Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 5031–5032, August 1982
Genetics

Fourteen nucleotides in the second complementarity-determining region of a human heavy-chain variable region gene are identical with a sequence in a human *D* minigene

(independent assortment/gene conversion/antibody diversity and complementarity)

TAI TE WU* AND ELVIN A. KABAT†‡

†National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205; *Department of Biochemistry and Molecular and Cell Biology and Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60201; and ‡Departments of Microbiology, Human Genetics and Development, and Neurology, Columbia University, New York, New York 10032

Contributed by Elvin A. Kabat, May 11, 1982

ABSTRACT **A sequence of 14 nucleotides in one human diversity (*D*) minigene (*D2*) is identical with a sequence in complementarity-determining region 2 (*CDR2*) of one human gene for the variable region of the heavy chain of immunoglobulin. The finding that nucleotide segments present in a *D* minigene can appear in *CDR2* raises the possibility that other minigene segments may be involved in the generation of antibody diversity and complementarity or that nucleotide segments may move from one *CDR* to another by a gene conversion mechanism.**

Ten and five diversity (*D*) minigene segments for heavy-chain variable (V_H) regions of immunoglobulins have been identified in the mouse (1, 2) and human (3) genomes, respectively. In the mouse some of the *D* minigenes code for portions of complementarity-determining region 3 (*CDR3*); matches of 9 to 18 nucleotides with those of assembled V_H genes have been found. In both species, recombination signal sequences (1–8) are involved in assembly of complete V_L and V_H regions by *V–J* and *V–D–J* joining, respectively (L indicates light chain; *J*, joining segment). Most of the *D* segments have been identified with probes involving a nonamer, a 12-nucleotide spacer, and a heptamer 5′ to the presumed coding sequence and a heptamer, a 12-nucleotide spacer, and a nonamer 3′ to this sequence (1–3). Thus the coding sequence may or may not be related to already known *CDR3* segments, and indeed the four human *D* segments reported (3)—*D1*, *D2*, *D3*, and *D4*—do not correspond to any known *CDR3* (9).

Although many nucleotide sequences of mouse V_H regions are known (1, 4, 5, 7, 10–12), only one human genomic sequence, V_H*26*, has been reported (6). We translated the four *D* segments and their complementary strands in all three reading frames into amino acid sequences and compared them in our data bank with the known amino acid sequences of human and other species of heavy chains (9). One of the reading frames for the human *D2* minigene gave the sequence Ser-Gly-Gly-Ser ( )Tyr, which matched with residues 53 to 58 of a translated amino acid sequence of V_H*26* (6) and of two V_H chains of type III anti-pneumococcal antibody from rabbit 3381 (refs. 13 and 14; see also ref. 9) (antibodies 3381 and 3381-2). Although the sequence of antibody 3381-2 has not been completely determined, it differs from that of antibody 3381 by serine replacing arginine at position 31 in CDR1. This sequence has not been reported in nonimmunoglobulins in Dayhoff's protein data bank (15), nor has it been reported in immunoglobulin constant regions or light chains, or in HLA, H-2, Ia, and Thy-1 antigens, complement, C-reactive protein, etc. (9, 15). The *D* nucleotide sequences were then examined. Table 1 shows a stretch of 14 contiguous nucleotides identical in V_H*26* and *D2* with the segment in solid boxes.

The location of this match is remarkable, because it is in the middle of the *CDR2* (amino acids 50–65) of V_H*26*. Moreover, 5′ to this 14-nucleotide match (dashed boxes) there is T-A-T-T, the A-T-T of which codes for isoleucine, which is found frequently at position 51, occurring 42 times in 61 V_H sequences in all species examined (9). On the 3′ side there is a T-A-C-T (dashed boxes), the T-A-C coding for tyrosine, which occurred in 21 of 62 reported V_H chains at position 58 and in 55 of 63 chains at position 59 (9). If the nucleotide sequences of unrelated genes from the Dayhoff (16) and Goad (17) data banks are compared with the 14-nucleotide stretch in *D2*, the best match, in the primary origin of replication, genes *1.1* and *1.2* of the complementary strand of bacteriophage T7, is 12 nucleotides broken up into segments of 10 and 2 by a mismatched nucleotide, and without the bracketing T-A-T-T and T-A-C-T. It is of interest that the V_H chains of the mouse NP^b family of heteroclitic monoclonal antibodies specific for (4-hydroxy-3-nitrophenyl)acetyl have the largest number of base changes in *CDR2*; the reason for this is obscure and some or all of these might arise from a mechanism other than somatic (7) or a germline mutation.

If these findings are not coincidental and additional instances occur as other V_H gene and *D* minigene sequences become available, they could add additional facets to the already extraordinary complex of mechanisms hypothesized as involved in the generation of antibody diversity. Thus this finding might be an instance of the insertion (18) or assortment (19, 20) of minigenes by gene conversion mechanisms (21, 22) involving framework region (*FR*) and *CDR* segments of the *V* region other than *D* and *J*. On this basis the human *D2* may not be a *CDR3* minigene. Alternatively, if it is indeed a *CDR3* minigene, it might indicate that gene conversion may result in the insertion of *D* sequences into *CDR2*.

The extent to which mechanisms such as minigene assortment, somatic mutation, and a repertoire of germ-line genes contribute to the generation of antibody complementarity as distinct from antibody diversity will be the main question of the next few years. The findings of Kranz and Voss (23) that they could not generate functional sites by recombining heavy and light chains of six murine monoclonal anti-fluorescyl antibodies except for the homologous recombinants would appear to have

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: V_H region, variable region of the immunoglobulin heavy chain; V_H gene, nucleotides coding for the V_H region through *FR3* (excludes the *D* and *J* minigenes); *FR*, framework segment of the *V* region; *D*, diversity; *J*, joining; V_L region, variable region of the light chain; CDR, complementarity-determining region.

eliminated the generation of different antibody sites by random assortment of V_H and V_L chains.

Table 1. Comparison of nucleotide sequences of *CDR2* of a human V_H gene, V_H26, with four human *D* minigene segments

Residue no.	Amino acid	Nucleotide sequence				
		V_H26	*D1*	*D2*	*D3*	*D4*
50	Ala	G				
		C				
		T	A	A	A	A
51	Ile	A	G	G	G	G
		T	G	G	C	G
		T	A	A	A	A
52	Ser	A	T	T	T	T
		G	A	A	A	A
		T	T	T	T	T
52A	Gly	G	T	T	T	T
		G	G	G	G	G
		T	T	T	T	T
53	Ser	A	A	A	G	A
		G	C	G	G	G
		T	T	T	T	T
54	Gly	G	G	G	G	A
		G	G	G	G	G
		T	T	T	T	T
55	Gly	G	G	G	G	A
		G	G	G	A	C
		T	T	T	T	C
56	Ser	A	G	A	T	A
		G	T	G	G	G
		C	A	C	C	C
57	Thr	A	T	T	T	T
		C	G	G	A	G
		A	C	C	T	C
58	Tyr	T	T	T	T	T
		A	A	A	C	A
		C	T	C	C	T
59	Tyr	T	A	T		G
		A	C	C		C
		C	C	C		C
60	Gly	G				
		G				
		A				
61	Asp	G				
		A				
		C				
62	Ser	T				
		C				
		C				
63	Val	G				
		T				
		G				
64	Lys	A				
		A				
		G				
65	Gly	G				
		G				
		C				

Note Added in Proof. G. Rechavi and D. Givol of the Weizmann Institute of Science have determined the sequence of a human V_HII gene, *HG3* (personal communication). *HG3* has 13 of the 14 identical nucleotides (boxed in Table 1) in the *D2* minigene and in V_H26 (V_HIII) with C replacing the G of the first of the 14 nucleotides.

We thank Dr. David Baltimore for helpful discussion. Work with the PROPHET computer system is supported by the National Cancer Institute, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases, the National Institute of General Medical Sciences, and the Division of Research Resources (Contract N01-RR-8-2118) of the National Institutes of Health. This work was aided by Grant PCM 81-02321 from the National Science Foundation to E.A.K., by Cancer Center Support Grant CA 13696 to Columbia University, and by Grant 5-R01-GM21482 from the National Institutes of Health to T.T.W.

1. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. (1981) *Nature (London)* **290**, 562–565.
2. Kurosawa, Y. & Tonegawa, S. (1982) *J. Exp. Med.* **155**, 201–218.
3. Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Nature (London)* **294**, 631–635.
4. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) *Cell* **19**, 981–992.
5. Bernard, O. & Gough, N. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3630–3634.
6. Matthyssens, G. & Rabbitts, T. H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6561–6565.
7. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
8. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66.
9. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains. Tabulation and analysis of amino acid sequences of precursors, V-regions, C-regions, J-chain, and β_2-microglobulin* (Government Printing Office Publication NIH 80-2008, Washington, DC).
10. Zakut, R., Cohen, J. & Givol, D. (1980) *Nucleic Acids Res.* **8**, 3591–3601.
11. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. B. (1981) *Nature (London)* **292**, 426–430.
12. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286**, 676–683.
13. Margolies, M. N., Cannon, L. E., Kindt, T. J. & Fraser, B. (1977) *J. Immunol.* **119**, 287–294.
14. Haber, E., Margolies, M. N., Cannon, L. E. & Rosemblatt, M. S. (1975) *Miami Winter Symp.* **9**, 303–338.
15. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*; 1973 Supplement I; 1976 Supplement II; 1978 Supplement III; 1981 Data Tape (National Biomedical Research Foundation, Washington, DC).
16. Dayhoff, M. O., Schwartz, R. M., Chen, H. R., Hunt, L. T., Barker, W. C. & Orcutt, B. C. (1981) *Nucleic Acid Sequence Database* (National Biomedical Research Foundation, Washington, DC).
17. Fickett, J., Goad, W. & Kanehisa, M. (1982) *Los Alamos National Laboratory Report LA-9274-MS*, pp. 1–82.
18. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–250.
19. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2429–2433.
20. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1980) *J. Exp. Med.* **152**, 72–84.
21. Egel, R. (1981) *Nature (London)* **290**, 191–192.
22. Baltimore, D. (1981) *Cell* **24**, 592–594.
23. Kranz, D. M. & Voss, E. W., Jr. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5807–5811.