58 STADLER SYMP. Vol. 14 (1982) University of Missouri, Columbia

STADLER SYMP. Vol. 14 (1982) University of Missouri, Columbia 59

THE ONTOGENETIC AND EVOLUTIONARY ORIGINS OF ANTIBODY DIVERSITY

(Antibody diversity, ontogeny, evolution, transposable elements)

HENRY HUANG AND LEE HOOD

Department of Microbiology & Immunology, Washington University School of Medicine, St. Louis, MO. 63110 and Div. of Biology, California Inst. of Technology, Pasadena, CA 91125

SUMMARY

Three ontogenetic origins of antibody diversity have been identified. First, there is a large number of germline gene segments. Second, multiple combinatorial strategies are used to amplify antibody diversity. These include combinatorial association of heavy and light chains, and heavy chain class switching. Third, somatic alterations include junctional diversity and somatic hypermutation.

Finally, we present a new model for the evolutionary origin of V gene formation. This model postulates that the heavy chain family is the first to appear, from which are derived the light chain families. The model readity explains the location of the conserved sequences believed to be used as recognition sites in V gene formation.

INTRODUCTION

The vertebrate organism is able to respond to immunization with any arbitrarily chosen antigen by the synthesis of large amounts of antibodies specific for the antigen. This ability is based upon the presence of a large diversity of cells, each capable of synthesizing a single kind of antibody. The antigen stimulates only those cells that synthesize, and bear on their surfaces, antibodies that tightly bind the antigen. Clonal selection results in specific cells that are stimulated to proliferate and to synthesize increased amounts of their antibodies, these then constitute the specific response to the antigen. We will briefly review the properties of antibody polypeptides and genes. We will then discuss the ontogenetic origins of

the enormous diversity of antibodies that an animal can make, and present a new model for the evolutionary origin of the DNA rearrangements that are central to the generation of antibody diversity.

ANTIBODY MOLECULES AND GENES

The typical antibody molecule consists of two polypeptide chains, heavy (H) chain, and a smaller light (L) chain. Each chain is divided into two structurally and functionally distinct portions. The sequence of the amino-terminal portion is extremely variable and is therefore called the variable (V) region. The antigen-binding site is formed by the V regions of both H and L chains. The carboxyl-terminal portion of each chain is called the constant (C) region, since only relatively few types of C regions are found in an animal. The C region of heavy chains determines the class of the antibody, e.g., IgM antibodies have μ heavy chains and IgG have γ heavy chains. The C regions serve as structural support for the V regions, as well as to provide important effector functions. For example, when antigen is bound to the V regions, the C regions are triggered to activate processes (e.g., complement activation) that destroy or eliminate the foreign antigen.

The antibody polypeptides are encoded by three multigene families: k and λ families for the two types of light chains, and the H family. Each polypeptide is encoded by several separate gene segments. L chain genes consist of variable (V_L), joining (J_L) and constant (C_L) gene segments, H chain genes consist of V_H, diversity (D), J_H and C_H gene segments. The V_L and J_L gene segments, and the V_H, D and J_H gene segments are separate in the germline. They are joined together by DNA rearrangements to form the complete V_L and V_H genes during the differentiation of antibody-producing cells (DREYER & BENNETT 1966; BRACK et al. 1978; SEIDMAN & LEDER 1978; EARLY et al. 1980a). The DNA rearrangement that results in the construction of the V genes is called V gene formation.

Two conserved sequences, 7 and 10 base pairs (bp) in length, are located immediately 3' to all germline V gene segments, 5' to all germline J gene segments, and on both 5' and 3' sides of all germline D gene segments examined (Table 1). The location and the conservation of these sequences in all three families and in all vertebrates examined to date suggest that they are recognized by the enzymes that catalyze V gene formation (MAX et al. 1979; SAKANO et al. 1979; EARLY et al. 1980a). The conserved sequences are separated from each other either by 11 bp, approximately 1 DNA helical turn (J_λ, V_k and D gene segments), or by 20-23 bp, approximately 2 DNA helical turns (V_λ, J_k, V_H and J_H gene segments). In every case, correct V gene formation requires that gene segments flanked by "1 turn" recognition sites to be joined to gene segments flanked by "2 turn" recognition sites (Table 1). This is the "1-turn to 2-turn" joining rule (EARLY et al. 1980a; SAKANO et al. 1980).

Table 1

Recognition site for V gene formation
The I-turn to 2-turn joining rule

	location relative to gene segment	type (1 or 2 turn)
V_λ	3'	2
J_λ	5'	1
V_K	3'	1
J_K	5'	2
V_H	3'	2
D	5' and 3'	1
J_H	5'	2

Successful V gene formation requires the joining of V_L to J_L, and the joining of V_H to D, and D to J_H.

A second kind of DNA rearrangement is involved in the expression of antibody genes. Since all J_H gene segments are 5' to the C_μ gene, V gene formation in the H family places the V gene 5' to the C_μ gene (EARLY et al. 1980a; NEWELL et al. 1980; SAKANO et al. 1980). At this stage the cell synthesizes μ chains, and therefore IgM molecules (RAFF 1976; COOPER et al. 1976). Later in development the cell can make μ chains as well as heavy chains of a different class. At least initially this is due to alternate RNA processing. Primary transcripts are known to be processed to mRNA's encoding either membrane-bound or secreted IgM molecules (EARLY et al. 1980b). Long primary transcripts that include the sequence of the V_H gene, the C_μ gene and other C_H genes 3' to the C_μ gene may be processed to give mRNA's that encode different classes of H chains (MOORE et al. 1981). Finally, the cell can rearrange its DNA such that any C_H gene can be placed 3' to the V_H gene (HONJO & KATAOKA 1978; DAVIS et al. 1980; SAKANO et al. 1980). This DNA rearrangement is called class switching.

ONTOGENETIC GENERATION OF ANTIBODY DIVERSITY

Three sources of antibody diversity are known.

First, there are perhaps 100-300 germline V_H and V_k gene segments, 10 or more D gene segments, and 4 germline J_H and 4 germline J_k gene segments in the mouse (WEIGERT et al. 1978; MAX et al. 1979; NEWELL et al. 1980; SAKANO et al. 1980; KUROSAWA et al. 1981). Thus one source of antibody diversity is in having multiple germline genes.

Second, the immune system has evolved to employ powerful

combinatorial strategies for generating antibody diversity. Gene segments are joined in a combinatorial fashion to amplify V gene diversity. For example, if any V_k can join with any J_k gene segment, and any V_H can join with any D, and any D with any J_H gene segment, then 800 V_k (200 V_k x 4 J_k) and 8000 V_H (200 V_H x 10 D x 4 J_H) genes can be made by combinatorial joining. Since the antigen-binding site of antibody molecules is made up of both V_H and V_L sequences, combinatorial association of H and L chains greatly amplifies the number of different antigen-binding sites that can be made. For example if every k chain can associate with every H chain, then 800 V_k x 8000 V_H = 6.4 x 10^6 antigen-binding sites can be made. In addition, each V_H sequence can associate with each of a number of C_H sequences by alternate RNA processing and by class switching. Therefore each of the many possible antigen-binding specificities can be associated with a number of different sets of effector functions.

The third source of antibody diversity includes two types of somatic alteration of V gene sequences. The first type is junctional diversity (MAX et al. 1979; SAKANO et al. 1979; WEIGERT et al. 1980). V gene formation is imprecise, such that the joining of gene segments can occur at a number of positions at the boundaries of the gene segments. The imprecision results in variability of the sequence and the number of codons that are included at the junctions. Although imprecise joining can occasionally result in frame shifts and the creation of termination codons, the increased diversity at the V_L-J_L, V_H-D, and D-J_H junctions is functionally significant, since these junctions encode sequences in the antigen-binding site (AMZEL & POLJAK 1979). The second type of somatic alteration of V gene sequences is somatic hyper-mutation (WEIGERT & RIBLET 1976; BOTHWELL et al. 1981; CREWS et al. 1981; KIM et al. 1981; BOTHWELL et al. 1982). The mutations are localized in the V_H and V_L coding sequences and their immediate flanking sequences. In the coding sequences both silent changes and changes that result in amino acid replacement are found. They are very extensive (at least 44 substitutions in and around the M167 V_H gene; KIM et al. 1981). Finally, somatic hypermutation correlates well with class switching: V_H and V_L sequences of IgM are predominantly germline, most of the mutated V sequences are found in antibodies of other classes (CREWS et al. 1981; GEARHART et al. 1981). It is not clear if somatic hypermutation is a functional source of antibody diversity. We have speculated that somatic hypermutation allows the immune system to fine-tune the antibody response (CREWS et al. 1981). Somatic hypermutation might also have a predictive function, by facilitating the specific response to subsequent invasion by foreign organisms that are genetically, and therefore structurally and antigenically related to, but not identical to the invading organism that originally stimulated the immune response (HUANG & HOOD 1982).

THE EVOLUTIONARY ORIGIN OF ANTIBODY DIVERSITY

Some of the mechanisms for generating antibody diversity are also found in other systems. Higher eukaryotes in general have multiple, linked copies of most genes. Indeed, non-multigenic systems constitute exceptions rather than the rule in higher eukaryotes. Thus the antibody families are not exceptional in having multiple germline genes. Similarly, the combinatorial association of dissimilar polypeptide subunits to generate structural and functional diversity can be found in many other systems, the most obvious examples are hemoglobin and lactate dehydrogenase. The strategy of alternate RNA processing is also not uncommon (e.g., α Amylase, YOUNG et al. 1981; calcitonin, AMARA et al. 1982).

The remaining mechanisms for generating antibody diversity, i.e., combinatorial joining, junctional diversity, class switching and somatic hypermutation, are so far unique to the antibody system (the antigenic variation of trypanosomes might involve both combinatorial DNA rearrangements as well as "somatic" mutation, see BORST & CROSS 1982 for review). All of these mechanisms are closely coupled to developmentally programmed DNA rearrangements. Therefore much of the problem of the evolutionary origin of antibody diversity centers on the problem of the origin of DNA rearrangements.

We will not speculate on the evolutionary origins of class switching, since the sequence requirements for class switching are not clear (see HOOD et al. 1981 and MARCU et al. 1982 for two different models for class switching). The sequence requirements and properties of V gene formation are clearer, and we will present two models for its origin.

DNA rearrangements occur in a number of organisms (see HOOD et al. 1977, for an early catalogue of such systems). The most familiar examples are the large number of different transposable elements found in a number of organisms. This led SAKANO et al. (1979) to postulate that the insertion of a transposable element into a primordial V gene gave rise to the primordial λ family. The insertion split the primordial V gene into two portions corresponding to the V_{λ} and J_{λ} gene segments. Expression of the split gene requires the developmentally programmed excision of the transposable element coupled with the joining of the gene segments. This is an attractive model, since bacterial and eukaryotic transposable elements can insert into and inactivate or modify the activity of a gene; subsequent excision can lead to the restoration of gene activity (CALOS & MILLER 1980). The transposed segment need not contain its own transposase gene since the gene-regulating activity and transposition of the segment can be regulated in trans by linked or unlinked genetic elements (e.g., the Spm system in maize, McCLINTOCK 1965). Furthermore, the movement of some transposable elements is correlated with normal developmental events (e.g., the _crown_ and _flow_ alleles in the _En_ system of maize,

PETERSON 1966). Other developmentally regulated DNA rearrangements are known. These include chromosomal diminution, the amplification of ribosomal genes, and the production of macronuclear DNA in flagellates (see HOOD et al. 1977). Thus the model invokes processes and features of transposable element-mediated DNA rearrangement that are known to occur. However, the model cannot be easily generalized to all the antibody families. To account for the different positions of the 1-turn and 2-turn recognition site in the k and λ families (Table 1), the model requires the insertion of the same or a related transposable element into the k precursor, in an orientation opposite to that in the λ precursor. Alternately, the transposable element in λ must invert itself to give rise to the k family. At least one more insertion is required to generate the H family configuration. We find this unattractive for the following reasons. First, the V gene segments of the 3 families are clearly homologous, as are the J gene segments. Thus the model requires that the insertion of the transposable element occurred repeatedly at about the same relative location in the primordial V gene of each family. Although transposable elements can show site specificity, e.g. the integration of lambda phage into the _E. coli_ genome, we feel that it is an unnecessary complication of the model (see below). Second, the model minimally requires two, and as many as four, insertion events: once into λ, followed by inversion to give k, and at least once more to give the H family. The successful insertion of a transposable element into a primordial V gene requires that its excision to be developmentally regulated, and that the excision be reasonably precise so as to restore proper gene activity. Each of these requirements is not too unlikely. However, the successful conjunction of both requirements must be quite rare. It is rather implausible for it to occur two or more times.

We propose a new model to overcome these objections. We propose that the H family is the first to appear: a composite transposable element, with a structure analogous to that of bacterial transposons, inserted into a primordial V_H gene (Figure 1). The transposon has two flanking insertion sequences (IS), labeled _a_ and _b_ in Figure 1. Each IS bears a 1-turn and a 2-turn recognition site at its ends. The two IS are each a transposable element, and they are in inverted orientation to each other, properties common in bacterial transposons. The two IS flank a central sequence, structurally analogous to the central antibiotic resistance genes of many bacterial transposons (CALOS & MILLER 1980). The central sequence is the ancestor of the D gene segment. The primordial H chromosome now has the structure V, IS_a_, D, IS_b_, J, as the comtemporary H family is believed to be organized. The light chain families are then derived from the H family by the excision, in the _germline_, of IS_a_ to give rise to the k family, and the excision of IS_b_ to give rise to the λ family. Note that after the creation of each family, the gene segments can be freely duplicated. Gene expression in the

families would involve the developmentally programmed, somatic excision of the appropriate IS to constitute V gene formation.

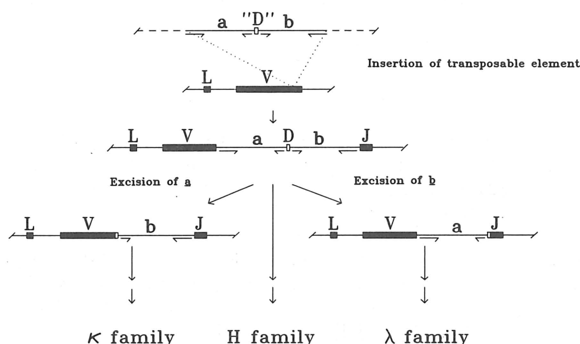

Figure 1. model for the origin of V gene formation. The insertion of a composite transposable element, ISa-D-ISb, created the primordial H family. Excision of ISa or ISb in the composite transposable element then gave rise to the primordial k and λ families respectively. L: leader sequence, V: V gene segment, D: D gene segment, J: J gene segment. Long arrow: 2-turn recognition sequence. Short arrow: 1-turn recognition sequence. For clarity, the C gene is not shown, but would be present 3' to the J gene segments. Note that V gene formation is the developmentally programmed excision of ISa or ISb in the L chain families, and both ISa and ISb in the H family.

This model therefore postulates a single, successful insertion event. The subsequent excision events postulated by the model are expected to occur readily, given the properties of contemporary transposable elements and given that excision can occur at least in the soma. Our model is a parsimonius explanation for the origin of the three antibody families, especially since it directly explains the positions of the 1-turn and 2-turn recognition sites in the three families (Table 1). The model invokes common structural and transposing properties of contemporary transposable elements, and requires only a single primordial event that is likely to be rare. One intriguing feature of the model is that it requires the excision (from some other location) and insertion (into the primordial V_H gene) of a transposable element flanked by 2-turn recognition sites on its extreme ends, thus predicting that the transposition of the composite transposable element need not obey the 1-turn to 2-turn rule, while excision of the component IS does obey the rule. If this is indeed true, then the transposition of a large chromosomal region, containing one or more V_H gene segment, the D gene segments, and one or more J_H gene segments might occur during evolution, to give rise to other gene families that use DNA rearrangements for gene expression (HOOD et al. 1977).

ACKNOWLEDGEMENTS

This work was supported by NIH grant AI 16913.

LITERATURE CITED

AMARA, S.G., V. JONES, M.G. ROSENFELD, E.S. ONG and R.M. EVANS 1982 Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature 298: 240-244.

AMZEL, L.M. and R. J. POLJAK 1979 Three dimensional structure of immunoglobulins. Ann. Rev. Biochem. 48: 961-967.

BORST, P. and G.A.M. CROSS 1982 Molecular basis for trypanosome antigenic variation. Cell 29: 291-303.

BOTHWELL, A.L.M., M. PASKIND, M. RETH, T. IMANISHI-KARI, K. RAJEWSKY and D. BALTIMORE 1981 Heavy chain variable region contribution to the NP^b family of antibodies: Somatic mutation evident in a γ2a variable region. Cell 24: 625-637.

BOTHWELL, A.L.M., M. PASKIND, M. RETH, T. IMANISHI-KARI, K. RAJEWSKY and D. BALTIMORE 1982 Somatic variations of murine immunoglobulin λ light chains. Nature 298: 380-382.

BRACK, C., M. HIRAMA, R. LENHARD-SCHULLER and S. TONEGAWA. 1978 A complete immunoglobulin gene is created by somatic recombination. Cell 15: 1-14.

CALOS, M.P. and J.H. MILLER 1980 Transposable elements. Cell 20: 579-595.

COOPER, M.D., J.F. KEARNEY, P.M. LYDYARD, C.E. GROSSI, and A.R. LAWTON. 1976 Studies of generation of B-cell diversity in mouse, man and chicken. Cold Spring Harbor Symp. Quant. Biol. 41: 139-145.

CREWS, S., J. GRIFFIN, H. HUANG, K. CALAME and L. HOOD 1981 A single V_H gene segment encodes the immune response to phosphorylcholine: Somatic mutation is correlated with the class of the antibody. Cell 25: 59-66.

DAVIS, M.M., K. CALAME, P.W. EARLY, D.L. LIVANT, R. JOHO, I.L. WEISSMAN and L. HOOD 1980 An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature 283: 733-738.

DREYER, W.J. and J.C. BENNETT 1965 The molecular basis of antibody formation: a paradox. Proc. Natl. Acad. Sci. USA 54: 864-868.

EARLY, P., H. HUANG, M. DAVIS, K. CALAME and L. HOOD 1980a An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. Cell 19: 981-992.

EARLY, P., J. ROGERS, M. DAVIS, K. CALAME, M. BOND, R. WALL and L. HOOD 1980b Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell 20: 313-319.

GEARHART, P.J., N.D. JOHNSON, R. DOUGLAS and L. HOOD 1981 IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. Nature 291: 29-34.

HONJO, T. and T. KATAOKA 1978 Organization of immunoglobulin heavy chain genes and allelic deletion model. Proc. Natl. Acad. Sci. USA 75: 2140-2144.

HOOD, L., H.V. HUANG and W.J. DREYER 1977 The area-code hypothesis: The immune system provides clues to understanding the genetic and molecular basis of all recognition during development. J. Supramolec. Struct. 7: 531-559.

HOOD, L., M. DAVIS, P. EARLY, K. CALAME, S. KIM, S. CREWS and H. HUANG 1981 Two types of DNA rearrangements in immunoglobulin genes. Cold Spring Harbor Symp. Quant. Biol. 45: 887-898.

HUANG, H. and L. HOOD 1982 The expression of antibody genes. Pp.169-172. Comparative Research on Leukemia and Related Diseases, (Yohn, D.S. and J.R. Blakeslee, Eds) Elsevier North Holland Inc., New York.

KIM S., M. DAVIS, E. SINN, P. PATTEN and L. HOOD 1981 Antibody diversity: Somatic hypermutation of rearranged V_H genes. Cell 27: 573-581.

KUROSAWA, Y., H. VON BOEHMER, W. HAAS, H. SAKANO, A. TRAUNEKER and S. TONEGAWA 1981 Identification of D segments of immunoglobulin heavy chain genes and their rearrangement in T lymphocytes. Nature (London) 290: 565-570.

MARCU, B.M., R.B. LANG, L.W. STANTON and L.J. HARRIS 1982 A model for the molecular requirements of immunoglobulin heavy chain class switching. Nature (London) 298: 87-89.

MAX, E.E., J.G. SEIDMAN and P. LEDER 1979 Sequence of five potential recombination sites encoded close to an immunoglobulin k constant region gene. Proc. Natl. Acad. Sci. USA. 76: 3450-3454.

McCLINTOCK, B. 1965 The control of gene action in maize. Brookhaven Symp. Biol. 18: 162-184.

MOORE, K.W., J. ROGERS, T. HUNKAPILLER, P. EARLY, C. NOTTENBURG, I. WEISSMAN, H. BAZIN, R. WALL and L.E. HOOD (1981) Expression of IgD may use both DNA rearrangement and RNA splicing mechanisms. Proc. Natl. Acad. Sci. USA 78: 1800-1804.

NEWELL, N., J.E. RICHARDS, P.W. TUCKER and F.R. BLATTNER 1980 J genes for heavy chain immunoglobulins of mouse. Science 209: 1128-1132.

PETERSON, P.A. 1966 Phase variation of regulatory elements in maize. Genetics 54: 249-266.

RAFF, M.C. 1976 Development and modulation of B lymphocytes: studies on newly formed B cells and their putative precursors in the hemopoietic tissues of mice. Cold Spring Harbor Symp. Quant. Biol. 41: 159-162.

SAKANO, H., K. HUPPI, G. HEINRICH and S. TONEGAWA 1979 Sequences at the somatic recombination sites of immunoglobulin light chain genes. Nature 280: 288-294.

SAKANO, H., R. MAKI, Y. KUROSAWA, W. ROEDER and S. TONEGAWA 1980 Two types of somatic recombination are necessary for generation of complete immunoglobulin heavy-chain genes. Nature 286: 676-683.

SEIDMAN, J.G. and P. LEDER 1978 The arrangement and rearrangement of antibody genes. Nature (London) 276: 790-795.

WEIGERT, M. and R. RIBLET 1976 Genetic control of antibody variable regions. Cold Spring Harbor Symp. Quant. Biol. 41: 837-846

WEIGERT, M., R. PERRY, D. KELLEY, T. HUNKAPILLER, J. SCHILLING and L. HOOD. 1980 The joining of V and J gene segments creates antibody diversity. Nature 283: 497-499.

YOUNG, R.A., O. HAGENBUCHLE and U. SCHIBLER 1981 A single mouse α-amylase gene specifies two different tissue specific mRNA. Cell 23: 451-458.