subclasses are depicted as long and short thick lines, respectively. Interchain disulfide bridges are shown as thin lines.

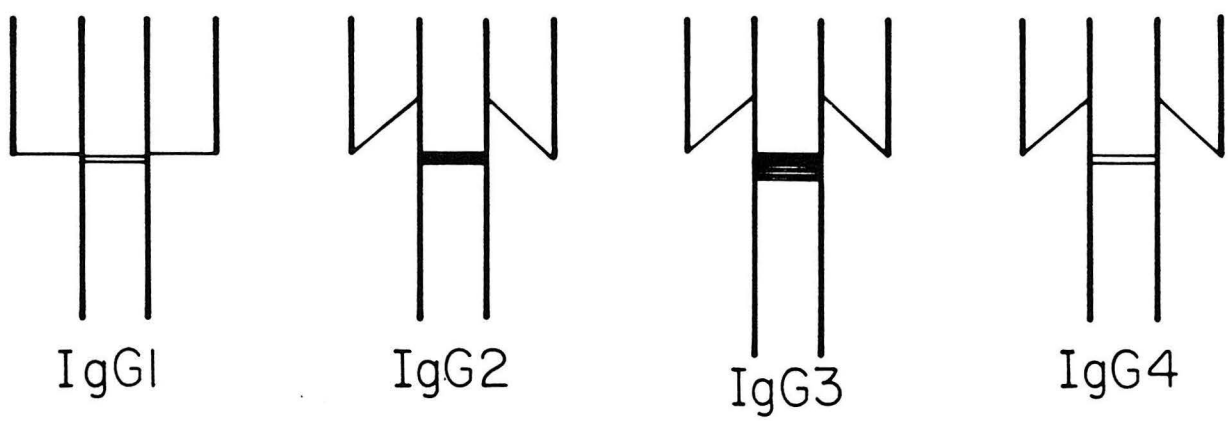

Figure 1

Figure 2. Evolutionary schemes for C_γ genes. See text for details.

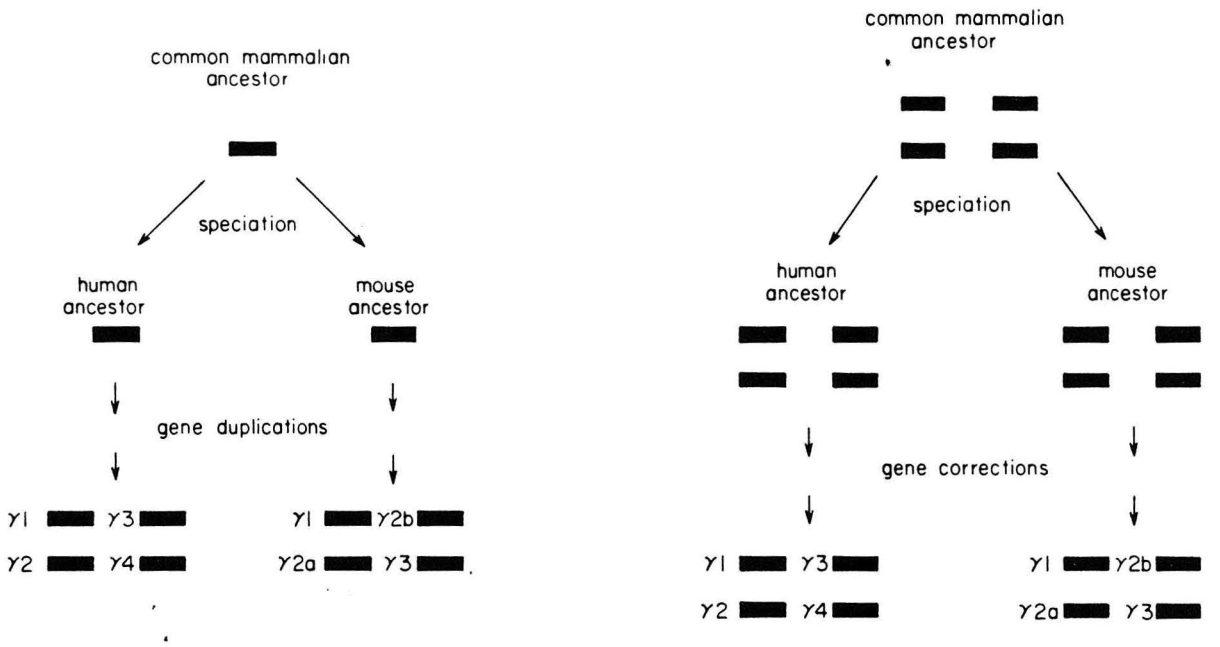
common mammalian ancestor
speciation
human ancestor
mouse ancestor
gene duplications
γ1 γ3
γ2 γ4
γ1 γ2b
γ2a γ3
common mammalian ancestor
speciation
human ancestor
mouse ancestor
gene corrections
γ1 γ3
γ2 γ4
γ1 γ2b
γ2a γ3

Figure 2

CHAPTER 2

Nucleotide Sequence of a Human Immunoglobulin $C_{\gamma 4}$ Gene

(Published in DNA)

DNA
Volume 1, Number 1, 1981
Mary Ann Liebert, Inc., Publishers

Nucleotide Sequence of a Human Immunoglobulin $C_{\gamma 4}$ Gene

J. ELLISON, J. BUXBAUM,* and L. HOOD†

ABSTRACT

We report the nucleotide sequence of a gene encoding the constant region of a human immunoglobulin γ4 heavy chain ($C_{\gamma 4}$). These data represent the first complete sequence determination of a human C_H gene. As expected from structural studies of mouse C_γ genes, the coding sequences for the C_H domains and hinge segment are separated from one another by intervening DNA sequences. Comparison with genomic sequences of the mouse $C_{\gamma 1}$, $C_{\gamma 2a}$, and $C_{\gamma 2b}$ genes shows conservation of the sequences in the constant region domains and the 3′ untranslated region surrounding the presumed site of poly(A) addition.

INTRODUCTION

IMMUNOGLOBULINS are composed of light and heavy polypeptides. These polypeptides consist of an N-terminal variable (V) portion involved in antigen binding, and a C-terminal constant (C) portion involved in effector functions such as antigen elimination. The immunoglobulin class is determined by the type of heavy-chain C region (C_H) in the molecule; e.g., C_μ in IgM, C_γ in IgG, C_α in IgA, C_δ in IgD, and C_ϵ in IgE. Most of our knowledge concerning the structure and expression of the genes coding for immunoglobulin heavy chains has resulted from studies using mice, in which immunoglobulin-producing tumors can be induced in the laboratory (Potter, 1972). These studies have shown that gene segments encoding the variable and constant regions are physically separate in germline DNA, and that through DNA rearrangements these genes are brought into closer proximity in the antibody-producing cells in which they are expressed (Early *et al.*, 1979). It is generally agreed that IgM is the first immunoglobulin class to be expressed in the ontogeny of a B lymphocyte (Raff, 1976). Several studies indicate that progeny of this B cell can undergo a class switch in which other C_H regions are associated with the same V_H domain initially expressed in the IgM-producing B cell (Fudenberg *et al.*, 1971; Levin *et al.*, 1971; Sledge *et al.*, 1976; Wang *et al.*, 1970, 1973). Hence molecules with the same antigen specificity and a broader range of effector functions are synthesized as the immune response matures. The C_H genes in the mouse are organized in a tightly linked cluster (Mage *et al.*, 1973), and this class switch results from a DNA rearrangement in which a second C_H gene displaces the C_μ gene from its position near the V_H gene (Davis *et al.*, 1980).

IgG is the quantitatively dominant immunoglobulin class in the blood of mammals. The population of IgG molecules in many species can be divided into distinct subclasses, each encoded by a distinct C_γ gene. These subclasses show differences in serum concentrations and effector functions. In evolutionary terms the C_γ genes are interesting for several reasons. First, the C_γ genes apparently are present in differing numbers in various mammals—e.g., human has four ($C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\gamma 3}$, $C_{\gamma 4}$), mouse four ($C_{\gamma 1}$, $C_{\gamma 2a}$, $C_{\gamma 2b}$, $C_{\gamma 3}$), guinea pig two, and rabbit one. Second, the γ chains of mouse and man exhibit species-associated residues that distinguish all mouse from all human C_γ regions. This species specificity suggests that the mouse (and human) C_γ genes all arose by gene duplication from a common ancestor after mammalian speciation. Moreover, the C_γ genes in humans diverged more recently than their mouse counterparts in that the human subclasses are >90% homologous to one another (Edelman *et al.*, 1969; Frangione *et al.*, 1980; Pink *et al.*, 1970; Ponstingl and Hilschmann, 1972; Wang *et al.*, 1980), whereas the mouse C_γ genes are much less homologous (Dunnick *et al.*, 1980; Honjo *et al.*, 1979; Obata *et al.*, 1980; Sikorav *et al.*, 1980; Tucker *et al.*, 1979a; Yamawaki-Kataoka *et al.*, 1980, 1981).

Buxbaum *et al.* have recently isolated cDNA clones containing sequences coding for a human γ3 heavy chain disease protein (Buxbaum *et al.*, unpublished). We have used one of these clones as a hybridization probe to isolate C_γ genes from a recombinant library of human DNA. Genomic clones for the mouse $C_{\gamma 1}$, $C_{\gamma 2a}$, and $C_{\gamma 2b}$ genes have been sequenced. Comparisons of the human and mouse C_γ genes should allow us to determine conserved coding and noncoding regions that may be of functional importance and, as well, to begin to delineate the nature of the genetic events that have led to the recent evolution of these small C_γ multigene families. In this paper we report the cloning and complete nucleotide sequence of the human

Division of Biology, California Institute of Technology, Pasadena, California 91125.
*Veterans Administration Medical Center, New York, New York 10010.
†To whom correspondence should be addressed.

$C_{\gamma 4}$ gene. This sequence agrees extremely well with the partial amino acid sequence for a γ4 protein reported by Pink *et al.* (1970).

MATERIALS AND METHODS

Materials

Restriction endonucleases, T4 DNA ligase, and *E. coli* DNA polymerase I large fragment were purchased from New England Biolabs. Bacterial alkaline phosphatase and *E. coli* DNA polymerase I were obtained from Bethesda Research Laboratories. T4 polynucleotide kinase was obtained from Boehringer-Mannheim. The γ-[^{32}P]ATP (~9000 Ci/mmole) was purchased from ICN, and α-[^{32}P]deoxynucleoside triphosphates were obtained either from New England Nuclear (835 Ci/mmole) or Amersham (410 and 2-3000 Ci/mmole). The X-ray film (XR5 or XAR5) was purchased from Kodak, and intensifying screens were obtained from DuPont (Cronex Lighting Plus). Human fetal liver DNA and the *Hae* III-*Alu* I human library were generously provided by T. Maniatis.

Preparation of cDNA hybridization probe

The pOMMC plasmid DNA, which contains sequences coding for the constant region of a human γ3 chain, was digested with *Hae* II and electrophoresed on a preparative horizontal 1.5% agarose gel. A 1.9 kb fragment containing 1050 nucleotides of cDNA insert and 890 nucleotides of flanking pBR322 DNA was eluted by electrophoresis into hydroxyapatite as described previously (Tabak and Flavell, 1978), and further purified by chromatography on DE52. The DNA was labeled by nick-translation (Maniatis *et al.*, 1975) using all four α-^{32}P-labeled deoxynucleoside triphosphates to a specific activity of approximately 7×10^8 cpm/μg.

Genome blot hybridization and library screening

Ten μg of human fetal liver DNA were digested for 3 hr at 37° with 100 units of *Hind* III, followed by another hour with an additional 50 units of enzyme. The DNA was electrophoresed on a horizontal 0.5% agarose gel and transferred to a nitrocellulose filter (Southern, 1975). The filter was preincubated for 2 hr at 68°C in "blot buffer." Blot buffer consists of 6X SET/0.1% ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin/0.1% SDS/0.1% Na pyrophosphate/0.1 *M* phosphate buffer/10 μg/ml each of poly(r)C and poly(r) A/50 μ/ml salmon sperm DNA; 1X SET = 0.15 *M* NaCl/30 m*M* Tris, pH 8.0/2.5 m*M* EDTA. Following preincubation, the heat-denatured nick-translated probe was reacted with the filter in blot buffer for 18 hr at 68°C. The reaction mix was then removed and the filter was washed at 68°C for 4 hr with two changes each of high salt (6X SET) and low salt (1X SET) wash buffer (these solutions also contained 0.1% SDS/0.1% Na pyrophosphate/0.025 *M* phosphate buffer). The filter was autoradiographed at −70°C using pre-flashed film and an intensifying screen.

The partial *Hae* III-*Alu* I human DNA library (Lawn *et al.*, 1978) cloned in the bacteriophage Charon 4A (Blattner *et al.*, 1977) was plated at a density of about 25,000 plaques per plate, using *E. coli* strain DP50supF as host. The library was screened essentially as described by Benton and Davis (1977), using the same solutions as used for the genome blot experiment. Filters were autoradiographed as described above. Hybridizing plaques were purified by rescreening (Maniatis *et al.*, 1978).

Restriction mapping and subcloning

Positions of cleavage of the restriction enzymes *Eco RI*, *Bam* HI, and *Hind* III in the clones of Fig. 2A were determined by analyses of single and double digests with these enzymes. Fragments hybridizing to the cDNA probe were identified by blot hybridizations performed as described above. Following appropriate digestion and agarose gel electrophoresis, the 2 kb *Eco* RI-*Hind* III fragment of clone 24B was isolated by the $NaClO_4$-GFC method of Chen and Thomas (1980). The fragment was directionally subcloned into pBR322 by insertion between the *Eco* RI and *Hind* III sites of the plasmid, followed by transformation of *E. coli* strain HB101 as described by Kushner (1978). Insert-containing transformants were identified by the size (as determined by gel electrophoresis) of supercoiled plasmid DNA present in lysates of single colonies. The subcloned human DNA insert was mapped by analyses of single and double restriction digests with the enzymes indicated in Fig. 2B. This was done using either the chimeric plasmid or the isolated 2 kb insert, which was purified as described above for the hybridization probe.

DNA sequence analysis

The purified 2 kb insert of p24BRH was sequenced essentially as described by Maxam and Gilbert (1977, 1980). Restriction fragments were labeled at their 5′ ends with polynucleotide kinase and γ-[^{32}P]ATP or at their 3′ ends with the *E. coli* DNA polymerase I large fragment and α-[^{32}P]deoxynucleoside triphosphates. In every case but one, the fragments used for the chemical degradation reactions were double-stranded fragments labeled at only one end, generated by cleavage with a second enzyme following labeling. The single exception was a fragment which was strand-separated on an 8% polyacrylamide gel. Details of the procedures are given in Maxam and Gilbert (1980). Degradation products were electrophoresed on 40 cm and 80 cm thin (0.4 mm) sequencing gels. The acrylamide concentrations ranged from 5% to 25%.

RESULTS AND DISCUSSION

Isolation of recombinant clones complementary to a human γ cDNA

The cloning of human C_γ genes was facilitated by the availability of a human cell line (OMM) which synthesizes a γ3 heavy chain disease protein. RNA prepared from these cells was used to construct a set of cDNAs cloned in the plasmid pBR322. The hybridization probe used in these experiments is derived from one of these clones, and contains

sequences coding for a human $C_{\gamma 3}$ region. Details of the construction and characterization of this clone, designated pOMMC, will be presented elsewhere (Buxbaum *et al.*, unpublished). A 1.9 kb *Hae* II restriction fragment containing roughly equal lengths of insert and pBR322 flanking sequences was used in these experiments. Figure 1 shows the results of an experiment in which genomic DNA from human fetal liver was digested with the restriction endonuclease *Hind* III, electrophoresed on a 0.5% agarose gel, transferred to a nitrocellulose filter, and hybridized with the ^{32}P-labeled pOMMC probe. Several bands in the genomic DNA hybridize to the probe, although because of the close band spacing it is difficult to obtain an exact number. Since the four human C_γ regions share over 90% amino acid sequence homology, we believe that the $\gamma 3$ probe should hybridize to all four C_γ genes.

A library of human genomic DNA fragments cloned in the bacteriophage Charon 4A was obtained from T. Maniatis. The source of the cloned DNA is the same as that used in the genomic blot hybridization experiment of Fig. 1. The library was constructed from fragments resulting from the ligation of synthetic *Eco* RI linkers onto the ends of partial cleavage products generated by the restriction enzymes *Hae* III and *Alu* I (Lawn *et al.*, 1978; Maniatis *et al.*, 1978). Approximately 10^6 plaques were screened with the ^{32}P-labeled $\gamma 3$ cDNA probe, and over 40 strongly hybridizing clones were isolated following the plaque-purifying rescreens.

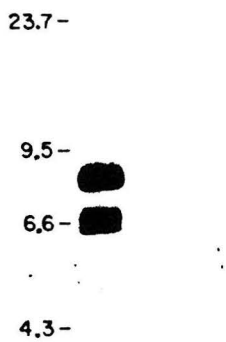

FIG. 1. Hybridization of a human $\gamma 3$ cDNA to genomic DNA. Ten μg of human fetal liver DNA were digested with *Hind* III, electrophoresed on a 0.5% agarose gel, and transferred to a nitrocellulose filter, which was reacted with the nick-translated $\gamma 3$ cDNA probe. Details are given in Materials and Methods. Numbers refer to the sizes in kilobases of *Hind* III fragments of phage λ_{cI857} DNA run in a parallel lane.

The primary structure of a human $C_{\gamma 4}$ gene

Figure 2A shows the DNA inserts of three clones isolated from the human genomic library with the cleavage sites for the restriction enzymes *Eco* RI, *Bam* HI, and *Hind* III indicated. The positions of these sites indicate that the inserts overlap as shown. The restriction fragments hybridizing to the $\gamma 3$ cDNA probe in clone blots are cross hatched. The 2 kb *Eco* RI-*Hind* III fragment of clone 24B was subcloned by insertion into pBR322. This subclone, designated p24BRH, was mapped by analyses of single and double digests with several restriction enzymes. The resulting restriction map is shown in Fig. 2B.

The entire DNA sequence of the 2 kb insert of p24BRH was determined by the methods of Maxam and Gilbert (1977, 1980) and is presented in Fig. 3. This sequence begins 214 nucleotides 5′ to the C_H1 coding region and continues approximately 100 nucleotides past the putative poly(A) addition site. The human C_γ gene is seen to be comprised of separate exons coding for each of the constant region domains and the hinge segment. Translation of the coding regions indicates that this gene encodes the $C_{\gamma 4}$ region. We draw this conclusion because the predicted protein sequence is 90–92% homologous to sequenced C_H domains of human $\gamma 1$, $\gamma 2$, and $\gamma 3$ heavy chains, whereas it is virtually identical to the partially determined sequence of a human $\gamma 4$ protein (see below). Furthermore, the subclasses have distinct hinge segments that are quite different from one another, and the predicted hinge sequence for this gene clearly corresponds to that of the $\gamma 4$ heavy chain. The translated amino acids appear above the corresponding codons in Fig. 3. A total of 327 amino acids are encoded by this gene. An analysis of codon usage reveals that the great majority (83.5%) of amino acids are encoded in triplets ending in C or G. This phenomenon is characteristic of many eucaryotic genomes (Grantham *et al.*, 1980). Another feature of this gene shared with mosaic eucaryotic genes is the presence of sequences at the intron–exon junctions that obey the "GT-AG" rule (Breathnach *et al.*, 1978; Lewin, 1980).

The complete sequence of a human $\gamma 4$ heavy chain protein has not been determined, but Pink *et al.* (1970) reported the sequence of 229 amino acids in the human $C_{\gamma 4}$ region of the myeloma protein Vin. The protein encoded by the gene analyzed in this study is identical to the Vin partial sequence at all but one residue. The $C_{\gamma 4}$ gene encodes a serine at the tenth residue of the hinge region, whereas the Vin protein sequence has a proline at that position. This difference could be due to either a protein sequencing error or to a polymorphism in the human population. (It is perhaps not surprising that a polymorphism be found in the hinge gene segment, since this is the region showing the greatest variation among C_γ subclasses.) In addition, a lysine residue is encoded at the C-terminus which is apparently not present in the mature $\gamma 4$ chain. A C-terminal lysine also is encoded in the mouse $C_{\gamma 1}$, $C_{\gamma 2a}$, and $C_{\gamma 2b}$ genes

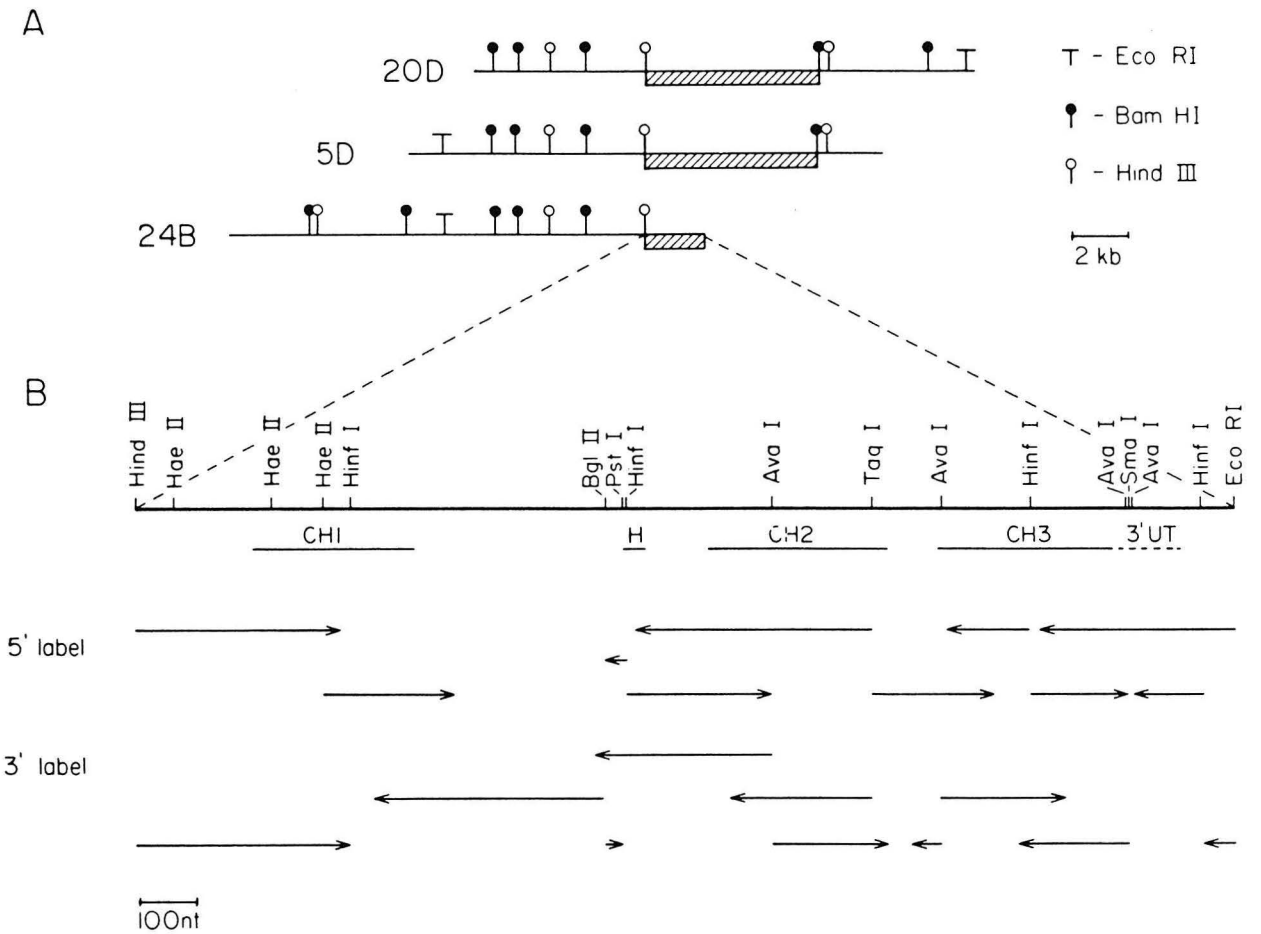

FIG. 2. A. Restriction maps of cloned chromosomal fragments complementary to a human γ3 cDNA. The indicated symbols are placed at the cleavage sites of the respective enzymes. The ends of the fragments contain *Eco* RI recognition-cleavage sequences not present in human chromosomal DNA. *Hind* III-*Bam* HI or *Hind* III-*Eco* RI fragments containing sequences complementary to the human γ3 cDNA are cross hatched. B. Restriction map and sequencing strategy of the subcloned DNA insert of p24BRH. The DNA segment shown is an expanded view of the 2 kb *Eco* RI-*Hind* III fragment of clone 24B of Fig. 2A. The fragment was subcloned as described in Materials and Methods. The cleavage sites of several restriction enzymes are indicated. The arrows below the restriction map indicate independent sequence determinations. The "blunt" end of the arrowed lines represent cleavage and labeling at either the 5′ or 3′ end, and the lengths of the lines represent the number of determined nucleotides from the labeled end.

and also fails to appear in the mature polypeptide. This lysine residue is apparently removed post-translationally.

The overall structural features of the human $C_{\gamma 4}$ gene can be compared with those of the mouse $C_{\gamma 1}$, $C_{\gamma 2a}$, and $C_{\gamma 2b}$ genes, whose complete sequences have been determined (Honjo *et al.*, 1979; Tucker *et al.*, 1979a; Yamawaki-Kataoka *et al.*, 1980, 1981). Figure 4 schematically shows the general similarity that exists between human and mouse genes with regard to the lengths and positions of introns and coding regions. We expect that human C_γ genes have an additional exon(s) 3′ to the C_H3 exon coding for a protein segment which anchors the membrane-bound form of the IgG molecule on the lymphocyte cell surface. This membrane exon sequence would presumably be spliced at the RNA level to the 3′ end of the C_H3 coding sequence in lymphocytes producing membrane-bound IgG molecules. In fact, a potential splice donor dinucleotide (GT) is seen in the penultimate C_H3 codon of the human C_γ gene reported here. The length of the 3′ untranslated region of the human γ4 mRNA has not been directly determined, but we have estimated the length of this region indicated in Fig. 4 on the basis of homologies to corresponding regions in mouse genes. We note the presence of the sequence AATAAA beginning 101 nucleotides 3′ to the stop codon (marked by an asterisk in Fig. 3). This sequence precedes the poly(A) addition site of most eukaryotic mRNAs (Proudfoot and Brownlee, 1976), including those of mouse C_γ (Dunnick *et al.*, 1980; Silorav *et al.*, 1980; Tucker *et al.*, 1979a; Yamawaki-Kataoka *et al.*, 1981), C_μ (Rogers *et al.*, 1980), and C_κ genes (Hamlyn *et al.*, 1978). Figure 5 shows homologies of human and mouse C_γ genes in the region 3′ to this sequence. We can assign possible sites of polyadenylation of the human γ4 mRNA in this region, as detailed in the legend to Fig. 5.

Sequence homologies among the C_H domains of immunoglobulins have been recognized for some time,

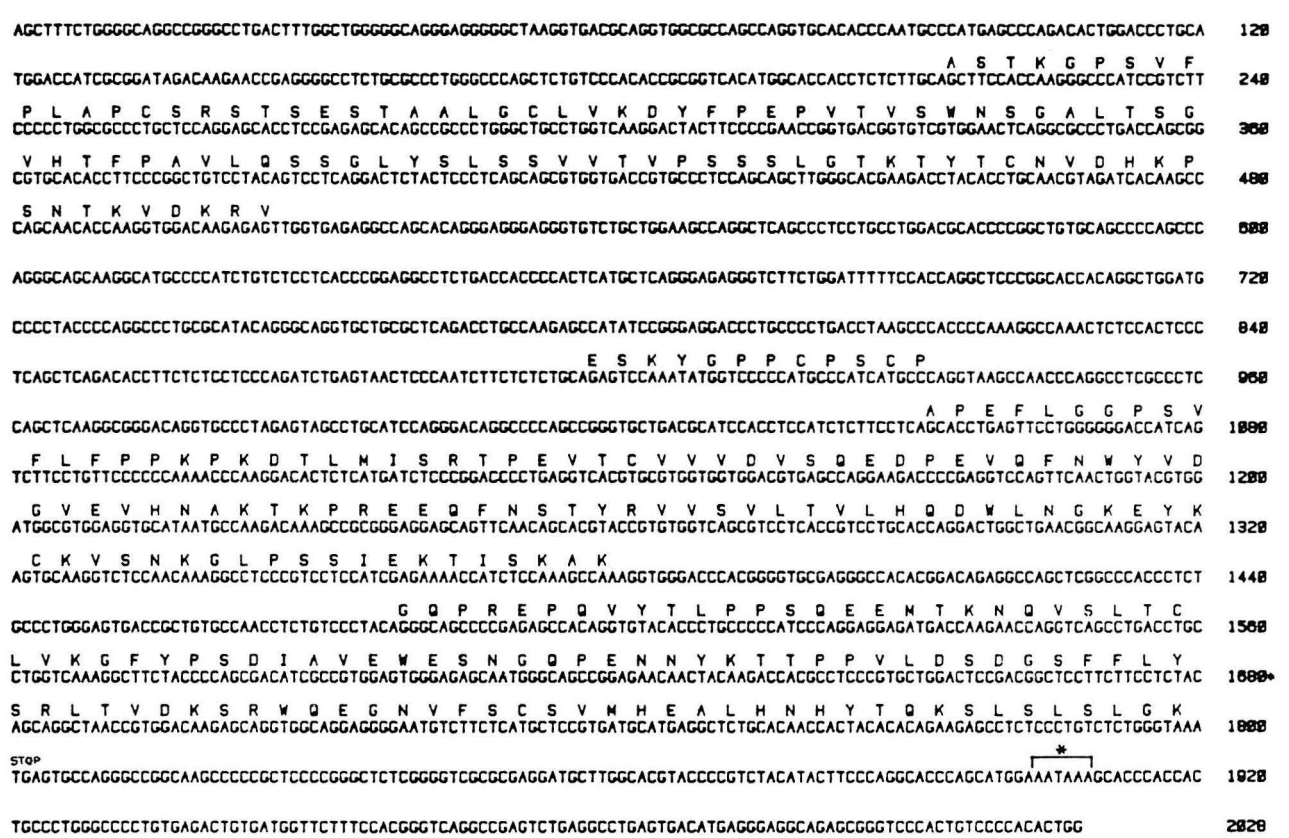

FIG. 3. The nucleotide sequence of a human $C_{\gamma 4}$ gene and its corresponding protein sequence. The complete sequence of the mRNA synonymous strand of the subcloned 2 kb insert of p24BRH is shown 5′ to 3′. Amino acids predicted by the nucleotide sequence are shown above the first base of the respective codons. "Stop" indicates the termination codon UGA. The presumptive poly(A) addition signal sequence is marked by an asterisk. A potential ambiguity in our exon assignments is our placement of glycine at the N-terminal end of the C_H3 domain. This residue also is encoded by the triplet immediately following the C-terminus of the C_H2 domain (GGT). We believe our assignment to be correct, however, since placement of glycine in the C_H2 domain would lead to a violation of the GT-AG rule for splice junction sequences. The one-letter code for amino acids is as follows: (A) alanine; (C) cysteine; (D) aspartic acid; (E) glutamic acid; (F) phenylalanine; (G) glycine; (H) histidine; (I) isoleucine; (K) lysine; (L) leucine; (M) methionine; (N) asparagine; (P) proline; (Q) glutamine; (R) arginine; (S) serine; (T) threonine; (V) valine; (W) tryptophan; (Y) tyrosine.

leading to the suggestion that they arose from tandem duplication of a primordial gene unit (Edelman *et al.*, 1969; Hill *et al.*, 1966). The sequence of the human $C_{\gamma 4}$ gene reported here allows precise domain boundaries to be drawn for human γ heavy chains, on the basis of the boundaries of the exons (see Fig. 3). When the three C_H domains of the $C_{\gamma 4}$ gene are aligned with the introduction of gaps as noted by Edelman *et al.* (1969), the interdomain protein homologies are seen to range from 26% to 30% when gaps are excluded from the comparison. At the nucleotide level, these homologies are 43–50%. These homology relationships, as well as the presence of a centrally placed disulfide bridge within each domain, are consistent with the existence of an ancient common ancestor for the C_H exons.

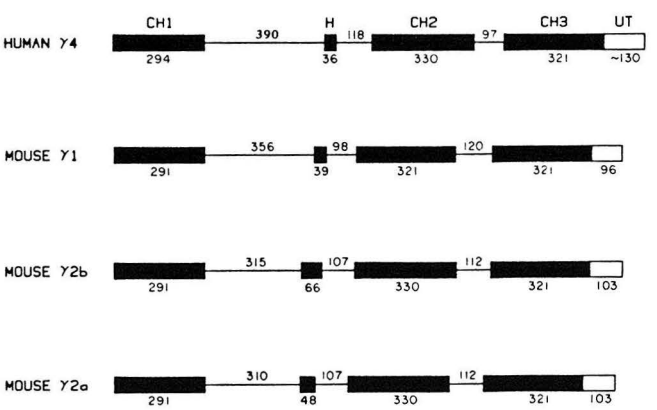

FIG. 4. Domain organization of human and mouse C_γ genes. The lengths of and distances between C_H domains and hinge segments are schematically compared. Coding regions are denoted by solid blocks and intervening sequences by lines connecting the blocks. Open blocks depict 3′ untranslated regions in the corresponding mRNAs (see legend to Fig. 5).

Homologies of the human and mouse Cγ regions

Complete amino acid sequences have been determined for proteins of two human C_γ genes: the $\gamma 1$ proteins Eu (Edelman *et al.*, 1969) and Nie (Postingl and Hilschmann, 1972), and the $\gamma 2$ protein Til (Wang *et al.*, 1980). The se-

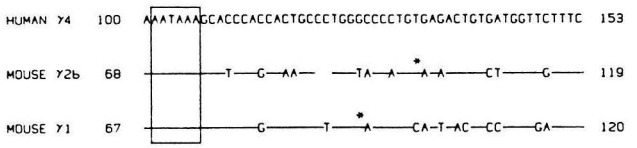

FIG. 5. Homologies near the presumed site of poly(A) addition. Sequences occurring 3′ to the termination codons of the indicated genes are compared. Numbers at the ends of the sequences refer to the number of nucleotides from the codon to the first and last residues listed. The solid lines in the mouse sequences represent identity to the human sequence. Where differences occur the nucleotides in the mouse sequences are indicated. The poly(A) addition signal is boxed. A two nucleotide gap is introduced into the mouse $C_{\gamma 1}$ sequence to maintain homology to the human sequence. Additional homologies can be observed on both sides of the stretches shown above by introduction of gaps into the mouse sequences. Asterisks mark the presumed sites of addition of poly(A) to the corresponding mouse mRNAs. cDNA sequence data leading to these conclusions appear in Tucker *et al.* (1979b) (γ2b) and Obata *et al.* (1980) (γ1). Sikorav *et al.* (1980) determined the site of poly(A) addition in the mouse γ2a gene to be at the position corresponding to that of the γ2b gene above, and interestingly, Dunnick *et al.* (1980) assigned the site to this (γ2b) position in a mouse γ1 mRNA which carried a deletion of the C_H1 coding region. The discrepancies in the γ1 data may very well be real, since preliminary sequence data for the human γ3 cDNA indicate that poly(A) is added at the leftward (γ1) site (Buxbaum *et al.*, unpublished). Thus alternative sites for addition of poly(A) to human and mouse gamma chain mRNAs may exist. The scale of Fig. 4 accommodates either of these sites for the human γ4 gene.

quence of the hinge, C_H2, and C_H3 domains have been determined for the γ3 heavy chain disease protein Wis (Frangione *et al.*, 1980). When the $C_{\gamma 4}$ gene reported here is included in a comparison of these sequences, it is seen that the human γ subclasses share 90–95% amino acid sequence homology in the C_H region if the hinge segment is excluded from comparison. This homology suggests that a recent gene duplication event(s) gave rise to the four human C_γ genes. On the other hand, the hinge segments of these heavy chains are quite dissimilar in both length and sequence. The hinge exon apparently has evolved very rapidly since divergence of the C_γ genes. It has been postulated that the hinge exon may have evolved from a degenerate domain (Tucker *et al.*, 1979a). However, its short length makes it difficult to detect any significant homology between the $C_{\gamma 4}$ hinge exon and the C_H domain exons. Sequence analysis of other human C_γ genes may reveal common structural features which will provide unambiguous clues to the evolutionary origin of this gene segment.

Mice, like humans, have four subclasses of C_γ regions. In Fig. 6 we compare the nucleotide sequence of the mouse $C_{\gamma 1}$ gene with that of the human $C_{\gamma 4}$ gene reported here, using a dot matrix homology program (for discussion of this program see Hunkapiller and Hood, 1981). This method of analysis allows one to easily see homology relationships between two sequences which may be difficult to determine by manual sequence alignment. Homology relationships are indicated by diagonal lines, nucleotide differences by gaps in these lines, and sequence gaps (insertions or deletions) by offset diagonals. Figure 6 demonstrates that the human and mouse genes share significant homology in the C_H domains and 3′ untranslated region, but have diverged extensively in the introns and hinge region. A similar result is seen when the mouse $C_{\gamma 2a}$ and $C_{\gamma 2b}$ genes are compared to the human $C_{\gamma 4}$ gene (data not shown). A close inspection of this figure shows that the sequence divergence in the C_H domains is nonuniform; that is, divergent nucleotides are nonrandomly clustered along the lengths of the exons. This pattern also is seen in the amino acid sequences. This observation implies that certain sequences are more highly conserved than others, presumably because of functional constraints.

A comparison of the amino acid sequences encoded by the mouse $C_{\gamma 1}$, $C_{\gamma 2a}$, and $C_{\gamma 2b}$genes has shown that the proteins are 65–80% homologous to one another in the C_H regions when hinge segments are not included in the comparison (Sikorav *et al.*, 1980). This is significantly less than the homology shared by human γ chains, and suggests that the mouse genes arose from a separate and much earlier duplication event than that which gave rise to the human C_γ genes.

The idea of C_γ gene duplications occurring independently in mammalian species was noted by Pink *et al.* (1970). If

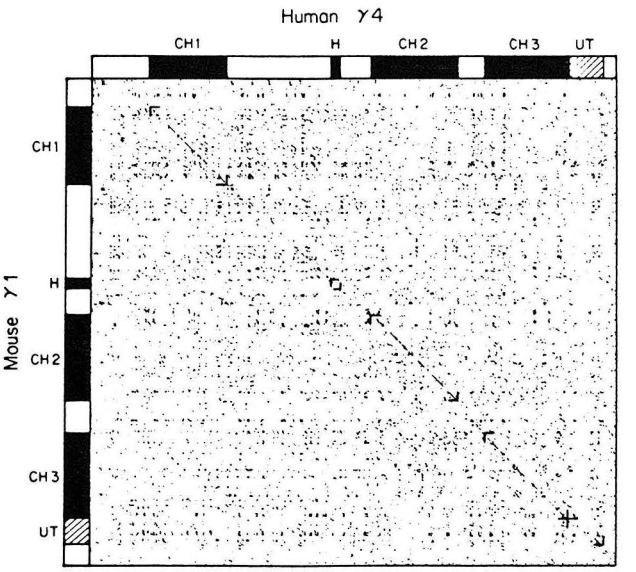

FIG. 6. Dot matrix homology analysis of human $C_{\gamma 4}$ and mouse $C_{\gamma 1}$ genes. A computer program (Hunkapiller and Hood, 1981) was used to generate a matrix of dots indicating segments of homology between the two sequences being compared. Different matrix patterns can be generated from such an analysis, depending on the desired homology criterion. In this figure a dot was scored when a four out of five match was seen in the regions of comparison of the two sequences. The human $C_{\gamma 4}$ and mouse $C_{\gamma 1}$ sequences are represented along the horizontal and vertical axes, respectively; coding and 3′ untranslated regions are indicated. Areas on the matrix corresponding to the ends of these regions are enclosed within pairs of short perpendicular lines which, if extended, lead to the borders of the exons and untranslated regions of the respective genes.

this is correct, a particular C_γ gene of one species would be evolutionarily equidistant from the C_γ genes of other species. In Table 1 we compare the protein and DNA sequences of the human $C_{\gamma 4}$ domains with the corresponding domains of the mouse C_γ genes. Also included are the 3′ untranslated sequences near the poly(A) addition site. Since no significant homology was seen between the introns by dot matrix analysis, these were not compared. Gaps were introduced in the mouse genes where necessary to maintain homology with the human sequence (see table legend). The mouse $C_{\gamma 3}$ sequence has not been published and so was not included in the comparisons. The results clearly show that the human $C_{\gamma 4}$ gene is equally related to each of the mouse C_γ genes. (Note that a 3-nucleotide gap is needed in an identical position in all the mouse genes in order to maintain homology to the human sequence in the C_H1 domain.) The failure to see a mouse C_γ counterpart to the human $C_{\gamma 4}$ gene thus supports the notion of separate C_γ gene duplications in the two species.

It is interesting to compare the relatedness of human and mouse C_γ genes with that shown by C_κ genes. Hieter *et al.* (1980) recently compared the sequences of human and mouse C_κ genes. These sequences are 81% homologous, which is higher than the levels indicated in Table 1 for C_γ genes of the two species. This difference may be related to the evolution of specialized functions shown by the proteins encoded by C_H genes.

TABLE 1. HOMOLOGY COMPARISONS OF HUMAN AND MOUSE γ GENES.

	% Homology to Human γ4			
	C_H1	C_H2	C_H3	3′
Mouse γ1	71	76	73	75 [57]
	(67)	(68)	(59)	
Mouse γ2b	69	75	70	77 [57]
	(62)	(65)	(58)	
Mouse γ2a	69	75	70	73 [59]
	(64)	(65)	(63)	

The human $C_{\gamma 4}$ domains and the region surrounding the presumed site of poly(A) addition are compared with the corresponding sequences in the indicated mouse genes. Numbers on the top lines of the coding-region comparisons refer to the nucleotide homologies, and numbers below those in parentheses refer to amino acid homologies. Sequences were aligned with the following gaps introduced to maintain homologies: (1) A 3 nucleotide gap was inserted between amino acid position 59 and 60 of all the mouse C_H1 domains. (2) A 9 nucleotide gap was inserted between amino acids 3 and 4 of the mouse γ1 C_H2 domain. (3) A 2 nucleotide gap was inserted in the 3′ untranslated region of the mouse γ1 gene, as indicated in Fig. 5. Sequences compared in the 3′ column correspond generally to the regions shown in Fig. 5; the boundaries of the compared sequences were chosen where the introduction of additional gaps would have been necessary to maintain homology to the human gene. Numbers in brackets in this column refer to the number of residues compared. In all the comparisons the percent homology to the human sequence was calculated with gaps excluded from the comparison. Mouse sequences used in these calculations were taken from Tucker *et al.* (1979a) (γ2b), Honjo *et al.* (1979) (γ1), and Yamawaki-Kataoka *et al.* (1981) (γ2a).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health. J.B. was supported by funds from the Veterans Administration and a grant from the American Cancer Society. We thank T. Hunkapiller and M. Douglas for help with the data analysis and figure construction and T. Maniatis for providing us with the human fetal liver DNA and genomic library.

REFERENCES

BENTON, W.D., and DAVIS, R.W. (1977). Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**, 180-182.

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. (1977). Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **196**, 161-169.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F., and CHAMBON, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. USA **75**, 4853-4857.

BUXBAUM, J., ALEXANDER, A., STEINMETZ, M., and HOOD, L. Unpublished.

CHEN, C.W., and THOMAS, C.A. JR. (1980). Recovery of DNA segments from agarose gels. Anal. Biochem. **101**, 339-341.

DAVIS, M.M., CALAME, K., EARLY, P.W., LIVANT, D.L., JOHO, R., WEISSMAN, I.L., and HOOD, L. (1980). An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature **283**, 733-739.

DUNNICK, W., RABBITTS, T.H., and MILSTEIN, C. (1980). A mouse immunoglobulin heavy chain deletion mutant: isolation of a cDNA clone and sequence analysis of the mRNA. Nucleic Acids Res. **8**, 1475-1485.

EARLY, P.W., DAVIS, M.M., KABACK, D.B., DAVIDSON, N., and HOOD, L. (1979). Immunoglobulin heavy chain gene organization in mice: analysis of a myeloma genomic clone containing variable and alpha constant regions. Proc. Natl. Acad. Sci. USA **76**, 857-861.

EDELMAN, G.M., CUNNINGHAM, B.A., GALL, W.E., GOTTLIEB, P.D., RUTISHAUSER, U., and WAXDAL, M.J. (1969). The covalent structure of an entire γG1 immunoglobulin molecule. Proc. Natl. Acad. Sci. USA **63**, 78-85.

FRANGIONE, B., ROSENWASSER, E., PRELLI, F., and FRANKLIN, E.C. (1980). Primary structure of human γ3 immunoglobulin deletion mutant: γ3 heavy-chain disease protein Wis. Biochemistry **19**, 4304-4308.

FUDENBERG, H.H., WANG, A.C., PINK, J.R.L., and LEVIN, A.S. (1971). Studies of an unusual biclonal gammopathy: implications with regard to genetic control of normal immunoglobulin synthesis. Ann. N.Y. Acad. Sci. **190**, 501-507.

GRANTHAM, R., GAUTIER, C., and GHOY, M. (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res. **8**, 1893-1913.

HAMLYN, P.H., BROWNLEE, G.G., CHENG, C.-C., GAIT, M.J., and MILSTEIN, C. (1978). Complete sequence of constant and 3′ noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. Cell **15**. 1067-1075.

HIETER, P.A., MAX, E.E., SEIDMAN, J.G., MAIZEL, J.V. JR., and LEDER, P. (1980). Cloned human and mouse

kappa immunoglobulin constant and J region genes conserve homology in functional segments. Cell **22**, 197-207.

HILL, R.L., DELANEY, R., FOLLOW, R.E., and LEBOWITZ, H.E. (1966). The evolutionary origins of the immunoglobulin. Proc. Natl. Acad. Sci. USA **56**, 1762-1769.

HONJO, T., OBATA, M. YAMAWAKI-KATAOKA, Y., KATAOKA, T., KAWAKAMI, T., TAKAHASHI, N., and MANO, Y. (1979). Cloning and complete nucleotide sequence of mouse immunoglobulin γ1 chain gene. Cell **18**, 559-568.

HUNKAPILLER, T., and HOOD, L. (1981). Manuscript in preparation.

KUSHNER, S.R. (1978). In *Genetic Engineering*, H.W. Boyer and S. Nicosia, eds. (Elsevier/North-Holland Biomedical Press, New York).

LAWN, R.M., FRITSCH, E.F., PARKER, R.C., BLAKE, G., and MANIATIS, T. (1978). The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell **15**, 1157-1174.

LEVIN, A.S., FUDENBERG, H.H., HOPPER, J.E., WILSON, S., and NISONOFF, A. (1971). Immunofluorescent evidence for cellular control of synthesis of variable regions of light and heavy chains of immunoglobulins G and M by the same gene. Proc. Natl. Acad. Sci. USA **68**, 169-171.

LEWIN, B. (1980). Alternatives for splicing: recognizing the ends of introns. Cell **22**, 324-326.

MAGE, R., LIEBERMAN, R., POTTER, M., and TERRY, W.D. (1973). In *The Antigens*, M. Sela, ed. (Academic Press, New York). pp. 299-376.

MANIATIS, T., JEFFREY, A., and KLEID, A.G. (1975). Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. USA **72**, 1184-1188.

MANIATIS, T., HARDISON, R.C., LACY, E., LAUER, J., O'CONNELL, C., QUAN, D. SIM, G.K., and EFSTRATIADIS, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**, 687-701.

MAXAM, A.M., and GILBERT, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**, 560-564.

MAXAM, A.M., and GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. **65**, 499-560.

OBATA, M., YAMAWAKI-KATOAKA, Y., TAKAHASHI, N., KATAOKA, T., SHIMIZU, A., MANO, Y., SEIDMAN, J.C., PETERLIN, B.M., LEDER, P., and HONJO, T. (1980). Immunoglobulin γ1 heavy chain gene: structural gene sequences cloned in a bacterial plasmid. Gene **9**, 87-97.

PINK, J.R.L., BUTTERY, S.H., DeVRIES, G.M., and MILSTEIN, C. (1970). Human immunoglobulin subclasses: partial amino acid sequence of the constant region of a γ4 chain. Biochem. J. **117**, 33-47.

PONSTINGL, H., and HILSCHMANN, N. (1972). Die Primarstruktur eines monoklonalen IgG1-Immunoglobulin (Myeloprotein Nie). II. Aminosauresequenz des konstanten Teils der H-Kette, zuordnung genetischer Faktoren. Hoppe-Seyler's Z. Physiol. Chem. **353**, 1369-1372.

POTTER, M. (1972). Immunoglobulin-producing tumors and myeloma proteins of mice. Physiol. Rev. **52**, 631-719.

PROUDFOOT, N., and BROWNLEE, G.G. (1976). 3' Non-coding region sequences in eucaryotic messenger RNA. Nature **263**, 211-214.

RAFF, M.C. (1976). Development and modulation of B lymphocytes: studies on newly formed B cells and their putative precursors in the hemopoietic tissues of mice. Cold Spring Harbor Symp. Quant. Biol. **41**, 159-162.

ROGERS, J., EARLY, P., CARTER, C., CALAME, K., BOND, M., HOOD, L., and WALL, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell **20**, 303-312.

SIKORAV, J.-L., AUFFRAY, C., and ROUGEON, F. (1980). Structure of the constant and 3' untranslated regions of the murine BALB/c γ2a heavy-chain messenger RNA. Nucleic Acids Res. **8**, 5983-5991.

SLEDGE, C., FAIR, D.S., BLACK, B., KRIEGER, R.G., and HOOD, L. (1976). Antibody differentiation: apparent sequence identity between variable regions shared by IgA and IgG immunoglobulins. Proc. Natl. Acad. Sci. USA **73**, 923-927.

SOUTHERN, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-517.

TABAK, H.F., and FLAVELL, R.A. (1978). A method for the recovery of DNA from agarose gels. Nucleic Acids Res. **5**, 2321-2332.

TUCKER, P.W., MARCU, K.B., NEWELL, N., RICHARDS, J., and BLATTNER, F.R. (1979a). Sequence of the cloned gene for the constant region of murine γ2b immunoglobulin heavy chain. Science **206**, 1303-1306.

TUCKER, D.W., MARCU, K.B., SLIGHTON, J.L., and BLATTNER, F.R. (1979b). Structure of the constant and 3' untranslated regions of the murine γ2b heavy chain messenger RNA. Science **206**, 1299-1303.

WANG, A.C., WILSON, S.K., HOPPER, J.E., FUDENBERG, H.H., and NISONOFF, A. (1970). Evidence for control of synthesis of the variable regions of the heavy chains of immunoglobulins G and M by the same gene. Proc. Natl. Acad. Sci. USA **66**, 337-343.

WANG, A.C., GRISELY, J., and FUDENBERG, H.H. (1973). Amino acid sequences at constant and variable regions of heavy chains of monotypic immunoglobulins G and M of a single patient. Biochemistry **12**, 528-534.

WANG, A.-C., TUNG, E., and FUDENBERG, H.H. (1980). The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and functional implications. J. Immunol. **125**, 1048-1054.

YAMAWAKI-KATAOKA, Y., KATAOKA, T., TAKAHASHI, N., OBATA, M., and HONJO, T. (1980). Complete nucleotide sequence of immunoglobulin γ2b chain gene cloned from newborn mouse DNA. Nature **283**, 786-789.

YAMAWAKI-KATAOKA, Y., MIYATA, T., and HONJO, T. (1981). The complete nucleotide sequence of mouse immunoglobulin γ2a gene and evolution of heavy chain genes: further evidence for intervening sequence-mediated domain transfer. Nucleic Acids Res. **9**, 1365-1381.

Address Reprint Requests to:
Jay Ellison
Biology Division
California Institute of Technology
Mail Code 156-29
Pasadena, CA 91125

Received for publication, May 8, 1981, and in revised form, July 1, 1981.

CHAPTER 3

Linkage and Sequence Homology of Two Human Immunoglobulin γ Heavy Chain Constant Region Genes

(Published in Proceedings of the National Academy of Sciences, U.S.A.)

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1984–1988, March 1982
Immunology

Linkage and sequence homology of two human immunoglobulin γ heavy chain constant region genes

(evolution/gene duplication/unequal crossover/gene conversion)

JAY ELLISON AND LEROY HOOD

Division of Biology, California Institute of Technology, Pasadena, California 91125

Communicated by Norman Horowitz, November 23, 1981

ABSTRACT **We report the nucleotide sequence of a gene encoding a human immunoglobulin $C_{\gamma 2}$ region. Comparison with the previously determined $C_{\gamma 4}$ sequence reveals that these two genes share extensive (≈95%) homology in the three C_H domain exons and adjacent noncoding regions. In contrast, hinge exons have diverged to a much greater degree, implying that natural selection has favored the generation of diversity in these coding regions. We have used the noncoding nucleotide differences to estimate that approximately 6–7 million years have elapsed since the occurrence of the gene duplication or correction event which generated the two identical ancestral genes. In addition we show that the two C_γ genes are arranged in human chromosomal DNA in the configuration 5′-$C_{\gamma 2}$-17 kilobase pairs -$C_{\gamma 4}$-3′.**

IgG is the major class of antibody molecule in the serum of mammals, representing 70–80% of the total serum immunoglobulin in humans. It is distinguished from other classes of antibodies by a γ heavy chain constant region (C_γ). The C_γ regions of human IgG molecules are divided into four subclasses ($C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\gamma 3}$, and $C_{\gamma 4}$) encoded by distinct germ-line genes (1). Protein sequence studies (2–5) have shown that the subclasses are highly homologous, indicating that the corresponding genes derive from a common ancestral C_γ gene.

Subclasses of IgG have been observed in several other mammals, although the number varies for different species (6). This observation suggested that C_γ gene duplications occurred independently in various mammalian evolutionary lines after their divergence from a common ancestor. The model assumes that members of a C_γ gene family evolve independently and that the accumulated differences in the C_γ genes reflect the time elapsed since the duplication event. Molecular analyses of mouse C_γ genes suggest that this simple model may not be correct (7, 8). These studies indicate that, during evolution, genetic information has been exchanged between nonallelic mouse C_γ genes. This implies the existence of mechanisms that prevent the C_γ genes from freely diverging from one another, so that sequence homology is continually renewed within the gene family. This postulated type of mechanism has been termed "gene correction" (9).

We are interested in determining the structural characteristics of human C_γ genes to provide insights into the evolution of the C_γ gene family. We previously determined the complete nucleotide sequence of one human C_γ gene, that encoding the $C_{\gamma 4}$ region (10). In this paper we report the sequence of a human $C_{\gamma 2}$ gene and compare it to the $C_{\gamma 4}$ sequence. In addition we provide molecular evidence that these genes lie adjacent to one another in human chromosomal DNA.

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

Materials. The human genomic DNA library was obtained from T. Maniatis. Sources of nucleic acid enzymes, reagents for DNA sequence analysis, bacteria, and the phage M13mp2 were those described by Steinmetz *et al.* (11).

Isolation and Subcloning of Cloned Human Chromosomal Fragments. All cloning experiments were carried out in accordance with the recommended National Institutes of Health guidelines for recombinant DNA research. Isolation and restriction mapping of DNA fragments from a human genomic DNA library cloned in λ Charon 4A bacteriophage were done as described (10). Either the entire Charon 4A recombinant or the C_γ-containing 6.4-kilobase-pair (kb) *Hin*dIII fragment of clone 5A was subcloned into the phage M13mp2. The DNA was first digested with *Hin*dIII plus either *Ava* II or *Alu* I. *Eco*RI-cleaved M13mp2 DNA and the fragments to be cloned were made flush-ended by treatment with T4 DNA polymerase and then were blunt-end ligated (12). The ligation mixture was used to transform *Escherichia coli* strain JM101, and C_γ-containing clones were isolated after screening of plaques (13) with the subcloned $C_{\gamma 4}$ gene from clone 24B (see Fig. 1 and ref. 10).

DNA Sequence Analysis. Individual M13 subclones were analyzed by the dideoxy technique (14, 15) essentially as described by Steinmetz *et al.* (11), except that [α-^{32}P]dATP was used as the labeled precursor. Alignment of the analyzed fragments yielding the composite C_γ DNA sequence was determined by either overlaps of *Ava* II and *Alu* I fragments or by homology of the translated DNA sequence to existing sequence data for a human γ2 protein (3).

RESULTS

Human $C_{\gamma 2}$ and $C_{\gamma 4}$ Genes Are Linked. Human genomic DNA clones hybridizing to a human $C_{\gamma 3}$ cDNA probe were isolated as described (10). Restriction maps for the inserts of five of these clones indicate that the corresponding chromosomal fragments overlap (Fig. 1). From clone blot hybridization experiments with the $C_{\gamma 3}$ probe, we deduce that two separate regions on the composite stretch of human DNA contain C_γ sequences. We have previously determined the nucleotide sequence of the gene on the right of Fig. 1 and found that it encodes a $C_{\gamma 4}$ region (10). The sequence analysis of the $C_{\gamma 4}$ gene indicated that it is transcribed from left to right in Fig. 1, allowing us to orient the mRNA synonymous strand as shown. Below we show that the gene lying 17 kb 5′ to $C_{\gamma 4}$ is a $C_{\gamma 2}$ gene and that it also is transcribed from left to right in Fig. 1. This intergenic distance is comparable to the distances found be-

Abbreviations: C_H, constant region of heavy chain; kb, kilobase pair(s).

tween mouse C_γ genes, which have been shown to range from 17 to 34 kb (16).

Primary Structure of a Human $C_{\gamma 2}$ Gene. The nucleotide sequence of the C_γ gene contained in clone 5A is presented in Fig. 2. Translation of the coding regions indicates that this gene is transcribed in the same direction as the $C_{\gamma 4}$ gene and that it encodes the constant region of a γ2 protein. This conclusion is based on comparison with the complete sequence of the myeloma γ2 protein Til (3). The predicted protein sequence agrees with the Til sequence at all but three positions. Two of these are differences in amide assignment: we found a glutamic acid codon for position 20 of the C_H1 coding region and an asparagine codon for position 84 of the C_H2 exon, whereas Wang *et al.* (3) found glutamine and aspartic acid, respectively, for these residues. The other difference is a serine ↔ alanine interchange at position 60 of the C_H1 domain. The hinge segment, which most clearly distinguishes the human C_γ subclasses, is identical for the putative $C_{\gamma 2}$ gene and protein sequences. Thus we feel confident in classifying this gene as a $C_{\gamma 2}$ gene.

Our sequence of the $C_{\gamma 2}$ gene, which begins 214 nucleotides 5′ to the C_H1 coding region and continues 207 residues past the termination codon, contains the same general structural features that we previously observed for the human $C_{\gamma 4}$ gene (10). The constant region and hinge exons are separated from one another by intervening DNA sequences (whose lengths are virtually identical in the two genes), and characteristic residues are present at the intron–exon junctions which presumably play a role in determining the proper splicing of the coding segments in the nuclear RNA precursor (17). The hexanucleotide A-A-T-A-A-A, which has been implicated as a signal sequence for the polyadenylylation of eukaryotic structural gene transcripts (18), is centered 103 nucleotides 3′ to the translation stop codon. As observed for the human $C_{\gamma 4}$ gene (10) and mouse C_γ genes (8, 19, 20), a lysine residue not present in the mature protein is encoded at the COOH terminus of the C_H3 exon.

Comparison of $C_{\gamma 2}$ and $C_{\gamma 4}$ Gene Sequences. In Fig. 2 the nucleotide sequences of the human $C_{\gamma 2}$ and $C_{\gamma 4}$ genes are aligned for direct comparison. The substitutions leading to coding differences are indicated by listing the distinguishing amino acids of the $C_{\gamma 4}$ gene. The $C_{\gamma 2}$ region contains one less amino acid than its $C_{\gamma 4}$ counterpart; the apparent deletion is located near the NH_2 terminus of the C_H2 domain [we refer to the difference as a deletion because the C_H2 domains of the human γ1 and γ3 heavy chains contain the same number of amino acids as does the corresponding domain of the $C_{\gamma 4}$ region (2, 4)]. The nucleotide sequence alignment giving maximal homology in this region suggests that the coding difference arose from deletion or insertion events at two sites.

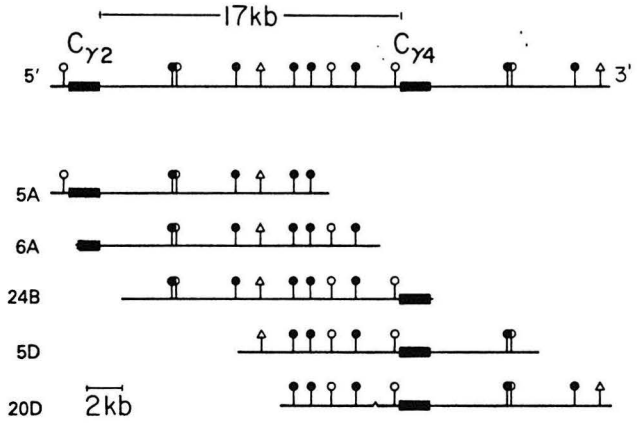

FIG. 1. Restriction map of a chromosomal region containing two human C_γ genes. The five lowest lines represent the human DNA inserts of recombinants cloned in Charon 4A. Positions of cleavage of the restriction enzymes *Hin*dIII (⚲), *Bam*HI (🡇), and *Eco*RI (🡅) are indicated. The restriction maps indicate that the cloned fragments overlap as shown. Solid bars represent the regions encompassed by C_γ coding sequences; introns are not shown. A composite map of the chromosomal region spanned by the cloned fragments appears at the top. Arrows indicate the 5′-to-3′ orientation of the mRNA synonymous strands of the two C_γ genes (see text). The *Bam*HI–*Hin*dIII fragment immediately 5′ to the $C_{\gamma 4}$ gene in clone 20D is approximately 100 nucleotides longer than the corresponding fragment in clones 24B and 5D. This conceivably could represent allelic variation.

Table 1 lists the nucleotide differences in the various coding and noncoding segments of the two human genes. Two notable features of the homology relationship are evident from the data. First, the noncoding regions show nearly as much homology as do the C_H domain exons (≈95%). Second, the hinge exons are only about 70% homologous and thus are far more divergent than any of the other coding or noncoding regions.

In studies of recently diverged genes, Perler *et al.* (21) determined that noncoding nucleotide substitutions appear at a rate of approximately 7×10^{-9} nucleotide substitutions per site per year. Assuming that these substitutions are phenotypically silent and thus not subjected to natural selection, this rate approximates the actual mutation rate and is presumed to be linear over a relatively short evolutionary period [about 100 million years according to these authors (21)]. Thus we should be able to use the data of Table 1 to estimate the time of divergence of the human $C_{\gamma 2}$ and $C_{\gamma 4}$ genes from a common ancestral sequence. Using the total percentage divergence in noncoding regions (4.6%) and the above substitution rate, we estimate that approximately 6.6 million years have elapsed since divergence of the human $C_{\gamma 2}$ and $C_{\gamma 4}$ genes.

DISCUSSION

Human C_γ Coding Sequences. We previously observed that the predicted protein sequence encoded in our $C_{\gamma 4}$ gene differs by a single residue from the partially determined sequence of a human γ4 protein (10). The $C_{\gamma 2}$ region encoded by the gene reported here is seen to differ by three amino acid residues from the complete sequence determined for the human γ2 heavy chain Til (3). Most of the sequence of the constant region of another human γ2 chain has been determined (22), and comparison of the three $C_{\gamma 2}$ sequences (two protein and one DNA) reveals three to four interchanges between all pairs of compared sequences, most of which are due to differences in amide assignment. We cannot be certain that any of these differences reflect genetic polymorphisms rather than technical artifacts. Thus, the protein polymorphisms seen in this small sample of $C_{\gamma 2}$ and $C_{\gamma 4}$ sequences are quite limited and possibly nonexistent. This observation is not surprising, given the paucity of different allotypes observed for human γ2 and γ4 chains (23). In contrast, the human γ1 and γ3 chains exhibit a large number of distinct genetic variants (24).

One striking feature of the coding sequences compared here is the extensive divergence in the hinge exons. Fig. 2 and Table 1 reveal that the hinge exons are situated between two introns showing the same high degree of homology as the C_γ exons. Assuming that changes in intron sequences are not acted on by natural selection, we expect the rate of appearance of these changes to reflect the true mutation rate. The hinge sequences, however, show a much greater rate of genetically fixed change than the observed presumed basal level of the introns. We believe that the high rate of base substitution in hinge exons is due to rapid fixation of the substituted nucleotides by natural selection. There apparently is a selective advantage to generating diversity in the hinge coding region. The nature of this advantage is not obvious, although others have speculated about dif-

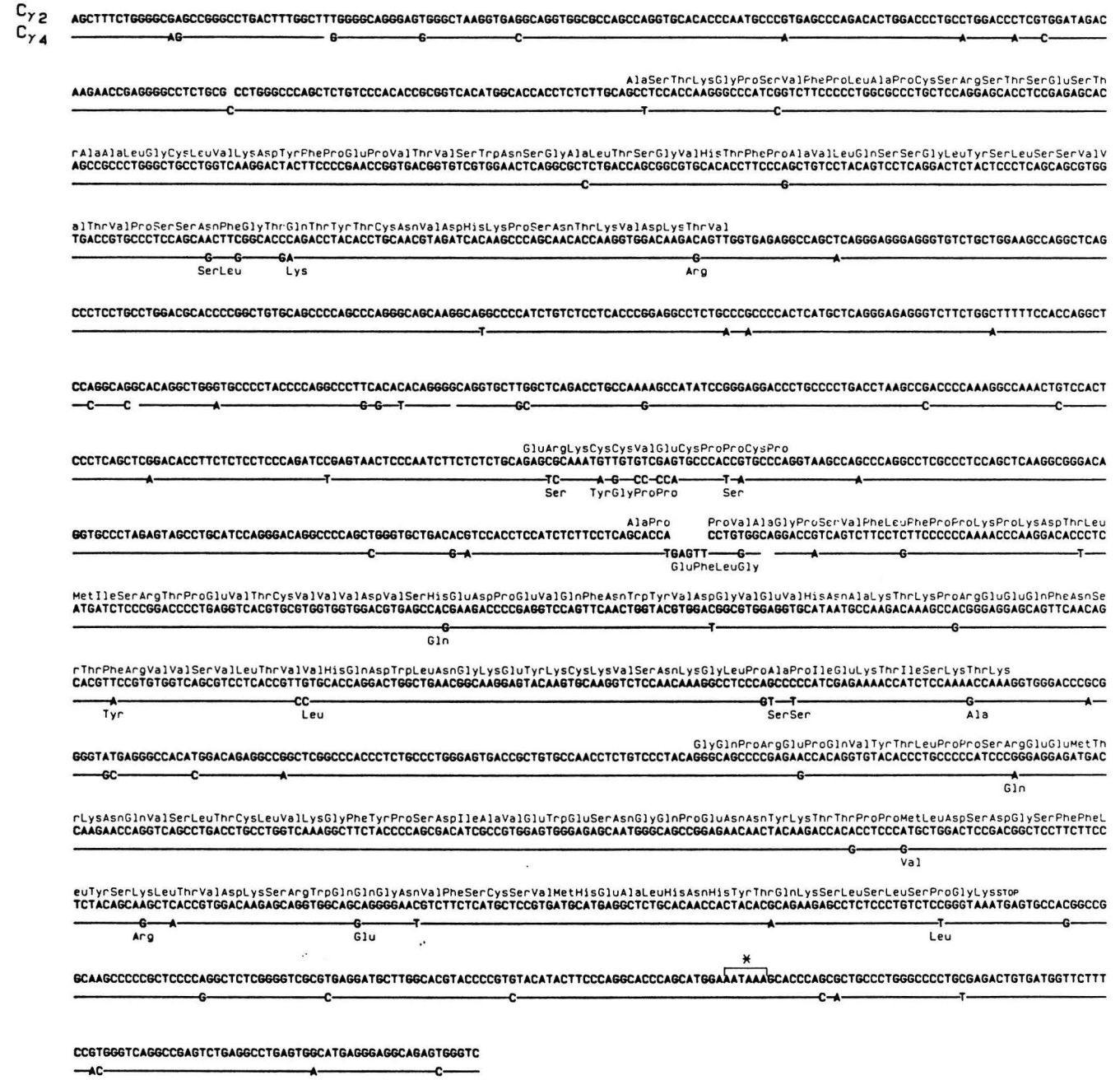

FIG. 2. Nucleotide sequences of the human $C_{\gamma 2}$ and $C_{\gamma 4}$ genes. The complete $C_{\gamma 2}$ protein sequence is listed above the corresponding DNA sequence. Solid lines represent identity of the $C_{\gamma 4}$ gene to the $C_{\gamma 2}$ sequence; where differences occur, the $C_{\gamma 4}$ sequence is indicated. A total of 2015 nucleotide positions are compared, including sequence gaps introduced into both genes to maintain sequence homology. Amino acid differences encoded in the $C_{\gamma 4}$ sequence are indicated by listing the $C_{\gamma 4}$-specific amino acids below the corresponding codons. Note that seven nucleotides near the 5′ end of the C_H2 exon are read in different translational reading frames in the two genes. "Stop" indicates the termination codon UGA. The presumptive poly(A) addition signal sequence is marked by an asterisk.

ferences in hinge sequences leading to different effector functions for the IgG subclass molecules (3, 25–28).

Other Genetic and Evolutionary Considerations. Yamawaki-Kataoka *et al.* (8) performed a detailed sequence comparison of three of the four mouse C_γ genes, those encoding the C_H regions of γ1, γ2a, and γ2b proteins. The homologies among these genes are much less than the homology seen for the human C_γ genes reported here [on the order of 50–80%, depending on the genes compared (8)]. The mouse C_γ genes, however, are more similar to one another than they are to the human genes (refs. 8 and 10, and this work). One way to explain the greater intraspecies homologies is to postulate that the common ancestor of humans and mice had a single C_γ gene and that multiple C_γ genes in the two species arose from independent gene duplication events occurring after speciation. Another explanation is that the common ancestor possessed multiple C_γ genes and that gene correction mechanisms have operated to maintain sequence homology within the respective C_γ gene families. According to these views, because the human C_γ genes are more similar to one another than are the mouse genes, duplication or correction events have occurred more recently in the C_γ genes of humans than of mice.

Gene correction can be defined, in the context of a multigene family, as a genetic process whereby the sequence of all or part

Table 1. Sequence differences between $C_{\gamma 2}$ and $C_{\gamma 4}$ genes

Gene segment	Residues compared, no.	Gaps*	Substitutions, no.		Difference, %		
			Noncoding or synonymous coding	Coding replacement	Silent or noncoding	Replacement	Total
Coding:							
C_H1	294		5	4	1.7	1.4	3.1
Hinge	36		2	9	5.6	25.0	30.6
C_H2	325†	7‡	9	6	2.8	1.8	4.6
C_H3	321		5	5	1.6	1.6	3.2
Noncoding:							
5′ flanking	213	2	9		4.2		4.2
C_H1–hinge intron	390	2	18		4.6		4.6
Hinge–C_H2 intron	118		4		3.4		3.4
C_H2–C_H3 intron	97		5		5.1		5.1
3′ untranslated§	125		6		4.8		4.8
3′ flanking	82		5		6.1		6.1

* These were introduced as noted in Fig. 2; the relevant residues were not compared and do not contribute to the calculation of % difference.

† The homologous residues read in different reading frames are not compared here.

‡ See Fig. 2.

§ This region extends from the residue immediately 3′ to the termination codon to the site of poly(A) addition. We have tentatively placed this latter site 125 nucleotides 3′ to the stop codon (see figure 5 of ref. 10).

of the corrected gene is replaced by the sequence of a homologous nonallelic gene (9). Two mechanisms for gene correction among tandemly linked genes are homologous unequal crossover and gene conversion. Both models assume that the genetic recombination takes place between tandem gene arrays that are in phase but out of register, so that apposing DNA sequences are homologous but not identical. When unequal crossover events occur in intergenic regions, the result is expansion and contraction of the size of the gene family. Repeated events of this type during evolution can result in fixation of the sequence of a single family member at the expense of the other members (29). The result of an unequal crossover event within nonallelic structural genes is the production of a hybrid gene, the classic example being that of hemoglobin Lepore (30). Unlike unequal crossovers, gene conversion events do not change the size of the gene family. Rather, a given stretch of DNA sequence of one gene is replaced by the sequence of another nonallelic gene through a recombination event.

In light of these considerations and our determination of the $C_{\gamma 2}$ and $C_{\gamma 4}$ gene order, it is interesting to examine the case of a human IgG molecule that apparently resulted from a recombination between the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes. Natvig and Kunkel (31) described a myeloma protein in which the C_H1 and C_H2 domains were characteristic of a γ4 chain and the C_H3 domain resembled that of a γ2 polypeptide. Fig. 3*A* schematically diagrams an unequal crossover involving $C_{\gamma 2}$ and $C_{\gamma 4}$ genes linked in the order we have determined. Both products of the recombination event contain hybrid genes; chromosome I contains only a Lepore-like gene, whereas chromosome II contains normal $C_{\gamma 2}$ and $C_{\gamma 4}$ genes in addition to a Lepore-like gene. In studies of sera from normal volunteers which apparently contained the normal counterpart of the hybrid myeloma protein, Natvig and Kunkel found that these individuals expressed normal IgG2 molecules, although they lacked IgG4 proteins. Chromosome II of Fig. 3*A* conceivably could represent the configuration in the DNA of these individuals, if one assumes that the normal $C_{\gamma 4}$ gene present is not expressed. Upstream (5′) to both the normal $C_{\gamma 4}$ gene and the hybrid gene are DNA sequences that mediate class switching. If C_H switching is subclass-specific, the relevant sequences 5′ to the normal $C_{\gamma 4}$ gene may be excluded from the switching event by virtue of their being downstream from the same sequences present near the hybrid gene. If this explanation for the failure to express the normal $C_{\gamma 4}$ gene is correct, it may shed some light on the mechanisms whereby C_H genes are selected for class-switch recombination.

A second explanation for the above clinical observations is

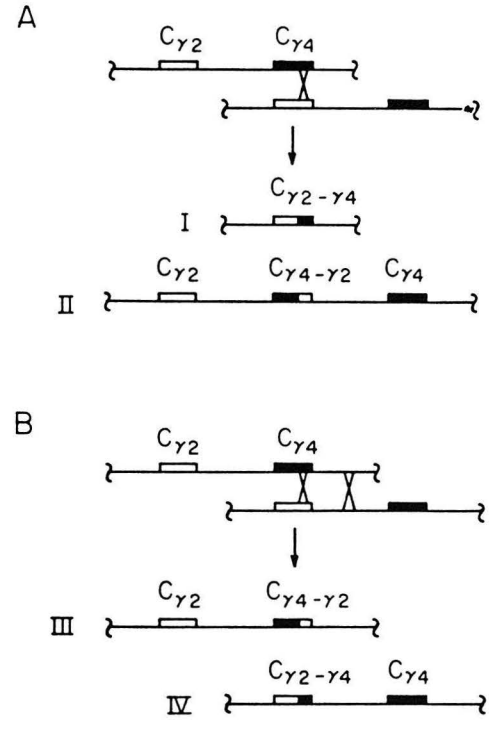

FIG. 3. Two models for the generation of a hybrid $C_{\gamma 4}$–$C_{\gamma 2}$ protein. (*A*). Unequal crossing over. A crossover occurs between misaligned chromosomes. The site of reciprocal exchange is indicated by the crossed lines. The two resulting chromosomes are labeled I and II. For simplicity we refer to the species participating in the genetic exchange as homologous chromosomes, but sister chromatids could equally well be involved. (*B*). Gene conversion. Gene conversion results in genetic information from one gene being copied onto a second gene, thus leading to gene correction. The chromosomes resulting from the gene conversion are labeled III and IV.

presented schematically in Fig. 3*B*. Here, a gene conversion process is proposed which involves a genetic exchange between aligned $C_{\gamma 2}$ and $C_{\gamma 4}$ genes. Chromosome III of Fig. 3*B* is consistent with the phenotype expressed by the volunteer individuals mentioned above. The gene conversion model is the simpler of the two proposals in that it does not require an *ad hoc* explanation for the failure to express a normal $C_{\gamma 4}$ gene.

Several examples of apparent gene correction have been reported for mammalian genes (8, 32, 33). In all of these cases, recombination points are proposed to lie within the structural genes rather than in intergenic regions. This conclusion follows from the observation that the levels of homology are different in different parts of the genes involved. Relatively sharp boundaries (representing presumed recombination break points) are observed, on either side of which are regions of greater and lesser homology. This pattern is evident when one examines rodent C_γ sequences. For the case of the two most similar mouse genes ($C_{\gamma 2a}$ and $C_{\gamma 2b}$), the percentage of silent site substitutions in the C_H1 and C_H3 exons is 2–3 times greater than the corresponding neutral changes seen in the region between these exons (8). Comparison of guinea pig IgG subclass sequences also reveals disparate levels of homology in different C_H domains. The $C_{\gamma 1}$ and $C_{\gamma 2}$ protein sequences are >90% identical in C_H1 domains, whereas the homology is 64–70% in C_H2 and C_H3 domains (28). In contrast, the two human C_γ genes compared here show a nearly constant level of homology among the different exons and introns (except for the hinge region, as noted above). Thus, if a gene correction mechanism is responsible for the extensive homology between these genes, we would propose that the recombination break points of the most recent correction event lie outside the regions we have analyzed.

Clinical studies have uncovered a number of examples of apparent duplications and deletions of human C_γ genes (34–37), as well as several cases of hybrid human γ heavy chains (31, 38, 39). Thus, it appears that unequal crossover or gene conversion events occur at a significant frequency among human C_γ genes. These types of events can be envisioned to play a part in both gene duplication and gene correction. It seems reasonable to view duplication and correction as two different manifestations of the same fundamental genetic process, one which has played an important role in the evolution of the human C_γ gene family.

We thank Karyl Minard for excellent technical assistance and Bernita Larsh for the preparation of the manuscript. This work was supported by Grant GM 20927 from the National Institutes of Health and National Research Service Award 1T32 GM07616 from the National Institute of General Medical Sciences to J.E.

1. Kunkel, H. G., Allen, J. C., Grey, H. M., Martensson, L. & Grubb, R. (1964) *Nature (London)* **203**, 413–414.
2. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 78–85.
3. Wang, A.-C., Tung, E. & Fudenberg, H. H. (1980) *J. Immunol.* **125**, 1048–1054.
4. Frangione, B., Rosenwasser, E., Prelli, F. & Franklin, E. C. (1980) *Biochemistry* **19**, 4304–4308.
5. Pink, J. R. L., Buttery, S. H., DeVries, G. M. & Milstein, C. (1970) *Biochem. J.* **117**, 33–47.
6. Nisonoff, A., Hopper, J. E. & Spring, S. B. (1975) in *The Antibody Molecule*, eds. Dixon, F. J., Jr., & Kunkel, H. G. (Academic, New York), pp. 314–318.
7. Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2143–2147.
8. Yamawaki-Kataoka, Y., Miyata, T. & Honjo, T. (1981) *Nucleic Acids Res.* **9**, 1365–1381.
9. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Annu. Rev. Genet.* **9**, 305–353.
10. Ellison, J., Buxbaum, J. & Hood, L. (1981) *DNA* **1**, 11–18.
11. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. & Hood, L. (1981) *Cell* **25**, 683–692.
12. Wartell, R. M. & Reznikoff, W. S. (1980) *Gene* **9**, 307–319.
13. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
14. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
15. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321.
16. Roeder, W., Maki, R., Traunecker, A. & Tonegawa, S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 474–478.
17. Lewin, B. (1980) *Cell* **22**, 324–326.
18. Proudfoot, N. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
19. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
20. Tucker, P. W., Marcu, K. B., Newell, N., Richards, J. & Blattner, F. R. (1979) *Science* **206**, 1303–1306.
21. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20**, 555–566.
22. Connell, G. E., Parr, D. M. & Hofmann, T. (1979) *Can. J. Biochem.* **57**, 758–767.
23. Schanfield, M. S. (1978) in *Basic and Clinical Immunology*, eds. Fudenberg, H. H., Stites, D. P., Caldwell, J. L. & Wells, J. V. (Lange, Los Altos, CA), p. 59.
24. W.H.O. meeting announcement (1976) *J. Immunol.* **117**, 1056–1058.
25. Brunhouse, R. & Cebra, J. J. (1979) *Mol. Immunol.* **16**, 907–917.
26. Isenman, D. E., Dorrington, K. J. & Painter, R. H. (1975) *J. Immunol.* **114**, 1726–1729.
27. Boackle, R. J., Johnson, B. J. & Caughman, G. B. (1979) *Nature (London)* **282**, 742–743.
28. Cebra, J. J., Brunhouse, R. F., Cordle, C. T., Daiss, J., Fecheimer, M., Ricardo, M., Thunberg, A. & Wolfe, P. B. (1977) *Prog. Immunol.* **3**, 264–277.
29. Smith, G. P. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 507–514.
30. Flavell, R. A., Kooter, J. M., De Boer, E., Little, P. F. R. & Williamson, R. (1978) *Cell* **15**, 25–41.
31. Natvig, J. B. & Kunkel, H. G. (1974) *J. Immunol.* **112**, 1277–1284.
32. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
33. Liebhaber, S. A., Goossens, M. & Kan, Y. W. (1981) *Nature (London)* **290**, 26–29.
34. van Loghem, E., Sukernik, R. I., Osipova, L. P., Zegers, B. J. M., Matsumoto, H., de Lange, G. & Lefranc, G. (1980) *J. Immunogenet.* **7**, 285–299.
35. Lefranc, G., Dumitresco, S.-M., Salier, J.-P., Rivat, L., de Lange, G., van Loghem, E. & Loiselet, J. (1979) *J. Immunogenet.* **6**, 215–221.
36. Lefranc, G., Rivat, L., Rivat, C., Loiselet, J. & Ropartz, C. (1976) *Am. J. Hum. Genet.* **28**, 51–61.
37. van Loghem, E. & Natvig, J. B. (1970) *Vox Sang.* **18**, 421–434.
38. Kunkel, H. G., Natvig, J. B. & Joslin, F. G. (1969) *Proc. Natl. Acad. Sci. USA* **62**, 144–149.
39. Arnand, P., Wang, A.-C., Gianazza, E., Wang, I. Y., Lasue, Y., Creyssel, R. & Fudenberg, H. H. (1981) *Mol. Immunol.* **18**, 379–384.

CHAPTER 4

The Nucleotide Sequence of a Human Immunoglobulin $C_{\gamma 1}$ Gene

(Published in Nucleic Acids Research)

Volume 10 Number 13 1982 Nucleic Acids Research

The nucleotide sequence of a human immunoglobulin $C_{\gamma 1}$ gene

Jay W.Ellison, Bennett J.Berson and Leroy E.Hood

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received 12 April 1982; Revised and Accepted 2 June 1982

ABSTRACT

We report the nucleotide sequence of a gene encoding the constant region of a human immunoglobulin γ1 heavy chain ($C_{\gamma 1}$). A comparison of this sequence with those of the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes reveals that these three human C_γ genes share considerable homology in both coding and noncoding regions. The nucleotide sequence differences indicate that these genes diverged from one another approximately 6-8 million years ago. An examination of hinge exons shows that these coding regions have evolved more rapidly than any other areas of the C_γ genes in terms of both base substitution and deletion/insertion events. Coding sequence diversity also is observed in areas of C_H domains which border the hinge.

INTRODUCTION

Immunoglobulin G (IgG) molecules in humans are divided into four subclasses based on the presence of particular gamma heavy chain constant regions (C_γ). These C_γ regions ($C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\gamma 3}$, and $C_{\gamma 4}$) are encoded by distinct germline genes (1) which are presumed to be the products of gene duplication of an ancestral C_γ gene. Several species of mammals have been shown to possess IgG subclasses, although the number of subclasses varies for different species. For example, both humans and mice have four subclasses, while guinea pigs have two and rabbits have only a single type of IgG. Structural studies at the protein and DNA level have been carried out with several species, and have shown that the homology relationships within the C_γ gene families are different for different mammals (2-9). For example, human C_γ protein regions are over 90% homologous (2-5), while mouse C_γ genes share significantly less homology (70-80% at the nucleotide level (6-8)). Moreover, cross-species comparisons reveal no clear correspondence between individual human and mouse genes. These intra- and interspecies homology relationships, as well as the different numbers of C_γ genes found in different mammals, indicate that the various mammalian C_γ gene families have evolved quite differently since the time of mammalian speciation.

We are interested in studying structural features of human C_γ genes in order to gain insights into the evolution of the human C_γ gene family. We have previously

0305-1048/82/1013-4071$ 2.00/0

characterized the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes (10,11). In this paper we report the complete nucleotide sequence of a $C_{\gamma 1}$ gene and compare the three human C_{γ} sequences.

MATERIALS AND METHODS

Materials

The human fetal liver DNA library was obtained from T. Maniatis. Sources of nucleic acid enzymes, reagents for DNA sequencing, E. coli strain JM101, and the phage M13mp2 were those described by Steinmetz et al. (12).

Isolation and restriction mapping of a human $C_{\gamma 1}$ genomic clone

Screening of a human fetal liver DNA library cloned in lambda Charon 4A bacteriophage with a human $C_{\gamma 3}$ cDNA probe was done as previously described (10). Mapping of restriction sites for the enzymes Eco RI, Bam HI, Hind III, Xba I, Bgl II, and Pvu II was done by analysis of single and double digests with these enzymes.

Subcloning and DNA sequence analysis

The 3.0 kb Hind III-Pvu II fragment of clone HG3A (see Fig. 1) was digested separately with frequent-cutting restriction enzymes and the products were subcloned into the phage M13mp2 as described (11). Subclones were chosen for sequence analysis following screening of plaques with a labelled genomic fragment containing a full-length $C_{\gamma 4}$ gene (see refs. 10 and 11). DNA sequencing of individual subclones was carried out as described (11). The composite C_{γ} DNA sequence was determined either by overlaps of sequenced regions or by homology of the translated DNA sequence to existing sequence data for a human immunoglobulin γ1 protein (2).

RESULTS AND DISCUSSION

The primary structure of a human $C_{\gamma 1}$ gene

We have previously described the isolation of human C_{γ} genes from a recombinant phage library of fetal liver DNA, using as hybridization probe a cDNA encoding part of a $C_{\gamma 3}$ gene (10). One of these clones, HG3A, is shown diagrammatically in Fig. 1. The restriction map of this clone indicated that it is a distinct species from the clones shown to contain $C_{\gamma 2}$ and $C_{\gamma 4}$ genes (10,11). A 2.0 kb region from clone HG3A containing sequences hybridizing to a full-length $C_{\gamma 4}$ gene was sequenced by the dideoxynucleotide chain-termination method in the phage M13mp2. The sequence obtained is shown in Fig. 2, where we see that the gene has the same basic exon-intron organization that has been previously observed for both human (10,11) and mouse (6-8) C_{γ} genes. The three C_H domains and the hinge segment of the polypeptide are encoded in individual exons that are separated from one another by introns, the largest one lying between the C_H1 and hinge exons. The predicted amino acid residues are listed above the corresponding codons in Fig. 2, and

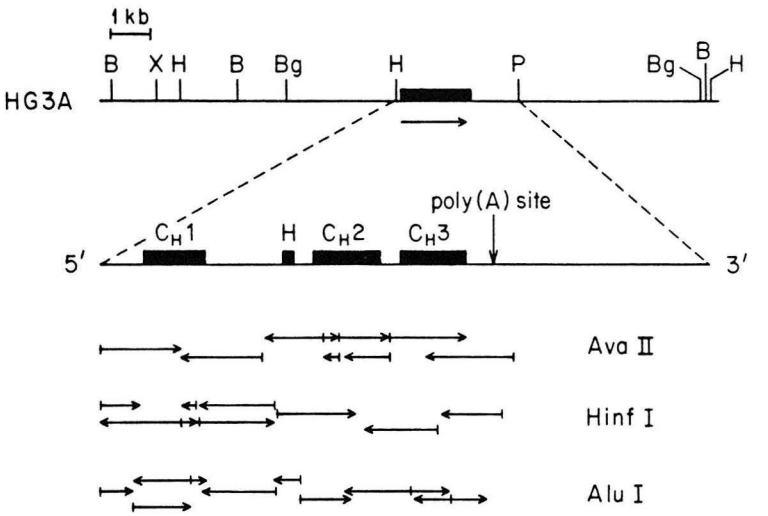

Figure 1. Restriction map and sequencing strategy of a cloned human DNA fragment containing a $C_{\gamma 1}$ gene. Letters on the top line refer to cleavage sites for the following restriction enzymes: B, Bam HI; H, Hind III; Bg, Bgl II; P, Pvu II; X, Xba I. Only the indicated Pvu II site was mapped, although this enzyme also cuts in other places in the clone. The arrow under the solid block indicates the direction of transcription. The dashed lines lead to an enlarged view of the region which was sequenced. Individual exons are shown here as solid blocks, whereas introns are not indicated at the top of the Figure. The arrowed lines represent the extent and direction of sequence determinations of individual subclones generated using the indicated enzymes.

a comparison of this protein sequence with that of the heavy chains of the two human IgG1 molecules Eu (2) and Nie (13) lead to an unambiguous designation of the cloned sequence as a $C_{\gamma 1}$ gene. Except for differences in amide assignments of several residues, the encoded protein sequence differs from the Eu sequence at just three of 329 compared residues, and only one difference is seen in a comparison with the Nie heavy chain. These differences do not include the lysine encoded at the C-terminus of the C_H3 domain, which has been observed in mouse (6-8) and human (10,11) C_γ genes but does not appear in the mature polypeptides. Table 1 compares the lengths of the exons and introns of the human and mouse C_γ genes that have been sequenced to date. Although some variation is seen in the lengths of noncoding regions and hinge exons, the overall organization of the C_γ genes is conserved in humans and mice.

Antigenic determinants have been found on human IgG molecules which can serve as genetic markers for C_H regions (14). Some of these allelic variants, called allotypes, have been correlated with specific amino acid residues in the heavy chains

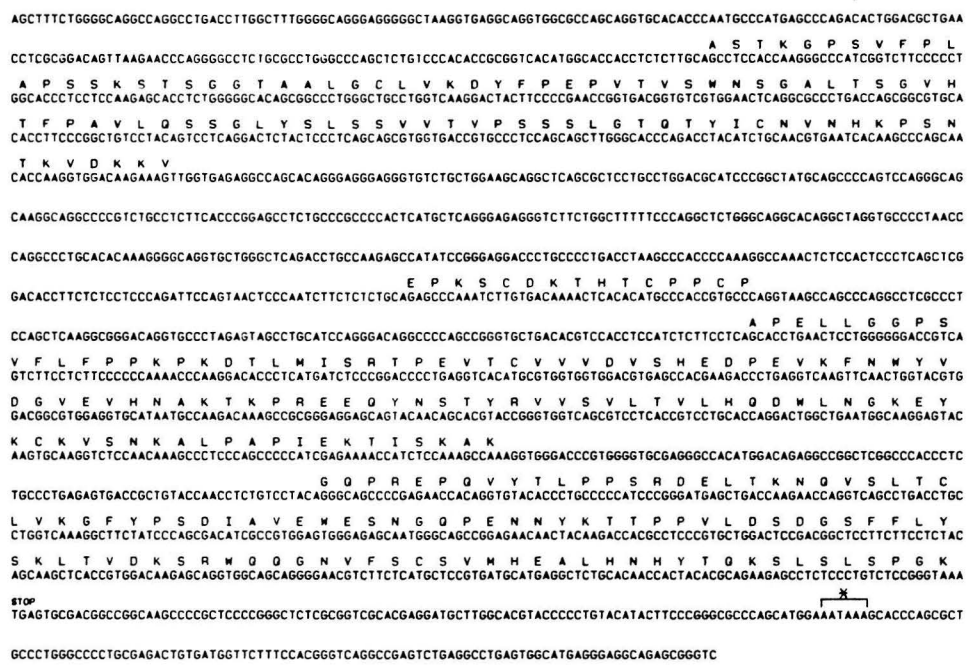

Figure 2. The nucleotide sequence of a human $C_{\gamma 1}$ gene and its corresponding protein sequence. The sequence of the mRNA synonymous strand is listed 5' to 3'. Amino acids predicted by the DNA sequences are listed in one-letter code above the respective codons. "Stop" indicates the termination codon UGA. The presumptive poly(A) addition signal sequence is marked by an asterisk.

(15). We find that the discrepant residues in the Eu heavy chain and the encoded polypeptide reported here can be correlated with certain of these allotypic markers. The lysine encoded at position 97 of the C_H1 domain (Fig. 2) correlates with the Gm (17) determinant, while the arginine at the corresponding place in the Eu heavy chain is associated with the Gm (3) marker. Similarly, the asp-glu-leu sequence at positions 16-18 of the C_H3 domain of the cloned gene are believed to represent the Gm (1) allotypic determinant, whereas the glu-glu-met present in Eu correlates with the Gm (non-1) variant. Thus the cloned gene reported here encodes a polypeptide with the genetic markers Gm (1,17). The Nie heavy chain also carries these markers, yet differs at amino acid number 41 of the C_H3 domain (Nie has arginine as compared to a tryptophan codon for the cloned sequence).

Sequence divergence among three human C_γ genes

We have previously reported the nucleotide sequences of genes encoding C_H regions of human $\gamma 2$ and $\gamma 4$ heavy chains (10,11). Our analysis of a $C_{\gamma 1}$ gene allows a

Table 1 Intron and exon lengths in C_γ genes

	length of gene segment (nucleotides)							
C_γ gene	C_H1	C_H1-hinge intron	hinge	hinge-C_H2 intron	C_H2	C_H2-C_H3 intron	C_H3	3' UT
human γ1	294	388	45	118	330	96	321	∿ 130
human γ2	294	392	36	118	327	97	321	∿ 130
human γ4	294	390	36	118	330	97	321	∿ 130
mouse γ1	291	356	39	98	321	121	321	93
mouse γ2a	291	310	48	107	330	112	321	103
mouse γ2b	291	316	66	107	330	112	321	103

The data for the mouse genes are from reference 8. The human γ2 and γ4 numbers come from references 11 and 10, respectively. The lengths of the 3' untranslated (UT) regions in the human genes are determined by homology to the corresponding regions in mouse C_γ genes (see Fig. 5 of reference 10).

comparison of three members of the human C_γ gene family. A summary of the nucleotide sequence comparisons is shown in Table 2. Nucleotide differences in the various noncoding regions are similar, and so values are listed for the total divergence in noncoding DNA. Similarly, each of the C_H exons show similar homologies among the three genes, and the total observed differences for these exons are given. Hinge exons, on the other hand, show much greater variation than any other gene segment, and these regions are separately compared. Table 2 shows that the level of nucleotide substitution (not including gaps) in noncoding areas is not much greater than the total (silent plus amino acid replacement) seen in the C_H coding regions. Except for areas surrounding the site of polyadenylation of the mRNA (16) and splice junctions (17), the noncoding segments of these genes have no known function. If these sequences are without any function, they are presumably not subjected to natural selection and are free to diverge. Estimates of the rate of appearance of nucleotide substitutions in unselected noncoding DNA (18) lead us to conclude that approximately 6-8 million years have elapsed since any two of these genes shared an identical sequence. The similar homology levels seen in the three pairwise comparisons make it difficult to determine which two genes shared the most recent

Table 2 Nucleotide sequence comparisons of three human immunoglobulin C_γ genes

	% nucleotide difference*				
		C_H exons		Hinge exons	
genes compared	total noncoding areas	silent	replacement	silent	replacement
γ1 vs. γ2	4.7 (14 gaps)‡	1.6	1.9	2.7	11.1
γ1 vs. γ4	5.4 (18 gaps)	2.3	2.2	2.7	16.7
γ2 vs. γ4	4.6 (4 gaps)	2.0	1.6	3.3	16.7

* This is calculated as (number of substitutions/number of residues compares) x 100. Gaps were not compared.

‡ These were introduced into one or another of the compared sequences to maintain the homology alignment.

common ancestor. However, significantly fewer gaps need to be placed in the noncoding areas of the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes to maintain the homology alignment of the two sequences. This observation along with the determined linkage of these genes (11) suggests that they diverged more recently from each other than from the $C_{\gamma 1}$ gene.

Coding sequence divergence in and near the hinge

The most interesting areas of these genes in evolutionary terms are the hinge exons, which Table 2 indicates are the most divergent gene segments. The differences listed do not reflect the fact that the $C_{\gamma 2}$ and $C_{\gamma 4}$ hinge exons encode three fewer amino acids than the $C_{\gamma 1}$ hinge exon, which codes for 15 residues. The DNA sequence alignment giving maximum homology among the three genes in this exon is shown in Fig. 3. Here we see that distinct nine-nucleotide gaps are placed in the $C_{\gamma 2}$ and $C_{\gamma 4}$ sequences. On either side of these gaps are small coding stretches which are homologous in the three C_γ genes. Every nucleotide substitution indicated in the $C_{\gamma 2}$ and $C_{\gamma 4}$ sequences is in a triplet which encodes an amino acid unique to that hinge region. The combination of nucleotide substitution and insertion/ deletion events leads to quite different coding properties in the hinge exons for the three C_γ genes. Fig. 4 shows the predicted amino acid sequences for the three hinge segments, as well as some contiguous residues in the C_H1 and C_H2 domains. The alignment shows that coding sequence diversity is not limited to the hinge exon itself, but is also

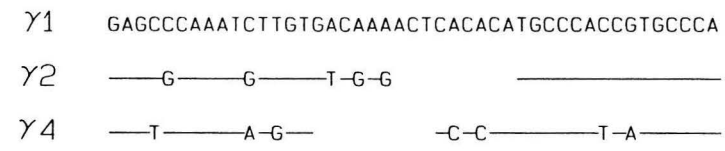

Figure 3. Comparison of hinge exon nucleotide sequences. Solid lines represent identity of the γ2 and γ4 sequences to the γ1 sequence. Where differences occur in the γ2 and γ4 exons, the relevant residues are listed. Gaps are introduced into the γ2 and γ4 listings to maximize homology to the γ1 sequence.

found in areas of the C_H domains which are adjacent to the hinge. Again both base substitution and insertion/deletion events produce coding differences; the latter type of event leads to nucleotides in the C_H2 exon of the $C_\gamma2$ gene being read in a different translational reading frame than their homologous counterparts in the other two genes (see Fig. 2 of ref. 11). Thus although the three genes encode polypeptides which are at least 95% identical over most of their length, amino acid substitutions are clustered in the hinge areas of the proteins. We believe that the high level of divergence in this region exists because natural selection favors the generation of diversity in this part of the molecule. This is not to say that the rate of nucleotide substitution is greater in the hinge than in the more conserved noncoding regions, but rather that substitutions in the hinge area are more rapidly fixed by selection. The nature of the selective advantage offered by hinge variation is not obvious, although it has been suggested that divergent hinges may be responsible for the differences in effector functions carried out by IgG subclasses (3,19,20). If this view is correct, then the generation of new and diverse effector functions may be the selective force which fixes nucleotide changes in the hinge area and the hinge exon itself.

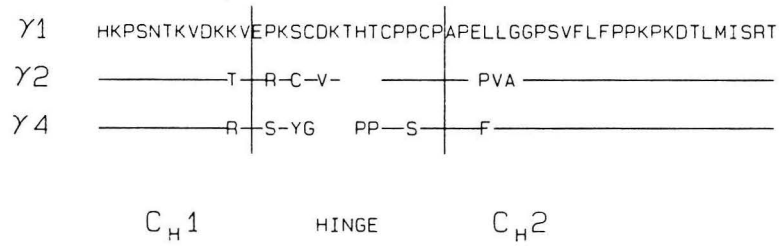

Figure 4. Comparison of amino acid residues in the hinge area of three C_γ polypeptides. Vertical lines separate the hinge residues from those contiguous amino acids which are encoded in the C_H1 and C_H2 exons. Amino acids are listed in the one-letter code. Solid lines represent identity of the γ2 and γ4 sequences to the γ1 sequence. The C_H2 domain of the $C_{\gamma2}$ sequence contains one less amino acid than is found in the other genes.

An unresolved evolutionary issue

Our current picture of human C_γ genes is that they have diverged recently from one another, and that hinge regions have evolved rapidly since that divergence. What is not clear is the nature of the genetic event(s) giving rise to the identical C_γ genes which were the ancestors of the present-day genes. There are two likely alternatives for the generation of two or more identical sequences: (1) a duplication of a single gene sequence, thus producing a gene de novo, and (2) a gene correction process (21) in which all or part of the sequence of one gene is replaced by the sequence of a nonallelic but homologous gene. The latter explanation implies that members of a multigene family do not evolve independently of one another, but rather that genetic information can be exchanged between nonallelic members of a gene cluster. Molecular evidence for the occurrence of such events has been cited for human (22,23) and mouse (24) globin genes and for mouse immunoglobin genes (8). Such evidence consists of the finding of a presumed recombination breakpoint which separates areas of a gene which either were or were not involved in a genetic exchange with another member of the gene family. This breakpoint defines a relatively sharp boundary on either side of which two nonallelic genes share different levels of homology. A boundary of this kind is not found in a comparison of the three human C_γ genes, since except for the extensive divergence found in the hinge region, the nucleotide differences are distributed rather evenly over the length of the genes. If evidence exists for recombination between any two of these nonallelic genes, it is most likely to be found in regions flanking the coding areas that we have characterized.

Thus we are unable to distinguish between the above two alternatives, although we have argued (11) that gene duplication and gene correction are not mutually exclusive concepts. The same kinds of fundamental genetic processes that result in gene duplication can also bring about gene correction. We think it likely that these genetic processes have continued to act on human C_γ genes since the occurrence of the initial duplication event(s). According to this view, our estimated time of divergence of human C_γ genes represents the time elapsed since the most recent correction event. Thus we believe that the human C_γ gene family is probably much older than indicated by the extensive homology shared by its members.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health and a National Research Service Award from the National Institute of General Medical Sciences to J.E. We thank Karyl Minard for excellent technical assistance and Stephanie Canada for the preparation of the manuscript.

REFERENCES

1. Kunkel, H. G., Allen, J. C., Grey, H. M., Martensson, L. and Grubb, R. (1964) Nature (London) 203, 413-414.
2. Edelman, G. M., Cummingham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. and Waxdal, M. J. (1969) Proc. Natl. Acad. Sci. USA 63, 78-85.
3. Wang, A.-C., Tung, E. and Fudenberg, H. H. (1980) J. Immunol. 125, 1048-1054.
4. Frangione, B., Rosenwasser, E., Prelli, F. and Franklin (1980) Biochemistry 19, 4304-4308.
5. Pink, J. R. L., Buttery, S. H., DeVries, G. M. and Milstein, C. (1970) Biochem. J. 117, 33-47.
6. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. and Mano, Y. (1979) Cell 18, 559-568.
7. Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M. and Honjo, T. (1980) Nature 283, 786-789.
8. Yamawaki-Katoaka, Y., Miyata, T. and Honjo, T. (1981) Nucleic Acids Res. 9, 1365-1381.
9. Brunhouse, R. and Cebra, J. J. (1976) Mol. Immunol. 16, 907-917.
10. Ellison, J., Buxbaum, J. and Hood, L. (1981) DNA 1, 11-18.
11. Ellison, J. and Hood, L. Proc. Natl. Acad. Sci. USA 79, 1984-1988.
12. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. and Hood, L. (1981) Cell 25, 683-692.
13. Ponstingl, H. and Hilschmann, N. (1972) Hoppe-Seyler's Z. Physiol. Chem. 353, 1369-1372.
14. W.H.O. meeting announcement (1976) J. Immunol. 117, 1056-1058.
15. Mage, R., Lieberman, R., Potter, M. and Terry, W. D. (1973) in The Antigens. (Sela, M. Ed.), Vol. I, p. 354, Academic Press, New York.
16. Proudfoot, N. and Brownlee, G. G. (1976) Nature 263, 211-214.
17. Lewin, B. (1980) Cell 22, 324-326.
18. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
19. Klein, M., Haeffner-Cavaillon, N., Isenman, D. E., Rivat, C., Navia, M. A., Davies, D. R. and Dorrington, K. J. (1981) Proc. Natl. Acad. Sci. USA 78, 524-528.
20. Cebra, J. J., Brunhouse, R. F., Cordle, C. T., Daiss, J., Fecheimer, M., Ricardo, M., Thunberg, A. and Wolfe, P. B. (1977) Prog. Immunol. 3, 264-277.
21. Hood, L., Campbell, J. H. and Elgin, S. C. R. (1975) Ann. Rev. Genet. 9, 305-353.
22. Slightom, J. L., Blechl, A. E. and Smithies, O. (1980) Cell 21, 627-638.
23. Liebhaber, S. A., Goossens, M. and Kau, Y. W. (1981) Nature (London) 290, 26-29.
24. Konkel, D. A., Maizel, J. V., Jr. and Leder, P. (1979) Cell 18, 865-873.

CHAPTER 5

Human Antibody Genes: Evolutionary and Molecular Genetic Perspectives

(Submitted to Advances in Human Genetics)

Human Antibody Genes: Evolutionary and Molecular Genetic Perspectives

Jay W. Ellison* and Leroy E. Hood

Division of Biology, California Institute of Technology

Pasadena, California 91125

*Present address: University of California, San Francisco, School of Medicine, San Francisco, California 94143.

INTRODUCTION

In the past several years a vast amount of information has been accumulated regarding the genes encoding immunoglobulin (Ig) molecules. Studies in the murine system have shown that the expression of these genes in lymphocytes involves a number of genetic processes, including DNA rearrangements, alternative patterns of RNA splicing, and somatic mutation, all of which serve to amplify the information encoded in the germline. The cloning of mouse genes has been facilitated by the availability of inducible myeloma tumors which produce large quantities of homogeneous immunoglobulins (Potter, 1972). These myeloma tumors represent clonal cell populations which have undergone the DNA rearrangements necessary for antibody gene expression, and have furnished large amounts of Ig mRNA, which has served as the starting material for the isolation of the corresponding genes.

Due in large part to sequence homology between mouse and human Igs, the antibody genes of man have recently become amenable to molecular analysis. The study of human Ig genes can be expected to provide us with valuable information pertinent to both basic and clinical issues. As well as complementing the body of knowledge existing for human Ig proteins, the analysis of human genes allows comparisons with mouse Ig genes. Such comparisons reveal functionally important protein and DNA regions, as implied by the conservation of structure between the two species. Detailed analyses of the genes and flanking regions may give clues to the molecular mechanisms of the genetic processes acting on these and other genes during evolution. Molecular characterization of human Ig genes should eventually help us to understand the events occurring in some clinical disorders in which abnormal expression of Ig genes is observed. In addition, the DNA rearrangements that occur during lymphocyte ontogeny can be used as molecular markers to characterize aberrant (e.g., neoplastic) cell populations with respect to their state of differentiation.

One promising practical application is the manipulation of cloned human antibody genes in order to obtain large quantities of homogeneous antibodies of desired specificities, which can be put to diagnostic and therapeutic use.

In this paper we summarize the recent work done on human Ig genes, and discuss some of the genetic, evolutionary, and clinical implications that have arisen from these studies.

MOUSE IMMUNOGLOBULIN GENES

Since studies with mice have furnished most of the information on antibody genes, a brief review of the salient features of the mouse genes is in order (for more detailed reviews, see Adams, 1980; Hood et al., 1981; Early and Hood, 1981). We may then consider the human genes in light of the described features of the murine system.

The gene segments encoding Ig molecules are divided into three unlinked gene families—two for light chains, lambda and kappa, and one for heavy chains (see Fig. 1). Heavy and light chains are composed of an amino-terminal variable (V) region that binds antigen, and a constant (C) region that is involved in immunological effector functions. Light chain V regions (V_L regions) are encoded by two physically separate germline elements: a V_L gene segment, encoding approximately the first 98 amino acids, and a J_L gene segment, which encodes the remaining approximately 13 residues of the V_L region (Brack et al., 1978). Heavy chain V regions (V_H regions) are encoded by germline V_H and J_H gene segments, as well as a D gene segment, which encodes from one to 15 amino acids lying between the V_H and J_H residues (Early et al., 1980a; Sakano et al., 1980). Constant regions of both heavy and light chains are encoded in blocks of nucleotides containing approximately 100-110 codons. Heavy chains are larger as a result of duplication of the C_H coding unit,

giving rise to the multidomain structure of C_H regions. Variable and constant domains of the heavy and light chain families all show structural homology to one another, and are believed to be derived from a common primordial gene unit. Sequence homology to antibody gene domains has also been observed for the cell-surface molecules β_2-microglobulin (Peterson et al., 1972), thy-1 (Williams and Gagnon, 1982), and class I (Steinmetz et al., 1981) and class II (Larhammar et al., 1982) histocompatibility antigens, indicating that the primordial gene unit is the ancestor of a large pool of mammalian genes related to a cell's interaction with its environment.

An obligatory step for the production of an Ig polypeptide is the process of V-gene formation, a DNA rearrangement occurring in lymphocytes in which the germline V region components are brought together in the proximity of a C gene (Brack et al., 1978; Early et al., 1980a). The rearrangement process for heavy chains is shown schematically in Fig. 2A. The three germline segments are brought into contiguity, with the deletion of the intervening DNA. Models for how this process occurs have been formulated based on the observations of specific conserved structures flanking those portions of the germline coding segments to be joined (Early et al., 1980a; Sakano et al., 1980). The conserved DNA sequences are proposed to serve as recognition signals for proteins which bring the germline segments into correct juxtaposition for the splicing event which makes them contiguous. An analogous process occurs during light chain V gene formation, except that D segments are not involved (Brack et al., 1978). Conservation between both heavy and light chain gene families of the recognition sequences flanking V and J segments suggests that the same or very similar enzymes mediate the DNA rearrangement for both types of V genes. This recombination system is thus quite ancient, since heavy and light chain families diverged more than 500 million years ago (Marchalonis and Cone, 1973).

Not all V gene rearrangements lead to expression of an Ig polypeptide. Such nonproductive rearrangements fall into three categories. The first consists of out-of-frame joinings of V_L-J_L, V_H-D, or D-J_H gene segments, in which these sequences are brought together in different translational reading frames (Altenburger et al., 1980; Max et al., 1980; Early and Hood, 1981). Another type of nonproductive rearrangement occurs when a germline V gene component is joined to an abnormal DNA sequence, such as the case where a V_K gene was rearranged to a site in the intron between J_K and C_K, with resulting deletion of the J_K gene segments (Choi et al., 1980; Schnell et al., 1980; Seidman and Leder, 1980). The third type of nonproductive rearrangement is restricted to heavy chain V-gene formation. Since three elements are involved in this rearrangement (V_H, D, and J_H), sometimes only V_H-D or D-J_H joinings occur (Ravetch et al., 1981; Sakano et al., 1981). Such incomplete rearrangements, as well as the other two classes of nonproductive events, have been observed on the unexpressed chromosomes of antibody-producing cells. This observation has prompted suggestions that nonproductive rearrangements may be the basis for allelic exclusion (Early and Hood, 1981), whereby the expression of antibody genes is limited to only one of the two homologous chromosomes carrying these genes.

One of the major incentives for studying antibody genes was to understand how an organism is able to assemble an enormous number of binding sites (V regions) to specifically recognize millions of different antigenic determinants. Several different factors now appear to be responsible for this V region diversity (reviewed in Huang et al., 1981). One of these factors is a large number of germline gene segments: mice have several hundred V_K and V_H gene segments, several different J gene segments for each of the three antibody families, and probably at least 10 D gene segments (Barstad et al., 1978; Seidman et al., 1978; Valbuena et al.,

1978; Kemp et al., 1979; Max et al., 1979; Sakano et al., 1979, 1980, 1981; Newell et al., 1980). The possibility of forming many different V-J and V-D-J combinations provides a large number of diverse V region structures. Furthermore, the exact points of joining of the germline segments during V-gene formation can generate unique structures at the junctions of the joined segments (Max et al., 1979; Sakano et al., 1979; Early et al., 1980a). Since Igs are comprised of both heavy and light chains, different pairwise combinations of these polypeptides serve to multiple the binding site variability of intact antibody molecules. Perhaps the most intriguing source of V region diversity is the occasional mutation of germline V gene segments that occurs during lymphocyte differentiation. Examples of these somatic sequence changes have been documented for both V_H (Bothwell et al., 1981; Crews et al., 1981) and V_L (Bernard et al., 1978; Selsing and Storb, 1981) genes. The mechanism generating these nucleotide substitutions is not presently understood, although one study has shown that the somatic mutations extend into regions flanking the germline V gene (Kim et al., 1981).

Through the occurrence of different kinds of duplication events, the germline gene segments of the three mouse gene families (lambda, kappa, and heavy) have become quite different. In the kappa family, any one of the several hundred germline V_κ gene segments may be joined to one of the four functional J segments situated near the single C_κ gene (Max et al., 1979; Sakano et al., 1979; Seidman et al., 1979; Valbuena et al., 1978). In contrast, there appear to be four J_λ and C_λ sequences and only two V_λ gene segments, and their arrangement with respect to one another is very different from the kappa case (Blomberg et al., 1981; Miller et al., 1981). Each of the C_λ gene segments has its own J_λ adjacent to it; these $J_\lambda C_\lambda$ units are organized into two pairs which are not closely linked (see Fig. 1). The current picture is that each of the V_λ genes lies immediately upstream of and can associate

with only one of the two C_λ loci. Thus V, J, and C segments are interspersed in the lambda gene family, whereas V_κ and J_κ clusters occur separately from each other and the C_κ gene. Both this different structural organization and number of germline components are probably related to the almost exclusive expression of kappa light chains in the mouse.

The five recognized classes of antibody molecules—IgM, IgD, IgG, IgA, and IgE—are distinguished from one another by the presence of particular heavy chain constant regions. These C_H regions are encoded by individual germline genes—C_μ, C_δ, C_γ, C_α, and C_ε—which are presumed to be the descendants of a single ancestral C_H gene. Mouse C_γ regions are in fact encoded by four related C_γ genes—$C_{\gamma 1}$, $C_{\gamma 2a}$, $C_{\gamma 2b}$, and $C_{\gamma 3}$. The C_H genes of mice are clustered on a 200-kilobase segment of chromosomal DNA (Shimizu et al., 1982). The process of V-gene formation brings a functional V gene close to, but not contiguous with, either a C_L or a C_μ gene. A mature mRNA is produced following transcription of the entire region and splicing of the coding regions in the nuclear RNA precursor. During the course of lymphocyte differentiation a given cell can change from expression of μ chains to a different heavy chain class with the same binding specificity (Levin et al., 1971; Wang et al., 1973; Raff, 1976). This class switch is accompanied by a second kind of somatic DNA rearrangement, depicted schematically in Fig. 2B, in which a downstream C_H gene displaces the C_μ gene from its position near the rearranged V_H gene (Davis et al., 1980; Katoaka et al., 1980). The recombination takes place between the switch (S) sequences 5' of the C_H gene to be expressed and the corresponding S sequences 5' of C_μ, and results in the deletion of the C_μ and other intervening C_H genes. DNA sequence analyses of these S regions have shown that they contain tandemly repeated oligonucleotides (Davis et al., 1980; Nikaido et al., 1981; Obata et al., 1981). Although the precise mechanism of the class switch recombination is unknown, the distinct

structural features of the S regions imply that C_H switching occurs by a fundamentally different process than V-gene formation.

In addition to the DNA rearrangements outlined above, alternative patterns of RNA splicing also play roles in the expression of Ig genes. This was first demonstrated for the case of the μ heavy chain gene (Early et al., 1980b). Two exons encoding intramembrane segments of μ chains are located 3' of the C_μ gene. As shown in Fig. 3A, following transcription of the entire region, alternate RNA splicing strategies appear to be used to produce mRNAs encoding either secreted or membrane-bound forms of μ polypeptides. Separate membrane exons for γ chains also have been characterized, indicating that a similar mechanism is employed for production of cell-surface IgG (Rogers et al., 1981; Tyler et al., 1981).

Differential RNA splicing also is implicated in the dual expression of IgM and IgD by B cells at an early stage of differentiation (Liu et al., 1980; Tucker et al., 1980; Moore et al., 1981). Figure 3B shows the proposed model in which the transcribed V gene sequences can be spliced to either C_μ or C_δ sequences to generate, respectively, μ or δ mRNAs. Evidence exists that a C_δ gene may be expressed in mature plasma cells after a typical class switch leading to deletion of the C_μ gene (Moore et al., 1981). Thus both RNA splicing and DNA rearrangement mechanisms may be employed in the expression of IgD molecules. These dual strategies also may be available for the expression of other heavy chain classes, as suggested by the recent observations that ε and γ chains can be produced in cells which apparently have not deleted the upstream C_H genes (Alt et al., 1982; Yaoita et al., 1982).

We have seen that several different genetic processes operate in somatic cells (i.e., lymphocytes) to bring about the expression of mouse Ig genes. The regulation of these processes remains a major unresolved question. In fact, the

precise mechanisms also are not completely clear; models have been proposed based almost entirely on the observed structures of the genes and flanking regions in differentiated and undifferentiated cells. Since the inferred molecular processes (V-gene formation, class switching, alternative RNA splicing) appear to be fundamental to the expression of Ig genes, we would expect that certain structural features would be conserved between human and mouse genes, and as outlined below, these expectations have been borne out. Thus far we have concentrated our discussion on the genetic processes taking place in somatic cells, while only briefly mentioning one of the types of events occurring in the germline: duplications of genes and gene segments. We shall see how variations in the frequency of these and related events during evolution results in different germline repertoires of Ig gene segments which serve as substrates for the somatic processes, with resultant variations in the outcomes of antibody gene expression.

HUMAN V GENES

Humans differ from mice in that while only 5% of mouse serum light chains are of the lambda type, about 30-40% of the antibody molecules in human serum contain lambda chains (Milstein, 1965). Correlated with the dominance of mouse kappa chains is the presence of several hundred germline V_κ gene segments, as compared to two V_λ gene segments. Human germline V_λ gene segments have not yet been characterized, so at this time we have no information concerning the relationship between V_λ gene number and the level of lambda chain expression in humans. However, Rabbitts and co-workers have studied human V_κ genes, and have concluded that humans have an order of magnitude fewer germline V_κ gene segments than mice (Bentley and Rabbitts, 1981), which is consistent with the limited number of kappa chain subgroups revealed by protein sequence data

(Milstein, 1967; Wang et al., 1973). It will be interesting to see if the putatively small (relative to mouse) human V_κ repertoire is compensated for by the presence of a large number of germline V_λ genes.

Human V_H genes are just beginning to be characterized (Matthyssens and Rabbitts, 1980), and it is unclear at present whether the human V_H pool is as large as that of mice, which appears to consist of hundreds of germline gene segments (Barstad et al., 1978; Kemp et al., 1979). The cloned human V_H and V_κ genes that have been studied share structural features with their mouse counterparts (Bentley and Rabbitts, 1980; Matthyssens and Rabbitts, 1980). At the 5' ends of these genes are regions coding for about twenty amino acids comprising the signal peptide characteristic of secreted proteins. These coding blocks are separated by intervening DNA from the contiguous residues of the germline V segment. The recognition sequences immediately 3' of germline V_H and V_κ gene segments, which are presumably involved in functional V-gene formation, also are highly conserved between humans and mice.

A major difference between V_H and V_L regions in both species is that a third element (a D segment) encodes part of V_H regions, while V_L regions are encoded only by germline V_L and J_L sequences. A family of four tandemly arranged D gene segments has been characterized in humans (Siebenlist et al., 1981). These elements have recognition sequences flanking them which allow the D gene segments to be joined on either side to V_H and J_H gene segments, respectively. It has been suggested that both human and mouse D gene segments may recombine among themselves to generate new D sequences (Ravetch et al., 1981; Kurosawa and Tonegawa, 1982). If D-D recombination can in fact take place, it could serve as yet another mechanism of antibody binding site diversification.

HUMAN C_L GENES

Constant region genes for human kappa and lambda light chains have been studied by Hieter, Leder and co-workers (Hieter et al., 1980, 1981a, 1982). Like mice, humans have a single C_κ gene, with multiple J_κ segments clustered on its 5' side (Fig. 4). The relatedness between individual members of the J_κ clusters of the two species indicates that loss and gain of J_κ gene segments has occurred during evolution (we will discuss this phenomenon in more detail later). Interestingly, the sequence homology between corresponding human and mouse J_κ coding regions (average, 82%) is significantly greater than that seen beween C_κ genes (68%). The strong conservation implies an important role for the resultant J structures of kappa light chains.

As expected from protein studies (Ein, 1968; Gibson et al., 1971; Fett and Deutsch, 1975; Solomon, 1977), multiple germline C_λ genes are found in the human genome. Six C_λ genes have been shown to be closely linked, separated by about 3 to 6 kilobases (kb) of DNA (Hieter et al., 1981a) (Fig. 4). Three of these correspond to specific lambda chain subtypes (indicated in Fig. 4). Although all the relevant DNA sequences have not yet been determined, initial comparisons indicate that C_λ genes are more conserved between humans and mice than are C_κ coding sequences (Hieter et al., 1980, 1981a). It has not been established whether each of the human C_λ genes has its own J_λ segment, as observed in mice. In addition to the six tandemly arranged genes, three more human C_λ genes exist which are not linked to the major cluster. At least one of these is a pseudogene, and is discussed below. It will be interesting to see whether the organization of human C_λ genes reflects the interspersion of V, J, and C gene segments that characterizes the mouse lambda family, and whether differences between the gene families in the two species can shed any light on the differential expression of lambda light chains in humans and mice.

One of the unlinked C_λ genes differs from the active genes in an interesting way (Hollis et al., 1982). It appears to be a pseudogene, as a result of nonsense and frameshift mutations which would block the expression of a C_λ polypeptide region. The C gene is joined to a sequence resembling a J gene segment, and the junction of these joined sequences is reminiscent of the splicing of C and J sequences in an RNA precursor. The homology between this gene and one of the active C_λ genes extends in the 3' direction to a site expected for termination of transcription and addition of a poly(A) tail to the RNA. Interestingly, a very A-rich sequence follows this termination site in the pseudogene. It has been proposed (Hollis et al., 1982) that these structural features reflect a series of events beginning with the transcription of a J_λ-C_λ region, followed by "processing" of the resulting RNA; i.e., splicing of the J_λ and C_λ sequences, and addition of a poly(A) stretch to the 3' end of the transcript. The next steps in the scenario proposed to account for the generation of this pseudogene are: the integration of the transcript into the RNA genome of a retrovirus, conversion of the RNA to DNA by viral enzymes, and integration of the resulting sequence into a human chromosome. The presence of short direct repeated sequences flanking the pseudogene (reminiscent of similar sequences in retroviruses) and its location on a different chromosome than the major C_λ locus are consistent with the proposed model for the formation of such a "processed" gene.

It is noteworthy that sequences resembling a V gene are not found adjacent to the presumed J residues in the "processed" C_λ pseudogene. According to the model this may reflect aberrant initiation of transcription just upstream of the J_λ segment, followed by normal J-C RNA splicing and poly(A) addition. Such an aberrant transcript would be generated in germ cells, which of course do not express antibodies. Transcription of antibody gene segments in non-Ig-producing

cells has been documented in a number of cases (Kemp et al., 1980a,b; Zúñiga et al., 1982).

HUMAN C_H GENES

Most, if not all, of the human C_H genes have recently been studied, at least preliminarily, at the molecular level. The C_μ gene and associated flanking sequences have been analyzed by several groups (Takahashi et al., 1980; Rabbitts et al., 1981; Ravetch et al., 1981). The C_μ coding region is comprised of four exons corresponding to duplication products of the primordial Ig domain. Several kilobases 5' of the C_μ gene is a cluster of nine J_H gene segments (Fig. 5). Three of these are pseudogenes as indicated by the absence of a particular sequence at their 3' ends which is necessary for correct RNA splicing to the downstream C_H sequences (Ravetch et al., 1981). The human J_H family is larger than that of the mouse, an observation that was previously noted for J_κ genes. Thus events related to gene duplications have had different consequences in the two species with regard to the composition of J gene clusters. At the 5' ends of the active human J_H gene segments are the recognition signals necessary for variable region gene formation. Interestingly, the pseudo-J segments appear to have retained these sequences, suggesting that these nonfunctional J segments may be able to participate in V-gene formation. Unlike the case in mice, a single D gene segment is found within the human J_H cluster (Ravetch et al., 1981). This is not closely linked to the tandemly arranged family of D segments described earlier. It will be interesting to see where other human D segments reside, and whether the D segment within the J_H cluster is actually used to construct heavy chains.

As expected by analogy to the mouse, gene segments encoding peptide stretches which anchor the μ heavy chain in the lymphocyte membrane are found 3' of the

C_μ domain exons, and a few kilobases further downstream is the C_δ gene (Rabbitts et al., 1981; Ravetch et al., 1981). The expression of the C_μ membrane exons and C_δ gene presumably involves the RNA splicing patterns shown in Fig. 3. Use of such a splicing strategy has recently been rather directly demonstrated for the production of surface IgG in human cells (Cushley et al., 1982). The region between the J_H cluster and the C_μ gene contains the sequences involved in the class switching event in which a downstream C_H gene displaces the C_μ gene from its position near the rearranged V gene. This switch region has localized sequence homology to the corresponding area near the mouse C_μ gene (Ravetch et al., 1980; Takahashi et al., 1980). DNA sequence homology in these switch regions have also been observed for human and mouse γ (Takahashi et al., 1982), α (Ravetch et al., 1980), and ε (Nishida et al., 1982) constant region loci.

Gamma heavy chains are shorter than μ polypeptides, reflecting different patterns of evolution of C_μ as opposed to C_γ genes. While C_μ regions are encoded in four domain-sized exons, C_γ genes are composed of three of the larger exons and a separate coding block for the hinge segment of the polypeptide, which lies between the C_H1 and C_H2 domains (Honjo et al., 1979; Liu et al., 1980). Protein and DNA sequence comparisons indicate that the C_H1, C_H2, and C_H3 domains of C_γ regions are most closely related to, respectively, the C_H1, C_H3, and C_H4 domains of μ chains (Barker et al., 1980). No C_γ counterpart to the $C_\mu2$ domain is found. This is explained by the model shown in Fig. 6. Following duplication of the original four-domain C_H gene to give rise to the ancestors of the major C_H classes, the C_H2 domain of the primordial C_γ gene degenerated so that it no longer was used to encode protein. This occurred in conjunction with the utilization of new RNA splice sites flanking small coding regions for the hinge segments. The distances between exons in C_γ genes, and the existence of sequence homology

between the C_H1-hinge intron and the other C_γ exons (Tucker et al., 1979) are consistent with the scheme shown in Fig. 6.

In addition to the five major classes of antibodies, subclasses of IgG and IgA exist in humans. The molecules within these groups are very closely related in structure although they show differences in serum concentrations and some of the biological effector functions they mediate. These subclasses are the manifestations of multiple germline C_γ and C_α genes, which are the products of duplications of ancestral genes. The C_γ genes of both humans and mice have been rather well characterized. Both species express four subclasses of IgG, but the relationships among the members of the respective C_γ gene families are quite different. The C_γ genes of humans are over 90% homologous (Ellison and Hood, 1982; Ellison et al., 1982), while mouse C_γ genes are much less similar (about 70-80% homologous [Miyata et al., 1980; Yamawaki-Kataoka et al., 1981]). One interpretation of these observations is that the duplication event(s) yielding the human C_γ genes occurred more recently than the corresponding events in murine evolution. This view implies that the individual C_γ genes evolve independently of one another, and that the degree of similarity between genes reflects the time elapsed since the duplication event. Another possibility is posed by the analysis of the mouse C_γ gene family, which indicates that these genes have exchanged genetic information during their evolution (Miyata et al., 1980; Ollo et al., 1981; Yamawaki-Kataoka et al., 1981). The human C_γ subclasses are most clearly distinguished by their hinge regions, which are far more divergent than the C_H domains. Molecular analysis of these DNA segments has recently shown that genetic exchanges between nonallelic C_γ genes have been an important part of the evolutionary history of the human C_γ family. The $\gamma 3$ polypeptides have a unique type of hinge which is about four times larger than that found in the other subclasses, and includes three nearly identical repeating units (Michaelson et al., 1977). These repeating

units are related to the γ1 hinge, while the remaining γ3 hinge exon resembles the hinge of a fifth human C_γ gene that is probably a pseudogene (Takahashi et al., 1982). Thus it appears that the $C_{\gamma 3}$ gene was assembled from the ancestral $C_{\gamma 1}$ and pseudo-C_γ genes, according to the scheme shown in Fig. 7. Duplication of the γ1-like hinge exon followed to result in the four-hinge exons that $C_{\gamma 3}$ genes now have. The model shown in Fig. 7 predicts a specific organization of the $C_{\gamma 1}$, $C_{\gamma 3}$, and pseudo-C_γ genes. Thus far, however, the only linkage relationship established for the human C_γ genes is the linkage of $C_{\gamma 2}$ to $C_{\gamma 4}$, with the $C_{\gamma 2}$ gene lying on the upstream (5') side (Ellison and Hood, 1982; Krawinkel and Rabbitts, 1982; Takahashi et al., 1982).

Although subclasses of IgE have not been inferred from protein studies, it has recently been shown that humans have three nonallelic C_ε genes (Flanagan and Rabbitts, 1982; Max et al., 1982; Nishida et al., 1982). One is apparently a "processed" pseudogene analogous to the $J_\lambda C_\lambda$ structure described earlier (Max et al., 1982). Interestingly, both of the other C_ε genes appear to lie adjacent to C_α genes, and thus C_ε and C_α genes may have duplicated as a unit. Only one of these C_ε genes is functional; the other is a pseudogene containing only about half the C_ε coding sequences. Although the two C_α genes have not been studied in detail, the existence of two IgA subclasses in humans would predict that they both are functional genes.

HOMOLOGOUS NONALLELIC RECOMBINATION BETWEEN IMMUNOGLOBULIN GENE SEGMENTS

We have noted that humans and mice have different numbers of germline elements for the three antibody gene families. Thus different gene duplications and/or deletions have occurred in the two species since their divergence approximately 80 million years ago (Romero-Harrera et al., 1973). Other observations follow

this trend of variation in mammalian antibody gene repertoires: rabbits have multiple C_K genes (Heidmann and Rougeon, 1982), rats have more J_K segments than mice (Sheppard and Gutman, 1982), and different numbers of IgG subclasses are found in different mammals (Nisonoff et al., 1975). The most popular model for how these various gene duplications occur is homologous unequal crossover. In this process the strand breaks occur in different places in the two recombining regions; schematically this is most easily represented by having the participating species in a misaligned configuration, as shown in Fig. 8A. This misalignment is presumably promoted by DNA sequence homology between noncorresponding (i.e., nonallelic) parts of the chromosomal segments. The products of an unequal crossover are an expanded chromosome and a contracted chromosome. When the strand breaks occur in intergenic regions, the result is a gene duplication on one chromosome and a gene deletion on the other. Thus it is easy to see how unequal crossovers can facilitate either expansion or contraction of a tandemly arranged gene family, depending on which products are fixed by natural selection. Recombination events of this type are envisioned to have been involved in many of the duplications of immunoglobulin gene segments throughout evolution, beginning with the primordial antibody domain segment. Among the most recent duplication products are (1) the families of C_H subclass genes; (2) J_H, J_K, and D clusters; and (3) subfamilies of germline V genes. It is important to note that since gene deletion accompanies duplication in an unequal crossover, a consideration of the varying numbers of germline elements in different mammals does not always allow one to deduce in which species the event(s) took place.

A slightly different consequence of an unequal crossover event comes about when the breakpoints are within, rather than between, structural genes. In this case hybrid genes are created and, as shown in Fig. 8B, the expanded and contracted

chromosomes exhibit either hybrid orientation, respectively. The hybrid nature of the human $C_{\gamma 3}$ hinge suggest that it might be the product of such an intragenic unequal crossover. This inference is further supported by the fact that the ancestral components of the $\gamma 3$ hinge (those of the $\gamma 1$ and pseudo-γ genes) are also present (Takahashi et al., 1982); thus the expanded chromosome was apparently fixed in the human population. Recombination breakpoints within mouse C_{γ} genes have also been inferred (Miyata et al., 1980; Ollo et al., 1981; Yamawaki-Kataoka et al., 1981). Nucleotide sequence analyses have localized the breakpoints in both human and mouse C_{γ} genes to intron regions. Breakage-and-reunion events in these areas allows new combinations of exons to be assembled without the possibility of a shift in translational reading frame, as might happen if the strand breaks occurred within coding sequences. Thus intron breakpoints promise greater chances of survival, which probably accounts for their appearance in the above cases. Miyata et al. (1980) have termed this phenomenon "intervening sequence-mediated domain transfer."

Another type of genetic interaction between nonallelic DNA segments is gene conversion, which is shown schematically in Figs. 8C and D for an event involving two tandem homologous genes. This event is by definition a unidirectional transfer of genetic information. It has in common with the unequal crossovers shown in Figs. 8A and B the characteristic of the replacement (on at least one chromosome) of the sequence of all or part of a given gene with the sequence of its homologous neighbor. This matching or replacement of sequences has been termed "gene correction" (Hood et al., 1975). One product of a gene conversion (the "corrected" chromosome) is formally indistinguishable from the product of a double unequal crossover event, thus further emphasizing the fundamental similarity of the two processes. It seems reasonable to view gene conversion and unequal crossover as two different versions of the same fundamental genetic process,

which we shall refer to as "homologous nonallelic recombination." Manifestations of this process can include both gene duplications and gene corrections. A molecular earmark of homologous nonallelic recombination is depicted in Fig. 9; a comparison of the sequences of two homologous nonallelic genes shows that a boundary exists which separates regions having different levels of sequence homology. The areas showing greater homology are those involved in the most recent recombination event, and the boundary represents the presumed recombination breakpoint. If the recombination has occurred recently enough, we observe a hybrid gene. Several examples of hybrid human IgG molecules have been characterized which appear to have resulted from recombinations between nonallelic C_γ genes (Kunkel et al., 1969; Natvig and Kunkel, 1974; Arnaud et al., 1981). In addition an IgA molecule has been described which has an α chain that resembles a hybrid product of the $C_{\alpha 1}$ and $C_{\alpha 2}$ genes (Tsuzukida et al., 1979). Molecular analyses of what resides on the chromosomes along with the presumed hybrid genes would probably tell us whether unequal crossover or gene conversion events are responsible for the hybrid molecules. Other outcomes of this class of events include cases of apparent duplications and deletions of C_γ genes which have been reported (van Loghem et al., 1970, 1980; Lefranc et al., 1976, 1979). Thus homologous nonallelic recombination between tandem gene arrays appears to be an important genetic mechanism operating on human immunoglobulin gene families.

SEQUENCE CONSERVATION AND DIVERGENCE OF IMMUNOGLOBULIN GENES

The nucleotide sequence homology between human and mouse Ig genes [on the order of 65-85% in most cases (Bentley and Rabbitts, 1980; Hieter et al., 1980, 1981a; Matthyssens and Rabbitts, 1980; Takahashi et al., 1980; Dolby et al., 1981; Ellison et al., 1981)] has facilitated the isolation of most of the cloned human genes that have been studied. This sequence conservation between species of

course reflects selective pressure to maintain functionally important structures. These structures can be classified into two categories: (1) features of immunoglobulin molecules relating to their biological activities; and (2) the noncoding nucleotide sequences involved in the rearrangement and expression of antibody genes. Particularly strong homology is seen in these latter structures: the sequences flanking the 3' region of germline V region components are nearly identical in humans and mice (Tonegawa et al., 1978; Max et al., 1979; Bentley and Rabbitts, 1980; Matthyssens and Rabbitts, 1980), and the tandemly repeated oligonucleotides comprising the C_H switch regions show a great deal of similarity (Rabbitts et al., 1981). These conserved structures are implicated as the binding and/or catalytic sites of the enzymes carrying out the corresponding somatic DNA rearrangements. Interestingly, the switch sequences 5' of human and mouse C_ε genes show more homology than the coding sequences (Nishida et al., 1982).

We might expect that we could deduce the timing during evolution of many of the gene duplications of antibody genes, based on homology relationships both between human and mice genes and within the various human gene families and subfamilies. However, the capacity of homologous nonallelic recombination to mediate gene duplication, deletion, and correction events makes such deductions difficult, if not impossible. To illustrate this: it is not clear whether the extremely similar human C_γ genes are the products of very recent de novo gene duplications, or whether gene correction events acted on the human C_γ family to homogenize the genes about 6-10 million years ago [this is the estimated divergence time based on the sequence differences] (Ellison and Hood, 1982; Ellison et al., 1982; Takahashi et al., 1982). Pertinent to this issue is the inference, based on preliminary hybridization experiments with a human C_γ probe, of multiple C_γ genes in various nonhuman primates (B. Chapman and J. Ellison, unpublished). Further molecular

analysis of primate C_γ genes may shed some light on the evolutionary past of the human C_γ gene family, perhaps by revealing which gene duplication(s) occurred before and after the divergence of humans from other primates.

An intriguing feature of the human C_γ genes is the extensive divergence found among the various hinge exons. These short (36-51 nucleotide) segments show both length and sequence heterogeneity, while the C_H domain exons as well as intervening DNA regions are about 95% identical among the C_γ subclass genes (Ellison et al., 1982; Ellison and Hood, 1982; Krawinkel and Rabbitts, 1982; Takahashi et al., 1982). It is possible to imagine that this disparity exists because gene correction events have homogenized the C_γ genes, but hinge exons have been excluded from these events. This view would predict the existence of remnants of recombination breakpoints surrounding the hinge exons (of the type indicated in Fig. 9). However, the high level of homology four among C_γ genes extends in the introns right up to the hinge exon splice junctions; no other distinct boundary is found separating areas of different homologies (Ellison and Hood, 1982; Ellison et al., 1982). Since the high degree of divergence exists exclusively in coding sequences, we favor the idea that structural diversity in hinge regions of IgG molecules is the result of natural selection. The nature of the selective advantage to having different hinge structures is not immediately obvious. One possibility that has been suggested is that divergent hinge regions may in part be responsible for the variations in effector functions carried out by the IgG subclasses (Cebra et al., 1977; Wang et al., 1980; Klein et al., 1981). According to this view the generation of new and diverse effector functions for IgG molecules may be the selective force which fixes nucleotide substitutions in hinge exons. Just as it is beneficial to have diverse binding site structures for a broad range of antigenic determinants, so may it be advantageous to have multiple strategies for handling the foreign bodies once they are recognized.

ANTIBODY GENE EXPRESSION IN RELATION TO LYMPHOCYTE DIFFERENTIATION

A description of the sequence of events pertaining to antibody gene expression during B-lymphocyte ontogeny is still being worked out. Our current picture (see Fig. 10) is that Ig gene activity begins in the progenitors of B cells (pre-B cells) with a V-D-J joining event which brings a functional V_H gene close to C_μ. This rearranged locus becomes transcriptionally active, and the production of cytoplasmic μ heavy chains follows (Levitt and Cooper, 1980; Maki et al., 1980). Light chain gene rearrangement follows. Analyses of various human cell lines has shown that in lambda-producing cells both allelic C_κ genes are either deleted or have undergone nonproductive V gene rearrangements which present their expression, while cells making kappa chains have both their C_λ genes in the germline (unrearranged) configuration (Hieter et al., 1981b). These observations are consistent with the interpretation that the enzymes involved in V-gene formation choose their substrates in the order heavy, followed by kappa, and then lambda. The implication is that these DNA rearrangements are programmed facets of B lymphocyte differentiation. The production of light chains in conjunction with μ chains leads to the appearance of IgM molecules on the surfaces of B cells. At a later stage of differentiation, IgD and IgM molecules are often expressed simultaneously by the same cell (Knapp et al., 1973; Rowe et al., 1973a,b), reflecting the alternative splicing of C_μ and C_δ sequences to the rearranged V_H transcription products. As we noted earlier, this alternative splicing strategy may also be employed to result in the expression of other C_H genes downstream of C_μ. The most typical way of expressing other Ig classes, however, is probably through the class switch recombination event, which follows the above processes along the pathway of B-cell differentiation.

The general features of the somatic genetic processes leading to the production of immunoglobulin molecules have been deduced largely on the basis of structural analyses of the genes themselves. This approach of course has its limitations, for we know little if anything of the detailed molecular mechanisms or the enzymes mediating them. Consequently many unanswered questions remain relating to the nature of the regulatory signals which trigger these events. For example: What regulates the alternative production of either membrane-bound or secreted forms of immunoglobulins? How do T cells exert their influence on the differentiation of B cells? What determines the levels of expression of the different subclasses of IgG, and why are the two IgA subclasses distributed differently in serum and secretory fluids? How do B cells become committed to expression of a given Ig isotype? These and many other questions are analogous to those confronted by the many developmental and molecular biologists working in other systems.

Certain immunodeficiency disorders may be of help in understanding key stages in B-cell differentiation. One such disorder, dysgammaglobulinemia, is characterized by elevated levels of serum IgM (and sometimes IgD) and absence of other Ig classes. Lymphocytes in the afflicted individuals are thus apparently blocked at a critical stage of differentiation: they are unable to undergo class switching. Although this failure could be envisioned to result from a variety of physiologic factors, it was recently shown that at least one defect was at the level of the B cell itself (Schwaber et al., 1981). This was demonstrated by establishing permanent B cell lines from patients with the disorder, and showing that such cell lines exhibited a pattern of Ig expression that closely resembled that seen in vivo (i.e., production only of IgM and IgD). These cell lines were generated by infecting primary B lymphocytes with Epstein-Barr virus (EBV). Human and other primate B lymphocytes are easily transformed by this virus, allowing them

to be maintained indefinitely in culture (Miller, 1971). By providing such cell lines, EBV transformation is a valuable tool for investigating B cell function.

One promising source of information concerning the expression of antibody genes are the abnormal lymphoid cells characteristic of certain disease states. Leukemic lymphocytes were used to delineate the hierarchy of light chain gene rearrangements in human pre-B cells (Korsmeyer et al., 1981). Those analyses were feasible because the malignant cells represented monoclonal populations which were apparently arrested at discrete stages of differentiation. Analogous characterizations of aberrant cells in such disorders as selective IgA deficiency (Goldberg et al., 1969) and irregular IgG subclass expression (Terry, 1968; Yount et al., 1970) may soon be possible; the cloning and long-term culturing of human B lymphocytes (Sredni et al., 1981) should aid in studies of this kind. Molecular genetic analysis of one type of lymphoproliferative disorder, heavy chain disease (HCD), has already begun. A permanent HCD cell line has been established (Buxbaum et al., 1978) that synthesizes a $\gamma 3$ heavy chain having an internal deletion; interestingly, no light chains are produced. Complementary DNA (cDNA) clones have been constructed and characterized that encode the HCD protein (Alexander et al., 1982), and analysis of the expressed genomic $C_{\gamma 3}$ gene should soon reveal the basis for the production of the abnormal $\gamma 3$ heavy chain. Perhaps more importantly, this cell line may be helpful in understanding the regulation of light chain gene expression.

SOME FUTURE PROSPECTS

We have noted how a few clinical disorders related to antibody-producing cells may soon be better understood, at least with regard to the programmed rearrangements of Ig genes that normally occur. Our knowledge of the structural organization of antibody genes and the genetic processes taking place in germ cells should

also help in interpreting certain hereditary defects in immunoglobulin production. Another possible practical application of our current knowledge involves the use of both recombinant DNA and hybridoma technologies. Specific segments of human (and even mouse) Ig genes can be grafted together, and the resulting product then introduced into cells so the gene is expressed. The fusion of these cells with myeloma cells would produce a hybridoma which would be the source of a human immunoglobulin of any desired class and binding specificity. The recent establishment of a human myeloma cell line (Karpas et al., 1982) makes the production of human monoclonal antibodies a possibility in the very near future. Such preparations would be useful in both diagnostic and therapeutic applications.

The three unlinked families of human antibody genes—kappa, lambda, and heavy—have recently been assigned to specific human chromosomes, through the techniques of somatic cell genetics and/or in situ hybridization to chromosome spreads. Kappa variable (Malcolm et al., 1982) and constant region (McBride et al., 1982) genes have been mapped to chromosome 2, lambda constant region genes to chromosome 22 (McBride et al., 1982), and heavy chain genes to chromosome 14 (Croce et al., 1979; Hobart et al., 1981). These locations are of considerable interest because of the occurrence of specific translocations involving these chromosomes in certain malignancies of B-cell origin. For example, most Burkitt's lymphoma cells undergo a highly specific reciprocal 8;14 translocation (Zech et al., 1976; Kaiser-McCaw, 1977; Berger et al., 1979a), and intriguingly, the breakpoint on chromosome 14 is in the same region (q32) to which the immunoglobulin C_H genes have been mapped (Kirsch et al., 1982). Other Burkitt's lymphomas and some B-cell leukemias show translocations involving the light chain chromosomes 2 and 22 (Miyoshi et al., 1979, 1981; Van Den Berghe, 1979; Berger et al., 1979b; Rowley et al., 1981). In virtually all these cases the other chromosome involved

in the translocation is chromosome 8 (reviewed by Rowley, 1982). Another example is the well-known "Philadelphia" chromosome, an aberrant chromosome 22, which is frequently seen in chronic myelogenous leukemia (Whang-Peng et al., 1970; Rowley et al., 1978, 1980). The participation in these translocations of chromosomes which are involved in somatic DNA rearrangements raises the interesting possibility that the enzymatic machinery working on antibody genes may be directly or indirectly connected to the translocation events. The relationship between the chromosomal events themselves and the neoplasm is not clear, although others have speculated that the movement of certain cellular genes into actively expressed areas (Ig gene loci) may cause these genes to be expressed in such a fashion as to induce or maintain the malignant state (Klein, 1981; Rowley, 1982). In this regard it will be interesting to determine what genes reside on chromosome 8. Probes from antibody gene regions will no doubt be used in the isolation of these sequences.

This review of the current knowledge of human antibody genes has pointed out the many expected structural similarities to the well-studied genes of the mouse. As studies in both species progress, we may expect that observations of certain evolutionarily-conserved features of Ig gene loci will aid in directing research toward crucial aspects of antibody gene expression and lymphocyte differentiation. As we have discussed above, the study of human genes also provides opportunities to better understand genetic mechanisms operating during evolution. In addition, cloned antibody DNA fragments can be used for finding DNA polymorphisms to serve as genetic markers in family and population studies. Finally, the potential clinical benefits that exist from studies of human Ig genes clearly justifies continued work in this area.

ACKNOWLEDGMENTS. We thank Bernita Larsh and Connie Katz for the preparation of the manuscript, and Ellen Kraig and Roger Perlmutter for their comments. This work was supported by a grant from the National Institutes of Health and by a National Research Award (1 T32 GM07616) to J.E. from the National Institute of General Medical Sciences.

REFERENCES

Adams, J., 1980, The organization and expression of immunoglobulin genes, Immunol. Today, July, 10-17.

Alexander, A., Steinmetz, M., Barritault, D., Frangione, B., Franklin, E. C., Hood, L., and Buxbaum, J. N., 1982, γ heavy chain disease in man: cDNA sequence supports partial gene deletion model, Proc. Natl. Acad. Sci. USA **79**:3260-3264.

Alt, F. W., Rosenberg, N., Casanova, R. J., Thomas, E., and Baltimore, D., 1982, Immunoglobulin heavy-chain expression and class switching in a murine leukemia cell line, Nature **296**:325-331.

Altenburger, W., Steinmetz, M., and Zachau, H. G., 1980, Functional and non-functional joining in immunoglobulin light chain genes of a mouse myeloma, Nature **287**:603-607.

Arnaud, P., Wang, A.-C., Gianazza, E., Wang, I. Y., Lasne, Y., Creyssel, R., and Fudenberg, H. H., 1981, Gamma heavy chain disease protein CHA: immunological and structural studies, Mol. Immunol. **18**:379-384.

Barker, W. C., Ketcham, L. K., and Dayhoff, M. O., 1980, Origins of immunoglobulin heavy chain domains, J. Mol. Evol. **15**:113-127.

Barstad, P., Hubert, J., Hunkapiller, M., Goetze, A., Schilling, J., Black, B., Eaton, B., Richards, J., Weigert, M., and Hood, L., 1978, Immunoglobulins with hapten-binding activity: structure-function correlations and genetic implications, Eur. J. Immunol. **8**:497-503.

Bentley, D. L., and Rabbitts, T. H., 1980, Human immunoglobulin variable region genes—DNA sequences of two V_K genes and a pseudogene, Nature **288**:730-733.

Bentley, D. L., and Rabbitts, T. H., 1981, Human V_K immunoglobulin gene number: implications for the origin of antibody diversity, Cell **24**:613-623.

Berger, R., Bernheim, A., Brouet, J. C., Daniel, M. T., and Flandrin, G., 1979a, t(8;14) Translocation in a Burkitt's type of lymphoblastic leukemia (L3), Br. J. Haematol. **43**:87-90.

Berger, R., Bernheim, A., Weh, H. J., Flandrin, G., Daniel, M. T., Brouet, J. C., and Colbert, N., 1979b, A new translocation in Burkitt's tumor cells, Hum. Genet. **53**:111-112.

Bernard, O., Hozami, N., and Tonegawa, S., 1978, Sequences of mouse immunoglobulin light chain genes before and after somatic changes, Cell **15**:1133-1144.

Blomberg, B., Traunecker, A., Eisen, H., and Tonegawa, S., 1981, Organization of four mouse λ light chain immunoglobulin genes, Proc. Natl. Acad. Sci. USA **78**:3765-3769.

Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K., and Baltimore, D., 1981, Heavy chain variable region contribution to the NP^b family of antibodies: somatic mutation evident in a γ2a variable region, Cell **24**:625-638.

Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S., 1978, A complete immunoglobulin gene is created by somatic recombination, Cell **15**: 1-14.

Buxbaum, J. N., Alexander, A., and Olivier, O., 1978, Gamma heavy chain disease in man: synthesis of a deleted γ3 immunoglobulin by lymphoid cells in short and long term tissue culture, Clin. exp. Immunol. **32**:489-497.

Cebra, J. J., Brunhouse, R. F., Cordle, C. T., Daiss, J., Fecheimer, M., Ricardo, M., Thunberg, A., and Wolfe, P. B., 1977, Isotypes of guinea pig antibodies: restricted expression and basis of interactions with other molecules, Prog. Immunol. **3**:264-277.

Choi, E., Kuehl, M., and Wall, R., 1980, RNA splicing generates a variant light chain from an aberrantly rearranged κ gene, Nature **286**:776-779.

Crews, S., Griffin, J., Huang, H., Calame, K., and Hood, L., 1981, A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody, Cell **25**:59-66.

Croce, C. M., Shander, M., Martinis, J., Cicurel, L., D'Aucona, G. G., Dolby, T. W., and Koprowski, H., 1979, Chromosomal location of the genes for human immunoglobulin heavy chains, Proc. Natl. Acad. Sci. USA **76**:3416-3419.

Cushley, W., Coupar, B. E. H., Mickelson, C. A., and Williamson, A. R., 1982, A common mechanism for the synthesis of membrane and secreted immunoglobulin α, γ, and μ chains, Nature **298**:77-79.

Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L., and Hood, L., 1980, An immunoglobulin heavy-chain gene is formed by at least two recombinational events, Nature **283**:733-739.

Dolby, T. W., Devuono, J., and Croce, C. M., 1980, Cloning and partial nucleotide sequence of human immunoglobulin μ chain cDNA from B cells and mouse-human hybridomas, Proc. Natl. Acad. Sci. USA **77**:6027-6031.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L., 1980a, An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H, Cell **19**:981-992.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L., 1980b, Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways, Cell **20**:313-319.

Early, P., and Hood, L., 1981, Mouse Immunoglobulin Genes, in: Genetic Engineering (J. K. Setlow and A. Hollaender, eds.), pp. 157-188, Plenum, New York.

Ein, P., 1968, Nonallelic behavior of the Oz groups in human lambda immunoglobulin chains, Proc. Natl. Acad. Sci. USA **60**:982-985.

Ellison, J., Buxbaum, J., and Hood, L., 1981, Nucleotide sequence of a human immunoglobulin $C_{\gamma 4}$ gene, DNA **1**:11-18.

Ellison, J., and Hood, L., 1982, Linkage and sequence homology of two human immunoglobulin γ heavy chain constant region genes, Proc. Natl. Acad. Sci. USA **79**:1984-1988.

Ellison, J. W., Berson, B. J., and Hood, L. E., 1982, The nucleotide sequence of a human immunoglobulin $C_{\gamma 1}$ gene, Nucleic Acids Res. **10**:4071-4079.

Fett, J., and Deutsch, H. F., 1975, A new λ-chain gene, Immunochemistry **12**:643-652.

Flanagan, J. G., and Rabbitts, T. H., 1982, The sequence of a human immunoglobulin epsilon heavy chain constant region gene, and evidence for three non-allelic genes, EMBO J. **1**:655-660.

Gibson, D., Levannon, M., and Smithies, O., 1971, Heterogeneity of normal human immunoglobulin light chains. Nonallelic variation in the constant region of λ chains, Biochemistry **10**:3114-3122.

Goldberg, L. S., Douglas, S. S., and Fudenberg, H. H., 1969, Studies on salivary γA in agammaglobulinemia, Clin. exp. Immunol. **4**:579-584.

Heidmann, O., and Rougeon, F., 1982, Multiple sequences related to a constant-region kappa light chain gene in the rabbit genome, Cell **28**:507-513.

Hieter, P. A., Max, E. E., Seidman, J. G., Maizel, J. V., Jr., and Leder, P., 1980, Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments, Cell **22**:197-207.

Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A., and Leder, P., 1981a, Human immunoglobulin κ light-chain genes are deleted or rearranged in λ-producing B cells, Nature **290**:368-372.

Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A., and Leder, P., 1981b, Clustered arrangement of immunoglobulin λ constant region genes in man, Nature **294**:536-540.

Hieter, P. A., Maizel, J. V., Jr., and Leder, P., 1982, Evolution of human immunoglobulin κ J region genes, J. Biol. Chem. **257**:1516-1522.

Hobart, M. J., Rabbitts, T. H., Goodfellow, P. N., Solomon, E., Chambers, S., Spurr, N., and Povey, S., 1981, Immunoglobulin heavy chain genes in humans are located on chromosome 14, Ann. Hum. Genet. **45**:331-335.

Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D., and Leder, P., 1982, Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing, Nature **296**:321-325.

Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N., and Mano, Y., 1979, Cloning and complete nucleotide sequence of mouse immunoglobulin γ1 chain gene, Cell **18**, 559-568.

Hood, L., Campbell, J. H., and Elgin, S. C. R., 1975, The organization, expression, and evolution of antibody genes and other multigene families, Ann. Rev. Genet. **9**:305-353.

Hood, L., Davis, M., Early, P., Calame, K., Kim, S., Crews, S., and Huang, H., 1981, Two types of DNA rearrangements in immunoglobulin genes, Cold Spring Harbor Symp. Quant. Biol. **45**:887-897.

Huang, H., Crews, S., and Hood, L., 1981, The arrangement, rearrangement, and diversification of antibody genes, in: Frontiers in Immunogenetics (W. H. Hildemann, ed.), pp. 63-74, Elsevier North Holland, Inc., New York.

Kaiser-McCaw, B., Epstien, A. L., Kpalan, H. L., and Hecht, F., 1977, Chromosome 14 translocation in African and North American Burkitt's lymphoma, Int. J. Cancer **19**:482-486.

Karpas, A., Fischer, P., and Swirsky, D., 1982, Human myeloma cell line carrying a Philadelphia chromosome, Science **216**:997-999.

Kataoka, T., Kawakami, T., Takahashi, N., and Honjo, T., 1980, Rearrangement of immunoglobulin γ1-chain gene and mechanism for heavy-chain class switch, Proc. Natl. Acad. Sci. USA **77**:919-923.

Kemp, D. J., Cory, S., and Adams, J. M., 1979, Cloned pairs of variable region genes for immunoglobulin heavy chains isolated from a clone library of the entire mouse genome, Proc. Natl. Acad. Sci. USA **76**:4627-4631.

Kemp, D. J., Wilson, A., Harris, A. W., and Shortman, K., 1980a, The immunoglobulin μ constant region gene is expressed in mouse thymocytes, Nature **286**, 168-170.

Kemp, D. J., Harris, A. W., Cory, S., and Adams, J. M., 1980b, Expression of the immunoglobulin C_{μ} gene in T and B lymphoid and myeloid cell lines, Proc. Natl. Acad. Sci. USA **77**:2876-2880.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L., 1981, Antibody diversity: somatic hypermutation of rearranged V_H genes, Cell **27**:573-581.

Kirsch, I. R., Morton, C. C., Nakahara, K., and Leder, P., 1982, Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes, Science **216**:301-303.

Klein, G., 1981, The role of gene dosage and genetic transpositions in carcinogenesis, Nature **294**:313-318.

Klein, M., Haeffner-Cavaillon, N., Isenman, D. E., Rivat, C., Navia, M. A., Davies, D. R., and Dorrington, K. J., 1981, Expression of biological effector functions by immunoglobulin G molecules lacking the hinge region, Proc. Natl. Acad. Sci. USA **78**:524-528.

Knapp, W., Bolhuis, R. L. J., Radl, J., and Hijmans, W., 1973, Independent movement of IgD and IgM molecules on the surface of individual lymphocytes, J. Immunol. **111**:1295-1298.

Konkel, D. A., Maizel, J. V., Jr., and Leder, P., 1979, The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes, Cell **18**:865-873.

Korsmeyer, S. J., Hieter, P. A., Raretch, J. V., Poplack, D. G., Waldmann, T. A., and Leder, P., 1981, Developmental hierarchy of immunoglobulin gene rearrangements in human leukemic pre-B-cells, Proc. Natl. Acad. Sci. USA **78**:7096-7100.

Krawinkel, U., and Rabbitts, T. H., 1982, Comparison of the hinge coding segments in human immunoglobulin gamma heavy chain genes and the linkage of the gamma 2 and gamma 4 subclass genes, EMBO J. **1**:403-408.

Kunkel, H. G., Natvig, J. B., and Joslin, F. G., 1969, A "Lepore" type of hybrid γ globulin, Proc. Natl. Acad. Sci. USA **62**:144-149.

Kurosawa, Y., and Tonegawa, S., 1982, Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments, J. Exp. Med. **155**:201-218.

Larhammar, D., Gustaffson, K., Claesson, L., Bill, P., Wiman, K., Schanning, L., Sundelin, J., Widmark, E., Peterson, P. A., and Rask, L., 1982, Alpha chain of HLA-DR transplantation antigens is a member of the same protein superfamily as the immunoglobulins, Cell **30**:153-161.

Lefranc, G., Dumitresco, S.-M., Salier, J.-P., Rivat, L., de Lange, G., van Loghem, E., and Loiselet, J., 1979, Familial lack of the IgG3 subclass. Gene elimination or turning off expression and neutral evolution in the immune system, J. Immunogenet. **6**:215-221.

Lefranc, G., Rivat, L., Rivat, C., Loiselet, J., and Roparts, C., 1976, Evidence for "deleted" or "silent" genes homozygous at the locus coding for the constant region of the γ3 chain, Am. J. Hum. Genet. **28**:51-61.

Levin, A. S., Fudenberg, H. H., Hopper, J. E., Wilson, S., and Nisonoff, A., 1971, Immunofluorescent evidence for cellular control of synthesis of variable regions of light and heavy chains of immunoglobulins G and μ by the same gene, Proc. Natl. Acad. Sci. USA **68**:169-171.

Levitt, D., and Cooper, M. D., 1980, Mouse pre-B cells synthesize and secrete μ heavy chains but not light chains, Cell **19**:617-625.

Liebhaber, S. A., Goossens, M., and Kan, Y. W., 1981, Homology and concerted evolution at the α1 and α2 loci of human α-globin, Nature (London) **290**:26-29.

Liu, C.-P., Tucker, P. W., Mushinski, J. F., and Blattner, F., 1980, Mapping of heavy chain genes for mouse immunoglobulins M and D, Science **209**:1348-1353.

Maki, R., Kearney, J., Paige, C., and Tonegawa, S., 1980, Immunoglobulin gene rearrangement in immature B cells, Science **209**:1366-1369.

Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L., and Rabbitts, T. H., 1982, Localization of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by in situ hybridization, Proc. Natl. Acad. Sci. USA, in press.

Marchalonis, J. J., and Cone, R. E., 1973, The phylogenetic emergence of vertebrate immunity, Aust. J. Exper. Biol. Med. Sci. **51**:461-488.

Matthyssens, G., and Rabbitts, T. H., 1980, Structure and multiplicity of genes for the human immunoglobulin heavy chain variable region, Proc. Natl. Acad. Sci. USA **77**:6561-6565.

Max, E. E., Seidman, J. G., and Leder, P., 1979, Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene, Proc. Natl. Acad. Sci. USA **76**:3450-3454.

Max, E. E., Seidman, J. G., Miller, H., and Leder, P., 1980, Variation in the crossover point of kappa immunoglobulin gene V-J recombination: evidence from a cryptic gene, Cell **21**:793-799.

Max, E. E., Battey, J., Ney, R., Kirsch, I. R., and Leder, P., 1982, Duplication and deletion in the human immunoglobulin ε genes, Cell **29**:691-699.

McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C., and Leder, P., 1982, Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes, J. Exp. Med. **155**:1480-1490.

Michaelsen, T. E., Frangione, B., and Fraaklin, E. C., 1977, The amino acid sequence of a human immunoglobulin G3m(g) pFc' fragment, J. Immunol. **119**:558-563.

Miller, G., 1971, Human lymphoblastoid cell lines and Epstein-Barr virus: a review of their interrelationships and their relevance to the etiology of leukoproliferative states in man, Yale J. Biol. Med. **43**:358.

Miller, J., Bothwell, A., and Storb, U., 1981, Physical linkage of the constant region genes for immunoglobulins λ_I and λ_{III}, Proc. Natl. Acad. Sci. USA **78**:3829-3833.

Milstein, C., 1965, Interchain disulfide bridge in Bence-Jones proteins and in γ-globulin B chains, Nature **205**:1171-1173.

Milstein, C., 1967, Linked groups of residues in immunoglobulin κ chains, Nature **216**:330-332.

Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M., and Honjo, T., 1980, Nucleotide sequence divergence of mouse immunoglobulin γ1 and γ2b chain genes and the hypothesis of intervening sequence-mediated domain transfer, Proc. Natl. Acad. Sci. USA **77**:2143-2147.

Miyoshi, I., Hiraki, S., Kimura, I., Miyamoto, K., and Sato, J., 1979, 2/8 translocation in a Japanese Burkitt's lymphoma, Experientia **35**:742-743.

Miyoshi, I., Hamazaki, K., Kubonishi, I., and Yoshimoto, S., 1981, A variant translocation (8;22) in a Japanese patient with Burkitt lymphoma, Gana **72**:176-177.

Moore, K. W., Rogers, J., Hunkapiller, T., Early, P., Nottenberg, C., Weissman, I., Bazin, H., Wall, R., and Hood, L. E., 1981, Expression of IgD may use both DNA rearrangement and RNA splicing mechanisms, Proc. Natl. Acad. Sci. USA **78**:1800-1804.

Natvig, J. B., and Kunkel, H. G., 1974, A hybrid IgG4-IgG2 immunoglobulin, J. Immunol. **112**:1277-1284.

Newell, N., Richards, J. E., Tucker, P. W., and Blattner, F. R., 1980, J genes for heavy chain immunoglobulins of mouse, Science **209**:1128-1132.

Nikaido, T., Nakai, S., and Honjo, T., 1981, Switch region of immunoglobulin C_{μ} gene is composed of simple tandem repetitive sequences, Nature **292**:845-848.

Nishida, Y., Miki, T., Hisajima, H., and Honjo, T., 1982, Cloning of human immunoglobulin chain genes: evidence for multiple C genes, Proc. Natl. Acad. Sci. USA **79**:3833-3837.

Nisonoff, A., Hopper, J. E., and Spring, S. B., 1975, Immunoglobulins of the rabbit, mouse, guinea pig, and horse, in: The Antibody Molecule (F. J. Dixon Jr. and H. G. Kunkel, eds), pp. 314-318, Academic Press, New York.

Obata, M., Kataoka, T., Nakai, S., Yamagishi, H., Takahashi, N., Yamawaki-Kataoka, Y., Nikaido, T., Shimizu, A., and Honjo, T., 1981, Structure of a rearranged $\gamma 1$ chain gene and its implication to immunoglobulin class-switch mechanism, Proc. Natl. Acad. Sci. USA **78**:2437-2441.

Ollo, R., Auffray, C., Morchamps, C., and Rougeon, F., 1981, Comparison of mouse immunoglobulin $\gamma 2a$ and $\gamma 2b$ chain genes suggests that exons can be exchanged between genes in a multigenic family, Proc. Natl. Acad. Sci. USA **78**:2442-2446.

Peterson, D. A., Cunningham, B. A., Berggard, I., and Edelman, G. M., 1972, β_2-microglobulin—a free immunoglobulin domain, Proc. Natl. Acad. Sci. USA **69**:1697-1701.

Potter, M., 1972, Immunoglobulin-producing tumors and myeloma proteins of mice, Physiol. Rev. **52**:631-719.

Rabbitts, T. H., Forster, A., and Milstein, C. P., 1981, Human immunoglobulin heavy chain genes: evolutionary comparisons of the C_{μ}, C , and C_{γ} genes and associated switch sequences, Nucleic Acids Res. **9**:4509-4524.

Raff, M. C., 1976, Development and modulation of B lymphocytes: studies on newly formed B cells and their putative precursors in the hemopoietic tissues of mice, Cold Spring Harbor Symp. Quant. Biol. **41**:159-162.

Ravetch, J. V., Kirsch, I. R., and Leder, P., 1980, Evolutionary approach to the question of immunoglobulin heavy chain switching: evidence from cloned human and mouse genes, Proc. Natl. Acad. Sci. USA **77**:6734-6738.

Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T., and Leder, P., 1981, Structure of the human immunoglobulin μ locus: characterization of embryonic and rearranged J and D genes, Cell **27**:583-591.

Rogers, J., Choi, E., Souza, L., Carter, C., Word, C., Kuehl, M., Eisenberg, D., and Wall, R., 1981, Gene segments encoding transmembrane carboxyl termini of immunoglobulin γ chains, Cell **26**:19-27.

Romero-Herrera, A. E., Lehmann, H., Joysey, K. A., and Friday, A. E., 1973, Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis, Nature **246**:389-395.

Rowe, D. S., Hug, K., Faulk, W. P., McCormick, J. N., and Garber, H., 1973a, IgD on the surface of peripheral blood lymphocytes of human newborn, Nature (New Biol. London) **242**:155- .

Rowe, D. S., Hug, K., Forni, L., and Pernis, B., 1973b, Immunoglobulin D as a lymphocyte receptor, J. Exp. Med. **138**:965-972.

Rowley, J. D., 1978, Chromosomes in leukemia and lymphoma, Semin. Hematol. **15**:301-319.

Rowley, J. D., 1980, Ph^1 positive leukemia, including chronic myelogenous leukaemia, Clinics Haematol. **9**:55-86.

Rowley, J. D., 1981, Nonrandom chromosome changes in human leukemia, in: Genes, Chromosomes, and Neoplasia (F. Arrighi, P. Rao, and E. Stubblefield, eds.), pp. 273-296, Raven Press, New York.

Rowley, J. D., 1982, Identification of the constant chromosome regions involved in human hematologic malignant disease, Science **216**:749-751.

Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S., 1979, Sequences at the somatic recombination sites of immunoglobulin light chain genes, Nature **280**:288-294.

Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S., 1980, Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes, Nature **286**:676-683.

Sakano, H., Kurosawa, Y., Weigert, M., and Tonegawa, S., 1981, Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes, Nature **290**:562-565.

Schnell, H., Steinmetz, M., and Zachau, H. G., 1980, An unusual translocation of immunoglobulin gene segments in variants of the mouse myeloma MPC11, Nature **286**:170-173.

Schwaber, J. F., Lazarus, H., and Rosen, F. S., 1981, IgM-restricted production of immunoglobulin by lymphoid cell lines from patients with immunodeficiency with hyper IgM (dysgammaglobulinemia), Clin. Immunol. Immunopathol. **19**:91-97.

Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D., and Leder, P., 1978, Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis, Proc. Natl. Acad. Sci. USA **75**:3881-3885.

Seidman, J. G., Max, E. E., and Leder, P., 1979, A κ-immunoglobulin gene is formed by site-specific recombination without further somatic mutation, Nature **280**:370-375.

Seidman, J. G., and Leder, P., 1980, A mutant immunoglobulin light chain is formed by aberrant DNA- and RNA-splicing events, Nature **286**:779-783.

Selsing, E., and Storb, U., 1981, Somatic mutation of immunoglobulin light-chain variable-region genes, Cell **25**:47-58.

Sheppard, H. W., and Gutman, G. A., 1982, Rat kappa-chain J-segment genes: two recent gene duplications separate rat and mouse, Cell **29**:121-127.

Shimizu, A., Takahashi, N., Yaoita, Y., and Honjo, T., 1982, Orgnaization of the constant-region gene family of the mouse immunoglobulin heavy chain, Cell **28**:499-506.

Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T., and Leder, P., 1981, Human immunoglobulin D segments encoded in tandem multigenic families, Nature **294**:631-635.

Slightom, J. L., Blechl, A. E., and Smithies, O., 1980, Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes, Cell **21**:627-638.

Solomon, A., 1977, Bence Jones proteins and light chains of immunoglobulins. XVI. Immunochemical recognition of the human lambda light-chain constant-region isotype Mcg, Immunogenetics **5**:525-533.

Sredni, B., Sieckmann, D. G., Kumagai, S., House, S., Green, I., and Paul, W. E., 1981, Long-term culture and cloning of nontransformed human B lymphocytes, J. Exp. Med. **154**:1500-1516.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S., and Hood, L., 1981, Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes, Cell **24**:125-134.

Takahashi, N., Nakai, S., and Honjo, T., 1980, Cloning of human immunoglobulin μ gene and comparison with mouse μ gene, Nucleic Acids Res. **8**:5983-5991.

Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S., and Honjo, T., 1982, Structure of human immunoglobulin gamma genes: implications for evolution of a gene family, Cell **29**:671-679.

Terry, W. D., 1968, Variations in the subclasses of IgG, in: Immunologic Deficiency Diseases in Man, vol. 4, (D. Bergsma, ed), pp. 357-363, The National Foundation, New York.

Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W., 1978, Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain, Proc. Natl. Acad. Sci. USA **75**:1485-1489.

Tsuzukida, Y., Wang, C.-C., and Putnam, F. W., 1979, Structure of the A2m(1) allotype of human IgA—a recombinant molecule, Proc. Natl. Acad. Sci. USA **76**:1104-1108.

Tucker, P. W., Marcu, K. B., Newell, N., Richards, J., and Blattner, R. W., 1979, Sequence of the cloned gene for the constant region of murine γ2b immunoglobulin heavy chain, Science **206**:1303-1306.

Tucker, P. W., Liu, C.-P., Mushinski, J. F., and Blattner, F. R., Mouse immunoglobulin D: messenger RNA and genomic DNA sequences, Science **209**:1353-1360.

Tyler, B. M., Cowman, A. F., Adams, J. M., and Harris, A. W., 1981, Generation of long mRNA for membrane immunoglobulin γ2a chains by differential splicing, Nature **293**:406-408.

Valbuena, O., Marcu, K. B., Weigert, M., and Perry, R. B., 1978, Multiplicity of germline genes specifying a group of related mouse κ-chains with implications for the generation of immunoglobulin diversity, Nature **276**:780-784.

van den Berghe, H., Parloir, C., Gosseye, S., Englebienne, V., Cornu, G., and Sokal, G., 1979, Variant translocation in Burkitt-lymphoma, Cancer Genet. Cytogenet. **1**:9-14.

van Loghem, E., and Natvig, J. B., 1970, Uncommon Gm gene complexes, Vox Sang. **18**:421-434.

van Loghem, E., Sukernik, R. I., Osipora, L. P., Zegers, B. J. M., Matsumoto, H., de Lange, G., and Lefranc, G., 1980, Gene deletion and gene duplication within the cluster of human heavy-chain genes. Selective absence of IgG sub-classes, J. Immunogenet. **7**:285-299.

Wang, A.-C., Geisely, J., and Fudenberg, H. H., 1973, Amino acid sequences at constant and variable regions of heavy chains of monotypic immunoglobulins G and M of a single patient, Biochemistry **12**:528-534.

Wang, A.-C., Tung, E., and Fudenberg, H. H., 1980, The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and functional implications, J. Immunol. **125**:1048-1054.

Whang-Peng, J., Henderson, E. S., Knutsen, T., Freireich, E. J., and Gart, J. J., 1970, Cytogenetic studies in acute myelocytic leukemia with special emphasis on the occurrence of Ph^1 chromosome, Blood **36**:448-457.

Williams, A. F., and Gagnon, J., 1982, Neuronal cell thy-1 glycoprotein: homology with immunoglobulin, Science **216**:696-703.

Yamawaki-Kataoka, Y., Miyata, T., and Honjo, T., 1981, The complete nucleotide sequence of mouse immunoglobulin γ2a gene and evolution of heavy chain genes: further evidence for intervening sequence-mediated domain transfer, Nucleic Acids Res. **9**:1365-1381.

Yaoita, Y., Kumagai, Y., Okumura, K., and Honjo, T., 1982, Expression of lymphocyte surface IgE does not require switch recombination, Nature **297**:697-699.

Yount, W. J., Hong, R., Seligmann, M., Good, R., and Kunkel, H. G., 1970, Imbalances of gamma globulin subgroups and gene defects in patients with primary hypogammaglobulinemia, J. Clin. Invest. **49**:1957-1966.

Zech, L., Haglund, U., Nilsson, K., and Klein, G., 1976, Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas, Int. J. Cancer **17**:47-56.

Zúñiga, M., D'Eustachio, P., and Ruddle, N. H., 1982, Immunoglobulin heavy chain gene rearrangement and transcription in murine T cell hybrids and T lymphomas, Proc. Natl. Acad. Sci. USA **79**:3015-3019.

FIGURE LEGENDS

Fig. 1. The three antibody gene families of mice. A, The light chain families, kappa and lambda, are drawn to scale on their respective chromosomes. The vertical lines immediately 5' to the V_L gene segments represent the coding sequences for the leader peptides. B, The heavy chain gene family. Distances between elements are not drawn to a precise scale, although the relative distances between C_H genes are approximated. The nine documented germline D segments are indicated, although several more are probably also present. The precise number of germline V_H and V_κ gene segments is not known, though both are thought to number several hundred (see text).

Fig. 2. DNA rearrangements leading to immunoglobulin production. A, V-gene formation. The process is shown for the heavy chain family, which is drawn schematically on the top line. Arrows flanking the regions to be joined on the V, D, and J gene segments represent the recognition sequences which mediate the recombination. B, Class switching. This event is shown for a switch from production of μ polypeptides to α heavy chains. The C_α gene displaces the C_μ gene from its position near the rearranged V_H gene.

Fig. 3. Alternative pathways of RNA splicing of immunoglobulin gene sequences. A, Strategies for the production of either membrane-bound or secreted forms of μ chains. P and M refer to the sequences coding for the signal peptide and transmembrane amino acid residues, respectively. The top line represents the DNA region involved, and the arrows lead to diagrams of the two alternate RNA products resulting from transcription. The blocks represent the contiguous sequences in the mature mRNAs, and the lines are the sequences spliced out of the nuclear

Fig. 3 (continued)

RNA precursor (redrawn from Early and Hood, 1981). B, Strategy for dual expression of μ and δ chains by a single lymphocyte. Following transcription of V_H, C_μ, and C_δ sequences into a single RNA precursor, either of the latter two sequences can be spliced to the V_H sequences to generate μ or δ mRNAs, respectively. Since IgM and IgD molecules appear on the surface of these cells, the membrane exons are presumably spliced onto the messages.

Fig. 4. The human light chain constant region loci. The single C_κ gene and the five J_κ gene segments are shown on the top line. The lower line shows the six linked C_λ genes. Sequences of the three 5' C_λ genes have correlated them with the indicated genetic variants Mcg, Ke^-Oz^-, and Ke^-Oz^+. The other three C_λ genes have not yet been correlated with known human lambda chain proteins. J_λ sequences have not yet been localized within this cluster.

Fig. 5. The human μ locus. The four C_μ exons are indicated by filled boxes. The μ membrane exons are shown as a single unfilled box, as are the C_δ sequences. Neither of these have been precisely localized relative to the C_μ gene. The J_H gene segments and single D gene segment are represented by short vertical lines. The sequences taking part in the class switch recombination lie between the J_H cluster and the C_μ gene. See details in the text.

Fig. 6. Model for the establishment of heavy chain hinge exons. The original 4-domain C_H gene was duplicated to give rise to the ancestors of the present-day C_μ and C_γ genes (other classes of C_H genes are not included here). The C_μ gene has retained the 4-domain organization, while as a result of the indicated mutations the C_γ gene has no counterpart to the ancestral C_H2 domain. In its place is a large noncoding intervening sequence and a small hinge exon.

Fig. 7. Model to account for the relationships among three human C_γ genes. The ancestral γ1 and pseudo-γ (Ψγ) genes underwent an unequal crossover; only the expanded chromosome is shown. The γ1-like hinge exon of the hybrid gene then was duplicated to give rise to the modern $C_{\gamma 3}$ gene.

Fig. 8. Two classes of homologous nonallelic recombination events. A and B, Unequal crossovers. The filled and open boxes represent two tandem homologous genes. A and B show events in which the strand breaks occur either in intergenic regions or within the structural genes, respectively. C and D, Gene conversions. The two cases shown both have one strand break in the region between the two tandem homologous genes. The other breakage-and-reunion events occur either within or between the two genes.

Fig. 9. Molecular earmark of homologous nonallelic recombination. Two tandem homologous genes are shown at the top. An alignment of the nucleotide sequences of the two genes reveals a relatively sharp boundary separating two areas showing different levels of sequence homology. The region having the higher level of homology was involved in the most recent recombination. The difference in homology levels reflects the timing of the recombination event relative to the original gene duplication or an earlier recombination event. The boundary represents the presumed recombination breakpoint. This type of structural relationship between homologous nonallelic genes has been described for human (Slightom et al., 1980; Liebhaber et al., 1981) and mouse (Konkel et al., 1979) globin genes and for mouse immunoglobulin genes (Miyata et al., 1980; Ollo et al., 1981; Yamawaki-Kataoka et al., 1981).

Fig. 10. Some key steps in B lymphocyte ontogeny. Several stages in the developmental pathway of antibody-producing cells are shown, with special reference to the molecular genetic processes discussed in the text.

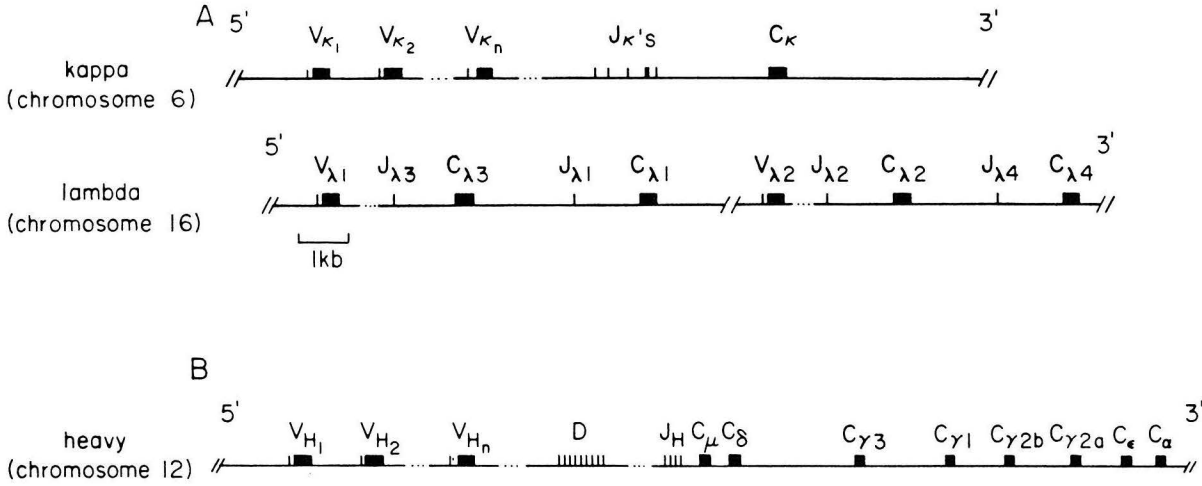
A
5'
V_{κ_1}
V_{κ_2}
V_{κ_n}
J_κ's
C_κ
3'
kappa
(chromosome 6)
5'
$V_{\lambda 1}$
$J_{\lambda 3}$
$C_{\lambda 3}$
$J_{\lambda 1}$
$C_{\lambda 1}$
$V_{\lambda 2}$
$J_{\lambda 2}$
$C_{\lambda 2}$
$J_{\lambda 4}$
$C_{\lambda 4}$
3'
lambda
(chromosome 16)
1kb
B
5'
V_{H_1}
V_{H_2}
V_{H_n}
D
J_H C_μ C_δ
$C_{\gamma 3}$
$C_{\gamma 1}$
$C_{\gamma 2b}$
$C_{\gamma 2a}$
C_ϵ
C_α
3'
heavy
(chromosome 12)

Figure 1

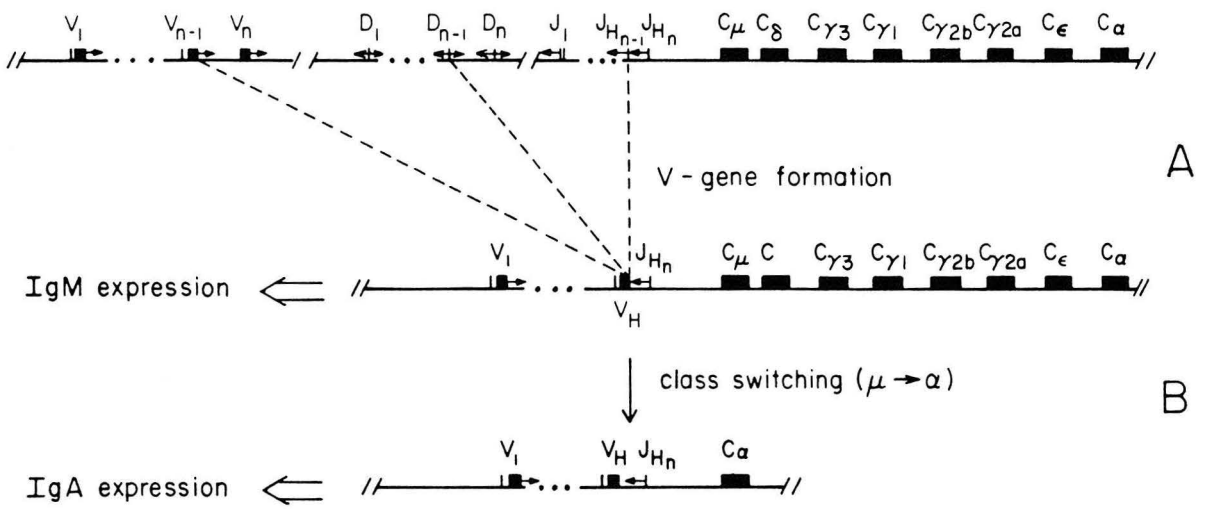
V1
Vn-1
Vn
D1
Dn-1
Dn
J1
JHn-1
JHn
Cμ Cδ Cγ3 Cγ1 Cγ2b Cγ2a Cε Cα
V - gene formation
A
IgM expression
V1
JHn
Cμ C Cγ3 Cγ1 Cγ2b Cγ2a Cε Cα
VH
class switching (μ → α)
B
IgA expression
V1
VH JHn
Cα

Figure 2

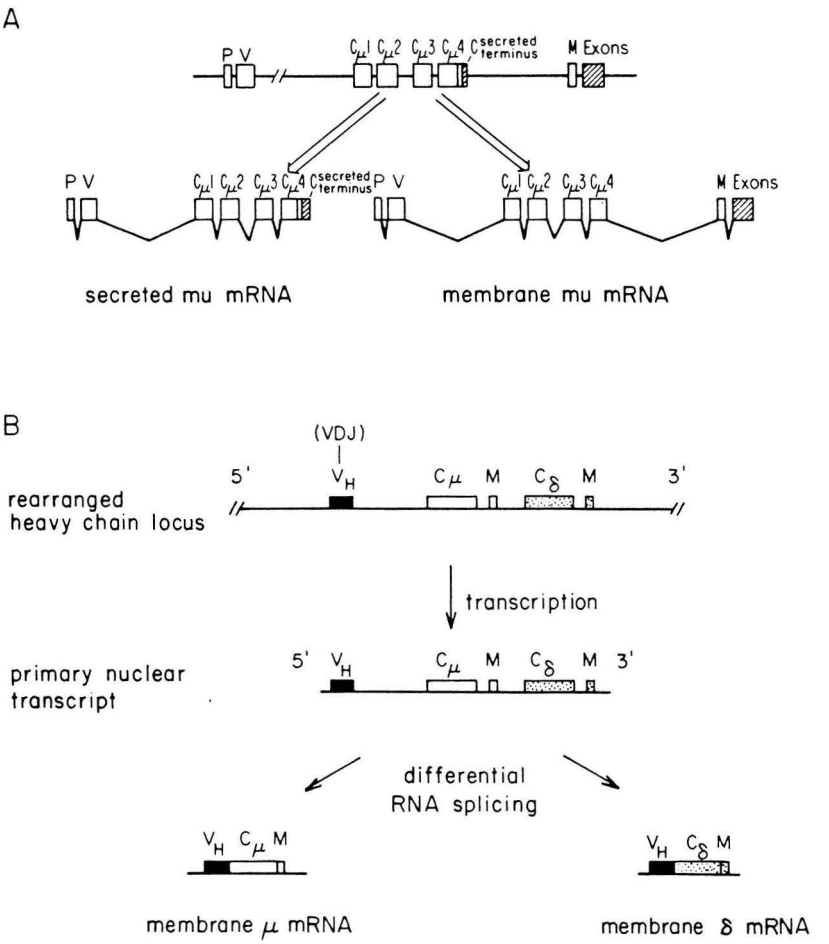
A
P V
Cμ1 Cμ2 Cμ3 Cμ4 Csecreted terminus
M Exons
secreted mu mRNA
membrane mu mRNA
B
(VDJ)
5'
VH
Cμ
M
Cδ
3'
rearranged heavy chain locus
transcription
primary nuclear transcript
differential RNA splicing
membrane μ mRNA
membrane δ mRNA

Figure 3

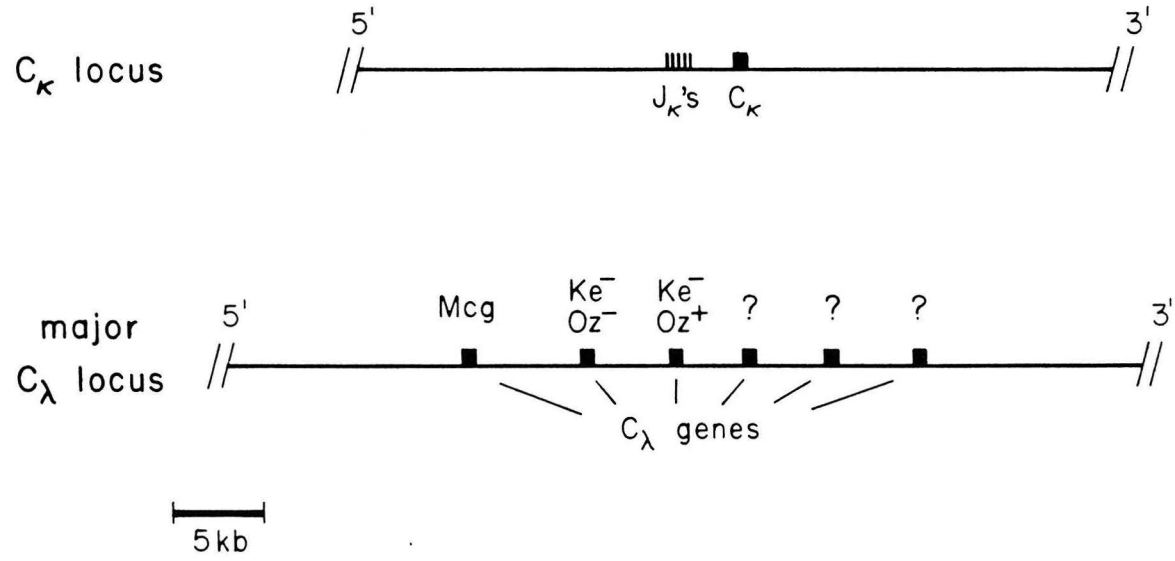
C_κ locus
5'
3'
J_κ's
C_κ
major
C_λ locus
5'
Mcg
Ke^- Oz^-
Ke^- Oz^+
?
?
?
3'
C_λ genes
5kb

Figure 4

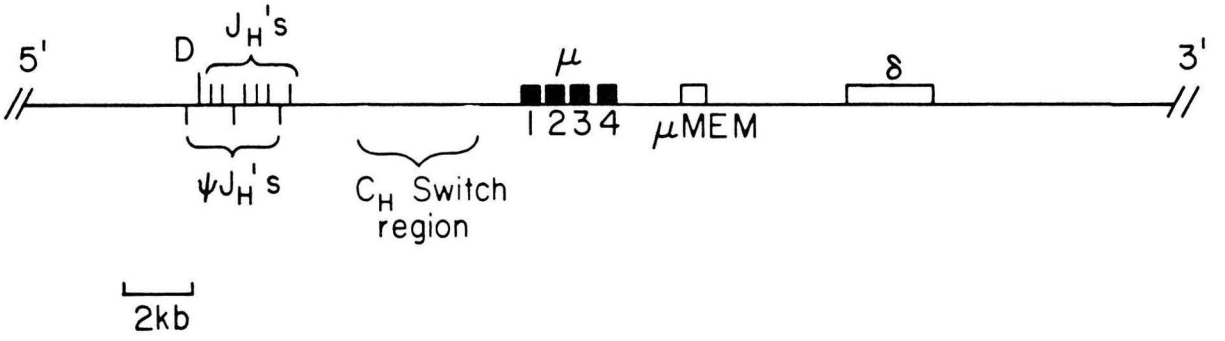

Figure 5

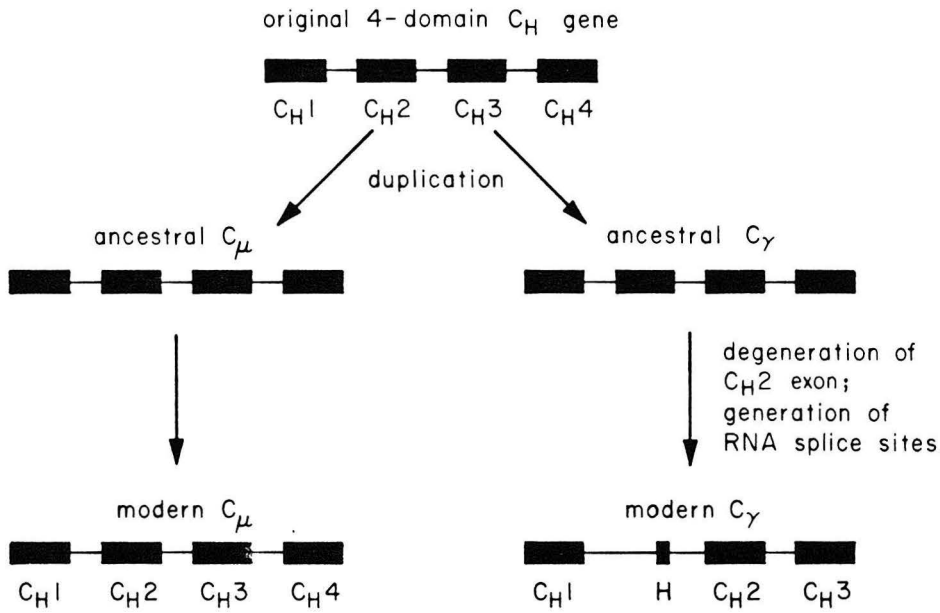
original 4-domain C_H gene
C_H1 C_H2 C_H3 C_H4
duplication
ancestral C_μ
ancestral C_γ
degeneration of C_H2 exon; generation of RNA splice sites.
modern C_μ
C_H1 C_H2 C_H3 C_H4
modern C_γ
C_H1 H C_H2 C_H3

Figure 6

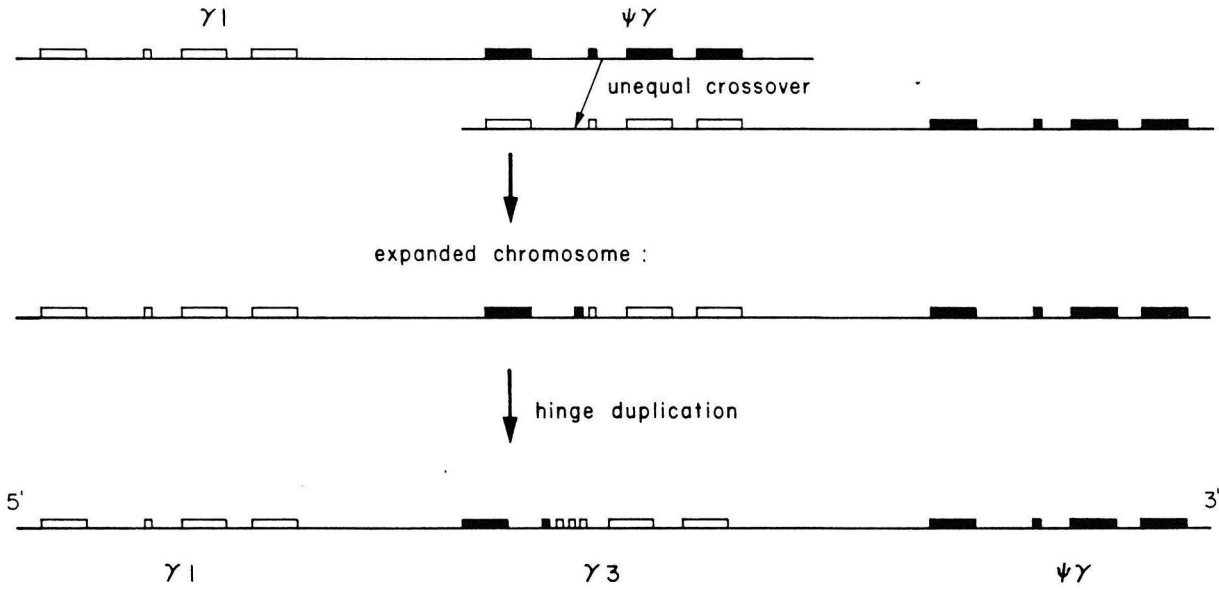
γ1
ψγ
unequal crossover
expanded chromosome :
hinge duplication
5'
3'
γ1
γ3
ψγ

Figure 7

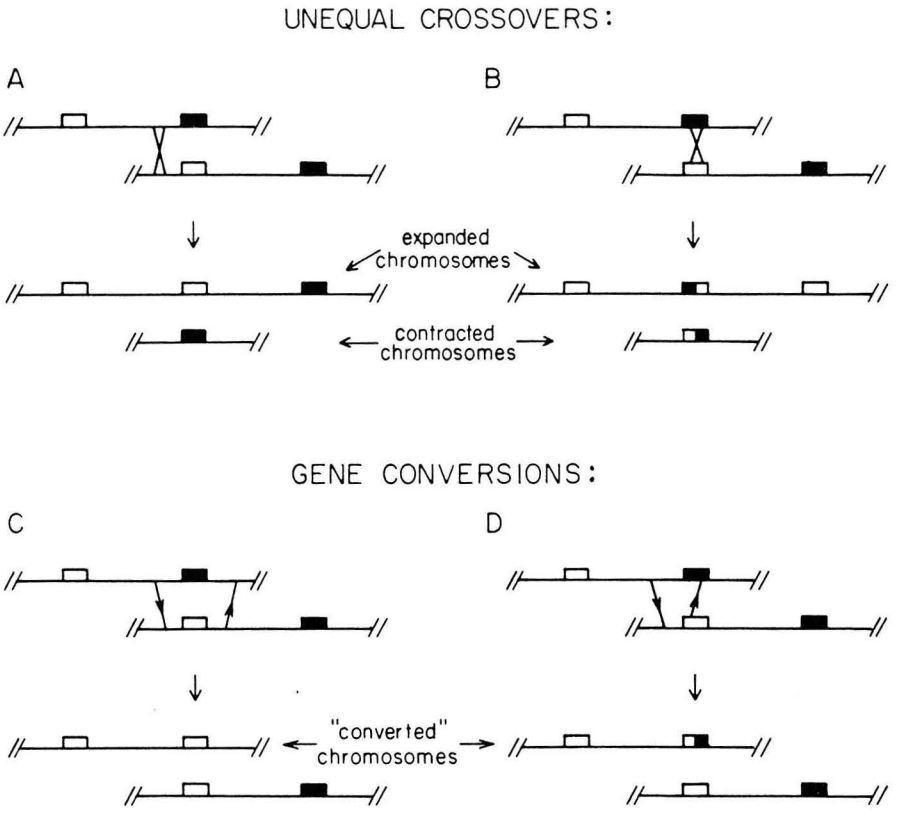
UNEQUAL CROSSOVERS:
A
B
expanded chromosomes
contracted chromosomes
GENE CONVERSIONS:
C
D
"converted" chromosomes

Figure 8

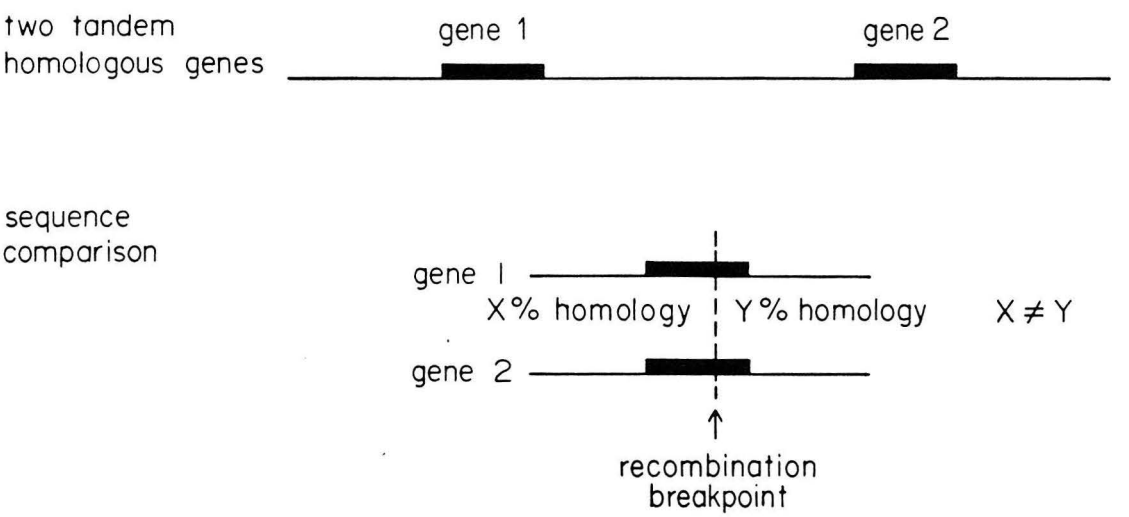
two tandem
homologous genes
gene 1
gene 2
sequence
comparison
gene 1
X% homology
Y% homology
X ≠ Y
gene 2
recombination
breakpoint

Figure 9

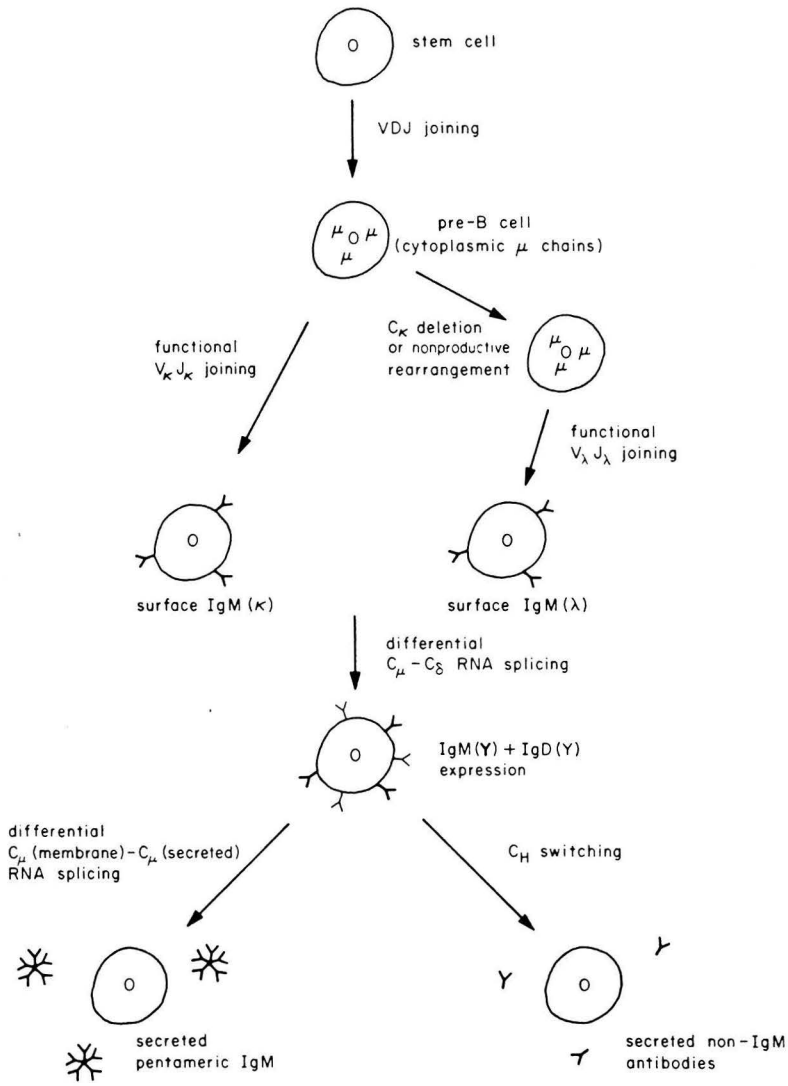
stem cell
VDJ joining
pre-B cell
(cytoplasmic μ chains)
functional
$V_\kappa J_\kappa$ joining
C_κ deletion
or nonproductive
rearrangement
functional
$V_\lambda J_\lambda$ joining
surface IgM (κ)
surface IgM (λ)
differential
$C_\mu - C_\delta$ RNA splicing
IgM(Y) + IgD(Y)
expression
differential
C_μ (membrane) - C_μ (secreted)
RNA splicing
C_H switching
secreted
pentameric IgM
secreted non-IgM
antibodies

Figure 10

CHAPTER 6

Conclusions

Chapters 2, 3, and 4 describe the structural analyses of three members of the human immunoglobulin C_γ gene family: the genes encoding constant regions of γ4, γ2, and γ1 heavy chains, respectively. As expected from earlier protein studies, these genes share extensive nucleotide sequence homology in their C_H domain exons. The DNA studies described here show that they share almost as much homology in their introns and other noncoding regions. It is generally believed that intron regions have no function dependent on their sequence, and thus are free to diverge without selective constraints. Thus the rate of fixation of these substitutions is expected to be rather constant for all genes, and the level of difference in these areas between any two compared genes reflects the time elapsed since they shared an identical sequence. Indeed, estimates have been made of the rate of appearance of these "unselected" nucleotide substitutions. Perler et al. (1980) place this value at 7×10^{-9} nucleotide substitutions per site per year, while a different type of analysis led Miyata et al. (1980a) to estimate this rate at about 5×10^{-9}. By determining the sequence divergence in introns (and/or other noncoding, presumably unselected, regions), one may use one of the above values to estimate the time of divergence of two homologous sequences. I have used the value of Perler et al. to calculate that approximately 6-8 million years have elapsed since any two of the $C_{\gamma 1}$, $C_{\gamma 2}$, or $C_{\gamma 4}$ genes were identical in sequence. As noted in Chapter 4, the need for fewer gaps in the $C_{\gamma 2}$-$C_{\gamma 4}$ sequence alignment, as well as the determined linkage of these two genes (Chapter 3), makes it seem likely that they diverged more recently from each other than from the $C_{\gamma 1}$ gene.

The most divergent parts of the C_γ genes are the hinge exons. These coding regions have evolved more rapidly than any other areas of the C_γ genes in terms of both base substitution and deletion/insertion events. The clearly nonrandom clustering of nucleotide changes in hinge exons seems inconsistent with their fixation by random

drift, as neutralists would argue. As discussed in Chapters 3-5, I believe the sequence differences in hinge exons reflect the selection of particular structures in the corresponding regions of gamma heavy chains. The divergence in C_H exons in residues lying near the hinge in the polypeptide chain (Chapter 4) further suggests that the relevant amino acid sequences have been selected for. This view implies that the different hinge regions each have beneficial effects on IgG function, and the rapid changes in these sequences suggest that strong selective pressure has continually acted to fix new hinge exons in the human population. The functional importance of hinge regions is not well understood at present, although others have suggested that different hinge structures have large effects on conformational states of IgG molecules (Cebra et al., 1977; Klein et al., 1981; Wang et al., 1980). These conformational states may be envisioned as being important for interactions with other molecules, such as complement proteins or cell-surface receptors. This selectionist view argues that the selective force for the fixation of nucleotide substitutions in hinge regions is the beneficial impact of the particular hinge structures on the functions of IgG molecules.

Besides the three C_γ genes described here, two other members of the gene family have recently been described by other workers. In addition to the expected $C_{\gamma 3}$ gene encoding the C_H region of IgG3 molecules, a fifth C_γ gene has been partially characterized by Takahashi et al. (1982). Two structural features of this latter gene have led to its tentative designation as a pseudogene: (1) the predicted amino acid sequence of the hinge region does not correspond to that of any reported human gamma chain; and (2) although the $C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\gamma 3}$, and $C_{\gamma 4}$ genes have sequence homology to the switch (S) regions of mouse C_γ genes, the putative pseudogene apparently lacks this region. Regardless of whether this fifth gene is functional, the available sequence data indicate that its structural relationship to the other

members of the gene family follows the pattern described in this work: extensive homology to the other C_γ genes exists in introns and C_H exons, and the hinge exon shows a high level of divergence.

The analyses of the C_γ hinge exons by Takahashi et al. revealed some interesting implications regarding the evolution of the gene family. As expected from the protein sequence data showing repeating amino acid sequences in the hinge regions of IgG3 molecules, the $C_{\gamma 3}$ hinge region is encoded by multiple exons. The 5'-most hinge exon is most homologous to the pseudogene hinge, while the three other exons (which are identical to one another) are more similar to the $C_{\gamma 1}$ hinge exon than to any other C_γ hinge. These observations led to the proposal of unequal crossover followed by hinge duplication to account for the generation of the modern $C_{\gamma 3}$ gene (see Chapter 5). These proposed events fall into the category of genetic processes discussed in Chapters 3 and 4, and referred to in Chapter 5 as "homologous nonallelic recombination." Thus although my analysis of the $C_{\gamma 1}$, $C_{\gamma 2}$, and $C_{\gamma 4}$ genes revealed no structural evidence of these events, homologous nonallelic recombination has evidently have played an important role in the evolution of human C_γ genes, and is not necessarily restricted to the events proposed by Takahashi et al. I think it likely that homologous nonallelic recombination acted on the human C_γ gene family 6-8 million years ago, thus accounting for the extreme similarity observed for the modern genes.

The evolutionary scheme proposed by Takahashi et al. for the $C_{\gamma 1}$, $C_{\gamma 3}$, and pseudo-C_γ genes predicts a specific linkage arrangement for these genes: 5'-$C_{\gamma 1}$-$C_{\gamma 3}$-$\Psi\gamma$-3'. This configuration has not been directly demonstrated, however, and thus far the only linkage relationship established is for the $C_{\gamma 2}$ and $C_{\gamma 4}$ genes (Chapter 3). In an effort to extend the map of the chromosomal region(s) containing C_γ genes, a cosmid library of (my own) sperm DNA was constructed and screened as

described by Steinmetz et al. (1982). Only a single cosmid clone, containing part of a $C_{\gamma 2}$ gene, was isolated from this library. Its position with respect to the $C_{\gamma 2}$-$C_{\gamma 4}$ cluster, along with that of a recombinant phage clone, are shown in Figure 1. Although no other C_{γ} clones were found in the region encompassed by these "walking" clones, I believe I can tentatively map the $C_{\gamma 2}$ gene at the 5' end of the C_{γ} cluster for two reasons. First, the distance along the chromosome 5' of $C_{\gamma 2}$ is about 46 kilobases; this correlates most closely with the distance between the 5'-most mouse C_{γ} gene and the C_{δ} gene. All of the other established distances between either human or mouse C_{γ} genes are significantly shorter. A more solid reason for my mapping of the $C_{\gamma 2}$ gene regards its expression: in humans and mice, the C_H region expressed in the antibody response to carbohydrate antigens (e.g., on bacterial cell walls) are of the $\gamma 2$ and $\gamma 3$ subclasses, respectively. Since the mouse $C_{\gamma 3}$ gene has been localized at the 5' end of the mouse C_{γ} cluster (Shimizu et al., 1982), I think it is likely that the human $C_{\gamma 2}$ gene occupies an analogous position with respect to the other members of the human C_{γ} gene family. The mechanism whereby the human $C_{\gamma 2}$ and mouse $C_{\gamma 3}$ genes are selected for the response to carbohydrate antigens is unknown. Future work on human C_{γ} genes may be expected to focus on aspects such as this: the differential expression of individual members of a gene family.

REFERENCES

Cebra, J. J., Brunhouse, R. F., Cordle, C. T., Daiss, J., Fecheimer, M., Ricardo, M., Thunberg, A., and Wolfe, P. B. (1977). Isotypes of guinea pig antibodies: Restricted expression and basis of interactions with other molecules. Prog. Immunol. **3**, 264-277.

Klein, M., Haeffner-Cavaillon, N., Isenman, D. E., Rivat, C., Navia, M. A., Davies, D. R., and Dorrington, K. J. (1981). Expression of biological effector functions by immunoglobulin G molecules lacking the hinge region. Proc. Natl. Acad. Sci. USA **78**, 524-528.

Miyata, T., Yasunaga, T., and Nishida, T. (1980). Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**, 7328-7332.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980). The evolution of genes: The chicken preproinsulin gene. Cell **20**, 555-566.

Shimizu, A., Takahashi, N., Yaoita, Y., and Honjo, T. (1982). Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell **28**, 499-506.

Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S., and Honjo, T. (1982). Structure of human immunoglobulin gamma genes: Implications for evolution of a gene family. Cell **29**, 671-679.

Wang, A.-C., Tung, E., and Fudenberg, H. H. (1980). The primary structure of a human IgG2 heavy chain: Genetic, evolutionary, and functional implications. J. Immunol. **125**, 1048-1054.

Figure 1. Restriction enzyme map of the human $C_{\gamma 2}$-$C_{\gamma 4}$ region. Positions of cleavage of various restriction enzymes are indicated at the bottom by short vertical lines. Parentheses enclose regions which were not mapped for the relevant enzymes. The eight unmarked horizontal lines represent cloned human chromosomal fragments isolated either from a cosmid library of human sperm DNA (next-to-highest line) or a Charon 4A bacteriophage library of human fetal liver DNA (all other lines). The top line shows the length of DNA encompassed by the cloned fragments. Solid rectangles represent the C_{γ} coding regions (introns are not indicated). kb = kilobase pairs.

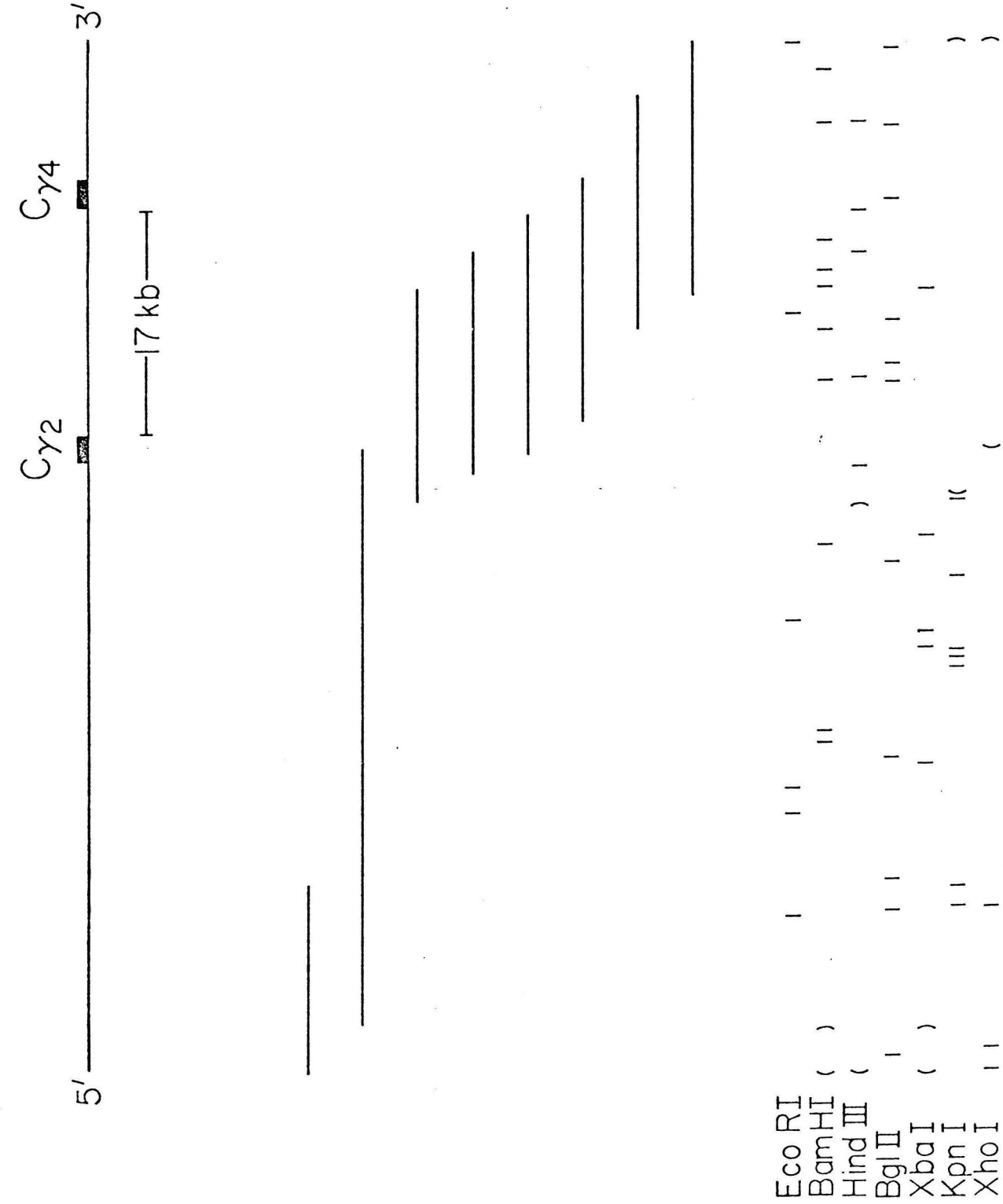

Figure 1