FEBS Letters 580 (2006) 2985–2992

Minireview

Post-genomic vaccine development

Davide Serruto, Rino Rappuoli*

Novartis Vaccines, Via Fiorentina, 53100 Siena, Italy

Received 27 April 2006; accepted 28 April 2006

Available online 4 May 2006

Edited by Horst Feldmann

Abstract For over a century, vaccines were developed according to Pasteur's principles of isolating, inactivating and injecting the causative agent of an infectious disease. The availability of a complete microbial genome sequence in 1995 marked the beginning of a genomic era that has allowed scientists to change the paradigm and approach vaccine development starting from genomic information, a process named reverse vaccinology. This can be considered as one of the most powerful examples of how genomic information can be used to develop therapeutic interventions, which were difficult or impossible to tackle with conventional approaches. As the genomic era progressed, it became apparent that multi-strain genome analysis is fundamental to the design of universal vaccines. In the post-genomic era, the next challenge of the vaccine biologist will be the merging of the vaccinology with structural biology.

Keywords: Vaccines; Reverse vaccinology; Proteomics; Microarray; Pan-genome; Structural biology

1. Introduction

Louis Pasteur, who developed the first vaccine against rabies, established in 1881 the basic paradigm for vaccine development, which included the isolation, inactivation and injection of the causative microorganism. These basic principles have guided vaccine development during the twentieth century. All existing vaccines are based on killed or live-attenuated microorganisms or subunits purified from the microorganism such as toxins detoxified by chemical treatment, purified antigens or polysaccharide conjugated to proteins (Table 1). Vaccines produced following Pasteur's principles allowed the control and, in some cases, the eradication of many important infectious diseases. Despite several successes, the Pasteur's approach to vaccine development took a long time to generate vaccines against those pathogens for which the solution was feasible, but failed to produce vaccines for those bacteria and parasites that do not have obvious immunodominant protective antigens or for as yet uncultivable microorganisms.

In the last decade genomics has revolutionized vaccine research. Since the publication of the first complete genome sequence of a living microorganism in 1995 [1], the rate of genomic discoveries has grown exponentially. To date, more than 300 bacterial species have been sequenced and analyzed, including those of most major human pathogens [2]. New DNA sequencing technologies are emerging [3,4] which will likely enable faster DNA sequencing projects and further expansion of genomic information. The study of genomes by both computational and experimental approaches has significantly advanced our understanding of the physiology and pathogenicity of many microbes and has provided insights into the mechanisms and history of genome evolution [5]. While genomic and genome-based technologies applied to viral, bacterial and parasite pathogens are important from a scientific perspective, they also have significant potential to aid in the development of novel diagnostics, therapeutics and vaccines.

The new approach of the genomic era, to develop vaccines starting from the genomic information rather than growing the causative microorganism, was named reverse vaccinology [6] and can be used for the development of vaccines against pathogen for which the applications of Pasteur's principles have failed.

In the post-genomic era, pathogen genome sequencing efforts have expanded in order to include multi-representatives of the same species and this pan-genome approach has shown tremendous potential for making vaccines that once might have been impossible to design. This review will focus on recent reports that have contributed to the discovery of novel vaccine candidates providing the proof of concept of genome-based approaches such as DNA microarray analysis, pan-genome investigation and proteomics (Fig. 1). A future view of the vaccinology field and its fusion with structural biology will be also discussed.

2. The pioneering work of Meningococcus B

The concept of reverse vaccinology was applied for the first time in the attempt to develop a vaccine against serogroup B *Neisseria meningitidis* (MenB), the major cause of sepsis and meningitis in children and young adults.

At the end of the 1990s it was apparent that a universal meningococcal vaccine was beyond the reach of conventional vaccinology for two reasons: firstly, the polysaccharide which has been successfully used to make conjugate vaccines against other meningococci serogroups [7] is not immunogenic for serogroup B (and, indeed, is a potential cause of auto immunity, having an identical structure to a self human antigen), and secondly, protein-based vaccines have focused on

*Corresponding author. Fax: +39 0577 243564.
E-mail address: rino_rappuoli@chiron.com (R. Rappuoli).

0014-5793/$32.00
doi:10.1016/j.febslet.2006.04.084

Table 1
Different approaches to vaccine design in the pre-genomic era: application of Pasteur's principles

Type	Description	Advantages	Drawbacks	Examples
Killed microorganisms	The causative agent is inactivated by chemical or physical treatments	Efficacious	• Some pathogens are difficult or almost impossible to cultivate in a scalable setting • Regulatory authorities require high safety and quality standards for all new vaccine formulations; obtaining approval might be difficult	Polio virus vaccine (Salk) Influenza vaccine Rabies vaccine Oral cholera vaccine
Live attenuated microorganism	The causative agent is live but it has lost the ability to cause the disease	Efficacious, induce a protective immune response	As above	Polio virus (Sabin) Intranasal influenza vaccine (cold adapted) Measles, mumps and rubella (MMR)
Subunit	Vaccines contain purified portions of the causative agents	• There is no risk that vaccines can provoke the disease • If recombinant form of the selected components are utilized, the pathogen need not be cultivated	The identification of the few protective components from the pool of molecules present in the pathogen is usually complex and time consuming	Diphtheria toxoid Tetanus toxoid Pertussis toxoid Hepatitis B vaccine
Subunit – onjugated	A polysaccharide component of the causative agent is chemically linked to a protein carrier	The conjugated polysaccharide that is poorly immunogenic on its own becomes immunogenic	• Need to grow the pathogen in vitro to obtain the capsule polysaccharide • Capsule not always immunogenic • Too many capsule types	*Haemophlius influenzae* Meningococcus A, C, Y, W135 Pneumococcus

antigenically variable proteins that do not confer cross-protection against other strains except the one used to make the vaccine.

In a field where traditional technologies and Pasteur's principles had failed to supply a vaccine, the advent of genomics was considered a logical and promising resource that could lead to a vaccine. Therefore, the complete genome sequence of a virulent MenB strain was determined using a shotgun strategy in collaboration with The Institute for Genomic Research (TIGR) [8].

Based on the concept that surface-exposed antigens are susceptible to antibody recognition and are therefore the most suitable vaccine candidates, the complete MenB genome was screened using bioinformatic algorithms in order to select ORFs coding for putative surface-exposed or secreted proteins. Genome mining allowed the prediction of approximately 600 novel vaccine candidates, 350 of which were then expressed in *Escherichia coli* and tested for their ability to elicit protective immunity. This approach identified 28 novel protective antigens, some of which are conserved in a panel of strains representative of the meningococcal population and are likely able to induce immunity against all meningococcal isolates [9].

In essence, in less than 18 months, reverse vaccinology applied to MenB enabled the identification of more vaccine candidates than had been discovered during the past 40 years by conventional methods. The genome-derived vaccine candidates were prioritized and a selection of them have been developed in a formulation suitable for human use in order to evaluate their ability to induce immune responses in human beings.

The success of reverse vaccinology in the development of a new protein-based vaccine against *N. meningitidis* indicates that this approach is very powerful and implied that any future vaccine discovery project would strongly benefit from taking genome information into account (Fig. 1). Indeed, following the approach described for MenB, other research groups have recently used genome-based approaches to identify vaccine antigens against several human pathogens. Table 2 reports representative examples of genomes explored for vaccine candidates. Other approaches that have recently written new chapters in the reverse vaccinology concept are reported in more detail below.

3. The first application of the pan-genome concept in vaccine design: the GBS example

While the genome sequence of many different bacteria has been determined, rarely more than one or two isolates of each species have been sequenced. While the genome sequence of a single strain reveals many aspects of the biology of a species, it fails to address how genetic variability drives pathogenesis within a bacterial species and also limits genome-wide screens for vaccine candidates or for antimicrobial targets to a single strain.

The advantage of multiple genome analysis in vaccine design is highlighted by the discovery of universal vaccine candidates against *Streptococcus agalactiae*, group B *Streptococcus* (GBS). GBS is the leading cause of illness and death among

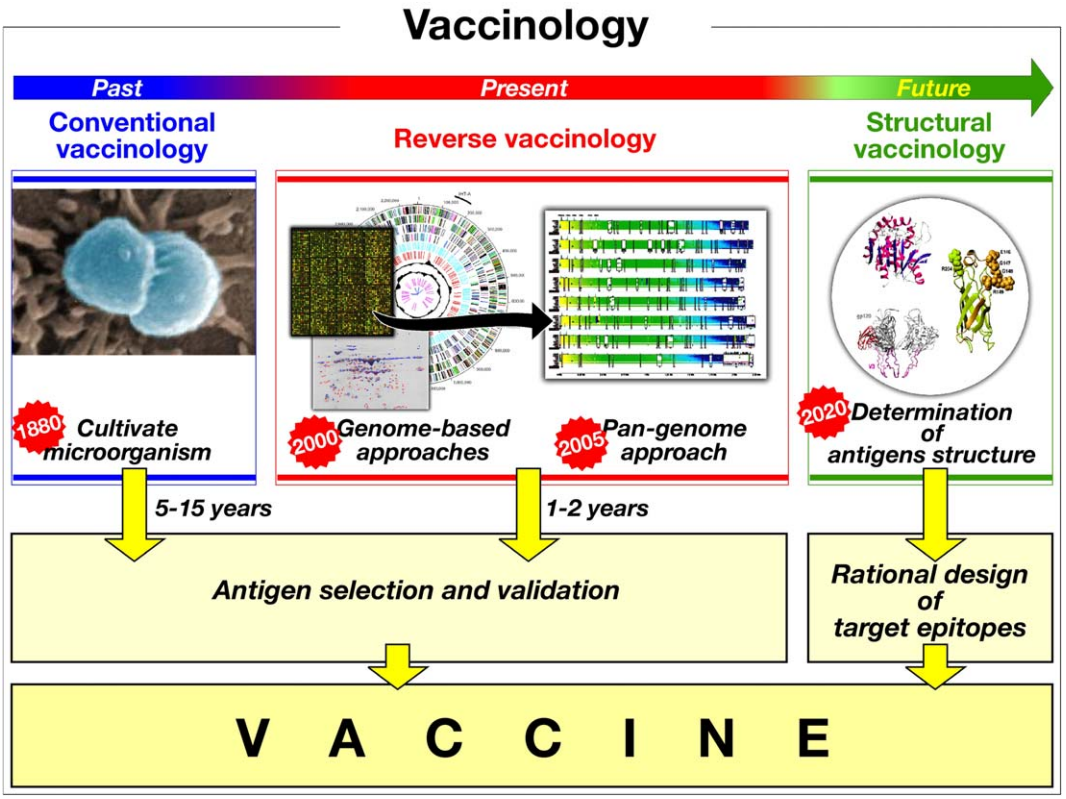

Fig. 1. The vaccinology field. **Past**: Conventional approaches to produce vaccines are based on the cultivation of the microorganism in laboratory conditions from which single components are isolated individually by using biochemical, microbiological and serological methods. Each antigen is produced in pure form either directly from the bacterium or using the recombinant DNA technology, and finally tested for its ability to induce an immune response. However, although successful in many cases, this approach presents several limitations. This approach needs to grow the pathogen in vitro, so it is not applicable to uncultivable microorganisms, and in many cases the antigens expressed during infection are not produced in laboratory conditions. Moreover, the proteins that are most abundant and easily purified are not necessarily protective antigens and, in any case, only a few molecules can be isolated and tested simultaneously. In conclusion, this method can employ many years to identify a protective and useful antigen, and has failed to provide a vaccine against those pathogens that did not have obvious immunodominant protective antigens. **Present**: The genomic era has completely changed the way to design vaccines. The availability of the complete genome of microorganisms combined with novel genome-based approaches (proteomic and microarray analysis) has introduced new perspectives in vaccine research. For the first time, the genome sequences represent an inclusive virtual catalogue of all the potential vaccine candidates from which it is possible to select the molecules that are likely more effective, regardless of their abundance, or if they are expressed in vivo or in vitro conditions. Pan-genome approach represents an advance in the use of reverse vaccinology, since it highlights the potential of looking at more than one genome for the same bacterial species to overcome the problems represented by gene presence and variability. **Future**: The structural biology applied to vaccinology (structural vaccinology) might give a tremendous help in the development of vaccines against diseases that we could not defeat using other approaches. Structural biology studies will allow the atomic resolution of the structures of potential antigens and, through the structure, the rational design of target epitopes to use as vaccine candidates.

newborn infants [10]. Nine distinct capsular serotypes of GBS have been described; however, the major disease-causing isolates in Europe and US belong to only five serotypes: Ia, Ib, II, III and V. Work on GBS vaccines has primarily focused on the capsular carbohydrates but, unfortunately, antibodies against any of these fail to confer protection against the others [11].

A genomic approach to identify candidate vaccines against GBS was conducted determining the complete genome sequence of a virulent GBS strain [12]. However, DNA microarray analysis revealed that there was a great deal of variation in gene content among clinical isolates of GBS [12]. From this analysis it was clear that only one genome sequence was not enough to develop a universal vaccine against GBS.

Hence, in order to explore gene variability within the GBS species, Tettelin et al. determined the complete genome sequence of a type I strain and draft genome sequences of five additional strains, representing the five major serotypes. Comparative analysis of the newly sequenced genomes, together with two genomes already available in the databases, revealed that a bacterial species can be described by its "pan-genome", which includes two subgenomes: a "core genome" containing genes present in all strains and a "dispensable genome" composed of genes absent from one or more strain and genes that are unique to each strain [13]. In general, the core genome includes all genes responsible for the basic aspects of the biology of a given species. In contrary, dispensable genes are responsible for species diversity and might encode additional biochemical pathways and functions that can confer selective advantages, such as antibiotic resistance or colonization of the host [14].

Maione et al. have applied the pan-genome concept to GBS vaccine discovery [15]. Bioinformatic algorithms were used to select genes from the two subgenomes that encode putative surface-associated and secreted proteins. Among the identified

Table 2
Examples of genomes that have been explored for vaccine components using genome-based approaches

Pathogen	Disease	Brief description of the approaches	References
Neisseria meningitidis B	Major cause of bacterial septicemia and meningitis	**Reverse vaccinology, microarray** – see text for details	[9,26]
Streptococcus pneumoniae	Most common cause of fatal community-acquired pneumonia in the elderly and is also one of the most common causes of middle ear infections and meningitis in children.	**Reverse vaccinology**: all the ORFs of the genome sequence of a clinical isolate of *S. pneumoniae* were evaluated to determine whether the gene products contained sequence motifs predictive of their localization on the surface of the bacterium. This led to the identification of 130 ORFs. Mice were immunized with 108 of these proteins and 6 were shown to confer protection against disseminated *S. pneumoniae* infection. All the 6 protective antigens were broadly distributed among several pneumococcus strains and showed immunogenicity during human infection	[44]
		Comparative genomics: the genome of an avirulent strain (R6) of *S. pneumoniae* has been sequenced. Comparative genome hybridization using DNA arrays reveled differences between the genomes of avirulent and virulent *S. pneumoniae*, which could contribute to differences in virulence and antigenicity. This comparison might lead to the identification of some specific proteins as potential target for vaccine development	[45]
Staphylococcus aureus	Infects wounds and causes severe infections. Following acquisition of resistance to most available antibiotics has emerged as an important opportunistic pathogen.	**Genomic peptide libraries**: *S. aureus* peptides were displayed on the surface of *E. coli* via fusion to one or two outer membrane proteins (LamB and FhuA) and probed with sera selected for high antibodies titers and opsonic activity. The exhaustive screening of these libraries by magnetic cell sorting determines the profile of antigens, which are expressed in vivo and elicit an immune response in humans. A total of 60 antigenic proteins were identified	[46,47]
		Serological proteome analysis: A surface proteins preparation from *S. aureus* was resolved by 2D electrophoresis and analyzed in immunoblotting using two pools, each consisting of five sera coming from healthy donors or patients. Twenty-one spots were isolated analyzed in mass spectrometry allowing the identification of 15 proteins including known and new vaccine candidates	[31]
Porphyromonas gingivalis	Periodontal pathogen that has been implicated in the etiology of chronic adult periodontitis	**Reverse vaccinology**: applying a series of bioinformatics tools 120 putative new antigens have been identified from the genome of *P. gingivalis*. The selected genes were cloned and expressed in *E. coli* and screened by Western blot using sera from human periodontitis patients. These candidates were reduced to a set 40 proteins, which were purified and used to immunized mice that were subsequently challenged with live bacteria in a subcutaneous abscess model. Two antigens demonstrated protection in this model of infection and therefore could represent potential vaccine candidates	[48]
Streptococcus agalactiae (Group B streptococcus)	Leading cause of bacterial sepsis, pneumonia and meningitis in neonates in US and Europe	**Proteomics**: Proteome analysis of the outer surface proteins of this pathogen allowed the discovery of novel surface proteins. Sera, raised against some of these proteins were protective in a neonatal animal model against a lethal dose of the pathogen	[49]
		Reverse vaccinology – see text for details	[12,15]
Streptococcus pyogenes (Group A streptococcus)	Causes many human infections ranging from mild pharyngitis to severe diseases, including toxic shock syndrome, necrotizing fasciitis and rheumatic fever	**Comparative genomics**: Analysis of the genome of four GAS strains led to the discovery of four new extracellular proteins. These proteins are very well conserved as observed applying sequencing and genetic population analysis. Western immunoblot confirmed that all four proteins are made during the course of distinct GAS infections and immunization with the purified form of one of these can confer protection in a murine model of infection	[50,51]
		Surface proteome – see text for details	[32]
Chlamydia pneumoniae	Causes pneumonia and is also associated with atherosclerotic and cardiovascular disease	**Reverse vaccinology, proteomic**: As a result of in silico analysis of *C. pneumoniae* genome, 157 putative surface-exposed proteins have been identified. Recombinant forms were expressed in *E. coli*, purified and used to immunize mice. Antisera were used to detect cell-surface localization by FACS analysis. 2D-gel electrophoresis and mass spectrometry were used to confirm the expression of the FACS-positive antigens in the elementary body phase of development. The result of these systematic genome-proteome combined approaches allowed the identification of 28 new vaccine candidate antigens	[52]
Plasmodium falciparum	Major causative agent of human malaria	**Post-genomic approaches**: the availability of the genome of *Plasmodium falciparum*, together with that of the vector *Anopheles gambiea* and the human host provides for the first time the entire genomic information of the three living organisms whose interaction is responsible for malaria. Kooij et al. reviewed several post-genomic approaches recently undertaken, to develop vaccines and drugs against this devastating infectious disease	[53,54]

putative surface-exposed proteins, 396 were core genes and 193 were variable genes. Authors were able to express soluble recombinant gene products in *Escherichia coli* from nearly half of the identified genes and the corresponding purified proteins were tested for protection against a GBS strain using an active maternal immunization/neonatal pup challenge model. Four antigens were capable of significantly increasing the survival rate among challenged infant mice. Only one of these antigens, the Sip protein that had been previously identified as a potential vaccine antigen [16], was part of the core genome. Each of the genes encoding the other three antigens was present in approximately 75% of strains. When the four antigens were given simultaneously as a vaccine, nearly universal protection was observed against challenge with a panel of strains representing the major pathogenic GBS serotypes [15]. Levels of protection were similar to that seen when using capsular carbohydrate-based vaccines [17].

Characterization of the newly identified vaccine antigens revealed that one of them is able to form a pilus-like structure extending from the bacterial surface [18]. This likely represents an essential virulence factor of Gram-positive bacteria that has been missed by conventional technologies for a century.

Eleven years after the first bacterial genome sequence was published, the pan-genome concept has shown that the initial strategy of sequencing one or two genomes per species is not sufficient and that multiple strains need to be sequenced to understand the basics of bacterial species and to overcome the problem represented by gene variability (Fig. 1).

The successful use of multi-strain genome analysis and the screening described for GBS provides the basis for the potential development of universal protein-based vaccines against other important and highly variable bacterial pathogens. Some important examples are already reported: multiple genome analysis was instrumental in discovering pilus-like structures in group A *Streptococcus* (GAS) and *S. pneumoniae* [19,20].

Interestingly, GAS pili correspond to Lancefield T antigens and have been unsuspectingly used during more than five decades to characterize GAS isolates. Moreover, immunization of mice with a combination of recombinant pilus proteins confers protection against mucosal challenge with virulent GAS bacteria, underling the key role of pili for infection. The data indicate that induction of a protective immune response against these structures may be a useful strategy for development of a vaccine against disease caused by GAS infection [19].

Regarding *S. pneumoniae* pili, Barocchi et al. showed that pilus-like structures (present in some but not all clinical pneumococcal isolates) are important for pneumococcal adherence to lung epithelial cells as well as for colonization, pneumonia and bacteremia in a murine model of infection. In addition, pilus-expressing pneumococci evoke higher TNF response during systemic infections than nonpiliated mutants. These data highlight the pneumococcus pili as an important factor in colonization and in the severity of disease [20].

4. Genome-based approaches: proteomic and microarray in vaccine design

4.1. Microarray analysis

Global genomic profiling of gene expression using ordered DNA or oligonucleotide microarrays has become in the last few years a very powerful technology, which has revolutionized the study of genes that are involved in microbial pathogenesis.

Microarray analysis has been used to genotype bacteria, viruses and parasites via comparative genome hybridization (CGH). CGH involves the use of a microarray containing DNA from a sequenced reference microorganism. These arrays can be used to compare genomes of different unsequenced isolates by detecting genes that are conserved between them. However, this highlights one intrinsic technical limitation of microarray: detection is limited to the DNA spotted on the array. This method also fails to detect acquisition events with respect to the reference strains.

Alternatively, microarrays can be used to study gene expression. In this case they are hybridized with cDNA prepared from mRNA isolated from microorganism grown under different growth conditions (for example in vivo versus in vitro growth). Researchers are using microarray technology to identify genes that are differently expressed in response to alteration in environmental parameters and to evaluate mutations or key factors in regulatory and metabolic pathways. Another purpose is to capture the transcriptome of bacteria growing within infected cells, tissues or animal models (for review see: [21–24]). Gene expression can be analyzed in either pathogen or host, thus allowing investigation of both sides of the host–pathogen interaction [25].

For vaccine discovery programs it is of key importance to know what genes are expressed during host infection. In fact, proteins that are expressed during disease represent the most likely protective vaccine components. The first example where microarray technology was successfully used to identify potential vaccine candidates, as well as new virulence genes, was in the case of *N. meningitidis*, where DNA microarray technology was used to study gene regulation after interaction of *N. meningitidis* to human epithelial cells [26]. RNA was isolated from adherent and non-adherent bacteria and comparatively analyzed on DNA microarrays carrying the entire collection of PCR-amplified meningococcus B genes. The authors found that bacterial adhesion to epithelial cells altered the expression of approximately 350 genes: 189 genes were upregulated and 151 genes were downregulated. Most of the regulated genes can be grouped into five major categories: adhesion genes, host–pathogen cross-talk genes, amino acids and selenocysteine biosynthesis genes, DNA metabolism genes and hypothetical genes. Moreover, of the 12 adhesion induced surface-exposed antigens identified, five were able to induce bactericidal antibodies. It is of interest to note that none of the 12 genes identified by transcriptional profiling were identified by in silico mining of strain MC58. In conclusion, this study shows that DNA microarray technology is able to identify potential vaccine candidates and complement other genome mining methods such as reverse vaccinology.

4.2. Proteomic analysis

While the availability of the complete genome sequence permits the identification of all potential protein products, this information is not sufficient to allow the identification of the subset of proteins (the proteome), which are actually expressed at any stage of the life of the bacteria in particular compartments or under different growth conditions. Recently, advances in protein separation technologies, combined with mass spectrometry and genome sequencing, have made the elucidation of total protein components of a given cellular population a

feasible task [27–29]. Moreover, the combination of proteomics with serological analysis has recently led to the development of a new valuable approach defined as serological proteome analysis (SERPA) for the identification of in vivo immunogens suitable as vaccine candidates (Table 2, [30,31]).

An attractive and powerful application of proteomics, was recently described by Grandi and colleagues who analyzed the surface proteome of *Streptococcus pyogenes* (Group A *Streptococcus*, GAS) to identify new vaccine candidate proteins [32].

This new approach, consisting of the surface digestion of live bacteria with different proteases, allowed fast and consistent identification of proteins that are expressed on the bacteria surface and thus exposed to the immune system. The cell-surface peptide fragments generated after protease treatment of GAS strain SF370 were recovered, concentrated and analyzed by tandem mass spectrometry and identified using bioinformatic examination of the publicly available genome sequence. Seventy-two proteins were identified, of which only four were predicted by the PSORT algorithm to be cytoplasmic proteins indicating that the method was highly specific for surface-exposed proteins. The power of this proteomics-based approach to recover surface proteins with exposed domains was confirmed by FACS analysis. Mouse polyclonal antibodies were produced against 51 recombinant proteins selected from the surface-exposed, and 43 of these gave a positive result in the FACS assay, confirming the surface exposure of the proteins. Considering that 7 of the 11 reported GAS protective antigens were part of the surface proteome, the authors decided to investigate whether some of the proteins identified could elicit protective responses in a mouse model of infection. To do this, they first defined the cell-surface proteins of serotype M23 strain DSM2071, a GAS strain that, unlike SF370, is highly virulent in the mouse model. Seventeen proteins were identified in this virulent strain, all of which have a homologue in strain SF370. Of the 17 proteins identified, 14 were expressed in *E. coli* as recombinant forms and used to immunize mice. Intranasal challenge of the immunized mice with a lethal dose of strain DSM2071 revealed that two of the proteins conferred protection in this model: the M protein and a putative cell envelope proteinase. M protein has been known for decades to be a protective antigen, while the protective activity of the putative proteinase was never been reported before [32].

The proteomic analysis of bacterial membranes is a key step towards the resolution of membrane proteins: it supplies experimental support to in silico prediction of protein compartmentalization and allows investigation of new aspects of the topological organization of prokaryotic membranes. From a vaccine point of view, a reliable list of surface-exposed proteins offers the possibility to identify and screen additional antigens for their capacity to elicit a protective immune response.

5. Looking at the future of vaccinology: the structural approach

The explosion of genome sequence and protein sequence data has led to the growth of the field of structural genomics, a high-throughput application of structural biology, which utilizes X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy for determination of protein structure or folding.

Structural biology studies in the past few decades have allowed the identification and understanding of many basic principles of proteins, nucleic acid structures, molecular machines and viruses. Besides, structural biology demonstrated to have enormous biomedical applications; indeed structure-based development of therapeutics is already well integrated into the pharmaceutical industry and lead to the identification of important drugs directed at the active sites of enzymes [33] such as the inhibitors for HIV-1 protease [34] and influenza virus neuraminidase [35].

We believe that structural biology can also significantly aid the rational design of future vaccines. Following the paradigm that in order to defeat the enemy, one must first know or even better see the enemy, the application of structural biology principles in the vaccinology field may result in the creation of a powerful new approach, the structural vaccinology (Fig. 1). It will allow the atomic resolution of the structures of potential antigens and, through the structure, the rational design of target epitopes to use as vaccine candidates.

Several examples of the application of structural biology to the design of effective vaccines are already in place against pathogens that most heavily afflict global health, HIV and *Mycobacterium tuberculosis*.

For HIV, structural and antigenic characterizations of the virus envelope have allowed the discovery of unique mechanisms that the virus possesses for evading the host antibody response. Nabel and co-authors excellently reviewed the rational design of an effective AIDS vaccine applying structural information [36]. HIV represents an evident example where the structural vaccinology approach might be useful in the development of a vaccine eliciting a broadly neutralizing response.

M. tuberculosis has recently become a model system for initiatives in structural genomics that resulted in solving the structure of new potential drug targets and vaccine candidates [37–39].

The antigen 85C (ag85C) story is an example of how basic structural research can help in the development of an effective treatment for tuberculosis. Ronning et al. reported the X-ray crystal structure of the *M. tuberculosis* ag85C protein (one of the three components of the antigen 85 complex), which is highly immunogenic (recognized by the immune system of infected individuals) and was shown to play an important role in cell wall synthesis and in fibronectin binding [40,41]. The three-dimensional structure sheds light on the regions of this antigen that are responsible for fibronectin binding and that may act as epitopes. At the same time, the structure potentially allows the rational design of peptide mimetics that promote antibody production against surface-exposed sequences [42]. Antibodies raised against the ag85C fibronectin-binding peptide could interfere with the ability of mycobacteria to interact with and enter host macrophages. Therefore, epitopes based on the ag85C fibronectin-binding domain as well as other surface exposed epitopes containing high sequence similarity between the three ag85 homologues, could represent candidates for vaccine design [42].

The structural approach in vaccine design has also been applied to meningococcus. NMR spectroscopy was used to obtain the solution structure of the immunodominant domain of GNA1870, a protective antigen of *N. meningitidis* identified by reverse vaccinology. Mapping of bactericidal epitopes on the solved structure will help to predict the mechanism of immune recognition. Moreover, the sequence similarity of

GNA1870 with members of the bacterial transferrin receptor family has allowed to predict the folding of this class of well known bacterial antigens, providing the basis for the rational engineering of high affinity B cell epitopes [43].

Acknowledgements: We are grateful to Kate L. Seib for the critical reading of the manuscript, to Giorgio Corsi for artwork and to Catherine Mallia for manuscript editing.

References

[1] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spiggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512.

[2] Fraser, C.M. and Rappuoli, R. (2005) Application of microbial genomic science to advanced therapeutics. Annu. Rev. Med. 56, 459–474.

[3] Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. and Rothberg, J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

[4] Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728–1732.

[5] Raskin, D.M., Seshadri, R., Pukatzki, S.U. and Mekalanos, J.J. (2006) Bacterial genomics and pathogen evolution. Cell 124, 703–714.

[6] Rappuoli, R. (2000) Reverse vaccinology. Curr. Opin. Microbiol. 3, 445–450.

[7] Balmer, P. and Miller, E. (2002) Meningococcal disease: how to prevent and how to manage. Curr. Opin. Infect. Dis. 15, 275–281.

[8] 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. and Venter, J.C. (2000) Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 287, 1809–1815.

[9] 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. and Rappuoli, R. (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287, 1816–1820.

[10] Schuchat, A. and Wenger, J.D. (1994) Epidemiology of group B streptococcal disease. Risk factors, prevention strategies, and vaccine development. Epidemiol. Rev. 16, 374–402.

[11] Davies, H.D., Raj, S., Adair, C., Robinson, J. and McGeer, A. (2001) Population-based active surveillance for neonatal group B streptococcal infections in Alberta, Canada: implications for vaccine formulation. Pediatr. Infect. Dis. J. 20, 879–884.

[12] Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Van Aken, S.E., Gill, J., Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G., Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R., Telford, J.L., Kasper, D.L., Grandi, G. and Fraser, C.M. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc. Natl. Acad. Sci. USA 99, 12391–12396.

[13] Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R. and Fraser, C.M. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. USA 102, 13950–13955.

[14] Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. (2005) The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589–594.

[15] Maione, D., Margarit, I., Rinaudo, C.D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E.T., Rosini, R., D'Agostino, N., Miorin, L., Buccato, S., Mariani, M., Galli, G., Nogarotto, R., Nardi Dei, V., Vegni, F., Fraser, C., Mancuso, G., Teti, G., Madoff, L.C., Paoletti, L.C., Rappuoli, R., Kasper, D.L., Telford, J.L. and Grandi, G. (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309, 148–150.

[16] Brodeur, B.R., Boyer, M., Charlebois, I., Hamel, J., Couture, F., Rioux, C.R. and Martin, D. (2000) Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. Infect. Immun. 68, 5610–5618.

[17] Paoletti, L.C., Peterson, D.L., Legmann, R. and Collier, R.J. (2001) Preclinical evaluation of group B streptococcal polysaccharide conjugate vaccines prepared with a modified diphtheria toxin and a recombinant duck hepatitis B core antigen. Vaccine 20, 370–376.

[18] Lauer, P., Rinaudo, C.D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A.R., Mora, M., Rappuoli, R., Grandi, G. and Telford, J.L. (2005) Genome analysis reveals pili in Group B Streptococcus. Science 309, 105.

[19] Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A.G., Maggi, T., Taddei, A.R., Grandi, G. and Telford, J.L. (2005) Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc. Natl. Acad. Sci. USA 102, 15641–15646.

[20] Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A.R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D.W., Normark, S., Rappuoli, R. and Henriques-Normark, B. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. Proc. Natl. Acad. Sci. USA.

[21] Cummings, C.A. and Relman, D.A. (2000) Using DNA microarrays to study host-microbe interactions. Emerg. Infect. Dis. 6, 513–525.

[22] Schoolnik, G.K. (2002) Microarray analysis of bacterial pathogenicity. Adv. Microb. Physiol. 46, 1–45.

[23] Conway, T. and Schoolnik, G.K. (2003) Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Mol. Microbiol. 47, 879–889.

[24] Bryant, P.A., Venter, D., Robins-Browne, R. and Curtis, N. (2004) Chips with everything: DNA microarrays in infectious diseases. Lancet Infect. Dis. 4, 100–111.
[25] Rappuoli, R. (2000) Pushing the limits of cellular microbiology: microarrays to study bacteria-host cell intimate contacts. Proc. Natl. Acad. Sci. USA 97, 13467–13469.
[26] Grifantini, R., Bartolini, E., Muzzi, A., Draghi, M., Frigimelica, E., Berger, J., Ratti, G., Petracca, R., Galli, G., Agnusdei, M., Giuliani, M.M., Santini, L., Brunelli, B., Tettelin, H., Rappuoli, R., Randazzo, F. and Grandi, G. (2002) Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. Nat. Biotechnol. 20, 914–921.
[27] Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. Nature 405, 837–846.
[28] Grandi, G. (2001) Antibacterial vaccine design using genomics and proteomics. Trends Biotechnol. 19, 181–188.
[29] Righetti, P.G., Hamdan, M., Reymond, F. and Rossier, J.L. (2004) The proteome, Anno domini Two Zero Zero Three. in: Genomics Proteomics and Vaccines (Grandi, G., Ed.), pp. 103–134, Wiley.
[30] Klade, C.S. (2002) Proteomics approaches towards antigen discovery and vaccine development. Curr. Opin. Mol. Ther. 4, 216–223.
[31] Vytvytska, O., Nagy, E., Bluggel, M., Meyer, H.E., Kurzbauer, R., Huber, L.A. and Klade, C.S. (2002) Identification of vaccine candidate antigens of Staphylococcus aureus by serological proteome analysis. Proteomics 2, 580–590.
[32] Rodriguez-Ortega, M.J., Norais, N., Bensi, G., Liberatori, S., Capo, S., Mora, M., Scarselli, M., Doro, F., Ferrari, G., Garaguso, I., Maggi, T., Neumann, A., Covre, A., Telford, J.L. and Grandi, G. (2006) Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. Nat. Biotechnol. 24, 191–197.
[33] Harrison, S.C. (2004) Whither structural biology? Nat. Struct. Mol. Biol. 11, 12–15.
[34] Kaldor, S.W., Kalish, V.J., Davies 2nd, J.F., Shetty, B.V., Fritz, J.E., Appelt, K., Burgess, J.A., Campanale, K.M., Chirgadze, N.Y., Clawson, D.K., Dressman, B.A., Hatch, S.D., Khalil, D.A., Kosa, M.B., Lubbehusen, P.P., Muesing, M.A., Patick, A.K., Reich, S.H., Su, K.S. and Tatlock, J.H. (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. J. Med. Chem. 40, 3979–3985.
[35] Kim, C.U., Lew, W., Williams, M.A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M.S., Mendel, D.B., Tai, C.Y., Laver, W.G. and Stevens, R.C. (1997) Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. J. Am. Chem. Soc. 119, 681–690.
[36] Douek, D.C., Kwong, P.D. and Nabel, G.J. (2006) The rational design of an AIDS vaccine. Cell 124, 677–681.
[37] Goulding, C.W., Apostol, M., Anderson, D.H., Gill, H.S., Smith, C.V., Kuo, M.R., Yang, J.K., Waldo, G.S., Suh, S.W., Chauhan, R., Kale, A., Bachhawat, N., Mande, S.C., Johnston, J.M., Lott, J.S., Baker, E.N., Arcus, V.L., Leys, D., McLean, K.J., Munro, A.W., Berendzen, J., Sharma, V., Park, M.S., Eisenberg, D., Sacchettini, J., Alber, T., Rupp, B., Jacobs Jr., W. and Terwilliger, T.C. (2002) The TB structural genomics consortium: providing a structural foundation for drug discovery. Curr. Drug. Targets Infect. Disord. 2, 121–141.
[38] Rupp, B. (2003) High-throughput crystallography at an affordable cost: the TB Structural Genomics Consortium Crystallization Facility. Acc. Chem. Res. 36, 173–181.
[39] Smith, C.V. and Sacchettini, J.C. (2003) Mycobacterium tuberculosis: a model system for structural genomics. Curr. Opin. Struct. Biol. 13, 658–664.
[40] Belisle, J.T., Vissa, V.D., Sievert, T., Takayama, K., Brennan, P.J. and Besra, G.S. (1997) Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. Science 276, 1420–1422.
[41] Schorey, J.S., Carroll, M.C. and Brown, E.J. (1997) A macrophage invasion mechanism of pathogenic mycobacteria. Science 277, 1091–1093.
[42] Ronning, D.R., Klabunde, T., Besra, G.S., Vissa, V.D., Belisle, J.T. and Sacchettini, J.C. (2000) Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. Nat. Struct. Biol. 7, 141–146.
[43] Cantini, F., Savino, S., Scarselli, M., Masignani, V., Pizza, M., Romagnoli, G., Swennen, E., Veggi, D., Banci, L. and Rappuoli, R. (2006) Solution structure of the immunodominant domain of protective antigen GNA1870 of Neisseria meningitidis. J. Biol. Chem. 281, 7220–7227.
[44] Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S. and Koenig, S. (2001) Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. Infect. Immun. 69, 1593–1598.
[45] Hoskins, J., Alborn Jr., W.E., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck Jr., P.R., Skatrud, P.L. and Glass, J.I. (2001) Genome of the bacterium Streptococcus pneumoniae strain R6. J. Bacteriol. 183, 5709–5717.
[46] Etz, H., Minh, D.B., Henics, T., Dryla, A., Winkler, B., Triska, C., Boyd, A.P., Sollner, J., Schmidt, W., von Ahsen, U., Buschle, M., Gill, S.R., Kolonay, J., Khalak, H., Fraser, C.M., von Gabain, A., Nagy, E. and Meinke, A. (2002) Identification of in vivo expressed vaccine candidate antigens from Staphylococcus aureus. Proc. Natl. Acad. Sci. USA 99, 6573–6578.
[47] Nagy, E., Henics, T., von Gabain, A. and Meinke, A. (2004) Identification of the "Antigenome" – a novel tool for design and development of subunit vaccines against bacterial pathogens. in: Genomics Proteomics and Vaccines (Grandi, G., Ed.), pp. 223–244, Wiley.
[48] Ross, B.C., Czajkowski, L., Hocking, D., Margetts, M., Webb, E., Rothel, L., Patterson, M., Agius, C., Camuglia, S., Reynolds, E., Littlejohn, T., Gaeta, B., Ng, A., Kuczek, E.S., Mattick, J.S., Gearing, D. and Barr, I.G. (2001) Identification of vaccine candidate antigens from a genomic analysis of Porphyromonas gingivalis. Vaccine 19, 4135–4142.
[49] Hughes, M.J., Moore, J.C., Lane, J.D., Wilson, R., Pribul, P.K., Younes, Z.N., Dobson, R.J., Everest, P., Reason, A.J., Redfern, J.M., Greer, F.M., Paxton, T., Panico, M., Morris, H.R., Feldman, R.G. and Santangelo, J.D. (2002) Identification of major outer surface proteins of Streptococcus agalactiae. Infect. Immun. 70, 1254–1259.
[50] Reid, S.D., Green, N.M., Buss, J.K., Lei, B. and Musser, J.M. (2001) Multilocus analysis of extracellular putative virulence proteins made by group A Streptococcus: population genetics, human serologic response, and gene transcription. Proc. Natl. Acad. Sci. USA 98, 7552–7557.
[51] Reid, S.D., Green, N.M., Sylva, G.L., Voyich, J.M., Stenseth, E.T., DeLeo, F.R., Palzkill, T., Low, D.E., Hill, H.R. and Musser, J.M. (2002) Postgenomic analysis of four novel antigens of group a streptococcus: growth phase-dependent gene transcription and human serologic response. J. Bacteriol. 184, 6316–6324.
[52] Montigiani, S., Falugi, F., Scarselli, M., Finco, O., Petracca, R., Galli, G., Mariani, M., Manetti, R., Agnusdei, M., Cevenini, R., Donati, M., Nogarotto, R., Norais, N., Garaguso, I., Nuti, S., Saletti, G., Rosa, D., Ratti, G. and Grandi, G. (2002) Genomic approach for analysis of surface proteins in Chlamydia pneumoniae. Infect. Immun. 70, 368–379.
[53] Long, C.A. and Hoffman, S.L. (2002) Parasitology. Malaria—from infants to genomics to vaccines. Science 297, 345–347.
[54] Kooij, T.W., Janse, C.J. and Waters, A.P. (2006) Plasmodium post-genomics: better the bug you know? Nat. Rev. Microbiol. 4, 344–357.