Journal of Supramolecular Structure 7:531–559 (1977)
Cell Surface Carbohydrates and Biological Recognition 407–435

The Area-Code Hypothesis: The Immune System Provides Clues to Understanding the Genetic and Molecular Basis of Cell Recognition During Development

L. Hood, H. V. Huang, and W. J. Dreyer
Division of Biology, California Institute of Technology, Pasadena, California 91125

Numerous studies of embryogenesis have provided evidence for highly specific cell-surface recognition phenomena. These include both the interactions of neighboring cells and the specific cellular migrations which occur as the developmental program of the embryo progresses. The area-code hypothesis elaborated here is an attempt to provide a framework for understanding cell-recognition phenomena in development.

This hypothesis is based on extensive genetic, molecular, and cellular studies of the immune system. These studies suggest that the following events occur during the differentiation of antibody-producing cells. 1) Somatic cell lines of antibody-producing cells undergo a modification of their DNA as they become committed to synthesize a particular type of antibody molecule. This chromosomal modification event is probably a DNA translocation which leads to a somatic rearrangement of certain antibody genes. 2) In each of the specific cell lineages the new arrangement of DNA is inherited by all subsequent generations of cells. 3) The developmental programs which control these genetic alterations may be employed in a programmed and reproducible fashion. This programming of antibody development is suggested because different embryos appear to become committed to the production of identical antibody molecules in the same developmental sequence. 4) Antibody molecules are initially displayed on the cell surface where they serve as highly specific receptors to trigger the cell to proliferate and differentiate upon interacting with appropriate external molecular signals. 5) Antibody-producing cells display combinations of different molecules on their surfaces which cause each of a very large number of different cells to interact differently with their environment. 6) The genes which code for many of these cell-surface molecules are organized into multigene families.

These observations as well as information from other developmental systems have led us to propose the area-code hypothesis. This hypothesis is concerned with the structure, function, and regulation of cell-surface molecules that mediate recognition phenomena during embryogenesis. Area-code molecules are cell-surface molecules which are involved in the specific recognition phenomena during growth and development. These molecules provide cells with distinct cell-surface addresses or phenotypes, and provide the basis for the specificity in cell-cell recognition during cell migrations and cell-cell interactions, as well as serving as receptors for diffusible

Received for publication June 22, 1977; accepted August 15, 1977

differentiation signals. The area-code hypothesis has 3 main postulates. i) There is a progressive display of specific combinations of area-code molecules on the surfaces of cells during development. ii) The genetic programs which determine the specific expression of area-code molecules are in part controlled by DNA modifications. These chromosomal modifications are believed to channel cells into specific lineages with progressively restricted developmental options. iii) Many of the area-code systems are organized into multigene families. Rapid evolutionary increases in complexity may proceed by the duplication and subsequent independent evolution of multigene families. In short, many of the remarkable events which occur during the development of the immune system may form a basis for understanding other developmental systems. Some experimental approaches toward testing this hypothesis are discussed.

Key words: **area-code hypothesis, combinations of cell-surface recognition molecules, chromosomal modifications, DNA translocation, multigene families, immune system as developmental model**

Development is the orderly process whereby a single cell, the zygote, generates a large diversity of cell types (Fig. 1). These cells migrate to appropriate locations and interact with one another to give rise to the supracellular organization of the adult organism. The adult human has about 10^{14} cells. If all cells were to divide at an equal rate, the average adult cell would be separated from the zygote by at least 48 cell divisions ($2^{47} \simeq 10^{14}$). All adult cells have a cell lineage whose origin can be traced back to the zygote. As embryogenesis proceeds, cell lineages develop which become increasingly limited in their future development options (Fig. 1). Individual cells acquire specific developmental programs which limit the developmental fate of their progeny cells. Individual cells may become committed to a particular developmental program long before they differentiate to acquire the phenotypic characteristics of that cell lineage. These committed cells may later be induced to differentiate by hormones or other external signals. Little is known about the nature of the developmental programs or the mechanisms of cellular commitment. However, information has accumulated on the nature of cell-surface changes which occur as a cell lineage differentiates.

The unfolding of the developmental programs of individual cells leads to the expression of new gene products including molecules on the cell surface. Some of these cell-surface molecules are involved in cell recognition processes and may encompass a variety of functions in the developing embryo. Combinations of these molecules displayed on the surface may play a vital role in providing an address system for the massive cellular migrations that are characteristic of the developing embryo (Fig. 2). They are also vital to the myriad of specific cell-cell interactions occurring during growth and development. In the next section we propose a hypothesis that describes the general features of area-code molecules and the genes which encode them.

THE AREA-CODE HYPOTHESIS

The area-code hypothesis was formulated from an analysis of the vertebrate immune system. This hypothesis deals both with the role played by cell-surface recognition or area-code molecules in development and with the genetic events which place this address system on cell surfaces. The area-code hypothesis has several interrelated postulates: i) During development combinations of area-code molecules are displayed on the surface of cells of specific lineages. These cell-surface displays provide the specificity for cell-cell interactions. They also provide cell-surface receptors essential for sensing diffusible differentiation

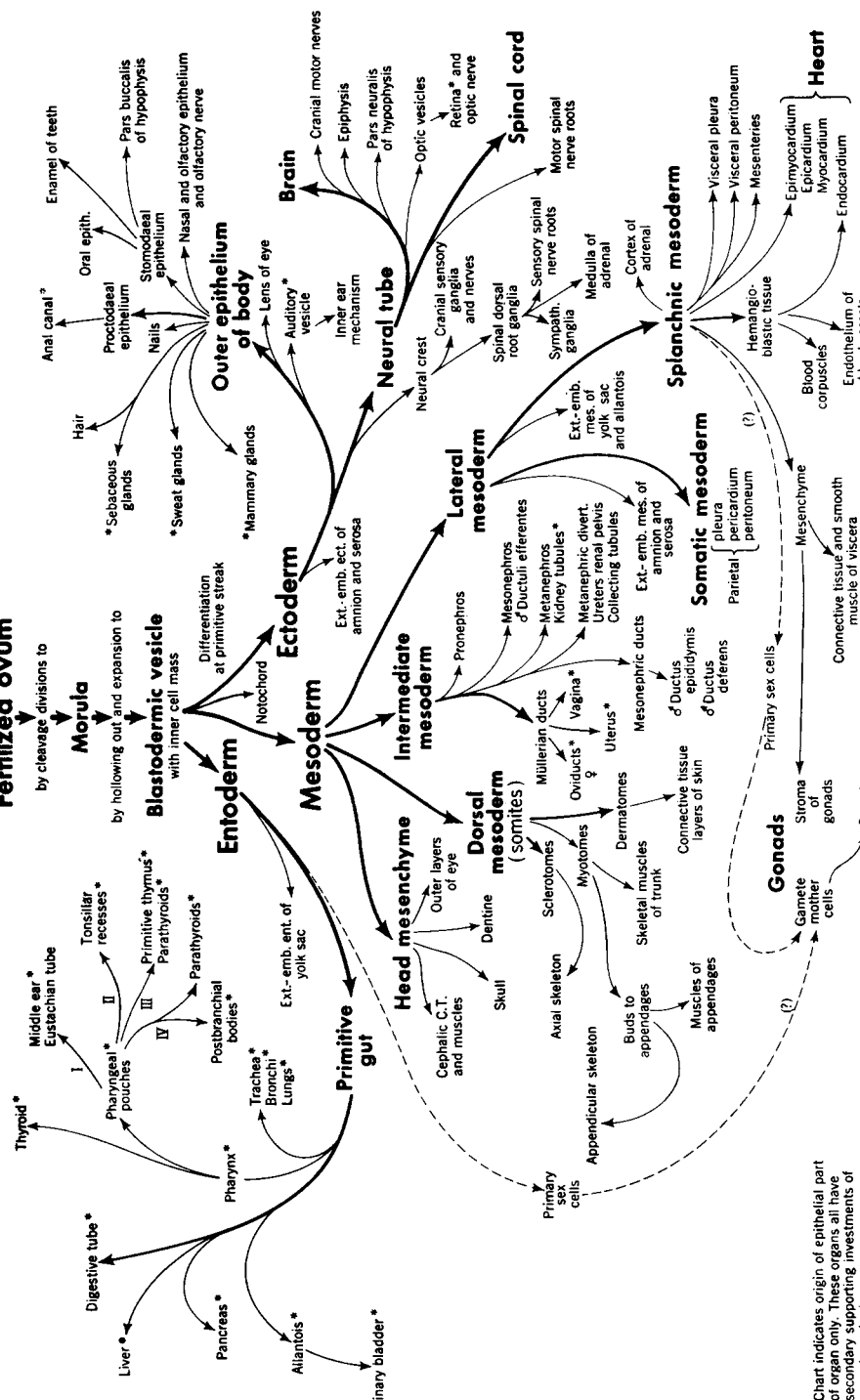

Fig. 1. Cell lineages of a developing organism. [Reprinted from Patten BM, Carlson BM: "Foundations of Embryology." 3rd Ed. San Francisco: McGraw-Hill, 1974, p 141, with permission.]

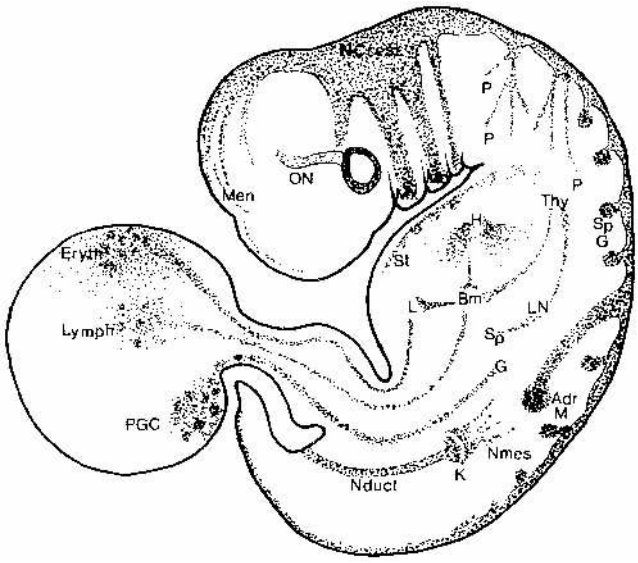

Fig. 2. Cellular migrations in the embryo. This diagram displays events taking place at different stages of development.

	Origin	Route	Target(s)
Erythrocyte precursors (Eryth)	Yolk sac	Blood stream	Liver (L), Bone marrow (Bm)
Lymphoid cell precursors (Lymph)	Yolk sac	Blood stream	Bone marrow, Thymus (Thy), Lymph nodes (LN), Spleen (Spl)
Primordial germ cells (PGC)	Yolk sac		Gonadal ridges, forming gonads (G)
Mesenchyme cells	Dorsal embryo	Migrate ventrally	Form sternum (St)
Retinal ganglion cell axons	Eye	Optic nerve (ON)	Brain visual centers
Neural crest cells (NCrest)	Neural crest	Migrate ventrally	Form meninges (Men), embryonic skull cartilages (Mx,Md), Pigment cells (P), Spinal ganglia (SpG), Adrenal medulla (AdrM)
Heart cells	"Heart-forming territory"	Aggregate to	Form heart (H)
Nephric ducts	Elongate and meet aggregations of nephric mesenchyme cells (Nmes) to form kidney rudiments (K)		
Not included in figure:	The intricate cell translocations in histogenesis of the nervous system, and other examples of cell translocations and aggregations (e.g., those resulting in the formation of hair follicles, teeth, limb rudiments).		

[Reprinted from Ref. 99, with permission.]

signals such as those provided by hormones. Thus differentiating cells acquire distinct combinations of area-code molecules which serve as specific cellular addresses, not unlike those of the telephone or postal systems. ii) The genetic programs which govern the expression of area-code molecules are controlled in part by alterations in somatic cell chromosomes such as the translocation of DNA sequences. These chromosomal modifications occur at branch points of differentiation which define cell lineages and limit the

TABLE I. Features of the Immune System Facilitating Its Study

1. Lymphocytes are freely wandering cells easily separated from other cell types and fractionated into functional subclasses.
2. Developmental decisions made in the immune system generate specific cell lineages in which the future options are genetically programmed.
3. Lymphocyte tumors represent clones derived from single progenitor cells. Many different stages of lymphocyte differentiation are represented in the available tumor lines.
4. Serologic reagents are available to identify different lymphocyte lineages. Thus specific tags are available for studying the differentiation, migration, interactions, and triggering of lymphocytes.
5. The cells of the immune system undergo programmed migrations during development and interact in highly specific ways within various tissues such as the liver, spleen, bone marrow, thymus, and lymph nodes.
6. Lymphocytes form a regulatory net work with one another and with other cells. This network can be studied in vitro as well as in vivo. Thus it is perhaps easier to study cell cooperation in this system than in any other.
7. Triggering of lymphocytes to differentiate and divide by mitogens, antigens, and hormones can readily be studied in vitro and in vivo.
8. Antibody molecules and mRNAs have been isolated from myeloma tumors and studied by chemical serological, genetic, and functional techniques.

future options open to particular cell lineages. iii) In many cases the genes encoding area-code systems are organized into closely linked multigene families. During evolution, multigene families can be duplicated to produce new multigene families. These new families can assume new functions in programming development. Presumably these reprogrammed multigene families can produce major evolutionary changes in complex systems such as the brain. Thus the area-code gene families from diverse developmental systems may share common evolutionary originals.

It is our belief that area-code molecules play an important role in the recognition phenomena of embryogenesis for a variety of simple and complex systems including the neuroectoderm (1), the nervous system (2), the optic nerve (3), and the immune system.

THE IMMUNE SYSTEM IS A MODEL FOR STUDYING DEVELOPMENT IN COMPLEX EUKARYOTIC SYSTEMS

A. General Comments

The immune system is the first complex eukaryotic system that has been investigated in depth at the genetic, molecular, and cellular levels. Its study has revealed some remarkable mechanisms and strategies for development that are embodied in the area-code hypothesis. Although some of these mechanisms appear unique to the immune system, we feel that this is so only because no other complex eukaryotic system has been studied in similar detail. We suggest that many of the developmental strategies of the immune system will be employed by other area-code systems.

The immune system has several features which facilitate its experimental study (4). These are summarized in Table I. The major cells of the immune system, lymphocytes, are freely wandering cells that can be readily separated from other cell types and even fractionated into functional subclasses. Migration pathways of embryonic lymphocytes can be traced as development proceeds in vivo. The differentiation of lymphocytes also can be followed in vitro. The availability of large quantities of antibody molecules and antibody mRNA from

myeloma tumors has permitted a detailed analysis of the antibody molecules and their genes. The antibody molecule appears to play an important role in mediating cellular interactions among lymphocytes and thus serves as a model for the prototype area-code molecule. Cell-cell interactions and receptor-mediated triggering of lymphocytes to differentiate can be studied in vivo or in vitro. We will discuss in some detail the developmental, cellular, molecular, genetic, and evolutionary strategies of the immune system as they have formed the basis for our thinking about the area-code hypothesis.

B. The Immune System Employs Antibody Molecules to Recognize Foreign Molecular Patterns

The vertebrate immune system shows 2 cardinal features of area-code systems, namely it utilizes recognition molecules exhibiting specificity and diversity.

Specificity. The immune system recognizes and destroys foreign molecules or antigens (5). The fundamental unit of recognition in this process is the antibody molecule. This molecule can be affixed to a lymphocyte as a specific cell-surface receptor or it can be secreted into the blood or lymphatic circulations. The antibody binds antigen through a molecular complementarity similar to that which an enzyme exhibits for its specific substrate. This interaction leads by a variety of mechanisms to the specific destruction or elimination of the antigen.

Diversity. Lymphocytes and their antibody molecules recognize a virtually limitless number of different antigenic determinants because almost any macromolecule that is foreign to a particular vertebrate organism can evoke an immune response. The average man has approximately 10^{12} lymphocytes circulating throughout his body and 10^{20} antibody molecules in his circulation. Estimates as to the number of different molecular species of antibody molecules a vertebrate can synthesize range between 10^5 and 10^8. Hence the immune system is capable of generating an enormous array of different types of specific cell-surface recognition molecules. Let us consider how distinct cell lineages develop in the immune system.

C. Lymphocytes Differentiate Along One of Two Discrete Developmental Pathways to Produce B Cells and T Cells

Differentiation in lymphocytes requires specific cellular migration and hormonal induction, two features shared by other developmental systems.

B- and T-cell lineages. The development of the immune system begins with stem cells arising in the yolk sac and later migrating to the fetal liver and finally to the bone marrow (Fig. 2) (6). In the adult, stem cells for lymphocytes divide in the bone marrow and there become committed to the B- or T-cell pathway. These are termed pre-B or pre-T cells. Later these committed lymphocytes migrate to an appropriate microenvironment where hormonal inducers trigger the subsequent expression of the precommitted B- or T-cell developmental programs.

The pre-T cells migrate to the thymus. Under the influence of thymic hormones, they differentiate and later enter the circulation as mature T cells (7). T cells undergo additional differentiation steps on interaction of antigen with the antibody-like receptors on their cell surface.

In birds, the pre-B cell migrates to the bursa of Fabricius and presumably under hormonal induction differentiates to a mature B cell (8). In mammals, the pre-B cell probably differentiates in the bone marrow (9). B cells migrate to the blood and lymphatic circulation. There the interaction of antigen with antibody receptors induces terminal differentiation to the plasma cell, a highly efficient factory for the synthesis of antibody molecules.

B and T cells both employ antibody or antibody-like molecules as specific cell-surface receptors (10). The library of T-cell antibody-like receptors is believed to be comparable in diversity to those of its B-cell counterpart.

Functions of B and T cells. B cells synthesize antibody molecules. These molecules are employed as cell-surface receptors and they are also secreted into the serum. B cells constitute the basis of the humoral immune response which depends on the secreted antibody molecules to fight acute viral and bacterial infections. T cells synthesize antibody-like molecules which are employed as cell-surface receptors for the diverse reactions of the cellular immune response. These include the surveillance for and destruction of cells altered by neoplastic transformation or viral infection. T and B cells display characteristic cell-surface molecules, some of which mediate cell-cell interactions.

D. As Lymphocytes Differentiate, Distinct Combinations of Cell-Surface Molecules Are Displayed

The area-code hypothesis suggests that cell-surface recognition molecules play a fundamental role in growth and development. The successive acquisition of cell-surface molecules during differentiation has been clearly demonstrated in lymphocytes.

Cell-surface molecules on T and B cells. The display of a variety of cell-surface molecules on mouse lymphocytes at various stages of development has been studied by detailed genetic and serological analyses (11). In the bone marrow all precursor cells of lymphocytes express the transplantation (H-2) antigens (Fig. 3). The pre-T cell migrates to the thymus and there is induced by thymic hormones to express at least 4 cell-surface nolecules: TL, Ly 1, Ly 2, and Thy 1 (12). The T-cell receptors, designated IgT, probably appear at this stage. As the T cell migrates to the periphery, TL disappears, and the cell-surface concentration of Thy 1 decreases while that of H-2 increases (13). There are 3 Ly phenotypes of T cells in the periphery circulation, Ly 1, Ly 2, and Ly 1, 2, and it appears likely that the Ly 1 and Ly 2 cells are derived from the Ly 1,2 cells (4). The pre-B cell migrates to the bursa or its equivalent and there presumably acquires the antibody receptor, designated IgB (Fig. 3). Antigen induces the expression of the PC-1 antigen in plasma cells (14).

These cell-surface molecules are designated differentiation antigens (11) because they are molecules that have been expressed during successive developmental stages and have been studied by serological techniques.

Expression of differentiation antigens. Differentiation antigens show various modes of expression. i) The differentiation antigens induced in pre-T cells by thymic hormones can be expressed within 5 h of induction. Moreover, this expression can occur without cell division (12). Accordingly, inducers trigger the expression of previously committed developmental programs. ii) Certain differentiation antigens are lost during the course of subsequent differentiation (e.g., TL), and others change markedly in their cell-surface concentrations (e.g., Thy 1 and H-2). Hence genes coding for differentiation antigens can be turned off or altered in their rate of expression.

Sets of genes determine phenotype. The genes encoding the differentiation antigens of lymphocytes are present on at least 8 different chromosomes (Fig. 3b). Thus, an array of cell-surface molecules encoded by unlinked genes determines the cell-surface state of differentiation in B and T cells. Different subsets of these genes must be expressed in a coordinated fashion by the developmental programs for each distinct cellular phenotype.

Differentiation antigens and lymphocyte-lineage relationships. Differentiation antigens provide clues as to the lineage relationships of various lymphocyte clones. T cells can be readily distinguished from B cells based on their cell surface molecules (Fig. 3a). Indeed, 3 distinct subclasses of T cells can be distinguished in the peripheral circulation

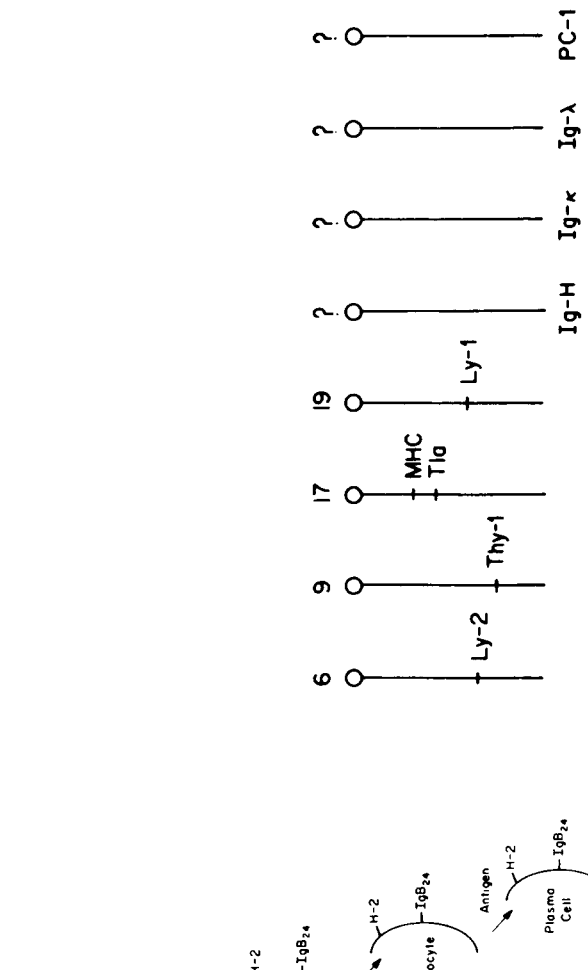

Fig. 3. a) Differentiation antigens on mouse lymphocytes. The bone marrow lymphocyte precursor, which gives rise to the T- and B-lymphocyte lineages, expresses H-2 antigens on its surface. In the respective inductive microenvironments, lymphocytes display new sets of differentiation antigens. On seeding to the periphery some differentiation antigens may be selectively lost (e.g., TL, Ly1, or Ly2). Terminal differentiation may be correlated with new differentiation antigens (e.g., PC-1). H-2 designates the H-2 K and D transplantation antigens. Thy-1 denotes thymus-derived lymphocyte antigen. IgT and IgB designate the antigen receptors of T and B cells, respectively, with subscripts of arbitrary numbers used to emphasize the great diversity of these molecules, and to stress the one cell-one antigen-binding site feature of these cells. Ly 1 and Ly2 represent lymphocyte antigens 1 and 2 which are present exclusively on T cells. TL denotes thymus leukemia antigen. PC-1 represents plasma cell antigen. Other known lymphocyte differentiation antigens (e.g., Ia, Qa-2, Ala-1) are not shown. b) Chromosomal map of genes coding for differentiation antigens on mouse lymphocytes. The positions on the chromosomes are approximate for loci whose position is known. Chromosomal assignment for the immunoglobulin (Ig) families (H, λ, κ) and PC-1 are not known.

based on the Ly phenotypes (Fig. 3a). Finally, clones of lymphocytes acquire their unique functional specificities based on their expression of individual antibody or antibody-like cell-surface receptors. Thus, each individual lymphocyte clone expresses a unique set of cell-surface molecules beginning, for example, with those shared by all lymphocytes (e.g., H-2), to those shared by T cells (e.g., Thy 1), to those shared by T-cell subclasses (e.g., Ly1, Ly 2, or Ly 1,2), and finally to those conferring clonal individuality (e.g., IgT_{117}) (Fig. 3). These combinations of cell-surface molecules serve as molecular addresses to distinguish individual clones of lymphocytes.

Area-code molecules and differentiation antigens. Area-code molecules are defined as those involved in cell-surface recognition processes. On the other hand, the differentiation antigens are any cell-surface molecule that appears during the course of differentiation in a particular cell lineage. Some differentiation antigens may be area-code molecules, though presumably not all. Many differentiation antigens may carry out cell-surface roles unrelated to cell recognition such as enzymatic reactions, structural support, and transport functions. The functions of the differentiation antigens of lymphocytes, apart from those of the IgT or IgB receptors, are unknown. Since the B-cell system serves as a model for an area-code system, let us consider how it is stimulated to undergo the final stages of differentiation by antigen.

E. Antigen Triggers Clones of Lymphocytes With Complementary Antibody Receptors

One of the unresolved questions about complex eukaryotic systems employing cell-surface recognition molecules is how are they triggered to differentiate. While the molecular details of this triggering process are not understood for lymphocytes, a reasonable phenomenological description of this process is available.

In a system that encompasses 10^{12} lymphocytes expressing 10^5 to 10^8 different antibody molecules, how are appropriate antibody molecules expressed in response to individual antigens? There are several aspects to the triggering of a specific immune response to antigen (15) (Fig. 4). i) Individual lymphocytes can synthesize only one molecular species of antibody molecule. This commitment of each lymphocyte to the synthesis of one type of antibody molecule is an antigen-independent process. ii) Antigen triggers the clonal expansion of individual lymphocytes through interaction with complementary antibody receptors at the cell surface. iii) The clonal descendants of a particular lymphocyte are all committed to the expression of antibody molecules of precisely the same specificity as those of the parent lymphocyte. iv) Antigenic triggering of a lymphocyte results in 2 general classes of clonal descendants. Effector cells are terminally differentiated and mediate the immediate response to antigen. Memory cells constitute a greatly expanded specific lymphocyte compartment for enhanced secondary immune responses on reencounter with antigen. Thus antigen is one of the final inductive triggers in lymphocyte differentiation. Moreover, antibody molecules, the prototype area-code molecules, play a fundamental role in this differentiation process.

Once antigen triggers clones of specific lymphocytes, antibody production from the individual lymphocyte clones must be regulated. Specific antibody molecules and antigen play an important role in this process. The effects of lymphocytes interacting with one another in a regulatory network are also very important in regulating the immune response.

F. Lymphocyte Interactions Regulate the Immune Response

One might argue that the immune system fails as a model for many potential area-code systems because lymphocytes are mobile and do not exhibit the fixed cellular

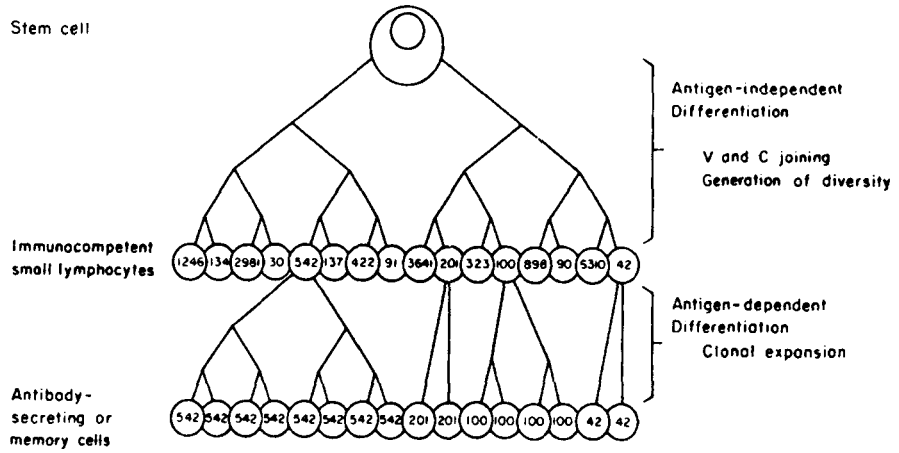

Fig. 4. A model for clonal selection in the immune system. Adapted from Ref. 34.

TABLE II. Cellular Interactions Among Sets of Lymphocytes*

Cooperating sets		
Effector	Inducer	Function
B cell	Ly 1:H[a]	↑ antibody
Ly 2:KS[a]	Ly 1:H	↑ cytotoxicity
Ly 1,2:ARC[a]	Ly 2:KS	↑ suppressing
Macrophage	Ly 1:H	Delayed hypersensitivity
Macrophage	Ly 2:KS	Macrophage killers

*Adapted from Ref. 31.
[a]H indicates helper; KS denotes killer-suppressor; ARC designates antigen receptor cell.

interactions characteristic of most differentiated tissues. However, lymphocyte interactions with one another and with other tissues play a fundamental role in lymphocyte differentiation.

Lymphocyte interactions in the immune response. Clones of lymphocytes interact with one another to regulate the immune response and produce a finely balanced lymphocyte network (16–18). The cellular basis for these interactions rests in part on the presence of 3 functionally distinct subclasses of T cells with differing Ly phenotypes (19). The Ly 1 helper-T cells cooperate with B cells or other T cells to produce an immune response. The Ly 2 cells fall into 2 distinct categories. The Ly 2 suppressor-T cells inhibit immune responses of B cells or other T cells. The Ly 2 killer-T cells destroy foreign cells by lysing them through unknown mechanisms. The Ly 1,2 cell also plays a role in the suppression process. These cellular interactions are summarized in Table II. Indeed, a third type of cell, the macrophage, also interacts with T cells to mediate certain aspects of the immune response (20). Macrophages share a common progenitor lineage with lymphocytes and, accordingly, are closely related in an embryological sense (21). Thus lymphocytes interact with one another and with macrophages to facilitate or suppress immune responses in a delicately balanced network of cellular regulation. Accordingly, lymphocyte networks may be an ideal model system for studying various aspects of cell-cell interaction – a cardinal feature of development during embryogenesis and of the area-code hypothesis.

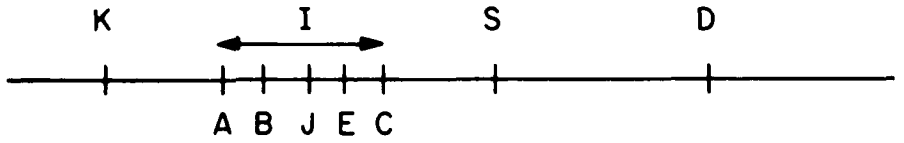

Fig. 5. A genetic map of the major histocompatibility complex or H-2 complex of the mouse.

The lymphocyte network and cell-surface molecules. Lymphocyte interactions are mediated by several classes of cell-surface molecules including those coded by the major histocompatibility complex (22) and the antibody genes (17).

i) In the mouse, the major histocompatibility complex, or H-2 complex, was defined on the basis of its ability to mediate strong graft rejections (13). This chromosomal complex encodes at least 2 different categories of cell-surface molecules (Fig. 5). First, the K and D regions encode the classical transplantation antigens, cell-surface glycoproteins which appear to play a fundamental role in T-cell surveillance of infected or transformed cells (23). Second, the I region encodes a number of cell surface glycoproteins collectively designated the Ia antigens (see Ref. 24). These molecules are present on B cells and, at lower concentrations, on at least some T cells. The I region regulates a bewildering array of immune-related traits. They include the control of the ability of mice to respond to a wide variety of different antigens (immune responsiveness) and some of the cell-surface interactions of T cell-B cell cooperation. The role of the Ia antigens in these functions is uncertain.

Both the transplantation antigens and the I region gene products mediate specific lymphocyte interactions. The key to understanding these interactions is the realization that the genes of the K, D, and I regions are extremely polymorphic in mice. For example, inbred mice have at least 11 different K alleles and 10 different D alleles (25). Thus cells with one combination of K, D, and I phenotype can be tested against those with other combinations. Some experiments demonstrate that in order for specific lymphocyte interactions to occur, these gene products must be identical on the 2 interacting cell types (23). For example, killer-T cells and target cells must share the same K or D gene products in order for the destruction of infected or transformed cells to occur. Also, certain helper-T cells must share I region identity with the cells that stimulate their participation in the immune response (26). The molecular basis for these cell-cell interactions is uncertain; however, it does not appear to be a simple like-like interaction of identical molecules coded by the H-2 complex. Thus, gene products from the K, D, and I regions are involved in cellular recognition in the immune system and, accordingly, are area-code molecules.

ii) The antibody molecules displayed on B cells and the antibody-like molecules on T cells also mediate cellular interactions among lymphocyte clones. The collection of antigenic determinants on the receptor portion (V domain) of an individual antibody molecule is known as its idiotype (27). An idiotype defines serologically each distinct molecular species of antibody molecule, Three interesting observations related to cellular interactions in the immune system have been made about idiotypes. First, animals can make antibodies to their own idiotypes (antiidiotype antibodies) (28). Second, antiidiotype antibodies may enhance or suppress immune response (29). Third, the network of B cells, helper-T cells, and suppressor-T cells involved in a particular immune response exhibit antibody receptors with related idiotype and antiidiotype specifities (30). Thus cellular interactions in the immune response also are mediated by idiotype-antiidiotype interactions. Since the antibody molecule participates in cellular recognition phenomena, it also is an area-code molecule.

iii) Specific cellular interactions occur between lymphocytes and the stromal and endothelial cells of various lymphoid organs (31). For example, the lymph nodes and the spleen appear to have specific regions to which B cells or T cells migrate. Indeed, lymph nodes appear to have 3 separate regions, respectively, for B, Ly 1, and Ly 2 cells (32). Presumably specific sets of lymphocytes migrate to these areas because of specific cell-surface reactions with the underlying endothelial or stromal cells.

In summary, cellular interactions as well as precise cellular migrations are fundamental features of the immune system. Lymphocytes form a network of interacting cells that presumably play a fundamental role in regulating the immune response. Moreover, lymphocytes can specifically interact with other cell types such as macrophages and the stromal or endothelial cells of specific lymphocyte regions. Many of these interactions are mediated by cell-surface molecules. The programs for differentiation that lead to these specific cellular interactions are contained in T cells, in B cells, and in the other cell types with which lymphocytes specifically interact. Clearly these programs and their expression of appropriate area-code molecules must be coordinated with one another to produce these functionally interacting cellular networks.

Since the antibody molecule serves as a model for our thinking about area-code molecules, it is important to understand the molecular strategies it employs for its functions.

G. The Antibody Molecule Is Composed of Discrete Globular Domains Which Carry Out Distinct Functions

Light and heavy chains (33). Individual antibody molecules are composed of 2 identical light and 2 identical heavy polypeptide chains associated by noncovalent and disulfide interactions (Fig. 6). Each antigen-binding site requires a single pair of light and heavy chains.

Variable and constant regions (34). All immunoglobulin polypeptides can be divided into an amino-terminal portion, the variable (V) region, and a carboxy-terminal portion, the constant (C) region (Fig. 6). The V regions exhibit extensive amino acid sequence diversity and the C regions far more limited diversity. The variable regions encode the antigen-binding or recognition function, whereas the C regions encode a more limited number of effector functions such as complement fixation.

Homology units (35). Antibody polypeptides can be divided into homology units about 110 residues in length on the basis of amino acid sequence similarities. The heavy chain in Fig. 6 is comprised of 4 homology units (V_H, C_H1, C_H2, C_H3) and the light chain has 2 homology units (V_L and C_L). The V_H and V_L homology units exhibit extensive sequence homology with one another as do the C region homology units, These homology relationships indicate that the homology units of antibody genes share a common evolutionary ancestry.

Domains (35). Each pair of homology units (e.g., V_H-V_L, C_H1-C_L, C_H 2-C_H2 and C_H3-C_H3) folds into a compact globular domain (Fig. 6). The V_H and V_L homology units fold together to form a large crevice for antigen binding, whereas individual C domains carry out various effector functions. For example, the carboxy-terminal domain affixes the antibody molecule to the lymphocyte cell surface and is involved in the antigen-stimulated triggering of differentiation. Likewise, the C_H2 domain mediates complement fixation (36). Thus the antibody molecule is a sophisticated molecular machine that folds into discrete globular domains each of which may carry out different functions.

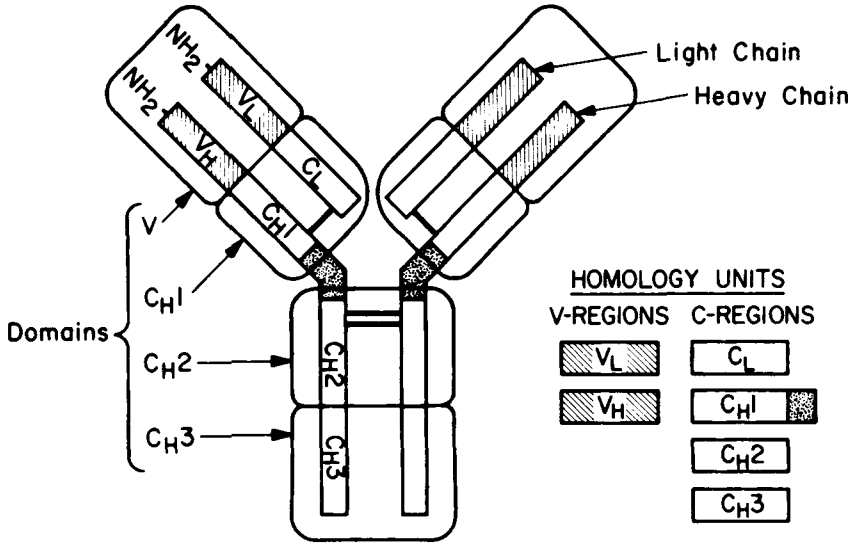

Fig. 6. Structure of an antibody molecule. [Reprinted from Ref. 74, with permission.]

The antibody molecule has fused together 2 discrete types of functions – recognition and effector. This functional dichotomy of the antibody molecule is reflected in the structure and organization of the antibody genes.

H. Antibody Genes Are Encoded as Three multigene Families With Two Distinct Types of Genes – Variable and Constant

The organization of the antibody gene families provides important insights into the possible organization of gene families for other area-code systems.

Three gene families. Classical genetic studies have demonstrated that 3 clusters or families of antibody genes, λ, κ, and H, for B cells are present in all mammals studied to date (Fig. 7). These gene families are genetically unlinked to one another. The λ and κ gene families code for light chains whereas the heavy gene family codes for heavy chains. Thus multiple gene families encode the antibody-receptor molecules.

Antibody gene families are multigenic (38). The number of antibody genes present in the germ line (or zygote) of a vertebrate organism is still a matter of controversy. However, most immunologists agree that the average antibody gene family has somewhere between 20 and thousands of genes. The antibody gene families are multigenic in nature.

Separate V and C genes. The variable and constant regions of antibody polypeptides appear to be encoded by separate germ line genes (Fig. 7). The evidence for this surprising gene organization is compelling. Initially this supposition was based on genetic, serological, and amino acid sequence analyses (40). Subsequently this problem has been approached directly by the use of restriction endonucleases which cleave DNA at specific recognition sites (41, 42). When genomic DNA is digested by such an enzyme and the resulting fragments are separated by size, a particular gene present as a single copy in the genome will be found only in one or a few fractions. Accordingly, linkage relationships between 2 or more genes (or the lack thereof) can be determined by examining these fractions with appropriate nucleic acid probes (radio-labeled mRNA or cDNA). In this manner, it has

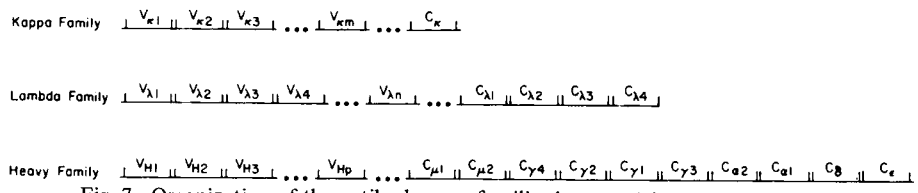

Fig. 7. Organization of the antibody gene families in man. Adapted from Ref. 34.

been shown that the V and C genes for light chains from the mouse are in separate restriction fragments in undifferentiated mouse embryo DNA, but on the same restriction fragment (and presumably joined) in differentiated myeloma tumor DNA. More recently, DNA sequence analysis of a V gene isolated from mouse embryonic DNA has confirmed that the C gene is not adjacent to the V gene in undifferentiated DNA (43). Obviously, the separate organization of V and C genes has important implications for mechanisms of lymphocyte differentiation.

I. The Translocation of V and C Genes in Lymphocytes During Their Differentiation Appears to be a Fundamental Mechanism of Commitment

V-C translocation and antibody polypeptides. Since the V and C genes are separted in the embryo, these genes or mRNAs must undergo a rearrangement during somatic differentiation to form a contiguous VC gene or mRNA that is translated into a single polypeptide chain. This probably occurs by a DNA translocation event that joins the V and C genes (Fig. 8).[1]

V-C translocation and cellular commitment. Each mature lymphocyte expresses one type of antibody molecule. We feel that DNA translocation is a fundamental component of the molecular mechanism for committing a single lymphocyte to the expression of one type of antibody molecule (44) (Fig. 8). The implications of this hypothesis are extremely interesting with regard to development. The new arrangement of genetic material, including the V and C genes which have been joined as a part of the developmental process, becomes a heritable property of the daughter cell lines. The process of joining a specific V and a specific C gene is a definitive example of a commitment event. It provides a simple mechanism for limiting the future options open to a given cell lineage. It also provides a mechanism to explain how the memory of developmental decisions can be maintained throughout cell division and passed on to subsequent generations of somatic cell lines of the same lineage. The new arrangement of DNA in differentiated lymphocytes is simply replicated and passed to the daughter cells along with the newly activated genetic programs which control future events in the lineage.

[1] The nucleic acid studies on antibody genes provide unequivocal evidence for 2 suppositions. First, the DNA sequence data on the embryonic V gene (43) provide compelling evidence that the V and C genes are distinct in the germ line. Second, the restriction enzyme studies on embryonic and differentiated DNA (44, 45) argue that a DNA modification event has occurred during lymphocyte differentiation. This DNA modification could be anything which alters the restriction enzyme sites in the adult DNA with respect to embryonic DNA. The obvious DNA modification to explain these restriction enzyme patterns would be a translocation of the V and C gene sequences (Fig. 7). Other DNA modifications that do not rearrange V and C gene sequences require 2 assumptions. i) DNA modification such as base methylation changes multiple restriction enzyme sites. ii) The separate V and C genes are transcribed as a single mRNA. In this regard, recent observations suggest that a single mRNA can be obtained from separate DNA segments on the adenovirus chromosome (105, 106). These alternative possibilities for DNA modification can be tested by examining the organization of V and C genes in myeloma DNA. However, we tend to favor the simpler hypothesis of DNA translocation and assume throughout this paper that the DNA modification event in lymphocyte differentiation is a joining of the V and C genes.

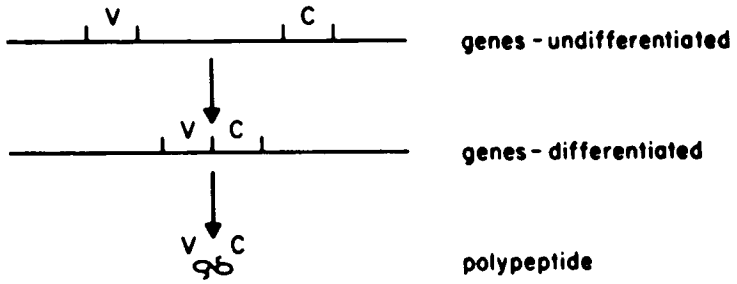

Fig. 8. Illustration of the differentiation of a lymphocyte through V-C joining. [Reprinted from Ref. 38, with permission.]

V-C translocation may alter developmental programs. The V-C translocation during somatic differentiation in lymphocytes has 2 distinct functions. First, it joins together separate V and C genes into a single, contiguous V-C sequence which codes for a complete antibody chain.[1] Second, it changes the developmental program of the lymphocyte so that it is hereafter committed to the expression of a single antibody polypeptide. Thus V-C translocation may alter the organization of control as well as structural elements and thereby alter the future developmental options of that particular cell lineage. We view this DNA modification event as a fundamental feature of differentiation in the antibody system and as we shall discuss subsequently in the differentiation of other area-code systems. Moreover, DNA translocation may be involved in the orderly read out of antibody genes during development.

J. Antibody Genes May Be Expressed During Differentiation in an Orderly and Programmed Manner

Development requires the orderly expression of phenotypic information. For example, during the differentiation of the T-cell lineage, the differentiation antigens are expressed at precise developmental stages (Fig. 3a). Accordingly, the set of unlinked genes that codes for these molecules (Fig. 3b) must be expressed in a coordinated and orderly manner.

Antibody V genes may also be expressed during differentiation in a coordinated and orderly manner (Fig. 9). Although the issue is controversial (see Ref. 45), several lines of evidence suggest that animals acquire the ability to respond to different antigens at different stages of development of the immune system. i) The development of individual clones of specific antibody-producing cells can be followed using x-irradiated mice as an in vivo tissue culture system (46). When lymphocytes are taken from neonatal mice, the ability to respond to the dinitrophenyl and trinitrophenyl groups appears by the first day after birth, to fluorescein by 3 days after birth, and to phosphorylcholine at 6–7 days after birth (47). All mice in an inbred line seem to follow this same developmental progression. Moreover, isoelectric-focusing analyses of these antibody molecules suggest that the same major molecular species are expressed in each mouse at each developmental stage. ii) In the bursa of Fabricius of individual chicken embryos, specific antigen-binding lymphocytes for keyhole limpet hemocyanin and poly-L(Tyr,Glu)-poly-D,L-Ala-poly-L-Lys appeared earlier than those binding sheep erythrocytes (48). Once again this implies that certain antibody molecules are expressed during differentiation before others. iii) Since the

bursa appears to be the central organ for B-cell differentiation in chickens, an attempt has been made to remove this organ at various stages of chicken development and determine whether there is a corresponding loss of the ability to express certain antibody polypeptides (Huang and Dreyer, in preparation). Early ablation should delete most of the antibody repertoire, whereas later ablations should permit the reproducible expression of an ever increasing fraction of the total repertoire. Preliminary studies of this type employing the very sensitive technique of two-dimension gel electrophoresis bear out these predictions and further suggest that there may be an ordered and reproducible expression of light chains.

Each of these individual studies can be given alternative interpretations (45), but taken together they raise the intriguing possibility that antibody genes may be read out during development in a programmed and orderly fashion (Fig. 9). If so, several interesting questions are raised. During the differentiation of lymphocytes, are the V genes in a given multigene family sequentially translocated to their corresponding C genes? Is the V-C translocation mechanism an integral part of the development program for reading out V genes? Is there some type of mechanism for coordinating the readout of 2 multigene families (e.g., light and heavy chain gene families) so that particular molecular combinations of these 2 distinct polypeptide chains may be expressed in a programmed and reproducible fashion? How might genetic translocation help set up future developmental programs? Clearly those area-code systems mediating cell-cell recognition during embryogenesis must be capable of expressing their information in an orderly and programmed manner consistent with the orderly nature of growth and development.

K. Antibody Genes Appear to Evolve From a Common Precursor Gene

The evolution of antibody genes gives important insights into the possible evolutionary pathways of other area-code systems and raises the intriguing possibility that some distinct area-code systems may share common gene ancestors.

Antibody polypeptides are linearly differentiated into homology units of about 110 amino acid residues (Fig. 6). The constant homology units demonstrate significant amino acid sequence homology to one another as do the variable homology units (35). The existence of variable and constant region homology units and the observation by x-ray crystallographic analysis that the tertiary structures for the V and C homology units are very similar (49, 50) indicate that antibody genes probably evolved from a precursor gene coding for a single ancestral homology unit. One hypothetical evolutionary scheme is depicted in Fig. 10. The hypothetical precursor gene duplicated at a very early time to produce ancestral V and C genes. These gene products presumably assumed primitive area-code functions on membranes. Once a V-C translocation mechanism evolved, the V gene library could be expanded to generate a primordial multigene family.

This original family may have coded for primitive membrane area-code molecules (Fig. 10). This multigene family was in turn duplicated either by polyploidization or duplication and translocation of a chromosomal fragment to produce primitive antibody gene families. Subsequent duplication of this primitive multigene family produced the 3 families that evolved to become contemporary λ, κ, and H families. Contiguous or fused gene duplication in the heavy chain family led to the different C_H genes, each comprised of 3 or 4 homology units. Homology units, folding to comprise domains with different functions, can be added to (or deleted from) C_H genes in the course of evolution. Accordingly, the evolution of antibody genes employed all the major mechanisms of gene evolution – point mutation, discrete duplication, contiguous duplication, polyploidization,

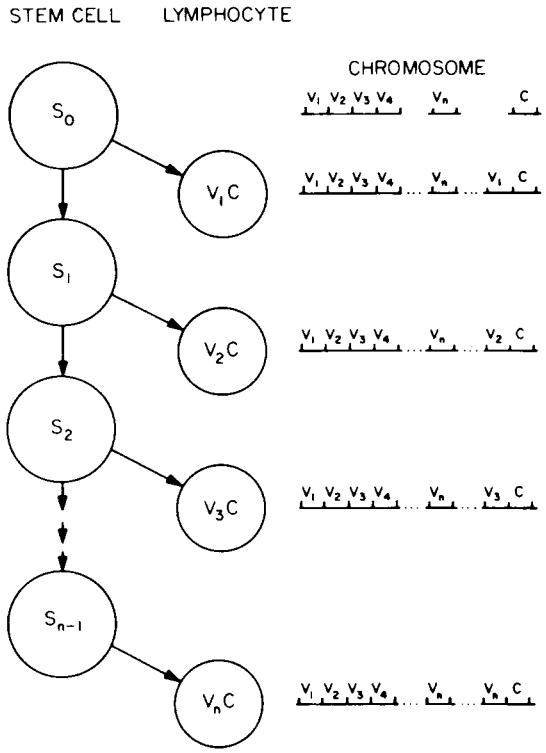

Fig. 9. A model of the linear and programmed read out of V genes in a multigene family.

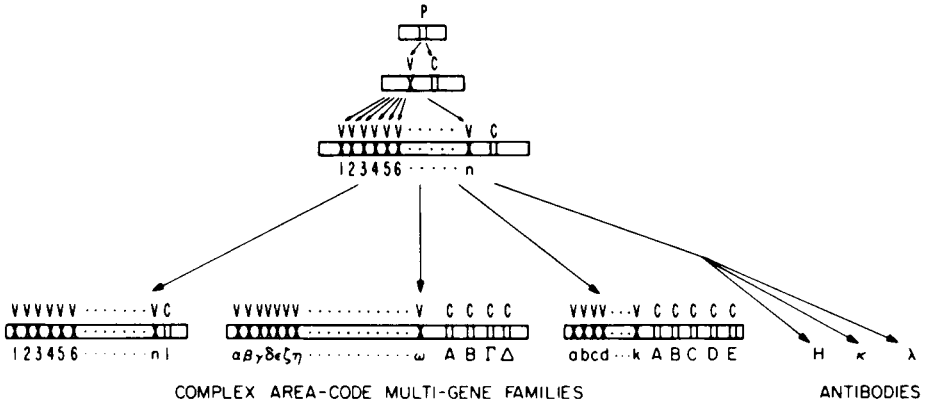

Fig. 10. A hypothetical model for the evolution of the antibody and other area-code gene families.

and/or translocation. As we shall discuss subsequently, the antibody gene families may share a common evolutionary origin with other cell-surface recognition systems (Fig.10).

L. The Immune System Employs a Variety of Mechanisms for the Amplification of Information

An enormous amount of information is required to develop a eukaryotic organism. Thus it is of interest to analyze those strategies the immune system employs for the amplification of information. Once again, many of these strategies will almost certainly be employed by other area-code systems.

The vertebrate immune system can respond specifically to a universe of different antigenic determinants by virtue of molecular interactions with complementary antibody molecules. How then can the gene products from a finite number of antibody genes react with untold numbers of different antigenic determinants? The basic strategies for the amplification of antibody information fall into 2 broad categories – genetic and molecular (Table III). Genetic strategies amplify information by producing multiple antibody genes, whereas molecular strategies amplify information by employing certain fundamental characteristics of antibody molecules themselves. i) The antibody gene families are multigenic and thus encode directly multiple receptor molecules (38). ii) Somatic mutation may occur in germ line antibody genes during the differentiation of individual lymphocytes to produce additional antibody genes (52, 53). iii) At the genetic level different V genes may associate with the same C gene (e.g., V_1C_μ, V_2C_μ, . . .V_nC_μ). In addition, the same V gene may associate with different C genes (e.g., $V_1\,C_\mu$, $V_1\,C_{\gamma 1}$, $V_1C_{\gamma 2}$, $V_1\,C_{\gamma 3}$, . . .$V_1\,C_\epsilon$). Thus the combinatorial-joining mechanism of DNA translocation allows one library of recognition sites to be combinatorially associated with a second library of effector functions. Moreover, during evolution crossing-over can add (or delete) homology units to C_H genes. Thus various combinations of domains can be associated in a single molecule by evolutionary mechanisms. iv) The diversity of antigen-binding sites can be increased by the combinatorial association of light and heavy chains. For example, if 10^3 different L chains could associate with 10^3 different heavy chains, 10^6 different antibody molecules will be produced ($10^3 \times 10^3 = 10^6$). Thus unrestricted light and heavy chain interactions will generate an amplification factor of $p \times q$, where p equals the number of light chains and q the number of heavy chains. v) Multispecificity is defined as the ability of a single antibody molecule to interact with a variety of different antigens, some presumably related in tertiary structure and others possibly unrelated. Thus the inherent degeneracy of the antigen-binding site is an important mechanism for amplifying the information contained in a discrete number of antibody V genes. vi) At the supramolecular level, cell-cell recognition phenomena may involve combinations of 2 or more distinct species of cell-surface molecules. For example, T-cell surveillance of virus-transformed cells requires the simultaneous recognition of both viral and transplantation antigens on the transformed cells (23). One explanation for this simultaneous dual recognition of 2 distinct molecules is that the viral antigen and the transplantation antigen associate at the cell surface to form a supramolecular complex. Obviously, the combinatorial association of cell-surface molecules can lead to significant amplification of the number of distinct cell-surface recognition units.

Thus the immune system displays a variety of strategies at the gentic, protein, and surface-display levels for the amplification of information (Table III). Other membrane recognition systems will certainly employ similar strategies.

TABLE III. Levels at Which Information Amplification Occurs

Genetic level
1. Multiple germ line genes
2. Somatic mutation
3. Combinatorial joining of V_H and C_H genes
4. Association of different homology units by crossing over during evolution.
Protein level
1. Combinatorial association of subunits
2. Multispecificity
Cell-surface displays
1. Combinations of 2 or more distinct species of cell-surface molecules

M. Summary of Features of the Immune System Which Relate to the Area-Code Hypothesis

Area-code molecules. The immune system is an intriguing microcosm of the differentiating organism. Lymphocytes develop along 2 separate cell lineages – B cells and T cells. Cell-surface molecules are expressed at various developmental stages of these lineages. Area-code molecules confer upon lymphocytes specific cell-surface addresses that direct 2 important recognition phenomena – cell-cell interactions and migration to specific tissues. The antibody molecule, a prototype area-code molecule, is divided into discrete molecular domains which carry out distinct recognition and effector functions. The immune system employs combinatorial mechanisms at the genetic, evolutionary, molecular, and supramolecular levels to amplify information.

DNA modification. The commitment of a lymphocyte to express a particular antibody polypeptide appears to require the DNA translocation of distinct V and C genes. This somatic modification of chromosomes can be passed on to progeny in a stable and heritable fashion. The DNA translocation even may create new structural genes and also alter the organization of control elements, thus modifying the future developmental options of a particular cell lineage.

Multigene families. Antibody molecules are coded for by 3 multigene families. Multigene families are a fundamental unit of chromosomal organization and evolution in the immune system and presumably in other complex eukaryotic systems.

Let us now consider the area-code hypothesis in more general terms.

IMPLICATIONS OF THE AREA-CODE HYPOTHESIS

A. The Area-Code Hypothesis Suggests That Cell-Surface Recognition Molecules Form a Recognition Code on the Cell Surface

General. As a cell undergoes successive stages of differentiation, changing patterns of area-code molecules are displayed on the cell surface (Fig. 11). These molecules, singly or in groups, constitute a cell-surface display system that gives a cell or a group of cells an individual address much as the collective digits in a phone number or postal zip code identify individual locations. Area-code molecules expressed early in the differentiation of cell lineages if not lost, will constitute a portion of the cell-surface address that identifies those cell lineages with an earlier embryonic origin than other area-code molecules ex-

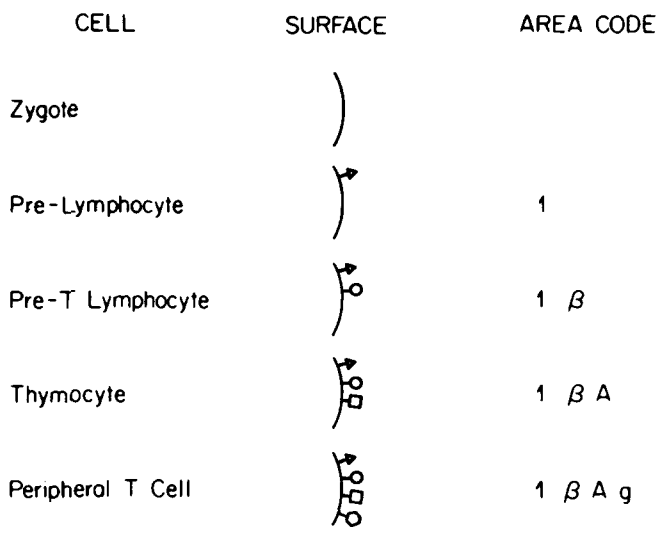

Fig. 11. The model of the display of area-code molecules. For the sake of simplicity the possibilities of loss of area-code molecules, simultaneous appearance of 2 or more area-code molecules, sharing of area-code molecules, etc., are not indicated. The use of the T lymphocyte lineage is for illustration purposes only.

pressed late in differentiation. For example, the α molecules expressed early in the differentiation of the hypothetical cells in Fig. 12 identify earlier lineage relationships than do the A molecules expressed somewhat later. Thus one can understand how the early digits in the area-code address specify general tissue relationships, whereas the later digits confer individual specificity (Fig. 12).

The area-code molecules which constitute these cell-surface displays may serve as molecular addresses to direct the migrating embryonic cells to appropriate locations or to permit specific cell-cell interactions. From Fig. 12 it can be seen how a cell which specifically recognizes, for example, the α, 1, B cell, could search and find that cell. The searching process may be a random one, or it may take advantage of the adjacent cells with slightly different area codes. For example, the cell could first find α, 10, B, then α, 9, B, and so on up the "cell-surface gradient" until it contacts α, 1, B. Thus a cell's search might be towards an increasingly better match between its area-code molecules and those of the target cell. If it strays off the path, to, e.g., α, 8, A, then the matching would decrease, and the cell could act accordingly by returning to the appropriate cell-surface gradient. Once at the target site, with its area-code maximally matched, the cell would lose its motility, perhaps through a process analogous to contact inhibition. Thus cell-surface as well as diffusible molecule gradients may play an important role in development.

Diverse cell-surface addresses may be generated by the various mechanisms employed by the immune system (Table III). These include genetic, protein, and cell-surface display strategies. These mechanisms are capable of generating an enormous array of distinct cell-surface addresses from relatively few germ-line genes. Cells may interact with one another via area-code molecules by 1 of 2 general mechanisms – self-self recognition or lock-and key recognition (Fig. 13). Thus area-code molecules play a fundamental role in the highly precise cell-cell interactions of embryogenesis.

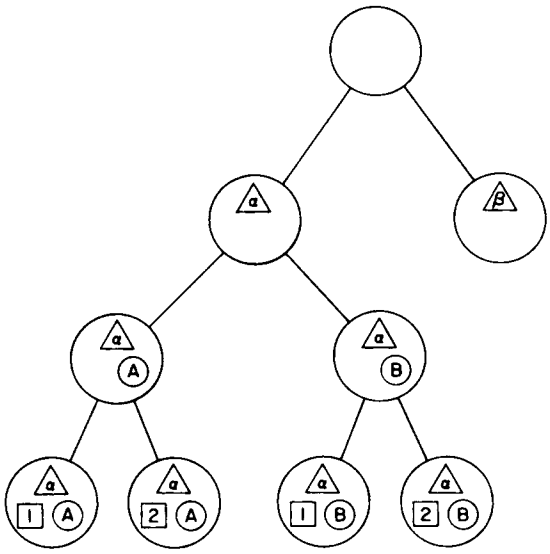

Fig. 12. A model of the lineage of a differentiation cell line. α, β, A, B, 1, and 2 denote area code molecules.

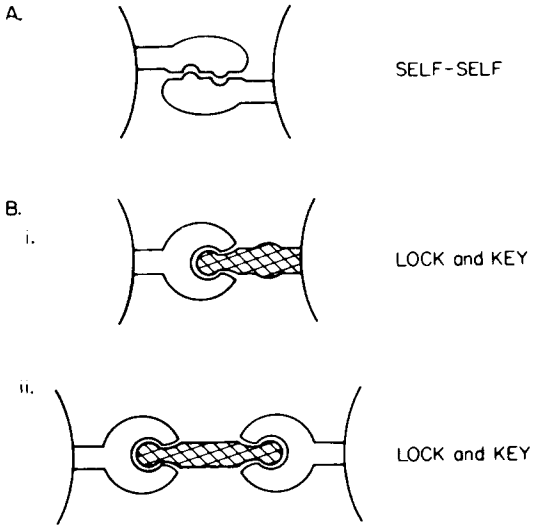

Fig. 13. Models for specific cell-cell interactions. A) Self-self. B) Lock-and-key.

B. DNA Modification May Be a Fundamental Mechanism for Cellular Differentiation

The generality of DNA modification in somatic cell lines. In addition to the translocation of V and C genes, there is evidence for DNA modification during differentiation in a variety of developmental systems. For example, chromosomal diminution or loss occurs in organisms as diverse as dipteran insets (43), ascarid nematodes (55), and copepodes

(56). In these cases, somatic cells discard much of the DNA carried by germ-line cells. This loss is due to the fragmentation or reduction in size of chromosomes when somatic cell lineages diverge from the germ line. In addition, various types of chromosomal modifications are seen in the somatic cells of a variety of organisms such as maize (57, 58), snapdragons (59), and Drosophila (60). These genetic events, apparently due to the insertion and excision of DNA elements, include the turning on and off of genes, and the transposition, deletion, and inversion of chromosomal segments. Some of these DNA modifications appear to occur in a more or less programmed fashion while others are viewed as genetic instabilities leading to random genetic changes during development. Some of these phenomena appear to be very similar to those exhibited by the insertion sequences of prokaryotes (61). Insertion sequences are specific DNA sequences which insert into and excise from bacterial chromosomes presumably by crossing over (62). Indeed, the alternative expression of 2 genes in Salmonella for flagellar proteins is mediated by DNA inversions (63). Accordingly, it appears that DNA modifications may play a role in gene expression in a wide variety of systems. Admittedly few examples have been described for eukaryotes, but the search for DNA modification during somatic differentiation is an extremely active area of research (64). Our feeling is that DNA modification will be a fundamental mechanism of gene expression in simple as well as complex gene systems (65, 66).

DNA modification and regulatory elements. B cells never express a V gene or a C gene unless they have been joined. Thus the translocation process creates a complete structural gene as well as activating (and rearranging) the regulatory elements necessary for its expression. So far the modification or rearrangement of structural genes has been found only in the immune system. However, the modification or rearrangement of regulatory elements may be more common. Single genes can be turned on or off by appropriate modification of its regulatory element, as seen with the Salmonella flagellar proteins (63). In addition, a whole gene complex may be activated by the modification of a regulatory element. For example, a reversible inversion of a promoter element has been postulated for the alternate and mutually exclusive expression of 2 closely linked gene complexes which determine alternate mating types in the fission yeast (67). If one of the activated genes has a regulatory product (e.g., an inducer), then genes controlled by that product would be affected, even if they are on another part of the genome. This appears to be the case with the *Spm* system in maize (58). In general, whole batteries of genes, linked or unlinked, could be controlled by the DNA modification of control elements. Thus area-code systems may regulate gene expression through the DNA modification of control elements.

Do the nuclear transplantation experiments of Gurdon and others argue against DNA modification as a general mechanism of differentiation? It is possible to transplant a somatic nucleus into an enucleated egg and then to stimulate this chimeric cell to begin embryogenesis (68, 69). When nuclei from blastula cells are transplanted to enucleated eggs, a high fraction of the eggs develop normally to form fertile, adult frogs. When nuclei from the intestinal cells of young embryos are used, most become defective embryos but a very small number produced fertile, adult frogs (70). These experiments suggest to some that differentiation is a reversible process and therefore the DNA in all somatic cells is identical to that of the germ cells. However, an important reservation about these experiments should be noted. There are striking differences in the results obtained with blastula and somatic cell nuclei. Clearly these nuclei have quite different potentials for generating adult frogs. Perhaps germ cells (or early stem cells), which are known to migrate through

the intestinal mucosa, are the source of the nuclei which produce the few fertile adults in the nuclear transplantation experiments carried out with putative intestinal nuclei. To rule out this possibility, nuclei from lymphocytes and differentiated epidermal cells of adults have been transplanted into enucleated eggs (71, 72). In these cases, no fertile adult frogs are obtained. All embryos died at or before the early tapole stage. Indeed, the nuclei from an aneuploid liver cell line from a frog supports development almost as well as the adult somatic cell nuclei (73). These results are not inconsistent with the proposition that DNA modifications occur as somatic cell lineages undergo development. Perhaps somatic nuclei support development in nuclear transplantation experiments up until the stage at which the genes which have been modified are needed.

We feel that the nuclear transplantation experiments do not rule out the possibility that DNA modifications occur as a general phenomena in the differentiation of somatic cells. Certainly the experiments already discussed in relation to the development of the immune system provide compelling evidence for DNA modification in lymphocytes. The evidence for DNA modifications is also compelling in a variety of other systems as discussed above.

C. Multigene Families Appear to be a Fundamental Unit of Eukaryotic Gene Organization and Evolution

General. Multigene families display 4 fundamental characteristics: multiplicity, close linkage, sequence homology, and similar or overlapping functions (see Ref. 74 for review). A variety of eukaryotic genes exhibit these properties including the ribosomal RNA genes (75), tRNA genes (76), histone genes (77), the β-like globin genes (78), and the DNA satellites (79). Multigene families may range in size from a few gene members to thousands of gene members. Indeed, simple multigene families appear to code for several of the differentiation antigens found on lymphocytes including TL (80), Qa (81), and Ia (82). Accordingly, area-code families may range from few to many gene members.

Evolution of primordial multigene families. Multigene families are found in eukaryotes but not in prokaryotes (74). Multigene families may have evolved in response to the informational requirements of differentiation in multicelluar organisms (see Fig. 1). This supposition suggests that multigene families will be employed in cell-recognition systems and accordingly, will code for a variety of area-code molecules.

New evolutionary unit. New multigene families can arise in 2 ways: 1) gene duplication from a single gene, or 2) duplication of all or part of a preexisting multigene family. The duplication of a preexisting multigene family may occur in 2 ways: tetraploidization (see Ref. 83) or duplications and translocation of a portion of a chromosome. It obviously requires a long period of evolutionary time to produce a large multigene family starting with a single gene. In contrast, the duplication of an entire multigene family presumably generates new gene families by a single genetic event. Thus evolution can proceed at a more rapid pace by the instantaneous generation of entire families of new genes. The duplication of multigene families is illustrated by the homology relationships of the antibody gene families which suggest that all 3 gene families were derived from a common ancestral multigene family. Thus the multigene family is a basic unit of eukaryotic evolution. New and complex area-code gene families can obviously be created rapidly.

Homologies among multigene families. A new multigene family arising by duplication from a primordial gene family may interact with its sister-gene family as occurs between the light and heavy chain gene families of antibodies. Alternatively, it may evolve to assume new functions. Evolutionary relationships among multigene families might be

discerned from 2 aspects of homology: i) amino acid sequence homology among gene products and ii) shared complex control mechanisms such as V-C translocation. Once more data are available on a variety of area-code systems, these homology analyses should allow us to construct a genealogical tree of relationships among the multigene families encoding these area-code systems.

POTENTIAL SYSTEMS EMPLOYING THE AREA-CODE STRATEGY

A. The Immune System

Antibodies. Antibody molecules and genes provide a model system for posing questions about other complex area-code systems (Table IV). Other area-code systems may employ some but not necessarily all of the strategies of the antibody system. Indeed, some will probably present novel strategies and characteristics. However, the questions posed in Table IV help us to approach experimentally other area-code systems, both simple and complex, at the protein, genetic, regulatory, and evolutionary levels.

The study of antibody molecules and genes is currently under intensive investigation in many different laboratories (see Ref. 84). Important unresolved experimental questions include the following: i) What are the mechanisms responsible for antibody diversity? ii) What is the molecular mechanism of V-C translocation or joining? iii) How are antibody genes and their control elements organized on vertebrate chromosomes? iv) How many V genes are present in various antibody gene families? v) Are V genes translocted to C genes according to an orderly developmental program? vi) What developmental strategy is used for programming specific combinations of light chain genes and heavy chain genes? Certainly the investigation of these and many other questions will provide important insights into this model area-code system.

T-cell receptors. The T-cell receptor has been an elusive entity. Recently, however, serological and genetic studies suggest that the T-cell receptor for antigen binding employs a V_H gene from the B-cell H chain family presumably associated with a new C_H gene (see Ref. 85). The presence of a light chain in the T-cell receptor is still a matter of uncertainty (86). Thus T and B cells both employ the same V_H gene family. However, the expression of V_H genes on T or B cells appears to be controlled by the nature of the C_H gene to which they are translocated. The various C_H genes given in Fig. 7 are expressed on B cells, whereas the putative C_H gene(s) for T cells is expressed only on that cell lineage. Thus the same area-code gene library can be employed for 2 distinct receptor systems by virtue of the DNA translocation mechanism and the differential expression of C_H genes in the T- and B-cell lineages.

The study of T-cell receptors is in its infancy. Very little is known about the structure of these molecules. All of the questions raised for antibody molecules and genes of B cells can be asked of the T-cell receptor (see above and Table IV).

H-2 complex. Chromosome 17 of the mouse appears to code for a series of multigene families (87) (Fig. 14). For example, by simplistic genetic calculations the region between K and Tla has sufficient DNA to code for about 16,000 polypeptides 100 residues in length (e.g., V region size). The T/t family with at least 6 complementation groups appears to encode cell-surface proteins regulating neuroectodermal development during embryogenesis (1). The S family codes for the structural and/or regulatory elements of several complement components (88). The Qa region appears to have 2 or more genes coding for differentiation antigens on lymphocytes (89). The D (90), I (87), and Tla (91)

TABLE IV. Questions About Potential Area-Code Molecules and Genes Based on an Analysis of the Immune System

A. Protein level
   1. Are the molecules composed of multiple subunits?
   2. If so, do they exhibit combinatorial association?
   3. Are homology units present? How large are they?
   4. Do they show homology to immunoglobulins? To other known gene systems?
   5. Are V and C regions present?
   6. What type of molecular recognition functions do they perform?

B. Genetic level
   1. Are the polypeptides encoded by multigene families?
   2. Are V and C genes present?
   3. How large is the gene family?
   4. How are the various control and structural genes organized?

C. Regulatory level
   1. Is there V-C translocation?
   2. What is the pattern of clonal expression of individual area-code molecules?
   3. What is the nature of the programs which govern development?

D. Evolutionary level
   1. Are the multigene families within the system homologous to one another?
   2. Are they related to other known multigene systems?

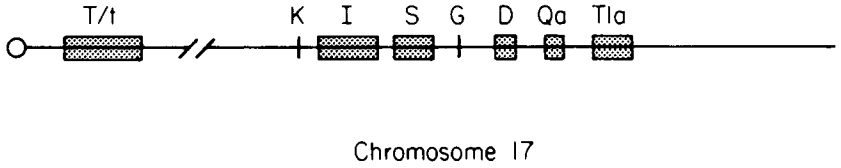

Fig. 14. Chromosome 17 of the mouse. Shaded bars represent 2 or more closely linked homologous genes.

gene families also appear to have 2 or more homologous gene members. Moreover, indirect evidence suggests that the genes coding for the transplantation antigens may be present in multiple copies (92). Accordingly, chromosome 17 of the mouse appears to be a library of gene families coding for cell-surface molecules.

The gene products of the K, D, Qa, and Tla gene families show 2 interesting relationships to one another. i) All are cell-surface glycoproteins of about 45,000 molecular weight (89, 93). ii) Each is associated with β_2-microglobulin, a polypeptide of about 100 residues that demonstrates significant homology with homology units of antibody constant regions (94, 95). These relationships suggest that these differentiation antigens may be homologous and thereby share a common evolutionary origin. This supposition is supported by preliminary amino acid sequence analyses of the N-terminal regions of the K and D molecules that demonstrate sequence homology. A similar analysis has not yet been carried out on the Qa and Tla gene products. The association between these differentiation antigens and β_2-microglobulin provides a fascinating relationship between the antibody system and these cell-surface molecules. Indeed, amino acid sequence data have demonstrated a striking homology between an internal 17-residue portion of a human transplantation antigen and a portion of the antibody V region (C. Terhorst and J. Strominger,

personal communication). If further sequence analyses confirm this observation, it appears that antibodies and transplantation antigens may be evolutionarily r[illegible]ed. It will be interesting to determine whether the same is true of the Qa and Tla gene products.

The functions of these cell-surface molecules are uncertain, but the Qa, Tla, and transplantation antigens are expressed in different patterns on different populations of lymphocytes. Thus these cell-surface molecules may be encoded by small multigene families that diverged from a common ancestor and which have diverged to carry out distinct cell-surface functions.

One gene product from the T/t system has been characterized and it also appears to be 45,000 in molecular weight and associated with a β_2-microglobulin-like molecule (96). The I region molecules differ in molecular weight from the transplantation antigens and do not show obvious sequence hemology at their N termini (97). Preliminary amino acid sequence studies indicate that the non-membrane-associated S region gene products also fail to show homology to immunoglobulins (98). Clearly additional data are necessary before any convincing relationships can be established among these gene families.

Certain of these differentiation antigens carry out cell-surface recognition functions (e.g., K, Ia, and D) and, accordingly, are area-code molecules. The others may be area-code molecules, although their functions are generally unknown. We are in a position to ask of these gene families many of the questions listed in Table IV.

B. Other Developmental Systems

General. There are a variety of developmental systems where area-code molecules appear to play an important role. For example, embryonic tissues appear to have cell-surface recognition molecules that mediate tissue-specific cellular interactions (99). The imaginal disks of Drosophila larva can be disassociated and reaggregated in a manner that suggests area-code molecules may be involved (100). The retinal nerves of the goldfish upon sectioning appear capable of reattaching to their specific tectal counterparts (101). One possibility is that these specific interactions are mediated through the cell-surface addresses provided by area-code molecules. The difficulty with each of these systems has been that large quantities of homogeneous area-code molecules are difficult to obtain. A promising approach is to create homogeneous cell lines from individual cells in these systems. Homogeneous cell lines could be produced from cells transformed by viral, chemical, or physical agents. In addition, permanent cell lines may be generated by fusing primary cells with established cell lines. This latter approach has been successful in creating cell lines secreting homogeneous anti-H2 antibodies by fusing normal B cells synthesizing anti-H-2 antibodies with B-cell tumors (102). The production of large quantities of cells from homogeneous cell lines appears to be one of the primary prerequisites for the detailed biochemical analysis of many interesting developmental systems.

Tumor-specific antigens. Certain tumor-specific antigen systems appear to be potential area-code systems which can be studied currently. One of the key experimental approaches to the study of the immune system was the use of tumors cultured in vivo and in vitro. These tumors were homogeneous clones of lymphoid cells and provided large amounts of experimental material. Since we believe that different embryonic cells have different cell-surface arrays of area-code molecules, we might ask if there are presently available tumor cell lines which display a large diversity of line-specific surface molecules or antigens. In fact, such individually specific cell-surface antigens have been found on a wide variety of tumors induced by physical or chemical agents. In certain inbred strains of mice both methylcholanthrene (103) and ultraviolet light (104) induce tumors with in-

dividually unique tumor antigens. The repertoire of these unique tumor antigens appears very large (103). Moreover, large quantities of tumor cells can be raised for analysis of the tumor antigens and their genes. It will be important to determine whether these antigens are abnormal molecules related only to tumor transformation or whether they might indeed be area-code molecules playing a fundamental role in some stage of development. For these systems it will be interesting to answer many of the questions raised in Table IV.

CODA

An analysis of the immune system has led to the area-code hypothesis. The strategies employed by the antibody system for gene organization, regulation and evolution provide a model for asking experimentally testable questions about other potential area-code systems (Table IV). Investigations of area-code systems have 2 basic requirements. First, homogeneous clones of cells expressing the potential area-code molecules must be available. Second, microchemical techniques must be employed to characterize these molecules which are generally available in very small quantities. Our laboratories have been engaged in developing microsequencing techniques for the last several years. We are now applying these techniques to the analysis of several potential area-code systems (93, 97) including the ultraviolet light-induced tumors of C3H mice. This tumor system may provide a second detailed view of a complex area-code system.

ACKNOWLEDGMENTS

We thank the Caltech secretaries for their patience. This work was supported by grants NSF PCM71-00770, USPHS A1-10781, USPHS CA15926, and GM06965. H.V.H. is a Danforth Fellow.

REFERENCES

1. Bennett D: Cell 6:441, 1975.
2. Changeux J-P, Danchin A: Nature (London) 264:705, 1976.
3. Sperry RW: Proc Natl Acad Sci USA 50:703, 1963.
4. Cantor H, Boyse EA: Immunol Rev 33:105, 1977.
5. Jerne NK: Sci Am 229:52, 1973.
6. Metcalf D, Moore MAS: In Neuberger A, Tatum EL (eds): "Haemopoietic Cells." Amsterdam: North Holland Research Monographs, 1971, Frontiers of Biology Series, vol 24, p 172.
7. Cantor H, Weisman I: Prog Allergy 20:1, 1976.
8. Houssaint E, Belo M, Le Douarin NM: Dev Biol 53:250, 1976.
9. Good RA: Harvey Lect 67:1, 1973.
10. Cazenave P-A, Cavaillon JM, Bona C: Immunol Rev 34:34, 1977.
11. Boyse EA, Old LJ: Annu Rev Genet 3:269, 1969.
12. Storrie B, Goldstein C, Boyse EA, Hammerling U: J Immunol 116:1358, 1976.
13. Klein J: "Biology of the Mouse Histocompatibility-2 Complex." New York: Springer-Verlag, 1975, p 135.
14. Takahashi T, Olds J, Boyse EA: J Exp Med 131:1325, 1970.
15. Raff M: Nature (London) 242:19, 1973.
16. Jerne NK: Ann Immunol (Inst Pasteur) 125C:373, 1974.
17. Jerne NK: Harvey Lect 70:93, 1976.
18. Jerne NK: "27th Mosbacher Colloquium." Berlin: Springer-Verlag, 1976, p 259.
19. Cantor H, Shen FW, Boyse EA: J Exp Med 143:1391, 1976.
20. Shevach EM: Fed Proc Fed Am Soc Exp Biol 35:2048, 1976.

21. Metcalf D, Moore MAS: In Neuberger A, Tatum EL (eds): "Hemopoietic Cells." Amsterdam: North Holland Research Monographs, 1971, Frontiers of Biology Series, vol 24, p 70.
22. Katz DH, Benacerraf B (eds): "The Role of Products of the Histocompatibility Complex in Immune Responses." New York: Academic Press, 1976.
23. Doherty PC Götze D, Trinchieri G, Zinkernagel RM: Immunogenetics 3:517, 1976.
24. Möller G (ed): "Biochemistry and Biology of Ia Antigens." Immunol Rev 30, 1976.
25. Klein J: Annu Rev Genet 8:63, 1974.
26. Rosenthal AS, Shevach EM: J Exp Med 138:1194, 1973.
27. Cosenza H, Julius MH, Augustin AA: Immunol Rev 34:3, 1977.
28. Rodkey LS: J Exp Med 139:712, 1974.
29. Eichmann K, Rajewsky K: Eur J Immunol 5:661, 1975.
30. Cantor H: In Sercarz EE, Herzenberg LA, Fox CF (eds): "The Immune System. II. Regulatory Genetics." New York: Academic Press (In press).
31. Boyse EA, Cantor H: In "First International Conference on the Molecular Basis of Cell-Cell Interaction" (San Diego). New York: Academic Press (In press).
32. Huber B, Cantor H, Shen FW, Boyse EA: J Exp Med 144:1128, 1976.
33. Gally J: In Sela M (ed): "The Antigens." New York: Academic Press, 1973, vol 1, p 162.
34. Gally J, Edelman GM: Annu Rev Genet 6:1, 1972.
35. Edelman GM, Cunningham BA, Gall W, Gottlieb P, Rutishauser U, Waxdal M: Proc Natl Acad Sci USA 63:78, 1969.
36. Kehoe JM, Fougereau M: Nature (London) 224:1212, 1969
37. Mage R, Lieberman R, Potter M, Terry WD: In Sela M (ed): "The Antigens." New York: Academic Press 1973, vol 1, p 300.
38. Hood L, Kronenberg M, Early P, Johnson N: In Sercarz EE, Herzenberg LA, Fox CF (eds): "The Immune System. II. Regulatory Genetics." New York: Academic Press (In press).
39. Hood L, Loh E, Hubert J, Barstad P, Eaton B, Early P, Fuhrman J, Johnson N, Kronenberg M, Schilling J: Cold Spring Harbor Symp Quant Biol 41:817, 1976.
40. Hood L: Fed Proc Fed Am Soc Exp Biol 31:177, 1972.
41. Hozumi N, Tonegawa S: Proc Natl Acad Sci USA 73:3632, 1976.
42. Tonegawa S, Hozumi N, Matthyssens G, Schuller R: Cold Spring Harbor Symp Quant Biol 41:877, 1976.
43. Tonegawa S, Brack C, Hirama M, Hozumi N, Matthyssens G: Cold Spring Harbor Symp Quant Biol (In press).
44. Dreyer WJ, Gray W, Hood L: Cold Spring Harbor Symp Quant Biol 32:353, 1967.
45. Edelman GM: Cold Spring Harbor Symp Quant Biol 41:891, 1976.
46. Klinman NR, Press JL: J Exp Med 141:1133, 1975.
47. Sigal NH, Gearhart PJ, Press JL, Klinman NR: Nature (London) 259:51, 1976.
48. Lydyard PN, Grossi CE, Cooper MD: J Exp Med 144:79, 1976.
49. Amzel L, Poljak R, Saul F, Varga J, Richards F: Proc Natl Acad Sci USA 71:1427, 1974.
50. Padlan EA, Segal DM, Cohen GH, Danes DR: In Sercarz E, Williamson A, Fox CF (eds): "The Immune System: Genes, Receptors, Signals." New York: Academic Press, 1974, p 7.
51. Edmunson AB, Ely KR, Girling RL, Abola EF, Schiffer M, Westholm FA, Fausch ND, Deutsch HF: Biochemistry 13:3816, 1974.
52. Cohn M, Blomberg B, Geckeler W, Raschke W, Riblet R, Weigert M: In Sercarz E, Williamson A, Fox CF (eds): "The Immune System: Genes, Receptors, Signals." New York: Academic Press, 1974, p 89.
53. Tonegawa S: Proc Natl Acad Sci USA 73:203, 1976.
54. Bantock CR: J Embryol Exp Morph 24:257, 1970.
55. Goldstein P: J Morphol 152:141, 1977.
56. Beermann S: Chromosoma (Berlin) 60:297, 1977.
57. McClintock B: Brookhaven Symp Biol 18:162, 1965.
58. Nevers P, Saedler H: Nature (London) 268:109, 1977.
59. Fincham JRS, Harrison BJ: Heredity 22:211, 1967.
60. Green NM: Genetics (Suppl) 73:187, 1974.
61. Starlinger P: Crit Rev Microbiol (In press).
62. Cohen SN: Nature (London) 263:731, 1976.
63. Zieg J, Silverman M, Hilmen M, Siman M: Science 196:170, 1977.
64. Bukhari AJ (ed): "DNA insertion elements, plasmids and episomes." Cold Spring Harbor, New York: Cold Spring Harbor Laboratories (In press).

65. Holliday R, Pugh JE: Science 187:226, 1975.
66. Sager R, Kitchin R: Science 189:426, 1975.
67. Egel R: Nature (London) 266:172, 1977.
68. Gurdon JB, Laskey RA: J Embryol Exp Morphol 24:227, 1970.
69. Briggs R, King TF: Proc Natl Acad Sci USA 38:455, 1972.
70. Gurdon JB, Uehlinger V: Nature 210:1240, 1966.
71. Gurdon JB, Laskey RA, Reeves OR: J Embryol Exp Morphol 34:93, 1975.
72. Wabl MR, Brun RB, DuPasquier L: Science 190:1310, 1975.
73. Kobel HR, Brun RB, Rischberg M: J Embryol Exp Morphol 29:539, 1973.
74. Hood L, Campbell JH, Elgin SCR: Annu Rev Genet 9:305, 1975.
75. Birnstiel ML, Chipchase M, Spiers J: Prog Nucleic Acid Res Mol Biol 11:351, 1971.
76. Clarkson SG, Birnstiel ML: Cold Spring Harbor Symp Quant Biol 38:451, 1973.
77. Schaffner W, Gross K, Telford J, Birnstiel M: Cell 8:471, 1976.
78. Kitchen H, Bayer S (eds), "Hemoglobin: Comparative Molecular Biology Models for the Study of Disease." Ann NY Acad Sci 24, 1974.
79. Gall JG, Cohen EH, Atherton DD: Cold Spring Harbor Symp Quant Biol 38:417, 1973.
80. Boyse EA, Stockert E, Olds LJ: In Rose NR, Milgrom F (eds): "International Convocation on Immunology." Basel, Switzerland: S Karger, 1968, p 353.
81. Flaherty L: Immunogenetics 3:533, 1976.
82. McDevitt HO, Delovitch TL, Press JL, Murphy DB: Transplant Rev 30:197, 1976.
83. Ohno S: "Evolution by Gene Duplication," New York: Springer-Verlag, 1970.
84. "Origins of Lymphocyte Diversity." Cold Spring Harbor Symp Quant Biol 41:627, 1977.
85. Janeway CA, Wigzell H, Binz H: Scand J Immunol 5:993, 1976.
86. Rajewsky K: In Sercarz EE, Herzenberg LA, Fox CF (eds): "The Immune System. II. Regulatory Genetics." New York: Academic Press (In press).
87. McDevitt HO: Fed Proc Fed Am Soc Exp Biol 35:2168, 1976.
88. Schreffler DC: Transplant Rev 32:140, 1976.
89. Michaelson J, Flaherty L, Vitetta E, Poulik MD: J Exp Med 145:1066, 1977.
90. Hansen TH, Cullen SE, Sachs DH: J Exp Med 145:438, 1977.
91. Klein J: "Biology of the Mouse Histocompatibility-2 Complex." New York: Springer-Verlag, 1975, p 241.
92. Garrido F, Festenstein H, Schirrmacher V: Nature (London) 261:705, 1976.
93. Silver J, Hood L: In Eisen HN, Reisfeld RA (eds): "Contemporary Topics in Immunobiology." New York: Plenum Press, 1976, vol 5, p 35.
94. Smithies O, Poulik MD: Proc Natl Acad Sci USA 69:2914, 1972.
95. Cunningham BA, Wang JL, Berggård I, Peterson PA: Biochemistry 12:4811, 1973.
96. Jacob F: Immunol Rev 33:3, 1977.
97. Silver J, Cecka JM, McMillan M, Hood L: Cold Spring Harbor Symp Quant Biol 41:369, 1976.
98. Bolotin C, Morris S, Tack B, Prahl J: Biochemistry 16:2008, 1977.
99. Moscona AA, Housman RE: In Lash J, Burger M (eds): New York: Raven Press (In press).
100. Garcia-Bellido A: In Ursprung H, Nöthiger R (eds): Berlin: Springer-Verlag, 1972, vol 5, p 59.
101. Sperry RW: In DeHaan, Urspring H (eds): "Organogenesis." New York: Holt, Reinhart, and Winston, 1965, p 161.
102. Galfre A, Howe SC, Milstein C, Butcher AW, Howard JC: Nature (London) 266:550, 1976.
103. Basombrio MA: Cancer Res 30:2458, 1970.
104. Kripke ML: J Natl Cancer Inst 53:1333, 1974.
105. Berget SM, Berk AJ, Harrison T, Sharp P: Cold Spring Harbor Symp Quant Biol (In press).
106. Broker TR, Chew LT, Dunn A, Gelinas RE, Hassell J, Kleosig DF, Lewis J, Roberts RJ: Cold Spring Harbor Symp Quant Biol (In press).