INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University
Microfilms
International

300 N. Zeeb Road
Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

8302558

Perlin, Mark William

RANDOM GRAPHS APPLIED TO THE IMMUNE NETWORK

City University of New York PH.D. 1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

RANDOM GRAPHS APPLIED TO THE IMMUNE NETWORK

by

MARK PERLIN

A dissertation submitted to the Graduate Faculty
in Mathematics in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
The City University of New York.

1982

This manuscript has been read and accepted for the University Committee in Mathematics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

5/27/82 Stanley Kaplan
date Chairman of Examining Committee

5/27/82
date Executive Officer

Michael Anshel

Stanley Kaplan
Supervisory Committee

The City University of New York

Abstract

RANDOM GRAPHS APPLIED TO THE IMMUNE NETWORK

by

Mark Perlin

Adviser: Professor Stanley Kaplan

Chapter I treats the classical pre-1970 clonal selection theory of immune memory. An explanation of the genetic restrictions on the immune response is given. We develop some estimates of clonal connectivity which are used in later chapters.

In Chapter II we discuss Jerne's network theory of the immune system. We look at the Boolean matrix products of large 0-1 matrices. A variance estimate shows us how changes in the probability p of a matrix entry equalling 1 produce changes in the immune network's memory. We also apply some results of Erdos and Renyi to the immune network.

In Chapter III, we look at the specificity labelling, or rank, problem as it originally arose in immunogenetics. We develop some probabilistic formulas for this lattice embedding problem which we use in Chapter IV. In the final

section we extend some of these ideas to matrices with entries in an arbitrary lattice, not just {0,1}. We show that such an extension is possible if and only if the lattice is a chain lattice.

We introduce a probabilistic version of the rank lattice embedding problem in Chapter IV. In particular, we conjecture that the probabilistic rank function rises from a value of 0 at p=0 to a maximum of rank n at p=log n/n, and then decreases linearly to the value 1 at p=1. This is applied to immunology and discrete mathematics.

In Chapter V we discuss the population dynamics and stability of the interacting clones in a suppressive immune network. We use the May-Wigner stability theorem, a result about the eigenvalues of random matrices. This stability theorem is used to describe the aging of the immune system in an individual and to explain the evolution of the antigen receptor.

Preface

1.Motivation

The immune system is a large, complex collection of cells whose main task is to recognize and eliminate foreign invaders. We would like to understand how it recognizes that an entity is foreign, and how it processes and stores this information. Some insight into the dynamics of its functioning would be useful as well.

The immune system is too large for a deterministic analysis, being comprised of some 10^6 to 10^7 distinct clones. The key feature of the immune system which enables it to recognize virtually any foreign biological invader is that the clones are constructed as a randomly generated basis set of receptors. We can use the large size of the system, together with this ingredient of randomness, to model the immune system using random variables. Most of our analysis will be with random graphs.

The mathematical models used and developed here will help us to better understand the biology, particularly the network theory of the immune system. As importantly, however, some of these methods are quite general and can be applied to problems in other fields, such as computer science.

2.Results

We prove some results about directed random graphs, developing, for example, the probabilistic rank of a 0-1 matrix. Besides giving us insight into the network theory, this rank notion is applied to circuit design and Boolean independent rows of a 0-1 matrix.

On the biological side, some of our random graph results will be used to describe the effects of aging on the immune system. Other immune phenomena, such as the observed genetic restriction on the immune response and the effect of pharmocological agents which enhance or reduce immune suppression will be discussed using random variables.

This dissertation is, to a large extent, an experiment in applying the random graph concept to the immune system. Although we do examine the standard clonal selection theory, most of our attention will be focused on the network theory. We will suggest that there is an optimal clone-clone connectivity which maximizes the information processing capacity of the immune network. We will use a stability theorem about random matrices to discuss the population dynamics of lymphocyte clones.

3. Overview of chapters

Chapter I treats the classical pre-1970 clonal selection theory of immune memory. An explanation of the genetic restrictions on the immune response is given. We will develop some estimates of clonal connectivity which will be used in later chapters.

In Chapter II we will look at the Boolean matrix products of large 0-1 matrices. Later on, a variance estimate will show us how changes in the probability p of a matrix entry equalling 1 produce changes in the immune network's memory. We will also apply some results of Erdos and Renyi to the immune network.

In Chapter III, we look at the specificity labelling, or rank, problem as it originally arose in immunogenetics. We develop some probabilistic formulas for this lattice embedding problem which we use in Chapter IV. In the final section we extend some of these ideas to matrices with entries in an arbitrary lattice, not just {0,1}. We show that such an extension is possible if and only if the lattice is a chain lattice, e.g., the natural numbers.

We introduce a probabilistic version of the rank lattice embedding problem in Chapter IV. In particular, we conjecture that the probabilistic rank function rises from a value of 0 at $p=0$ to a maximum of rank n at $p=\log n/n$, and then decreases linearly to the value 1 at $p=1$. This is applied to immunology and discrete mathematics.

In Chapter V we discuss the population dynamics and stability of the interacting clones in a suppressive immune network. We use the May-Wigner stability theorem, a result about the eigenvalues of random matrices. This stability theorem is used to describe the aging of the immune system in an individual and to explain the evolution of the antigen receptor.

4. Suggestions for reading

Those unfamiliar with immunology should begin by reading the Appendix. Chapter I should at least be skimmed over to get a taste of how the immune system comes into being and learns to distinguish "self" from "not-self". The mathematics, per se, begins in Chapter II and continues through Chapter V.

Table of Contents

Chapters

Appendix

List of figures

I. Clonal Selection

A human being does not inherit enough genetic information from its parents to discern what chemistry is "self" and what is "not-self". Obviously, however, an individual does inherit the arbitrary cell surface chemistry resulting from random selection from the parental genes. The clonal selection theory [Bu] sketched below will explain how each individual learns to chemically distinguish "self" from "not-self" by using random receptors.

This learning process will be modelled using a simple random walk. The model will then be used to account for the genetic restrictions experimentally observed in the immune response. It will also predict an increase in the total number of clones as the body ages, a result used in the sequel.

1. Antigen recognition

If an individual cannot learn to recognize and eliminate foreign invading organisms, it is only a matter of time before it eventually succumbs to colonizing microorganisms or tumor cells. As discussed in Appendix A, an individual's immune system can recognize a foreign antigen (Ag) by performing a sort of factor analysis on the Ag with its basis set of 10^6 to 10^7 clones. Recall that the Ag is operationally identified by the immune system as a particular small subset of these cellular clones.

The question arises: Why is "self" chemistry not recognized by the immune system? The answer is that the clonal basis set contains only Igs with binding sites having low chemical affinity for self.

Clearly this information cannot be inherited directly from the genes. An individual's cell surface chemistry results from the arbitrary juxtaposition on the cell membrane of many gene products inherited at random from both parents. Just by changing the direction of approach to a single quantum chemical configuration one obtains an infinite number of distinct possible chemical affinity measurements. There is simply not enough information in the finite genome.

2. Clonal elimination

A B-lymphocyte clone is defined by its unique Ig. Although the information of whether a particular Ig molecule has high affinity for some "self" chemical configuration is not encoded in the genome, the answer is easily determined. The Ig molecule needs only to empirically test its binding affinity for the body surface chemistry.

Thus, each clone can perform the following experiment: just construct the unique Ig it codes for and display it on its cell surface. Let it be exposed to the surrounding surface chemistry. If it has high affinity and binds some body chemistry, eliminate it. If not, let it live.

Performing this test on the set of all B-cell clones will eliminate the subset of clones which bind to "self". What survives is the complement of the eliminated set, and every element of the surviving set has the property that it does not have a high affinity for "self". That is, the surviving set detects "not-self".

The set of "not-self" clones was pruned from a very large initial set of B-cells. Furthermore, whether or not a particular Ig binding site combines with some Ag chemistry, while completely determined after an empirical test, is virtually unknowable beforehand. Only if a chemical moiety is very similar to another one can any prediction be made. We will therefore assume that even after the sieving procedure, the remaining set is somehow dense in not-self space. (These spaces will be defined below.)

3. Generation of diversity

Where does the basis set of clonal Igs come from? There are two classical theories in immunology to explain this, and the real situation appears to be somewhere between them.

The somatic mutation theory, see [Bu] and [J i)], attributes the diversity of B-cell clones to random variation of (hypervariable region) Ig DNA during an individual's embryogenesis. A highly idealized version would run as follows. We start with a single cell- the fertilized egg. After many cell divisions, at a certain stage of development, the progenitor immune cell appears. (This cell is deterministically derived from many successive variations on the original single cell.) The progenitor immune cell differentiates into many other immune cell types, one of which is a progenitor B-cell. This B-cell manufactures some genetically encoded Ig molecule. Now somatic mutation begins.

The DNA for this B-cell's Ig is randomly changed (somewhere in the hypervariable region of the gene). Now we have two B-cell types, those expressing Ig and those expressing Ig'. This splitting process continues, and by the random DNA changes which occur, we (geometrically) arrive at millions of different B-cell clones. Each clone has the DNA for and thus manufactures exactly one version of the Ig molecule.

In order to explain the restriction of B-cell Ig recognition to "not-self", we claim that at each split a decision is made. A split yields the old Ig B-cell and a new one Ig'. The new Ig' is tested for chemical affinity to "self" chemistry. If it binds self too avidly, the new clone is eliminated; otherwise it survives. Assuming that the initial progenitor B-cell Ig has low affinity for self, by this process we obtain a basis set of Igs, each of which recognizes only not-self.

The key feature of the somatic mutation theory is that the genome of the early embryo does not contain the DNA for the Igs of the adult. These DNA words are produced by random mutations within the body (i.e., somatic mutation) rather than by some kind of evolutionary process. In a sense then, the immune system evolves anew in each individual, subject to the (ecological) constraints of the individual's cell surface chemistry.

In slight constrast to the somatic mutation hypothesis, the germ line theory holds that all the DNA required for an immune system is present at fertilization. Hence somatic gene mutation need not be invoked.

The idea here is that the Ig binding site is formed from two protein chains. So if n hypervariable genes are coded for, there are n^2 possible combinations. To explain the 10^6 observed distinct Igs then, only 10^3 need be present in the genome. As in the somatic mutation theory, empirical testing of the clonal Ig combinations for chemical affinity to self determines which B-cells survive in the adult.

4. Generation of Diversity, continued

Why should either of these schemes work? What do random changes in DNA, as in the somatic mutation theory, have to do with foreign antigen (Ag) recognition? In the germ line theory, why does random association of two DNA encoded protein chains produce diversity in chemical recognition? Answer: the uncoupling of DNA words and the chemistry of the protein binding sites they code for.

In the appendix we discuss how small changes in DNA encodings produce small, random changes in a protein's three dimensional configuration, and how these small changes, in turn, yield small, random changes in the binding activity of the active site. (If a DNA change has no effect on binding activity or somehow renders a B-cell dysfunctional, the body does not notice the change. Therefore our model will consider only functional changes in affinities.) To make this notion of "small change" precise, we now develop metrics for DNA changes and for changes in chemical affinity.

The DNA metric on the hypervariable region encoding the Ig binding site is the usual one. Let $\Sigma = \{G,A,C,T\}$ be the alphabet of bases. Say that accounting for both the heavy chain and the light chain of an Ig there are n amino acids which comprise the active site. Thus there are 3n bases coding for the binding site. Put S equal to the set of all 3n-tuples with values in Σ. Then S is the set of all possible DNA words coding for the Ig binding sites. The distance between two binding region DNA words is then defined to be the <u>number of positions where the bases differ</u>.

5. Chemical affinity space

Our problem is to measure the binding properties of receptors and other molecules in some rigorous way. We want a space which identifies to every chemical a point describing its binding characteristics. We would like binding to be continuous in this space, that is, moving a small distance affects binding behavior in a small way. A likely candidate might be some kind of quantum mechanical space, since waveform functions tend to be nicely behaved. This is not practical however: setting up such a space may be impossible and, even if we had it, the calculations involved in computing affinities are impractical.

Instead, we will construct our space from _empirical_ binding affinity measurements. As we commented before, Nature (and laboratory biologists) performs experiments, not calculations. Say we are given a receptor h. Let * be some test molecule. We can _measure the binding affinity_ of h to * and record our result as a real non-negative number in R. Since it is order of magnitude of binding which is important in chemistry (most equations use products, not sums of concentrations), the number actually recorded will be the logarithm of the affinity. (If we have the result of many such measurements between different receptors h' and the test *, we can renormalize the result for h and * according to some *-scale.)

Given a receptor h then, we can choose n such test molecules. (These tests may be chosen to suit the given problem, e.g., n variants on the insulin molecule for testing a certain insulin receptor.) For each of the n tests, we measure the binding affinity of h to the test molecule and record the (renormalized) result as a number in R. In this way, we can associate to each receptor h a point in R^n which records h's affinity with every test. Conversely, each point in R^n corresponds to the binding properties of a possible chemical configuration. We call this Euclidean space *chemical affinity space*.

For a particular test, we would expect that small changes in the receptor's three dimensional configuration and binding site would cause only small changes in the receptor's binding affinity to the test molecule. In assessing protein configurations then, we can use the usual Euclidean metric on R affinity space as a measure of distance.

Because of the uncoupling between DNA descriptions and the resulting protein configurations, we now assert: *small changes in the DNA description produce small random changes in the protein's chemical affinity space representation*. The changes in the DNA are quantified by the usual DNA metric, and the changes in affinities are measured by the usual Euclidean distance.

6. Random Walk

When a B-cell clone's Ig DNA undergoes a point mutation to DNA', what can we say about the transition Ig -> Ig' in chemical affinity space? Assuming that the new Ig' has different binding chemistry than Ig, and that the Ig' molecule and its B-cell are functional, we would like to know where in space the new receptor has moved to.

The simplest case is when n=1, and affinity space consists of only one test molecule. Then a transition Ig -> Ig' moves Ig' to the left or right of Ig on the real line R. Because DNA descriptions are uncoupled from protein binding affinities, we cannot predict the direction of chemical affinity change. We therefore assume that the probability of a move to the left or to the right is one half in either direction.

Let Δt be some time increment for Ig -> Ig' splittings, and let Δx be an affinity increment on R. Then if $w(x,t)$ is the position of an Ig obtained by splittings, we observe that

$$w(x,t+\Delta t) = \frac{1}{2} w(x-\Delta x,t) + \frac{1}{2} w(x+\Delta x,t)$$

A Taylor polynomial expansion and subsequent linearization for small Δx and Δt space and time increments yields (in the limit) the diffusion equation $w_t = -Dw_{xx}$. This derivation of Ig position in time generalizes to R^n. We will therefore feel free to speak about Igs *diffusing* through R^n.

For n=1, we prefer to look at the discrete set N of natural numbers instead of R. Let $\Delta x = \Delta t = 1$ be unit increments. Then the above equation describes a *symmetric random walk* on the line. We now generalize to n>1.

For n=2, we have a random walk in the lattice plane N^2. Since we cannot predict chemistry from description, we assume that the walk is symmetric. So Ig' moves up, down, left or right from Ig with probability 1/4. For arbitrary n>0, we have a symmetric random walk on the lattice N^n. Ig' is permitted to move +1 or -1 relative to Ig in one of the n coordinates. So there are 2n possible moves, each with probability 1/2n.

-> Boundary conditions

We assume that there are two types of regions in affinity space. *Type A regions* have predominantly "not-self" chemistry. Thus the Igs are relatively free to diffuse through these regions. Type A space is denoted by the presence of lattice points. A *type B region*, on the other hand, has an abundance of "self" chemistry and this severely restricts the diffusion of Igs. Type B space corresponds to *holes* in N^n where lattice points have been removed. (Note that in a given individual, the partition into types A and B space is arbitrarily determined by the inherited cell surface chemistry. A major task of the immune system is to chart the A regions while avoiding the potential auto-immune catastrophe of wandering into a B region.)

-> The process

We start with

Ig_1.

This eventually produces a mutant Ig', which we relabel Ig :

Ig_1-> Ig', and we get

Ig_1, Ig_2.

Continuing,

Ig_1-> Ig'_1 , Ig_2-> Ig'_2 , to yield

Ig_1, Ig_2, Ig_3, Ig_4.

And so on.

In this way we obtain a geometric expansion of superimposed random walks. The questions arise, What happens in the limit? What does this mean biologically? How do the lattice holes affect the process?

7. Random walk with reflecting barriers in bounded regions of N^n.

The graph of our process is a tree with many bifurcations of the form Ig -> Ig + Ig'. We will start with an Ig, and trace just one path along the tree. This path describes a random walk.

Every time it hits a hole, the new Ig' is absorbed. But then $Ig \rightarrow Ig + Ig' = Ig$ alone. In following our path, it appears as if nothing has happened. So instead of considering holes as absorbing, for a single path a hole is essentially reflecting. That is, when a particle hits a hole, it returns to the lattice point from which it came.

Recall that we are using logarithmic units to measure binding affinity. Beyond some upper bound K, the affinity values in our space are physically unrealizable. So instead of looking at all of N^n, we restrict our attention to the bounded region $B = \{(a_1, a_2, \ldots, a_n) \in N^n : 0 \leq a_i \leq K\}$. We assume that all points outside of B are holes.

8. Analysis in a connected component

The B-type regions (holes) will divide B into a certain number of connected component regions. We consider one such region C.

C may contain holes. To understand the dynamics of the sum total of all paths appearing in the process, some kind of memory is needed to keep track of which paths encountered which holes. We are examining only the limit distribution of a single path, therefore no memory is required.

We call each point in C a <u>state</u>. Without holes, each interior state has 2n neighbors. Allowing for holes, each state has 2n-h neighbors, h being equal to the number of holes directly adjacent to the state. (We may consider points exterior to C as holes as well.)

If E' is a neighbor of E, then the transition probability Prob{E->E'} = 1/2n. Also Prob{E->E} = h/2n. Clearly, the probability of ending up in some new state is 1, so we have a legitimate probability distribution.

Since the transition probability depends only on current state, there is no memory and the process is a <u>Markov chain</u>. If M is the matrix of transition probabilities, we put each entry m_{ij} = Prob{E_i->E_j}. The $\lim_{k\to\infty} M^k$ gives the limit distribution of the random walk.

Observe that the sum of each column of M is 1. This is because E' has 2n-h neighbors and transition probabilities Prob{E->E'} = 1/2n and Prob{E'->E'} = h/2n <u>into</u> E'.

Obviously, $(2n-h)\times 1/2n + h/2n = 1$. Hence M is a doubly stochastic matrix, i.e., both its row sums and column sums add to 1.

Since C is connected and contains only a finite number of states, the basic theorem on invariant distributions [F Thm.XV.7] holds.

This guarantees that the limits $u_j = \lim_{k\to\infty} p_{ij}^{(k)}$ exist, where $p_{ij}^{(k)}$ is the transition probability of moving from E_i to E_j in exactly k steps.

These limits, independent of the initial distribution of states, satisfy

(1) $\sum_j u_j = 1$, and

(2) $u_j = \sum_i u_i p_{ij}$.

Since M is doubly stochastic, the column sums $\sum_i p_{ij} = 1$.

So if $u_j = c$, a constant, equation (2) is satisfied, since $c = \sum_i c p_{ij}$.

If N is the total number of states in C, then $u_j = 1/N$ satisfies equation (1).

Thus $\{u_j\} = \{1/N\}$ is the limit distribution of M.

9. Genetic restriction of the immune response

What is the immunological significance of this? In a connected component of chemical affinity space containing "self" regions, every possible "not-self" Ig configuration occurs. Furthermore, assuming no outside antigenic stimulation, no matter <u>where</u> in the region the first Ig was located, eventually the average size of each B-cell clone will be the same. (That is, a uniform "not-self" coverage of the region.)

So far we have restricted our model to the somatic mutation theory. We now combine our results with some insights from the germ line theory. It appears to be the case (experimentally) that the genes code for several dozens of different Ig molecules. (This is less than the thousands required by a strict germ line approach and more than the single Ig of idealized somatic mutation.)

Assume now that <u>the number of connected type A components</u> of affinity space is <u>greater than the number of initial germ line Igs</u>. If an initial inherited Ig is in a connected region, eventually the component will be filled and an immune response is then possible against every "not-self" moiety in it. If a germ line Ig does not lie in the region, however, this entire range of "not-self" chemistries will be <u>inaccessible to immune recognition</u>. This concurs with the experimentally observed genetic restriction on the Ig repertoire and the immune response.

10. Aging of the immune system

R.K. Gershon has described the generation of immunologic diversity (i.e., random receptors) as arising from three distinct mechanisms [G]. The first mechanism is the inheritance of the particular handful of initial DNA-encoded Ag-binding site genes from the parents. The second is how the not-self random walk unfolds during embryonic development. (This depends as well on the arbitrary surface chemistry inherited from both parents.)

Gershon's third mechanism is the fine tuning to Ag that occurs as new clones are made during an immune response. When Ag is introduced, clones residing in the corresponding regions of chemical affinity space undergo an intense positive selection to optimize their affinity. The end result of such a localized random walk is to add more clones to the immune system. Since whether or not a randomly constructed receptor will bind to any given molecule is unknowable in advance (which is why we treat it as a random variable), the overall influence of these new clones on immune recognition goes far beyond the initial stimulating Ag.

The net effect of this third mechanism is to increase the total number of lymphocyte clones over the life span of an individual. It is reasonable to assume that this constant antigenic challenge results in a ten-fold increase in clones as the immune system matures from birth to old age. (Such an increase in circulating T-cells has been observed experimentally by Hallgren [H].) Jerne has estimated the total number of clones to be around 10^6 to 10^7 [J ii)]. We will therefore take 10^6 to be our estimate of *the number of clones in a young immune system*, and 10^7 as our estimate in an old one.

In new-born mice, 10 to 20 B-cell clones respond to a given Ag [C]. Since the Ag-binding molecules of the B-cells (or the T-cells) are all of the same class, we would expect the newer clones to have the same probability of antigen recognition as the older ones. That is, the probability p of clones-Ag interaction remains *fixed* over the lifetime of an individual. An average of $\mu = np$ clones respond to a given Ag, where n is the total number of clones. In a young immune system, $p=10/10^6=10^{-5}$. So as n changes with age from 10^6 to 10^7, μ increases from 10 to 100.

It is interesting to note that these new clones have not undergone the intense negative selection that occurs during embryogenesis. They would be expected, therefore, to cause more auto-immune cell destruction than the older clones. The presence of these newer clones arising out of continual immune responses supports the immunologic theory of aging [W]: low levels of these clones attack the body over an individual's lifetime and eventually weaken it.

This thesis, however, is more concerned with the aging of the immune system itself. We will show in Chapter II, and by different methods in Chapter IV, that as the number of clones increases in an immune system, the efficiency of Ag recognition goes down. (This agrees with experiment [H].)

11. Immunological note

What motivated this chapter was Baruj Benacerraf's analysis of Norman Klinman's B-cell ontogeny data in his classic "Textbook of Immunology" [B]. Klinman used the haptens DNP (dinitrophenol), TNP (trinitrophenol) and fluorescein (a dye) to understand how new Igs appear in early embryogenesis [Kl]. (A hapten is a small chemical configuration with very high specificity, hence useful in analyzing immune responses.)

Klinman observed that early in ontogeny, B-cell clones appear with Igs which can bind DNP and TNP. It is more than a week after this that clones appear which can recognize fluorescein. Benacerraf interprets this experiment as proof that the "Ig library unfolds in a deterministic way" [K1].

Let R^3 be a chemical affinity space built from the three test chemicals DNP, TNP and fluorescein. Assume an idealized somatic mutation theory where the B-cell random walk starts from just one Ig. It is a very reasonable hypothesis that the Ig starts from a coordinate with high DNP and TNP affinities, but low fluorescein affinity. Let the random walk begin. Lymphocytes divide (roughly) every 24 hours. It is plausible that after 7 or 8 time periods some Ig' ends up at a coordinate with high fluorescein affinity.

Observe that this is not a deterministic model.

II. Network Theory

We present the main features of immune network theory here and model the network as a random graph. We use the results of our analysis to explain the breakdown of immune recognition that accompanies aging.

1. Immunology

In the early 1970's a study was conducted on the survival of kidney transplants. The patients were divided into two populations, those who had received multiple blood transfusions, and those who had not. It was expected that the patients without the transfusions would have a lower rate of graft (i.e., kidney) rejection. This prediction was consistent with the clonal selection theory, since multiple alloantigenic challenges would be expected to increase the populations of B-cell clones with Ig directed against the major histocompatibility complex (MHC). (An alloantigen is an antigen which is widely distributed within a species. The immune response of an individual is much more pronounced against an alloantigen than it is against an arbitrary foreign chemistry. The MHC is a gene complex coding for many such cell surface alloantigens, and is the classic barrier to organ transplantation.)

In fact, however, the study showed that kidney graft recipients who had received multiple blood transfusions had a greater survival rate of their grafts.

Another indication that something was amiss with the strict clonal selection theory was experimental phenomena of low-zone and high-zone tolerance.[3] An animal is said to be tolerant of an antigen if its B-cells can recognize it, but no immune response is mounted against it. Low-zone tolerance occurs when very low concentrations of an immunogenic Ag are introduced to an organism, causing a complete absence of future immune responses to the Ag. The body has learned to forget the Ag. Similarly, high-zone tolerance happens when an individual is overwhelmed with high doses of an Ag, somehow making the animal insensitive to further Ag challenges.

The clonal selection theory predicts a more linear type of response. If physiologic doses of Ag prompt an immune response, very high Ag concentrations should cause an even greater response. Instead, high-zone tolerance is observed. Low-zone tolerance is more problematic. As Jerne observed, at low doses the Ag concentration in the blood is less than the concentration of Igs and B-cells which can bind to it.

How then can the Ag interact with the B-cell clones, much less cause tolerance?

2. Jerne's Network Theory

Jerne was able to explain these, and other, anomalies with a network theory of the immune system [J iii)]. We make the simplifying assumption that of all the (infinitely) possible chemical configurations of the Ig binding site, only a handful of these are actually used in biology. We call such an Ig binding site configuration a paratope. Simlilarly, we assume that there are only a small, finite number of chemical specificities on an Ag that come into play in an immune response. We call such a chemical moiety an epitope.

We now make the further assumption (which is actually all we need to assume for the sequel) that lymphocyte clones' Ag binding sites can be presented as epitopes to the paratopes of other lymphocytes. That is, lymphocytes can chemically recognize one another via the arbitrary chemistry of their binding sites. In this way, the immune sytem is built from a network of lymphocyte clones, and responds to antigen not only by directly recognizing it, but by internal interactions resulting from which clones see the clones recognizing Ag, which clones see those clones, and so on.

This situation is described in figure 1.

On the far left of the diagram we have the stimulating Ag. The antigen's presenting epitopes will be recognized by a <u>subset</u> of clones having chemically complementary paratopes. This subset is designated as subset 1. Immediately above and to the right of subset 1 is the collection of lymphocyte clones with binding region paratopes recognizing the epitopes of the clones in subset 1. These clones form subset 2, and are the lymphocytes which <u>recognize</u> the clones of subset 1. To the bottom right of clonal subset 1 is subset 3. The lymphocytes in this subset are <u>recognized by</u> the paratopes in subset 1. As we move to the right in the diagram, we see that every clonal subset has two immediate neighbors to its right, an upper subset which recognizes it, and a lower subset which is recognized by it. This diagram branches out in this way indefinitely.

Several features of this diagram are especially noteworthy. First, it is clear that immunologic memory in this scheme is not completely determined by the populations of the lymphocyte clones which directly recognize Ag, in contrast to the strict clonal selection theory. Memory also depends on the populations of clones which are linked via the immunological network of clones to the lymphocytes directly detecting Ag.

Secondly, note in the diagram the arrow from set 1 to set 3. Subset 3 is the collection of clones recognized by subset 1. Now subset 1 has paratopes which, by definition, recognize the epitopes on the presenting Ag. Since the paratopes of subset 1 also recognize the epitopes of subset 3, and we have bounded the number of possible epitopes, there must be considerable overlap between the epitopes of subset 3 and the epitopes of the antigen. So we see that subset 3 serves as an internal image of the antigen within the immune system.

Our last comment on the diagram is word about the arrows. In the figure they are only that, arrows. They may, however, take on a variety of biologically meaningful mathematical forms in actual modelling of the network. The next section will discuss Richter's interpretation of the arrows as non-linear sigmoidal functions, and the conclusions he thereby draws about network dynamics. In all the discussion after that, we will take the arrows to be matrices relating the set of lymphocyte clones with itself.

(Note that if each arrow is the 0-1 matrix R, then the reversed arrows become ^{T}R, the reversal of the graph R represents. In this way, a relationship between one point in the diagram and another becomes the matrix obtained by taking a matrix product of R's and ^{T}R's. For example, how set 2 sees set 6 is described by the word RR^TR, read from left to right.)

3. Explanations with the network theory

How does the network theory explain the enhanced survival of kidney grafts in patients with multiple blood tranfusions? Jerne was the first to point out that if the immune response is that of a network of clones, not just independent populations, then one ongoing response may mask or prevent the initiation of a second response. A patient receiving blood transfusions is constantly mounting immune responses against them. This may then perturb the immune network and effectively shield the kidney graft from a new response.

A far more rigorous approach is Richter's model of low-zone high-zone tolerance using the network concept [F]. Following Jerne, we make the assumption that when a clone recognizes, its population is enhanced, and that when a clone is recognized, its population is suppressed. We also simplify the branching network diagram to a linear one containing only left-pointing arrows (see figure 2). The arrows are taken to represent functions describing how one clone suppresses (and is enhanced by) another one, and the functions used are steep sigmoid shaped curves around some threshold value. Richter wrote down the non-linear dynamical equations governing this system, and obtained numerical solutions to them. We qualitatively describe his results.

Low-zone tolerance. When a low dose of Ag is administered, this triggers clonal subset 1 to increase its population. This increase is always sufficient to trigger subset 2's expansion. Since the dose is low, subset 2 does not increase enough to trigger population 3. We therefore end up with a high number of subset 2 clones, which act to suppress subset 1, driving its population down to a small number. Should Ag be introduced again at a later time, there aren't enough clones in subset 1 to initiate an immune response against it. Thus the immune network has become insensitive to Ag, and we have low-zone tolerance.

Normal immune response and memory. Given a moderate dose of Ag, subset 1 recognizes it and increases its population. This stimulates subset 2 to grow. Since there is a moderate amount of Ag, there is enough driving force in the coupled equations to trigger subset 3, which recognizes subset 2, to expand. There is not enough Ag, however, to trigger subset 4. This results in a high level of subset 3, which suppresses subset 2, the suppressor of subset 1. Since the suppressor of subset 1 is reduced, subset 1 can greatly increase its population. Subsequent introduction of Ag causes a pronounced immune response because subset 1, which directly recognizes Ag, is now at a high level.

High-zone tolerance. We go through the diagram now with a very large dose of antigen. There is enough driving force now from the Ag now to trigger subset 4, but not subset 5. Subset 4 suppresses subset 3. This frees subset 2 to proliferate. High levels of subset 2 now completely suppress subset 1. Thus the direct recognizers of Ag (i.e., subset 1) are not present when Ag is introduced at a later time. So the network has learned to forget Ag, whence high-zone tolerance.

How good is this approach? At each network level in the above explanation, what is the likelihood that a clone in subset k indirectly recognizes Ag via some path through the lower levels? If it turns out that by the fourth level virtually every clone is detecting Ag via the network, then the above description of high-zone tolerance might need some modification. We begin to look at such questions in the next section.

4. Graph theoretic approach

We now begin to look at the problem of which clones recognize (and are recognized by) which other clones. Assume for the moment that we have this information; how do we represent it and what do we do with it?

Say X is the set of clones which recognize Ag directly and Y is the lymphocyte set recognizing clones in X. If we arrange the elements of these sets as in figure 3, as two lists with arrows directed from clones in set Y to clones in X, we obtain a graph of the information. The elements of sets X and Y form the vertices of this graph and the arrows are the directed edges. Note that there may be more than one edge entering or leaving a given vertex. Alternatively, we can represent the information of which elements in Y recognize (and suppress) which elements in X by a matrix. We do this by defining the entries in the matrix R as

$r_{ij} = 1$, if clone $i \in X$ is recognized by $j \in Y$, and
$r_{ij} = 0$, otherwise.

We will use both graphs and matrices in the work below.

So far we have represented how set Y recognizes set X. If set X contains the Ag-recognizing clones, how do we represent the Ag? Recall from the Introduction that Ag is operationally defined by the immune system as the subset of (B-cell) clones which chemically recognize it. In a graph, then, we can label the clones of X which have affinity for

Ag as distinguished vertices. In questions of which elements of the clonal network are connected to which others, we can then restrict the first level (subset 1) to these distinguished vertices.

Similarly, with the matrix notation X is an ordered set. We can therefore view subsets of X as vectors :X->{0,1}. To determine which elements of Y suppress Ag-recognizing clones in X, we then need only multiply the Ag vector times the recognition matrix R. If we use the Boolean rule 1+1=1, and ordinary multiplication, we then obtain the characteristic vector for the recognition subset in Y.

In a real immune system, Ag is recognized by B-cells, which in turn are recognized by helper-T-cells, both of which are recognized by suppressor-T-cells. These suppressor cells are further recognized by other suppressor cells. Some Ags, such as viruses and tumor cells, are directly recognized (and eliminated) by T-cells. It makes sense then to speak of a stricly <u>suppressor network</u>, as Richter analyzed above.

Since we are now assuming that all clones are suppressors, it is reasonable to ask: if clone y suppresses clone x, need clones x and y belong to the <u>same</u> set of suppressors? Arguments can be made both ways (as we do below), so we will use various techniques to analyze both cases.

We first assume that all the supressor clones belong to the same set X. This means that at each network level of antigen recognition the same cells are suppressing and counter-supressing each other. In a sense, this version of the network theory holds that the network of clones generates immunologic memory by "introspectively" looking at itself.

Since all network clones are assumed to belong to a single set X, subsets of clones are seen by other subsets of clones by the map $R:2^X \to 2^X$. That is, for $A \subset X$, the set of clones which sees A is some some subset $B \subset X$. This function R, as mentioned above is representable by a matrix or as a graph. The second level of the network, i.e., which clones sense the clones which see the Ag-recognizing clones, is then $R(R(A)) = R^2(A)$. Under Boolean matrix multiplication, this is representable by the square of the matrix R (times vector A). Similarly, the k-th network level is can be described by the matrix R^k. We will apply some results of Erdos and Renyi to analyze this case later on.

We also have the other situation in which each network level has its own set of suppressor clones. This corresponds physiologically to the idea that suppressor cells come in specialized subsets, and that suppression requires a second cell surface signal in addition to epitope recognition.

(The network level of a given T-cell clone is then determined by the particular non-epitope recognition molecules on its surface.)

Each level then has its own subset X_k of suppressor clones. X_1 corresponds to the Ag-recognizing clones, X_2 to the clones recognizing X_1, and so on. Between every consecutive pair of levels is the function $R^{(k)}: X_{k+1} \to X_k$. These functions can be represented as matrices, and their composition is equivalent to Boolean matrix multiplication.

5. Random graphs

Since we want to know the connectedness of the immune network, we are looking at it as a graph. If the number of clones n were relatively small, we could examine the connectedness of the network graph directly. We would do this by computing the powers R^k of the connection matrix R, and thereby obtain each k-th step connectivity. For small n, even if the exact connections between suppressing clones were not exactly known, from even a little data we could estimate the average number N of 1's (connections) in the matrix R. We could then look at the k-th powers of each of the $\binom{n^2}{N}$ 0-1 matrices with N 1's, and average the results together to get an estimate (with variance) of the k-th step connectivity.

Unfortunately n is not small. Estimates of the number of distinct Ag combining sites in the immune system range from 10^6 to 10^7 clones. All is not lost however: recall that the chemistries of Ag binding sites are decoupled from their encoding DNA. As we have argued in Chapter I and in the Appendix, it is then reasonable to describe recognition of an epitope by some paratope as a _random variable_.

We would like an estimate of the probability p that one T-cell clone recognizes another. There are many experiments in the immunological laboratory (such as CML-Cell Mediated Lymphocytotoxicity and MLR-Mixed Lymphocyte Reaction) which measure the extent to which one lymphocyte recognizes another. This data is usually obtained for cell surface recognition structures other than Ag binding sites. (For example, histocompatibility matching tests for organ transplants test lymphocyte response to the HLA transplantation antigens on donor cells.)

To obtain this probability p, such an experiment should be performed on the lymphocytes of an individual _against his own lymphocytes_, rather than against those of another person. Because an immune system will not recognize self-Ags, all the responses will be of paratopes versus the binding site epitopes. The number of positive responses, divided by the total number of experiments, estimates the probability of one T-cell recognizing another.

These experiments will be carried out in the next year or two, after the appropriate T-cell clones are generated [D]. However, we would still like to have some estimate of interclonal recognition to use here. To get this probability, we will exploit the fact that the Ag-binding site of the lymphocyte may be regarded as just another antigen.

Jerne's network theory holds that the Ag binding site of an immune receptor not only has paratopes for Ag binding, but also *presents epitopes* which can be recognized by other lymphocytes. These T-cell epitopes are ordinary quantum chemical protein moieties, and they are recognized by the immune system in exactly the same way as foreign antigens. (The difference between these epitopes and foreign epitopes is the action taken by the immune system subsequent to recognition.) Knowing the probability of a T-cell clone recognizing a typical Ag will therefore provide us with an estimate of p, the probability that one clone recognizes another.

We estimate the values of n,p and μ needed in the next section for young and old immune systems. The average number of B-cell populations appearing in an immune response of neonatal mice is about 10 [C]. Assuming that the Ag binding properties of T-cells is similar to those of B-cells, we can set $\mu = 10$. Since there are approximately $n = 10^6$ clones in the young immune system [J ii)], we will take p to be $10/10^6 = 10^{-5}$.

In Chapter I we argued that the number of clones increases over a lifetime to $n=10^7$ clones. Since $p=10^{-5}$ does not change with time, we take $\mu = np = 100$ as an estimate of the average number of clones directly recognizing Ag in the aged immune system.

6. Random graphs - independent populations

We now explore the situation in which the set of T-cell clones is partitioned into subsets $A_1, A_2, \ldots$. (Recall that this is a functional partitioning based on cell surface molecules other than the Ag binding site. This allows the same Ag combining site to appear in the clones of more than one equivalence class.) For every pair of sets A_k and A_{k+1}, we have the set relation $R^{(k)} : 2^{A_{k+1}} \to 2^{A_k}$ which tells which clones in A_k are suppressed by which clones in A_{k+1}. As noted above, each $R^{(k)}$ has a 0-1 matrix representation, and composition of relations is equivalent to Boolean matrix multiplication.

The key feature of this model is that the matrices $R^{(k)}$ are <u>independent</u>. This will permit us to use some properties of probabilistic independence in the analysis. We first calculate the probability of one clone recognizing another indirectly via one intermediary (a second network level). After that, we will look at recognition via k intermediate clones (the k+1st network level) by computing the mean and variance of the expected number of paths of length k+1.

6'. Random graphs. independent populations. one step multiplication

The combinatorics of multiplying k nxn matrices together is difficult. If each matrix $M^{(q)}$ has entries $m^{(q)}_{\alpha,\beta}$, $1 \leq \alpha, \beta \leq n$, a typical entry in the product of the M 's has the form

$$\sum_{\alpha_1=1}^{n} \sum_{\alpha_2} \cdots \sum_{\alpha_{k-1}} m^{(1)}_{i\alpha_1} m^{(2)}_{\alpha_1\alpha_2} \cdots m^{(k)}_{\alpha_{k-1}j}, \quad 1 \leq i, j \leq n.$$

Multiple appearances of the terms $m^{(q)}_{\alpha\beta}$ in the summands make this a sum of highly non-independent variables. Observe that in the case of k=2, however, all the summands happen to be independent.

Let M be an nxn matrix whose entries are independent random variables defined as follows:

$$m_{ij} = 1, \text{ with probability } p, \text{ and}$$
$$= 0, \text{ with probability } 1-p, \text{ where } 0<p<1.$$

We call M a *random nxn matrix of probability p*. Let N be a different random nxn matrix of probability p' which is entirely independent of M. Then we may ask what the probability of a non-zero entry in the product of M and N is.

Proposition. The product under Boolean multiplication MN is a random nxn matrix of probability $1-(1-pp')^n$ whose entries are not independent.

Proof.

We first note that $(MN)_{11}$ and $(MN)_{12}$ both contain all the entries of the first row of M, hence the entries of MN are not independent.

Put $s_{ij} = \sum_{\alpha=1}^{n} m_{i\alpha} n_{\alpha j}$.

Observe that $(MN)_{ij} = 1$ iff $s_{ij} > 0$.

$\text{Prob}\{s_{ij} > 0\} = 1 - \text{Prob}\{s_{ij} = 0\}$.

$$\text{Prob}\{s_{ij}=0\} = \text{Prob}\{m_{i1} n_{1j} = 0,\ m_{i2} n_{2j} = 0, \ldots, \text{ and } m_{in} n_{nj} = 0\}.$$
$$= \text{Prob}\{m_{i1} n_{1j} = 0\} \times \ldots \times \text{Prob}\{m_{in} n_{nj} = 0\},$$

by independence of the $m_{i\alpha} n_{\alpha j}$.

$\text{Prob}\{m_{i\alpha} n_{\alpha j} = 0\} = 1 - \text{Prob}\{m_{i\alpha} n_{\alpha j} = 1\}$.

$\text{Prob}\{m_{i\alpha} n_{\alpha j} = 1\} = \text{Prob}\{m_{i\alpha} = 1 \text{ and } n_{\alpha j} = 1\}$,

$$= \text{Prob}\{m_{i\alpha}=1\} \times \text{Prob}\{n_{\alpha j}=1\},$$

by independence of $m_{i\alpha}$ and $n_{\alpha j}$,

$$= pp'.$$

Substituting back in, we obtain

$\text{Prob}\{s_{ij}>0\} = 1- (1-pp')^n$. //

Corollary. If M and N are independent nxn random matrices of probability p, then their product under Boolean multiplication is an nxn random matrix of probability $1-(1-p^2)^n$ (with dependent entries).

We investigate the properties of the function $y= f(x)= 1-(1-x^2)^n$ which takes $[0,1]$ into itself. Note that $f(0)= 0$, and $f(1)= 1$.

The derivative $f'(x)= -n(1-x^2)^{n-1}(-2x)$,

$$= 2n(1-x)^{n-1}(1+x)^{n-1}x.$$

So $f'(x) >0$ on $(0,1)$ and is 0 at the endpoints 0 and 1. Thus $f(x)$ is a monotonically increasing function of the interval onto itself.

Differentiating $f'(x)$ gives us the second derivative

$$f''(x)= 2n[(n-1)(1-x^2)^{n-2}(-2x)x + (1-x^2)^{n-1}]$$

$$= 2n(1-x^2)^{n-2}[(n-1)(-2x^2) + (1-x^2)]$$

$$= 2n(1-x^2)^{n-2}[1 - (2n-1)x^2].$$

Setting $f''(x)$ equal to 0, we find that the sole inflection point of $f(x)$ on $(0,1)$ occurs when $x^2 = 1/2n-1$. That is, $x_i = 1/\sqrt{2n-1}$. From the formula for $f''(x)$ we have that $f(x)$ has positive curvature for $0<x<x_i$, and has negative curvature on $x_i<x<1$. Substituting x_i back into $f(x)$, for large n we obtain

$$
\begin{aligned}
f(x_i) &= 1- (1-x_i^2)^n ,\\
&= 1- (1-1/2n-1)^n ,\\
& 1- (1-(1/2)/n)^n ,\\
& 1- e^{-1/2}.
\end{aligned}
$$

The curvature of $f(x)$ changes only once on $(0,1)$, so the function has exactly one fixed point x_f satisfying $x=f(x)$. We approximate x_f using $\log(1-x) \sim -x$ for small x. (This is legitimate because we know that for large n, since $x_i \to 0$ while $f(x_i)$ remains constant, x_i is to the right of the curve crossing. Hence $x_f < x_i$, which goes to 0. Since x_f is small, it is appropriate to use the logarithm approximation.)

$$
\begin{aligned}
x &= 1- (1-x^2)^n ,\\
1-x &= (1-x^2)^n ,\\
\ln(1-x) &= n \ln(1-x^2),\\
x &\sim n\, x^2,\\
x &\sim 1/n.
\end{aligned}
$$

So we have a general picture of the behavior of $f(x)$

(see figure 4). For large n, $f(x) < x$ if $x < x_f$, and $f(x) > x$ if $x > x_f$. What does this tell us about the immune network? In a young immune system, $n=10^6$, $\mu=10$ and $p=10^{-5}$. Since $p > 1/10^{-6}$, the probability f(p) of one clone recognizing another through some other clone is greater than p. We calculate

$$
\begin{aligned}
f(p) &= 1-(1-p^2)^n \\
&= 1-\left(1-\frac{\mu^2/n}{n}\right)^n, \text{ since } p=\mu/n, \\
&= 1-\left(1-\frac{(100)(10^6)}{n}\right)^n \\
&\sim 1-e^{-.0001} \\
&\sim .0001.
\end{aligned}
$$

This predicts that at the second network level $(.0001)(10^6) = 10^2$ clones recognize an Ag-recognizing clone.

This value of 100 clones tells us that by the second network level, there is more redundancy in which clones suppress the suppressors of the Ag-recognizing T-cells. (A more precise formulation of information and redundancy will be presented in Chapter IV when we treat the rank of a 0-1 matrix.)

In an aged immune system, $n=10^7$, $\mu=100$ and $p=10^{-5}$. As we calculate later in this chapter:

$f(p) = .001$, and

$f(p)n = (.001)(10^7) = 10{,}000$ clones.

Thus there is considerably more redundancy and information loss in the old immune system.

We would expect the ability of the network to distinguish T-cells to be even less at the third and fourth network levels required for the normal immune response and for high-zone tolerance. What does our equation f(x) tell us about these cases?

It would be convenient if we could analyze f(x) as an iterated map of the interval. That is, for x in (0,1), examine the sequence [x,f(x),f(f(x)),...]. This would give us the probabilities of immune recognition at the 1st, 2nd, 4th, 8th, etc. network levels. It is then clear from the properties of f(x) discussed above that the sequence tends to 0 for $x<x_f$ and tends to 1 for $x>x_f$. The amount of information (i.e., the non-redundancy as measured by the closure operation discussed in the next chapter) in the network would therefore drop to 0 as we proceeded to further levels.

Unfortunately, we can not do this. The random matrix obtained by multiplying two independent random matrices together has <u>dependent</u> entries, so the above proposition does not apply and iteration is not permitted. We therefore turn our attention now to arithmetic (not Boolean) sums of these 0-1 matrix products. We will calculate the expected value of an entry in a product matrix and its variance. In determining whether an entry equals or is greater than zero, we will find that the picture suggested by the iterated map analysis is qualitatively correct.

6''. Random graphs. independent populations. calculation of mean and variance.

Let $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ be k independent random nxn matrices each of probability p. A typical entry in the product matrix is $S_k = \sum_{\alpha_1,\ldots,\alpha_{k-1}}^{n} X^{(1)}_{i\alpha_1} X^{(2)}_{\alpha_1\alpha_2} \cdots X^{(k)}_{\alpha_{k-1}j}$. Put $W_\alpha = X^{(1)}_{i\alpha_1} X^{(2)}_{\alpha_1\alpha_2} \cdots X^{(k)}_{\alpha_{k-1}j}$. Since each $X^{(q)}$ is a (probabilistic) 0-1 matrix, an entry $X^{(q)}_{\alpha_{q-1}\alpha_q}$ may be interpreted as describing the occurrence of a path between the points α_{q-1} in A_{q-1} and α_q in A_q, where A_{q-1} and A_q are distinct sets, each of size n. W_α can then be understood to be a random variable describing the probability of the particular path $\alpha = \langle i, \alpha_1, \ldots, \alpha_{k-1}, j \rangle$ occurring (see figure 5).

From this perspective, S_k describes the sum of all possible paths between set 0 and set k. Put $\mu = np = E(\sum_{\alpha_{q-1}} X^{(q)}_{\alpha_{q-1}\alpha_q})$, the expected number of points in A_{q-1} to which each point in A_q is connected. We now calculate the expected value of S_k, an entry in the product matrix. By the preceding discussion, this is the same as the expected number of paths between a point in A_0 and A_k, via the sets $A_1, A_2, \ldots A_{k-1}$.

$$E(S_k) = E(\sum_\alpha W_\alpha)$$

$$
\begin{aligned}
&= \sum E(W_{\alpha}) \\
&= \sum E(X^{(1)}_{i,\alpha_1})E(X^{(2)}_{\alpha_1,\alpha_2})\cdots E(X^{(k)}_{\alpha_{k-1},j}) \\
&= \sum p^{k} \\
&= n^{k-1}p^{k} \\
&= \mu^{k-1}p.
\end{aligned}
$$

We record this as

Proposition. Let $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ be independent random matrices of probability p. Putting $\mu = pn$, the expected value of each entry of the product of the $X^{(q)}$ is then $\mu^{k-1}p$.

We thus obtain immediate qualitative agreement with our previous result on iterated maps. When $p<1/n$, $E(S_k)$ goes to zero as the number (of iterations) k gets large, so we expect to find a 0 with high probability in each entry of the product matrix. Similarly, for $p>1/n$, $E(S_k)$ becomes unbounded and the probability of at least one path (a 1 in the Boolean product matrix) is high.

What are the exact probabilities that $S_k=0$ or $S_k>0$? To get some kind of estimate we need to calculate the variance of S_k. We do this now by computing the sum of the covariances.

From the definition of variance we have that

$\mathrm{Var}(S_k) = E(S_k^2) - E(S_k)^2$. Since we know $E(S_k)$, it remains to compute $E(S_k^2)$. Observe that

$$S_k^2 = (\sum_{\alpha\in P} W_\alpha)(\sum_{\beta\in P} W_\beta)$$
$$= \sum_{\alpha,\beta} W_\alpha W_\beta ,$$

where $\alpha = <\alpha_1, \alpha_2, \ldots, \alpha_{k-1}>$, and

$\beta = <\beta_1, \beta_2, \ldots, \beta_{k-1}>$.

α and β are paths through $A_1, A_2, \ldots, A_{k-1}$, starting at i in A_0, passing through the points α_q, β_q in A_q, and terminating at A_k (see figure 6).

Since the expected value of a sum is the sum of the expected value of each summand, we need to calculate $E(W_\alpha W_\beta)$. Letting P be the product space $A_1 x A_2 x \ldots x A_{k-1}$, α and β are arbitrary paths originating at i in A_0, traversing P, and terminating at some j in A_k. As observed before, by independence $E(W_\alpha) = p^k$. Is there a similar simple expression for $E(W_\alpha W_\beta)$? We develop a usable approximation to it.

Definition. For $\alpha_q, \beta_q \in A_q$, define

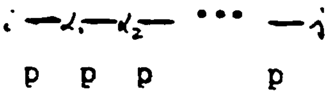
$|\alpha_q - \beta_q| = 0$, if $\alpha_q = \beta_q$, and

1, if $\alpha_q \neq \beta_q$.

Then for $\alpha, \beta \in P$, the distance between α and β is defined to be

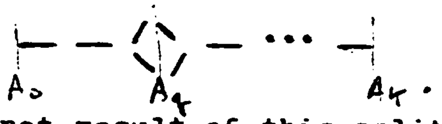
$$|\alpha - \beta| = \sum_{q=1}^{k-1} |\alpha_q - \beta_q| . \qquad //$$

We now classify the summands of $\sum W_\alpha W_\beta$ on the basis of the distances between α and β. Consider the case in which $\alpha = \beta$. Here $|\alpha - \beta| = 0$ and both paths start at $i \in A_0$, go through identical points in product set P, and terminate at point $j \in A_K$. Each link of the path has a probability p of occurring, so the extreme dependencies reduce $E(W_\alpha W_\beta)$ to p^k, where k is the length of the path. Its graph might look like:



where the line segments represent links of the path and each p another probability to be considered.

Observe that in every case, the endpoints are fixed ($i \in A_0$ and $j \in A_K$) and that α and β are defined on, hence can only differ on, interior points of the path. For the next simplest case we consider $|\alpha - \beta| = 1$: this means that somewhere in the interior of paths α and β there is a single set A_q at which α and β differ. The graph of this case looks like:



The net result of this split at set A_q is to add two additional links to the graph of the $|\alpha - \beta| = 0$ case. Thus the probability of this 1-split situation $E(W_\alpha W_\beta)$ is p^{K+2}.

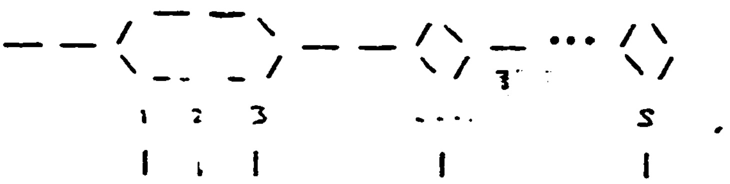
We seek an upper bound on the probability that two paths α and β both connect i in A_0 and j in A_K. Because W_α can only assume the values 0 and 1 (the absence or the presence of a path),

$$E(W_\alpha W_\beta) = 0\text{xProb}\{W_\alpha W_\beta = 0\} + 1\text{xProb}\{W_\alpha W_\beta = 1\}$$

$$= \text{Prob}\{\alpha \text{ and } \beta \text{ both connect } i \in A_0 \text{ and } j \in A_k\}$$

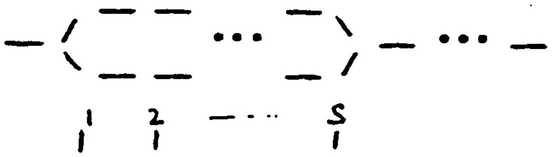
Such an upper bound will then provide an upper bound on the variance.

Given s splits in the $\alpha - \beta$ path graph of some case (i.e., $|\alpha - \beta| = s$), the diagram looks like:



This is a graph of the two paths, each of length k. At the sets A_q where $\alpha \neq \beta$ a split occurs and this is pointed out underneath the graph by vertical bar. If we just trace along the bottom path, since the paths are of length k we have k links. Each link has probability p, so it is immediate that the probability of the two paths W_α and W_β is at least p^k. By counting the number of links in the top path which have not been counted already in the bottom path, we find the <u>number of additional links</u> which must be added to k. The sum total of all the links gives the power to which p must be raised to in order to obtain $E(W_\alpha W_\beta)$.

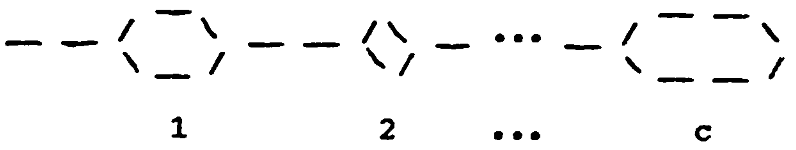
Consider the case in which all the splits are contiguous. Its graph is:



Say there are s splits, depicted above by vertical bars. The total number of links is then k+s+1. We claim that for s splits, this number is minimal.

Lemma. Let α and β be two paths in $P=A_0 x \ldots x A_k$ which share common initiating and terminating points in A_0 and A_k. If α and β differ from each other in exactly s places (i.e., there are exactly s splits), and if there are precisely c distinct regions of contiguous splits, then the combined graph of the two paths contains a total of k+s+c links.

Proof. The combined graph of and looks like:



The numbers under the graph label the contiguous regions. If there are no splits, the two paths are identical and there are only k links.

If there is exactly one contiguous region of splits, then as discussed above the number of links is k+s+1: k links for the combined bottom path, and s+1 for the non-redundant top links (1 to go up, s-1 across, and 1 to come back down).

Say there are exactly two contiguous regions of top links, with s_1 splits in the first and s_2 splits in the second,

$s_1+s_2=s$. Since each region has 1 link to go up to the first split, s_h-1 links across the top, and 1 link to go back down, a region contains s_h+1 top links, h=1,2. The number of top links for two regions is then $s_1+1 + s_2+1 = s+2$. Say there are exactly c distinct contiguous regions. Reasoning as above, there are $s+c=\sum_{h=1}^{c}(s_h+1)$ non-redundant top links. Hence there are a total of k+s+c links in the graph. //

To calculate $E(W_\alpha W_\beta)$ we need to know the power to which link probability p is raised to (the number of links) for each pair α,β. The combinatorics of the distribution of the number of contiguous regions for s splits is tedious. The above lemma gives us a useful bound, however. In all but one case ($\alpha=\beta$) there is at least one split. If s>0, then there is at least one contiguous component and c>0. So for s>0, the number of links $k+s+c \geq k+s+1$. Thus the probability $E(W_\alpha W_\beta) = p^{k+s+c} \leq p^{k+s+1}$. This proves the

<u>Corollary</u>. If $|\alpha-\beta| = s$, $\alpha\neq\beta$, then $E(W_\alpha W_\beta) \leq p^{k+s+1}$.

We now compute the variance.

$$\begin{aligned}
E(S_k^2) &= \sum_{\alpha,\beta\in P} E(W_\alpha W_\beta)\\
&= \sum_{s=0}^{k-1}\Big[\sum_{|\alpha-\beta|=s} E(W_\alpha W_\beta)\Big]\\
&= \sum_{s=0}^{k-1}\sum_{|\alpha-\beta|=s} E(W_\alpha W_\beta).
\end{aligned}$$

How many paths are there such that $|\alpha-\beta|=s$? Let

$Q=A_{q_1}\times A_{q_2}\times\cdots\times A_{q_s}$ be the sub-product set of P in which $\alpha_q \neq \beta_q$ (split). The endpoints in A_0 and A_k are fixed and equal. In choosing the first path, there are n choices for each of the k-1 interior sets in P. In selecting the second path, there n-1 choices remaining for each of the s sets in Q where $\alpha_q \neq \beta_q$. For every configuration Q then, there are $n^{k-1}(n-1)^s$ paths.

How many arrangements of Q are there? We can arrange s objects (the indices of Q) in k-1 places in $\binom{k-1}{s}$ ways. So

$$\sum_{|\alpha-\beta|=s} E(W_\alpha W_\beta) = n^k\ (n-1)^s\binom{k-1}{s}E(W_\alpha W_\beta),\ |\alpha-\beta|=s.$$

$$E(S_k^2) = n^{k-1}\sum_{s=0}^{k-1}\sum_{|\alpha-\beta|=s}(n-1)^s\binom{k-1}{s}E(W_\alpha W_\beta)$$

$$= n^{k-1}\sum_{|\alpha-\beta|=0} E(W_\alpha W_\beta) + n^{k-1}\sum_{s=1}^{k-1}\sum_{|\alpha-\beta|=s}(n-1)^s\binom{k-1}{s}E(W_\alpha W_\beta)$$

Applying the above corollary,

$$\leq n^{k-1}p^k + n^{k-1}\sum_{s=1}^{k-1}(n-1)^s\binom{k-1}{s}p^{k+s+1}$$

$$< n^{k-1}p^k + n^{k-1}p^{k+1}\sum_{s=1}^{k-1}n^s\binom{k-1}{s}p^s.$$

Setting $\mu = np$ as before,

$$= \mu^{k-1}p + \mu^{k-1}p^2\sum_{s=1}^{k-1}\binom{k-1}{s}\mu^s.$$

Therefore

$$\mathrm{Var}(S_k) = E(S_k^2) - E(S_k)^2$$

$$\leq [\mu^{k-1}p + \mu^{k-1}p^2\sum_{s=1}^{k-2}\binom{k-1}{s}\mu^s + \mu^{k-1}p^2(\mu^{k-1})] - (\mu^{k-1}p)^2$$

$$= \mu^{k-1}p + \mu^{k-1}p^2\sum_{s=1}^{k-2}\binom{k-1}{s}\mu^s$$

$$= \mu^{k-1}p(1 + p.\sum_{s=1}^{k-2}\binom{k-1}{s}\mu^s).$$

We can show that $\sum_{s=1}^{k-2}\binom{k-1}{s}\mu^s \leq (k-1)(1+\mu)^{k-2}$:

$$\sum_{s=1}^{k-2}\binom{k-1}{s}\mu^s \leq c_k\sum_{s=1}^{k-2}\binom{k-2}{s}\mu^s$$

$$\leq c_k(1+\mu)^{k-2},$$

where $c_k \equiv \max_{1\leq s\leq k-2}\binom{k-1}{s}/\binom{k-2}{s} = \max_{1\leq s\leq k-2}\frac{k-1}{k-s-1} = k-1.$

Thus we have proved

<u>Theorem</u>. If S_k is the random variable describing the number of paths connecting two points via k nxn random graphs of probability p, letting $\mu = np$, we have that the variance of S_k

$$\begin{aligned} Var(S_k) &\leq \mu^{k-1} p[1+(k-1)p(1+\mu)^{k-2}] \\ &= E(S_k)[1+(k-1)p(1+\mu)^{k-2}]. \end{aligned}$$

The variance is a measure of the dispersion of a distribution about its mean. We now use this measure to calculate a bound on just how much of the distribution mass is concentrated away from zero.

By Chebyshev's inequality, for any random variable X and $m=E(X)$, it is true that

$$P\{|X-m|\geq t\} \leq t^{-2}\, Var(X).$$

Also, observe that if $m=E(S\ )$,

$$P\{S_k=0\} \leq P\{|S_k-m|\geq m\}.$$

Thus we have:

$$\begin{aligned} P\{S_k=0\} &\leq P\{|S_k-m|\geq m\} \\ &\leq m^{-2}\, Var(S_k) \\ &\leq \frac{(\mu^{k-1}p)[1+(k-1)p(1+\mu)^{k-2}]}{(\mu^{k-1}p)^2} \\ &= \frac{1}{\mu^{k-1}p} + \frac{(k-1)p(1+\mu)^{k-2}}{\mu^{k-1}p} \\ &= \frac{1}{E(S_k)} + \frac{(k-1)}{\mu}\left(\frac{1+\mu}{\mu}\right)^{k-2}. \end{aligned}$$

We record this as

<u>Corollary</u>. $P\{S_k=0\} < \frac{1}{E(S_k)} + \frac{k-1}{\mu}\left(\frac{1+\mu}{\mu}\right)^{k-2}$.

6'''. Random graphs. independent populations. connectivity of the immune network.

In this section we apply the results of the preceding sections in order to find out how many clones at each level of the immune network are involved in suppressing the 1-level Ag-recognizing clones. The implications for low-zone tolerance, normal immune memory, and high-zone tolerance are discussed. To accomplish this, we will have to eventually calculate a modified one-step random matrix probability.

We give a brief review. There are n T-cell clones at each network level. An antigen Ag is recognized by about μ clones. The probability of a particular T-cell clone recognizing an Ag is then $p=\mu/n$. From chapter I we have that for a young immune system, $n=10^6$, $\mu=10$ and $p=10^{-5}$. For an old immune system, $n=10^7$, $\mu=100$ and $p=10^{-5}$. The antigen recognizing clones are the 1-level of the network. Kth level clones are assumed to recognize only (k-1)st level clones. These levels are independent of one another. We consider the epitopes on the Ag combining site of the T-cell to behave just like any other Ag, so the numbers n, μ and p have the above values.

When foreign Ag is presented to the immune system, about 10^6 clones chemically recognize it. The particular clones recognizing Ag form a subset which label the Ag in the immune system. In older theories of immunologic memory, this 1-level sufficed. In the network theory, however, we must now ask: how many k-level clones recognize the 1-level clones which recognize Ag?

We define c_k=Prob{a particular k-level clone recognizes Ag}. Our objective is to determine $n_k = nc_k$, the average number of clones in the kth network level indirectly recognizing Ag, k>1. In the section above we introduced the random variable $W_k = X^{(1)}_{i,\alpha_1} \times X^{(2)}_{\alpha_1,\alpha_2} \times \cdots \times X^{(k)}_{\alpha_{k-1},j}$ which describes the event that a path of length k $\alpha = \langle \alpha_1, \alpha_2, \ldots, \alpha_{k-1} \rangle \in P$, the set of such paths, exists between two points. $X^{(q)}_{\alpha_{q-1},\alpha_q}$ is the random nxn 0-1 matrix of probability p relating network levels k-1 and k. We then introduced the random variable $S_k = \sum_{\alpha \in P} W_\alpha$ which describes the number of such paths. How does this help answer the question of caculating c_k?

Observe that a summand $X^{(1)}_{i,\alpha_1} X^{(2)}_{\alpha_1,\alpha_2} \times \cdots \times X^{(k)}_{\alpha_{k-1},j}$, for fixed i and j, may be written instead as the product of a random n-vector of probability p $Y^{(1)}$ and k-1 matrices:

$Y^{(1)}_{\alpha_1} X^{(2)}_{\alpha_1 \alpha_2} x \ldots x X^{(k)}_{\alpha_{k-1} i}$. That is, for fixed i, $X^{(1)}_{i \alpha_1}$ is just a row vector $Y^{(1)}_{\alpha_1}$. <u>All the theorems from the last section apply to this situation</u>. As we have said before, the 1-level (direct Ag recognition) detects Ag with probability p (see figure 7). We hereafter regard the kth level recognition of Ag as the <u>product of a random Ag vector and k-1 random matrices</u>.

What, then, is n_1, the number of clones directly recognizing Ag? Note that c_1 is just p. Taking expected values, $n_1 = E(Y^{(1)}) = np$. So for a young immune system,

$$n_1 = np = (10^6)(10^{-5}) = 10,$$

while for an old immune system

$$n_1 = np = (10^7)(10^{-5}) = 100.$$

We now determine n_2. This will give us the expected number of T-cell clones involved in a low-zone tolerance response. (Recall that the final immune state after low-zone tolerance is induced is a suppression of the populations of level 1 and an enhancement of level 2. We denote this by [S,E], where S means "suppressed" and E means "enhanced".)

Observe that the probability c_2 here is for the event $[0 < \sum_{\alpha_1} Y^{(1)}_{\alpha_1} X^{(2)}_{\alpha_1 i}]$. This is equivalent to the one-step matrix multiplication analyzed in section 6'. The theorem there tells us that $c_2 = 1-(1-p^2)^n$.

Therefore

$$
\begin{aligned}
c_2 &= 1-(1-p^2)^n \\
&= 1-(1-\tfrac{p^2 n}{n})^n \\
&\doteq 1-e^{-p^2 n} \\
&\doteq 1-(1-p^2 n) \\
&= p^2 n.
\end{aligned}
$$

For a young immune system, then,

$$c_2 = (10^{-5})^2 (10^6) = 10^{-10}\ 10^6 = 10^{-4},$$

and $n_2 = nc_2 = (pn)^2 = (10^{-5}\ 10^6)^2 = (10)^2 = 100.$

For an old immune system,

$$c_2 = (10^{-5})^2 (10^7) = 10^{-10}\ 10^7 = 10^{-3},$$

and $n_2 = nc_2 = (pn)^2 = (10^{-5}\ 10^7)^2 = (100)^2 = 10{,}000.$

By the second network level, there are many more suppressor clones stimulated in an old immune system than in a young one.

What can we say about recognition by the third and fourth levels of the network (and the associated regular immune memory and low-zone tolerance, respectively)? To calculate the probability of 3-level recognition we need a slight extension of the one step multiplication technique just used. We defer this until after we have computed the 4-level high-zone tolerance case.

The final state of the immune system after network high-zone tolerance has been achieved is [S,E,S,E]. We want the probability that a 4-level suppressor clone is enhanced after an immune response to a given Ag.

A 4-level clone recognizes Ag exactly when there is a path through levels 1, 2 and 3 connecting it to Ag. So the event that it recognizes, is the event $[S_4>0]$, where $S_4=\sum_{\alpha} Y^{(1)}_{\alpha_1} X^{(2)}_{\alpha_1\alpha_2} X^{(3)}_{\alpha_2\alpha_3} X^{(4)}_{\alpha_3 j}$. In section 6'' we computed the variance of S and used Chebyshev's inequality to obtain the Corollary

$$P\{S_k=0\} \leq \frac{1}{E(S_k)} + \frac{k-1}{\mu}\left(\frac{1+\mu}{\mu}\right)^{k-2}.$$

Therefore the probability of at least one path from the clone to Ag (i.e., detection by any given 4-level clone) $c_4 = 1-P\{S_4=0\} \geq 1-$ this upper bound.

Since $E(S_k)=\mu^{k-1}p$, in a young immune system

$$\begin{aligned} P\{S_4=0\} &\leq \frac{1}{\mu^3 p} + \frac{3}{\mu}\left(\frac{1+\mu}{\mu}\right)^2 \\ &< \frac{1}{(10^3)(10^{-5})} + d,\ d>0, \\ &= 100 + d. \end{aligned}$$

This yields no useful upper bound, since $P\{S_4=0\}<1$. In an old immune system, however, $\mu = 100$ and

$$\begin{aligned} P\{S_4=0\} &\leq \frac{1}{\mu^3 p} + \frac{3}{\mu}\left(\frac{1+\mu}{\mu}\right)^2 \\ &= \frac{1}{(10^6)(10^{-5})} + \frac{3}{100}\left(\frac{101}{100}\right)^2 \\ &= .1 + (.03)(1.01) \\ &< .15. \end{aligned}$$

Therefore

$$c_4 \geq 1-P\{S_4=0\}$$

$> 1-.15$

$= .85,$

or about 85% of the 10 T-cell clones are involved in the suppressive response to Ag. This massive response is not observed experimentally.

We return now the case k=3 and compute the probability of 3-level network recognition of Ag in the regular immune response. Observe that $c_3 = 1-\text{Prob}\{S_3=0\}$.
$S_3 = \sum_{\alpha_1,\alpha_2} Y^{(1)}_{\alpha_1} X^{(2)}_{\alpha_1,\alpha_2} X^{(3)}_{\alpha_2,\delta}$, the probability of each entry in each summand equalling p. Direct computation of this triple product (as in the one-step multiplication case) is very difficult. An analogous computation can be made, however, by reducing the problem to a one-step multiplication with <u>two different probabilities p and p'</u>.

Direct computation of S_3 requires keeping track of non-independent paths of length 3 (see figure 8): one leg from set (level) 0 to set 1, one from set 1 to set 2 and one link from set 2 to set 3. This computation is too difficult for us to do directly, so we use a rougher method to estimate S_3. Instead, we will consider multiplication by the random Ag vector $Y^{(1)}$ to <u>select out</u> a subset A from set 1 of average size μ. We can then compute the probability p' of a path connecting A to a point in set 2 (see figure 9).

Assuming $\#A=\mu$,

$$p' = \text{Prob}\{A \text{ is connected to a point in set } 2\}$$

$$= 1 - \text{Prob}\{A \text{ is not connected to a point in set } 2\}$$

$$= 1 - (1-p)^{\mu} .$$

Since $p=\mu n$, and μ/n is small, we have

$$p' = 1 - (1-\tfrac{\mu}{n})^{\mu}$$

$$= 1 - (1-\tfrac{\mu^2/n}{\mu})^{\mu}$$

$$\doteq 1 - e^{-\mu^2/n}$$

$$\doteq 1 - (1-\mu^2/n)$$

$$= \mu^2/n.$$

We thus obtain

$$c_3 = 1 - (1-pp')^n$$

$$= 1 - (1-(\tfrac{\mu}{n})(\tfrac{\mu^2}{n}))^n$$

$$= 1 - (1-\tfrac{(\mu^3/n)}{n})^n$$

$$\doteq 1 - e^{-\mu^3/n}.$$

In a young immune system, $\mu^3 = (10)^3$ and $n= 10^6$, so

$$c_3 \doteq 1 - e^{-\mu^3/n}, \mu^3/n \text{ small},$$

$$\sim 1 - (1-\mu^3/n)$$

$$= \mu^3/n$$

$$= 10\ /10$$

$$= .001.$$

In an old immune system, however,

$\mu^3 = (10^2)^3$ and $n= 10^7$, so

$$c_3 = 1 - e^{-10^6/10^7}$$

$$= 1 - e^{-.1}$$

$$\doteq 1 - (1-.1)$$

$= .1.$

This would indicate that about 10% of the 3-level T-cells would be expected to be involved in each regular [E,S,E] immune response. As in the case of high-zone tolerance, such an overwhelming response to foreign Ag is not observed experimentally. Such a high fraction of total clonal involvement in aged immune system responses eliminates specific immune recognition of Ag. Different Ags would be stimulating most of the clones at the third and fourth network levels, and the large overlap would destroy the specific basis for immunologic memory.

This loss of information as $p(\text{new}) = \frac{10}{10^6}$ ages to $p(\text{old}) = \frac{100}{10^7}$ is a general property of 0-1 matrices. In Chapter IV we will use a different approach which suggests that immune information is optimized when $p(\text{optimal}) = \log n/n$.

7. Random graphs - single population

In the last section we looked at the immune network under the assumption that each level of the network was composed of an independent collection of T-cell clones, with no repetition of clones between levels. We now examine the case where there is only one population of many distinct clones, with each level of the network being another copy of all the clones.

This is a reasonable point of view. It derives from the experimental observation that there is a large, finite number of distinct clones, each having its own Ag binding site. If there is then further splitting of this population into distinct functional subsets, there is no reason why *the entire population* of the immune repertoire could not be replicated with every subset.

To simplify the mathematics, we assume in this section that *a clone is recognized by another if and only if it recognizes it as well*. An equivalent way of stating this would be that *paratopes are to be identified with epitopes*. Persuasive arguments for this have been made elsewhere [Ro].

The questions we will address in this section do not concern what happens at each level of the network, but instead what *ultimately* happens. The question of connectivity of clones becomes: what are the chances that one clone ultimately is recognized by another, by some path of finite length? Similarly, we may ask: what is the probability that a clone eventually sees itself (i.e., forms a cycle) through the network?

As before, let R be the 0-1 matrix which describes the recognition relationships between the clones. By the assumption above that epitopes and paratopes are identified, R becomes a *symmetric* matrix. R is then the adjacency matrix of an undirected graph. Our problem of eventual connections and cycles thus becomes the mathematical problem of understanding the behavior of R^k as k gets large.

We do not know the exact relations between all the T-cell clones. Even if we did, since n is about 10^6, multiplying the nxn matrix R times itself many times would be a computationally difficult problem. As we have argued before, however, neither Nature nor we know in advance of the empirical test which clones will recognize which. It makes sense then to analyze R as a *random graph*.

Say it is known that about μ clones are recognized by any given clone. Then we would expect our graph to have about $n\mu$ edges. If we proceeded as in the previous section, we would let each of the edges of the undirected graph be an independent random variable with probability $p=\mu/n$ of being present and 1-p of being absent. We will take a different approach, however.

Put $N=[\mu n]$. There are $\binom{\binom{n}{2}}{N}$ different (undirected) graphs on n vertices having N edges. Let this collection be our sample space, assigning to each member an equal probability weight. What can we say about various properties of members of this set selected at random from it when n is large? This analysis has been carried out by Erdos and Renyi [E].

We have $N=[\mu n]$. Depending on how μ behaves as a function of n, we have N(n) as a function of n. Let $\Gamma(n,N(n))$ be the collection of all graphs on n vertices having N(n) (undirected) edges. $\Gamma(n,N(n))$ has $\binom{\binom{n}{2}}{N(n)}$ elements. If there is a property A of a graph $g \in \Gamma(n,N(n))$ such that lim Prob{g has property A} = 1, we say that the random graph $\Gamma(n,N(n))$ has property A almost surely (a.s.).

A graph may be partioned into connected components. (Each component corresponds to an irreducible diagonal block of its adjacency matrix.) Using formulas given in Erdos and Renyi's 1960 paper, we now show that the graph of the network relationships R is expected to consist of a single giant component.

(In usual applications of Erdos and Renyi's results on random graphs, the value of n ranges from 50 to 1,000. On these empirical grounds we feel justified in applying their limit theorems to the immune system, where $n=10^6$ to 10^7.)

7'. Random graphs. single population. connectivity.

Theorem 9b in [E] describes the size of the greatest component of a random graph. Let N(n) be asymptotic to cn, where n is the number of vertices, and let x(c) be the unique root in (0,1) of the equation $xe^{-x} = 2ce^{-2c}$. Then the size of (i.e., the number of vertices in) the greatest component of $\Gamma(n,N(n))$ tends to nG(c), a.s., where $G(c) = 1-x(c)/2c$.

Using a binary search APL computer program, we find that $x(10) = 4.47\text{x}10^{-8}$ and $x(100) = 7.45\text{x}10^{-9}$. Letting $\mu = c$, for $\mu = 100$ we have

$$
\begin{aligned}
G(c) &= 1 - x(c)/c \\
G(100) &= 1 - (7.5\text{x}10^{-9})/100 \\
&> 1 - 10^{-10} \\
&= 1, \text{ relative to } n=10^{7}.
\end{aligned}
$$

Similarly, even for a low estimate of $\mu = 10$,

$$
\begin{aligned}
G(10) &= 1 - (4.5\text{x}10^{-8})/10 \\
&> 1 - 10^{-8} \\
&= 1, \text{ relative to } n=10^{6}.
\end{aligned}
$$

That is, the expected number of clones in the largest connected component of T-cells which recognize each other is G(c)n= n. This is true regardless of the maturity of the immune system.

The above computation shows that the immune network consists of a single giant component, a.s.. Moderate doses of Ag, therefore, are capable of ultimately exciting all the clones.

7''. Random graphs. single population. cycles.

The number of cycles of length 3 in the network graph tells us how many 4-level clones recognize themselves at the Ag-recognizing 1-level. Recall that in high-zone tolerance, the final state of the immune network is [S,E,S,E] (that is, the 1-level clones are suppressed, the 2-level one are enhanced, and so on.) How many of the 1-level clones which are suppressed will turn out to also be 4-level clones which must, paradoxically, be enhanced?

Let $N(n) \sim cn$, $c>0$. Theorem 3a in [E] states that the number of cycles of length k in the random graph $\Gamma(n,N(n))$ has a Poisson distribution in k, with mean $(2c)^k/2k$. Set $c=\mu$. In an old immune system, $\mu=100$, and

$$\begin{aligned} E[\text{number of 3-cycles}] &= (2\text{x}100)^3/2\text{x}3 \\ &= 8\text{x}10^6/6 \\ &> 10^6 \\ &= .1\text{x}n. \end{aligned}$$

So by the fourth level of the immune network, we would expect over 10% of the clones which recognize Ag to also recognize themselves via network levels 2 and 3 (see figure 10).

In high-zone tolerance [S,E,S,E], the tolerance suppression S of level 1 is achieved by the enhancement E of 4-level clones. If the 1-level suppressed clones are also enhanced as 4-level clones by the network, this would indicate that the aged immune system has an unstable response to Ag. For this immune response to function properly, cellular regulatory mechanisms outside the network interactions must be invoked.

On the other hand, in the young immune system $\mu = 10$, and

$$
\begin{aligned}
E[\text{number of 3-cycles}] &= (2\text{x}10)^3 / 2\text{x}3 \\
&= (8/6)\text{x}10^3 \\
&<< 10^6 = n.
\end{aligned}
$$

Since the probability of an Ag-recognizing clone also being a suppressing 4-level clone is then negligible, there is no enhancing clonal feedback and the network functions properly.

III. Specificity labelling

1. The problem

An insulin hormone binds to a liver cell surface and transmits the message that the blood sugar is elevated; the phospho-fructo-kinase enzyme recognizes, transforms and directs a 6-P-glucose molecule into the glycolytic pathway. A transfer-RNA's three base codon specifically recognizes a complementary three base signal on a messenger-RNA and then contributes its amino acid to a nascent protein chain.

These are several of the tens of thousands of known examples of specificity in biology. Specificity is central to the transmission and processing of biological signals and materials. We will concern ourselves in this section with the role of specificity in the transmission of information in biology, particularly in immunology and the immune system.

How are specific signals transmitted in Nature? All such processes use chemical recognition. (Even the nervous system, which is often thought of primarily as an electrical system, has an underlying chemical basis.) Communication is initiated between or within cells when one chemical configuration combines with a complementary one with a sufficiently high chemical affinity. The signal is usually transmitted by some change of chemical state in the non-recognizing part of one or both of the molecules. Our interest here will be restricted to recognition between complementary chemical configurations.

When two configurations combine, we will assume that the combination results from matching of complementary three dimensional localized quantum chemical regions which we will call specificities. For example, when an insulin molecule binds to its receptor, we may interpret this as the binding of two or three small chemical patches on the molecule locally binding to complementary patches on the receptor. When the insulin molecule or receptor evolves, some small change in the binding affinity of one of these specificities with its partner would be expected.

We would like to uncover these physiologic "specificities". Presumably they tell us something about the history and mechanism of various biological processes. That is, each instance of biological communication is presumed to be due to one or more underlying specificities. We therefore define the specificity to be the unit of biological information. The more specificities an organism's subsystem uses, the greater we would expect its complexity to be.

In what follows, we will present a method of obtaining underlying specificities which has been used recently in the field of immunogenetics. We will extend the method to some new cases and prove that, in fact, these are the only cases to which it can be extended. We will also look at it probabilistically, and use this as a measure of information in the immune system.

2. Immunogenetics

Genetics is the branch of biology which traces the inheritance of chromosomal genes (genotypes) and the physiologies and morphologies derived from those genes (phenotypes). Cellular biologists are interested in the genes coding for cell surface molecules, since these molecules are responsible for intercellular communication in the normal and diseased states. As these genes undergo random evolutionary change in a population, the surface molecules they code for are altered somewhat in their quantum chemical configuration. This phenotypic change (i.e., binding properties of the gene product) can be measured by test molecules which can either bind or not bind to the surface molecule. Thus the problem of the genetics of cell surface molecules can sometimes be reduced to a panel of yes-no type binding experiments; all we need are the test molecules.

When an animal mounts an immune response against some foreign antigen, its blood will contain large concentrations of antibody (Igs) which have a high binding affinity for the Ag. These Igs represent the dozens of B-cell clones expressing Ig which recognize Ag, that is, the antibody obtained is a pooled mixture of different Igs. If we let the Ag be some biochemically purified cell surface molecule we are interested in, we thereby obtain a test blood serum for the surface molecule.

This method of tracing cell surface phenotypes by using immune anti-sera induced against a surface product of interest as the panel of test reagents is called immunogenetics. A single test consists in observing whether or not a test serum reacts with a sample of cells from the individual whose phenotype we are investigating. If there is a positive reaction, we assume that this is because a specificity on the individual's cell surface was recognized by a complementary anti-specificity present on an Ig in the immune serum. If one individual's cells react with the serum, but another's does not, we say that the former group of cells contains a specificity not contained in the latter.

(A warning, however: most biochemical changes, such as in "specificities", do not lead to functional changes [Wil],[K]. Most observed differences in immunogenetic specificities are of no physiological significance to the organism being studied.)

A number of different test sera may be reacted against a variety of individual's cells. In this way we obtain a table of 0's and 1's which record the results of the test: the (i,j)th entry in the table is a 1 if the ith test serum reacted with the jth individual's cells, and is 0 otherwise. The game is to uncover the specificities which can explain the pattern of binding observed in the reaction table.

A related question is: how many specificities are needed to account for the reaction table? In particular, we can ask what the <u>minimum number</u> r of specificities required is. If the reaction table correlates with some physiological parameters, this minimum value gives a lower bound on the number of distinct receptor chemistries which underlie the cellular interactions. In this way, r is a measure of the complexitiy of the particular system, and can therefore be used as a measure of <u>information</u>.

3. Specificity labelling

In figure 11 we give an example of a reaction table of two test sera against the cells of three different individuals. The first row of the table indicates that groups 1 and 3 had a positive chemical reaction with the first test serum, while the group of cells from individual 2 did not react. Similarly, we can regard the second row of the table as a 0-1 row vector which records the results of test serum 2 against the 3 individuals. (Alternatively, we could have chosen to view the table as a collection of 3 0-1 column vectors, each of length 2.)

Figure 12 gives a labelling of this reaction table. The rule is: the (i,j)th entry equals 1 only when there is some specificity s such that s labels the ith row of the table and $\bar{s}$ labels the jth column. We will regard the table as a 0-1 matrix. In general, the labellings are not unique, even accounting for permuations of the labels s_q.

What would happen if we pooled together test sera 1 and 2? We would then be reacting specificities s_1 and s_2 against the cells $\langle \bar{s}_1, \bar{s}_2, \bar{s}_1 \bar{s}_2 \rangle$. Since there is then an $s\bar{s}$ combination, $\bar{s} = \bar{s}_1$ or $\bar{s}_2$, for each group of cells, by the above rule the row vector in the matrix would be <1,1,1> (see figure 13). In other words, the adding together of rows is a Boolean (or set) operation; 1+1=1, for each row vector entry.

There is an easy upper bound on the minimal number of specificities required to label a reaction matrix. If the matrix has m rows and n columns, set q=min{m,n}. Then for, say, $m \leq n$, we can label the m rows of the matrix by specificities $s_1, s_2, \ldots, s_m$. We then label the columns by the rule: column j is labelled by specificity $\bar{s}_i$ precisely when entry (i,j)=1. An example is given in figure 14.

The number of specificities required to label a reaction matrix is less than min{m,n} whenever a row (or column) can be written as a *Boolean combination* of other rows (columns). For example, the 3x3 0-1 matrix in figure 13 can be labelled with only 2 specificities because the third row can be written as a Boolean combination of rows 1 and 2. Specifically, <1,1,1> = 1x<1,0,1> + 1x<0,1,1>.

In general, if we knew something about the space generated by all the Boolean combinations of the rows (columns), we might be able to draw some conclusions about the minimal number of specificities. What we need to know turns out to be the size of the *smallest Boolean lattice* into which the row space lattice can imbed. The number of generators of this Boolean lattice is then the minimal number of specificities. (A Boolean lattice on q generators is equivalent to the power set of a set of size q, ordered by set inclusion.)

The proof of the remarks in the preceding paragraph may be found in [M ii)]. An algorithm has been given for computing the row space [Wo i)],[N]. After some notational remarks, we will present and discuss Wohlgemuth's form of the algorithm. We will develop a probabilistic approach to the lattice and apply it to the network theory of the immune system. We will then extend the method to reaction tables with entries different from 0 and 1.

3'. Specificity labelling. Relation notation.

A 0-1 table or matrix may be viewed as a relation between two sets. We give some ways of writing this down.

Let $X:I\to\{x_i\}$, $x(i)=x_i$, be an ordered set , $I=\{1,2,\dots,m\}$. Similarly, let $Y:J\to\{y_j\}$, $y(i)=y_j$, be another ordered set, $J=\{1,2,\dots,n\}$. A relation is a subset of the Cartesian product XxY. We say that x is related to y, xRy, whenever $(x,y)\in R$.

We define the transpose or reflection of R, ${}^{T}R$, to be the subset of YxX such that $y\,{}^{T}R\,x \Leftrightarrow xRy$. All remarks made below about R also hold for ${}^{T}R$. Similarly, statements about row vectors can be reinterpreted for column vectors.

We may write R as the function $R:X\to 2^Y$, $R(x)=\{y\in Y \mid xRy\}$. We may extend R to the set function $R:2^X\to 2^Y$, $R(A)=\{y\in Y \mid xRy \text{ for some } x\in A\}$, where $A\subset X$. This extension is said to preserve unions, since $R(A\cup A') = R(A)\cup R(A')$.

We may also consider R as the matrix defined by

$$(R)_{ij} = r_{ij} = \begin{cases} 1, & \text{if } x_i R y_j, \\ 0, & \text{otherwise}. \end{cases}$$

If $A\subset X$, $A=\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$. Since X is ordered, A can be represented by a characteristic vector. For example, $\{x_1, x_3\}\subset\{x_1, x_2, x_3\}$ can be written as $\langle 1,0,1\rangle$. We define

$$(A)_i = a_i = \begin{cases} 1, & \text{if } x_i \in A, \\ 0, & \text{otherwise}. \end{cases}$$

We can interpret image sets under the relation, $AR\subset Y$, as characteristic vectors of Y in the same way.

Put $B=AR$, $A\subset X$. We can define the 0-1 vector B by the Boolean inner product (i.e., 1+1=1, multiplication as usual) of the 0-1 valued vector A and matrix R

$$b_j = (AR)_j = \sum_i a_i r_{ij}.$$

This is true because

$$\begin{aligned} y_j \in AR &\iff \exists\, x_i \in A\subset X \text{ such that } x_i R y_j \\ &\iff \exists\, i \in I \text{ such that } a_i r_{ij} = 1. \end{aligned}$$

Say we want to take a Boolean combination of rows of R. Let

$A \subset X$ be the collection of rows we wish to combine. In set notation, the resulting image set in 2^Y is AR. In matrix notation, the image vector in 2^Y is obtained by taking the Boolean inner product of the 0-1 characteristic vector A with the 0-1 matrix R, as defined above.

Since we will be working with lattices, we need to represent our partial ordering of vector <u>inclusion</u>. The set $A' \subset A$ if and only if the vector $A' \leq A$, so that

$$A' \leq A \iff a'_i \leq a_i, \text{ for every } i \in I.$$

We will feel free to interchange the set $A \subset X$ with its vector and the relation R with its matrix when our meaning is clear from the context.

3''. Specificity labelling. The method.

Markowsky's approach [M i)] is to look at the lattice of sets obtained as images under R of some subset of X. This is the same as looking at all the Boolean combinations of rows. The collection of objects in this lattice is then notated as:

$$\begin{aligned} L &= \{AR \subset Y \mid A \subset X\} \\ &= \{\text{vectors } B \text{ in } 2^Y \mid AR=B, \text{ for some 0-1 vector } A \text{ in } 2^X\}. \end{aligned}$$

Instead of using the image sets under R, Wohlgemuth considers the equivalence classes of the preimages of these image sets (vectors). By preservation of unions, if $AR=A'R=B$, then $(A\cup A')R=(AR)\cup(A'R)=B$. So the union of all the sets in an equivalence class is again in the equivalence class. Wohlgemuth takes this largest member of each preimage equivalence class as its representative, and calls it a closed set. (The set is called closed because it is closed under the operation $(R(AR)^c)^c$, where ^{c} denotes set complementation. That is, $(R(AR^T)^c)^c=A$ if $A=\bar{A}$. This is discussed in section 4.)

Is Wohlgemuth's lattice of preimages the same as Markowsky's lattice of image sets? Is the lattice generated by rows of R equivalent to the one generated by columns? The answer to both these questions, for a Boolean relation, is "yes"; the proof is given in section 4.

We give an example now in the matrix notation. Let $R:2^X \to 2^Y$ be represented by the matrix

$$R = \begin{array}{cc} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \quad Y \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \\ X & \end{array}$$

We work Wohlgemuth's approach in detail. Recall that $A\subset X$

is closed whenever it is the largest subset of X mapping into its image set. Thus (A is not closed) iff ($\exists x \in X, x \notin A$, such that $xR \subset AR$). That is, if the image of set A under R is not changed when for some $x \notin A$, $(A \cup \{x\})R = AR$, then A cannot be closed. If this occurs, we say that A <u>subsumes</u> the element x. In terms of immunogenetics, this means that the test sera $\{x_{i_1}, \ldots, x_{i_k}\} = A$ when pooled together is <u>indistinguishable</u> from the pooled sera $\{x_{i_1}, \ldots, x_{i_k}, x\} = (A \cup \{x\})$ when tested against the cells $\{y_1, \ldots, y_n\}$.

If all the information in the powerset of X were transmitted through the 0-1 relation matrix (such as when R is the identity matrix I), then each subset of X would be mapped into its own image set. The lattice of preimage (equivalence class) sets would then consist of all 2^m subsets of X, i.e., the full Boolean lattice on m generators. The smallest Boolean lattice into which this could embed would the one on m generators (itself), and m specificities would be required to label R=I.

Each time one subset of X subsumes another with respect to the relation R, the number of lattice points is reduced by 1. If enough lattice points are lost, it may be possible to embed the (pre-)image lattice L into a Boolean lattice on fewer than m generators. In this example, R is only slightly different from the 3x3 identity matrix. As we will now compute, only two subsets of X are subsumed (hence lost to the lattice L), and L cannot be embedded in a Boolean lattice on fewer than 3 generators. Therefore 3 specificities are needed to label R.

We now compare the images of x and x under R.

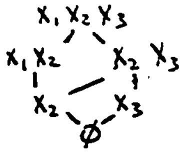
$\{x_1\}R = \{y_1,y_2\}$ $\quad\quad$ $\langle 1,0,0\rangle R = \langle 1,1,0\rangle$

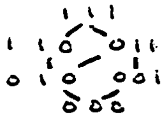
$\{x_2\}R = \{y_2\}$ $\quad\quad$ $\langle 0,1,0\rangle R = \langle 0,1,0\rangle$

Since x_1 subsumes x_2, x_1 is not closed. (Notation: we write $\{x\}$ as x when the meaning is clear.) As noted above, if A subsumes A', then $(A \cup A')R=AR$. So we ask: what is the largest subset of X with image $\{x_1,x_2\}R=\langle 1,1,0\rangle$? Since x introduces a third position $\langle 0,0,1\rangle$ (y_3) into the image vector, we cannot add it. So $\{x_1,x_2\}$ is a closed set, with image vector $\langle 1,1,0\rangle R$ = 1xfirst row of R + 1xsecond row + 0xthird row = $\langle 1,1,0\rangle + \langle 0,1,0\rangle = \langle 1,1,0\rangle$.

If any other row is added to $x_2R=\langle 0,1,0\rangle$, a different row vector is obtained: x_2 is closed. The only other set which is not closed is $\{x_1,x_3\}$, since $\{x_1,x_3\}R = \langle 1,1,1\rangle = XR$.

Writing down the preimage closed sets on the left and their row vector images on the right, we have the lattice L





Can L be embedded in the Boolean lattice on two generators? Since it has lattice length 4, it cannot.

Since m=#X=3, L can of course be embedded in the Boolean lattice on three generators.

One way of labelling R with the three specificites $\{s_1, s_2, s_3\}$ would be

	s_1	s_1s_2	s_3	Y
$s_1 s_2$	1	1	0	
s_2	0	1	0	
s_3	0	0	1	
X				

If s_i labelling the ith row matches its complementary s labelling the jth column, then a 1 appears in r_{ij} ; otherwise a 0 appears. (Since each s is paired with exactly one complementary $\bar{s}$, we may write $\bar{s}$ as s.)

3'''. Specificity labelling. factoring and covers.

This section is expository and is of interest primarily to mathematicians desiring to read the related literature.

Another way of looking at specificity labelling is as the problem of *factoring* the 0-1 matrix R into the Boolean product of two or more matrices. This was Wohlgemuth's original approach [W ii) and iii)]. and we follow his presentation in the exposition below.

Factoring into the product of two Boolean matrices assumes that each specificity s can recognize one and only one $\bar{s}$. Factoring into three matrices can account for *crossreactivity*, where a given s is not assigned a unique complementary $\bar{s}$. (This concept corresponds to the real world fact that a particular chemical configuration is not restricted to bind to only one other complementary configuration. This is especically true in biological chemistry, where molecules may be very similar.) We examine the crossreactive case first.

Let $X=\{x_1,\ldots,x_m\}$ be a set of sera.
Let $S=\{s_1,\ldots,s_k\}$ be a set of specificities.
Let $\bar{S}=\{\bar{s}_1,\ldots,\bar{s}_k\}$ be a set of anti-specificities.
Let $Y=\{y_1,\ldots,y_n\}$ be a set of cells.
The labelling problem may be described as finding a pair of labelling functions ϕ and ψ, where

$$\phi : X \to 2^S$$

$$\phi(x) = \{s_{i_1}, s_{i_2}, \ldots, s_{i_p}\}$$

is the labelling function for the rows,

$\psi : Y \to 2^{\bar{S}}$

$\psi(y) = \{\bar{s}_{j_1}, \bar{s}_{j_2}, \ldots, \bar{s}_{j_q}\}$

is the labelling function for the columns and

$C : S \to 2^{\bar{S}}$

$C(s) = \{\bar{s}_{j_1}, \bar{s}_{j_2}, \ldots, \bar{s}_{j_r}\}$

describes the crossreactivity relation matrix. If C is the identity matrix (i.e., each s is paired with its own unique s) then there is no crossreactivity.

With crossreactivity, we have the picture

$$\begin{array}{cc} & \begin{array}{cccc} s_1 & s_2 & \cdots & S \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ \\ X \end{array} & \left(\begin{array}{c} \\ \phi \\ \\ \end{array} \right) \end{array} \qquad \begin{array}{cc} & \begin{array}{cccc} \bar{s}_1 & \bar{s}_2 & \cdots & \bar{S} \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ \vdots \\ \\ S \end{array} & \left(\begin{array}{c} \\ C \\ \\ \end{array} \right) \end{array} \qquad \begin{array}{cc} & \begin{array}{cccc} y_1 & y_2 & \cdots & Y \end{array} \\ \begin{array}{c} \bar{s}_1 \\ \bar{s}_2 \\ \vdots \\ \\ \bar{S} \end{array} & \left(\begin{array}{c} \\ \psi \\ \\ \end{array} \right) \end{array}$$

Since $x_i R y_j \iff \exists\, s_\alpha, \bar{s}_\beta$ such that

$x_i \ \phi \ s_\alpha$,

$s_\alpha \ C \ \bar{s}_\beta$, and

$\bar{s}_\beta \ \psi \ y_j$;

i.e., $\phi(x_i, s_\alpha) \times C(s_\alpha, \bar{s}_\beta) \times \psi(\bar{s}_\beta, y_j) = 1$,

we have the Boolean matrix product

$R = \phi C \psi$.

If there is no crossreactivity, then $:S \to \bar{S} : s \to \bar{s}$ is an isomorphism, and C=I, the identity matrix. The picture is then

$$\begin{array}{cc} & \begin{array}{ccc} s_1 & s_2 & \cdots \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \end{array} & \left(\begin{array}{c} \\ \phi \\ \\ \end{array} \right) \end{array} \qquad \begin{array}{cc} & \begin{array}{ccc} y_1 & y_2 & \cdots \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ \vdots \end{array} & \left(\begin{array}{c} \\ \psi \\ \\ \end{array} \right) \end{array}$$

and $x_i R y_j \Leftrightarrow \exists\, s \in S$ such that $x_i \phi s$ and $s \psi y_j$,

$$\text{i.e., } \phi(x_i,s) \times \psi(s,y_j)=1.$$

Thus $R=\phi\psi$.

A similar construction shows that if

ϕ is an mxk relation matrix and

ψ is a kxn relation matrix,

such that $R=\phi\psi$,

then this describes a unique specificity labelling of R with k specificities. (The rows of ϕ define the function $\phi: X \rightarrow 2^S$ and the columns of ψ give $\psi: Y \rightarrow 2^S$.) So the minimal number of specificities which can label R is also the size of the *smallest set S* such that R can be *factored through S* as $R: 2^X \rightarrow 2^S \rightarrow 2^Y$. As Wohlgemuth showed in [Wo iii)], this factoring of relations is equivalent to the *specificity covers* introduced by Nau and Markowsky [N]. We will explain what a specificity cover is and, following Wohlgemuth's treatment, show its equivalence to relation factoring.

The specificity cover problem is to express a reaction matrix R as a Boolean sum of a minimal number of *elementary specificity matrices* (ESM). An ESM is a 0-1 mxn matrix which can be expressed as the Boolean matrix product of an mx1 and a 1xn matrix (i.e., a column matrix and a row matrix). We show that this differs only notationally from relation factoring, and then give an example.

(All sums and products in this section are Boolean.) Put $R_s = U_{\cdot s} V_{s \cdot}$, where $U_{\cdot s}$ is the sth column of U, taken as a mx1 matrix, and $V_{s \cdot}$ is the sth row of V, taken as a 1xn matrix. Then $R=UV \iff R=\sum_s U_{\cdot s}V_{s \cdot} \iff R=\sum_s R_s$, that is, every factoring is equivalent to a specificity cover.

Example:

$$R = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

Factoring of relations. R=UV.

$$\begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1 & 1 & 0 \\ x_2 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 1 \end{array} = \begin{array}{c|ccc} & s_1 & s_2 & s_3 \\ \hline x_1 & 1 & 1 & 0 \\ x_2 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 1 \end{array} \times \begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline s_1 & 1 & 1 & 0 \\ s_2 & 0 & 1 & 0 \\ s_3 & 0 & 0 & 1 \end{array}$$

Specificity cover as sum of relations. $R=R_{s_1}+R_{s_2}+R_{s_3}$.

$$\begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1 & 1 & 0 \\ x_2 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 1 \end{array} = \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\!\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4. Closure (and subclosure) operators.

Note: in this section the notation will read from right to left in order to be consistent with the APL implementation of the ideas. For example, instead of the row vector A multiplying R on the left, AR, the column vector A will multiply R on the left as RA.

Markowsky studied the lattice of image sets $\{RA \mid A \in 2^{X}\}$, ordered by set inclusion. In Wohlgemuth's notation, the lattice of sets $\{(\not{R}A)\not{R} \mid A \in 2^{X}\}$, which corresponds to the preimage equivalence classes of the image sets. Since $A' \subset A \Rightarrow RA' \subset RA$, the lattice of preimages contains the lattice of image sets. Under what conditions is the converse true? That is: (1) Are the two lattices the same, and the methods therefore equivalent? We can form the (pre-)image set lattice by using either rows or columns of the matrix R. (2) Is the row space lattice the same as the column space lattice?

We will define closure and subclosure set operations (with respect to R) in terms of sets and relations. We will then obtain an operator notation which can be applyed to both sets and vectors. We will use these operators to show that the answer to questions (1) and (2) is "yes".

<u>Definition</u>. For $A \subset X$, the <u>closure</u> of A is $\overline{A} = \{x \in X \mid Rx \subset RA\}$. Since $Rx \subset RA$, $\forall\ x \in A$, $A \subset \overline{A}$. If $A = \overline{A}$, we say that A is <u>closed</u>.

Note 1. $RA=R\bar{A}$.

This follows since $A\subset\bar{A}$ implies $RA\subset R\bar{A}$, since R preserves unions. By definition, $RA\supset Rx, \forall\ x\in\bar{A}$.

Note 2. $\bar{\bar{A}}=\bar{A}$.

Immediate from Note 1 and the definition.

Observe that for any $A\subset X$, with $RA=B\subset Y$, the closure of A is $\bar{A}=\bigcup_{I} A'$, $I=\{A'\subset X\,|\,RA'=B\}$. So a closed set is the *largest set* in X which is taken into a *particular* B Y. So this definition is consistent with the closed set as the representative of the equivalence class of preimages of some image set in Y.

Definition. For $B\subset Y$, the *subclosure* of B is $\underline{B}=R\{x\in X\,|\,Rx\subset B\}$. Since $\underline{B}=\bigcup_{Rx\subset B} Rx$, $\underline{B}\subset B$. If $B=\underline{B}$, we say that B is *subclosed*.

Note 3. If B is subclosed, then $\exists!$ closed $A\subset X$ such that $B=RA$. This is seen by setting $A=\{x\in X\,|\,Rx\subset B\}$.

(This fact is the reason why R is a one-to-one map from the preimage lattice to the image lattice, which makes R an isomorphism. Thus the answer to question (1) is "yes".)

Note 4. $\underline{\underline{B}}=\underline{B}$.

$$\begin{aligned}\underline{\underline{B}} &= R\{x\in X\,|\,Rx\subset\underline{B}\}\\ &= R\{x\in X\,|\,Rx\subset RA\}, \text{ for some closed } A\subset X\\ &= R\bar{A}\end{aligned}$$

$$= RA$$

$$= \underline{B}.$$

Note 5. $B=\underline{B} \Leftrightarrow B=RA$ for some $A \subset X$.

This is because $B \subset Y$ is subclosed if and only if

$B = \bigcup_{Rx \subset B} Rx = R(\bigcup_{Rx \subset B} x) = RA$, for $A=\{x \in X \mid Rx \subset B\}$.

One consequence (used below) is: $RA=\underline{RA}$.

Note 6. $\underline{B}$ is the largest subset of B which is an image set under R.

This is because $\underline{B}=RA=R\bar{A}$, and $\bar{A}$ is of maximal size for its particular image set.

We now introduce the operator notation for set relations.

RA is the image of $R:2^X \to 2^Y$, $R(A)=RA$.

${}^{T}RB$ is the image of ${}^{T}R:2^Y \to 2^X$, ${}^{T}R(B)={}^{T}RB$.

cA is the complement of the set A, $cA=X-A$.

Operations are performed from right to left on sets.

For example, $(cRc)A$ would be the set which is "not related to the complement of A".

Proposition. $\bar{A}=c\,{}^{T}RcRA$, for $A \subset X$.

Proof: $\bar{A} = \{x \in X \mid Rx \subset RA\}$

$c\bar{A} = \{x \in X \mid Rx \cap cRA \neq \emptyset\}$

$= {}^{T}R(cRA)$, by def. of ${}^{T}R$.

Taking complements of both sides,

$\bar{A} = c\,{}^{T}RcRA.$ //

This closure operation is Wohlgemuth's starting point: $\bar{A} =$

$\cancel{R}\cancel{R}A = c^{T}RcRA.$

Proposition. $\underline{B} = Rc^{T}RcB$, for $B \subset Y$.

Proof: $\quad {}^{T}RcB = \{x \in X \mid cB \cap Rx \neq \emptyset\}$

$$c^{T}RcB = \{x \in X \mid cB \cap Rx = \emptyset\}$$

$$= \{x \in X \mid Rx \subset B\},$$

hence

$$Rc^{T}RcB = R\{x \in X \mid Rx \subset B\}$$

$$= \underline{B}. \qquad //$$

Example. We illustrate the use of the closure operator c RcR by tracing through the matrix of R.

$$\begin{array}{cccccl}
 & x_1 & x_2 & x_3 & X & \\
y_1 & 1 & 0 & 0 & & A = \{x_1\} \text{ or } \langle 1,0,0\rangle \\
y_2 & 1 & 1 & 0 & & \\
y_3 & 0 & 0 & 1 & R. & \\
Y & & & & &
\end{array}$$

$$\begin{array}{llcccl}
 & & \overbrace{}^{A} & & & \\
RA\ \big\{ & y_1 & 1 & 0 & 0 & \\
 & y_2 & 1 & 1 & 0 & \\
 & y_3 & 0 & 0 & 1 & ,
\end{array}$$

$$\begin{array}{llcccl}
RA\ \big\{ & y_1 & 1 & 0 & 0 & \\
 & y_2 & 1 & 1 & 0 & \\
cRA\ \big\{ & y_3 & 0 & 0 & 1 & ,
\end{array}$$

$${}^{T}R \subset RA$$

$$\begin{array}{ccc} x_1 & x_2 & \overline{x_3} \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{array},$$

$$\subset {}^{T}R \subset R\,A$$

$$\begin{array}{ccc} \multicolumn{2}{c}{\overline{x_1 \quad x_2}} & x_3 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{array}.$$

So $\overline{\{x_1\}} = \{x_1, x_2\}$.

We now prove that the lattice of closed sets and the lattice of subclosed (image) sets, whether arising from the row space or the column space, are the same.

<u>Definitions</u>. For the finite sets X and Y and the relation $R: 2^X \to 2^Y$ between them, we define

1. $\overline{R}(X) = \{A \subset X \mid A = \overline{A}\}$,

   the lattice of closed subsets of X with respect to R, ordered by set inclusion.

Similarly,

2. $\underline{R}(Y) = \{B \subset Y \mid B = \underline{B}\}$

   is the lattice of subclosed sets.

3. We define the lattices $\overline{R}(Y)$ and $\underline{R}(X)$ in the same way. where ${}^{T}R: 2^Y \to 2^X$ is the reflection of R.

We now show that

$$\begin{array}{cccc} & \overline{R}(X) & \cong & \underline{R}(Y) & \\ (\text{dual}) & \rotatebox{90}{$\cong$} & & \rotatebox{90}{$\cong$} & (\text{dual}) \\ & \underline{R}(X) & \cong & \overline{R}(Y). & \end{array}$$

<u>Lemma</u> <u>1</u>. $R:2^X \rightarrow R(Y)$ is a lattice homomorphism.

Proof. As observed above in Note 5, $R:A \rightarrow RA=\underline{RA}$, and R preserves the lattice ordering (since it preserves unions).

<u>Lemma</u> <u>2</u>. A is closed if and only if cA is subclosed.

Proof: A is closed <=> $A= c\,{}^{T}RcRA$,

$cA= {}^{T}RcRA$,

and $cA= {}^{T}RcRc(cA)$,

<=> cA is subclosed.

(Note that complementation reverses the lattice ordering.) //

<u>Theorem</u>. If $R:2^X \rightarrow 2^Y$ is a relation between the finite sets X and Y, ${}^{T}R$ its reflection, and c denotes set complementation, then

$$\begin{tikzcd} \overline{R}(X) \arrow[r, "R"] & \underline{R}(Y) \arrow[d, "c"] \\ \underline{R}(X) \arrow[u, "c"'] & \overline{R}(Y) \arrow[l, "{}^{T}R"] \end{tikzcd}$$

are lattice isomorphisms.

<u>Proof</u>. The diagram makes sense: the horizontal arrows by Lemma 1 and the vertical arrows by Lemma 2.

For $A \in \overline{R}(X)$, $c\,{}^{T}RcRA=A$, by the closure operation.

A similar argument giving the identity map at each of the nodes $\underline{R}(Y)$, $\overline{R}(Y)$ and $\underline{R}(X)$ using the closure (or subclosure) operation can be made.

Thus the maps above are isomorphisms.//

This theorem answers questions (1) and (2).

(1) c^TRc is given above as the inverse of $R{:}\overline{R}(X) \to \underline{R}(Y)$. For closed sets A, A' then, $RA \subset RA' \Rightarrow A \subset A'$. Therefore $\overline{R}(X) \xrightarrow[\cong]{R} \underline{R}(Y)$, that is, Wohlgemuth's preimage equivalence classes give the same lattice as Markowsky's image sets.

That c^TRc gives the inverse of RA is clear from 0-1 matrix tracing (and was known to Wohlgemuth). Any row vectors in R which have only 0's in the coordinates given by $c(RA)$ *will be subsumed* by RA. So all the $x \in X$ which are not related (c^TR) to any y's having a 1 in these coordinates (cRA) will be in the closure of A, giving $A = c^TRcRA$.

(2) From the isomorphisms above, the row space and column space lattices (as either closed or subclosed sets) are dual to one another. Since the complete Boolean lattice is symmetric (i.e., $\binom{m}{k} = \binom{m}{m-k}$), the embeddings of L will yield the same minimal number of specificities. (Markowsky demonstrated this in [M ii)] by proving that $cR{:}\overline{R}(X) \to \overline{R}(Y)$ gave an antiisomorphism of the row space to the column space. The inverse map is, of course, c^TR.)

5. Probabilistic closure

There are $\binom{mn}{k}$ mxn reaction tables with exactly k "1" entries. What does a typical lattice of closed (or image) sets selected from this sample space look like?

One approach would be to compute the $\binom{mn}{k}$ lattices and average them together, say by averaging at each order level (i.e., set size). This approach would also let us examine how each lattice embeds into a minimal complete Boolean lattice, giving the specificity number for each lattice. We could then take the average of these minimal specificities and use it as a measure of information for the sample space. For fixed m and n, we would then understand how information changed as a function of k, the number of 1's.

We will use a different, though analogous, approach which has the merit of being easier to analyze mathematically. We let p=k/mn be the average number of 1's in a matrix in the sample space. We will understand p to be the <u>probability</u> that a 1 occurs in an arbitrary entry of the matrix. We assume that the matrix entries are mutually independent random variables taking on the value 1 with probability p and 0 with probability x=1-p: we will study the closure, lattice and specificity properties of these random 0-1 matrices.

Let $R:2^X \to 2^Y$ be a random mxn 0-1 matrix of probability p. If we knew the expected number of subsets of X of size k which were closed, we could use these values, $0 \leq k \leq m$, to estimate a typical lattice. How this "typical" lattice embedded into Boolean lattices would enable us to estimate the minimal specificity number for reaction tables of probability p.

5'. Probabilistic closure. computation.

Let $R:2^X \to 2^Y$ be a random relation matrix of probability p, $\#X=m$ and $\#Y=n$. The goal of this section is to compute the probability that a subset of X of cardinality k is closed. (I thank Don Coppersmith of IBM, Yorktown Heights for considerable assistance in the derivation below.)

Let $Z_{k,i}$, $0 \leq k \leq m$, $i=1,\ldots,\binom{m}{k}$, be the random variable describing the event that [the ith subset of X of cardinality k is closed]. $Z_{k,i}=1$ if the event occurs and $=0$ if it does not.

Then $E(Z_{k,i}) = 0 \times \mathrm{Prob}\{A_{k,i} \neq \bar{A}_{k,i}\} + 1 \times \mathrm{Prob}\{A_{k,i} = \bar{A}_{k,i}\}$

$= \mathrm{Prob}\{A_{k,i} = \bar{A}_{k,i}\}$.

This gives us the expected number of closed sets of size k E_k as

$$E_k = E\left(\sum_{i=1}^{\binom{m}{k}} Z_{k,i}\right)$$

$$= \sum_{i=1}^{\binom{m}{k}} E(Z_{k,i})$$

$$= \binom{m}{k} \text{Prob}\{A_{k,i} = \bar{A}_{k,i}\}.$$

We now compute $\text{Prob}\{A_{k,i} = \bar{A}_{k,i}\}$.

We partition the event that a subset of size k is closed into the disjoint union of the events that the subset is closed and has an image in Y of size 1, $0 \le 1 \le n$.

$$[A_{k,i} = \bar{A}_{k,i}] = \bigcup_{\ell=0}^{n} [A_{k,i} = \bar{A}_{k,i} \text{ and } \#RA_{k,i} = 1]$$

$$\text{Prob}\{A_{k,i} = \bar{A}_{k,i}\} = \sum_{\ell=0}^{n} \text{Prob}\{A_{k,i} = \bar{A}_{k,i} \text{ and } \#RA_{k,i} = 1\}$$

$$= \sum_{\ell=0}^{n} \text{Prob}\{A_{k,i} = \bar{A}_{k,i} \mid \#RA_{k,i} = 1\} \text{Prob}\{\#RA_{k,i} = 1\}.$$

We calculate the two probabilities appearing in each summand.

(1) $\text{Prob}\{\#RA_{k,i} = 1\}$. (See figure 15.)

We want the probability that a set of cardinality k has an image of cardinality 1.

Referring to figure 15, we see that this event occurs when there are 1 subcolumns containing at least one 1, and n-1 subcolumns which are all 0, each subcolumn being of length k.

There are $\binom{n}{\ell}$ ways of choosing the columns.

Setting x=1-p and using the independence of the matrix entries, the probability of at least one 1 appearing in a subcolumn is $(1-x^k)$ and that of no 1's appearing is x^k.

Putting this all together, we have

$$\text{Prob}\{\#RA=1\} = \binom{n}{\ell} (1-x^k)^{\ell} (x^k)^{n-\ell}.$$

That is, the probability is binomially distributed in $(1-x^k)$.

(2) Prob$\{A_{k,i} = \bar{A}_{k,i} \mid \#RA=1\}$. (See figure 16.)

We want the probability of the event

[a subset A of size k is closed given that $\#RA=1$]

Assuming that $\#A=k$ throughout now, this is equivalent to

$[x \in \bar{A} \Rightarrow x \in A]$,

which is the same event as

$[\forall x \in A^c, (Rx) \wedge (RA)^c \pm \emptyset]$,

that is, the rows of A do not <u>subsume</u> any rows of R outside of A. This is the same as the event

[there is at least one 1 in every row of $x \in A^c$ lying under the 0-block of A]

(refer to the figure). The probability that this event occurs for each row of A^c is $(1-x^{n-\ell})$, and since there are m-k such rows, we have that

$$\text{Prob}\{A_{k,i} = \bar{A}_{k,i} \mid \#RA_{k,i} = 1\} = (1-x^{n-\ell})^{m-k}.$$

<u>Theorem</u>. For $R:2^X \to 2^Y$ a random mxn 0-1 matrix of probability p, setting $x=1-p$, the expected number of closed sets of cardinality k is

$$E_k = \binom{m}{k} \sum_{\ell=0}^{n} \binom{n}{\ell} (1-x^k)^\ell x^{k(n-\ell)} (1-x^{n-\ell})^{m-k}.$$

<u>Proof</u>. From the discussion above,

$$\begin{aligned} E_k &= \binom{m}{k} \text{Prob}\{A_{k,i} = \bar{A}_{k,i}\} \\ &= \binom{m}{k} \sum_{\ell=0}^{n} \text{Prob}\{A_{k,i} = \bar{A}_{k,i} \mid \#RA_{k,i} = 1\} \text{Prob}\{\#RA_{k,i} = 1\} \\ &= \binom{m}{k} \sum_{\ell=0}^{n} [(1-x^{n-\ell})^{m-k}][\binom{n}{\ell} (1-x^k)^\ell (x^k)^{n-\ell}] \\ &= \binom{m}{k} \sum_{\ell=0}^{n} \binom{n}{\ell} (1-x^k)^\ell x^{k(n-\ell)} (1-x^{n-\ell})^{m-k}. \quad // \end{aligned}$$

<u>Corollary</u>. The expected number of closed sets (image sets) is

$$E = \sum_{k=0}^{m} \binom{m}{k} \sum_{\ell=0}^{n} \binom{n}{\ell} (1-x^k)^{\ell} x^{k(n-\ell)} (1-x^{n-\ell})^{m-k} .$$

<u>Corollary</u>. The expected number of image sets of an mxn relation $R:2^X \to 2^Y$ of probability $p=1-x$ is

$$\binom{n}{\ell} \sum_{k=0}^{m} \binom{m}{k} (1-x^k)^{\ell} x^{k(n-\ell)} (1-x^{n-\ell})^{m-k} .$$

<u>Proof</u>. The considerations are identical to those in the calculation of the expected number of closed sets. The difference is that here we sum over the subsets of X for a particular subset of Y. (It was the other way around in the closed set calculation.) //

5''. Probabilistic closure. analysis.

We write the summands of E_k as $\binom{n}{\ell} f_{k,\ell}(x)$; where $f_{k,\ell}(x) = (1-x^k)^{\ell} x^{k(n-\ell)} (1-x^{n-\ell})^{m-k}$. By inspection, for $0<k<m$ and $0<l<n$, $f_{k,\ell}(x)$ is a positive valued function on (0,1) and assumes the value 0 at the endpoints 0 and 1. It therefore has at least one local maximum on (0,1); does it have more than one? We now compute the derivative of $f_{k,\ell}(x)$, $0<k<m$, $0<l<n$. By setting this derivative equal to 0, we will show that $f_{k,\ell}(x)$ is a unimodal function on [0,1]: it has exactly

one maximum. (The behavior of $\{f_{k,\ell}(x) \mid k=0,m;\ 1=0,n\}$ is a special case and we defer its analysis to the end of this section.)

$$f_{k,\ell}(x) = (1-x^{k})^{\ell}\, x^{k(n-\ell)}\, (1-x^{n-\ell})^{m-k},$$

$$\begin{aligned} f'_{k,\ell}(x) = {} & [1(1-x^{k})^{\ell-1}\, (-kx^{k-1})\,]x^{k(n-\ell)}\, (1-x^{n-\ell})^{m-k} \\ & + (1-x^{k})^{\ell}\, [k(n-1)\,x^{k(n-\ell)-1}](1-x^{n-\ell})^{m-k} \\ & + (1-x^{k})^{\ell}\, x^{k(n-\ell)}[\,(m-k)\,(1-x^{n-\ell})^{m-k-1}\, (-(n-1)\,x^{n-\ell-1}\,). \end{aligned}$$

Factoring:

$$\begin{aligned} f'_{k,\ell}(x) = {} & (1-x^{k})^{\ell-1}\, x^{k(n-\ell)-1}\, (1-x^{n-\ell})^{m-k-1} \\ & [1(-k)\,x^{k-1}\, x(1-x^{n-\ell}) \\ & \quad + (1-x^{k})\,k(n-1)\,(1-x^{n-\ell}) \\ & \quad + (1-x^{k})\,x(m-k)\,(-(n-1)\,x^{n-\ell-1}\,)\,] \end{aligned}$$

Set $f'_{k,\ell}(x)=G(x)H(x)$, $H(x)$ being the sum in the brackets. Since $G(x)$ has no roots in $(0,1)$, we restrict our attention to $H(x)$.

$$\begin{aligned} H(x) = {} & -\,k1x^{k}(1-x^{n-\ell}) \\ & + k(n-1)\,(1-x^{k})\,(1-x^{n-\ell}) \\ & - (m-k)\,(n-1)\,(1-x^{k})\,x^{n-\ell}, \\ = {} & k(n-1) \\ & + x^{k}[-k1\ -k(n-1)\,] \\ & + x^{n-\ell}[-k(n-1)\ -(m-k)\,(n-1)\,] \\ & + x^{n-\ell+k}[k1\ +k(n-1)\ +\ (m-k)\,(n-1)\,], \\ = {} & k(n-1)\ -knx^{k}\ -m(n-1)\,x^{n-\ell} + (k1+m(n-1))\,x^{n-\ell+k}. \end{aligned}$$

The four coefficients of $H(x)$ have sign +--+. Since there are exactly two changes of sign (+ -> - and - ->+), by Descartes' Rule of Signs there are at most two postitive roots of $H(x)$. One of these occurs at $x=1$: $H(1)=0$. There is then at most one root of $H(x)$, hence of $f'_{k,\ell}(x)$, on

(0,1). So $f_{k,\ell}(x)$ has at most one maximum on (0,1). Since $f_{k,\ell}(x)$ has at least one maximum on (0,1), we have proved

__Theorem__. $f_{k,\ell}(x)$, $0<k<m$, $0<l<n$, has exactly one local maximum on (0,1). That is, $f_{k,\ell}(x)$ is a unimodal function (with 0-valued endpoints) on the unit interval.

This theorem underlies our work in Chapter IV on probabilistic rank. From the $f_{k,\ell}(x)$, we can calculate the expected number of closed subsets of X of size k $E_k(x) = \binom{m}{k}\sum_{\ell=0}^{n}\binom{n}{\ell} f_{k,\ell}(x)$. Using these mean values we will look at a __probabilistic__ __lattice__, which has $E_k(1-p)$ lattice points at the kth lattice level, $0\leq k\leq n$. We will do an analogue of the embedding problem to study lattice lengths, and exploit the fact that $E_k(1-p)$ is a binomial averaging of unimodal functions on [0,1]. We will suggest that the probabilistic rank function rank(p), which gives the expected number of specificities as a function of the probability p that a matrix entry is 1, is again a unimodal function on [0,1].

In the remainder of this section we will look at the behavior of $f_{k,\ell}(x) = (1-x^k)^\ell x^{k(n-\ell)}(1-x^{n-\ell})^{m-k}$ for k=0,m and l=0,n. We do this by examining each of the eight cases (see figure 17). (This analysis is included here for completeness only and is not essential to the presentation.)

1. $k=0,\ l=0$: $f_{k,\ell}(x) = (1-x^n)^m$.

This is the probability that the empty set in X is taken to the empty set in Y by R. The functional form tells us that this probability decreases monotonically from 1 to 0 as $x=1-p$ goes from 0 to 1. That is, when the matrix R is all 1's ($p=1$), $\emptyset=\bar{\emptyset}$; when it is all 0's ($p=0$), there is an all 0-row present and $\emptyset$ is not closed.

2. $0<k<m,\ l=0$: $f_{k,\ell}(x) = x^{kn}(1-x^n)^m$.

$f_{k,\ell}(x)$ here gives the probability of the event that a subset of X of size k is the largest subset to be mapped into the empty set in Y by R. In other words, the probability that exactly k rows of the matrix R are all 0's. We show that these are unimodal functions with the maximum occurring at $x_0=(k/m)^{1/n}$.

$$\begin{aligned}\frac{d}{dx}\, x^{kn}(1-x^n)^m &= (knx^{kn-1})(1-x^n)^{m-k} - x^{kn}(m-k)(1-x^n)^{m-k-1}(nx^{n-1})\\ &= x^{kn-1}(1-x^n)^{m-k}[kn(1-x^n) - (m-k)nxx^{n-1}]\\ &= nx^{kn-1}(1-x^n)^{m-k}[k - kx^n - (m-k)x^n]\\ &= nx^{kn-1}(1-x^n)^{m-k}[k - mx^n].\end{aligned}$$

This equals 0 on (0,1) when $[k-mx^n]=0$, i.e., when

$$x_0 = (k/m)^{1/n}.$$

3. $k=m,\ l=0$: $f_{k,\ell}(x) = x^{mn}$.

This is the probability that R takes X into $\emptyset$, or that every entry in the matrix is 0. The function increases monotonically form 0 to 1 as x goes from 0 to 1. When $x=1$,

$p=0$, and we have all 0-rows.

4. $0<l<m$, $k=m$: $f_{k,\ell}(x) = (1-x^{m})^{\ell} x^{m(n-\ell)}$.

The analysis here is similar to that in case 2. The only zero of f's derivative on (0,1) occurs when x is the root of $[-nx^{m}+(n-1)]=0$, so unimodal f's unique maximum is at

$$x_0 = ((n-1)/n)^{1/m}.$$

5. $k=m$, $l=n$: $f_{k,\ell}(x) = (1-x^{m})^{n}$.

This case is analogous to case 1. $f_{k,\ell}(x)$ decreases monotonically from 1 to 0 as x moves from 0 to 1.

In the remaining three cases, f(x) is identically 0.

6. $k=0$, $0<l<n$.

This corresponds to the impossible situation in which the empty set is taken by R to a non-empty set in Y. Observe that

$$(1-x^{k})^{\ell} = (1-x^{0})^{\ell} = (1-1)^{\ell} = 0 \ , \ l>0,$$
$$= 0.$$

7. $k=0$, $l=n$.

The analysis is identical to that above: $f=0$.

8. $0<k<n$, $l=n$.

This is the situation where a proper subset A of X, which is closed, is mapped by R into the entire set Y. Whence, $A=\bar{A}=\bar{X}=X$, but $A \subsetneq X$. The analysis is as in case 6:

$$(1-x^{n-l})^{m-k} = (1-x^{0})^{m-k} = 0, \quad k<n.$$

6. Non-Boolean lattices

The motivation for understanding Boolean matrices was to analyze data obtained as tables of 0's and 1's under the experimentally observed rule that $1+1=1$. Toward that end we looked at the (pre-)image space lattice generated by Boolean combinations of row vectors. What can we say about matrices whose entries are ordered more generally than $0<1$?

Experiments can give rise to other, non-Boolean, kinds of ordered data arranged in reaction tables. In immunogenetics, for example, we may be interested in the *strengths* of the serum-cell interactions. We can record these strengths as 0,1,2,... . As in the 0-1 case, we can "pool together" matrix rows. Combinations of, say, several sera against one cell type will mask the lower strength interactions and interact as the *maximum* of all the interactions. Addition of numbers would then be defined by their maximum, e.g., $0+2+3=3$. If m is the number of rows that the table has, we could look at all the 2^m combinations of rows under the max operation.

Sometimes the data entries in our matrix can show more complicated ordering relations, as in the case of <u>pairs</u> of responses. For example, suppose that each interaction was recorded as a pair of "yes-no" responses from the list (yes,yes), (yes,no), (no,yes), (no,no). Assume that for each coordinate a "yes" response masks a "no" response. Put 1="yes" and 0="no". Then each entry in our matrix would be one of the four elements selected from the lattice

```
   11
   / \
 10   01
   \ /
   00
```

<u>Definition</u>. An <u>upper lattice</u> is a set T with a partial ordering $\leq$ and the sup lattice operation + such that $a+b= c \in T$, where c is the smallest element satisfying $a\leq c$ and $b\leq c$.

Let $R:X\times Y\to T$ be a matrix, where T is an upper lattice, $\#X=m$, $\#Y=n$. If $r=\langle r_1, \dots, r_n\rangle$ and $s=\langle s_1, \dots, s_n\rangle$ are rows of R, we define r+s to be $\langle r_1+s_1, \dots, r_n+s_n\rangle$. The object we will study in this section will be the set of the 2^m upper lattice combinations of the rows of R

$L= \{a_1 r + a_2 r + \dots + a_m r_m \mid a_i = 0$ or $1,$

r_i is the ith row of R, and

+ is the upper lattice sup operation.}

The <u>main objective</u> of this section is to take an mxn matrix with upper lattice entries, and obtain a Boolean matrix which we can analyze by the methods presented above. To do this, we will take such a matrix

$R:X\times Y\to T$

and transform it to a Boolean mx(nxt) matrix

$R':X\times(Y\times T)\to\{0,1\}$,

where t=#T and T is the underlying set of $(T,\leq,+)$. With R' the particular Boolean matrix which we define below, we will prove the

<u>Theorem</u>. Let $R:X\times Y\to T$ be a matrix with entries in an upper lattice $(T,\leq,+)$ and let $R':X\times(Y\times T)\to\{0,1\}$ be the Boolean-transformed matrix of R. Then the row space lattice $\overline{R}'(X)$ of R' is the same as the row space lattice $L=\overline{R}(X)$ if and only if the upper lattice T is a chain.

6'. Non-Boolean lattices. the proof.

The proofs in this section are straightforward, but are rather tedious. With the exception of the proof of the main theorem, they are omitted.

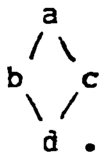
Definition. Let $(T,\leq)$ be a partially ordered set. Then for every $\alpha \in T$, we may define the threshold function

$$\theta_\alpha : T \to \{0,1\}$$
$$t \to 1, \text{ if } \alpha \leq t, \text{ and}$$
$$0, \text{ otherwise.}$$

This definition is easily extended to functions having images in T. For example, given $R: X \times Y \to T$,

$$\theta_\alpha : T^{X\times Y} \to \{0,1\}^{X\times Y}$$
$$R \mapsto \theta_\alpha(R), \text{ defined by}$$
$$\theta_\alpha(R)(x,y) = \theta_\alpha(R(x,y)).$$

Example. Let T be the complete Boolean lattice on 2 generators B_2,



If $\alpha = b$ and

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$\theta_b(R) = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

If R_1 and R_2 are $: X \times Y \to \{0,1\}$ matrices, then we define

$$R_1 \le R_2 \iff R_1(x,y) \le R_2(x,y),\ \forall\ (x,y) \in X \times Y.$$

That is, $R_1 \le R_2$ whenever every entry in R_1 is $\le$ its corresponding entry in R_2.

__Definition__. The __threshold__ __set__ lattice is defined to be the collection

$$\Theta(R) = \{\, \Theta_\alpha(R) \mid \alpha \in T \},$$

ordered as a collection of 0-1 matrices.

As an abstract lattice, $\Theta(R)$ is contained (dually) in T. This is immediate from the observation that for α, β and $m \in T$,

$$\alpha \le \beta \implies \Theta_\alpha(m) \le \Theta_\alpha(m),$$

so we cannot have both

$(m \ge \alpha)$ and $(m \text{ is not } \ge \beta)$,

since this implies $1 \le 0$.

__Example__. As before, let $T = B_2$. Then

$$\Theta(b) = \begin{tikzcd}[row sep=small, column sep=small] & 0 \arrow[dl, dash] \arrow[dr, dash] & \\ 1 \arrow[dr, dash] & & 0 \arrow[dl, dash] \\ & 1, & \end{tikzcd} = \begin{tikzcd}[row sep=small, column sep=small] & \Theta_a(b) \arrow[dl, dash] \arrow[dr, dash] & \\ \Theta_b(b) \arrow[dr, dash] & & \Theta_c(b) \arrow[dl, dash] \\ & \Theta_d(b). & \end{tikzcd}$$

With $R: X \times Y \to T$ as above, another example is

$$\Theta(R) = \begin{tikzcd}[row sep=small, column sep=small] & \begin{matrix} 1 & 0 \\ 0 & 0 \end{matrix} \arrow[dl, dash] \arrow[dr, dash] & \\ \begin{matrix} 1 & 1 \\ 0 & 0 \end{matrix} \arrow[dr, dash] & & \begin{matrix} 1 & 0 \\ 1 & 0 \end{matrix} \arrow[dl, dash] \\ & \begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}. & \end{tikzcd}$$

For every R, the above definition gives $\theta(R)$ in a well defined way. We show in Lemma a that the map θ from the set $T^{X\times Y}$ to the set $\left(\{0,1\}^{X\times Y}\right)^T$ is injective. In constructing the inverse map on the image of θ, we will make use of the following property.

Property *. The set $T|_t = \{\alpha \in T \mid \theta_\alpha(t)=1\}$ forms a sub-upper lattice of T under $\leq$.

Lemma a. For $R: X\times Y \to T$, T an upper lattice,

$$\theta : T^{X\times Y} \to \left(\{0,1\}^{X\times Y}\right)^T$$
$$R \mapsto \theta(R)$$

is an injective map.

This is proved usinge the inverse map

$$\theta^{-1} : \theta(T) \to T$$
$$T|_m \to m = \sup T|_m .$$

For θ to be a homomorphism between the lattice T and the Boolean threshold representation $\theta(T)$, θ must preserve the order $\leq$ and the operation +. We now show that $\theta(T)$ preserves order.

Lemma b. Let R,S:XxY->T be two matrices with values in (T,≤). Then

$$R \underset{L}{\leq} S \iff \theta_\alpha(R) \underset{B}{\leq} \theta_\alpha(S),\ \forall \alpha \in T.$$

($\underset{L}{\leq}$ and $\underset{B}{\leq}$ denote the (upper) lattice order relation in T and the Boolean ordering on {0,1}.)

We now state the key lemma that whenever (T,≤) is a chain lattice, the image under θ preserves the (upper) lattice sup operation with respect to ≤. That is, $\theta : (T,\leq,+) \to (2^T,\leq,v)$ is a lattice homomorphism, where v is the Boolean "or" {0,1} operation. (If R and S:XxY->{0,1} are Boolean matrices, then the Boolean matrix (RvS):XxY->{0,1} is defined by (RvS):(x,y)->R(x,y)vS(x,y). Extending θ to matrices by performing the operation separately on each entry then extends the result stated here for an element of a lattice to matrices with lattice entries.) In fact, we show that θ is a lattice homomorphism if and only if T is a chain.

Lemma c. Let R_1,R_2,S_1 and S_2:XxY->T be matrices with entries in the upper lattice T. Then the following are equivalent.

(1) T is a chain.

(2) For any R_1 and R_2, and for all $\alpha \in$ T

$$\theta_\alpha(R_1+R_2) = \theta_\alpha(R_1)\, v\, \theta_\alpha(R_2).$$

(3) For all R_1,R_2,S_1,S_2

$$R_1+R_2 \le S_1+S_2 \iff$$

$$\Theta_\alpha(R_1) \vee \Theta_\alpha(R_2) \le \Theta_\alpha(S_1) \vee \Theta_\alpha(S_2), \ \forall \alpha \in T. \ //$$

(2) says that $\Theta(R_1+R_2) = \Theta(R_1) \vee \Theta(R_2)$, where the operation v is performed at each lattice point. (Formally, $(\Theta(R_1) \vee \Theta(R_2)) : T \to 2^{X\times Y}$, defined by $(\Theta(R_1) \vee \Theta(R_2)) = \Theta(R_1) \vee \Theta(R_2)$.)

(3) explicitly shows how Θ preserves order, with $R_1+R_2 \le S_1+S_2 \iff \Theta(R_1) \vee \Theta(R_2) \le \Theta(S_1) \vee \Theta(S_2)$.

Combining Lemmas a, b and c proves the

<u>Theorem</u>. Let $R : X\times Y \to T$ be a matrix with entries in the (upper) lattice T. Then

$$\Theta : (T^{X\times Y}, \le, +) \to (\{0,1\}^{X\times Y\times T}, \le, \vee)$$

$$R \mapsto \Theta(R)$$

is a lattice isomorphism onto its image if and only if T is a chain.

Let $(T, \le, +)$ be an upper lattice and $R : X\times Y \to T$ a matrix. The object we are trying to understand in this section is the row space lattice L, the set of the $2^{\#X}$ upper lattice combinations of the rows of R. Our method is to apply the threshold function Θ to R, and then investigate the 0-1 object $\Theta(R)$. The above theorem tells us that if T is not a chain, then we cannot reduce the study of the row space L of R to an analysis of $\Theta(R)$ for all R. We now see what exactly can be learned about L from $\Theta(R)$ in the case where

T is a chain.

Let $I=\{1,2,\ldots,\#T\}$ be the (ordered) set underlying $(T,\leq)$, and let $i:T\to I$ $(t\to i(t))$, $0\leq i(t)\leq\#T$ be the bijection. We define

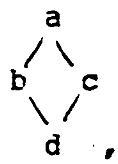
$$i^*:\{0,1\}^T \to \{0,1\}^I$$

$$(T\to\{0,1\})\to(I\to\{0,1\})$$

as the function which transforms the lattice arrangement of 0's and 1's to a 0-1 vector arrangement.

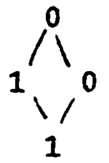
<u>Example</u>. If $i:T=(\{a,b,c,d\},\leq)\to I=\{1,2,3,4\}$, where T is ordered as

T=



,

an example of the action of i^* on $\Theta(b)$ is



$\to$ $\langle 0,1,0,1\rangle$.

Instead of simply T, consider the array $R:X\times Y\to T$ having entries in the lattice T. We use i^* to modify the lattice $\Theta(R)$ of 0-1 matrices to the string $i^*\circ\Theta(R)$ of 0-1 matrices, and can reshape this triply indexed object $i^*\circ\Theta(R)$: $X\times Y\times I\to\{0,1\}$ into the matrix $r\circ i^*\circ\Theta(R): X\times(Y\times I)\to\{0,1\}$. r here is just the natural reindexing of the two singly indexed vectors Y and I into the doubly indexed matrix (YxI). How does the lattice of closed sets of $R:X\times Y\to(T,\leq,+)$ relate to the closed sets of the Boolean matrix $r\circ i^{*}\circ\theta(R)$?

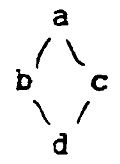
The diagram of our maps is

$$T^{X\times Y}\xrightarrow{\theta}\left(\{0,1\}^{X\times Y}\right)^{T}\xrightarrow{i^{*}}\left(\{0,1\}^{X\times Y}\right)^{I}\xrightarrow{r}\{0,1\}^{X\times(Y\times I)}$$

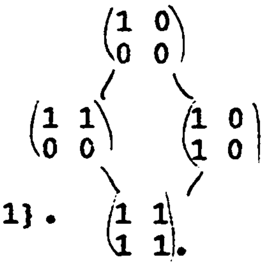
<u>Definition</u>. The <u>Boolean</u> <u>reshape</u> of the matrix $R:X\times Y\to(T,\leq,+)$ is the 0-1 matrix $R':X\times(Y\times I)\to(\{0,1\},\leq,\vee)$, where R' is defined by

$$R'=r\circ i^{*}\circ\theta(R).$$

<u>Example</u>. As before, put

T= 

$$R=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$$

$\theta(R)=$ 

$\theta(R):(X\times Y)^{T}\to\{0,1\}$.

Here, $i:T=\{a,b,c,d\}\to I=\{1,2,3,4\}$,

and the picture of the Boolean reshape of R,

$R'=i^{*}\circ\theta R:X\times(Y\times I)\to\{0,1\}$, is

$$\left(\begin{array}{cc:cc:cc:cc} 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right)$$

$$\alpha = a, \quad b, \quad c, \quad d.$$

Example.

Here, let

$$T = \begin{array}{c} 2 \\ | \\ 1 \\ | \\ 0 \end{array} \qquad R = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.$$

In this case, the threshold of R is

$$\theta(R) = \begin{array}{c} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ | \\ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ | \\ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \end{array} \quad .$$

The Boolean reshape matrix is

$$R' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\alpha = 0, \quad 1, \quad 2.$$

The theorem below says that if T is a chain, then the lattice L of closed sets under R:XxY->T is the same as the lattice of closed sets of the Boolean reshape R':Xx(YxI)->{0,1}. This lets us reduce rank (specificity) problems on arbitrary finite chain lattices to the 0-1 case described earlier in this chapter.

Definitions. Let R:XxY->(T,≤,+) be a matrix having entries in the upper lattice T, $\#X=m$ and $\#Y=n$.

Let r_i denote the ith row of R, $1\leq i\leq m$.

As before r_1+r_2 is defined by term-by-term sup.

Let {0,1} multiply (row) vectors in T as follows:

1xr= r, while

0xr removes the vector r from further consideration.

(if T were a lattice, we would send 0xr to the vector with entries all equal to the inf of T.)

Let $A=\langle a_1,\ldots,a_m\rangle$, $a_i= 0$ or 1, be the *characteristic vector* of one of the 2^m subsets of X.

Then we define:

1. The *image* of A under R as $AR=\sum_i a_i r_i$.
2. The *image vector* lattice (ordered by 0-1 vector inclusion) to be $\underline{R}(X)=\{AR \mid A\in\{0,1\}^X\}$.
3. The *closure* of A, $\bar{A}$, to be the 0-1 vector with the maximum number of 1's such that $\bar{A}R= AR$.

A vector A is said to be *closed* under R if $A=\bar{A}$.

Clearly, each closed vector is the preimage equivalence class representative for the image vector AR.

4. The lattice of *closed vectors* $\bar{R}(X)= \{A=\bar{A} \mid A\subset\{0,1\}^X\}$.

As before, $R:\bar{R}(X)\to\underline{R}(X)$, $R(A)=AR$, is a lattice isomorphism.

Example. Let T= N, the natural numbers up to and including the highest entry in R, and

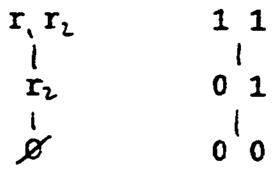
$$R=\begin{pmatrix}2&1\\0&1\end{pmatrix}.$$

Excluding the all-0 and all-1 matrices, the reshape $R':X\times(Y\times N)\to\{0,1\}$ is

$$\begin{array}{c} r_1 \\ r_2 \end{array}\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right)$$

$\alpha = 1, \quad 2.$

R(X) =



closed sets; closed vectors (in X).

Theorem. Let R:XxY->N be a matrix with (truncated) natural number entries, where N={0,1,...,max(R)} is the chain lattice ordered by $\leq$ with the max operation. Let R':Xx(YxN)->{0,1} be the Boolean reshape matrix of R. Then the lattice $\bar{R}(X)$ of closed vectors under R in (N,$\leq$,max) is the same as the lattice $\bar{R}'(X)$ of closed vectors under R' in ({0,1},$\leq$,v). That is,

$$\bar{R}(X) = \bar{R}'(X).$$

Proof. We show that the equivalence class of inverse images

$E_A = \{B \in \{0,1\}^X \mid BR=AR\}$ equals

$E'_A = \{B \in \{0,1\}^X \mid BR'=AR'\}$, where $A \in \{0,1\}^X$.

This implies that the closed vector lattices are the same.

(Notation: we say $a_i \in A$ if the ith entry of A is 1.)

$$AR = BR'$$

by Lemma a, this is true <=>

$$\Theta(AR) = \Theta(BR)$$

by the definition of $\Theta_\backslash$, <=>

$$\theta_\alpha(AR) = \theta_\alpha(BR), \forall \alpha \in N$$

by the definition of AR, this is <=>

$$\theta_\alpha\Big(\sum_{a_i \in A} a_i r_{ij}\Big) = \theta_\alpha\Big(\sum_{b_i \in B} b_i r_{ij}\Big), \forall \alpha \in N$$

Since N is a chain, by Lemma c this <=>

$$\bigvee_{a_i \in A} \theta_\alpha(a_i r_{ij}) = \bigvee_{b_i \in B} \theta_\alpha(b_i r_{ij}), \forall \alpha \in N$$

Since N is a chain, we can reshape, and this is <=>

$$\bigvee_{a_i \in A} (a_i R') = \bigvee_{b_i \in B} (b_i R')$$

which in the Boolean case holds <=>

$$AR' = BR'. \qquad //$$

By this theorem, we can reduce the lattice embedding problem on matrices with entries in N under the maximum operation to the Boolean case of 0-1 matrices under logical "or". Note that in the reshape part of this reduction we arbitrarily chose to reindex the lattice N with the set Y, giving Xx(YxN). Had we reindexed N with the left hand set X, obtaining (XxN)xY, a different Boolean matrix would have been constructed. In general, the closure lattices of these two reshapes *are not the same*.

This means that, in general, $\overline{R}(X) \neq {}^{T}\overline{R}(X)$. This is unlike the 0-1 matrix case, for which we proved that the lattices are equal. We give a counterexample to demonstrate this loss of invariance.

Counterexample.

Put $R = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$.

Reshaping as Xx(YxN) gives

$$R' = \begin{array}{c} r_1 \\ r_2 \end{array}\begin{pmatrix} 1 & 1 & \vdots & 1 & 0 \\ 0 & 1 & \vdots & 0 & 0 \end{pmatrix} \qquad \overline{R}'(X) = \begin{tikzcd}[column sep=tiny, row sep=small] r_1 r_2 \arrow[d, dash] \\ r_2 \arrow[d, dash] \\ \emptyset \end{tikzcd} .$$

$$\alpha = 1, \quad = 2$$

Reshaping R as (XxN)xY, though, gives

$$R' = \begin{array}{cc} c_1 & c_2 \\ \hline \end{array}$$

$$R' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ \hdashline 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{array}{l} \alpha = 1, \\ \\ \\ = 2. \end{array} \qquad \overline{R}'(X) = \begin{tikzcd}[column sep=tiny, row sep=small] & c_1 c_2 \arrow[dl, dash] \arrow[dr, dash] & \\ c_1 \arrow[dr, dash] & & c_2 \arrow[dl, dash] \\ & \emptyset & \end{tikzcd}$$

IV. Probabilistic rank

In Chapter III we introduced the specificity labelling lattice embedding problem. The idea was to look at the set of image vectors obtained by taking the Boolean inner product of a given mxn 0-1 matrix R with every 0-1 vector of length n. This collection of characteristic image vectors, ordered by set inclusion, then formed a lattice L; the smallest complete Boolean lattice on r generators B_r into which L embedded then gave the minimal number of specificities r with which to label the matrix. We called r the rank of the Boolean matrix R, i.e., the rank of the corresponding set relation R. Throughout this chapter, m is set equal to n.

This Boolean rank of a 0-1 matrix has applications to immunology, discrete mathematics and computer science. Its calculation for a particular matrix R is, however, an NP-complete problem (Dana et al,1978). For the study of the immune system, where $n=10^6$ to 10^7, this makes a deterministic computational approach unworkable. Is some less hard, probabilistic approach possible? That is, for a given density $p \in [0,1]$ of 1's in R, what can we say about an expected rank?

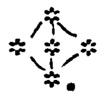
In this chapter we will model a probabilistic approach (the meaning of "model" will be made clearer in the next section.) We will conjecture that our expected rank function rises from r=0 at p=0 to full rank r=n at p=log n/n, then descends no faster than a linear function down to 1 over the remainder of the unit interval. We will then apply this conjecture to some problems in immunology and mathematics.

2. The approach

The lattice embedding problem searches for the smallest r such that the image set lattice L is embedded as an <u>abstract lattice</u> into the Boolean cube B_r. By "embedding", we mean that for every lattice level k of L there is a corresponding lattice level j(k) of B_r such that the number of lattice points L(k) at lattice level k of L is less than or equal to the number of lattice points $B_r(j(k))$ at level j(k) of B_r, where $j:\{0,\ldots,\text{length}(L)\}\to\{0,\ldots,\text{length}(B_r)\}$ is a strictly increasing function.

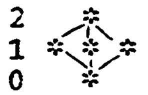
<u>Example</u>. We put

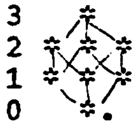
L=



Since L has lattice girth 3, it is too large to embed into

B_2, but can be embedded into B_3 in four different ways:





The four possible j_q's are:

	3	2	3	2	3	2	3
2→	2		2	1	2	1	2
1→	1	1	1		1	0	1
0→	0	0	0	0	0		0

$k \to j_1(k)$, $k \to j_2(k)$, $k \to j_3(k)$, $k \to j_4(k)$.

Instead of using actual lattices, for which the combinatorics can become very complicated, we will <u>model</u> the embedding problem using <u>ordered lists</u> (vectors) of <u>integers</u>.

<u>Definition</u>. Let $V:I \to N$ and $B:J \to N$ be vectors, where $I=\{0,1,\ldots,\#I-1\}$ and $J=\{0,1,\ldots,\#J-1\}$ are index sets and $N=\{0,1,\ldots\}$ is the set of natural numbers. V is said to <u>embed</u> into B if there exists a function $j:I \to J$ satisfying:

1. $v_i \leq b_{j(i)}$, and
2. $j(i) < j(i+1)$.

<u>Example</u>. Put $V=(1\ 3\ 1)$. Then V does not embed into $B_2=(1\ 2\ 1)$ because $V(1)=3 >$ every entry in B_2. V does, however, embed into $B_3=(1\ 3\ 3\ 1)$ in four different ways. These four maps j_q are:

```
   3      2  3      2  3      2  3
2  2         2      1  2      1  2
1  1      1  1         1      0  1
0  0      0  0      0  0         0
```

$k \to j_1(k)$, $k \to j_2(k)$, $k \to j_3(k)$, $k \to j_4(k)$.

How does this model of the lattice embedding problem help us to find a probabilistic rank function? In Chapter III we calculated $E_k(p)$, the expected number of closed sets (i.e., preimage equivalence classes of a 0-1 matrix R) at the kth level of the lattice L=R(X); p here is the probability of a 1 appearing in an entry of R. The idea is to use the vector $E=(E_0\ E_1\ \ldots E_n)$ of expected values to represent the average behavior of the vector of lattice level sizes (L(0),L(1),...,L(n-1)) for a random nxn 0-1 vector of probability p. We then solve the vector embedding problem to find the smallest r such E embeds into B_r. In this context, B_r is the vector giving the number of points at the kth levels of the complete Boolean lattice. That is, $B_r=(B_r(0)\ B_r(1)\ldots B_r(r))$, which is the vector of binomial coefficients $(\binom{r}{0}\ \binom{r}{1}\ \ldots \binom{r}{r})$.

We can calculate the E_k in E by using the formulas developed in Chapter III. We would like to round these expected numbers to integers, and then solve the vector embedding problem. To make the embedding problem a little harder, i.e., to force higher expected ranks instead of lower ones, we will round by using the least integer of E_k instead of

$[E_k]$. (This also lets us include expected closed sets which have $\lim_{n\to\infty} E_{k_n} \to 1$ from below.)

This causes a problem, however, with very small E_k. For most k and l, however small E_k may become, since it arises fom certain probabilities, E_k is always strictly greater than zero. Taking the least integer will round these E_k to 1, and give just about full lattice length in every case. To circumvent this, we set all E_k which are less than a given threshold value equal to zero. (In our computer work, this threshold was set equal to 1/e, which is a limiting probability for some $f_{k,l}(x)$.)

Example. Let R be a 5x5 random 0-1 matrix of probability .1 Then from our formula for E_k, we calculate E to be (.01 .11 .45 1.05 1.44 1). Ignoring the very small E_k values and rounding up to the next integer gives the vector (0 0 1 2 2 1). The smallest vector of binomial coefficients into which this embeds is (1 3 3 1). We therefore say that the probabilistic rank of R is r=3.

Example. Let R be the 5x5 random 0-1 matrix of probability .9. Then E is (1 1.47 .36 .04 .00 1), which rounds to (1 2 1 0 0 1). Since the minimal B_r into which this embeds is B_3=(1 3 3 1), the probabilistic rank is 3.

If the numbers E_k are much less than $\binom{n}{k}$, the probabilistic vector embedding problem for ranks reduces to the question of lattice length. This is because we need not then concern ourselves with how large E_k becomes, but only with whether or not it is non-vanishing, i.e., greater than or less than the threshold. (In the example with V=1 3 1, E_1 was not $<<\binom{2}{1}$, and so the V did not embed into B_2. That is, the lower bound of lattice length did not give a sharp estimate of the rank. We would prefer vectors such as V not to arise in our probabilistic model, and to be able to estimate the rank using lattice length.)

For small p(n), small subsets of rows (i.e., k=0,1,2...<<n) tend to subsume one another and the all-0 vector and to therefore not be closed. Boolean combinations of larger numbers of rows, on the other hand, have a greater tendency to form unique images, hence be closed. We would therefore expect the lower lattice levels (0<k<<n) to vanish and the higher levels (0<<k<n) to be preserved. This is seen in the example above where n=5 and p=.1: the vector is (- - * * * *), where * denotes preservation and - disappearance.

When p(n) is large, Boolean combinations of many rows (k large) will tend to give the all-1 vector, so these lattice levels will disappear. Combinations of smaller numbers of rows (k small), though, will tend to be unique and thus produce closed sets, preserving the corresponding values of E_K. This concurs with the example in which n=5 and p=.9. There we had E=(* * * - - *): the lower k-valued levels are kept, while the higher levels vanish. (Note that E_n=1; this is because the combination of all the rows cannot subsume any additional rows, hence is always closed.)

At p=0, the all-0 matrix has rank=0. As we increase slowly to small values of p(n), computer calculations indicate that we begin to add in lattice levels corresponding to the higher values of k. When p is large enough, we have the lattice levels of the small k, but begin to lose the high k levels. Finally, at p=1 even these small k levels are gone and only the trivial k=n level remains, so that the rank has decreased down to 1 (see figure 18). Is there any value of p(n) in this passage for which we have both the small and high k levels? If so, do we get the full lattice length n+1, hence attain the full rank r=n?

We will suggest below that full rank might be achieved at p=logn/n. Thus on [0,log n/n] we would start with the lattice levels corresponding to the highest values of k and add in ever smaller k levels until we get all values of k to be non-vanishing and full lattice length. On [log n/n,1], we begin to remove the higher k levels until finally, at

$p=1$, even the smallest k values disappear and the rank is reduced to 1. We will argue heuristically that this loss of the higher k levels on $[\log n/n, 1]$, hence the decrease in rank, is no faster than linear in p.

3. Roots: the maxima of the $f_{k,l}(x)$

Say we have a random 0-1 nxn matrix of probability p. Recall from Chapter III that $E_k(x)=\binom{n}{k}\sum_{l=0}^{n}\binom{n}{l} f_{k,l}(x)$, where $f_{k,l}(x)= (1-x^k)^l x^{k(n-l)}(1-x^{n-l})^{n-k}$, $x=1-p$. Each $f_{k,l}(x)$ was proved to be a unimodal function, and so $E_k(x)$ is a binomial averaging of these unimodal functions. If the kth lattice level is to make an appearance for some values of $p(n)=1-x$, then $E_k(x)>1$ for these values of x. This can happen if the maxima of the $f_{k,l}(x)$ *superimpose* for these values of $x(n)$, $l=0,1,\ldots,n$, and k fixed. Otherwise (i.e., the maxima of the $f_{k,l}(x)$ do not superimpose) the high values of the $f_{k,l}(x)$ may be dispersed and it may happen that nowhere along $[0,1]$ will the functions add together sufficently to give $E_k(x)>1$ for some x.

We would like then to locate these maxima of the $f_{k,l}(x)$. This will give us some idea of where the reinforced values of $p(n)$ lie on $[0,1]$ as a function of k and l. We find these maxima by extracting the roots of $f'(x)$.

In Chapter III we wrote $f'(x)=G(x)H(x)$, where only $H(x)$ had roots on $(0,1)$. This function was

$$H_{k,l}(x) = k(n-l) - knx^{k} - n(n-l)x^{n-l} + (kl+n(n-l))x^{n-l+k}.$$

There is no easy way of computing the roots of $H(x)$ directly. However, we can rewrite $H(x)$ as the sum

$$\begin{aligned} H_{k,l}(x) &= k(n-l) - knx^{k} - n(n-l)x^{n-l} + n^{2}x^{n-l+k} - (n-k)lx^{n-l+k} \\ &= (k-nx^{n-l})((n-l)-nx^{k}) - (n-k)lx^{n-l+k} \\ &= a(x) + b(x). \end{aligned}$$

The roots of $a(x)$ are easy to compute and for large n and most values of k and l, $b(x)$ is only a small perturbation of $a(x)$. Thus we can use the roots of $a(x)$ to approximate those of $H(x)$, and thereby obtain a good estimate of the location of the maxima of the $f_{k,l}(x)$ on $(0,1)$.

Observe that the solutions to $a(x)=0$ are

$$x_1 = \left(\frac{k}{n}\right)^{\frac{1}{n-l}} \quad \text{and} \quad x_2 = \left(\frac{n-l}{n}\right)^{\frac{1}{k}}.$$

Before describing how these maximizing values of $p(n)$ vary in k and l, we must prove that $a(x)$ is a good approximation to $H(x)$.

Assume we are given k and l satisfying

$$1 \leq k \leq (n-l) \leq n.$$

(Since $a(x)$ and $H(x)$ are symmetric in k and $(n-l)$ we need not consider the other case $(n-l)<k$.) Assume that n is fixed as well. Let $\alpha=k/n$ and $\beta=l/n$. Since $a(x)$ and $H(x)$ will be set equal to 0, we divide throughout by n^{2}. We

denot by x_1 and x_2 the two roots of

(1) $$(\alpha - x^{n-\ell})((1-\beta) - x^k) = 0$$

that is, $x_1 = \alpha^{\frac{1}{\ell}}$ and $x_2 = (1-\beta)^{\frac{1}{k}}$. We define $x_0^* = \min\{x_1, x_2\}$. (As discussed below, it does not follow from $k<(n-1)$ that $x_1 < x_2$.) Let x_0 be the root on $(0,1)$ of

(2) $$(\alpha - x^{n-\ell})((1-\beta) - x^k) - (1-\alpha)\beta\, x^{n-\ell+k} = 0.$$

It is obvious that $0 < x_0 < x_0^*$; we shall in fact show that

<u>Lemma</u>. For α, β, k, x_1 and x_2 as above,

(*) $$\left(\frac{1}{(1+\theta)^{1/k}}\right) x_0^* \le x_0 < x_0^*,$$

where $\theta = \sqrt{(1-\alpha)\beta}$.

<u>Proof</u>. Suppose that $\rho > 0$ is any number such that

$$x_0^{n-\ell} < \rho\alpha$$

and $$x_0^k < \rho(1-\beta).$$ Then

$$\begin{aligned}(1-\rho)^2 \alpha (1-\beta) &< (\alpha - x_0^{n-\ell})((1-\beta) - x_0^k)\\ &= (1-\alpha)\beta\, x_0^{n-\ell+k} \quad \text{by (2)}\\ &< \rho^2 \alpha (1-\beta)\theta^2\end{aligned}$$

so $\frac{1-\rho}{\rho} < \theta$, and therefore $\rho > \frac{1}{1+\theta}$. It follows that we must have at least one of the inequalities

$$x_0^{n-\ell} \ge \frac{\alpha}{1+\theta} \qquad \text{i.e., } x_0 \ge \frac{x_1}{(1+\theta)^{1/n-\ell}}$$

or $$x_0^k \ge \frac{1-\beta}{1+\theta} \qquad \text{i.e., } x_0 \ge \frac{x_2}{(1+\theta)^{1/k}}.$$

Thus

$$x_0 \ge \min\left\{\frac{x_1}{(1+\theta)^{1/n-\ell}}, \frac{x_2}{(1+\theta)^{1/k}}\right\} \ge \frac{1}{(1+\theta)^{1/k}} x_0^*,$$

which was to be shown.//

Since the last inequality in the proof is stronger than (*), we record it as

<u>Corollary</u>. $x_0 \geq \min\{ \frac{x_1}{(1+\theta)^{1/n-l}}, \frac{x_2}{(1+\theta)^{1/k}} \}$

(**) $\qquad = \min\{ \left(\frac{\alpha}{1+\theta}\right)^{1/n-l}, \left(\frac{\beta}{1+\theta}\right)^{1/k} \}$.

<u>Example</u>. If k=l=1, we have

$\alpha = \beta = 1/n$,

$x_1 = (1/n)^{\frac{1}{n-1}}$, $x_2 = 1-(1/n)$,

and, for $n \geq 2$, $x_0^* = (1/n)^{\frac{1}{n-1}}$.

So $\quad \theta = -\sqrt{(1/n)(1-1/n)} = -\sqrt{n-1/n}$

$1+\theta = \frac{n+\sqrt{n}}{n}$

$\left(\frac{\alpha}{1+\theta}\right)^{1/n-l} = \left(\frac{1}{n+\sqrt{n-1}}\right)^{\frac{1}{n-1}}$

but $\left(\frac{1-\beta}{1+\theta}\right)^{\frac{1}{k}} = \frac{n-1}{n+\sqrt{n-1}}$, so for large n, (**) gives

$\left(\frac{1}{n+\sqrt{n-1}}\right)^{1/n-1} \leq x_0 \leq \left(\frac{1}{n}\right)^{\frac{1}{n-1}}$.

The following derivation is suggestive:

$x_0 \sim (1/n)^{\frac{1}{n-1}}$

$\sim (1/n)^{1/n}$

$= \exp(-(\log n)/n)$

$\doteq 1-(\log n)/n$, or

$p = 1-x_0 = \log n/n$.

This could indicate that 1-p is close to the maximizing probability of $f_{k,l}(x)$ when k and l are both small. We prove this in the

<u>Theorem</u>. Let x_0 be the root on (0,1) of

(3) $(x^{n-\ell} - (k/n))(x^{k} - (n-1)/n) - (1-(k/n))(1/n)x^{n-\ell+k} = 0$

($x_0 = x_0(k,l;n) = x_0(n)$ for k,l fixed.)

Then for each <u>fixed</u> k,l

$$1-x_0(n) \sim \log n/n \text{ as } n\to\infty.$$

<u>Proof</u>. We have already seen that

$$\frac{1}{(1+\theta)^{1/k}} x_0^* < x_0 < x_0^*,$$

where $x_0^* = x_0^*(n) = \min\{(k/n)^{\frac{1}{n-\ell}}, (n-1/n)^{\frac{1}{k}}\}$

$= (k/n)^{\frac{1}{n-\ell}}$ for all n sufficiently large,

and $\theta = \theta(n) = \sqrt{(1/n)(1-k/n)}$.

Thus we have

$$1-x_0^* < 1-x_0 < 1-\frac{1}{(1+\theta)^{1/k}} x_0^*.$$

The left hand side of this last inequality is clearly asymptotic to $\log n/n$ as $n\to\infty$. The right hand side, unfortunately, is not:

$$1-\frac{1}{(1+\theta)^{1/k}} x_0^* = 1 - (1-\theta/k)(1-\log n/n) + \text{lower order terms}$$

$$= \frac{\sqrt{\ell}}{k}\frac{1}{\sqrt{n}} + \text{lower order terms.}$$

Thus, we need another argument. Note that we have $x_0 < x_0^*$.

It follows that

$$(n-1)/n - x_0^k > (n-1)/n - (x_0^*)^k$$

$$= 1 - 1/n - (k/n)^{\frac{k}{n-\ell}} \text{ for n sufficiently large.}$$

Then $|(n-1)/n - x_0^k| \geq 1 - 1/n - (1-\log n/n)^k - o(\log n/n)$

$= 1 - 1/n - (1 - k\log n/n) - o(\log n/n)$

$= k\log n/n - o(\log n/n)$

so that for all n sufficiently large we have

$$|(n-1)/n - x_0^k| \geq k\log n/2n.$$

From (3) it follows that

$$|k/n - x_0^{n-\ell}| \leq \frac{\ell/n\,(1-k/n)}{\frac{k\log n}{n}} x^{n-\ell+k}$$

$$\leq \frac{2\ell}{k\log n}(1-k/n)(x_0^*)^{n-\ell}\ (x_0^*)^k$$

$$\leq (k/n)[21/k\log n]$$

so that

$$x_0^{n-\ell} \geq (k/n)[1-21/k\log n]$$

$$x_0 \geq [k/n]^{\frac{1}{n-\ell}}[1-21/k\log n]^{\frac{1}{n-\ell}}$$

$$= (1-\log n/n)(1-21/kn\log n) + o(\log n/n)$$

so

$$1-x_0 \leq \log n/n + o(\log n/n)$$

and our proof is complete.//

From the above theorem we see that for _any_ fixed values of k and 1, for sufficently large n the maxima of $f_{k,\ell}(x)$ are near x=1-p, p=logn/n. Since the individual summands of the E_k(1-logn/n) are near their maximum values for these small (as compared with n) k, we anticipate that the E_k for small _k_ are nonvanishing. This would assure us low lattice levels for E in our lattice embedding rank problem.

Observe that since $f_{k,\ell} = (1-x^k)^\ell x^{k(n-\ell)}(1-x^{n-\ell})^{n-k} = f_{n-k,n-\ell}$, the above theorem also proves the

Corollary. Let x_0 be as in the theorem. Then for each n-k, n-1, k and 1 fixed,

$1-x_0(n) \sim \log n/n$ as $n \to \infty$.

Reasoning similar to that in the discussion following the theorem indicates that $E_k(1-\log n/n)$ is nonvanishing for large (i.e., near n) values of k and l. What can be said about $E_k(1-\log n/n)$ for nonextreme values of k and l? How does $E_k(x)$ behave when $p=1-x=\log n/n$? A heuristic discussion of these questions follows.

4. Heuristic approach

In this section we make several conjectures and suggest possible avenues of proof. We stress at the outset that there are no theorems proved in this section.

In section 3 we wrote $E_k(x)=\binom{n}{k}\sum_{l=0}^{n}\binom{n}{l} f_{k,l}(x)$. We showed that for k and l both low and k and l both high the $f_{k,l}(1-\log n/n)$ were near their maximum values on (0,1). That is, we studied the summands of E_k in order to better understand the behavior of E_k. We will use this approach throughout the section.

Conjecture. Full lattice length, hence full rank=n is attained when $p=\log n/n$.

To prove this, we must show that for middle range values of k, $E_k(1-\log n/n)>1$. One approach would be to look at the distribution of (k,l) in the table for a <u>fixed</u> x which maximize $f_{k,\ell}(x)$. We can do this by setting $x=\min\{ \alpha^{\frac{1}{n-\ell}}, (1-\beta)^{\frac{1}{k}} \}$, and seeing where the level curve (for our fixed x) lies. Since these approximate roots are equal when $\alpha \log \alpha = (1-\beta)\log(1-\beta)$, the table is divided into four regions (see figure 19). Note the symmetry with respect to the line l=n-k.

In each region of the table, setting x equal to the appropriate root gives the loci

$$x=\alpha^{\frac{1}{n(1-\beta)}} \quad \Rightarrow \quad x^{n(1-\beta)}=\alpha \quad \text{and}$$

$$x=(1-\beta)^{\frac{1}{n\alpha}} \quad \Rightarrow \quad \beta = 1-x^{n\alpha} .$$

For each fixed x then, we obtain a monotonically increasing function $\beta(\alpha)$ which has negative curvature (see figure 20). (These facts are immediate from looking at the first and second derivatives of the functions.)

Rough asymptotics suggest that the values assumed by the $f_{k,\ell}$ along the level curve $\beta(\alpha)$ are nonvanishing. If this estimate can be made sharp, we could have a proof of the conjecture: Since the curve passes near some l for every k, then for each k there is a set of l values for which $f_{k,\ell}$ is nonvanishing. Then $E_k=\binom{n}{k}\sum_{\ell=0}^{n}\binom{n}{\ell}f_{k,\ell}$ would be high for all the k, and we would have full lattice length.

The rank (vector embedding) problem has been implemented in the APL programming language. These computer studies show that for $10<n<100$ the tables entries assuming a values greater than 1/e are distributed as in figure 21, and that maximum rank n is attained when $p=\log n/n$.

Computer studies also indicate that as p increases along $[\log n/n,1]$, rank(p) decreases as the linear function $(1-p)n$ (see figure 22). We give a heuristic rationalization of this linear descent.

We begin by rescaling the table of values $f_{k,\ell}(x)$ as $F_{k,\ell}(x)=\binom{n}{k}\binom{n}{\ell} f_{k,\ell}(x)$. ($E_k=\sum_{\ell} F_{k,\ell}$.) Observe that $F_{k,\ell}=F_{n-\ell,n-k}$; that is, the rescaled table has the same axis of symmetry $l=n-k$. This symmetry will be heavily exploited in the heuristic arguments below.

We shift our attention to a family of level curves slightly different from maxima loci we were looking at. Instead, we combine the factors $\binom{n}{\ell}$, $(1-x^k)^{\ell}$, and $x^{k(n-\ell)}$ in $E_{k,\ell}$ to form the binomial distribution in l $B(l;1-x^k,n)$ of probability $1-x^k$. We can then write $E_k=\binom{n}{k}\sum_{\ell=0}^{n} B(l;1-x^k,n)\,(1-x^{n-\ell})^{n-k}$, and look at the mean curves of the center of the binomial mass $l=(1-x^k)n$. Since the variance of $B(l;1-x^k,n)$ $\neg\sqrt{n(1-x^k)x^k}$ gets smaller as $p=1-x$ increases, it may be reasonable to look at these mean curves for larger values of

p. (Note that in the rescaled table of $F_{\alpha,\beta}$, this curve is $\beta=1-x^{n\alpha}$. We saw this curve before in two of the four regions of the maximizing loci discussed above.)

Some preliminary rough asymptotics indicate that with $1\leq k, l\leq n-1$, $\mathrm{rank}(p)\geq(1-p)n$ for p in $[\log n/n, 1]$. If the mean level curves give a good approximation to the location of high values of $F_{k,l}$, such asymptotics could tell us where points in the table are nonvanishing. This would provide a lower bound on the expected lattice length. We now seek a rationale for an upper bound.

One such argument which works for only the <u>center</u> of the binomial mass, but not the spread around it, is presented here. It shows that $\mathrm{rank}(p)\leq(1-p)n$ if it true that $E_k\leq\binom{nx}{k}$. (To get the conjectured rank result, the values of $F_{k,l}$ can not get too large. Otherwise the large values would require binomial coefficents of larger size in the embedding problem, which would increase the rank.) Until the values spread around the curve are controlled as well, this is not a proof.

Observe that for $k=0, l>0$, $F_{k,l}=0$, and

$k<n, l=n$, $F_{k,l}=0$,

i.e., the left and top borders of the table are identically 0. By the symmetry of the table about $l=n-k$, where the level curve $l_p(k)$ hits the left border shows where the curve hits the top border.

From $l= (1-x^k)n$, the value of l for which the curve hits the left border $k=1$ is

$$l_0= (1-x)n= (1-(1-p))n= pn.$$

By symmetry of $F_{k,\ell}$, the corresponding value of k is

$$k_0= n-l_0 = (1-p)n.$$

This means that the only values of k such that E_k can be high are $k \leq k_0$. The expected lattice length would then be $\leq k_0 =(1-p)n$.

Plans for future research with this material are discussed in the next section.

5. Future work

The first order of business, of course, is to turn the conjectures and their rough sketches of proofs into precise theorems with proofs. This section looks at some of the problems remaining even after this has been done.

Computer calculations suggest that for a fixed p, as the distribution of (k,l) values approaches the axis of symmetry

$l=n-p$, the values of $F_{k,\ell}$ increase. We would like to turn this into some kind of monotonicity statement about the level curves. The conjecture could be stated as: Let p be in (0,1) and let c be the intersection point of the level curve $\beta_p(\alpha)$ with the axis of symmetry $\beta=1-\alpha$. Then the values of $F_{\alpha,\beta}$ along the level curve $\beta_p(\alpha)$ are strictly increasing on [0,c] and strictly decreasing on [c,1].

One consequence of this conjecture would be that full rank=n is achieved at p=logn/n. This is because we proved in section 3 that the F values are nonvanishing for extreme (both high and low) values of k and l, and this monotonicity conjecture fills in the intermediate values of k. This would mean that $E_k=\sum_{\ell} F_{k,\ell}$ would be greater than 1 for each k, and the full lattice length would then imply full rank.

We would also like to prove that the rank function is linear for small p, i.e., on [0,logn/n] or on [0,1/n]. We are trying to do this using the same approach of understanding the behavior of the table entries. There may be some additional symmetry in p (perhaps by reflection across the point logn/n) which would let us use any results from the large p case to to work with the case of small p.

Analyzing the rank function would be easier if we had explicit expressions for the E_k, and could avoid using the

$F_{k,\ell}$ altogether. We list three unsuccessful attempts at this. Instead of writing E_k as a sum of a function against a binomial kernel, for large n we can approximate this by integrating the function against the Gaussian kernel. A second approach would be to work with different indices of summation. We could do this by summing over k instead of l to obtain expected numbers of image sets, or by binomially expanding the function $(1-x^{n-\ell})^{n-k}$ as a series. A third line of attack might be to apply the Fourier tranform to the convolution E_k. What appears to be rendering these attempts unsuccessful is that the varying index of summation l is present *inside* the exponential term $(1-x^{n-\ell})^{n-k}$.

6. Applications

In this section we apply the conjecture which says that rank(p) is maximized at logn/n to immunology and discrete mathematics. In the discussion, we use the conjecture of linear descent in p as well. We begin with the network theory of the immune system.

In a young immune system, the number of clones $n=10^6$. It is experimentally observed that $\mu=10$ to 20 clones respond to a given Ag. The probability p of a given clone responding is μ/n. Since $\log 10^6 = 13.8$, the rank maximizing probability

$p=\log n/n$ sits squarely in the middle of the the range of actual probabilities μ/n. If the conjecture is true, this means that the recognition of nth level network clones by (n+1)st level network clones optimizes the rank function rank(p). If we take the rank of a 0-1 matrix to be a measure of the amount of information it can transmit, this would show that the young immune network has just the right recognition probability to maximize the transfer of information from one level to the next.

In the old immune system, however, since $p=10^{-5}$ is fixed, the increase of n to 10^7 causes a linear increase in the average number of responding clones to $\mu=100$. Logn increases much more slowly: $\log 10^7=16.1$. Since $\log n/n < \mu/n$ by a factor of 10, the old immune network does not optimize the collective recognition of one level by the next as well as a young network. The aged immune network should retain its function, however. Since rank(p) is conjectured to be bounded below by $n(1-p)$ in this range, the expected rank should decrease from n to n-10. Since n is a very large number, the rank wouldn't decrease by all that much. We can speculate that only in the third or fourth network levels will the effects of aging be apparent.

Our last comment on the immune system concerns the effect of drugs which alter network connectivity. From the conjectured sharp descent of rank(p) in [0,logn/n] as we move to the left of p=logn/n, agents such as cortisone which decrease the suppressive network tone would cause a dramatic reduction in the expected rank. The resulting loss of information flow might render the network inoperative.

Consider a pharmocological agent having the effect oppcsite to that of cortisone, namely that of increasing suppressor tone. This would bring the rank function over to the right side of p=logn/n. By the conjectured linear descent, the slope in this region would be far more gradual. Such a drug might give a physician finer control over the amount of information passed through the immune network. We shall see in Chapter V that, unfortunately, such a drug would also have the effect of decreasing the stability of the immune system.

We now direct our attention to some applications of the rank function to discrete mathematics. Recall that the specificity labelling problem, the Boolean matrix factoring problem and the 0-1 matrix covering problem were shown to be equivalent to one another in Chapter III. Since rank(p) is a model of the expected number of specificities needed to label a 0-1 matrix, it is applicable to these other problems (and their applications) as well.

One application of the matrix covering problem is in the use of Karnough maps (truth tables) in the design of digital circuits. Briefly, a Boolean output function f which assumes the value 0 or 1 and depends on n 0-1 variables x can be written in disjunctive normal form (DNF). (A function is said to be in DNF when it is written as the logical sum of products of the form $x_1\bar{x}_2\ldots x_n$, where $\bar{x}$ means the negation of x.) For n=s+t, the values f assumes can be written out in a $2^s \times 2^t$ 0-1 matrix, each entry in the matrix corresponding to $f(x_1,\ldots,x_n)$, $(x_1,\ldots,x_n)$ in $\{0,1\}^n$. The minimal number of elementary specificity covers (ECM's were discussed in Chapter III) needed to cover this matrix, i.e., its rank, is also the minimal number of terms in any DNF of f. Our rank conjecture tells us that the maximum number n terms would be needed when there are about nlogn 1 entries (out of a possible n^2) in the matrix. This has application to the wiring density of logical circuits and semiconductor chip design. The linear descent conjecture could also be of some importance.

As our last application, we look at the maximum number of Boolean independent rows ind(M) of a 0-1 matrix M. By the lemma below ind(M)$\geq$rank(M). Since rank(M) $\geq$ lattice length(M), this lower bound gives some knowledge of the ind function. In particular, we can define the function ind(p), p in (0,1), for any n, which is the expected maximal number of Boolean independent rows in an nxn 0-1 matrix with about pn^2 1's. If the conjecture of this chapter is true, this would show that at $p=\log n/n$, ind assumes the full value n. Furthermore, this number would descend no faster than the function $n(1-p)$ on the interval $[\log n/n,1]$.

Definition. A collection of 0-1 vectors is said to Boolean independent if no one of them can be written as a Boolean combination of the others.

Example.

Put $v_1 = (1\ 1\ 0)$,

$v_2 = (0\ 1\ 1)$, and

$v_3 = (1\ 1\ 1)$.

Since $v_3 = 1 \times v_1 + 1 \times v_2$ (Boolean operations), $\{v_1, v_2, v_3\}$ is not Boolean independent.

Lemma. The number of Boolean independent rows of a 0-1 nxn matrix is greater than or equal to its rank.

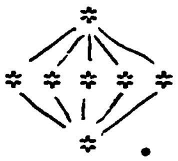
Proof. Let R be an nxn 0-1 matrix, and call its rows $r_1, \ldots, r_n$. If r_n is a Boolean combination of some collection of rows $\{r_i \mid i \neq n\}$, the lattice of image sets is the same as that of the $(n-1) \times n$ matrix R' formed by deleting the last row of R. From Chapter III we have that $\text{rank}(R') \leq \min\{(n-1), n\} = n-1$. By induction, it is clear that every

Boolean dependence reduces the rank by at least 1.//

Observe that the converse to the corollary is false: independence of rows does not guarantee full rank. As a counterexample, consider the matrix

$$R=\begin{matrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0. \end{matrix}$$

The Boolean sum of any two rows is (1 1 1 1 1), and since every row of R has at least one 0, no row is the sum of any of the others and the five rows are independent. The only image sets of R though are these five row vectors, the all-1 vector and the all-0 vector. Its lattice looks like



Since the number of different 0-1 vectors containing exactly k={0 1 2 3 4} 1's is {1 4 6 4 1}, respectively, the lattice L of R can be embedded into the complete Boolean lattice on only 4 generators. (That is, (1 5 1) embeds into (1 4 6 4 1).) Thus rank(R)=4<5=n.

V. Population dynamics in an immune network

When one T-cell clone recognizes another, the population of the recognized clone is reduced and the recognizers' population is enhanced. Thus far in our work, we have restricted ourselves to static interactions between clones. We now introduce a time component to the immune network and study the dynamics of clonal populations. Our principal concern will be those features of the network related to its stability as a dynamical system.

For any individual, the strengths of the chemical interactions between the clones can be written down as an nxn interaction matrix M, where n is the number of clones. *A priori*, nothing can be said about the exact value of each entry m_{ij}, but once an individual's immune system develops, the m_{ij} are fixed. Which clones recognize which can be represented by our nxn 0-1 matrix R, where $r_{ij}=1$ exactly when $|m_{ij}|$ is greater than some prescribed threshold chemical affinity. p=(number of 1 entries in R)/n^2 is the fraction of non-zero (above threshold) values of M. We can use M to write down n coupled predator-prey (recognizers-recognized) equations as a first-order non-linear system. For n>2, there are no easy mathematical solutions to this system; in the immune network, $n=10^6$ to 10^7.

For a large linear system of equations, however, we can regard M as an element randomly chosen from a very large sample space of suitable matrices. We are then able to use the May-Wigner Stability Theorem [Ma ii)] to discuss its almost surely stability properties. We begin then by formulating network dynamics in the framework of this stability theorem. Our exposition closely follows and paraphrases [Ma ii)].

Consider an immune network with n clones which obeys some nonlinear set of first-order differential equations. The stability of a possible equilibrium of the network may be studied by Taylor-expanding in the neighborhood of the equilibrium point, so that the stability of the possible equilibrium is characterized by the equation

(1) $dx/dt = Ax$

Here x is the column vector of the disturbed lymphocyte clonal populations x_i, and the nxn interaction matrix A has elements a_{ij} which characterize the effect of clone j on clone i near equilibrium [Ma i)]. The network interaction matrix R determines which a_{ij} are zero (no recognition), i.e., $r_{ij} = 0 \Rightarrow a_{ij} = 0$, and the type of interaction determines the sign and magnitude of a_{ij}. (If clone i is recognized by clone j, then the interaction is suppressive, while if clone i recognizes clone j, the population of clone i is enhanced.)

Following May, we suppose that each of the n clones would by itself have a density dependent or otherwise stabilized form, so that if it were disturbed from equilibrium it would return with some characteristic damping time. To set a time scale, these damping times are all chosen to be unity: $a_{ii}=-1$. When the network interaction are "switched on", it is assumed that each interaction between two clones is is equally likely to be positive or negative, i.e., enhancing or suppressive. Each of the matrix entries is a random variable, which is distributed with mean 0 and variance α^2. α is interpreted here as the average interaction "strength", that is, the average binding affinity which one clone's Ag-binding site has for the epitope of another clone.

We have then

(2) $A= B-I$

where B is a random matrix and I is the unit matrix. Our infinite sample space is the collection of all matrices with real valued entries, with a measure on the space determined by α and the rules described above. As in our work with random matrices in the previous chapters, randomness only enters in the initial choice of the coefficents a_{ij}. Once they have been chosen to get a specific system, all subsequent analysis is deterministic.

The system (1) is stable if and only if all eigenvalues of A have negative real parts. For fixed n and α we may ask what is the probability $P(n,\alpha)$ that a particular matrix in our sample space will correspond to a stable system. May used the techniques introduced by Wigner [Wi] to show that such a matrix will almost certainly be stable ($P\text{->}1$) if

$$\alpha < 1/\sqrt{n}$$

and almost certainly unstable ($P\text{->}0$) if

$$\alpha > 1/\sqrt{n}$$

May then introduced the connectance p, which in our setting is the probability that any one lymphocyte clone recognizes another. It is a measure of the fraction of non-zero entries in the matrix B. The matrix entries in B now either, with probability p, are drawn from the previous distribution, or, with probability 1-p, are zero. For large n, $\alpha^2 p$ takes the place of α^2, and letting $P(\alpha,n,p)$ be the probability that a particular matrix in our sample space is stable, we state the

<u>May-Wigner</u> <u>Stability</u> <u>Theorem</u>. Let $\epsilon > 0$ be given, $\alpha = \alpha(n)$, $p=p(n)$. Then

$P(\alpha,n,p)\text{->}1$ as $n\text{->}\infty$ provided $\alpha^2 np < 1-\epsilon$, for sufficiently large n,

and, conversely,

$P(\alpha,n,p)\text{->}0$ as $n\text{->}\infty$ when $\alpha^2 np > 1+\epsilon$, for sufficiently large n.

Put μ =np. Roughly, this result says that if the average interaction strength is greater than a threshold value determined by μ, then the system becomes unstable. This threshold value becomes smaller as the size of the system and its connectance increase, scaling as $1/\sqrt{\mu}$. We now use this theorem to draw some conclusions about the stability of the immune network.

The aging of an individual's immune network was discussed at length in Chapter II. There we defined μ =np, the average number of clones recognized by a given T-cell. We argued in Chapter I that over an individual's lifetime, μ increases from 10 to about 100. What does this tell us about the stability of the immune network as a person ages?

α is relatively fixed number in a species. So as μ moves from 10 to 100, the stability constraint

$$\alpha^2 > 1/np = 1/\mu$$

changes from $\alpha^2 > 1/10$ to $\alpha^2 > 1/100$.

That is, the stability threshold drops by an order of magnitude as the immune system ages. So as a person becomes older, the chance of an instability arising in his immune network greatly increases.

An instability leads to the proliferation of one or more lymphocyte clones. Disruption of a network equilibrium may impair immunologic memory. An example of such an impairment would be an inability to recognize tumor cells and other invaders.

As a second application of the above theorem, we look at the long term biological evolution of the immune network. In Chapter IV we used the rank of the 0-1 interaction matrix R as a measure of network information. We conjectured that rank, hence information, is maximized for a random Boolean matrix when $p=\log n/n$. We observed that this is exactly the value of p in the modern mammalian immune network. We now use the May-Wigner stability theorem to discuss the evolution of the Ag-combining site.

Assuming that evolution has optimized the rank measure of information in the immune network, the stability criterion becomes

$$\begin{aligned} \alpha^2 &< 1/np \\ &= 1/n(\log n/n) \\ &= 1/\log n, \text{ or} \\ \alpha &< 1/\sqrt{\log n}. \end{aligned}$$

Over the course of evolution the number of clones n has increased. Therefore the stability threshold $1/\sqrt{\log n}$ has decreased. For the network to remain stable, then, the average interaction strength has decreased as well. This would imply that the average chemical affinity strength of the Ag-binding site for Ag (e.g., epitopes) has been steadily decreasing over evolutionary time. (This is consistent with the biochemistry of proteins. When the earliest Ig-like antibodies emerged, they were probably derived from some ordinary high-affinity, small binding site protein. Although it is highly specific, the modern Ig forms a very loose non-covalent bond to Ag, and has an unusually large binding site.)

Our last application of the theorem is to the effect of certain drugs on the immune system. We remarked in Chapter IV the slow descent of the rank function for random matrices on [logn/n,1] might have therapeutic application. It was suggested tht agents which caused a nonspecific increase in immune suppression would give a physician fine control over the extent of information transmitted in the immune network. (This is because an increase in suppression means an increase in the network recognition probability p, which moves us from p=logn/n over to the slow descent region of rank(p).) Agents such as cortisone, which have the opposite effect of causing a decrease in network suppressor tone, move p to the left and cause a sharp reducation of rank. The stability theorem shows us why drugs such as cortisone are the preferred treatment.

As we move to right on [0,1], starting from p=logn/n in increments s of logn/n, the stability criterion becomes

$$\alpha^2 < 1/np$$
$$= 1/n(s\log n/n)$$
$$= 1/s\log n.$$

So linearly increasing p causes a steady decrease in the stability threshold. Suppression enhancing drugs cause little reduction in network rank (information), but make instabilities inevitable. Cortisone-like drugs, on the other hand, may severely impair the immune network's functioning, but have a strong stabilizing effect. In the long run, this is better for the patient.

(Note: a proof of the May-Wigner stability theorem for a.s. connected systems has been given by Hastings [Ha] which is based on the random graph results of Erdos and Renyi we used in Chapter II. This is applicable to our situation because the immune network is a.s. connected.)

A. Immunology

1. The problem

The warm-blooded vertebrate is a perfect culture medium for a wide variety of micro organisms. The purpose of the immune system is to recognize and eliminate those bacteria, viruses and tumor cells which try to invade the body. Immunology is the study of how the immune system accomplishes this task, particularly how it "remembers" the highly virulent invaders which repeatedly attempt to reinfect the body.

In dealing with these foreign organisms, the immune system must take care to minimize the damage to the host body it is trying to protect. What is required, therefore, is a highly specific mechanism to recognize what entities are foreign. Once foreign invaders are specifically recognized, the lytic (i.e., cell killing) action can be directed mainly against these offenders and not against the host cells.

The problem of recognizing (and remembering) foreign infectious agents is not trival. After all, most biological life is built up from the same building blocks. Viruses and tumor cells are direct descendants of our own DNA and cells. Bacteria and viruses play the coevolution game to escape detection by the immune system. Clearly then, the immune system needs to make many measurements, each of which should be "narrow range" and specific. Otherwise, there is no hope of distinguishing "self" from "not-self".

Once recognition has been achieved, some action must be taken to eliminate the invaders. There are any number of mechanisms which have evolved to kill or impair cells; these will not concern us in the sequel. Our topic here is that first step of specific recognition. We will now explore the nature of the specific measurements by which "self" is distinguished from "not-self".

2. Chemical recognition

The immune system detects foreign invaders by their chemistry. Some kind of molecule, then, is needed to perform this task. We would expect a stable molecule, held together by covalent bonds. The recognition would be effected by a loose, non-covalent association of the molecule and a chemical moiety on the foreign invader. The non-covalent bond would form only within a narrow range of specificity, and its formation should act as a switch to transmit the information that a foreign chemistry has been detected.

Note: A <u>chemical bond</u> is a quantum mechanical association between two (or more) molecules. The degree to which the molecules are held together, or <u>affinity</u>, is determined largely by the attractive forces of the electromagnetic fields induced by negatively charged electrons moving around the positively charged nuclei of each molecule and by the three dimensional geometry of the associating molecules. One kind of very stable (under biological conditions) bond is the covalent bond. Lower affinity bonds between biomolecules are called non-covalent bonds.

3. Proteins

Proteins are molecules which have these properties: A protein is a covalently bound macromolecule which is put together from simple amino acid building blocks. Each of the twenty amino acids has its unique special functional group which contributes to the overall chemistry of the protein molecule, and left and right linking parts which permit it to be connected to adjacent amino acids. Thus, a protein with n amino acids may be considered as a word of length n constructed from an alphabet with 20 letters. The real-world chemical properties of the linear word (i.e., the amino acid sequence) cause it to fold up into a particular three dimensional biomolecule.

The active site or binding site of the protein can form a loose non-covalent bond with a narrow range of other molecules. This narrow range of specific binding is determined by the quantum chemical charge configuration and the geometry of the active binding site. (These properties of the binding site are determined by the arrangement of the particular few amino acids which form the active site.)

Proteins are often two state molecules. When some substrate (i.e., a second molecule) binds with high affinity at the active site, the three dimensional configuration of the entire protein molecule changes. In this way, the information that the substrate has been bound can be transmitted to the other end of the molecule and thus affect some other process. This two state situation which depends on specific binding is sometimes called the "lock and key" model, since binding of some particular substrate serves as the key which unlocks (or changes the state of) the molecule.

4. DNA

The linear arrangement of the sequence of amino acids for a given protein is stored in the DNA (Deoxyribo-Nucleic Acid) of a cell. DNA is an essentially linear molecule comprised of sequences of the four bases Guanine, Adenine, Cytosine and Thymine. That is, a DNA molecule with n bases is a word of length n built from the alphabet {G,A,C,T}.

The DNA codes for protein by coding for each amino acid in the order in which it appears in the protein. Observe that $16=4^2 < 20$ (the number of amino acids) $< 64=4^3$, where 4 is the number of letters in the DNA base alphabet. This requires that there be at least 3 bases coding for each amino acid, and in fact, there are exactly 3. With the exception of a few start and stop codings, the set of all base pair triplets is mapped onto the 20 amino acids. This map is called the genetic code. Thus every linear DNA sequence determines a unique amino acid sequence which then folds into a protein molecule. The mechanism: DNA -> mRNA -> amino acid sequence and protein. DNA is the permanent record, mRNA (messenger Ribo-Nucleic Acid) is a carbon copy of the DNA record which can be spliced and edited and ultimately used as the template for the functional protein molecule.

5. Uncoupling of DNA coding and the functional protein

Given the DNA code for a particular protein, the questions arise (1) what is the protein's three dimensional configuration and (2) what are its chemical binding properties? What physiological function does it perform? It may be that the three dimensional folding of a sequence of amino acids produces some abortive, dysfunctional biomolecule. A viable binding site may not even be present. These are among the most difficult problems in theoretical chemistry. While there has been some progress in computing protein folding, computing the binding properties of the active site remains an essentially unsolved problem.

Nature, however, does not make quantum mechanical calculations. It physically constructs a protein and then empirically tests its binding affinities to various substrates. If the protein (usually a slight variant on a protein from the last generation) does no harm to the organism, then the protein's DNA code may be preserved for its descendants.

So Nature's answer to the question, "Given a protein's DNA code, what is the chemistry of its binding site?", is that this is unknowable in advance of some empirical test. We will therefore <u>assume</u> in the sequel that in advance of any experiment, the binding properties of a protein with some given substrate is a <u>random</u> <u>variable</u>. It is also reasonable to assert that a small change in the code (say by one amino acid near the binding site) will result in smaller change in binding activity than a larger change. We will therefore also assume that a small change in the code will produce a small random change in the binding chemistry.

6. Immunoglobulin

The recognition protein actually used in the immune system is the immunoglobulin (Ig) molecule. Ig is a symmetric molecule comprised of two heavy chains and two light chains (see figure 23). The heavy chain has four domains, each of which contains about 100 amino acids. (A domain of an Ig chain is a physiologically and evolutionarily useful way of understanding the molecule.) Three of the heavy chain domains are called constant, and are written as C-H (constant domain, heavy chain). The end domain of the heavy chain is the variable one, designated V-H, and is used to form the Ig binding site in association with the light chain. The light chain has two domains, C-L and V-L. The variable regions are responsible for recognition, while the constant regions are involved in transmitting and acting on the chemical binding recognition signal.

The variable domain is so named because the many Igs in an individual differ considerably in the particular amino acids occupying a given position (1-100) in the variable domain. (In a constant domain, the amino acids appearing in each position are almost invariant.) The number of different amino acids appearing in a position is especially marked in the hypervariable region of the variable domain. The two hypervariable regions of the V-H and V-L chains associate to form the binding site of the Ig molecule. Thus by changing amino acids in the hypervariable region, the binding affinities of the active site can be changed considerably.

The binding site of the Ig molecule is a large one, with each V domain contributing 10 to 20 amino acids. A lower bound on the number of possible binding regions is then $(20^{10}) \times (20^{10}) = 20^{20}$, since there are 20 different amino acids. Even if a large fraction of these possibilities are chemically similar or form nonfunctional molecules, we are still left with a large set of Igs. This provides a large basis set of specific, narrow range chemical binding measurements.

7. Cells and receptors

The cell is the smallest unit of biological organization. It is made up of thousands of protein, DNA, and other molecules. Cells are derived from other cells, and most can split in two to produce a new one. Separating the cell from the outside world is the cell membrane or cell surface. Among other functions, it controls the flow of chemicals and information between the inside and outside of the cell. Inside the cell is the nucleus, which contains the DNA arranged as genes (protein encoding functional units) on chromosomes (strings of genes). The genes, and the extent to which each is turned on, regulate the metabolic activity of the cell. (For example, Ig can be produced only when the gene encoding it in the nucleus is turned on.) The details of the other cellular organelles will not be needed for our discussion.

A receptor is a cell surface recognition unit. Upon chemically sensing a specific substrate outside of the cell, the receptor (often a protein molecule) transmits the signal across the membrane that this specific chemical configuration has been recognized. This message is then relayed to the nucleus (often via some other specialized molecule), which can then act on the information.

By using intercellular messenger molecules, receptors allow the cells of a multicellular organism to communicate and thus coordinate their activity. For example, the insulin hormone released by pancreas cells informs other cells carrying the appropriate receptor that the blood sugar level is elevated, and these other cells can then act accordingly. As another example, consider the immune response. Once a foreign organism has been detected, its elimination requires the coordination of B-cells, T-cells, macrophages and many other cell types. Furthermore, the lymphocytes (B- and T-cells) involved will not be uniform in type, but come in thousands of different subpopulations. An effective immune response will use hundreds of different short-range and long-range intercellular signals, each of which is communicated by cell surface receptors.

All of the receptors discussed above have had their binding sites shaped by considerable coevolution of the receptor with the specific chemical messenger it binds to. What makes the Ig molecules so remarkable is that these receptors arise <u>de novo</u> each generation, and yet are able to detect foreign chemistries <u>with which they have had no prior experience</u>. As we will discuss in the next section, it is precisely because the Igs are created each generation as random receptors that they can do this.

8. Factor analysis

When a foreign invader is introduced to the body, it is the job of the immune system to detect and eliminate it. Since this detection is accomplished chemically, we call the invader antigen (Ag), by which we will mean that chemistry of the invader responsible for initiating an immune response.

The average mammal has 10^6 to 10^7 different Ig molecules with which to test and chemically measure the Ag. Of these, approximately 10^{1-2} will actually bind the Ag with some (greater than threshold) affinity. This small subset of the individual's Igs is the operational definition of the Ag for the individual. It is by this collection of Igs that the body can name and identify a foreign Ag, and it is the proliferation of this subset that is responsible for immunologic memory.

In essence then, a factor analysis is performed on the Ag with a very large basis set. Although the probability of a high affinity interaction of Ag with any particular Ig is low, there are enough basis elements to (probabilistically) guarantee that a moderate number of Igs (basis factors) will "tag" the Ag. This subset of Igs is a equivalent to a cluster in factor space on the basis of the individual's complete set of Igs.

9. B-lymphocytes (B-cells)

For the 10^6 to 10^7 dimensional factor analysis of foreign chemistry to work well, the Igs need to be maintained in relative populations. (The relative proportions will depend on the history of immune system exposure to different Ags.) The immune system does this by identifying each Ig with a unique clone of cells.

The cells used for this task are a class of <u>lymphocytes</u> (lymph- or immune- cells). The particular class which manufacures and secretes Ig is the B-lymphocyte (Bursa derived lymphocyte) or <u>B-cell</u>. Each clone can produce exactly one Ig binding site type. This is because a B-cell has the DNA coding for only one V-H region and one V-L domain. The proteins coded by these two DNA strings associate to form the binding site of the Ig molecule.

Thousands of these identical Ig molecules with their unique V-H V-L binding site are displayed on the surface of a clonal B-cell. When the binding site of these molecules recognize a complementary chemical configuration (usually a foreign Ag), the cell is triggered into a new state.

Upon triggering, the B-cell will secrete free Ig molecules into the surrounding tissues to bind to and injure the Ag. It may send out a variety of different chemical signals to other cells to inform them that it has detected an Ag. The B-cell will also proliferate to expand its relative clone size to improve the immune response to the invader over the next few weeks and in the future (immunologic memory).

10. T-lymphocytes (T-cells)

Not all immune responses are effected by the chemical secretion of Igs. Some immunity is mediated by direct cell-cell contact. This is especially true in the killing of tumor cells and viruses. The lymphocytes of cell mediated immunity are the T-cells (Thymus derived- lymphocytes).

As an example, consider the cell mediated killing of virus. The macrophage is a large non-specific scavenger cell which can ingest other cells and viruses. When activated, it can efficiently kill ingested virus particles. A T-cell with surface receptors similar to Ig can detect the presence of foreign viral chemistry and give this information to the macrophage by direct cell-cell contact. This will then activate the macrophage, thereby enabling it to better ingest and to kill the virus. Note that whereas the B-cell secretes a chemical (Ig) to kill an invading bacteria, the

T-cell uses direct cell contact to directly or indirectly (as in our example) impair the offending organism.

There are various subpopulations of T-cells, such as suppressor cells, helper cells and killer cells. Suppressor cells suppress the action of other T-cells and B-cells, while T-helper cells can greatly enhance the B-cell response to Ag. Killer T-cells perform direct cell-cell killing of their targets. Each subpopulation can be further divided into clonal subpopulations by various other properties. One of these properties, of course, is the unique Ag binding site which a T-cell clone expresses on its cell surface. As we shall examine more closely later, T-cells can regulate (e.g., suppress) one another not just by generic subpopulation properties, but by the specific nature of their Ag binding sites as well.

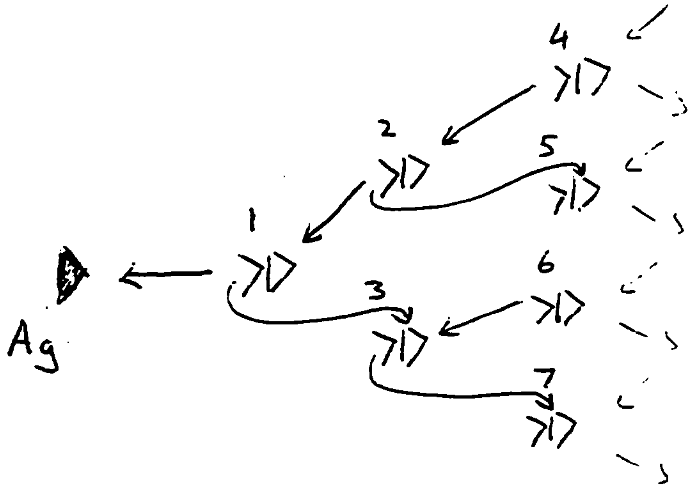


figure 1

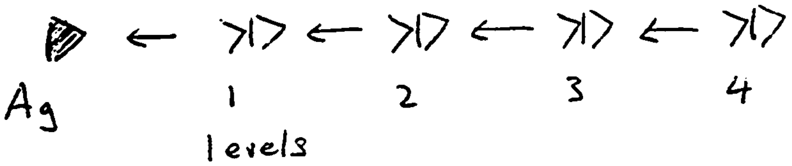


figure 2

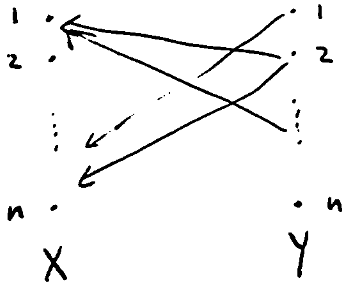


figure 3

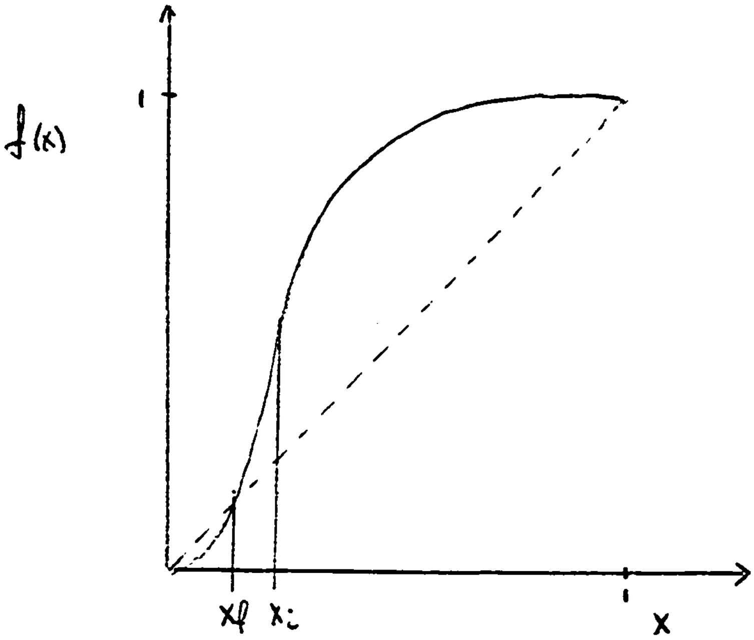


figure 4

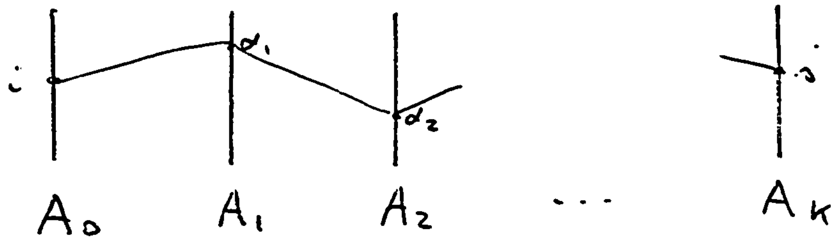


figure 5

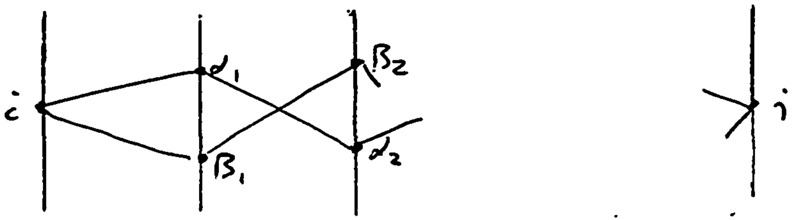


figure 6

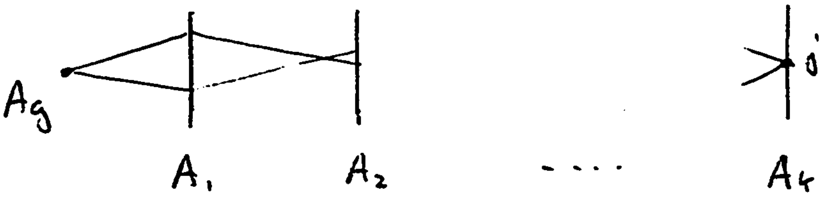


figure 7

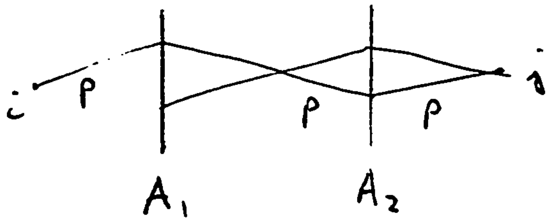


figure 8

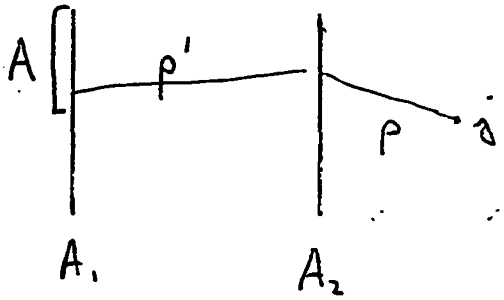


figure 9

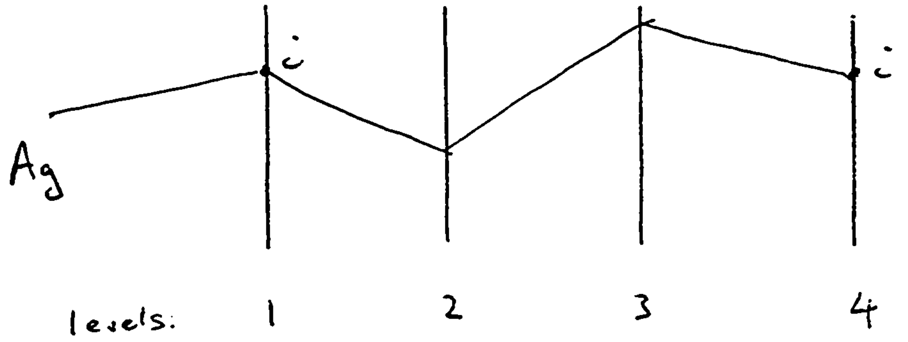


figure 10

$$\begin{array}{rl} & \begin{array}{ccc}1 & 2 & 3\end{array}\ \text{cell groups} \\ \text{sera}\ \begin{array}{c}1\\2\end{array} & \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\end{pmatrix}\end{array}$$

figure 11

$$\begin{array}{rl} & \begin{array}{ccc}\bar{s}_1 & \bar{s}_2 & \bar{s}_1\bar{s}_2\end{array} \\ \begin{array}{c}s_1\\s_2\end{array} & \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\end{pmatrix}\end{array}$$

figure 12

$$\begin{array}{rl} & \begin{array}{ccc}\bar{s}_1 & \bar{s}_2 & \bar{s}_1\bar{s}_2\end{array} \\ \begin{array}{c}s_1\\s_2\\s_1 s_2\end{array} & \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\\1 & 1 & 1\end{pmatrix}\begin{array}{l}\\ \\ -\ \text{serum } 1 + \text{serum } 2\end{array}\end{array}$$

figure 13

$$\begin{array}{c|cccc} & \bar{s}_2 & \bar{s}_1\bar{s}_2\bar{s}_3 & \bar{s}_2\bar{s}_3 & \bar{s}_1\bar{s}_3 \\ \hline s_1 & 0 & 1 & 0 & 1 \\ s_2 & 1 & 1 & 1 & 0 \\ s_3 & 0 & 1 & 1 & 1 \end{array}$$

figure 14

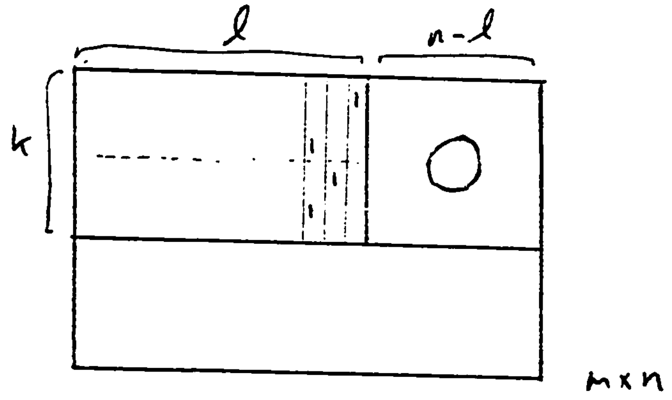


figure 15

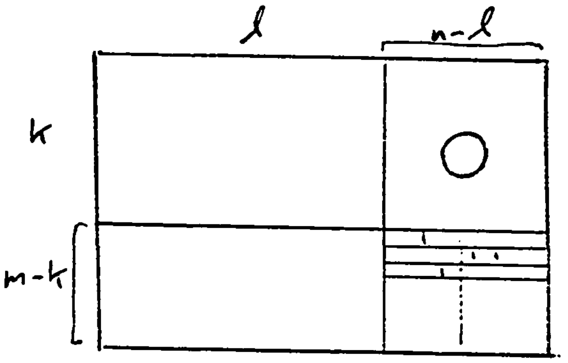


figure 16

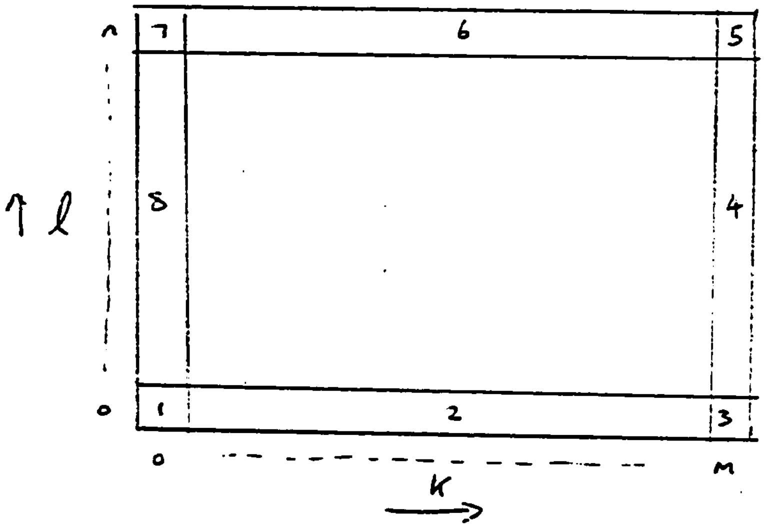
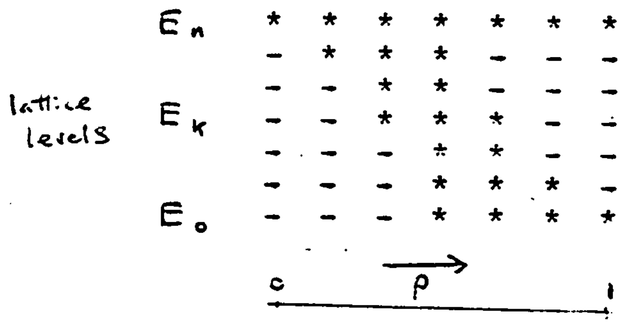


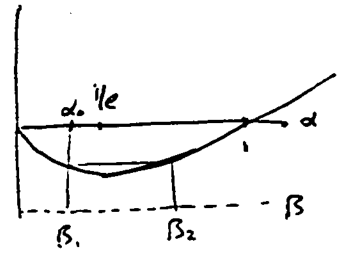
figure 17



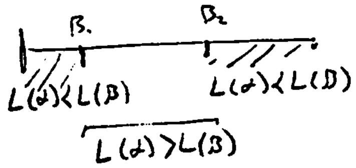
E_n	*	*	*	*	*	*	*
	-	*	*	*	-	-	-
	-	-	*	*	-	-	-
E_k	-	-	*	*	*	-	-
	-	-	-	*	*	-	-
	-	-	-	*	*	*	-
E_0	-	-	-	*	*	*	*

figure 19

Fix α_0.



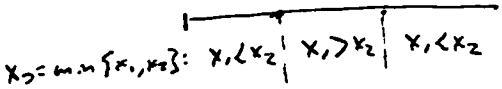
put $L(y) = y \log y$.

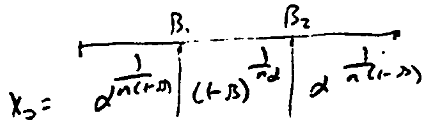


$L(\alpha) < L(\beta)$

$\iff x_1 < x_2$.

Similarly for $>$ and $=$.





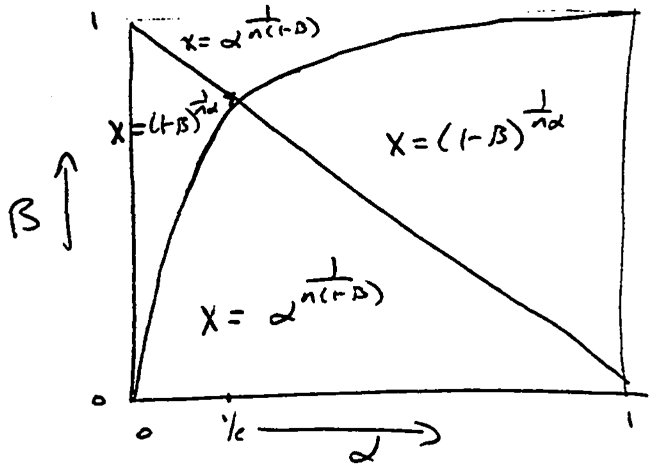


figure 19

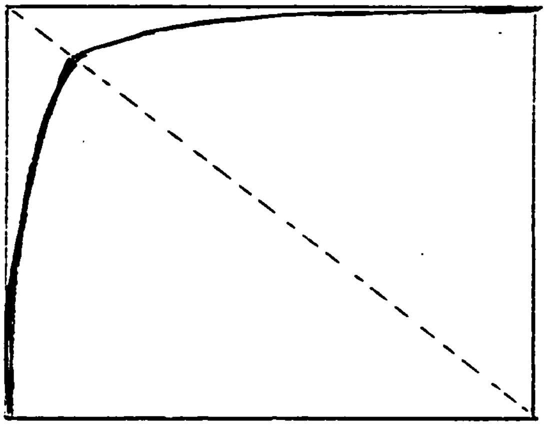


figure 20

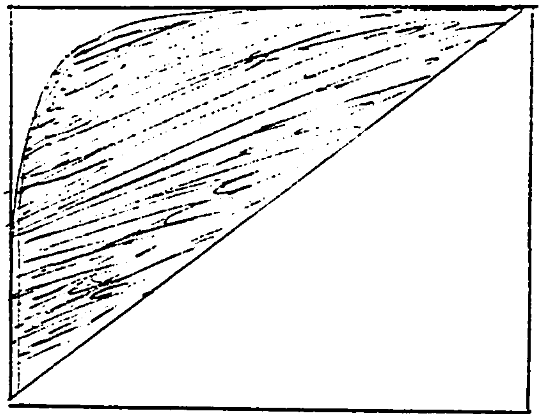


figure 21

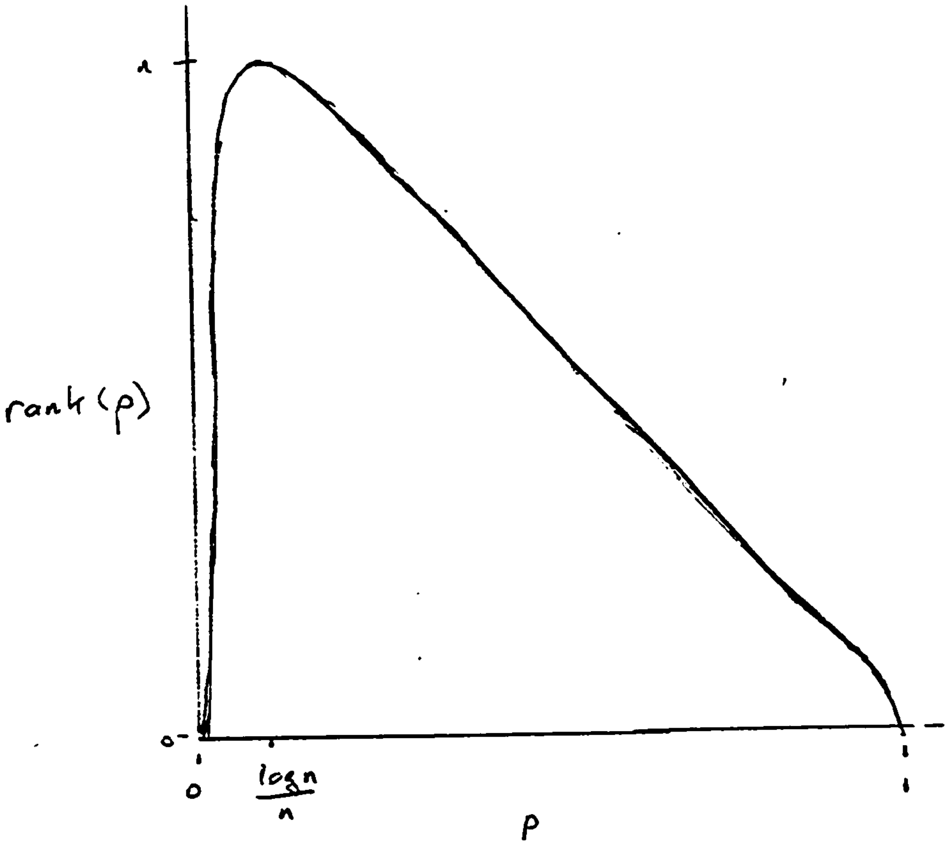


figure 22

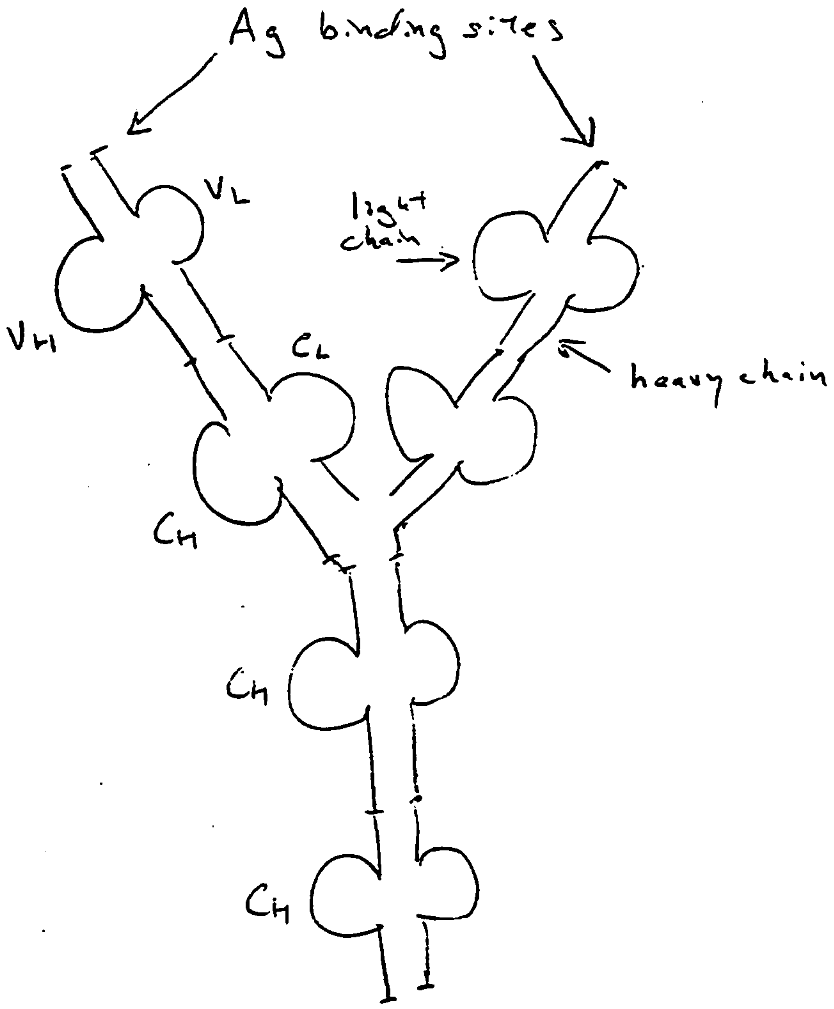


figure 23

The antibody (Ig) molecule.

Bibliography

[B] B. Benacerraf and E.R. Unanue, *Textbook of Immunology*, Williams and Wilkins, Baltimore, Maryland, 1979.

[Bu] M. Burnet, *The Clonal Selection Theory of Acquired Immunity*, Cambridge University Press, 1959.

[C] J.E. Cohen, P. D'Eustachio and G.M. Edelman, "The specific antigen-binding cell populations of individual fetal mouse spleens: repertoire composition, size, and genetic control. *J. Exp. Med.*, Vol.146, pp.394-411, (1977).

[D] D. Dubey, personal communication, (1982).

[E] P. Erdos and A. Renyi, "On the evolution of random graphs", *Magyar Tud. Akad. Mat. Kutato Intez Kozl.*, 5, pp.17-61, (1961).

[F] W. Feller, *An Introduction to Probability Theory and its Applications, Vol.I*, 3rd edition, John Wiley and Sons, New York, 1968.

[G] R.K. Gershon, "The G.O.D.-like role of the T-cell", in: *The Generation of Antibody Diversity* (A.J. Cunningham, ed.), pp.105-126, Academic Press, New York, 1976.

[H] H.M. Hallgreen, J.H. Kersey, D.P. Dubey and E.J. Yunis, "Lymphocyte subsets and integrated immune function in aging humans", *Clin. Immunol. and Immunopath.*, 10, pp.65-78, (1978).

[Ha] H.M. Hastings, "The May-Wigner stability theorem for interaction and competition matrices in ecology", preprint, Hempstead, NY, (1982).

[J] N.K. Jerne, i) "The somatic generation of immune recognition", *Eur. J. Immunol*, 1, pp.1-9, (1971).
ii) "The immune system", *Sci. Amer.*, 229(1), pp.52-60, (1973).
iii) "Towards a network theory of the immune system", *Ann. Immunol.*, 125C, pp.373-389, (1974).

[K] M. Kimura, "The neutral theory of molecular evolution", *Sci. Amer.*, 241(5), pp.94-104, (1979).

[Kl] N.R. Klinman, J.L. Press, N.H. Sigal, and P.J. Gearhart, "The acquisition of the B-cell specificity repertoire: the germ line theory of predetermined permutation of genetic information", in: *The*

Generation of Antibody Diversity (A.J. Cunningham, ed.), pp.105-126, Academic Press, New York, 1976.

[M] G. Markowsky, i) "The representation of posets and lattices by sets", *Research Report* RC6140, IBM Watson Research Center, Yorktown Heights, New York, (1976). ii) "The representation of posets and lattices by sets", *Algebra Universalis*, 11, pp.173-192, (1980).

[Ma] R.M. May, i) "Stability in multispecies community models", *Math. Biosci.*, 12, p.59-79, (1971). ii) "Will a large complex system be stable?", *Nature*, 238, pp.413-414, (1972).

[N] D.S. Nau, G. Markowsky, M.A. Woodbury, and D.E. Amos, "A mathematical analysis of human leukocyte antigen serology", *Math. Biosci.*, 40, pp.243-270, (1978).

[R] P.H. Richter, "A network theory of the immune system", *Eur. J. Immunol.*, 5, pp.350-354, (1975).

[Ro] D.A. Rowley, H. Kohler, and J.D. Cowan, "An immunologic network", in: *Contemporary Topics in Immunobiology* (J.J. Marchalonis and N. Cohen, eds.), vol.9, pp.205-230, (1980).

[W] R.L. Walford, *The Immunologic Theory of Aging*, Williams and Wilkins, Baltimore, 1969.

[Wi] E.P. Wigner, "Statistical properties of real symmetric matrices with many dimensions", *Proc. Fourth Canad. Math. Cong.*, Toronto, pp.174-184 (1959).

[Wil] A.C. Wilson, S.S. Carlson, and T.J. White, "Biochemical Evolution", *Ann. Rev. Biochem.*, 46, pp.573-639, (1977).

[Wo] A. Wohlgemuth, i) "Labelled reaction matrices, a histocompatibility model", abstracted in *Notices Am. Math. Soc.*, 23, pp.A422-A423, (1976). ii) "Abstract immunogenetic systems", *J. Theor. Biol.*, 73, pp.469-508, (1978). iii) "Modelling immunogenetic specificities", *Math. Biosci.*, 45, pp.175-177, (1979).