

genomic insult. This integrated field is now deciphering the complex regulatory pathways transduced by signalling mechanisms that detect DNA damage and/or arrested DNA replication. As these pathways become better understood, parallel technological gains in gene therapy and therapeutic intervention by rational drug design will offer new strategies for blocking the unwanted consequences of DNA damage, especially cancer.

We must remember, however, that while evolution could not have transpired without robust cellular mechanisms to ameliorate the most serious consequences of spontaneous and environmental DNA damage, the process of evolution mandates that the genetic diversification on which Darwinian selection operates be maintained constantly. Thus, life is necessarily a delicate balance between genomic stability and instability — and of mutation and repair. □

doi:10.1038/nature01408

1. Crick, F. The double helix: a personal view. *Nature* **248**, 766–769 (1974).
2. Friedberg, E. C. *Correcting the Blueprint of Life: An Historical Account of the Discovery of DNA Repair Mechanisms* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
3. Friedberg, E. C. The intersection between the birth of molecular biology and the DNA repair and mutagenesis field. *DNA Repair* **1**, 855–867 (2002).
4. Pontecorvo, G. *Trends in Genetic Analysis* (Columbia Univ. Press, New York, 1958).
5. Timoféeff-Ressovsky, N. W., Zimmer, K. G. & Delbrück, M. Über die Natur der Genmutation und der Genkonstruktion. *Nachr. Ges. Wiss. Göttingen FG VI Biol. N.F.* **1**, 189–245 (1935).
6. Muller, H. J. Artificial transmutation of the gene. *Science* **66**, 84–87 (1927).
7. Zimmer, K. G. in *Phage and the Origins of Molecular Biology* Expanded edn (eds Cairns, J., Stent, G. S. & Watson, J. D.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992).
8. Hollaender, A. & Curtis, J. T. Effect of sublethal doses of monochromatic ultraviolet radiation on bacteria in liquid suspension. *Proc. Soc. Exp. Biol. Med.* **33**, 61–62 (1935–36).
9. Kellner, A. Effect of visible light on the recovery of *Streptomyces griseus conidia* from ultraviolet irradiation injury. *Proc. Natl Acad. Sci. USA* **35**, 73–79 (1949).
10. Dulbecco, R. Reactivation of ultra-violet-inactivated bacteriophage by visible light. *Nature* **162**, 949–950 (1949).
11. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (American Society of Microbiology Press, Washington DC, 1995).
12. Watson, J. D. & Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).
13. Watson, J. D. *The Double Helix. A Personal Account of the Discovery of the Structure of DNA* (Atheneum, New York, 1968).
14. Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
15. Hanawalt, P. C. & Haynes, R. H. The repair of DNA. *Sci. Am.* **216**, 36–43 (1967).
16. Setlow, R. B. & Carrier, W. L. The disappearance of thymine dimers from DNA: an error-correcting mechanism. *Proc. Natl Acad. Sci. USA* **51**, 226–231 (1964).
17. Rasmussen, D. E. & Painter, R. B. Evidence for repair of ultra-violet damaged deoxyribonucleic acid in cultured mammalian cells. *Nature* **203**, 1360–1362 (1964).
18. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
19. Le Page, F. *et al.* Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101**, 159–171 (2000).
20. Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71**, 17–50 (2002).
21. Zhou, B.-B. S. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
22. Cory, S. & Adams J. M. The Bcl2 family: regulators of the cellular life-or-death switch. *Nature Rev. Cancer* **2**, 647–656 (2002).
23. Boveri, T. *The Origin of Malignant Tumors* (Williams and Wilkins, Baltimore, MD, 1929).
24. Vogelstein, B. & Kinzler, K. W. (eds) *The Genetic Basis of Human Cancer* (McGraw-Hill, New York, 1998).
25. Cleaver, J. E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652–656 (1968).
26. De Weerd-Kastelein, E. A., Keijzer, W. & Bootsma, D. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nature New Biol.* **238**, 80–83 (1972).
27. Wood, R. D. DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167 (1996).
28. Sancar, A. DNA excision repair. *Annu. Rev. Biochem.* **65**, 43–81 (1996).
29. Masutani, C. *et al.* Xeroderma pigmentosum variant: from a human genetic disorder to a novel DNA polymerase. *Cold Spring Harbor Symp. Quant. Biol.* **65**, 71–80 (2000).
30. Levinson, G., & Gutman, G. A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**, 5323–5338 (1987).
31. Aaltonen, L. A. *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816 (1993).
32. Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–1038 (1993).
33. Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215–1225 (1993).
34. Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227–1236 (1993).

Acknowledgements

I thank numerous individuals for their constructive comments. This article is dedicated to James Watson with thanks for his inspiration as a writer and historian.

The double helix and immunology

Gustav J. V. Nossal

Department of Pathology, The University of Melbourne, Victoria 3010, Australia

The immune system can recognize and produce antibodies to virtually any molecule in the Universe. This enormous diversity arises from the ingenious reshuffling of DNA sequences encoding components of the immune system. Immunology is an example of a field completely transformed during the past 50 years by the discovery of the structure of DNA and the emergence of DNA technologies that followed.

“This short history of research in one area, lymphocyte receptors, is yet another witness to the power of DNA technology, and to the ability of this approach not only to explain known biological phenomena, but also to contribute to the discovery of new biological systems.”

Susumu Tonegawa, Nobel lecture, 8 December 1987.

The double helix is all about biological information: how it is encoded, stored, replicated and used when required. Immunology, too, is about information. What genetic processes control the vast array of synthetic potential within an immune system capable of reacting specifically to virtually any microbe or foreign molecule? The secret lies in unique DNA processing that occurs during the development of lymphocyte cells, which are responsible for the specific immune response to a foreign agent (antigen). B lymphocytes produce antibodies (on their surface as well as secreted) and T lymphocytes mount cellular attacks on pathogenic infiltrators.

As lymphocytes develop, an array of short genes are rearranged and assembled together at the DNA level to form genes whose products recognize distinct antigens. As the process is mostly random, each lymphocyte makes different choices and thus the result is a vast repertoire of lymphocytes reactive to different antigens. This process has implications for antibody formation, cell-mediated immunity and malignancies of the immune system.

One B cell produces one antibody

At the beginning of the last century, Paul Ehrlich¹ recognized that the specificity of antibodies lay in the complementarity of their shapes to the antigen (s) on the microbe being recognized. He saw antibodies as cellular ‘side chains’, which budded out from the cell surface as what today we would term receptors. Karl Landsteiner² then demonstrated the exquisite specificity of antibodies, showing that animals could make antibodies to almost anything, including small synthetic organic molecules that had never previously existed in nature. Moreover, tiny structural changes in the antigen could lead to the production of a different antibody. It beggared belief that there could be so many different side chains. When antibodies were shown to be proteins, it seemed natural to conclude that a specific antibody molecule was shaped in close proximity to an antigen molecule much as plastic or sheet metal is moulded against a template. This ‘direct template’ hypothesis³ held sway for several decades.

In 1955, Niels Jerne⁴ published his natural selection theory of antibody formation, which postulated the random synthesis of a million or more different sorts of antibodies. When an antigen enters the body, it unites with an antibody that just happens to fit it, the antigen–antibody complex is taken up by a cell and the antibody

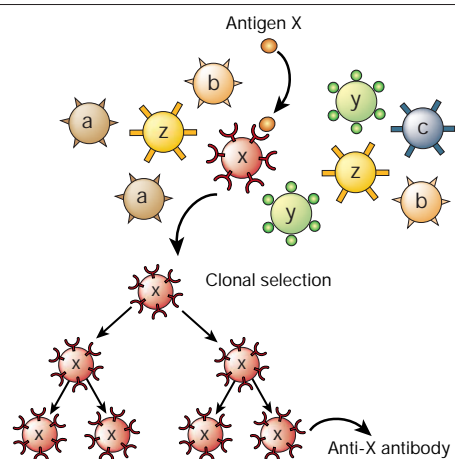


Figure 1 The clonal selection theory of antibody formation. Each B lymphocyte produces only one type of antibody receptor on its surface. An antigen recognizes one B lymphocyte out of a large repertoire. This triggers the rapid division and differentiation of the B cell to become a 'plasma cell', producing and secreting antibodies specific to the original antigen.

somehow acts as the template for the formation of more of itself. David Talmage⁵ and Macfarlane Burnet⁶ recognized that this theory would make more sense if the postulated natural antibodies were located on the surface of what we now call B lymphocyte cells. If each cell were endowed with only one sort of antibody specificity, then the antigen could select one lymphocyte out of a repertoire, cause its clonal division and stimulate antibody production and secretion (Fig. 1). In 1958, Joshua Lederberg⁷ and I provided the first evidence for the clonal selection theory, namely that one B cell always produces only one antibody.

DNA shuffling in antibody formation

Antibodies are multichain proteins that come in different forms. The most abundant, immunoglobulin- γ (IgG), consists of two identical light (L) chains and two identical heavy (H) chains⁸ (Fig. 2a). The

carboxy-terminal halves of the two light chains are identical to each other, but the amino-terminal halves differ in more than 50 residues, called the 'variable' region⁹. The heavy chains, too, consist of a variable (V) part and a constant (C) part.

In 1965, William J. Dreyer and J. Claude Bennett¹⁰ bucked the dogma at the time that 'one gene makes one protein', and put forward the revolutionary concept that the carboxy-terminal C region of the L chain was always encoded by a single gene, but that the amino-terminal V half could be encoded by multiple separate genes, perhaps as many as 100,000 in number. It followed that a chosen V gene must then somehow become associated with a C gene by a DNA rearrangement event in each lymphocyte cell, because only when the V-C regions were spliced together could a functional protein be expressed.

At that time, there was no way to interrogate the genome directly to test this concept, and for a decade debate and controversy raged. One school, led by Leroy Hood, favoured the idea that a large array of germline-encoded V genes for the L and H chains underwent rearrangement. At the other extreme were proponents of a single, very highly mutable V gene that was extensively mutated in emerging B cells. In the middle were those who favoured the idea of a handful of V genes that were subject to extensive recombination in somatic cells (the cells of the body, excluding the sex cells). This compromise was supported by luminaries such as Oliver Smithies and Gerald Edelman. Francis Crick was quite taken with the idea that just two V genes undergo rearrangement in the germ line, with further mutation in somatic cells.

Before arriving at the solution brought by advances in molecular biology, one more fact is worthy of note. Elvin Kabat astutely pointed out that in the V regions of both heavy and light chains there were also three short stretches of amino acids where variation was considerably greater than elsewhere in the molecules, and these so-called hypervariable regions were deemed likely to be sites of union with the antigen¹¹. Could it be that there was actually an assembly of several, rather than just two, genes encoding each chain?

The new tools for manipulating and sequencing DNA came to the rescue. In 1976, Nobumichi Hozumi and Susumu Tonegawa¹² conducted a landmark experiment. They used a DNA cutting enzyme known as a restriction endonuclease to digest the DNA extracted from a mouse embryo and from an antibody-secreting tumour. The resulting DNA fragments were then separated on the basis of size, and

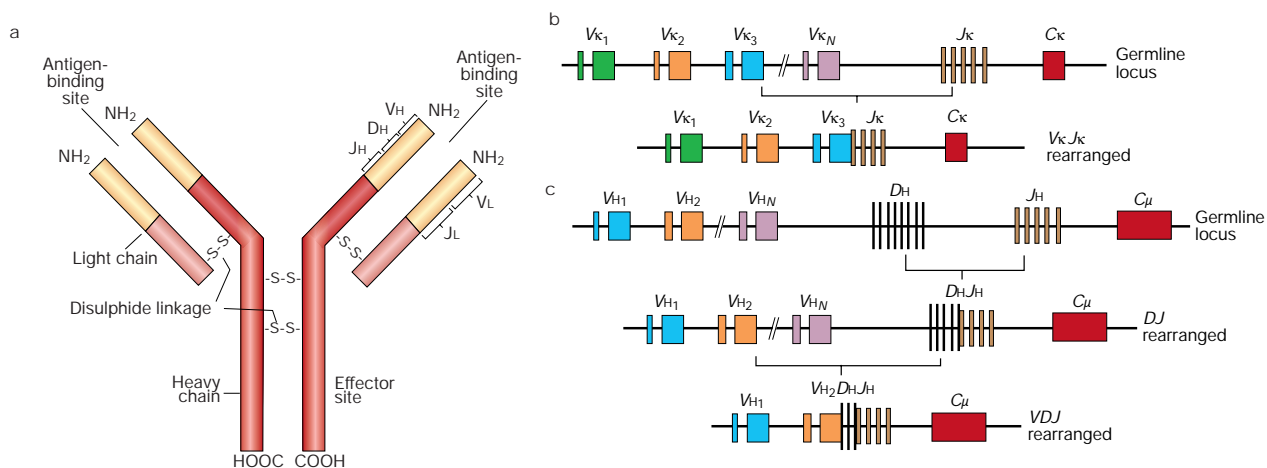


Figure 2 Antibody formation. **a**, Structure of an antibody (immunoglobulin). Two identical heavy chains are connected by disulphide linkages. The antigen-recognizing site is composed of the variable regions (yellow) of the heavy and light chains, whereas the effector site (which determines its function) is determined by the amino-acid sequence of the heavy chain constant region (red). **b**, Assembly of the light (κ -) and heavy (H)-chain genes of antibodies by somatic recombination during B-lymphocyte development. The L

chain is encoded by variable (V), joining (J) and constant (C) genes. While the developing B cell is still maturing in the bone marrow, one of the 30–40 V genes combine with one of the five J genes and is juxtaposed to a C gene. The recombining process involves deletion of the intervening DNA between the selected genes. **c**, The H chain is encoded by V, D, J and C genes. The assembly of the H chain occurs in two stages: one of the D genes joins with a J gene, then one of the V genes joins with that DJ assembly.

were reacted with radioactive probes, one corresponding to the whole L chain, the other to just the C portion. In the tumour, both probes lit up the identical fragment, whereas in the embryonic extract, two different fragments hybridized to the full-length probe, but only one of them to the C-region probe. Both fragments from the embryo sample were different in size from the single hybridizing fragment in the tumour sample. The experiment strongly argued for the V and C genes being some distance from each other in the embryo, but having been rearranged and assembled during development of antibody-forming cells in adults to form a continuous DNA sequence constituting the full L-chain gene.

Generating more diversity

Definitive elucidation of immunoglobulin gene structure depended on molecular cloning and subsequent sequencing of the genes themselves¹³. Here came another surprise, and one that could really not have been anticipated. V-region genes in the germline were found to be significantly shorter than is required to code for the V region of the L chain. It turned out that there is a series of 'minigenes' known as 'joining' (J) genes, which code for about 13 amino acids of the L chain. Thus, the full L chain is actually encoded by V, J and C genes (Fig. 2b). For the H chain, it is still more complicated, as there exists a series of 'diversity' (D) genes that encode up to eight amino acids that lie between the V and J regions. Thus, the H chain is encoded by V, D, J and C genes (Fig. 2c). The assembly of a complete H-chain V region occurs in two separate steps: first, one of the D regions joins with one of the J regions, then one of many V regions joins with that DJ assembly (Fig. 2c). The joining process is followed by deletion of the intervening DNA between the chosen minigenes. This is the first example of a somatic cell possessing a different genome from its fellow cells.

This minigene assembly process has important implications for antibody diversity. In humans, there are two types of L chains, κ and λ , each with its own sets of V and J genes. For the κ light chain, there are 40 functional V genes and 5 functional J genes; for the λ chain, there are 31 and 4, respectively. There is only one kind of variable region for the H chain, encoded by 51 V genes, 25 D genes and 6 J genes. To a first approximation, therefore, there are $(40 \times 5) + (31 \times 4) = 324$ different possible assemblies of L chains, and $51 \times 25 \times 6 = 7,650$ combinations for H chains. Thus, together, there are potentially 2,478,600 different types of germline-encoded antibodies.

But this is a considerable underestimate for two reasons. Recombination junctions can occur at different positions and this junctional diversity increases variability. Furthermore, a few extra nucleotides, called N regions, can be inserted between D–J junctions and V–D junctions in many H chains, and in a smaller percentage of L-chain V–J junctions. These nucleotides are not present in the germline and add to antibody diversity.

Yet further diversity can be generated by DNA mutations in dividing B cells. B cells expressing newly assembled immunoglobulin genes, each with its own unique specificity, constitute the 'primary repertoire'. When an antigen stimulates a chosen B cell to divide (Fig. 1), a proportion of the progeny migrate into the vicinity of antigen-capturing follicular dendritic cells (FDCs) and gradually form a 'germinal centre'. FDCs retain antigen on their surface for long periods and stimulate further rounds of division. Within the germinal centre the B cells display an extraordinarily high rate of somatic mutation in V genes, estimated at 10^{-3} per nucleotide per division¹⁴. As antibody production accumulates, only those B cells with heightened affinity for the antigen gain access to FDC-bound antigen and thus are further stimulated to divide. As a result, B-cell clones secreting higher affinity antibody are selected in an iterative manner.

The 'memory' B cells that emerge from the germinal centre constitute the 'secondary repertoire', which is even more diverse than the primary one. Twenty mutations per chain are not uncommon; nor are thousand-fold increases in affinity. Thus, as it turned out, two early theories of antibody diversification proved to be correct: rearrangement of germline genes gives the naive B-cell

Box 1 Tolerating self

How does the immune system distinguish foreign molecules from the cells of its own body? The failure to form antibodies against self components confounded scientists until Macfarlane Burnet proposed that an antigen introduced during embryonic development would 'trick' the immune system into regarding it as self, so that even in later life it could not react against that antigen²². Peter Medawar's group proved the prediction correct by inducing immunological tolerance to transplantation antigens. Injecting foreign cells from a donor into late-stage embryo mice within the uterus led to a situation where, when these mice had grown up, they could not reject skin grafts from the donor strain²³.

What happens if the antibody generated by a newly created B cell is reactive with a self component? The cell either dies immediately, if the stimulus is one capable of strongly cross-linking the cell-surface immunoglobulin receptors, or else is given a negative signal which impedes its full function²⁴. This negative selection contributes to immunological tolerance.

For a T cell, this is more of a juggling act. Although it must recognize the self-molecule major histocompatibility complex (MHC), it must avoid mounting an abnormal immune response to the host's own cells. For the T-cell repertoire to be fashioned correctly within the thymus, two sorts of selection are required. If the T cell, having assembled its receptor genes, has the capacity to recognize self MHC, it is allowed to mature. This step is known as positive selection. But the T cell must not leave the thymus if it has strong potential to cause autoimmunity. This would be the case if there were too high an affinity of binding with self MHC, or with a peptide from a normal cellular component bound to the MHC groove. Such cells undergo programmed cell death within the thymus, the process of negative selection.

Defects in the immune system's ability to recognize self are the basis of many autoimmune diseases, such as juvenile diabetes and multiple sclerosis.

repertoire, and somatic mutation ensures further diversification during memory B-cell development.

Switching function

There are several different classes of antibodies, all of which have distinct roles that are also produced by rearrangements at the DNA level. There are eight different genes for the C region of the H chain, which specify different antibody functions. Each B cell first links the chosen VDJ assembly to a C gene known as μ , creating an antibody class called IgM. If that cell is propelled into a pathway favouring the production of an antibody prominent in mucus secretions (such as in the gastrointestinal tract), the VDJ section is switched over to a C region encoded by C gene α , and the cell produces IgA. If, on the other hand, the antigen is of parasite origin, or an allergen such as a pollen grain, the cell may be stimulated to produce IgE, in which case the VDJ region associates with the product of the C gene ϵ . All of this occurs without any change in the specificity of the antibody being secreted. Although the detailed molecular mechanisms are still being investigated, the class switching again involves sequential excision of portions of the genetic material. Cytidine deaminase induced by B-cell-specific activation may be significant in both class switching and somatic hypermutation.

Assembly of T-cell receptors

Whereas B cells make antibodies against antigens, the thymus-derived or T cells also respond to foreign agents, specializing in a more localized form of combat. Cytotoxic T cells are capable of killing virus-infected cells or cells displaying cancer-specific antigens. Other

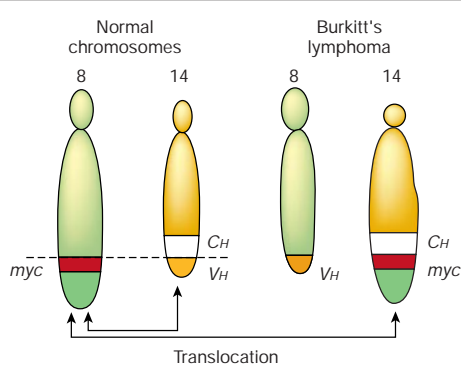


Figure 3 Reciprocal chromosomal translocations in Burkitt's lymphoma, a solid tumour of B lymphocytes. The genes for making the heavy chains of antibodies (C_H) are located on chromosomes 14, whereas those for making the light chains are on chromosomes 2 and 22. These genes are expressed exclusively in B lymphocytes, because only these cells have the necessary transcription factors to switch on their expression. In most (over 90%) of Burkitt's lymphoma cases, a reciprocal translocation moves the proto-oncogene *c-myc* from its normal position on chromosome 8 to a location close to the antibody heavy-chain genes on chromosome 14 (ref. 18). In other cases, *c-myc* is translocated close to the antibody genes on chromosome 2 or 22. In every case, *c-myc* now finds itself in a region of active gene transcription, and it may simply be the overproduction of the *c-myc* product (a transcription factor essential for cell division) that propels the lymphocyte down the pathway towards cancer.

T cells secrete powerful stimulatory and inflammatory molecules, most of which act in a strictly localized context. T cells can also help guide B cells down appropriate pathways of differentiation.

In common with B cells, T cells also have one-receptor specificity and, for simplicity, I shall mention only the $\alpha\beta$ T-cell receptor (TCR), a heterodimer consisting of two subunits, the α - and β -chains, joined by disulphide bonds^{15,16}. The strategy for generating T cells with different receptors is strikingly similar to that used by B cells to produce different antibodies and, in fact, the TCR binding surface looks much like that of an antibody. The β -chain of the TCR is assembled in somatic cells from V, D, J and C genes; the α -chain from V, J and C genes. There are additions of N-region nucleotides between V and D, as well as between D and J on the β -chain; and between V and J on the α -chain. A similar rearrangement also takes place for the $\gamma\delta$ TCR.

But something peculiar about T-cell recognition was noted by Rolf Zinkernagel and Peter Doherty¹⁷, who demonstrated that cytotoxic T cells could recognize viral antigens only if a specific 'self' molecule were also present on the target cell (see Box 1). The key part of the T-cell recognition puzzle fell into place when it was discovered that the TCR recognized short antigenic peptides bound to the groove of a self molecule known as the major histocompatibility complex (MHC), as well as surrounding portions of the MHC molecule itself. Cells have special mechanisms for fragmenting proteins into peptides of 8–24 amino acids in length, attaching these to MHC molecules and transporting the entire complex to the cell surface. TCRs then 'see' these short linear portions of antigens, be these of viral, bacterial or parasitic origin, or even portions of normal intracellular components. Such a system can help to control infections where the pathogen goes 'underground' inside a cell, and can also eliminate cells with mutated self antigens, such as cancer cells.

Lymphocytes and cancer

Lymphocytes have been a favourite tool in cancer research. A notable example of DNA science applied in this way relates to the B-cell tumour of humans known as Burkitt's lymphoma. Occasionally DNA strands break and are incorrectly repaired. Thus, a piece of a chromosome becomes attached to the broken end of another one, and vice versa, in a process known as reciprocal translocation (Fig. 3).

In the case of Burkitt's lymphoma, a tumour-promoting gene or oncogene called *myc* is translocated from its normal position on chromosome 8 right into the middle of the IgH chain locus on chromosome 14 (ref. 18). In this highly active transcriptional environment, *myc* expression is switched on, and eventually cancer develops.

It has been possible to create lymphoma-prone transgenic mice, which express *myc* in aberrantly high amounts. Because cancer is typically a multistage process, if further oncogenes are expressed simultaneously in transgenic mice, the onset of cancer can be dramatically accelerated. One such example involves the gene *bcl-2*. When this gene is expressed, it stops cells from undergoing natural programmed death (apoptosis)¹⁹. Mice expressing *myc* and *bcl-2* showed very rapid development of tumours. An enormous amount of literature has accumulated related to the expanding family of *bcl-2*-related genes and their roles in the regulation of programmed cell death. Models derived from lymphocytes and their malignancies have led to insights with implications well beyond immunology.

DNA vaccines

DNA research has been of immense value to vaccine research. Through gene cloning and expression, candidate antigens can be identified and tested. In an era of rapid nucleotide sequencing, the whole genome of a pathogen can be determined, and computer programs can search for sequences likely to encode outer membrane proteins, which can be assessed as candidate vaccine molecules (see, for example, the series of papers published recently in *Nature* (419, 489–542, 2002) on the genomics of the malaria parasite).

Amazingly, DNA itself can serve as a vaccine. DNA vaccines work on the principle that the gene sequence for one or more candidate antigens is introduced into an animal or person via a delivery vehicle known as a vector, together with a strong promoter that can switch on its expression in mammalian cells. Cells that take up the injected DNA transcribe and translate the gene and release the relevant antigen protein, which the body can in turn manufacture antibodies against. Thus, the body itself becomes a vaccine factory²⁰. Unfortunately, so far this approach has worked better in mice than in humans, but many avenues are being pursued to improve this situation.

To strengthen the immune response to a vaccine, it may be necessary to use an adjuvant substance. Here, DNA may also be of potential use. Scavenger cells, which capture antigens, have evolutionarily conserved receptors, known as Toll-like receptors (TLRs), which recognize antigens common to many pathogens. One such receptor is TLR-9, which recognizes unmethylated CpG motifs commonly found in bacterial but not mammalian DNA. Accordingly, unmethylated CpG-rich DNA sequences represent a promising new category of adjuvant²¹.

Future directions

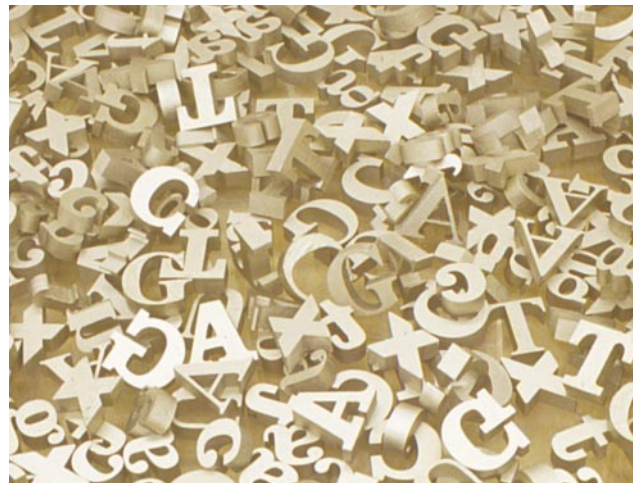
The solutions to the puzzle of antibody diversity and mystery of T-cell recognition of antigenic peptides are among the brightest chapters of biology in the last quarter of the twentieth century. The future of immunology will be all about how the system is regulated and how it makes decisions: whether to respond or not; whether to direct efforts towards antibody formation or cell-mediated immunity; and, if the latter, whether more towards cytokine-secreting T cells or cytotoxic T cells.

As in the past, the future will be about information and thus about DNA science. All the complex signalling pathways, the feedback loops, and the intricate rules governing cell division on the one hand or programmed cell death on the other, will be progressively revealed. As this happens, the possibilities for applied research and development will be immense. In particular, new therapeutic targets will be identified. The 'miracle' drug for chronic myelogenous leukaemia, Glivec, was made possible after the characterization of the extraordinary cancerous potential of the chimaeric oncogene *bcr-abl*. This will surely be only the first of a plethora of more intelligently designed

anti-cancer drugs. Potent cytokines and monoclonal antibodies directed against cell surface-associated structures are already prominent within a radically revised pharmaceutical armamentarium in areas including cancer, autoimmunity, allergy and transplantation. DNA research is therefore crucial to a new generation of immunologists, from those striving towards the development of novel vaccines to those seeking to understand and control autoimmune diseases, allergy and transplant tolerance. □

doi:10.1038/nature01409

- Ehrlich, P. The Croonian Lecture. On immunity with special reference to cell life. *Proc. R. Soc. Lond. B* **66**, 424–448 (1900).
- Landsteiner, K. *The Specificity of Serological Reactions* (Harvard Univ. Press, Boston, 1945).
- Breinl, F. & Haurowitz, F. Untersuchungen des Präzipitates aus Hämoglobin und anti-Hämoglobin-Serum und Bemerkungen über die Natur der Antikörper. *Hoppe-Seyler's Z. Physiol. Chem.* **192**, 45–57 (1930).
- Jerne, N. K. The natural selection theory of antibody formation. *Proc. Natl Acad. Sci. USA* **41**, 849–857 (1955).
- Talmage, D. W. Allergy and immunology. *Annu. Rev. Med.* **8**, 239–256 (1957).
- Burnet, F. M. A modification of Jerne's theory of antibody production using the concept of clonal selection. *Aust. J. Sci.* **20**, 67–69 (1957).
- Nossal, G. J. V. & Lederberg, J. Antibody production by single cells. *Nature* **181**, 1419–1420 (1958).
- Edelman, G. M. & Gall, W. E. The antibody problem. *Annu. Rev. Biochem.* **38**, 415–466 (1969).
- Hiltschmann, N. & Craig, L. C. Amino acid sequence studies with Bence-Jones proteins. *Proc. Natl Acad. Sci. USA* **53**, 1403–1409 (1965).
- Dreyer, W. J. & Bennett, J. C. The molecular basis of antibody formation: a paradox. *Proc. Natl Acad. Sci. USA* **54**, 864–869 (1965).
- Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132**, 211–250 (1970).
- Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl Acad. Sci. USA* **73**, 3628–3632 (1976).
- Bernard, O., Hozumi, N. & Tonegawa, S. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15**, 1133–1144 (1978).
- Kocks, C. & Rajewsky, K. Stable expression and somatic hypermutation of antibody V regions in B-cell developmental pathways. *Annu. Rev. Immunol.* **7**, 537–559 (1989).
- Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308**, 149–153 (1984).
- Yanagi, Y. *et al.* A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* **308**, 145–149 (1984).
- Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis by a syngeneic or semiallogeneic system. *Nature* **248**, 701–702 (1974).
- Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J. M. Sequence of the murine and human cellular myc oncogenes and two modes of myc transcription resulting from chromosome translocation in B lymphoid tumours. *EMBO J.* **2**, 2375–2383 (1983).
- Vaux, D. L., Cory, S. & Adams, J. M. *Bcl-2* gene promotes haemopoietic cell survival and cooperates with *c-myc* to immortalize pre-B cells. *Nature* **335**, 440–442 (1988).
- Donnelly, J. J., Ulmer, J. B., Shiver, J. W. & Liu, M. A. DNA vaccines. *Annu. Rev. Immunol.* **15**, 617–648 (1997).
- Krieg, A. M. CpG motifs in bacterial DNA and their immune effects. *Annu. Rev. Immunol.* **20**, 709–760 (2002).
- Burnet, F. M. & Fenner, F. J. *The Production of Antibodies* (Macmillan, Melbourne, 1949).
- Billingham, R. F., Brent, L. & Medawar, P. B. Actively acquired tolerance of foreign cells. *Nature* **172**, 603–606 (1953).
- Nossal, G. J. V. & Pike, B. L. Clonal anergy: persistence in tolerant mice of antigen-binding B lymphocytes incapable of responding to antigen or mitogen. *Proc. Natl Acad. Sci. USA* **77**, 1602–1606 (1980).



The digital code of DNA

Leroy Hood* & David Galas†

*Institute for Systems Biology, 4225 Roosevelt Way NE, Seattle, Washington 98105, USA (e-mail: lhood@systemsbiology.org)

†Keck Graduate Institute of Applied Sciences, 535 Watson Drive, Claremont, California 91711, USA (e-mail: david_galas@kgi.edu)

The discovery of the structure of DNA transformed biology profoundly, catalysing the sequencing of the human genome and engendering a new view of biology as an information science. Two features of DNA structure account for much of its remarkable impact on science: its digital nature and its complementarity, whereby one strand of the helix binds perfectly with its partner. DNA has two types of digital information — the genes that encode proteins, which are the molecular machines of life, and the gene regulatory networks that specify the behaviour of the genes.

"Any living cell carries with it the experiences of a billion years of experimentation by its ancestors." Max Delbruck, 1949.

The discovery of the double helix in 1953 immediately raised questions about how biological information is encoded in DNA¹. A remarkable feature of the structure is that DNA can accommodate almost any sequence of base pairs — any combination of the bases adenine (A), cytosine (C), guanine (G) and thymine (T) — and, hence any digital message or information. During the following decade it was discovered that each gene encodes a complementary RNA transcript, called messenger RNA (mRNA)², made up of A, C, G and uracil (U), instead of T. The four bases of the DNA and RNA alphabets are related to the 20 amino acids of the protein alphabet by a triplet code — each three letters (or 'codons') in a gene encodes one amino acid³. For example, AGT encodes the amino acid serine. The dictionary of DNA letters that make up the amino acids is called the genetic code⁴. There are 64 different triplets or codons, 61 of which encode an amino acid (different triplets can encode the same amino acid), and three of which are used for 'punctuation' in that they signal the termination of the growing protein chain.

The molecular complementarity of the double helix — whereby each base on one strand of DNA pairs with its complementary base on the partner strand (A with T, and C with G) — has profound implications for biology. As implied by James Watson and Francis Crick in their landmark paper¹, base pairing suggests a template-copying mechanism that accounts for the fidelity in copying of genetic material during DNA replication (see article in this issue by Alberts, page 431). It also underpins the synthesis of mRNA from the DNA template, as well as processes of repairing damaged DNA (discussed by Friedberg, page 436).

Tools to modify DNA

The enzymes that function in cells to copy, cut and join DNA molecules were also exploited as key tools for revolutionary new techniques in molecular biology, including the cloning of genes and expression of their proteins, and mapping the location of genes on chromosomes. The ability to recreate the process of DNA replication artificially in the laboratory led to the development of two techniques