J. theor. Biol. (1983) **102**, 425–437

Mathematical Immunogenetics II
Antibody Incidence Structure

GEORGE MARKOWSKY

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

AND

ANDREW WOHLGEMUTH

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

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

Mathematical Immunogenetics I argued for the development of mathematics as a language for immunogenetics. A three-fold factorization of a reaction matrix was seen to be the important form of a model of a first order immunogenetic system. In the present paper, results of the authors on determining this factorization are reworked from a physical perspective and presented in an algorithmic form that can be used to compute a labeling matrix from data. Computer programs to perform these computations are in preparation.

1. Introduction

In Mathematical Immunogenetics I it was seen that the important form of a model for a first order immunogenetic system is a 3-fold Boolean factorization $\mathbf{M} = \mathscr{C} \times \mathscr{D} \times \mathscr{E}$ of a reaction matrix $\mathbf{M}$. The problem of determining this factorization depends for its solution on applying information other than that contained only in the reaction matrix $\mathbf{M}$. We shall see in this paper how information about stimulator/responder combinations used in producing reagents or their cell-mediated analogs can be used to gain information on this factorization.

In this paper the mathematical results of Markowsky & Wohlgemuth (1980) and Wohlgemuth & Markowsky (1981) will be reworked from a physical (biological) standpoint and we shall see their implications for using mathematics as a language for immunogenetics. In addition to these two papers the works of Denniston (1976) and Nau *et al.* (1978) also consider absorption information. In all of these works, and indeed in all work known to us, it is assumed that "absorption"—or its equivalent for cell-mediated

0022-5193/83/110425+13 $03.00/0

phenomena—subtracts factors. While the latter two works refer to these factors as "specificities" an example in Mathematical Immunogenetics I shows that only recognition factors subtract. Recognized factors do not. Thus we consider it dangerous to use the term "specificity". In all immunogenetic systems where cross-reactivity cannot be ruled out *a priori* "specificity" must be interpreted as "recognition factor" when subtracted, as in absorption. We will use the word "antibody" to denote such a recognition factor in this paper. Thus we use serologic language but the model is more general.

Thus if $\mathbf{M} = \mathscr{G} \times \mathscr{D} \times \mathscr{E}$, $\mathscr{G}$ labels individuals with antigens (recognized factors), $\mathscr{E}$ labels reagents with antibodies, and $\mathscr{D}$ defines antigens in terms of antibodies and vice versa. In application, $\mathscr{D}$ represents the set of producible antibodies for the system under investigation. If there is no cross-reactivity then $\mathscr{D}$ is just a one-to-one correspondence between the antigens in the system and their corresponding antibodies. In this case there is no mathematical (or linguistic) distinction necessary between recognized and recognition factors. If there are n antigens there could imaginably be a total of 2^n different antibodies in a system as they are identified by their reactions with the antigens. This does not seem to be the case for real systems although it is difficult to say this for certain since in many cases the real systems in question have been conceptualized assuming $\mathscr{D}$ to be merely a one-to-one correspondence. In this paper we make no assumptions about $\mathscr{D}$. The producible antibodies may be any subset whatever of the set of imaginable antibodies. There are also no restrictions on cross-reactivity.

The matrix product $\mathscr{G} = \mathscr{C} \times \mathscr{D}$ labels individuals with producible antibodies and the factorization $\mathbf{M} = \mathscr{G} \times \mathscr{D} \times \mathscr{E}$ becomes $\mathbf{M} = \mathscr{G} \times \mathscr{E}$. Since this factorization labels both individuals and reagents with recognition factors, it is this factorization that absorption information uncovers. For the sake of illustration we will consider as our Example 1 the example of Wohlgemuth and Markowsky in Mathematical Immunogenetics I for $\mathscr{G} = \mathscr{C} \times \mathscr{D}$ reproduced here in Fig. 1.

$\mathscr{G}$:					$\mathscr{C}$:					$\mathscr{D}$:			
	α	β	γ			a	b	c			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. 1. The labeling matrix. $\mathscr{G} = \mathscr{C} \times \mathscr{D}$.

We will explain our results using the language of sets. No mathematical arguments will be presented. Results that do not follow from the development but come from Markowsky & Wohlgemuth (1980) or Wohlgemuth & Markowsky (1981) are marked by a (*).

For an individual i, the set of all producible antibodies labeling i is denoted by $i\mathcal{G}$. If S is a set of individuals, the set of all producible antibodies labeling any individual in S is denoted by $S\mathcal{G}$. Recall that an antibody labels an individual i if and only if that antibody can be produced and reacts with some antigen of i.

Let

$$\delta\langle j, l\rangle = j\mathcal{G} - l\mathcal{G}.$$

In words $\delta\langle j, l\rangle$ is the set of all producible antibodies labeling j except those labeling individual l.

Let l anti-j denote the reagent obtained by using individual j as stimulator and individual l as responder. We take as an hypothesis for this paper that $\delta\langle j, l\rangle$ is the antibody content of reagent l anti-j. It is clear that the content of l anti-j must be a subset of $\delta\langle j, l\rangle$ (in any model worth investigating). Some justification for our hypothesis will be given below after we have introduced some more terminology.

In mouse serology it is found that the content of the reagent $m \times l$ anti-j is the same as the content of l anti-j with m absorbed or m anti-j with l absorbed. There is an equivalence of using F_1 responders and absorption. Our hypothesis is consistent with this. Also notice that the equivalence of m anti-j with l absorbed and l anti-j with m absorbed demands some sort of explanation in the model. If m anti-j contains only a subset of $\delta\langle j, m\rangle$ and l anti-j contains only a subset of $\delta\langle j, l\rangle$ there would seem to be no reason why absorbing with l and m respectively would lead to reagents that had the same reaction patterns.

In keeping with the ideas of the previous paragraph we extend our hypothesis as follows: If $S = \{s_1, \ldots, s_k\}$ (a set of individuals) we consider the reagent s_1 anti-j with $s_2, \ldots, s_k$ absorbed to have antibody content $\delta\langle j, S\rangle$. Note that this should be the same as s_2 anti-j with $s_1, s_2, \ldots, s_k$ absorbed if our hypothesis is correct. And this would seem to agree with experimental results.

Another paper which investigates a weaker hypothesis is in preparation. Many of the ideas of this present paper carry through with some modification but we prefer the current hypothesis largely because of the symmetry it gives in the notation (which we will shortly explain) and because we have seen this symmetry in the raw data we have seen. The weakened hypothesis would not seem to support this symmetry.

2. The Boolean Model

We now work backward from a fixed but unknown $\mathcal{G}$ to data required to determine $\mathcal{G}$. For an integer $k \geq 1$ let the reaction table $R(\mathcal{G}, k)$ be defined as follows.

(1) Rows are all individuals $i \in \mathcal{I}$.

(2) Columns are all pairs $\langle j, S\rangle$, $j \in \mathcal{I}$, $S \subseteq \mathcal{I}$ where $|S| \leq k$ (S is a set with k or fewer elements).

(3) the entry in row i and column $\langle j, S\rangle$ is the set of antibodies labeling row i and also column $\langle j, S\rangle$. This set is denoted by $\Delta(i, j, S)$.

Thus, $\Delta(i, j, S) = i\mathcal{G} \cap \delta\langle j, S\rangle = i\mathcal{G} \cap j\mathcal{G} - S\mathcal{G}$. The symmetry in our hypothesis lacking in the weaker hypothesis is that $\Delta(i, j, S) = \Delta(j, i, S)$.

Let $\mathcal{I}$ be the set of individuals tested and define the Boolean reaction matrix $\mathbf{M}(\mathcal{G}, k)$ with rows corresponding to each $i \in \mathcal{I}$ and columns corresponding to each $\langle j, S\rangle$, $|S| \leq k$. $\mathbf{M}(\mathcal{G}k,)$ has a *one* in row i, column $\langle j, S\rangle$ if $\Delta(i, j, S) \neq \emptyset$ (is not empty) and *zero* otherwise. If the entry in row i column $\langle j, S\rangle$ is one we write $i\mathbf{M}\langle j, S\rangle$—if it is zero we write $\neg i\mathbf{M}\langle j, S\rangle$. It is easily seen that $i\mathbf{M}\langle j, S\rangle$ if and only if an antigen labeling i reacts with an antibody in $\delta\langle j, S\rangle$. Thus $\mathbf{M}(\mathcal{G}, k)$ plays the role of a zero/one reaction matrix which can be taken as data. In histo-typing using F_1 responders, $k = 2$. In serological tests for a given k we are considering $k - 1$ absorptions.

Note that $\mathbf{M}(\mathcal{G}, 1)$ can be considered as a subset of the results given in $M(\mathcal{G}, 2)$. If n is the number of individuals in the data, then $\mathbf{M}(\mathcal{G}, n-1)$ gives information on all possible tests involving any number of absorptions. We call $\mathbf{M}(\mathcal{G}, n-1) = \mathbf{M}(\mathcal{G}, \omega)$. Thus

$$M(\mathcal{G}, 1) \subseteq M(\mathcal{G}, 2) \subseteq \ldots \subseteq M(\mathcal{G}, \omega).$$

In the same way we get

$$\mathbf{R}(\mathcal{G}, 1) \subseteq \mathbf{R}(\mathcal{G}, 2) \subseteq \ldots \subseteq \mathbf{R}(\mathcal{G}, \omega).$$

$\mathbf{R}(\mathcal{G}, 1)$ and $\mathbf{M}(\mathcal{G}, 1)$ for the $\mathcal{G}$ of Example 1 are given in Fig. 2. Antibodies labeling each row and column are given.

Suppose part of a reaction matrix is given as in Fig. 3.

Here 1 is used as stimulator for all reagents considered and n is 7. Because of the 1 in row 4 column $\langle 1, 2\rangle$, individual 4 and individual 1 must be labeled by a common antibody say α. Further, α must also label 5, 6 and 7 since absorbing by 5, 6 and 7 gives zeros in columns $\langle 1, 5\rangle$, $\langle 1, 6\rangle$ and $\langle 1, 7\rangle$. Therefore, the set $\{1, 4, 6, 7\}$ must be a subset of the reaction range of α. The reaction pattern in Fig. 3 implies the existence of an antibody whose reaction range includes $\{1, 4, 5, 6, 7\}$. $\{1, 4, 5, 6, 7\}$ is called a 1-*fragment* (from $k = 1$). If there is an antibody α whose reaction

(a)

		$\delta\langle j, l\rangle$								
		$\emptyset$	γ	α	$\emptyset$	$\emptyset$	α	β	$\beta\gamma$	$\emptyset$
		$\langle 1,1\rangle$	$\langle 1,2\rangle$	$\langle 1,3\rangle$	$\langle 2,1\rangle$	$\langle 2,2\rangle$	$\langle 2,3\rangle$	$\langle 3,1\rangle$	$\langle 3,2\rangle$	$\langle 3,3\rangle$
	$\alpha\gamma$ 1	$\emptyset$	γ	α	$\emptyset$	$\emptyset$	α	$\emptyset$	γ	$\emptyset$
$i\mathcal{G}$	α 2	$\emptyset$	$\emptyset$	α	$\emptyset$	$\emptyset$	α	$\emptyset$	$\emptyset$	$\emptyset$
	$\beta\gamma$ 3	$\emptyset$	γ	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	β	$\beta\gamma$	$\emptyset$

(b)

		$\emptyset$	γ	α	$\emptyset$	$\emptyset$	α	β	$\beta\gamma$	$\emptyset$
		$\langle 1,1\rangle$	$\langle 1,2\rangle$	$\langle 1,3\rangle$	$\langle 2,1\rangle$	$\langle 2,2\rangle$	$\langle 2,3\rangle$	$\langle 3,1\rangle$	$\langle 3,2\rangle$	$\langle 3,3\rangle$
	$\alpha\gamma$ 1	0	1	1	0	0	1	0	1	0
$i\mathcal{G}$	α 2	0	0	1	0	0	1	0	0	0
	$\beta\gamma$ 3	0	1	0	0	0	0	1	1	0

FIG. 2. (a) $\mathbf{R}(\mathcal{G}, 1)$ and (b) $\mathbf{M}(\mathcal{G}, 1)$ for the $\mathcal{G}$ of Example 1.

	$\langle 1,2\rangle$	$\langle 1,3\rangle$	$\langle 1,4\rangle$	$\langle 1,5\rangle$	$\langle 1,6\rangle$	$\langle 1,7\rangle$
4	1	1	0	0	0	0

FIG. 3. Part of some $\mathbf{M}(\mathcal{G}, k)$.

range is precisely $\{1, 4, 5, 6, 7\}$, α is called 1-*tractable*. This is the biological view of the idea of a fragment.

Recall that if there is a one in row i and column $\langle j, m\rangle$ of $\mathbf{M}(1)$ we denote this by $i\mathbf{M}\langle j, m\rangle$. A zero in this position is denoted by $\neg i\mathbf{M}\langle j, m\rangle$. For a fixed i, j if $\neg i\mathbf{M}\langle j, m\rangle$ for all m then no individual makes an antibody attacking i and j: One way to have $\neg i\mathbf{M}\langle j, m\rangle$ for all m is that there is an antibody α labeling i and j but α labels every other individual as well. α would then label no reagent and have no effect on reactions. We make the assumption in this paper that no antibody has a universal reaction range—a clearly uninteresting situation in our model. The only other way to have $\neg i\mathbf{M}\langle j, m\rangle$ for all m is that i and j have no common antibody label.

Definition. For any pair i, j of individuals define $F(i, j)$ as follows:

(1) If $\neg i\mathbf{M}\langle j, m\rangle$ for all m, let $F(i, j) = \emptyset$.

(2) Otherwise, let $F(i, j)$ be the set of all m such that $\neg i\mathbf{M}\langle j, m\rangle$. Non-empty sets $F(i, j)$ are called 1-*fragments*.

Note that $F(i, j) = F(j, i)$ and $i, j \in F(i, j)$ if $F(i, j) \neq \emptyset$.

Definition. For any pair of individuals i, j and any subset S of individuals of size $|S| \leq k - 1$ define $F(i, j, S)$ as follows:

(1) If $\neg iM\langle j, S\rangle$, let $F(i, j, S) = \emptyset$.

(2) Otherwise let $F(i, j, S)$ be the set of all m such that $\neg iM\langle j, S + m\rangle$. (Here $S + m = S \cup \{m\}$).

Non-empty sets $F(i, j, S)$ for $|S| \leq k - 1$ are called *k-fragments*. Note that i, j are in the fragment $F(i, j, S)$ and $F(i, j, S) = F(j, i, S)$. Suppose we have some fragment $F(i, j, S)$. Then there is an antibody α labeling i and j which is not absorbed out by S since $iM\langle j, s\rangle$. Since $\neg iM\langle j, S + m\rangle$ for $m \in F(i, j, S)$, α must label m. Thus $\mathcal{G}\alpha$ (the reaction range of α) must include the set $F(i, j, S)$ and not intersect S. An antibody that is equal to a k-fragment is called *k-tractable* and is called *tractable* if it is k-tractable for some k.

Note that k-fragments can be computed from the matrix $\mathbf{M}(\mathcal{G}, k) = \mathbf{M}(k)$ and in fact the matrix $\mathbf{M}(k)$ can be recovered if all the k-fragments are known. The model that uses the matrices $\mathbf{M}(k)$ as data from which to determine $\mathcal{G}$ will be called the "Boolean model". In order to use the Boolean model it is necessary to distinguish the presence or absence of a reaction in each row i and column $\langle j, S\rangle$.

If we delete an antibody α from the relation $\mathcal{G}$ (denoted by $\mathcal{G} - \alpha$) the effect on any matrix $\mathbf{M}(k)$ would be to possibly change some ones to zeros. We denote this by $\mathbf{M}(\mathcal{G} - \alpha, k) \leq \mathbf{M}(\mathcal{G}, k)$. If equality is preserved, then α is called *k-Boolean-undetectable. Thus* α is *k-Boolean-detectable* if $\mathbf{M}(\mathcal{G} - \alpha, k) \neq \mathbf{M}(\mathcal{G}, k)$. α is called *Boolean-detectable* if it is k-detectable for some k, that is, if $\mathbf{M}(\mathcal{G} - \alpha, \omega) \neq \mathbf{M}(\mathcal{G}, \omega)$. An undetectable antibody is therefore one that has no effect on a reaction matrix obtained by any number of absorptions—its action will always be masked by the reactions of other

(a)

	α	β	γ	δ	ϵ
1	1	1	1	0	0
2	1	1	0	1	0
3	1	0	1	1	0
4	0	0	0	0	1

(b)

	β	γ	δ	ϵ
1	1	1	0	0
2	1	0	1	0
3	0	1	1	0
4	0	0	0	1

		γ $\langle 1, 2\rangle$	β $\langle 1, 3\rangle$	$\alpha\beta\gamma$ $\langle 1, 4\rangle$	δ $\langle 2, 1\rangle$	β $\langle 2, 3\rangle$	$\alpha\beta\delta$ $\langle 2, 4\rangle$	δ $\langle 3, 1\rangle$	γ $\langle 3, 2\rangle$	$\alpha\gamma\delta$ $\langle 3, 4\rangle$	ϵ $\langle 4, 1\rangle$	ϵ $\langle 4, 2\rangle$	ϵ $\langle 4, 3\rangle$
$\alpha\beta\gamma$	1	1	1	1	0	1	1	0	1	1	0	0	0
$\alpha\beta\delta$	2	0	1	1	1	1	1	1	0	1	0	0	0
$\alpha\gamma\delta$	3	1	0	1	1	0	1	1	1	1	0	0	0
ϵ	4	0	0	0	0	0	0	0	0	0	1	1	1

FIG. 4. Example 2, a Boolean undetectable antibody α in $\mathcal{G}$. (a) $\mathcal{G}$; (b) $\mathcal{G} - \alpha$; (c) $M(\mathcal{G}, 1) = M(\mathcal{G} - \alpha, 1)$.

antibodies. The condition that an antibody be undetectable is given in Theorem 1 below. In the following example (2) antibody α is undetectable.

In $\mathbf{M}(1)$ of Fig. 4(c) we have omitted the useless columns $\langle j, j\rangle$. Observe that deleting α from $\mathcal{G}$ has no effect on $\mathbf{M}(1)$. It is clear that deleting α has no effect on $\mathbf{M}(\omega)$ so α is Boolean-undetectable.

The condition on $\mathcal{G}$ that an antibody be undetectable is given in terms of the reaction range of antibodies. For any antibody α in $\mathcal{G}$ the set of all individuals that are labeled by α is written $\mathcal{G}\alpha$. In Example 2 $\mathcal{G}\alpha = \{1, 2, 3\}$, $\mathcal{G}\beta = \{1, 2\}$. $\mathcal{G}\varepsilon = \{4\}$.

In fact for us an antibody is defined by its reaction range so we may think of α as the set $\{1, 2, 3\}$. With this terminology notice that α is the union of β, γ and δ. Moreover, this union is of a special kind since for any pair i, j of individuals in α (i.e., in the reaction ragne of α), i and j will be in β, γ or δ. Theorem 1 (*) below states that this is precisely the way in which an antibody must be Boolean-undetectable.

Definition. A subset $S \subseteq \{1, 2, \ldots, N\}$ is a *pair-undetectable union* of its proper subsets $T_1, T_2, \ldots, T_k$ if

(1) $\bigcup_{i=1}^{k} T_i$ and

(2) for any pair $j, l \in S$ there is some T_i such that $j, l \in T_i$.

Theorem 1. An antibody is Boolean-undetectable if and only if it is a pair-undetectable union of other antibodies. □

3. The Fragment–Cofragment Model

Fragment $F(2, 1, 4)$ can be calculated from the piece of $R(\mathcal{G}, 2)$ found in Fig. 5. We define the *cofragment* $C(2, 1, 4)$ as the set of all individuals Dm for which $\Delta(2, 1, 4) \neq \Delta(2, 1, 4 + m)$, that is, we compare the number of antibodies that row 2 (α, β, δ) has in common with reagent $\langle 1, 4\rangle$ and with reagent $\langle 1, 4 + m\rangle$ for each $m = 1, 2, 3, 4$.

We see that for $m = 1, 2, 3$, $\Delta(2, 1, 4 + m) = \emptyset, \emptyset, \{\beta\}$ respectively, and for $m = 4$ $\Delta(2, 4 + m) = \{\alpha, \beta\} = \Delta(2, 1, 4)$. Under appropriately designed experimental conditions one would expect to see a reaction in row 2 column $\langle 1, \{4, 3\}\rangle$ which is a reduction of the reaction in row 2 column $\langle 1, 4\rangle$, since $\Delta(2, 1, 4) = \{\alpha, \beta\}$ but $\Delta(2, 1, \{4, 3\}) = \{\beta\}$. Cofragments are defined as sets of individuals for which we see reduction in reaction strength. These include fragments where we see a reduction to no reaction at all. Formally we have the following.

(a)

		$\alpha\beta\gamma$ $\langle 1,4\rangle$	$\emptyset$ $\langle 1,\{4,1\}\rangle$	γ $\langle 1,\{4,2\}\rangle$	β $\langle 1,\{4,43\}\rangle$	$\alpha\beta\gamma$ $\langle 1,\{4,4\}\rangle$
$\alpha\beta\delta$	2	$\alpha\beta$	$\emptyset$	$\emptyset$	β	$\alpha\beta$

(b) $F(2,1,4)=\{1,2\}$
since $\neg 2M\langle 1,\{4,1\}\rangle$
and $\neg 2M\langle 1,\{4,2\}\rangle$
but $2M\langle 1,\{4,3\}\rangle$
and $2M\langle 1,\{4,4\}\rangle$

(c) $C(2,1,4)=\{1,2,3\}$
since $\emptyset=\Delta(2,1,\{4,1\})\subset\Delta(2,1,4)=\{\alpha,\beta\}$
and $\emptyset=\Delta(2,1\{4,2\})\subset\Delta(2,1,4)=\{\alpha,\beta\}$
and $\{\beta\}=\Delta(2,1\{4,3\})\subset\Delta(2,1,4)=\{\alpha,\beta\}$
but $\{\alpha,\beta\}=\Delta(2,1\{4,4\})=\Delta(2,1,4)=\{\alpha,\beta\}$

FIG. 5. (a) Part of the reaction table $R(\mathcal{G},2)$ for $\mathcal{G}$ of Example 2 plus calculations of (b) $F(2,1,4)$ and (c) $C(2,1,4)$.

Definition. For $|S|\le k-1$ define

$$F(\mathcal{G},i,j,S)=\{m\,|\,\Delta(i,j,S)\neq\emptyset \text{ and } \Delta(i,j,S+m)=\emptyset\}$$

$$C(\mathcal{G},i,j,S)=\{m\,|\,\Delta(i,j,S)\neq\Delta(i,j,S+m)\}$$

Fragments and cofragments are defined mathematically, working back from $\mathcal{G}$, in terms of antibodies in the entries $\Delta(i,j,S)$. In using the model to determine $\mathcal{G}$, the sets of fragments $F(i,j,S)$ and cofragments $C(i,j,S)$ are used as data. The table $\mathbf{R}(\mathcal{G},k)$ is not available. The matrix $\mathbf{M}(\mathcal{G},k)$ can be used to determine fragments as in the Boolean model but cofragments must be calculated from data using reaction strength comparisons: full-reaction, partial-reaction, no-reaction.

Mathematically there is a dual relation between fragments and cofragments. Fragments are intersections of antibody reaction ranges—cofragments are unions.

Theorem 2(*).

(a) $F(i,j,S)=\bigcap_{\alpha\in\Delta(i,j,S)}\mathcal{G}\alpha$.

(b) $C(i,j,S)=\bigcup_{\alpha\in\Delta(i,j,S)}\mathcal{G}\alpha$.

Since α in Example 2 is Boolean-undetectable, deleting α from $\mathcal{G}$ will have no effect on fragments. But deleting α does affect cofragments. In particular $C(\mathcal{G},2,1,4)=\{1,2,3\}$ and $C(\mathcal{G}-\alpha,2,1,4)=\{1,2\}$ for the preceding example. Thus detectability is somewhat sharper using the fragment–cofragment model.

Definition. An antibody α is called *k-fragment–cofragment undetectable* if $\mathcal{G}$ and $\mathcal{G}-\alpha$ give the same sets $F(i,j,S)$ and $C(i,j,S)$ for $|S|\le k-1$. An antibody is called *fragment–cofragment undetectable* if it is k-fragment–cofragment undetectable for all k.

Definition. A subset $S \subseteq \{1, \ldots, n\}$ is a *triple-undetectable union* of its proper subsets $T_1, T_2, \ldots, T_k$ if

(1) $S = \bigcup_{i=1}^{k} T_i$ and

(2) for any triple $j, l, m \in S$ there is some T_i such that $j, l, m \in T_i$.

Theorem 3(*). An antibody is fragment–cofragment undetectable if and only if it is a triple-undetectable union of other antibodies. □

To provide input for the fragment–cofragment model an experimenter would need to decide which reactions in row i, column $\langle j, S\rangle$ are positive and for each of these which reactions in row i, column $\langle j, S+m\rangle$ are as strong as those in row i, column $\langle j, S\rangle$. Note that since $\Delta(i, j, S+m)$ is a subset (perhaps proper) of $\Delta(i, j, S)$ we avoid comparing reaction strengths of different antibodies. In order to use cofragments it is therefore necessary to have an experimental procedure that makes comparison of reaction strengths for various columns and the same row meaningful. It is perhaps more natural experimentally to obtain a meaningful comparison between various rows for the same column (reagent). However, detectability in models using the latter comparisons is no sharper than in the Boolean model.

Tractable antibodies are those that are by definition obtained as fragments. The following theorem states that if we are given enough information (i.e. given $\mathbf{M}(\omega)$) all Boolean detectable antibodies are obtained in this way.

Theorem 4(*) An antibody is Boolean-detectable if and only if it is tractable. □

Working with fragments is then a conservative approach to finding $\mathcal{G}$ which will reveal all detectable antibodies in time. Using the fragment–cofragment model gives more information in that at any stage ($k = 1, 2, \ldots$) we can have some idea of how far away we are from an antibody. In particular, Theorem 2 shows that $F(i, j, S)$ is a lower bound for the reaction ranges of all antibodies labeling iG and $\delta\langle j, S\rangle$, and $C(i, j, S)$ is an upper bound. Thus if $F(i, j, S) = C(i, j, S)$ there must be some antibody α such that $\mathcal{G}\alpha = F(i, j, S)$. Also if $|C(i, j, S)|$ is one larger than $|F(i, j, S)|$ then both $C(i, j, S)$ and $F(i, j, S)$ are $\mathcal{G}\alpha$ and $\mathcal{G}\beta$ for antibodies α and β.

If complete information is not available we would like to find a best possible solution to the problem of finding $\mathcal{G}$ and to know how close we are to finding the "real" $\mathcal{G}$.

We describe this first for the Boolean model when only $\mathbf{M}(1)$ is known as data.

4. Solutions in the Boolean Model from M(1) as Data

Let Ω_1 be the set of all 1-fragments obtained from $\mathbf{M}(1)$. We wish to make precise the sense in which Ω_1 can be considered a "solution" to the problem of discovering $\mathcal{G}$. First, Ω_1 is a set of subsets of individuals. Since from our viewpoint an antibody is completely determined by its reaction range we can view each fragment in Ω_1 as an antibody that reacts with the individuals that are contained in the fragment when we view it as a set. Thus Ω_1 can be viewed as a relation from the set of individuals to the set of antibodies. Second, if we calculate $\mathbf{M}(\Omega_1, 1)$ we get precisely $\mathbf{M}(\mathcal{G}, 1)(*)$.

Definition. A relation R from the set of individuals is a *Boolean solution for* k to the problem of finding $\mathcal{G}$ if $\mathbf{M}(R, k) = \mathbf{M}(\mathcal{G}, k)$. If every "antibody" in R is detectable we call R a *detectable solution* for k.

We have the following result.

Theorem 5(*). Ω_1 is the largest detectable solution for $k = 1$. If R is any other solution for $k = 1$, the elements of R are either in Ω_1 or undetectable unions of the fragments in Ω_1. □

A word about the interpretation of Theorem 5. If an element in R is "built up" as an undetectable union of elements in Ω_1 these elements need not all appear in R. For example, $\{1, 2, 3\}$ is the undetectable union of $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$. If all four of these sets occur in R, then $\{1, 2, 3\}$ is undetectable in R. If, however, the last three sets occur in Ω_1 and only $\{1, 2, 3\}$, $\{1, 2\}$ and $\{1, 3\}$ occur in R, then $\{1, 2, 3\}$ is *detectable in* R. It is therefore possible to have many detectable solutions for $k = 1$. This (mathematical) difficulty is ultimately resolved (in theory) by enough absorptions as the next theorem shows.

Theorem 6(*). There is one and only one detectable solution for $k = \omega$. □

The unique solution of Theorem 6 is obtainable using fragments. The procedure is also used to find largest detectable solutions for all $k \geq 2$.

5. Boolean Solutions from M(k), k ≥ 2

For $k \geq 2$ it is not true that Ω_k the set of all k-fragments is a solution. The procedure for determining the largest detectable solution for a given k can be given algorithmically:

(1) Start with Ω_k the set of all k-fragments. If $\mathbf{M}(\Omega_k, k) = \mathbf{M}(\mathcal{G}, k)$, then undetectable unions can be deleted from Ω_k to obtain the desired result. In general, however, $\mathbf{M}(\mathcal{G}, k) \leq \mathbf{M}(\Omega_k, k)$ and there may be zeros in $\mathbf{M}(\mathcal{G}, k)$ that correspond to ones in $\mathbf{M}(\Omega_k, k)$. If $\neg i\mathbf{M}(\mathcal{G}, k)\langle j, S\rangle$, but $i\mathbf{M}(\Omega_k, k)$

$\langle j, S\rangle$, then it can be shown that S can be taken minimal in the sense that $i\mathbf{M}(\Omega_k, k)\langle j, S-t\rangle$ for all $t \in S$. Further, there is some fragment B in Ω_k that contains i and j and misses S. In this case change Ω_k by deleting B from it and adding to it all sets $B+t$ for each $t \in S$ to obtain a new set Ω_k'.

(2) If $\mathbf{M}(\Omega_k', k) < \mathbf{M}(\mathcal{G}, k)$, then the procedure of (1) can be repeated replacing Ω_k with Ω_k'.

(3) Repeating 1 and 2 above will eventually yield a set Ω_k^* where $\mathbf{M}(\Omega_k^*, k) = \mathbf{M}(\mathcal{G}, k)$. Deleting undetectable unions from Ω_k^* yields the largest detectable solution.

Theorem 7(*). For a given k, the largest detectable solution Ω is obtained by deleting detectable unions from the set Ω_k^* obtained by the preceding algorithm. If R is any other solution for k, the elements of R are undetectable unions of the elements in Ω. □

If the complete matrix $\mathbf{M}(\omega)$ is known there is a more direct method of determining Ω.

Theorem 8(*). Let $F = (i, j, S)$ be a fragment. $F(i, j, S)$ is $\mathcal{G}\alpha$ for some antibody α if and only if $i\mathbf{M}\langle j, \mathcal{I} - F\rangle$ where $\mathcal{I}$ is the set of all individuals not in the fragment F. □

Thus if $\mathbf{M}(\omega)$ is known we can calculate a list of all fragments. For each fragment F we can look to see if $iM\langle j, \mathcal{I} - F\rangle$ and delete F from the list if not. This procedure will give all detectable antibodies in $\mathcal{G}$.

6. Solutions in the Fragment–Cofragment Model

We need to have $k \geq 2$ in order to use the fragment–cofragment model. In this case we have the set of k-fragments as data. The notion of detectability is somewhat sharper in this model but we can use the same definitions in the present context.

Definition. A relation R from the set of individuals is a *fragment–cofragment solution for* k to the problem of finding $\mathcal{G}$ if $F(R, i, j, S) = F(\mathcal{G}, i, j, S)$ and $C(R, i, j, S) = C(\mathcal{G}, i, j, S)$ for all i, j and $|S| \leq k-1$. If every "antibody" in R is (fragment–cofragment) detectable we call R a *detectable solution* for k.

Analogous to Theorem 6 we have also in this model:

Theorem 9(*). There is one and only one detectable solution for $k = \omega$. □

If not enough is known to determine $\mathcal{G}$ uniquely (up to detectability) it is still possible to define a best solution in the manner of the two previous

sections. The procedure is somewhat easier conceptually but may be slower computationally in some cases.

Definition. For any $k \geq 2$ let Ω_k be the set of $X \subseteq \mathcal{S}$ for which we can find i, j, S such that $X = F(i, j, S) = C(i, j, S)$ for $|S| < k - 1$ plus the set of all $Y \subseteq \mathcal{S}$ such that $F(i, j, S) \subseteq Y \subseteq C(i, j, S)$ for $|S| = k - 1$.

Definition. $H = \Omega_k - \{Y \in \Omega_k |$ for some x, y, T we have $x, y \in Y$, $Y \cap T = \emptyset$ and $\neg F(x, y, T) \subseteq Y \subseteq C(x, y, T)\}$.

Theorem 10(*). Deleting triple-undetectable unions from H gives the desired best possible solution. □

H is the only solution of Theorem 9 if $k = \omega$. For other k we have that the elements of any other solution are obtained as triple-undetectable unions of the elements in our best solution. In both models, best solutions converge to the essentially unique solution for increasing k.

7. Experimental Design

If $\mathcal{G}$ is a solution labeling individuals with detectable recognition factors (antibodies) and if $\mathcal{G}$ is used to determine recognized factors in $\mathcal{C}$ according to $\mathcal{G} = \mathcal{C} \times \mathcal{D}$ then since any undetectable antibody is a union of columns of $\mathcal{G}$ it is a union of columns of $\mathcal{C}$. Hence, $\mathcal{C}$ is still a factor of the matrix $\mathcal{G}$ augmented with any combination of undetectable antibodies. Thus *undetectable antibodies have no effect on the determination of recognized factors* (antigens or genes or antigenic determinants).

The preceding material can influence experiments designed to determine $\mathcal{G}$. There is no design using our data that will reveal undetectable antibodies. On the other hand undetectable antibodies have no effect on determining antigenic or recognized factors. Complete data is sufficient to uniquely determine detectable antibodies, but in general if only data for $k < \omega$ is known some ambiguity about detectable solutions is left unresolved. Even for $k < \omega$ a best possible solution does however present itself and this may be sufficient in a practical situation.

$\mathbf{M}(k)$ has $n \sum_{i=1}^{k} \binom{n-1}{i}$ columns and n rows. Here $\binom{n-1}{i}$ is the binomial coefficient "$n-1$ choose i." This gives us $n^2 \sum_{i=1}^{k} \binom{n-1}{i}$ tests which potentially might need to be done to determine $M(k)$. Even for small n and k this number grows rapidly. For $n = 10$ and $k = 3$ we have 4600 potential tests. Much of the information resulting from these tests is repetitive and could be used to check the consistency of the data. Alternatively, a systematic way of avoiding the need to do all tests is suggested in Markowsky & Wohlgemuth (1980).

It is seen from the models that if use is made of reaction strength, comparisons should be meaningful between reactions of a single individual with various sera. Due to the lack of symmetry in the model, comparisons between reactions of a single reagent with various individuals do not provide equally useful data.

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

DENNISTON, C. (1976). *Anim. Blood Groups biochem. Genet.* **7**, 101.
MARKOWSKY, G. & WOHLGEMUTH, A. (1980). *Math. Biosci.* **52**, 141.
NAU, D., MARKOWSKY, G., WOODBURY, M. & AMOS, D. B. (1978). *Math. Biosci.* **40**, 243.
WOHLGEMUTH, A. & MARKOWSKY, G. (1981). *Math. Biosci.* **53**, 265.