JCTC Journal of Chemical Theory and Computation

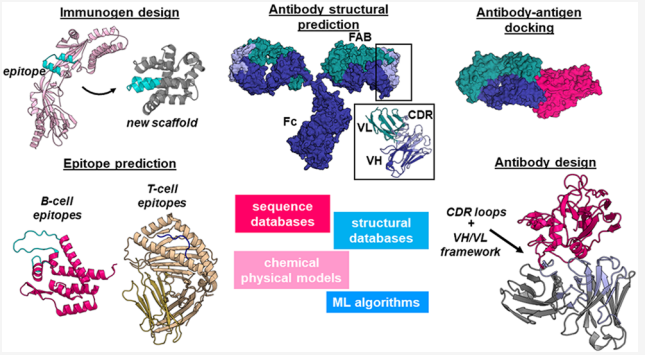
pubs.acs.org/JCTC
Review

Computational Methods in Immunology and Vaccinology: Design and Development of Antibodies and Immunogens

Federica Guarra and Giorgio Colombo*

Cite This: J. Chem. Theory Comput. 2023, 19, 5315−5333
Read Online
ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The design of new biomolecules able to harness immune mechanisms for the treatment of diseases is a prime challenge for computational and simulative approaches. For instance, in recent years, antibodies have emerged as an important class of therapeutics against a spectrum of pathologies. In cancer, immune-inspired approaches are witnessing a surge thanks to a better understanding of tumor-associated antigens and the mechanisms of their engagement or evasion from the human immune system. Here, we provide a summary of the main state-of-the-art computational approaches that are used to design antibodies and antigens, and in parallel, we review key methodologies for epitope identification for both B- and T-cell mediated responses. A special focus is devoted to the description of structure- and physics-based models, privileged over purely sequence-based approaches. We discuss the implications of novel methods in engineering biomolecules with tailored immunological properties for possible therapeutic uses. Finally, we highlight the extraordinary challenges and opportunities presented by the possible integration of structure- and physics-based methods with emerging Artificial Intelligence technologies for the prediction and design of novel antigens, epitopes, and antibodies.



1. INTRODUCTION

The rational development of new strategies to control cellular networks is a key challenge and a trove of opportunities for the scientific community.

The ways and mechanisms through which proteins interact are what shape such networks and determine their functions. The levels of involved proteins, their interaction strengths, as well as alterations in interaction partners build network connectivity at a proteome-wide scale ultimately defining cell phenotypes.[1,2]

Such phenomena are epitomized by mechanisms by which protein−protein recognition, interactions, and (re)-organizations shape the functional networks that control immune responses.[3−6]

In this general framework, molecular designers have the opportunity to address general and specific issues that will have important returns in the development of new diagnostics and therapeutics with potential in personalized medicine.

Vaccination is arguably the most successful medical discovery that has benefited human health globally. Worldwide vaccination campaigns have contained the recent Covid-19 pandemic and eradicated life-threatening diseases.[7−9] The development of new immunotherapies is now realizing its potential in combating drug resistance.[10] Moreover, exciting results are reported in the treatment of cancer, where immunooncology is becoming an increasingly important tool for patients. In this context, the discovery of the mechanisms of negative immune regulation and approval of immune checkpoint inhibitors had a primary role in revealing the potentialities of the immune system cells in counteracting tumor development,[11,12] and the prognostic value of the tumor-associated immune landscape is now becoming more and more apparent.[13] More recently, chimeric antigen receptor-based therapies (CAR-T) were approved for the treatment of some hematologic cancers (acute lymphoblastic leukemia, multiple myeloma, and different type of lymphomas)[14,15] while showing clinically significant antitumor activities in other malignancies.[16,17] Interestingly, in these therapies, the unpaired targeting specificity of monoclonal antibodies (mAbs) is combined with the cytotoxicity and long-term persistence provided of T-cells.

Following the genome era advent, conventional immunology has been overtaken by a more rapid and effective *in silico*-led approach to protein-antigen (Ag) selection, termed Reverse

Received: May 17, 2023
Published: August 1, 2023

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

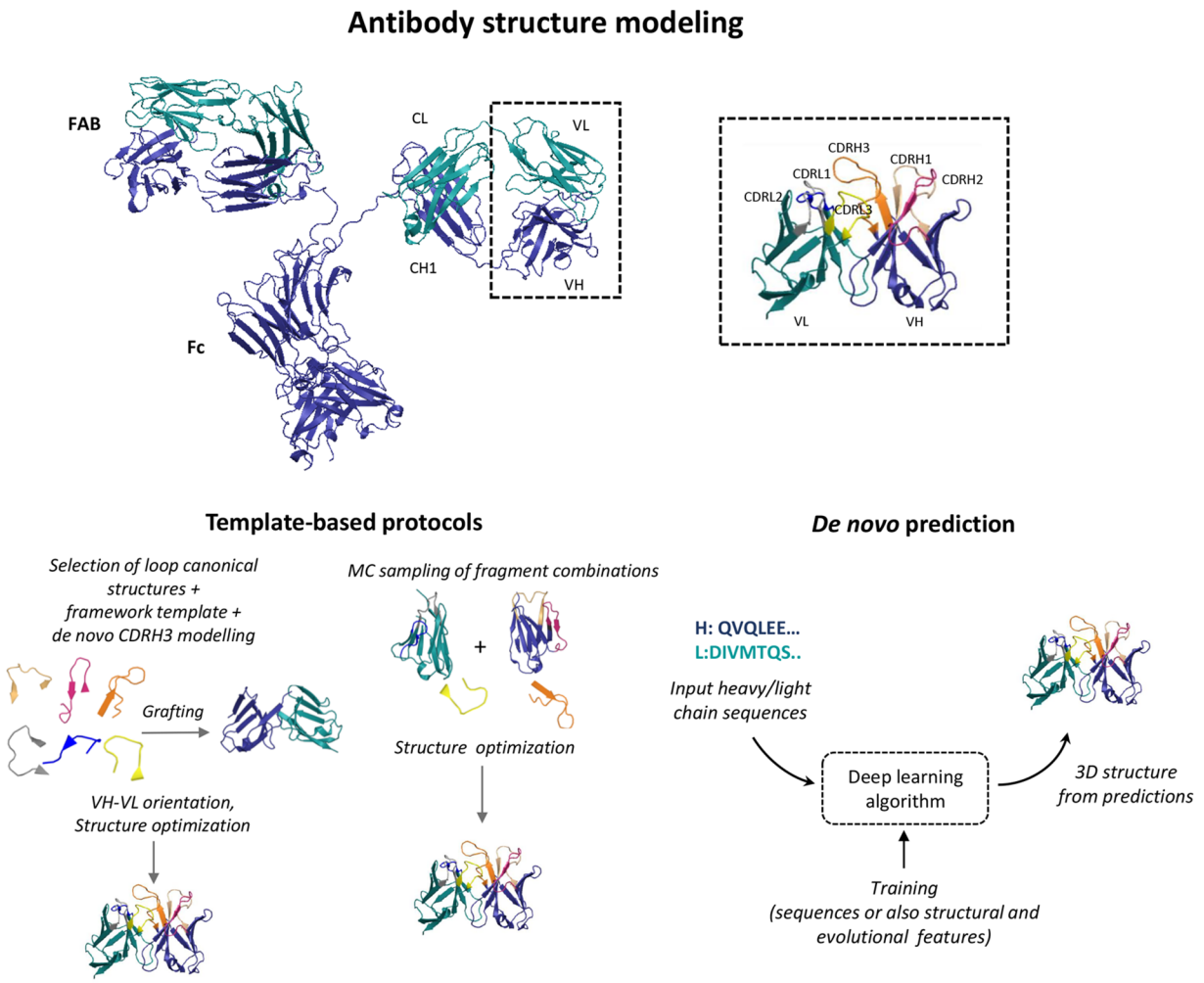


Figure 1. Antibody modeling and design. (Top) Full-sequence structure of an antibody and detail of VH/VL domains with CDR loop regions (PDB codes 1IGT[58] and 6OZB[59]). (Bottom) Overview of the antibody structure prediction methods. Template-based protocols rely on template construction from databases of fragments or framework regions onto which loop regions (usually obtained from canonical clusters) are grafted, followed by structural refinement steps. In some protocols the CDRH3 region is modeled *de novo*. In the figure, differences between two of the available protocols, Rosetta Antibody and AbPredict, are exemplified. *De novo* prediction methods can infer 3D structure from sequence and rely on deep learning algorithms trained on features encoding sequence and, in some cases, structural and evolutionary information.

Vaccinology (RV), and subsequently, pan-genomic RV,[18,19] which resulted in the first vaccine, Bexsero (approved by the European Medicines Agency in 2013), against meningococcal serotype B infections.[20] The most recent development of RV, termed Structural Vaccinology (SV), exploits atomic-level three-dimensional information to engineer Ags with improved immunological and/or biochemical properties.[21]

In a time where medical challenges are numerous, the value of being able to design protein antigens and antibodies with rational methods is clearly manifold. Besides providing new approaches to diagnostics and therapeutics development, this will generate new concepts and methods of general applicability in the realm of protein-interaction studies in health and disease, favoring the development of novel precision tools for chemical and cell biology.

In a different context, the possibility to design protein binders that may act as therapeutic tools, engaging specific molecular targets and hijacking pathologic protein pathway, has the potential to provide novel approaches to block the detrimental activities of proteins that still remain largely undruggable by classical interventions.[22,23] One particularly interesting field of application entails the manipulation of the activation of cell-surface receptors in response to extracellular signals. Proteins can, in fact, be designed to engage the extracellular domains of these receptors, in some cases driving their clustering on the cell surface, and ultimately generating different downstream cell cascades different from the ones observable in the absence of the binder protein.[24,25]

Our ability to design molecules that target proteins in specific networks thus has important implications for basic science and applied research in chemical biology, biochemistry, and molecular medicine. From the fundamental point of view, designed biomolecules can help advance our understanding of key interactions for diseases (which entail both host−pathogen interactions and derailed networks in nontransmissible pathologies) at the atomic level. Indeed, selective perturbation of well-defined complexes can aptly translate into an impact on the functional activities of the involved biomolecules, shedding light on their roles in biochemical pathways. From an applicative point of view, this knowledge can be translated into new rules for the efficient discovery of new candidates, such as antigens for vaccine development, proteins as

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

biological therapeutics, and new molecular probes for diagnostic development.

To make progress along this avenue, it is first necessary to understand what makes a protein substructure a potential (immune) interaction surface and then translate this information into the design of new systems that mimic desired interactions.

This is an exciting time for this field of research: fundamental contributions have been presented over the past few years in the structure-based design of new antigens, antibodies, and binders in general. The advent of Artificial Intelligence (AI) now enables protein structure prediction and analysis on an unprecedented scale and with accuracy. It is clear that the combination of AI with biophysics- and biochemistry-based methods will shape the next few years, speeding up the discovery of novel biologics with a wide range of applications.

Here, we will review some of the main advancements in these fields with a specific focus on methods involving the development and application of physical chemistry and computational chemistry approaches. A more general overview of methods relying on bioinformatic (and AI-powered) tools is provided, and, where needed, the reader is directed to more specialized reports. We will then provide some (personal) perspectives on the evolution and potential future impacts of this field in chemical, cell, and molecular biology.

2. ANTIBODY STRUCTURE PREDICTION

Antibodies are not only important investigative tools in (bio)chemistry, molecular biology, and medicine but are now largely employed as therapeutics in cancer, autoimmune and infectious diseases, and as diagnostic tools.[26−28]

Established antibody production processes rely on long and costly experimental procedures, often employing inoculated animals. This approach is sometimes limited by the capacity of animals to elicit antibodies targeting the desired epitope. In this framework, computational methods hold the potential of making antibody development more customizable, faster, and cheaper.[29] Breakthroughs in computational structure prediction methods, high-throughput sequencing, and data analysis have indeed greatly contributed to *in silico* antibody development.[30,31]

Computational methods for antibody development entail structure prediction, design, and antigen−antibody complex characterization with the end goal of designing an optimal Ab sequence complementary to a desired antigen.

Since immunoglobulins share a common Y-shaped fold, most traditional methods for modeling antibody structure rely on homology modeling strategies. Human antibody structure consists of two heavy and two light chains characterized by a well-conserved constant domain and the fragment antigen binding (FAB) domain, where conserved framework regions alternate with the highly variable loop regions named complementary determining regions (CDRs). The sequence/conformational variability in CDR regions, resulting from VDJ gene recombination and somatic hypermutation, is what makes affinity maturation and diverse antigen targeting possible. Yes, from the molecular designer's point of view, this also represents one of the major challenges in antibody structure prediction. Fortunately, making the issue more manageable, canonical structural clusters of loop conformations/orientations have been established[32] and usually, five of the six CDR regions fall into these representative structures. The same does not hold for the CDRH3 loop region, whose extended length and conformational variability make structural prediction more elusive. Small changes in heavy/light chain variable domain (VH-VL) orientations have also been shown to have a significant impact on CDR position and thus on antibody/antigen affinity. Therefore, efficacious modeling of the VH-VL interface is also a highly relevant matter. Figure 1 depicts an overview of the current strategies for antibody structural prediction, which will be described below.

Rosetta is one of the leading computational tools in biomolecular design and protein fold prediction. Initially developed in the group of David Baker at the University of Washington, it is now managed and developed by RosettaCommons, an international community counting more than 70 research groups.[33,34] Among the many computational frameworks included in Rosetta, there are several protocols specific for the computational modeling and design of antibodies.[35,36]

Rosetta Antibody[37] is a template-based protocol for structural prediction from sequence. First, selected templates for the framework and five canonical loops are grafted onto a preliminary model; at this point, the HCDR3 loop is modeled *de novo*, while the heavy/light chain variable region (VH-VL) orientations are refined. Template selection relies on a BLAST sequence search of the PyIgClassify database[38] and includes the sampling of diverse VH-VL orientations. The CDRH3 *de novo* modeling step employs the next-generation kinematic loop closure algorithm (KIC) in increasing resolution steps together with side-chain packing and minimization. With time, more accurate conformational constraints were introduced to overcome the limited capacity of the modeling protocol to sample the so-called kinked conformations of CDRH3 which were, on the contrary, observed more frequently in antibodies' native structures.[39,40] Applicability to nonhuman or nonmurine antibodies is limited by the availability of structures in the PDB database. For such modeling tasks, templates could be entered manually or selected from a custom-made database. Similarly, the modeling of rare CDR conformations will be more reliable when more of those structures will be experimentally solved. Sphinx, a hybrid method combining *ab initio* and knowledge-based loop structural prediction, has been shown to outperform Rosetta.[41]

A second protocol for antibody modeling implemented in Rosetta is AbPredict,[42] which does not rely on homologous templates. Experimentally determined antibody structures are fragmented into four backbone regions (light and heavy chain CDR3, heavy and light chain variable domains, which comprise the framework, and CDR1-2 region). These fragments are randomly recombined also considering different rigid-body VH-VL orientations to have a database of structures with a target sequence length. Starting from a randomly selected initial conformation, a Monte Carlo simulation is performed for sampling combinations of backbone fragments, packing of side chains, and minimization of the entire structure which outputs a single chain variable fragment (scFv) structure. Given the independence from sequence homology, it allows one to select templates with low sequence identity but high structural compatibility. On the other hand, the capacity of representing CDR loops of rare length is limited also due to the requirement that the target sequence and template match in length. The more recent implementation AbPredict2 is available as a Web server.[43] Improvements consist of reduced computational cost, a decrease in the stereochemical strain of

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

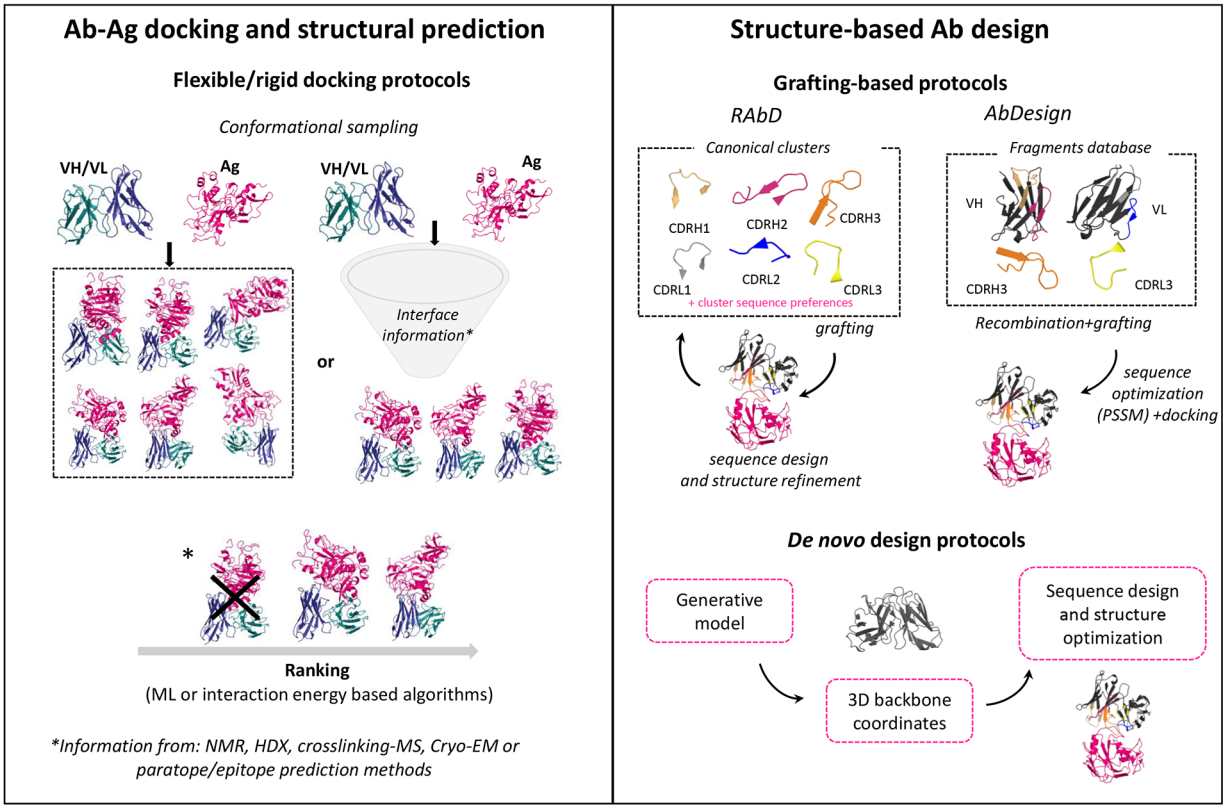


Figure 2. Antibody−antigen interactions and structure-based antibody design (left panel) overview of antibody antigen docking protocols. Rigid docking protocols only allow for rigid body movements of the binding partners whereas flexible docking protocols also include backbone or side chain refinement steps. Most of the docking programs envisage an integrative modeling approach meaning that information on the interface residues can be employed in the conformational sampling step or in the ranking/scoring step, or both. Structures are reproduced from PDB code 2Q8A.[26] (Right panel) Summary of Antibody Design strategies. Rosetta Antibody Design and AbDesign are taken as illustrative examples of two different grafting-based protocols. Rosetta Antibody Design includes a routine for assembling of a template from a selected scaffold+CDR loop regions from a canonical cluster database followed by sequence and structure refinement in the presence of the binder. AbDesign builds a template from sampling of combinations of VH and VL domains (including all non-CDR3 loops) plus CDR3 regions followed by sequence optimization and docking onto the binder. *De novo* design protocols rely on generative models for obtaining 3D backbone coordinates, followed by sequence design and structural refinements. Structures are reproduced from PDB code 6OC3.[59]

the model, and better sampling of light-heavy chain orientations.

Other open-source automated similar workflows are ABodyBuilder,[44] Kotai Antibody Builder,[45] and PIGS.[46] ABodyBuilder is a fully automated computational pipeline comprising steps similar to those already described for Rosetta Antibody with template selection relying on sequence similarity, using FREAD[47] for all CDR loops modeling and SCWRL4[48] for side chains.[44] It has the advantage of being a fast prediction tool, favoring high-throughput screenings, and implements a data-driven estimation of model accuracy as a measure of the probability of a region to be modeled within a defined threshold. Kotai Antibody Builder selects templates for framework and non-H3 CDR region, privileging statistically derived rules together along with sequence identity evaluation. For CDRH3 loop modeling, it relies on previously established rules for base region type prediction. Subsequent refinement can employ the SPANNER module for fragment-based structural prediction, followed by energy minimization steps and scoring.[45] PIGS privileges template construction from a unique structure and relies on canonical structures for loop modeling.[46] The main packages for full antibody modeling together with others that focus on more specific modeling tasks have been listed in a recent dedicated review.[30]

The AMA II benchmarking of many of these protocols revealed that their performance is overall similar, with an average RMSD of about 1.1 Å for the entire Fv and with the most critical region being the CDRH3 loop region, modeled with an average accuracy of about 2.2 Å but up to >5 Å RMSD in some cases.[49]

In addition to the above-described grafting-based tools, deep learning-based methods are emerging and are being heavily applied to fast and accurate antibody *de novo* structural prediction.[50,51] Methods optimized for immunoglobulin structural prediction perform better than AlphaFold2, originally developed for general protein structural prediction.[52,53] DeepAb, an improved version of DeepH3, was tested on the Rosetta Antibody benchmark and shown to outperform grafting-based methods.[54] The method is based on a deep learning model trained using cross-entropy loss on pairs of heavy/light chain sequences derived from the Observed Antibody Space[55] and can predict relative distances and orientations between pairs of residues of Fv domain from sequence information and previous information on Ab structural organization. The final step is a fast Rosetta-based protocol for the generation of the 3D protein structure. Insights about relevant amino-acid interactions and mutations improving binding affinity could also be derived. ABlooper,[53]

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

based on an equivariant graph neural network is currently the fastest method for predicting CDR conformations and was shown to achieve on average lower RMSD than ABodyBuilder but not DeepAb on models from the Rosetta Antibody Benchmark. The same group recently released IgFold which also embeds information from templates and has lower runtime for structural predictions.[56] These methods circumvent the lack of experimentally solved structures on which grafting methods rely and, on the other hand, take profit from the large amount of sequence data currently available. Their short runtime (on the order of minutes for structural prediction) opens the possibility of high-throughput structural screenings. An extended description of AI-based methods is beyond the scope of this review. For further details, the reader is directed to other reports.[50,57]

3. STRUCTURE PREDICTION OF ANTIBODY−ANTIGEN COMPLEXES

While great improvements to apo-antibody structural prediction have been brought about by DL-based methods, reliable structural prediction of the antibody−antigen complex is still highly desirable for the future. Indeed, atomic-level structural information on antibody−antigen complexes is instrumental to investigate the interactions involved in forming the complex and providing a rational framework for the design of optimized interactors. Given the difficulty of directly predicting the multimer structures from sequences[60] and of experimentally solving high-resolution structures, developing more and more accurate protein−protein docking methods is of paramount value. While general-purpose methods could be used, accurate prediction of the antibody−antigen interface is challenging given the conformational variability of the Ab loop regions involved. In an attempt to overcome such inherent challenges, tailored docking protocols were developed. They can be classified into rigid docking methods such as those implemented in ClusPro[61] and ZDock,[62] only allowing rigid body movements of the binding partners, and flexible/semiflexible methods that also include backbone/side chain refinement steps such as SnugDock,[37,63] HADDOCK,[64] SwarmDock,[65] and LightDock.[66] These tools implement an integrative modeling approach allowing the use of experimentally derived or sequence-conservation knowledge about the interface (Figure 2, left panel).[67] Benchmarks of docking protocols were recently reported.[68,69] The methods differ in their conformational space sampling strategies and in the way they employ previous knowledge: for instance, HADDOCK and LightDock employ information on the interface both for guiding the docking in the sampling step and for ranking and scoring the obtained models, whereas other tools like ZDock and ClusPro employ previous knowledge only in the scoring and filtering of the generated models. These aspects result in a high number of high-quality models when some knowledge is provided in the case of HADDOCK but a lack of good models when no good quality information is available. Other methods perform a more extensive sampling with a low percentage of near-native models. Thus, an accurate scoring function is needed in these cases. Improvements in the accuracy of the models come therefore from a more extensive knowledge of the interface or more accurate scoring functions/methods able to sort out near-native models from the generated poses. Experimental information about important residues or interface conformations can be derived, for example, from mutagenesis, cross-linking mass spectrometry, NMR or hydrogen−deuterium exchange experiments, Cryo-EM, and SAXS.[64,70−72] In this context, knowledge derived from the computational prediction of paratopes and epitopes leveraging the great availability of sequence data is a very attractive resource in terms of its high-throughput potentialities and broad applicability. Epitope prediction remains the more challenging task and will be treated in detail in the next paragraphs (*vide infra*). Antibody-specific ML paratope predictors include PINet,[73] PECAN,[74] Parapred,[75] proABC2,[76] and others.

Another challenging aspect is that once docking poses have been generated, a good scoring function is needed for discriminating non-native from near-native complexes. In this regard, scoring functions can be derived more traditionally from interaction energy evaluation[77] or also from statistical potentials obtained from known complexes. In the latter case, ML-based methods have been recently employed for improving pose ranking.[78,79]

4. STRUCTURE-BASED ANTIBODY DESIGN

A computational problem complementary to the prediction challenges described above is antibody design: whereas structure prediction demands modeling of the protein fold starting from knowledge of the sequence, antibody design aims at defining the best sequence for a target 3D structure. The topic has been recently reviewed in depth elsewhere.[30,31] In general, methods can be classified into *de novo* or grafting-based design protocols, and different criteria in the design process need to be adopted depending on whether the goal is to obtain a single antigen specific antibody or a protein capable of binding a broader spectrum of antigens (Figure 2, right panel).

Concerning the single-state design problem, Rosetta also includes antibody-tailored design protocols. RosettaAntibody-Design (RAbD)[80] is the design equivalent of RosettaAntibody. Starting from an antibody−antigen complex, first it is possible to design different CDRs from the canonical cluster database based on North−Dunbrack clustering[32] and then the sequence is designed/optimized based on canonical CDR sequence preferences. Briefly, RAbD consists of a graft design routine which then passes the generated model to a protocol performing sequence design, side chain (re)packing, CDR minimization, and, if needed, docking with epitope- or paratope-based constraints. Minimization employs CDR cluster-derived constraints, and a Metropolis Monte Carlo criterion is used for optimization in both the graft design and subsequent steps. In this context, the total energy or interface energy can be considered as the guiding criterion. AbDesign[3,81] applies to antibody design an approach similar to AbPredict: experimentally determined structures are fragmented analogously and then recombined and grafted onto the target structure. After constrained optimization of segments, dihedral and coordinate templates where each segment deviates <1 Å from the original structure are kept for subsequent sequence optimization according to conformation-dependent position-specific scoring matrices (PSSMs) for each segment. At this point, the predicted apo structure is docked against the target antigen to allow further backbone conformation/sequence optimization. Designs are scored based on different criteria concerning protein structure quality and shape complementarity/binding affinity with the antigen. Benchmarking of the method highlighted issues related to the expression of the designed antibodies and the deviation of the experimentally

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

determined structure from the predicted one, indicating that rounds of optimization were required.

Protocols for multistate-design tasks aiming at the development of antibodies that can bind to multiple antigens, optimizing specificity through negative selection, or binding to different antigen conformations are also implemented. Broadly neutralizing antibodies have the potential of retaining neutralizing efficacy despite the escape mechanisms engaged by pathogens by targeting conserved residues.[82] A suitable protocol is RECON MSD, which has the goal of optimizing the sequence so that it can adopt various conformations for binding more antigens or more conformational states of a single antigen.[83] BROAD is an implementation that reduces computational costs.[84]

Vendruscolo and collaborators recently reported a novel strategy for designing antibodies that target epitopes whose structure is known (computationally or experimentally).[85] Starting from the target epitope structure, the so-called AbAg database comprising linear CDR-like motifs and antigen-like sequences, derived from PDB structures, is searched for complementary CDR-like fragments. After CDR-like fragments are combined, contacts with the epitope are optimized, and CDRs are grafted onto an antibody scaffold. This approach circumvents the need for approximate energy calculations and sampling of the conformational and mutational space. The applicability is, of course, dependent on the availability of paratope-like fragments compatible with the target epitope.

A different approach to the structure-based antibody design computational problem was adopted by Huang and colleagues, resorting to the use of emerging AI methods.[86] Ig-VAE is a Variational Autoencoder-based generative model that directly outputs 3D backbone coordinates that can then be combined with various constraint and optimization approaches for applications in antibody design. For example, this method was applied to the design of a SARS-Cov-2 ACE2 epitope binder. The latent space of this model can be exploited to expand the capacity of generating novel proteins that cannot result from the sampling of existing structures. The protein design problem is thus reformulated as a constrained optimization in the latent space of a generative model.

Another recent report moved in the direction of a generative model capable of codesigning the antibody CDR 3D structure and sequence, given a framework region restraint, through iterative refinement steps that optimize one in response to the other. This represents an advance since the optimal 3D antibody structure for a target antigen is seldom known *a priori*.[87] A further important improvement to this model would come from applying the antibody design problem constraints and conditions derived from the stereoelectronic properties of the specific target antigen.

Similarly, other generative models such as RF Diffusion[88] and Chroma,[89] which have previously been applied to the design of protein binders hold the promise to bring advances to the Ab field.

The paradigm is thus shifting from an unconstrained structural generation complemented with a binding affinity predictor and conformational/sequence search to a constrained *de novo* structural generation of the antibody in the presence of the target epitope/antigen.

Currently, experimentally or computationally designed antibodies often need to undergo other rounds of optimization. In this regard, *in silico* methods that mimic natural affinity maturation have been developed. Both physics-based and ML-based methods aim at predicting the mutations that will enhance antibody stability and affinity for the antigen.[90−92] Automated tools aiming at increasing antibody developability (i.e immunogenicity, solubility) can also be employed in the design process.[51]

In addition to classical antibodies, single-domain antibodies are also being developed since they present some advantages with respect to immunoglobulins. For instance, they may be easier to produce and manage and, given their reduced dimensions, they may encounter fewer adverse reactions when administered as biological drugs.[93−95]

5. ANTIGEN AND EPITOPE DISCOVERY

5.1. Immunogen Discovery and Reverse Vaccinology.

While antibody therapeutics are responsible for passive immunization, a vaccine is defined as a preparation capable of stimulating an active immune response in the host. Active immunity includes both the humoral response (antibody-mediated) and the cellular-mediated immune response.[96] The ideal vaccine should cause a prolonged neutralizing immune response in the host and is composed of an antigen complemented with adjuvants.

Traditional vaccine discovery approaches relied on the direct investigation of pathogen components hypothesized to have a role in eliciting responses to identify protective antigens. This approach has led to the development of many lifesaving vaccines but has shown limitations in targeting infectious diseases or other pathologies characterized by immune evasion mechanisms. Examples are HIV, influenza, or malaria pathogens whose surface proteins elicit extreme sequence variation or are shielded by glycans (HIV, hepatitis C). The search for anticancer vaccines, a hot field of study at the moment, may also need different approaches.[97,98]

In 2000[99] Rappuoli and colleagues spurred a dramatic change in the field by introducing the paradigm of Reverse Vaccinology, which has led to the development of a vaccine against meningococcus B. Taking advantage of whole genome sequencing, possible antigenic species are sorted out with the aid of immuno-bioinformatics tools, subject to different criteria such as surface exposition, conservation, toxicity, and allergenicity. The most promising species are then screened experimentally for their immunogenic activity. More recently, the term Reverse Vaccinology 2.0 has been coined to refer to the advancements brought to the field by the analysis of human protective antibodies, sequencing and cloning of B-cell repertoires together with structural characterization of antigens and epitopes.[97,100] The investigation of effective antibody response allows the identification of promising corresponding antigens/epitopes.

Various open-source tools for complete RV workflows have been designed mainly for bacterial antigen discovery. They can be classified in filtering based methods such as Vaxign,[101] NERVE,[102] Jenner-predict,[103] and VacSol,[104] or machine learning-based methods including VaxiJen,[105] Vacceed,[106] Vaxign-ML,[107] and the Bowman-Heinson[108,109] method. Filtering-based methods give as a result putative vaccine candidates selecting them only based on predetermined threshold values on features like the probability of exposure to the immune system, to be an adhesion molecule or more generally involved in virulence, limited homology to host proteins and number of transmembrane domains predicted using external computational tools or databases (pSort,

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

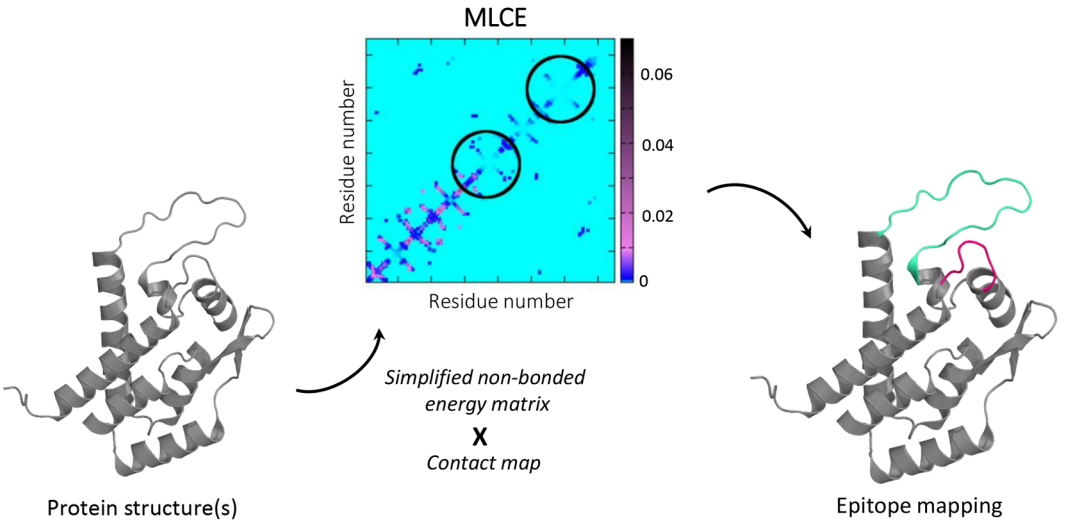


Figure 3. Epitope prediction. Schematic representation of the energy-decomposition-based epitope prediction method. The protein shown is reproduced from PDB code 1GWP.[26]

CELLO2GO, SPAAN, HMMTOP, BLASTp, OrthoMCL, DEG, Vaxitope, Pfam, VFDB).[110]

Machine learning-based methods are trained on a set of bacterial protective antigens and nonprotective proteins and are employed to perform a ranking of the probability of all searched proteomes being a protective antigen based on the predicted properties. VaxiJen classifies proteins based on physicochemical properties only and could be suitable for predicting also viral or cancer antigens. The Bowman-Heinson method employs a support vector machine classifier and was trained on a data set of bacterial protective antigens (BPAs) and non-bacterial protective antigens annotated with 10 features related to antigen properties.

A comparison of six of these RV tools (Vaxign, NERVE, Vac-sol, Jenner-predict, VaxiJen, and Bowman-Heinson) in the recognition of BPAs in 11 pathogens has been reported recently.[111] Results highlighted that none of the methods recognized all the known BPAs, with the ML-based Heinson-Bowman method being the best performing according to the fraction of proteome identified as a good candidate and agreement with the testing set. Moreover, there is poor accordance between the outputs of the different tools, suggesting the need to interrogate more than one tool in antigen discovery research. Overall, the urge for improvements in protein annotation tools and incorporation in databases of proteins with negative experimental results to be employed as negative test sets for ML-based methods is highlighted.

Since this benchmark was reported, updated versions of the tested RV tools such as VaxyJen3[112] and Vaxign2[113] were implemented to address possible limitations. For instance, the Vaxign2 Web server includes a predictive framework and postprediction computational tools and can be also employed for the development of candidate viral vaccines. In the predictive part it incorporates both the filtering-based Vaxign and the ML-based Vaxign-ML[107] tools. Postprediction analysis entails the prediction of MHC-I and II epitopes with MEME, epitope mapping onto the input proteins based on IEDB[114] immune epitope database including B- and T-cell experimentally tested epitopes, calculation of population coverage, and prediction of protein function and orthologs.

5.2. Prediction of B-Cell Epitopes. Computational epitope prediction is of key importance in vaccine development. Indeed, humoral antibody-mediated immune responses rely on interactions between B-cell receptors and antigen epitopes, whereas cellular responses include, among others, T-cell mediated responses that rely on interaction between T-cell receptors (TCRs) and MHC-I and MHC-II epitopes present on antigen presenting cells. Different tools were developed for T- or B-cell epitope predictions.

B-cell epitope predictions can be classified as sequence-based or structure-based methods. The latter have been demonstrated to be more reliable,[115] but their application is limited by the availability of 3D antibody−antigen structures. Indeed, B-cell epitopes are more frequently discontinuous (also named conformational epitopes), and their prediction poses great challenges to current predictive approaches, especially sequence-based ones that are more suitable for linear epitopes. Moreover, doubts have emerged on whether the question of searching for all the possible epitopes of an antigen for all possible antibodies is well posed.[116] Some evidence suggests that in principle all exposed protein surface patches could represent an epitope, given the "right" interacting partner.[117,118] Therefore, antibody-specific prediction methods were proposed.[119,120] Clearly, these implicate that the target antibody is known, which is not always the case.

Recently, two sequence-based (Bepi-pred 2.0,[121] CBTope[122]), some structure-based methods (SEPPA3,[123] Disco Tope 2.0,[124] ElliPro,[125] EPSVR,[126] BEpro,[127] epitope3D[128]), and an antibody-specific B-epitope prediction method (EpiPred[120]) have been benchmarked.[116] Results evidenced that while performances support the reliability and usability of the methods, they may be suboptimal most likely because they were trained on limited (and now potentially outdated) data sets.

To overcome current limitations, the authors suggest considering aspects that range from taking into account oligomerization properties, conformational changes, the possibility of incorporating residues that should be excluded from prediction, the prediction of glycosylated surfaces, the potential for other post-translational modifications, and introducing antibody sequences to guide epitope search.

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

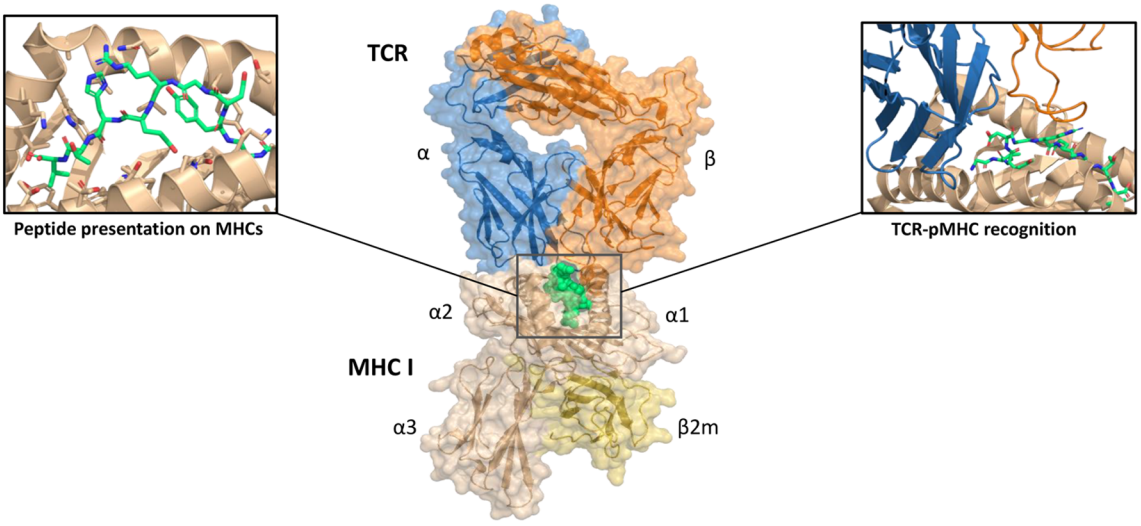


Figure 4. T-cell epitopes. Interaction between a peptide-presenting class I MHC molecule and a T-cell receptor, reproduced from PDB code 6TRO.[183] After antigen processing, peptide immunogenicity is determined by its effective presentation on MHC molecules and binding of the pMHC complex by TCRs, which triggers downstream activation of CD4 or CD8 T-cells. Sequence- and structure-based computational tools for peptide binding affinity prediction and TCR-pMHC recognition are described in the text.

In this framework, it has been shown that the introduction of protein language models for the representation of protein sequences can improve the performance of sequence-based methods in the prediction of both linear and conformational epitopes.[129,130] The importance of training the models on updated larger data sets characterized by low redundancy was also pointed out. A further advance would be the incorporation of the target antibody in the prediction using the same encoding strategy of the antigen.

Overall, the role of structure-based computational approaches in the prediction of B-cell epitopes has significantly increased. Also, to improve the spectrum of properties of a newly discovered antigen, the fine physicochemical determinants of its immune reactivity need to be characterized. This is important for both applicative and fundamental research. From a fundamental point of view, this knowledge can unveil the basic mechanisms underlying molecular recognition mechanisms in immune responses. On the other hand, this would improve our capacity to design antigens or antigen mimics with optimized reactivities and pharmacological properties. Nowadays, structure-based methods can take advantage of advances in the accuracy of predicted protein structures and in the increasing number of experimentally solved structures of antibody−antigen complexes.

In this context, we have developed a structure-based method (BEPPE) for the prediction of interface regions in proteins based only on energetic determinants and available as a Web server.[131−134] The method was proven successful for epitope prediction and reverse vaccinology applications.[135−138]

This prediction method is based on an energy decomposition approach: given a protein 3D structure (or also several structures extracted from MD simulations), the matrix (or average matrix) of pairwise nonbonded interactions is calculated. After eigenvalue decomposition, the most stabilizing eigenvector (or few eigenvectors)[139] is employed to reconstruct a simplified matrix which is then filtered for retaining only local interactions and to obtain the so-called MLCE (matrix of local coupling energies). From this, patches that have the lowest energetic couplings with the surrounding structures are identified as putative interface regions with the potential to be recognized by Abs (see Figure 3). This method is based on the idea that epitopes (and interface regions in general) should be more prone to undergo conformational changes to interact with binding partners while contributing less to the overall protein stability and being mutation tolerant to facilitate escape from the host immune system.[140]

Another method relying on energy considerations was reported by Fiorucci and Zacharias.[141] While other methods estimated desolvation penalties of protein surfaces by means of surface-area calculations, here the desolvation free energy penalty is obtained by probing the protein surface with a neutral low-dielectric sphere, therefore also considering the neighboring environment and long-range electrostatic effects. It was found that interface regions largely coincide with those characterized by lower free energy losses.[141]

Among ML-based partner specific methods, PINet, already mentioned as a paratope prediction tool, can predict antibody specific interface residues (both the paratope and epitope side) through a geometric deep learning approach aimed at dissecting the properties of complementary surfaces. It has been reported to have state-of-the-art performances in both antibody−antigen and protein−protein interaction cases.[73]

Recently, AxIEM has considered the issue of predicting "cryptic epitopes" of class I viral fusion proteins exposed in metastable conformations important for virulence and that are often the target of broadly neutralizing antibodies.[142] The method leverages residue specific and environment features to determine the probability of a certain residue belonging to an epitope based on the conformation of the protein.

Interestingly, the SARS-CoV-2−human interactome was predicted with the computational pipeline PEPPI.[143] It ranks interaction probability based on a naive Bayesian consensus classifier that combines outputs from independent modules that evaluate sequence and structural similarity to known protein−protein interactions, functional association, and a neural network classification. While results should be considered also based on other considerations not taken into account, such as subcellular colocalization of the predicted

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

interactors, this and other similar tools can be useful for predicting relevant immunogens on a proteome-wide scale.

5.3. Prediction of T-Cell Epitopes and Neoantigens. T-cell epitope immunogenicity encompasses three fundamental steps: antigen processing resulting in the immunogenic peptide, binding of the peptide to MHC class I or II molecules on Antigen Presenting Cells (APCs), and binding of peptide-presenting MHC molecules (p-MHC) to TCR.

MHC class I molecules are expressed on the surface of all somatic cells and present peptides from the intracellular environment, whereas MHC class II molecules are found on immune cells such as B-cells, dendritic cells (DCs), and sample peptides from the extracellular milieu. CD8 T cells only interact with class I p-MHC and once primed and activated, they generate cytotoxic T cells. On the other hand, class II p-MHC complexes are recognized by CD4 T-cells that can give rise to different types of T helper cells. Among T-cell epitopes, neoepitopes are defined as those distinctive of cancer cells resulting from tumor-related genetic aberrations. The identification and targeting of neoepitopes is one of the most actively evolving fields of cancer immunotherapy aiming at the development of neoantigen based vaccines and T-cell based therapies.[144−148] Current strategies for the identification of neoantigens rely on advanced sequencing techniques, including next-generation sequencing or -omics data aimed at individuating peptides distinguished from self-peptides. These are coupled with computational tools including those that we describe below for general purpose prediction of T-cell epitopes immunogenicity.[98,149−151]

Importantly, along with the general IEDB epitope database, T-cell epitope databases such as SYFPEITHY[152] and ATLAS[153] as well as the CEDAR cancer epitope database have been curated.[154]

In this section, we will give an overview of computational tools predicting peptide binding to MHC molecules and effective interaction between p-MHC and TCR. Figure 4 depicts the interaction between a peptide presenting the MHC molecule and a T-cell receptor.

Binding to MHC I and II is similar but has some relevant distinctive features: MHC I groove is more closed, it typically hosts peptides ranging from 8 to 14 residues (more often 9), and binding site pockets have selective physical-chemical features. On the other hand, predicting binding to MHC II molecules is more challenging, as peptides are typically longer, even though only a 9-residue core typically sits in the more open binding groove featured with less discriminating interaction properties. A great number of sequence-based predictors have been developed that can either distinguish a binder from a nonbinder or also give an estimation of the binding affinity. These are listed in recent reviews[155,156] and mostly rely on existing experimental elution or binding affinity data. While the first developed methods only took into account the presence of the so-called anchoring residues with suitable spacing, current methods consider fixed-length peptides (mainly 9 residues long) and quantify the contribution of each peptide position to the binding considering each residue independently in the case of linear methods (such as BIMAS),[157] or also take into account nonlinear effects (for instance methods based on artificial neural networks (ANNs)[158] or support vector machines (SVMs).[159] Interestingly, it was demonstrated that algorithms taking into account nonlinear effects do not outperform linear ones in the prediction of binding to MHC class I molecules.[160] The hypothesized explanation is that since peptides bind in the groove in an extended conformation, the contribution of each residue is dependent on its own interactions in the binding pocket with other nonlinear effects related to residue−residue interactions having a lower magnitude. Also, no significant gain of performance was obtained with prediction methods taking into account the 3D structure of the complexes with respect to sequence-based methods.[161] However, MHC molecules in humans (called human leukocyte antigens (HLAs)) are characterized by great allelic variability. Indeed, different HLA molecules from different individuals will have different peptide preferences, and HLA supertypes with similar binding preferences were individuated and clustered. Since experimental data on which the above-mentioned models were trained are available for only a few allelic variants, they have a limited capability of predicting binding for all allelic variants. To overcome this issue, pan-MHC methods able to predict the binding of a peptide to uncharacterized HLA molecules have been developed. For example, ML models capable of performing such predictions for MHC I molecules such as NetMHCpan[162] and a structure-based threading approach[163] were developed by training the algorithm on information related to the MHC peptide binding site.[164] Similar predictors for MHC class II molecules have been developed and improved versions of the initial models have been implemented more recently.[155,156,165] Computational tools for predicting promiscuous sequences with high population coverage are also available.[166,167] Interestingly, a method for predicting allele-specific anchor residue positions has been proposed for complementing neoantigen prioritization pipelines.[168]

Predictive tools have mostly taken into peptide presentation on MHC molecules and indeed, binding affinity to MHCs correlated well with immunogenicity for viral epitopes.[163] However, the fact that good presentation on HLAs is necessary but not sufficient for stimulating T-cell response is particularly true for responses to neoantigens: owing to their endogenous nature, corresponding T-cells could have been deleted or tolerized to avoid self-immune reactions. Therefore, either mutations with respect to self-peptides should increase MHC binding (which is not so common) or (more likely) there should be structural differences in T-cell recognition with respect to wild type p-MHCs. In general, in the structural prediction of TCR-pMHC binding challenges are posed by the degree of TCR diversity and conformational flexibility of the TCR-pMHC interface. Therefore, at first, efforts were devoted to predicting TCR-pMHC from the protein sequences. Among sequence-based methods of notice is NetTepi[169] which integrates previously developed methods for the evaluation of peptide-MHC stability (NetMHCstab) and binding (NetMHCcons), and a model for TCR affinity prediction reported by Calis and colleagues.[170] The latter scores the peptide sequence based on rules derived from experimental data evidencing, for example, the relevance of positions 4−6. Other computational tools relying only on sequence have been described in more detail in a recent review.[171] Sequence data driven methods have outlined some amino acidic physico-chemical features related to immunogenicity such as the enrichment of aromatic hydrophobic and aromatic residues in the binding interface. However, physics-based and structure-based models can provide a rational background to these findings: for instance, the energetic gain of burying a hydrophobic residue by complex formation strongly depends

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

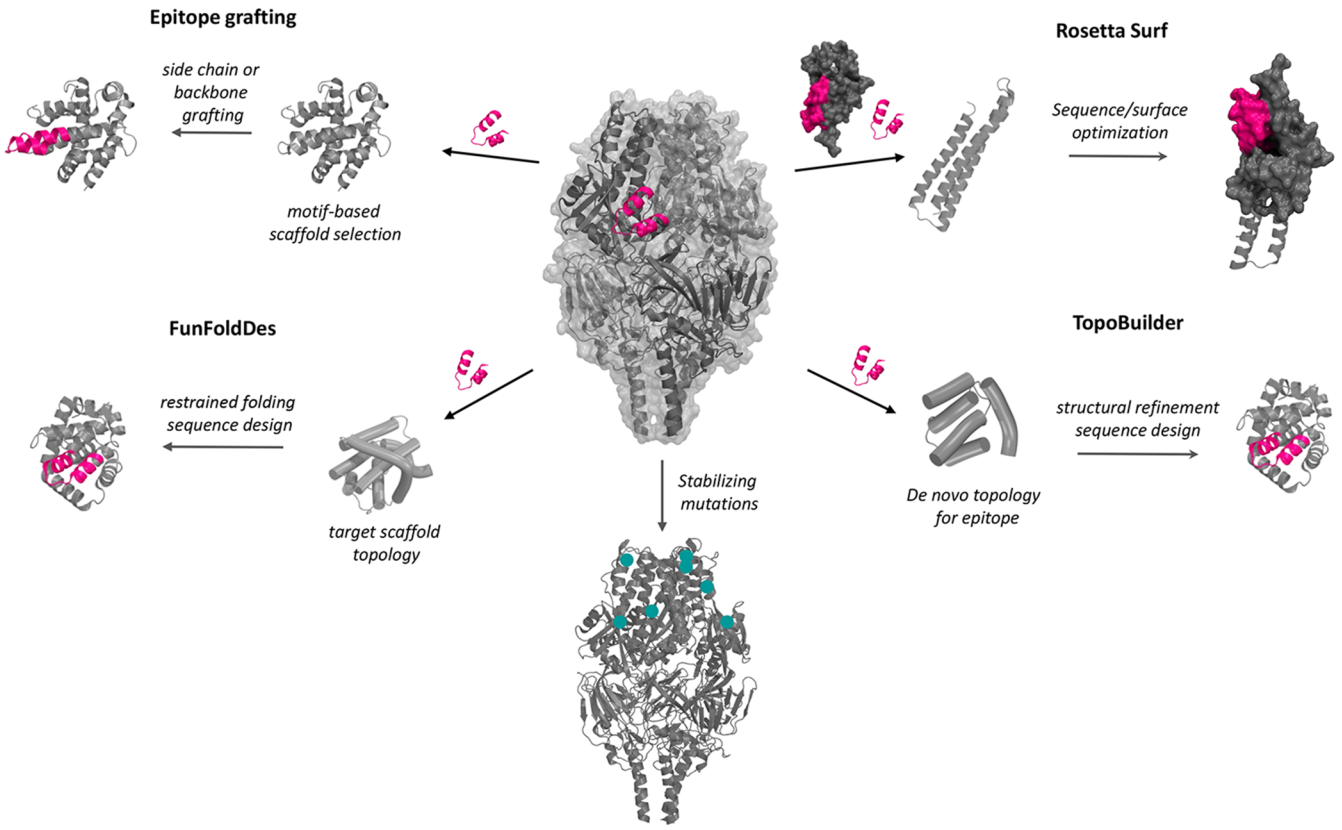


Figure 5. Overview of the main immunogen design approaches. Epitope grafting. Starting from a known epitope on the antigen, this can be transplanted through side-chain or backbone grafting onto a suitable scaffold chosen for structural compatibility with the motif to be transplanted. FunFolDes allows transplantation onto a defined topology instead of a specific scaffold protein; epitope grafting is followed by refolding of the rest of the protein into the target topology and sequence design. Rosetta Surf can follow the grafting of an epitope onto a scaffold and allows for surface optimization to mimic the native epitope environment and optimize complementarity with a target binder. TopoBuilder is a protocol for building *de novo* a topology around the epitope to be transplanted and has been complemented with FunFoldDes for the actual resulting protein design. Interestingly, these protocols allow biasing of the design process with the target binder. Another possible approach is to stabilize a particular conformation of the antigen exposing the desired immunogenic epitopes; an example are prefusion-stabilized variants of fusion proteins from viral pathogens.

on the 3D structural features. Given these premises, considering structural aspects alongside sequence holds the promise of giving better prediction of peptide immunogenicity and TCRs binding specificities.[172−174]

Baker and co-workers developed a neural network-based method trained on structural and energetic features derived from HLA-A2/peptide complexes.[175] It outlines biophysical determinants of neoantigen immunogenicity such as mutations increasing the exposition of hydrophobic residues and overall peptide energy. Moreover, by including well-presented but not immunogenic as well as non-MHC binding peptides in the training data set it is able to capture both binding to HLA and TCR, thus predicting cases where low affinity to MHC is compensated by strong TCR binding and *vice versa*.

Modeling protocols specialized for pMHC-TCR structural prediction similar to those described for antibodies have been developed.[176−178] Structural modeling was employed for optimizing force field-based scoring for evaluating the involved interactions. This modeling and scoring approach was employed to predict corresponding pMHC/TCR pairs given α and β subunits of MHC and TCR molecules and a pool of possible peptide binders.[179,180] Very recently, a customized version of Alpha Fold specifically trained for pMHC-TCR modeling was employed for the same task with prediction success correlating well with 3D model accuracy.[181]

These results demonstrate recent advances in identifying interacting p-MHC and TCR among possible candidates, but further developments are needed for a more general prediction of the target peptides of a given TCR or if and which TCRs could bind to a novel epitope.[182]

6. ANTIGEN DESIGN

With the advent of the new paradigms of reverse vaccinology 1.0 and 2.0, antigen/epitope design strategies have been envisaged to selectively present epitopes that are able to optimally elicit the desired immune response. The main design approaches are illustrated in Figure 5 and described in this section.

In this frame of thought, the idea of presenting only subdominant epitopes is related to the predominance of responses to immunodominant epitopes in the presence of the whole antigen species and was conceived to circumvent the capacity of pathogens (or malignancies) to mutate and/or mask immunodominant epitopes to evade the immune system.[184] This approach may be particularly relevant in the development of vaccines for those diseases where traditional

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

approaches are not effective such as HIV, respiratory syncytial virus (RSV), and cancer.

To this end, structure-based computational methods in the field of protein design have been implemented for the grafting of the desired epitope into novel scaffolds:[185] the idea is to present the epitope in the context of a nonantigenic protein or of an already reactive antigen, thus generating a superantigen. This aim is more straightforward for linear epitopes, while it presents significant challenges for conformational discontinuous epitopes. Rosetta includes different protocols for the so-called epitope transplantation tasks.[36] The first methods implemented allowed side-chain grafting when only side chains of binder interacting residues were grafted or backbone grafting when the entire peptide sequence was transplanted onto the scaffold. The choice of the scaffold protein was the critical step in that there should be a good superposition of the grafted motif backbone (side chain grafting) or C- and N-terminal regions (backbone grafting) and the scaffold to allow retaining the desired epitope conformation and overall protein stability. Our group developed the protocol SAGE, a computational pipeline that automates the prediction of the best grafting positions for linear epitopes onto a selected scaffold protein based on structure and sequence alignment, scoring of the secondary structure compatibility between the transplanted epitope and the scaffold, and scoring of surface exposition of the epitope.[186]

FunFoldDes follows previous approaches (i.e., FoldFromLoops) and involves the grafting of the epitope onto a scaffold topology instead of a specific protein, followed by structure folding (with some topology constraints) and sequence design.[187] Importantly, the protocol also implements the possibility of biasing the design process in the presence of the binder. The latter outlines how proteins designed in the presence of the binder deviate from the accessible energy minimum of the protein fold and highlights the relevance of introducing "function constraints" in the process. This method is particularly useful for grafting loop regions for which proper protein scaffolds are more difficult to find and allows for a global adaptation of the protein backbone to motif incorporation. It has been employed for the design of scaffolds presenting antigenic site II of the respiratory syncytial virus fusion protein, which is a good proof-of-principle result. The resulting antigen was however shown to induce only low levels of site-specific neutralizing antibodies in nonhuman primates, indicating the space for additional optimization of the design approach.[188] Applicability of this grafting strategy to more complex conformational epitopes remains difficult.

Within this frame, TopoBuilder is a protocol that reverses the computational pipeline: given the desired motif a stable protein is built around it as a carrier.[189] First, a customized topology is assembled to accommodate the motif; then structural refinements and sequence design are performed with Rosetta FunFoldDes. This computational pipeline was applied for the design of a trivalent vaccine for RSV that combined the above-mentioned scaffold with the epitope from site II and two newly designed immunogens carrying structurally complex epitopes from sites 0 and IV.[189] This construct was shown to induce an improved antibody response in immunized mice and nonhuman primates and importantly the possibility of guiding the induction of epitope-specific antibodies. Interestingly, the recently implemented protocol RosettaSurf[190] allows sequence design for protein interfaces by employing information derived from the representation of the protein as a near-continuous surface featured by shape and electrostatic properties. After epitope grafting onto a novel scaffold, protein surfaces can then be compared by the implemented surface similarity or complementarity scores, which guide the surface/sequence design process. Importantly, the surface optimization step can be biased by the presence of a binder.

In general, such immunogen design tasks usually require the exploration of a great number of possible models and subsequent rounds of optimization guided by experimental results. Moreover, the scaffold presenting the desired epitope could also present some other new epitopes that could also cause some immunodominance issues that should be taken into account. Therefore, the current efforts now focus on improving or enhancing antibody responses. One possibility that is being explored is the use of nanoparticles as multivalent antigen carriers. Such nanosystems exploit inherent immune system features, optimized to detect exogenous particles. Indeed, for their shape and size they display advantageous biodistribution in the lymphatic system and uptake in lymphatic cells and can show increased immune response with respect to soluble antigens thanks to effects such as B-cell receptor cross-linking.[191,192] Alongside more traditionally employed natural protein assemblies, self-assembling nanoparticles have been designed computationally by symmetrical docking of building blocks, followed by the design of protein−protein interacting surfaces.[193] These nanoparticles were shown to be versatile and promising platforms for displaying several antigen molecules on the surface[193] including heterotypic antigens[194,195] leading the way to more customizable and easy-to-obtain platforms. The topic has been reviewed recently.[196−198]

Another immunogen design strategy includes the thermostabilization of known poorly stable antigenic proteins. Viral fusion proteins usually have a metastable prefusion conformation that rearranges after fusion of the viral envelope with the host cells and that often does not expose the epitope targeted by neutralizing antibodies. Therefore, stabilization of this class of proteins aims at retaining the immune-active prefusion conformations. These approaches have been employed for the development of prefusion stabilized forms of RSVF,[199] SARS-Cov-2 spike protein,[200] and HIV envelope.[201] For achieving these results, structural knowledge has guided various rounds of optimization mostly relying on experimental techniques. Indeed, the application of computational structure-based methods for the design of thermostabilized immunogens seems still scarce, but as in the design of antibodies, there is great potential in the use of high-throughput protocols for the prediction of stabilizing mutations and in the computational screening of the properties of stabilized mutants. For example, PROSS is a tool available as a Web server that leverages information from proteins' evolutionary sequence variability and is able to predict mutations that will increase protein stability based on $\Delta\Delta G$ calculations.[202]

Molecular dynamics and free energy calculations have been employed to identify mutations increasing the stability of interactions between the pentameric subunits in the icosahedral capsid of the foot and mouth disease virus. This resulted in a safer experimental vaccine with an improved shelf life.[203]

The protein structures are reproduced from PDB codes 4JHW[204] (RSV fusion protein in its prefusion form), 2WH6,[205] and 3LHP[206] (examples of scaffold proteins).

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

7. CONCLUSIONS AND PERSPECTIVES

Structural immunology and vaccinology are among the research areas that are likely to benefit most from the current explosion in computing power and the advent of increasingly potent AI approaches.

The increase in computing power, combined with more and more efficient algorithms and improved force-field parameters, is already ushering in the era of whole-virion models and simulations. Approaches simulating multiple copies of an antigenic protein in a realistic environment have, in fact, recently started to appear. In the context of Covid research, the Hummer group, for instance, set out to identify possible antibody binding sites, with multimicrosecond molecular dynamics simulations of a 4.1 million atom system containing a patch of viral membrane with four full-length, fully glycosylated and palmitoylated S proteins. The authors combined the analysis of steric accessibility, structural rigidity, sequence conservation, and generic antibody binding signatures, to identify efficaciously the known epitopes on S. Moreover, this work revealed additional promising epitope candidates for structure-based vaccine design.[207]

The Amaro group recently reported impressive mesoscale, all-atom MD simulations of two evolutionary-linked glycosylated influenza A virions.[208] The simulations reveal the main molecular motions of the principal surface antigenic targets of the influenza virus, hemagglutinin (HA), and neuraminidase (NA). Indeed, NA head tilting, HA ectodomain tilting, and HA head breathing are characterized at full atomistic resolution. The flexibility of NA reveals a cryptic epitope targeted by a novel convalescent human-donor-derived monoclonal antibody. Importantly, this type of work, combining extensive structural characterizations of the proteins in a realistic crowded environment and the possibility to analyze multiple copies of the same molecule in its native conditions, shows the possibility to obtain previously unappreciated (and inaccessible) views on the dynamics of antigenic proteins HA and NA. In this context, MD simulations could unveil transient intermediates able to induce antibodies against epitopes that, although conserved, are exposed only transiently in native molecules and may not be visible in a static 3D representation of the antigen. This, combined with the epitope prediction approaches described above, will help develop antibodies and vaccine components efficiently, for instance, by guiding the stabilization of specific conformations via mutagenesis.

Approaches like these will become more and more common as we begin entering the era of exascale computing. In this framework, efforts to adapt and port MD simulations to these new architectures will also play a key role.

Significant advances will likely stem from the integration of AI with computational biology and biotechnology for the development of new therapies and treatments.

One promising area of research is of course the use of AI to predict structures as well as design and optimize proteins as antigens, antigen-binders, and/or antibodies.[209,210]

If AI were to maintain its current promise, this may also find translation in the development of personalized medicine protocols. By analyzing a patient's genetic data, medical history, and other factors, AI algorithms could predict which antigens will be most important to target and, as a consequence, which treatments are most likely to be effective for that individual.

In cancer treatment, AI supported interventions may be translated into the identification and molecular profiling of patient-specific cancer cells and then designing antibodies that can selectively bind to and destroy those cells while leaving healthy cells unharmed. Similarly, one could design novel and specific antigens that are able to elicit a strong protective immune response.

The combination of AI, protein design, and CAR-T cell therapy is an exciting area of research that has the potential to significantly improve cancer treatment outcomes. By identifying specific proteins or antigens that have specific signatures in cancer cells but not on healthy cells, researchers can design new CARs with improved safety and efficacy.

The same type of philosophy can be applied to the design of new proteins that can be used as CARs.

Clearly, limitations still have to be considered with a critical eye. Indeed, AI-based antigen and antibody design will depend on the correct understanding and modeling of the biophysical principles that underpin complex cell molecular recognition phenomena, a problem that is still far from being fully solved.

In this framework, it is important to note that similar to experimentally derived immunologically active molecules, unsolved issues remain. *In silico* developed antibodies, engineered T-cells, and immunogens could cause unwanted inflammatory reactions, toxicity, resistance phenomena, and other side effects. One advantage of (AI-powered) *in silico* screenings of increasingly rich databases is that they could help reduce or predict these risks. Given the complex interplay of the factors involved and the inherent variation of responses of different individuals, however, risks cannot be fully eliminated but can be mitigated by an improved knowledge of adverse response predictors. Furthermore, the computational design of immunologically active molecules remains difficult and usually involves several rounds of optimization with experimental testing to obtain molecules of good fitness and efficacy.

To advance along these fascinating avenues, an ever increasing integration of structural biology, computational biochemistry, computer and data science, and algorithm design will be necessary. New researchers and professionals that already embody this necessity are starting to appear. Importantly, novel (and at present mostly unpredictable) opportunities for the development of new research lines will begin to emerge.

We envision that the above-described developments will come of age in the next few years, opening new avenues for fundamental and applied research and facilitating the creation of new biomolecules with specific modes of action to be used both as therapeutics in the treatment of diverse diseases and as molecular tools to disentangle the intricacies of immune molecular mechanisms.

AUTHOR INFORMATION

Corresponding Author

Giorgio Colombo − *Department of Chemistry, University of Pavia, 27100 Pavia, Italy;* orcid.org/0000-0002-1318-668X; Email: g.colombo@unipv.it

Author

Federica Guarra − *Department of Chemistry, University of Pavia, 27100 Pavia, Italy;* orcid.org/0000-0002-7301-3000

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00513

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding from AIRC under IG 2022 - ID. 27139 project and from PRIN (grant 20209KYCH9) to Giorgio Colombo. Federica Guarra is supported through the AIRC fellowship Love Design 2021 (ID 26647-2021). Funding is also acknowledged through the IMMUNO-HUB, T4-CN-02, project from the Ministero della Salute (Italy).

■ REFERENCES

(1) Pricer, R.; Gestwicki, J. E.; Mapp, A. K. From Fuzzy to Function: The New Frontier of Protein−Protein Interactions. *Acc. Chem. Res.* **2017**, *50*, 584−589.
(2) Hakes, L.; Pinney, J. W.; Robertson, D. L.; Lovell, S. C. Protein-Protein Interaction Networks and Biology—What's the Connection? *Nat. Biotechnol.* **2008**, *26*, 69−72.
(3) Lapidoth, G. D.; Baran, D.; Pszolla, G. M.; Norn, C.; Alon, A.; Tyka, M. D.; Fleishman, S. J. AbDesign: An Algorithm for Combinatorial Backbone Design Guided by Natural Conformations and Sequences. *Proteins Struct. Funct. Bioinforma.* **2015**, *83*, 1385−1406.
(4) Yin, Q.; Fu, T.-M.; Li, J.; Wu, H. Structural Biology of Innate Immunity. *Annu. Rev. Immunol.* **2015**, *33*, 393−416.
(5) Dyer, M. D.; Murali, T. M.; Sobral, B. W. The Landscape of Human Proteins Interacting with Viruses and Other Pathogens. *PLOS Pathog.* **2008**, *4*, No. e32.
(6) Shah, P. S.; Link, N.; Jang, G. M.; Sharp, P. P.; Zhu, T.; Swaney, D. L.; Johnson, J. R.; Von Dollen, J.; Ramage, H. R.; Satkamp, L.; Newton, B.; Hüttenhain, R.; Petit, M. J.; Baum, T.; Everitt, A.; Laufman, O.; Tassetto, M.; Shales, M.; Stevenson, E.; Iglesias, G. N.; Shokat, L.; Tripathi, S.; Balasubramaniam, V.; Webb, L. G.; Aguirre, S.; Willsey, A. J.; Garcia-Sastre, A.; Pollard, K. S.; Cherry, S.; Gamarnik, A. V.; Marazzi, I.; Taunton, J.; Fernandez-Sesma, A.; Bellen, H. J.; Andino, R.; Krogan, N. J. Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **2018**, *175*, 1931−1945.
(7) Plotkin, S. History of Vaccination. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12283−12287.
(8) Fang, E.; Liu, X.; Li, M.; Zhang, Z.; Song, L.; Zhu, B.; Wu, X.; Liu, J.; Zhao, D.; Li, Y. Advances in COVID-19 MRNA Vaccine Development. *Signal Transduct. Target. Ther.* **2022**, *7*, 94.
(9) Chen, X.; Huang, H.; Ju, J.; Sun, R.; Zhang, J. Impact of Vaccination on the COVID-19 Pandemic in U.S. States. *Sci. Rep.* **2022**, *12*, 1554.
(10) Wallis, R. S.; O'Garra, A.; Sher, A.; Wack, A. Host-Directed Immunotherapy of Viral and Bacterial Infections: Past, Present and Future. *Nat. Rev. Immunol.* **2023**, *23*, 121−133.
(11) Twomey, J. D.; Zhang, B. Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics. *AAPS J.* **2021**, *23*, 39.
(12) Sharma, P.; Allison, J. P. Immune Checkpoint Targeting in Cancer Therapy: Toward Combination Strategies with Curative Potential. *Cell* **2015**, *161*, 205−214.
(13) Galon, J.; Bruni, D. Tumor Immunology and Tumor Evolution: Intertwined Histories. *Immunity* **2020**, *52*, 55−81.
(14) Chen, Y.-J.; Abila, B.; Mostafa Kamel, Y. CAR-T: What Is Next? *Cancers (Basel).* **2023**, *15*, 663.
(15) Holstein, S. A.; Lunning, M. A. CAR T-Cell Therapy in Hematologic Malignancies: A Voyage in Progress. *Clin. Pharmacol. Ther.* **2020**, *107*, 112−122.
(16) Lee, D. W.; Barrett, D. M.; Mackall, C.; Orentas, R.; Grupp, S. A. The Future Is Now: Chimeric Antigen Receptors as New Targeted Therapies for Childhood Cancer. *Clin. Cancer Res.* **2012**, *18*, 2780−2790.
(17) DeRenzo, C.; Gottschalk, S. Genetic Modification Strategies to Enhance CAR T Cell Persistence for Patients with Solid Tumors. *Front. Immunol.* **2019**, *10*. DOI: 10.3389/fimmu.2019.00218.
(18) Moxon, R.; Reche, P. A.; Rappuoli, R. Editorial: Reverse Vaccinology. *Front. Immunol.* **2019**, *10*, 1−2.
(19) Bidmos, F. A.; Siris, S.; Gladstone, C. A.; Langford, P. R. Bacterial Vaccine Antigen Discovery in the Reverse Vaccinology 2.0 Era: Progress and Challenges. *Front. Immunol.* **2018**, *9*, 1−7.
(20) Serruto, D.; Bottomley, M. J.; Ram, S.; Giuliani, M. M.; Rappuoli, R. The New Multicomponent Vaccine against Meningococcal Serogroup B, 4CMenB: Immunological, Functional and Structural Characterization of the Antigens. *Vaccine* **2012**, *30*, B87−B97.
(21) Dormitzer, P. R.; Grandi, G.; Rappuoli, R. Structural Vaccinology Starts to Deliver. *Nat. Rev. Microbiol.* **2012**, *10*, 807−813.
(22) Marchand, A.; Van Hall-Beauvais, A. K.; Correia, B. E. Computational Design of Novel Protein−Protein Interactions − An Overview on Methodological Approaches and Applications. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102370.
(23) Linsky, T. W.; Vergara, R.; Codina, N.; Nelson, J. W.; Walker, M. J.; Su, W.; Barnes, C. O.; Hsiang, T.-Y.; Esser-Nobis, K.; Yu, K.; Reneer, Z. B.; Hou, Y. J.; Priya, T.; Mitsumoto, M.; Pong, A.; Lau, U. Y.; Mason, M. L.; Chen, J.; Chen, A.; Berrocal, T.; Peng, H.; Clairmont, N. S.; Castellanos, J.; Lin, Y.-R.; Josephson-Day, A.; Baric, R. S.; Fuller, D. H.; Walkey, C. D.; Ross, T. M.; Swanson, R.; Bjorkman, P. J.; Gale, M.; Blancas-Mejia, L. M.; Yen, H.-L.; Silva, D.-A. De Novo Design of Potent and Resilient HACE2 Decoys to Neutralize SARS-CoV-2. *Science* **2020**, *370*, 1208−1214.
(24) Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Sheffler, W.; Park, J. S.; Jude, K. M.; Marković, I.; Kadam, R. U.; Verschueren, K. H. G.; Verstraete, K.; Walsh, S. T. R.; Bennett, N.; Phal, A.; Yang, A.; Kozodoy, L.; DeWitt, M.; Picton, L.; Miller, L.; Strauch, E. M.; DeBouver, N. D.; Pires, A.; Bera, A. K.; Halabiya, S.; Hammerson, B.; Yang, W.; Bernard, S.; Stewart, L.; Wilson, I. A.; Ruohola-Baker, H.; Schlessinger, J.; Lee, S.; Savvides, S. N.; Garcia, K. C.; Baker, D. Design of Protein-Binding Proteins from the Target Structure Alone. *Nature* **2022**, *605*, 551−560.
(25) Silva, D.-A.; Yu, S.; Ulge, U. Y.; Spangler, J. B.; Jude, K. M.; Labão-Almeida, C.; Ali, L. R.; Quijano-Rubio, A.; Ruterbusch, M.; Leung, I.; Biary, T.; Crowley, S. J.; Marcos, E.; Walkey, C. D.; Weitzner, B. D.; Pardo-Avila, F.; Castellanos, J.; Carter, L.; Stewart, L.; Riddell, S. R.; Pepper, M.; Bernardes, G. J. L.; Dougan, M.; Garcia, K. C.; Baker, D. De Novo Design of Potent and Selective Mimics of IL-2 and IL-15. *Nature* **2019**, *565*, 186−191.
(26) Coley, A. M.; Gupta, A.; Murphy, V. J.; Bai, T.; Kim, H.; Anders, R. F.; Foley, M.; Batchelor, A. H. Structure of the Malaria Antigen AMA1 in Complex with a Growth-Inhibitory Antibody. *PLOS Pathog.* **2007**, *3*, No. e138.
(27) Sormanni, P.; Aprile, F. A.; Vendruscolo, M. Third Generation Antibody Discovery Methods: In Silico Rational Design. *Chem. Soc. Rev.* **2018**, *47*, 9137−9157.
(28) Kaplon, H.; Chenoweth, A.; Crescioli, S.; Reichert, J. M. Antibodies to Watch in 2022. *MAbs* **2022**, *14*, 2014296.
(29) Fischman, S.; Ofran, Y. Computational Design of Antibodies. *Curr. Opin. Struct. Biol.* **2018**, *51*, 156−162.
(30) Norman, R. A.; Ambrosetti, F.; Bonvin, A. M. J. J.; Colwell, L. J.; Kelm, S.; Kumar, S.; Krawczyk, K. Computational Approaches to Therapeutic Antibody Design: Established Methods and Emerging Trends. *Brief. Bioinform.* **2020**, *21*, 1549−1567.
(31) Hummer, A. M.; Abanades, B.; Deane, C. M. Advances in Computational Structure-Based Antibody Design. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102379.
(32) North, B.; Lehmann, A.; Dunbrack, R. L. A New Clustering of Antibody CDR Loop Conformations. *J. Mol. Biol.* **2011**, *406*, 228−256.
(33) *Rosetta*. https://www.rosettacommons.org/about.
(34) Leman, J. K.; Weitzner, B. D.; Renfrew, P. D.; Lewis, S. M.; Moretti, R.; Watkins, A. M.; Mulligan, V. K.; Lyskov, S.; Adolf-Bryfogle, J.; Labonte, J. W.; Krys, J.; Bystroff, C.; Schief, W.; Gront,

D.; Schueler-Furman, O.; Baker, D.; Bradley, P.; Dunbrack, R.; Kortemme, T.; Leaver-Fay, A.; Strauss, C. E. M.; Meiler, J.; Kuhlman, B.; Gray, J. J.; Bonneau, R. Better Together: Elements of Successful Scientific Software Development in a Distributed Collaborative Community. *PLoS Comput. Biol.* **2020**, *16*, e1007507.

(35) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; Basanta, B.; Bender, B. J.; Blacklock, K.; Bonet, J.; Boyken, S. E.; Bradley, P.; Bystroff, C.; Conway, P.; Cooper, S.; Correia, B. E.; Coventry, B.; Das, R.; De Jong, R. M.; DiMaio, F.; Dsilva, L.; Dunbrack, R.; Ford, A. S.; Frenz, B.; Fu, D. Y.; Geniesse, C.; Goldschmidt, L.; Gowthaman, R.; Gray, J. J.; Gront, D.; Guffy, S.; Horowitz, S.; Huang, P. S.; Huber, T.; Jacobs, T. M.; Jeliazkov, J. R.; Johnson, D. K.; Kappel, K.; Karanicolas, J.; Khakzad, H.; Khar, K. R.; Khare, S. D.; Khatib, F.; Khramushin, A.; King, I. C.; Kleffner, R.; Koepnick, B.; Kortemme, T.; Kuenze, G.; Kuhlman, B.; Kuroda, D.; Labonte, J. W.; Lai, J. K.; Lapidoth, G.; Leaver-Fay, A.; Lindert, S.; Linsky, T.; London, N.; Lubin, J. H.; Lyskov, S.; Maguire, J.; Malmström, L.; Marcos, E.; Marcu, O.; Marze, N. A.; Meiler, J.; Moretti, R.; Mulligan, V. K.; Nerli, S.; Norn, C.; Ó'Conchúir, S.; Ollikainen, N.; Ovchinnikov, S.; Pacella, M. S.; Pan, X.; Park, H.; Pavlovicz, R. E.; Pethe, M.; Pierce, B. G.; Pilla, K. B.; Raveh, B.; Renfrew, P. D.; Burman, S. S. R.; Rubenstein, A.; Sauer, M. F.; Scheck, A.; Schief, W.; Schueler-Furman, O.; Sedan, Y.; Sevy, A. M.; Sgourakis, N. G.; Shi, L.; Siegel, J. B.; Silva, D. A.; Smith, S.; Song, Y.; Stein, A.; Szegedy, M.; Teets, F. D.; Thyme, S. B.; Wang, R. Y. R.; Watkins, A.; Zimmerman, L.; Bonneau, R. Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665−680.

(36) Schoeder, C. T.; Schmitz, S.; Adolf-Bryfogle, J.; Sevy, A. M.; Finn, J. A.; Sauer, M. F.; Bozhanova, N. G.; Mueller, B. K.; Sangha, A. K.; Bonet, J.; Sheehan, J. H.; Kuenze, G.; Marlow, B.; Smith, S. T.; Woods, H.; Bender, B. J.; Martina, C. E.; Del Alamo, D.; Kodali, P.; Gulsevin, A.; Schief, W. R.; Correia, B. E.; Crowe, J. E.; Meiler, J.; Moretti, R. Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design. *Biochemistry* **2021**, *60*, 825−846.

(37) Weitzner, B. D.; Jeliazkov, J. R.; Lyskov, S.; Marze, N.; Kuroda, D.; Frick, R.; Adolf-Bryfogle, J.; Biswas, N.; Dunbrack, R. L.; Gray, J. J. Modeling and Docking of Antibody Structures with Rosetta. *Nat. Protoc.* **2017**, *12*, 401−416.

(38) Adolf-Bryfogle, J.; Xu, Q.; North, B.; Lehmann, A.; Dunbrack Jr, R. L. PyIgClassify: A Database of Antibody CDR Structural Classifications. *Nucleic Acids Res.* **2015**, *43*, D432−D438.

(39) Finn, J. A.; Leman, J. K.; Willis, J. R.; Cisneros, A.; Crowe, J. E.; Meiler, J. Improving Loop Modeling of the Antibody Complementarity-Determining Region 3 Using Knowledge-Based Restraints. *PLoS One* **2016**, *11*, e0154811.

(40) Weitzner, B. D.; Gray, J. J. Accurate Structure Prediction of CDR H3 Loops Enabled by a Novel Structure-Based C-Terminal Constraint. *J. Immunol.* **2017**, *198*, 505−515.

(41) Marks, C.; Nowak, J.; Klostermann, S.; Georges, G.; Dunbar, J.; Shi, J.; Kelm, S.; Deane, C. M. Sphinx: Merging Knowledge-Based and Ab Initio Approaches to Improve Protein Loop Prediction. *Bioinformatics* **2017**, *33*, 1346−1353.

(42) Norn, C. H.; Lapidoth, G.; Fleishman, S. J. High-Accuracy Modeling of Antibody Structures by a Search for Minimum-Energy Recombination of Backbone Fragments. *Proteins Struct. Funct. Bioinforma.* **2017**, *85*, 30−38.

(43) Lapidoth, G.; Parker, J.; Prilusky, J.; Fleishman, S. J. AbPredict 2: A Server for Accurate and Unstrained Structure Prediction of Antibody Variable Domains. *Bioinformatics* **2019**, *35*, 1591−1593.

(44) Leem, J.; Dunbar, J.; Georges, G.; Shi, J.; Deane, C. M. ABodyBuilder: Automated Antibody Structure Prediction with Data−Driven Accuracy Estimation. *MAbs* **2016**, *8*, 1259−1268.

(45) Yamashita, K.; Ikeda, K.; Amada, K.; Liang, S.; Tsuchiya, Y.; Nakamura, H.; Shirai, H.; Standley, D. M. Kotai Antibody Builder: Automated High-Resolution Structural Modeling of Antibodies. *Bioinformatics* **2014**, *30*, 3279−3280.

(46) Marcatili, P.; Olimpieri, P. P.; Chailyan, A.; Tramontano, A. Antibody Modeling Using the Prediction of ImmunoGlobulin Structure (PIGS) Web Server. *Nat. Protoc.* **2014**, *9*, 2771−2783.

(47) Choi, Y.; Deane, C. M. FREAD Revisited: Accurate Loop Structure Prediction Using a Database Search Algorithm. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 1431−1440.

(48) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 778−795.

(49) Almagro, J. C.; Teplyakov, A.; Luo, J.; Sweet, R. W.; Kodangattil, S.; Hernandez-Guzman, F.; Gilliland, G. L. Second Antibody Modeling Assessment (AMA-II). *Proteins Struct. Funct. Bioinforma.* **2014**, *82*, 1553−1562.

(50) Graves, J.; Byerly, J.; Priego, E.; Makkapati, N.; Parish, S. V.; Medellin, B.; Berrondo, M. A Review of Deep Learning Methods for Antibodies. *Antibodies.* **2020**, *9*, 12.

(51) Kim, J.; McFee, M.; Fang, Q.; Abdin, O.; Kim, P. M. Computational and Artificial Intelligence-Based Methods for Antibody Development. *Trends Pharmacol. Sci.* **2023**, *44*, 175−189.

(52) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

(53) Abanades, B.; Georges, G.; Bujotzek, A.; Deane, C. M. ABlooper: Fast Accurate Antibody CDR Loop Structure Prediction with Accuracy Estimation. *Bioinformatics* **2022**, *38*, 1877−1880.

(54) Ruffolo, J. A.; Sulam, J.; Gray, J. J. Antibody Structure Prediction Using Interpretable Deep Learning. *Patterns* **2022**, *3*, 100406.

(55) Kovaltsuk, A.; Leem, J.; Kelm, S.; Snowden, J.; Deane, C. M.; Krawczyk, K. Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. *J. Immunol.* **2018**, *201*, 2502−2509.

(56) Ruffolo, J. A.; Gray, J. J. Fast, Accurate Antibody Structure Prediction from Deep Learning on Massive Set of Natural Antibodies. *Biophys. J.* **2022**, *121*, 155a−156a.

(57) AlQuraishi, M. Machine Learning in Protein Structure Prediction. *Curr. Opin. Chem. Biol.* **2021**, *65*, 1−8.

(58) Harris, L. J.; Larson, S. B.; Hasel, K. W.; McPherson, A. Refined Structure of an Intact IgG2a Monoclonal Antibody,. *Biochemistry* **1997**, *36*, 1581−1597.

(59) Bangaru, S.; Lang, S.; Schotsaert, M.; Vanderven, H. A.; Zhu, X.; Kose, N.; Bombardi, R.; Finn, J. A.; Kent, S. J.; Gilchuk, P.; Gilchuk, I.; Turner, H. L.; García-Sastre, A.; Li, S.; Ward, A. B.; Wilson, I. A.; Crowe, J. E. A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. *Cell* **2019**, *177*, 1136−1152.

(60) Yin, R.; Feng, B. Y.; Varshney, A.; Pierce, B. G. Benchmarking AlphaFold for Protein Complex Modeling Reveals Accuracy Determinants. *Protein Sci.* **2022**, *31*, No. e4379.

(61) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein-Protein Docking. *Nat. Protoc.* **2017**, *12*, 255−278.

(62) Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins* **2003**, *52*, 80−87.

(63) Sircar, A.; Gray, J. J. SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models. *PLOS Comput. Biol.* **2010**, *6*, No. e1000644.

(64) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333

(65) Torchala, M.; Moal, I. H.; Chaleil, R. A. G.; Fernandez-Recio, J.; Bates, P. A. SwarmDock: A Server for Flexible Protein−Protein Docking. *Bioinformatics* **2013**, *29*, 807−809.
(66) Jiménez-García, B.; Roel-Touris, J.; Romero-Durana, M.; Vidal, M.; Jiménez-González, D.; Fernández-Recio, J. LightDock: A New Multi-Scale Approach to Protein−Protein Docking. *Bioinformatics* **2018**, *34*, 49−55.
(67) van Noort, C. W.; Honorato, R. V.; Bonvin, A. M. J. J. Information-Driven Modeling of Biomolecular Complexes. *Curr. Opin. Struct. Biol.* **2021**, *70*, 70−77.
(68) Ambrosetti, F.; Jiménez-García, B.; Roel-Touris, J.; Bonvin, A. M. J. J. Modeling Antibody-Antigen Complexes by Information-Driven Docking. *Structure* **2020**, *28*, 119−129.
(69) Guest, J. D.; Vreven, T.; Zhou, J.; Moal, I.; Jeliazkov, J. R.; Gray, J. J.; Weng, Z.; Pierce, B. G. An Expanded Benchmark for Antibody-Antigen Docking and Affinity Prediction Reveals Insights into Antibody Recognition Determinants. *Structure* **2021**, *29*, 606−621.
(70) Tran, M. H.; Schoeder, C. T.; Schey, K. L.; Meiler, J. Computational Structure Prediction for Antibody-Antigen Complexes From Hydrogen-Deuterium Exchange Mass Spectrometry: Challenges and Outlook. *Front. Immunol.* **2022**, *13*, 1−22.
(71) Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. Efficient and Robust Proteome-Wide Approaches for Cross-Linking Mass Spectrometry. *Nat. Protoc.* **2018**, *13*, 2964−2990.
(72) Xia, B.; Mamonov, A.; Leysen, S.; Allen, K. N.; Strelkov, S. V.; Paschalidis, I. C.; Vajda, S.; Kozakov, D. Accounting for Observed Small Angle X-Ray Scattering Profile in the Protein−Protein Docking Server Cluspro. *J. Comput. Chem.* **2015**, *36*, 1568−1572.
(73) Dai, B.; Bailey-Kellogg, C. Protein Interaction Interface Region Prediction by Geometric Deep Learning. *Bioinformatics* **2021**, *37*, 2580−2588.
(74) Pittala, S.; Bailey-Kellogg, C. Learning Context-Aware Structural Representations to Predict Antigen and Antibody Binding Interfaces. *Bioinformatics* **2020**, *36*, 3996−4003.
(75) Liberis, E.; Velickovic, P.; Sormanni, P.; Vendruscolo, M.; Lio, P. Parapred: Antibody Paratope Prediction Using Convolutional and Recurrent Neural Networks. *Bioinformatics* **2018**, *34*, 2944−2950.
(76) Ambrosetti, F.; Olsen, T. H.; Olimpieri, P. P.; Jiménez-García, B.; Milanetti, E.; Marcatilli, P.; Bonvin, A. M. J. J. ProABC-2: PRediction of AntiBody Contacts v2 and Its Application to Information-Driven Docking. *Bioinformatics* **2020**, *36*, 5107−5108.
(77) Moal, I. H.; Torchala, M.; Bates, P. A.; Fernández-Recio, J. The Scoring of Poses in Protein-Protein Docking: Current Capabilities and Future Directions. *BMC Bioinformatics* **2013**, *14*, 286.
(78) Schneider, C.; Buchanan, A.; Taddese, B.; Deane, C. M. DLAB: Deep Learning Methods for Structure-Based Virtual Screening of Antibodies. *Bioinformatics* **2022**, *38*, 377−383.
(79) Renaud, N.; Geng, C.; Georgievska, S.; Ambrosetti, F.; Ridder, L.; Marzella, D. F.; Réau, M. F.; Bonvin, A. M. J. J.; Xue, L. C. DeepRank: A Deep Learning Framework for Data Mining 3D Protein-Protein Interfaces. *Nat. Commun.* **2021**, *12*, 1−8.
(80) Adolf-Bryfogle, J.; Kalyuzhniy, O.; Kubitz, M.; Weitzner, B. D.; Hu, X.; Adachi, Y.; Schief, W. R.; Dunbrack, R. L. RosettaAntibodyDesign (RAbD): A General Framework for Computational Antibody Design. *PLoS Comput. Biol.* **2018**, *14*, No. e1006112.
(81) Lipsh-Sokolik, R.; Listov, D.; Fleishman, S. J. The AbDesign Computational Pipeline for Modular Backbone Assembly and Design of Binders and Enzymes. *Protein Sci.* **2021**, *30*, 151−159.
(82) Corti, D.; Lanzavecchia, A. Broadly Neutralizing Antiviral Antibodies. *Annu. Rev. Immunol.* **2013**, *31*, 705−742.
(83) Sevy, A. M.; Jacobs, T. M.; Crowe, J. E., Jr; Meiler, J. Design of Protein Multi-Specificity Using an Independent Sequence Search Reduces the Barrier to Low Energy Sequences. *PLOS Comput. Biol.* **2015**, *11*, No. e1004300.
(84) Sevy, A. M.; Panda, S.; Crowe Jr, J. E.; Meiler, J.; Vorobeychik, Y. Integrating Linear Optimization with Structural Modeling to Increase HIV Neutralization Breadth. *PLOS Comput. Biol.* **2018**, *14*, No. e1005999.
(85) Rangel, M. A.; Bedwell, A.; Costanzi, E.; Taylor, R. J.; Russo, R.; Bernardes, G. J. L.; Ricagno, S.; Frydman, J.; Vendruscolo, M.; Sormanni, P. Fragment-Based Computational Design of Antibodies Targeting Structured Epitopes. *Sci. Adv.* **2022**, *8*, 1−14.
(86) Eguchi, R. R.; Choe, C. A.; Huang, P. S. Ig-VAE: Generative Modeling of Protein Structure by Direct 3D Coordinate Generation. *PLoS Comput. Biol.* **2022**, *18*, e1010271.
(87) Jin, W.; Wohlwend, J.; Barzilay, R.; Jaakkola, T. Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-Design. *arXiv:2110.04624* **2021**. DOI: 10.48550/arXiv.2110.04624
(88) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; Bortoli, V. De; Mathieu, E.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. Broadly Applicable and Accurate Protein Design by Integrating Structure Prediction Networks and Diffusion Generative Models. *bioRxiv* **2022**, DOI: 10.1101/2022.12.09.519842.
(89) Ingraham, J.; Baranov, M.; Costello, Z.; Frappier, V.; Ismail, A.; Tie, S.; Wang, W.; Xue, V.; Obermeyer, F.; Beam, A.; Grigoryan, G. Illuminating Protein Space with a Programmable Generative Model. *bioRxiv* **2022**, DOI: 10.1101/2022.12.01.518682.
(90) Warszawski, S.; Katz, A. B.; Lipsh, R.; Khmelnitsky, L.; Nissan, G. B.; Javitt, G.; Dym, O.; Unger, T.; Knop, O.; Albeck, S.; Diskin, R.; Fass, D.; Sharon, M.; Fleishman, S. J. Optimizing Antibody Affinity and Stability by the Automated Design of the Variable Light-Heavy Chain Interfaces. *PLoS Comput. Biol.* **2019**, *15*, 1−24.
(91) Wang, M.; Cang, Z.; Wei, G. W. A Topology-Based Network Tree for the Prediction of Protein−Protein Binding Affinity Changes Following Mutation. *Nat. Mach. Intell.* **2020**, *2*, 116−123.
(92) Myung, Y.; Pires, D. E. V.; Ascher, D. B. MmCSM-AB: Guiding Rational Antibody Engineering through Multiple Point Mutations. *Nucleic Acids Res.* **2020**, *48*, W125−W131.
(93) Muyldermans, S. A Guide to: Generation and Design of Nanobodies. *FEBS J.* **2021**, *288*, 2084−2102.
(94) Yang, J.; Zhang, Z.; Yang, F.; Zhang, H.; Wu, H.; Zhu, F.; Xue, W. Computational Design and Modeling of Nanobodies toward SARS-CoV-2 Receptor Binding Domain. *Chem. Biol. Drug Des.* **2021**, *98*, 1−18.
(95) Sircar, A.; Sanni, K. A.; Shi, J.; Gray, J. J. Analysis and Modeling of the Variable Region of Camelid Single-Domain Antibodies. *J. Immunol.* **2011**, *186*, 6357−6367.
(96) Pollard, A. J.; Bijker, E. M. A Guide to Vaccinology: From Basic Principles to New Developments. *Nat. Rev. Immunol.* **2021**, *21*, 83−100.
(97) Kwong, P. D. What Are the Most Powerful Immunogen Design Vaccine Strategies?: A Structural Biologist's Perspective. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, a029470.
(98) Roudko, V.; Greenbaum, B.; Bhardwaj, N. Computational Prediction and Validation of Tumor-Associated Neoantigens. *Front. Immunol.* **2020**, *11*, 1−11.
(99) Rappuoli, R. Reverse Vaccinology Rino Rappuoli. *Curr. Opin. Microbiol.* **2000**, *3*, 445−450.
(100) Rappuoli, R.; Bottomley, M. J.; D'Oro, U.; Finco, O.; De Gregorio, E. Reverse Vaccinology 2.0: Human Immunology Instructs Vaccine Antigen Design. *J. Exp. Med.* **2016**, *213*, 469−481.
(101) He, Y.; Xiang, Z.; Mobley, H. L. T. Vaxign: The First Web-Based Vaccine Design Program for Reverse Vaccinology and Applications for Vaccine Development. *J. Biomed. Biotechnol.* **2010**, *2010*, 297505.
(102) Vivona, S.; Bernante, F.; Filippini, F. NERVE: New Enhanced Reverse Vaccinology Environment. *BMC Biotechnol.* **2006**, *6*, 35.
(103) Jaiswal, V.; Chanumolu, S. K.; Gupta, A.; Chauhan, R. S.; Rout, C. Jenner-Predict Server: Prediction of Protein Vaccine Candidates (PVCs) in Bacteria Based on Host-Pathogen Interactions. *BMC Bioinformatics* **2013**, *14*, 211.

(104) Rizwan, M.; Naz, A.; Ahmad, J.; Naz, K.; Obaid, A.; Parveen, T.; Ahsan, M.; Ali, A. VacSol: A High Throughput in Silico Pipeline to Predict Potential Therapeutic Targets in Prokaryotic Pathogens Using Subtractive Reverse Vaccinology. *BMC Bioinformatics* **2017**, *18*, 106.
(105) Doytchinova, I. A.; Flower, D. R. VaxiJen: A Server for Prediction of Protective Antigens, Tumour Antigens and Subunit Vaccines. *BMC Bioinformatics* **2007**, *8*, 4.
(106) Goodswen, S. J.; Kennedy, P. J.; Ellis, J. T. Vacceed: A High-Throughput in Silico Vaccine Candidate Discovery Pipeline for Eukaryotic Pathogens Based on Reverse Vaccinology. *Bioinformatics* **2014**, *30*, 2381−2383.
(107) Ong, E.; Wang, H.; Wong, M. U.; Seetharaman, M.; Valdez, N.; He, Y. Vaxign-ML: Supervised Machine Learning Reverse Vaccinology Model for Improved Prediction of Bacterial Protective Antigens. *Bioinformatics* **2020**, *36*, 3185−3191.
(108) Bowman, B. N.; McAdam, P. R.; Vivona, S.; Zhang, J. X.; Luong, T.; Belew, R. K.; Sahota, H.; Guiney, D.; Valafar, F.; Fierer, J.; Woelk, C. H. Improving Reverse Vaccinology with a Machine Learning Approach. *Vaccine* **2011**, *29*, 8156−8164.
(109) Heinson, A. I.; Gunawardana, Y.; Moesker, B.; Hume, C. C. D.; Vataga, E.; Hall, Y.; Stylianou, E.; McShane, H.; Williams, A.; Niranjan, M.; Woelk, C. H. Enhancing the Biological Relevance of Machine Learning Classifiers for Reverse Vaccinology. *Int. J. Mol. Sci.* **2017**, *18*, 312.
(110) Goodswen, S. J.; Kennedy, P. J.; Ellis, J. T. A Guide to Current Methodology and Usage of Reverse Vaccinology towards in Silico Vaccine Discovery. *FEMS Microbiol. Rev.* **2023**, *47*, 1−22.
(111) Dalsass, M.; Brozzi, A.; Medini, D.; Rappuoli, R. Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery. *Front. Immunol.* **2019**, *10*, 1−12.
(112) Dimitrov, I.; Zaharieva, N.; Doytchinova, I. Bacterial Immunogenicity Prediction by Machine Learning Methods. *Vaccines* **2020**, *8*, 1−10.
(113) Ong, E.; Cooke, M. F.; Huffman, A.; Xiang, Z.; Wong, M. U.; Wang, H.; Seetharaman, M.; Valdez, N.; He, Y. Vaxign2: The Second Generation of the First Web-Based Vaccine Design Program Using Reverse Vaccinology and Machine Learning. *Nucleic Acids Res.* **2021**, *49*, W671−W678.
(114) Fleri, W.; Paul, S.; Dhanda, S. K.; Mahajan, S.; Xu, X.; Fleri, W. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Front. Immunol.* **2017**, *8*, 278.
(115) Haste Andersen, P.; Nielsen, M.; Lund, O. Prediction of Residues in Discontinuous B-Cell Epitopes Using Protein 3D Structures. *Protein Sci.* **2006**, *15*, 2558−2567.
(116) Cia, G.; Pucci, F.; Rooman, M. Critical Review of Conformational B-Cell Epitope Prediction Methods. *Brief. Bioinform.* **2023**, *24*, 1−9.
(117) Kunik, V.; Ofran, Y. The Indistinguishability of Epitopes from Protein Surface Is Explained by the Distinct Binding Preferences of Each of the Six Antigen-Binding Loops. *Protein Eng. Des. Sel.* **2013**, *26*, 599−609.
(118) Kringelum, J. V.; Nielsen, M.; Padkjær, S. B.; Lund, O. Structural Analysis of B-Cell Epitopes in Antibody:Protein Complexes. *Mol. Immunol.* **2013**, *53*, 24−34.
(119) Jespersen, M. C.; Mahajan, S.; Peters, B.; Nielsen, M.; Marcatili, P. Antibody Specific B-Cell Epitope Predictions: Leveraging Information from Antibody-Antigen Protein Complexes. *Front. Immunol.* **2019**, *10*, 1−10.
(120) Krawczyk, K.; Liu, X.; Baker, T.; Shi, J.; Deane, C. M. Improving B-Cell Epitope Prediction and Its Application to Global Antibody-Antigen Docking. *Bioinformatics* **2014**, *30*, 2288−2294.
(121) Jespersen, M. C.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes. *Nucleic Acids Res.* **2017**, *45*, W24−W29.
(122) Ansari, H. R.; Raghava, G. P. Identification of Conformational B-Cell Epitopes in an Antigen from Its Primary Sequence. *Immunome Res.* **2010**, *6*, 6.
(123) Zhou, C.; Chen, Z.; Zhang, L.; Yan, D.; Mao, T.; Tang, K.; Qiu, T.; Cao, Z. SEPPA 3.0-Enhanced Spatial Epitope Prediction Enabling Glycoprotein Antigens. *Nucleic Acids Res.* **2019**, *47*, W388−W394.
(124) Kringelum, J. V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. *PLOS Comput. Biol.* **2012**, *8*, No. e1002829.
(125) Ponomarenko, J.; Bui, H.-H.; Li, W.; Fusseder, N.; Bourne, P. E.; Sette, A.; Peters, B. ElliPro: A New Structure-Based Tool for the Prediction of Antibody Epitopes. *BMC Bioinformatics* **2008**, *9*, 514.
(126) Liang, S.; Zheng, D.; Standley, D. M.; Yao, B.; Zacharias, M.; Zhang, C. EPSVR and EPMeta: Prediction of Antigenic Epitopes Using Support Vector Regression and Multiple Server Results. *BMC Bioinformatics* **2010**, *11*, 381.
(127) Sweredoski, M. J.; Baldi, P. PEPITO: Improved Discontinuous B-Cell Epitope Prediction Using Multiple Distance Thresholds and Half Sphere Exposure. *Bioinformatics* **2008**, *24*, 1459−1460.
(128) Da Silva, B. M.; Myung, Y.; Ascher, D. B.; Pires, D. E. V. Epitope3D: A Machine Learning Method for Conformational B-Cell Epitope Prediction. *Brief. Bioinform.* **2022**, *23*, 1−8.
(129) Clifford, J. N.; Høie, M. H.; Deleuran, S.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-3.0: Improved B-Cell Epitope Prediction Using Protein Language Models. *Protein Sci.* **2022**, *31*, 1−11.
(130) Park, M.; Seo, S.; Park, E.; Kim, J. EpiBERTope: A Sequence-Based Pre-Trained BERT Model Improves Linear and Structural Epitope Prediction by Learning Long-Distance Protein Interactions Effectively. *bioRxiv* **2022**, 2022.02.27.481241.
(131) Peri, C.; Solé, O. C.; Corrada, D.; Gori, A.; Daura, X.; Colombo, G. Prediction of Antigenic B and T Cell Epitopes via Energy Decomposition Analysis: Description of the Web-Based Prediction Tool BEPPE. *Methods Mol. Biol.* **2015**, *1348*, 13−22.
(132) Scarabelli, G.; Morra, G.; Colombo, G. Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping. *Biophys. J.* **2010**, *98*, 1966−1975.
(133) Morra, G.; Colombo, G. Relationship between Energy Distribution and Fold Stability: Insights from Molecular Dynamics Simulations of Native and Mutant Proteins. *Proteins Struct. Funct. Genet.* **2008**, *72*, 660−672.
(134) Capelli, R.; Serapian, S. A.; Colombo, G. Computational Epitope Prediction and Design for Antibody Development and Detection. *Methods Mol. Biol.* **2023**, *2552*, 255−266.
(135) Peri, C.; Gagni, P.; Combi, F.; Gori, A.; Chiari, M.; Longhi, R.; Cretich, M.; Colombo, G. Rational Epitope Design for Protein Targeting. *ACS Chem. Biol.* **2013**, *8*, 397−404.
(136) Gourlay, L. J.; Peri, C.; Ferrer-Navarro, M.; Conchillo-Solé, O.; Gori, A.; Rinchai, D.; Thomas, R. J.; Champion, O. L.; Michell, S. L.; Kewcharoenwong, C.; Nithichanon, A.; Lassaux, P.; Perletti, L.; Longhi, R.; Lertmemongkolchai, G.; Titball, R. W.; Daura, X.; Colombo, G.; Bolognesi, M. Exploiting the Burkholderia Pseudomallei Acute Phase Antigen BPSL2765 for Structure-Based Epitope Discovery/Design in Structural Vaccinology. *Chem. Biol.* **2013**, *20*, 1147−1156.
(137) Lassaux, P.; Peri, C.; Ferrer-Navarro, M.; Gourlay, L. J.; Gori, A.; Conchillo-Solé, O.; Rinchai, D.; Lertmemongkolchai, G.; Longhi, R.; Daura, X.; Colombo, G.; Bolognesi, M. A Structure-Based Strategy for Epitope Discovery in Burkholderia Pseudomallei OppA Antigen. *Structure* **2013**, *21*, 167−175.
(138) Capelli, R.; Matterazzo, E.; Amabili, M.; Peri, C.; Gori, A.; Gagni, P.; Chiari, M.; Lertmemongkolchai, G.; Cretich, M.; Bolognesi, M.; Colombo, G.; Gourlay, L. J. Designing Probes for Immunodiagnostics: Structural Insights into an Epitope Targeting Burkholderia Infections. *ACS Infect. Dis.* **2017**, *3*, 736−743.
(139) Genoni, A.; Morra, G.; Colombo, G. Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J. Phys. Chem. B* **2012**, *116*, 3331−3343.

(140) Gourlay, L. J.; Colombo, G.; Soriani, M.; Grandi, G.; Daura, X.; Bolognesi, M. Why Is a Protective Antigen Protective? *Hum. Vaccin.* **2009**, *5*, 872−875.
(141) Fiorucci, S.; Zacharias, M. Prediction of Protein-Protein Interaction Sites Using Electrostatic Desolvation Profiles. *Biophys. J.* **2010**, *98*, 1921−1930.
(142) Fischer, M. F. S.; Crowe, J. E.; Meiler, J. Computational Epitope Mapping of Class I Fusion Proteins Using Low Complexity Supervised Learning Methods. *PLoS Comput. Biol.* **2022**, *18*, 1−18.
(143) Bell, E. W.; Schwartz, J. H.; Freddolino, P. L.; Zhang, Y. PEPPI: Whole-Proteome Protein-Protein Interaction Prediction through Structure and Sequence Similarity, Functional Association, and Machine Learning. *J. Mol. Biol.* **2022**, *434*, 167530.
(144) Ott, P. A.; Hu, Z.; Keskin, D. B.; Shukla, S. A.; Sun, J.; Bozym, D. J.; Zhang, W.; Luoma, A.; Giobbie-Hurder, A.; Peter, L.; Chen, C.; Olive, O.; Carter, T. A.; Li, S.; Lieb, D. J.; Eisenhaure, T.; Gjini, E.; Stevens, J.; Lane, W. J.; Javeri, I.; Nellaiappan, K.; Salazar, A. M.; Daley, H.; Seaman, M.; Buchbinder, E. I.; Yoon, C. H.; Harden, M.; Lennon, N.; Gabriel, S.; Rodig, S. J.; Barouch, D. H.; Aster, J. C.; Getz, G.; Wucherpfennig, K.; Neuberg, D.; Ritz, J.; Lander, E. S.; Fritsch, E. F.; Hacohen, N.; Wu, C. J. An Immunogenic Personal Neoantigen Vaccine for Patients with Melanoma. *Nature* **2017**, *547*, 217−221.
(145) Wang, J.; Mamuti, M.; Wang, H. Therapeutic Vaccines for Cancer Immunotherapy. *ACS Biomater. Sci. Eng.* **2020**, *6*, 6036−6052.
(146) Chen, S.; Xu, X.; Zhang, Y.; Ye, L.; Zhang, L.; Li, L.; Chen, G. Nanovaccine Based on a Biepitope Antigen to Potentiate the Immunogenicity of a Neoantigen. *ACS Macro Lett.* **2023**, *12*, 281−287.
(147) Poorebrahim, M.; Mohammadkhani, N.; Mahmoudi, R.; Gholizadeh, M.; Fakhr, E.; Cid-Arregui, A. TCR-like CARs and TCR-CARs Targeting Neoepitopes: An Emerging Potential. *Cancer Gene Ther.* **2021**, *28*, 581−589.
(148) June, C. H.; O'Connor, R. S.; Kawalekar, O. U.; Ghassemi, S.; Milone, M. C. CAR T Cell Immunotherapy for Human Cancer. *Science* **2018**, *359*, 1361−1365.
(149) Xie, N.; Shen, G.; Gao, W.; Huang, Z.; Huang, C.; Fu, L. Neoantigens: Promising Targets for Cancer Therapy. *Signal Transduct. Target. Ther.* **2023**, *8*, 9.
(150) Tran, N. H.; Xu, J.; Li, M. A Tale of Solving Two Computational Challenges in Protein Science: Neoantigen Prediction and Protein Structure Prediction. *Brief. Bioinform.* **2022**, *23*, 1−13.
(151) Shi, Y.; Guo, Z.; Su, X.; Meng, L.; Zhang, M.; Sun, J.; Wu, C.; Zheng, M.; Shang, X.; Zou, X.; Cheng, W.; Yu, Y.; Cai, Y.; Zhang, C.; Cai, W.; Da, L. T.; He, G.; Han, Z. G. DeepAntigen: A Novel Method for Neoantigen Prioritization via 3D Genome and Deep Sparse Learning. *Bioinformatics* **2020**, *36*, 4894−4901.
(152) Rammensee, H.; Bachmann, J.; Emmerich, N. P.; Bachor, O. A.; Stevanović, S. SYFPEITHI: Database for MHC Ligands and Peptide Motifs. *Immunogenetics* **1999**, *50*, 213−219.
(153) Borrman, T.; Cimons, J.; Cosiano, M.; Purcaro, M.; Pierce, B. G.; Baker, B. M.; Weng, Z. ATLAS: A Database Linking Binding Affinities with Structures for Wild-Type and Mutant TCR-PMHC Complexes. *Proteins Struct. Funct. Bioinforma.* **2017**, *85*, 908−916.
(154) Koşaloğlu-Yalçın, Z.; Blazeska, N.; Vita, R.; Carter, H.; Nielsen, M.; Schoenberger, S.; Sette, A.; Peters, B. The Cancer Epitope Database and Analysis Resource (CEDAR). *Nucleic Acids Res.* **2023**, *51*, D845−D852.
(155) Sanchez-Trincado, J. L.; Gomez-Perosanz, M.; Reche, P. A. Fundamentals and Methods for T- and B-Cell Epitope Prediction. *J. Immunol. Res.* **2017**, *2017*, 1.
(156) Peters, B.; Nielsen, M.; Sette, A. T Cell Epitope Predictions. *Annu. Rev. Immunol.* **2020**, *38*, 123−145.
(157) Parker, K. C.; Bednarek, M. A.; Coligan, J. E. Scheme for Ranking Potential HLA-A2 Binding Peptides Based on Independent Binding of Individual Peptide Side-Chains. *J. Immunol.* **1994**, *152*, 163−175.
(158) Nielsen, M.; Lund, O. NN-Align. An Artificial Neural Network-Based Alignment Algorithm for MHC Class II Peptide Binding Prediction. *BMC Bioinformatics* **2009**, *10*, 296.
(159) Jandrlić, D. R. SVM and SVR-Based MHC-Binding Prediction Using a Mathematical Presentation of Peptide Sequences. *Comput. Biol. Chem.* **2016**, *65*, 117−127.
(160) Peters, B.; Tong, W.; Sidney, J.; Sette, A.; Weng, Z. Examining the Independent Binding Assumption for Binding of Peptide Epitopes to MHC-I Molecules. *Bioinformatics* **2003**, *19*, 1765−1772.
(161) Zhang, H.; Wang, P.; Papangelopoulos, N.; Xu, Y.; Sette, A.; Bourne, P. E.; Lund, O.; Ponomarenko, J.; Nielsen, M.; Peters, B. Limitations of Ab Initio Predictions of Peptide Binding to MHC Class II Molecules. *PLoS One* **2010**, *5*, No. e9272.
(162) Nielsen, M.; Lundegaard, C.; Blicher, T.; Lamberth, K.; Harndahl, M.; Justesen, S.; Røder, G.; Peters, B.; Sette, A.; Lund, O.; Buus, S. NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS One* **2007**, *2*, No. e796.
(163) Jojic, N.; Reyes-Gomez, M.; Heckerman, D.; Kadie, C.; Schueler-Furman, O. Learning MHC I—Peptide Binding. *Bioinformatics* **2006**, *22*, e227−e235.
(164) Zhang, L.; Udaka, K.; Mamitsuka, H.; Zhu, S. Toward More Accurate Pan-Specific MHC-Peptide Binding Prediction: A Review of Current Methods and Tools. *Brief. Bioinform.* **2012**, *13*, 350−364.
(165) Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* **2020**, *48*, W449−W454.
(166) Reche, P. A.; Reinherz, E. L. PEPVAC: A Web Server for Multi-Epitope Vaccine Development Based on the Prediction of Supertypic MHC Ligands. *Nucleic Acids Res.* **2005**, *33*, W138−W142.
(167) Zhang, G. L.; DeLuca, D. S.; Keskin, D. B.; Chitkushev, L.; Zlateva, T.; Lund, O.; Reinherz, E. L.; Brusic, V. MULTIPRED2: A Computational System for Large-Scale Identification of Peptides Predicted to Bind to HLA Supertypes and Alleles. *J. Immunol. Methods* **2011**, *374*, 53−61.
(168) Xia, H.; McMichael, J.; Becker-Hapak, M.; Onyeador, O. C.; Buchli, R.; McClain, E.; Pence, P.; Supabphol, S.; Richters, M. M.; Basu, A.; Ramirez, C. A.; Puig-Saus, C.; Cotto, K. C.; Freshour, S. L.; Hundal, J.; Kiwala, S.; Goedegebuure, S. P.; Johanns, T. M.; Dunn, G. P.; Ribas, A.; Miller, C. A.; Gillanders, W. E.; Fehniger, T. A.; Griffith, O. L.; Griffith, M. Computational Prediction of MHC Anchor Locations Guides Neoantigen Identification and Prioritization. *Sci. Immunol.* **2023**, *8*, No. eabg2200.
(169) Trolle, T.; Nielsen, M. NetTepi: An Integrated Method for the Prediction of T Cell Epitopes. *Immunogenetics* **2014**, *66*, 449−456.
(170) Calis, J. J. A.; Maybeno, M.; Greenbaum, J. A.; Weiskopf, D.; De Silva, A. D.; Sette, A.; Keşmir, C.; Peters, B. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLOS Comput. Biol.* **2013**, *9*, No. e1003266.
(171) Schaap-Johansen, A.-L.; Vujovic, M.; Borch, A.; Hadrup, S. R.; Marcatili, P. T Cell Epitope Prediction and Its Application to Immunotherapy. *Front. Immunol.* **2021**, *12*, 1−11.
(172) Blevins, S. J.; Pierce, B. G.; Singh, N. K.; Riley, T. P.; Wang, Y.; Spear, T. T.; Nishimura, M. I.; Weng, Z.; Baker, B. M. How Structural Adaptability Exists alongside HLA-A2 Bias in the Human Aβ TCR Repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E1276−E1285.
(173) Toor, J. S.; Rao, A. A.; McShan, A. C.; Yarmarkovich, M.; Nerli, S.; Yamaguchi, K.; Madejska, A. A.; Nguyen, S.; Tripathi, S.; Maris, J. M.; Salama, S. R.; Haussler, D.; Sgourakis, N. G. A Recurrent Mutation in Anaplastic Lymphoma Kinase with Distinct Neoepitope Conformations. *Front. Immunol.* **2018**, *9*. DOI: 10.3389/fimmu.2018.00099.
(174) Pang, Y.-P.; Elsbernd, L. R.; Block, M. S.; Markovic, S. N. Peptide-Binding Groove Contraction Linked to the Lack of T Cell Response: Using Complex Structure and Energy To Identify Neoantigens. *ImmunoHorizons* **2018**, *2*, 216−225.

(175) Riley, T. P.; Keller, G. L. J.; Smith, A. R.; Davancaze, L. M.; Arbuiso, A. G.; Devlin, J. R.; Baker, B. M. Structure Based Prediction of Neoantigen Immunogenicity. *Front. Immunol.* **2019**, *10*. DOI: 10.3389/fimmu.2019.02047.
(176) Gowthaman, R.; Pierce, B. G. TCRmodel: High Resolution Modeling of T Cell Receptors from Sequence. *Nucleic Acids Res.* **2018**, *46*, W396−W401.
(177) Leimgruber, A.; Ferber, M.; Irving, M.; Hussain-Kahn, H.; Wieckowski, S.; Derré, L.; Rufer, N.; Zoete, V.; Michielin, O. TCRep 3D: An Automated In Silico Approach to Study the Structural Properties of TCR Repertoires. *PLoS One* **2011**, *6*, No. e26301.
(178) Klausen, M. S.; Anderson, M. V.; Jespersen, M. C.; Nielsen, M.; Marcatili, P. LYRA, a Webserver for Lymphocyte Receptor Structural Modeling. *Nucleic Acids Res.* **2015**, *43*, W349−W355.
(179) Lanzarotti, E.; Marcatili, P.; Nielsen, M. T-Cell Receptor Cognate Target Prediction Based on Paired α and β Chain Sequence and Structural CDR Loop Similarities. *Front. Immunol.* **2019**, *10*, 1−10.
(180) Lanzarotti, E.; Marcatili, P.; Nielsen, M. Identification of the Cognate Peptide-MHC Target of T Cell Receptors Using Molecular Modeling and Force Field Scoring. *Mol. Immunol.* **2018**, *94*, 91−97.
(181) Bradley, P. Structure-Based Prediction of T Cell Receptor:-Peptide-MHC Interactions. *Elife* **2023**, *12*, No. e82813.
(182) Moris, P.; De Pauw, J.; Postovskaya, A.; Gielis, S.; De Neuter, N.; Bittremieux, W.; Ogunjimi, B.; Laukens, K.; Meysman, P. Current Challenges for Unseen-Epitope TCR Interaction Prediction and a New Perspective Derived from Image Classification. *Brief. Bioinformatics* **2021**, *22*, 1−12.
(183) Coles, C. H.; McMurran, C.; Lloyd, A.; Hock, M.; Hibbert, L.; Raman, M. C. C.; Hayes, C.; Lupardus, P.; Cole, D. K.; Harper, S. T Cell Receptor Interactions with Human Leukocyte Antigen Govern Indirect Peptide Selectivity for the Cancer Testis Antigen MAGE-A4. *J. Biol. Chem.* **2020**, *295*, 11486−11494.
(184) Caradonna, T. M.; Schmidt, A. G. Protein Engineering Strategies for Rational Immunogen Design. *npj Vaccines* **2021**, *6*, 1−11.
(185) Castro, K. M.; Scheck, A.; Xiao, S.; Correia, B. E. Computational Design of Vaccine Immunogens. *Curr. Opin. Biotechnol.* **2022**, *78*, 102821.
(186) Capelli, R.; Marchetti, F.; Tiana, G.; Colombo, G. SAGE: A Fast Computational Tool for Linear Epitope Grafting onto a Foreign Protein Scaffold. *J. Chem. Inf. Model.* **2017**, *57*, 6−10.
(187) Bonet, J.; Wehrle, S.; Schriever, K.; Yang, C.; Billet, A.; Sesterhenn, F.; Scheck, A.; Sverrisson, F.; Veselkova, B.; Vollers, S.; Lourman, R.; Villard, M.; Rosset, S.; Krey, T.; Correia, B. E. Rosetta FunFolDes − A General Framework for the Computational Design of Functional Proteins. *PLoS Comput. Biol.* **2018**, *14*, e1006623.
(188) Correia, B. E.; Bates, J. T.; Loomis, R. J.; Baneyx, G.; Carrico, C.; Jardine, J. G.; Rupert, P.; Correnti, C.; Kalyuzhniy, O.; Vittal, V.; Connell, M. J.; Stevens, E.; Schroeter, A.; Chen, M.; MacPherson, S.; Serra, A. M.; Adachi, Y.; Holmes, M. A.; Li, Y.; Klevit, R. E.; Graham, B. S.; Wyatt, R. T.; Baker, D.; Strong, R. K.; Crowe, J. E.; Johnson, P. R.; Schief, W. R. Proof of Principle for Epitope-Focused Vaccine Design. *Nature* **2014**, *507*, 201−206.
(189) Sesterhenn, F.; Yang, C.; Bonet, J.; Cramer, J. T.; Wen, X.; Wang, Y.; Chiang, C. I.; Abriata, L. A.; Kucharska, I.; Castoro, G.; Vollers, S. S.; Galloux, M.; Dheilly, E.; Rosset, S.; Corthésy, P.; Georgeon, S.; Villard, M.; Richard, C. A.; Descamps, D.; Delgado, T.; Oricchio, E.; Rameix-Welti, M. A.; Más, V.; Ervin, S.; Eléouët, J. F.; Riffault, S.; Bates, J. T.; Julien, J. P.; Li, Y.; Jardetzky, T.; Krey, T.; Correia, B. E. De Novo Protein Design Enables the Precise Induction of RSV-Neutralizing Antibodies. *Science* **2020**, *368* (6492), No. eaay5051, DOI: 10.1126/science.aay5051.
(190) Scheck, A.; Rosset, S.; Defferrard, M.; Loukas, A.; Bonet, J.; Vandergheynst, P.; Correia, B. E. RosettaSurf—A Surface-Centric Computational Design Approach. *PLoS Comput. Biol.* **2022**, *18*, e1009178.
(191) Brinkkemper, M.; Sliepen, K. Nanoparticle Vaccines for Inducing HIV-1 Neutralizing Antibodies. *Vaccines* **2019**, *7*, 76.
(192) Nguyen, B.; Tolia, N. H. Protein-Based Antigen Presentation Platforms for Nanoparticle Vaccines. *npj Vaccines* **2021**, *6*, 70.
(193) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012**, *336*, 1171−1174.
(194) Kang, Y.-F.; Sun, C.; Sun, J.; Xie, C.; Zhuang, Z.; Xu, H.-Q.; Liu, Z.; Liu, Y.-H.; Peng, S.; Yuan, R.-Y.; Zhao, J.-C.; Zeng, M.-S. Quadrivalent Mosaic HexaPro-Bearing Nanoparticle Vaccine Protects against Infection of SARS-CoV-2 Variants. *Nat. Commun.* **2022**, *13*, 2674.
(195) Boyoglu-Barnum, S.; Ellis, D.; Gillespie, R. A.; Hutchinson, G. B.; Park, Y.-J.; Moin, S. M.; Acton, O. J.; Ravichandran, R.; Murphy, M.; Pettie, D.; Matheson, N.; Carter, L.; Creanga, A.; Watson, M. J.; Kephart, S.; Ataca, S.; Vaile, J. R.; Ueda, G.; Crank, M. C.; Stewart, L.; Lee, K. K.; Guttman, M.; Baker, D.; Mascola, J. R.; Veesler, D.; Graham, B. S.; King, N. P.; Kanekiyo, M. Quadrivalent Influenza Nanoparticle Vaccines Induce Broad Protection. *Nature* **2021**, *592*, 623−628.
(196) Hsieh, C.-L.; McLellan, J. S. Protein Engineering Responses to the COVID-19 Pandemic. *Curr. Opin. Struct. Biol.* **2022**, *74*, 102385.
(197) Castro, K. M.; Scheck, A.; Xiao, S.; Correia, B. E. Computational Design of Vaccine Immunogens. *Curr. Opin. Biotechnol.* **2022**, *78*, 102821.
(198) Olshefsky, A.; Richardson, C.; Pun, S. H.; King, N. P. Engineering Self-Assembling Protein Nanoparticles for Therapeutic Delivery. *Bioconjugate Chem.* **2022**, *33*, 2018−2034.
(199) Joyce, M. G.; Zhang, B.; Ou, L.; Chen, M.; Chuang, G.-Y.; Druz, A.; Kong, W.-P.; Lai, Y.-T.; Rundlet, E. J.; Tsybovsky, Y.; Yang, Y.; Georgiev, I. S.; Guttman, M.; Lees, C. R.; Pancera, M.; Sastry, M.; Soto, C.; Stewart-Jones, G. B. E.; Thomas, P. V.; Van Galen, J. G.; Baxa, U.; Lee, K. K.; Mascola, J. R.; Graham, B. S.; Kwong, P. D. Iterative Structure-Based Improvement of a Fusion-Glycoprotein Vaccine against RSV. *Nat. Struct. Mol. Biol.* **2016**, *23*, 811−820.
(200) Hsieh, C.; Goldsmith, J. A.; Schaub, J. M.; Divenere, A. M.; Kuo, H.; Javanmardi, K.; Le, K. C.; Wrapp, D.; Lee, A. G.; Liu, Y.; Chou, C.; Byrne, P. O.; Hjorth, C. K.; Johnson, N. V.; Ludes-meyers, J.; Nguyen, A. W.; Park, J.; Wang, N.; Amengor, D.; Lavinder, J. J.; Ippolito, G. C.; Maynard, J. A.; Finkelstein, I. J.; Mclellan, J. S. Structure-Based Design of Prefusion-Stabilized SARS-CoV-2 Spikes. *Science* **2020**, *369*, 1501−1505.
(201) Sanders, R. W.; Derking, R.; Cupo, A.; Julien, J.-P.; Yasmeen, A.; de Val, N.; Kim, H. J.; Blattner, C.; de la Peña, A. T.; Korzun, J.; Golabek, M.; de los Reyes, K.; Ketas, T. J.; van Gils, M. J.; King, C. R.; Wilson, I. A.; Ward, A. B.; Klasse, P. J.; Moore, J. P. A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 Gp140, Expresses Multiple Epitopes for Broadly Neutralizing but Not Non-Neutralizing Antibodies. *PLOS Pathog.* **2013**, *9*, No. e1003618.
(202) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63*, 337−346.
(203) Kotecha, A.; Seago, J.; Scott, K.; Burman, A.; Loureiro, S.; Ren, J.; Porta, C.; Ginn, H. M.; Jackson, T.; Perez-Martin, E.; Siebert, C. A.; Paul, G.; Huiskonen, J. T.; Jones, I. M.; Esnouf, R. M.; Fry, E. E.; Maree, F. F.; Charleston, B.; Stuart, D. I. Structure-Based Energetics of Protein Interfaces Guides Foot-and-Mouth Disease Virus Vaccine Design. *Nat. Struct. Mol. Biol.* **2015**, *22*, 788−794.
(204) McLellan, J. S.; Chen, M.; Leung, S.; Graepel, K. W.; Du, X.; Yang, Y.; Zhou, T.; Baxa, U.; Yasuda, E.; Beaumont, T.; Kumar, A.; Modjarrad, K.; Zheng, Z.; Zhao, M.; Xia, N.; Kwong, P. D.; Graham, B. S. Structure of RSV Fusion Glycoprotein Trimer Bound to a Prefusion-Specific Neutralizing Antibody. *Science* **2013**, *340*, 1113−1117.
(205) Kvansakul, M.; Wei, A. H.; Fletcher, J. I.; Willis, S. N.; Chen, L.; Roberts, A. W.; Huang, D. C. S.; Colman, P. M. Structural Basis

for Apoptosis Inhibition by Epstein-Barr Virus BHRF1. *PLOS Pathog.* **2010**, *6*, No. e1001236.
(206) Correia, B. E.; Ban, Y.-E. A.; Holmes, M. A.; Xu, H.; Ellingson, K.; Kraft, Z.; Carrico, C.; Boni, E.; Sather, D. N.; Zenobia, C.; Burke, K. Y.; Bradley-Hewitt, T.; Bruhn-Johannsen, J. F.; Kalyuzhniy, O.; Baker, D.; Strong, R. K.; Stamatatos, L.; Schief, W. R. Computational Design of Epitope-Scaffolds Allows Induction of Antibodies Specific for a Poorly Immunogenic HIV Vaccine Epitope. *Structure* **2010**, *18*, 1116−1126.
(207) Sikora, M.; von Bülow, S.; Blanc, F. E. C.; Gecht, M.; Covino, R.; Hummer, G. Computational Epitope Map of SARS-CoV-2 Spike Protein. *PLOS Comput. Biol.* **2021**, *17*, No. e1008790.
(208) Casalino, L.; Seitz, C.; Lederhofer, J.; Tsybovsky, Y.; Wilson, I. A.; Kanekiyo, M.; Amaro, R. E. Breathing and Tilting: Mesoscale Simulations Illuminate Influenza Glycoprotein Vulnerabilities. *ACS Cent. Sci.* **2022**, *8*, 1646−1663.
(209) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning−Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378*, 49−56.
(210) Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; DiMaio, F.; De Munck, S.; Savvides, S. N.; Baker, D. Improving de Novo Protein Binder Design with Deep Learning. *Nat. Commun.* **2023**, *14*, 2625.

https://doi.org/10.1021/acs.jctc.3c00513
J. Chem. Theory Comput. 2023, 19, 5315−5333