Science

AAAS

Reverse Vaccinology and Genomics
Rino Rappuoli, *et al.*
Science **302**, 602 (2003);
DOI: 10.1126/science.1092329

Reverse Vaccinology and Genomics

Rino Rappuoli and Antonello Covacci

The genomic revolution has had a dramatic effect on our ability to find new vaccine targets and develop effective vaccines.

Ever since Jenner successfully used a cowpox virus to vaccinate against human smallpox in 1796, biologists have focused on vaccination as the best defense against numerous bacterial and viral pathogens. In vitro–grown pathogens have been used to develop killed, live attenuated, or subunit vaccines (vaccines containing whole killed microorganisms, live microorganisms that have lost the ability to cause disease, or purified components of microorganisms, respectively). These vaccines are among the most important medical interventions ever developed and are powerful tools against biological weapons. However, in spite of the impressive results obtained with these conventional approaches, not all pathogens have been successfully grown in vitro. This endeavor has been brought to a new level with the advent of genomic sequencing and the wealth of information about vaccine targets that it provides.

Hepatitis B and C viruses are examples of pathogens that do not grow in vitro and cannot be approached through conventional vaccinology. The availability of their small genomes made it possible for researchers to identify genes coding for the viral envelope proteins and to develop recombinant vaccines that are now used for the universal immunization of children (hepatitis B) or are at an initial stage of clinical testing (hepatitis C) (*1*, *2*). Group B meningococcus is an example where several decades of conventional vaccine development had been unsuccessful (even though this bacterium could be grown in the lab), because the components identified by conventional approaches were identical to self-antigens or were hypervariable in sequence.

The availability of the complete genome sequence of a free-living organism (*Haemophilus influenzae*) in 1995 (*3*) marked the beginning of a "genomic era" that opened the eyes of vaccine biologists to a new approach to vaccine design for the treatment of bacterial infections. This "reverse vaccinology" was not based on growing microorganisms but on running algorithms to mine the information contained in the blueprint of the bacterium (*4*). Within 18 months of the beginning of the sequencing of meningococcus B, over 600 potential vaccine candidates had been predicted by computer analysis of the genome, and 350 of them were expressed in *Escherichia coli*, purified, and used to immunize mice (*5*, *6*). Many novel antigens with properties that could overcome the limits of previous vaccine candidates were discovered and are now being tested in clinical trials. Today, the genome-based approach is routine in vaccine development and is being applied to streptococci, Chlamydiae, staphylococci, *Plasmodium falciparum*, and bioterrorism-associated agents such as *Yersinia pestis*. In most cases, the new technology has identified treasure troves of novel vaccine candidates.

The recent emerging disease SARS is a perfect example of the speed with which genomic information can have an impact on public health. In less than a month from the first suggestion that a coronavirus might have been implicated in the disease, the nucleotide sequence of the virus was available (*7*, *8*) and provided instant answers to a number of pressing questions. It was clear that the agent was a natural (and not a laboratory-fabricated) coronavirus, diagnostic tests were set up, and vaccine targets were identified. Today, some of these vaccines are already being tested in animal models. None of this would have been possible without the public release of the genome sequence.

The 140 sequenced bacterial genomes and 1600 sequenced viral genomes, comprising potentially over 400,000 encoded proteins, already exceed by 10-fold the complexity of the human genome, which less than 3 years ago was seen as a major challenge for biocomputing. The analysis of single genomes is no longer satisfactory; comparisons of multiple genomes to provide insights into conserved or unique families of proteins or functional domains are needed to continuously improve the precision of annotation and to identify the basic building blocks of proteins, trace the evolution of virulence mechanisms, potentially reconstruct complex structures, and identify and design novel immunogens.

These increasing needs are helping to drive the beginning of the next phase of reverse vaccinology. It will take advantage of the new computing infrastructure proposed by Ian Foster to solve problems of large-scale computation by connecting independent supercomputing centers, which is already being implemented by several institutions, including Argonne National Laboratory and CERN, and is spreading worldwide (*9*, *10*). The system will be based on a grid of supercomputers connecting major scientific institutions, with decentralized databases containing a repository of nucleotide and protein sequences, three-dimensional structures, expression profiles, immunological properties, and functional data.

Today, a scientist working in an advanced research institution, with a cluster of Unix servers and workstations, needs 48 hours to compare one genome against all other available genomes and 2 weeks to compare all available genomes against all others. The time for the analysis could be reduced to 16 minutes and 40 hours, respectively, by a 10-node grid architecture and to 30 seconds and 6 hours, respectively, with a 100-node grid. Finally, all these operations will be performed in real time, when the grid system will be ubiquitous and available to any scientist able to formulate fundamental questions using delocalized databases and computing power.

Scientists who are not making use of the widening universe of genomic information available in databases are wasting some of the power of today's science. There have been some discussions lately about the appropriate balance between access to genomic data and global security (*11*). The potential of genomic information is enormous for combating microbial agents (both through vaccines and through antimicrobials) that could be used as weapons of bioterrorism. In our view, the awareness that we have the technology to develop vaccines that will render any biological weapon inoffensive is a strong deterrent for bioterrorism. Conversely, restrictions on the sharing of genomic information would represent a recognition of weakness and only serve to encourage the development of biological weapons.

References

1. For hepatitis B, see www.cdc.gov/mmwr/preview/mmwrhtml/mm5236a5.htm.
2. R. Rappuoli, unpublished data.
3. R. D. Fleischmann *et al.*, *Science* **269**, 496 (1995).
4. R. Rappuoli, *Curr. Opin. Microbiol.* **3**, 445 (2000).
5. H. Tettelin *et al.*, *Science* **287**, 1809 (2000).
6. M. Pizza *et al.*, *Science* **287**, 1816 (2000).
7. P. A. Rota *et al.*, *Science* **300**, 1394 (2003).
8. M. Marra *et al.*, *Science* **300**, 1399 (2003).
9. I. Foster, C. Kesselman, in *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster, C. Kesselman, Eds. (Morgan Kaufman, San Francisco, 1998), pp. 15–52.
10. See http://eu-datagrid.web.cern.ch/eu-datagrid/.
11. *Biotechnology Research in an Age of Terrorism: Confronting the "Dual Use" Dilemma*, in press (available at www4.nationalacademies.org/news.nsf/isbn/0309089778?OpenDocument).

IRIS, Chiron Vaccines, 53100 Siena, Italy.

Reverse Vaccinology and Cellular Immunity. As discussed above, reverse vaccinology relies on the combined use of immunological and genomic information to identify relevant protein antigens for diagnostic or vaccine purposes. In this context, the identification of the epitopes recognized by CD4+ T cell or CD8+ T cells can be utilized in "reverse" as a tool to identify new antigens. Reverse vaccinology uses the expressed genomic sequences to find new potential vaccines. Normal vaccines are created using the pathogenic organism. The term reverse refers to the use of expressed DNA over the purified proteins from the organism itself. The reverse vaccinology approach is hard to imagine without comparative genomics and structural biology of bacterial pathogens. By comparing whole genomes of pathogens with their less-virulent variants, nonpathogenic relatives, unrelated pathogens, and even commensals, it is possible to identify either unique, disease-related genes or conserved vaccine targets. Reverse vaccinology accelerates the discovery of potential vaccine candidates (PVCs) prior to experimental validation. Current programs typically use one bacterial proteome to identify PVCs through a filtering architecture using feature prediction programs or a machine learning approach. Filtering approaches may eliminate potential antigens based on limitations in the accuracy of prediction tools used. Machine learning approaches are heavily dependent on the selection of training datasets with experimentally validated antigens (positive control) and non-protective-antigens (negative control). Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. Immunity. 2010;33(4):530–41. CAS PubMed PubMed Central Article Google Scholar. 3. Reverse Vaccinology is a new approach to design vaccines from genomic information with the help of bioinformatics. This 'top-down', computer data-based approach to vaccine design contrasts with the classical, 'bottom up' laboratory-based, hypothesis-driven analysis of microbes to identify components that could elicit protective immunity. Reverse Vaccinology was first used in the 1990's as a discovery tool to identify candidate antigens for a meningococcal vaccine. Since then, much has changed through the explosion of whole genome sequencing (WGS) data and technologies that facilitate vaccine design. These include contributions from transcriptomics, proteomics, immune profiling, structural biology, systems biology and mathematical modelling.