Proc. Natl. Acad. Sci. USA
Vol. 78, No. 2, pp. 785–789, February 1981
Biochemistry

Complete amino acid sequence of the Fc region of a human δ chain

(immunoglobulin D/internal homology/domain origin/evolution)

TOMOTAKA SHINODA*, NOBUHIRO TAKAHASHI*, TATSUNORI TAKAYASU*, TSUNEO OKUYAMA*, AND AKIRA SHIMIZU†

*Department of Chemistry, Tokyo Metropolitan University, Setagaya-ku, Tokyo 158, Japan; and †Department of Medicine, University of Osaka Medical School, Osaka 553, Japan

Communicated by Frank W. Putnam, October 20, 1980

ABSTRACT **The complete amino acid sequence of an Fc-like fragment designated $Fc_\delta(t)$ and obtained by limited proteolysis with trypsin of an intact myeloma IgD protein (NIG-65) has been determined. The fragment contains 226 amino acid residues and has a molecular weight of 32,000 per monomeric unit. It has three glucosamine oligosaccharides at asparagine residues 68, 159, and 210. Of these, glucosamine-159 is characteristic of the δ chain and has no counterpart position in any of the other classes. On the other hand, glucosamine-68 is shared by γ, μ, and ε, and glucosamine-210 is shared by α and μ. Although the $Fc_\delta(t)$ has the common framework structure of immunoglobulins, its sequence has many individual characteristics when its two domains are compared separately with the counterpart domain of other heavy chains. Such comparison has shown that the two Fc domains of the δ chain should be placed in an independent branch in topology; for all the other classes, the Fc domains are paired well with their counterparts. The comparison has also shown that there are three prominent gaps by which each domain can be divided into two homologous halves. For each class of immunoglobulin, a moderate degree of internal homology exists between the first half and the second half of each domain of the Fc, suggesting that the primordial gene may have coded for a unit about the size of a half domain. Based on this observation together with sequence comparisons, a possible genetic mechanism is proposed for the origin and evolution of the genes for immunoglobulin domains.**

Since the discovery of IgD as a fourth major class of human immunoglobulins (2), studies have been undertaken to elucidate not only its biological functions but also its physicochemical and chemical characteristics. Although preliminary studies have shown some characteristics of IgD (3), detailed analyses have been hindered mainly because of (*i*) the low incidence of IgD myeloma, and the relatively low concentration of IgD in such plasma, (*ii*) the extreme susceptibility of IgD to various proteolytic enzymes (4), which hinders preparation of intact IgD in quantity, and (*iii*) great charge heterogeneity, mainly due to high carbohydrate content (12–15%) which causes difficulty in purification of IgD. However, because IgD has been demonstrated to be a major class of surface immunoglobulins on human B lymphocytes (5) as well as on B cells of animals (6), studies have been carried out to learn whether the IgD molecule plays an essential role during cell differentiation, including antibody synthesis. Accumulating data have shown that IgD is closely involved in some cellular events, although its key role is still to be clarified (7).

Little progress has been reported on structural studies of the IgD molecule, but preliminary work has unveiled some of its characteristic features (8, 9). These data show that, unlike other classes, IgD has only one disulfide bridge between heavy chains, and its heavy chain (δ) is likely to have either a five-domain structure (10) or a somewhat extended hinge structure (4). Some sequence studies were also done on the sites of spontaneous cleavage with various proteolytic enzymes and an intact human myeloma IgD (9). Recently, some amino acid sequences have been determined for an Fc_δ fragment generated by short-time trypsin digestion (11) which revealed some characteristics of the δ chain. More recently, a four-domain structure for human IgD molecules and processes of its fragmentation with proteolytic enzymes have been reported by Lin and Putnam (12).

Despite great interest in the structure of the IgD molecule in order to elucidate its relationship to cellular events, no relevant data have been available so far. In this context, we undertook structural analysis of a human myeloma IgD (13) and now have completed the amino acid sequence of the entire $Fc_\delta(t)$ fragment obtained by limited proteolysis of the intact IgD with trypsin.

MATERIALS AND METHODS

For purification of IgD, 2 units of frozen plasma (NIG-65, approximately 600 ml) were melted in 400 ml of 0.15 M NaCl/0.12% NaN_3, pH 7.2/25 mM ε-aminocaproic acid/10 mM $CaCl_2$. Saturated $(NH_4)_2SO_4$ solution/0.05% NaN_3/10 mM ε-aminocaproic acid, pH 7.2, was added, and the precipitate obtained between 1.5 and 1.8 M $(NH_4)_2SO_4$ was collected. The precipitate was dissolved in 10 mM phosphate/0.05% NaN_3, pH 8.0/10 mM ε-aminocaproic acid and dialyzed against the same buffer. The solution was applied to a DEAE-Sephadex A-50 column, and most of the IgD was eluted with the same buffer containing 80 mM NaCl. The IgD fraction was concentrated on a column of DEAE-Sepharose CL-6B and was purified further by gel filtration with a Bio-Gel A-5 m column in 0.1 M Tris·HCl/0.15 M NaCl/0.05% NaN_3, pH 8.0/10 mM ε-aminocaproic acid. Fractions were assayed by immunodiffusion and immunoelectrophoresis for immunoglobulins, α_1-acid glycoprotein, prealbumin, albumin, group-specific component, transferrin, fibrinogen, ceruloplasmin, and κ and λ Bence Jones proteins. The purity of IgD-positive fractions was also tested by polyacrylamide gel electrophoresis (4.5–7.5%) before and after treatment with sodium dodecyl sulfate and also by NH_2- and COOH-terminal analyses.

Purified IgD (1.2 g) in 60 ml of 10 mM phosphate/0.05% NaN_3, pH 8.0, was digested with 12 mg of trypsin (treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone, twice crystallized, Worthington) at 37°C for 2 min. The reaction was stopped by addition of 24 mg of lima bean trypsin inhibitor

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: $Fc_\delta(t)$; an Fc-like fragment of IgD resulting from limited proteolysis with trypsin; PAMs, accepted point mutations per 100 residues. Abbreviations for classes, fragments, regions, and domains of immunoglobulins are in accordance with official World Health Organization recommendations for human immunoglobulins (1).

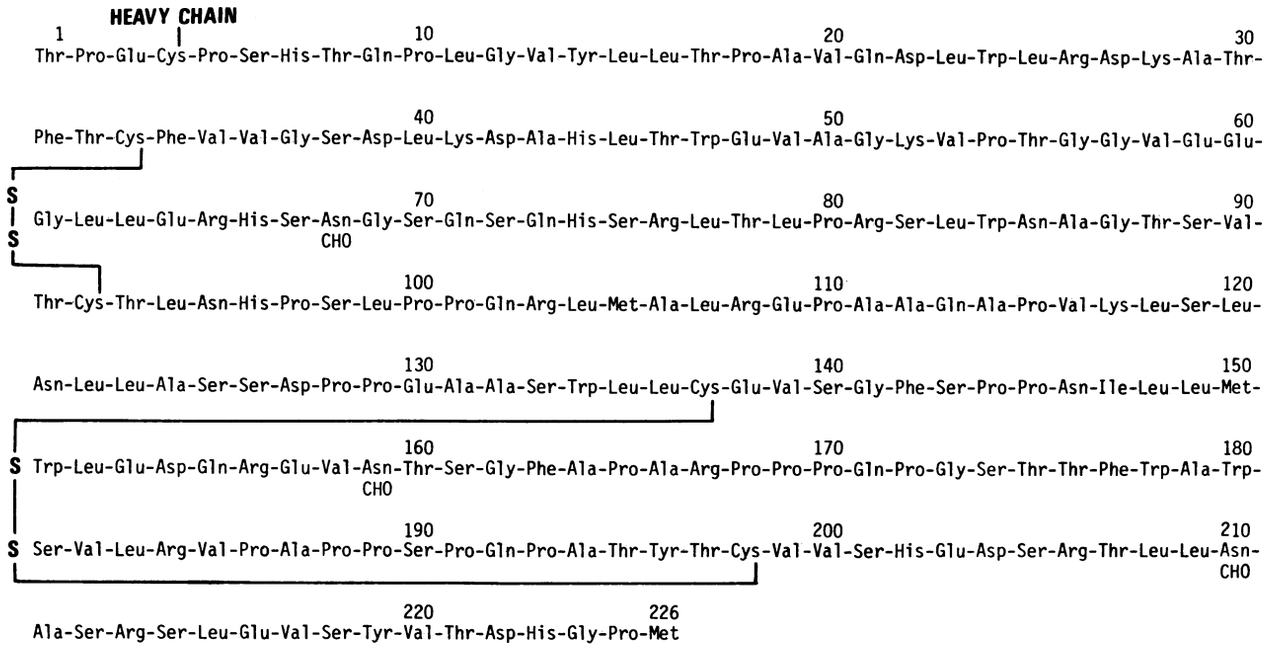

FIG. 1. Amino acid sequence of entire Fc region of the NIG-65 δ chain. The residues are numbered consecutively from Thr-1 to Met-226. This numbering will be changed as the complete sequence of the δ chain becomes available. A tentative alignment for disulfide bridges is given by the solid line; this is based on homology to other immunoglobulins. Glucosamine oligosaccharides (designated CHO) are indicated under the residues to which they are attached.

(twice crystallized, Worthington). The digestion mixture was fractionated into the Fab and Fc fragments by column chromatography on DEAE-Sepharose CL-6B (14). For reduction and aminoethylation, the purified $Fc_\delta(t)$ (420 mg) was dissolved in 20 ml of 1 M Tris·HCl, pH 8.6/6.8 M guanidine hydrochloride/2 mM EDTA and treated as described (14). Aminoethylated $Fc_\delta(t)$ (100 mg) was digested with 1.3 mg of trypsin at 37°C for 3 hr, and the digest was chromatographed on a column of DEAE-Sephadex A-25. Peptides were further purified by gel chromatography with Bio-Gel P-6 (14). Digestion of aminoethylated $Fc_\delta(t)$ (120 mg) was done with 4 mg of staphylococcal proteinase to obtain necessary overlaps for the tryptic peptides. Cyanogen bromide cleavage and purification of the fragments were carried out essentially as described (13).

The methods for molecular weight determination, for amino acid and carbohydrate analysis, for NH_2- and COOH-terminal end group determination, for digestion with carboxypeptidases A, B, and P, for hydrazinolysis of the $Fc_\delta(t)$ and of some peptides, and for sequence determination by manual Edman degradation have been described (14, 15). An overlap for residues 178–181 was done with the Beckman model 890B sequenator by courtesy of F. W. Putnam (Department of Biology, Indiana University, Bloomington, IN).

RESULTS AND DISCUSSION

Purity and Enzymatic Digestion of NIG-65 IgD. On polyacrylamide gel electrophoresis, the isolated myeloma protein NIG-65 showed a single relatively broad band accompanied by an additional faint component with a slightly larger mobility and averaging about 2% of the total material. The purified protein was tested with specific antisera at concentrations of 0.5–20.0 mg/ml and was found to react only with anti-δ and -λ antisera, indicating that it was homogeneous in this criterion. In all, approximately 8.4 g of purified IgD was obtained from 2 units of NIG-65 plasma by this procedure. The molecular weight of the major component of the purified NIG-65 was estimated to be between 168,000 and 172,000 (mean, 170,000). The minor contaminant in the purified IgD preparation reacted only with anti-δ antiserum, and its mean molecular weight was estimated to be 84,000.

Quantitative NH_2-terminal analysis of the intact IgD gave 1.86 mol of alanine (from the δ chain) and 1.82 mol of asparagine (from the λ chain) per 170,000 g. After reduction and alkylation, the IgD could be separated into heavy and light chains by gel filtration on a Bio-Gel A-5m column. The molecular weight of the heavy chain was found to be 63,000, whereas that of the light chain was 23,000. The δ chain contained approximately 480 residues, without correction for the destruction of certain amino acids dring acid hydrolysis. Virtually all the carbohydrates found in the purified NIG-65 IgD, including sialic acid, glucosamine, and galactosamine, were recovered from the δ chain. On the other hand, the λ chain contained 220 residues and had a homogeneous NH_2-terminal sequence of Asn-Leu-Met-Leu-. Its complete sequence has recently been determined, giving a new λ chain subgroup of $V_\lambda VI$ (14).

Amino Acid Sequence. The complete amino acid sequence, including assignments of all the amides and locations of three glucosamine oligosaccharides of the $Fc_\delta(t)$, of NIG-65-IgD is given in Fig. 1. The sequence is the same as that of the $Fc_\delta(t)$ of IgD protein WAH, which was determined independently by Lin and Putnam (16) using different approaches and techniques. The striking evidence that the $Fc_\delta(t)$ sequence is identical in the two δ chains does not necessarily exclude the possible existence of allotypes, subclasses, or isotypes of the δ chain.

The $Fc_\delta(t)$ fragment has 226 residues with a single homogeneous NH_2-terminal threonine as well as homogeneous COOH-terminal methionine, and it has the common framework structure shared by all the other immunoglobulin classes so far analyzed. That is, it has five half-cystine residues and they are distributed in a fashion similar to that in the other heavy chain classes. The half-cystine residues are cysteine-4, which is involved in the inter-heavy chain disulfide bridge and intra-chain bridges Cys-33/Cys-92 and Cys-137/Cys-198.

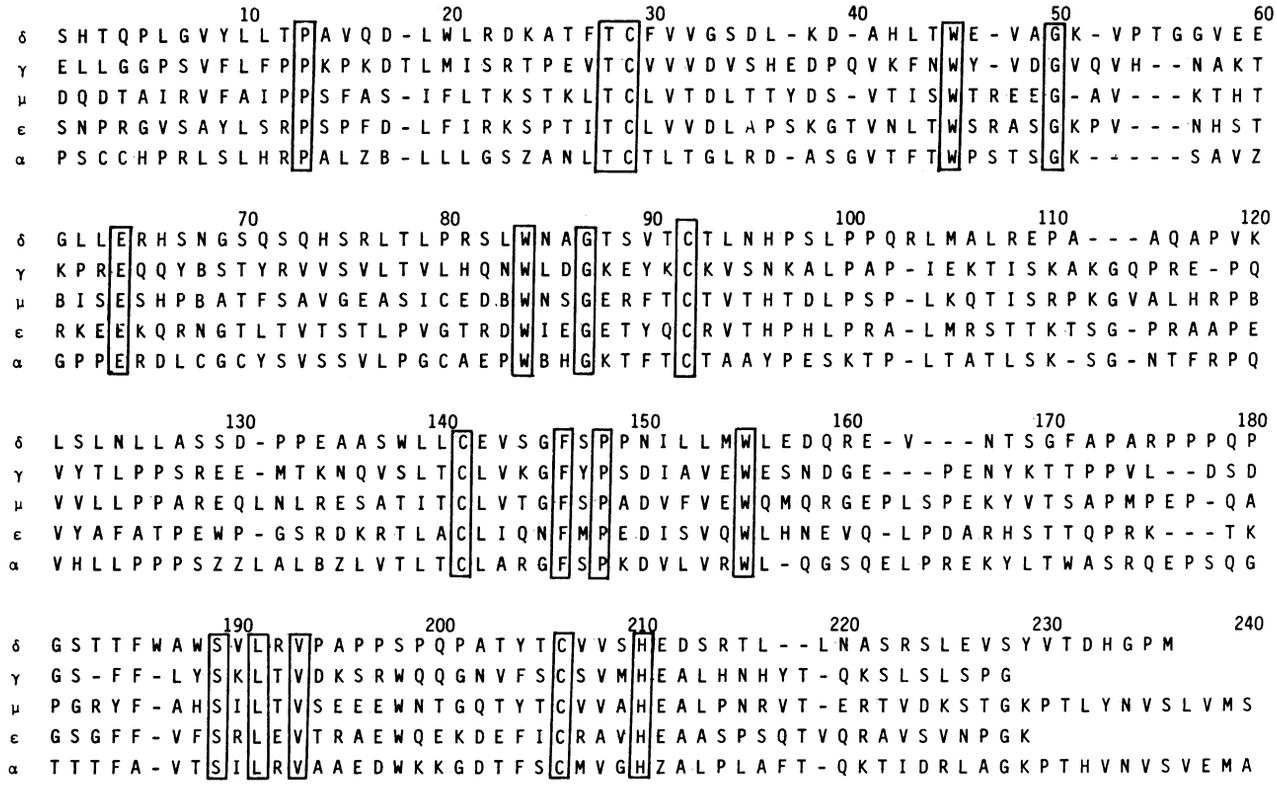

FIG. 2. Amino acid sequences [standard one-letter code (21)] of the Fc regions of human IgD, IgG1, IgM, IgE, and IgA1, denoted by δ, γ, μ, ε, and α, respectively. The sequence is not given for the hinge region or for the last seven residues of μ and α but only for the middle domain (the first domain of Fc—namely, $C_\delta 2$, $C_\gamma 2$, $C_\mu 3$, $C_\varepsilon 3$, and $C_\alpha 2$)—and the last domain (the second domain of Fc—namely, $C_\delta 3$, $C_\gamma 3$, $C_\mu 4$, $C_\varepsilon 4$, and $C_\alpha 3$). The sequence of each chain is given continuously from left to right moving down the figure, but the chains are aligned to show maximal symmetry of the domains and maximal homology of the sequences. Residues identical in all chains are boxed. The sequences are aligned to place all half-cystine residues and invariant tryptophan residues in register. Gaps have been inserted to maximize the homology. The first residue of the δ chain in the figure is equivalent to Ser-6 in Fig. 1; the first residues for γ, μ, ε, and α are glutamate-233, aspartate-340, serine-321, and proline-240, respectively. Sources of sequence data are ref. 17 for γ, ref. 18 for μ, ref. 19 for ε, and ref. 20 for α; the sequence and tentative numbering for the δ chain are from the present work.

In contrast to the γ_1 chain, the $Fc_\delta(t)$ contains three glucosamine oligosaccharides. The first one, at asparagine-68 (designated GlcN 1), is shared by the γ_1 chain and also by μ and ε chains. On the other hand, the second one, at asparagine-159 (designated GlcN 2), appears to be characteristic of the δ chain. It has no counterpart in any of the other classes. Although the last glucosamine oligosaccharide, at asparagine-210 (designated GlcN 3), is absent from the γ_1 chain, it has a counterpart in the COOH-terminal extra sequences of both μ and α chains. The latter two heavy chains have the same acceptor sequence, Asn-Val-Ser; but the δ chain has a different one, Asn-Ala-Ser. This interchange in the triplet seems to be fortuitous because there is little homology between δ and μ or between δ and α in the tail piece (about 26% in each case), whereas that between μ and α is about 58%. This suggests that there are small DNA segments that either have already been attached to the C_H3 domain exon or originally exist as discrete segments in introns and are joined with the C_H3 exon prior to expression into a functional domain gene.

It is a characteristic feature of the δ chain that, whereas the other chains have 1 to 3 proline residues in the loop of the second half of the C_H3 domain, there are 10 prolines in this loop in the δ chain, suggesting that they may cause a large conformational distortion in this domain. This might expose two nearby hydrophobic regions—namely, 147-Ile-Leu-Leu-Met-Trp-Leu-152 and 177-Phe-Trp-Ala-Trp-180—and this will enable the domain to interact with other components such as those of the cell membrane of B lymphocytes. This, together with an extreme susceptibility to proteolytic enzymes, might enable IgD to play an essential role as a regulatory molecule in cell differentiation during antibody synthesis. It is of interest that, when the sequence of residues 165 through 172 is compared with that of residues 186 through 193, 6 of the 8 are identical. This strongly suggests that these sequences have arisen by partial duplication during evolution.

Some Implications of Sequence Homology. The $Fc_\delta(t)$ sequence is compared with the counterparts of IgG1 (17), IgM (18), IgE (19), and IgA (20) in Fig. 2. Many gaps are introduced to obtain maximal symmetry of the domains and maximal homology of the sequences among the chains. The hinge regions are omitted from the comparisons because they have least identity and that of the δ chain has not yet been established. Furthermore, the hinge region is encoded separately in mouse γ and α chains by discrete exons (22). There are 18 loci where all the chains share common residues. These loci are distributed nonuniformly within the loops of the domains, and none is present in the interdomain area. This suggests that they are highly conserved throughout evolution, probably because they are related to the basic framework structure and also to the functions of immunoglobulins at the molecular level. As expected, the number of residues common to all human heavy chains of known sequence (γ1, α, μ, and ε) is diminished by addition of the sequence of new chain. When the δ chain is compared to the others it is striking that the number of loci at which all the chains have a common residue is diminished from 30 to 18. There also are loci where δ is identical to three of the four other

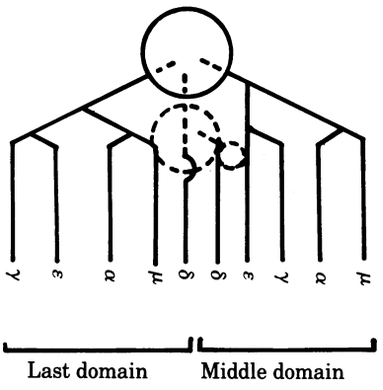

FIG. 3. Phylogenetic tree of constant region domains of human immunoglobulins. This was derived from a matrix (not shown) based on the sequence alignments in Fig. 2 by using the procedure of Dayhoff (23). Although the constant region of the δ, γ, and α chains has three domains, whereas the constant region of the μ and ε chains has four domains, the Fc fragment of all five chains has only two domains—the last two in each chain. Here (and in the text) the term "middle" domain refers to the first Fc domain of each chain ($C_\mu 3$, $C_\alpha 2$, $C_\gamma 2$, $C_\varepsilon 3$, and $C_\delta 2$, respectively) and "last" refers to $C_\mu 4$, $C_\alpha 3$, $C_\varepsilon 4$, $C_\gamma 3$, and $C_\delta 3$. Except for δ, the middle and the last domains form two major branches. Except for those of δ, the topological connections of the middle and the last domains will always be γ with ε and μ with α. The two domains of δ seem to form independent branches by themselves, but a significant correlation is seen between the middle domain of δ and the counterpart of ε and also between it and the last domain of μ. On the other hand, the last domain of δ appears to have no counterpart in this scheme. The two domains of δ are separated from each other by 210 accepted point mutations per 100 residues [PAMs (23)]. The corresponding values for γ, μ, ε, and α are 208, 192, 278, and 235 PAMs, respectively. Because the δ chain is placed in the middle, topological arrangements for μ/α and γ/ε pairs are interchanged in the figure for both domains. The branch lengths shown do not correlate to an actual distance in PAMs in each pair but indicate only the topological connections. Dotted circles between branches ($C_\delta 2$ and $C_\varepsilon 3$, and $C_\delta 2$ and $C_\mu 4$) show high correlation among them.

chains but different from the fourth. At these loci, the α chain is different from the other four in seven instances, μ chain in five, ε in two, and γ in one.

By comparison of the sequences in Fig. 2, a phylogenetic tree (Fig. 3) was constructed according to the procedure of Dayhoff (23). In this case, we compared each domain separately because homology between any pair of the counterpart domains varied considerably and because each domain might have different evolutionary history (that is, it may reflect different amounts and rates of change). In general, except for the δ chain, we obtained relatively good correlations in each pair between the counterpart domains of the C region. For instance, for the γ chain, the highest homology for the middle domain of the C region was obtained with the third domains of ε and μ chains, and that for the last domain of γ was with the last domains of μ, ε, and α. Degrees of homology in these pairs are in the range 34–36%. In these comparisons, highest homology was always obtained with pairs of μ/α and ε/γ for both the domains. Thus, they are placed in pairs in the phylogenetic tree. On the other hand, when we compared the two domains of Fc_δ with those of the Fc region of other chains, the figures were different. As indicated by Fig. 3, the middle domain of the δ chain ($C_\delta 2$) has highest homology with the third domain of ε ($C_\varepsilon 3$), but $C_\delta 2$ is more homologous with the last domains of both the γ chain ($C_\gamma 3$) and the μ chain ($C_\mu 4$). Furthermore, $C_\delta 2$ is almost equally homologous to $C_\alpha 2$ and $C_\alpha 3$. These facts strongly suggest that $C_\delta 2$ has no counterpart either in γ or μ. Furthermore, the last domain of the δ chain ($C_\delta 3$) has little homology with the last domains of the other chains. Therefore, it is reasonable to assume that both $C_\delta 2$ and $C_\delta 3$ diverged independently from their counterparts in the other chains. Thus, we placed these domains of δ as a separate branch in the tree (Fig. 3). The distance between the two domains is approximately the same as those in each class.

Origin and Evolution of Immunoglobulin Constant Domains. On inspection of the sequences shown in Fig. 2, it is noted that there are three prominent gaps: (*i*) in the loop of the second domain involving residues 51–56, (*ii*) residues 111–118, just in the middle of the interdomain area, and (*iii*) residues 162–178, in the middle of the last domain loop. In this way, the Fc region of all chains is divided symmetrically into four segments of similar size. In other words, each domain can be divided into two nearly equal parts, the first half and the second half of the domain. Each segment has about 60 residues, and has one half-cystine residue nearly in the middle. On comparison of this half-domain sequence, there are moderate degrees of internal homology in each pair of the segments within the same domain (18–28%). These values are roughly the same as those observed between any pair of domains.

Although there are many plausible ways to interpret this finding, one can make an assumption that these gaps were introduced through genetic events that had taken place in the early history of evolution of a precursor immunoglobulin domain-coding gene. Such events might have required either fusion of two discrete DNA segments or random duplication of a half-domain-encoding DNA segment immediately adjacent to itself. Either of these mechanisms could have generated a new DNA segment, a primordial domain-encoding gene from which both the present-day light chain and heavy chain domain-encoding genes are evolved.

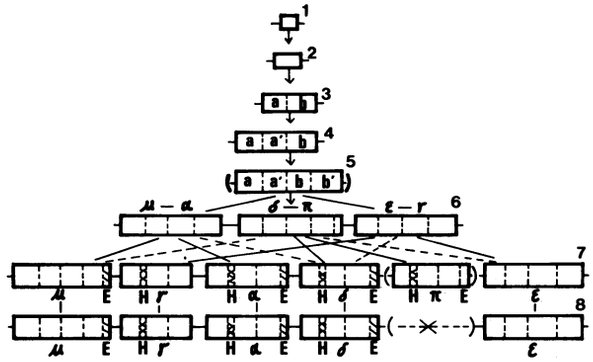

FIG. 4. Proposed model for the origin and evolution of constant domain genes of the heavy chains. The model consists of three major stages. (*i*) Steps 1 and 2 involve either duplication of element 1 (a half-domain-encoding DNA segment) immediately adjacent to itself, or fusion of two discrete DNA segments (24). The process could generate a primordial domain-encoding DNA segment. (*ii*) Steps 2–5 require successive duplications, from which basic four-domain DNA segment(s) could arise in which a' is a duplicate of a and b' is a duplicate of b. (*iii*) Steps 6–8 include various rearrangements in which each domain rearranges either independently or restrictedly with either of the domain-encoding DNA segments. Inefficient rearrangement as well as abortive rearrangement might be operative at a certain frequency in this stage. Furthermore, the hinge exon (designated *H*) and extra tail exons (designated *E*) which code for the extra tail sequences of μ and α and of δ could be introduced through the process. As an analogy to mouse heavy chain genes (22), certain lengths of intervening sequences might be present between domain-encoding DNA segments as well as on both sides of the H exon and probably also the E exon. An immunoglobulin gene designated π is an example for a putative gene(s) generated by inefficient or abortive combinations among domain-encoding DNA segments. The proposed gene order does not necessarily show the actual order in the human lymphocyte genome.

A proposed model to account for the mechanism is illustrated in Fig. 4. After the generation of the primordial domain-encoding gene, successive duplications could have occurred immediately adjacent to themselves, to result in a precursor immunoglobulin-encoding gene(s) with a four-domain structure (step 5 in Fig. 4). The next step requires stochastic rearrangement among the domain-encoding genes, in which each of them rearranged independently at a certain frequency. Besides many inefficient and abortive rearrangements, there should have to be some efficient rearrangements from which undifferentiated precursor genes for μ/α, δ/π and ε/γ evolved. During the process of rearrangements, γ, α, and δ genes independently deleted the second domain-encoding gene, giving a three-domain structure. The hinge was likely introduced into α, γ, and δ in place of the second (or extra) domain, or it could be a residual portion of a flanking sequence of a second domain-encoding DNA segment. An additional rearrangement that was probably restricted might have been involved in the next step, and it eventually could have generated present-day embryonic immunoglobulin genes. The model is applicable for all the immunoglobulin classes so far identified and might also be applicable for any yet to be discovered.

We thank Dr. F. W. Putnam for giving us an opportunity to have the necessary overlap for residues 178–181 and for his critical reading of the manuscript, S. Dorwin, J. Dwulet, E. Dodge, and J. Madison for technical assistance, and Dr. A. Yamada for general discussion. This work was supported by a grant-in-aid from the Ministry of Education of Japan and by research funds from the Ito Science Foundation.

1. World Health Organization (1972) *Biochemistry* **11,** 3311–3312.
2. Rowe, D. W. & Fahey, J. L. (1965) *J. Exp. Med.* **121,** 171–184.
3. Griffiths, R. W. & Gleich, G. J. (1972) *J. Biol. Chem.* **217,** 4543–4548.
4. Spiegelberg, H. H., Prahl, J. W. & Grey, H. M. (1970) *Biochemistry* **9,** 2115–2122.
5. Rowe, D. S., Hug, K., Forni, L. & Pernis, B. (1973) *J. Exp. Med.* **138,** 965–972.
6. Leslie, G. A. & Martin, L. N. (1978) *Contemp. Top. Mol. Immunol.* **7,** 1–49.
7. Vitetta, E. S. & Uhr, J. W. (1977) *Immunol. Rev.* **37,** 50–88.
8. Spiegelberg, H. H. (1977) *Immunol. Rev.* **37,** 1–24.
9. Jefferis, R. & Matthews, J. B. (1977) *Immunol. Rev.* **37,** 25–49.
10. Leslie, G. A., Clem, L. W. & Rowe, D. S. (1971) *Immunochemistry* **8,** 565–568.
11. Spiegelberg, H. H. (1975) *Nature (London)* **254,** 723–725.
12. Lin, L.-C. & Putnam, F. W. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6572–6576.
13. Shinoda, T., Takahashi, N., Takayasu, T. & Shimizu, A. (1979) *Proc. Jpn. Immunol. Soc.* **9,** 451 (abst.).
14. Takahashi, N., Takayasu, T., Isobe, T., Shinoda, T., Okuyama, T. & Shimizu, A. (1979) *J. Biochem.* **86,** 1523–1535.
15. Shinoda, T. (1973) *J. Biochem.* **73,** 417–431.
16. Lin, L.-C. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 504–508.
17. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. O., Rutishauser, U. & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. USA* **63,** 78–85.
18. Putnam, F. W., Florent, G., Paul, C., Shinoda, T. & Shimizu, A. (1973) *Science* **182,** 287–291.
19. Bennich, H. & Bahr-Lindstrom, H. (1974) in *Progress in Immunology II*, eds. Brent, L. & Holborow, J. (North-Holland, Amsterdam), Vol. 1, pp. 49–58.
20. Low, T. L. K., Liu, Y. S. V. & Putnam, F. W. (1976) *Science* **191,** 390–392.
21. IUPAC–IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem.* **243,** 3557–3559.
22. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277,** 627–633.
23. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5.
24. Shinoda, T. (1977) *Rinsho Kagaku* **13,** 844–851.