J Appl Genetics (2015) 56:381–391
DOI 10.1007/s13353-014-0265-2

MICROBIAL GENETICS • REVIEW

Overview of computational vaccinology: vaccine development through information technology

Nishita Vaishnav · Aparna Gupta · Sneha Paul · Georrge J. John

Received: 9 August 2014 / Revised: 17 November 2014 / Accepted: 8 December 2014 / Published online: 23 December 2014

Abstract Pathogenic organisms, causes of various infectious diseases, possess a rich repository of antigenic proteins that engender an immune response in a host. These types of diseases are usually treated with the use of pharmaceuticals; unfortunately, many of these also have a potential to induce fatal side effects, especially allergic responses in the diseased host. In addition, many pathogens evolve (by selective survival) single or multi-drug resistance (MDR). Therefore, a means to prevent the host from becoming susceptible to the pathogen from the onset, rather than trying to devise pharmacologic protocols to treat an ongoing infection, are increasingly seen as desirable to reduce the incidence of infectious diseases altogether. To this end, cost-effective development and use of "safe" vaccines is key. This paper provides an overview on the new and expanding area of computational vaccinology and a brief background on pathogen antigenicity, identification of pathogen-specific antigens, and screening of candidate antigens using various tools and databases developed in the recent past.

Keywords Infection · In silico · Prediction · Vaccine

Communicated by: Agnieszka Szalewska-Palasz

N. Vaishnav · A. Gupta · S. Paul · G. J. John (✉)
Department of Bioinformatics, Christ College, Rajkot, Gujarat, India
e-mail: johnjgeorrge@gmail.com

G. J. John
Department of Biochemistry and Molecular Biology, University Clinic of Bonn (UKB), Bonn, Germany

Present Address:
S. Paul
Centennial College, Toronto, Toronto, ON, Canada

Introduction

Every year, approximately 3.5 million people die worldwide due to infectious diseases (WHO 2014). Infection is defined as an invasion of a host's tissues by disease-causing organisms, their multiplication, and the reaction of host tissues to these organisms/toxins they produce. Infections can be caused by a multitude of microbes and macro-parasites, and are classified on the basis of the causative agents, i.e., bacterial, protozoal, viral, fungal, etc., transferred among hosts by numerous means. For example, respiratory diseases are often acquired by a naive host when in contact with aerosolized droplets containing the organism that are released from an infected host during sneezing, coughing, or talking. On the other hand, gastrointestinal diseases are routinely acquired by consuming food\water directly contaminated with pathogens, or by the inadvertent transmission by inappropriate usage of sanitary techniques during the handling of food and liquid meant for consumption.

These types of diseases, when diagnosed, are treated with various pharmaceuticals such as penicillin, Cephalosporin, Glycopeptide, and Macrolide; unfortunately, many of these drugs also have the potential to induce fatal side effects (Poppe 2001; Nebeker et al. 2004; Kronman et al. 2012), especially those that induce allergic responses, in the diseased host. Nevertheless, even with a plethora of drugs to choose from, many pathogens can develop single/multi-drug resistance (MDR) through selective survival (Sood and Arti 2008). Therefore, a means to prevent the host from becoming susceptible to the pathogen from the onset, rather than trying to devise pharmacologic protocols to treat an ongoing infection, have increasingly been seen as desirable throughout the world to reduce the incidence of infectious diseases altogether. To this end, the practical/cost-effective development and use of "safe" vaccines is key.

Vaccine: a shielding measure

When a pathogen/antigen is introduced in a host, the vaccine tricks (i.e., using a "non-real" version of the organism) the host immune system and educates it about how to respond at that moment/in the future against the same agent (NIAID 2013), thus inducing a cell-mediated/humoral immune response. For example, a viral vaccine contains an undermined form of the virus that neither causes a disease nor replicates. Since the host macrophages are unaware that the vaccine virus is undermined, it processes it as it would any other antigen/pathogen. The macrophages then induce the viral antigen to T- and B-cells in the lymph nodes. In turn, after several steps including activation and differentiation, virus-specific T-cells activate B-cells to ultimately produce anti-viral antibodies. While the weakened viruses in the vaccine are quickly eliminated, the host is left with a set of memory T- and B-cells which are imperative to opposing the specific viral disease should the host ever actually encounter the live organism.

In the past, vaccines have been developed using different strategies, each with their own disadvantages and advantages. These vaccines are of the following types:

Live or attenuated These vaccines enclose the living microbe in a weakened form so that it cannot induce the actual disease. For example, vaccines against measles, mumps, and chickenpox are live, attenuated vaccines. They provoke more long-lasting humoral and cellular immune responses and are preferred for healthy adults, not immunocompromised people (Sinha and Bhattacharya 2014). The limitation here is that while the microbes are live and attenuated, they still have the tendency to change and potentially relapse to a virulent form. In addition, the habituated live attenuated vaccine necessitates refrigeration to stay viable; this makes their use costly and problematic in parts of the world where optimal facilities are not available. It is more difficult to develop these vaccines for bacteria as they have thousands of genes and are hence much harder to control (Alexandersen 1996; NIAID 2013). The other concern with live or attenuated vaccines is that they can confound disease investigation, based on serological testing which leads to false positives (Pluimers 2004).

Inactivated Since dead microbes in an inactivated vaccine are treated with chemicals, heat, or radiation, they lose the ability to replicate and cause infection. This inactivated vaccine contains pathogen recognition patterns capable of initiating innate immune responses (van Duin et al. 2006). Vaccines for polio, plague, TB, etc. are of inactivated type (Immunization 2014). The storage of these vaccines doesn't require refrigeration, making them useful and safer for larger populations or easier to deliver to remote areas of the world. However, the major constraint for the use of such vaccine is that they lead to generation of weaker immune responses, and therefore, there is a need for recurring dosages or booster shots due to diminishing immunity. In order to overcome this limitation, inactivated vaccines are often administered along with an adjuvant, which induces an inflammatory response and can stimulate the cellular immune system (Plotkin 2003; Principles of Vaccination 2014).

Subunit A vaccine candidate causes an immune response due to the presence of an antigen as a small part of the pathogen protein, i.e., a subunit. These subunits bear the antigens that correspond to the major antigenic sites, which is followed to generate hepatitis B and C vaccines. Sub-unit-based vaccines use semi-pure antigens as epitopes targeted by the antibodies; use of these vaccines tends to cause a reduction in adverse reactions compared to other types of vaccines. The advantage of using such vaccines is the significant increase in safety and reduction in antigenic competition faced by the host immune cells. The potential disadvantage of this class of vaccines is the usage of strong adjuvants having low sustainable immunity in order to attain maximal specificity (Plotkin 2003; NIAID 2013). Subunit vaccines can be effective at inducing humoral immunity, but may not effectively induce cellular immunity and will not mimic the natural route of infection (Foged 2011).

Toxoid These types of vaccines are best suited against bacterial toxins wherein the antigens themselves are key components in the pathologies observed in an ongoing infection. Vaccines administered for diphtheria and tetanus are examples of toxoid vaccines. Tetanus toxoid is innately a stronger immunogen than that for diphtheria, which becomes apparent when immunocompetence is more limited, such as in preterm infants (Faldella et al. 1998). In this type of vaccine, an agent such as formalin is used to inactivate purified toxins and these "detoxified" toxins (now termed toxoids) are used to generate the vaccine. Toxoids may elicit immune response, but with no guarantee of retention of memory for future response; in other words, it produces a less effective immune response compared to live-attenuated vaccines (Clem 2011; NIAID 2013).

DNA vaccines These forms of vaccines contain fragments of DNA (generally a gene) from the pathogen. A DNA vaccine would induce a strong antibody response to the free-floating antigen secreted by cells, and would also kindle a strong cellular retort against the microbial antigens displayed on cell surfaces. It is not possible for the DNA vaccine to cause the disease, even in immunocompromised people, since it would not contain the microbe, but just copies of a few of its genes capable of producing antigenic proteins. In addition, DNA

vaccines are relatively easy and inexpensive to design and produce, and there is also no need of an adjuvant. Vaccines against influenza and herpes are DNA vaccines, which are currently in the experimental stage (Donnelly et al. 2005; NIAID 2013).

Recombinant vector vaccines These are similar to DNA vaccines, but employ an attenuated virus or bacterium to commence microbial DNA to cells of the body used as haulier and defined as vectors. This vaccine induces both humoral and cellular immune response. Vaccines for HIV, rabies, and measles are of this type, which are also currently in the experimental stage (NIAID 2013). In this type of vaccine, it is possible to introduce genes from more than one pathogenic virus into the same vector that protects against several pathogens. A potential problem with recombinant viruses is that they could cause disease in immunocompromised people (Plotkin 2003, 2005). A bacteriophage is a vector used to introduce a multivalent antigen into a cell or an organism. This phage is constructed to display an antigen on its surface, resulting in the presentation of the antigen to the immune response cells (Pande et al. 2010). The advantages of using this type of bacteriophage-based recombinant vaccine delivery include high immunogenicity because of the addition of foreign CD4 T-cell epitopes to pIII or pVIII, high stability due to the relatively low surface complexity of the page, and low production cost (van Houten et al. 2010). Recombinant Bacillus subtilis spores are a durable delivery vehicle for the heterogenous vaccine antigens. As it has various coat proteins (CotB, CotC, CotG, CotZ, and CgeA) on the surface, it is feasible to express and fuss the heterogenous antigen (Iwanicki et al. 2014). The tuberculosis antigen, MPT64, fussed with spore coat gene cotB, was tested for the ability to protect mice against tuberculosis (Sibley et al. 2014). In another study, the tetanus toxin (TTFC) was expressed with spore coat gene CotB, which is not experimentally proven in any animal studies (Isticato et al. 2011).

As noted here, each of these commonly-used vaccine types have inherent problems associated with their generation and/or their subsequent utilization. Generation of a novel vaccine through conventional approaches requires concerted efforts, huge costs, and quite prolonged procedures. Among the latter are the issues of isolating, inactivating, and injecting antigenic agents of various infectious pathogens into systems to create the vaccine itself (André 2003).

Computational vaccine development

Computational vaccinology holds the promise of changing vaccine development in that it challenges the conventional "price-and-period" approach. In this genome era, with the bulk inflow of constructive genome sequence information from various hosts and numerous pathogens, there has been an increase in the use of computational approaches for vaccine development based on genomic information; this was termed "reverse vaccinology" (Bambini and Rappuoli 2009; De Groot et al. 2002; Major et al. 2011; Plotkin 2008; Serruto et al. 2009). Vaccine informatics is a growing arena that draws attention to the expansion and relevance of bioinformatics techniques in pre-clinical vaccine development ventures. Diverse immunoinformatics algorithms have also been developed to envisage T- and B-cell immune epitopes for epitope vaccine development. Virtual screening of potential vaccine candidates directly from genome sequences is made possible through reverse vaccinology approaches. Computational simulations for host–pathogen interactions are also increasingly used in vaccine development protocols and for other purposes like development of immunization registries, appraisal of vaccine safety and worth, and immunization modeling (He et al. 2010a, b). The instances of successful vaccines and probable vaccine candidates for a few pathogens are provided in Tables 1 and 2.

The tools and databases used for computational vaccine development are provided in Tables 3 and 4. The first step is data collection, which included retrieving proteomes from proteomics databases. Several protein sequence databases are available in this bioinformatics era. Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data [http://www.uniprot.org/], which includes UniProt Knowledgebase (UniProtKB),

Table 1 Vaccine developed through reverse vaccinology approach

Sr. No	Vaccines	References
1.	*Anaplasma marginale*	PMID: 22066488
2.	*Schistosoma spp*	PMID: 21458582
3.	*Neisseria meningitides*	PMID: 21993656
4.	*Listeria monocytogenes*	PMID: 21791233
5.	*Trypanosoma brucei*	PMID: 21349321
6.	*Corynebacteriumpseudo tuberculosis*	PMID: 23201561
7.	*Stryptococcus spp*	PMID: 23181284
8.	*Herpes simplex virus type 2*	PMID: 22649465
9.	*Mycobacterium tuberculosis*	PMID: 21756938
10.	*Human adenovirus type 55*	PMID: 23087107
11.	*Pasteurella multocida*	PMID: 22540951
12.	*Rhipicephalus microplus*	PMID: 22521592
13.	*Brachyspira hyodysenteriae*	PMID: 19179021
14.	*Leptospira borgpetersenii*	PMID: 23176980
15.	*Pasteurella multocida*	PMID: 22792202
16.	*Cryptosporidium spp*	PMID: 21918117
17.	*Ehrlichia ruminantium*	PMID: 20566221
18.	*Teladorsagia circumcincta*	PMID: 23282110

Table 2 Vaccine candidate prediction through reverse vaccinology approach

Sr. No	Vaccine	References
1.	*Leishmania spp.*	PMID: 22434357
2.	*Campylobacter fetus*	PMID: 22890137
3.	*E coli 536*	PMID: 22493535
4.	*Plasmodium vivax*	PMID: 21713006
5.	*Mannheimiahaemolytica*	PMID: 20336679
6.	*Human influenza H1N1*	PMID: 22450272
7.	*Halicobacter pylori*	PMID: 22130156
8.	*Mycoplasma genitalium*	PMID: 22057004
9.	*Human papilloma virus 16 E6*	PMID: 21699918

UniProt Reference Clusters (UniRef), UniProt Archive (UniParc), and UniProt Metagenomic and Environmental Sequences (UniMES). The species specific sequences in fasta format can be downloaded through a web server and file transfer protocol (FTP) (Magrane and Consortium 2011). The protein database of the National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov/protein] has translation sequences from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB, which could be retrieved through text query search and taxonomy search [http://www.ncbi.nlm.nih.gov/taxonomy/]. Then, clustering these retrieved sequences using CD-HIT tool [http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit] (Huang et al. 2010) results in non-redundant sequences. MaxCluster (Herbert 2014) and USEARCH (Robert 2014) are other tools that allow sequence clustering, but these have a lower accuracy (Fu et al. 2012). CD-HIT run is based on a greedy incremental clustering algorithm in which the input sequences are first sorted in the order of decreasing length. The longest sequence becomes the representative of the first cluster. Then, each residual sequence is compared with the representatives of existing clusters. If similarity is found above a given threshold, it is grouped into that cluster. Else, a new cluster is defined with that sequence as the representative. Short word filtering is used for sequence comparison to confirm whether the similarity is below the clustering threshold. If this is not confirmed, an actual sequence alignment is performed (Li et al. 2001). The sequence post-CD-HIT is then screened by eliminating homologous proteins between rodents and humans using a BLASTP tool. Basic Local Alignment Search Tool (BLAST) directly approximates alignments that optimize a measure of local similarity, i.e., the maximal segment pair (MSP) score. BLAST performs faster than existing sequence comparison tools with high sensitivity. (Altschul et al. 1990). The data from BLASTP yields homologous and non-homologous protein sequences to the pathogen, human, and rodents, and then the homologous sequence is discarded as it generates an autoimmune response in the host (John et al. 2012).

The non-homologous proteins are then characterized for different properties, a step that is a prerequisite for it to become a vaccine candidate. Identification of membrane proteins that are necessary for pathogenic interactions with the host is performed using sub-cellular localization tools such as PSORTb for bacteria and archaea and CELLOfor eukaryotes. PSORTb is a stand-alone, web-based tool which analyses sequence based on multiple analytical modules such as SCL-BLAST & SCL-BLASTe or SubCellular Localization BLAST, Support Vector Machine (SVM), Motif & Profile Analysis, Outer Membrane Motif Analysis, ModHMM, and Signal Peptide. Each module analyzes one biological feature known to influence or be a characteristic of subcellular localization. The final results are derived based on probabilistic method and fivefold cross validation to assess the likelihood of a protein being at a specific localization. These likelihoods are used to generate a probability value for each of the five major localizations for Gram-negative bacteria (cytoplasmic, inner membrane, periplasmic, outer membrane, and extracellular) and four localizations for Gram-positive bacteria (cytoplasmic, cytoplasmic membrane, cell wall, and extracellular) with 80–95 % accuracy (Yu et al. 2010). CELLO uses four types of sequence coding schemes: the amino acid composition, the di-peptide composition, the partitioned amino acid composition, and the sequence composition based on the physico-chemical properties of amino acids. It predicts four localizations for eukaryote (cytoplasmic, extracellular, mitochondrial, and nuclear) with 87 % accuracy and five localizations for bacteria (cytoplasmic, inner membrane, periplasmic, outer membrane, and extracellular) with 88.9 % accuracy (Yu et al. 2004).

Subsequently, plasma membrane proteins and extracellular proteins are screened to predict transmembrane helices using a high-efficiency TM-HMM software. The hidden Markov model predicts transmembrane helices in the protein sequences with 77 % accuracy (Sonnhammer et al. 1998). Those proteins that yield transmembrane (TM) values equal to 1 are chosen for subsequent steps as they can be easily cloned and expressed; if proteins have a TM value greater than 1, this would likely result in errors in cloning and expression in vitro (Pizza et al. 2000).

The selected transmembrane proteins are then analyzed for binding epitopes using ProPred1 and ProPred programs that predict 47 MHC class I (Singh and Raghava 2001) and several class II alleles (Singh and Raghava 2003), respectively. These are web-based tools which run on a matrix-based approach that provides an accuracy of approximately 80 %. The simultaneous prediction of MHC class I binders and proteasome cleavage sites in an antigenic sequence leads to the identification of potential T-cell epitopes. B- and T-cell epitopes are

Table 3 Immunoinformatics databases

Name	Purpose	URL
AntigenDB	Sequence, structure, and other data on pathogen antigens	http://www.imtech.res.in/raghava/antigendb/index.html
AntiJen	Quantitative binding data for peptides and proteins of immunological interest	http://www.ddg-pharmfac.net/antijen/AntiJen/antijenhomepage.htm
BCIpep	Warehouses information of all experimentally determined B-cell epitopes of antigenic proteins	http://bioinformatics.uams.edu/mirror/bcipep/
dbMHC	The dbMHC database provides an open, publicly accessible platform for DNA and clinical data related to the human major histocompatibility complex (MHC)	http://www.ncbi.nlm.nih.gov/gv/mhc/main.fcgi?cmd=init
DIGIT	Database of ImmunoGlobulin sequences and Integrated Tools	http://www.biocomputing.it/digit4/
GPX-Macrophage Expression Atlas	A database for expression profiles of macrophages challenged with aa variety of pro-inflammatory, anti-inflammatory, benign, and pathogen insults	http://gpxmea.gti.ed.ac.uk/
HPTAA	HPTAA is a database of potential tumor-associated antigens that uses expression data from a range of expression platforms, including vigilantly selected overtly accessible microarray expression data, GEO SAGE data and Unigene expression data	http://www.bioinfo.org.cn/hptaa/
IEDB	The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species	http://www.iedb.org
IEDB-3D	Structural data within the Immune Epitope Database	http://www.iedb.org/bb_structure.php
IL2Rgbase	X-linked severe combined immunodeficiency mutations	http://www.ncbi.nlm.nih.gov/lovd/home.php?select_db=IL2RG
IMGT	IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility (MH) of human and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MhSF), and related proteins of the immune system (RPI) of vertebrates and invertebrates	http://www.imgt.org/
IMGT_GENE-DB	IMGT/GENE-DB is the IMGT® comprehensive genome database for immunoglobulins (IG) and T cell receptors (TR) genes from human and mouse, and, in development, from other vertebrate species (e.g., rat)	http://www.imgt.org/IMGT_GENE-DB/GENElect?livret=0/
IMGT/HLA	The IMGT/HLA Database provides a specialist database for sequences of the human major histocompatibility complex (HLA)	http://www.ebi.ac.uk/ipd/imgt/hla/
IMGT/LIGM-DB	IMGT/LIGM-DB is the first and the largest database of IMGT®, the international ImMunoGeneTics information system®, the high-quality integrated knowledge resource specialized in IG, TR, major histocompatibility complex (MHC) of human and other vertebrate species, and related proteins of the immune system (RPI) that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF)	http://www.imgt.org/cgi-bin/IMGTlect.jv/
Interferon Stimulated Gene Database	Contains over more than 400 microarray based interferon stimulated gene information	http://www.lerner.ccf.org/labs/williams/xchip-html.cgi
IPD-ESTDAB	ESTDAB aims to provide a service enabling investigators to search on-line for HLA-typed, immunologically-characterized tumor cells available for distribution from a central bank	http://www.ebi.ac.uk/ipd/estdab/
IPD-HPA - Human Platelet Antigens	The database provides a centralized repository for the data which define the human platelet antigens (HPA)	http://www.ebi.ac.uk/ipd/hpa/
IPD-KIR - Killer-cell Immunoglobulin-like Receptors	The database provides a centralized repository for human KIR sequences. Killer-cell immunoglobulin-like receptors (KIR) have been shown to be highly polymorphic at the allelic and haplotypic level.	http://www.ebi.ac.uk/ipd/kir/
IPD-MHC	The IPD-MHC Database provides a centralized repository for sequences of the major histocompatibility complex (MHC) from a number of different species	http://www.ebi.ac.uk/ipd/mhc/

Table 3 (continued)

Name	Purpose	URL
MHCBN	The MHCBN is a curated database consisting of detailed information about major histocompatibility complex (MHC) binding, non-binding peptides and T-cell epitopes	http://www.imtech.res.in/raghava/mhcbn/
MHCPEP	This data repository bridges the gap between immunological and computer science/machine learning communities by providing preprocessed and scaled immunological data sets suitable for use in machine learning applications	http://bio.dfci.harvard.edu/DFRMLI/
MUGEN Mouse Database	Welcome to the MUGEN Mouse Database (MMdb) a fully searchable database of murine models of immune processes and immunological diseases	http://bioit.fleming.gr/mugen/mde.jsp
Protegen	Protective antigens are specifically targeted by the acquired immune response of the host and are able to induce protection in the host against infectious and non-infectious diseases	http://www.violinet.org/protegen/
SuperHapten	The current version of the database compiles 2D/3D structures, physicochemical properties and references for about 7500 haptens and 25,000 synonyms	http://bioinformatics.charite.de/superhapten/
VBASE2	VBASE2 is an integrative database of germ-line variable genes from the immunoglobulin loci of human and mouse	http://www.vbase2.org/
Epitome	Epitome is a database of all known antigenic residues and the antibodies that interact with them, including a detailed description of residues involved in the interaction and their sequence/structure environments	http://rostlab.org/services/epitome/

predicted next, as these are critical elements for generation of specific humoral and cell-mediated responses. BCPREDS is a widely used web-based tool for predicting B-cell epitopes based on the support vector machine algorithm with an accuracy of 74.5 % (El-Manzalawy et al. 2008). NetChop is another tool (based on a machine-learning approach using neural networks) meant for predicting T-cell epitopes; it has a prediction accuracy of 70 %. It combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope (Saxova et al. 2003; Nielsen et al. 2005).

Sequences predicted to have all the above properties are then analyzed for epitope topology which is helpful in interpreting the likely antigen binding site. Epitope binding and mapping are done using the Pepitope software to accurately identify discontinuous epitopes from any given protein sequence. This approach is based on two algorithms, PepSurf and Mapitope, to yield 3D structures of antigenic proteins, a step necessary to study the binding efficacy with antibodies. PepSurf aligns each peptide to a graph that represents the surface of the 3D structure. Each aligned peptide corresponds to a path of residues on the 3D structure that exhibits a high similarity to the input peptide. The resulting paths are clustered and the epitope location is inferred. The Mapitope first identifies pairs of residues that are significantly overrepresented in the panel of peptides, compared to their expected frequencies, and then searches for patches on the surface that are enriched with these pairs. The resulting patches are presented on the 3D protein structure (Mayrose et al. 2007). Binding efficacy can be predicted only for the available 3D protein structure. It is also an indispensable tool for modeling the structure in order to identify the potential relevancy using diversified approaches, such as homology modeling, threading, and ab initio methods. Homology modeling predicts 3D models for a protein that are related to at least one known protein sequence/structure in the data repositories. Tools used for homology Modeling are MODELLER (Sali and Blundell 1993), HHPred (Soding 2005), YASARA (Krieger et al. 2002), RaptorX (Kallberg et al. 2012), SWISS-MODEL (Arnold et al. 2006), WHAT IF (Vriend 1990), BHAGEERATH-H (Jayaram et al. 2012), and CPHModels (Nielsen et al. 2010).

Among these tools, MODELLER automatically calculates a model containing all non-hydrogen atoms, making it a highly accurate tool for comparative protein structure modeling. It satisfies spatial restraints, and also performs many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure according to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, and comparison of protein structures (Sali and Blundell 1993). When homologous or weakly homologous sequences of a known structure are unavailable, the alternative successful structure prediction method has been the threading method used to predict secondary structure and local structure motifs. This deals with modeling the query protein based on similar secondary

Table 4 Immunoinformatics tools

Name	Function	URL	Algorithm	Accuracy	Stand-alone/Web based
VaxiJen	It is the first server for alignment-independent prediction of protective antigens	http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html	Auto cross covariance (ACC).	70–89 %	Web-based
Vaxign	Vaccine target prediction and analysis tool	http://www.violinet.org/vaxign/	Pipeline of software programs	–	Web-based
PSortb	Used for bacterial protein subcellular localization prediction	http://www.psort.org/psortb	Support vector machine (SVM)	80–95 %	Stand-alone and web-based
CELLO	Used for bacterial and eukaryotic protein subcellular localization prediction	http://cello.life.nctu.edu.tw/	SVM	87–88.9 %	web-based
WoLF PSORT	Wolf PSORT converts protein amino acid sequences into numerical localization features	wolfpsort.org	k-nearest neighbor	70 %	Stand-alone (Apache and Linux) and Web-based
BEST	Predicts B cell epitopes from antigen sequences	http://biomine.ece.ualberta.ca/BEST/	SVM	74.5 %	Stand-alone
SVMTriP	It is capable of recognizing viral peptides from a human protein sequence background	http://sysbio.unl.edu/SVMTriP	SVM	80.1 %	Web-based
BCPREDS	A tool for predicting linear B-cell epitopes	http://ailab.ist.psu.edu/bcpreds/predict.html	SVM	74.5 %	Web-based
TMHMM	Used for transmembrane helix prediction	http://www.cbs.dtu.dk/services/TMHMM/	Hidden Markov model (HMM)	77 %	Stand-alone (UNIX) and Web-based
LocateP	This pipeline is designed such that it mimics protein targeting and secretion processes. Used for prediction of sub-cellular localization of proteins	http://www.cmbi.ru.nl/locatep-db/cgi-bin/locatepdb.py	TMHMM 2.0, SignalP 3.0, PrediSi, Phobius, CELLO and Psortb	90 %	Web-based
ProPred	Predicting MHC class II binding regions in antigenic protein sequences	http://www.imtech.res.in/raghava/propred/	Matrix based prediction algorithm	80 %	Web-based
T-Epitope Designer	It facilitates HLA-peptide binding prediction	http://www.bioinformation.net/ted/	Uses 3D models of MHC peptides and Q matrix	60 %	Web-based
Pepitope	It predicts discontinuous epitopes	http://pepitope.tau.ac.il/	PepSurf and Mapitope		Web-based
MimoPro	Used for epitope prediction	http://informatiin.nenu.edu.cn/MimoPro	Dynamic programming and branch and bound optimization	54 %	Web-based
ProPred 1	ProPred1 is an on-line web tool for the prediction of peptide binding to MHC class-I alleles	http://www.imtech.res.in/raghava/propred1/	Matrix based approach	80 %	Web-based
NetChop	It combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope	http://tools.immuneepitope.org/stools/netchop/netchop.do?app=netchop	Neural network	70 %	Web-based

structures available from secondary protein structure databases. Tools used for this analysis, among the many available, are Raptor X (Kallberg et al. 2012), HHPred (Soding 2005), SUPERFAMILY (Gough et al. 2001), and Phyre2 (Kelley and Sternberg 2009). PHYRE2 is the most popular and widely used one owing to its accuracy; it was designed to ensure a user-friendly interface for inexperienced users in protein structure prediction methods and is free for noncommercial users. Phyre2 uses the alignment of hidden Markov models via HHsearch (Söding 2005) to significantly improve the accuracy of alignment and detection rate. It also incorporates a new ab initio folding simulation called Poing (Jefferys and Kelley 2010) to model the regions of proteins that have no detectable homology to known structures. Poing is also used to combine multiple templates.

If, in the end, no similarity is determined between the query protein sequence and the available protein structures, either at the secondary or tertiary level, the ab initio strategy is employed. This strategy applies energy values and global energy minimizations after model building. The tools used for protein modeling based on this approach are QUARK (Xu and Zhang 2012), I-TASSER (Zhang 2008), and PEP-FOLD (Thevenet et al. 2012). CASP9 accredits QUARK as the best online tool for ab initio protein modeling with the limitation that it accepts only proteins with less than 20 amino acids. QUARK is an online server that works on replica-exchange Monte Carlo simulation under the guidance of an atomic-level knowledge-based force field method. I-TASSER runs on threading fragment structure reassembly methods on an online server and also as a standalone. It provides five full-length models with the confidence score, the estimated TM-score and RMSD, and the standard deviation of the estimations. Though I-TASSER is a highly used tool owing to its accuracy, its limitation is that it accepts only a single task at a time for a single registered user and requires an academic e-mail ID for registration.

Understanding the binding of the antigenic protein to a specific MHC class allele or a T-/B-cell receptor is of paramount importance as the strength and specificity of the immune response is ultimately based on these events. This critical information can be obtained using protein–protein docking tools, including PatchDock (Schneidman-Duhovny et al. 2005), PIPs (McDowall et al. 2009), PIPE (Pitre et al. 2006), APID (Prieto and De Las Rivas 2006), Protopia (Real-Chicharro et al. 2009), and Cluspro (Comeau et al. 2004). Among these, PatchDock has been successfully tested in three rounds of CAPRI and found to generate near-native results. PatchDock algorithm is adapted from object recognition and image segmentation techniques used in Computer Vision. This tool executes docking of proteins with three main steps: molecular shape representation, surface patch matching and filtering, and scoring (Schneidman-Duhovny et al. 2005).

Ultimately, prediction of a protein structure provides a constructive way of searching for the best conformations of antigen–antibody, antigen–MHC class, and antigen–receptor binding. Lowest energy values during binding assure optimal binding. A variety of tools are available for such molecular dynamics implementation, such as AMBER Package (Case et al. 2005), TINKER (Jay 2014), MDAPI (Hardy 2014), MINDY (Justin 2014), and FeMD (Glavaš 2014). AMBER is an effective tool that involves two different applicative modes: a set of molecular mechanical force fields for simulation of biomolecules and a package of molecular simulation programs that includes source code and demos (Case et al. 2005).

Apart from the above-mentioned tools and databases for computational vaccine approach, very few vaccine database or vaccine candidate prediction pipelines are currently available, among which VIOLIN [http://www.violinet.org/] is a well-curated database that includes 3453 vaccines for 210 pathogens or non-infectious diseases (e.g., cancer). In addition, it allows the user to perform various simple and advanced searches to retrieve vaccine information (Xiang et al. 2008). The Web-based vaccine prediction tool, VaxiJen, is the first server for alignment-independent prediction of protective antigens and subunit vaccines, which is based on auto cross covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. This tool predicts protective antigens and subunit vaccines from the selected virus, bacteria, parasite, fungi, and tumor with 70 to 89 % accuracy (Doytchinova and Flower 2007). On the other hand, Vaxign is a vaccine target prediction and analysis tool for microbial genomic and protein sequences. This tool includes a pipeline of software programs, such as PSORTb: bacterial localization prediction tool, TMHMM: Prediction of transmembrane helices in proteins, SPAAN: Prediction of adhesins and adhesin-like proteins, BLAST: NCBI sequence similarity alignment and analysis program, and IEDB: The Immune Epitope Database and Analysis Resource. The predictable features in the Vaxign pipeline include antigen sublocation, adhesion, epitope binding to MHC class I and class II, and sequence similarities to human, mouse, and/or pig proteins (He et al. 2010a, b).

Computational vaccine adjuvant designing

Vaccine adjuvants are compounds that enhance the specific immune responses against coadministered antigens in vaccines (Petrovsky and Aguilar 2004; Marciani 2003). The various adjuants used are aluminum salts, Freund's complete adjuvant (FCA) which an emulsion of water and mineral oil containing killed mycobacteria, Freund's incomplete adjuvant (FIA) which does not have mycobacteria and is less toxic, lipopolysaccharides (LPSs) from Gram-negative bacteria, and

muramyldipeptide (MDP). Currently, several hundred natural and synthetic compounds are known to have adjuvant activity (Petrovsky and Aguilar 2004; Sayers et al. 2012). However, in the USA, alum salts and AS04 (composed of aluminum salt and MPL) are only licensed for use in humans (Giannini et al. 2006). The mode of action of adjuvants is (1) immunostimulation, i.e., the ability to modify the cytokine network, (2) presentation of the antigen to the appropriate immune effector cells, (3) CTL induction of CD8+ cytotoxic T-lymphocyte (CTL) responses, (4) targeting, i.e., the ability to deliver an immunogen to immune effector cells, and (5) depot generation, i.e., the generation of a short-term or long-term depot to give a continuous or pulsed release (Cox and Coulter 1997).

Bioinformatics approach to vaccine adjuvant designing is a potential new research area for the researchers as it is in an early developmental stage. No significant tools have thus far been developed in this field; however, the currently available bioinformatics tools and databases are used for manual analysis, which includes literature mining, ontology data integration, and analysis, and bioinformatics data analysis (He 2014). Vaxjo (http://www.violinet.org/vaxjo/) a web-based well-curated vaccine adjuvant database is developed by a group of researchers from the United States of America. This database stores the information of more than 100 vaccine adjuvants and approximately 400 vaccines that use the adjuvants. To improve data integration and automated reasoning, the vaccine adjuvant-related data have been logically represented in the Vaccine Ontology (VO). VO has been found to be able to improve literature mining of gene–gene integration (Ozgür et al. 2011). Omics and informatics are also critical to the designing of a personalized vaccine (including vaccine adjuvant), which reduce the adverse reactions (Poland et al. 2008).

Conclusion

The computational vaccinology approach can be widely used and developed as it provides a variety of advantages over traditional approaches to vaccine development. Some of the advantages include virtual screening of every antigen, approaching non-cultivable microorganisms, determining non-profuse antigens, identifying nonimmunogenic antigens, knowing antigens expressed during infectious stages, finding antigens not expressed in vitro, and implementing non-structural proteins in the prediction. However, there still exist some disadvantages such as difficulty in predicting non-proteinous antigens and the requirement of biological determination for the prediction of antigenic proteins (Rappuoli 2000). On the other hand, computational vaccinology is yet to be the primary procedure followed for vaccine development in this era, despite the significant advantage of least time taken in mining the antigenic protein of a pathogen (Fig. 1).

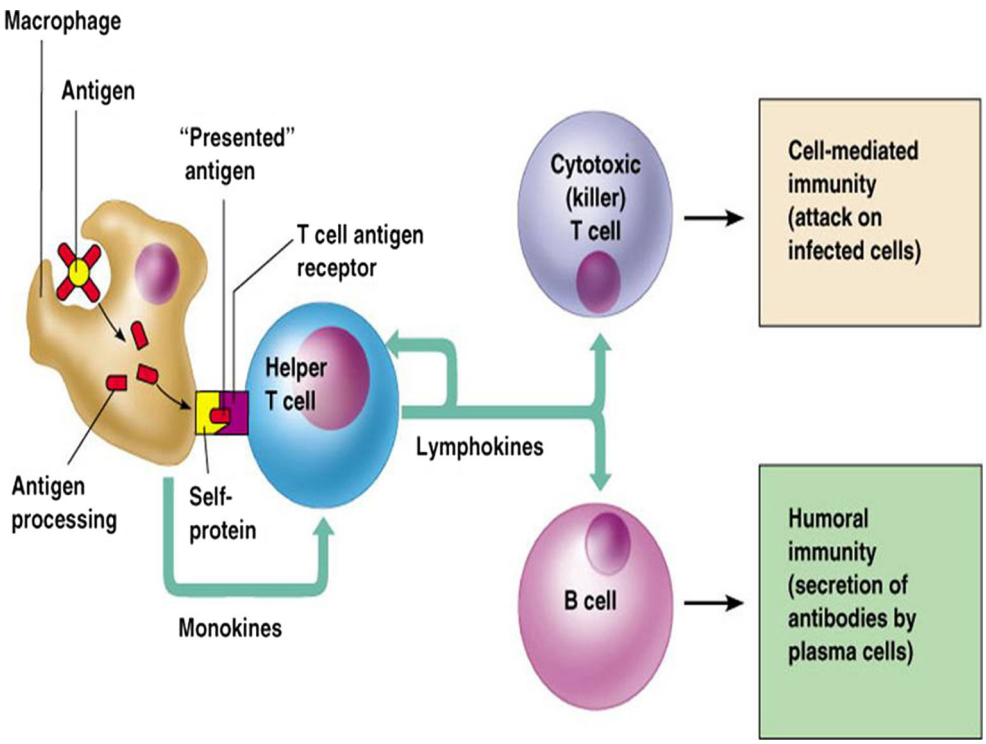

Fig. 1 Antigen presentation

References

Alexandersen S (1996) Advantages and disadvantages of using live vaccines risks and control measures. Acta Vet Scand 90:89–100

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2

André FE (2003) Vaccinology: past achievements, present roadblocks and future promises. Vaccine 21(7–8):593–5

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. doi:10.1093/bioinformatics/bti770

Bambini S, Rappuoli R (2009) The use of genomics in microbial vaccine development. Drug Discov Today 14:252–260. doi:10.1016/j.drudis.2008.12.007

Case DA et al (2005) The amber biomolecular simulation programs. J Comput Chem 26:1668–1688. doi:10.1002/jcc.20290

Clem AS (2011) Fundamentals of vaccine immunology. J Glob Infect Dis 3(1):73–8. doi:10.4103/0974-777X.77299

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 32:W96–99. doi:10.1093/nar/gkh354

Cox JC, Coulter AR (1997) Adjuvants–a classification and review of their modes of action. Vaccine 15(3):248–56

de Groot AS, Sbai H, Aubin CS, McMurry J, Martin W (2002) Immunoinformatics: mining genomes for vaccine components. Immunol Cell Biol 80:255–269. doi:10.1046/j.1440-1711.2002.01092.x

Donnelly JJ, Wahren B, Liu MA (2005) DNA vaccines: progress and challenges. J Immunol 175(2):633–9

Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinforma 8:4

El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit : JMR 21:243–255. doi:10.1002/jmr.893

Faldella G, Alessandroni R, Magini GM et al (1998) The preterm infant's antibody response to a combined diphtheria, tetanus, acellular pertussis and hepatitis B vaccine. Vaccine 16(17):1646–9

Foged C (2011) Subunit vaccines of the future: the need for safe, customized and optimized particulate delivery systems. Ther Deliv 2(8):1057–77

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. doi:10.1093/bioinformatics/bts565

Giannini SL, Hanon E, Moris P et al (2006) Enhanced humoral and memory B cellular immunity using HPV16/18 L1 VLP vaccine formulated with the MPL/aluminium salt combination (AS04) compared to aluminium salt only. Vaccine 24(33–34):5937–49

Glavaš G (2014) Front-end for Molecular dynamics (FeMD). http://www.globulus.net/femd/. Accessed 18 January 2014

Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol 313:903–919. doi:10.1006/jmbi.2001.5080

Hardy DJ (2014) MDAPI, Molecular Dynamics Applications Programming Interface;. http://www.ks.uiuc.edu/Development/MDTools/mdapi/. Accessed 18 January 2014

He Y (2014) Vaccine adjuvant informatics: from data integration and analysis to rational vaccine adjuvant design. Front Immunol 5:32. doi:10.3389/fimmu.2014.00032

He Y, Rappuoli R, de Groot AS, Chen RT (2010a) Emerging vaccine informatics. J Biomed Biotechnol 2010:218590. doi:10.1155/2010/218590

He Y, Xiang Z, Mobley HL (2010b) Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. J Biomed Biotechnol 2010:297505. doi:10.1155/2010/297505

Herbert A (2014) MaxCluster: A tool for Protein Structure Comparison and Clustering. http://www.sbg.bio.ic.ac.uk/~maxcluster/. Accessed 5 January 2014

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26(5):680–2. doi:10.1093/bioinformatics/btq003

Immunization (2014) http://pathmicro.med.sc.edu/ghaffar/immunization-ver2.htm. Accessed 18 January 2014

Isticato R, Cangiano G, Tran HT, Ciabattini A et al (2011) Surface display of recombinant proteins on Bacillus subtilis spores. J Bacteriol 183(21):6294–301

Iwanicki A, Piątek I, Stasiłojć M et al (2014) A system of vectors for Bacillus subtilis spore surface display. Microbiol Cell Fact 13(1):30. doi:10.1186/1475-2859-13-30

Jay P (2014) TINKER Molecular Modeling. http://dasher.wustl.edu/ffe. Accessed 30 March 2014

Jayaram B, Dhingra P, Lakhani B, Shekhar S (2012) Bhageerath—Targeting the near impossible: pushing the frontiers of atomic models for protein tertiary structure prediction. J Chem Sci 124:83–91. doi:10.1007/s12039-011-0189-x

Jefferys BR, Kelley LA (2010) Sternberg MJ. Protein folding requires crowd control in a simulated cell. J Mol Biol 397(5):1329–38. doi:10.1016/j.jmb.2010.01.074

John L, John GJ, Kholia T (2012) A reverse vaccinology approach for the identification of potential vaccine candidates from Leishmania spp. Appl Biochem Biotechnol 167:1340–1350. doi:10.1007/s12010-012-9649-0

Justin G (2014) Mindy - A 'minimal' molecular dynamics program. http://www.ks.uiuc.edu/Development/MDTools/mindy/. Accessed 18 January 2014

Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7:1511–1522. doi:10.1038/nprot.2012.085

Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371. doi:10.1038/nprot.2009.2

Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. Proteins 47:393–402

Kronman MP, Zaoutis TE, Haynes K, Feng R, Coffin SE (2012) Antibiotic exposure and IBD development among children: a population-based cohort study. Pediatrics 130:e794–803. doi:10.1542/peds. 2011-3886

Li W, Jaroszewski L, Godzik A (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics 17(3):282–3

Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011:bar009. doi: 10.1093/database/bar009.Print 2011

Major M, Moss S, Gold R (2011) From genes to vaccine: a breakthrough in the prevention of meningococcal group B disease. Paediatr Child Health 16:e61–64

Marciani DJ (2003) Vaccine adjuvants: role and mechanisms of action in vaccine immunogenicity. Drug Discov Today 8(20):934–943

Mayrose I et al (2007) Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics 23:3244–3246. doi:10.1093/bioinformatics/btm493

McDowall MD, Scott MS, Barton GJ (2009) PIPs: human protein-protein interaction prediction database. Nucleic Acids Res 37:D651–656. doi:10.1093/nar/gkn870

Nebeker JR, Barach P, Samore MH (2004) Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med 140:795–801

Nielsen M, Lundegaard C, Lund O, Kesmir C (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics 57:33–41. doi:10.1007/s00251-005-0781-7

Nielsen M, Lundegaard C, Lund O, Petersen TN (2010) CPHmodels-3.0–remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res 38:W576–581. doi:10.1093/nar/gkq535

NIAID NIoAaID (2013). http://www.niaid.nih.gov/topics/vaccines/Pages/Default.aspx. Accessed 27 December 2013

Ozgür A, Xiang Z, Radev DR, He Y (2011) Mining of vaccine-associated IFN-γ gene interaction networks using the Vaccine Ontology. J Biomed Semantics 17;2 Suppl 2:S8. doi:10.1186/2041-1480-2-S2-S8

Pande J, Szewczyk MM, Grover AK (2010) Phage display: concept, innovations, applications and future. Biotechnol Adv 28(6):849–58. doi:10.1016/j.biotechadv.2010.07.004

Petrovsky N, Aguilar JC (2004) Vaccine adjuvants: current state and future trends. Immunol Cell Biol 82(5):488–496

Pitre S et al (2006) PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. BMC Bioinforma 7:365. doi:10.1186/1471-2105-7-365

Pizza M et al (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820

Plotkin SA (2003) Vaccines, vaccination, and vaccinology. J Infect Dis 187(9):1349–59

Plotkin SA (2005) Vaccines: past, present and future. Nat Med 11(4 Suppl):S5–11

Plotkin SA (2008) New vaccination strategies. Bull Acad Natl Med 192:511–518, discussion 518–519

Pluimers FH (2004) Foot-and-Mouth disease control using vaccination: the Dutch experience in 2001. Dev Biol (Basel) 119:41–9

Poland GA, Ovsyannikova IG, Jacobson RM (2008) Personalized vaccines: the emerging field of vaccinomics. Expert Opin Biol Ther 8(11):1659–67. doi:10.1517/14712598.8.11.1659

Poppe WA (2001) Nitroimidazole-resistant vaginal trichomoniasis treated with paromomycin. Eur J Obstet Gynecol Reprod Biol 96:119–120

Prieto C, de las Rivas J (2006) APID: agile protein interaction dataanalyzer. Nucleic Acids Res 34:W298–302. doi:10.1093/nar/gkl128

Principles of Vaccination (2014) http://www.cdc.gov/vaccines/pubs/pinkbook/downloads/prinvac.pdf. Accessed 18 January 2014

Rappuoli R (2000) Reverse vaccinology. Curr Opin Microbiol 3:445–450

Real-Chicharro A et al (2009) Protopia: a protein-protein interaction tool. BMC Bioinforma 10(Suppl 12):S17. doi:10.1186/1471-2105-10-S12-S17

Robert E (2014) USEARCH: Ultra-fast Sequence analysis. http://drive5.com/usearch. Accessed 5 January 2014

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815. doi:10.1006/jmbi.1993.1626

Saxova P, Buus S, Brunak S, Kesmir C (2003) Predicting proteasomal cleavage sites: a comparison of available methods. Int Immunol 15:781–787

Sayers S, Ulysse G et al (2012) Vaxjo: a web-based vaccine adjuvant database and its application for analysis of vaccine adjuvants and their uses in vaccine development. J Biomed Biotechnol 2012:831486. doi:10.1155/2012/831486

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:W363–367. doi:10.1093/nar/gki481

Serruto D, Serino L, Masignani V, Pizza M (2009) Genome-based approaches to develop vaccines against bacterial pathogens. Vaccine 27:3245–3250. doi:10.1016/j.vaccine.2009.01.072

Sibley L, Reljic R, Radford DS, Huang JM, Hong HA, Cranenburgh RM (2014) Cutting SM.Recombinant Bacillus subtilis spores expressing MPT64 evaluated as a vaccine against tuberculosis in the murine model. FEMS Microbiol Lett 358(2):170–9. doi:10.1111/1574-6968.12525

Singh H, Raghava GP (2001) ProPred: prediction of HLA-DR binding sites. Bioinformatics 17:1236–1237

Singh H, Raghava GP (2003) ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 19:1009–1014

Sinha JK & Bhattacharya S (2014) A Text Book of Immunology (Google Book Preview). Academic Publishers. p. 318. ISBN 978-81-89781-09-5. Retrieved 2014-01-09

Soding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21:951–960. doi:10.1093/bioinformatics/bti125

Söding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21(7):951–60

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc / Int Conf Intell Sys Mol Biol; ISMB Int Conf Intell Sys Mol Biol 6:175–182

Sood S, Arti K (2008) An update on Trichomonas vaginalis. Indian J Sex Transm Dis 29:7–14

Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P (2012) PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res 40:W288–293. doi:10.1093/nar/gks419

van Duin D, Medzhitov R, Shaw AC (2006) Triggering TLR signaling in vaccination. Trends Immunol 27(1):49–55

van Houten NE, Henry KA, Smith GP, Scott JK (2010) Engineering filamentous phage carriers to improve focusing of antibody responses against peptides. Vaccine 28(10):2174–85. doi:10.1016/j.vaccine.2009.12.059

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8:52–56

WHO (2014) Tropical Disease Research (TDR). http://www.who.int/tdr/en/. Accessed 10 January 2014

Xiang Z, Todd T, Ku KP, Kovacic BL et al. (2008) VIOLIN: Vaccine Investigation and Online Information Network. Nucleic Acids Res 36(Database issue):D923-8

Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80:1715–1735. doi:10.1002/prot.24065

Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci : a Publ Protein Soc 13:1402–1406. doi:10.1110/ps.03479604

Yu NY et al (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. doi:10.1093/bioinformatics/btq249

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinforma 9:40. doi:10.1186/1471-2105-9-40