Proc. Natl. Acad. Sci. USA
Vol. 76, No. 1, pp. 452–456, January 1978
Immunology

Human heavy chain disease protein WIS: Implications for the organization of immunoglobulin genes

(intervening sequences/gene splicing/mutations/deletions)

EDWARD C. FRANKLIN, FRANCES PRELLI, AND BLAS FRANGIONE

Departments of Medicine and Pathology, Irvington House Institute, New York University Medical Center, 550 First Avenue, New York, New York 10016

Communicated by Henry G. Kunkel, October 10, 1978

ABSTRACT **Protein WIS is a human γ3 heavy (H) chain disease immunoglobulin variant whose amino acid sequence is most readily interpreted by postulating that three residues of the amino terminus are followed by a deletion of most of the variable (V_H) domain, which ends at the variable–constant (VC) joining region. Then there is a stretch of eight residues, three of which are unusual, while the other five have striking homology to the VC junction sequence. This is followed by a second deletion, which ends at the beginning of the quadruplicated hinge region. These findings are consistent with mutations resulting in deletions of most of the gene coding for the V region and C_H1 domain followed by splicing at the VC joining region and at the hinge. These structural features fit well the notion of genetic discontinuity between V and C genes and also suggest similar mechanisms of excision and splicing in the interdomain regions of the C gene of the heavy chain.**

Amino acid sequence studies of heavy chain disease (HCD) proteins and other immunoglobulin variants have demonstrated that most often these molecules contain deletions and that the deletions are not random but frequently involve immunoglobulin domains (1). In particular, in all but one of the γ HCD proteins for which degradation can be definitively excluded, the deletion encompasses part of the variable (V) region and all of the C_H1 domain of the constant (C) region, with resumption of normal sequence at the beginning of an unusual region known as the hinge, which contains the interchain disulfide bridges. The hinge varies for each class and subclass, is rich in cysteine and proline, has little homology with any heavy (H) chain domain, and can undergo duplications. Three γ1 myeloma proteins with internal deletions were recognized because of their dissociability under nonreducing conditions, and it was shown that only the hinge region from residues 216–230 is missing (2–4; L. A. Steiner, personal communication). Two other myeloma proteins with internal gaps in the light (L) chains (SAC and SM) have deletions of much of the variable region; the deletions end at points that may correspond to the VC joining region (5, 6).

The present report presents detailed amino acid sequence studies of a γ3 HCD protein WIS, which is of interest because the amino acid sequence is most consistent with the existence of an apparently normal amino-terminal sequence followed by two large deletions of V_H and C_H1, which are separated by a small stretch with striking homology to the VC joining region of the H chain. The second deletion ends at the beginning of the quadruplicated hinge (7). Though the precise mechanism for the generation of these deletions remains to be worked out, these findings are consistent with the possible existence of transcriptional units corresponding to immunoglobulin domains.

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

Protein WIS was isolated from the urine of a patient with HCD by precipitation with 60% saturated ammonium sulfate and purified by ion-exchange chromatography on DEAE-cellulose (Whatman DE-52). Immunoelectrophoresis and Ouchterlony double diffusion in agar were performed with antisera to γ chains, Fd, Fab, and Fc fragments, and light chains (8). Molecular weights were determined in 10% polyacrylamide gels containing 0.1% sodium dodecyl sulfate (9). Reduction was performed with 5 mM dithiothreitol and alkylation with iodo[^{14}C]acetic acid (0.7 Ci/mol; 1 Ci = 3.7×10^{10} becquerels) (10). Cyanogen bromide (CNBr) cleavage was carried out overnight at room temperature with a 5-fold excess (wt/wt) of CNBr to protein in 70% (wt/wt) formic acid (11). Reduced and alkylated CNBr fragments were separated by gel filtration on a (2 cm × 40 cm) Sephadex G-25 column in 35% formic acid. The hinge fragment was purified on Sephadex G-100 followed by isoelectric focusing in an LKB 810 (Bromma, Sweden) (110 ml) column, in 1% Ampholine (LKB) at pH 3.5–5 (12). Chemical typing and tryptic fingerprints were performed as described (10). Digestion by trypsin treated with L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone (Worthington) was carried out in 0.2M NH_4HCO_3, pH 8.2, for 1 hr at room temperature at an enzyme-to-substrate ratio of 1:100 (wt/wt). Digestion by Pronase (Sigma) of the completely reduced and alkylated protein was performed in 0.1M NH_4HCO_3, pH 8.2, for 3 hr at 37°C at an enzyme-to-substrate ratio of 1:50 (wt/wt). The digest was applied to a Dowex AG 50W-X2 (Bio-Rad) column in water to separate the peptides containing pyrrolidone carboxylic acid (<Glu, pyroglutamate) and carbohydrate (CHO) (13).

Digestion by calf liver pyroglutamate aminopeptidase (Boehringer Mannheim) of completely reduced and alkylated protein was done in 0.1M sodium phosphate buffer, pH 8.0, made 5 mM in dithiothreitol, 10 mM in Na_2 EDTA, and 5% (vol/vol) in glycerol, for 9 hr at 4°C followed by 14 hr at room temperature at a crude enzyme-to-substrate ratio of 1:10 (wt/wt) (active enzyme, 1:200) (14).

Isolation and purification of peptides and amino-terminal analyses by the dansyl chloride method were done as described (8). Amino acid sequences were determined automatically with a Beckman model 890C sequencer by the method of Edman and Begg. Amino acid residues were identified by gas/liquid chromatography performed with a 7620A Hewlett–Packard gas chromatograph, by thin-layer chromatography of phenylthiohydantoin derivatives, and by amino acid analysis using a Durrum D500 amino acid analyzer after hydrolysis with HCl (7). The position of carboxymethylcysteine (CmCys) was confirmed by measuring radioactivity.

Abbreviations: H chain, heavy chain of immunoglobulins; L chain, light chain; V region, variable region; C region, constant region; HCD, heavy chain disease; CHO, carbohydrate (oligosaccharide).

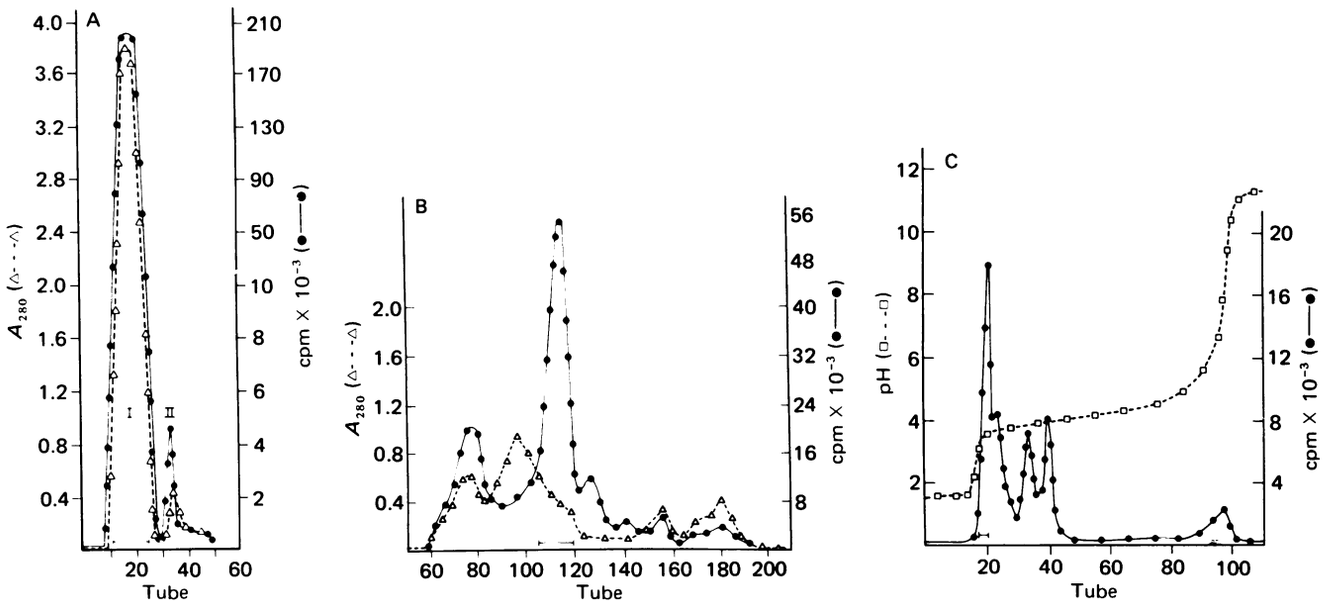

FIG. 1. Purification of CNBr fragments of protein WIS. (*A*) Chromatography on Sephadex G-25. Peak II is the amino terminus. (*B*) Chromatography on Sephadex G-100 of peak I from *A*. (*C*) The contents of tubes 105–120 from *B* were further purified by isoelectric focusing. Tubes 15–20 were used for further analyses.

RESULTS AND DISCUSSION

On sodium dodecyl sulfate/polyacrylamide gel electrophoresis, protein WIS migrated as a homogeneous band with a molecular weight of 85,000 in the unreduced state and 43,000 after reduction and alkylation. The protein was typed chemically and immunologically as belonging to the γ3 H chain subclass and found to contain the intact hinge with antisera specific for the hinge of γ3 (15). Fingerprint analysis demonstrated all the peptides characteristic of the Fc fragment, including the cysteine-containing peptides of the hinge. Amino-terminal analysis was negative. Because this was probably due to an amino-terminal <Glu, the protein was subjected to Pronase digestion. The composition of the three major peptides—Pr1: Glu,Met; Pr2: Cys,Asp,Thr,Glu,Gly,Val,CHO; Pr3: Cys,Asp,Thr, Glu_2,Gly,Val,Met,CHO—suggests that the amino-terminal sequence is <Glu-Met. Therefore, 250 mg of protein was completely reduced and alkylated, desalted on Sephadex G-25 to remove free iodoacetic acid, and then subjected to cleavage with CNBr. Separation on a column of Sephadex G-25 in 35% formic acid (Fig. 1*A*) yielded two major components. The second small peak, which coeluted with free iodoacetic acid, was purified by paper electrophoresis at pH 6.5 and shown to consist of glutamic acid and homoserine (Hsr) as predicted from the Pronase peptides. Because peak I contained the rest of the molecule, it was subjected to gel filtration on Sephadex G-100 in 35% formic acid (Fig. 1*B*). One major radioactive peak, which was presumed to contain the region of the molecule starting with residue 3 and going past the hinge to residue 252 (Eu numbering) (16), was recovered between tubes 105 and 120. It was further purified by isoelectric focusing (Fig. 1*C*). Amino acid analysis indicated that all three fractions obtained by isoelectric focusing were identical and that the observed heterogeneity resulted from incomplete alkylation (data not shown). Only the major radioactive peak, indicated by a bar in Fig. 1*C*, was used for further studies. Amino-terminal analysis was negative due to the fact that the Gln in position 3 had cyclized during the preparative procedures (17).

To obtain the amino-terminal sequence, two types of studies

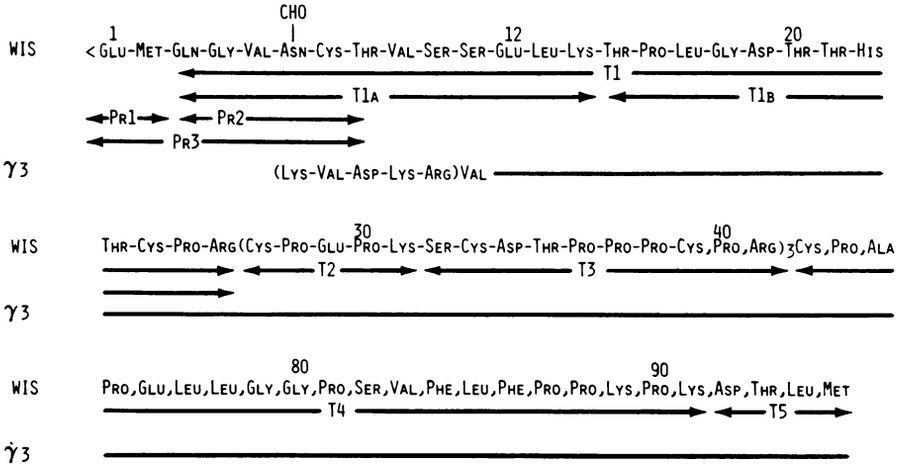

FIG. 2. Amino acid sequence of amino terminus and hinge fragment of protein WIS compared to hinge of γ3 (7). Automatic sequence analysis after removal of <Glu yielded a sequence to residue 39. Tryptic peptides T_4 and T_5 were placed by homology. Pr, Pronase peptides. Evidence for triplication of peptides T_2 and T_3 is provided in Tables 1 and 2. Amino acid sequence identity (—) starts with Glu at position 12.

Table 1. Amino acid compositions of CNBr hinge fragment and its tryptic peptides

	CNBr fragment (Fig. 1C)	Peptide T_1	T_2	T_3	T_4	T_5
CmCys*	10	1.5	0.3	0.8	0.9	
Asp	6.0	2.1		1.0		1.0
Thr	8.6	4.6		1.0		1.0
Ser	6.8	2.2		1.0	1.0	
Glu	6.6	2.3	1.1		1.1	
Pro	28.0	3.1	2.0	4.3	6.5	
Gly	4.3	1.8			2.3	
Ala	1.0				0.9	
Val	3.2	1.6			1.0	
Leu	6.4	2.0			3.0	1.1
Phe	5.4†				1.9	
His	1.0	1.4				
Lys	5.2	1.2	1.0		1.9	
Arg	4.6	0.9		1.0		
Hsr	+					0.5
CHO	+	+				

Compositions are expressed as residues per mole.
* Yields low due to poor recovery.
† Falsely high due to CHO.

were performed: (*i*) The intact molecule was treated with calf liver pyroglutamate aminopeptidase, and after removal of <Glu it was possible to obtain the sequence of the first 39 residues by automatic sequencer analysis (Fig. 2). (*ii*) The CNBr fragment containing the hinge region was digested with trypsin and the peptides were purified by high-voltage electrophoresis at pH 6.5, 3.5, and 2.1. Table 1 shows the amino acid analysis of the major CNBr fragment and its tryptic peptides. In Fig. 2 an attempt is made to align the sequence data and the Pronase and tryptic peptides with the established sequence of the γ3 hinge (7). Starting with residue 12, there is identity with the γ3 hinge (7). However, residues 1–11 do not correspond to the sequence preceding the γ3 hinge or the homologous region of the γ1 heavy chain. When the first 11 residues are compared to various regions of the H chain (Fig. 3), optimal homology is achieved by aligning residues 1, 2, and 3 with the amino terminus of the V_H1 region; Met can result from Val by a single base substitution. Positions 4–7 do not match with any of the published V_H amino terminal sequences, but positions 8–11 show striking homology to the end of the V_H1 sequence from residues 114–117, commonly considered as the VC joining region (18). It should be noted that the presence of a typical complex oligosaccharide linked to Asn and the presence of a Cys residue in this region are unique. To account for these unusual residues, we have chosen to place the start of the deletion at position 3 and we included Gly as an invariant residue after the deletion near the beginning of the VC joining region. However, it is equally possible to have the deletion begin at residue 7 and end with Thr-Val-Ser-Ser. This is then followed by a deletion of the entire C_H1 domain, with resumption of

Table 2. Radioactivity associated with tryptic peptides obtained from CNBr fragment (Figs. 1C and 2)

Peptide (Fig. 2)	cpm	Ratio
T_1a	40,000	0.6
T_1b	65,000	1.0
T_2	187,000	3.0
T_3	370,000	6.0
T_4	Not recovered	—

normal sequence at the beginning of the hinge. The other possibility, that the molecule begins somewhere within the variable region, was examined by attempting to establish homology to other parts of V_H, especially near the VC joining region. Although methionine has been observed previously in five proteins at or near residue 100 (16), comparison of the sequence in protein WIS with the amino acid sequence preceding or following that methionine makes it unlikely that the amino terminus corresponds to this point. Similarly, alignment with the three residues preceding Gly (residue 110, Eu numbering) yields poorer homology. The third possibility, that the first 11 residues represent the precursor peptide, is unlikely in view of the sparsity of hydrophobic residues (19, 20).

The hinge of γ3 has been shown to be unusually large due to the quadruplication of a 15-residue subunit (7). Three types of analysis were carried out on protein WIS to demonstrate that the same holds true for its hinge. (*i*) The molecular weight of the CNBr fragment was shown to be approximately 10,000. (*ii*) Comparison of the amino acid analysis of the CNBr fragment (Table 1) to various combinations of the four peptides gives the best fit when peptides T_2 and T_3 are triplicated and T_1, T_4, and T_5 are included only once (Table 1). (*iii*) The radioactive peptides were isolated by two-dimensional chromatography and high-voltage electrophoresis from a tryptic digest of 25 mg of completely reduced and alkylated CNBr fragment. Under the conditions of this experiment there was an additional cleavage site after the lysine at residue 14, giving two peptides known as T_1a and T_1b. Table 2 lists the radioactivity recovered in each of these peptides and indicates that the cpm in peptides T_1a and T_1b are approximately one-third of peptide T_2 and one-sixth of peptide T_3, which contains two Cys residues. This indicates that peptides T_2 and T_3 are triplicated. In this experiment peptide T_4 was not recovered, but, on the basis of other studies, it can be assumed not to be triplicated.

Fig. 4 summarizes the structural data on eight γHCD proteins, three molecules with a deletion limited to the hinge, and protein WIS. In the myelomas with a deleted hinge, the defect is limited to the 15-residue hinge. In most γ HCD proteins in which degradation can be excluded, a normal or at times an abnormal amino terminus (23) is followed by a large deletion that ends at the beginning of the hinge with the sequence Glu-Pro-Lys. However, in the case of γ3 HCD proteins the

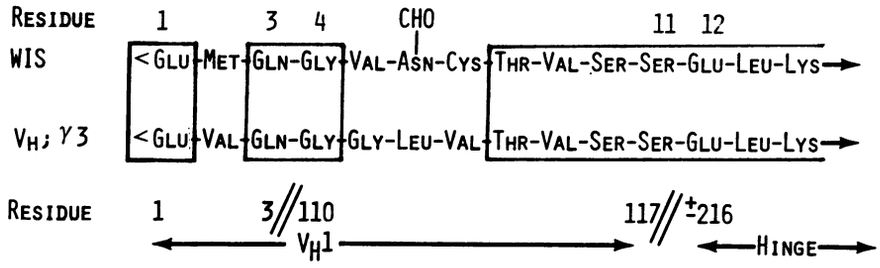

FIG. 3. Alignment of amino terminus of protein WIS with V_H1 and hinge of γ3. Boxes indicate identity. //, Sites of deletion.

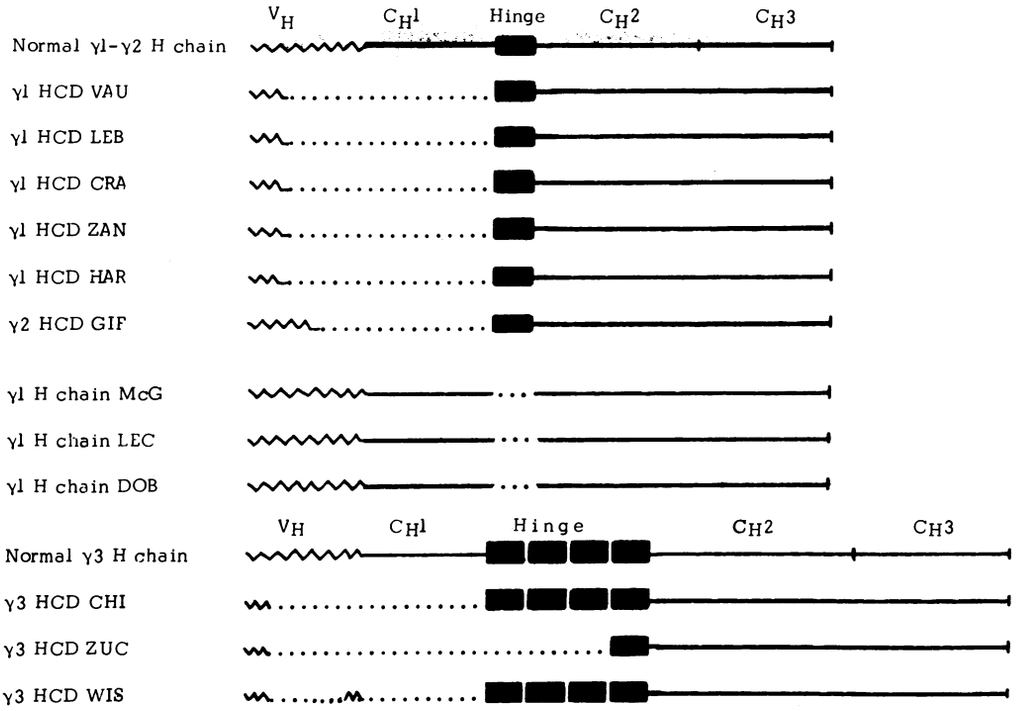

FIG. 4. Diagram of normal γ1, 2, and 3 H chains, five γ1 and one γ2 HCD proteins with internal gaps to residue 216, three myeloma proteins with deletions of hinge only (216–230), two γ3 HCD proteins with deletions ending at the hinge, and protein WIS. Structures of most proteins are summarized in ref. 1; VAU, LEB, and ZAN are unpublished; McG (2), LEC (3), DOB (4), CHI (21), ZUC (22), HAR (23). ⁓, V region; . . ., deletion; —, C region.

deletion can end at the beginning of the extended hinge with the corresponding sequence Glu-Leu-Lys or within it at one of the quadruplicated regions (ZUC). These findings have given rise to the idea that the hinge represents an unusual site for genetic events (1–3, 8). Protein WIS is of interest because it suggests the preservation of another normal recognition site at the VC joining region where splicing may take place after an internal deletion of about 107 residues. For unknown reasons, perhaps as a result of a second mutation, this is followed by a second deletion, which ends at the beginning of the hinge.

Possible mechanisms that can result in these unusual structural features remain to be established. The most likely possibilities are either some mutational events resulting in the transcription of an altered RNA molecule or faulty processing of the heterogeneous nuclear RNA, because studies of several proteins such as ovalbumin (24), hemoglobin (25), immunoglobulin light chains (26), and a number of viral proteins (27, 28) have shown that DNA is transcribed first into a heterogeneous nuclear RNA, from which sections are looped out and excised in the formation of a smaller mRNA that serves as a template for translation. In order to assure translational fidelity, certain regions of the RNA must serve as sites for the splicing process.

In discussing these two possibilities, it seems appropriate to consider two myeloma proteins with abnormal light chains (5, 6) for which protein structure can now be correlated with the DNA sequence obtained by cloning (26). Fig. 5 shows that, in the cloned DNA from a murine λ_{II} producing plasmacytoma, there is a large untranslated region of DNA at the end of the precursor peptide and another one at residue 98 in the embryo (26) and at residue 110 in the myeloma DNA (29). In comparing this to the deletions of the only two L chain mutants known, it becomes apparent that the gap in the L chain of SM ends at residue 110 (which corresponds to the myeloma untranslated region and is close to the chemically defined VC joining region of the L chain) and resumes with the sequence Gln-Pro-Lys, while that of the other protein, SAC, ends at residue 99 [which corresponds exactly to the position of the second intervening

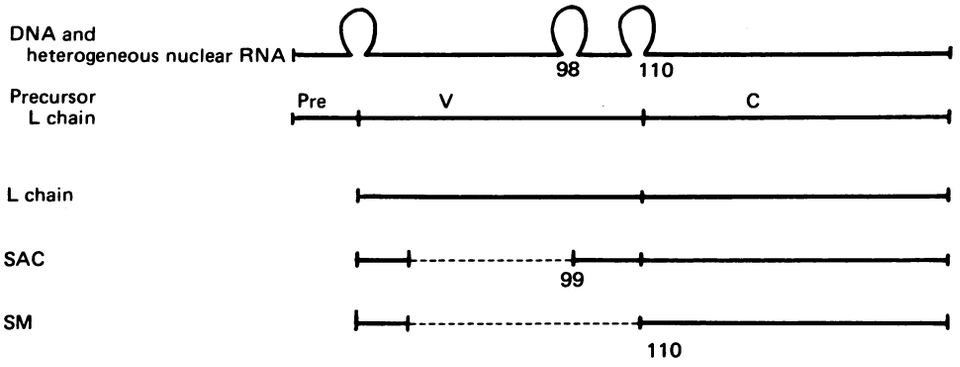

FIG. 5. Diagram of L chain DNA (based on data in refs. 26 and 29), the precursor L chain, the L chain, and two L chain mutants (5, 6) whose deletions (broken lines) end at the sites of the introns (intervening sequences) (Ω) defined by DNA sequencing. The stretch from position 99 to 110 has recently been called the J sequence (29).

sequence (intron) in the embryonal λ_{II} DNA defined by Tonegawa *et al.* (26)]. Thus, splicing seems to have taken place at the next available recognition site, which is the VC joining region. Because, depending on the L chain class or subclass, and perhaps on the state of differentiation of the cell, this may be somewhere in the vicinity of residue 110 and not always exactly at the chemically defined VC joining region, it is difficult to define precisely the beginning of the C region.

If one attempts to predict the structure of H chain DNA on the basis of the characterized H chain mutants, it seems likely that recombination or splicing sites may exist not only at the VC joining region but also at the beginning and end of the hinge. In addition, results with other rarer or less well characterized human variants (1) and several murine proteins (30) suggest that each of the domains may correspond to a separate transcriptional unit defined in terms of excision and splicing of heterogeneous nuclear RNA.

The possibility that protein WIS starts around residue 110 as a consequence of degradation of a larger precursor should be mentioned but appears unlikely for two reasons. First, attempts to align the amino-terminal sequence with this region in each of the V_H subclasses is not possible without deviating from at least one invariant residue (18). Second, and perhaps more important, in each of the 12 HCD proteins presumed to be the result of degradation, proteolysis occurs at various sites in the hinge region and has never been noted elsewhere in the molecule (31).

We thank Dr. L. G. Selbowitz for protein WIS, Ms. Ann Chapman and Dr. Stuart Kornfeld for carbohydrate analyses, Dr. George Abraham for introducing us to the use of the pyroglutamate aminopeptidase, Ms. Joan Zaretzky for technical assistance, and Ms. Marlene Chavis and Ms. Beverly Coopersmith for secretarial help. This work is supported by U.S. Public Health Service Research Grants AM 01431 and AM 02594 and The Irvington House Institute.

1. Franklin, E. C. & Frangione, B. (1976) in *Contemporary Topics in Molecular Immunology*, eds. Inman, F. P. & Mandy, W. J. (Plenum, New York), Vol. 4, pp. 89–126.
2. Fett, J. W., Deutsch, H. F. & Smithies, O. (1973) *Immunochemistry* **10**, 115–118.
3. Rivat, C., Schiff, C., Rivat, L., Ropartz, C. & Fougereau, M. (1976) *Eur. J. Immunol.* **6**, 545–551.
4. Lopes, A. D. & Steiner, L. A. (1973) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32**, 1003.
5. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921.
6. Garver, F. A., Chang, L., Mendicino, J., Isobe, T. & Osserman, E. F. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4559–4563.
7. Michaelsen, T. E., Frangione, B. & Franklin, E. C. (1977) *J. Biol. Chem.* **252**, 883–889.
8. Franklin, E. C. & Frangione, B. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 187–191.
9. Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412.
10. Frangione, B., Milstein, C. & Franklin, E. C. (1969) *Nature (London)* **221**, 145–151.
11. Gross, E. (1967) *Methods Enzymol.* **11**, 238–254.
12. Vesterberg, O. & Svensson, H. (1966) *Acta Chem. Scand.* **20**, 820–834.
13. Press, E. M., Piggot, P. J. & Porter, R. R. (1966) *Biochem. J.* **99**, 356–366.
14. Podel, D. N. & Abraham, G. (1978) *Biochem. Biophys. Res. Commun.* **81**, 176–185.
15. Franklin, E. C., Frangione, B. & Adlersberg, J. (1975) *J. Immunol.* **115**, 314–315.
16. 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.
17. Blombäck, B. (1967) *Methods Enzymol.* **11**, 398–411.
18. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1976) *Variable Regions of Immunoglobulin Chains—Tabulations and Analyses of Amino Acid Sequence* (National Institutes of Health, Bethesda, MD).
19. Jilka, R. L. & Pestka, S. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5692–5696.
20. Burstein, Y. & Schechter, I. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 716–720.
21. Frangione, B. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1552–1555.
22. Frangione, B. & Milstein, C. (1969) *Nature (London)* **224**, 597–599.
23. Frangione, B., Franklin, E. C. & Smithies, O. (1978) *Nature (London)* **273**, 400–401.
24. Weinstock, R., Sweet, R., Weiss, M., Cedar, H. & Axel, R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1299–1303.
25. Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P. & Weissmann, C. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1309–1313.
26. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1485–1489.
27. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature (London)* **272**, 113–120.
28. Berget, S. M., Moore, C. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **74**, 3171–3175.
29. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell*, in press.
30. Secher, D. S., Milstein, C. & Adetugbo, K. (1977) *Immunol. Rev.* **35**, 51–72.
31. Franklin, E. C. (1978) *J. Immunol.*, in press.