Published in final edited form as:

Curr Opin Virol. 2015 April ; 11: 103–112. doi:10.1016/j.coviro.2015.03.013.

Computational tools for epitope vaccine design and evaluation

Linling He¹ and Jiang Zhu^{1,2,3,*}

¹Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA

²Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California 92037, USA

³Department of Integrative Structural and Computational Biology, The Scripps Research Institute La Jolla, California 92037, USA

Abstract

Rational approaches will be required to develop universal vaccines for viral pathogens such as human immunodeficiency virus, hepatitis C virus, and influenza, for which empirical approaches have failed. The main objective of a rational vaccine strategy is to design novel immunogens that are capable of inducing long-term protective immunity. In practice, this requires structure-based engineering of the target neutralizing epitopes and a quantitative readout of vaccine-induced immune responses. Therefore, computational tools that can facilitate these two areas have played increasingly important roles in rational vaccine design in recent years. Here we review the computational techniques developed for protein structure prediction and antibody repertoire analysis, and demonstrate how they can be applied to the design and evaluation of epitope vaccines.

Rational vaccine design: a brief background

Empirical approaches have resulted in a rich catalog of efficacious human vaccines. However, these approaches have failed for infectious diseases such as human immunodeficiency virus type-1 (HIV-1), hepatitis C virus (HCV), and influenza [1•,2]. Driven by the continuous discovery of broadly neutralizing antibodies (bnAbs) [3•], an antibody-based rational approach has begun to emerge in HIV-1 vaccine research [4–6•]. Structures of bnAbs in complex with epitope peptides [7,8], envelope (Env) glycoproteins [9–13], and the native viral spike [14–17] have provided a detailed picture of vaccine targets. Next-generation sequencing (NGS) has enabled an in-depth understanding of the diversity and evolution of bnAbs during chronic infection [18–24•]. Previous attempts using the rational approach to develop immunogens targeting the membrane proximal external region (MPER) [25–28•] and the CD4-binding site (CD4bs) of HIV-1 [29]

© 2015 Published by Elsevier B.V.

*To whom correspondence should be addressed: JZ: Phone (858)784-8157; jiang@scripps.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

reported no neutralization. However, a similar approach towards respiratory syncytial virus (RSV) was successful in that RSV-neutralizing antibodies were elicited in rhesus macaques [30••]. This study demonstrated what can be potentially achieved by computational design. A general strategy for epitope vaccine design is illustrated in Figure 1A.

Computational tools for structure-based immunogen design

Protein structure prediction can be divided into template-based and free modeling, with a large number of computational tools available (Figure 1B) [31–37•]. Template-based modeling aims to build a model based on the structures of evolutionarily related proteins [32–34], and a typical workflow involves template selection, sequence alignment, model building, quality assessment, and structure refinement [35]. Free modeling, however, often relies on complicated procedures to render an initial model [38–41]. As fold recognition has become increasingly more effective in detecting remote homologs, the boundaries between the two prediction methods are often blurred [42,43]. Individual scoring functions or a composite score that combines multiple terms with machine learning can be used to identify problematic regions in a predicted model or select the best model from a pool of candidates [44–46]. The predicted model can be further refined using a range of modeling and simulation techniques to improve the local or global quality.

Side chain modeling tools

In conventional protein design, the combinatorial space of side chain conformations (rotamers) of twenty amino acids is exhaustively searched to identify the global minimum [47–50]. Extensive efforts have been devoted to developing energy functions [51–53], search algorithms [54], and rotamer libraries [55,56]. Programs such as SCWRL [57•] and SCAP [58•] provide convenient tools to model, predict, and mutate protein side chains *in silico*. Side chain modeling tools have been used to (1) optimize the antigen-interacting surface of the antibody (paratope); (2) resurface the antigen by varying the sequence of non-epitope regions; and (3) optimize the antibody-interacting surface of the antigen (epitope) (Figure 2A).

Side chain modeling tools have been used to optimize antibody affinity and specificity [59,60]. Clark *et al.* improved the affinity of an antibody targeting the I-domain of integrin VLA1 by an order of magnitude using a hierarchical procedure that combines energy functions and search algorithms of different resolutions [61]. Lippow *et al.* observed a 10 to 140-fold improvement in affinity for two antibodies using a physical energy function in conjunction with exhaustive search algorithms [62•]. Despite these successes, incorporating backbone flexibility into protein design remains a challenge [63]. To tackle this problem, ensemble-based methods can be used to generate a large number of backbone conformations in either Cartesian or torsional space [64].

Resurfacing of non-epitope regions has been applied to increase the solubility and stability of a designed antigen and to create antigen variants with an intact epitope. Correia *et al.* combined resurfacing and flexible backbone modeling to design antigens for the HIV-1 MPER epitope recognized by bnAb 4E10, using RosettaDesign to identify distinct mutations [25]. Wu *et al.* adopted a similar strategy in engineering a resurfaced stabilized gp120 core

(RSC3), which was used as a B-cell sorting probe to isolate VRC01, a potent CD4bs-directed bnAb [65], and a class of VRC01-like bnAbs from HIV-1-infected individuals [21].

Side chain modeling tools can be adapted to model glycan epitopes. As shown by recent studies, *N*-linked glycans are major structural elements of several HIV-1 epitopes recognized by bnAbs [8,12,14–16]. Computational tools have been developed for building a rigid glycan structure on the protein surface but these force field methods are often computationally costly and parameter-sensitive [66–68]. Therefore, a robust modeling program based on empirical principles would be highly desirable in designing and engineering glycan epitopes on immunogens. Structurally, an *N*-linked glycan can be considered as an extension of the asparagine side chain, and as such, many techniques originally developed for side chain modeling can be applied to glycan modeling. To illustrate this possibility, we glycosylated an HIV-1 gp120 core (PDB ID: 3NGB) *in silico* with an in-house program that combines a simple scoring function, a stochastic search algorithm [58], and a discrete glycan rotamer library (Figure 2B).

Backbone modeling tools

Scaffolding, the grafting of an epitope of interest onto a heterologous protein scaffold, has been proposed as a solution for epitope-focused immunogen design [69]. The concept of scaffold has long been adopted in the development of protein therapeutics and diagnostics, such as cysteine knots [70•,71], DARPins [72], and immunoglobulin-like proteins [73]. Only recently has this concept been extended into vaccine design [69]. For the HIV-1 4E10 epitope, Correia *et al.* designed 10 scaffold antigens using a “side-chain grafting” method [26]. In a follow-up study, the authors resurfaced and adopted the fragment-based approach in Rosetta, termed “flexible-backbone remodeling”, to optimize a 4E10 epitope scaffold [25]. For the HIV-1 2F5 epitope, Ofek *et al.* applied MAMMOTH [74] to identify protein scaffolds and Rosetta to design scaffold antigens [28]. A similar strategy was applied to the antigenic site II of the RSV fusion protein recognized by motavizumab [75]. In this study, the search algorithm, termed “multi-segment side-chain grafting”, was used to match two helices in a consecutive manner. A variant of this method, termed “multigraft design”, was utilized to graft the CD4-binding loop and the outer-domain exit loop of HIV-1 gp120 onto a scaffold [29]. In a breakthrough study, a computational procedure termed “fold from loops” was used to design an epitope scaffold that elicited RSV-neutralizing antibodies in rhesus macaques [30]. Recently, a large-scale scaffolding study was reported for three major HIV-1 epitopes, with up to 50% of the designs bound to their respective bnAbs [76•]. This study highlighted the value of streamlined computational design and antigen screening in the early phase of immunogen discovery.

Here we elaborate on the criteria for selecting suitable scaffolds for epitope grafting in the design process (Figure 3A). Smaller scaffolds are preferred in that they can focus antibody responses to the grafted epitope while minimizing unwanted immunogenicity (Figure 3B, left). A small epitope scaffold can also be considered as a display unit in the context of a larger, multivalent carrier. For example, a scaffold with the N- and C-termini in close proximity can be either inserted into a surface loop or fused to the terminus of a carrier protein providing crucial T-helper epitopes in immunization. Flexibility is another criterion

in scaffold selection (Figure 3B, middle). Many RNA viruses have evolved an evasion strategy using variable loops as “decoy epitopes” to distract the immune response from conserved neutralizing epitopes. In a similar manner, target epitopes can be displayed using inherently flexible scaffolds to improve their immunogenicity. This possibility was illustrated with the analysis of the conformational entropy – a metric of flexibility – for five previously reported 2F5 epitope scaffolds [28]. The 1D3B-based design, which yielded the highest entropy value, elicited high titers of epitope-specific antibody responses, suggesting a possible correlation between flexibility and immunogenicity. The third criterion for scaffold selection is the structural environment of the graft (Figure 3B, right). Correia *et al.* designed a set of 4E10 epitope scaffolds by grafting the helical epitope onto protein scaffolds with long extended helices [26]. Such designs can elicit non-specific antibodies to exposed regions overlapping the epitope and the scaffold along the helical axis. Alternatively, presenting an epitope either in a different secondary structural environment or with a well-defined structural boundary may enhance the specificity of elicited antibodies and facilitate the NGS-based dissection of epitope-specific antibody lineages (see below). The 3C8I- and 3MHS-based 10E8 epitope scaffolds provided examples of these two design scenarios [76].

Meta-server and consensus-based design

Previous scaffolding studies have been carried out using primarily a single algorithm, resulting in a pool of scaffolds specific only to the algorithm used [25–30,75]. To illustrate the limitations of relying on any single scaffolding algorithm, we compared MAMMOTH [74] to TM-align [77], two algorithms used in recent design practices [28,76]. Using a C_α-RMSD cutoff of 1.5Å, TM-align identified significantly more scaffolds for the 2F5 epitope compared to MAMMOTH (539 versus 273), with an overlap of 133 scaffolds. Although many structural alignment tools can be used for scaffold search [78], each method will generate a different but overlapping set of candidates. Rather than relying on a single method, a meta-server can render a consensus-based prediction based on multiple methods. This approach has been found superior to any individual method when applied to difficult protein structure prediction problems [79,80]. To evaluate the concept of a scaffolding meta-server, we combined the output from six structural alignment algorithms: two TM-align implementations [77], SPalign [81], CLICK [82], FAST [83], and MAMMOTH [74]. A database of 23,576 protein chains generated by a sequence culling server, PISCES [84], was used in the scaffold search. For the 2F5 epitope, the six scaffolding algorithms identified 262, 539, 331, 266, 152, and 273 scaffolds, respectively, with varying overlaps (Figure 3C). The combined pool of scaffolds can be screened based on the selection criteria described above followed by a clash score determined after rigid-body docking of each scaffold into the antibody-antigen complex.

Scaffold family and sequence profile

Another key concept in protein structure prediction is fold family. Protein domains can be classified into families based on their structural similarities [85–88]. A high sequence similarity is often indicative of a short genetic distance, whereas a low sequence similarity (<30%) accompanied with high structural homology may suggest conserved structure and function in evolution. As the protein structure database often contains multiple members of

the same fold family, it is not uncommon that structurally similar scaffolds can be identified by one or more scaffolding algorithms. For example, TM-align identified three scaffolds of the same fold family for the MPER epitope (Figure 3D), all of which showed 10E8 specificity upon epitope grafting (unpublished data). These three epitope scaffolds will be useful in a heterologous prime-boost strategy to focus the immune responses to the invariable epitope. An experimentally tested scaffold can also be used as a template to identify structural homologs as potential carriers of the same epitope. Another related concept in protein structure prediction is sequence profile, which tabulates the probability of each amino acid type for each position of a polypeptide chain. A sequence profile can be derived from multiple sequence alignment (MSA) of evolutionarily related proteins. A “scaffold profile” can therefore be used as a guideline as to which mutations can be introduced outside the grafted epitope in order to improve the properties of an epitope scaffold and to create resurfacing variants with minimal adverse effects.

Multi-graft and multivalent scaffolding

Composite modeling, the use of multiple templates to build a composite model, has become a common practice in protein structure prediction as it often yields a more accurate model than the use of a single template [89]. The same modeling procedure can be utilized to design multi-graft scaffold immunogens. The unique challenge lies in the identification of scaffolds that can accommodate multiple epitopes and allow antibodies to access these epitopes without occluding one another. To examine the possibility of designing multi-graft immunogens, we transplanted the E1 antigenic site (aa 314–324, an α -helix) and the E2 antigenic site (aa 412–423, a β -hairpin) of HCV [90] onto a unique scaffold, which exhibited high affinity for their respective neutralizing antibodies (unpublished data).

As monomeric proteins can be utilized as scaffolds to present an epitope in the bnAb-bound conformation, nanoparticles and virus-like particles (VLPs) can be considered as multivalent scaffolds to display an antigen in a highly ordered and repetitive array in order to elicit potent immune responses [91–94•]. Attempts to rationally design chimeric nanoparticle immunogens have been previously reported. Jardine *et al.* displayed an engineered gp120 outer domain on a 60-mer particle of the lumazine synthase from the hyperthermophilic bacterium *Aquifex aeolicus* to target the germline precursors of VRC01 [95]. Kanekiyo *et al.* used a 24-mer ferritin particle from *Helicobacter pylori* to present the hemagglutinin (HA) of the H1N1 strain of influenza virus [96]. Zhou *et al.* demonstrated the possibility of presenting PGT128 epitope scaffolds with a ferritin particle, which showed increased binding affinity for PGT128 (Figure 3E) [76••]. We have applied the multivalent scaffolding approach to the two HCV antigenic sites, resulting in a series of bnAb-binding nanoparticles (unpublished data).

Quantitative evaluation of vaccine-induced antibody responses

NGS has enabled unprecedented access to the antibody repertoire [97••]. Novel bioinformatics tools have been developed to identify somatic variants, dissect maturation pathways, and infer lineage intermediates for HIV-1 bnAbs [18–24]. The impact of antibody NGS to rational vaccine design has been highlighted in recent reviews [98••,99••]. Given the rapid development of NGS technologies [100••] and bioinformatics tools, antibody

repertoire analysis is likely to expand from its current role of characterizing bnAb lineages to other critical aspects of vaccine research such as temporal monitoring of animal immunization and human vaccine trial.

Antibodyomics tools

The bioinformatics analysis of an NGS-derived antibody repertoire can be divided into a primary stage consisting of data processing, filtering, and annotation and a second stage of more in-depth analysis (Figure. 4A). For the primary analysis, a computational pipeline termed “*Antibodyomics 1.0*” has been developed and validated using the 454 sequencing data generated for several classes of HIV-1 bnAbs [20–24••]. During the pipeline processing, raw NGS reads are assigned to putative germline genes, error-corrected, compared to known bnAbs, and subjected to a detailed CDR3 analysis, producing a high-quality database for further inquiry. Novel bioinformatics tools have been developed to interrogate the repertoire. Identity/divergence plot can be used to visualize a repertoire along two critical parameters: the sequence identity to a known bnAb (Y-axis) and the sequence divergence from putative germline genes (X-axis) [19–24]. On such plots, closely related somatic variants often form “islands” which can be visually distinguished from the main sequence population. However, distant somatic variants can only be identified by more sophisticated bioinformatics tools such as intra-donor phylogenetic analysis, which searches for sequences with the same evolutionary pattern as the template bnAbs [22,23]. Zhu *et al.* reported the similarity between the heavy and light chain phylogenetic trees for the PGT141–145 lineage, providing a potential solution to the heavy/light pairing problem within an antibody lineage [23••]. Cross-donor phylogenetic analysis was originally developed to illustrate the converged evolution of VRC01-like bnAbs [20,21]. Zhu *et al.* further demonstrated the utility of this method in the *de novo* identification of VRC01-like bnAbs from HIV-1-infected donors [24••]. Phylogenetic tools have also been adapted for inferring ancestral and intermediate antibodies [19,20], providing valuable insights into the critical events of the maturation process. Additionally, CDR3-based lineage analysis can be used to directly identify germline precursors and intermediate antibodies from an NGS-derived repertoire [20,21].

Implications for vaccine design strategy

Structural design and antibody NGS can be integrated into a coherent vaccine strategy in which rationally designed immunogens serve as an “input” to the immune system while NGS offers a quantitative “readout” of the induced antibody responses. Once antigenicity is confirmed for a candidate immunogen, a control immunogen can be created through epitope manipulation – either mutation of bnAb-interacting residues or removal of the entire epitope (Figure 4B). During immunization, blood sample can be collected at the peak of the immune response after each injection with B cell repertoires captured by antibody NGS. In principle, the positive immunogen would elicit epitope-specific antibodies that would not be present in the repertoires of the control group. Independently, longitudinal analysis can be used to trace the maturation process of the immunogen-specific antibody lineages (Figure 4C). As the immunogen-primed B cells expand and affinity mature, the frequencies of the corresponding antibody lineages will increase over time. Of note, epitope manipulation and longitudinal tracing are not mutually exclusive and can in fact be combined to dissect the antibody

lineages with high resolution. Such analysis will further benefit from the novel antibody NGS technologies such as unbiased repertoire capture and single-molecule barcoding [100•]. Germline gene usage, degree of somatic hypermutation, CDR3 signature, and other repertoire properties, together, provide a quantitative antibody profile for the rational evaluation of vaccine immunogens.

Conclusion

Vaccine development against antigenically variable viruses has called for innovative approaches [4–6•]. Owing to the advances in structural biology, genomics and, in particular, computational biology, rational approaches can now be translated from concept to practice. Computational tools developed for structure-based immunogen design and antibody repertoire analysis will likely play an indispensable role in the future development of epitope vaccines.

Acknowledgements

We thank Dennis Burton, Ian Wilson, Michael Zwick, and Mansun Law for helpful discussions and comments on the manuscript. We also thank Arthur Kim for proofreading. Funding was provided by the Scripps Center for HIV/AIDS Immunology & Immunogen Discovery (CHAVIID) UM1 AI-100663).

References and recommended reading

Papers of particular interest, published with the period of review, have been highlighted as:

- of special interest

- of outstanding interest

1. Nabel GJ. Designing Tomorrow's Vaccines. *New Engl J Med*. 2013; 368:551–560. [PubMed: 23388006] • This review covered a wide range of topics in vaccine research, including past successes and current challenges, new conceptual and technological advances, clinical translation and implementation, and future vaccine development.
2. Rappuoli R, Mandl CW, Black S, De Gregorio E. Vaccines for the twenty-first century society. *Nat Rev Immunol*. 2011; 11:865–872. [PubMed: 22051890]
3. Kwong PD, Mascola JR, Nabel GJ. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat Rev Immunol*. 2013; 13:693–701. [PubMed: 23969737] • The authors discussed the role of broadly neutralizing antibodies in structural analysis of antibody-epitope interactions and rational design of immunogens to target the intermediate antibodies during B-cell development.
4. Walker LM, Burton DR. Rational antibody-based HIV-1 vaccine design: current approaches and future directions. *Curr Opin Immunol*. 2010; 22:358–366. [PubMed: 20299194] •• This was one of the earliest articles describing the concept and major components of an antibody-based rational vaccine design approach. The importance of broadly neutralizing antibody identification was highlighted in the review.
5. Haynes BF, Kelsoe G, Harrison SC, Kepler TB. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol*. 2012; 30:423–433. [PubMed: 22565972] •• The authors extended the rational vaccine approach of Walker *et al.* by including the intermediate antibodies derived from a neutralizing B-cell lineage as templates to design a series of immunogens to guide the B-cell development during the vaccination.
6. Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. Antibodies in HIV-1 Vaccine Development and Therapy. *Science*. 2013; 341:1199–1204. [PubMed: 24031012]

7. Huang J, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature*. 2012; 491:406–412. [PubMed: 23151583]
8. McLellan JS, Pancera M, Carrico C, Gorman J, Julien J-P, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*. 2011; 480:336–343. [PubMed: 22113616]
9. Zhou T, Georgiev I, Wu X, Yang Z-Y, Dai K, Finzi A, Kwon YD, Scheid JF, Shi W, Xu L, et al. Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science*. 2010; 329:811–817. [PubMed: 20616231]
10. Zhou T, Xu L, Dey B, Hessel AJ, Van Ryk D, Xiang S-H, Yang X, Zhang M-Y, Zwick MB, Arthos J, et al. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*. 2007; 445:732–737. [PubMed: 17301785]
11. Scharf L, Scheid JF, Lee JH, West AP, Chen C, Gao H, Gnanapragasam PNP, Mares R, Seaman MS, Ward AB, et al. Antibody 8ANC195 Reveals a Site of Broad Vulnerability on the HIV-1 Envelope Spike. *Cell Rep*. 2014; 7:785–795. [PubMed: 24767986]
12. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang P-S, Wang S-K, Stanfield RL, Julien J-P, Ramos A, Crispin M, et al. A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the HIV Glycan Shield. *Science*. 2011; 334:1097–1103. [PubMed: 21998254]
13. Joyce MG, Kanekiyo M, Xu L, Biertuempfel C, Boyington JC, Moquin S, Shi W, Wu X, Yang Y, Yang Z-Y, et al. Outer Domain of HIV-1 gp120: Antigenic Optimization, Structural Malleability, and Crystal Structure with Antibody VRC-PG04. *J Virol*. 2013; 87:2294–2306. [PubMed: 23236069]
14. Blattner C, Lee JH, Sliepen K, Derking R, Falkowska E, de la Pena AT, Cupo A, Julien J-P, van Gils M, Lee PS, et al. Structural Delineation of a Quaternary, Cleavage- Dependent Epitope at the gp41-gp120 Interface on Intact HIV-1 Env Trimers. *Immunity*. 2014; 40:669–680. [PubMed: 24768348]
15. Julien J-P, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse P-J, Burton DR, Sanders RW, Moore JP, et al. Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science*. 2013; 342:1477–1483. [PubMed: 24179159]
16. Lyumkis D, Julien J-P, de Val N, Cupo A, Potter CS, Klasse P-J, Burton DR, Sanders RW, Moore JP, Carragher B, et al. Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. *Science*. 2013; 342:1484–1490. [PubMed: 24179160]
17. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. 2014; 514:455–461. [PubMed: 25296255]
18. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, Ernandes MJ, Georgiev IS, Kim HJ, Pancera M, et al. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*. 2014; 509:55–62. [PubMed: 24590074]
19. Sok D, Laserson U, Laserson J, Liu Y, Vigneault F, Julien J-P, Briney B, Ramos A, Saye KF, Le K, et al. The Effects of Somatic Hypermutation on Neutralization and Binding in the PGT121 Family of Broadly Neutralizing HIV Antibodies. *Plos Pathog*. 2013; 9:e1003754. [PubMed: 24278016]
20. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K, et al. Focused Evolution of HIV-1 Neutralizing Antibodies Revealed by Structures and Deep Sequencing. *Science*. 2011; 333:1593–1602. [PubMed: 21835983] •• In this article, the authors introduced basic concepts and novel bioinformatics tools for the next-generation sequencing analysis of HIV-1 antibody repertoire, which revealed the converged evolution of VRC01-like antibodies from multiple HIV-1-infected donors.
21. Zhou T, Zhu J, Wu X, Moquin S, Zhang B, Acharya P, Georgiev IS, Altae-Tran HR, Chuang G-Y, Joyce MG, et al. Multidonor Analysis Reveals Structural Elements, Genetic Determinants, and Maturation Pathway for HIV-1 Neutralization by VRC01-Class Antibodies. *Immunity*. 2013; 39:245–258. [PubMed: 23911655]
22. Zhu J, O'Dell S, Ofek G, Pancera M, Wu X, Zhang B, Zhang Z, Mullikin JC, Simek M, Burton DR, et al. Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. *Front Microbiol*. 2012; 3:315. [PubMed: 23024643]

23. Zhu J, Ofek G, Yang Y, Zhang B, Louder MK, Lu G, McKee K, Pancera M, Skinner J, Zhang Z, et al. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc Natl Acad Sci USA*. 2013; 110:6470–6475. [PubMed: 23536288] •• This was the first report of using phylogenetic similarity to pair heavy and light chains within an antibody lineage (PGT141-145). Heavy and light chains co-evolve during the maturation process, resulting in two phylogenetic trees with matching branching patterns.
24. Zhu J, Wu X, Zhang B, McKee K, O'Dell S, Soto C, Zhou T, Casazza JP, Mullikin JC, Kwong PD, et al. De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc Natl Acad Sci USA*. 2013; 110:E4088–E4097. [PubMed: 24106303] •• This was the first report of *de novo* identification of broadly neutralizing antibodies from an HIV-1-infected donor solely by next-generation sequencing and bioinformatics, suggesting a new and more effective approach for antibody discovery.
25. Correia BE, Ban Y-EA, Friend DJ, Ellingson K, Xu H, Boni E, Bradley-Hewitt T, Bruhn-Johannsen JF, Stamatatos L, Strong RK, et al. Computational Protein Design Using Flexible Backbone Remodeling and Resurfacing: Case Studies in Structure-Based Antigen Design. *J Mol Biol*. 2011; 405:284–297. [PubMed: 20969873]
26. Correia BE, Ban Y-EA, Holmes MA, Xu H, Ellingson K, Kraft Z, Carrico C, Boni E, Sather DN, Zenobia C, et al. Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. *Structure*. 2010; 18:1116–1126. [PubMed: 20826338]
27. Azoitei ML, Ban Y-EA, Julien J-P, Bryson S, Schroeter A, Kalyuzhnyi O, Porter JR, Adachi Y, Baker D, Pai EF, et al. Computational Design of High-Affinity Epitope Scaffolds by Backbone Grafting of a Linear Epitope. *J Mol Biol*. 2012; 415:175–192. [PubMed: 22061265]
28. Ofek G, Guenaga FJ, Schief WR, Skinner J, Baker D, Wyatt R, Kwong PD. Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Natl Acad Sci USA*. 2010; 107:17880–17887. [PubMed: 20876137] •• The authors combined scaffolding, antigenic characterization, immunization, and X-ray crystallography to demonstrate that epitope-focused immunogens can be designed with high precision to elicit epitope-specific antibody responses *in vivo*.
29. Azoitei ML, Correia BE, Ban Y-EA, Carrico C, Kalyuzhnyi O, Chen L, Schroeter A, Huang P-S, McLellan JS, Kwong PD, et al. Computation-Guided Backbone Grafting of a Discontinuous Motif onto a Protein Scaffold. *Science*. 2011; 334:373–376. [PubMed: 22021856]
30. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhnyi O, Vittal V, et al. Proof of principle for epitope-focused vaccine design. *Nature*. 2014; 507:201–206. [PubMed: 24499818] •• In this study, the authors demonstrated that a computationally designed epitope scaffold can elicit RSV-neutralizing antibodies in rhesus macaques, providing the first proof of principle for rational epitope vaccine design.
31. Zhang Y. Progress and challenges in protein structure prediction. *Curr Opin Struct Biol*. 2008; 18:342–348. [PubMed: 18436442] • This article offered a concise but comprehensive review of the progress and challenges in protein structure prediction.
32. Baker D, Sali A. Protein Structure Prediction and Structural Genomics. *Science*. 2001; 294:93–96. [PubMed: 11588250] • The two leading experts provided a detailed review of template-based and *ab initio* protein structure prediction methods as well as their implications for structural genomics.
33. Ginalski K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol*. 2006; 16:172–177. [PubMed: 16510277]
34. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000; 29:291–325. [PubMed: 10940251]
35. Petrey D, Honig B. Protein structure prediction: Inroads to biology. *Mol Cell*. 2005; 20:811–819. [PubMed: 16364908]
36. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys*. 2008; 37:289–316. [PubMed: 18573083]
37. Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Protein folding: The endgame. *Annu Rev Biochem*. 1997; 66:549–579. [PubMed: 9242917]
38. Das R, Baker D. Macromolecular modeling with Rosetta. *Annu Rev Biochem*. 2008; 77:363–382. [PubMed: 18410248]

39. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using rosetta. *Methods Enzymol.* 2004; 383:66–93. [PubMed: 15063647]
40. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010; 5:725–738. [PubMed: 20360767]
41. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008; 9:40. [PubMed: 18215316]
42. Chivian D, Kim DE, Malmstrom L, Bradley P, Robertson T, Murphy P, Strauss CEM, Bonneau R, Rohl CA, Baker D. Automated prediction of CASP-5 structures using the Robetta server. *Proteins.* 2003; 53:524–533. [PubMed: 14579342]
43. Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins.* 2007; 69:108–117. [PubMed: 17894355]
44. Kihara D, Chen H, Yang YD. Quality Assessment of Protein Structure Models. *Curr Protein Pept Sci.* 2009; 10:216–228. [PubMed: 19519452]
45. Kryzhtafovich A, Fidelis K. Protein structure prediction and model quality assessment. *Drug Discov Today.* 2009; 14:386–393. [PubMed: 19100336]
46. Kryzhtafovich A, Barbato A, Fidelis K, Monastyrskyy B, Schwede T, Tramontano A. Assessment of the assessment: Evaluation of the model quality estimates in CASP10. *Proteins.* 2014; 82:112–126. [PubMed: 23780644]
47. Lippow SM, Tidor B. Progress in computational protein design. *Curr Opin Biotechnol.* 2007; 18:305–311. [PubMed: 17644370]
48. Park S, Xi Y, Saven JG. Advances in computational protein design. *Curr Opin Struct Biol.* 2004; 14:487–494. [PubMed: 15313244]
49. Saven JG. Combinatorial protein design. *Curr Opin Struct Biol.* 2002; 12:453–458. [PubMed: 12163067]
50. Street AG, Mayo SL. Computational protein design. *Struct Fold Des.* 1999; 7:R105–R109.
51. Gordon DB, Marshall SA, Mayo SL. Energy functions for protein design. *Curr Opin Struct Biol.* 1999; 9:509–513. [PubMed: 10449371]
52. Vizcarra CL, Mayo SL. Electrostatics in computational protein design. *Curr Opin Chem Biol.* 2005; 9:622–626. [PubMed: 16257567]
53. Poole AM, Ranganathan R. Knowledge-based potentials in protein design. *Curr Opin Struct Biol.* 2006; 16:508–513. [PubMed: 16843652]
54. Desjarlais JR, Clarke ND. Computer search algorithms in protein modification and design. *Curr Opin Struct Biol.* 1998; 8:471–475. [PubMed: 9729739]
55. Dunbrack RL. Rotamer libraries in the 21(st) century. *Curr Opin Struct Biol.* 2002; 12:431–440. [PubMed: 12163064]
56. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins.* 2000; 40:389–408. [PubMed: 10861930]
57. Canutescu AA, Shelenkov AA, Dunbrack RL. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* 2003; 12:2001–2014. [PubMed: 12930999] • SCWRL is one of the most widely-used side chain modeling tools and can be applied to design and engineer antibody-antigen interactions.
58. Xiang ZX, Honig B. Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol.* 2001; 311:421–430. [PubMed: 11478870] • Based on a set of different principles, SCAP provides an alternative side chain modeling tool for the design and engineering of antibody-antigen interactions and antigen resurfacing.
59. Kuroda D, Shirai H, Jacobson MP, Nakamura H. Computer-aided antibody design. *Protein Eng Des Sel.* 2012; 25:507–521. [PubMed: 22661385]
60. Caravella JA, Wang D, Glaser SM, Lugovskoy A. Structure-Guided Design of Antibodies. *Curr Comput Aided Drug Des.* 2010; 6:128–138.
61. Clark LA, Boriack-Sjodin PA, Eldredge J, Fitch C, Friedman B, Hanf KJM, Jarpe M, Liparoto SF, Li Y, Lugovskoy A, et al. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* 2006; 15:949–960. [PubMed: 16597831]

62. Lippow SM, Wittrup KD, Tidor B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat Biotechnol.* 2007; 25:1171–1176. [PubMed: 17891135] • This study demonstrated the application of side chain modeling tools to antibody engineering.
63. Mandell DJ, Kortemme T. Backbone flexibility in computational protein design. *Curr Opin Biotechnol.* 2009; 20:420–428. [PubMed: 19709874]
64. Friedland GD, Kortemme T. Designing ensembles in conformational and sequence space to characterize and engineer proteins. *Curr Opin Struct Biol.* 2010; 20:377–384. [PubMed: 20303740]
65. Wu X, Yang Z-Y, Li Y, Hogerkorp C-M, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, et al. Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science.* 2010; 329:856–861. [PubMed: 20616233]
66. Woods RJ, Tessier MB. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr Opin Struct Biol.* 2010; 20:575–583. [PubMed: 20708922]
67. Frank M, Schloissnig S. Bioinformatics and molecular modeling in glycobiology. *Cell Mol Life Sci.* 2010; 67:2749–2772. [PubMed: 20364395]
68. Fadda E, Woods RJ. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today.* 2010; 15:596–609. [PubMed: 20594934]
69. Burton DR. Scaffolding to build a rational vaccine design strategy. *Proc Natl Acad Sci USA.* 2010; 107:17859–17860. [PubMed: 20937874]
70. Craik DJ, Cemazar M, Daly NL. The cyclotides and related macrocyclic peptides as scaffolds in drug design. *Curr Opin Drug Discovery Dev.* 2006; 9:251–260. •• The authors reviewed the use of cystine knots and related protein folds as scaffolds to design peptide drugs. By using diverse protein scaffolds to display a neutralizing epitope, scaffolding or epitope grafting, can be considered as an extension of this approach.
71. Craik DJ, Simonsen S, Daly NL. The cyclotides: Novel macrocyclic peptides as scaffolds in drug design. *Curr Opin Drug Discovery Dev.* 2002; 5:251–260.
72. Stumpp MT, Binz HK, Amstutz P. DARPin: A new generation of protein therapeutics. *Drug Discov Today.* 2008; 13:695–701. [PubMed: 18621567]
73. Jefferis R. The antibody paradigm: present and future development as a scaffold for biopharmaceutical drugs. *Biotechnol Genet Eng Rev.* 2010; 26:1–42. [PubMed: 21415874]
74. Ortiz AR, Strauss CEM, Olmea O. MAMMOTH (Matching molecular models obtained from theory): An automated method for model comparison. *Protein Sci.* 2002; 11:2606–2621. [PubMed: 12381844]
75. McLellan JS, Correia BE, Chen M, Yang Y, Graham BS, Schief WR, Kwong PD. Design and Characterization of Epitope-Scaffold Immunogens That Present the Motavizumab Epitope from Respiratory Syncytial Virus. *J Mol Biol.* 2011; 409:853–866. [PubMed: 21549714]
76. Zhou T, Zhu J, Yang Y, Gorman J, Ofek G, Srivatsan S, Druz A, Lees CR, Lu G, Soto C, et al. Transplanting supersites of HIV-1 vulnerability. *PLoS ONE.* 2014; 9:e99881. [PubMed: 24992528] •• In this study, the authors combined an efficient scaffolding algorithm and a high-throughput antigen screening procedure to design ~150 epitope scaffolds for three HIV-1 bnAb epitopes.
77. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005; 33:2302–2309. [PubMed: 15849316]
78. Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol.* 2009; 19:341–348. [PubMed: 19481444]
79. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics.* 2003; 19:1015–1018. [PubMed: 12761065] • Meta-server was proposed as a solution to tackle difficult protein structure prediction problems by combining output from multiple methods to obtain a consensus-based prediction.
80. Fischer D. 3D-SHOTGUN: A novel, cooperative, fold-recognition meta-predictor. *Proteins.* 2003; 51:434–441. [PubMed: 12696054]

81. Yang Y, Zhan J, Zhao H, Zhou Y. A new size-independent score for pairwise protein structure alignment and its application to structure classification and nucleic-acid binding prediction. *Proteins*. 2012; 80:2080–2088. [PubMed: 22522696]
82. Nguyen MN, Tan KP, Madhusudhan MS. CLICK-topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res*. 2011; 39:W24–W28. [PubMed: 21602266]
83. Zhu JH, Weng ZP. FAST: A novel protein structure alignment algorithm. *Proteins*. 2005; 58:618–627. [PubMed: 15609341]
84. Wang G, Dunbrack RL. PISCES: a protein sequence culling server. *Bioinformatics*. 2003; 19:1589–1591. [PubMed: 12912846]
85. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJP, Chothia C, Murzin AG. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res*. 2007; 36:419–425.
86. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton J, Orengo CA. The CATH classification revisited--architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res*. 2009; 37:D310–D314. [PubMed: 18996897]
87. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. The Pfam protein families database. *Nucleic Acids Res*. 2010; 38:D211–D222. [PubMed: 19920124]
88. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38:W545–W549. [PubMed: 20457744]
89. Peng J, Xu J. A multiple-template approach to protein threading. *Proteins*. 2011; 79:1930–1939. [PubMed: 21465564]
90. Kong L, Giang E, Robbins JB, Stanfield RL, Burton DR, Wilson IA, Law M. Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc Natl Acad Sci USA*. 2012; 109:9499–9504. [PubMed: 22623528]
91. Rodriguez-Limas WA, Sekar K, Tyo KEJ. Virus-like particles: the future of microbial factories and cell-free systems as platforms for vaccine development. *Curr Opin Biotechnol*. 2013; 24:1089–1093. [PubMed: 23481378] • An excellent review on the use of virus-like particles in vaccine development.
92. Grgacic EVL, Anderson DA. Virus-like particles: Passport to immune recognition. *Methods*. 2006; 40:60–65. [PubMed: 16997714] • An excellent review on the use of virus-like particles in vaccine development.
93. Ludwig C, Wagner R. Virus-like particles - universal molecular toolboxes. *Curr Opin Biotechnol*. 2007; 18:537–545. [PubMed: 18083549]
94. Jennings GT, Bachmann MF. Coming of age of virus-like particle vaccines. *Biol Chem*. 2008; 389:521–536. [PubMed: 18953718] • An excellent review on the use of virus-like particles in vaccine development.
95. Jardine J, Julien J-P, Menis S, Ota T, Kalyuzhniy O, McGuire A, Sok D, Huang P-S, MacPherson S, Jones M, et al. Rational HIV Immunogen Design to Target Specific Germline B Cell Receptors. *Science*. 2013; 340:711–716. [PubMed: 23539181]
96. Kanekiyo M, Wei C-J, Yassine HM, McTamney PM, Boyington JC, Whittle JRR, Rao SS, Kong W-P, Wang L, Nabel GJ. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature*. 2013; 498:102–106. [PubMed: 23698367]
97. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. 2014; 32:158–168. [PubMed: 24441474] •• The authors provided a comprehensive review on the technologies and applications in the emerging field of antibody repertoire analysis by next-generation sequencing.
98. Koff WC, Burton DR, Johnson PR, Walker BD, King CR, Nabel GJ, Ahmed R, Bhan MK, Plotkin SA. Accelerating Next-Generation Vaccine Development for Global Disease Prevention. *Science*. 2013; 340:1232910. [PubMed: 23723240] •• The authors reviewed the rational vaccine design approach and elaborated on how the recent technological advances can accelerate future vaccine development. The use of next-generation sequencing technology to decipher the human immune responses is highlighted.

99. Koff WC, Gust ID, Plotkin SA. Toward a human vaccines project. *Nat Immunol.* 2014; 15:589–592. [PubMed: 24940943] •• In this article, the authors proposed a big-science initiative aiming to delineate the complex human "immunome" (all genes and proteins associated with the immune system) in the context of vaccine development, focusing on B and T cell antigen receptors.
100. He LL, Sok D, Azadnia P, Hsueh J, Landais E, Simek M, Koff WC, Burton DR, Zhu J. Towards a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Sci Rep.* 2014; 4:6778. [PubMed: 25345460] •• The authors reported several critical advances in antibody repertoire sequencing, including the use of Ion Torrent PGM platform for long-read antibody sequencing, unbiased repertoire capture, single-molecule barcoding, and high-fidelity library preparation and sequencing techniques.

Highlights

- Computational tools can facilitate rational design of epitope vaccines
- Protein structure prediction tools can be adapted for structure-based immunogen design
- Antibody repertoire analysis can be used to evaluate vaccine-induced responses

Figure 1.

(A) A general strategy for epitope vaccine design consisting of epitope identification, immunogen design by epitope grafting, particulate presentation of designed immunogen, animal immunization, next-generation sequencing (NGS) analysis of antibody responses, and functional characterization of elicited broadly neutralizing antibodies (bnAbs). (B) Three categories of protein structure prediction tools: sequence analysis, structural modeling, and machine learning. Potential utility in epitope vaccine design is indicated by asterisk (*).

Figure 2.

(A) Application of side chain modeling tools to resurfacing of non-epitope regions of an antigen, optimization of an antibody paratope, and engineering of an antigen epitope. (B) A glycan modeling tool based on a clash-based scoring function, a stochastic search algorithm, and a glycan rotamer library. As an example, HIV-1 gp120 core (PDB ID: 3NGB) is glycosylated *in silico* using a high-mannose rotamer library.

Figure 3.
(A) Epitope scaffold design consisting of scaffold search, epitope grafting, and design optimization. (B) Three key criteria for scaffold selection including size and topology, flexibility, and the structural environment of the graft. (C) Application of a “scaffolding meta-server” to the HIV-1 2F5 epitope. The number of scaffolds identified by each algorithm and the overlap between any two algorithms are listed in an upper-diagonal matrix. (D) Concept of “scaffold family” using three HIV-1 10E8 epitope scaffolds as an example. Three proteins of the same fold family are matched to the 10E8 epitope with an

average C_α RMSD of 1.5Å. **(E)** Structural model of a ferritin nanoparticle presenting 24 copies of an HIV-1 PGT128 epitope scaffold.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

(A) A general strategy for antibody repertoire analysis consisting of sample collection, next-generation sequencing (NGS) of antibody library, and bioinformatics analysis, which can be divided into two consecutive stages – primary analysis and in-depth analysis. (B) Dissection of epitope-specific antibody responses through epitope manipulation. (C) Longitudinal tracing of immunogen-specific antibody lineages. Epitope manipulation and longitudinal

tracing can be used in combination to analyse the vaccine-induced antibody responses with high resolution.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript