RESEARCH NEWS AND VIEWS

In silico predictions; in vivo veritas

Anne S. De Groot and Frank G. Rothman

In this issue, Sturniolo et al.¹ take an important step toward the new era of applied bioinformatics. They report a major advance in the use of computers that will contribute to the development of vaccines by predicting which amino acid sequences on proteins are likely to be effective immunogens.

Some (and we count ourselves among them) would argue that the marriage of computers and biology and the creation of a new field—bioinformatics—will bear abundant fruit in the years to come. Why so? Because computers are extremely adept at finding patterns in masses of data. New rapid-sequencing and gene chip technologies are providing that mass of data, and, fortunately, “nature” provides the patterns. The task we have before us in the early years of the next millennium is to prove how accurately computers can detect natural patterns and put this new knowledge to use in curing human disease.

Building on well-documented experimental work by their group² and others^{3–6}, Juergen Hammer, Francesco Sinigaglia, and their colleagues describe a new matrix-based prediction algorithm for peptides that bind to major histocompatibility complex (MHC) class II molecules, which in humans are designated HLAs. And, in a more speculative departure, they have combined their new matrix-based prediction algorithm directly to information derived from DNA chip based sequencing of expressed genes, forging a link between two of the most powerful new tools of biotechnology. The end product of this work is a set of peptides, which the authors believe to be unique to colon cancer and highly likely to be immunogenic. They propose that these peptides form the basis for a new colon cancer vaccine.

In fact, this method may overcome a major stumbling block for vaccine development—namely, that not all components of tumors and infectious agents elicit immune responses equally well. Each individual's ability to respond to antigenic challenges is genetically restricted by the involvement of the

MHC, a family of cell surface glycoproteins⁷. T cells only recognize antigen when it is displayed on a cell surface bound to MHC molecule. The human MHC genes comprise six highly polymorphic genetic loci; individuals therefore differ in the repertoire of their MHC molecules. Three MHC loci are devoted to presentation of class I molecules (which usually interact with peptides from self or from viruses and the T-cell receptor), and three are devoted to presentation of class II molecules (usually derived from bacterial

pathogens). Class II molecules are involved in presenting antigen peptides to T-helper cells and are therefore involved in both B-cell and T-cell immune responses.

The entire antigen is not displayed on the cell surface. Instead, protein antigens are degraded into short peptides, 8–18 amino acids in length, and the cell selects a peptide to display based on its ability to bind to the host's MHC. Thus, the ability of a protein antigen to elicit an immune response will be modulated by the degree to which peptide degradation

Figure 1. For a peptide epitope to elicit a T-cell response, it must bind to an HLA class I or II molecule on an antigen-presenting cell (APC) and interact with a T cell via the T cell receptor. Computer algorithms, such as TEPITOPE, are now available to assist the researcher to rapidly search and identify the subset of peptides that bind to HLA and stimulate T cells. Protein sequences (e.g., identified in pathogens or cancer cells via DNA chip methodology) are input into the computer, which parses them into overlapping peptides. These are then screened for matches to a predetermined pattern thought to be representative of peptides that bind to one or several HLA molecules. The next step is to test in vitro the candidate T-cell epitopes for binding to the HLA molecule of interest (in which case the peptides are termed “ligands”) and/or for stimulation of T cells (“epitopes”). Good candidates can then be used in vaccine development.

Anne S. De Groot is assistant professor of medicine and director, TB/HIV Research Lab (Anne_DeGroot@Brown.edu) (http://www.Brown.edu/Research/TB-HIV_Lab/); Frank G. Rothman is professor, department of molecular biology, cell biology and biochemistry, Brown University, Providence, RI 02912.

ANALYSIS

products derived from it will be bound to MHC molecules of each particular host. Since such binding is a necessary, though not sufficient, condition for a good vaccine epitope, knowledge of which peptides are bound to a variety of MHC molecules is important for the design of synthetic vaccines.

Rotzschke and Falk⁸ showed that the bound peptides have patterns, or motifs, that are unique to each MHC molecule. Understanding of the MHC-peptide binding was further enhanced by the crystallographic work of Strominger and Wiley⁹ showing the structure of the peptide-binding groove of one class II MHC molecule. For researchers who came of age in the era of bioinformatics, the publication of "MHC binding motifs" and the visual image of a peptide binding tightly in the MHC-binding groove were seeds cast on fertile ground. It was not long before a number of laboratories were writing computer-driven algorithms for finding the patterns described as MHC motifs in proteins, and proposing that these tools for pattern recognition could be used to design vaccines.

What Sturniolo et al. have now done is to greatly expand earlier contributions to the field of T-cell epitope prediction by developing a new matrix-based prediction algorithm for the MHC class II molecules. The paper takes an innovative leap that has been much discussed but not, until now, attempted: it links the profiles of peptides bound to selected pockets within the binding grooves of various MHC class II molecules to the amino acid residues that reside there, and uses this technique to generate a much broader range of predictive matrices from a limited set of binding profiles. Their algorithm can be applied to analyze the immunogenic potential of any protein whose sequence is known. Although the method is limited by the approximation that binding of a peptide to each pocket along the groove is independent, this approximation is supported by extensive experimental evidence.

In addition to creating their well-supported prediction algorithm, Sturniolo et al. are proposing to employ the predictive algorithm on a genome-wide level by linking it to DNA chip based technology. They illustrate the potential of this linkage with a study of proteins specifically expressed or upregulated in tumor cells. Although the data presented indicate that the algorithm will succeed in greatly reducing the number of candidate peptides, it is premature to predict that any of them will be successful T-cell epitopes for two reasons: first, the specificity of expression of the proteins under investigation may not be absolute; and second, class II-restricted responses to cancer antigens are considered by most immunologists to be less important than class I-restricted ones.

Nevertheless, in proposing that the new

computer algorithms be applied directly to information derived from DNA chip based sequencing of cancer genes, the authors have linked the new discipline of bioinformatics to the older ones of molecular biology and immunology. While new applications of these algorithms to vaccine research need to be moved from in silico science to experiments in vitro and in vivo, that advance can be expected in the first few years of the new millennium.

Delivering tissue regeneration

W. Mark Saltzman

Gene therapy and tissue engineering are relatively new concepts in the treatment of human disease. Gene therapy uses DNA as a superpharmaceutical, an agent with the potential to alter cell function for an extended period of time in relation to more established therapeutic agents. Tissue engineering aims to regrow tissue structure lost as a result of trauma or disease through the application of engineered materials. Both fields can point to initial successes, such as the replacement of adenosine deaminase genes in children with severe immunodeficiency and the use of synthetic materials to accelerate healing of burns and skin ulcers. Nevertheless, major limitations have been encountered: transient, low-level expression limits gene therapy of chronic diseases and tissue engineers have not yet learned how to reproduce complex architecture, such as vascular networks, which are essential for normal tissue function. In this issue, Shea and co-workers¹ demonstrate a new strategy that should encourage collaboration between these two emerging fields in the efforts to overcome these problems.

Recent reports have described DNA delivery in situ from polymer coatings³, microspheres^{4,5}, and synthetic matrices⁴. DNA-material hybrid systems promise to enable forms of gene therapy not possible with other gene delivery systems. For example, polymeric microspheres facilitate the expression of orally administered genes⁶. In some ways, these new matrix materials are a logical extension of earlier work demonstrating that certain materials, such as polymers⁷ and lipids⁸, form complexes with DNA to facilitate cellular uptake (perhaps by mimicking physical and chemical attributes of

1. Sturniolo, T. et al. *Nat. Biotechnol.* **17**, 555–561 (1999).
2. Hammer, J. et al. *J. Exp. Med.* **180**, 2352–2358 (1994).
3. Davenport, M.P. et al. *Immunogenetics* **42**, 392–397 (1995).
4. Fleckenstein, B. et al. *Eur. J. Biochem.* **240**, 71–77 (1996).
5. Meister, G.E. et al. *Vaccine* **13**, 581–591 (1995).
6. Schafer, J.R. et al. *Vaccine* **16**, 1880–1884 (1998).
7. Zinkernagel, R.M. & Doherty, P.C. *Nature* **248**, 701–702 (1974).
8. Falk, K. et al. *Nature* **351**, 290–296 (1991).
9. Stern, L.J. et al. *Nature* **368**, 215–221 (1994).

viruses). Modern polymer science permits synthesis of new materials that complex with DNA and are quickly degraded⁹, perhaps facilitating DNA uptake with reduced cytotoxicity.

To the general category of synthetic gene transfer agents, Shea et al. add solid, porous materials that slowly release DNA in situ, or directly into a localized region of the body. The report provides an important demon-

Figure 1. Duration of biological effect after administration of low-molecular-weight drugs (red), proteins (green), or genes (blue). The expected duration of the biological effect is shown for bolus delivery (top) and sustained delivery (bottom): the orange box indicates the period of delivery. The biological effects of proteins are often less dramatic than expected because of rapid clearance of proteins from the body; therefore, sustained delivery is an important and accepted strategy for enhancing protein therapeutics. Controlled DNA delivery systems will enable the better definition of the advantages of slow, continuous gene delivery.

W. Mark Saltzman is professor at the School of Chemical Engineering, Cornell University, Ithaca, NY 14850 (saltzman@cheme.cornell.edu).