J. theor. Biol. (1983) **102,** 411–424

Mathematical Immunogenetics I
Mathematics as Language

ANDREW WOHLGEMUTH

Department of Mathematics, University of Maine at Orono, Orono, Maine 04469, *U.S.A.*

AND

GEORGE MARKOWSKY

Computer Sciences Department, IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, *U.S.A.*

(*Received* 10 *July* 1981, *and in final form* 20 *December* 1982)

This paper summarizes approaches to developing mathematics that can act as a language for immunogenetics. The need for this has been documented by showing inadequacies of the standard symbolism. Apparent distinctions in symbolizing and conceptualizing factors involved in immunogenetics are seen to disappear in the mathematical models presented here. One model, a three-fold Boolean matrix factorization, subsumes all approaches to the idea of specificity and yet is general enough to incorporate data beyond that found only in a reaction matrix.

1. Symbolic Representation

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 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 in as much as it has been able to deal with discrete

0022–5193/83/110411+14 $03.00/0

conceptual entities that *can* be denoted by letter symbols and subject 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 (1965) who has also described the complex–simple and complex–complex codes (Hirschfeld, 1972*b*) which we will describe later.

Hirschfeld (1980) has carefully documented the experimental evidence for cross-reacting antibodies and therefore the inappropriateness of always using the simple–complex code. Mobraaten (1972, 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 (1972*d*, 1975). 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.

2. Artifacts

As an example, the human Ag blood group system considered in Hirschfeld (1972*a*), Hirschfeld & Wohlgemuth (1978) and Wohlgemuth (1978*a*, pp. 489–97) has been thought to be a five locus system with two alleles at each locus—the alleles called x, y; a_1, d; c, g; t, z; h, i. Nine reagents are considered: anti-x, anti-y, . . . , anti-h. That is, each reagent is considered monospecific and as detecting the product of a single gene from among

$x, \ldots, h$. The matrix obtained by testing 362 individuals with the nine reagents is found in (Wohlgemuth, 1978*a*, p. 491).

Using the 362×9 matrix mentioned above as raw data for the Ag system with no assumption made about the reagents each defining a single gene product, the technique in Wohlgemuth (1978*a*) or Hirschfeld and Wohlgemuth (1978) leads to an alternate definition of the genes in the Ag system where there are six alleles $A, \ldots, F$ at a single locus. The definition of each of these genes is now seen in terms of the complete panel of (cross-reacting) reagents. Thus the definition of genes in the system is part of a resolution of the whole system and is, of course, consistent with this resolution. The first mentioned definition of genes $x, y, \ldots$ in the system is consistent with a scientific philosophy that maintains that the best definition is always provided by "isolating" the particular gene. In this case, monospecific reagents or simple antibodies are thought to identify each gene, quite apart from the system as a whole. The properties of the whole system then become a result of the way that the defined entities fit together. For example, there is severe linkage disequilibrium in the Ag systems—with the $\{x, y, a, \ldots, i\}$-definition of genes—most of the expected genotypes not occurring at all. For the single locus model of the system with genes $A, B, \ldots, F$ there is of course not even the possibility of linkage disequilibrium.

It is of course a truism that properties of basic entities in a system depend on the definition of the entities. Mathematically, this is clear. Biologically, this may be obscured by the overly optimistic identification of operationally defined entities with real world counterparts. As long as operationally defined entities do correspond to real world counterparts the properties derived from the definitions can be interpreted as properties of the system. But without this correspondence the properties will be artifacts of the mindset and language used to describe the system. For example, if the six allele $A, \ldots, F$ single locus theory of the Ag system is "true"—$A, \ldots, F$ correspond to real world counterparts—then the linkage disequilibrium seen in the system with the alternate definitions would be an artifact produced by inappropriate notation and a mindset toward reagent or antibody specificity.

As another example, consider the Rh system with alternate definitions given in (Hirschfeld, 1965, 1972*b*, 1973) and (Wohlgemuth, 1978*a*, pp. 499–502). Traditional notation considers Rh as three closely linked loci with gene pairs D, d; C, c; E, e at the three loci. In terms of this notation it is cited in Wohlgemuth (1978*a*) that a man of type CDe/cDE was found with anti-*ce* in his serum. The explanation offered for this has been that the gene product *ce* is produced only when *c* and *e* are on the same

chromosome (in *cis* position) and not when they are in *trans*. Thus the man would not have produced the antigens corresponding to *ce* and could have anti-*ce* without autoimmunity. In terms of the alternate definitions of the Rh genes—given by numbers 5, . . . , 9—the same fact is recorded as: a man of type 7, 9/6, 9 was found with anti-8 in his serum. This is now quite what one would expect. If the latter definitions of genes are "true" then the cis-trans effect would not be a property of the Rh system but an artifact produced by inappropriate notation.

Thus linkage relations and other "genetic laws" may be the result of the way we have used letters to identify antigens and genes—artifacts of an inappropriate notational system. That this is of more than academic interest is seen by the enormous amount of work done in linkage studies. The National Cancer Institute's ICRDB Cancergram devotes an entire section to abstracting such work each month. The writers of the abstracted articles denote antigens with literal symbols that have with little doubt been assigned according to the simple–complex code, that is, each antigen has been defined by a reaction with a single antibody species or monospecific reagent. (For cell-mediated phenomena the language is different but the principle is the same.)

Thus if antibodies (or killer cells) do *not* cross-react the choice of a notational system is easy: an antigen and the corresponding antibody that defines its presence are represented by a single letter. The antigen is X. The antibody is anti-X. In this case the assignment of antigens to an individual or sample is straightforward. If an individual reacts with anti-X, anti-Y and anti-Z the individual must be of type XYZ. We therefore end up with a labeling by the simple–complex code—attributing all complexity to antigens by virtue of our notation. But the experimental evidence summarized in Hirschfeld (1980) suggests that antibodies do cross-react. Further, the exclusive use of the simple–complex code produces almost bizarre "genetic laws" which are completely explained when alternate notation is used.

To summarize our argument so far, we have pointed out that, whatever the complexities involved, it is necessary to refer to the antigens and antibodies in immunogenetic systems by symbols. Monoclonal antibodies ought to be the same, and denoted by the same symbol (not necessarily a single letter). If single letters are used to symbolize these antibodies and antigens are assigned letters to account for reactions (i.e. by the simple–complex code) then this *defines* antigens. It must then be true, just by the way we are using symbols to define things, that the same antibody can never react with two different antigens. (If the antibody is said to define a single antigenic specificity, antigenic determinant, etc., then we would say

that the antibody can never react with two different antigenic specificities, antigenic determinants, etc.) Notice that this is a way of using language. We are using symbols in such a way that it is impossible for an antibody to react with two different antigens. Notice that antigens and antibodies are in a one-to-one correspondence by definition.

We use the term cross-reactivity (regardless of its other possible historic uses) to refer to the opposite of the foregoing. That is, we say our symbolism allows for cross-reactivity when the same antibody can react with antigens denoted by different symbols. Note that this agrees with the definition in Woodbury, Ciftan & Amos (1979).

3. Cross Reactivity and Symbolism

The first symbolic approach that did not preclude complexity in the relation between antigens and antibodies was that of Hirschfeld (1965, 1972*b*, 1975). Hirschfeld considered antigens as sets of more primitive units called *ants* (for antigenic specificity) and antibodies as sets of more primitive units called *antis* (for antibody specificity). Ants and antis are in a one-to-one correspondence. An antigen and antibody react if they share a specificity—that is, if some ant in the antigen corresponds to an anti in the antibody. This is an example of labeling by the complex–complex code: Each antigen is assigned labels from a set of all labels, each antibody is assigned labels from the set, and the antigen and antibody react if they have a label in common.

Hirschfeld was the first to point out the effect that a symbolic language had on the way we see the immunogenetics described by the language. Until recently all mathematical and statistical models and techniques were directed as tools to *uncover* factors (antigens, say) that might account for reactions in a data matrix. The world of immunogenetics was still to be *understood* by assigning a single letter or label to operationally monospecific reagents (the closest approach at the time to a reagent of monoclonal antibodies) and then *defining* an antigen by its reaction with the reagent. A sample that reacted with more than one reagent was considered complex, that is, labelled with all the letters needed to account for reactions with the various reagents. Thus the simple–complex code became part of the way immunogenetics was conceptualized.

4. Cross-Reactivity and Mathematics—The Model

Early mathematical and statistical approaches to immunogenetics all used a symbolism that did not permit cross-reactivity. For example, Elandt-

Johnson (1968) denotes antigens with letters (e.g. A) and "their corresponding antibodies" with barred letters (e.g. $\bar{A}$). There is a tacit assumption of a one-to-one correspondence between antigens and antibodies. Similarly Suttle and Ciftan (1973) state, "Let ... α_i, $i = 1, \ldots, n$ be n antigens and α_i', $i = 1, \ldots, n$ be their corresponding antibodies". The antigen/antibody bipartite graph in Wohlgemuth (1975) was motivated by Suttle and Ciftan's approach and merely allowed for cross-reactivity in its use of symbols. Extension of this work led to the Labeled Reaction Matrix (LRM) model of Wohlgemuth (1976, 1978*a*, 1979). It was shown that the LRM model is equivalent to a three-fold Boolean factorization of a reaction matrix for the first order immunogenetic systems defined in Wohlgemuth (1978*a*). This latter article was motivated by an attempt to use mathematics as a consistent language to describe immunogenetics instead of merely a tool. We will now describe the model of the three-fold factorization and later compare some other approaches with it.

The overwhelmingly predominant pattern of the genetics of histocompatibility and blood grouping is that of co-dominant alleles. Indeed, "well-known" exceptions to this, such as the *cis-trans* effect in the Rh system, may turn out after deeper investigation not to be exceptions at all. In (Wohlgemuth, 1978*a*) a first-order abstract immunogenetic system is defined as a model to cover this predominant genetic pattern. Second and third order models are defined which allow for more genetic complexity, such as dominance. Since most real world systems are of the first order and since the possible exceptions may turn out to be only artifacts it is reasonable to develop the first order models at the outset.

The most convenient description of a first order system is in terms of a matrix product. For an illustration consider the following as our hypothetical example. Let $\mathcal{I}$ be the set of individuals $\{1, 2, 3\}$ each of which is labeled by, i.e. has, some of the antigens a, b, c. This labeling or typing can be given as a matrix $\mathcal{C}$ as in Fig. 1.

In our models antigens are defined by a panel $\mathcal{A}$ of antibodies using a matrix that gives, for each antigen, the complete set of antibodies that

		Antigens			
		a	*b*	*c*	
Individuals	1	1	0	1	Individual 1 has antigen(s) *a, c.*
	2	1	0	0	Individual 2 has antigen *a*
	3	0	1	0	Individual 3 has antigen *b*
		(a)			(b)

FIG. 1. The labeling matrix $\mathcal{C}$ of our hypothetical example. (a) $\mathcal{C}$ as a matrix; (b) $\mathcal{C}$ as a list.

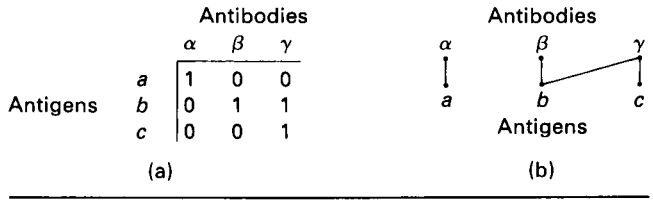

FIG. 2. The definition matrix $\mathscr{D}$ of our example. (a) $\mathscr{D}$ as a definition matrix; (b) $\mathscr{D}$ as a (bipartite) graph.

react with the antigen. As part of our example consider the definition matrix $\mathscr{D}$ of Fig. 2. For example, in $\mathscr{D}$ antigen b reacts with antibodies β and γ, while antibody β reacts only with antigen b. The definition matrix $\mathscr{D}$ is equivalent to and easily visualized as the antigen/antibody bipartite graph given in Fig. 2(b), see (Wohlgemuth, 1975). Cross-reactivity is easily seen by the edge connecting b and γ. (As an aside for mathematicians note that the group $C_2^3 \otimes S_3$ mentioned in Suttle & Ciftan (1973) is the automorphism group of this *un*directed bigraph with edge $\{b, \gamma\}$ deleted.)

The relation between individuals $\mathscr{I}$ and antibodies $\mathscr{A}$ is given by a labeling matrix $\mathscr{G}$ which is the Boolean product of the matrices $\mathscr{C}$ and $\mathscr{D}$. This matrix product is the same as the normal matrix product except that we always take $1+1=1$. The best introduction for the immunologist to matrix notation and this product is probably found in (Wohlgemuth, 1978*b*). The matrix $\mathscr{G} = \mathscr{C} \times \mathscr{D}$ is given in Fig. 3. The rule for determining $\mathscr{G}$ is that,

$\mathscr{G}$	α	β	γ		$\mathscr{C}$	a	b	c		$\mathscr{D}$	α	β	γ
1	1	0	1		1	1	0	1		a	1	0	0
2	1	0	0	=	2	1	0	0	×	b	0	1	1
3	0	1	1		3	0	1	0		c	0	0	1

FIG. 3. The labeling matrix. $\mathscr{G} = \mathscr{C} \times \mathscr{D}$.

for example, a 1 is found in row 3 column β because β reacts with some antigen labeling individual 3; in general there is a 1 in row i, column j of $\mathscr{G}$ if and only if individual i is labelled by some antigen recognized by antibody j. In this case we say antibody j is a label for individual i so that $\mathscr{G}$ is viewed as a labeling matrix.

In the serological application of our model a "reagent" is considered as a combination of antibodies. In application to cell-mediated phenomena

this is analogous to combinations of whatever factors (possibly cross-reacting) recognize antigens. The importance of the distinction between recognized and recognition factors will be brought out later in this paper. The distinction between cell-mediated and serological recognition factors is not important in our model and we will use serological language calling these merely "antibodies" and combinations used in testing "reagents". The particular combination of antibodies in some reagent is given by a labeling matrix. The matrix $\mathscr{E}$, part of our example, is given in Fig. 4.

		Reagents			
		1	2	3	4
Antibodies	α	1	0	0	0
	β	1	1	0	1
	γ	0	1	1	0

(a)

Reagent 1 has antibodies α, β
Reagent 2 has antibodies β, γ
Reagent 3 has antibody γ
Reagent 4 has antibody β

(b)

FIG. 4. The labeling matrix $\mathscr{E}$. (a) $\mathscr{E}$ as a matrix; (b) $\mathscr{E}$ as a list.

The example is completed by the computation of the reaction matrix $\mathbf{R}$ as $\mathscr{C} \times \mathscr{D} \times \mathscr{E}$ which is given in Fig. 5.

R:

		Reagents					$\mathscr{C} \times \mathscr{D} \times \mathscr{E}$:									
		1	2	3	4		a	b	c			α	β	γ		1	2	3	4
	1	1	1	1	0		1	1	0	1	a	1	0	0	α	1	0	0	0
Individuals	2	1	0	0	0	=	2	1	0	0	$\times$ b	0	1	1	$\times$ β	1	1	0	1
	3	1	1	1	1		3	0	1	0	c	0	0	1	γ	0	1	1	0

FIG. 5. The reaction matrix. $\mathbf{R} = \mathscr{C} \times \mathscr{D} \times \mathscr{E}$.

The matrix product gives the following rule for $\mathbf{R}$: a 1 is seen in $\mathbf{R}$ denoting a reaction between individual i and reagent j if and only if some antigen labeling i is recognized by an antibody labeling j. The factorization of $\mathbf{R}$ into $\mathscr{C} \times \mathscr{D} \times \mathscr{E}$ gives a labeling of the rows and columns of $\mathbf{R}$ with antigens and antibodies and $\mathbf{R}$ together with the factorization is called a "labeled reaction matrix". The problem of defining the basic entities antigens and antibodies responsible for data as seen (in part) in a reaction matrix $\mathbf{R}$ is equivalent to the problem of factoring $\mathbf{R}$ into three factors.

5. Specificity

The point we wish to make now is that the three-fold matrix factorization is general enough to incorporate the idea of specificity. The first abstract approach to specificity was given by the letter codes of Hirschfeld (1975). That these codes are equivalent to a two-fold factorization of a matrix is shown in (Wohlgemuth, 1978*b*). This paper is easily read and there is no need to reproduce the results here. Note that if the antigen/antibody reactions given by a matrix (say $\mathscr{D}$ of our example) are "explained" in terms of specificity, this amounts to a factorization of the reaction matrix. If, in the future, experimental techniques are adequate for determining this additional set of factors, the factorization may prove useful. At this time it is quite difficult just to identify antigens and antibodies in an unbiased and unambiguous way. Note also that genes are another set of factors that "label" individuals and code for antigens. Incorporating these too would give a five-fold factorization. Hirschfeld (1980) lists 27 different words or phrases used to describe factors that could be called antigenic specificities. Since the matrix product is associative, it is possible to group those factors not of interest together (see Wohlgemuth, 1978*a*, p. 507). In fact from a mathematical perspective it makes no difference what names we give the factors. The only thing that matters is the role that they play. Since the two-fold matrix factorization is subsumed in a three-fold factorization simply by assuming one of the relations involved to be a one-to-one correspondence, the question is whether accounting for a reaction matrix in terms of a single set of factors is adequate. We shall see that the experimental technique of absorption makes the distinction between recognized and recognition factors important. Hence at least a three-fold factorization is needed.

A mathematical definition of specificity is given by Nau *et al.* (1978) and is shown to be equivalent again to a two-fold factorization in (Wohlgemuth, 1979). As an example of the latter we can take the factorization of $\mathscr{G}$

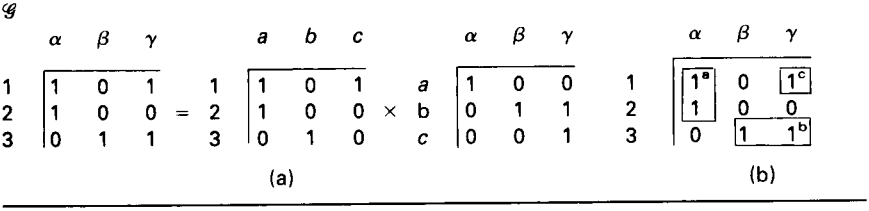

FIG. 6. The specificity cover corresponding to a factorization. (a) $\mathscr{G} = \mathscr{C} \times \mathscr{D}$ from Fig. 3. (b) The specificity cover of $\mathscr{G}$.

originally given in Fig. 3 and here reproduced in Fig. 6. Consider, a, b and c in this context as "specificities" where 1, 2, 3 would now be considered as distinct *antigens* and α, β, γ distinct antibodies. An antigen and antibody react if and only if they share a specificity. The specificities of the antigens 1, 2, 3 are given in terms of the (labeling) matrix $\mathscr{C}$. The specificities of the antibodies α, β, γ are given in terms of the labeling matrix $\mathscr{D}$.

Observe that the set of all rows (antigens) of $\mathscr{G}$ labeled by specificity a (i.e., 1 and 2) and the set of all columns (antibodies) of $\mathscr{G}$ labeled by a (i.e., α) form a submatrix of $\mathscr{G}$ consisting of a solid block of 1's. The same is true for the submatrices defined by b and c. Further all 1's in $\mathscr{G}$ are covered by these blocks of solid 1's. Such a covering of a reaction matrix is called a specificity covering (Nau *et al.*, 1978).

Thus, explaining any one of the matrices in $\mathbf{R} = \mathscr{C} \times \mathscr{D} \times \mathscr{E}$ by specificities, either by Hirschfeld's codes or by a specificity cover, is equivalent to a further factorization of the matrix in question. There is therefore a lack of meaningful distinction among the various sorts of "factors" as they are used in the model. For example, in the model, rows (individuals) of $\mathbf{R}$ could equally well be labeled by genes, antigens, antigenic determinants or antigenic specificities. The name is not important to the model. (In a second or third order system, however, the distinction between labeling with genes and antigens must be maintained.) The fact that the superficially different ways of defining specificity are all essentially the same but obtained independently argues for the appropriateness of the matrix model which subsumes them all.

6. Finding Factorizations

The practice of labeling operationally monospecific reagents or reagents of monoclonal antibodies with a single letter and individuals or samples with all letters needed to account for reactions (the simple–complex code) has become the predominant symbolic "language" of immunogenetics. Although, as we have stated before, this practice is not based on sound evidence on the nature of antibodies and although its use leads to bizarre results in some cases, it nevertheless has two very strong points in its favor. It is easy to understand and it is easy to use. Mathematical approaches, if they are to replace the traditional language, will have to be workable as well as faithful to reality. If the unbiased way of symbolizing an immunogenetic system involves a factorization then the way or ways of obtaining this factorization needs to be studied.

Pioneering approaches, to using mathematics in immunogenetics, such as Ciftan (1970), usually aimed at determining the number of antigens or

factors needed to account for a reaction matrix **R** of data. Although we have no expertise in statistics it would seem to us that the same is true for the statistical approaches. For example Brown, Davis & Goodman (1970, p. 703) state, "Our problem is first to determine how many antigens and corresponding antibodies must be postulated". (The precise effect that assuming a one-to-one correspondence between antigens and antibodies has on the statistics may be worth investigating.) This problem of finding the smallest number of antigens needed to account for **R** is very complex. For example we could ask the following question about factorizations: For a given whole number n and a given **R** does there exist a factorization of **R** involving n antigens? This question falls into the category of NP-complete problems—problems which are computationally intractable, or for which there is really no computational approach better than a case by case exhaustive search. For a discussion see (Nau *et al.*, 1978). This reference also contains results on the bounds of possible n (giving a "yes" answer) for a given **R**. There are of course usually more pressing requirements for the definitions of genes or antigens involved in a system than merely minimizing the number required to account for a reaction matrix. Nau & Woodbury (1977) provide a heuristic computer program for determining labelings of **R** (with say antigens). This program is run interactively with input from a biologist. Another example of heuristic approach is given in the uncovering of the Ag and Rh labelings in (Wohlgemuth, 1978*a*).

If only **R** is available as data then, remarkably little can be said with certainty about any factorization—two-fold or three-fold. What is needed is to incorporate more information into the model/language. The second part of this paper is aimed at precisely this. In the next section we will give an example to show why a three-fold factorization is needed when absorption (or an equivalent technique for cell-mediated reactions) information is included in the model.

7. Recognized and Recognition Factors

In the further development of the model, information on reagents is incorporated. Recall that the matrix $\mathscr{G}$ (Fig. 3) of our example, with its original meaning, labels individuals with antibodies: α and γ label individual 1, α labels individual 2 and β and γ label individual 3. Recall that to say, α labels individual 1 is to say that α reacts with some antigen labeling individual 1. Consider the reagent 2 anti-1, i.e., (individual 2) anti-(individual 1). (Such known combinations for producing reagents occur in working with mouse strains where individuals 1 and 2 stand for two different but known strains). 2 anti-1 is a reagent obtained by challenging individual

2 with cells from individual 1 and would normally, and in our model, contain only antibodies that recognize cells in individual 1 but do not react with cells from individual 2. Thus the only antibody in our example that could be contained in reagent 2 anti-1 is γ; β does not recognize any antigen in 1 and α recognizes an antigen in 2 itself. Under any conditions assuring us of immunocompetence, 2 anti-1 would be $\{\gamma\}$. In a reaction matrix then, 2 anti-1 should react with individuals 1 and 3 but not with 2.

In models incorporating absorption (Nau *et al.*, 1978; Denniston, 1976) it is taken that specificities subtract. Thus 2 anti-1 would be labeled with all those specificities of individual 1 minus the specificities of individual 2. If we take antibodies or antibody labels as specificity, this agrees with what we have seen so far: all antibody labels of individual $1 = \{\alpha, \gamma\}$; all antibody labels of individual $2 = \{\alpha\}$; and $\{\alpha, \gamma\} - \{\alpha\} = \{\gamma\} = 2$ anti-1.

Notice, however, that we cannot take *antigen* labels as specificities in this sense; all antigen labels of $1 = \{a, c\}$, all antigen labels of $2 = \{a\}$, and $\{a, c\} - \{a\} = \{c\}$. But reagent 2 anti-1 cannot be labeled with $\{c\}$ since this would provide no explanation for the reaction of 2 anti-1 with individual 3. The only antigen labeling individual 3 is b (recall figure 3). Since 2 anti-$1 = \{\gamma\}$ and γ does react with b we see the need for labeling reagents with antibodies or *recognition factors* when we take absorption of stimulator/responder information into account.

This makes good physical sense. What is absorbed out? Antibodies, entities that play the role of recognition factors. Thus when we incorporate absorption information or its equivalent, it is important to make a distinction between recognized (antigen) factors and recognition factors. That is, a three-fold factorization $\mathbf{R} = \mathscr{C} \times \mathscr{D} \times \mathscr{E}$ of a reaction matrix is the important form of a model of a first order immunogenetic system.

As we pointed out in a previous section it is not important from a mathematical point of view what we call recognition factors or recognized factors.

We note that the work of Sheehy & Denniston (1980) and Denniston & Sheehy (1980) also uses known stimulator/responder combinations to determine factors involved in mixed lymphocyte culture testing. This work is fundamentally different from our work in that it draws *inferences* about recognized factors from information on whether a given stimulator–responder combination can produce any recognition factor. Factors are not subtracted.

8. Antigens, Antibodies and Genes

The abstract immunogenetic system of Wohlgemuth (1978*a*) is an attempt to provide a context in which to examine the relationships between

the symbols used to represent genes, antigens, antibodies and specificities. This model is general enough to include all symbolic descriptions in immunogenetics as special cases. Earlier, Hirschfeld (1975) had related genes, antigens, and antibodies by a double use of the complex–complex code. Thus genes are considered as sets of more primitive units called *aggs*. Aggs, ants, and antis are all in a one-to-one correspondence whereas the correspondence between genes and antigens is no longer one-to-one but is determined by the complex–complex code. Similarly the correspondence between antigens and antibodies is determined by another use of the complex–complex code. This approach does place some restriction on the "second" use of the code however since the determination of antigens as sets of ants is accomplished by one use of the code (say to relate antigens to antibodies) and thus we are not free to redefine antigens in terms of ants in the "second" used of the code (to relate genes and antigens.) Of course there is no present evidence to suggest that this model of Hirshfeld is not adequate to cover actual immunogenetic systems. But nevertheless this does introduce a bias in the language. That is, it puts some prior constraint on the possible relations between genes and antigens and antibodies.

9. Summary

Immunogenetics deals with antigens and genes which are denoted by letter symbols. The practice of always identifying a particular antigen by an operationally monospecific reagent or reagent of monoclonal antibodies leads to an assignment of letter symbols by what is called the simple–complex code. The use of this code in cases where there is cross-reactivity can result in apparent genetic laws which in reality are no more than artifacts of the notational system. Thus cross-reactivity is more than just an annoyance. Its existence has been proven and necessitates a descriptive language for immunogenetics that is more general than the simple–complex code. There have been several independently developed symbolic and mathematical approaches to providing this descriptive language. All of these are subsumed in a model called an abstract immunogenetic system and are equivalent (for a first order system) to a Boolean matrix factorization. If the only experimental results available to identify antigens consist of a table or reaction matrix, then a two-fold factorization has enough generality to describe the factors involved. These factors may be called specificities, antigens, antibodies or even genes. If a reaction matrix is the only data available there is no reason to expect that the factors involved in the two-fold factorization correspond to one rather than another of these

names. There are an enormous number of two-fold factorizations—some identifying genes, some antibodies, etc., and some having no physical counterpart whatever.

If absorption information is available, a distinction must be made between recognized factors and recognition factors. Symbols assigned to recognition factors subtract in correspondence with absorption data but the same is not true (due to cross-reactivity) of recognized factors. Thus a three-fold factorization is the important form of a descriptive language when absorption information is incorporated in the data. A precise mathematical connection between absorption information and the factorization is given in the second part of this paper.

We are grateful for the interest and suggestions of the referees.

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

REFERENCES

BROWN, R., DAVIS, H. & GOODMAN, H. (1970). *Biometrics* **26,** 701.
CIFTAN, M. (1970). *Math. Biosci.* **6,** 487.
DENNISTON, C. (1976). *Anim. Blood Groups biochem Genet.* **7,** 101.
DENNISTON, C. & SHEEHY, M. (1980). *Tissue Antigens* **16,** 127.
ELANDT-JOHNSON, R. (1968). *Am. J. hum. Genet.* **20,** 213.
HIRSCHFELD, J. (1965). *Science* **148,** 968.
HIRSCHFELD, J. (1972*a*). In: *Protides of the Biological Fluids.* Oxford: Pergamon Press.
HIRSCHFELD, J. (1972*b*). *Nature* **239,** 385.
HIRSCHFELD, J. (1972*c*). *Hum. Hered.* **22,** 124.
HIRSCHFELD, J. (1972*d*). *Hum. Hered.* **22,** 130.
HIRSCHFELD, J. (1973). *Vox Sang.* **24,** 21.
HIRSCHFELD, J. (1975). *Prog. Allergy* **19,** 275.
HIRSCHFELD, J. (1980). *Afr. J. clin. exp. Immunol.* **1,** 23.
HIRSCHFELD, J. & WOHLGEMUTH, A. (1978). In *Progress in Theoretical Biology.* New York: Academic Press.
MARKOWSKY, G. & WOHLGEMUTH, A. (1980). *Math. Biosci.* **52,** 141.
MOBRAATEN, L. (1972). *Ph.D. Thesis.* Orono, Maine: University of Maine.
NAU, D., MARKOWSKY, G., WOODBURY, M. & AMOS, D. E. (1978). *Math. Biosci.* **40,** 243.
NAU, D. & WOODBURY, M. (1977). *Comp. Biomed. Res.* **10,** 259.
SHEEHY, M. & DENNISTON, C. (1980). *Tissue Antigens* **16,** 118.
SUTTLE, J. & CIFTAN, M. (1973). *Math. Biosci.* **16,** 315.
WOHLGEMUTH, A. (1975). *Math. Biosci.* **25,** 51.
WOHLGEMUTH, A. (1976). *Notices AMS* **23,** A422.
WOHLGEMUTH, A. (1978*a*). *J. theor. Biol.* **73,** 469.
WOHLGEMUTH, A. (1978*b*). *Immunogenetics* **7,** 481.
WOHLGEMUTH, A. (1979). *Math. Biosci.* **45,** 175.
WOHLGEMUTH, A. & MARKOWSKY, G. (1981). *Math. Biosci.* **53,** 265.
WOODBURY, M. A., CIFTAN, E. A. & AMOS, D. B. (1979). *Comp. Prog. Biomed.* **9,** 263.