Annu. Immunol. (Inst. Pasteur)
1976, 127 C, 439-448

SOMATIC GENERATION OF ANTIBODY DIVERSITY

by G. Matthyssens, N. Hozumi and S. Tonegawa

Basel Institute for Immunology, Postfach 1005, Basel 4 (Switzerland)

INTRODUCTION

The genetic origin of antibody diversity is one of the central issues of immunology. Does every cell of an organism contain structural genes for the full repertoire of antibody molecules which can be produced by that organism? As long as the conventional definition of a « gene » (*i. e.* the genetic element coding for a polypeptide chain) is applied, it is generally conceded that the answer to the above question is no. Both light and heavy immunoglobulin chains can be considered to consist of a variable (V) and a constant (C) region. The genetic element coding for these regions are termed V and C genes. The diversity issue is now whether or not structural V genes for the full repertoire of antibody V regions are present in every cell of an organism. The alternative is that a large proportion of this diversity is generated by somatic processes, such as somatic mutation. Until recently the immense complexity of the vertebrate genome has hindered direct approaches to the issue of antibody diversity. Instead, arguments for and against these contrasting views have been based on indirect experimental observations (primarily of amino acid sequences of immunoglobulin chains) and on general principles of genetics and evolution [2, 11]. Recent advances in the technology of nucleic acid biochemistry, in particular the availability of highly purified light and heavy chain mRNAs from mouse plasmacytomas, have enabled several groups of researchers to study immunoglobulin genes at the molecular level. In this communication, we will review the evidence that mouse λ-chains, at least those represented among myelomas, are coded for by genes derived from no more than three germ line V genes, possibly from a single one.

RESULTS AND DISCUSSION

The principles of the approach.

The ultimate approach is probably to isolate the immunoglobulin genes from both germ line and plasma cells and to determine whether or

Manuscrit reçu le 1er juin 1976.

zed with the particular mRNA used in the Cot curve experiment. If this number is significantly larger than the experimentally determined reiteration frequency, it constitutes formal evidence for the somatic generation of antibody diversity.

Evidence for the somatic generation of antibody diversity.

We have applied the approach described in the last section to both κ- and λ-chains of the mouse [9, 10]. Heavy chains have also been studied somewhat [1, 7]. Because they are simple and straightforward we will describe here the experimental results obtained with mouse λ-chains in some detail. Figure 1 summarizes amino acid sequence data for the V region of mouse λ chains. These sequences were determined by Weigert and his collaborators using myelomas induced in Balb/c mice [12, 3].

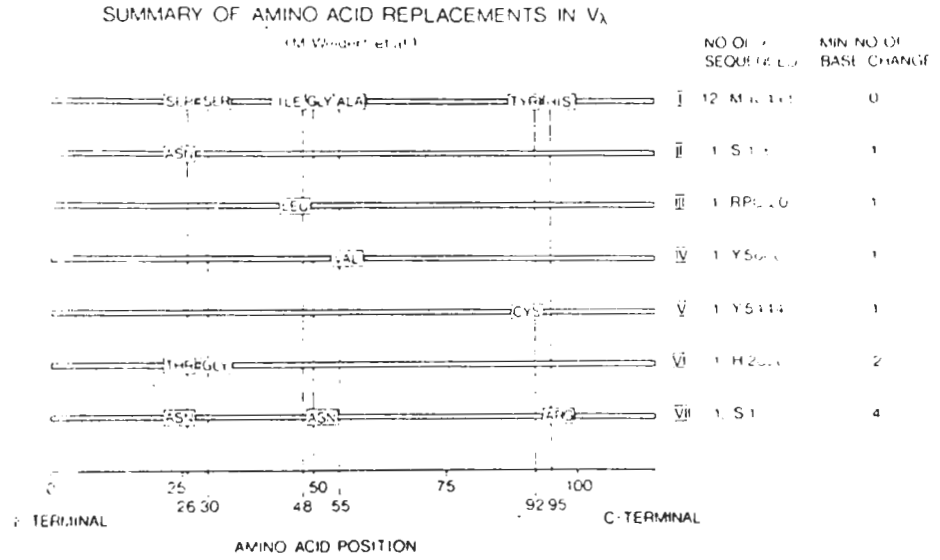

FIG. 1. — *Schematic representation of differences among λ-chains* (after Cohn *et al.* [3]).

Twelve out of eighteen λ-chains had an identical sequence (λ_0 type). The other six were not only mutually different but also differed from the λ_0 type by one to three amino acids. These substituted amino acids are without exception in the hypervariable regions. With one exception, the triplet codon for each of these amino acids is related to the codon for the corresponding amino acid in the λ_0 type by a single base change. Thus there seems to be an extensive nucleotide sequence homology among genes coding for the λ-V regions. If there is no somatic generation of diversity, there must be seven different λ-V genes in the mouse genome, and the DNA in all of these genes should hybridize with a single purified

mRNA of the λ_0 type. Analysis of the Cot curve would yield a reiteration frequency (F) of 7. On the other hand, if the λ_1-λ_6 sequences are generated from λ_0 type by a somatic process, then a Cot curve corresponding to a unique gene (F = 1) would be obtained. In reality, however, the number of different λ-V region sequences in the Balb/c mouse should be much higher than seven. This is clear if one makes a simple statistical calculation on the data shown in figure 1. Among the chain producing myelomas, those myelomas making the λ_0 type constitute the major group, whereas the rest constitute the minor group. The latter group is represented by the types λ_1-λ_6. One might ask what is the total number of different λ-V region sequences belonging to the minor group. This number of course cannot be less than six. Weigert and his collaborators have yet to find two independently induced tumors which synthesize an identical chain belonging to the minor group. If, however, there were only six different λ-V sequences in this group, the chance of having found six myelomas each with a different sequence would be 1.5 % $\left(\frac{6}{6} \times \frac{6}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times \frac{1}{6} = 0.015\right)$: a highly unlikely event. This probability is 50 % if there are 25 different λ-V sequences. In order to increase the chance to 90 %, as many as 145 different λ-V sequences are required.

Our task then is to purify mRNA coding for a λ_0 sequence and to hybridize this RNA to total mouse DNA of various sources and determine the number of cross-hybridizing genes per haploid genome. As pointed out in the previous section, we should also experimentally verify the assumption that λ_0 mRNA would cross-hybridize with other λ-V genes if they existed in each cell. If the number of cross-hybridizing genes turns out to be smaller than the expected number of different λ-V region sequences (25 or more), it will provide us with formal evidence for the somatic generation of antibody diversity.

Two mRNAs coding for two different Vλ regions were prepared to a purity of ninety per cent or higher. The sources of the mRNAs were the Balb/c plasmacytomas MOPC104E and HOPC 2020. The two V regions differ in two amino acids (fig. 1). Each amino acid substitution requires only a single base substitution. The Cot curve of M104E mRNA is illustrated in figure 2. A mixture of mouse α- and β-globin mRNAs as well as 18S ribosomal RNA were used as kinetic standards. The gene reiteration frequencies of these RNAs have previously been determined at less than three and at 200-250, respectively [4, 5]. The reiteration frequency of M104E λ chain gene is calculated to be two to three from the results represented in figure 2.

In the competition hybridization experiments illustrated in figure 3, a fixed amount of radioactive M104E mRNA and an excess of denatured liver DNA were annealed to a high Cot value (Cot = 10,000) in the presence of varying amounts of unlabeled H2020 mRNA (heterologous competition) or M104E mRNA (homologous competition). Radioactive M104E mRNA in the hybrid is plotted as a function of the ration of total

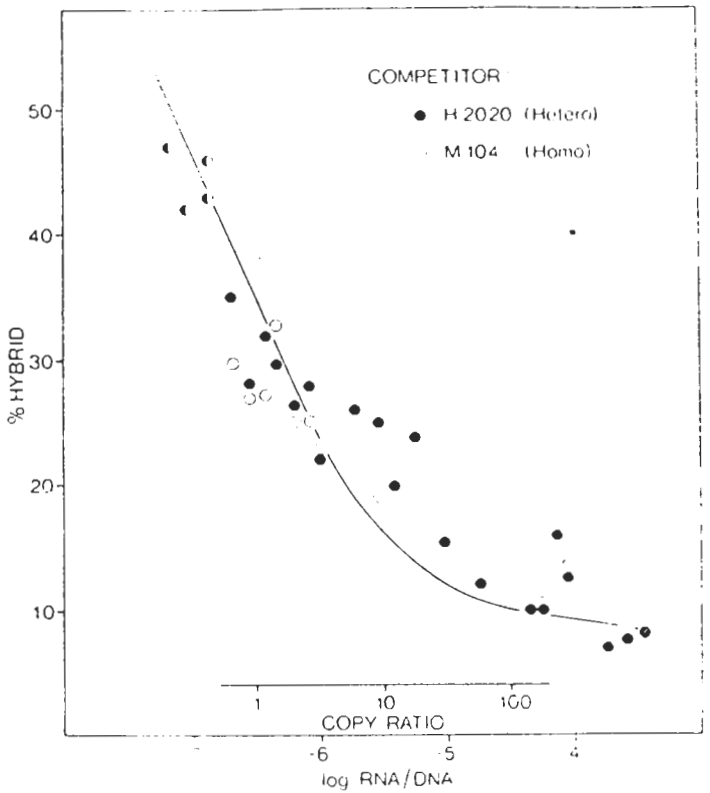

FIG. 3. — *Competition hybridization of λ-mRNAs.* The theoretical curve was derived by a computer program and represents homologous competition of a hypothetical RNA mixture containing 85 % unique, major component, and 15 % impurities assumed to consist of 1,500 different equimolar species.

genes are on the 5'-end and 3'-end halves of the mRNA molecules, respectively. The details and justification of such assay procedures are described elsewhere (Hozumi and Tonegawa, submitted).

Three DNA components of 4.6, 2.7 and 2 million MW hybridized with whole H2020 λ-mRNA. Only the largest one hybridized with the 3'-end half message (λ-mRNA lacking both the 5'-end untranslated region and the V region) (fig. 4A). Using λ-mRNA lacking the 5'-end untranslated region (5'E), only the two larger DNA fractions hybridized to this 3'-end three quarter message (fig. 4B). Hence, the Cλ gene is in the 4.6 million component, whereas the Vλ and 5'E gene are in the 2.7 and 2 million component respectively (Hozumi and Tonegawa, manuscript in preparation). The fact that the extent of hybridization in the larger component is

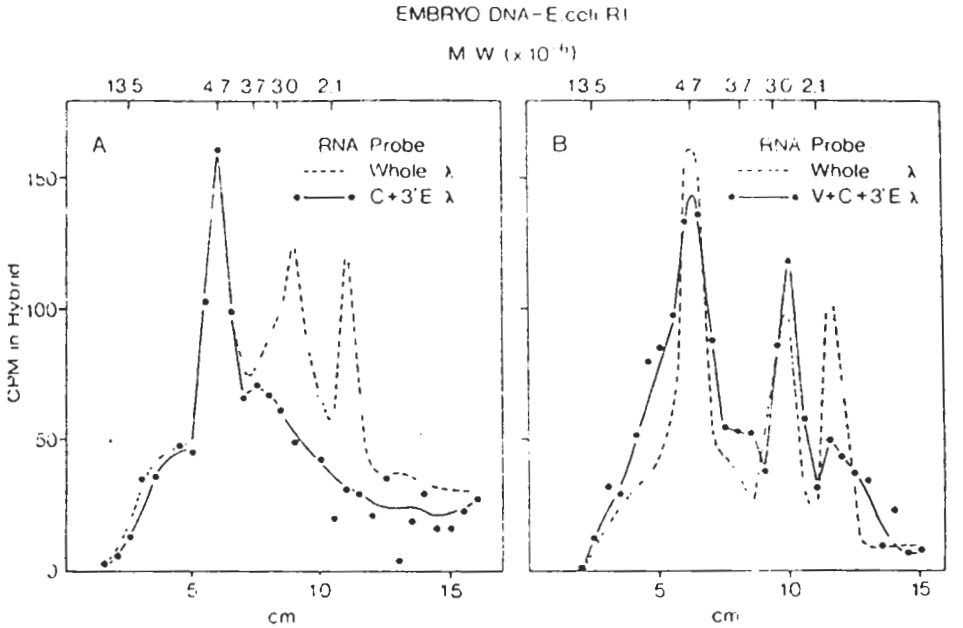

FIG. 4. — *Hybridization pattern of* E. coli *R-I digest of embryo DNA with H2020 λ-mRNA and the 3'-end half message (A), and with H2020 λ-mRNA and the 3'-end three quarter message (B).*

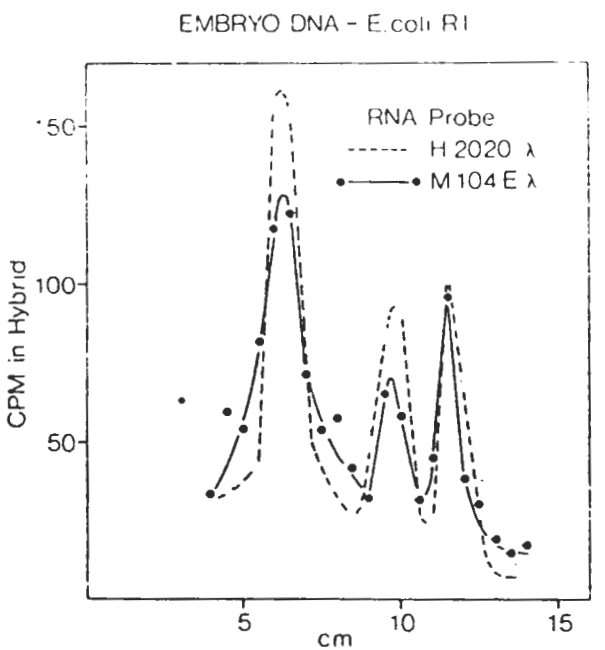

FIG. 5. — *Hybridization pattern of* E. coli *R-I digest of embryo DNA with H2020 and M104E λ-mRNA.*

In the third experiment, the DNA fragment carrying the Vλ gene and the Cλ gene were isolated and hybridized with an excess of M104E or H2020 λ-mRNA. The hybrids were then treated with S1-nuclease, which specifically digests away all single stranded polynucleotides. In this way, all DNA carrying sequence complementary to the λ-mRNA it was hybridized to, is protected from being digested. The protected DNA carrying the sequence of one of the two λ-mRNAs was then hybridized, to the other λ-mRNA. Therefore, if both mRNAs cross-hybridize, the extent of hybridizability should be identical when using homologous or heterologous λ-mRNA. That this is the case is shown in table I.

TABLE I. – Cross-hybridization of two lambda mRNAs.

DNA fragment	RNA		^{125}I cpm in 2nd hybridization
	1st hybridization	2nd hybridization	
C	M104E	M104E	299
		H2020	255
	H2020	M104E	285
		H2020	286
	None	M104E	81
		H2020	62
V	M104E	M104E	216
		H2020	192
	H2020	M104E	228
		H2020	210
	None	M104E	62
		H2020	61

Mouse embryo DNA (7 mg) was digested with Eco R-1 and fractionated on 0.9 % agarose gel. DNA fragments containing Cλ and Vλ genes were first hybridized to an excess of λ-mRNA in 6 × SSC/50 % formamide at 15° C ($Crt = 4.2 \times 10^{-2}$; $Cot = 10^2$, where C_0 is the equivalent concentration of unfractionated DNA). The hybrid was treated with 0.6 kU/ml of S1-nuclease in 0.3 M NaCl, 0.05M NaAc. 1 mM $ZnSO_4$, pH 4.6, for 30 min at 35°. The second hybridization was to λ-mRNA-I^{125} (5×10^7 cpm/μg RNA) and was carried out at 70° in 2 × SSC up to a Cot of 5,000. Input^{125} I-RNA was 1,500 cpm. The hybrid was then treated with RNAse-mix (RNAse A 20 μg/ml + T-1 RNAse 2 units/ml) at 37° for 20 min and TCA precipitated. Data represent the average of two measurements.

These results indicate that the nucleotide sequences coding for the V regions of the two λ-chains are so similar that if there were two separate germ line V genes, either one of the two λ-chain mRNAs must have cross-hybridized to the gene for the other. Given the extensive homology in the amino acid sequences (fig. 1), one can extrapolate this conclusion to other λ chains. Hence, the sequence homology among λ-mRNAs is so high that any λ-mRNA should cross-hybridize with all or most of the germ line Vλ genes, at least for those V regions which are represented in myelomas.

We have previously concluded [9] that the number of germ line V_κ genes is too small to account for the sequence diversity. The present result with λ reinforces and extends this conclusion: there are too few germ line genes for the « myeloma group » of λ chains to account for the observed diversity in this group.

SUMMARY

Hybridization of λ-mRNAs to excess liver DNA yielded results compatible with gene reiteration frequencies of three or less. Purified mRNA from tumors producing structurally different λ chains were used in competition hybridization experiments. An unlabeled λ-mRNA competed with another, labelled mRNA to the same extent as homologous unlabelled λ-mRNA. Mouse DNA was digested with Eco R-I restriction endonuclease and fractionated by gel electrophoresis. A DNA fragment carrying the V λ-gene(s) was identified in this digest. This fragment hybridized with λ-mRNAs coding for two different λV regions equally well. These results indicate that base sequence homology among λ-mRNAs is so high that any λ-mRNA should cross-hybridize with all or most of germ line V λ genes. From amino acid sequence data, it is argued that there are probably more than 25 different λV regions. Hence it is concluded that the number of germ line genes is too small to account for the diversity of λ chains.

KEY-WORDS: Antibody, V region, Immunogenetics; Hybridization, RNA, DNA, Diversity, Myeloma, Mouse.

ACKNOWLEDGMENTS

We acknowledge the expert technical assistance offered by Ms B. Durrer, Mr G. B. Dastoorniko, Ms S. Stutz and Ms R. Schuller.

REFERENCES

[1] BERNARDINI, A., & TONEGAWA, S., Hybridization studies with an antibody heavy chain m-RNA. *FEBS Letters*, 1974, 41, 73.

[2] CAPRA, J. D. & KEHOE, J. M., Antibody diversity: is it all coded for by the germ line genes? *Scand. J. Immunol.*, 1974, 3, 1.

[3] COHN, M., BLOMBERG, B., GECKELER, W., RASCHKE, W., RIBBLET, R. & WEIGERT, M., First order considerations in analyzing the generator of diversity, *in* « The immune system: genes, receptors, signals », (E. E., Sercarz, A. R. Williamson & C. R. Fox) (p. 89), Academic Press Inc., New York and London, 1974.

[4] HARRISON, R. P., HELL, A., BIRNIE, G. D. & PAUL, J., Evidence for single copies of globin genes in the mouse genome. *Nature* (Lond.), 1972, 239, 219.

[5] Melli, M., Whitfield, C., Rao, K. V., Richardson, M. & Bishop, J. O., DNA-RNA hybridization in vast DNA excess. *Nature New Biol.* (Lond.), 1971, 231, 8.

[6] Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M. & Helling, R. B., Replication and transcription of eukaryotic DNA in *Escherichia coli*. *Proc. nat. Acad. Sci.* (Wash.), 1974, 71, 1743.

[7] Schuller, R., Purification of immunoglobulin heavy chain m-RNA's and their use in hybridization studies. M. Sc. Thesis, University of Alberta, 1976.

[8] Thomas, M., Cameron, J. R. & Davis, R. W., Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. *Proc. nat. Acad. Sci.* (Wash.), 1974, 71, 4579.

[9] Tonegawa, S., Steinberg, C., Dube, S. & Bernardini, A. Evidence for somatic generation of antibody diversity. *Proc. nat. Acad. Sci.* (Wash.), 1974, 71, 4027.

[10] Tonegawa, S., Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proc. nat. Acad. Sci.* (Wash.), 1976, 73, 203.

[11] Wigzell, H. Antibody diversity: is it all coded for by the germ line genes? *Scand. J. Immunol.*, 1973, 2, 199.

[12] Weigert, M., Cesari, M., Yonkovich, S. J. & Cohn, M. Variability in the lambda light chain sequences of mouse antibody. *Nature* (Lond.), 1970, 228, 1045.