The generative grammar of the immune system

Niels K.Jerne M.D. FRS.

Institute for Immunology, 487 Grenzacherstrasse, 4058 Basel, Switzerland

Grammar is a science that is more than 2000 years old, whereas immunology has become a respectable part of biology only during the past hundred years. Though both sciences still face exasperating problems, this lecture attempts to establish an analogy between linguistics and immunology, between the descriptions of language and of the immune system. Let me first recall some of the essential elements of the immune system, with which I shall be concerned. In 1890, von Behring and Kitasato (12) were the first to discover antibody molecules in the blood serum of immunized animals, and to demonstrate that these antibodies could neutralize diphtheria toxin and tetanus toxin. They also demonstrated the specificity of antibodies: tetanus antitoxin cannot neutralize diphtheria toxin, and vice versa. During the first 30 years, or more, after these discoveries, most immunologists believed that all cells of our body are capable of producing antibodies, and it took until the 1950s before it became clear, and until 1960 before it was demonstrated (13), that only white blood cells, named lymphocytes, can produce antibodies. The total number of lymphocytes represent little more than 1% of the body weight of an animal. Thus, it would not be wrong to say that our immune system is an organ consisting of about 10¹² lymphocytes:

(human) immune system = 10^{12} lymphocytes or in a mouse that is 3000 times smaller than we:

(mouse) immune system = 3×10^8 lymphocytes

This brief description of the immune system disregards the fact that lymphocytes interact with most other cells in the body, which in my definition do not belong to the immune system *sensu strictu*.

Let me draw attention to the fact that this number of lymphocytes in the immune system is at least one order of magnitude larger than the number of neurons in the nervous system. Also, we should note that lymphocytes travel among most other cells of our body, that they circulate in blood and lymph, and occur in large concentrations in spleen, lymph nodes, appendix, thymus and bone marrow. Strangely enough, however, they seem to be excluded from the brain. The 1960s was a very fruitful decade of immunological discoveries, of which I shall name a few: in the beginning of the decade, the primary structure of antibody molecules was clarified (14,15); then followed the demonstration that the dictum of Burnet (2,16) was correct, namely that all antibody molecules synthesized by one given lymphocyte are identical, and finally, towards the end of that decade, lymphocytes were shown to fall into two classes, called T cells and B cells, existing in the body in almost equal numbers (17,18,5). Only B lymphocytes, or B cells, however, can produce and secrete antibody molecules.

Schematically, I could picture this as in Figure 1.

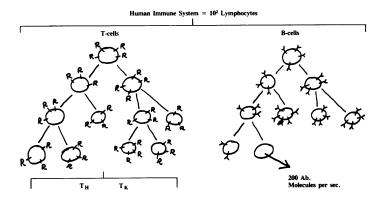


Figure 1.

What I should like you to retain from this picture is both what we know as well as what we do not know at present. Our knowledge is increasing at an accelerating pace: what was unknown yesterday may be known tomorrow, and what is correct today may turn out to be wrong tomorrow. Thus, B lymphocytes are known to carry so-called receptor molecules on their surface (about 105 identical receptors per B cell), and when such a 'resting' B cell is properly stimulated to divide and to mature, it will end up excreting about 2000 antibody molecules per second, all of which are identical, and similar or identical to the receptors that the resting B cell originally displayed. This clonal nature of antibody formation was clearly demonstrated in the early 1970s (19,20). The normal antibody response of an animal to a foreign antigen involves a large number of different clones, however, and is characterized by the production, usually of several hundreds, of different antibody molecules (21). T lymphocytes are also known to carry receptor molecules on their surface, but firstly these molecules are not well known yet because they have been discovered only during the past two years and secondly, the T cells do not excrete such molecules: these T cell receptors are antigen-recognizing molecules, but they do not contribute to the population of freely circulating antibody molecules which are purely B cell products. Furthermore, there are at least two different types of T cells, one of which is called T helper cells (T_H) (22) because they help B cells to become stimulated (or, in their absence, they prevent B cells from receiving a proper stimulus); the other type is called T killer cells (T_K) (23) because they are capable of killing other cells which they consider undesirable (such as virus-infected cells); moreover, as suppressor cells, they may prevent B cells from being stimulated (6).

Thus, in this picture, B cells have the single-minded desire to express their antibody language, but are subordinate to the T cells, that can either enhance or suppress this capacity.

Before looking at grammar, I must briefly describe the struc-

©The Nobel Foundation 1985

ture of an antibody molecule. No matter what you try to investigate in biology turns out to become increasingly complex — thus also the structure of antibody molecules. The basic element of all antibody molecules has been shown to be a Y-shaped protein structure (26) of about 150 000 daltons molecular weight. This is a three-dimensional structure, like all molecules and cells that biology has to deal with. Three-dimensionality tends to perplex our mind, which is most at ease with one-dimensional, linear sequences, but I shall try to make a rough two-dimensional sketch of an antibody in Figure 2.

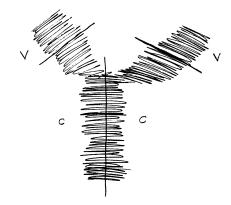


Figure 2.

We can make some important cuts through this molecule. The vertical cut shows symmetry. It divides the antibody molecule in two identical halves. The other other cuts separate a so-called 'constant' part (c) from a 'variable' part (v). By constant part, we mean that molecules of different antibody specificity have this part in common (such as, for example, diphtheria antitoxin and tetanus antitoxin). By variable part, we mean that this is the region of the antibody molecule that determines its 'specificity'. The two variable parts are identical: that is to say, that the molecule, with regard to specificity, is divalent. The difficulty we face is not to transform this sketchy two-dimensional picture into a one-dimensional primary structure, but into a threedimensional tertiary structure. The primary structure (14,15) has been clarified: the half molecule is made up of a light polypeptide chain of about 214 amino acid residues, and a heavy polypeptide chain of a little more than 400 amino acid residues, as in Figure 3.

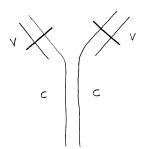


Figure 3.

It turned out that antibody molecules of different specificity have identical amino acid sequences in their carboxyl terminal regions, but they vary with respect to the amino acid sequences in the amino-terminal regions of both heavy and light chains (24). It became obvious immediately that the great diversity of antibody

molecules, the great number of different molecules which antibodies can recognize, or in other words, the great repertoire of antibody specificities, must result from an enormous number of varieties in the variable regions with respect to amino acid sequences. This insight does not solve our problems, however. It is like saying that the great variety of words or sentences in a language results from the enormous number of varieties with respect to the sequences of letters or of phonemes.

Interpretation in immunology remained practically as it had traditionally been, namely that the variable region of an antibody molecule forms a three-dimensional 'combining site', and that 'specificity' simply means that this combining site is complementary in shape to the shape of part of the three-dimensional profile of an antigen molecule. Antigen is the word that was given, and is still in use, for molecules that can induce the immune system to produce specific antibodies which can recognize these antigens. The antibody combining site was conceived as a cleft which recognizes a protruberance on the outer shape of an antigenic molecule, and all antibodies were named after the antigens that they recognized, such as diphtheria antitoxin, anti-sheep red blood cell antibodies, anti-TNP, etc. (25). I shall try now to give you an impression of the size of this system of antigens and specific antibody molecules. Let us first consider macromolecules of molecular weights exceeding 10 000 daltons; they may be polysaccharides, proteins, lipoproteins, nucleic acids, viruses, bacteria – in fact any such molecule or particle existing in the world is an antigen to which the immune system can make specific antibodies. Moreover, molecules such as nitrophenol, or arsonate, or any inorganic molecules you care to mention are antigenic when attached to a so-called carrier molecule, for example to a protein: the immune system will then produce antibodies that specifically recognize these molecules, even if they have been synthesized in a chemical laboratory without ever before having existed in the world (1). How is this possible? For example, the immune system of a mouse possesses no more than about 108 B lymphocytes, which would be the maximal available repertoire of variable regions on its antibody molecules. We realize that 'recognition' need not be perfect, and that the same 'combining site' might recognize, with more or less precision, a number of similar antigens.

I shall now turn to some remarkable discoveries, made during the past 25 years, showing that the variable regions of antibody molecules are themselves antigenic and invoke the production of anti-antibodies. Kunkel (27) showed that monoclonal myeloma antibodies, when injected into another animal, induce specific antibodies which recognize the particular myeloma antibodies used, but which do not recognize any other myeloma antibodies isolated from other myeloma patients. This work was extended by others, but mainly by Jacques Oudin and his colleagues in Paris, who showed that ordinary antibody molecules that arise in an immunized animal are antigenic and invoke the formation of specific anti-antibodies (28,29,30,31). In other words, the variable region of an antibody molecule constitutes not only its 'combining site', but also presents an antigenic profile (named its idiotype) against which anti-idiotypic antibodies can be induced in other animals. Moreover, it turned out that this antigenic, idiotopic profile of the variable region of a given antibody molecule is not a single site, but consists of several distinct sites against which a variety of different anti-idiotypic antibody molecules can be made. These individual sites are now named idiotopes, implying that the the idiotype of one antibody molecule can be described as a set of different immunogenic idiotopes. And finally, it has been shown that the immune system of a single animal, after producing specific antibodies to an antigen, continues to produce antibodies to the idiotopes of the antibodies which it has itself made. The latter anti-idiotopic antibodies likewise display new idiotypic profiles, and the immune system turns out to represent a network of idiotypic interactions (7,8,9,10,11).

I shall now show, in Figure 4, a preliminary picture of how we might vaguely try to imagine the shape of the variable region of an antibody molecule.

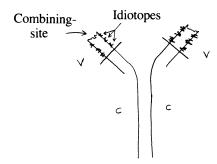


Figure 4.

This picture is a historical compromise: we uphold our antigencentered tradition (25) by retaining the notion of a 'combining site' which enables the antibody molecule to recognize the antigenic molecule that induced its production, and we simply add a number of idiotopes on the same variable region, which are capable of inducing the production of other antibody molecules that have 'combining sites' recognizing these idiotopes.

We are now getting into trouble, however, when trying to interpret our experimental results. As I said earlier, a resting B lymphocyte displays on its surface about 100 000 identical receptor molecules which are representations of the type of antibody molecules this B cell will produce if stimulated by an antigen.

Let us return to this picture for a moment, as in Figure 5,

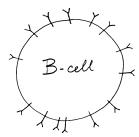


Figure 5.

and let us make an enlargement, in Figure 6, of a small part of the surface of this B cell, focusing on one receptor molecule only,

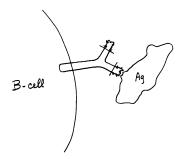


Figure 6.

making the usual cuts between constant and variable regions: As you see, I have retained the traditional distinctions between the antigen-recognizing combining site, and a number of antigenic idiotopes. I have also added an imaginary profile of an antigenic molecule, part of which is recognized by the combining site of the B cell receptor. This is the basic picture of the selective theories of antibody formation, as most clearly formulated by Burnet (16,2). The antigen 'selects' the lymphocytes by which it is recognized, and stimulates these cells to proliferate, to mature, and to secrete antibodies with fitting combining sites. Clearly, T cell control is also involved, as well as growth factors, maturation factors, etc., but this picture remains the basic idea of antibody induction. Figure 7 shows one of these

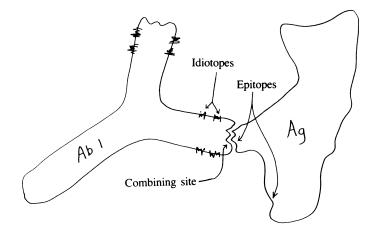


Figure 7.

antibody molecules, which recognizes part of the surface profile ('epitope') of the antigen.

Now imagine, however, that this antibody molecule (Ab1), displaying both its combining site as well as its antigenic idiotopes, is itself used as antigen. It is then possible to envisage two situations, shown in Figure 8.

Figure 8 shows, schematically, a freely circulating Ab1 molecule which recognizes, and sticks to, a receptor of a B cell.

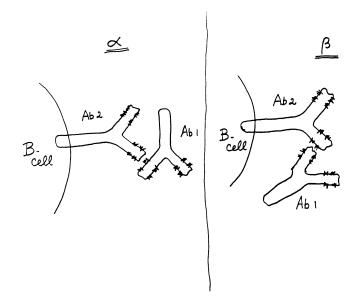


Figure 8.

The figure envisages the stimulation of two different B cells by Ab1 molecules now acting as antigens. In the α case, the combining site of the receptor on a B cell recognizes an idiotope of Ab1 and the cell may be stimulated to produce the corresponding anti-idiotopic antibodies (Ab2). In the β case, however, it is the combining site of the Ab1 molecule which recognizes an idiotope of the receptor on a B cell which may thus be stimulated to produce antibodies which possess idiotopes that have a shape that is similar to the epitope displayed by the original antigen. Experiments have shown both these situations actually to occur. For example, if the orignal antigen is insulin, and Ab1 is an antiinsulin antibody, then some of the anti-idiotypic antibodies (Ab2) of the β types show a similarity to insulin and can even be shown to function like insulin (32). Similar results in other systems have been obtained by Cazenave and Roland, by Strosberg, by Urbain, and their colleagues, and by others (33,10,34,35).

The point I wish to make, however, is to consider whether the two situations α and β are fundamentally different, or not. Is there a difference between saying that Ab1 recognizes Ab2 or that Ab2 recognizes Ab1? Can we, at this three-dimensional molecular level, distinguish between 'recognizing' and 'being recognized'? If not, it becomes meaningless to distinguish between idiotopes and combining sites, and we could merely say that the variable region of an antibody molecule displays several equivalent combining sites, a set of idiotopes, and that every antibody molecule is multi-specific. I do not have to belabour this point which has been made repeatedly (36,37,38,25). Instead, I should now like to introduce some numerology into this discussion. How large is the number of different antibodies that the immune system of one single animal (be it a human or a mouse) can make? This number, during the past few decades, has been estimated, on more or less slender evidence, to exceed ten million, and this enormous diversity has been designated as the 'repertoire' of the B lymphocytes. Such a 'repertoire' has been characterized by Coutinho as 'complete' (39). 'Completeness' means that the immune system can respond, by the formation of specific antibodies, to any molecule existing in the world, including, as I said earlier, to molecules that the system has never before encountered. Immunologists sometimes use words they have borrowed from linguistics, such as 'immune response'. Looking at languages, we find that all of them do make do with a vocabulary of roughly a hundred thousand words, or less. These vocabulary sizes are a hundred-fold smaller than the estimates of the size of the antibody repertoire available to our immune system. But if we consider that the variable region that characterizes an antibody molecule is made up of two polypeptides, each about 100 amino acid residues long, and that its threedimensional structure displays a set of several combining sites, we may find a more reasonable analogy between language and the immune system, namely by regarding the variable region of a given antibody molecule not as a word but as a sentence or a phrase. The immense repertoire of the immune system then becomes not a vocabulary of words, but a lexicon of sentences which is capable of responding to any sentence expressed by the multitude of antigens which the immune system may encounter.

At this point, I shall make a few quotations from Noam Chomsky (3) concerning linguistics: "The central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them . . . Grammar is a device that specifies the infinite set of well-formed sentences and assigns to each of these one or more structural descriptions.

Perhaps we should call such a device a generative grammar . . . which should, ideally, contain a central syntactic component . . ., a phonological component and a semantic component." That is the end of my quotation. For the size of the set of possible sentences in a language, Chomsky uses the word 'openendedness', and I now think that 'open-ended' is the best description also of the 'completeness' of the antibody repertoire. As for the components of a generative grammar that Chomsky mentions, we could with some imagination equate these with various features of protein structures. Every amino acid sequence is a polypeptide chain, but not every sequence will produce a wellfolded stable protein molecule with acceptable shape, hydrophobicity, electrostatics, etc. Some grammatical rules would seem to be required. It is harder, however, to find an analogy to semantics: does the immune system distinguish between meaningful and meaningless antigens? It would seem, at first sight, that the immune response to a sentence presented by an invading protein molecule is merely to select, from amongst its enormous preformed antibody repertoire, a suitable mirror image of part of this antigenic sentence. As you will know, Leonardo da Vinci wrote his private journal in the mirror image of ordinary handwriting. It is difficult, without considerable practice, to write and read mirror handscript. Let me show you an example on the next figure (Figure 9).



Figure 9.

On the following two figures (Figures 10 and 11), I shall use the device of showing ordinary letters in black, and of using zones marked in grey to indicate the mirror images of these letters.

I should perhaps again emphasize that the sentences representing antibodies possess partial mirror images of an antigenic sentence. These antibodies are not echoes of the invading antigen, but were already available to the animal in its repertoire of B cells before the antigen arrived. This is the important insight that followed the introduction into immunology of the selective theories in the 1950s. Also, I must emphasize another important quantitative aspect of the situation facing the immune system. It has been estimated that one human individual produces about 10 000 different proteins, such as enzymes, hormones, cell surface proteins, etc. At the same time, we estimate that the immune system maintains a repertoire exceeding 10 million different proteins, namely antibody molecules. This is a thousandfold more than all other body proteins taken together. Man and mice, normally, have about ten milligrams of antibodies in a milliliter of their blood. Thus, a normal human possesses between about 50 to 100 grams of freely circulating antibodies, called immunoglobulins. If we divide this figure by 107 different specificities, we are still left with an average of 5 to 10 micrograms of every specificity in the available repertoire, representing an average of about 3 x 10¹³ monoclonal antibodies of every specificity. For mice that are 3000 times smaller, we would have to divide these figures by 3000, which would still leave a mouse

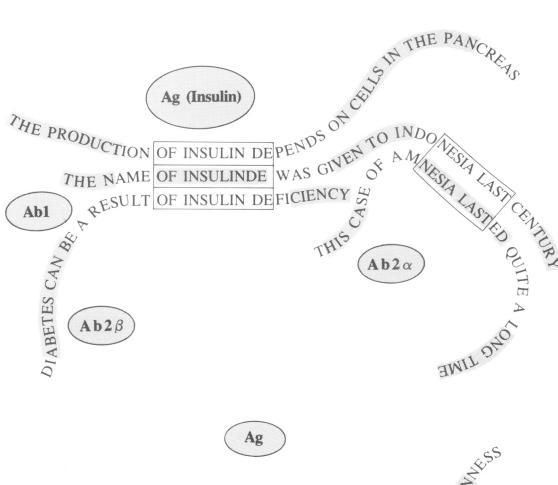


Figure 10.

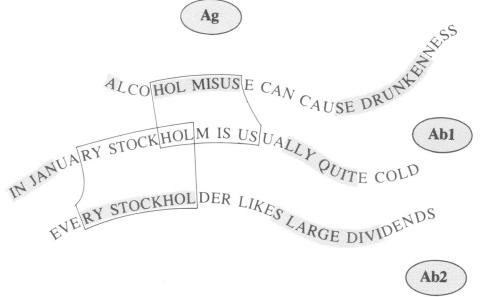


Figure 11.

with an average of 2 to 3 nanograms of antibodies of every of 10⁷ specificities. That even such nanogram quantities of monoclonal anti-idiotypic antibodies, when introduced into mice, produce remarkable effects has been shown by Rajewsky and his colleagues (40,41,42).

I should therefore like to conclude that in its dynamic state our immune system is mainly self-centered, generating antiidiotypic antibodies to its own antibodies, which constitute the
overwhelming majority of antigens present in the body. The
system also somehow maintains a precarious equilibrium with
the other normal self-constituents of our body, while reacting
vigorously to invasions into our body of foreign particles, proteins, viruses, or bacteria, which incidentally disturb the dynamic

harmony of the system.

The inheritable 'deep' structure of the immune system is now known: certain chromosomes of all vertebrate animals contain DNA segments which encode the variable regions of antibody polypeptides. Furthermore, experiments in recent years have demonstrated the generative capacities of this innate system. In proliferating B lymphocytes, these DNA segments are the targets for somatic mutations, which result in the formation of antibody variable regions which differ, in amino acid sequences, from those encoded by the stem cell from which these B cells have arisen (43,44,45,46,47). The experiments showed that it was still possible, however, to identify the original stem cell genes that must have undergone these mutations. Expressed in linguistic

terms, such investigations belong to the etymology of the immune

As immunologists, we should like to know the semantics of the inheritable gene structures. What is the meaning of the basic lexicon, or what are the specificities of the antibodies, B cell receptors and T cell receptors as encoded in the genes of our germ cells? It is known that B cells recognize the language of the T cell receptors. I have said so little about the latter because T cell receptorology is still in too early a stage of development. An immune system of enormous complexity is present in all vertebrate animals. When we place a population of lymphocytes from such an animal in appropriate tissue culture fluid, and when we add an antigen, the lymphocytes will produce specific antibody molecules, in the absence of any nerve cells (48). I find it astonishing that the immune system embodies a degree of complexity which suggests some more or less superficial though striking analogies with human language, and that this cognitive system has evolved and functions without assistance of the brain.

It seems a miracle that young children easily learn the language of any environment into which they are born. The generative approach to grammar, pioneered by Chomsky (4), argues that this is only explicable if certain deep, universal features of this competence are innate characteristics of the human brain. Biologically speaking, this hypothesis of an inheritable capability to learn any language means that it must somehow be encoded in the DNA of our chromosomes. Should this hypothesis one day be verified, then linguistics would become a branch of biology, and the humanities, perhaps, some day, part of the natural sciences.

References

Books

- 1. Landsteiner, K. (1947) The Specificity of Serological Reactions, Howard University Press, Washington.
- 2. Burnet, F.M. (1959) The Clonal Theory of Acquired Immunity, Cambridge University Press.
- 3. Chomsky, N. (1964) Current Issues in Linguistic Theory, Janua Linguarum, Series minor, Mouton, The Hague.
- 4. Chomsky, N. (1972) Language and Mind, Harcourt Brace Jovanovich, New
- 5. Greaves, M.F., Owen, J.J.T. and Raff, M.C. (1974) T and B Lymphocytes, American Elsevier, New York.
- 6. Golub, E.S. (1977) The Cellular Basis of the Immune Response, Sinauer Associates, Massachusetts.
- 7. Janeway, C., Sercarz, E.E. and Wigzell, H., eds. (1981) Immunoglobulin Idiotypes, Academic Press, New York.
- 8. Westen-Schnurr, I., ed. (1982) Idiotypes: Antigens on the Inside, Editiones Roche, Basel.
- 9. Bona, C.A. and Köhler, H., eds. (1983) Immune Networks, Annals N.Y. Acad. Sci., Volume 418.
- 10. Köhler, H., Urbain, J. and Cazenave, P.-A., eds. (1984) Idiotype in Biology and Medicine, Academic Press, New York.
- 11. Greene, M.J. and Nisonoff, A., eds. (1984) The Biology of Idiotypes, Plenum Press, New York.

- 12. Behring, E. and Kitasato, S. (1890) Dtsch. Med. Wochenschr., 16, 1113.
- 13. Gowans, J.L. and McGregor, D.D. (1965) Prog. Allergy, 9, 1.
- 14. Porter, R.R. (1959) Biochem. J., 73, 119.
- 15. Edelman, G.M. (1959) J. Am. Chem. Soc., 81, 3155.
- 16. Burnet, F.M. (1957) Aust. J. Sci., 20, 67.
- 17. Miller, J.F.A.P. and Mitchell, G.F. (1968) Proc. Natl. Acad. Sci. USA, 59, 296.
- 18. Raff, M.C. (1970) Immunology, 19, 637.
- 19. Bosma, M. and Weiler, E. (1970) J. Immunol., 104, 203.
- 20. Lefkovits, I. (1972) Eur. J. Immunol., 2, 360.
- 21. Fazekas de St. Groth, S., Underwood, P.A. and Scheidegger, D. (1980) in Peeeters, H. (ed.), Protides of the Biological Fluids, Pergamon Press, Oxford,
- 22. Claman, H.N., Chaperon, E.A. and Triplett, R.F. (1966) Proc. Soc. Exp. Biol. Med., 122, 1167.

- 23. Govaerts, A. (1960) J. Immunol., 85, 516.
- 24. Hilschmann, N. and Craig, L.C. (1965) Proc. Natl. Acad. Sci. USA, 53, 1403.
- 25. Coutinho, A., Forni, L., Holmberg, D., Ivars, F. and Vaz, N. (1984) Immunol. Rev., 79, 151.
- 26. Valentine, R.C. and Green, N.M. (1967) J. Mol. Biol., 27, 615.
- 27. Slater, R.J., Ward, S.M. and Kunkel, H.G. (1955) J. Exp. Med., 101, 85.
- 28. Oudin, J. and Michel, M. (1963) C.R. Hebd. Seances Acad. Sci. Paris, 257,
- 29. Oudin, J. and Michel, M. (1969) J. Exp. Med., 130, 595, 619
- 30. Kunkel, H.G., Mannik, M. and Williams, R.C. (1963) Science (Wash.), 140,
- 31. Gell, P.G.H. and Kelus, A.S. (1964) Nature, 201, 687.
- 32. Sege, K. and Peterson, P.A. (1978) Proc. Natl. Acad. Sci. USA, 75, 2443.
- 33. Jerne, N.K., Roland, J. and Cazenave, P.-A. (1982) EMBO J., 1, 243.
- 34. Strosberg, D.A., Couraud, P.-O. and Schreiber, A. (1981) Immunol. Today, 2. 75.
- 35. Gullet, J.G., Kaveri, S.V., Durieu, O., Delavier, C., Hoebeke, J. and Strosberg, A.D. (1984) Proc. Natl. Acad. Sci. USA, in press.
- 36. Richards, F.F. and Konigsberg, W.H. (1973) Immunochemistry, 10, 545.
- 37. Varga, J.M., Konigsberg, W.H. and Richards, F.F. (1973) Proc. Natl. Acad. Sci. USA, 70, 3269.
- 38. Jerne, N.K. (1984) Immunol. Rev., 79, 5.
- 39. Coutinho, A. (1980) Ann. Immunol. (Inst. Pasteur), 131D, 235.
- 40. Kelsoe, G., Reth, M. and Rajewsky, K. (1981) Eur. J. Immunol., 11, 418.
- 41. Rajewsky, K. and Takemori, T. (1983) Annu. Rev. Immunol., 1, 569.
- 42. Müller, C.E. and Rajewsky, K. (1984) J. Exp. Med., in press.
- 43. McKean, D.M., Hüppi, K., Bell, M., Staudt, L., Gerhard, W. and Weigert, M. (1984) Proc. Natl. Acad. Sci. USA, 81, 3180. 44. Rüdikoff,S., Pawlita,M., Pumphrey,J. and Heller,M. (1984) Proc. Natl. Acad.
- Sci. USA, 81, 2162.
- 45. Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) Cell, 24, 625.
- 46. Sablitzly, F., Wildner, G. and Rajewsky, K. (1984) EMBO J., in press.
- 47. Sims, J., Rabbitts, T.H., Estess, P., Slaughter, C., Tucker, P.W. and Capra, J.D. (1982) Science (Wash.), 216, 309.
- 48. Mishell, R.I. and Dutton, R.W. (1967) J. Exp. Med., 126, 423.