Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin γ chains

(conserved hydrophobic segment/intervening sequence-mediated domain transfer)

YURIKO YAMAWAKI-KATAOKA*, SUMIKO NAKAI*, TAKASHI MIYATA†, AND TASUKU HONJO*

*Department of Genetics, Osaka University Medical School, Nakanoshima, Kita-ku, Osaka 530, Japan; and †Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Communicated by Motoo Kimura, December 28, 1981

ABSTRACT The nucleotide sequences of the exons encoding membrane-bound IgG1, IgG2a, and IgG2b of mouse were determined and compared with the sequence of membrane-bound IgM. The sequences indicate that membrane-bound γ chains bear an additional 71 residues at the COOH termini, including 17-residue acidic, 26-residue hydrophobic, and 28-residue hydrophilic portions. The hydrophobic portion, which seems to be anchored in the lipid bilayer of the membrane, is highly conserved among membrane-bound μ and γ chains. We propose the presence of a membrane protein that recognizes the conserved hydrophobic segment and anchors the membrane-bound immunoglobulin. Comparison of the nucleotide sequences revealed another example of intervening sequence-mediated domain exchange around one of the membrane exons.

Immunoglobulin molecules such as IgM, IgG, and IgA exist in two different forms, as a surface antigen receptor and as a secreted antibody (1–10). The membrane-bound forms of the heavy (H) chains— μ_m , γ_m , and α_m —are larger than the secreted forms of the heavy chains— μ_s , γ_s , and α_s . Recent analyses of cloned immunoglobulin genes demonstrated that mRNAs encoding membrane-bound and secreted forms of the μ chain are transcribed from the same C_μ gene, and the two forms of the γ_{2a} chain are transcribed from the same $C_{\gamma_{2a}}$ gene (3–6, 9, 10). (C indicates the gene for the constant region of the heavy chain.) There are two membrane exons in the 3' flanking regions of the C_μ and $C_{\gamma_{2a}}$ genes. The alternative pathways of RNA splicing give rise to two mRNAs encoding separately the membrane-bound and secreted forms of the heavy chains.

The amino acid as well as nucleotide sequence of the μ_m chain revealed a 26-residue hydrophobic segment at the COOH terminus, which seems to be anchored in the lipid bilayer of the membrane (3, 4). Previously, we showed by heteroduplex analyses that the nucleotide sequences of the membrane exons are conserved among four C_γ subclass genes (9, 11).

We have now determined the nucleotide sequences of the membrane exons of the $C_{\gamma_{2a}}$, $C_{\gamma_{2b}}$, and C_{γ_1} genes; the sequences clearly demonstrate that the highly conserved 26-residue hydrophobic segments are also present in these C_γ genes. Comparison of the nucleotide sequences of the membrane exons and neighboring introns revealed that not only the M1 exon but also about 160 base pairs (bp) of the intron immediately 5' to the M1 exon are highly conserved between the C_{γ_1} and $C_{\gamma_{2b}}$ (or $C_{\gamma_{2a}}$) genes. Such sequence homology at limited portions of a gene indicates that another intervening sequence (IVS)-mediated domain exchange took place around the M1 exon, in addition to the one proposed around the CH1 exon (12, 13), between the C_{γ_1} and $C_{\gamma_{2b}}$ genes.

MATERIALS AND METHODS

Materials. Recombinant DNA clones that contain membrane exons of mouse immunoglobulin genes were previously obtained and characterized: IgH2 bearing the C_{γ_1} gene (14), Ig γ_{2a} -9 bearing the $C_{\gamma_{2a}}$ gene (15), and IgH22 bearing the $C_{\gamma_{2b}}$ gene (16). All the clones were obtained from BALB/c mouse DNA. An *Eco*RI fragment [4.9 kilobases (kb)] of Ig γ_{2a} -9 was inserted into pBR322 and subcloned. The insert (6.6 kb) of IgH22 was subcloned in pBR322 previously (17). Phage and plasmid DNAs were prepared as described (14, 17). Sources of restriction enzymes were as described previously except that *Hind*III, *Hinc*II, *Pst* I, and *Acc* I were obtained from Takara Shuzo Co. (Kyoto, Japan). [γ - 32 P]ATP was purchased from the Radiochemical Centre (Amersham, England).

Methods. Nucleotide sequence was determined according to Maxam and Gilbert (18). Thin gels (0.5 mm) were used as described (14). Nucleotide sequences of three γ subclass genes were aligned by using a computer program (12). The extent of the nucleotide sequence divergence was calculated as described (13).

RESULTS AND DISCUSSION

Nucleotide Sequences of C_γ Membrane Exons. Locations of membrane exons in the C_γ genes had been determined by R-loop and heteroduplex studies (9). Restriction enzyme cleavage maps around membrane exons of the C_{γ_1} , $C_{\gamma_{2a}}$, and $C_{\gamma_{2b}}$ genes were constructed by conventional procedures. The ranges and restriction sites used to determine the nucleotide sequences of the membrane exons are shown in Fig. 1.

The nucleotide sequences of the membrane exons of the C_{γ_1} , $C_{\gamma_{2a}}$, and $C_{\gamma_{2b}}$ genes were determined, compared, and aligned to maximize homology as shown in Fig. 2. We looked for an open coding frame that can be translated into the amino acid sequence homologous to that in the C_μ membrane exon. Such a region is shown by amino acid residues predicted from DNA sequences of the $C_{\gamma_{2a}}$ gene. The precise splice sites as indicated by vertical lines in Fig. 2 were determined by comparison to the nucleotide sequence of the cDNA coding for the γ_{1m} chain (19). The amino acid as well as nucleotide sequences of three C_γ genes are so homologous to each other that the M1 exons of the $C_{\gamma_{2a}}$, $C_{\gamma_{2b}}$, and C_{γ_1} genes were unequivocally identified. The M2 exons of the $C_{\gamma_{2a}}$ and $C_{\gamma_{2b}}$ genes were also easily assigned by comparison with the γ_1 cDNA sequence (19). Although the M2 exon of the C_μ gene encodes only two amino acid residues, the M2 exons of the $C_{\gamma_{2a}}$ and $C_{\gamma_{2b}}$ genes encode 27 amino acid residues. Because the M2 exon is about 1,000 bp

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: subscripts m and s, membrane-bound and secreted forms of immunoglobulin heavy chains, respectively; IVS, intervening sequence; bp, base pair(s); kb, kilobase(s).

The Membrane Domain Is a Common Structure of the Immunoglobulin C_H Gene. The membrane exons of the C_{γ1}, C_{γ2b}, and C_{γ2a} genes, which were indicated by R-loop and hetero-

[illegible]

duplex analyses (9), were unequivocally identified by nucleotide sequence analysis. It is very likely, therefore, that the conserved regions in the 3' flanking region of the $C_{\gamma 3}$ gene also contain the membrane exons. We have also identified the membrane domains of the C_{δ} gene by R-loop analysis and Southern blot hybridization (unpublished data). The membrane exons of the C_{μ} gene were identified recently by nucleotide sequence determination (3, 4). The membrane exons of the C_{δ} gene seem to be present as indicated by R-loop analyses (20). The mRNA encoding the α_m chain was recently identified (8). We have also shown by Southern blot analyses that the 3' flanking region of the C_{α} gene has a restriction fragment homologous to the membrane exons of the $C_{\gamma 1}$ gene (unpublished data). Taken together, these results indicate that all the C_H genes have membrane exons that are transcribed and spliced to form mRNAs encoding membrane-bound immunoglobulin heavy chains.

Amino Acid Sequences of Transmembrane Segments Are Conserved. The amino acid sequences of the membrane domains of the $C_{\gamma 2a}$, $C_{\gamma 2b}$, $C_{\gamma 1}$, and C_{μ} genes were compared as shown in Fig. 3. The membrane domains were separated into three parts as suggested by Rogers *et al.* (3)—i.e., the acidic, hydrophobic, and hydrophilic parts, which seem to be located in the outside of the cell, the intramembrane layer, and the inside of the cell, respectively. It is clear that the hydrophobic part is the most conserved; 1 divergence out of 26 residues among three C_{γ} genes and 9 differences out of 26 residues between the C_{μ} and any one of the C_{γ} genes.

This conservation of the amino acid sequence in the transmembrane segment is more striking than that in any other part of the C_H gene. When we compared the amino acid sequence of each domain of the $C_{\gamma 1}$, $C_{\gamma 2b}$, and $C_{\gamma 2a}$ genes with that of the C_{μ} gene, we found that the homology ranged between 22% and 37%, whereas the homology of the membrane domains of the C_{μ} and C_{γ} genes is about 50% (Table 1). The results indicate that the transmembrane segment is under a strong constraint of selection pressure. The stronger constraint on the membrane exon was also confirmed by comparison of human and mouse μ chains. The amino acid sequences of the M1 exons are identical in these two species (22), whereas those of other secreted domains are calculated to be 48–78% conserved on the basis of the published sequences (21, 22).

The 26-residue hydrophobic region can form α -helices, which require 3.6 residues per turn. A proper alignment of the hydrophobic residues to form α -helix can bring the diverged residues to two clustered patches as shown in Fig. 4. The other

FIG. 3. Comparison of amino acid sequences of IgM and IgG membrane domains. The one-letter code is given in the legend of Fig. 2. Amino acid sequences of membrane domains of the μ_m (3), $\gamma 1_m$ (this paper; ref. 19), $\gamma 2a_m$ (this paper), and $\gamma 2b_m$ (this paper) chains are presented. Homologous residues are boxed. A, B, and C indicate acidic, hydrophobic, and hydrophilic portions, respectively.

Table 1. Divergence of amino acid sequences between the μ and γ chains of mouse*

Heavy chains compared	C_H1	$C_H2^{\dagger}$	C_H3	M
μ and $\gamma 1$	0.76	0.75	0.63	0.47
μ and $\gamma 2a$	0.78	0.75	0.68	0.47
μ and $\gamma 2b$	0.76	0.73	0.63	0.47

* Amino acid residues predicted from nucleotide sequences (21) were compared. Numbers are difference per residue.

[†] The C_H2 domain of the γ chain was compared with the C_H3 domain of the μ chain.

side of the helix is, therefore, extremely conserved among the C_{μ} and C_{γ} genes. The striking conservation of the hydrophobic segment in the three-dimensional structure leads us to propose the presence of a receptor protein in the membrane that anchors the μ_m and γ_m chains. When antigens bind to surface receptor immunoglobulin molecules, either IgM or IgG, a similar signal may be transmitted through the membrane anchor protein to proliferate B lymphocytes. The difference in the lengths of the intracellular portions of the μ_m and γ_m chains may indicate that other qualitatively distinct types of signals may also be transmitted by different surface receptors, IgM and IgG.

Comparison of Nucleotide Sequences of Membrane Exons of C_{γ} Genes. We have compared the nucleotide sequences of membrane exons and their surrounding regions of the $C_{\gamma 1}$, $C_{\gamma 2b}$, and $C_{\gamma 2a}$ genes. The sequences were aligned with use of a computer program (12) as shown in Fig. 2. The divergence of the coding region was calculated separately for synonymous (K_s) and

FIG. 4. Putative structure of transmembrane segment. (A) A putative α -helical structure of the γ_m chain is shown. Residues different from those in the μ_m chain are boxed. Oblique lines indicate a unit of one turn (3.6 residues). (B) Hypothetical membrane protein that anchors membrane-bound immunoglobulin. Hatched regions indicate diverged residues.

amino acid-substituting (K_a) sites as shown in Table 2. It should be noted that nucleotide divergence at the synonymous sites of the coding regions is free from the selective constraint operating at the protein level. The divergence of the noncoding region (K_n) was calculated by comparing every nucleotide position.

Because K_n and K_s values are free from functional constraint, they are directly comparable (13). It is clear that the K_s and K_n values of the M1 exon and of the 3' portion of IVS 4 indicate that these sequences are much more conserved between the $C_{\gamma 1}$ and $C_{\gamma 2b}$ (or $C_{\gamma 2a}$) genes than any other parts are. The divergence values (0.36–0.51) in the less conserved regions (the 5' part of IVS 4, the IVS 5, and the M2 exons) are roughly equivalent to those (0.37–0.53) in the exons and introns encoding secreted domains (13). Because it is difficult to explain such varied divergence among different segments of the C_γ gene by selective constraint, the homology region might have been generated by the exchange of the DNA segment due to relatively recent recombinational events between C_γ subclass genes as proposed before (12). We do not know whether the recombination took place before or after the segregation of the $C_{\gamma 2a}$ and $C_{\gamma 2b}$ genes, which seems to be the latest event during the evolution of the C_γ subclass genes. If it occurred afterwards, the ancestral $C_{\gamma 2b}$ gene is a more likely partner of the $C_{\gamma 1}$ gene because the $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes are located next to each other (15).

The recombination required for such domain exchange is either double unequal crossing-over or gene conversion (24). Because the sequence divergence changes drastically at the middle of IVS 4, one site of the recombination seems to be located within IVS 4. The precise crossing-over point was estimated by comparison of every 50-bp segment of IVS 4 along the gene sequence. The 3' recombination site is apparently somewhere around the junction of the M1 exon and IVS 5. Because the M1 exon is significantly more conserved than the 3' portion of IVS 4, there might have been another recombination at the junction of the M1 exon and IVS 4.

The K_n and K_s values of the $C_{\gamma 2b}$ and $C_{\gamma 2a}$ genes are almost constant throughout the whole region whose sequence has been

FIG. 5. Schematic representation of IVS-mediated domain exchange between ancestors of $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes. Wider rectangles represent exons. Filled rectangles are homologous regions that were exchanged between the two genes.

determined. The values are equivalent to those of the most conserved regions (IVS 2, CH 2, and IVS 3) in the secreted exons. The results are consistent with models of the mouse C_γ gene evolution that were proposed previously (13). As expected from amino acid sequence comparison, the K_a value of the M1 exon is more conserved than that of any other exon (12, 13).

Two IVS-Mediated Domain Exchanges Between $C_{\gamma 1}$ and $C_{\gamma 2b}$ Genes. We have already shown that a contiguous DNA segment of the exon coding for the C_H1 domain and the 5' portion of IVS 1 must have undergone a recombinational exchange of genetic information between the ancestors of $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes (12). We have now shown that another DNA segment containing the M1 exon and the 3' portion of IVS 4 was also exchanged during evolution of the C_γ genes. As schematically shown in Fig. 5, the ancestral $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes have exchanged at least two segments of the gene, probably within a short interval because the K_n values of the two conserved segments were almost identical (0.26 and 0.23).

Since we reported the IVS-mediated domain exchange in the $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes of mouse (12), similar observations have been made in several other gene families of the eukaryote system (24–27). It is likely that IVS-mediated domain exchange is a universal phenomenon, probably characteristic of clustered families of genes.

At the time of submission of this manuscript the nucleotide sequences of membrane exons of the $C_{\gamma 2b}$ and $C_{\gamma 1}$ genes were published independently (23).

We thank Dr. Y. Kyogoku of Osaka University for valuable suggestions about the structure of the transmembrane segment, Drs. B. Tyler, O. Bernard, and J. Adams for sending unpublished data, Ms. F. Oguni for preparation of the manuscript, and Ms. Y. Sakagami and Ms. S. Nishida for their excellent technical assistance. This investigation was supported by grants from the Ministry of Education, Science and Culture of Japan, from the Toray Science Foundation, from the Mitsubishi Foundation, and from the Naito Foundation.

Table 2. Nucleotide sequence differences in membrane exons and introns of $C_{\gamma 1}$, $C_{\gamma 2b}$, and $C_{\gamma 2a}$ genes

		Comparison		
Region in which C_γ genes are compared*	Nucleotide position†	$C_{\gamma 2a}$ and $C_{\gamma 2b}$	$C_{\gamma 2b}$ and $C_{\gamma 1}$	$C_{\gamma 2a}$ and $C_{\gamma 1}$
		K_n or K_s		
IVS 4				
5' portion	8–150	0.024	0.515	0.507
3' portion	151–313	0.033	0.271	0.265
M1 exon	314–445	0.032	0.128	0.096
IVS 5	446–962	0.072	0.476	0.448
M2 exon				
Coding region	963–1,043	0.057	0.370	0.363
3' UT	1,047–1,077	0	0.410	0.410
		K_a		
M1 exon	314–445	0.02	0.06	0.06
M2 exon	963–1,046	0.016	0.134	0.118
M1 + M2 exons	314–1,046	0.018	0.088	0.082

* IVS sequences are numbered from 5' to 3' in the whole C_γ gene. UT, untranslated sequence.

† Nucleotide position shown in Fig. 2. The discontinuous point in IVS 4 was determined by calculation of divergence of every 50-bp segment starting within the sequence. Divergence values were calculated from published sequences (19, 23).

- Williams, P. B., Kubo, R. T. & Grey, H. M. (1978) *J. Immunol.* **121**, 2435–2439.
- Kehry, M., Ewald, S., Douglas, R., Sibley, C., Raschke, W., Fambrough, D. & Hood, L. (1980) *Cell* **21**, 393–406.
- Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. & Wall, R. (1980) *Cell* **20**, 303–312.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980) *Cell* **20**, 313–319.
- Singer, P. A., Singer, H. H. & Williamson, A. R. (1980) *Nature (London)* **285**, 294–300.
- Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., Koshland, M. & Baltimore, D. (1980) *Cell* **20**, 293–301.
- Oi, V. T., Bryan, V. M., Herzenberg, L. A. & Herzenberg, L. A. (1980) *J. Exp. Med.* **151**, 1260–1274.
- Kikutani, H., Sitia, R., Good, R. A. & Stavnezer, J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6436–6440.
- Nakai, S., Oi, V. T., Herzenberg, L. A., Yamagishi, H. & Honjo, T. (1982) *Biomed. Res.* **3**, 37–45.
- Tyler, B. M., Cowman, A. F., Adams, J. M. & Harris, A. W. (1981) *Nature (London)* **293**, 406–408.
- Takahashi, N., Shimizu, A., Obata, M., Nishida, Y., Nakai, S., Nikaido, T., Kataoka, T., Yamawaki-Kataoka, Y., Yaoita, Y., Ishida, N. & Honjo, T. (1981) in *Immunoglobulin Idiotypes*,

- ICN-UCLA Symposia on Molecular and Cellular Biology, eds. Janeway, C., Sercarz, E. E. & Wigzell, H. (Academic, New York), Vol. 20, pp. 123–134.
12. Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2143–2147.
 13. Yamawaki-Kataoka, Y., Miyata, T. & Honjo, T. (1981) *Nucleic Acids Res.* **9**, 1365–1381.
 14. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
 15. Shimizu, A., Takahashi, N., Yamawaki-Kataoka, Y., Nishida, Y., Kataoka, T. & Honjo, T. (1981) *Nature (London)* **289**, 149–153.
 16. Kataoka, T., Yamawaki-Kataoka, Y., Yamagishi, H. & Honjo, T. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4240–4244.
 17. Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M. & Honjo, T. (1980) *Nature (London)* **283**, 786–789.
 18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
 19. Tyler, B. M., Cowman, A. F., Gerondakis, S. D., Adams, J. M. & Bernard, O. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2008–2012.
 20. Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W. & Tonegawa, S. (1981) *Cell* **24**, 353–365.
 21. Kawakami, T., Takahashi, N. & Honjo, T. (1980) *Nucleic Acids Res.* **8**, 3933–3945.
 22. Rabbitts, T. H., Forster, A. & Milstein, C. P. (1981) *Nucleic Acids Res.* **9**, 4509–4524.
 23. Rogers, J., Choi, E., Souza, L., Carter, C., Word, C., Kuehl, M., Eisenberg, D. & Wall, R. (1981) *Cell* **26**, 19–27.
 24. Baltimore, D. (1981) *Cell* **24**, 592–594.
 25. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
 26. Liebhhaber, S. A., Goossens, M. & Kan, Y. W. (1981) *Nature (London)* **290**, 26–29.
 27. Miyata, T. & Yasunaga, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 450–453.