Cloning the Genes of the MHC

The cloning of histocompatibility genes moves ahead, permitting rapid advances in understanding the nature of the genes and how they work

St. Hilda's College, Oxford. Barely a year and a half have elapsed since gene cloners got their first toehold in the major histocompatibility complex (MHC), a large genetic region that controls many of the activities of immune cells. A meeting held in Oxford on 21 to 24 March to discuss the cloning results amply documents the rapid progress that is occurring. Walter Bodmer of the Imperial Cancer Research Fund Laboratory in London, who was the meeting organizer, says of the gene cloning, "It's going very quickly. It started with hemoglobin because you could get messenger RNA quickly and went to immunoglobulins [antibodies] because of the diversity problem. Now it has moved into the MHC."

Investigators have cloned and determined the complete nucleotide sequences of several MHC genes and are on the verge of doing the same for many more. Within the next year or two, they expect to have the complete molecular structures of representatives of the major classes of MHC genes and consequently of the protein products encoded by the genes. This will be the first look at some of these protein structures.

The information should shed light on the evolution of the MHC genes and on the origin of the diversity that is their hallmark. But more than that, the ability to clone the genes means that their activities, and particularly the role of their products in mediating the myriad interactions between immune cells, can be studied directly. "In the broadest sense," explains Leroy Hood of the California Institute of Technology, "the reason why these gene products are interesting is because this is the first system for which we can study cell-cell recognition at the molecular level. . . . We'll be able to do very powerful experiments to sort out how these genes work. It will revolutionize cellular immunology."

The MHC, which in the mouse comprises on the order of 2 million base pairs, contains several types of genes (see the figure on p. 401 for the organization of the MHC). Most advanced is the cloning of the class I genes, which include the classic transplantation antigens, substances present on the surfaces of all cells that elicit rejection of transplanted tissue by the immune system of the recipient. About a decade ago, investigators discovered that these same antigens are an essential component of normal immune responses, such as those in which killer T cells recognize and attack virus-infected target cells. The killer cell is "restricted" in that it can interact only with virus-infected cells that carry the same transplantation antigen as it does.

To get the cloning under way, investigators first had to solve a major problem, that is, how to prepare a probe that could pick the desired gene out of the 100,000 or so present in the mammalian genome. In the case of hemoglobin, this had been relatively easy because the cells that synthesize the protein make large quantities of hemoglobin messenger RNA (mRNA), which can be enzymatically copied into complementary DNA (cDNA). The cDNA binds to, and thus identifies, the corresponding gene. Very little messenger for any of the histocompatibility proteins is present in a cell, however, and this made preparation of appropriate probes much more difficult.

Two groups of investigators, using different approaches, solved the problem at almost the same time. In the October 1980 issue of the *Proceedings of the National Academy of Sciences*, Jack Strominger of Harvard University, Hidde Ploegh, who has since moved to the University of Cologne, and Harry Orr, who is now at the University of Minnesota, described a cDNA probe for the larger of the two subunits of a human class I antigen. Sherman Weissman and his colleagues at Yale University School of Medicine reported their probe, also a cDNA for the heavy chain of a human class I antigen, in the January 1981 issue of the *Proceedings*. These cDNA's provided an entry not just to the human MHC but also to that of the mouse.

At the Oxford meeting, several researchers* described the complete nucleotide sequences of heavy chain genes of both human and mouse origin. As their descriptions made clear, there is good agreement among the findings of the various laboratories. The genes for the larger of the two class I protein subunits, whether they come from mice or men, have similar overall designs. Not surprisingly, these correspond quite neatly with the designs of the proteins.

Studies on the transplantation antigens themselves had shown that the heavy chains have molecular weights of about 45,000 and contain about 350 amino acid residues. The molecules are situated on the cell surface with roughly the first 300 of the amino acid residues projecting outward. The next 25 or so are imbedded in the membrane, and the remainder dangle in the cytoplasm. The external portion of the heavy chain is further subdivided into structural regions called domains, each consisting of about 90 amino acid residues.

The complexity of the proteins is reflected in the complexity of the genes. Michael Steinmetz, in Hood's group, has shown that the mouse heavy chain genes, which contain about 5000 base pairs, consist of eight exons that code for the amino acid sequence of the protein. The exons are separated by seven noncoding introns. The first exon codes for the leader sequence, a segment of about 20 amino acids that helps to direct the newly synthesized protein to its proper location in the cell but is clipped off the final product. The next three exons code for the three external domains and the fifth codes for the transmembrane segment. The last three, which are very short, code for the cytoplasmic region.

The genes for the human heavy chains have similar structures, although there may be modest differences. For example, a human gene that was sequenced by Bertrand Jordan's group at the Centre D'Immunologie de Marseille closely resembles the mouse genes except that it has only two exons coding for the cytoplasmic region.

The smaller component of class I histocompatibility antigens is a protein called β_2-microglobulin, which is encoded by a gene that is not a member of the MHC. In contrast to the larger chain, many variants of which exist, β_2-microglobulin is an invariant component of the antigens. Its molecular weight is 12,000.

*Among the laboratories that have cloned these genes are those of Walter Bodmer of the Imperial Cancer Research Fund, London; Richard Flavell of the National Institute for Medical Research, Mill Hill; Leroy Hood of the California Institute of Technology; Bertrand Jordan of Centre D'Immunologie de Marseille; Philippe Kourilsky of the Institute Pasteur, Paris, and Bernard Dobberstein of the European Molecular Biology Laboratories, Heidelberg; Hugh McDevitt of Stanford University; Jonathan Seidman of the National Institute of Child Health and Human Development; and Sherman Weissman of Yale University School of Medicine.

0036-8075/82/0423-0400$01.00/0

Jane Parnes, who works with Jonathan Seidman at the National Institute of Child Health and Human Development, has cloned and determined the complete nucleotide sequence of the mouse β_2-microglobulin gene. This gene contains only four exons and three introns, with most of the amino acid sequence of the protein encoded within the second exon.

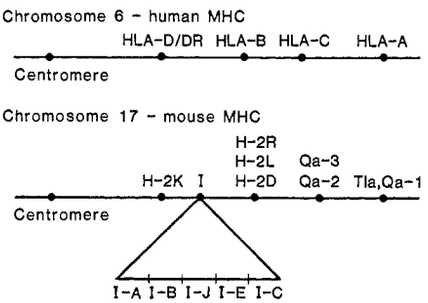
Analysis of the class II genes, which are encoded in the I region of the mouse MHC and the DR region of the human MHC, is not as far along as that of the class I genes. Unlike the classic histocompatibility antigens, the class II antigens are present only on certain types of immune cells where they participate in the cellular interactions needed to control antibody production by B cells. The class II antigens also contain two nonidentical protein chains, however.

So far no complete gene for a class II protein has been sequenced, although the Strominger group has determined almost all of the nucleotide sequence coding for the heavy chain of a human DR antigen. As described at Oxford by Alan Korman, this gene contains at least five exons and four introns. The first exon codes for the signal sequence, the next two for most of the two external domains of this protein, the fourth for the transmembrane and cytoplasmic regions, and the fifth for a nontranslated segment that appears near the end of the messenger for the chain.

Additional sequences can be expected soon, as other investigators have cloned and are beginning to characterize class II genes. The Bodmer and Weissman groups, for example, also have genes for a human DR heavy chain. Moreover, many investigators have cDNA's that will serve as probes for identifying either heavy or light chains. As Seidman notes, "Once you have the cDNA, it's just a matter of time until you get the gene."

One gene clone that attracted particular interest was described by Steinmetz. The I region of the mouse MHC encodes antigens needed for activation of antibody production by B cells and also contains a region thought to encode molecules involved in suppressing antibody production. The identity of the suppressor molecules and how they work are matters of some uncertainty, however.

Using a probe that was originally prepared in the laboratory of Bernard Mach at the University of Geneva, Steinmetz was able to produce a cloned DNA that extends into the suppressor region, which may enable him to locate and clone the suppressor genes. Bodmer says of this development, "We will find these other products [the suppressors] by walking along with the clones, and this will happen very quickly now."



The MHC's of mice and men

The human MHC is located on chromosome 6. There are three regions (HLA-A, -B, and -C) coding for the transplantation antigens of class I. The class II antigens, which are involved in regulating antibody production by B cells, map to the HLA-DR region. The mouse MHC is located on chromosome 17. The class I transplantation antigens map to the H-2K, -2D, -2L, and -2R regions. The Qa and Tla differentiation antigens also belong to class I. The mouse class II antigens map to the I region, which is subdivided as shown.

A major reason why the gene cloning successes are generating so much excitement is that they will enable investigators to study directly how histocompatibility antigens mediate immune cell interactions. In another display of converging research, several investigators presented results at Oxford showing that cloned genes, all for class I heavy chains at the moment, can be introduced into cultured cells. The gene products are made in the recipient cells and integrated into the cell membranes just as histocompatibility antigens are supposed to be. They can be detected there with appropriate monoclonal antibodies. And, says Hood, "The genes work in a perfectly normal physiological sense."

Both the Hood and Seidman groups have introduced the gene for the heavy chain of the L^d transplantation antigen into mouse cells of a different histocompatibility type. The cells carrying the transferred gene could be killed by allogeneic cytotoxic T cells directed against the L^d antigen, whereas they are not normally attacked by killers of that specificity. Very similar experiments, with a heavy chain gene of a different histocompatibility type, were performed by Richard Flavell and his colleagues at the National Institute for Medical Research, Mill Hill.

In addition, the Hood group has shown that the transformed cells participated in restricted interactions with killer T cells. Cells bearing the introduced L^d gene and also infected with lymphocytic choriomeningitis (LCM) virus were killed by T cells specific for both the virus and the L^d histocompatibility antigen. Ordinarily, T cells restricted for L^d antigen would not interact with the recipient cell type.

This type of experiment may also shed some light on the varying degrees of susceptibility to viral infections among different members of a population. Experiments by the Hood group show that a particular virus may be recognized by killer T cells only when presented on the surfaces of infected cells together with a specific histocompatibility antigen. LCM virus was recognized only in conjunction with the L^d, and not with the K^d, antigen. This finding suggests that there is an interaction between the virus and histocompatibility antigen.

Studying the interactions of the histocompatibility antigens at the molecular level may help resolve an issue that has bedeviled cellular immunologists ever since restriction was discovered, that is, the nature of the T cell receptor. The debate has centered around whether a restricted T cell has two receptors that recognize virus (or other foreign molecule) and histocompatibility antigen separately, or whether it has a single receptor that recognizes them as a unit.

Now researchers can introduce specific changes into a cloned gene, transfer the altered gene into cultured cells and see how the mutation affects the ability of the antigen to perform its normal functions. It should be possible to see, for example, whether a particular mutation abolishes the antigen's ability to interact with restricted T cells.

Nevertheless, Robert Goodenow from Hood's laboratory sounded a cautionary note about the gene transfer experiments. In one experiment, he found that a transferred histocompatibility gene fragment was expressed on the cell surface in an apparently full-sized molecule. Goodenow postulates that the DNA fragment may have been incorporated into the cellular MHC and then expressed in conjunction with cellular genetic information. An event of this type, which would complicate interpretation of gene transfer experiments, is not an unlikely event because of the structural similarities of the histocompatibility genes.

In fact, one of the most striking features of the gene structures determined so far is the degree with which their nucleotide sequences resemble one another. All of the class I heavy chains have related structures, often having 70 to 80 percent of their nucleotide sequences in common, even when mouse and human chains are compared. Moreover, there are significant resemblances even among molecules of different types. Each of the three external domains of the

class I heavy chains resembles the β_2-microglobulin molecule, with the greatest similarity in the third domain with which the smaller protein interacts. This means that these structures are also related to the constant region domains of antibody heavy chains, and lends additional credence to the hypothesis, long extant among immunologists, that all of these molecules evolved from a common ancestral gene. As Hood puts it, "I think you have to say that for gene families that diverged as long ago as these did that there are striking homologies."

The structural similarities of the genes for β_2-microglobulin and the heavy chains of the immunoglobulins and the class I antigens had been expected from the results of studies of these proteins. Less predictable was the resemblance between the second external domain of the human DR chain and β_2-microglobulin and the other proteins, which was described by Korman. Unlike the case of the class I heavy chains for which a few amino acid sequences are available, largely from the work of Strominger's laboratory and those of John Coligan at the National Institute of Allergy and Infectious Diseases and Stanley Nathenson of Albert Einstein College of Medicine, little had been known about the structures of the much rarer class II proteins. As it is now far easier to clone and sequence genes than it is to isolate and sequence proteins, especially those present in such small amounts, the gene cloning will greatly facilitate analysis of the class II antigens.

Because of the redundancy of the nucleotide code for amino acids it is difficult to tell whether the resemblances result from convergent evolution, in which different genes evolve to form similar products, or from divergent evolution, in which a common ancestral gene undergoes duplication followed by changes in the resulting genes. Analysis of the gene sequences suggests that the immunoglobulins and histocompatibility antigens arose from divergent evolution.

Despite their resemblances, there is a great deal of diversity among the molecules that are encoded in the MHC. The best studied of these are, as usual, the classic transplantation antigens. Any of at least 50 and perhaps 100 different genes (alleles) may occur at the K and D sites of the mouse MHC. The corresponding sites in the human MHC (designated HLA-A, -B, and -C) may carry any of 30 or so different allelic genes.

The question of how this large number of genes originated came in for a good deal of attention in Oxford. Comparisons of both protein and gene structures show that one molecule often contains short segments of a few amino acid residues or nucleotides from another molecule. The findings, Seidman says, "suggest a mechanism for generating polymorphism in H-2 [mouse MHC] genes. One gene may exchange sequences with another."

There are a number of ways in which this might come about and there is currently insufficient evidence to unequivocally rule out or confirm any of them, but much of the Oxford discussions centered around the possibility that gene conversion might be involved. In gene conversion, which is better known as a way of conserving multiple copies of similar genes than as a way of generating diversity, all or part of a gene is converted into another.

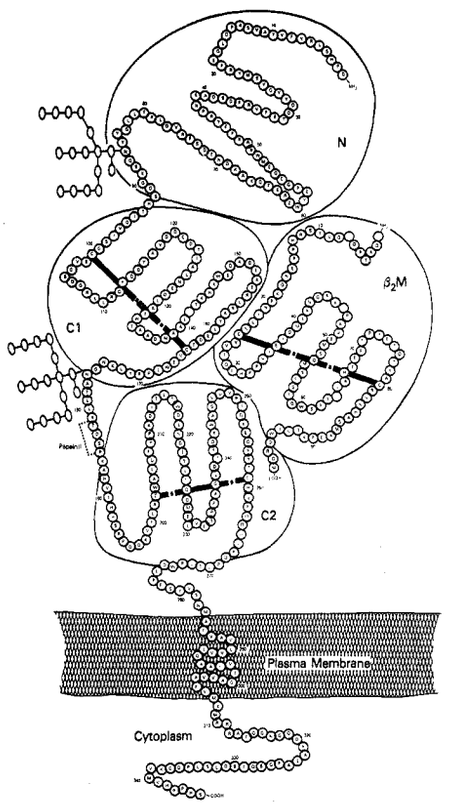
Bodmer favors gene conversion among the histocompatibility antigen genes for two reasons, "It is the only way you can explain these islands of diversity in seas of homogeneity. And you have to ask not only how new variants are produced but how they are propagated in the population. Gene conversion is the only way I can think of that can propagate a mutation in the absence of selective pressures." In yeast, the organism for which there is the best evidence for the existence of gene conversion, one change is often found to be preferred, for unknown reasons, to others. In that case the frequency of the particular altered gene in the population cannot help but increase.

One possible way of generating diversity in the class I heavy chains for which there is currently little evidence is the type of rearrangement seen in antibody genes during the development of antibody-producing cells. Because investigators find the same arrangement patterns of the class I genes in sperm or fetal DNA as in the DNA of mature cells, such gene rearrangements do not appear to be involved here.

Another conclusion emerging from the gene studies is that the number of genes in the MHC is much greater than might have been expected. Classic transplantation antigens have been mapped to four genetic loci in the mouse MHC and to three in the human. But investigators find from 15 to 35 class I heavy chain genes per haploid genome in the mouse MHC. Often these genes are located outside the classic transplantation loci, mapping in such regions as the T1a and Qa loci. These regions code not for transplantation antigens, which are expressed on all cells, but for differentiation antigens that are found only on certain types of immune cells. The differentiation antigens also contain a β_2-microglobulin subunit and are considered class I antigens, however. As Nathenson points out, "We now have to think in terms of different families of class I molecules." The products of many of the genes have not yet been identified, however, and even where they have been there is little evidence regarding what function they may have. Transfer experiments with cloned genes may help to resolve these issues.

The final business of the participants in the Oxford meeting was to begin plans for their next gathering. This is to be held in the spring of 1983, although there were a few suggestions, not altogether facetious, that the rapid progress justified holding the next conclave in a month or so. It seems safe to conclude that after another year's work investigators will have the class II regulatory genes in hand, too, and the MHC will be well on the way to giving up its mysteries.

—JEAN L. MARX



A class I antigen

*The structure shows a schematic representation of a transplantation antigen of the mouse, which consists of two protein chains. The larger chain is divided into three external domains (*N, C1, *and* C2*), a transmembrane region, and a cytoplasmic region. Carbohydrate moieties (trident-like structures) are attached at two points. The large chain is associated with a smaller protein subunit called β_2-microglobulin (β_2M). Human transplantation antigens have similar structures. [Source: Reprinted with permission from J. E. Coligan, T. J. Kindt, H. Uehara, J. Martinko, S. G. Nathenson,* Nature (London) ***291****, 35 (1981)]*