Reverse Vaccinology: Basics and Applications

Amol M Kanampalliwar*, Rajkumar Soni, Amandeep Girdhar and Archana Tiwari
School of Biotechnology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Airport Bypass Road, Bhopal, Madhya Pradesh, India

Abstract
Vaccinomics is a new branch of bioinformatics that deals with designing a candidate vaccine against a pathogen that can be used for production of the vaccine in less time as that of conventional vaccinology. Reverse vaccinology is a part of vaccinomics which starts with the genome of pathogen and is used for the predicting the epitope. Epitope prediction is the heart of reverse vaccinology. Reverse vaccinology was used for designing vaccines against some diseases eg. Malaria, Anthrax, Endocarditis, Meningitis etc. Some approaches against viruses have also been done by reverse vaccinology.

Keywords: Conventional vaccinology, Reverse vaccinology; Epitope prediction

Introduction
Disease prevention is the most effective approach for health and can easily be achieved by administration of the biological preparation called vaccines [1]. Development of vaccines has proved a milestone in prevention of diseases for which the cure is not available. In many countries mortality rate for various diseases like small pox, polio, measles, diphtheria etc [2] was very high but due to invention of vaccines against these diseases, it has fallen to negligible. In many developing countries vaccines have played an important role in decreasing the mortality rate due some major killer diseases. Vaccines are biological preparations that are helpful in improving the immunity of a person against a particular disease [3]. Vaccines can be prepared by various means depending on the pathogenicity of microbes.

The concept of vaccination was given by Edward Jenner in 1796 by developing vaccine against smallpox and averting the infection by isolating the materials from cow. He also introduced the term vaccine. The widespread progressive use of vaccines for many infectious pathogens has been the milestone in medical sciences. When it was found that micro organisms are the cause of infectious diseases, Louis Pasteur gave the rules of vaccinology. The rules given by Pasteur were followed by Salk and Sabin. They prepared the vaccine against polio that is killed and attenuated live polio virus as a vaccine respectively. Measles is a severely communicable disease that mainly infects the children. Rubella is another serious disorder that causes severe child birth defects. Hilleman developed vaccine against measles, mumps and rubella [4] with the help of attenuated viruses and focused light on development of vaccine against diphtheria, tetanus, N. meningitides, S. pneumonia and so on. In case of hepatitis the vaccine was prepared using inactivated viral antigen. So, development of vaccines with the use of rules of Pasteur has indeed been the most powerful tool in the history of medicine (Figure 1) [5].

The chart shows that how the vaccine can be prepared by various means and each one of them has their own importance for preventing the disease. Vaccine may be prophylactics [6] that are used for decreasing the effect of the disease occurring in the future while some may be curative.

Till 20th century vaccines were prepared by using traditional means but now a days various other techniques are available those are being used for the development of recombinant DNA vaccine, conjugated proteins and advances in novel adjuvant. There have been many innovations in the field of vaccines, the first being against hepatitis B [7] and Bordetella pertussis with the introduction of the molecular biology and genetic engineering [8].

With the start of genomic era new revolutions have been taking place in the field of vaccines [9]. The application of shotgun sequencing has been introduced in giving the whole genomic sequences of several pathogens. With the completion of the sequence of the first living organism, the genomic data was used for the preparation of the vaccines against the organism. The complete genomic sequence of an organism is the reservoir of genes encoding the proteins that can act as potential antigens that can be used as vaccine candidates. This technique of identifying the proteins that are exposed on the surface by using genome instead of the microorganism, this novel approach is known as “reverse vaccinology” [10].

The present review focuses on the idea of the reverse vaccinology.
how it is important for the preparation of vaccines against pathogens which are difficult to culture in laboratory. Until now vaccines against some diseases like AIDS, hepatitis C are not present. Reverse vaccinology can make a breakthrough for this. Is reverse vaccinology a hope for preparation of the vaccine against such pathogens or a myth?? Evidently, vaccines against some disease have been provoked due to infection and death of the individuals after the vaccination; these vaccines were prepared by using the conventional approach that contained the inactivated pathogens. So to prepare a successful vaccine reverse vaccinology can be a much better approach.

Reverse Vaccinology vs. Conventional Vaccinology

The use of genomic information with aid of computer for the preparation of vaccines without culturing microorganism is known as reverse vaccinology. The first revolution in field of vaccination is the use of genetic engineering to produce vaccines. In this approach the pathogenic components of organisms were identified by culturing the organism which can be used to explore the protein coding sequences that can be used as a potential target for vaccine preparation (Figure 2) [11].

The 2nd revolution took place in 20th century with the aid of genomic technology [4]. Now a days various technologies are available which can be helpful in determining the whole genome sequence of the organism which can be used to explore the protein coding sequences that can be used as a potential target for vaccine preparation (Figure 3) [11].

The genome sequences provide at once all protein antigens that the pathogen can express at any time. This approach contains

1. Genome sequences
2. Computer analysis
3. Prediction of epitope/ antigen
4. Candidate vaccine

But the question arises whether any of the potential antigen candidate acts to provide the properties of immunity without prior knowledge of that antigen present in sufficient amount and could it develop a potential immunity after infection or expressed in vitro. By this approach one may discover a new antigen that can work on a different pattern. High throughput screening is required for the production of feasible candidate vaccine. For achieving this, all genes of pathogens are studied that can efficiently act as candidate vaccine but there are some limitations that it can’t predict polysaccharides, lipids which are some active compounds of vaccine.

Modification in Reverse Vaccinology

Pan genomic reverse vaccinology

In this approach the genome of the different isolates of same organism is compared with each other by using computer analysis. The first pan genome approach was done against Streptococcus agalactica. (Figure 4) [12,13].

Comparative reverse vaccinology

In this approach the pathogenic and non pathogenic strains of one species are compared at their genetic level. It deals with the differences in structure of proteins of different organisms.

Role of Epitope Prediction in Reverse Vaccinology

When the conventional ways fail to develop a vaccine, then one has to follow the non conventional ways for the preparation of vaccine. Until now the genomic sequences of more than 500 pathogens including bacteria and viruses are available on NIH list. As the techniques are available for studying host-patogen interactions, whole genome study and every unique gene, the work is now focused on the development of epitope driven vaccines that are target specific.

An epitope is an antigenic determinant that plays an important role in immunity of an organism. These are present on the surface of organisms that can be detected by the antibody [14]. Reverse vaccinology deals with computational analysis of genome that can be used for the prediction of the epitopes that are surface proteins. So the epitopes play an important role in development of a candidate vaccine. The major role played in immune system is by B and T lymphocyte. B cells are important in recognizing the epitopes of the antigens that can be identified by the paratopes of antibody. In some cases, T cells play a role in cell mediated immunity as the processed antigenic peptides interact with the T cell when they are presented in context of T cell.

**Figure 2:** Pipeline for the development of the vaccines by conventional means.

**Figure 3:** Pipeline for the development of the vaccines by reverse vaccinology.

**Figure 4:** Flow chart of the process of Pan Genomic Reverse Vaccinology.
the prediction of the epitopes of T and B cell plays an important role in determination of the candidate vaccine. The epitope prediction plays an important role in designing of epitope based vaccine.

**T-Cell Epitope Mapping and Prediction**

A successful peptide-based vaccine must include immune dominant epitopes [15,16]. One of the problems facing traditional vaccines is the lack of a broad cell-mediated immune response against variable pathogens [17-19].

Humoral immunity may help prevent infection, but to date only a limited number of antibodies with neutralizing capability have been identified for viruses such as HIV. The induction of cell-mediated immune responses with a large repertoire of immune specificities has emerged as an essential characteristic for the clearance or control of hypervariable viral infections such as HCV and HIV [19-21]. T cell recognizes the antigenic peptides only when they are presented by MHC I or II, with the help of the CD4 and CD8 molecule. Given the importance of T-cell responses in controlling viral infections, the larger number of T-cell epitope mapping and prediction algorithms available today comes as no surprise [22,23].

One of the more comprehensive programs seems to be EpiMatrix from EpiVax Inc. The EpiMatrix tool set is able to predict epitopes against over 100 different MHC class I and class II alleles. In a typical EpiMatrix analysis the target protein sequence is broken down into overlapping 9-mer frames in which each frame overlaps the last by eight amino acids. Each of the derived 9-mer frames is then screened for predicted affinity against a panel of MHC class I and/or class II alleles. The resulting scores fall on a common scale that can be directly compared across HLA alleles. The ability to rate putative epitopes on a common scale is described as an exclusive feature of the EpiMatrix system [24]. The EpiMatrix platform is also closely tied with additional computational tools such as ClustiMer (scans EpiMatrix results for T-cell epitope ‘clusters’), BlastiMer (automated BLAST search tool), OptiMatrix (involved in de-immunizing sequences), Conservatrix (involved in finding conserved epitopes) and Vaccine CAD (an in silico vaccine design algorithm) [25].

**B-Cell Epitope Mapping and Prediction**

The antigen antibody interaction plays an important role in immunity, binding takes place at antigenic determinant also known as B-cell epitopes. The B-cell epitopes are defined by a specific surface region of an antigenic protein, and may be divided into two different types of epitopes: linear epitopes and conformational epitopes [26].

The linear epitopes are short peptides while conformational epitopes composed of amino acid folded in 3-dimensional protein structure [27]. The mapping of the B cell epitopes can be done by various techniques. The focus of the scientist is only on the determination of linear B cell epitope [28]. The propensity value of amino acid plays an important role in determination of its position in B cell epitopes. It was introduced by Hopp and Woods. They utilized the Levitt hydrophobicity scale for the determination of the propensity value for each amino acid [27].

Today, several tools are available for the prediction of linear B cell epitopes such as PREDIITOPE [29], PEOPLE [30], BEPITOPE [29] and BcePred [31] with the determination of the propensity value. ABC pred [31] uses the machine learning based method for the prediction of the linear B cell epitopes. The conformational B cell epitope prediction can be done by following:

**Sequence based prediction method**

It does not require the target antigen structure to be known [27].

**Structure based prediction method**: It depends on the determination of the antigen antibody complexes using X-ray crystallography. DiscoTope is used for the determination of the conformational B cell epitope prediction. PEPITOPE uses combination of propensity value and half sphere exposure value of amino acid residues [27].

**Mimotopoe analysis based prediction method**

It combines both computational and experimental techniques for B cell epitope mapping. It determines the organization of the genuine epitopes [27].

Examples of B-cell epitope-mapping algorithms include 3DEX (3D-Epitope-Explorer) [32], CEP (conformational epitope predictor) [33] and DiscoTope [34]. 3DEX software is designed to allow the localization of linear peptide sequences within the three-dimensional structure of a protein. CEP predicts epitopes of proteins with known structures using accessibility of residues and spatial distance cutoffs to predict antigenic determinants, conformational epitopes and sequential epitopes [33,34]. DiscoTope was designed specifically for the prediction of conformational B-cell epitopes [14].

**Developments in B cell epitope prediction**

Prediction of the protective linear B cell epitopes [6], hybrid and consensus prediction of B cell epitopes [27], improved conformational B cell epitope prediction [27], critical assessment of B cell epitope prediction [27], immune epitope database and analysis resources [35].

With the help of these databases one can easily identify and predict the B cell epitope very correctly.

**Applications of Reverse Vaccinology**

**Meningitides**

The first pathogen against which the vaccine was prepared with the aid of reverse vaccinology was for serogroup B Nisseria meningitidis. The in silico approach for designing the vaccine by reverse vaccinology came into play due to high rate of mortality due to cause of meningitides by bacteria and N. meningitis. The process of preparation of vaccine against meningitides was a difficult task during the impact of disease. Meningitis is the swelling of the membranes around the spinal cord and brain showing the symptoms like fever, stiff neck and back, confusion, coma; caused by bacteria, viruses, or other microorganisms [36,37].

The classical way for the production of the vaccine against meningitis failed due to similarity of the proteins to humans and also because of the hypervariable nature of the pathogen. The whole genome of N.meningitidis was analyzed and with computer aid, the specific sequences were selected that are surface protein and can act as a vaccine candidate. The expression of protein was expressed in E.coli. The proteins which are successfully expressed in E.coli were confirmed by using ELISA, FACS etc. These proteins are outer membrane protein, surface proteins, and surface associated lipoproteins. These proteins act as potential source for development of vaccine against Men B. The analysis of many protein sequences were done to check its antigenity but only the few could act as good vaccine candidate which has bactericidal activity that can induce the protective immunity against Men B strains. The successful vaccine was then introduced in human.
volunteers under phase III trials [38]. The vaccine was also prepared with the combination of the other proteins of streptococcus.

**Listerosis**

Listerosis is infectious food borne disease caused by *Listeria monocytogenes* in animals and human. It leads to septicemia, encephalitis, pneumonia in humans. *L. monocytogenes* classified under gram positive bacillus belongs to *Firmicutes* which is named by Joseph Lister. It was recognized as the threat to food industry. The DNA vaccine for the disease is under progress which uses the MHC-I peptide chain. Reverse vaccinology approach is used for the development of the vaccine against listeriosis [39]. The whole genome sequence of the *L. monocytogenes* is available which is now used for the development of the vaccine by the reverse vaccinology approach. The signal peptides, LPXTG motifs, transmembrane helices and many surface proteins can be easily identified with the help of the various web based tools [40]. Various tools are used that have their own specific role.

**SignalP 3.0:** The tool was used for checking the position and the presence of the signal peptide cleavage positions in protein with the references of Gram positive bacteria [40].

**TMHMM:** The tool was used for determining the number of the transmembrane helices in proteins with the help of the hydrophobic amino acid [40].

**LipoP:** The tool was used for the determining the no. of the lipoproteins and it distinguishes between the lipoprotein signal peptides, other signal peptides and n terminal membrane helices in accordance with the gram positive bacteria [40].

**PSORTb:** The sub cellular localization of the proteins can be done by PSORTb, a valuable tool for the genome analysis [40].

By the above tools, one can easily identify the surface proteins which can be easily used as antigenic epitopes against which the vaccine was produced.

**Malaria**

Malaria is one of the most dangerous diseases that cause large number of deaths. The spread of malaria is due to bite of the mosquitoes that transfer the plasmodium into blood stream. The cause and spread of the disease is due to resistance of pathogens against various drugs. In 2010, near about 1.2 million people died due to malaria [41].

There are several drugs available that act primarily against *Plasmodium falciparum*. It has become resistant to these drugs, so vaccination could be the good approach for prevention of malaria. Vaccination is an efficient way for controlling spread of malaria. The reverse vaccinology approach is in progress for development of vaccine [42]. The sequences of 2 chromosomes of *Plasmodium falciparum* out of 14 are now available. The genome sequences were analyzed using computer approach and the ORFs were selected that can be translated in proteins. By studying the properties of these proteins and localization they can be used for the preparation of a successful vaccine candidate.

**Endocarditis**

It is the inflammation of the valves of the heart due to bacterial infection; characterized by scint inflammatory cells, vegetation. The causative organism is *Streptococcus mutans*. It is classified under gram positive bacteria and is a normal flora of the human oral cavity. So the reverse vaccinology approach is under progress for the preparation of a successful vaccine candidate. The evolutionary conserved sequences of the surface proteins are selected that are used for vaccine preparation [43].

**Anthrax**

* Bacillus anthracis *is the causative agent of anthrax which infects animals and humans. The organism was used as a weapon for bioterrorism. So the development of a vaccine is an excellent approach for the prevention of spread of the disease. Until 20th century many people and animals had been killed by anthrax. First vaccine was prepared against anthrax by Pasteur [44-46]. The reverse vaccinology approach was used for the formulation of a candidate vaccine. The antigenic determinants were found out by using EMBOS. The antigenic determinant with greater LCV value was used for designing a molecule for vaccine candidate. Docking was also done with MHC I molecule showing stable interaction [47].

**Conclusion**

The conventional way of vaccine development includes culturing of pathogens in laboratory but this is not possible in case of highly infectious pathogens that are hazardous to culture in laboratory. Hence a new approach has been discovered to design an efficient vaccine i.e. reverse vaccinology. The only requirement of the technique is the availability of whole genome sequence of the organism. With the advent of reverse vaccinology, the genome sequences of a large number of isolates can be screened for homology. The whole genome sequence is required for the prediction of epitopes and other surface protein; which is the important part of reverse vaccinology for the designing of a successful candidate vaccine. The examples given in this review article show that reverse vaccinology is useful for the preparation of epitope based vaccines against the most dangerous pathogens.

**References**

4. LM L: New strategies for vaccine development, SPCV 2010 2: e4
5. Ananya Manda, Vaccines History.
7. A vaccine produced by genetic engineering? Why not?


25. www.epivax.com/platform


45. Cohn DV (1996) Life and Times of Louis Pasteur”. School of Dentistry, University of Louisville.
