Review Article

Comparative Pathogenesis and Systems Biology for Biodefense Virus Vaccine Development

Gavin C. Bowick and Alan D. T. Barrett

Departments of Pathology and Microbiology & Immunology, Center for Biodefense and Emerging Infectious Diseases, Sealy Center for Vaccine Development, and Institute of Human Infections & Immunity, University of Texas Medical Branch, Galveston, TX 77555-0609, USA

Correspondence should be addressed to Gavin C. Bowick, gabowick@utmb.edu

1. Vaccines for Biodefense

We have traditionally associated the term “Biodefense” with military applications. However, since October 2001 when anthrax spores were sent in envelopes through the US Postal Service, our understanding of biodefense has shifted significantly. We now see biodefense as the process to protect both civilian and military populations. It has become clear that many highly pathogenic microorganisms can be considered as either agents of biological warfare or naturally occurring emerging disease threats. In addition, we consider bioterrorism from varying points of view, including public health threats, veterinary threats, and agricultural threats. Taken together, they can be considered to be “biotreats”.

The biodefense field represents a unique challenge for vaccine development as the traditional economic models for vaccine development are based on large populations purchasing a vaccine to protect against common infectious diseases in order that a vaccine pharma can make a profit. The situation in biodefense is very different where the goal is to stockpile vaccines with the hope that they will never be used. Relatively small markets, combined with the difficulties of working with many of these agents—many of which require biosafety level 3 or 4 containment—means that there are currently no vaccines licensed for general use in the United States and most other countries for nearly all of the biotreat agents. There are major hurdles to the development of biodefense vaccines. In addition to the traditional issues of identifying protective immunogens and platforms to deliver the vaccines, there are major difficulties in undertaking efficacy studies as most of the diseases are rare, occur sporadically, or are not found naturally. Accordingly, emphasis is being placed on appropriate animal models to demonstrate efficacy in support of licensure. However, at this time, no vaccine has been approved based on animal efficacy trials only. Nonetheless, a significant amount of data are being accumulated describing the molecular basis of disease, which provides a strong foundation for continuing product development. The current state of biodefense vaccines in clinical development is summarized in Table 1.
The advent of the 21st century has seen our expertise in molecular biology increase exponentially. With the increasing affordability of high-throughput sequencing, concepts such as reverse vaccinology—using pathogen genome sequences to produce peptides and nucleic acids for vaccine testing—have become more established [11, 12]. As our understanding of the molecular basis of pathogenesis continues to grow, coupled with large numbers of genome sequences for various pathogens becoming available and novel bioinformatic tools to facilitate their analysis, we are heading towards a position where vaccines can be rationally designed based on our molecular knowledge. This includes selection of an optimum immunogen and delivery system to maximize the host protective immune response rather than empirical approaches that were used for much of the 20th century. The continuing generation of high-throughput datasets characterizing high-priority biodefense and public health pathogens, and their host responses in the post-genomic era, will require novel bioinformatic techniques. These computational approaches will allow a more complete and thorough understanding of infection and pathogenesis, and greatly facilitate rational vaccine design. This review describes some of these methods and how they can be used in the development of novel vaccines against biothreat viruses.

2. Global Host-Pathogen Interactions of Biothreat Agents

Techniques such as microarray expression profiling, 2D gels, automated spot picking and peptide sequencing, and other high-throughput methods have revolutionized the study of virus interactions with the host cell. In particular, microarray analysis has proven a very useful tool in the emerging virus and biodefense field as the RNA extraction process facilitates removal of samples from the biosafety level 3 (BSL-3) or BSL-4 laboratory where only inactivated samples are allowed to leave biocontainment.

A successful vaccine must be both safe and immunogenic. Many pathogenic viruses interfere with the development of a protective immune response by preventing...
the activation of key aspects of the immune response, from inhibition of IRF-3 phosphorylation to block type-I interferon synthesis to sequestering MHC-I in the cell to prevent display of virus-derived peptides. A central feature of many bioterror pathogens is the infection of macrophages and dendritic cells as the primary target for replication. Infection of these cell types enables viruses to prevent immune response development at an early, central stage. For example, dendritic cells infected with *Lymphocytic choriomeningitis virus* (LCMV), an Old World arenavirus used to infect macaques as a model for Lassa fever, lead to downregulation of MHC-II on the dendritic cell surface, inhibiting the ability of these cells to present antigen and develop the adaptive immune response [13] while *Marburg virus* infection of macaques leads to downregulation of MHC-II on dendritic cells in the spleen [14]. Effects such as these are common following infection with these viruses resulting in significantly impairment of the host to develop a protective immune response at an early stage of infection. The downregulation of MHC-II has also been observed in a microarray study of PBMCs from LCMV-infected macaques with pathogenic virus inducing a greater downregulation than infection with attenuated virus [15]. Interestingly, MHC-II was upregulated in liver tissues [16], highlighting the importance of animal models and role of the host as having different functional systems, which may respond differently to a similar challenge at the molecular level.

The contributions that host-pathogen interaction profiling can make to vaccine development are exemplified by a study which determined that substitution of a single amino acid in the VP35 protein of *Ebola virus* is sufficient to disrupt the viral inhibition of innate immune signaling, while maintaining the ability to replicate to wild-type levels in cell culture [17]. Similarly, deletion of an entire multigene family, with a hypothesized role in modulation of the interferon system, still results in viral replication [18]. If the viral proteins responsible for mediating inhibition of the host immune response can be determined, then targeted mutations can be made to disrupt these effects, counteracting the inhibition of the host immune response and allowing the host to induce either an innate and/or a protective immune response. In this way, live-attenuated viruses may be developed for further characterization as novel vaccine candidates.

Large DNA viruses, such as *Variola virus*, the causative agent of smallpox, contain large genomes of almost 200 kb encoding several genes that modulate the host immune response. In these situations, individual proteins may have a single function only, allowing entire genes to be knocked out to attenuate the virus, as is seen with vaccinia virus. In some cases, sufficient redundancy may exist requiring the deletion of multiple genes or whole genome sections to sufficiently attenuate the virus. For example, *African swine fever virus*, from another family of large DNA viruses, causes a hemorrhagic disease in domestic pigs and encodes a number of genes that modulate the host immune response. One such gene, A238L, encodes a protein which inhibits the transcription factor NF-kB, a central regulator of inflammation and the innate immune response [19]. However, deleting this gene does not cause any difference in porcine disease [20]. Similarly, deletion of an entire multigene family with a hypothesized role in modulation of the interferon system still results in viral replication [21, 22].

However, all of the current bioterror viruses, with the exception of *Variola virus*, are RNA viruses, with limited coding capacity and proteins that have multiple functions. In these cases, the deletion of entire genes may result in a nonviable virus. This illustrates the importance of identifying the individual residues/regions responsible for specific virus-host interactions. If specific residues are identified that are important in determining virulence or in modulating the host immune response, but do not interfere in the critical functions of viral replication, mutation of these sites may lead to novel vaccine candidates. For example, recent mapping of the *Nipah virus* P gene has identified a region that is required for inhibition of STAT-1 signaling, but without disrupting its function as a cofactor for the viral polymerase [23].

The use of global approaches, particularly microarray gene expression technology, has provided a wealth of information detailing the host responses to infection. While these data are often limited to transcription rather than expression at the protein level, they have provided significant insight into which classes of host genes are up- and downregulated in response to virus infections. A novel application of this technology is to use the differentially expressed genes as a marker of upstream transcription factor activation using bioinformatics tools such as CARRIE [24, 25]. In this way, host signaling pathways activated or repressed by virus infection can be inferred and studied further, in particular those inhibited by virus-induced proteins that are critical in the pathogenesis of the virus (see below).

3. Comparative Molecular Pathogenesis

The study of virus-host interactions has helped to identify host proteins required for particular stages in the virus life cycle. These studies have also identified innate immune mechanisms that are suppressed during infection. Identifying such cellular proteins may allow the development of novel therapeutics, for example, by inhibiting a cellular kinase required for viral entry or replication [26–28]. However, it is likely that these studies may not provide the key information required to aid the development of novel vaccines as elucidating correlates of protection is not possible if the study compares candidate vaccine-infected versus mock-infected groups without including wild-type virulent virus-infected groups or vice versa. Currently, there are few high-throughput systems-level studies employing these types of comparisons in vivo. Analyses of human transcriptional responses to the yellow fever 17D vaccine have provided significant insights into the host response to this attenuated virus [29], but such studies, not surprisingly, lack inclusion of the human response to Asibi virus (the wild-type parent to vaccine strain 17D) and limit our knowledge of the molecular basis of attenuation of 17D virus. However, studies using virulent/attenuated pairs, such
as LCMV WE/Armstrong infection in macaques [15, 16], allow one set of responses to be “subtracted” from the other, filtering the dataset down to a more manageable size for further analysis and potential correlation of transcriptional profiles with pathogenesis or protection.

As higher-throughput sequencing becomes ever more affordable, whole genome sequencing of many virus strains and species is becoming more commonplace. Many of these types of study have been undertaken from an epidemiological perspective in order to trace the natural history of virus strains and determine and predict their spread. However, these approaches have potential important applications in the field of vaccinology. Genomic sequence analysis of large numbers of naturally occurring field strains of viruses will be important in the process to identify specific mutations that correlate with differences in pathogenicity of these viruses, and subsequently the process of attenuation.

Comparing the cellular responses to infection by virulent and attenuated viruses has provided a large amount of information on the molecular basis of pathogenesis that may have application to developing correlates of protection. A significant trend that seems to be emerging is that infection with attenuated viruses promotes a strong immune response, while these events are suppressed in virulent infection. A number of virus pairs have been used for these types of analyses. A microarray study of mouse brain tissue following infection with wild-type or a laboratory-adapted strain of Rabies virus revealed extensive activation of the inflammatory response and the type-I interferon system in mice infected with the attenuated virus [30].

Studies using Pichinde virus, a guinea pig model for Lassa fever, have taken advantage of two passage variants of the virus which cause either a severe hemorrhagic fever or a mild, self-limiting infection from which animals recover [31, 32]. Proteomic and kinomic level studies using this system have been analyzed using pathway analysis and have shown that infection with the attenuated virus induced significantly more cellular signaling events and immune response activation than infection with the virulent virus, which more closely resembles patterns of protein expression and kinase activity seen following mock infection [33–35]. These systems-level analyses can be used to generate hypotheses and tested using pathway-focused assays. Analysis of NF-κB family activity showed increased expression and DNA binding of a transcriptionally repressive NF-κB protein following virulent infection [36] consistent with observations that pathogenic arenavirus infection fails to activate cells.

The genomes of the virulent and attenuated Pichinde viruses have been sequenced and the amino acid differences between them identified [37]. The striking differences in phenotype are caused by a small number of mutations, with three amino acid differences in the envelope glycoprotein precursor, one in the nucleoprotein, and five in the polymerase. Interestingly, while the nucleoprotein has been shown to play an important role in the inhibition of interferon induction [38, 39], the mutation observed in the Pichinde virus nucleoprotein does not inhibit this function [37]. Taken together with the observation described above that a single amino acid in the VP35 protein of Ebola virus is sufficient to disrupt the viral inhibition of innate immune signaling, it suggests that modulation of immune function is not the sole determinant of pathogenicity for this virus. The recent development of an infectious clone system [40] will allow further characterization of the roles of these mutations and shed light on the relation of individual amino acid changes to arenavirus pathogenesis.

Similar findings have been reported using non-high-throughput assays. A number of studies have compared Lassa virus with the related naturally nonvirulent arenavirus Mopeia virus. Infection of monocytes, targets of arenavirus infection in vivo and central mediators of the development of a protective immune response, with Lassa virus does not result in cell activation or production of TNF-α or IL-8, and consequently the host immune response to virus infection is inhibited. In contrast, IL-8 production is not suppressed and interferon signaling is activated in Mopeia virus-infected cells, allowing the host immune response to be induced [41, 42]. Dendritic cells, which are professional antigen-presenting cells, are also targets of hemorrhagic fever virus infection. Lassa virus infection of dendritic cells inhibits upregulation of the correct cell-surface expression of proteins involved in antigen presentation, adhesion, or activation, again inhibiting the development of the adaptive immune response [43]. Observations of Lassa fever patients demonstrated higher IL-8 and interferon-inducible protein 10 levels in patients who recovered from infection compared to those with fatal outcome [44], which corroborates in vitro studies suggesting that functional impairment of immune responses contributes to pathogenesis. However, given the ability of runaway immune responses to cause an immunopathology or a “cytokine storm”-like disease following infection by some viruses, immune mechanisms need to be appropriate and controlled to allow the development of the correct responses and lead to clearance of the virus. In the case of SARS coronavirus, a microarray study demonstrated that early induction of host responses and proinflammatory signaling was correlated with a fatal outcome of infection in aged mice [45]. Two attenuated variants of Rift Valley fever virus (RVFV), a Bunyavirus and category A priority pathogen, have also provided insights into how differences in viral genome, host interactions, and pathogenesis can identify important viral virulence determinants. The NSs protein mediates the inhibition of the type-I interferon response, a critical mediator of the innate response to viral infection important in establishing appropriate secondary immune developments. The clone 13 strain of RVFV has a large deletion in the NSs coding region, removing its ability to inhibit the type-I interferon response, which leads to a highly attenuated phenotype in mice, but retains immunogenicity, and is a potential candidate vaccine [46–49].

Integrating the virus infection data using pathway analysis and functional systems biology tools can provide an overview of the similarities and differences between infections by either a virulent or an attenuated variant of a virus. For example, inferring the upstream transcription factors that are differentially activated between attenuated
and virulent infections from microarray data may allow the responsible signaling pathways to be identified. If activation of a specific signaling pathway is associated with the expression of a subset of genes associated with the development of a protective immune response, this pathway could be targeted for stimulation as a novel adjuvant strategy, or its activity assayed as a proxy for potential efficacy if screening several vaccine candidates.

While the continued discovery of the types of virulence determinants discussed above by the study of virus strains that differ in pathogenesis illustrates the power and simplicity of this approach, the above examples demonstrate the important point that characterization of the appropriate immunological events associated with protection differs from pathogen to pathogen and from model to model. As datasets from global host-pathogen interaction studies continue to become available, novel meta-analyses at the systems level may allow particular patterns associated with pathogen, host, and disease pathology to be observed, which can feed back into models for vaccine development. At present, most studies are concentrating on the innate immune response, in particular the interferon signaling pathway, rather than the adaptive immune response. The major difficulty of all these studies is undertaking studies on wild-type virulent virus in human cells due to biocontainment requirements for studies in cell culture and animal models, while studies in humans are very hard to undertake as patient samples do not usually become available until late in the disease course or from deceased individuals.

4. Applications of Systems Biology and Machine Learning in Virus-Host Interactions

High-throughput “omics” experiments provide an almost overwhelming amount of data for validation and further studies, which makes interpretation difficult. As such, analysis is often limited to a heatmap following hierarchical clustering analysis to obtain a broad overview of the dataset and the relationships between the treatment groups, with the transcripts showing the highest fold-changes selected for confirmation by RT-PCR and further experiments. With the majority of recent work focusing on either the underlying molecular biology of virus infection and cellular responses or disease pathology at the histological level, emerging bioinformatics tools, and the continued use of experimental methods that characterize cellular regulation at a number of levels, from mRNA expression to metabolomics, will allow these two extremes to be linked, providing a holistic “systems-level” view of pathogenesis [50]. Importantly, these levels of increased understanding will be beneficial to many areas of research, including vaccinology, as we will ultimately be able to screen vaccine candidates based on the effects on particular host genes and/or proteins following infection (e.g., lack of inhibition of interferon signaling pathway) of a particular cell type, and later in appropriate animal models, with the vaccine candidates.

Techniques such as k-means clustering can be of use in “filtering” large numbers of genes down to manageable sizes for further analysis. In this approach, the differentially expressed genes are portioned into a number of clusters determined arbitrarily. An iterative algorithm then refines the clustering until each transcript is in the cluster with the closest mean average. In this way, transcripts with similar expression characteristics will collect in the same cluster. If gene expression analysis is used to compare the cellular responses following infection with mock, wild-type and attenuated or candidate vaccine strain viruses, groups of genes that may be specifically up or downregulated by a particular virus variant, will become apparent. These genes may be selected for further investigation using other methods. Figure 1 shows a generic example of the type of output produced following k-means clustering. As can be seen, the three clusters show groups of genes that are upregulated following wild-type but not attenuated virus infection, genes upregulated as a consequence of both infections, and a cluster of genes which are upregulated by attenuated but not wild-type infection. This latter group of genes may represent those that are required to initiate a protective immune response and indicate potential biomarkers of protection.

An approach that is becoming more widely used in this field is that of pathway analysis. Using knowledgebases of known interactions between proteins, DNA, RNA, and small molecules, large datasets can be placed into a more “functional” context on the basis of their known interactions. Pathway analysis tools such as Cytoscape [51, 52] and the Ingenuity Pathway Analysis software [53] are increasingly used to understand protein and gene interactions and construct so-called “signaling networks”, often following characterization of gene and protein expression by microarray or proteomics analysis. Figure 2 shows an example network constructed from microarray analysis of viral infection.

These types of approaches have also been used to analyze interactions and network relationships between viral proteins as well as to characterize the interaction of viral proteins with cellular proteins [54–56]. However, to date, these types of mapping analyses have required high-throughput yeast two-hybrid type assays to generate the input data. Consequently, this type of approach may not yet be suited to the high-throughput identification of protein interaction networks for large numbers of virus strains, including potentially attenuating mutations. Furthermore, the yeast 2-hybrid system does not always generate practical meaningful interactions and a significant number of experiments are required to substantiate protein interactions. However, machine learning approaches have been applied to the determination of protein-protein interactions. In particular, support vector machines have been used to predict protein-protein interactions on the basis of sequence information alone.

Support vector machines (SVMs) are a machine learning approach which allows binary classification following training of an algorithm on a “test-set” of data. The SVM is trained on a dataset which includes a number of parameters where the classification is known. The SVM develops an algorithm that, when given a set of parameters (n) of unknown classification, can be used to predict the class on the basis of the training set. The SVM transforms the data,
Mock Wild-type Attenuated
Cluster 1
(a)

Mock Wild-type Attenuated
Cluster 2
(b)

Mock Wild-type Attenuated
Cluster 3
(c)

Figure 1: A generic example of analysis using k-means clustering. Data, such as from an mRNA expression microarray following mock, virulent, and attenuated infection of cell cultures or animals, are placed into clusters, on the basis of a similar pattern of expression to other genes in that cluster; lines represent an individual transcript. As can be seen, this type of analysis can quickly determine groups of transcripts which may associate with pathogenesis or protection.

places it into n-dimensional space, and attempts to separate the data by class by defining a hyperplane through this “feature space” (Figure 3). The advantage of these types of methods is that the number of parameters can be tailored to strike a balance between what is experimentally feasible and predictive accuracy. As the number of parameters increases, classification accuracy may improve, but collecting the data may become less feasible experimentally. With training sets constructed from known protein-protein interactions defined experimentally, SVM-based approaches have shown increasing accuracy for predicting these interactions [57–59]. Using these approaches, it may be possible to screen large numbers of protein sequences in silico to elucidate amino acid mutations to disrupt protein-protein interactions that may attenuate a virus, such as the single amino acid substitution that significantly attenuated Ebola virus [17].

Computational approaches have also demonstrated their power in de novo attenuation of viruses. Coleman et al. coined the term “synthetic attenuated virus engineering”, SAVE, to describe their approach for attenuating virulent viruses [60]. Their approach takes advantage of the redundancy of the triplet genetic code and the fact that there is often a codon usage bias, where particular codons are used more or less often to code for a particular amino acid than would be predicted by chance. They developed a computer algorithm that recoded a given amino acid sequence at the genetic level, changing the codon pairs to those which were under or overrepresented in the wild-type sequence. Their algorithm also included other features such as RNA folding-free energy, a factor shown to be implicated in host adaptation of influenza virus [61]. By altering codon pair usage, they were able to construct a poliovirus that was attenuated in mice and provided protective immunity [60]. Of particular interest with this type of strategy is that the protein sequence remains unchanged from the wild-type virus. This reduces the likelihood of reduced protection by the attenuating mutations disrupting important epitopes.

As discussed above, the majority of high-throughput data to emerge from the biodefense field has been microarray mRNA expression data, rather than from high-throughput protein interaction data that directly lend itself to network analysis. However, genomics- and proteomics-level analyses have produced interaction networks that are beginning to yield some consistent and overlapping signaling pathways across different model systems, and novel computational approaches are providing alternative strategies for virus comparison and attenuation.

5. Towards Targeted Vaccine Design

Traditionally, live-attenuated viruses have been developed by serial passage of wild-type viruses in culture. The live-attenuated yellow fever 17D and the polio Sabin vaccines were developed in this way. However, the attenuated poliovirus vaccine can revert to virulence and has been associated with vaccine-associated poliomyelitis [62]. In recent years, the yellow fever 17D vaccine has caused a number of serious adverse effects with a significant case fatality rate [63–65]. Some of these outcomes may be associated with polymorphisms in the CCR5 chemokine receptor and RANTES chemokine genes in some vaccinees [66]. This finding complicates the vaccine safety field. If the vaccine-associated disease is associated with host polymorphisms rather than with virus reversion, then determining vaccine safety becomes increasingly complicated and suggests that in the 21st century each vaccine may become restricted to individuals with a particular host genetic background.
As high-throughput methods become more affordable, personalized vaccinology may be a strategy that becomes more feasible. Personalized vaccinology refers to the ability to tailor particular vaccines to specific individuals [67, 68]. If, for example, adverse responses to the yellow fever vaccine can be attributed to a specific genetic polymorphism, then the vaccinee can be screened for this polymorphism, and decisions whether or not to vaccinate can be made on that basis. If alternate vaccines are available, which perhaps offer a shorter-lived immunity but without reported adverse effects, appropriate risk assessments can be made and a course of action determined, in the best interests of the individual vaccinee.

The two basic requirements for vaccine design are the induction of a protective immune response and safety. Bioinformatics can assist with both of these. Studies of human responses to the yellow fever vaccine using mRNA microarrays have revealed the immune pathways responsible for protection [29, 69, 70]. Unfortunately, as these studies were performed in human volunteers, there is no directly comparable dataset following exposure to wild-type yellow fever virus infection. As discussed above, this raises one of the problems of utilizing system biology in vaccine development; namely, studies with candidate and licensed vaccines in humans are comparatively easy while obtaining samples from wild-type infections and serious adverse events are very difficult for a variety of reasons. However, these studies do reveal significant insights at the molecular level into how a historically very successful vaccine induces protection. These studies showed activation of several arms of the immune response but identified key transcription factors which acted as “master switches” in immune response development. Pathway analysis revealed these to be central “hubs”: highly connected “nodes” in a signaling network. There are a number of biodefense vaccines in Investigation New Drug (IND) Status, due to limited studies on safety and efficacy, which are administered to individuals who are at risk from the disease. Bioinformatic studies on vaccinees who receive these IND vaccines may help generate important data that have application to safety and efficacy.
These types of analysis may also assist in the development of novel adjuvants. Microarray studies characterizing the global host responses to the adjuvants alum, CpG, MF59, and LTK63 have identified transcript signatures common between and specific to these adjuvants [71, 72]. By understanding how these transcriptional responses correlate with the development of specific immune mechanisms, appropriate pathways could be stimulated with specific vaccines to induce the appropriate immune effectors functions for that pathogen: as efforts to design vaccines to be as safe as possible continue, a potential side effect could be over-attenuation leading to increased safety at the expense of immunogenicity and lasting protection. A possible solution could be the use of novel adjuvants to stimulate the signaling pathways associated with a strong immunogenic response. As an example, inflammasome activation was found to be a central regulator of the immune response to yellow fever vaccine [69]. Interestingly, a novel nanoparticle-based vaccine delivery method has recently been reported which includes lipopolysaccharide to activate the inflammasome alongside delivery of a protein subunit vaccine [73]. Combining findings such as these will be central to future themes of vaccinology. As our understanding of the correlates of protection for currently licensed vaccines increases, we may be able to target specific cellular pathways to tailor immune responses to a new generation of highly attenuated, safe vaccines.

Recently, a machine learning algorithm, based on a support vector machine, was used to identify potential correlates of immune protection, with potential transcript-based indicators of cell-mediated and antibody responses identified [29]. While these findings require further validation and comparison with potential indicators for responses to other vaccines, they suggest that bioinformatics could provide tools to predict immunization outcome soon after vaccination. Additionally, pathway analysis can be used to screen for transcriptional responses which are associated with other diseases or adverse effects, potentially providing a means to assay vaccine safety in vivo early in the development process.

Currently, a limiting factor in the rational attenuation of biodefense viruses for vaccine design is in obtaining datasets from equivalent platforms and model systems, facilitating meta-analyses to determine potential correlates of attenuation. However, while the virus-host interaction field may not yet be addressable in this way, efforts are being made to compile easily accessible databases of viral sequences for query. One such database has collected over 300 protein sequences from pathogenic arenaviruses to aid in epitope discovery studies [74]. Combining these types of databases with patient data such as antibody titer or disease severity may allow consistent peptide sequences which correlate with immune protection to be determined. This type of approach, combined with computational MHC epitope prediction, will significantly further our ability to rationally develop immunogenic vaccine candidates.

6. Summary and Conclusions

The increased availability and affordability of high-throughput technologies, such as whole genome sequencing and “omics” analyses, will allow the continuing development of large, multilevel datasets to characterize host responses to pathogens of differing virulence, and the changes in viral sequences responsible. As bioinformatics continues to evolve, and as microbiologists, immunologists, and vaccinologists become more familiar with the potential of these technologies, the stage will be set for novel approaches to vaccine design for biodefense, combining rational attenuation with immunogenicity and safety testing. By integrating these approaches early in the process, and using a combination of in vitro and in silico techniques, vaccine candidates with the most promise can be selected for further testing with a greater confidence, and funding can be better appropriated.

The use of comparative studies between virulent and attenuated strains could provide a wealth of data to assist in vaccine development. While many datasets have been published, comparison across different experimental systems remains an issue. Ideally, cell lines, timepoints, and other experimental criteria need to be standardized to allow informed comparison of the results. The ability to perform meta-analyses across different studies could allow a better understanding of what host responses are required to lead to a long-lasting protective immune response. If a dataset which characterized the host responses to yellow fever vaccine could be compared with similar datasets showing responses to other live vaccines, for example, smallpox vaccine and Junin vaccine, along with response to the appropriate wild-type agents, we will undoubtedly find several “groups” of similarities and differences which associate with protection,
virulence, or type of virus or disease pathology. These approaches to vaccine design illustrate how vaccinology can use the data from basic science, and particularly studies of molecular pathogenesis, to develop novel vaccine candidates. By identifying common cellular responses to attenuated virus vaccines, across different viral families and different wild-type disease pathologies, we may characterize a central "core" of responses which are always activated by a historically successful vaccine, regardless of virus type. These responses can guide us in the development of novel vaccine candidates or adjuvants.

As new viruses continue to emerge and the risks from emerging and remerging pathogens remain, the need for vaccines against these pathogens remains significant. The introduction and spread of West Nile virus to North America illustrates how quickly viruses can become established in a new area and present novel threats to public health [75]. The continuing investment in biodefense and emerging pathogens will result in even greater amounts of basic data to input into novel bioinformatics tools. As our understanding of the molecular determinants of protection and pathogenesis continues to grow, the development of safe, effective vaccines to these important pathogens will continue to accelerate.

Acknowledgments

The first author is supported by the Sealy Center for Vaccine Development, the Departments of Pathology and Microbiology & Immunology by the Department of Defense, through a National Biocontainment Training Center Fellowship. The authors thank Dr. Heidi Spratt for helpful discussions.

References


Submit your manuscripts at http://www.hindawi.com