Proceedings of the National Academy of Sciences
Vol. 66, No. 2, pp. 558–563, June 1970

Implications of Heavy Chain Disease Protein Sequences for Multiple Gene Theories of Immunoglobulin Synthesis

William D. Terry and Jack Ohms

IMMUNOLOGY BRANCH, NATIONAL CANCER INSTITUTE, BETHESDA, MARYLAND, AND BECKMAN INSTRUMENTS, INC., PALO ALTO, CALIFORNIA

Communicated by Robert J. Huebner, March 31, 1970

Abstract. The sequence of the amino-terminal 34 amino acids of γ-heavy chain disease (γ-HCD) protein Hi is homologous with the amino-terminal region of immunoglobulin heavy chains. γ-heavy chain disease is smaller than a normal γ-chain, but has the carboxy-terminal composition expected for γ-chains and must, therefore, contain an internal deletion. Comparison of the Hi sequence with that of γ-heavy chain disease Zu, which also has an internal deletion, indicates that the site of internal deletion is not a constant characteristic of γ-heavy chain disease proteins.

Heavy chains can be assigned to subgroups on the basis of variable region sequences. The variable regions of Hi and one other protein differ significantly from those determined for other heavy chains, and these two proteins are assigned to a new heavy chain variable region subgroup, V_{HIV}.

It has been suggested that single immunoglobulin heavy chains are the products of two separate structural genes and that variable region genetic information is translocated and integrated into common region information. These multiple gene theories make no prediction as to whether DNA or RNA is translocated. γ-heavy chain disease proteins provide unique information that indicates that if translocation is required for the production of immunoglobulin heavy chains, it is DNA, not RNA, that is translocated.

Introduction. Heavy chain disease (HCD) proteins are aberrant forms of human immunoglobulin heavy polypeptide chains. Examples of heavy chain disease proteins have been described for the classes IgG,[1–7] IgA,[8,9] and IgM.[10] Studies of IgG HCD (γ-HCD) proteins have shown that they consist of portions of the γ-chain and that they lack light polypeptide chains. The two γ-HCD proteins for which data are available have molecular weights of 26,000[11] and 40,000[7] as compared with a molecular weight of approximately 54,000 for normal γ-chains.[12]

The structure of γ-HCD proteins is of interest because, at least in some cases, they appear to be the intact products of mutated immunoglobulin structural genes, rather than the partially degraded products of normal immunoglobulin structural genes. It was first shown by Prahl[13] that the amino acid sequence of the amino-terminal (N-terminal) tripeptide of γ-HCD Zu was similar to the sequence determined for the amino-terminus of γ-chains.[14] Moreover, it was

demonstrated that the carboxy-terminal (C-terminal) sequence of Zu was homologous to that of other γ-chain C-terminal regions.[13] Zu therefore contained normal N-terminal and C-terminal regions but was smaller than an intact γ-chain, and Prahl concluded that this protein contained an internal deletion. Recent work of Frangione and Milstein has shown that this indeed is the case and that the first 18 residues of Zu are homologous with the amino-terminal sequence of γ-chains.[15] Beginning with position 19 of Zu, however, the sequence becomes homologous with that seen at about position 218 of other γ-chains and then apparently continues uninterruptedly to the C-terminus. Zu therefore contains an internal deletion of some 200 residues. The implications of these data for multiple gene models of immunoglobulin genetic control will be discussed below.

We have studied a γ-HCD protein from patient Hi. Antigenic and physical characteristics of Hi protein are published elsewhere.[7] Unlike most other γ-chains, Hi has a free amino-terminal residue. This makes it possible to directly determine the amino acid sequence of the amino-terminal portion of this protein.

Materials and Methods. Protein isolation: Hi protein was isolated from serum by a combination of ammonium sulfate precipitation, Pevikon electrophoresis, Sephadex G-200 gel filtration, and agar block electrophoresis. Isolated protein was reduced with 0.1 *M* dithioerythritol in 7 *M* guanidine-HCl buffered with 0.5 *M* Tris HCl, pH 8.1, and alkylated with 0.22 *M* iodoacetamide.

Amino acid sequence determination: The sequence of the reduced and alkylated protein was determined on three different occasions, using an automatic amino acid sequencer (Beckman Instruments, Inc., model 890A) in accordance with the methods and conditions described by Edman.[16] The thiazolinone derivatives obtained from the sequencer were converted to phenylthiohydantoin-amino acids and identified and quantitated according to the liquid-gas chromatographic methods of Pisano.[17] DC-560 support was used in the glass column (4 ft × 0.2 mm) of a gas chromatograph (GC-45, Beckman Instruments, Inc.).

Results and Discussion. The partial sequence of the amino-terminal 34 residues of γ-HCD Hi are compared in Figure 1 with the published amino acid sequences of γ-HCD Zu,[15] the γ-chains of four myeloma proteins,[18–21] and the μ-chain of a Waldenström macroglobulin.[22] Empty brackets in the Hi sequence indicate that, for technical reasons, the residue at that step has not been established with certainty.

The amino-terminus of γ-HCD Hi is homologous with the 34 amino-terminal residues of heavy chains (Fig. 1). There are at least 20 positions at which Hi is identical to Vin, 13 positions at which it is identical to Eu, Daw, and Ou, and 10 positions at which it is identical to He. The number of identical residues will probably increase when the amino acids at the 7 questionable positions are identified. At 7 of the 34 positions (2, 4, 8, 14, 22, 25, and 26), the four γ-chains, the μ-chain, and Hi have identical amino acids. The evidence for sequence homology is further strengthened by the presence of infrequently occurring amino acids, such as the cysteine at position 22, the phenylalanine at 27, and the methionine at 34.

Previous studies have shown that the C-terminal octadecapeptide of Hi is that expected for the C-terminus of γ-chains.[23] The molecular weight of completely

Class	Subgroup	Protein	1				5					10					15			18
γ-HCD	IV	Hi	Gly	Val	Leu	Leu	Val	Glu	Ser	Gly	*Gly*	Val	Ser	*Ileu*	[ ]	Pro	Gly	*Gly*	*Ser*	*Leu*
γ	IV	Vin	Glu	—	Gln	—	—	—	—	—	—	Gly	Leu	—	Gln	—	—	—	—	—
γ	II	Daw	*PCA*	*Val*	*Thr*	Leu	Arg	Glu	Ser	Gly	*Pro*	Ala	*Leu*	*Val*	Arg	Pro	Thr	Gln	Thr	Leu
γ	II	He	—	—	—	—	Lys	—	Asn	—	—	Thr	—	—	Lys	—	—	Glu	—	—
μ	II	Ou	—	—	—	—	Thr	—	—	—	—	—	—	—	Lys	—	Lys	—	Pro	—
γ	I	Eu	PCA	Val	Gln	Leu	Val	Gln	Ser	Gly	*Ala*	Glu	Val	Lys	*Lys*	*Pro*	*Gly*	Ser	Ser	Val
γ-HCD	I	Zu	—	—	—	Val	—	Glu	—	—	—	Asp	Leu	Val	—	—	—	Gly	—	Ser

Protein		20					25					30				34
Hi	[ ]	Leu	[ ]	Cys	[ ]	Ala	Ser	Gly	Phe	[ ]	Ileu	Gly	[ ]	Phe	[ ]	Met
Vin	Arg	—	Ser	—	Ala	—	—	—	—	Thr	Val	Ser	Thr	Asn	Tyr	—
Daw	*Thr*	*Leu*	*Thr*	*Cys*	*Thr*	Phe	Ser	Gly	Phe	Ser	Leu	Ser	Gly	Glu	Thr	Met
He	—	—	—	—	—	Leu	—	—	Leu	—	—	Thr	Thr	Asp	Gly	Val
Ou	—	—	—	—	—	—	—	—	—	—	—	—	Thr	Ser	Arg	—
Eu	Lys	Val	Ser	Cys	Lys	Ala	Ser	Gly	Gly	Thr	Phe	Ser	Arg	Ser	....	.....

FIG. 1.—N-terminal sequence of γ-HCD Hi is compared with sequences of four γ-chains, one μ-chain, and one γ-HCD protein. Residues 13, 19, 21, 23, 28, 31, and 33 of Hi have not been established with certainty. Proteins are grouped by sequence homology. Amino-terminal sequence of Hi compared with other heavy chain sequences.

Zu, ref. 15; Eu, ref. 18; He, ref. 19; Daw, ref. 20; Vin, ref. 21; Ou, ref. 22. ——, residue identical to that of other proteins in the same subgroup; PCA, pyrrolidone carboxylic acid; · · · ·, deletion at residues 33 and 34 of Eu to maximize homology.

reduced and alkylated Hi protein is about 40,000, indicating that Hi is smaller than a normal γ-chain and must be missing approximately 100 residues. Since it contains appropriate amino- and carboxy-terminal regions, Hi, like Zu, must contain an internal deletion.

Comparison of the Hi and Zu sequences shows that the position of the internal deletion is not constant among γ-HCD proteins. The Zu deletion begins with residue 19, while the Hi deletion begins at some point after residue 34. The extent of the deletion is also not constant, in that Zu lacks some 200 residues, while Hi is probably missing only about 100 residues. This suggests that the mechanism of the deletion does not depend upon a specific limited region of the gene (or genes) that is peculiarly prone to deletional events.

The γ-chains derived from human IgG consist of about 450 amino acids. Comparisons of several γ-chains show that they contain variable regions (amino-terminus to approximately position 118) which differ significantly in amino acid sequence, and common regions (119 to the carboxy-terminus) which have very similar sequences.[18,24] The amino acid sequences of heavy chain variable regions, like those of light chains, can be placed into subgroups on the basis of sequence homology.[19,24,25] Previous analyses have shown that Daw, Ou, and He have closely related sequences, and that their sequences differ significantly from that of Eu.[19,24] Eu has, therefore, been assigned to variable region subgroup V_{HI}, and Daw, Ou, and He assigned to subgroup V_{HII}. Zu is closely related to Eu and probably is a V_{HI} protein (Fig. 1, Table 1). On the other hand,

TABLE 1. *Comparison of identical residues within and between subgroups.*

	No. of identities	Percentage of identity
Eu (I) and Zu (I)	11/18*	61
Daw (II) and He (II)	22/34	65
Hi (IV) and Vin (IV)	20/27†	74
Hi (IV) and Eu (I)	12/27‡	44
Eu (I) and Daw (II)	10/32	31
Hi (IV) and Daw (II)	13/27	48

* Zu is only homologous for 18 residues.

† Seven undetermined residues in the first 34 of Hi; comparison therefore limited to 27 of 34 positions.

‡ To maximize homology with other heavy chains, a deletion has been placed at positions 33 and 34 in Eu. Comparison is therefore limited to 27 of 32 positions.

Hi differs from the V_{HI}, V_{HII}, and V_{HIII} proteins (V_{HIII}, not shown) but is highly homologous to Vin.[21] These two proteins, which have identical residues at 20 of the first 34 positions, are tentatively designated as belonging to subgroup V_{HIV} (Table 1). Subgroup assignments depend upon the distribution as well as the total number of identical residues. For example, the combination of glycine at position 9, associated with isoleucine at position 12 and the sequence glycine-serine-leucine at positions 16–18, appears to be characteristic of V_{HIV} proteins. Other examples of apparent subgroup-related amino acid combination are underlined in Figure 1. The significance of heavy chain variable region subgroups is discussed elsewhere.[19,24,25]

It has been proposed that in germ-line cells the variable region is encoded by a gene that is physically separate from the gene encoding the common region.[26]

It has further been proposed that in somatic cells differentiated for the production of immunoglobulins, variable region information (DNA or messenger RNA) is translocated and physically joined to the common region gene (or messenger). The formulations of this hypothesis do not distinguish between translocation of variable gene DNA or translocation of variable gene messenger RNA.

Three possible ways in which a translocational mechanism could operate are: (1) Variable region DNA is translocated and permanently integrated into common region DNA to give a stable, variable common gene. (2) Variable region DNA is translocated and temporarily integrated into common region DNA, and the integrated gene is transcribed. After transcription, variable region DNA dissociates and is retranslocated at the time of the next transcription. (3) Variable region messenger RNA is translocated and integrated into common region messenger RNA, and the integrated RNA is translated. This process would be repeated with each transcription. Whether it is DNA or RNA that is translocated, the hypothesis requires that recognition sites be present in the nucleotide sequences to permit the accurate joining together of variable and common region information.

Analysis of γ-HCD sequences suggests that, if there is translocation, DNA rather than RNA is the translocated material. If deletion were to occur prior to translocation, two separated regions of DNA would have to be affected simultaneously since the deletion includes portions of both the common and variable regions. In addition, the usual recognition site would be destroyed, and a new recognition site would have to be generated before integration could be effected. As both of these events are unlikely, it seems that translocation must occur before deletion, and that translocation of variable region information and integration into a given common region gene must occur only once in the differentiation of a cell.

These considerations rule out possibilities 2 and 3 and indicate that if a translocation mechanism operates for immunoglobulin heavy chains, and if this mechanism is the same in "normal" cells as it is in those cells synthesizing γ-HCD proteins, the synthesis of a γ-chain requires that variable region DNA be translocated to common region DNA where it is permanently integrated to form a "variable-common gene." In γ-HCD synthesizing cells, a subsequent genetic event leads to a large internal deletion. These conclusions are based on the Zu sequence data, since the exact location of the Hi deletion is not yet established.

There is no information concerning the frequency with which a deletion affecting portions of both common and variable regions occurs in immunoglobulin structural genes. A small number of cells carrying such a deletion may be present in all normal individuals, and heavy chain disease proteins may thus represent a normal serum immunoglobulin brought to the level of detectability by the heavy chain disease process. Alternatively, the genetic events precipitating heavy chain disease may include a deletion from a previously intact common-variable gene.

The skilled technical assistance of Mrs. Jan Rea and Mr. Matthew Miller, and the constructive comments of Drs. Daniel Ein, Leroy Hood, and John L. Fahey during preparation of this manuscript are gratefully acknowledged.

Abbreviations: γ-HCD, γ-heavy chain disease; IgG, immunoglobulin G; IgA, immunoglobulin A; IgM, immunoglobulin M; Hi, Zu, etc., are shortened forms of names of patients studied.

* Requests for reprints may be addressed to Dr. W. D. Terry, Immunology Branch, National Cancer Institute, Building 10, Room R4B08, National Institutes of Health, Bethesda, Md. 20014.

[1] Franklin, E. C., J. Lowenstein, B. Bigelow, and M. Meltzer, *Amer. J. Med.*, **37**, 332 (1964).

[2] Osserman, E. F., and K. Takatsuki, *Amer. J. Med.*, **37**, 351 (1964).

[3] Lebreton, J. P., C. Rivat, L. Rivat, L. Guillemot, and C. Ropartz, *Presse Med.*, **75**, 2251 (1967).

[4] Ellman, L. L., and K. J. Bloch, *New Engl. J. Med.*, **278**, 1195 (1968).

[5] Ein, D., D. N. Buell, and J. L. Fahey, *J. Clin. Invest.*, **48**, 785 (1969).

[6] Wager, O., J. A. Rasanen, L. Lindberg, and U. Makela, *Acta Path. Microbiol. Scand.*, **75**, 350 (1969).

[7] Woods, R., G. R. Blumenschein, and W. D. Terry, *Immunochemistry*, in press.

[8] Seligmann, M., F. Danon, D. Hurez, E. Mihaesco, and J-L. Preud'Homme, *Science*, **162**, 1396 (1968).

[9] Seligmann, M., E. Mihaesco, D. Hurez, C. Mihaesco, J-L. Preud'Homme, and J. C. Rambaud, *J. Clin. Invest.*, **48**, 2374 (1969).

[10] Josephson, A. S., E. Price, and L. Biro, *Clin. Res.* (Abstr.), **17**, 604 (1969).

[11] Franklin, E. C., *J. Exptl. Med.*, **120**, 691 (1964).

[12] Edelman, G. M., W. F. Gall, M. J. Waxdal, and W. H. Konigsberg, *Biochemistry*, **7**, 1950 (1968).

[13] Prahl, J. W., *Nature*, **215**, 1386 (1967).

[14] Press, E. M., P. J. Piggot, and R. R. Porter, *Biochem. J.*, **99**, 356 (1966).

[15] Frangione, B., and C. Milstein, *Nature*, **224**, 597 (1969).

[16] Edman, P., and G. Begg, *Eur. J. Biochem.*, **1**, 80 (1967).

[17] Pisano, J. J., and T. J. Bronzert, *J. Biol. Chem.*, **244**, 5597 (1969).

[18] Edelman, G. M., B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser, and M. J. Waxdal, these PROCEEDINGS, **63**, 78 (1969).

[19] Cunningham, B. A., M. N. Pflumm, U. Rutishauser, and G. M. Edelman, these PROCEEDINGS, **64**, 997 (1969).

[20] Press, E. M., and P. J. Piggot, in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 32 (1967), p. 45.

[21] Pink, J. R. L., and C. Milstein, in *FEBS Symposia Series* (New York: Academic Press), vol. 15, in press.

[22] Wikler, M. H. Kohler, T. Shinoda, and F. W. Putnam, *Science*, **163**, 75 (1969).

[23] Terry, W., D., and J. T. Harrington, unpublished observations.

[24] Press, E. M., and N. M. Hogg, *Nature*, **223**, 807 (1969).

[25] Hood L., A. Kaplan, H. Metzger, and W. D. Terry, in preparation.

[26] Dreyer, W. J., and J. C. Bennett, these PROCEEDINGS, **54**, 864 (1965).