13

Genome Based Vaccines Against Parasites

Yasser Shahein[1] and Amira Abouelella[2]
[1]*Department of Molecular Biology, National Research Centre,*
[2]*Department of Biochemistry, National Centre for Radiation Research and Technology*
Egypt

1. Introduction

In its original concept, vaccination aims to mimic the development of naturally acquired immunity by inoculation of nonpathogenic but still immunogenic components of the pathogen in question, or closely related organisms. Vaccination is one of the most effective tools allowing near or complete eradication of fatal diseases. In recent times, vaccination was based on conventional approaches that were successful in several diseases but they require the pathogen to be grown in laboratory conditions, are time-consuming and allow for the identification of only the most abundant antigens, which can be purified in quantities suitable for vaccine testing. The conventional approach to vaccine development uses two methods: first, attenuation of pathogens by serial passages *in vitro* to obtain live-attenuated strains to be used as vaccines, and second, identification of protective antigens to be used in non-living, subunit vaccines.

Using conventional vaccinology, vaccines targeting pathogens with low antigenic variability and those for which protection depends on antibody mediated immunity like polio, MMR, Tetanus, Influenza, and Diptheria, had been licensed. Furthermore, if we are dealing with non-cultivatable microorganisms or have high antigenic variability or need T cell-dependent induced immunity like MenB, TB, HIV, Hepatitis B & C, Malaria and parasite diseases, there is no advance to development a specific vaccine.

With the development of the DNA sequencing technologies in the 1970s till recent years, we observed the completion of the sequencing of the genome of humans, a number of invertebrate species including *Drosophila melanogaster* and *Caenorhabditis elegans* and expanding number of microbial pathogens of medical and economic importance (Dalton et al., 2003). Expansion of these sequence information developed new approaches to decode, analyze and share these massive genetic data. The most widely used term for these approaches is the genomics.

Genomics, studying the genome of organisms as a whole, and postgenomics technologies including investigating RNA (transcriptomics), proteins (proteomics), identification of immunogenic proteins (immunomics) and metabolites (metabolomics), have had a considerable impact in all areas of biological research (Bambini and Rappuoli, 2009), and the field of vaccinology is no exception. There are many examples using these approaches to develop vaccines but this chapter will focus on the possible genome based approaches to confer vaccine developed immunity against two examples of endo- and ecto-parasites such as malaria and ticks, respectively.

2. Reverse vaccinology

Genome sequencing is a powerful tool for understanding and controlling infectious pathogens. Using this technology, researchers can identify target genes for drug discovery and reveal small genetic variations between strains of a specific organism to define its virulence and improve the method of control. The approach of reverse vaccinology uses the genome sequences of viral, bacterial or parasitic pathogens of interest as starting material for the identification of novel antigens, whose activity should be subsequently confirmed by experimental biology (Fig. 1) (Rappuoli, 2001). One of the earlier applications of genomics to vaccinology (reverse vaccinology) had been the identification of vaccine candidates against serogroup B meningococcus by the completion of the whole-genome sequencing (Pizza et al., 2000). They had cloned the open reading frames (ORFs) that encode putative virulence factors and surface-localized proteins of meningococcus. Several hundred ORFs (350 surface-exposed protein coding frames) were cloned into expression vectors, purified and used to immunize mice. The antibodies binding properties to the products of ORF were analyzed using fluorescent activated cell sorter (FACS) analyses and Enzyme linked immunosorbent assay (ELISA). The primary vaccine candidates were then tested *in vitro* and/or animal models to provide an insight on the protective efficacy. Twenty nine of these surface-exposed proteins were found to be bactericidal. The selected candidates were then checked for sequence conservation across a panel of strains representing the genetic diversity of meningococcus allowing further selection of antigens capable of eliciting cross-bactericidal response against the majority of strains included in the panel.

Reverse vaccinology is based on the high throughput analyses of genome sequences. With continuous flow of new genomic sequence and functional annotation data from different taxonomic lineages permits scientists to confine correlations depending on the wide range of data bases, enabling the design of more reliable analytical and predictive tools. One of the most important tools is the alignment of multiple homologous sequences that permitted the identification of large number of structural and functional signatures including ligand-binding sites, sorting signals, protein domain profile, different motifs with catalytic sites and more (Vivona et al., 2008). This has generated several integrated databases containing combined sets of such signatures and of scanning tools capable of inferring possible functions or regulatory mechanisms from the presence of either canonical or degenerate signatures in a sequence of interest. Such 'sequence-to-function' approaches (Oliver, 1996) have enhanced the production of further functional evidence by prediction-driven experiments. Early examples included using the identification of specific signatures in a sequence to suggest functional assays able to identify otherwise long-sought functions (Filippini et al., 1996). They had shown that a plant oncogene *rolB* from *Agrobacterium rhizogenes*, induces differentiation and growth of neoplastic roots (Hairy-roots) in dicotyledonous plants. In addition, the prediction-driven experiments may imply functions for disease gene products (Emes and Ponting, 2001, Vacca et al., 2001). Later, more complex, genome-wide analyses have led to the identification of proteomic complements that underlie regulatory pathways or interaction network organization in model organisms (Carpi et al., 2002, Li et al., 2006).

More recently, bioinformatic approaches are used to uncover functional information and enable researchers to tackle biological and biotechnological problems that require the integration of diverse strategies of both *in silico* (on computer) and experimental evidences. Besides data analyses, a variety of algorithmic approaches have been used to develop novel

tools. The functional potential of these *in silico* approaches has found its pattern in reverse vaccinology.

Reverse vaccinology presents a revolution in both immunology and biotechnology and shows how a biological problem like designing a vaccine could be solved by applying integrating tools. However, reverse vaccinology presents a huge advance compared to the conventional vaccine production protocols. It takes advantage of the growing number of genome sequences available for many organisms. The approach uses computer analysis of the genomic sequence to predict suitable candidate vaccine molecules. Unfortunately, the approach does not provide certain evidence that the selected antigens are either immunogenic or protective. On the contrary, the approach permits the identification of novel protein antigens besides the antigens discovered by the traditional protocols.

3. Malaria vaccine research

*M*alaria is caused by parasites of the genus *Plasmodium* of which there are four species that infects humans, *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. *P. falciparum* is the most fatal and common causative species of malaria in humans. The parasite originally develops in the gut of an infected *Anopheles* mosquito and is passed into humans in the saliva of infected mosquitoes. Sporozoites; the pre-erythrocytic stage is transported to the liver in the blood or lymph of the human. In the liver, the parasite invades and replicates in hepatocytes and after one to two weeks, the sporozoite become mature into thousands of merozoites which penetrate the circulating red cells. The replication of the merozoites in blood cells causes rapid destruction of blood cells leading to fever and anemia. In the past years, antimalarial drugs were the first line of defense in endemic areas but unfortunately, there is increasing resistance to these drugs (Wykes and Good, 2007).

The 500 million new infections each year and 2.5 million annual deaths indicated that all measures used so far to control the disease have failed (WHO, 1997, Chaudhuri et al., 2008). This world malaria situation had directed researchers to conduct other strategies including the identification of only few dozen of malarial proteins (Duffy et al., 2005) by recombinant vaccine technology and more than 40 clinical trials were carried out but no vaccine has provided a strong and lasting immunity.

The accessibility of complete genome sequences of *Plasmodium falciparum* (Gardner et al., 2002), *P. yoelii* (Carlton et al., 2002) and *P. vivax* has provided new opportunity for applying the principles of reverse vaccinology. Reverse vaccinology uses bioinformatics in the initial steps to identify potential antigens, which are consequently examined for their effectiveness and toxicity. The use of algorithm for prediction of subcellular location improved the power of identifying potential vaccine candidates (Serruto and Rappuoli, 2006). Subsequently, developments have been proposed to reverse vaccinology by suggesting the use of additional algorithms to find probability of being an adhesin, of topology (transmembrane domains) and to find similarity with host protein (Vivona et al., 2006). Recently, integrative approaches are proposed for reverse vaccinology by including prediction of multiple features of proteins such as signal peptides, membrane spanning regions, functional motifs and differences in amino acid composition unique to specific cellular compartments (Vivona et al., 2008). Implementing this approach, the following predictions were included:

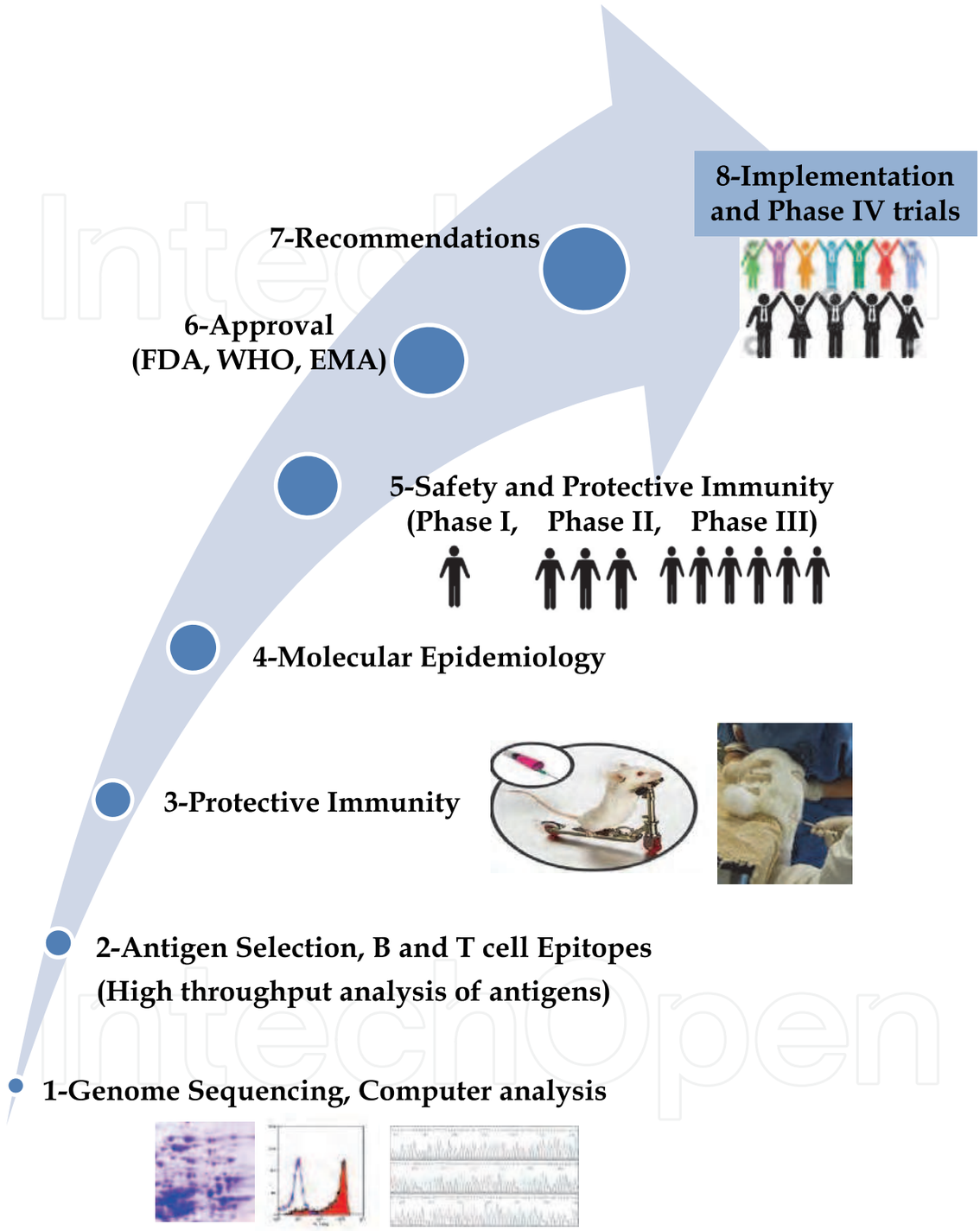


Fig. 1. Diagrammatic Illustration Summarizing the Vaccine Development Pathway Starting from Reverse Vaccinology. (1) First, computer analysis of the entire genome identifies the genes coding for predicted antigens and removes those antigens with homologies to human proteins. (2) Second, these identified antigens are screened for expression by the pathogen and for immunogenicity during infection. (3) Then the selected antigens are used to immunize

experimental animals and examine the protective efficacy of immunization. (4) The presence and conservation of protective antigens in a collection of strains representative of the species (molecular epidemiology), is examined. (5) Finally, large scale production of target antigens are manufactured for clinical trials and candidate vaccines are examined for their safety and protective immunity in humans. (6) Authorized agencies, such as the Food and Drug Administration (FDA), World Health Organization (WHO) or the European Medicinal Agency (EMA) analyze and approve the scientific, clinical, and technical information. (7) Policy-making organizations, such as the Advisory Committee on Immunization Practices (ACIP) and equivalent bodies from other nations, prepare the recommendation on how the vaccine should be used. (8) In this stage, the approved vaccine is then commercialized and used in large scale. At this point, phase IV clinical studies confirm safety.

3.1 Adhesins

One of the preventive approaches targets adhesion of parasites to host cells and tissues. Adhesion of parasites is mediated by proteins called adhesins. Abrogation of adhesion by either immunizing the host with adhesins or inhibiting the interaction using structural analogs of host cell receptors holds the potential to develop novel preventive strategies. The availability of complete genome sequence offers new opportunities for identifying adhesin and adhesin-like proteins. A nonhomology-based approach using 420 compositional properties of amino acid dipeptide and multiplet frequencies was used to develop Malarial Adhesins and Adhesin-like proteins Predictor (MAAP) Web server with Support Vector Machine (SVM) model classifier as its engine for the prediction of malarial adhesins and adhesin-like proteins. Several new predictions were obtained. This list includes hypothetical protein PF14 0040, interspersed repeat antigen, STEVOR, liver stage antigen, SURFIN, RIFIN, stevor (3D7-stevorT3-2), mature parasite-infected erythrocyte surface antigen or *P. falciparum* erythrocyte membrane protein 2, merozoite surface protein 6 in *P. falciparum*, circumsporozoite proteins, microneme protein-1, Vir18, Vir12-like, Vir12, Vir18-like, Vir18-related and Vir4 in *P. vivax*, circumsporozoite protein/thrombospondin related anonymous proteins, 28 kDa ookinete surface protein, yir1, and yir4 of *P. yoelii* (Ansari et al., 2008).

3.2 Paralogs

It means gene duplication, which have a major role in the evolution of new biological functions. Theoretical studies often assume that a duplication *per se* is selectively neutral and that, following a duplication, one of the gene copies is freed from purifying (stabilizing) selection, which creates the potential for evolution of a new function **(**Kondrashov et al., 2002).

3.3 Transmembrane topologies

Krogh et al. (2001), predicted a membrane protein topology prediction method; transmembrane protein topology with a hidden Markov model (TMHMM). The model is based on a hidden Markov model showing that it correctly predicts 97-98 % of the transmembrane helices. Additionally, TMHMM can discriminate between soluble and membrane proteins with both specificity and sensitivity better than 99 %. The method shows that proteins with N(in)-C(in) topologies are strongly preferred in all examined organisms, except *Caenorhabditis elegans*, where the large number of 7TM receptors increases the counts for N(out)-C(in) topologies.

3.4 β-helix supersecondary structural motifs

A program named BETAWRAP (http://theory.lcs.mit.edu/betawrap) implements the prediction of parallel beta -helices from primary sequences and recognizes each of the seven known parallel beta-helix families, when trained on the known parallel beta-helices from outside that family. BETAWRAP identifies 2,448 sequences among 595,890 screened from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) nonredundant protein database as likely parallel beta-helices. It identifies surprisingly many bacterial and fungal protein sequences that play a role in human infectious disease; these include toxins, virulence factors, adhesins, and surface proteins of Chlamydia, Helicobacteria, Bordetella, Leishmania, Borrelia, Rickettsia, Neisseria, and *Bacillus anthracis*.

3.5 Subcellular localization

A neural network-based tool, TargetP, for large-scale subcellular location prediction of newly identified proteins has been developed. Using N-terminal sequence information only, it discriminates between proteins destined for the mitochondrion, the chloroplast, the secretory pathway, and "other" localizations with a success rate of 85% (plant) or 90% (nonplant) on redundancy-reduced test sets. TargetP is available as a web-server at http://www.cbs.dtu.dk/services/TargetP/ (Emanuelsson et al., 2000).

3.6 Similarity against human proteins

Using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990).

3.7 Antigenic regions

Analysis of data from experimentally determined antigenic sites on proteins reveal that the hydrophobic residues Cys, Leu and Val, if they occur on the surface of a protein, are more likely to be a part of antigenic sites. The semi-empirical method which makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins. Application of this method to a large number of proteins has shown that the method can predict antigenic determinants with about 75% accuracy which is better than most of the known methods (Kolaskar and Tongaonkar, 1990).

3.8 Conserved domains

The conserved domain search results for protein sequences in Entrez are pre-computed to provide links between proteins and domain models, and computational annotation visible upon request. Protein–protein queries submitted to the BLAST search service are scanned for the presence of conserved domains by default (Marchler-Bauer et al., 2005).

3.9 Epitopes

The identification of B-cell and T-cell epitopes is a crucial step in peptide vaccine design. The experimental scanning of B-cell epitope active regions requires the synthesis of overlapping peptides, which span the entire sequence of a protein antigen. This process is time consuming and costly. The *in silico* approaches are the alternative procedures to figure out target regions of a protein with possible immunogenic capacity. In this respect, two new prediction tools; BcePred and ABCpred servers were developed. BcePred can predict

continuous B-cell epitopes, and physic-chemical scales used were hydrophilicity, flexibility/mobility, accessibility, polarity, exposed surface, turns and antigenicity. The ABCpred was developed to predict continuous or linear B-cell epitopes obtained from Bcipep database. This server is based on machine learning techniques (Saha and Raghava, 2007).

3.10 Allergens

The prediction of allergenic proteins is becoming very important due to the use of modified proteins in foods (genetically modified foods), therapeutics and bio-pharmaceuticals. The protein is considered allergen if it has one or more IgE epitopes. This kind of prediction can be achieved using the prediction tool created by Saha S, Raghava (2006). The predition tool is based on various approaches. First, a standard method for predicting allergens based on amino acid and dipeptide composition of proteins using support vector machine (SVM). In the second approach, motif-based technique has been used for predicting allergens using the software MEME/MAST. Third, they assigned a protein as an allergen, if it has a segment similar to allergen representative proteins (ARPs).

4. Data bases and predictive tools in malarial diseases

A community resource database; MalVac was created. The database is based on data analysis of available proteins including 161 adhesin proteins from *P. falciparum*, 137 adhesin proteins from *P. vivax* and 34 adhesin proteins from *P. yoelii*. The ORF identification tags (ORF ID) assigned to proteins of malaria parasites as given in PlasmoDB 5.4 which is is the official database of the *Plasmodium falciparum* genome sequencing consortium, were used as primary keys. The database was developed using MySQL version 4.1.20 at back end and operated in Red Hat Enterprise Linux ES release 4. The web interfaces have been developed in HTML and PHP 5.1.4, which dynamically execute the MySQL queries to fetch the stored data and is run through Apache2 server. The frame of MalVac consists of four basic concepts as follows:

4.1 Motif and topology: includes the transmembrane helices and right handed parallel bata helices.

4.2. Location: the position of the signal peptides and subcellular localization of the protein.

4.3 Immunoinformatics: includes T cell epitopes (MHC class I and II), B cell epitopes (conformationaland linear epitopes), Allergens and antigenic regions.

4.4 Homology: includes the paralogs, orthologs, conserved domains and similarity to host proteins.

The first step towards MalVac database creation is the collection of known vaccine candidates and a set of predicted vaccine candidates identified from the whole proteome sequences of Plasmodium species provided by PlasmoDB 5.4 release. These predicted vaccine candidates are the adhesins and adhesin-like proteins from Plasmodium species, *P. falciparum*, *P. vivax* and *P. yoelii* using MAAP server. Subsequently these protein sequences will be analysed with 20 algorithms important from the view of reverse vaccinology (Table 1).

Algorithm	**Principle Role in MalVac data base**
1. MAAP	Predicts Malarial adhesins and adhesins-like proteins.
2. BLASTCLUST	Clusters protein or DNA sequences based on pair wise matches found using the BLAST algorithm in case of proteins or Mega BLAST algorithm for DNA. Paralogs finding
3. TMHMM Server v. 2.0	Predicts the transmembrane helices in proteins based on Hidden Markov Model. Transmembrane helices prediction
4. BetaWrap (Betawrap finding)	Predicts the right-handed parallel beta-helix supersecondary structural motif in primary amino acid sequences by using beta-strand interactions learned from non-beta-helix structures.
5. TargetP1.1 (Localization Prediction)	Predicts the subcellular location of eukaryotic proteins based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP).
6. SignalP 3.0 (Signal Peptide Prediction)	Predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms.
7. BlastP	It uses the BLAST algorithm to compare an amino acid query sequence against a protein sequence database.
8. Antigenic (Antigenic region prediction)	Predicts potentially antigenic regions of a protein sequence, based on frequency occurrence of amino acid residue types in known epitopes..
9. Conserved Domain Database (Conserved Domain Finding)	It is used to identify the conserved domains present in a protein query sequence.
10. ABCPred (Linear B Cell Epitope Prediction)	Predict *B cell epitope(s)* in an antigen sequence, using artificial neural network.
11. BcePred (Linear B Cell Epitope Prediction)	Predicts linear B-cell epitopes, using physico-chemical properties. Linear B Cell Epitope Prediction.
12. Discotope 1.1 (Conformational B Cell Epitope Prediction)	Predicts discontinuous B cell epitopes from protein three dimensional structures utilizing calculation of surface accessibility (estimated in terms of contact numbers) and a novel epitope propensity amino acid score..
13. CEP (Conformational B Cell Epitope Prediction)	The algorithm predicts epitopes of protein antigens with known structures.
14. NetMHC 2.2 (HLA Class I Epitope prediction)	Predicts binding of peptides to a number of different HLA alleles using artificial neural networks (ANNs) and weight matrices.

Algorithm	Principle Role in MalVac data base
15. MHCPred 2.0 (MHC Class I and II epitope prediction)	MHCPred uses the additive method to predict the binding affinity of major histocompatibility complex (MHC) class I and II molecules and also to the Transporter associated with Processing (TAP).
16. Bimas (MHC Class I and II epitope prediction)	Ranks potential 8-mer, 9-mer, or 10-mer peptides based on a predicted half-time of dissociation to HLA class I molecules. The analysis is based on coefficient tables deduced from the published literature by Dr. Kenneth Parker, Children's Hospital Boston.
17. Propred (MHC Class I and II epitope prediction)	Predicts MHC Class-II binding regions in an antigen sequence, using quantitative matrices derived from published literature.
18. AlgPred (Allergen Prediction)	Predicts allergens in query protein based on similarity to known epitopes, searching MEME/MAST allergen motifs using MAST.
19. Allermatch (Allergen Prediction)	Predicts the potential allergenicity of proteins by bioinformatics approaches as recommended by the Codex alimentarius and FAO/WHO Expert consultation on allergenicity of foods derived through modern biotechnology.
20. WebAllergen (Allergen Prediction)	The query protein is compared against a set of pre-built allergenic motifs that have been obtained from 664 known allergen proteins.

Table 1. Different Algorithms used in MalVac predictions tool: Novel malarial candidate vaccines using reverse vaccinology approach.

5. Ticks

Ticks are obligate heamatophagous ectoparasite classified in the subclass Acari, order Parasitiformes, suborder Ixodida, which are distributed worldwide from the Arctic to tropical regions. Ticks include 899 species that parasitize terrestrial vertebrates, including amphibians, reptiles, birds and mammals. The three families of ticks include the Argasidae or soft ticks (185 species) which is divided into two subfamilies, Argasinae and Ornithodorinae, the Ixodidae or hard ticks (713 species) which is divided into the prostriata (Ixodinae; Ixodes) and the Metastriata (subfamilies Amblyomminae, Haemaphysalinae, Hyalomminae and Rhipicephalinae) and the Nuttalliellidae, which only contains one species. The recent classification of the genus *Boophilus* as a subgenus of *Rhipicephalus* is still controversial.

Ticks, as blood-feeding ectoparasites, affect their hosts both directly and as vectors of viral, bacterial and protozoal diseases. Their impact as disease vectors on human wellbeing is second only to that of mosquitoes, and their effect on livestock, wildlife and domestic animals is immeasurably greater. The veterinary importance of ticks compared to other ectoparasites is obvious as they consume large quantities of host blood during their lengthy attachment period (7-14 days), which may be extended depending on the tick species and unique host association.

5.1 Tick genomes

Ticks have large genomes and their estimated sizes vary from 1.04~7.1 x10(9) bp, about one third to over two times the size of the human genome. Karyotype studies have revealed a range in chromosome number and the sex determining system seems to be primarily driven by a XY or a XO format. Estimates for three species are currently available, the smallest being ~10^9 bp (*Amblyomma americanum*) and the largest ~7×10^9 bp [*Rhipicephalus* (*Boophilus*) *microplus*]; the *Ixodes scapularis* genome is estimated to be 2×10^9 bp (Jongejan et al., 2007).

Nevertheless, sequencing of the genome of *I. scapularis* is under way. Pilot scale studies on the genome of *R.* (*Boophilus*) *microplus* and *R. Annulatus* are progressing and it is expected that this will strengthen a future proposal to sequence the genome more fully. Information on ESTs is available for several tick species, including *R.* (*Boophilus*) *microplus*, *A. americanum* and *Amblyomma variegatum*. Some of these have been used to create a repository of clustered and auto-annotated data in the form of species-specific gene indices.

Ixodes scapularis, the black-legged tick or deer tick, is a hard (ixodid) tick vector of the causative agents of Lyme disease, babesiosis and anaplasmosis in the United States. The Ixodes genome project was initiated in 2004, and aimed to sequence the genome of a medically significant tick. The project represents two important scientific firsts; it is the first sequencing project for a tick and a chelicerate. The project is a multi-phase and multi-investigator undertaking. Current plans call for whole genome shotgun sequencing to approximately six fold coverage of the genome. Trace reads from small (2–4 Kbp), medium (10–12 Kbp) and large (40–50 Kbp) insert genomic clones will be the basis for assembly of the genome sequence. Also included are reads of ~160,000 bacterial artificial chromosome (BAC) clone ends (BAC-end sequencing), the complete end-to-end sequencing of 60 BAC clones and ~240,000 expressed sequence tag (EST) reads. Paired BAC end reads span large segments of the genome and will be used to help assemble whole genome shotgun sequence into scaffolds. Sequenced BACs will provide an early insight into the Ixodes genome and will have utility as probes for physical mapping of assembled sequence to chromosomes. ESTs are arguably one of the most valuable resources generated as part of any genome project. These ESTs will be used to identify expressed genes, confirm gene predictions and to train automated gene finding algorithms. To date, 20 Ixodes BAC clones have been shotgun sequenced and assembled. Also available are approximately 370,000 BAC-end reads and more than 80,000 ESTs have been sequenced from a normalized pooled stage/tissue library. All sequences have been deposited at the National Center for Biotechnology Information (NCBI) (Van Zee et al., 2007).

5.2 Tick proteins and proteomics approaches

The impact of genomics on the knowledge of expressed parasite proteins has barely been experienced, yet it is here that genome and gene sequences could have a major, short-term impact. Detailed molecular information is available for few proteins from tick-borne pathogens and, relatively, for even fewer tick proteins. As an indication of the current limitations, for 20 of the most abundantly expressed proteins from unfed *R.* (*Boophilus*) *microplus* larvae, only one could be identified from a mass fingerprint alone, whereas 18 others were tentatively identified through BLAST searches and the limited existing EST database, that is, using genomic resources. The importance of translated gene or genome

sequences is likely to be even greater for less abundant proteins. Probably the most studied tick proteins are those involved in the control of host hemostasis, thought to be critical to the success of the feeding tick.

Hemostasis occurs following vascular injury and comprises three distinct events: vascular constriction, platelet aggregation and blood coagulation. Following hemostasis, clot dissolution occurs, enabling the resumption of blood flow after tissue repair. Ticks have developed a diverse array of anti-hemostatic agents that are considered to be essential for successful feeding and tick survival. Ticks circumvent the host defense mechanism through the injection of saliva containing large array of bioactive molecule including immunomodulators, vasodilators, anticoagulants, inhibitorsof platelet adhesion and aggregation and also fibrinolytic and/or fibrinogenolytic agents (Maritz-Olivier et al., 2007).

These anti-hemostatics have been found in salivary glands, saliva, eggs and hemolymph, appearing not only to prevent blood clot formation in the host, as well as the ingested blood meal, but also to regulate hemolymph coagulation in the tick itself. Recently, alternative ways to control ticks have been developed, including the employment of anti-tick vaccines with either concealed or exposed antigens. Tick anti-hemostatics are predominantly exposed antigens because they are secreted from tick salivary glands by regulated exocytosis. The importance of these inhibitors to successful tick feeding is exemplified by the finding of many such compounds in the tick salivary gland transcriptomes of *Ixodes pacificus*, *Ixodes scapularis* and *Haemaphysalis longicornis*.

A large number of tick anti-hemostatics have been identified, isolated and characterized, however, sequence, kinetic and structural data are lacking for the vast majority. Examples of the possible hemostatic candidates for vaccine production are summarized in (Table 2).

Current information on tick protein sequences is extremely restricted, limiting the application of proteomics. Homology matches have limited usefulness. We urgently need more information on the complement of proteins expressed in a variety of tick tissues, life stages and species, ideally from more than one species. Sequencing of another tick species within the metastriata should be initiated.

Recent advances in vector biology open new possibilities in target identification and vaccine development. The efforts to characterize the genomes of *I. scapularis* and *B. microplus* and also the current studies implying the use of cDNA libraries of *R. annulatus* (Shahein et al., 2008, 2010), *R. microplus* (Guerrero et al., 2005), *A. variegatum* (Nene et al., 2002), *I scapularis* (Ribeiro et al., 2006) in the identification of novel molecule, will impact on the discovery of new tick-protective antigens. The use of the information in conjugation with functional analysis using RNAi, bioinformatics, mutagenesis, transcriptomics, proteomics and other technologies will allow for a rapid, systematic approach to tick vaccine discovery by addressing the sequencing, annotation and functional analysis. Vaccination trials can be designed to evaluate the effect of selected tick antigens in combination with other tick protective and pathogen-specific antigens, for improving the level of tick infestations and reducing transmission of tick-borne diseases in cattle and other domestic animals. The future of research on development of tick vaccines is exciting because of new and emerging technologies for gene discovery that facilitate the efficient and rapid identification of candidate vaccine antigens.

Anti-hemostatic candidate	Source	Target	Molecular Mass (kDa)
Thrombin inhibitors			
R. microplus (Boophilin (G2 and H2))	Unknown	Thrombin	G2=14, H2=14
Amblyomma variegatum	Salivary Glands	Thrombin	NI
R. microplus (Microphilin)	Saliva	Thrombin exosite 1	1.8
H. dromedarii NTI-1	Nymphs	Thrombin and FX	3.2
H. dromedarii NTI-2	Nymphs	Thrombin and FX	15
Ixodes ricinus Ixin	Whole tick Extract	NI	7
FX, FXa and Tissue Factor Pathway Inhibitors (TFPI)			
Ixodes ricinus	Whole tick	FXa	NI
I. scapularis Ixolaris	Salivary glands	FX (Heparin binding exosite)	15.7
I. scapularis Salp14	Saliva	FXa	14
R. appendiculatus	Salivary glands	FXa, not the active site	65
Ornithodoros moubata TAP	Salivary glands	FXa	6
O. savignyi	Salivary glands	FXa	7
Uncharacterized intrinsic (INPI) and Extrinsic (EXPI) pathway			
R. microplus	Eggs, larvae	INPI, EXPI	NI
H. longicornis HLS1	Gut	INPI	41
O. savignyi BSAP1	Salivary glands	EXPI	9.3
O. savignyi BSAP2	Salivary glands	EXPI	9.7
Kallikrein-Kinin inhibitors			
R. microplus	Larvae	NI	18
R. microplus BmTI-D	Larvae	NI	8
R. sanguineus RsTIQ2	Larvae	NI	12
H. longicornis	Salivary glands	FXII, FXIIa and HK	16
Platelet aggregation/Adhesion inhibitors			
D. variabilis Variabilin	Salivary glands	$\alpha_{IIb}\beta_3$	5
O. moubata Moubatin	Whole tick	NI	17
O. moubata TAI	Salivary glands	$\alpha_1\beta_2$	15

Anti-hemostatic candidate	Source	Target	Molecular Mass (kDa)
Fibrin (ogen) olytic agents			
I. scapularis MP1	Salivary glands	Fibrin or Fibrinogen	36.9
I. ricinus Iris	Saliva	Elastase	44

HK: High molecular weight kininogen; NI: Not Identified; TAI: Tick Adhesion Inhibitor; TAP: Tick anticoagulant peptide

Table 2. Properties of anti-hemostatic molecules from different tick species.

6. References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*. 215, 403-410.

Ansari, F. A., Kumar, N., Bala Subramanyam, M., Gnanamani, M. & Ramachandran, S. (2008). MAAP: Malarial adhesins and adhesin-like proteins predictor. *Proteins*, 70, 659-666.

Bambini, S. & Rappuoli R. (2009). The use of genomics in microbial vaccine development. *Drug Discovery Today,* 14(5-6), (2009 Mar), 252-260.

Carlton, J. M., Angiuoli, S. V., Suh, B. B., Kooij, T. W., Pertea, M., Silva, J. C., Ermolaeva, M. D., Allen, J. E., Selengut, J. D., Koo, H. L., Peterson, J. D., Pop, M., Kosack, D. S., Shumway, M. F., Bidwell, S. L., Shallom, S. J., van Aken, S. E., Riedmuller, S. B., Feldblyum, T. V., Cho, J. K., Quackenbush, J., Sedegah, M., Shoaibi, A., Cummings, L. M., Florens, L., Yates, J. R., Raine, J. D., Sinden, R. E., Harris, M. A., Cunningham, D. A., Preiser, P. R., Bergman, L. W., Vaidya, A. B., van Lin, L. H., Janse, C. J., Waters, A. P., Smith, H. O., White, O. R., Salzberg, S. L., Venter, J. C., Fraser, C. M., Hoffman, S. L., Gardner, M. J. & Carucci, D. J. (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*, 419, 512-519.

Carpi, A., Di Maira, G., Vedovato, M., Rossi, V., Naccari, T., Floriduz, M., Terzi, M. & Filippini, F. (2002). Comparative proteome bioinformatics: identification of a whole complement of putative protein tyrosine kinases in the model flowering plant Arabidopsis thaliana. *Proteomics.* 2(11), (2002 Nov), 1494-1503.

Chaudhuri, R., Ahmed, S., Ansari, F. A., Vir Singh, H. & Ramachandran, S. (2008). MalVac: Database of malarial vaccine candidates. *Malaria Journal,* 7, (2008 Sep), 184-190.

Dalton, J. P., Brindley, P. J., Knox, D. P, Brady, C. P., Hotez, P. J., Donnelly, S., O'Neill, S. M., Mulcahy, G. & Loukas, A. (2003). Helminth vaccines: from mining genomic information for vaccine targets to systems used for protein expression. *International Journal for Parasitology*, 33 (5-6), (2003 May), 621-640.

Duffy, P., Krzych, U., Francis, S. & Fried, M. (2005). Malaria vaccines: using models of immunity and functional genomics tools to accelerate the development of

vaccines against *Plasmodium falciparum*. *Vaccine*, 23 (17-18), (2005 March), 2235-2242.

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*. 300, 1005-1016.

Emes, R.D., Ponting, C.P. (2001). A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration. *Human Molecular Genetics*, 10 (24), (2001 Nov), 2813-2820.

Filippini, F., Rossi, V., Marin, O., Trovato, M., Costantino, P., Downey, P.M., Lo Schiavo, F. & Terzi, M. (1996). A plant oncogene as a phosphatase. *Nature*, 379 (6565), (1996 Feb), 499-500.

Gardner, M. J, Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. & Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419, 498-511.

Guerrero, F. D., Miller, R. J., Rousseau, M. E., Sunkara, S., Quackenbush, J., Lee, Y. & Nene, V. (2005). BmiGI: a database of cDNAs expressed in *Boophilus microplus*, the tropical/southern cattle tick. *Insect Biochemistry and Molecular Biology*. 35, 585–595.

Jongejan, F., Nene, V., de la Fuente, J., Pain, A. & Willadsen P. (2007).Advances in the genomics of ticks and tick-borne pathogens. *Trends in Parasitology*, 23 (9), 391-396.

Kolaskar, A. S & Tongaonkar, P. C. (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*. 276, 172-174.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biology*, 3,RESEARCH0008.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*. 305, 567-580.

Li, X., Zhou, L. & Gorodeski, GI. (2006). Estrogen regulates epithelial cell deformability by modulation of cortical actomyosin through phosphorylation of nonmuscle myosin heavy-chain II-B filaments. *Endocrinology*, 147(11), (2006 Aug), 5236-5248.

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. & Bryant, S. H. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research.* D192-196.

Maritz-Olivier, C., Stutzer, C., Jongejan, F., Neitz, A. & Gaspar, A. (2007). Tick anti-hemostatics: targets for future vaccines and therapeutics. *Trends in Parasitology*, 23 (9), 397-407.

Nene, V., Lee, D., Quackenbush, J., Skilton, R., Mwaura, S., Gardner, M. J. & Bishop, R. (2002). AvGI, an index of genes transcribed in the salivary glands of the ixodid tick *Amblyomma variegatum*. *International Journal for Parasitology*. 32, 1447–1456.

Oliver, S.G. (1996). From DNA sequence to biological function. *Nature*, 379, 597–600.

Rappuoli, R. (2001). Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine,* 19(17-19), (2001 Mar), 2688-26891.

Ribeiro, J. M., Alarcon-Chaidez, F., Francischetti, I. M., Mans, B. J., Mather, T. N., Valenzuela, J. G. & Wikel, S. K. (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochemistry and Molecular Biology*. 36, 111–129

Saha, S. & Raghava, G. P. (2007). Prediction methods for B-cell epitopes. *Methods in Molecular Biology*. 409, 387-394.

Serruto, D. & Rappuoli, R. (2006). Post-genomic vaccine development. *FEBS Letters*, 580, 2985-2992.

Shahein, Y. E. (2008). Molecular cloning and expression of an immunogenic larval protein from the cattle tick *Boophilus annulatus*. *Veterinary Immunology and Immunopathology*. 121(3-4), 281-289.

Shahein, Y. E., Abd El- Rahim, M. T., Hussein, N. A., Hamed, R. R., El Hakim, A. E., & Barakat, M. M. (2010). Molecular cloning of a small Heat Shock Protein (sHSPII) from the cattle tick *Rhipicephalus* (*Boophilus*) *annulatus* salivary gland. *International Journal of Biological Macromolecules* 47(5), 614-622.

Vacca, M., Filippini, F., Budillon, A., Rossi, V., Della Ragione, F., De Bonis, M.L., Mercadante, G., Manzati, E., Gualandi, F., Bigoni, S., Trabanelli, C., Pini, G., Calzolari, E., Ferlini, A., Meloni, I., Hayek, G., Zappella, M., Renieri, A., D'Urso, M., D'Esposito, M., Macdonald, F., Kerr, A., Dhanjal, S. & Hulten, M. (2001). MECP2 gene mutation analysis in the British and Italian Rett Syndrome patients: hot spot map of the most recurrent mutations and bioinformatic analysis of a new MECP2 conserved region. *Brain Development.,* 23 Suppl 1, (2001 Dec), S246-S250.

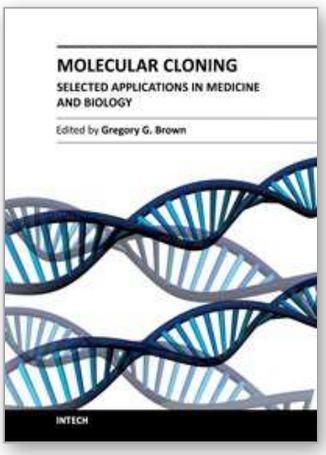
Van Zee, J. P., Geraci, N. S., Guerrero, F. D., Wikel, S. K., Stuart, J. J., Nene, V. M. & Hill, C. A. (2007). Tick genomics: The Ixodes genome project and beyond. *International Journal for Parasitology* 37, 1297–1305

Vivona, S., Gardy, J.L., Ramachandran, S., Brinkman, F.S., Raghava, G.P., Flower, D.R. & Filippini, F. (2008). Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends in Biotechnology,* 26(4), (2008 Feb), 190-200.

Vivona, S., Bernante, F. & Filippini, F. (2006). Nerve: New Enhanced Reverse Vaccinology Environment. *BMC Biotechnology*, 6, 35-42.

World Health Organization (1997). World malaria situation in 1994. *The Weekly Epidemiological Record.* 72, 269-274.

Wykes, M. & Good, M.F. (2007). A case for whole-parasite malaria vaccines. International Journal for Parasitology. 37(7), (2007 Feb), 705-712.



Molecular Cloning - Selected Applications in Medicine and Biology
Edited by Prof. Gregory Brown

ISBN 978-953-307-398-9
Hard cover, 324 pages
Publisher InTech
Published online 12, October, 2011
Published in print edition October, 2011

The development of molecular cloning technology in the early 1970s created a revolution in the biological and biomedical sciences that extends to this day. The contributions in this book provide the reader with a perspective on how pervasive the applications of molecular cloning have become. The contributions are organized in sections based on application, and range from cancer biology and immunology to plant and evolutionary biology. The chapters also cover a wide range of technical approaches, such as positional cloning and cutting edge tools for recombinant protein expression. This book should appeal to many researchers, who should find its information useful for advancing their fields.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yasser Shahein and Amira Abouelella (2011). Genome Based Vaccines Against Parasites, Molecular Cloning - Selected Applications in Medicine and Biology, Prof. Gregory Brown (Ed.), ISBN: 978-953-307-398-9, InTech, Available from: http://www.intechopen.com/books/molecular-cloning-selected-applications-in-medicine-and-biology/genome-based-vaccines-against-parasites

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen