Computer Design of Vaccines: Approaches, Software Tools and Informational Resources

Boris N. Sobolev[1,*], Ludmila V. Olenina[2], Ekaterina F. Kolesanova[1], Vladimir V. Poroikov[1] and Alexander I. Archakov[1]

[1]*Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Pogodinskaya Street, 10; Moscow, 119121, Russia*

[2]*Institute of Gene Biology of Rus. Acad. Sci., Vavilov Street, Moscow, 117334, Russia*

Abstract: Development of computer methods in molecular biology and fast growth of microbial genomics data enabled new approach based on selecting *in silico* antigenic components to design vaccine constructs. It is expected that application of this technology will eliminate side effects of new vaccines and reduce the time consumption and financial expenses. The bioinformatics methods of sequence analysis are used to reveal the most prospective proteins or protein fragments of infectious agents as candidates for vaccine design. In these studies the specialized molecular immunology databases are widely used. The new approach ("Reverse vaccinology") could help in designing vaccines against diseases where traditional methods are not successful, e.g. when the viral genome reveals the extreme variability and permanent changes of antigenic properties that make difficulties for selection of molecular targets for medicines and candidate vaccines. A number of informational resources are already designed to collect and provide genomic data on certain microbes or viruses. The peculiarity of such resources is presentation of data, characterizing the different genomic variants of the same infectious agents. These structural data coupled with information on functional/immune features and software tools have to compose basis for constructing a new generation of vaccines against "common" and new infections such as AIDS, Hepatitis C, and SARS. The approaches published in literature, as well as the authors' original results are discussed.

Keywords: Vaccine design, immune response, genomics, bioinformatics, immunoinformatics, sequence analysis, epitope prediction, informational resources.

1. INTRODUCTION: FROM TRADITIONAL VACCINES TO ARTIFICIAL VACCINE CONSTRUCTS

Vaccination was pioneered over 200 years ago and rapidly accepted throughout the Europe. However, its protective mechanisms remained unclear. New vaccines appeared in about a century, after profound studies of infection processes and microorganisms started. The development and application of effective vaccines enabled to control or even eliminate several dangerous diseases. Six vaccines against human and animal infections have been developed from 1980 to 1987 [1]. Since then, the number of vaccine-controlled infections increased with a middle rate. Currently, about 25 human diseases are controlled by vaccination. Since the beginning of the vaccination history, vaccine design technologies were significantly improved. Progress in immunology provides understanding for more and more subtle mechanisms of the immune protection. Wide-ranging studies of biopolymers *via* exploiting informational technologies enable to determine structural patterns of immunogenic components and design artificial constructs, expected to display the desired protective effects [2].

Traditionally, vaccines are subdivided into live (attenuated microbe or virus cultures), killed (inactivated infectious agents), and subunit ones. Vaccines of the latter type consist of individual chemically purified components, which can invoke immune responses sufficient to protect from an infection and avoid undesired effects following the inoculation of the whole pathogenic agents – even if they are attenuated or killed. Meanwhile, the subunit vaccine preparation and usage is associated with certain problems [3]:

- Long-term cultivation of pathogenic bacteria, viruses or protozoa for the industrial production of immunogenic components is very expensive.
- Purification and detoxification of vaccine products is also cost consuming.
- The risk of an infectious agent leakage always remains.
- Side effects cannot be excluded completely upon subunit vaccine inoculations.
- If a virus displays a high genetic variability, it is very difficult to isolate a chemical component, able to invoke an effective immune response against all strains of the virus.

Modern molecular biological techniques enable to clone, display and isolate biological macromolecules or their fragments, which can be used as immunogenic components. Molecular constructs composed of such components form a new generation of subunit vaccines. These preparations possess the following advantages:

*Address correspondence to this author at the Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Pogodinskaya Street, 10; Moscow, 119121, Russia; Tel: +7 095 247 30 29; E-mail: boris.sobolev@ibmc.msk.ru

1573-4099/05 $50.00+.00 © 2005 Bentham Science Publishers Ltd.

- Relatively cheap and safe production technologies. If a vaccine is based on short peptides – up to 30 amino acid residues – chemical synthesis is economically reasonable [4]. In other cases, recombinant products can be produced in reasonably priced expression systems – from bacterial and viral vectors to transgenic plants [3].
- Edible vaccines based on transgenic plants. Such preparations could be used as an input into the mucosa (the natural gate of many infections) that has the perfect immune control system [1, 5]. Furthermore, this way eliminates complications owing to injections, which became very unpopular due to the HIV/AIDS threat.
- At last, a researcher can define candidate molecular regions before starting the vaccine development. Rapid increase of information on microbial and viral genomes presented as symbol strings, provides the input data for computer analysis. By using software tools hypothetical proteins encoded by the genome nucleotide sequences can be defined and potential immunogenic components in decoded amino acid sequences can be detected.

Preliminary *in silico* studies enable to facilitate the experimental work and accelerate the vaccine development. The process of successive application of different technologies - computing (genome data analysis), experimental analysis, pre-clinical and clinical trials for vaccine development is defined now by a new term "Reverse Vaccinology" [6]. Hence, an artificial vaccine design is based on the genome sequence analysis. Computational technologies are used as a tool to design new vaccine products, based on a more rigorous scientific basis. Now any researcher can select immunogenic molecules or fragments, keeping in mind the immune mechanisms, which the developing vaccine should provoke. Obviously, effective artificial vaccines should be developed when traditional vaccine technologies fail. Moreover, a new approach can be applied to develop more effective and safe vaccines than those already available. Today, the detection of a new microbe or virus includes a partial or whole genome sequencing with an obligatory computational analysis. This fact provides a high motivation to apply the "Reverse vaccinology" approach.

Genome studies can significantly accelerate vaccine development. The recent story of Severe Acute Respiratory Syndrome (SARS) is just one example of such acceleration. First cases of SARS were reported in February 2003, though a single case of a similar disease was registered in November 2002 [7]. The genome sequence of SARS virus was published already in May 2003 [8]. The sequence analysis enabled to reveal the most conservative regions in the amino acid sequence and select candidate immunogenic components for the vaccine design. The results of studies on the experimental vaccine that invoked virus-neutralizing antibodies in animals were published in December 2003 [9-12]. Such rapid progress has become possible in the post-genomic era, when researchers can use computational methods to analyze quickly growing sequence data.

Now several vaccines obtained with genomics and bioinformatics methods are in pre-clinical or clinical trials.

It should be noted that development of vaccine faces several obstacles that could be overcome with the help of analytical methods. LYMERix vaccine designed against the Lyme disease pathogen (*Borrelia burgdorferi*) is such an example. The recombinant microbial outer surface protein (rOspA) of *B. burgdorferi* was used as the vaccine antigenic component. It caused an effective protection in vaccinated persons. However, a part of them complained of arthritic and muscle pain. These symptoms were more severe than those caused by the Borrelia natural infection, despite of the common opinion that artificial vaccines are specially designed to avoid side effects. Most patients suffering from side effects had HLA-DR4+ marker, which is found in one-third of the population. The firm withdrew LYMERix from the market. An immunogenic region responsible for autoimmune reactions was later predicted in the OspA sequence [28, 29]. Recently, Willett and co-authors reported that directed amino acid substitutions could eliminate the risk of autoimmune reactions. Since the mutant OspA

Table 1. Some Artificial Vaccines at Different Stages of Research and Development

Infectious agent	Vaccine	References
Hepatitis B virus	Several licensed vaccines	[13, 14]
Borrelia burgdorferi (Lyme disease)	Earlier licensed LYMERix vaccine was withdrawn from the market. Now a modified candidate antigen is proposed for new vaccine design	[15]
Bordetella pertussis	Licensed vaccine	[16]
Hepatitis E virus	III phase of clinical trials	[17]
HIV	I-III phases of clinical trials	[18, 19]
Malaria plasmodium	III phases of clinical trials	[20]
Influenza virus	Trials on animals	[21-23]
Mycobacterium tuberculosis	Trials on animals	[24, 25]
Streptococcus pneumoniae	Trials on animals, volunteers immunization with testing protective antibodies on mice	[26, 27]

protein retains the ability to elicit the anti-microbial immune response, authors consider it as an antigenic component of the second generation of Lyme disease vaccine [29].

2. EFFECTOR MECHANISMS OF IMMUNE RESPONSE AND VACCINE CONSTRUCT DEVELOPMENT

Prior to the discussion of computational methods used in vaccine design, let us consider briefly the immunological concepts that should be taken into account when a new vaccine is developed (Fig. **1**).

Immune response includes the interaction of B and T lymphocytes with an antigen or its part, resulting in a specific recognition of definite parts of a foreign molecule called antigenic determinants, or epitopes. Immunoglobulin B cell receptors (BCR) recognize surface antigen regions or B-epitopes. T cell receptors (TCR) interact with peptide fragments of a protein antigen (T epitopes) formed in the host cell by limited proteolysis. These peptides are presented at the host cell surface bound to the Major Histocompatibility Complex (MHC). Human MHC proteins are designated as Human Leukocyte Antigens (HLA). MHC class I molecules, expressed in all nuclear cells, bind fragments of proteins synthesized in the host cell, including intracellular parasites (viruses, rickettsiae, some bacteria) proteins, and cleaved in proteasomes. MHC class II molecules, expressed by several types of cells (B lymphocytes and plasma cells, macrophages, dendritic cells and some others), bind fragments of antigens that are captured by these cells *via* endocytosis and processed in endolysosomes. Each MHC class is encoded by a group of multiallele loci (genes). MHC polymorphism should be taken into account during the T epitope identification, as well as vaccine design.

TCRs bind epitopes associated with MHC proteins; hence the recognition of a protein fragment as a T epitope is defined both by TCR and MHC specificities. When a T lymphocyte interacts with a host cell that presents an antigen, the lymphocyte stimulation results in two possible events depending on the T cell type. Cytotoxic T lymphocytes (CTL) interact with peptides bound to MHC-I, initiating the complex reaction cascade, resulting in the death of cells presenting the respective T epitopes. T-helper lymphocytes (Th lymphocytes) start to proliferate after the interaction with MHC-II-bound peptides and give rise to new cell clones. Th cells are divided into subgroups discriminated with produced cytokines. It is considered that Th1 subtype cells participate in the stimulation of cytotoxic responses, and Th2 cells stimulate the antibody production *via* activating B and plasma cells. Th stimulation of the antibody production is achieved by the interaction of Th

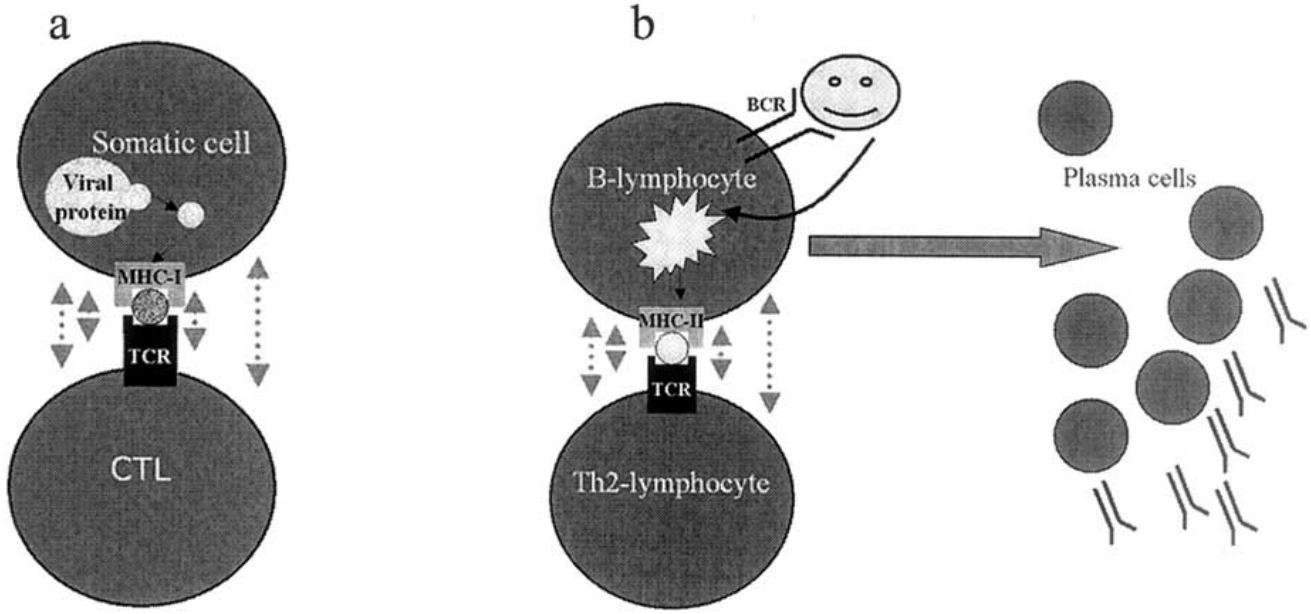

Fig. (1). Interactions of different types of lymphocytes.

a) Interaction between the cytotoxic lymphocyte (CTL) and somatic host cell, expressing foreign (e.g. viral) protein. A peptide fragment (CTL epitope) of processed protein (see text) coupled to MHC-I is recognized by lymphocyte receptor (TCR). Co-stimulatory molecule pairs involved into interaction of two cell are shown by arrows.

b) Interactions between the T-helper lymphocyte and B lymphocyte. The antigen bound by B-cell receptor (BCR) and captured by endocytosis is processed into peptide fragments (T helper epitopes) coupled to MHC-II molecules are recognized by TCRs. As a result of this interaction, the B-lymphocyte gives rise to plasma cell clone that produce soluble immunoglobulins (antibodies) revealing the BCR antigen-recognizing specificity.

lymphocytes with antigen-presenting B cells. TCR of Th cell recognizes the Th epitope bound by MHC-II at B cell surface. Some other ligand–receptor pairs also participate in the interaction between the B and T cells. This process results in mutual stimulation to proliferation of the interacting cells. A portion of the proliferating B cells gives rise to plasma cells producing antibodies; these free immunoglobulin (Ig) molecules have the same recognition specificity as the ancestor B cell receptor.

The portion of B and T lymphocytes stimulated during the primary immune response proliferate into memory cells, which provide a faster secondary immune response to the known antigen. Properly speaking, the vaccination procedure is aimed at producing sufficient pools of memory cells able to raise secondary T- and B-cell-based immune responses to the infection. For more details, see the perfect monograph by Roitt and co-authors [30].

Before starting development of the vaccine, one should define the effector immune mechanisms, which are the most efficient against the respective infection. Various vaccine constructs differ in the stimulation of individual immune processes. Free or coupled to biopolymer carriers synthetic peptides and isolated recombinant proteins usually stimulate the antibody production. In order to induce cytotoxic reactions, other vehicles such as viral vectors, plasmids, or peptide-loaded dendritic cells should be used. New technologies using artificial antigen-presenting cells and their cell-free substitutes are expected to have a major impact on investigation of T-cell immunity as well as immunotherapy [31].

It is necessary to keep in mind that a way of a vaccine administration defines the type of antibodies produced in response to the vaccination. If an infection gate is mucosa, IgA antibodies dominate in protective immune reactions. Hence, vaccines against such infections should be aimed at raising IgA-producing cells [5].

3. GENOMICS STUDIES AND BIOINFORMATICS APPLICATION TO THE DEVELOPMENT OF VACCINE CONSTRUCTS

Fast development of sequencing techniques enables to decode genomes of human, animals, plants as well as bacteria, viruses and protozoa parasites. By analyzing a genome sequence of an infectious agent, one can detect encoded proteins and use them in a vaccine design (Fig. **2**). Thus, protein immunogenic components can be predicted

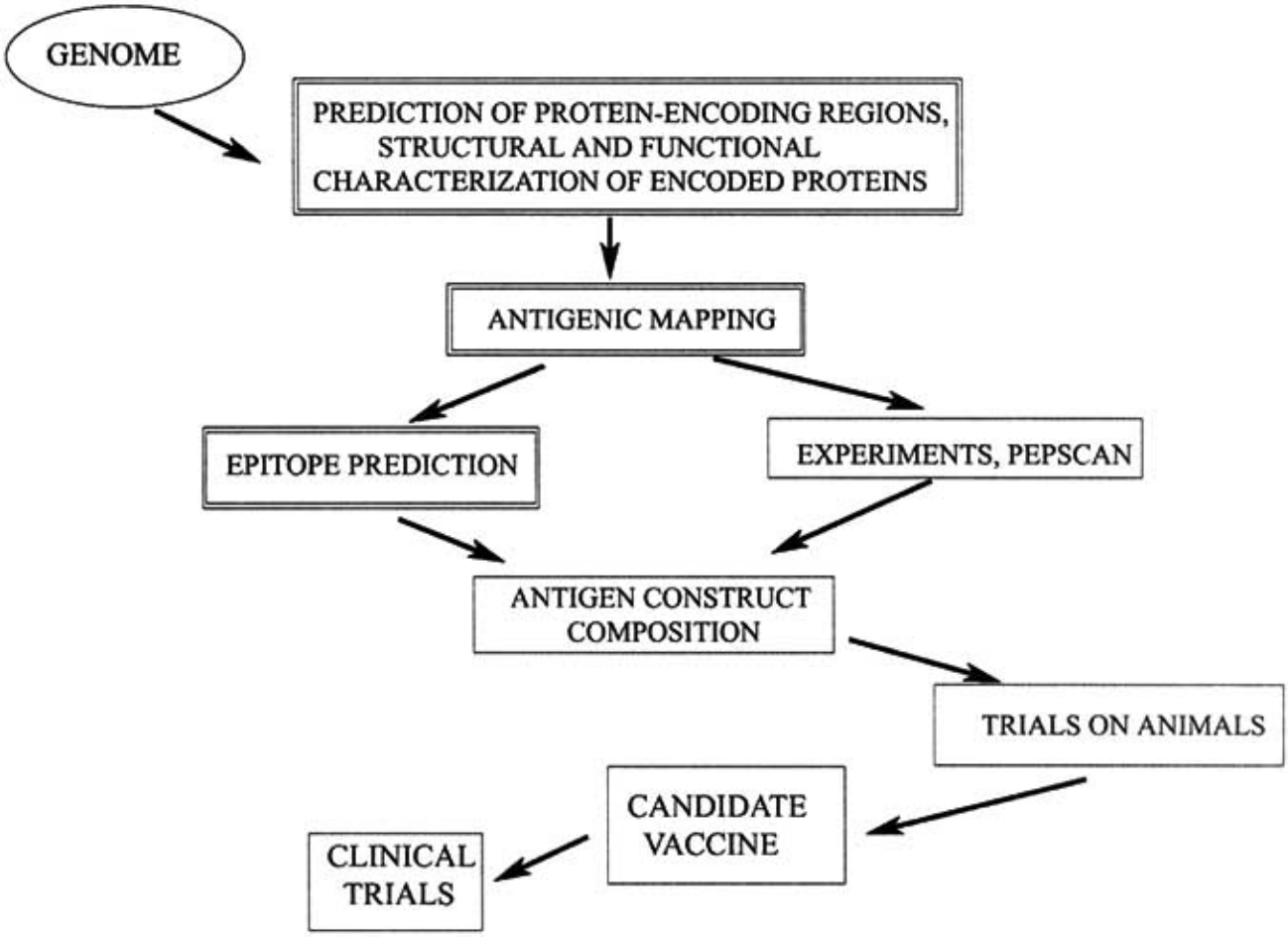

Fig. (2). From genome to vaccine. Units related to computational studies enclosed by double line.

without the time-consuming and highly expensive cultivation stage. Modern experimental technologies (proteomics or DNA microarrays) enable to ultimately validate the prediction. After that, the candidate immunogen can be rapidly manufactured to test its antigenic properties [32].

In certain cases, researchers consider the proteins expressed at different stages of the pathogen life cycle. Anti-malaria vaccine candidates preventing the infection are targeted to the mosquito or pre-erythrocytic parasite stages. Additionally, vaccines directed against the asexual blood-stage are suggested for preventing severe complications of disease such as cerebral malaria or anaemia [33]. Hence, the vaccine design requires a dedicated analysis in which the sequence data should be considered in relation to the existing knowledge of the microbe biology and the infectious process.

Many software tools were developed to store and analyze rapidly growing pools of biopolymer sequences. These new techniques underlie a new discipline known as "bioinformatics" [34].

Bioinformatics tools enable to solve the following problems:

- Detection of potential protein-encoding regions in genomes and "translating" amino acid sequences of putative proteins.
- Prediction of functional characteristics of proteins based on structure/sequence features. e.g., if one detects amino acid sequence fragments, corresponding to patterns of the surface localization, one can expect them to represent B epitopes capable of inducing antibodies specific to the whole protein.
- Comparison of amino acid sequences belonging to different strains of an infectious agent, in order to detect the most invariant proteins or protein parts. A vaccine construct bearing a conserved immunogen is expected to protect against the majority of strains.

If possible, minimal constructs composed of individual T and B epitopes are preferential. This approach is used if immunodominant regions in surface proteins are highly variable; less immunogenic, but more conserved sequence fragments are chosen as candidate antigens. The minimal construct design also seems to be reasonable, if a system expressing the candidate protein cannot provide its adequate modification. Bioinformatics methods predicting individual antigenic regions related to different immune response mechanisms are described below.

3.1. Detection of Protein-Encoding Areas in Genome Sequences

At present, alignment methods are usually applied to identify putative proteins: the detection is based on the similarity to known and characterized proteins. Translated amino acid sequences of proteins under study (taken from genomes) are compared with proteins from available databases, using various local alignment tools, e.g. Blast [35].

One should also make sure that a selected immunogen has no similarity to human proteins. Otherwise, the host organism will not be able to induce an adequate immune response, since a microbial antigen can be recognized as its own. Autoimmune reactions are also possible in this case.

If the similarity to known proteins is not established, translated regions, or open reading frames (ORFs) can be predicted by detecting regulatory regions in nucleotide sequences of infectious agents [32]. New proteins can compose up to 20% of ORFs in a newly sequenced genome. The use of modern proteomic technologies (2D electrophoresis or 2D HPLC followed by immunoblotting or ELISA and peptide fingerprinting based on mass-spectrometry) allows experimental verification of these predictions [36-38].

3.2. Prediction of the Functional Features of Proteins Based on their Amino Acid Sequences

Even if a new amino acid sequence (usually translated from a nucleotide genome sequence) reveals a moderate similarity with known proteins, the regions surrounding functionally significant residues retain specific sequence motifs that can be found with the local alignment procedure. The more reliable prediction is reached using templates that are built from the multiple alignments of homologous sequences. These templates are collected in the special informational resources described below.

The PROSITE database [39] was designed to collect the local motif patterns described as regular expressions. Profiles of protein families are constructed from the general multiple alignment using more sophisticated techniques such as Hidden Markov Models (HMM), and stored in informational resources such as PFAM [40] or SMART [41]. Specialized servers maintain the template databases and provide tools for search of the templates matched to the query sequence. Hence, one can detect the sequence regions in the protein under study that are related to certain functional characteristics.

Since the surface bacterial and viral proteins are considered as the targets for neutralizing antibodies, it is important to detect signals of surface location in the encoded amino acid sequences. For example, membrane-spanning or secreted proteins are predicted *via* detection of a signal peptide (which directs the translocation of the polypeptide chain through a membrane) by using the SignalP software [42, 43] based on the Artificial Neural Networks (ANN) and HMM. Other motifs of the protein location on a microbial (viral, etc.) surface can be also detected. Bacterial lipoproteins are recognized by the presence of the lipid attachment pattern [44]. The SPEPlip software based on ANN, predicts signal peptides and cleavage sites of microbial proteins [45]. Several non-enveloped and enveloped viruses contain outer proteins that process to lipid-anchor modification (myristoylation) at the N-terminus; in some viruses the lipid moiety is involved in viral entry into host cells [46]. These sites are accurately detected by the prediction based on the original method [47]. The similar algorithm is used to predict sites of GPI (glycosylphosphatidylinositol) [48]. C-terminal GPI-modification provides the anchoring of surface proteins in

some protozoan parasites [49, 50] and enveloped viruses [51, 52], therefore, GPI-sites are also considered as surface location determinants.

Bacterial membrane proteins may have several transmembrane beta-strands interleaving by inner and outer hydrophilic loops [53]. Several methods using training sets of resolved 3D structures and based on HMM [54-56], ANN and support vector machine [57] and other techniques [58] enable to predict transmembrane topology with accuracy about 80% and higher. A typical viral glycoprotein sequence contains one or more hydrophobic alpha-helical regions in its C-terminal part and potential N-glycosylation sites in the outer domain [59, 60] (Fig. **3**). The simplest way for locating transmembrane helices is to calculate the hydrophobicity profiles [61] and detect hydrophobic cluster(s) of length of about 20 amino acids. However, this approach gives many false-positive results detecting the hydrophobic clusters in non-membrane proteins. The methods based on the HMM technology (TMHMM [62], HMMTOP [63]) predict transmembrane helices more accurately. The DAS-TMfilter predictor [64] provides a significant decrease of false-positively matched transmembrane helices; this tool outputs the probability scores and matches the proposed signal peptide, which could be interpreted as a membrane anchor in other methods.

The development of a vaccine against *Neisseria meningitidis* group B [65, 66] is a good example of experimental and computational techniques combination. This bacterium exhibits capsular polysaccharides that are widely present in many human tissues. Therefore, the capsular preparation cannot induce a stable immunity unlike in the cases of *N. meningitidis* groups A and C. Protective epitopes (inducing neutralizing antibodies) were identified in the outer loops of membrane proteins. Outer membrane vesicles with inserted meningococcus B proteins were designed. Vaccines based on these constructs showed a good efficacy in clinical trials [67, 68]. However, they induced the strain-restricted immunity due to the high strain variability of loop regions [69]. A special project was launched that enabled the sequencing of the whole Meningococcus group B genome. 600 putative proteins were predicted by computational methods; and 350 of them were cloned in E. coli, purified and used for mouse immunizations. Induced antibodies evoked the complement-dependent lysis of bacteria. Some of the newly isolated proteins are lipoproteins with amino acid sequences conserved among different strains that make them attractive targets for the vaccine development [66].

The number of decoded microbial and viral genomes increases very rapidly [70]. Results of genome sequence analysis are applied in the studies directed to the development of vaccines against viral (Hepatitis C [71, 72], SARS [9-12]) and bacterial (Tuberculosis [73, 74], Syphilis [75], Meningitis [66]) infections. Moreover, genomics studies enable to design experimental vaccines against parasite infections (Malaria [76, 77], Chlamydiosis [78, 79]), which have been out of vaccine studies for a long time.

3.3. Antigen Mapping

Analysis of a genome and encoded amino acid sequences enables to select potential targets of immune reactions. These prediction results should be verified experimentally. Following this verification, there is the choice between a full-size protein and an artificial construct composed of individual antigenic fragments. The full-size candidate protein is chosen if its amino acid sequence is rather

Fig. (3). Predicted functional regions in amino acid sequences of HCV E2 envelope glycoprotein that shows its surface location. A signal peptide (grey shaded) detected by the SignalP program (http://www.cbs.dtu.dk/services/SignalP/) [43]. Potential glycosylation sites (framed) are depicted as matched to the PROSITE template PS00001. Transmembrane region (bolded) is indicated according to DAS-TMfilter predictor [64].

conserved among various strains of an infectious agent and the expression system for this protein provides its correct modification similar to that in the natural host cells [3]. The costs for the cell culture maintenance and protein purification are also taken into account. In the other case, small peptide synthesis can be preferable. Sometimes, production of a recombinant chimeric protein with insertions of conserved immunogenic regions, represents a successful and cost-saving compromise between the above-mentioned approaches.

Even if the first approach seems to be successful, the detailed antigen mapping is necessary for the exact localization of antigenic determinants, estimation of their variability/conservativity, studies of individual immune response variations, and testing the possibility of autoimmune reactions. Antigenic mapping is an obligatory step if a synthetic peptide or a chimeric recombinant vaccine is to be designed. In order to induce antibody response to a pathogen, a vaccine construct should contain B epitopes able to provoke the synthesis of pathogen-neutralizing antibodies, and Th epitopes recognizable by Th2 lymphocytes that will stimulate the B cell conversion into antibody-producing plasma and memory cells. If the cytotoxic immune response against a pathogen is a target, the vaccine construct should include cytotoxic as well as Th1 cell-stimulating T epitopes. The approaches and software for B and T epitope predictions are described below.

3.3.1. B Epitope Prediction

In general, B epitopes represent chemical structures able to interact with antigenic-binding regions of antibody or BCR. Any chemical structure of an infectious agent that is recognized as a foreign (polypeptide, glycan, lipid, glycolipid etc.) by the host B cells, in principle, is able to cause an antibody-dependent immune response against itself. However, in this paper we will consider only polypeptide B epitopes, for three principal reasons. First, these epitopes represent more diverse pool than others. Second, their recognition as non-selves is sometimes not as obvious as that of non-peptidic B-epitopes (see the above mentioned example of *N. meningitidis* group B antigen search). Third, a huge set of bioinformatics methods exists that should help, and really does in certain cases, reveal B epitopes in proteins.

B epitopes are formally classified into two groups: linear, or continuous, and conformational, or discontinuous. The former term corresponds to B epitopes that are composed of continuous stretches of amino acid residues representing fragments of protein primary structures. The latter term corresponds to B epitope structures that are formed with amino acid residues or their groups not adjacent to one another in the protein primary structure, but put in the vicinity after the polypeptide chain folding [80]. This classification is somewhat ambiguous since the recognition and presentation of the so-called linear, or continuous B epitopes may also depend on the protein conformation. Experimental results show that the majority of protein B epitopes is conformational. Linear B epitopes compose a portion of about 10% of all protein B epitopes usually detected.

Taking into account sizes of contact sites between immunoglobulin molecules and antigens, B epitope surface should be coincided to stretch to 4-8 amino acid residues [81]. This so-called contact B epitope makes the main thermodynamic contribution into the antigen-antibody interaction [82]. Another point of view takes into account the distance between the paratope and epitope atoms without regard for thermodynamic calculations. The B epitope structure defined by this way extends up to 22 amino acid residues [83]. Both short and long B epitopes can be present with either linear amino acid residue stretches or groups of residues assembled together upon protein folding. Nevertheless, it is obvious that only a part of atomic groups of a 22-residue stretch forms the contact B epitope.

While long B epitopes as well as conformational ones (of any length) can be assumed only from 3D structure data, short linear B epitopes can be predicted from amino acid sequences. These predictions are based on the fact that because of their physicochemical properties, different amino acid residues are differently distributed between the surface and the interior of the protein globule, as well as between various secondary structure elements. A certain index can be assigned to each amino acid residue, depending on its physicochemical properties or occurrence frequencies in certain structures. Indices are retrieved from special scales containing index values for all 20 natural amino acids, averaged for small overlapping stretches of the amino acid sequence of the protein under study, and plotted along the sequence [84]. Some scales used for B epitope predictions are listed below:

- Hydrophobicity / hydrophilicity scales are the most common. Scale values are obtained from experimental data and represent: free energy of transfer from water to ethanol [85] or from an organic solvent to water [86]; surface tension of amino acid solutions [87]; retention time of HPLC obtained for peptides of different composition [88].
- Flexibility scale is calculated from temperature factor values obtained for various regions in protein structures resolved by X-ray crystallography. The use of this scale is based on the suggestion that flexibility of protein molecule portions correlates with their antigenicity [89].
- Acrophilicity [90] scale is calculated from the distribution of residues standing away from protein 3D-structure core and the accessibility scale [91] also represents the degree of surface exposures of amino acid residues in resolved X-ray protein structures.
- Antigenicity scale [92] is based on statistically calculated occurrence frequencies of amino acid residues in linear epitopes experimentally detected in 20 proteins.

Secondary structure prediction is also applied for B epitope predictions. These methods are usually based on statistics of amino acid residue occurrences in certain secondary structure elements in solved 3D protein structures [93]. Sequence regions predicted as turns or loops of a polypeptide chain are considered as possible antigenic determinants.

Though these techniques were developed more than twenty years ago, the main principles of the antigenicity prediction by using the scales still remain unchanged. In general, all above-mentioned methods have been designed to predict surface-located regions of the protein molecule. Even the most accurate ones reveal about 65% residues that are really found in antigenic regions. However, all modern software for the B epitope predictions uses different combinations of the above mentioned scales that enable to increase the predictive power [94]. Nevertheless, these program tools still remain popular up to now since they allow a researcher to exclude regions with low prognostic scores from the following pre-experimental B epitope search [95].

More sophisticated method, which processes sequence alignment data with the help of ANN technology, enables to predict B epitopes stretches with average accuracy of 68 %. The latter value comparable with those obtained with the scale methods seems to provide an effective practical upper limit for the accuracy of predicting accessibility from a sequence only [96].

The most precise way is to locate loop regions in a model of 3D structure of a protein. Unfortunately, currently the number of solved 3D structures is much less than the number of determined amino acid sequences. However, one can build a 3D model if a sequence under study reveals high homology with one or several proteins with known 3D structures [97]. The sequence of a protein under study is folded into a 3D model by using the known 3D structures as templates. This approach was applied to the detection of antigenic determinants in the envelope protein of Japanese Encephalitis Virus; the homologous protein of Tick-born Encephalitis Virus was used as the 3D template. Minimal energy calculations for the revealed fragments showed that a vaccine construct should maintain the native conformation of an inserted epitope and induce antibodies against the whole protein from the native source [98].

3.3.2. T Epitope Prediction

T epitopes are fragments of proteins formed in host cells by limited proteolysis in proteasomes or endosomes, exposed at the cell surface bound to MHC proteins, leading to T lymphocyte activation *via* the interaction with TCR.

According to the crystallographic data, peptide ligands adopt extended conformations inside the MHC cavity and are associated with MHC by hydrogen bonds and van der Waals interactions. Docking of a peptide and MHC is possible, if the peptide sequence satisfies to specific positional motifs. T epitope motif is composed of sequential positions, each of them being specified in terms of amino acid type preferences. So-called anchor positions are the most restricted with regard to their amino acid composition and thus, are the most significant for MHC binding [99]. T epitope prediction methods are generally aimed at searching the motifs specific to MHC binding. In order to predict T epitopes, one should detect the cleavage sites in a processed protein and locate anchor amino acid positions in these cut stretches. The positional pattern methods were developed at first.

Structural features of MHC-I- and MHC-II–binding cavities differ significantly. The MHC-I binding pocket holds 8-11-mer peptides and is closed at both ends; the most stable hydrogen bonds are formed with amino acid residues close to the ligand peptide termini. A common MHC-I ligand (that is recognized by cytotoxic CD8+ lymphocytes) is a 9-mer peptide with two anchor residues in the second and ninth positions. The MHC-II cavity is opened at both ends and accommodates longer peptides, which can extend the cavity; anchor positions are located along the pocket [100]. Peptide ligands for MHC-II may vary in length from 12 to 25 residues, however, a typical MHC-II binding motif corresponds to a 12-15-mer peptide with two or four anchor residues depending on MHC–II allele pocket.

Each MHC locus or even locus allele has its specific peptide ligand motif, which represents the local differences in structure of the binding cavity including anchor residue allocation. These motifs can be calculated based on analysis of a vast amount of both cytotoxic and helper T epitope mapping data. Experimentally found sequences of peptides bound by MHC (MHC ligands) are stored in specialized freely available databases such as SYFPEITHI [99] and FIMM [101]. The corresponding servers provide tools for searching T epitope motifs in a query amino acid sequence based on the position-dependent matrices, which contain scores for 20 amino acid residues for each peptide ligand position. EPIPREDICT software uses matrices calculated from data obtained with the help of synthetic combinatorial peptide libraries; it enables to describe the MHC-peptide interactions in quantitative manner and predict peptides specific to the certain HLA-II alleles [102].

The other methods that used the ANN [103] and HMM technology [104], are also applied for MHC-ligand prediction.

The accuracy of these methods depends on the underlined data set quality. Yu and co-authors [105] carried out trials of prediction tools using positional matrices, ANN and HMM. They used datasets composed from experimentally established peptides that bind or do not bind to allele-specific MHCs. Integrative accuracy scores were calculated, accounting the sensitivity and specificity values accessed by the leave-one-out cross-validation. All used tools revealed the good predictive performance for HLA-A*2001, and poor performance (close to random choice) for HLA-B*3501. The ANN and HMM methods revealed the significant improvement of accuracy with increase of the data sets.

Another method implements the partial least squares-based, multivariate, statistical approach to the quantitative prediction of peptide binding to MHC-I [106] and MHC–II [107]. The authors used data collected in JenPep database [108] that contains experimental quantitative data on the MHC-peptide binding. The suggested approach showed the best results compared to four other methods revealing 24 from 25 known T-cell epitopes specified to MHC-II at the benchmark. The same authors applied their approach to design successfully series of high-affinity HLA-A2 peptides [109].

So, methods based on the sequence and functional data are limited to a small number of MHC alleles. However, this number still increases, and earlier defined motifs are corrected due to the continuous work on T epitope mapping using new high-throughput experimental techniques:

Table 2. Informational Resources on T Epitopes and MHC Ligands

Title and brief characterization	WWW address
HLA-Peptide Ligands. Web page contains links to the most popular resources on HLA peptides.	http://www.ihwg.org/components/peptider.htm
PROPREDICT. Server designed for predicting the MHC-II ligands.	http://www.imtech.res.in/raghava/propred/index.html
EPIPREDICT. Software for prediction of T epitopes and MHC-II ligands.	http://www.epipredict.de/index.html
EPIMATRIX. Commercial resource provides services on defining T epitopes.	http://epivax.com/
SYFPEITHI. Database on MHC ligands and T epitope motifs supplied with prediction tools.	http://www.syfpeithi.de/
MHCPEP. Database on T epitopes.	http://wehih.wehi.edu.au/mhcpep/
FIMM. Integrated database of functional immunology, focusing on MHC, antigens, T- and B-cell epitopes, and diseases.	http://sdmc.lit.org.sg/fimm/
BIMAS. Software for prediction of peptide fragment bound by human MHC-I.	http://bimas.dcrt.nih.gov/molbio/hla_bind/
VACCINOME. Non-commercial site provides the software for prediction of T epitopes (TEPITOPE).	http://www.vaccinome.com/
PAPROC. Software for prediction of proteosomal cleavage sites.	http://www.paproc.de/
MHC-Peptide (MPID). Database on 3D structures of MHC-peptide complexes.	http://surya.bic.nus.edu.sg/mpid/intro.html
MHCPred. Server for quantitative prediction of peptide-MHC binding.	http://www.jenner.ac.uk/MHCPred/
JenPep. Database contains binding data for the interaction of peptides with MHC, TAP transporter, and T cell receptor, annotated lists of B cell and T cell epitopes.	http://www.jenner.ac.uk/JenPep/
MHCBN. Curated database consisting of detailed information about binding, non-binding peptides and T-cell epitopes.	http://www.imtech.res.in/raghava/mhcbn/
MHC-Tools. Collection of MHC structures with links to related sites.	http://web.mit.edu/stern/www/mhctools.htm

combinatorial peptide libraries [102, 110-112] and peptidomic analysis of MHC-bound ligands [38].

The reliability of T epitope prediction is significantly increased due to the exact location of cleavage sites in the amino acid sequence. The PAProc program [113] was designed to predict the cleavage sites for proteasome enzymes that process intracellular proteins into ligands for MHC-I. The software uses a model based on the ANN trained on the experimental data. The PAProc program can also be applied to the vaccine design, because it helps to ensure the required processing of the vaccine construct. One should take into account that peptides obtained as a result of proteasome cleavage are subject to further N-terminal trimming [114].

Cytotoxic T epitope predictions can also be improved for certain HLA alleles *via* in silico assessing the ability of protein fragments to bind to the transporter protein TAP. Identification of TAP-binding peptides is performed with the help of ANN technique [115]. TAP-binding peptides can also be predicted at the JenPep server [108].

Links on servers with software tools for T epitopes prediction are given in Table **2**.

A quite different approach is based on 3D modeling of an MHC loaded with a peptide. The solved 3D structure of MHC-peptide complex is used to build a template, which a modeled peptide is fitted to. Six side pockets are specified in the 3D template so they can hold amino acid side chains of the ligand; pocket locations are invariable, but their physicochemical parameters can be changed. This approach enables to design a rough model. The EpiDock software [116] builds a more accurate model by the minimization of free energy scoring function (FRESNO) [117]. By scanning the Hepatitis B Virus (HBV) genome, the program detected 80% of known T-epitopes specified to HLA A*2001 with accuracy (true positive + true negative / total) from 77 to 85% for different HBV proteins. This accuracy is comparable with that obtained by sequence-based methods, but EpiDock enables to predict ligands of different MHC-I alleles including epitopes with undetermined motifs, using only the amino acid sequence of MHC allele, and promptly search the epitopes in the whole genomes [116].

Other authors have applied a statistical method of prediction using pooled peptide sequence data and three-dimensional modeling by molecular mechanics calculations. They developed a novel predictive model using information obtained from 29 human crystal structures of MHC loaded with peptides. The proposed method provides predictions for any given MHC allele whose sequence is defined [118].

4. VACCINE CONSTRUCT DESIGN AND CHOICE OF VEHICLE

When the immunogenic components are selected (see above), they can be used for designing a vaccine. Such characteristics as the construct structure, the way and vehicle for the delivery depend on which immune mechanisms should be induced by a new vaccine to provide an efficient protection. Keeping this in mind, in this section computational methods are discussed along with experimental approaches.

4.1. Main Principles of a Vaccine Construct Design

Computational methods are used to analyze amino acid sequences and select the candidate proteins or their fragments for further vaccine design. Vaccine constructs should induce either humoral or cellular immunity or both components of the immune response. Hence, a minimal vaccine construct should contain B-epitopes or/and CTL T epitopes depending on the required immune reactions, as well as Th epitopes that provide the maintenance of the immune response. As it was mentioned above, the cleavage sites inserted at the termini of T epitope regions provide the necessary antigen processing. However, lysosomal protease cleavage sites (that are not predicted reliably now) are not necessary for small peptide constructs, as the MHC-II pocket allows accommodating long peptides (see above).

Selecting Th epitopes, a researcher should take into account the distribution of HLA alleles in the population [119]. Vaccine developers can also use known T-helper epitopes isolated from other infectious agents. So-called universal or promiscuous epitopes reveal affinity to several HLA alleles most propagating in a certain population [120, 121]. Some of them are given in the Table **3**. Promiscuous epitopes are used to provide maintenance of the primary immune response; it is supposed that memory cells induced by vaccination are able to provide effective secondary response on the natural infectious agent, without specific T-helper maintenance. Perhaps, in this case, strictly specific T epitope of a pathogen binds T cell lines activated by the widely specific vaccine epitope. Evaluation of vaccine construct ability to induce the cell immune reactions can be carried out on transgenic animals, expressing HLA proteins [122, 123].

As it was shown in experimental studies, it is possible to increase immunogenicity of the candidates of CTL epitopes by site-directed mutagenesis of certain motif positions [128]. Thus, bioinformatics data helps to carry out the directed changes of antigenic features. Moreover, researchers can change replacement of separate immunogenic regions to select the more effective compositions.

4.2. Applying Minimal Constructs Against Extremely Variable Infectious Agents

High antigenic variability is one of the major obstacles for development of new vaccines against many microbial and viral infections. Certain viruses mutate so quickly that it is difficult to select the immunogenic molecule, which could not change its antigenic features.

Such viruses as HIV or HCV, are established to dramatically change their antigenic features. With modern sequencing techniques, researchers obtained large sets of homological sequences encoded by variants of the same virus. High variation in sequence causes significant antigenic diversity. Molecular portions containing immunodominant epitopes and enabling to induce effective virus-neutralizing antibodies reveal significant variability. Using this mechanism, the virus escapes from immune press.

It should be noted that the majority of effective vaccines protected from microbes and viruses, which reveal cytopathic effects. Agents that do not reveal cytopathic activity (e.g. HIV and HCV) can persist in the host cell, generating the multiple variants in the absence of significant immune response [129].

Envelope proteins of HIV and HCV contain hypervariable regions [130, 131] that give the major contribution into the virus diversity. Usually, the same patient is simultaneously infected by several viral variants (quasispecies). During the infectious process composition of quasispecies varies. Chimpanzees immunized by peptide representing HVR1 (hypervariable fragment of the HCV envelope protein), revealed the effective protection against

Table 3. Examples of Promiscuous T Epitopes

Source protein	Peptide	Refs	Specificity*
Tetanus toxin	QYIKANSKFIGITE	[124]	DR1, DRw15(2), DRw18(3), DR4Dw4, DRw11(5), DRw13(w6), DR7, DRw8, DR9, DRw52a,DRw52b
Pfg27 (*Plasmodium falciparim*, sexual stage)	IDVVDSYIIKPIPALPVTPD	[125]	DR15 (DRB1*1503), DR17 (DRB1*03011), DR18 (DRB1*03), DR51 (DRB5*0101), DR52 (DRB3*0101/DRB3*02)
PvMSP-1 (*Plasmodium vivax* merozoit)	LEYYLREKAKMAGTLIIPES	[126]	DRB1*0301,DRB1*0401,DRB1*1101,DRB1*0101.
Mce2 (*Mycobacterium tuberculosis*)	PRYISLIPVNVVAD	[127]	DRB1*0101 (DR1), DRB1*1501 (DR2),DRB1*0301 (DR3),DRB1*0401 (DR4),DRB1*1101 (DR5),DRB1*0701 (DR7),DRB1*0801 (DR8)
Mce2 (*Mycobacterium tuberculosis*)	VATRAGLVMEAGGSKVT	[127]	DRB1*0101 (DR1), DRB1*1501 (DR2),DRB1*0401 (DR4),DRB1*1101 (DR5),DRB1*0701 (DR7)

*Designations of the HLA variants according to referred papers

only a particular virus variant, remaining susceptible to other HCV variants [132].

In this case, a living vaccine is obviously inappropriate. Vectors expressing the whole envelope proteins do not solve the problem of antigenic variability. Therefore, researchers try to detect conservative immunogenic stretches in antigen molecules. The conservative regions bearing B epitopes were found even in permanently mutating HCV envelope proteins [133, 134].

The next step is the design of a construct, which could induce a protective response on the antigenic determinants that are not immunodominant at natural infection. The other way is to compose epitope set, inducing the protection against the majority of virus variants by cross immunity. The latter approach does not seem to be suitable, at least in the case of Hepatitis C. According to our estimation, the number of variants distinguished on the HVR1 region exceeds 2000.

Thus, studying amino acid sequences, we can find the variable regions of viral proteins as well as the regions, which remain conservative in all viral variants. This allows a researcher to select the most prospective immunogenic candidates.

4.3. Synthetic Peptide Vaccines

At present, several experimental vaccines are based on chemically synthesized antigenic peptides. The anti-malarial peptide vaccine is a single preparation of such kind that was subject to wide-ranging clinical trials in three parts of the world [20, 135]. Synthetic immunogenic constructs can be linear oligopeptides, branched dendrimers (so-called lysine trees), oligomers with attached functional groups and modified peptides. The more complex solutions are based on polymerization of modified peptides or their co-polymerization with other compounds, e.g. acrylamide. So, separated epitopes can be located in the linear mode or attached to the branched groups [121, 136]. Relative position of B and T epitopes can be selected in experiments to induce the most effective immune reactions. A bioinformation should recommend the localization of epitopes, keeping in mind the most probable cleavage sites and predicted secondary structure.

Peptides identified as immunogens are able to induce strong immune response if administered with the adjuvant. Long fatty chains and multicharged polymers are commonly used as adjuvants [121].

It should be noted that the modern peptide synthesis technology also includes the preliminary computational stage. Currently, the companies provide the preliminary calculations of custom peptide features using the scales mentioned above, representing the hydrophobicity, surface probability and antigenicity; results are delivered to a customer to confirm the order or change it. Another way is to provide the software to a customer directly. In this case, a user can compose a synthesis protocol at her/his desk prior to send an order.

4.4. Recombinant Antigens

Today, genetic engineers can create artificial biological system producing the required antigen proteins [3]. Bacteria, yeast and plants provide the base for the most effective producing systems. The choice is also dependent on the producing system ability to perform correct post-translational modifications. The latter is important for revealing native antigenic peculiarities.

Table 4. Bioinformatics Resources Related to Vaccinology

Title and brief characterization	WWW address
HLA Nomenclature. Interactive reference system.	http://www.anthonynolan.com/HIG/index.html
CMR (Comprehensive Microbial Resource). Bacterial genome database.	http://www.tigr.org/CMR
IMGT. International resource on immunogenetics.	http://imgt.cines.fr/
The Vaccine Page: Vaccine News & DataBase.	http://vaccines.org/
HCVMAP. Database on antigenic mapping of HCV proteins. T and B epitopes collection.	http://ibmc.msk.ru/hcvmap/
HIV Molecular Immunology. Collections of known T and B epitopes. Epitope maps. Searching tools.	http://hiv-web.lanl.gov/content/immunology
HIV / SIV Vaccine Trials Database. Contains information about vaccine studies using SIV and HIV in nonhuman primates.	http://www.hiv.lanl.gov/cgi-bin/vaccine/public/index.cgi
HCV Immunology Database. Collections of known T epitopes. Epitope maps. Searching tools.	http://hcv.lanl.gov/content/immuno/immuno-main.html
Hepatitis virus database. Sequence information on Hepatitis B, Hepatitis C and Hepatitis E viruses.	http://s2as02.genes.nig.ac.jp/
HCVDB. Database is designed for investigation of interrelations between viral sequences and pathogenic processes.	http://hepatitis.ibcp.fr/
SARS Bioinformatics Suite. Bioinformatics site provides in depth data, including a coronavirus database, and unique tools to analyze the genomes, genes and proteins of SARS and other related viruses.	http://athena.bioc.uvic.ca/sars/

Plasmid vectors are widely used for recombinant protein expression. It should be noted that plasmids are also designed with computational methods. A recombinant sequence can represent the entire protein or artificial minimal construct. Microbes that express surface proteins bearing the insertions of foreign antigenic region (e.g. *E. coli* or lactobacteria) are suitable to induce the humoral response. Attenuated Salmonella strains enabling to intracellular reproduction can be applied as vectors for constructs, which should induce both the cytotoxic and antibody-dependent response. Virus-based construction should activate the cytotoxic mechanisms [3].

Intracellular antigen expression is also provided by DNA vaccination, when the protein-encoding nucleotide sequences being inoculated into the organism are translated providing protein antigens [137]. Plasmid-based preparations are injected intramuscularly or subcutaneously. The bacteria presented as an input into the respiratory or intestine mucosa cells can be also used as DNA vaccine vehicles.

Sequencing of plant genes or whole genomes enables to apply bioinformatics for selecting the plant proteins, which could be appropriate for insertions of candidate antigens. It seems very attractive because the plant producing system is also used as a delivery vehicle. Several plant vaccines have already passed the clinical trials [138-146]. Existing technologies enable to process plant material to obtain elaborate lyophilized substances with precise antigen dosage [147]. However, the edible antigen taken into the intestine can suppress the immune response in animals, as a usual food antigen [148]. In such situation, special protocols are necessary to develop the edible vaccines [149].

5. BIOINFORMATICS RESOURCES USED FOR VACCINE DESIGN

The most popular general informational resources can be used for vaccine design. The IMGT databases [150] contain structural data on genes and proteins, which participate in the immune system functioning – HLA–loci, T cell receptors, immunoglobulins etc. Informational resources on microbial genomes provide the sequence data as well as bioinformatics tools (such as Blast).

The most interesting and useful information can be retrieved when we use the heterogeneous datasets related to microbes or viruses. Now the complex and flexible nets combine the general-purpose resources as well as specialized ones. Modern informational technologies enable to design a database rather quickly; so the universal principles and rules should be developed for including new databases into the large informational complex. Well known systems as SRS [151], ENTREZ [152], MIPS [153] and other servers provide navigation across the interlinked databases, including microbial genome resources. The specialized

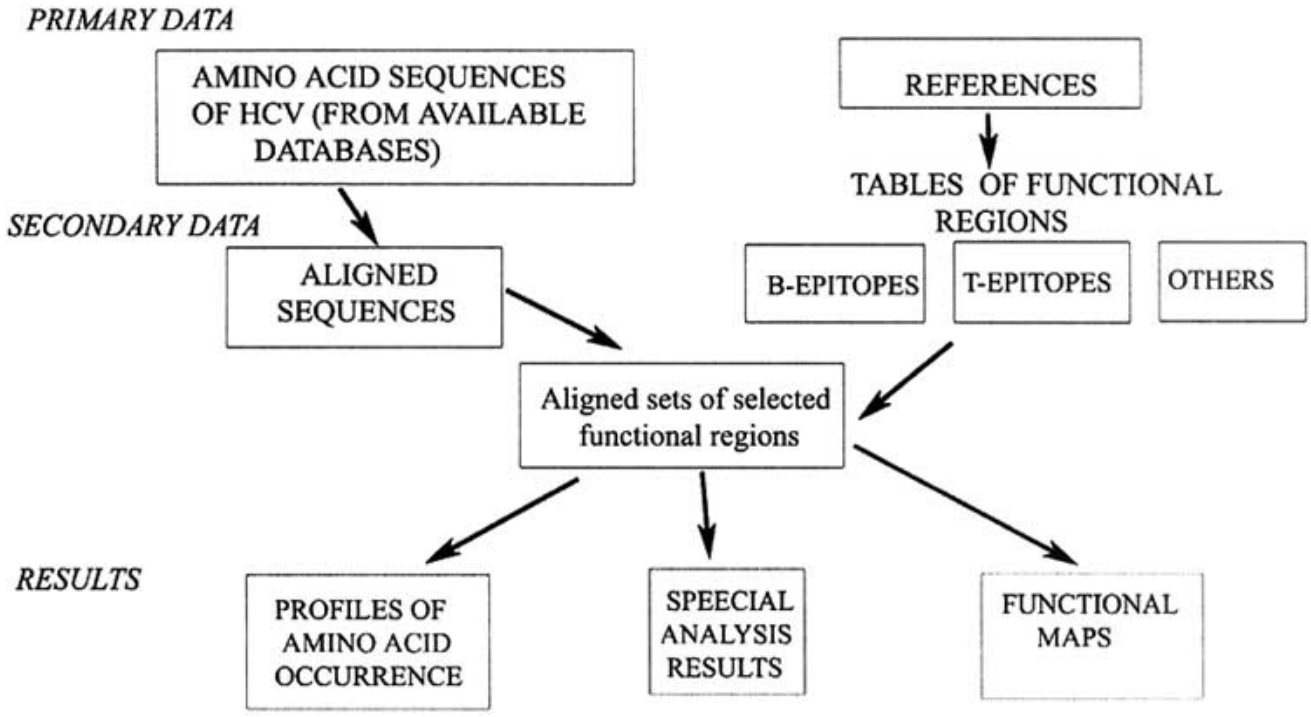

Fig. (4). Database on mapping the proteins, expressed by different variants of the same virus.

resources on the certain pathogens such as HIV, SARS coronavirus and HCV (see Table **4**), represent data on annotated nucleotide and amino acid sequence (annotations), variations of functional and antigenic regions in different strains. They also provide sequence analysis tools and store the obtained results (e.g. alignment) as secondary data. Maintenance of these resources is reasonable, because the number of different sequences related to the same virus reaches several tens of thousands.

The database on functional variability of proteins, encoded by the same virus, should be organized like scheme shown in Fig. (**4**). Now this scheme is partly implemented in the database HCVMAP designed by the Institute of Biomedical Chemistry (Russia) for antigenic mapping of the protein sequences (http://ibmc.msk.ru/hcvmap).

The primary dataset collects amino acid sequences and their functional characteristics. The sequences are retrieved from freely available resources of the UniProt knowledgebase [154]. Functional characteristics are collected using various publicly available resources (PubMed bibliographic system, epitope collections and others). The secondary data represent the aligned sequences and mapped regions related to the functional characteristics – epitope tables. These tables contain the peptide sequences mapped by their locations in HCV protein sequences excluding mimotopes. Each record contains brief characteristics, obtained for these regions. Most of the presented epitopes are human, but some of them are murine and chimpanzee ones. Browsing the epitope tables, one can invoke the reference page by clicking the "book" icons. Through this page, a user can access the Entrez bibliographic system (PubMed). By clicking labels at the table row, the user calls out the alignment set of a homologous fragment, including the corresponding region with data on occurrence frequencies for each peptide stretch. A click on the peptide string invokes all Swiss-Prot/TrEmble identifiers of sequences that contain it. A click on identifier can evoke a corresponding sequence with the colour-emphasised peptide. Besides the epitope tables, the frequency profile calculated on aligned sequences can be displayed. The corresponding page contains the HCV polyprotein scheme that allows the navigation along the different profile regions with amino acid occurrence frequencies.

6. CONCLUSION

Due to the vaccine application and improvement of sanitary conditions, several traditional infections wiped out. Moreover, such infections as Smallpox and Poliomyelitis are considered as practically beaten. At the same time, the effective vaccines against Malaria and other parasitical infections are not yet introduced in medical practice. Furthermore, new infectious agents (e.g. HIV, HCV) surprisingly appeared in the last quarter of the twentieth century, well known *Mycobacterium tuberculosis* presents new resistant variants causing hardly or non-curable sickness [155], etc.

It is expected that mankind may be faced with absolutely new infectious agents or new types of known ones. The first example and a big challenge for a new century is SARS. Today, we are also forced to take into account the threat of bioterrorism. The situation is quite possible when the traditional techniques will not be suitable to create effective vaccines. New approaches are necessary to design new kinds of vaccines – therapeutic, anticancer and contraceptive [156-159].

Bioinformatics, genomics and proteomics provide new opportunities to reveal and extract candidate immunogenic components that could be used in design of new vaccine generation. However, we should not forget about the dangers arising from the incompleteness of knowledge. The deeper understanding of molecular foundations of immunity can be reached, based on the integrative computational biology. Thus, further development of computational methods is necessary for theoretical studies, as well as for practical applications. It is clear that genomics and proteomics data and corresponding computational techniques will be widely used in this very promising direction.

ACKNOWLEDGEMENTS

We thank Drs. T.I. Kuzmina, M.A. Eldarov and S.A. Moshkovski for their help in selecting the papers and useful discussion.

This work is partly supported by the Russian Foundation of Basic Research, grant No 04-04-49390 and the Russian Ministry of Education and Science Interdisciplinary Program "New Generations of Vaccines".

ABBREVIATIONS

ANN = Artificial Neural Network

BCR = B cell receptor

HCV = Hepatitis C virus

HIV = Human Immunodeficiency Virus

HLA = Human Leukocyte Antigens

HMM = Hidden Markov Model

MHC = Major Histocompatibility Complex

SARS = Severe Acute Respiratory Syndrome

TCR = T cell receptor

REFERENCES

[1] Makela, P.H. *FEMS Microbiol. Rev.*, **2000**, *24*, 9-20.
[2] O'Hagan, D.T.; Rappuoli, R. *Pharm Res.*, **2004**, *9*, 1519-1530.
[3] Liljeqvist, S.; Stahl S.J. *Biotechnol.*, **1999**, *73*, 1-33.
[4] Ben-Yedidia, T.; Arnon, R. *Curr. Opin. Biotechnol.*, **1997**, *8*, 442-448.
[5] Rappuoli, R.; del Guidice, G. In *Vaccine. From Concept to Clinic;* Paoletti, L.C.; McInnes, P.M. Ed.; CRC Press LLC: Boca Raton, Boston, London, New York, Washington DC, **1999**; pp. 5-17.
[6] Rappuoli, R. *Curr. Opin. Microbiol.,* **2000**, *3*, 445-450.
[7] Schlagenhauf, P.; Ashraf, H. *Lancet,* **2003**, *361*, 1017.
[8] Rota, P.A.; Oberste, M.S.; Monroe, S.S.; Nix, W.A.; Campagnoli, R.; Icenogle, J.P.; Penaranda, S.; Bankamp, B.; Maher, K.; Chen, M.H.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J.L.; Chen, Q.; Wang, D.; Erdman, D.D.; Peret, T.C.; Burns, C.; Ksiazek, T.G.; Rollin, P.E.; Sanchez, A.; Liffick, S.; Holloway B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Gunther, S.; Osterhaus, A.D.; Drosten, C.; Pallansch, M.A.; Anderson, L.J.; Bellini, W.J. *Science,* **2003,** *300*, 1394-1399.

[9] Gao, W.; Tamin, A.; Soloff, A.; D'Aiuto, L.; Nwanegbo, E.; Robbins, PD.; Bellini, WJ.; Barratt-Boyes, S.; Gambotto, A. *Lancet*, **2003**, *362*, 1895-1896.

[10] Kondro, W. *Can. Med. Assoc. J.*, **2004**, *170*, 183.

[11] Xiong, S.; Wang, Y.F.; Zhang, M.Y.; Liu, X.J.; Zhang, C.H.; Liu, S.S.; Qian, C.W.; Li, J.X.; Lu, J.H.; Wan, Z.Y.; Zheng, H.Y.; Yan X.G.; Meng, M.J.; Fan, J.L. *Immunol. Lett.,* **2004**, *95*, 139-143.

[12] Pang, H.; Liu, Y.; Han, X.; Xu, Y.; Jiang, F.; Wu, D.; Kong, X.; Bartlam, M.; Rao, Z. *J. Gen. Virol.*, **2004**, *85*, 3109-3113.

[13] Andre, F.E. *Vaccine*, **1990**, 8 Suppl., S74-S78.

[14] Kao, J.H.; Chen, D.S. *Lancet Infect. Dis.*, **2002**, *2*, 395-403.

[15] Parenti, D. *Conn. Med.*, **1999**, *63*, 570.

[16] Greco, D.; Salmaso, S.; Mastrantonio, P.; Giuliano, M.; Tozzi, A.E.; Anemona, A.; Ciofi degli Atti, M.L.; Giammanco, A.; Panei, P.; Blackwelder, W.C.; Klein, D.L.; Wassilak, S.G. *N. Engl. J. Med.*, **1966**, *334*, 341-348.

[17] Hyams, K.C. *Curr. Gastroenterol. Rep.*, **2002**, *4*, 302-307.

[18] Schultz, A.M.; Bradac, J.A. *AIDS*, **2001**, *15* Suppl. 5, S147-S158.

[19] Mooij, P.; Heeney, J.L. *Vaccine*, **2001,** *20*, 304-321.

[20] Migasena, S.; Heppner, D.G.; Kyle, D.E.; Chongsuphajaisiddhi, T.; Gordon, D.M.; Suntharasamai, P.; Permpanich, B.; Brockman, A.; Pitiuttutham, P.; Wongsrichanalai, C.; Srisuriya, P; Phonrat, B.; Pavanand, K.; Viravan, C.; Ballou, W.R. *Acta Trop.*, **1997**, *67*, 215-227.

[21] Kandel, R.; Hartshorn, K.L. *BioDrugs*, **2001**, *15*, 303-323.

[22] Jackson, D.; Cadman, A.; Zurcher, T.; Barclay, W.S. *J. Virol.*, **2002**, *76*, 11744-11747.

[23] Jeon, S.H.; Ben-Yedidia, T.; Arnon, R. *Vaccine*, **2002**, *20*, 2772-2780.

[24] Agger, E.M.; Andersen, P. *Vaccine*, **2002**, *21*, 7-14.

[25] Doherty, T.M.; Olsen, A.W.; van Pinxteren, L.; Andersen, P. *Infect. Immun.*, **2002**, *70*, 3111-3121.

[26] Gor, D.O.; Ding, X.; Li, Q.; Schreiber, J.R.; Dubinsky, M.; Greenspan, N.S. *Infect. Immun.*, **2002**, *70*, 5589-5595.

[27] Briles, D.E.; Hollingshead, S.K.; King, J.; Swift, A.; Braun, P.A.; Park, M.K.; Ferguson, L.M.; Nahm, M.H.; Nabors, G.S. *J. Infect. Dis.*, **2000**, *182*, 1694-1701.

[28] Steere, A.C.; Gross, D.; Meyer, A.L.; Huber, B.T. *J. Autoimmun.,* **2001**, *16*, 263-268.

[29] Willett, T.A.; Meyer, A.L; Brown, E.L.; Huber, B.T. *Proc. Natl. Acad. Sci. USA,* **2004**, *101*, 1303-1308.

[30] Roitt, I.; Brostoff, J.; Male, D. *Immunology,* Mosby: London, Philadelphia, Sydney, Tokio, **1998**.

[31] Kim, J.V.; Latouche, J.-B.; Riviere, I.; Sadelain, M. *Nat. Biotechnol.,* **2004**, *22*, 403-410.

[32] Grandi, G. *Trends Biotechnol.*, **2001**, *19*, 181-188.

[33] Mahanty, S.; Saul, A.; Miller, L.H. *J. Exp. Biol.*, **2003,** *206*, 3781-3788.

[34] Mount, D.W. *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, **2001**.

[35] Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. *Nucleic Acids Res.*, **1997**, *25*, 3389-3402.

[36] McAtee, C.P.; Lim, M.Y.; Fung, K.; Velligan, M.; Fry, K.; Chow, T.P.; Berg, D.E. *J. Chromatogr. B. Biomed. Sci. Appl.*, **1998**, *714*, 325-333.

[37] Utt, M.; Nilsson, I.; Ljungh, A.; Wadstrom, T. *J. Immunol. Methods*, **2002**, *259*, 1-10.

[38] Purcell, A.W.; Gorman, J.J. *Mol. Cell Proteomics*, **2004**, *3*, 193-208.

[39] Falquet, L.; Pagni, M.; Bucher, P.; Hulo, N.; Sigrist, C.J.; Hofmann, K.; Bairoch, A. *Nucleic Acids Res.*, **2002**, *30*, 235-238.

[40] Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.; Studholme, D.J.; Yeats, C.; Eddy, S.R. *Nucleic Acids Res.*, **2004**, *32 Database issue*, D138-D141.

[41] Letunic, I.; Copley, R.R.; Schmidt, S.; Ciccarelli, F.D.; Doerks, T.; Schultz, J.; Ponting, C.P.; Bork, P. *Nucleic Acids Res.,* **2004**, *32 Database issue*, D142-D144.

[42] Nielsen, H.; Brunak, S.; von Heijne, G. *Protein Eng.*, **1999**, *12*, 3-9.

[43] Bendtsen, J.D.; Nielsen, H.; von Heijne, G; Brunak, S. *J. Mol. Biol.,* **2004**, *340*, 783-795.

[44] Hayashi, S.; Wu, H.C. *J. Bioenerg. Biomembr.*, **1990**, *22*, 451-471.

[45] Fariselli, P.; Finocchiaro, G.; Casadio, R. *Boinformatics,* **2003**, *19*, 2498-2499.

[46] Maurer-Stroh, S.; Eseinhaber, F. *Trends Microbiol.,* **2004**, *12*, 178-179.

[47] Maurer-Stroh, S.; Eseinhaber, B.; Eseinhaber, F. *J. Mol. Biol.,* **2001**, *317*, 541-557.

[48] Eisenhaber, B.; Bork, P.; Eisenhaber, F. *J. Mol. Biol.,* **1999**, *292*, 741-758.

[49] Acosta-Serrano, A.; Almeida, I.C.; Freitas-Junor, L.H.; Yoshida, N.; Shemckman, S. *Mol. Biochem. Parasitol.*, **2001**, *114*, 143-150.

[50] Naderer T.; Vince J.E., McConville M.J. *Curr. Mol. Med.,* **2004**, *4*, 649-665.

[51] van Hoven, N.S.; Miller, A.D. *J. Virol.*, **2005**, *79*, 87-94.

[52] Vigdorovich, V.; Strong R.K.; Miller, A.D. *J. Virol.*, **2005**, *79*, 79-86.

[53] van der Ley, P.; Heckels, J.E.; Virji, M.; Hoogerhout, P.; Poolman, J.T. *Infect. Immun.*, **1991**, *59*, 2963-2971.

[54] Martelli, P.L.; Fariselli, P.; Krogh, A.; Casadio, R. *Bioinformatics,* **2002**, *18*, 546-553.

[55] Bagos, P.; Liakopoulos, T.D.; Spyropoulos, I.C.; Hamodrakas, S.J. *BMC Bioinformatics,* **2004**, *5*, 29.

[56] Bigelow, H.R.; Petrey, D.S.; Liu, J.; Przybylski, D.; Rost, B. *Nucleic Acids Res.,* **2004**, *32*, 2566-2577.

[57] Natt, N.K.; Kaur, H.; Raghava, G.P. *Proteins,* **2004**, *56*, 11-18.

[58] Berven, F.S.; Flikka, K.; Jensen H.B.; Eidhammer, I. *Nucleic Acids Res.,* **2004**, *32 Web Server Issue*, W394-W399.

[59] Gavel, Y.; von Heijne, G. *Protein Eng.*, **1990**, *3*, 433-442.

[60] Yan, B.; Zhang, W.; Ding J.; Gao, P. *J. Protein Chem.*, **1999**, *18*, 511-521.

[61] Kyte, J.; Doolittle, R. *J. Mol. Biol.*, **1982**, *157*, 105-132.

[62] Krogh, A.; Larrson, B.; von Hejne, G. *J. Mol., Biol.*, **2001**, *395*, 567-580.

[63] Tusnady, G.E.; Simon, I. *Bioinformatics,* **2001**, *17*, 849-850.

[64] Grezko, M.; Eisenhaber, F.; Eisenhaber, B.; Simon, I. *Bioinformatics,* **2004**, *20*, 136-137.

[65] Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Ketchum, K.A.; Hood, D.W.; Peden, J.F.; Dodson, R.J.; Nelson, W.C.; Gwinn, M.L.; DeBoy, R.; Peterson, J.D.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Mason, T.; Ciecko, A.; Parksey, D.S.; Blair, E.; Cittone, H.; Clark, E.B.; Cotton, M.D.; Utterback, T.R.; Khouri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.; Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, J.C. *Science*, **2000**, *287*, 1809-1815.

[66] Pizza, M.; Scarlato, V.; Masignani, V.; Giuliani, M.M.; Arico, B.; Comanducci, M.; Jennings, G.T.; Baldi, L.; Bartolini, E.; Capecchi, B.; Galeotti, C.L.; Luzzi, E.; Manetti, R.; Marchetti, E.; Mora, M.; Nuti, S.; Ratti, G.; Santini, L.; Savino, S.; Scarselli, M.; Storni, E.; Zuo, P.; Broeker, M.; Hundt, E.; Knapp, B.; Blair, E.; Mason, T.; Tettelin, H.; Hood, D.W.; Jeffries, A.C.; Saunders, N.J.; Granoff, D.M.; Venter, J.C.; Moxon, E.R.; Grandi, G.; Rappuoli, R. *Science*, **2000**, *287*, 1816-1820.

[67] Drabick, J.J.; Brandt, B.L.; Moran, E.E.; Saunders, N.B.; Shoemaker, D.R.; Zollinger, W.D. *Vaccine*, **1999**, *18*, 160-172.

[68] Katial, R.K.; Brandt, B.L.; Moran, E.E.; Marks, S.; Agnello, V.; Zollinger, W.D. *Infect. Immun.*, **2002**, *70*, 702-707.

[69] Poolman, J.T.; Kriz-Kuzemenska, P.; Ashton, F.; Bibb, W.; Dankert, J.; Demina, A.; Froholm, L.O.; Hassan-King, M.; Jones, D.M.; Lind, I.; Prakash, K.; Xujing, H. *Clin. Diagn. Lab. Immunol.*, **1995**, *2*, 69-72.

[70] Peterson, J.D.; Umayam, L.A.; Dickinson, T.; Hickey, E.K.; White, O. *Nucleic Acids Res.*, **2001**, *29*, 123-125.

[71] Roccasecca, R.; Folgori, A.; Ercole, B.B.; Puntoriero, G.; Lahm, A.; Zucchelli, S.; Tafi, R.; Pezzanera, M.; Galfre, G.; Tramontano, A.; Mondelli, M.U.; Pessi, A.; Nicosia, A.; Cortese, R.; Meola, A. *Int. Rev. Immunol.*, **2001**, *20*, 289-300.

[72] Olenina, L.V.; Kuzmina, T.I.; Sobolev, B.N.; Kuraeva, T.E.; Kolesanova, E.F.; Archakov, A.I. *J. Viral. Hepat.* **2005**, *In press*.

[73] Montgomery, D.L. *Brief Bioinform.*, **2000**, *1*, 289-296.

[74] Louise, R.; Skjot, V.; Agger, E.M.; Andersen, P. *Scand. J. Infect. Dis.*, **2001**, *33*, 643-647.

[75] Pennisi, E. *Science*, **1998**, *281*, 324-325.

[76] Wang, R.; Doolan, D.L.; Le, T.P.; Hedstrom, R.C.; Coonan, K.M.; Charoenvit, Y.; Jones, T.R.; Hobart, P.; Margalith, M.; Ng, J.; Weiss, W.R.; Sedegah, M.; de Taisne, C.; Norman, J.A.; Hoffman, S.L. *Science*, **1998**, *282*, 476-480.

[77] Hoffman, S.L.; Rogers, W.O.; Carucci, D.J.; Venter, J.C. *Nat. Med.*, **1998**, *4*, 1351-1353.

[78] Stephens, R.S. *J. Infect. Dis.*, **2000**, *181* Suppl. 3, S521-S523.
[79] Igietseme, J.U.; Black, C.M.; Caldwell H.D. *BioDrugs*, **2002**, *16*, 19-35.
[80] Pellequer, J.-L.; Westhof, E.; van Regenmortel, M.H.V. In *Peptide antigens. A practical approach;* G.B. Wisdom, Ed.; IRL PRESS, Oxford University Press: Oxford, New York, Tokyo, **1994**; pp. 7-25.
[81] Stites, W.E. *Chem. Rev.*, **1997**, *97*, 1233-1250.
[82] Jin, L.; Wells, J.A. *Protein Sci.*, **1994**, *3*, 2351-2357.
[83] Janin, J.; Chothia, C. *J. Biol. Chem.*, **1990**, *265*, 16027–16030.
[84] van Regenmortel M.H.V.; Muller, S. *Synthetic peptides as antigens*, Elsevier: Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore-Tokyo, **1999**.
[85] Hopp, T.P.; Woods, K.R. *Proc. Natl. Acad. Sci. USA,* **1981**, *78*, 3824-3828.
[86] Nozaki, Y.; Tanford, C. *J. Biol. Chem.*, **1971**, *246*, 2211-2227.
[87] Bull, H.B.; Breese, K. *Arch. Biochem. Biophys.*, **1974**, *161*, 665-670.
[88] Parker, J.M; Guo, D.; Hodges, R.S. *Biochemistry*, **1986**, *25*, 5425-5432.
[89] Karplus, P.A.; Schulz, G.E. *Naturwissenschaften*, **1985**, *72*, 212-213.
[90] Hopp ,T.P. *Ann. Sclavo. Collana. Monogr.*, **1984**, *1*, 47-60.
[91] Lee, B.; Richards, F.M. *J. Mol. Biol.*, **1971**, *55*, 379-400.
[92] Welling, G.W.; Weijer, W.J.; van der Zee R.; Welling-Wester, S. *FEBS Lett.*, **1985**, *188*, 215-218.
[93] Chou, P.Y; Fasman, G.D. *Biophys. J.*, **1979**, *26*, 367-373.
[94] Odorico, M.; Pellequer, J.-L. *J. Mol. Recognit.*, **2003**, *16*, 20–22.
[95] Carter, J.M. *Methods Mol. Biol.*, **1994**, *36*, 207-223.
[96] Rost, B.; Sander, C. *Proteins*, **1994**, *20*, 216-226.
[97] Kopp, J.; Schwede, T. *Nucleic. Acids Res.*, **2004**, *32 Database issue*, D230-D234.
[98] Kolaskar, A.S.; Kulkarni-Kale, U. *Virology*, **1999**, *261*, 31-42.
[99] Rammensee, H.; Bachmann, J.; Emmerich, N.P.; Bachor, O.A.; Stevanovic, S. *Immunogenetics*, **1999**, *50*, 213-219.
[100] Dafforn, T. R.; Lesk, A.M. 2000 In *Protein-Protein Recognition;* C. Kleanthous, Ed.; Oxford University Press: Oxford, London, **2000**; pp. 163-188.
[101] Schonbach, C.; Koh, J.L.; Flower, D.R.; Wong, L.; Brusic, V. *Nucleic Acids Res.*, **2002**, *30*, 226-229.
[102] Jung, G.; Fleckenstein, B.; von der Mulbe, F.; Wessels, J.; Niethammer, D.; Wiesmuller, K.H. *Biologicals*, **2001**, *29*, 179-181.
[103] Bhasin, M.; Raghava, G.P. *Vaccine*, **2004**, *22*, 3195-3204.
[104] Mamitsuka, H. *Proteins,* **1998***, 33*, 460-474
[105] Yu, K., Petrovsky, N.; Schonbach, C.; Koh, J.Y.; Brusic V. *Mol. Med.,* **2002**, *8*, 137-148.
[106] Guan, P.; Doytchinova, I.A.; Zygori, C.; Flower, D.R. *Nucleic Acids Res.,* **2003**, *13*, 3621-3624.
[107] Doytchinova, I.A.; Flower, D.R. *Bioinformatics*, **2003**, *17*, 2263-2270.
[108] McSparron, H.; Blythe, M.J.; Zygouri, C.; Doytchinova, I.A.; Flower, D.R. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1276-1287.
[109] Doytchinova, I.A.; Walshe, V.A.; Jones, N.A.; Gloster, S.E.; Borrow, P.; Flower, D.R. *J. Immunol.,* **2004**, *172*, 7495-7502.
[110] Sung, M.H.; Zhao, Y.; Martin, R.; Simon, R. *J. Comput. Biol.*, **2002**, *9*, 527-539.
[111] Liu, R.; Enstrom, A.M.; Lam, K.S. *Exp. Hematol.*, **2003**, *31*, 11-30.
[112] Dogan, I.; Dorgham, K.; Chang, H.C.; Parizot, C.; Lemaitre, F.; Ferradini, L.; Reinherz, E.L.; Debre, P.; Gorochov, G. *Eur. J. Immunol.,* **2004**, *34*, 598-607.
[113] Nussbaum, A.K.; Kuttler, C.; Hadeler, K.P.; Rammensee, H.G.; Schild, H. *Immunogenetics*, **2001**, *53*, 87-94.
[114] Stoltze, L.; Schirle, M.; Schwarz, G.; Schroter C.; Thompson, M.W.; Hersh, LB.; Kalbacher, H.; Stevanovic, S.; Rammensee, H.G.; Schild, H. *Nat. Immunol.*, **2000**, *1*, 413-418.
[115] Brusic, V.; van Endert, P.; Zeleznikow, J.; Daniel S.; Hammer, J.; Petrovsky N. *In Silico Biol.*, **1999**, *1*, 109-121.
[116] Logean, A; Rognan, D. J. *Comput. Aided Mol. Des.,* **2002**, *16*, 229-243.
[117] Rognan, D.; Lauemoller, S.L.; Holm, A.; Buus, S.; Tschinke, V. *J. Med. Chem.*, **1999**, *42*, 4650-4658.
[118] Zhao, B.; Mathura, V.S.; Rajaseger, G.; Moochhala, S.; Sakharkar, M.K.; Kangueane, P. *Hum Immunol.,* **2003**, *64*, 1123-1143.
[119] Bodmer, W. *Ciba Found. Symp.*, **1996**, *197*, 233-253.
[120] Lairmore, M.D.; DiGeorge, A.M.; Conrad, S.F.; Trevino, A.V.; Lal, R.B.; Kaumaya, P.T. *J. Virol.*, **1995**, *69*, 6077-6089.
[121] Jackson, D.C.; Purcell, A.W.; Fitzmaurice, C.J.; Zeng, W.; Hart, D.N. *Curr. Drug Targets*, **2002,** *3*, 175-196.
[122] Taneja, V.; David, C.S. *J. Clin. Invest.*, **1998**, *101*, 921-926.
[123] Taneja V.; David, C.S. *Immunol. Rev.*, **1999**, *169*, 67-79.
[124] Panina-Bordignon, P.; Tan, A.; Termijtelen, A.; Demotz, S.; Corradin, G.; Lanzavecchia, A. *Eur. J. Immunol.*, **1989**, *19*, 2237-2242
[125] Contreras, C.E.; Ploton, I.N.; Siliciano, R.F.; Karp, C.L.; Viscidi, R.; Kumar, N. *Infect. Immun.*, **1998**, *66*, 3579-3590.
[126] Caro-Aguilar, I.; Rodriguez, A.; Calvo-Calle, J.M.; Guzman, F.; De la Vega, P.; Patarroyo, M.E.; Galinski, M.R.; Moreno, A. *Infect. Immun.*, **2002**, *70*, 3479-3492.
[127] Panigada, M.; Sturniolo, T.; Besozzi, G.; Boccieri, M.G.; Sinigaglia, F.; Grassi, G.G.; Grassi, F. *Infect. Immun.*, **2002**, *70*, 79-85.
[128] Sarobe, P; Pendleton, C.D.; Akatsuka, T.; Lau, D.; Engelhard, V.H.; Feinstone, S.M.; Berzofsky, J.A. *J. Clin. Invest.*, **1998**, *102*, 1239-1248.
[129] Hilleman, M.R. *Proc. Natl. Acad. Sci. USA*, **2004**, *101* Suppl. 2, 14560-14566.
[130] Kato, N.; Ootsuyama, Y.; Tanaka, T.; Nakagawa, M; Nakazawa, T.; Muraiso, K.; Ohkoshi, S.; Hijikata, M.; Shimotohno, K. *Virus Res.*, **1992**, *22*, 107-123.
[131] Hoffman N.G.; Seillier-Moiseiwitsch F.; Ahn, J.; Walker, J.M.; Swanstrom, R. *J. Virol.*, **2002**, *76*, 3852-3864.
[132] Farci, P.; Shimoda, A.; Wong, D.; Cabezon, T.; De Gioannis, D.; Strazzera, A.; Shimizu, Y.; Shapiro, M.; Alter, H.J.; Purcell, R.H. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 15394-15399.
[133] Sobolev, B.N.; Poroikov, V.V.; Olenina, L.V.; Kolesanova E.F.; Archakov, A.I. *J. Viral Hepat*., **2000**, *7*, 368-374.
[134] Olenina, L.V.; Nikolaeva, L.I.; Sobolev, B.N.; Blokhina, N.P.; Archakov, A.I.; Kolesanova, E.F. *J. Viral. Hepat.*, **2002**, *9*, 174-182.
[135] Patarroyo, G.; Franco, L.; Amador, R.; Murillo, L.A.; Rocha, C.L.; Rojas, M.; Patarroyo, M.E. *Vaccine*, **1992**, *10*, 175-178.
[136] Tam J.P. In: *Peptide Antigens. A Practical Approach;* Wisdom, G.B. Ed.; IRL PRESS, Oxford University Press: Oxford, New York, Tokyo, **1994**; pp. 83-115.
[137] Gregersen, J.P. *Naturwissenschaften*, **2001**, *88*, 504-513.
[138] Streatfield, S.J.; Lane, J.R.; Brooks, C.A.; Barker, D.K.; Poage, M.L.; Mayor, J.M.; Lamphear, B.J.; Drees, C.F.; Jilka, J.M.; Hood, E.E.; Howard, J.A. *Vaccine*, **2003**, *30*, 812-815.
[139] Haq, T.A.; Mason, H.S.; Clements, J.D.; Arntzen, C.J. *Science*, **1995**, *268*, 714-716.
[140] Mason, H.S.; Haq, T.A.; Clements, J.D.; Arntzen, C.J. *Vaccine*, **1998**, *16*, 1336-1343.
[141] Tacket, C.O.; Mason, H.S.; Losonsky, G.; Clements, J.D.; Levine, M.M.; Arntzen, C.J. *Nat. Med.*, **1998**, *4*, 607-609.
[142] Arakawa, T.; Chong, D.K.; Langridge, W.H. *Nat. Biotechnol.*, **1998**, *16*, 292-297.
[143] Richter, L.J.; Thanavala, Y.; Arntzen, C.J.; Mason, H.S. *Nat. Biotechnol.*, **2000**, *18*, 1167-1171.
[144] Pharmaprojects 2002, PJB Publications Ltd.
[145] Tackaberry, E.S., Dudani, A.K., Prior, F.; Tocchi, M.; Sardana, R.; Altosaar, I.; Ganz, P.R. *Vaccine*, **1999**, *17*, 3020-3029.
[146] Clough, J. *Drug Discov. Today*, **2002**, *7*, 886-887.
[147] Bonetta, L., *Nat. Med.*, **2002**, *8*, 94.
[148] Gotsman, I.; Alper, R.; Klein, A.; Rabbani E.; Engelhardt, D.; Ilan Y. *Cancer*, **2002**, *94*, 406-414.
[149] Tacket, C.O.; Mason, H.S. *Microbes Infect.*, **1999**, *1*, 777-783.
[150] Lefranc, M.P. *Methods Mol. Biol.*, **2004**, *248*, 27-49.
[151] Zdobnov, E.M.; Lopez R.; Apweiler, R.; Etzold, T. *Bioinformatics*, **2002**, *18*, 1149-1150.
[152] Wheeler, D.L.; Churc, D.M.; Edgar, R.; Federhen, S.; Helmberg, W.; Madden, T.L.; Pontius, J.U.; Schuler.; G.D.; Schriml, L.M.; Sequeira, E.; Suzek, T.O.; Tatusova, T.A.; Wagner L. *Nucleic Acids Res.*, **2004**, *32 Database issue*, D35-D40.
[153] Mewes, H.W.; Amid, C.; Arnold, R.; Frishman, D.; Guldener U.; Mannhaupt, G.; Munsterkotter, M.; Pagel, P.; Strack, N.; Stumpflen, V.; Warfsmann, J.; Ruepp, A. *Nucleic Acids Res.*, **2004**, *32 Database issue*, D41-D44.
[154] Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M.J.; Natale, D.A.; O'Donovan, C.; Redaschi, N.; Yeh, L.S. *Nucleic Acids Res.*, **2004**, *32 Database issue*, D115-D119.

[155] Kato-Maeda, M.; Bifani, P.J.; Kreiswirth, B.N.; Small, P.M. *J. Clin. Invest.*, **2001**, *107*, 533-537.
[156] Lollini, P.L.; Forni, G. *Trends Immunol.*, **2003**, *24*, 62-66.
[157] Davis, I.D.; Jefford, M.; Parente, P.; Cebon, J. *J. Leukoc. Biol.*, **2003**, *73*, 3-29.
[158] Meng, W.S.; Butterfield, L.H. *Pharm. Res.*, **2002**, *19*, 926-932.
[159] Espinoza-Delgado, I. *Oncologist*, **2002**, *7* Suppl. 3, 20-33.

Received: November 3, 2004 **Accepted: January 4, 2005**