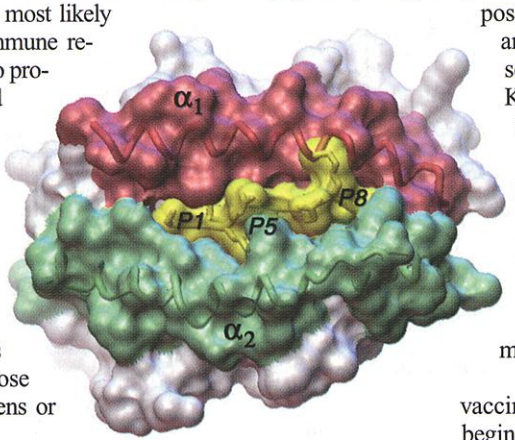
Computers Aid Vaccine Design

Immunologists are using computers to help them identify the antigen fragments that trigger immune responses and might thus provide vaccines for diseases ranging from malaria to cancer

Immunology may seem like the archetypal "wet" science, but, like most areas of research these days, it is rapidly going silicon. Immunologists are using computers in pursuit of one of the holy grails of their field: how to predict which snippets of a protein, out of hundreds or sometimes even thousands of possible candidates, are most likely to spark a strong immune response. The goal is to produce a vaccine based on tiny fragments of an antigen, such as a piece of protein from an invading pathogen or a cancer cell. Such "subunit vaccines," immunologists hope, will be less dangerous than those using whole pathogens or cancer cells.



Tight fit. The image shows a peptide fragment from an antigen (yellow) nestled in the cleft of an MHC protein.

The strategy is based on years of evidence showing that the immune system chops foreign proteins into small peptides, each containing about 10 amino acids. Some of these peptides, or epitopes, are then displayed on "antigen-presenting cells," where they are held in place by proteins of the major histocompatibility complex (MHC). These MHC proteins with their antigen fragments act like red flags, drawing the attention of the immune system's T cells, which either kill cells carrying the antigens outright or orchestrate an attack by various other immune players.

Only a few peptide fragments have the right shape to fit into the MHC proteins, however, and the challenge is to figure out which ones will lock into place. Finding these epitopes, says immunologist Vladimir Brusic of Kent Ridge Digital Labs in Singapore, is "one of the bottlenecks in vaccine research." To complicate matters even further, MHC proteins are highly diverse, with hundreds of slightly different forms that can vary in their peptide-binding preferences. And because an individual inherits only one variant from each parent, a vaccine based on just one peptide may not work for everyone. Immunologists have now enlisted computers to help them home in on the most promising pieces in this complex three-dimensional (3D) puzzle.

The idea is to develop computer algorithms that use information immunologists have already gathered about how peptides fit into an MHC protein to predict which epitopes in an untested protein will bind. The best candidates would then be tested in cell cultures or animals. "The alternative, a brute-force approach where you test every single peptide [in a given protein], is obviously not possible for financial and logistical reasons," says James Kazura, a malaria immunologist at Case Western Reserve University School of Medicine in Cleveland who is working with Brusic to find peptides for a badly needed malaria vaccine.

More and more vaccine researchers are beginning to apply these predictive computer tools to their favorite diseases. They've already used them to identify peptides that trigger immune responses to the malaria parasite and to various types of cancer cells, as well as those that spark allergic reactions. And some of these peptides are beginning to move into clinical trials as antimalaria or anticancer vaccines.

What makes for a good epitope?

Computational approaches to epitope prediction vary, but they all have their roots in the late 1980s, when researchers got the first 3D structures of MHC proteins. These revealed that the molecules have clefts on their surfaces that hold the antigenic peptides—a discovery that opened the door to uncovering the rules of engagement between MHC proteins and peptides. Although "the early pioneers got the details wrong," Brusic says, "their observations spurred a lot of interest in the field."

Over the following years, Hans-Georg Rammensee of the University of Tübingen in Germany compiled a huge database of the sequences of natural epitopes that he washed off different MHC class I proteins, which trigger the killer T cells of the immune system. Alessandro Sette, now at Epimmune, a biotech start-up in San Diego, California, did the same for epitopes for MHC class II proteins, which activate T helper cells, the master coordinators of the immune system.

These databases provided the fodder for the first computer models to predict epitopes. Based on the sequences of natural epitopes, Rammensee, Sette, and others in their wake extracted so-called binding motifs for individual MHC molecules. Akin to a signature, such a motif consists of two or three key amino acids within a peptide that are thought to be essential for binding to a particular MHC molecule. But although these motif-based approaches pick up the typical mainstream binders for each MHC molecule, Brusic says they may miss oddball peptides that don't conform to the consensus motifs.

To get around this potential shortcoming, Brusic and his colleagues have developed a computer algorithm based on artificial neural networks, an assembly of interconnected computer units akin to a simplified version of the human brain. When fed the sequences of numerous peptides known to bind to a given MHC protein, the neural net "learns" the features a peptide must have to be a good binder. Brusic then runs the sequence of an untested protein through the network, and the computer algorithm picks out the peptides that are likely to be bound by that particular MHC protein. Because neural networks do not simply recognize binding motifs, they "are particularly good at finding atypical epitopes," says Brusic.

By comparing his predictions for several proteins with experimental binding data for all the possible peptide fragments they contain, Brusic found that his neural nets can correctly predict up to 80% of the binding peptides. What's more, he adds, "they can improve as more data become available." A big disadvantage, though, is that training requires binding data from hundreds of peptides for each MHC variant, and the nets can't immediately identify a peptide that will bind to many different MHC variants—a must for a widely applicable vaccine.

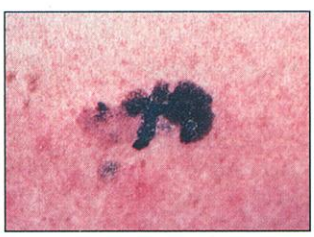
TEPITOPE, an algorithm developed by Jürgen Hammer of Hoffmann–La Roche Inc. in Nutley, New Jersey, and his colleagues, does have that capability. The binding cleft of an MHC class II protein contains nine so-called "pockets," minute indentations at the bottom of the cleft, each of which holds in place one of the binding peptide's amino acids. Comparing the sequences of different MHC variants, Hammer realized that their clefts contain only a limited number of these pockets in various combinations. In a painstaking series of some 10,000 binding experiments, published in the June 1999 issue of *Nature Biotechnology*, the Hammer

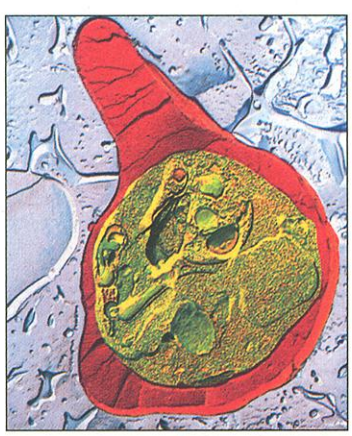
CREDIT: K. C. GARCIA *ET AL.*, *SCIENCE* **274**, 209 (1996)

team determined the affinity of 35 of these binding pockets for each of the 20 naturally occurring amino acids.

The researchers then use the TEPITOPE algorithm to calculate the binding strengths, or matrices, of all the peptides in a protein antigen for the 51 most common MHC variants consisting of various combinations of the 35 pockets tested. Those peptides calculated to have the greatest binding strengths would be presumed to be the immune-stimulating epitopes. This matrix-based method also allows the identification of so-called "promiscuous" peptides that bind to several different MHC variants and are therefore the most promising candidates for a vaccine.

The work of Darren Flower, a bioinformaticist at the Edward Jenner Institute for Vaccine Research in Compton, U.K., exemplifies a fourth type of approach to epitope examination, based on computer modeling of the 3D structures of peptide-MHC pairs. Because MHC molecules all have a very similar 3D architecture, his group, established just this summer, tries to model previously uncharacterized MHC structures and fit epitopes into their clefts. Flower says he hopes "to generate new information this way, especially for the large number of MHC alleles that haven't been properly characterized [by binding assays]."





Vaccine targets. Computer analysis has identified peptides from the skin cancer melanoma (*upper left*), ryegrass (*above*), and the malaria parasite (*left*, in a red blood cell) that might be used in vaccines.

Putting prediction to the test

Although this new, structure-based approach hasn't been tested yet, other prediction tools have already begun to prove their mettle. For example, researchers have used them to identify peptides that may be useful as cancer vaccines. Some of this evidence comes from cancer immunologist Walter Storkus of the University of Pittsburgh School of Medicine, who collaborates with Brusic. Storkus is focusing on several protein antigens that seem to distinguish melanoma cells from normal cells. Using Brusic's neural net, Storkus and his colleagues identified some 40 candidate epitopes from five different tumor antigens.

Not all of them were hits. Out of 36 peptides tested, only 11 activated T cells in lab culture or triggered a hypersensitivity reaction in skin-prick tests. His team has just begun a clinical trial with 20 metastatic melanoma patients in whom previous therapies had failed to see whether various combinations of the epitopes that tested positive can lead to tumor regression by boosting the patients' tumor-specific immune responses.

Similarly, Maria Pia Protti of the Scientific Institute H. San Raffaele in Milan, Italy, has applied Hammer's TEPITOPE algorithm to analyze a protein called MAGE-3, which is found on melanoma cells and also on lung and bladder cancer cells. It helped her identify 11 epitopes, all of which turned out to bind to all major MHC alleles. What's more, nine of the peptides also kicked off a strong T cell response when tested in lab cultures. Protti plans to start clinical trials in melanoma patients with some of her predicted peptides early next year.

It may even be possible to marry epitope prediction to one of the hottest technologies around: DNA microarrays—chips or slides containing DNAs from thousands of gene snippets that are used as probes to determine what genes are expressed in cells (see p. 82). Hammer, working with Protti and cancer immunologist Ugur Sahin of the University of Mainz in Germany, has used such a DNA microarray, harboring almost 20,000 human genes, to compare the gene expression patterns of tumor samples from 20 colon cancer patients with those of healthy colon tissue. The team found 34 genes that were highly active in at least half the patients, indicating that their protein products might be good vaccine targets.

However, the proteins made by those genes contain 19,000 overlapping peptide fragments—far too many to analyze to see whether they can evoke an immune response. But in less than a day TEPITOPE slashed that unworkable number to 130 candidate epitopes, all predicted to bind to many different MHC alleles. These can now be tested to see whether they evoke an immune reaction.

Computational immunologists have other diseases in their sights in addition to cancer. One prime target is malaria, for which researchers so far have been unable to come up with a vaccine, despite years of searching. In the July *Journal of Immunology*, Sette, in collaboration with Stephen Hoffman of the Naval Medical Research Center in Silver Spring, Maryland, reported that Epimmune's proprietary Epitope Identification System, a matrix-based algorithm similar to Hammer's, had turned up 11 epitopes from five proteins of the malaria parasite *Plasmodium falciparum*. All 11 triggered T cell responses in malaria-exposed individuals from Indonesia, Kenya, and the United States, three populations that differ greatly in their MHC proteins. "We were aiming at as broad a population coverage as possible, because for a vaccine to be acceptable you need to provide answers to global problems," Sette says.

In as-yet-unpublished work, Case Western's Kazura, using Brusic's neural networks, has collected a similar assembly of malaria epitopes, but geared toward the predominant MHC allele in malaria-infested Papua New Guinea. To find out which peptides actually confer resistance to *P. falciparum*, he wants to rid 200 to 300 individuals of the parasite, determine which epitopes their T cells recognize, and record how long it takes until each individual contracts the disease again. "The theory is that if you have a strong T cell response against peptide X and it takes a long time before you're reinfected, then peptide X is a really good candidate for conferring protective immunity," Kazura says.

Allergies and autoimmune diseases, in which the immune system attacks the body's own tissues, are coming in for their share of attention as well, although here researchers want to find peptide vaccines that damp down immune responses instead of beefing them up, a strategy that seems to work in animals with experimental autoimmune diseases. The computer models are now helping immunologists identify peptides that might be useful. Initial studies over the last year or so have already picked up immunogenic epitopes triggering pollen allergies, diabetes, and Lyme arthritis. "The clinical follow-up would actually be very simple. One just has to compare T cell responses to the natural allergens with and without epitope [vaccination]," says Paola Panina-Bordignon, an immunologist at Roche Milano Ricerche,

CREDITS: (CLOCKWISE FROM TOP) K. GREER/VISUALS UNLIMITED; S. DALTON/OSF/VISUALS UNLIMITED; T. BRAIN/PHOTO RESEARCHERS

who used Hammer's binding matrices to get at the allergy-causing epitopes in ryegrass.

With more and more fields beginning to use the various algorithms, Brusic predicts that "in a few years' time this will be a standard methodology for epitope [identification]." For Sahin and his fellow users, the tools have already become standard. "They're indispensable," he says.

There's still plenty of room for improvement, though. "On average only about 20% of all predicted epitopes are actually recognized by cytotoxic T cells," says Hansjörg Schild, a colleague of Rammensee's at Tübingen. One possible reason is that the unrecognized peptides are not produced by the cell's protein-cleaving machinery. Schild and Rammensee have recently developed another computer program that might help—an algorithm for predicting how a protein will be cleaved. Combining that with epitope prediction should further reduce the number of epitopes that need testing, says Schild.

Immunologists caution that, despite the number of promising preclinical studies, the algorithms still have to prove their worth on the real-life battlefield between pathogen and host, by pointing the way to new vaccines. Says vaccine researcher Anne DeGroot of Brown University in Providence, Rhode Island, who has herself developed a prediction algorithm based on MHC binding motifs, "You can do a lot of neat things on your computer, but ultimately you have to put [the predicted peptides] in vaccines and test whether they protect [from infection]." With clinical trials picking up speed on several fronts, the first answers are likely to come in before too long.

–MICHAEL HAGMANN

Michael Hagmann is a writer based in Zürich, Switzerland.

NEWS

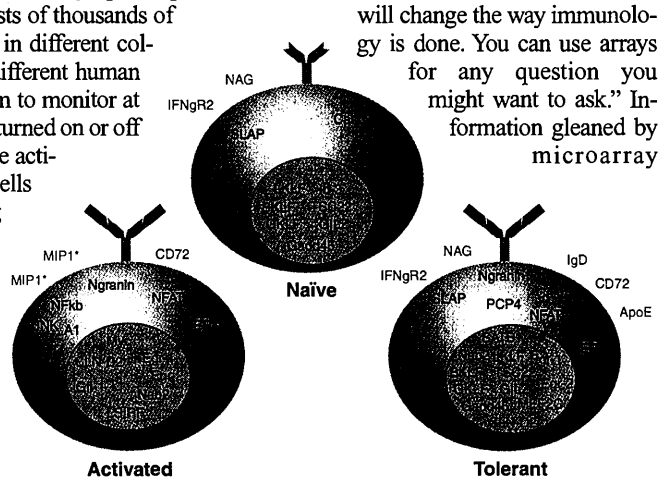
Doing Immunology On a Chip

Using microarrays to measure gene expression patterns may reveal the mysteries of normal immune cells and also of diseases in which immune cells go astray, such as autoimmunity and cancer

For Louis Staudt, an immunologist at the National Cancer Institute (NCI) in Bethesda, Maryland, the key to the enigmas of the human immune system lies buried in what looks like a miniature painting by Piet Mondrian, the Dutch abstractionist who elevated colored rectangles to high art. Dubbed Lymphochip, Staudt's work of art consists of thousands of tiny squares that light up in different colors, each representing a different human gene. This chip allows him to monitor at one time which genes are turned on or off when, say, killer T cells are activated by pathogens, or B cells turn into life-threatening leukemias.

Used by a small but growing number of immunologists, the Lymphochip and similar DNA microarrays mark the entry of immunology into the presumably golden age of genomics, a quantum leap that promises a new level of understanding of the genetic programs underlying immune responses, both normal and abnormal. "This approach allows you to look at the whole system at once instead of looking at one gene [at a time]. It's a 'Let the system tell you what's going on' approach," Staudt says.

Immunologists are already intrigued by what they are seeing. Microarray analysis is helping define the changes in gene activities that cause certain leukemias as well as those that govern the immune system's response to a range of pathogens. Microarrays are also illuminating the developmental paths leading from immature precursor cells to a plethora of immune warriors such as B and T cells. Indeed, says Lars Rogge, an immunologist at Roche Milano Ricerche in Milan, Italy, "this will change the way immunology is done. You can use arrays for any question you might want to ask." Information gleaned by microarray analysis should also help to both diagnose diseases and find new therapies.



Turned on. Microarrays reveal that lymphocyte activation has as much to do with down-regulating inhibitory genes (red) as turning up genes with activating functions (green). The genes' positions indicate the locations of their protein products.

Take Staudt's work. He and his colleagues have been focusing on identifying the gene changes that distinguish various malignancies of immune cells, or lymphocytes, from one another and from healthy cells. With this in mind, the NCI team, in collaboration with Stanford's Pat Brown, a founder of DNA chip technology, began developing the Lymphochip in 1998. The researchers collected 15,000 genes that are highly active in immune cells at various stages of development as well as in some leukemias. They added another 3500 genes known to be important in lymphocyte or cancer biology and started profiling a variety of B and T cell tumors.

The initial results have been surprising. Staudt and his colleagues reported in the 3 February issue of *Nature* that diffuse large B cell lymphoma is not one disease, but rather "two separate diseases with two distinct profiles hiding within one clinical category," says Staudt. What's more, patients with the disease vary in their responses to therapy, and the gene activity profiles correlate with that finding (*Science*, 8 September, p. 1670). The hope now is that clinicians can use such gene profiling to select the most appropriate treatment options for their patients and perhaps eventually to design better therapies.

And what works for immune cell cancers should work for other immune disorders as well, Staudt predicts. "It's fairly obvious that autoimmune diseases or immune deficiencies are pretty heterogeneous, and one can easily surmise that there are different subtypes. For example, only a subset of multiple sclerosis patients responds to interferon β treatment. I think we should look at their gene expression profiles" to see if they determine the patients' responses.

Infectious disease expert David Relman of Stanford University is pursuing similar goals. "The idea is to recognize specific [hard-to-diagnose] infections based on the gene expression profile in cells from the host," he says. That could be advantageous, because "the clinical sample doesn't even have to contain the bug you're after," says Relman. "With current diagnostics, if you're looking at the wrong time or place, you may end up empty-handed." A glimpse at the host response might also indicate the stage of infection, which could influence therapy.

Relman says his ultimate goal is to understand more precisely how microorganisms cause disease. In as-yet-unpublished work, he

ILLUSTRATION BY C. CAIN, ADAPTED FROM C. GOODNOW