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 $\gamma 1$ 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

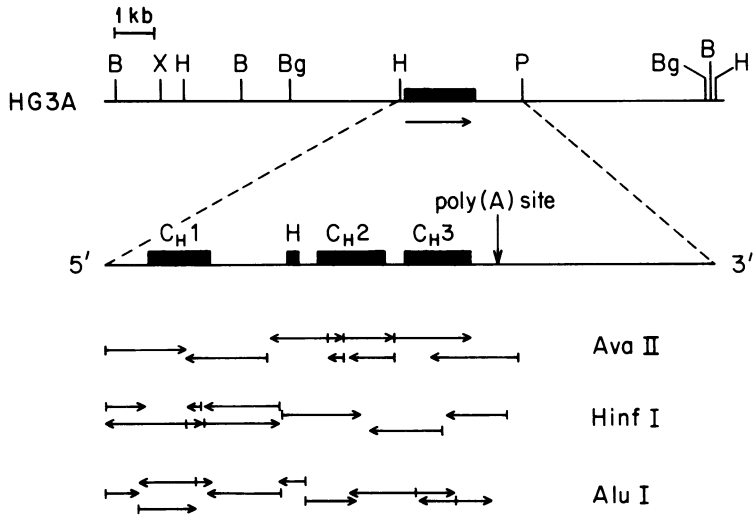

Figure 1. Restriction map and sequencing strategy of a cloned human DNA fragment containing a $C_{\gamma 1}$ gene. Letters on the top line refer to cleavage sites for the following restriction enzymes: B, Bam HI; H, Hind III; Bg, Bgl II; P, Pvu II; X, Xba I. Only the indicated Pvu II site was mapped, 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

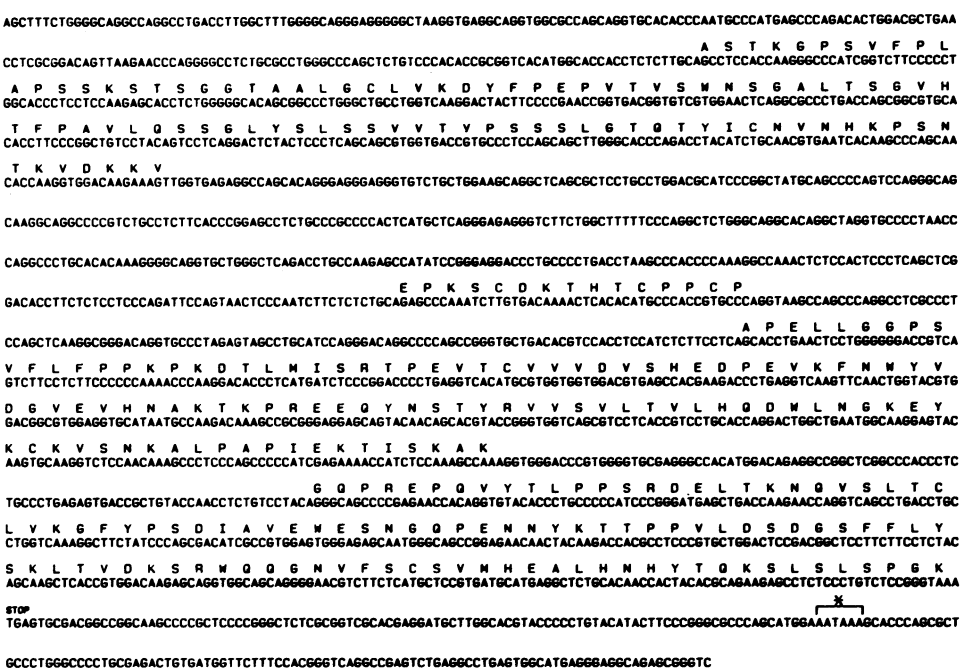

Figure 2. The nucleotide sequence of a human $C_{\gamma 1}$ gene and its corresponding protein sequence. The sequence of the mRNA synonymous strand is listed 5' to 3'. Amino acids predicted by the DNA sequences are listed in one-letter code above the respective codons. "Stop" indicates the termination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

C_γ gene	length of gene segment (nucleotides)							
	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

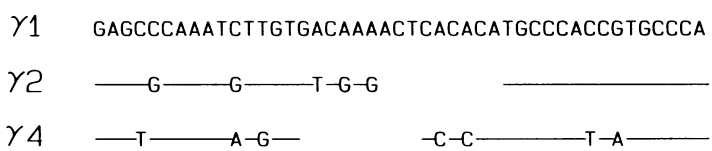

Figure 3. Comparison of hinge exon nucleotide sequences. Solid lines represent identity of the γ2 and γ4 sequences to the γ1 sequence. Where differences occur in the γ2 and γ4 exons, the relevant residues are listed. Gaps are introduced into the γ2 and γ4 listings to maximize homology to the γ1 sequence.

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.

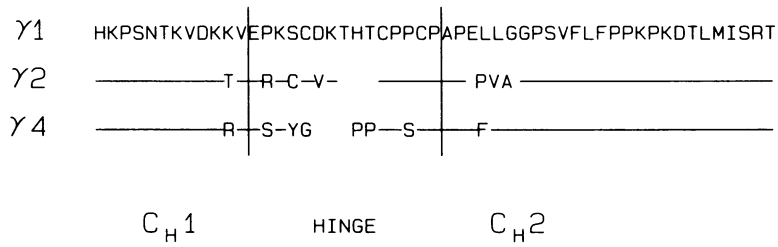

Figure 4. Comparison of amino acid residues in the hinge area of three C_γ polypeptides. Vertical lines separate the hinge residues from those contiguous amino acids which are encoded in the C_H1 and C_H2 exons. Amino acids are listed in the one-letter code. Solid lines represent identity of the γ2 and 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.