Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4484–4488, June 1986
Immunology

Structure of the human Ia-associated invariant (γ)-chain gene: Identification of 5′ sequences shared with major histocompatibility complex class II genes

(genomic library/nucleotide sequence)

DEIRDRE M. O'SULLIVAN*, DAN LARHAMMAR†, MICHAEL C. WILSON‡, PER A. PETERSON†, AND VITO QUARANTA*§

*Department of Immunology, ‡Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, CA 92037; and †Department of Cell Research, The Wallenberg Laboratory, Box 562, S-751 22 Uppsala, Sweden

Communicated by Frank J. Dixon, February 10, 1986

ABSTRACT **The human gene encoding the Ia-associated γ (or invariant) chain was isolated by screening a genomic library in phage λ with cDNA probes. The frequency of positive clones in the library, the overlapping restriction maps of the cloned fragments, and the patterns of genomic hybridization suggested that the γ-chain gene exists as a single copy per haploid genome. The gene consists of 8 exons, spanning approximately 12 kilobases of DNA. All exon sequences were in an open reading frame, contained appropriate splice junction sequences, and encompassed the entire sequence of full-length γ-chain mRNA, suggesting that the gene we isolated is most likely functional. Furthermore, "CAAT"-type and "TATA"-type promoter sequences were found at the expected positions upstream from the proposed cap site. The organization of the γ-chain gene has none of the distinctive features of the immunoglobulin superfamily of genes, of which Ia α and β chains are members. Therefore, the evolutionary origins, and perhaps the functions, of the Ia γ chains are distinct from those of the other two Ia subunits α and β. Despite the unrelatedness of these genes, consensus sequences found approximately 150 base pairs upstream from all the Ia α- and β-chain genes sequenced to date were also found in analogous positions in the γ-chain gene, suggesting a possible role in the coregulation of expression of these genes.**

Ia molecules are polymorphic cell surface glycoproteins expressed primarily by antigen-presenting cells and B lymphocytes (reviewed in ref. 1). They are of key importance in the immune system, since they restrict recognition of foreign antigen to T-helper lymphocytes sharing the appropriate Ia phenotypes. Ia molecules are formed by three noncovalently bound transmembrane glycoproteins, the highly polymorphic α and β chains, encoded by the major histocompatibility complex (MHC) class II genes, and the nonpolymorphic γ (or invariant) chain (reviewed in ref. 2). At the cell surface, they comprise only α–β dimers, as γ chains dissociate from Ia oligomers during Golgi processing (3, 4). The expression of α, β, and γ chains is strictly coregulated during development, most likely at the level of gene transcription (5, 6), as indicated by their virtually identical tissue distribution (7, 8) and their parallel induction by γ-interferon (5). An understanding of the regulatory mechanisms of Ia expression is of great interest, since expression of Ia molecules by immune cells is a key requirement for recognition of foreign antigens by T lymphocytes (9), and since Ia expression by nonimmune cells has been involved in the pathogenesis of autoimmune diseases (10). The cloning of α- and β-chain genes (reviewed in ref. 11) has provided necessary tools for studying the molecular basis for Ia function and expression. Thus, the detailed structure of several α and β chains from many haplotypes has been deduced from their respective genes, and the organization of the MHC class II gene family in mouse and man is almost entirely known. Expression of these genes by DNA-mediated transfer is yielding important information on the structure–function relationship of Ia α and β chains (1, 11).

More limited information is available on the γ subunit of Ia molecules. The characterization of cDNA clones corresponding to a human p31 γ chain (12, 13) showed that γ chains bear no structural relationship to α and β chains and are encoded outside of the MHC (2). Two-dimensional electrophoretic analyses revealed structural heterogeneity of human γ chains, since, in addition to the predominant p31 γ-chain form, two more basic and larger forms, named γ2 and γ3, were found in immunoprecipitates of anti-p31 and anti-β-chain monoclonal antibodies (2, 7). The function of γ chains is not known: it has been proposed that they may act as an intracellular transport piece for Ia α and β chains (2); a direct involvement in Ia-dependent immune recognition, however, has not been investigated.

Towards the goal of understanding the functional role of γ chains, the mechanisms for generating their structural diversity, and the molecular basis for their coordinate expression with Ia α and β chains, we have isolated and characterized the gene encoding the human γ chain. Here we report the structure of this gene, and the identification of upstream sequences shared with α- and β-chain genes that may contribute to coregulating their expression.

MATERIALS AND METHODS

Genomic DNA Library Screening. A human liver genomic library in the λ phage vector Charon 4A was kindly provided by T. Maniatis (Harvard University, Boston, MA) (14). Approximately 10^6 phages from this library were plated on *Escherichia coli* LE 392, blotted onto nitrocellulose filters, and screened as described (15) by hybridization to a human γ-chain cDNA (12) probe, pγ2, ^{32}P-labeled by nick-translation (15). Phage DNA from plaque-purified (15) positive clones was prepared in liquid cultures as described (15).

DNA Mapping, Subcloning, and Sequencing. The isolated positive phage clones were characterized by standard restriction mapping. The clone, λγ3, was digested with *Eco*RI and two fragments, 7.5 and 9.2 kilobases (kb), respectively, were

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.

Abbreviations: bp, base pair(s); kb, kilobase(s); MHC, major histocompatibility complex.

§To whom reprint requests should be addressed at: IMM5, Department of Immunology, Research Institute of Scripps Clinic, 10666 N. Torrey Pines Road, La Jolla, CA 92037.

inserted into the plasmid vector, pUC9. Exon-containing regions, identified by hybridization to nick-translated probes corresponding to the 5', the middle, and the 3' regions of pγ2, were sequenced using the method of Maxam and Gilbert (16).

For genomic hybridization (15), high molecular weight human DNA was prepared (15) from the lymphoblastoid cell line RAJI and from normal peripheral blood lymphocytes, digested with either *Eco*RI or *Hin*dIII, and transferred to nitrocellulose filters by blotting (15). The filters were hybridized (48 hr at 60°C) to the γ-chain cDNA, washed at high stringency, and analyzed by autoradiography.

RESULTS AND DISCUSSION

Isolation of a Human γ-Chain Genomic Clone. A human γ-chain cDNA probe (12) was used to screen approximately five genomic equivalents of a human genomic library in phage λ. Three positive clones were isolated and characterized by restriction endonuclease mapping. These analyses indicated that the three cloned inserts contained overlapping DNA fragments (Fig. 1*A*). The two *Eco*RI fragments from clone λγ3, inserted into the plasmid vector pUC9, were used in Maxam–Gilbert sequence determination (Figs. 1*B* and 2). Coding sequences were identified by comparison with the full-length cDNA sequence (12, 13), confirming that we had isolated the entire γ-chain gene in a single phage clone and allowing us to locate the exon-intron boundaries.

Hybridization of γ-chain cDNA probe to human genomic DNA digested with either *Eco*RI or *Hin*dIII (not shown) showed two hybridizing *Eco*RI fragments of >20 kb and 9.0 kb and four *Hin*dIII fragments of 18 kb, 2.5 kb, 2.25 kb, and 0.6 kb that corresponded exactly to the restriction maps of the isolated phage clones (Fig. 1 *A* and *B*). Together with the overlapping identity and the library frequency of the three genomic clones, this result strongly suggested that the γ-chain gene is a single-copy gene. However, one cannot rule out the possibility of additional gene copies yielding restriction fragments of identical sizes and/or of distantly related genes not detectable under the high stringency hybridization conditions used.

Organization of the Human Iγ γ-Chain Gene. The human γ-chain gene is approximately 12 kb long and consists of 8 exons (Figs. 1 and 2). Exons 1 and 2, respectively 135 and 173 base pairs (bp) long, are separated by a rather large 5.4-kb intron. The 30-amino acid residues representing the protein's cytoplasmic tail (12, 13) are encoded both by exons 1 (26 residues) and 2 (4 residues), and the 26 residues of the predicted transmembrane region (12, 13) are encoded by bp 14–91 of exon 2 (173 bp long). Exons 3 to 7 range in size from 63 to 96 bp and are scattered in the approximately 6 kb dividing exons 2 and 8 (Figs. 1*B* and 2). Exon 4 encodes both of the putative glycan addition sequences (dotted residues in Fig. 2) (12). The intron separating exons 4 and 5 contains an *Alu*-type repeat (underlined in Fig. 2). Exon 8 comprises the eight terminal bp of the coding sequence followed by the entire 587 bp of the 3'-untranslated sequence with one AATAAA polyadenylylation signal.

All intron-exon junctions contained the GT/AG donor/acceptor sequences. Because of codon redundancy, the exon boundaries were assigned by reference to this splice consensus. In addition, the consensus sequence YNYTRY, occurring 18–40 nucleotides proximal to the 3' splice site and thought to be necessary for correct mRNA splicing (18), was found close to the 3' splice sites of all introns. These results suggested that the γ-chain gene we isolated can be spliced correctly and is probably functional.

Organization of the 5' End of the γ-Chain Gene. Approximately 800 bp of the 5' flanking region was sequenced (Fig. 2), since this region is thought to be important in the regulation of gene expression. The transcription initiation site, or cap site, was tentatively assigned to the adenosine residue, labeled +1 in Fig. 2, based on the nucleotide structure of the γ-chain mRNA sequenced to its full length by primer extension (13) and on the fact that an adenosine

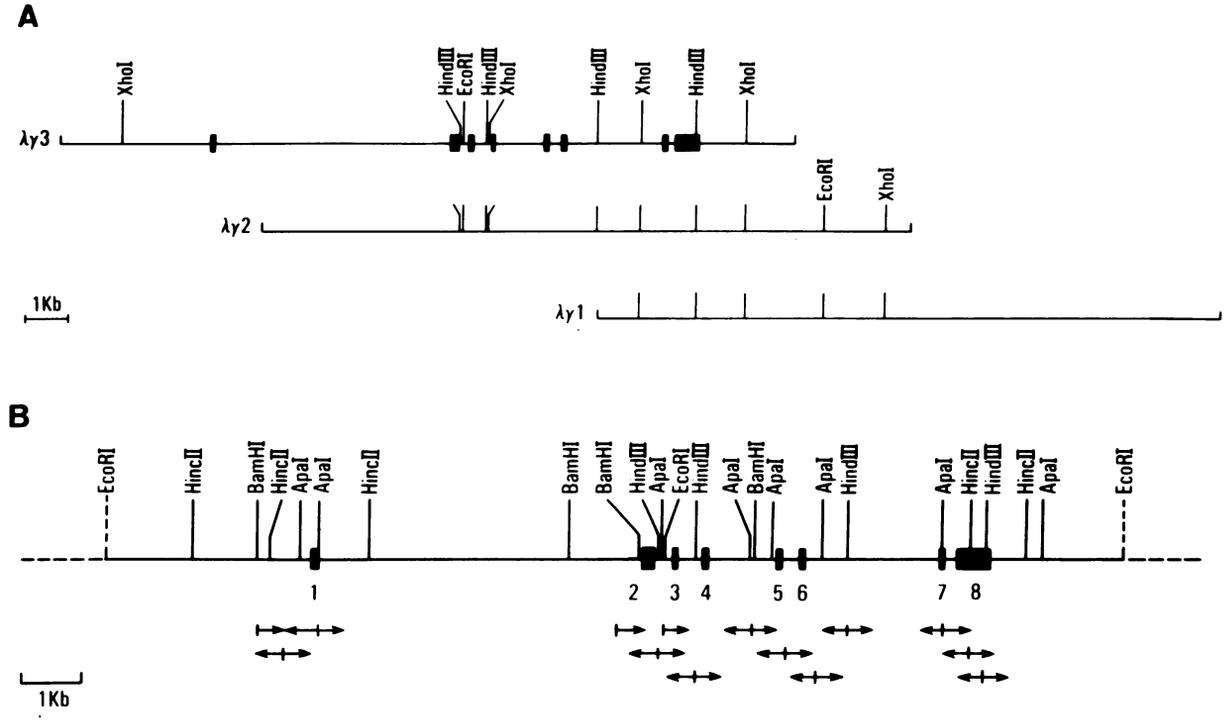

FIG. 1. (*A*) Restriction map of three overlapping positive clones isolated from a human genomic library in bacteriophage λ Charon 4A screened by hybridization to a radioactive human γ-chain cDNA (12) probe. Coding regions are indicated by black rectangles. (*B*) Detailed restriction map of the clone λγ3 with the general strategy used in Maxam and Gilbert sequencing indicated by arrows. Broken lines indicate the *Eco*RI linkers used for ligation into the phage arms, as described (14).

FIG. 2. Partial nucleotide structure of human Ia-associated γ-chain gene. In the region 5′ to exon 1, TATA- and CAAT-type sequences are boxed and 15-mer and 8-mer sequences sharing homology with upstream sequences of Ia α and β genes (see Table 1) are underlined. The approximate lengths of intronic regions that were not sequenced are indicated. Amino acids are numbered from the second ATG in exon 1, and the proposed cap site is indicated by +1. Glycan addition sites (12) in exon 4 are indicated by dots. An *Alu*-type repeat (17) in the intron separating exons 4 and 5 is underlined. The polyadenylylation signal AATAAA is also underlined.

preceded by a cytidine, in most cases, is the preferred cap site (19). The sequences CAAGT and TTTAA, found at −77 and −22 bp, respectively, from the proposed cap site (boxed sequences in Fig. 2) share sufficient homology to the "CAAT" and "TATA" consensus sequences (20) to be considered likely promoter sequences in the γ-chain gene.

As mentioned above, the γ-chain gene most likely is represented once per haploid genome. However, since multiple forms of γ chains, not due to posttranslational modifications, are found in human (7) and murine (21) Ia-positive cells, it is likely that the transcripts from this gene undergo alternate processing. Alternate initiation of translation is one possibility, since inspection of the 5′ end of exon 1 (Fig. 2) showed the existence of two in-frame ATG codons as possible start sites for translation, in agreement with the reported full-length mRNA sequence (13). The most 3′ ATG codon (marked 1 in Fig. 2), corresponds to the amino terminus of the predominant form of γ chain (p31), as determined by partial protein sequencing (12). (Note that γ chains are devoid of a signal peptide, ref. 12.) This site may be the one of choice since it fits the proposed initiation consensus RNNAUGG (22). The most 5′ ATG, if used to start translation, would give rise to a protein 16 amino acids larger and more basic (Fig. 2) than the p31 γ chain, characteristics compatible with the two-dimensional gel electrophoresis migration of two 33-kDa γ-chain forms designated γ2/γ3 (2, 7). According to which ATG codon is used, either an unusually short 10-bp or a 58-bp 5′-untranslated leader would be generated. Additional exon sequences may also exist in the human γ-chain gene to account for a γ chain-related 41-kDa protein (2).

Sequences on the 5′ Side of the γ-Chain Gene Shared with MHC Class II Genes. While Ia α and β chains are structurally related and encoded by linked genes (1, 11), characterized γ-chain cDNA clones (12, 13) have shown that the γ-chain protein bears no obvious structural relationship to α and β chains, has opposite transmembrane orientation (amino terminus facing the cytoplasm), and is encoded on a different chromosome both in mouse (23) and man (24). The structure of the γ-chain gene we present here has virtually none of the features distinctive of the immunoglobulin supergene family (25), of which the Ia α- and β-chain genes are members, such as protein structural domains delineated by exon-intron boundaries. Altogether these observations indicate that the γ chain has a separate evolutionary history from the other two Ia subunits, supporting the view that its functions, such as the proposed regulation of intracellular transport and/or of association of α and β chains (2), may not be strictly immunological.

Table 1. Homologous 5′ sequences flanking Ia α-, β-, and γ-chain genes

Gene	15-mer	Spacer, bp	8-mer	bp to AUG
γ chain	----------GA-----	44	T---C-----	227
Aβ2	----------A-----	16	T---------	119
Aβ	----------	18	---------	125
Eβ	A---------T------	18	---------	141
DQβ	----------T-	18	---------	133
DQ3β	----------	19	---------	130
DXβ	--------G--------	19	---------	
DPβ	---------GCA-----	18	TCC------	144
DRβ	A---------T------	18	-------CA	148
Eα	----------T	19	---------	99
DQα	G---------T---GAT	19	---A------	116
DXα	A---G-----ACA---AT	20	---------	
DZα	----------AC	20	G---------	117
DPα	----------GA-	19	------A---	138
DRα	----------C	19	---------	130
Consensus	CCYAGNRACAGATGA		CTGATTGG	

The consensus sequences for the 15-mer and the 8-mer were compiled on the basis of the reports by Larhammar *et al.* (26), Okada *et al.* (27), and Kelly and Trowsdale (28). These consensus sequences were used to search the γ-chain gene for homologies with the aid of the algorithm of Wilbur and Lipman (29) (WORDSEARCH program, University of Wisconsin Genetics Computer Group). Previous searches of the EMBO and GENBANK databases with the same algorithm had shown that sequences homologous to these consensus sequences are only found in the upstream regions of *E. coli* glutathione synthetase gene and in two histone genes (28). N, any nucleotide; Y, pyrimidine; R, purine.

Despite the lack of relatedness between γ-chain and the α- and β-chain genes and the fact that the γ-chain gene lies outside the MHC, we found a segment of sequence homology in the 5′-flanking sequence that is of potential interest. Thus, all human and murine Ia α- and β-chain genes sequenced to date possess sequences fitting the consensus CCYAGNRACAGATGA/16–20 bp/CTGATTGG, at a distance of 100–150 bp on the 5′ side of the ATG start codon for translation (26–28). The sequences CCCAGAAACAAGTGA and TTGCTTGG, separated by 44 bp and located 227 bp upstream of the ATG start codon of the γ-chain gene, are in excellent agreement with that consensus sequence, as outlined in Table 1. Since these sequences are located in a region (the 5′ flank) thought to direct selective gene expression in differentiated cells (30), it will be important to test their possible involvement in the regulatory mechanisms that coordinate the expression of α-, β-, and γ-chain genes in a tissue specific fashion or in response to lymphokines. Supporting this possibility is the fact that this consensus sequence is thus far virtually unique to Ia α, β, and γ genes (see legend to Table 1), and the reports that the tissue specificity (31) and γ-interferon inducibility (32) of class II gene expression may be mediated by sequences in the 5′-flanking region of those genes.

Note Added in Proof. Since submission of this manuscript, a similar study by Kudo *et al.* (33) has been reported.

The authors thank Alice Bruce for excellent secretarial assistance. This work was supported by Grant AI-10088 from the National Institutes of Health and by the Swedish Cancer Society. D.M.O'S. is a fellow, and V.Q. is a scholar of the Leukemia Society of America. This is publication 4244IMM from the Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, CA 92037.

1. Giles, R. C. & Capra, J. D. (1985) *Adv. Immunol.* **37,** 1–71.
2. Long, E. O. (1985) *Surv. Immunol. Res.* **4,** 27–34.
3. Moosic, J. P., Sung, E., Nilson, A., Jones, P. P. & McKean, D. J. (1982) *J. Biol. Chem.* **257,** 9684–9691.
4. Nowell, J. & Quaranta, V. (1985) *J. Exp. Med.* **162,** 1371–1376.
5. Paulnock-King, D., Sizer, K. C., Freund, Y. R., Jones, P. P. & Parnes, J. R. (1985) *J. Immunol.* **135,** 632–636.
6. Long, E. O., Mach, B. & Accolla, R. S. (1984) *Immunogenetics* **19,** 349–353.
7. Quaranta, V., Majdic, O., Stingl, G., Liszka, K., Honigsmann, H. & Knapp, W. (1984) *J. Immunol.* **132,** 1900–1905.
8. Koch, N. & Harris, A. W. (1984) *J. Immunol.* **132,** 12–15.
9. Benacerraf, B. (1981) *Science* **212,** 1229–1238.
10. Bottazzo, G. F., Pujol-Borrell, R., Hanafusa, T. & Feldmann, M. (1983) *Lancet* **ii,** 1115–1119.
11. Bell, J. I., Denny, D. W. & McDevitt, H. O. (1985) *Immunol. Rev.* **84,** 51–72.
12. Claesson, L., Larhammar, D., Rask, L. & Peterson, P. A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7395–7399.
13. Strubin, M., Mach, B. & Long, E. O. (1984) *EMBO J.* **3,** 869–872.
14. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15,** 1157–1174.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
17. Jelinek, W. R. & Schmidt, C. W. (1982) *Annu. Rev. Biochem.* **51,** 813–844.
18. Reed, R. & Maniatis, T. (1985) *Cell* **41,** 95–105.
19. Hentschel, E., Irminger, J. C., Burcher, P. & Birnstiel, M. L. (1980) *Nature (London)* **285,** 147–151.

20. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
21. Yamamoto, K., Koch, N., Steinmetz, M. & Hammerling, G. J. (1985) *J. Immunol.* **134,** 3461–3467.
22. Kozak, M. (1981) *Nucleic Acids Res.* **9,** 5233–5252.
23. Yamamoto, K., Floyd-Smith, G., Francke, U., Koch, N., Lauer, W., Dobberstein, B., Schafer, R. & Hammerling, G. J. (1985) *Immunogenetics* **21,** 83–90.
24. Claesson-Welsh, L., Barker, P. E., Larhammar, D., Rask, L., Ruddle, F. H. & Peterson, P. A. (1984) *Immunogenetics* **20,** 89–100.
25. Hood, L., Kronenberg, M. & Hunkapillar, T. (1985) *Cell* **40,** 225–229.
26. Larhammar, D., Servenius, B., Rask, L. & Peterson, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1475–1479.
27. Okada, K., Boss, J. M., Prentice, H., Spies, T., Mengler, R., Auffray, C., Lillie, J., Grossberger, D. & Strominger, J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3410–3414.
28. Kelly, A. & Trowsdale, J. (1985) *Nucleic Acids Res.* **13,** 1607–1621.
29. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
30. Edlund, T., Walker, M. D., Barr, P. J. & Rutter, W. J. (1985) *Science* **230,** 912–916.
31. Gillies, S. D., Folsom, V. & Tonegawa, S. (1984) *Nature (London)* **310,** 594–597.
32. Folsom, V., Gold, D. P., White, J., Marrack, P., Kappler, J. & Tonegawa, S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2045–2049.
33. Kudo, J., Chao, L., Narni, F. & Saunders, G. F. (1985) *Nucleic Acids. Res.* **13,** 8827–8841.