

## **Interactive Macromolecular Sites**

### **I. Basic Theory**

NORMAN G. ANDERSON

*Department of Surgery, and South Carolina Memorial Cancer Institute,  
Medical University of South Carolina, Charleston, S.C. 29401, U.S.A.*

*(Received 10 January 1975)*

A general theory of the evolution of interactive surfaces or sites on cell macromolecules is presented based on the assumption that all macromolecules present in cells are covered with reactive groups to which counter sites can theoretically be made. Initially such potentially reactive sites are considered to exist in pairs. Once one member of a pair is incorporated into a cell, its opposite number is forbidden except where it may be employed to advantage for structure formation. As more and more potentially interactive sites appeared in cells during evolution, a limit was approached where one member of each pair was present. Such a limit has been termed a perfect set, and would both define a "self" and recognize foreignness. The role of perfect sets in the development of the immune system, in differentiation, and in so-called random fixation of amino acid sequences is reviewed.

#### **1. Introduction**

"...the theory of Evolution in its ordinary technical sense is still one of the liveliest topics in biology, and the next great advance we may look forward to is a new theory of genetic variation, propounded by molecular biologists. ...it is impossible to predict what this advance will be, but it will clearly be something in the nature of a sentential calculus of genetic messages—a natural development of the glossary that is now being compiled of single words."

P. B. MEDAWAR, 1967

Biology admits of rather few central theories which give order and structure to masses of data and which also have predictive value. In this initial paper an attempt is made to summarize a group of problems and concepts concerning interactive sites on large molecules, and to present a unified theory of their evolution. The problems to be dealt with include the definition of self at the molecular level, the origin of active sites on antibodies and of cell mediated immunity, the limits of differentiation in multicellular organisms, mechanisms

of gene fixation, the molecular basis of subcellular structure formation, and the question of whether pathogenic micro-organisms from extraterrestrial sources could exist against which the human immune system can provide no defence. It is not at once apparent that these problems have any simple common denominators, and to demonstrate that this is indeed the case requires the assembly and integration of data and arguments from a wide variety of sources. Since this necessitates a very extended presentation, I have chosen to present a summary of some of the arguments and concepts in this first paper, and to treat each facet of the problem more exhaustively and quantitatively in subsequent papers in this series.

The molecular mechanisms underlying the morphogenesis of such complex organs as the human brain appear to involve vast arrays of specific molecular surfaces which engage in highly selective interactions. In fact, the entire program of embryogenesis appears to be orchestrated solely through the synthesis or cessation of synthesis of macromolecules bearing specific interactive surfaces. It is probable that most of what remains to be learned about cell self-assembly and differentiation will ultimately be stated in terms of those surfaces.

Unfortunately we are ill prepared to deal systematically with complex three-dimensional patterns. With the singular exception of the human face we find it difficult to classify, remember, and intercompare them. For example, who can inspect models of a half-dozen proteins and determine which will interact strongly and which will not? Who can remember arrays of surface interactive groups sufficiently well to deduce that a given new protein differs at one site from thousands of other existing sites—a feat routinely performed by the immune system? Possibly because these core problems are almost inaccessible to direct attack we have often failed to see the importance of attempting to explore the origin and classification of interactive molecular surfaces.

## **2. Definition of Interactive Sites**

The molecules of which cells are composed show a bimodal distribution of molecular weights (Anderson, 1968) reflecting the disproportion in size between proteins and other macromolecules on the one hand, and metabolites on the other. Binding sites for substances in both groups are found on macromolecules, however this discussion is limited almost entirely to interactions between macromolecules.

A widely held assumption is that such inter-macromolecular binding sites are relatively uncommon, and that large numbers of different proteins ought to coexist in solution with little interaction. When interaction does

occur it is often viewed as being due to specially evolved matching or paratactic (Jerne, 1960) sites. We adopt here precisely the opposite view which is that non-matching and non-interaction in protein mixtures is what is unique, and that all macromolecules are covered with surface groupings to which specific counter-surfaces can, in theory, be made. Since the interactions can cover a very wide range of binding energies it is evident that ultimately some threshold must be defined below which interaction will be considered inconsequential.

### 3. Postulates

The core concepts may be stated as propositions or postulates as follows:

(i) *The number of interactive sites which could theoretically be formed on macromolecules is infinite*

This proposition is believed to be true if no limit is placed on the size of the molecules in the theoretical population being considered. Once a restriction is placed on size, as is the case in living cells, the exact opposite is true.

(ii) *The number of potential binding sites found on macromolecules in one organism is limited because both the maximum molecular size and the number of unique molecules are limited*

Two simple facts suggest that limits exist to the number of different binding sites found in one organism. The first is the fact of immunity. If all possible sites normally exist in one higher vertebrate, then all antibodies could produce auto-immunity. The second is the fact of macromolecular toxicity. Many highly toxic and even lethal proteins are known. These possess specific surfaces which react with macromolecules normally present in cells, and one must assume that these toxic surfaces are normally absent. In a subsequent paper the experimental evidence for the existence of large numbers of potentially interactive or paratactic sites on cellular macromolecules is reviewed. Suffice it to note here that all cellular macromolecules can react sufficiently strongly with other macromolecules (enzymes) to allow all to be completely degraded, that large numbers of different antibodies to cellular proteins can be made, that many toxic proteins exist, and that protein enzyme inhibitors, lectins, and protein subunits all interact with high specificity. Quite obviously the size of interactive sites is limited by the radius of curvature of the molecules bearing them.

(iii) *When, during evolution, a new potentially interactive site appears, the complementary site is forbidden unless its presence is required for structure formation*

During early protocell evolution a limited number of proteins were present. Unlimited variation in the amino acid sequence in existing and in new proteins was allowed providing the new sites created did not interfere with the function of any pre-existing protein, or, in the case of a new protein, with its own function. Thus new proteins could not be allowed which were inhibitors of pre-existing important protocell enzymes. This restriction appears to be and was trivial at first. However it became of increasing importance as evolution proceeded because the number of interactive sites possible in cells is limited. For the sake of simplicity we assume that interactive sites (either existing or possible in biological systems) exist in pairs, and that a site is not interactive unless its opposite number can be made. Once a cell adopts one site from the universe of biologically possible interactive pairs, the matching one, except as required for structure formation, is forbidden. As evolution proceeded, and the number of intracellular sites increased very rapidly, a limit was approached, which was equal to one-half of all possible interactive sites plus additional ones equal to the number of paratactic pairs required for structure formation. Such a limiting number has been termed a *perfect set* (Anderson, 1969).

(iv) *No obligatory constraints of polarity govern the choice of which member of a pair is incorporated into a cell*

It has been proposed that antibody and antigen combining sites (which are subsets of the universe of interactive sites described here) are different in kind as keys are fundamentally different from locks (Jerne, 1960). This requires that for most of evolutionary history one class of surfaces was forbidden in anticipation of circulating immunity. It is difficult to imagine a mechanism for enforcing such a rule over millions of years—doubtless one could be postulated. Binding sites contain a variety of moieties whose pattern and topology combine to produce specificity. In certain instances the interacting molecules may be identical, as is the case in some proteins assembled from subunits. That convex sites may appear more often than concave ones, merely because most of the surface of protein molecules must be convex, does not mean that concave sites are automatically forbidden. The heart of the question is whether some separate selection mechanism exists which might, for example, choose sites bearing negative charges and exclude those bearing positive ones. The view adopted here is that no such mechanism exists.

(v) *The collection of potentially interactive surfaces found in a eukaryotic cell approaches a perfect set, and a perfect set defines a "self"*

If, during evolution, any surface possible may be chosen which does not interact with a pre-existing one (except to form an advantageous structure) then randomization of old and new sequences can occur until all or nearly all possibilities are gradually exhausted. The number of possible different perfect sets (or selves) is equal to  $2^N$  where  $N$  is equal to the total number of pairs possible given macromolecules of a limited range of molecular weights. Given tens of thousands, if not millions of genes in eukaryotic cells,  $2^N$  becomes large enough to account for the molecular uniqueness of all members of the animal kingdom, and especially of man.

(vi) *A perfect set automatically contains a mechanism for recognizing foreignness*

A set of reactive surfaces containing one member from each of all possible pairs of interactive sites must, by definition, contain recognition sites for all remaining possible binding sites. As has been frequently stressed (Burnet, 1968) some form of recognition of foreignness (and reciprocally definition of self) exists down through invertebrates to protozoa. The mechanism suggested here to account for recognition is the automatic end result of random variation. It is difficult to imagine a different and separate mechanism evolving from the very beginning solely to serve this function. The mechanism proposed here is thought to be the original and ancient basis for the recognition of food organisms and the origin of cell mediated immunity.

(vii) *No common denominators of either self or of foreignness exist or can exist*

The idea that signals, flags, or special surfaces can exist which are common denominators of foreignness appears to have originated with Burnet & Fenner (1949) and has appeared again and again in theoretical immunology, and is dealt with in detail in a subsequent paper. The basic notion is that an antigen can be recognized as foreign *non-specifically*, and that the recognition mechanism then triggers a specific response. An analogy clarifies the problem. Both the telephone and postal system use a unique number or address for each subscriber or addressee. Enormous savings could be made if the same results could be obtained with much less information. For each reader of this article the world population may be divided neatly into two groups—those with whom he wishes to communicate, and those with whom he does not. All incoming messages need only to be identified as being in one of these two groups to satisfy the requirements of one individual. One need only

contemplate for a moment the problem of doing the same thing for each and every member of the population to see the problem. No theory of the immune response which does not deal with this problem need be taken seriously. The obverse side of proposition seven is that specific recognition sites for any substances reacted to as foreign must pre-exist in the cell in which recognition occurs. This does not imply that the recognition site and the antibody counter-site produced by the immune system will be identical however.

(viii) *Possession of a perfect set does not prevent the evolution of new proteins*

New proteins may be made which possess the same sites found on pre-existing proteins, but in different arrangements. Also new structures may be formed by evolving proteins having sites paratactic to those already present.

(ix) *The restrictions inherent in the perfect set concept explain so-called "random fixation"*

Intercomparison of the amino acid sequences of homologous proteins from many different species suggests that a large number of amino acid substitutions can be made without affecting known functions (Margoliash, Barlow & Byers, 1970). However surprisingly little variation in sequence appears to be allowed within a species (King & Jukes, 1969; Arnheim & Taylor, 1969; Richmond, 1970). The explanation given is that certain sequences become "fixed" even though many of the possible substitutions are neutral (Kimura, 1968). This could easily occur in populations subject to great fluctuations in size. It is difficult to see how this could occur often enough and in enough species to produce the sequence fixation observed. The perfect set concept presented here suggests that very few modifications may be made in any protein in a given cell without producing deleterious interactions. In a special negative sense almost all surfaces have a function.

(x) *The very slow and apparently constant rate of amino acid substitutions in a given protein during evolution is due to the requirement for multiple simultaneous mutations to allow the appearance of a new interactive site and the disappearance of pre-existing sites which would bind the new site*

Fixation of an amino acid substitution occurs very infrequently when compared to the rate at which new substitutions are introduced by random mutations. Multiple simultaneous mutations required to allow persistence of a new site would occur much less often.

(xi) *The generation of "all possible" binding sites of some average size (for example composed of five amino acids) by a random process and with the certainty displayed by the immune system during the short interval required for the immune system to mature is not possible*

The somatic mutation theory of the generation of antibody variable sequences falls on the simple issue of the statistics of achieving each of a large number of surfaces with certainty by a random process. To discuss this problem quantitatively, as will be done in a subsequent paper, requires that estimates be made of the total number of antibodies necessary and possible, the probability of forming all in a limited period of time, and the total number of tries that must be made to achieve all with high probability. If  $N$  is the number of different sites ultimately required, then  $N$  mutational events cannot yield all sites. Some sites will be generated two or more times, many once, and some not at all. The total number of tries required to produce all (or conversely to be sure that a specified one is made) is very many times  $N$ , and when a high degree of probability is required, the number of tries needed can be astronomical. Matters become worse if some sequences are inherently less probable than others, for example if some require four or five base changes instead of one. We may ask, how can we be sure that a new mutant virus will not appear which requires from each and every citizen the most improbable series of mutations as the price of survival? Many games can be played with the variables mentioned in hopes of saving the somatic mutation theory of antibody formation. In my view, however, they all founder on the question of achieving specific surfaces with certainty in large populations in a short period of time by a purely random process. This is especially true if one considers small animals such as shrews weighing only a few grams where maturation is extremely rapid and where the immune system appears to be fully functional.

(xii) *The development of a complete set of variable region genes for all possible antigenic sites purely to serve immune functions is not possible*

If a set of genes for variable sequences produced purely by random mutations is postulated, and if completeness is a requirement, then the same problems occur here as were discussed for the somatic mutation theory. A very large number of genes must be produced to insure that one of each type ultimately required is present. Such a large number of genes would produce difficult storage and access problems. To save the simple germ line theory two postulates must be made. The first is some mechanism for deleting duplicates and preserving one copy of each gene. The second is some mechanism for insuring that a complete set is maintained and that it

spreads without loss throughout an entire population. Only if this occurs can we be assured of being able to develop an immune response against a new mutant virus. Unfortunately the requirements of the appearance, retention, and uniform distribution of genes having no previous function, but which may be crucial to survival in the future, are contrary to present evolutionary theory.

(xiii) *All basic information required for all immune responses resides naturally in the genome as a result of maximal randomization of all amino acid sequences during evolution on the one hand, and the restrictions inherent in the theory of perfect sets on the other*

In many respects this is a restatement of Ehrlich's (1900) original theory of immunity which proposed that antibody reactive sites were originally evolved to attach and assimilate foodstuffs, that is, the sites were originally evolved to serve a function not associated with an immune response. This circumvents the problem of how a primitive immune system, capable of a very restricted range of responses could have appeared, been retained, and gradually achieved the sophistication now evident.

(xiv) *The complexity of higher animals is such that their genomes may contain information sufficient for more than one perfect set, and the set of interactive surfaces in one differentiated cell may contain elements which interact with surfaces in one or more other differentiated cells of the same organism*

The amount of information required during all of the development of a higher vertebrate may be such that one perfect set is insufficient. Different differentiated cells in one animal may thus gradually come to express perfect sets which contain members which could interact in a deleterious fashion. In addition during development phase-specific antigens (Anderson & Coggin, 1974) may appear at different times which could not co-exist temporally. This could have two results. The first is the production of an effective barrier between differentiated states such that hybrids expressing both states completely are not viable. The second result is the capacity for autoimmunity against both tissue-specific and embryonic phase-specific antigens.

(xv) *To be effective an immune system need respond only to antigens which bind to sites normally in or on cells or tissues*

The so-called exobiology risk problem arises from the possibility that infectious agents could exist elsewhere in the universe against which the human immune system could produce no defense. (An extensive literature, not cited here, is concerned with this possibility.) This is another way of raising the question of completeness, of "all possible", which is the single



most difficult problem in immunological theory. Viruses require receptors or attachment sites on the cells they invade while pathogenic bacteria may produce toxins which react specifically with complementary sites in or on cells. The rule appears to be: no interaction, no damage. Either we must postulate what might be called the pathogen truce—an agreement that no pathogen will evolve a site against which we have no capacity to produce antibodies—or the cell receptor site is somehow linked to the immune response so that they appear and disappear together, and so that the information required to produce one is used to produce the other. This abolishes the absolute requirement for completeness. However the question arises at once, how can the information (i.e. a section of a gene) containing the requisite information be located and used by the immune system?

(xvi) *Binding of a cytoplasmic protein by an externally derived macromolecule results in a change in the rate of synthesis of that protein*

One would hesitate to postulate that a general mechanism exists which allows the rate of synthesis of each protein to be controlled by the fate of that protein, especially if the control is exerted at the transcriptional level. However such has been frequently shown to occur. Antibodies against a cellular protein may stop the synthesis of that protein (Sonneborn, 1957) suggesting that a mechanism for locating the gene for a protein paratactically bound exists. This mechanism would fulfil the first and most difficult requirement for using existing genomic information to mount a specific immunologic response, namely finding it. How the information, once located, is used is an unsolved problem. However if one is allowed free ranging speculation, one might postulate the production of small segments of RNA corresponding to portions of the located gene which, through reverse transcription, could give rise to segments of DNA subsequently attached to DNA coding for the constant regions of immunoglobulins. Such small RNA segments might serve as transfer factors transferring small specific bits of genetic information from one cell to another. These concepts are of considerable interest because transfer factors have not previously found a logical position in immunological theory, in addition to presenting difficult problems to experimentalists.

(xvii) *A major function of perfect sets is to assist in the destruction of denatured or mismade proteins*

Denatured or mismade proteins are rapidly and selectively destroyed (Goldschmidt, 1970; Platt, Miller & Weber, 1970). One selective advantage of evolving a perfect set of molecular surfaces would be that almost any deviant from the set could be identified. In the simplest instance a mechanism

for the destruction of dimers or larger aggregates could serve to remove proteins which, because of partial denaturation, have become attached to sites on proteins normally unattached within the cell.

(xviii) *Classical immunochemical methods are of little assistance in exploring the concept of perfect sets*

Immunochemistry is concerned with specific interactions, for the most part between large molecules. It might be thought therefore that immunochemical methods could be applied directly to exploring the concept of perfect sets. This is not the case because most of these methods depend on either the formation of insoluble precipitates or on the consumption of complement. Both are due to refinements evolved into antibody molecules. There is no reason to expect either precipitation or complement fixation to occur in the paratactic reactions described here. Protein inhibitors of enzymes for example rarely precipitate them. To explore the concepts presented here quite different methods will be required. Some of these are based on affinity chromatography and are described in detail elsewhere (Anderson *et al.*, 1975*a,b*).

#### 4. Discussion

The objective of this paper is to present in outline form a simple but general theory of the evolution of interactive macromolecular surfaces which simultaneously solves the problem of the origin of the immune system, the problem of the fixation of apparently neutral mutations in large numbers, the problem of the very slow and constant rate of amino acid substitution during evolution, and the problem of protein turnover. The data and arguments supporting various parts of the theory are drawn from several disciplines, and if included in this initial statement would make it difficult to follow and prohibitively long. Hence the strategy of an initial statement in the form of eighteen propositions, with succeeding papers devoted to setting out details.

At the outset we have assumed a universe of interactive surfaces of finite size which can be completely grouped into interactive pairs. This is a simplification adopted to allow the central concepts to be more easily grasped. As will be discussed subsequently, interactive families of surfaces correspond better with the real situation.

Several objections have been raised to the concept of perfect sets as presented here. The first stems largely from annoyance that a third alternative should be suggested when a large fraction of immunologists have taken

one side or the other in the germ line versus the somatic mutation theory argument. Three-way arguments have little appeal.

The second objection concerns the large number of surfaces postulated to exist in eukaryotic cells. This objection may be put in perspective as follows. If the total number of interactive pairs possible is  $N$ , then the total number of surfaces is  $2N$ . The somatic mutation theory assumes that  $2N$  surfaces are generated anew, with a high degree of certainty in each higher vertebrate in an interval of time ranging from days to weeks, and that cells making antibodies against self surfaces are destroyed. The conventional germ line theory allows a much longer time for the evolution of  $N$  surfaces, actually an interval measured in millions of years. In contrast the theory of perfect sets presented here allows all of bioevolutionary time to complete the task.

A third objection concerns cell fusion. How can cells fused from different species survive? In practice only a fraction of them do, and the genes expressed appear to be only a tiny fraction of all those present. Few characteristics of differentiated cells are present, and the chromosomes from one species of the two are generally lost during succeeding generations. Bare survival of cells in tissue culture is quite different from active participation in differentiation in an organism where the whole program of differentiation is played out, and where many requirements for special functions exist.

The most interesting aspects of a new theory however are the predictions which it makes, and the experiments which it suggests. These and the techniques for their experimental examination are described in subsequent papers. A very brief exposition of the theory of perfect sets has appeared previously (Anderson, 1969).

#### REFERENCES

- ANDERSON, N. G. (1968). In *Handbook of Biochemistry*, pp. K-28. Cleveland: Chemical Rubber Co.
- ANDERSON, N. G. (1969). *Exp. Hematol.* **19**, 45.
- ANDERSON, N. G. & COGGIN, J. H. (1974). *Cancer Res.* **34**, 2032.
- ANDERSON, N. G., WILLIS, D. D., HOLLADAY, D. W., CATON, J. E., HOLLEMAN, J. W., EVELEIGH, J. W., ATTRILL, J. E., BALL, F. L. & ANDERSON, N. L. (1975a). *Anal. Biochem.* **66**, 159.
- ANDERSON, N. G., WILLIS, D. D., HOLLADAY, D. W., CATON, J. E., HOLLEMAN, J. W., EVELEIGH, J. W., ATTRILL, J. E., BALL, F. L. & ANDERSON, N. L. (1975b). *Anal. Biochem.* **68**, 371.
- ARNHEIM, N. & TAYLOR, C. E. (1969). *Nature, Lond.* **223**, 900.
- BURNET, F. M. (1968). *Nature, Lond.* **218**, 426.
- BURNET, F. M. & FENNER, F. (1949). *The Production of Antibodies*, 2nd edn. Melbourne: Macmillan.
- EHRlich, P. (1900). *Proc. R. Soc., reprinted in Collected Papers of Paul Ehrlich*, Vol. II, p. 178. London: Pergamon Press.

- GOLDSCHMIDT, R. (1970). *Nature, Lond.* **228**, 1151.
- JERNE, N. K. (1960). *A. Rev. Microbiol.* **14**, 341.
- KIMURA, M. (1968). *Nature, Lond.* **217**, 624.
- KING, J. L. & JUKES, T. H. (1969). *Science, N.Y.* **164**, 788.
- MARGOLIASH, E., BARLOW, G. H. & BYERS, V. (1970). *Nature, Lond.* **228**, 723.
- MEDAWAR, P. B. (1967). *The Art of the Soluble*, p. 8. London: Methuen and Co., Ltd.
- PLATT, T., MILLER, J. H. & WEBER, K. (1970). *Nature, Lond.* **228**, 1154.
- RICHMOND, R. C. (1970). *Nature, Lond.* **225**, 1025.
- SONNEBORN, T. M. (1957). *The Species Problem*. Washington, D.C.: Am. Assoc. Adv. Sci