The question of how we acquire immunity has been investigated for a century or more. What have we learned from all of this endeavor? We asked Rolf Zinkernagel to provide, for the young investigator, food for thought about that which we still don't know—even if we think we do.

What is missing in immunology to understand immunity?

Rolf M. Zinkernagel

University of Zurich, Institute of Experimental Immunology, Department of Pathology, Schmelzbergstrasse 12, Zurich CH-8091, Switzerland. (rolf.zinkernagel@pty.usz.ch)

Immunology, as laid down in textbooks and reviews, reflects 150 years of intense research about protection against infectious disease and responses to chemically defined antigens. The data have been gathered through the use of various experimental animal species and ultimately embrace all of cell biology, down to the molecular level.

Many of the basic questions and ideas in immunology, however, were coined more than 100 years ago and have not changed much since; this includes concepts such as tolerance, the specificity of cells and antibodies, and immunological memory. Although these concepts are explained in textbooks, much remains unclear and the question arises: do we really know what is in our textbooks? And what may have been misleading us into thinking that we know what, in fact, we don't?

This commentary is an attempt to review ideas and seemingly established concepts in immunology on only a few pages. To do this is probably unrealistic, unfair to the subject and necessarily incomplete, but hopefully will help to point out some important open questions about immunity. Immune responses will be analyzed here from an evolutionary point of view, based on the following two assumptions: protection against infectious disease is the primary function of the immune system; and, like all biological processes, immune reactivity falls under a Gaussian-type (that is, bell-shaped) distribution and therefore can be selected for a survival advantage[1,2].

Compared to humans or mice, infectious agents have a very much shorter generation time and are more numerous by several orders of magnitude. The overall unfavorable numbers game between vertebrate hosts and infectious agents is somewhat compensated by the host's basic resistance mechanisms (such as interferons) and great numbers of rapidly expanding B and T lymphocytes. Nevertheless, for optimal cohabitation, adaptation to the host of infectious agents by mutation and selection must be relatively rapid to guarantee their survival, as it is also dependent on survival of the host species.

Specificity?

Specificity is defined by the interaction of an antibody with a three-dimensional folded antigen or interaction of a T cell with a complex of peptide and its presenter—a major histocompatibility complex (MHC) class I or II molecule. Protective binding avidities are about 1×10^{-8} M^{-1} and higher for antibodies, but are still somewhat unclear for T cells. The frequencies of B cells specific for a biologically relevant antigen, such as a neutralizing epitope on a virus or a bacterial toxin, and comprising about 8–15 amino acids, is roughly 1×10^{-4} –1×10^{-5} (refs. 3–5). T cell frequencies are about 1×10^{-5}–1×10^{-6} (refs 4–6). If one determines the frequency of B cells that are specific for a chemically defined hapten (for dinitrophenol or trinitrophenol, usually equivalent to the size of one amino acid), the B cell frequency is 1×10^{-2}–1×10^{-3} (binding with 1×10^{-6}–1×10^{-7} M^{-1}) (refs 3,5,7). Thus, B cells specific for haptens seem to be about 100 times more frequent than those specific for protective determinants or infectious agents. Therefore, estimates of repertoire size and specificity differ vastly, depending on the viewpoint—in infectious disease it is in the order of 1×10^{-4}, for a hapten it may range from 1×10^{-2}–1×10^{-13} (refs 3–5,8).

Specificity is also dependent on the strictness of epitope definition. Whereas a toxin or a neutralizing determinant on a virus has been defined by coevolution with respect to size, arrangement of the determinants (repetitiveness or not) and lack of crossreactivity with self-antigens, such limitations obviously do not apply to the common laboratory test antigens such as ovalbumin, lysozyme or haptens[3].

Although antibodies can bind to lysozyme or ovalbumin in numerous ways, protective antibodies binding to distinct serotypes of viruses or bacteria will bind to a single antigenic site[9,10]. For example, only neutralizing (protective) antibodies can bind to an intact rabies virus particle, simply because this is the only determinant exposed on the virus envelope; the glycoprotein is so densely packed that antibodies generated in the host that bind to other determinants (released from virally destroyed cells) simply cannot fit in between the glycoproteins on the virion[3,9,10]. Therefore, only neutralizing antibodies protect; other antibodies against internal antigens of the virus are usually wasted.

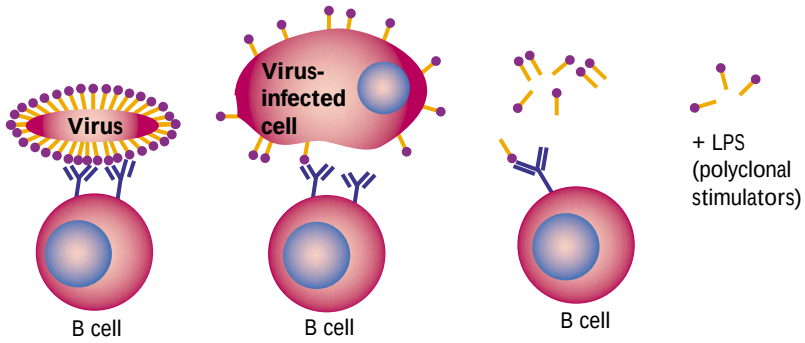
B cell "tolerance"?

Whether B cells that carry a specific receptor for self-antigens in the lymphohematopoietic system or in the periphery do exist or are deleted (negatively selected against), has been debated for a long time[5]. Of course, it is not easy to study B cell tolerance. Under conditions where the soluble self-antigen is at sufficient concentration, any antibody would be bound immediately and eliminated, rendering antibodies undetectable and the isolation of B cells with such specificity difficult.

Recent experiments with mice whose B cells express a transgenic antibody with specificity for either hen egg lysozyme or an alloantigen[11,12] seem to suggest that B cells are negatively selected upon exposure to cell membrane-bound determinants, but are not negatively selected against soluble forms of the same antigen. However these results contrast with the results of other experiments, which have suggested that autoantibody-producing B cells are not deleted in mice transgenic for a soluble or a membrane-bound viral glycoprotein[3,13]. These mice did not generate neutralizing antibodies when immunized with purified glycoprotein in adjuvant, but responded promptly with a neutralizing IgM response against intact virus (in which glycoprotein is in a highly repetitive form) that was independent of T cell help. Because virions also contain new helper T cell determinants (for instance, the viral nucleoprotein and matrix protein), the class switch to IgG was also efficient. This suggested that B cells are regulated by

antigen pattern and the availability of help from T cells—and not by B cell deletion[3–5].

Is there really a need for B cells to be tolerant? B cell induction depends on one of two conditions (**Fig. 1**). First, antigen either needs to be multimeric, highly repetitive and at best rigidly organized, so that antigens can optimally crosslink membrane immunoglobulin on B cells, or it must be recognized by B cells together with a polyclonal B cell activator such as bacteria or lipopolysaccharide (LPS)[14]—both trigger IgM production independent of T cell help[3,14–16]. Alternatively, B cells binding monomeric or oligomeric antigens cannot be induced by antigen binding alone, but need linked T cell help[5]; as T helper cells that recognize self-antigens are deleted by negative selection in the thymus and in the periphery, it is usually the case that no IgM or IgG responses are induced beyond the low titer of spontaneous (or natural) IgM antibody production. Also, because potentially autoreactive IgM antibodies are very short-lived ($t_{1/2}$ of 1–3 days), any autoantibody IgM response will be short-lived in the absence of T cell help. Therefore, disease-causing B cell IgM responses are not triggered against the usually monomeric or oligomeric self-determinants accessible to B cells independent of T cell help. In contrast to self-antigens, all infectious agents causing pathology have, as a hallmark, highly repetitive identical determinants expressed on the surface in a rigid or flexible form, offering exactly the right conditions to induce T cell–independent IgM responses. Such responses then amplify the rare specific B cells early, increase the targets for T help, and thus promote and enhance the switch to IgG[3]. Overall this results in coevolutionary equilibrium where both host species and infectious agents survive.



Cross-linking of membrane immunoglobulin	+ + +	+	——	——
Requirement for help	——	Bystander (nonlinked) help	Obligatory linked help	——
Examples	Neutralizing epitopes of viruses, surface antigen of bacteria, flagella and parasites	Viral glycoprotein on infected cell, transplantation antigens	Bacterial toxins, self antigens or viral antigens released from cells	Bacterial toxins, antigen plus polyclonal activators

Figure 1. Role of antigen pattern and T cell help in B cell responses.

Considering these various facets of B cell responsiveness, one may question whether B cells are really rendered tolerant, anergic or are deleted. Such mechanisms are probably not necessary and would perhaps render B cells very difficult to induce. The antigen pattern and LPS rule, and need for T cell help if the pattern or LPS rule is not fulfilled, would represent control and evolutionary safety enough against autoimmunity[3]. It is, however, interesting to note that autoantibodies against repetitive antigens not usually accessible to B cells, such as collagen or DNA, seem to be among the favorite autoantibodies encountered in a great variety of autoimmune diseases[5]—but also remember that autoimmune disease is relatively rare and usually associated with an age of 25 years or more (and female sex, see below), and therefore has probably not played a major role in evolution.

Positive selection of T cells?

What really happens in the thymus during T cell maturation is still somewhat unclear[17–19]. Is thymic epithelia important for rearrangement of the T cell receptor or is it important for T cell expansion and deletion (positive and negative selection)? Evidence for radio-resistant cells in the thymus, presumably thymic epithelial cells, to have a role in positive selection has been based largely on experiments with irradiated bone marrow chimeras or thymus chimeras[18, 19]. These experiments reveal largely that the MHC of the irradiated thymic graft determines the MHC-restriction of the maturing T cell repertoire. In contrast, nude mice reconstituted with irradiated or fetal thymic grafts exhibited MHC-restriction to the host only[20]. Obviously the nude reconstitution experiments do not match the data obtained with irradiated chimeras. One interpretation is that nude mice still have a thymic rudiment that can be rendered functional by the experimental thymic graft, but neither histological nor functional evidence really supports this hypothesis[20,21].

The other interpretation is that in nude mice reconstituted with an allogeneic thymus the majority of bone marrow–derived and other cells in the thymus and in the periphery are of host-MHC type and that the host-MHC–restricted functional T cell repertoire is therefore vastly expanded compared to the few thymocytes that are restricted and expanded by the donor thymic epithelial cells. Perhaps, then, so-called positive selection is usually confined to the thymus because that is where it all starts and where, usually, the MHC of the epithelial cells and other cells is the same. And, perhaps, the mechanisms of T cell expansion are largely driven by the MHC-presented antigens, available both in the thymus and the periphery[17,20,22].

T cell indifference *versus* induction *versus* deletion?

T cell responses can only be induced in organized lymphoid tissues; the same holds true for B cells. For example, mice lacking peripheral lymph nodes cannot mount an immune response after cutaneous or subcutaneous immunizations[23,24]. Nevertheless, current ideas involving the two-signal hypothesis and negative regulation of T cell responses via signal 1 (that is, antigen alone) have a different view[4,6,8,25]. According to the two-signal theory, T cell encounters with cells in the periphery that exhibit self-peptides or foreign peptides are not neutral, but negative (suppressive) events. Encountering sig-

nal 1 (antigen alone) in the absence of a signal 2 (CD40-CD40 ligand interaction, B7-1–CD80 or B7-2–CD80 interaction, interleukin 2, for example) results in a negative signal to the T cell, which renders it anergic, tolerant or deletes it[4,6,8,25]. It is well accepted, and not questioned here, that second signals play an important enhancing role in organized lymphoid organs. Although many experiments have been interpreted as compatible with the predictions of the signal 1 + signal 2 hypothesis (sometimes also called the "danger hypothesis")[6] other experiments have falsified them as follows.

A fibroblast expressing a defined foreign antigen, but not an acceptable signal 2, can induce cytotoxic T lymphocytes (CTLs) directly without crossprocessing if, and only if, such a fibroblast reaches organized lymphoid tissues[26–29]. In fact, the dose-response curve of CTL-induction in lymphatic organs by fibroblasts is comparable to that of the so-called professional antigen presenting cells, such as dendritic cells or macrophages. Induction of CTLs, therefore, depends on antigen reaching organized lymphoid tissues where (the admittedly enhancing) signal 2 is amply present. Thus antigens, either self or foreign, expressed exclusively on cells in solid tissues outside of lymphoid tissues, are immunologically ignored by both T and B cells[28,29] (**Fig. 2**).

Such immunologically ignored self-antigens are potential targets for T cell induction and immunopathology if these cells or the antigens on antigen-expressing cells reach draining lymph nodes in sufficient concentrations and for long enough periods of time. Again, this happens rarely and inefficiently and usually not in young hosts; induction of autoimmunity may be enhanced by some infections and unfavorable host-resistance conditions causing release of excessive amounts of ignored self-antigen. Such a process may be enhanced if lymphoid tissue is newly generated in a peripheral target organ, that is, when the secondary lymphoid organ is "brought" to the ignored antigen[30]. This causes a chronic autoimmune response, because the local antigen maintains it[30]. Examples of such processes are Hashimoto's thyroiditis or rheumatoid arthritis where lymphatic follicles are regularly found in the thyroid or synovial tissues, respectively.

Whether T cells are induced to become effector T cells, or are exhaustively induced and deleted, is a question of the balance between the limited number of precursor T cells and the kinetics of their induction, and the amount and distribution of antigen (**Fig. 2**). If only a few precursors are induced in an initially local infection, for example in a draining lymph node, T cells are induced and amplified very quickly, with an 8-h cycle[1,2,31,32]. If, however, the antigens or virus systemically spread very quickly and induce all available precursor T cells in all lymphoid organs within a few days, they may all mature and die off within a few days, so that no specific precursors are left[33,34]. Of course, such an event may only occur if the antigen or the infectious agent is noncytopathic. Negative selection or T cell deletion in the thymus, therefore, perhaps may not represent a special mechanism but could be viewed as overstimulation or exhaustive induction of all available precursor T cells. Obviously this process is most easily achieved against high avidity self-reactive T cells at a time when T cell receptors are assembled and when only few precursors are available—that is, in the thymus. Perhaps, therefore, negative selection or deletion of T cells in the thymus is nothing special and is quite comparable to that going on in the periphery[20].

Memory?

Like specificity and tolerance, memory is one of the key hallmarks of the immune system. What memory really means, however, is still being debated[1,3,4,28,31,32]. The original concept of memory was that it constituted a special altered status of the immune system that enabled it to deal, in a qualitatively and quantitatively distinct way, with re-encountered antigen, particularly infectious agents. Of course the complex interactions between helper T cells, CTLs, interleukins and antigen presenting cells, for example, make it difficult to clearly define this altered status. The alternative interpretation of immunological memory is as the permanent restimulation of a low-level response (**Table 1**). Depots of antigen, infections within the host, or periodic re-exposure to antigens from the outside, would activate T cells to become protective effectors, and B cells to mature to antibody-producing plasma cells. After all, why should the host need immunological memory? If the host survives the first infection, the host's immune system has proven itself fit to deal with repeat infections; if the host is killed by the first infection, there is definitely no need for immunological memory. Under what circumstances, then, is immunological memory of survival value?

Let us first ask: which types of vaccines are efficiently protective today and which fail? But also: which immune

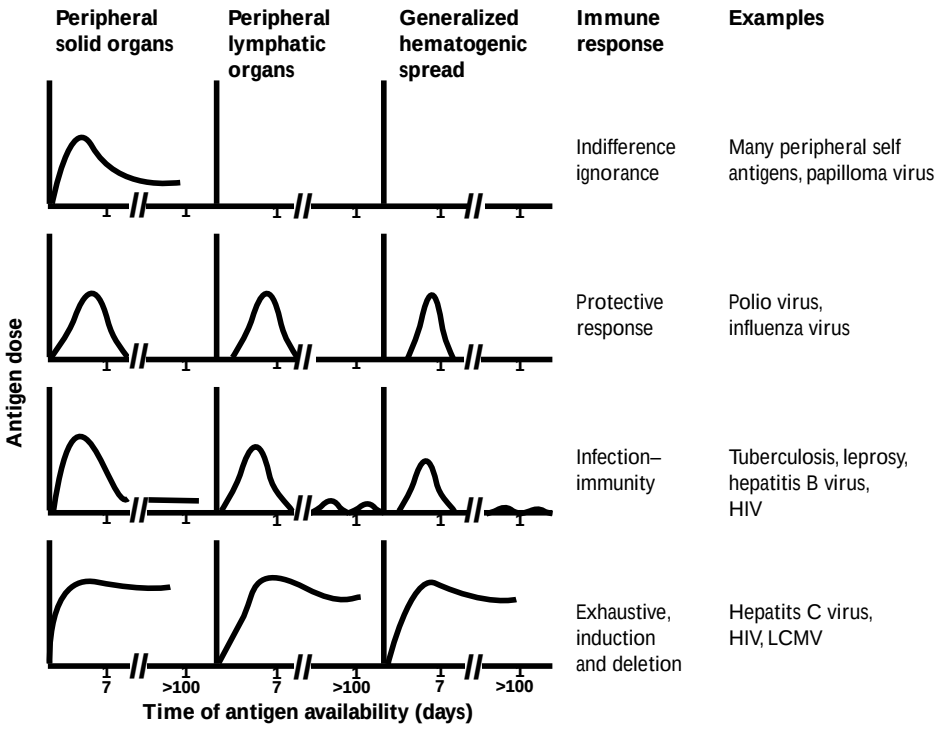


Figure 2. Correlation of T cell responses with antigen localization, dose and timescale of antigen availability. High or low precursor frequencies of T cells also influence reponsiveness by favoring protective response or exhaustive induction, respectively.

Table 1. Immunological memory

	B cells		T cells	
	Resting	Activated (antibody-producing)	Resting	Activated (with effector function)
Location	Blood Lymph node Spleen	Germinal center Bone marrow Lymph node Spleen	Blood Lymph node Spleen	Periphery Blood Lymph node Spleen
Function	Increase frequency	Antibody production	Increase frequency	Emigration into solid organs
Protection	Slow	Immediate	Slow	Immediate
Antigen-dependence	–	+	–	+

effector mechanisms mediate protection? Successful vaccines—against polio, measles, tetanus, diphtheria, pertussis and hepatitis B virus (HBV)—all protect via neutralizing antibodies. Vaccines that do not work well are BCG and the antiparasite vaccines. These vaccines seem to fail because they cannot induce protective effector T cells for extended periods—the antigenic stimulus apparently disappears[35]. The interesting point here is that a tremendous selective evolutionary pressure is exerted by acute cytopathic infections that kill children before they reach sexual maturity, quite in contrast to the much lesser evolutionary pressure exerted by non- or poorly cytopathic chronic types of infection, such as tuberculosis, leprosy, HBV and many parasitic infections[1,2,31,32,35].

When are children unprotected and therefore need immunological protection for survival? At birth, humans and mice are immunoincompetent[35–37]. This is probably because of MHC-restricted T cell recognition requiring MHC polymorphism within the species. It causes histo-incompatibility between mother and offspring, with a danger of mutual rejection. Thus the developing embryo remains immunologically immature until birth and the mother is immunosuppressed during pregnancy. To protect against the many lethal infectious diseases the newborn will be exposed to after birth, the immunological experience of the mother is transferred to the offspring. During pregnancy this immunity is conferred in the form of antibody transferred via the placenta, and after birth it is transferred via milk. But the possibility that all these protective antibodies to infectious diseases are accumulated by the mother and transferred to the developing offspring during pregnancy or right before birth is difficult, if not impossible, to imagine. The necessary infection to generate these antibodies would endanger the survival of the immunosuppressed mother and the fetus.

Accumulation of immunity by the mother therefore must occur earlier, during the first 1–12 years of life in humans. But this necessitates mechanisms that keep the induced antibodies at high enough titers for efficient transfer of protection to offspring. Two questions arise from this: how is it done and why do adult males have antibody memory? Antibody titers depend on the presence of antigen during plasma cell maturation; if antigen disappears, antibody titers dwindle. Antigen-antibody complexes on follicular dendritic cells[3,5] play an important role as antigen depots to maintain antibody responses, but nevertheless many antigens eventually disappear. Therefore booster infections or immunizations against tetanus and polio are necessary every 3–10 years.

Although preexisting protective antibody titers are obviously of some advantage to all immune hosts, including males, antibody concentrations are higher in females, leaving them with a 4:1 to 5:1 higher chance of developing autoantibody–mediated autoimmune diseases after the age of 30. Thus the key time period in which memory antibody is essential for survival is during the immunoincompetent early period of newborns. This is most dramatically illustrated by the fact that calves that are not given colostral milk within 18 h of birth, the short period when intact antibodies can be absorbed in the gut, do not possess protective maternal antibodies and die within the next 4–6 weeks of many infections (fetal calf serum has no immunoglobulins, which is why it is used for experimentation)[36].

What about T cell memory[27,28,31,35,38]? There is good evidence and a general consensus of opinion that T cell precursor frequencies increase after priming. The real question is whether this suffices for protection against an infectious challenge. Is protection dependent on how cytopathic the infection is? Many experimental results, along with epidemiological experience of infection with polio or influenza viruses, indicate that priming of T cell help or of CTLs is not sufficient to efficiently prevent reinfection with a distinct serotype weeks to months after the primary infection[1,3,35,39]. As stated earlier, the mere fact that viruses and bacteria have been able to coevolve with the hosts to express distinct serotypes indicates that the overall evolutionary balance between the two crucially depends on protection via neutralizing antibodies and not primed T cell help or primed CTLs[1,2,39].

T cells are, however, very important in keeping noncytopathic or poorly cytopathic chronic infections (such as tuberculosis or leprosy) or HIV infections under control. Experimental evidence suggests that once these latter types of infections are completely eliminated to a sterile state, a rapidly protective T cell response may fade rather quickly[27,35]. This is particularly well illustrated by the clinical experience that BCG (an attenuated tuberculosis bacillus) persists in vaccinated infants for anything from a few months to a few years and is then eventually lost, but that in parallel to BCG disappearing, protection fades[40].

A similar argument can be made for HIV or the variabilities of some vaccines (those against measles or mumps, for example). The coexistence of continuing immunity together with low-level infections is exemplified by tuberculosis, leprosy, HIV, hepatitis C virus, for example, and has appropriately been called infection-immunity[41]. Therefore, one may speculate that many low-cytopathic or noncytopathic types of infections (including HIV) never leave us, but are controlled for a long period of time by infection-immunity, at least through the period of procreation (when evolutionary selection can be exerted). Once immune reactivity diminishes, the residual infection reemerges, as is experienced by immunosuppressed patients or elderly people infected with tuberculosis[40]. Collectively, therefore, protective immunological memory may depend largely on persistent or reen-

countered antigen that keeps sufficient effector T cells activated and sufficient B cells maturing to become antibody-producing plasma cells that protect offspring against new infections from the outside and protect the host against expanding infections from within.

Conclusion

Usually biological processes are variable and can be approximately fitted to a bell-shaped curve. Modern analytical methods have enabled us to measure frequent and also very rare and low-level effects. These latter observations are interesting and may reveal the limits of the system. They may even be exploitable for preventive or therapeutic measures. However, it is the bulk of the observed phenomena that usually reflects the general physiology of the system. Our experimental approaches are still much too fragmented and biased, so we have difficulties in finding good answers to many of the open biological questions. Perhaps we should also be wary of hypotheses that are too attractive, such as networks or suppressor T cells or danger theories or cross-priming, which may be misleading because they attempt to build a generalizable theory on rare exceptions. From an evolutionary point of view, the quality control of our understanding the results of research on immune responses cannot be intellectual seduction, but should be protection against infection.

1. Mims, C.A. *Pathogenesis of Infectious Disease* (Academic Press, London,1987).
2. Zinkernagel, R. M. *Science* **271**, 173–178 (1996).
3. Bachmann, M. F. & Zinkernagel, R. M. *Annu. Rev. Immunol.* **15**, 235–270 (1997).
4. Cohn, M. & Langman, R. E. *Immunol. Rev.* **115**, 11–147 (1990).
5. Nossal, G. J. *Adv. Immunol.* **52**, 283–331 (1992).
6. Matzinger, P. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
7. Bos, N. A. *et al. Eur. J. Immunol.* **21**, 1893–1898 (1991).
8. Bretscher, P. & Cohn, M. *Science* **169**, 1042–1049 (1970).
9. Lafferty, K. J. *Virology* **21**, 76–81 (1963).
10. Roost, H. -P. *et al. Proc. Natl Acad. Sci. USA* **92**, 1257–1261 (1995).
11. Healy, J. I. & Goodnow, C. C. *Annu. Rev. Immunol.* **16**, 645–670 (1998).
12. Nemazee, D. *Annu. Rev. Immunol.* **18**, 19–51 (2000).
13. Bachmann, M. F. *et al. Science* **262**, 1448–1451 (1993).
14. Coutinho, A. & Moller, G. *Adv. Immunol.* **21**, 113–236 (1975).
15. Dintzis, H. M. & Dintzis, R. Z. *Immunol. Rev.* **115**, 243–250 (1990).
16. Feldmann, M. J., Howard, J. G. & Desaymard, C. *Transplant. Rev.* **23**, 78–97 (1975).
17. Matzinger, P. *Immunol. Rev.* **135**, 81–117 (1993).
18. Fink, P. J. & Bevan, M. J. *Adv. Immunol.* **59**, 99–133 (1995).
19. Kisielow, P. & von Boehmer, H. *Adv. Immunol.* **58**, 87–209 (1995).
20. Zinkernagel, R. M. & Althage, A. *Proc. Natl Acad. Sci. USA* **96**, 8092–8097 (1999).
21. Kindred, B. *Prog. Allergy* **26**,137–238 (1979).
22. Rocha, B. & von Boehmer, H. *Science* **251**, 1225–1228 (1991).
23. Frey, J. R. & Wenk, P. *Int. Arch. Allergy Appl. Immunol.* **11**, 81 (1957).
24. Karrer, U. *et al. J. Exp. Med.* **185**, 2157–2170 (1997).
25. Lafferty, K. J. & Woolnough, J. *Immunol. Rev.* **35**, 231–262 (1977).
26. Kundig, T. M. *et al. Science* **268**, 1343–1347 (1995).
27. Kundig, T. M. *et al. Immunol. Rev.* **150**, 63–90 (1996).
28. Ochsenbein, A. F. *et al. Proc. Natl Acad. Sci. USA* **96**, 2233–2238 (1999).
29. Ohashi, P. S. *et al. Cell* **65**, 305–317 (1991).
30. Ludewig, B., Odermatt, B., Ochsenbein, A. F., Zinkernagel, R. M. & Hengartner, H. *Immunol. Rev.* **169**, 45–54 (1999).
31. Ahmed, R. & Gray, D. *Science* **272**, 54–60 (1996).
32. Nathanson, N. & McFadden, S. F. *Viral Path.* 85–108 (1997).
33. Webb, S. R. & Sprent, J. *Adv. Exp. Med. Biol.* **292**, 149–157 (1991).
34. Moskophidis, D., Lechner, F., Pircher, H. P. & Zinkernagel, R. M. *Nature* **362**, 758–761 (1993).
35. Zinkernagel, R. M. *et al. Annu. Rev. Immunol.* **14**, 333–367 (1996).
36. Brambell, R. W. R. (North Holland, Amsterdam,1970).
37. Bernet, F. M. & Fenner, F. J. *The Production of Antibodies* 2nd edn (McMillan, Sidney 1949).
38. Murali-Krishna, K. *et al. Science* **286**, 1377–1381 (1999).
39. Gupta, S. C., Hengartner, H. & Zinkernagel, R. M. *Proc. Natl Acad. Sci. USA* **83**, 2604–2608 (1986).
40. Bloom, B. R. & McKinney, J. D. *Nature Med.* **5**, 872–874 (1999).
41. Mackaness, G. B. *Semin. Hematol.* **7**, 172–184 (1970).