

In: *The Generation of Antibody Diversity: A New Look*,  
A. Cunningham (ed.), New York, Academic Press, 1976,  
pp. 175-182.

## Too many chains—too few genes

S. TONEGAWA and C. STEINBERG

*Basel Institute for Immunology, Basel, Switzerland*

We have been attempting to use DNA-RNA hybridization techniques to study the problem of the generation of antibody diversity. We ask the question: Does there exist a germ-line V gene for every immunoglobulin (Ig) V region? To answer this question, we count the number of V genes which will hybridize with an Ig mRNA. Experience shows that our immunological colleagues are not familiar with the techniques involved in determining gene reiteration frequencies by hybridization. Hence we ask the indulgence of our molecular biological colleagues while we attempt to condense the principles involved into a few paragraphs.

If double-stranded DNA is denatured and then allowed to anneal in the presence of (radioactively labelled) RNA, some RNA-DNA hybrid molecules will be formed, provided of course that the DNA contains nucleotide sequences homologous to the RNA. The RNA in hybrid is resistant to degradation by RNAase and is thus easily measured. Under a given set of annealing conditions, the fraction of RNA hybridized ( $f$ ) is dependent upon the RNA concentration ( $R_0$ ), the DNA concentration ( $C_0$ ), and the time of incubation ( $t$ ). If DNA is present in large excess,  $f$  is independent of  $R_0$ , and depends only on the product  $C_0t$  (universally pronounced to rhyme with "dot").

The value of  $C_0t$  necessary to achieve 50 per cent hybridization,  $C_0t_{1/2}$ , will depend upon the fraction of the DNA which is complementary to the RNA. If the haploid genome of a eukaryotic cell contains say, 1000 genes (DNA sequences), which will hybridize with a given RNA,  $C_0t_{1/2}$  will be 1000 times *smaller* than would be the case if the haploid genome contains only one gene which will hybridize to the RNA. This method, due to Bishop (Melli *et al.*, 1971), allows one to measure the reiteration frequency of any gene for which one has a pure, complementary RNA.

Although simple in principle, the technical requirements are exceedingly stringent for nonreiterated (unique) genes of mammals. Excess DNA means a weight ratio of DNA to RNA of the order of  $10^7$ . Using highly

radioiodinated RNA, this can be achieved. A vast excess of DNA is impossible. With ratios of DNA to RNA much less than  $10^7$ , hybridization to unique genes will not be detectable.

We will now turn our attention to the specific problem of antibody diversity. Figure 1 summarizes amino acid sequence data for the V region of

Summary of amino acid replacements in  $V_\lambda$   
(M. Weigert *et al.*)

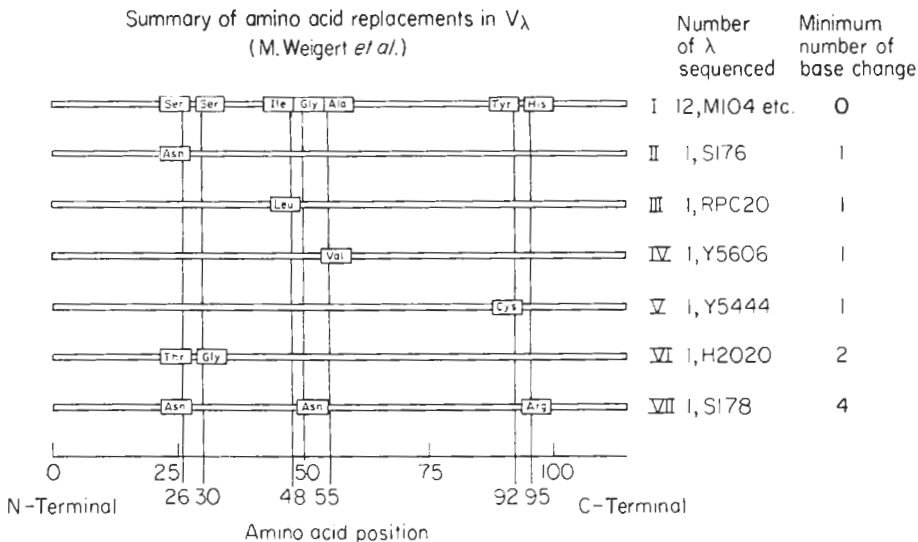


FIG. 1. Schematic representation of differences among  $\lambda$  chains. (After Cohn *et al.*, 1974.)

mouse  $\lambda$  chains. These sequences were determined by M. Weigert and his collaborators (Weigert *et al.*, 1970; Cohn *et al.*, 1974) using myelomas induced in Balb/c mice. Twelve out of eighteen  $\lambda$  chains had an identical sequence ( $\lambda_1$  type). The other six were not only mutually different but also differed from the  $\lambda_1$  type by one to three amino acids. These substituted amino acids are without exception in the hypervariable regions. The triplet codon for each of these amino acids except for one case is related to the codon for the corresponding amino acid in the  $\lambda_1$  type by a single base change. Thus there seems to be an extensive nucleotide sequence homology among genes coding for the  $V_\lambda$  regions. If the germ-line theory is correct, and there are seven different  $V_\lambda$  genes in each mouse genome, DNA in all of these genes will hybridize with a single purified mRNA coding for  $\lambda_1$  types. One would then obtain a  $C_0t$  curve which indicates that the sequence is seven times reiterated per haploid genome. On the other hand, if the  $\lambda_{II}$  to  $\lambda_{VII}$  sequences are generated from  $\lambda_1$  type by a somatic process, as suggested by the somatic theory of antibody diversity, then a  $C_0t$  curve

radioiodinated RNA, this can be achieved. A vast excess of DNA is impossible. With ratios of DNA to RNA much less than  $10^7$ , hybridization to unique genes will not be detectable.

We will now turn our attention to the specific problem of antibody diversity. Figure 1 summarizes amino acid sequence data for the V region of

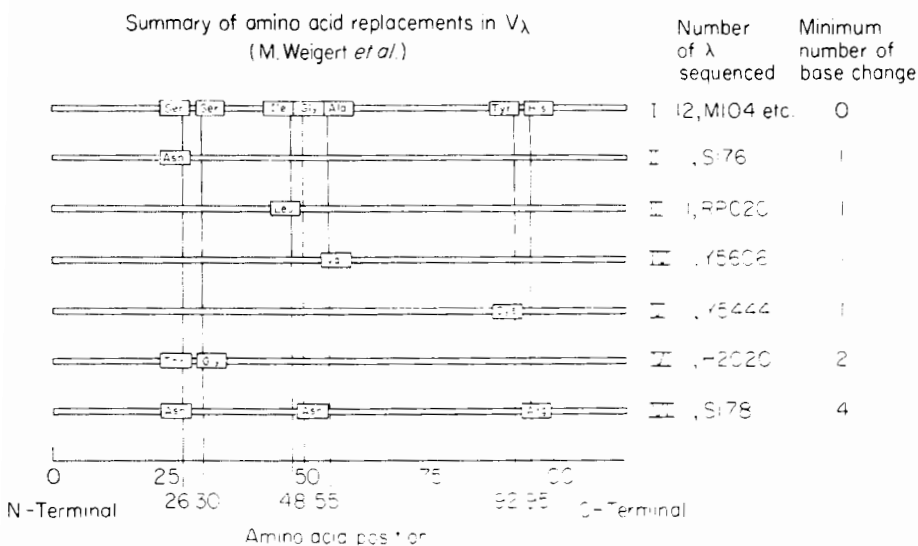


FIG. 1. Schematic representation of differences among  $\lambda$  chains. (After Cohn *et al.*, 1974.)

mouse  $\lambda$  chains. These sequences were determined by M. Weigert and his collaborators (Weigert *et al.*, 1970; Cohn *et al.*, 1974) using myelomas induced in Balb c mice. Twelve out of eighteen  $\lambda$  chains had an identical sequence ( $\lambda_I$  type). The other six were not only mutually different but also differed from the  $\lambda_I$  type by one to three amino acids. These substituted amino acids are without exception in the hypervariable regions. The triplet codon for each of these amino acids except for one case is related to the codon for the corresponding amino acid in the  $\lambda_I$  type by a single base change. Thus there seems to be an extensive nucleotide sequence homology among genes coding for the  $V_\lambda$  regions. If the germ-line theory is correct, and there are seven different  $V_\lambda$  genes in each mouse genome, DNA in all of these genes will hybridize with a single purified mRNA coding for  $\lambda_I$  types. One would then obtain a  $C_0t$  curve which indicates that the sequence is seven times reiterated per haploid genome. On the other hand, if the  $\lambda_{II}$  to  $\lambda_{VII}$  sequences are generated from  $\lambda_I$  type by a somatic process, as suggested by the somatic theory of antibody diversity, then a  $C_0t$  curve

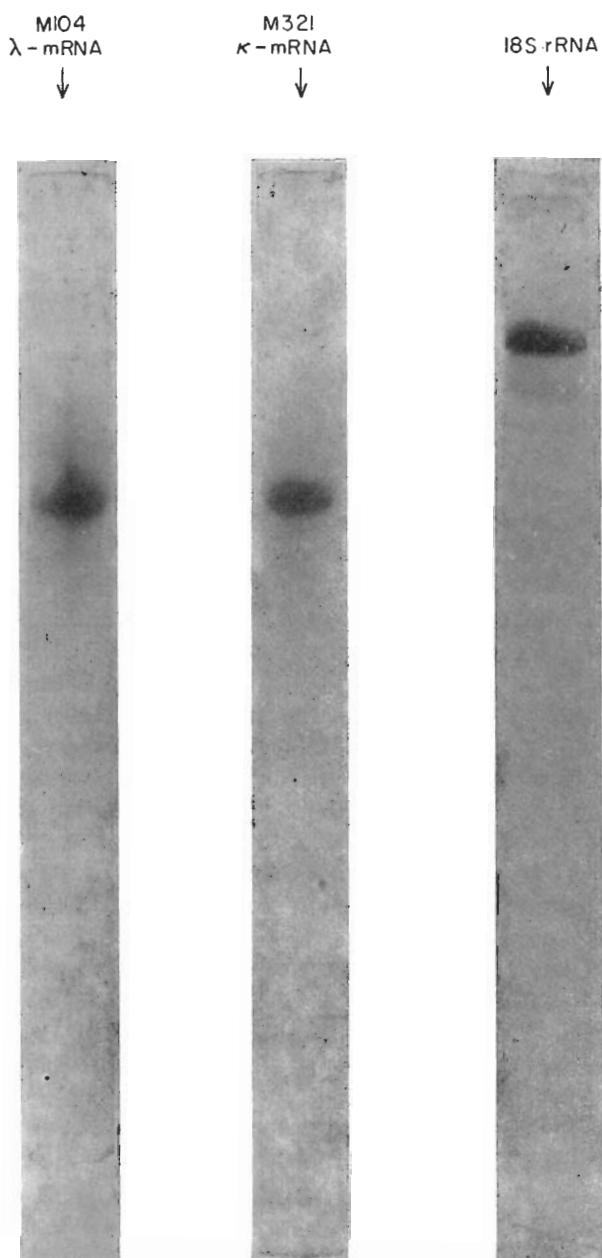


FIG. 2. Acrylamide gel patterns of mouse RNA preparations after electrophore

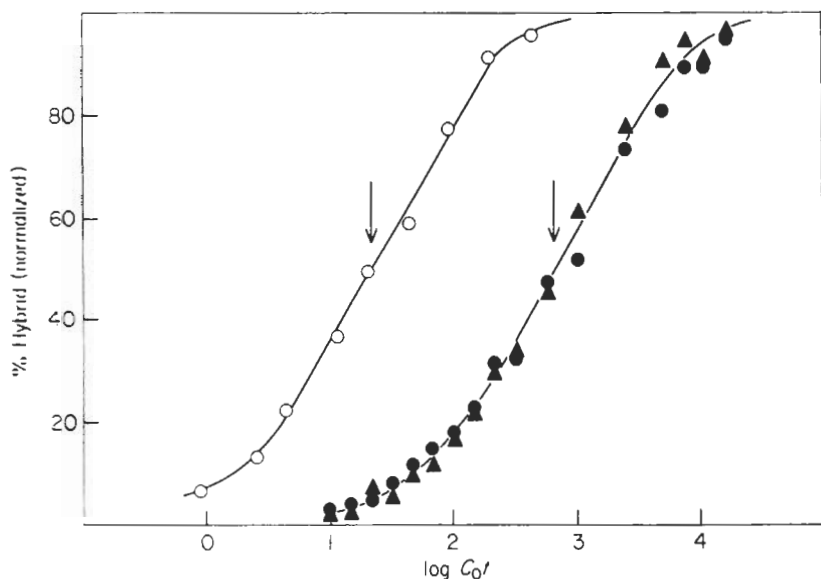


Fig. 3. Hybridization of iodinated mouse RNAs to mouse liver DNA. ● M-104Eλ mRNA; ▲ α and β globin mRNA; ○ 18S rRNA. The arrows mark values for  $C_0t_{1/2}$ .

and the real reiteration frequency is 2–3. In fact, when the experiments are carried out under similar conditions, the  $C_0t$  curves of λ mRNA and mouse globin mRNA are indistinguishable. The reiteration frequency of mouse globin genes were previously determined by Harrison *et al.* (1972) to be no more than 3. Figure 3 includes, as an additional standard, a  $C_0t$  curve with 18S rRNA. The nominal frequency is 200, in good agreement with reported estimates obtained by independent methods. These results indicate that the number of germ-line genes which cross-hybridize with MOPC-104E mRNA is no more than a few per haploid genome.

The remaining task then is to test the validity of the hypothesis that extensive nucleotide sequence homology exists among mRNAs coding for various chains. For this purpose a competition hybridization experiment was carried out. A mRNA coding for another λ chain was purified from HOPC-2020. As shown in Fig. 1, H-2020 λ chain is different from M-104E λ chain by two amino acids at only two positions. A series of tubes was set up. Each tube contained a fixed amount of  $^{125}\text{I}$  M-104 λ mRNA and excess denatured liver DNA. It also contained a varying amount of unlabelled H-2020 λ mRNA. The fraction of  $^{125}\text{I}$  RNA in the hybrid was determined after each mixture was incubated to a high  $C_0t$  ( $C_0t = 10\,000$ ). In Fig. 4 this was plotted as a function of the ratio of total RNA

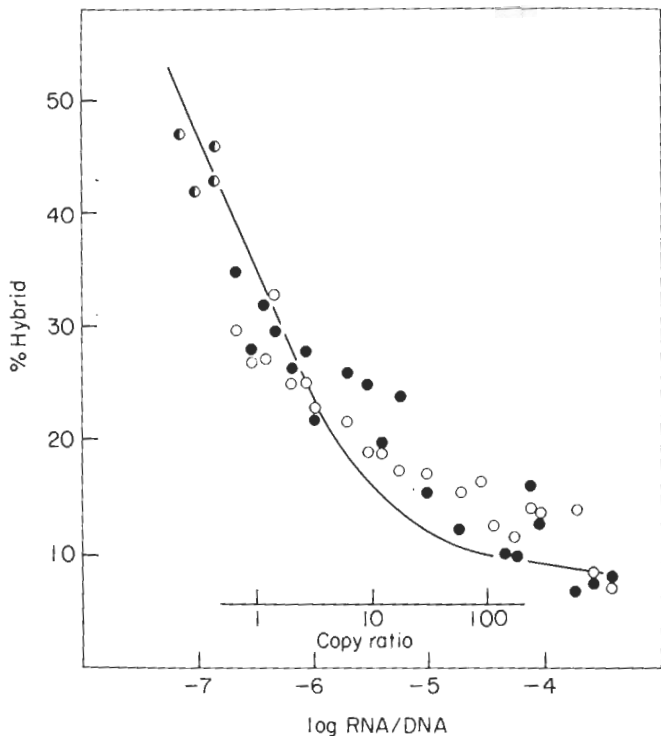


FIG. 4. Competition of hybridization of iodinated M-104E $\lambda$  mRNA by unlabelled mRNAs. ○ M-104E  $\lambda$  mRNA; ● H-2020  $\lambda$  mRNA.

to DNA. The points obtained with the homologous (M-104) and the heterologous competitors fell on the same line. We conclude that the nucleotide sequences coding for the V regions of the two chains are so similar that if there were two separate germ-line genes, either one of the two  $\lambda$ -chain mRNA 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  $\lambda$  chains. Hence the number of germ-line genes determined by nucleic acid hybridization (no more than a few) is too small to account for observed diversity of mouse  $\lambda$  chains (at least twenty-five, probably more). This is the evidence for the somatic theory of antibody diversity.

We have given a detailed presentation of the results of our work with mouse  $\lambda$  chain mRNA because they are the newest results and because they are free of complications. The  $C_0t$  curves shown in Fig. 3 show no indication of material which hybridizes to reiterated genes. In earlier work with a  $\kappa$  chain mRNA, some 15 per cent of the RNA in the prepara-

“recognition segments” flanking each hypervariable segment. Would such recognition segments be long enough for DNA-RNA hybrids to be stable? According to current ideas about recombination, the answer is yes and not coincidentally. For this reason, we feel that the insertion model is not likely to be compatible with our results.

## References

- CAPRA, T. D. and KEHOE, J. M. (1975). *Advan. Immunol.* **20**, 1-40.
- COHN, M., BLOMBERG, B., GECKELER, W., RASCHKE, W., RIBLET, R. and WEIGERT, M. (1974). In “The Immune System: Genes, Receptors, Signals” (E. E. Sercarz, A. R. Williamson and C. F. Fox, eds), p. 89. Academic Press, New York and London.
- HARRISON, R. P., HELL, A., BIRNIE, G. D. and PAUL, J. (1972), *Nature*, **239**, 219-221.
- MELLI, M., WHITFIELD, C., RAO, K. V., RICHARDSON, M. and BISHOP, J. O. (1971). *Nature New Biol.* **231**, 8-12.
- TONEGAWA, S. (1976). *Proc. Nat. Acad. Sci. U.S.A.* **73**, 203-207.
- TONEGAWA, S. and BALDI, I. (1973). *Biochem. Biophys. Res. Commun.* **51**, 81-87.
- TONEGAWA, S., STEINBERG, C., DUBE, S. and BERNARDINI, A. (1974). *Proc. Nat. Acad. Sci. U.S.A.* **71**, 4027-4031.
- WEIGERT, M., CESARI, M., YONKOVICH, S. J. and COHN, M. (1970). *Nature*, **228**, 1045-1047.