Immunogenetics

Lymphocyte development genes and immunodeficiency disease

from S. Darling and P. Goodfellow

UNLIKE AIDS, some human immuno-deficiency diseases are inherited rather than acquired. Several of these genetic diseases are linked to X-chromosome loci and are responsible for catastrophic illness[1,2]. As neither the immunoglobulin genes[3] nor the T-cell receptor genes are themselves sex linked[4-6], it has been suggested that the X-linked disease loci might include regulatory genes responsible for lymphocyte development. This hypothesis is supported by two papers in this issue[7,8], which report the isolation of an X-linked sequence that is developmentally expressed in lymphocytes and is aberrantly expressed in a mouse model for X-linked immune deficiency.

CBA/N mice have an X-linked mutation, *xid* (reviewed in ref.9), which renders them incapable of producing antibodies to soluble polysaccharides. Their immuno-deficiency is caused by a B-cell defect and the mice seem to lack a specific subpopulation of mature B cells. In humans, a similar defect in antibody production is present in the X-linked Wiskott-Aldrich syndrome[10], which, unlike the mouse disease, is accompanied by defects in T-cell and platelet function. Arguing that the mouse *xid* genes affected in X-linked immune deficiencies would be expressed in lymphocytes[7], Cohen *et al.* sought *xid*-related messenger RNA by the 'subtraction' cloning technology developed for isolating the T-cell receptor genes[11,12]. Of the 15 lymphocyte-specific cDNA clones thus prepared, one clone, which is apparently T-cell specific, recognized an X-linked gene family that might contain as many as 15-20 members. These genes have been named *XLR* (X-linked, lymphocyte-regulated)[7].

Two lines of circumstantial evidence suggest that the *XLR* genes are directly related to the *xid* locus. First, genetic studies with congenic strains of mice show there to be very close linkage between the *xid* locus and the *XLR* locus. Second, detailed expression studies detect *XLR*-derived

mRNA not only in T cells but also in plasma cells and other B cells at late stages of differentiation. Unlike plasmacytomas derived from normal mice, plasmacytomas from *xid* mutant mice do not contain *XLR* transcripts[8]. Independent evidence has implicated mature B cells as the target tissue for the *xid* locus[9]. So, either the *xid* locus is regulating the *XLR* locus or the two loci are identical. Final proof of identity will require demonstration of an alteration in an expressed *XLR* gene that is the result of the *xid* mutation.

The X-chromosome of eutherian mam-

mals has been conserved during evolution[13] and the presence of an *XLR* locus in humans can be predicted confidently. It will be of obvious interest to study *XLR* expression in Wiskott-Aldrich syndrome and in the other human X-linked, immunodeficiency syndromes. Even if mutation at the *XLR* locus is not directly responsible for disease, this gene family is certain to be studied in detail. With a developmentally-regulated expression that is apparently restricted to lymphocytes, and with the possibility that different members of the gene family are transcribed at different times, the *XLR* genes may well be responsible for regulating lymphocyte development and are therefore bound to be extensively examined in the coming months □

S. Darling and P. Goodfellow are at the Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK.

Molecular biology

Coordination of sequence data

from Arthur M. Lesk

MANY challenges are presented by the very rapid growth of DNA and protein sequence information, the opportunities arising from advances in computer data-base management facilities, and the variety and importance of the applications of this knowledge. Not least is the challenge of coordinating the various data-banking projects.

The main archives of nucleic acid sequences, at the European Molecular Biology Laboratory Library and GenBank (a collaborative effort between Los Alamos National Laboratory and Bolt, Beranek and Newman), now contain approximately 3×10^6 bases, about half of the total of published sequences. It is estimated that another 5×10^5 bases are currently being published every six months, and the process is still accelerating.

The major archive of amino-acid sequences of proteins (now often derived from gene sequences) is the Protein Information Resource, based at the National Biomedical Research Foundation in Georgetown, Virginia, now under the direction of W.C. Barker. It contains over 3,000 entries, comprising over 6×10^5 amino-acid residues. R.F. Doolittle of the University of California at San Diego, and S. Sakakibara of the Protein Reseach Foundation, Osaka, Japan, also maintain protein-sequence data banks. In addition, numerous specialized collections exist, such as the 'Sequences of Proteins of Immunological Interest', published by the US National Institutes of Health, and the International Haemoglobin Information Center, which is directed by R.N. Wrightstone of the Medical College of Georgia.

The Protein Structure Data Bank at Brookhaven National Laboratory, directed

by T.F. Koetzle, contains, in its latest release, 253 sets of atomic coordinates of proteins and nucleic acids, almost all derived from X-ray crystallography.

Current activities fall into four categories. First, there is data acquisition by an archive. The basic chain of events at present is: publication; surveillance of the literature and extraction of published data by the archiving project; quality control; editing, formating and addition to data base. Second, the archived information is distributed on magnetic tape or, if only small amounts of information are involved, over computer networks. Third, there is information retrieval, dependent on the development of computer program systems as interfaces between archived information and research projects. Finally, there are applications — for example, the search of a data base for sequences similar to the one that has been newly determined.

Those of us whose activities primarily involve applications depend, ever more crucially, on the quality, completeness and accessibility of the archives. Despite the stake of the scientific community at large in these resources, individual projects have so far borne the responsibility for collecting, maintaining and distributing the data. In contrast, many groups all over the world are developing software for information retrieval and research applications in molecular biology. To the extent that the computer programs are well described and readily distributed, they are subject to ordinary processes of natural selection from which, one may presume, the most effective will ultimately emerge in common use.

How can new generations of molecular biologists and new generations of com-

1. Waldman, T.A., Strober, W. & Blaese, R.M. in *Clinical Immunology* Vol. 1 314 (W.B. Saunders, Philadelphia, 1980).
2. McKusick, V.A. *Mendelian Inheritance in Man* (John Hopkins University Press, 1983).
3. *Human Gene Mapping* 7; *Cytogenet. Cell Genet.* 37, 1 (1984).
4. Collins, M.K., *et al. EMBO J.* 3, 2347 (1984)
5. Collins, M.K. *et al. Nature* (in the press).
6. Stinson, A. *et al.* (submitted).
7. Cohen, D.I. *et al. Nature* 314, 369 (1985).
8. Cohen, D.I., Steinberg, A.D., Paul, W.E. & Davis, M.M. *Nature* 314, 372 (1985).
9. Scher, I. *Adv. Immunol.* 33, 1 (1982).
10. Cooper, M.D., Chae, H.P., Lowman, J.T. Krivit, W. & Good, R.A. *Am. J. Med.* 44, 499 (1968).
11. Hedrick, S.M., Cohen, D.I., Nielsen, E.A. & Davis, M.M. *Nature* 308, 149 (1984).
12. Davis, M.M. *et al. Proc. natn. Acad. Sci. U.S.A.* 81, 2149 (1984).
13. Ohno, S. *Monograph Endocrinol.* 1 (Springer, 1967).