Immunological Rev. (1977), Vol. **36**
Published by Munksgaard, Copenhagen, Denmark
No part may be reproduced by any process without written permission from the author(s)

Dynamics of Immunoglobulin Genes

SUSUMU TONEGAWA, CHRISTINE BRACK, NOBUMICHI HOZUMI, GASTON MATTHYSSENS & RITA SCHULLER

INTRODUCTION

The immunoglobulin (Ig) genes provide a variety of problems which stimulate the curiosity of not only immunologists but also eucaryotic molecular biologists. These problems arise partly from the seemingly unorthodox manner in which these genes are expressed, and partly from the enormous diversity and unique characteristics observed in the primary structure of the gene products. One might include in these problems the genetic origin of antibody diversity, the two 'genes' – one polypeptide chain concept, the allelic exclusion, and the monoclonal 'switch' in the heavy chains. In spite of the availability of a large amount of information which had been accumulated by amino acid sequence analyses and genetic and cytochemical studies, it seemed certain that true understanding of these problems required direct analysis of Ig genes at the molecular level. It was with these considerations in mind that we started in 1972 to purify mRNA coding for Ig chains.

Purification of Ig mRNA

Readers are referred to our previous publications for details of purification procedures of Ig mRNA (Tonegawa & Baldi 1973, Tonegawa 1976). In Figure 1, we summarize the steps involved in the purification procedures. Among many factors which affect the purity of the final preparation, we consider the following to be most important.

Basel Institute for Immunology, 487 Grenzacherstrasse, Postfach, Basel 4005, Switzerland

Abbreviations: Ig, immunoglobulin; SSC, standard saline citrate (0.15M NaC1, 0.015M Na-citrate); *E coli, Escherichia coli*

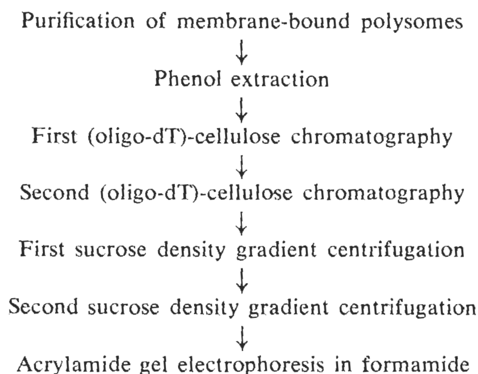

Figure 1. Steps involved in purification of Ig mRNA. See text for explanation.

1. *Myeloma lines*

The relative synthetic rate of Ig chains varies widely from one myeloma line to another. This presumably reflects the relative content of Ig mRNA among all mRNA species. Since the purification procedures depend on *non-specific* criteria such as sedimentation rate and electrophoretic mobility rather than *specific* criteria such as nucleotide sequence, relative content of Ig mRNA in the cell directly affects the purity of final preparations. We have found, for instance, that MOPC 21 and MPC 11 give final $\varkappa$ mRNA preparations which are of lower purity than MOPC 321.

2. *Integrity of Isolated Polysomes*

Ease or difficulty in preparing intact polysomes primarily depends on the growth conditions of tumors, which are to some extent characteristic of tumor lines. We dissect tumors while they are still relatively small (no more than one centimeter in diameter) and take great care in removing any necrotic parts. Before pooling, we examine each polysome preparation by sucrose density gradient centrifugation. With MOPC 321, HOPC 2020, MOPC 104E and several other myeloma lines, our success rate in getting 'good' polysomes is over 90 %.

3. *Heat Treatment*

One of the major causes of contamination is the tendency of Ig mRNA to aggregate with other RNAs, particularly ribosomal RNA. This aggregation is minimized by subjecting RNA samples to heat (70°C for 10 min) in a buffer of low ionic strength before oligo-(dT) cellulose chromatography

and sucrose density gradient centrifugation. Centrifugation profiles of the RNA fractions eluted from the second oligo-(dT) cellulose column are shown in Figure 2. The two peaks migrating at 13S and 16S contain $\varkappa$ and α chain mRNAs. Here, the effect of heat treatment is dramatically evident. The RNA fraction without heat treatment prior to two successive chromatographies clearly shows an additional peak of 18S ribosomal RNA, whereas the equivalent fraction with heat treatment shows no such peak. In addition, there is relatively more O.D.-absorbing material toward the bottom of the gradient in the untreated sample than in the treated sample.

4. *Gradient Centrifugation under Low Ionic Strength Conditions*

In order to minimize aggregation, we carry out centrifugation in a low salt buffer (Tris-HC1, 5 mM; Na_3 EDTA, 0.2 mM; Lithium dodecyl sulfate, 0.01 %) at low temperature (4°C). RNA samples are heat treated before layering on the gradient.

5. *Fraction Pooling*

We generally sacrifice yield for purity and pool fractions conservatively, both in the sucrose gradient centrifugation and in the gel electrophoresis. For instance, in a profile such as that shown in the middle panel of Fig-

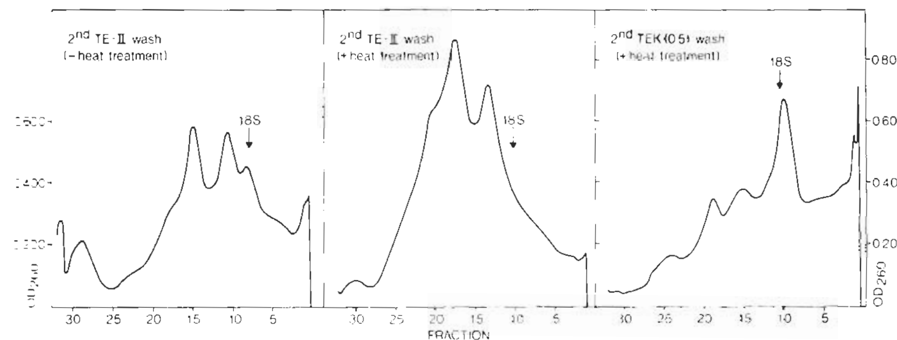

Figure 2. Effect of heat treatment on disaggregation of mRNA. Figures show profiles of sucrose density gradient centrifugation obtained with RNA fractions from the second oligo(dT)-cellulose column. Left panel; both first and second chromatography were carried out without prior heat treatment of RNA. The profile is of the fraction eluted with the low salt buffer (Tris HC1, 0.01M). Middle panel; both first and second chromatography were carried out with prior heat treatment of RNA. The profile is of the low salt fraction. Right panel; procedures are the same as for the middle panel. High salt (Tris-HC1, 0.01M; KC1, 0.5M) fraction. λ and α chain mRNA are contained in the OD peaks (fractions 15 and 11, respectively in the left panel, and fractions 17 and 13, respectively in the middle panel).

ure 2, we take no more than two fractions from the 13S $\varkappa$ mRNA peak.

During the mRNA purification procedures, we routinely use the cell-free translation system of Schreier & Staehelin (1973) for detection of Ig mRNA. With MOPC 321, HOPC 2020, and MOPC 104E the light chain mRNA preparation prior to the final gel electrophoresis step promotes synthesis of the corresponding pre-light chain. At this stage of purification, the pre-light chain is virtually the sole translation product when monitored by SDS-gel electrophoresis and tryptic peptide analysis (Schuller 1976, Bernardini & Tonegawa unpublished observation). When electrophoresed under denatured conditions, over 80 % of the RNA migrates as a single distinct band, the rest scattering on both sides of the band as faint background (Tonegawa 1976). In our earlier work with MOPC 70E, purity of $\varkappa$ mRNA preparation eluted from the major band was determined by two dimensional analysis of the T1-RNase digest (Tonegawa et al. 1974). We obtained a value of 70–80 % for this particular RNA preparation. Although we usually do not carry out such extensive purity analysis, we can get a purity estimate of a given light chain mRNA preparation relative to the 70E $\varkappa$ mRNA preparation. This is done by comparing the profiles of the second sucrose gradient centrifugation (the more dominant the 13S peak, the purer the final preparation) and the relative density of the major band in the acrylamide gel electrophoresis. With these criteria, most of the $\varkappa$ and λ chain mRNA preparations from MOPC 321, HOPC 2020, and MOPC 104E should be more than 90 % pure when eluted from the acrylamide gel. Most of the impurity consists of a large number of different mRNA species, each of which comprises a small fraction of total poly A-containing RNA in myeloma cells and happens to have sedimentation and migration properties very similar to those of Ig light chain. As already pointed out above, since our mRNA purification procedures rely on 'unspecific' criteria, it is not hoped to remove impurities of this nature from the final RNA preparation. Estimation of purity and characteristics of impurity described above are consistent with results of hybridization kinetics. Self competition experiments of a MOPC 104E λ mRNA preparation suggested that it contains an 85–90 % major RNA component and 10–15 % impurity which consists of over 1000 different RNA species (Tonegawa 1976). It is as important to remember these characteristics of mRNA preparations in designing various hybridization experiments as in evaluating data obtained from them.

Reiteration Frequency of $V\varkappa$ Genes

When mRNA preparations of sufficient purity were obtained, we studied

ure 2, we take no more than two fractions from the 13S κ mRNA peak.

During the mRNA purification procedures, we routinely use the cell-free translation system of Schreier & Staehelin (1973) for detection of Ig mRNA. With MOPC 321, HOPC 2020, and MOPC 104E the light chain mRNA preparation prior to the final gel electrophoresis step promotes synthesis of the corresponding pre-light chain. At this stage of purification, the pre-light chain is virtually the sole translation product when monitored by SDS-gel electrophoresis and tryptic peptide analysis (Schuller 1976, Bernardini & Tonegawa unpublished observation). When electrophoresed under denatured conditions, over 80 % of the RNA migrates as a single distinct band, the rest scattering on both sides of the band as faint background (Tonegawa 1976). In our earlier work with MOPC 70E, purity of κ mRNA preparation eluted from the major band was determined by two dimensional analysis of the T1-RNase digest (Tonegawa et al. 1974). We obtained a value of 70–80 % for this particular RNA preparation. Although we usually do not carry out such extensive purity analysis, we can get a purity estimate of a given light chain mRNA preparation relative to the 70E κ mRNA preparation. This is done by comparing the profiles of the second sucrose gradient centrifugation (the more dominant the 13S peak, the purer the final preparation) and the relative density of the major band in the acrylamide gel electrophoresis. With these criteria, most of the κ and λ chain mRNA preparations from MOPC 321, HOPC 2020, and MOPC 104E should be more than 90 % pure when eluted from the acrylamide gel. Most of the impurity consists of a large number of different mRNA species, each of which comprises a small fraction of total poly A-containing RNA in myeloma cells and happens to have sedimentation and migration properties very similar to those of Ig light chain. As already pointed out above, since our mRNA purification procedures rely on 'unspecific' criteria, it is not hoped to remove impurities of this nature from the final RNA preparation. Estimation of purity and characteristics of impurity described above are consistent with results of hybridization kinetics. Self competition experiments of a MOPC 104E λ mRNA preparation suggested that it contains an 85–90 % major RNA component and 10–15 % impurity which consists of over 1000 different RNA species (Tonegawa 1976). It is as important to remember these characteristics of mRNA preparations in designing various hybridization experiments as in evaluating data obtained from them.

Reiteration Frequency of Vκ Genes

When mRNA preparations of sufficient

the first problem listed above; namely, whether there is a separate germ line gene for every Ig chain. For this we measured the number of variable region (V) genes by RNA-DNA hybridization. In this approach the experimental steps and logic listed below were followed:

1. Purify mRNAs coding for particular light or heavy chains from mouse myeloma cells.

2. Radioiodinate the purified mRNA to a specific activity higher than 10 million cpm per microgram.

3. Determine the reiteration frequency of the DNA sequences complementary to the mRNA. This was done by annealing under proper conditions the radioiodinated RNA with denatured, sonicated liver DNA (liver is in this case a surrogate of germ line cells), and then determining, as a function of time, the fraction of RNA in the hybrid by measuring the resistance of hybridized RNA to RNase. Under a given set of annealing conditions, the fraction of hybridized RNA (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 of C and t (C_0t). The value of C_0t necessary to achieve 50 % hybridization, $C_0t_{1/2}$, will depend upon the fraction of the DNA which is complementary to the RNA. The reiteration frequency (F) of the gene in question is calculated by the formula

$$F = F^* \frac{C_0t_{1/2}^*}{C_0t_{1/2}} \frac{C}{C^*}$$

where C designates the genome complexity, and * designates a gene used for standardization and whose reiteration frequency F* is known by an independent method (Melli et al. 1971).

4. Carry out competition hybridization experiments by annealing the radioiodinated RNA used in the C_0t curve experiment in the presence of varying amounts of unlabeled mRNA coding for another V region. The purpose of the competition experiments is to determine the degree of base sequence homology between the two RNAs, and consequently between the two corresponding V genes. The results will tell us whether and how much the radioiodinated mRNA used in the C_0t curve experiment would cross-hybridize with other V genes if they existed as separate germ line genes.

5. Estimate, on the basis of such competition experiments and on the available amino acid sequence data, the number of different V regions whose genes would have cross-hybridized with the particular mRNA used in the C_0t curve experiment. If this number is significantly larger than the experimentally determined reiteration frequency, it constitutes formal evidence for somatic generation of antibody diversity.

We initially applied this approach to mouse $\varkappa$ chains. In this work we could define a group of $\varkappa$ chains, comprising 10 per cent of all such chains for which sequence information was available, within which there was over 80 % amino acid homology. DNA sequences coding for the V regions of one $\varkappa$ chain from this group were reiterated only 1–2 fold (Tonegawa et al. 1974). Competition hybridization experiments showed that there was extensive, though not complete, cross-hybridization. Hence, we argued that there are more chains in this group than there are germ line genes. The argument, however, was admittedly less forceful when judged by present standards. The primary problems stemmed from the impurity in the mRNA preparation (MOPC 70E mRNA was 70–80 % pure) and from the relatively high degree of variability in mouse $\varkappa$ chains. (In more recent experiments with purer ($>$ 90 %) MOPC 321 $\varkappa$ mRNA, we confirmed the conclusion that DNA sequences coding for entire $\varkappa$ mRNA are reiterated no more than a few times (Tonegawa 1976)).

An important observation made in this initial series of hybridization experiments is that, in conformity with the amino acid sequence data, while two polynucleotide chains coding for two V regions belonging to a single subgroup* or two closely related subgroups will cross-hybridize extensively, those coding for two V regions belonging to two different subgroups will in most cases give little cross-hybridization under the relatively stringent conditions (2 × SSC, 70°C) of hybridization. Relaxation of hybridization conditions will undoubtedly promote intersubgroup cross-hybridization. But hybrids formed under these conditions will contain many mismatched base pairs. Base mismatching causes reduction in hybrid formation, the degree of which would be difficult to estimate accurately, particularly when the extent of mismatching is large (Bonner et al. 1973). Furthermore, it would be difficult to make any meaningful estimate of the fraction of subgroups whose V genes would have been counted under the particular hybridization conditions employed. Thus, given a relatively high degree of variability in mouse $\varkappa$ chains, estimation of the *total* number of germ line V genes is virtually impossible. The strategy of choice then, was to retain the stringent hybridization conditions and estimate the number of V genes for a single subgroup.

Reiteration Frequency of Vλ Genes

With these considerations in mind, we turned our effort to mouse λ chains,

* See later section for definition of subgroup.

where the largest set of sequence information is available for a single subgroup, thanks to the extensive work of Weigert and his colleagues (Weigert et al. 1970). We could also increase the purity of mRNA preparation by judicious choice of myelomas and by improvement in purification procedures. The results demonstrated that the DNA sequence coding for the V region of MOPC 104E λ chain is reiterated no more than three fold. There were essentially complete cross-hybridizations between λ mRNAs from MOPC 104E and HOPC 2020 (Tonegawa 1976).

All hybridization experiments described above were carried out using whole λ mRNA and sonicated, whole mouse liver or myeloma DNA. Since mRNA molecules consist of both V and C region sequences in roughly equal amounts (Rabbitts & Milstein 1975), a maximum of only 50 % of the hybridized RNA are attributed to V region sequences. Although the C_0t curve obtained exhibited no sign of biphasicity, it was nevertheless desired for additional assurance to carry out hybridization exclusively for

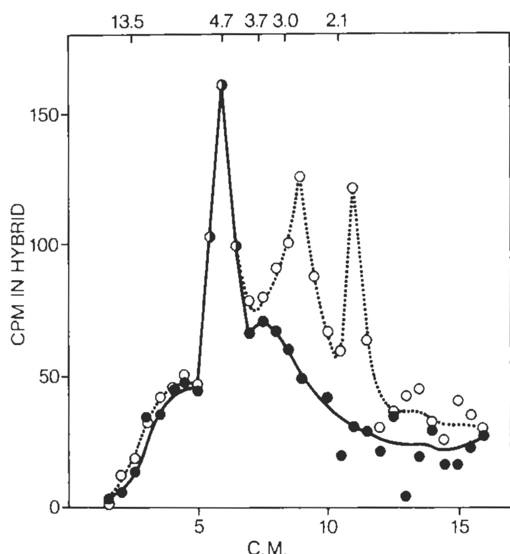

Figure 3. Gel electrophoresis patterns of embryo DNA digested with *Eco* RI enzyme and assayed with H2020 λ chain mRNA and its 3′ end-containing fragments. Embryo DNA was digested in R-I buffer (0.1M Tris-HC1 pH 7.5, 0.01M $MgSO_4$, 0.05M NaC1, 0.1 mM Na_2 EDTA) with *Eco* RI enzyme (purified according to Greene et al. (1974)) at 37°C for 4 h. Electrophoresis and hybridization were carried out as described elsewhere (Hozumi & Tonegawa 1976). A single preparation of ^{125}I-H2020 λ chain mRNA (specific activity 5×10^7 cpm/μg) was used for hybridization and for preparation of 3′-end containing fragments. Input ^{125}I counts per hybridization were 1300 cpm (whole mRNA) (○——○) and 620 cpm (3′-end half, ●——●). In this and Figures 5, 6a and 6b, numbers on top of the panel indicate the molecular weights (in millions) of DNA fragments used as migration markers.

V region sequences. This became possible when we found that embryonic DNA digested with a bacterial restriction enzyme *Eco* RI (see Figure 3) can be fractionated into portions which are free of C region sequences and others which are free of V region sequences. In Figure 3, hybridization patterns of electrophoretically fractionated DNA fragments are illustrated. Two RNA probes, whole 104E mRNA and its 3′ end-containing half, were used. The latter RNA lacks V gene sequences and contains a specific hybridization probe for C gene sequences (Tonegawa 1976). Figure 3 indicates that the slowest migrating component (M.W. = 4.7×10^6) carries only C genes while the other two components (M.W. = 2.8 and 1.8×10^6) carry V genes. The hybridization kinetics obtained using these fractionated DNA components are illustrated in Figure 4. The $C_0t_{1/2}$ values of these $C_0t_{1/2}$ curves indicate that neither V_λ nor C_λ genes are reiterated more than three fold per haploid genome. During this series of hybridization experiments, further evidence was obtained for extensive nucleotide sequence homology among the V regions of λ chains. In addition, it was demonstrated that C_0t curve analysis can distinguish between two genes

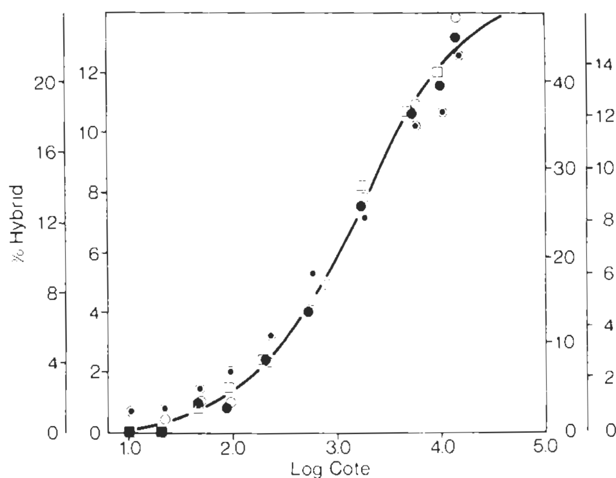

Figure 4. Kinetics of hybridization of ^{125}I-M104E λ mRNA with whole or fractionated embryo DNA. Unfractionated (◯) embryo DNA and the three DNA components described in Figure 3 (C, ⊙; V_1, ●; and V_2, □) were hybridized with a single preparation of ^{125}I-M104E mRNA (specific activity 4×10^7 cpm/μg) in 2×SSC plus 0.01M Tris-HC1, pH 7.0 at 70°C. In order to make the results obtained with all DNA fractions comparable, hybridization levels are plotted against C_0t values corrected by sequence enrichment factor *e* (Tonegawa et al. 1976). Ordinate scales are: unfractionated DNA, right inside; C gene fragment, left outside; V_1 gene fragment, right outside; and V_2 gene fragment, left inside.

which differ by as little as three fold in their reiteration frequency (Matthyssens et al. 1976, Tonegawa et al. 1976).

Weigert and his colleagues have determined the amino acid sequences of a number of Balb/c and NZB mouse myeloma λ chains. They have already found nine different V_λ region sequences (Weigert et al. 1970, Weigert personal communication). These V_λ region sequences are as homologous to that of M104E as is H2020, suggesting that the corresponding V genes cross-hybridize with M104E λ mRNA. We have shown that there are only a few germ line V_λ genes, using a method which can distinguish a three fold difference in gene reiteration frequency. Thus, the experimentally determined gene number seems to be smaller than the number of different V_λ regions we already know. Furthermore, a simple statistical calculation which can be applied to the sequence data, and which is described elsewhere (Tonegawa 1976), indicates that the repertoire of mouse V_λ regions is much larger than the nine known sequences. The results thus strongly suggest the existence of somatic processes by which V_λ region sequences are diversified from a few germ line genes.

Reiteration Frequency of Heavy Chain Genes

Our initial hybridization studies of heavy chains were carried out using an MPC 11 γ2b mRNA preparation, which was estimated by a self-competition experiment to be about 50 % pure (Bernardini & Tonegawa 1974). Contrary to the results of Premkumar et al. (1974) which were the only heavy chain data then available, we could not detect a fraction whose DNA sequences were reiterated several thousand fold. About 77 % of the hybridized RNA was complementary to unique DNA sequences, while the remainder seemed to be derived from genes reiterated about 300 times. These studies were later extended to μ and α chains using mRNA preparations of higher purity (Schuller 1976). Monophasic C_0t curves were obtained with MOPC 104E μ mRNA and HOPC 2020 α mRNA. Nominal reiteration frequencies of 8 and 7, respectively, were obtained with the μ and α genes. These results suggest that the moderately reiterated fraction previously observed with γ2b mRNA is due to impurity. Although paucity of heavy chain sequence data prevented us from carrying out more extensive hybridization experiments, the striking similarity of these results with those obtained using light chain mRNAs strongly suggests that there is no basic difference between light and heavy chains in the way in which V sequences are diversified.

How Many V Genes in total?

The hybridization studies described in the last few sections demonstrated that groups of closely related V regions are somatically generated from a few germ line genes. They do not, however, give us any reliable estimate of the *total* number of germ line genes. Given the enormous diversity of V regions, however, the existence of multiple germ line genes is a highly likely possibility. This notion is also suggested by the genetic analysis of idiotypic markers on mouse heavy chains (Eichmann 1975, Weigert et al. 1975).

A direct demonstration of multiple V genes for mouse $\varkappa$ chains is illustrated in Figure 5, which shows hybridization patterns obtained with mouse embryo DNA and whole $\varkappa$ RNAs from MOPC 21 and MOPC 321. These two $\varkappa$ chains show little homology in their V regions (Milstein & Svasti 1971, McKean et al. 1973) – they belong to two different 'subgroups' – whereas they have identical sequences in the C region. Both MOPC 321 and MOPC 21 mRNAs hybridized with the 6.0 million component. This component carries the $C_\varkappa$ gene as demonstrated by its hybridization with

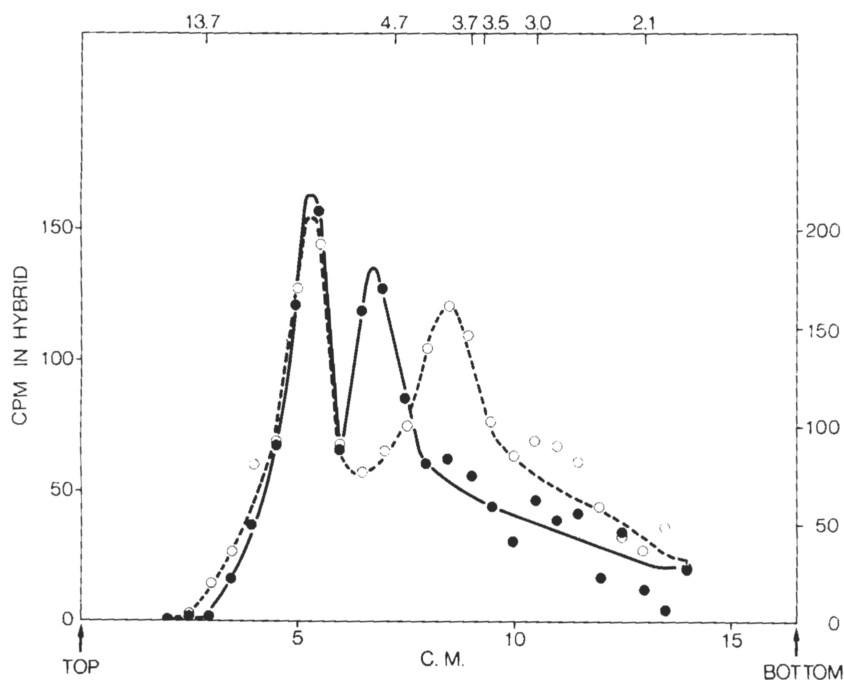

Figure 5. Gel electrophoresis patterns of embryo DNA digested with *Bam* H-I and analyzed with MOPC 321 (○——○) and MOPC 21 $\varkappa$ (●——●) mRNAs.

3′ end-containing half of MOPC 321 mRNA (Hozumi & Tonegawa 1976). MOPC mRNA hybridized also with a second component of 5.0 million to which MOPC 321 mRNA exhibited no hybridization above the background level. This component does not hybridize with the 3′ end-containing half (data not shown) and should contain MOPC 21 V gene. Conversely, the second component which hybridized with MOPC 321 mRNA (M.W. = 3.9×10^6), did not hybridize with MOPC 21 $\varkappa$ mRNA, and contains MOPC 321 V gene. Thus the results illustrated in Figure 5 not only confirm that there exists a single common $C_\varkappa$ gene, but also directly demonstrate that the two V regions are derived from two separate genes.

Amino acid sequences of immunoglobulin chain V regions can be classified into 'subgroups' such that the variability within a subgroup is less than in the group as a whole. Obviously, the greater the number of subgroups, the less the variability within the subgroups. The precise degree of variability tolerated within a subgroup is usually rather arbitrary. Cohn et al. (1974), however, have defined subgroups in a less arbitrary fashion: if two V regions differ by even one 'framework' residue (i.e., a residue outside the hypervariable regions), they belong to different subgroups. According to Cohn, all known V regions associated with the λ_I chain belong to a single subgroup. We have shown that there is probably only one germ line V gene for λ_I. MOPC 321 and MOPC 21 $\varkappa$ chains belong to different subgroups, and we have shown that their $V_\varkappa$ genes are derived from different germ line genes. It will be interesting to see how far this correlation between genes and subgroups is maintained. We think that the number of subgroups *à la* Cohn is an upper limit for the number of germ line V genes. If this is so, the mouse genome contains from 2 to 100 $V_\varkappa$ genes.

Two Separate DNA Segments, V and C, For One Polypeptide Chain

Bacterial restriction endonucleases described above recognize and cleave specific sequences of base pairs within DNA duplexes. For instance, the enzyme *Eco* RI cleaves DNA, regardless of its source, whenever it encounters the sequence,

(5′)GAATTC(3′)
(3′)CTTAAG(5′)

(Greene et al. 1974). The amino acid sequence studies by Weigert et al. indicate that this sequence cannot arise in the translatable part of λ_I mRNAs (Weigert et al. 1970). Yet the hybridization patterns shown in Figure 3 indicate that $V\lambda_I$ and $C\lambda_I$ sequences are on separate DNA fragments. A conclusion drawn from these results is that the two sequences are some

distance away from each other (see a later section for further e

Unfortunately, the same argument could not be applied to the lent results obtained with $\varkappa$ chains and illustrated in Figure 5. Here, the amino acid sequences of MOPC 321 nor MOPC 21 $\varkappa$ chain r cleavage of a continuously arranged whole $\varkappa$ gene. However, anoth of argument partly based on the results shown in Figure 5 suppo concept of separate $V_\varkappa$ and $C_\varkappa$ DNA segments. Namely, the total n of $V_\varkappa$ genes seems to be far greater than that of $C_\varkappa$ genes (see pre section). Unless an elaborate DNA network is postulated it is impo to arrange all $V_\varkappa$ genes immediately adjacent to fewer (probably single genes.

Somatic Joining of V and C Genes

The demonstration of separate V and C DNA segments in embryonic ce revives the now classical Dreyer & Bennett two 'gene'-one polypepti chain hypothesis (Dreyer & Bennett 1965). How is the information in th two separate DNA segments integrated? Since V and C gene sequence exist in a single mRNA molecule as a continuous stretch (Milstein et al 1974), such integration must take place either directly at the DNA or RNA level. Integration at the RNA level could result from a 'copy-choice' even during transcription or from joining of two RNA molecules after tran scription.

We studied this problem in a manner similar to the ones shown in Fig ures 3 and 5 (Hozumi & Tonegawa 1976). This time we compared hy bridization patterns obtained with DNA of various sources. Our hope wa that if there is rearrangement of specific Ig genes during differentiation o lymphocytes, we may be able to detect it as a difference in hybridizatio patterns of DNAs from embryonic cells (surrogate of germ line cells) and plasmacytomas (surrogate of plasma cells). A result of such an experimen is shown in Figures 6A and 6B. Here, DNAs from 12-day-old Balb/c mic embryos and MOPC 321 myeloma were separately digested to completio with *Bam* H-I enzyme. MOPC 321 $\varkappa$ mRNA and MOPC 104E λ mRNA were used as the RNA probes. As already mentioned in an earlier sectio the embryonic DNA patterns contain two components which hybridize with the $\varkappa$ mRNA. These are the 6.0 million $C_\varkappa$ gene component and th 3.9 million $V_\varkappa$ gene component. The pattern of hybridization with the $\varkappa$ mRNA was completely different in the MOPC 321 DNA. A new majo component of 2.4 millions appeared in the pattern. This component hy bridized with both whole MOPC 321 $\varkappa$ mRNA and its 3′ end-containin half (Hozumi & Tonegawa 1976). In contrast, hybridization patterns of th

two DNAs with the λ mRNA are virtually indistinguishable. These results are best interpreted in the following manner: the $V_{\varkappa}$ and $C_{\varkappa}$ genes, which are some distance away from each other in the embryo genome, are brought together, during differentiation of lymphocytes, to form a continuous DNA stretch.

An alternative explanation of the results, namely that accumulation of mutations or base modifications leading to either loss or gain of *Bam* H-I sites generated the observed pattern difference, is not impossible. On this view, there would have to be a *Bam* H-I site close to the V-C junction in embryo DNA. This *Bam* H-I site would have to be lost by mutation or by base modification in the MOPC 321 tumor. By itself, such an alteration

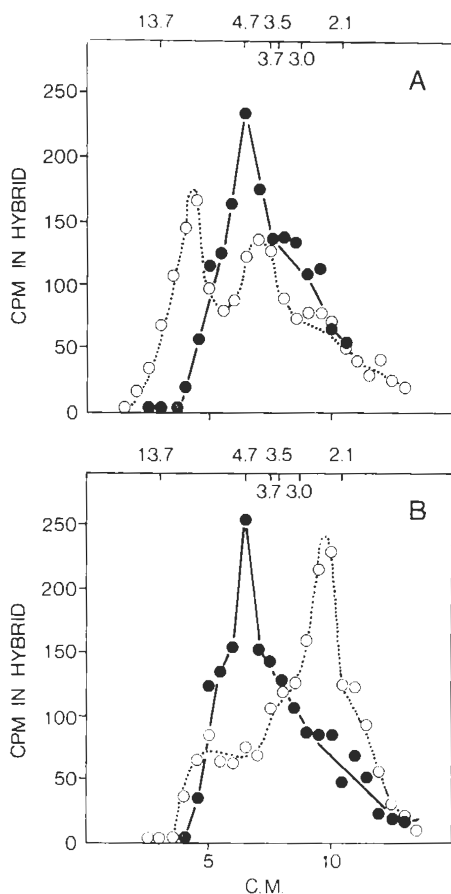

Figure 6. Gel electrophoresis patterns of embryo (a) and MOPC 321 (b) DNA's digested with *Bam* H-I and analyzed with MOPC 321 ϰ mRNA (○——○) or MOPC 104E λ mRNA (●——●).

TABLE I

Major Bam *H-I DNA fragments which hybridize with Ig mRNAs*

DNA / RNA	Embryo	κ myelomas			Kidney	λ myelomas	
		MOPC 321	TEPC 124	MOPC 21		J 558	MOPC 104E
MOPC 321 (whole)	6.0, 3.9[a]	2.4	6.0, 3.7, 2.4	3.9, 3.1	6.0, 3.9[a]	6.0, 3.9[a]	6.0, 3.9[a]
MOPC 321 (3′-end half)	6.0	2.4	6.0, 2.4	3.9, 3.1	6.0	N.D.	6.0
MOPC 21 (whole)	5.0, 5.0	5.0, 2.4	N.D.	3.9, 3.1	6.0, 5.0	N.D.	N.D.
MOPC 104E (whole)	4.8[a]	4.8[a]	N.D.	4.8[a]	N.D.	N.D.	4.8[a]

Numbers given are size of DNA fragments (in millions of daltons)
[a]: there may be additional minor fragments
N.D.: means *not determined*

would cause the appearance of a single 9.9 million M.W. component in the tumor. To achieve the M.W. of the single component actually observed in the tumor (2.4 million), there would have to be new *Bam* H-I sites created by mutation between the $V_{\varkappa}$ gene and the nearest site on either side. Since there is no reason why there should be any selective pressures involving *Bam* H-I sites, the occurence of three base alterations would seem to be quite unlikely. In addition, the constancy of λ mRNA patterns described above opposes the view that the dramatic difference in the two κ mRNA patterns arise from massive scrambling of DNA sequences in the myeloma cells.

In order to further eliminate such trivial explanations and gain more information on the putative translocation event, we extended the kinds of analysis described above to other combinations of DNA and RNA (Tonegawa et al. 1977). Although experiments along this line are incomplete, we summarize in Table I the present state of affairs. Below, we list points of special interest which emerged from these experiments.

1. Patterns of hybridization of DNA from adult non-lymphoid tissues are indistinguishable from the corresponding patterns of embryonic DNA.

2. Patterns of hybridization of DNAs from λ chain synthesizing myeloma

as obtained with a $\varkappa$ mRNA are also indistinguishable from the corresponding patterns of embryonic DNA. These results fortify the argument, presented above, against the hypothesis that the observed difference in $\varkappa$ mRNA hybridization is due to either fortuitous base changes in the cleavage sites or massive scrambling of DNA sequences. These results, together with those of the reciprocal experiment described in Figure 6, indicate that the V-C gene joining is mutually exclusive between $\varkappa$ and λ chains, just as the expression of these two light chains is.

3. The V gene of a $\varkappa$ chain of another subgroup, such as MOPC 21 $\varkappa$ chain, remains unjoined with the $C_\varkappa$ gene in MOPC 321 myeloma. Thus, mutual exclusiveness of V-C joining seems to be also applicable to two V genes coding for two V regions of different subgroups.

4. DNA from TEPC 124 myeloma (this and MOPC 321 myeloma synthesize $\varkappa$ chains of the same subgroup (McKean et al. 1973)), when analyzed with either homologous TEPC 124 or heterologous MOPC 321 $\varkappa$ mRNA, exhibited an identical hybridization pattern. The patterns contained a 2.4 million component which was indistinguishable from the one obtained in MOPC 321 DNA-RNA combination (see Figure 6B). In addition, these patterns contained two DNA components of 6.0 and 3.9 million M.W. which are again indistinguishable in size and hybridization patterns from those observed in embryo DNA-MOPC 321 $\varkappa$ mRNA combination (see Figure 6A. DNA from MOPC 21 myeloma (this and MOPC 321 myeloma synthesize $\varkappa$ chains of different subgroups), when analyzed with the homologous $\varkappa$ mRNA, exhibited two components of 3.9 and 3.1 millions M.W. Both of these components contain $C_\varkappa$ gene sequences and at least one contains $V_{\varkappa\text{-}21}$ gene sequences. Note that neither of these components was observed in embryonic DNA (Figure 5). Conversely, the 6.0 million C gene component observed in the embryonic DNA was not detected in MOPC 21 DNA.

One rule which can be extracted from this rather complicated body of information is that the new DNA component which appears in the various myeloma cells as a result of V-C joining, is subgroup specific. This is in perfect conformity with the conclusion drawn in earlier sections of this article; V regions belonging to a single subgroup share the same germ line V gene, and there are separate germ line V genes for two regions belonging to two different subgroups. There is a single $C_\varkappa$ gene which is shared by presumably all $V_\varkappa$ genes.

Mechanism of V-C Joining

In the past, several models have been proposed for the integration of and C gene sequences at the DNA level. We have discussed elsewhe each of these models with respect to its compatibility to our experiment observations (Tonegawa et al. 1976). Unfortunately, there still remain sev eral compatible models. Whichever model turns out to be correct, w believe that the joining event requires a site specific recombination systen analogous to the *int* system of phage λ (Campbell 1962).

V-C Gene Joining and Activation of V Gene

A committed B-lymphocyte or plasma cell produces antibody of only one specificity (Pernis et al. 1965, Mäkelä 1967, Raff et al. 1973). In particular, it expresses only one light chain V gene. Since there are not only V_{λ} and V_{κ} genes but also multiple V_{κ} genes, there must exist a mechanism for the activation of only one particular V gene. One of the most consistent observations made during the series of experiments summarized in Table I is that there is a strict correlation between expression of a particular V gene and its joining with the corresponding C gene. In the light of the correlation, one intriguing possibility is that activation of a V gene is directly coupled to its becoming joined to a C gene. For instance, in the 'inversion' model (see Tonegawa et al. 1976), a promoter site might be created by the insertion of the V gene fragment. This would activate that particular V gene for transcription.

Alternatively, all V genes and even the C genes might be activated for transcription early in lymphocyte differentiation. The RNAs synthesized would be rapidly degraded in the nucleus by a 5′-specific exonuclease before reaching the ribosomes. The specific nucleotide sequence created by the joining event at the insertion site could serve, directly or indirectly, as a signal for preventing the exonuclease from proceeding further down toward the 3′ end. This model looks rather wasteful at first sight, but it should not be discarded solely for this reason. It is, in fact, in agreement with the fact that in eucaryotic nuclei a large proportion of DNA is constitutively transcribed into large RNA molecules (HnRNA), most of which turn over rapidly before reaching the ribosomes. Only a small proportion (a few per cent) of these large RNA molecules are 'processed' to become mRNA.

In any case, in view of the above considerations, the current terminology of 'V genes' and 'C genes' is somewhat inappropriate. Rather, there are two types of segments of DNA, one specifying the V region and the other specifying the C region. The 'gene' is *created* by joining.

V-C Joining and Allelic Exclusion

For immunoglobulin loci, only one allele is expressed in any given lymphocyte (Pernis et al. 1965). This is not the case for any other autosomal genes studied until now. Our results obtained with MOPC 321 DNA (Figure 6) suggest an interesting explanation for this phenomenon of allelic exclusion – that the two homologous chromosofes are, in any given plasma cell, homozygous. Homozygosis could result from the loss of one homolog followed by reduplication of the other, or it could result from somatic recombination (presumably mediated by a specialized recombinase) between the centromere and the immunoglobulin locus (Ohno 1974). This explanation of allelic exclusion is independent of whether the joining takes place only in one of the two homologous chromosomes or in both, in a single lymphocyte. On the other hand, our results obtained with TEPC 124 DNA suggest that the joining takes place only in one of the two homologous chromosomes and no homozygosis is to take place. While this, too, can conveniently explain allelic exclusion, there is an apparent discrepancy between the two cases. The discrepancy could arise from the known abnormality in caryotypes of myeloma cells (Yoshida et al. 1968). Both MOPC 321 and TEPC 124 are subtetraploids (Tonegawa unpublished observation). At this time, we have no way of knowing how many of the particular chromosomes carrying V and C genes are present per tumor cell. MOPC 321 myeloma might have lost the homologous chromosome(s) on which V and C genes lie separate. Conversely, TEPC 124 myeloma might have acquired an additional chromosome of non lymphatic sources during generation or propagation of the tumor (Warner & Krueger 1975).

Resolution of the discrepancy and elucidation of the mechanism of allelic exclusion would require detailed analysis of DNA from Ig synthesizing cells, in particular DNA from natural lymphocyte clones.

Isolation of Ig Genes

The great difficulty often encountered during research on unique eucaryotic genes stems from the enormous complexity of the genomes. For instance, a mouse light chain V gene comprises only one in five million parts of the total haploid DNA. This requires one to use milligram amounts of total cellular DNA in each hybridization mixture, and limits the kinds of analysis described in this article to the case where large amounts of homogenous cellular materials are available. In addition, if one is to obtain a unique gene in a pure form for further studies, the problem one faces is two fold. It is virtually impossible to purify specific DNA sequences five million times

by any usual biochemical method. Even were this possible, one would require impractically large amounts of material. (To get 1 mg of pure V gene DNA, one must start with 1 kg of wet cells!).

Recently, a breakthrough was made in this field. A technique was developed with which virtually any eucaryotic DNA can be inserted into self-replicating bacterial plasmid or phage DNA by *in vitro* recombination (Morrow et al. 1974, Thomas et al. 1974, Wensink et al. 1974). The resulting hybrid plasmid or phage DNA is used to directly infect host bacteria for unlimited propagation. Since each bacterium receives only a single DNA molecule, the method provides us with powerful purification procedures. With the development of suitable selection or screening procedures for a particular gene sequence, one can isolate any eucaryotic gene in pure form and in large quantity.

We have recently applied this technique to Ig genes. Two kinds of experiments have been carried out. First, a microgram amount of double stranded DNA was synthesized *in vitro* using purified Ig mRNA as the initial template for a reverse transcriptase and a subsequent conversion of the single stranded DNA into a double stranded DNA with a DNA polymerase (Maniatis et al. 1976). This DNA was inserted into the *Eco* RI cleavage site of a plasmid, PCRI. Resulting hybrid DNA was propagated in *E coli* χ7776, a strain deliberately attenuated in order to prevent growth under out-of-laboratory conditions. Several clones obtained by these procedures are being characterized. We hope that these DNA clones will help in increasing sensitivity of detection in the analysis of restriction enzyme-digested DNA and allow us to extend studies to natural lymphocyte clones.

Second, we have also applied the technique to isolate Ig genes directly from chromosomal DNA. We enriched the DNA fragment carrying V_λ genes approximately 350 fold by preparative agarose gel electrophoresis and by hybridization of partially denatured DNA fragments with purified λ chain mRNA (R-loop formation (Thomas et al. 1976)). This DNA was inserted in the middle of DNA from phage λ WES, a derivative of phage λ attenuated for lower virulency in the out-of-laboratory environment (Enquist et al. 1976). After screening several thousand plaques which arose from such hybrid DNA, we could identify one phage which seems to carry a V_λ gene. Hybridization and electronmicroscopic studies indicate that the inserted mouse DNA piece lacks the C_λ gene, and that the putative V_λ gene is located internally. If these observations are confirmed by further characterization of the insert, it will constitute the most direct evidence for the concept of separate V and C gene sequences. We are currently determining the nucleotide sequence of the inserted DNA in order to see what a germ line V_λ gene looks like.

by any usual biochemical method. Even were this possible, one would require impractically large amounts of material. (To get 1 mg of pure V gene DNA, one must start with 1 kg of wet cells!).

Recently, a breakthrough was made in this field. A technique was developed with which virtually any eucaryotic DNA can be inserted into self-replicating bacterial plasmid or phage DNA by *in vitro* recombination (Morrow et al. 1974, Thomas et al. 1974, Wensink et al. 1974). The resulting hybrid plasmid or phage DNA is used to directly infect host bacteria for unlimited propagation. Since each bacterium receives only a single DNA molecule, the method provides us with powerful purification procedures. With the development of suitable selection or screening procedures for a particular gene sequence, one can isolate any eucaryotic gene in pure form and in large quantity.

We have recently applied this technique to Ig genes. Two kinds of experiments have been carried out. First, a microgram amount of double stranded DNA was synthesized *in vitro* using purified Ig mRNA as the initial template for a reverse transcriptase and a subsequent conversion of the single stranded DNA into a double stranded DNA with a DNA polymerase (Maniatis et al. 1976). This DNA was inserted into the *Eco* RI cleavage site of a plasmid, PCRI. Resulting hybrid DNA was propagated in *E coli* χ7776, a strain deliberately attenuated in order to prevent growth under out-of-laboratory conditions. Several clones obtained by these procedures are being characterized. We hope that these DNA clones will help in increasing sensitivity of detection in the analysis of restriction enzyme-digested DNA and allow us to extend studies to natural lymphocyte clones.

Second, we have also applied the technique to isolate Ig genes directly from chromosomal DNA. We enriched the DNA fragment carrying V_λ genes approximately 350 fold by preparative agarose gel electrophoresis and by hybridization of partially denatured DNA fragments with purified λ chain mRNA (R-loop formation (Thomas et al. 1976)). This DNA was inserted in the middle of DNA from phage λ WES, a derivative of phage λ attenuated for lower virulency in the out-of-laboratory environment (Enquist et al. 1976). After screening several thousand plaques which arose from such hybrid DNA, we could identify one phage which seems to carry a V_λ gene. Hybridization and electronmicroscopic studies indicate that the inserted mouse DNA piece lacks the C_λ gene, and that the putative V_λ gene is located internally. If these observations are confirmed by further characterization of the insert, it will constitute the most direct evidence for the concept of separate V and C gene sequences. We are currently determining the nucleotide sequence of the inserted DNA in order to see what a germ line V_λ gene looks like.

Conclusions and Prospects

Genes can mutate and recombine. These dynamic characteristics of genetic material are essential elements of evolution. Do they also play an important role during the development of a single multicellular organism? Our results strongly suggest that this is the case for the immune system. Here, presumably, mutation alters the coding sequence of the V DNA segment. This process is essential in amplifying the repertoire of genetic information in the germ line V DNA segments. Recombination brings two DNA segments, V and C, together to create a complete Ig gene. It is highly probable that this latter process is directly linked to activation of a specific V DNA segment. In Figure 7 we summarize these genetic events.

Many details in Figure 7 are yet to be studied. Research on Ig genes and eucaryotic genes in general has reached a new era with the development of *in vitro* recombinant DNA techniques. A variety of plasmids or phages carrying various parts of Ig gene loci will be isolated in the near future. These clones will directly or indirectly contribute greatly to elucidation of many of the unsolved problems, some of which were discussed

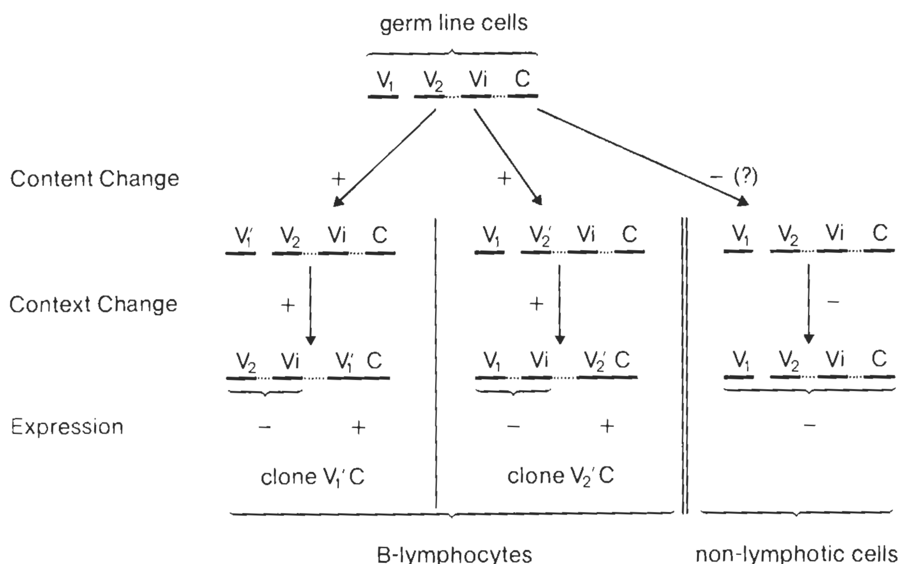

Figure 7. Differentiation of immunoglobulin genes. Content change refers to change in coding information. Context change refers to change in the relative location of a DNA sequence. We do not imply that the former takes place earlier than the latter during development.

in this article. For instance, characterization of these DNA clones by hybridization and electronmicroscopy will allow one to conclude which model is correct for V-C joining. In some cases, comparative nucleotide sequence analysis of such cloned DNA will give us a clue about how V gene sequences alter during development. The same sequence analysis might lead to discovery of eucaryotic promoters and provide us with evidence for the concept that joining and activation of specific DNA sequences are closely coupled events. Cloned Ig genes are also useful as pure and easily available hybridization probes. In combination with the cell sorting technique recently developed (Herzenberg et al. 1976) it may be possible to extend the restriction enzyme-analysis to various natural lymphocyte clones. Work along this line will be essential in uncovering the mechanism of allelic exclusion and the monoclonal switch.

Yet another exciting, but admittedly more difficult problem remains. What are the enzymatic mechanisms underlying the somatic changes in the content (coding information) and context (chromosomal location) of Ig genes? Here, too, cloned Ig genes will be very useful. We now have in our hand specific substrates with which these enzymes (if they exist) interact. Affinity columns prepared with the cloned DNA will be of great use in the search for such enzymes. Some of these enzymes, when identified and purified, may serve as markers for those cells which are at particular stages of differentiation.

ACKNOWLEDGEMENT

We thank Dr. M. O. Potter and Dr. M. Cohn for myeloma cells, Dr. R. Curtis for *E coli* strain χ1776, Dr. J. W. Beard for reverse transcriptase and Dr. C. Kung for terminal transferase. Cloning of Ig genes from enzymatically synthesized DNA was carried out in collaboration with S. Artavanis and P. Shedle at the Biocentrum, Basel, and with S. Drost and D. Baltimore at M.I.T., Cambridge, U.S.A.

We thank Mr. G. R. Dastoornikoo, Ms. S. Stutz, Ms. B. Durrer and Mr. A. Traunecker for their expert technical assistance. We also thank Dr. Richard Pink for his critical reading of the manuscript.

REFERENCES

Bernardini, A. & Tonegawa, S. (1974) Hybridization studies with an antibody heavy chain mRNA. *FEBS Lett.* **41**, 73.

Bonner, T., Brenner, D. J., Neufeld, B. R. & Britten, R. J. (1973) Reduction in the rate of DNA reassociation by sequence divergence. *J. Mol. Biol.* **81**, 123.

Campbell, A. (1962) Episomes. *Advan. Genet.* **11**, 101.

reiteration frequency of light-chain genes by hybridization procedures. *Eur. J. Biochem.* **52**, 125.

Raff, M. C., Feldmann, M. & De Petris, S. (1973) Monospecificity of bone marrow derived lymphocytes. *J. Exp. Med.* **137**, 1024.

Schreier, M. H. & Staehlin, T. (1973) Initiation of mammalian protein synthesis: the importance of ribosome and initiation factor quality for the efficiency of *in vitro* systems. *J. Mol. Biol.* **73**, 329.

Schuller, R. (1976) Purification of immunoglobulin heavy chain mRNA's and their use in hybridization studies. *Master of Science Thesis,* University of Alberta, Edmonton, Alberta.

Thomas, M., Cameron, J. R. & Davis, R. W. (1974) Viable molecular hybrids of bacteriophage lambda in eucaryotic DNA. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4579.

Thomas, M., White, R. L. & Davis, R. W. (1976) Hybridization of RNA to double-standard DNA: Formation of R-loops. *Proc. Nat. Acad. Sci., U.S.A.* **73**, 2294.

Tonegawa, S. (1976) Reiteration frequency of immunoglobulin light chain genes: Further evidence of somatic generation of antibody diversity. *Proc. Nat. Acad. Sci., U.S.A.* **73**, 203.

Tonegawa, S. & Baldi, I. (1973) Electrophoretically homogeneous myeloma light chain mRNA and its translation *in vitro. Biochem. Biophys. Res. Commun.* **51**, 81.

Tonegawa, S., Hozumi, N., Brack, C. & Schuller, R. (1977) Arrangement and rearrangement of immunoglobulin genes. *ICN-UCLA Symposium. Regulation of the immune system: genes and the cells in which they function* (in press).

Tonegawa, S., Hozumi, N., Matthyssens, G. & Schuller, R. (1976) Somatic changes in the content and context of immunoglobulin genes. *Cold Spring Harbor Symp. Quant. Biol.* **41** (in press).

Tonegawa, S., Steinberg, C., Dube, S. & Bernardini, A. (1974) Evidence for somatic generation of antibody diversity. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4027.

Warner, T. F. C. S. & Krueger, R. G. (1975) The multicellular (as opposed to the monoclonal) origin of the murine myeloma cell. *J. Theor. Biol.* **54**, 175.

Weigert, M. G., Cesali, I. M., Yonkovich, S. J. & Cohn, M. (1970) Variability in the lambda light chain sequences of mouse antibody. *Nature* **228**, 1045.

Weigert, M., Potter, M. & Sachs, D. (1975) Genetics of the immunoglobulin variable region. *Immunogenetics* **1**, 511.

Wensink, P. C., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) A system for mapping DNA sequences in the chromosomes of drosophila melanogaster. *Cell* **3**, 315.

Yoshida, T. H., Imai, H. T. & Potter, M. (1968) Chromosomal alteration and development of tumors. XIX. Chromosome constitution of tumor cells in 16 plasma cell neoplasms of Balb/c mice. *J. Nat. Cancer Inst.* **41**, 1083.