Identifying Antigens and Antibodies in Serology

GEORGE MARKOWSKY
Computer Sciences Department, IBM T. J. Watson Research Center, Yorktown Heights, New York 10598

AND

ANDREW WOHLGEMUTH*
Department of Mathematics, University of Maine at Orono, Orono, Maine 04469

Received 17 August 1982; revised 28 March 1983

ABSTRACT

The problem of whether it is theoretically possible to determine antigens and antibodies in a cross-reacting serological system is analyzed. Conditions are listed which will insure that the serological operations of absorption, elution, blocking, and counterblocking uniquely determine antigens and antibodies. There is no restriction on cross-reactivity. The protocol for counterblocking was given by Denniston and has not been developed by serologists. Denniston has shown that counterblocking can be replaced by combinations of the other operations in a system where there can be no cross-reactivity. This paper indicates the potential usefulness of such a technique.

1. INTRODUCTION

Reference [17] contains a detailed justification for the point of view taken in this paper. For the reader's convenience we have reproduced below the first few paragraphs of [17], but the interested or critical reader should consult the entire reference.

Immunogenetics has adopted the obvious and perhaps necessary practice of denoting genes or antigens by letter symbols. Thus, for example, A and B are ABO blood group antigens, D is an Rh blood group antigen, H-2^b is an allele at the mouse H-2 complex, and HLA-A is a human lymphocyte antigen. This symbolic representation of genes or gene products has certain consequences that are not obvious at all but are indeed logically necessary. A source of great difficulty is that the letters representing, say, antigens do not

*Work supported by Grant Number 5 R01 CA 20105-06 awarded by the National Cancer Institute, NIH.

52 Vanderbilt Ave., New York, NY 10017 0025-5564/83/$03.00

stand for correspondingly well-defined discrete biological entities but complicated residues that are (1) determined by inherited factors and (2) recognized by immunological mechanisms. There is nothing essentially discrete about either (1) or (2) above. Immunological recognition is certainly not a discrete phenomenon but depends on continuously varying reaction strengths. In spite of this difficulty it may fairly be said that immunogenetics has been successful only inasmuch as it has been able to deal with discrete conceptual entities that *can* be denoted by letter symbols and subjected to the rules of inheritance.

A problem immediately presents itself. If antigens X, Y, and Z have the property that Y always occurs with either X or Z but never alone, is this a genetic law: the gene for Y can only be expressed in the presence of X or Z? Or is it the result of our choice of letters? XY might better be called P, and YZ called Q. In the latter case the antibodies that were thought to *define* Y would now be said to cross-react—to react with P and Q. Immunologists have almost universally adopted a notational system that precludes cross-reactivity. If, in practice, a reagent of monoclonal antibodies reacts with antigens called P and Q, this fact will be explained by attributing to P and Q some common antigenic factor, say Y. Thus P is Y and more, i.e. XY; and similarly Q is YZ. In this view antibodies are considered as simple (recognizing only a single antigenic factor denoted by a single letter), and antigens are regarded as complex (denoted by the totality of all letters necessary to account for reactions observed with the different simple antibodies). This rule for making letter assignments has been called the simple-complex code by Hirschfeld [3], who has also described the complex-simple and complex-complex codes.

Hirschfeld [7] has carefully documented the experimental evidence for cross-reacting antibodies and therefore the inappropriateness of always using the simple-complex code. Mobraaten [11, p. 28] has stated that HLA serology developed according to a complex-simple interpretation. In the previous paragraph it was seen how different choices of letter symbols could lead to either a simple immunological situation with complicated genetic laws or a more complex immunological situation with simple genetic laws. The precise connection between these two extremes has been studied by Hirschfeld [5, 6]. Avoiding a notational system that allows for cross-reacting antibodies can lead to apparent genetic laws, such as inexplicable linkage relations and the cis-trans effect, which completely disappear when cross-reacting notation is allowed. Specific real-world examples of this are given in [17].

One notation in serology involves the idea of "specificity." In [1] Denniston presents operations on the sets of serological specificities associated with reagents and tested cells. Thus he denotes by C a set of specificities associated with cells, by R a set of specificities associated with some reagent, and (from Table 2 of [1]) the following sets obtained from C and R

corresponding to the serological operations:

Serological operation	Set operation	Result
Absorption	$R - C$	Reagent
Elution	$R \cap C$	Reagent
Blocking	$C - R$	Cell
Counterblocking	$R \cap C$	Cell

Thus, for example, absorption produces a new reagent with specificities $R - C$; counterblocking produces cells with a new set of specificities $R \cap C$.

Denniston states [1, p. 106] that "A more extensive discussion, including a treatment of the problem of 'cross-reactions,' is in preparation." The papers of Sheehy and Denniston [13] and Denniston and Sheehy [2] address the problem of assigning specificities to stimulating and responding cells in MLC (mixed lymphocyte culture) testing in a way that remains valid even in the presence of cross-reactivity. But cross-reactivity does indeed affect the set operations modeling the four serological operations given above, and this work has not appeared. It is the purpose of this paper to extend the operations to a model that allows for cross-reactivity and to analyze the model to determine how antigens and antibodies can be identified.

In [17] we have given an example that shows that if there is cross-reactivity in a system incorporating absorption data, a distinction must be made between recognized factors (antigens) and recognition factors (antibodies). The notion of specificity can be confusing in a system with cross-reactivity and in our opinion should not be used. In particular this example shows that specificities subtract (giving $R - C$ as above) only if by specificity we mean recognition factor. Thus there is a need to extend the operations to a model allowing for cross-reactivity.

Also in [17] it was shown that a threefold Boolean factorization of a reaction matrix M is the important form of a model of an immunogenetic system. We will denote this factorization here by $M = W \times Z \times E$. M gives observed reactions between cells and reagents (sera). W gives the antigen content of cells, and E gives the antibody content of reagents. Z defines antigens in terms of the antibodies recognizing them, and vice versa [14]. It is shown in [15] that the description of reactions in terms of specificities is subsumed in the matrix factorization.

It is important to realize here that the model operates on the level of notation or symbolization and not data. M is a zero-one matrix, and it is at this point that a reasonable fit must be made to real world data. See for example the technique in [14, p. 490]. Notice that every attempt at symbolizing immunogenetic phenomena implicitly or explicitly requires such a fit.

Describing the reactions in M in terms of a single set of factors—such as specificities or the letters in Hirschfeld's complex-complex code [4]—is

equivalent to a twofold factorization of M. The number of such factorizations is very large, and even determining whether a factorization exists involving factors of a given size is a computationally intractable (NP-complete) problem [12]. It is not feasible therefore to generate all solutions to the problem and choose from among these the solutions satisfying additional pertinent conditions. If effector-cell combinations or reagent combinations (e.g. 1 anti-2×3) are known, it is possible to say a good deal about the factorization of M into $G\times E$ (where $G=W\times Z$). The procedure for determining this factorization is given in [8], [16], and [17]. The problem of determining antigenic factors by factoring G remains and is again complex (although G will in general be smaller than M) if more information is not available.

In this paper we will give conditions that insure that W, Z, and E are uniquely determined by the four serological operations above.

2. NOTATION

Let C be a set of cells and R a set of reagents involved in some system. Each cell is considered as a set of antigens, and each reagent as a set of antibodies. For some $i\in C$ let iZ denote the set of all antibodies recognizing some antigen in i. If $r\in R$, let Zr denote the set of all antigens recognized by some antibody in r. Then Z defines each antibody by specifying the complete set of antigens with which it reacts, and vice versa. Z viewed as a matrix has a one in row k, column l if and only if antigen k is recognized by antibody l. If there is no cross-reactivity in the system, then z is a one-to-one correspondence between antigens and antibodies.

Analogously to table 2 of [1], we define the four (serological) functions:

Absorption	$\text{AB}(i,r)=r-iZ$
Elution	$\text{EL}(i,r)=r\cap iZ$
Blocking	$\text{BL}(i,r)=i-Zr$
Counterblocking	$\text{CB}(i,r)=i\cap Zr$

Notice that AB and EL are functions into the set of antibodies and that BL and CB are functions into the set of antigens. This is in contrast to the operations on a single set of factors (specificities) in [1]. Notice that our functions reduce to Denniston's operations when there is no cross-reactivity, i.e. when Z is the identity relation.

For the purposes of this paper we take an antibody to be completely described by its reaction range with antigens: since we consider data to be in the form of positive or negative test reactions, two antibodies that react in precisely the same way with each antigen are considered equivalent. Similarly an antigen is completely determined by its reaction range with antibodies.

This is formalized by our first basic principle, which is nothing more than a semantic convention:

PRINCIPLE 1

No two distinct antibodies react with exactly the same set of antigens, and no two distinct antigens react with exactly the same set of antibodies.

The next principle is again dictated by what an antigen and antibody are.

PRINCIPLE 2

An antibody must react with at least one antigen, and an antigen must react with at least one antibody.

Notice that by our definitions, if some antibody reacts with a combination of gene products but not with either one of the products, then this combination must be considered an antigen distinct from (but of course related to) the products themselves. It follows that an antibody reacts with a cell if and only if the antibody reacts with some antigen in the cell. With this understanding we can give our third principle, the first part of which states that certain cells can be excluded from consideration in the data with no loss of information.

PRINCIPLE 3

No cell reacts with every antibody, and no antibody reacts with every cell.

In situations where reagents are made by taking donor and host combinations, antibodies would not generally be made that reacted with the cells of the host. Any such antibodies can be easily recognized and excluded. In situations where less is known about the reagents, reagents that gave uniform positive reactions with all cells would have to be excluded from the study as being uninformative.

3. THE IMMUNOLOGICAL SEPARATION CONDITION

It is clear that testing with one or two cells can never reveal a complete system involving many antigens and antibodies. What is needed is a condition that insures there are enough cells from different individuals tested to reveal all the antigens and antibodies in the system.

SEPARATION CONDITION

For every pair of distinct antibodies, there is a cell in C that reacts with one of the pair but not the other.

In Section 5 we shall see that the four serological operations will reveal antigens and antibodies (with no assumptions about cross-reactivity) in any system satisfying the basic Principles 1–3 and Separation Condition. In the

next section we will give genetic conditions that are sufficient to satisfy these requirements.

4. THE BASIC GENETIC CONDITION

The predominant pattern in well-known immunogenetic systems is one of codominant alleles. In fact, examples have been given in [14] that show that a premature assumption that reagents or antibody species are monospecific can lead to a notation that would *incorrectly* suggest a system that violated the rule of codominant alleles. By a *simple* genetic system we mean a system where genes are codominant and each antigen is coded for by a single gene. This puts genes and antigens in a one-to-one correspondence. The scheme of the next section, which will uncover antigens and antibodies and their relation, will therefore also uncover the genetics of the system. A simple system is a special case of a first order system as defined in [14].

By a *complete* system we mean a system of haplotypes (or a system including all homozygous individuals) where all possible recombinants are present. This is the genetic condition that insures that the separation condition holds, i.e., that there are enough individuals (cells) being tested to reveal the system.

GENETIC CONDITION

The system is simple and complete.

We can now show that the genetic condition implies the immunological separation condition as follows. Suppose the simple genetic system has three loci and three alleles at each locus. Suppose a, b, c are the alleles at the first locus; d, e, f are the alleles at the second locus; g, h, i are the alleles at the third locus. (Note that the argument which follows works for any number of loci and any number of alleles at each locus. However, to avoid cumbersome notation the argument has been presented in this very concrete form.) Given two antibodies α and β, Principle 1 implies that they react with different sets of antigens. With no loss of generality, it may be assumed that α reacts with the antigen coded for by d, but β doesn't. By Principle 3 at least one of a, b, or c must code for an antigen which does not react with β, since each cell must contain an a, b, or c, and β cannot react with every cell. Again, without loss of generality it may be assumed that β does not react with the antigen produced by a. The third locus may be analyzed similarly, and it may be assumed that β does not react with the antigen produced by g. Observe now that the haplotype adg will produce at least one antigen reacting with α but no antigens reacting with β. Thus the separation condition has been satisfied.

5. RECOVERING ANTIGENS AND ANTIBODIES

The four operations absorption (AB), elution (EL), blocking (BL) and counterblocking (CB) will allow the total elucidation of the antigen-antibody

situation whenever the three Principles and Separation Condition hold. This process can be broken into the following two stages.

(1) *Use absorption and elution to reduce reagents until no new reagents are created. The final reduced reagents will contain equivalent antibodies.*

Observe that the reagents obtained at this stage should theoretically agree with reagents of monoclonal antibodies. There is a danger (elaborated below) in using monoclonal antibodies without going through procedure (1), since it is this procedure that theoretically guarantees enough diversity in (cross-reacting) antibodies to identify the antigens in the system. Thus cloning ought to take place after the procedure (1). The analysis of procedure (2), is dual to (1), so that the analysis we will offer for (1) can be applied for (2).

(2) *Use the antibodies produced in* (1) *together with blocking and counter-blocking to uncover the antigens.*

In stage (1), whenever a cell, i, reacts with a reagent, r, it is possible to construct two new reagents: $\text{AB}(i, r)$ and $\text{EL}(i, r)$. Here $\text{AB}(i, r)$ is defined to be the set of all antibodies in reagent r except those reacting with an antigen in cell i, and $\text{EL}(i, r)$ is defined to be the set of all antibodies in r that do react with i. If $\text{AB}(i, r)$ fails to react with any cell, then $\text{EL}(i, r) = r$ and no new reagents have been produced. When $\text{AB}(i, r)$ is not the empty reagent, so that neither $\text{AB}(i, r)$ nor $\text{EL}(i, r)$ is the empty reagent, then both $\text{AB}(i, r)$ and $\text{EL}(i, r)$ are reagents differing from r and are called the *sons of* r. Every reagent r gives rise to a *tree of descendants of* r with r at the root and each node being connected to its sons. Figure 1 is an example of a small tree of this kind. In this case, r has five descendants (counting r itself) which are of two kinds; *reducible* and *irreducible*.

Notice that r_1, r_3, and r_4 are irreducible in the sense that further absorption and elution were unable to yield any new reagents. However, r and r_2 are reducible, since r decomposes into r_1 and r_2, and r_2 decomposes into r_3 and r_4.

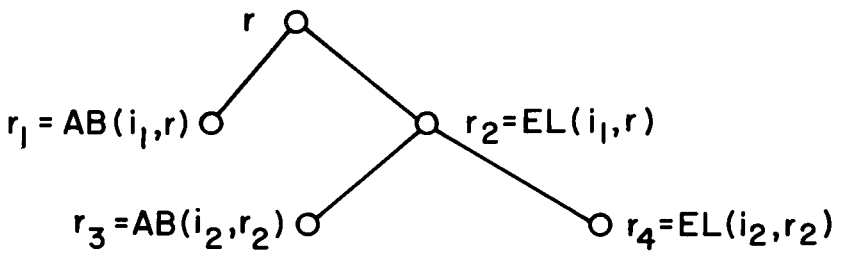

FIG. 1.

Irreducible descendants can only contain equivalent antibodies. To see this, suppose r is a reagent containing at least two distinct antibodies α and β. By the Separation Condition, there is a cell i reacting with exactly one of them, say α. Thus α is in EL(i, r), but not in AB(i, r), and β is not in EL(i, r) but is in AB(i, r). By Principle 2 every antibody must react with some cell, so that both EL(i, r) and AB(i, r) are recognized as sons of r, and r cannot be irreducible. Thus the procedure to reduce a reagent consists of performing absorptions and elutions until all the irreducible descendants have been found, at which point the reagent is totally decomposed into antibodies.

Several observations will help systematize this process and reduce the amount of work involved. First, in reducing a reagent, there is no need to use cells which do not react with the given reagent. Second, once a cell is tested against a reagent r, and perhaps even used to decompose r, there is no need to use the same cell in working with any of the descendants of r.

There are also two observations which are useful for estimating the amount of effort which is involved in reducing a reagent by using the process described above. First, the number of irreducible descendants is exactly equal to the number of antibodies in the original reagent. Second, the number of irreducible descendants is exactly 1 greater than the number of reducible descendants (this follows from a simple inductive argument). Thus in any tree, r will have $2n-1$ descendants, where n is the number of antibodies in r.

Given a reagent r which reacts with m cells $i_1,\ldots,i_m$, no more than roughly m^2 operations (depending on how operations are counted) are required either to decide that r is irreducible or to decompose r into two new reagents. In particular, first produce EL(i_1, r) and AB(i_1, r), and test the latter with $i_2,\ldots,i_m$ to determine whether r has been decomposed or not. If r has not been decomposed, i_1 can be dropped from further consideration, because every antibody in r reacts with i_1 and thus every descendant of r would react with i_1. Thus to eliminate i_1 requires constructing EL(i_1, r) and AB(i_1, r) and testing the latter against the $m-1$ cells $i_2,\ldots,i_m$. If AB(i_1, r) proves completely inactive, produce EL(i_2, r) and AB(i_2, r) and test the latter against the $m-2$ cells $i_3,\ldots,i_m$. By proceeding in this way r can either be decomposed or shown to be irreducible. In all this process requires at most $m-1$ elutions, $m-1$ absorptions, and at most $(m-1)+(m-2)+\cdots+1=m(m-1)/2$ reaction tests. Since r has $2n-1$ descendants (where n is the number of antibodies in r), the total number of operations is at most $(2n-1)(m-1)$ elutions, $(2n-1)(m-1)$ absorptions, and $(2n-1)m(m-1)/2$ reaction tests. Note that these numbers are just upper bounds, since the number of cells used decreases when we work with the descendants of r. Refining these estimates further would require much additional analysis, which would not be terribly appropriate here. The chief point of the preceding discussion is to provide a simple upper bound on the amount of work needed to completely decompose a reagent into antibodies.

Stage (2) is very similar to stage (1) with the roles of cells and reagents reversed, counterblocking replacing elution and blocking replacing absorption. Note that the appropriate versions of Principles 1–3 hold for antigens at the beginning of stage (2). To see this, observe that Principles 1 and 2 are perfectly symmetric with respect to antigens and antibodies. Since the reagents are now simply sets of equivalent antibodies, we have the appropriate form of Principle 3: *No antigen reacts with every antibody, and no antibody reacts with every antigen*. However, this is an immediate consequence of the original Principle 3. The appropriate form of the Separation Condition is: *Given two distinct antigens a and b, there is an antibody α which reacts with one of the antigens but not with the other*. This is now a consequence of Principle 1.

Thus in the same way that stage (1) reduced all the reagents to antibodies by using elution and absorption, stage (2) will reduce all the individuals to antigens by using counterblocking and blocking.

6. SUMMARY

The preceding section has shown that it is theoretically possible to determine the relation Z between antigens and antibodies in a system. This is accomplished as follows. By absorption and elution a set S of irreducible reagents is obtained, and by blocking and counterblocking a new set D of irreducible cells is obtained. Reactions between the cells in D and the reagents in S give the matrix Z defining antigens and antibodies. A cell $i \in C$ is labeled by an antigen provided that the irreducible cell corresponding to that antigen is found as a descendant of i. This gives the labeling matrix W. Similarly the labeling matrix E is obtained: those elements of S that are irreducible descendants of a given $r \in R$ give the antibody labels of r.

Thus the factorization $M = W \times Z \times E$ is obtained. It is seen that the information contained in this factorization can be uniquely determined by experimentation provided the diversity of cells tested is adequate for elucidating the system involved. An immunological separation condition is given that will insure the needed diversity. In turn it is seen that complete simple genetic systems satisfy these requirements.

REFERENCES

1 C. Denniston, A note on serological interpretations, *Anim. Blood Grps. Biochem. Genet.* 7:101–108 (1976).

2 C. Denniston and M. Sheehy, Determining antigenic relationships on the basis of MLC testing I. A matrix approach for listing all possible antigenic assignments, *Tissue Antigens* 16:118–126 (1980).

3 J. Hirschfeld, Serologic codes. Interpretation of immunogenetic system, *Science* 148:968–971 (1965).

4 J. Hirschfeld, Immunogenetic models, *Nature* 239:385–386 (1972).
5 J. Hirschfeld, Antibody relations—implications in genetic theory formation, *Human Heredity* 22:130–137 (1972).
6 J. Hirschfeld, Introduction to a conceptual framework in serology, *Prog. Allergy* 19:275–312 (1975).
7 J. Hirschfeld, The doctrine of monospecific antibodies: The history of an immunological myth, *Afr. J. Clin. Exp. Immunol.* 1:23–44 (1980).
8 G. Markowsky and A. Wohlgemuth, Uncovering antibody incidence structures, *Math. Biosci.* 52:141–156 (1980).
9 G. Markowsky and A. Wohlgemuth, Mathematical immunogenetics II—antibody incidence structure, *J. Theoret. Biol.*, to appear.
10 G. Markowsky and A. Wohlgemuth, Intersection-union systems, *Discrete Appl. Math.*, to appear.
11 L. Mobraaten, Ph.D. Thesis, Univ. of Maine, 1972.
12 D. Nau, G. Markowsky, M. Woodbury, and D. Amos, A mathematical analysis of human leukocyte antigen serology, *Math. Biosci.* 40:243–270 (1978).
13 M. Sheehy and C. Denniston, Determining antigenic relationships on the basis of MLC testing II. Mechanical algorithms, *Tissue Antigens* 16:127–139 (1980).
14 A. Wohlgemuth, Abstract immunogenetic systems, *J. Theoret. Biol.* 73:469–508 (1978).
15 A. Wohlgemuth, Modeling immunogenetic specificities, *Math. Biosci.* 45:175–177 (1979).
16 A. Wohlgemuth and G. Markowsky, A fragment-cofragment model of antibody incidence structures, *Math. Biosci.* 53:265–273 (1981).
17 A. Wohlgemuth and G. Markowsky, Mathematical immunogenetics I—mathematics as language, *J. Theoret. Biol.*, to appear.