




Review

# Leveraging Artificial Intelligence to Expedite Antibody Design and Enhance Antibody–Antigen Interactions

Doo Nam Kim , Andrew D. McNaughton  and Neeraj Kumar \* 

Pacific Northwest National Laboratory, 902 Battelle Blvd., Richland, WA 99352, USA; doonam.kim@pnnl.gov (D.N.K.); andrew.mcnaughton@pnnl.gov (A.D.M.)

\* Correspondence: neeraj.kumar@pnnl.gov; Tel.: +1-509-372-6422

**Abstract:** This perspective sheds light on the transformative impact of recent computational advancements in the field of protein therapeutics, with a particular focus on the design and development of antibodies. Cutting-edge computational methods have revolutionized our understanding of protein–protein interactions (PPIs), enhancing the efficacy of protein therapeutics in preclinical and clinical settings. Central to these advancements is the application of machine learning and deep learning, which offers unprecedented insights into the intricate mechanisms of PPIs and facilitates precise control over protein functions. Despite these advancements, the complex structural nuances of antibodies pose ongoing challenges in their design and optimization. Our review provides a comprehensive exploration of the latest deep learning approaches, including language models and diffusion techniques, and their role in surmounting these challenges. We also present a critical analysis of these methods, offering insights to drive further progress in this rapidly evolving field. The paper includes practical recommendations for the application of these computational techniques, supplemented with independent benchmark studies. These studies focus on key performance metrics such as accuracy and the ease of program execution, providing a valuable resource for researchers engaged in antibody design and development. Through this detailed perspective, we aim to contribute to the advancement of antibody design, equipping researchers with the tools and knowledge to navigate the complexities of this field.

**Keywords:** antibody; artificial intelligence; computer-aided drug discovery; computational modeling and simulations; deep learning; protein–protein interface; Rosetta; therapeutic design



**Citation:** Kim, D.N.; McNaughton, A.D.; Kumar, N. Leveraging Artificial Intelligence to Expedite Antibody Design and Enhance Antibody–Antigen Interactions. *Bioengineering* **2024**, *11*, 185. <https://doi.org/10.3390/bioengineering11020185>

Academic Editors: Yunfeng Wu, Jian Wu and Hongxia Xu

Received: 30 December 2023

Revised: 30 January 2024

Accepted: 6 February 2024

Published: 15 February 2024

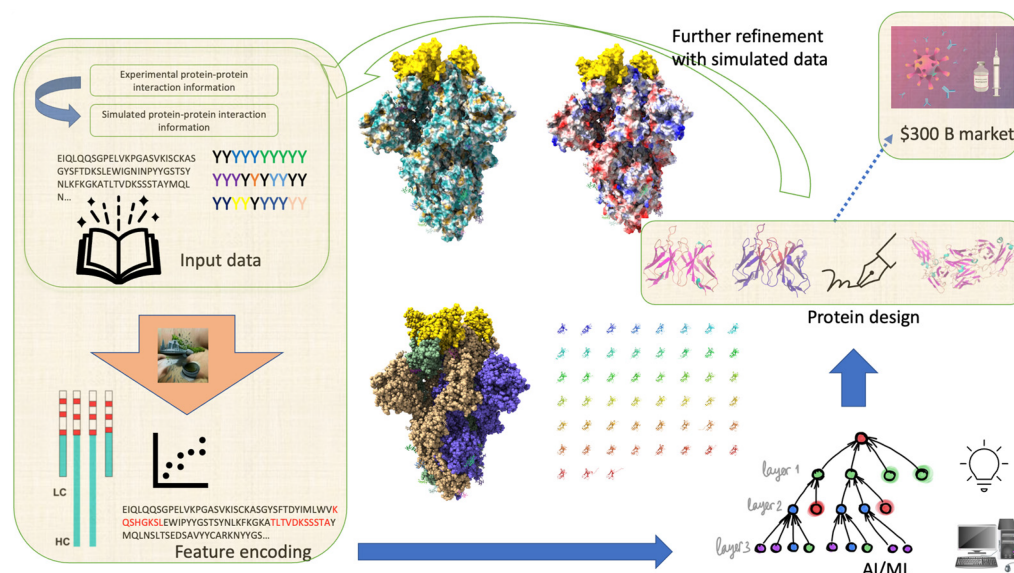


**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Living organisms can contain foreign substances, such as viruses or toxins, which are known as antigens (Ags). The presence of Ags triggers immune responses in the body, including the production of antibodies (Abs). The interaction between Abs and Ags has become increasingly important due to the expanding use of Ab-based therapeutics and diagnostics. With over 100 monoclonal Abs (mAbs) approved by the US Food and Drug Administration (FDA) [1,2], these approaches tend to exhibit lower toxicity and higher specificity control compared to traditional small molecule-based therapeutics. Consequently, the global therapeutic mAb market is anticipated to reach USD 300 billion by 2025 [3]. However, optimizing Abs experimentally is a laborious process because of the low-throughput nature of Ab screening in mammalian cells. For instance, yeast and phage display only cover between  $10^6$  and  $10^{10}$  Ab sequences [4]; while most Ab sequence screenings in industry often exceed  $10^{11}$ . Therefore, significant progress has been made in the application of deep learning (DL) towards Ab discovery, as reviewed in multiple studies [4–8]. These advancements encompass the modeling and prediction of Ab–target binding patterns for the identification of binding sequences [9], paratope prediction [10], complementarity-determining region (CDR) loop structure prediction [11], and target specificity. In this work, we present a review to discuss recent developments in

DL approaches for Ab sequence design (Figure 1). A promising trend in the field of DL is the widespread practice of researchers publicly sharing their code and publicly shared Ab affinity datasets [12,13]. This collaborative approach, apart from a few commercial entities, has facilitated synergy among disciplines. It is hoped that these collaborative efforts will continue to strengthen and eventually reach the level of successful collaborative groups such as the *Rosetta* community [14].



**Figure 1.** Generalized schematic of DL approaches for Ab sequence design. The three large middle molecules are SARS Corona virus 2 (SARS-CoV-2) spikes (PDB ID: 7kkl). Three bound nanobody (Nb) molecules are depicted with yellow caps. Hydrophobicity potential ranges from cyan for hydrophilic through white to goldenrod for hydrophobic. Electrostatic potential ranges from red for negative potential through white to blue for positive potential. The rainbow-colored array of molecules is Nb, a single-chain Ab.

### 1.1. Historical Perspective and Rise of Deep Learning in the Biomedical Field

DL serves as a critical universal approximator capable of generalizing complex, non-linear phenomena [15]. As a result, DL has been actively implemented in various fields related to protein design and small-molecule drug design. These advancements in DL approaches include the study of non-coding RNA–protein interactions [16], compound–protein interactions [17], the annotation of protein space [18] and gene ontology (GO) [19], and three-dimensional (3D) coordinates of drug-like molecules [20]. Likewise, DL methods facilitate Ab development. For example, a combination of a convolutional neural network (CNN) and a recurrent neural network (RNN) including a conditional random field [21] was used to predict the signal peptide cleavage site of recombinant mAbs to reduce product heterogeneity issues [22]. Most importantly, DL approaches often need to be integrated with other physics-based modeling and simulation methods. For instance, many modern DL protein modeling methods, including *AlphaFold* version 1 [23], typically refine the final structure using *Rosetta's FastRelax* [24]. However, a more precise DL-based amino acid (AA) packer would be ideal. In this context, utilizing *AttnPacker* [25] appears to be the optimal choice, as it reduces the inference time by over 100× compared to other DL-based methods, such as *DLPacker* [26], and physics-based method *FastRelax*.

When predicting protein–protein interactions (PPIs) or PPI complex structures, a DL-based tool called *AF2Complex* [27] is currently known to produce the most accurate results according to the DockQ score [28]. We discovered that using *AlphaFold-Multimer* [29] was simpler in terms of execution and analysis (see Supplementary Note S1 for details), which provides not only accurate outcomes but also produces a readily understandable iPTM report. On the other hand, the most recent docking and design model specifically

tailored for Ab–Ag interactions is the Hierarchical Structure Refinement Network. It has improved Ab docking success rates by 50%, outperforming other sequence-based and structure-based models [30]. As demonstrated with the transformer architecture [31], attention-based networks are powerful methods to capture interactions between input data and often result in better accuracy [12]. Since it is believed that most epitopes are somewhat discontinuous [32], it is critical to understand recent Ab–Ag interface-based DL models [12,15,33] that have adopted attention-based architectures. Ideally, attention-based methods that consider both the sequence and structural context would be more suitable for sequence generation [34]. The superior accuracy of DL over non-DL methods is often evident, as seen in immune status classification based on immune repertoire sequences [35] and Fv structure prediction [12].

Enhancements in Ab affinity often result from combined AA substitutions rather than individual site mutations [13]. Surprisingly, many affinity-boosting mutations are found in areas that do not directly interact with the Ag. These mutations can lead to voids, misaligned polar AA, and spatial conflicts, making them difficult to predict [36]. Therefore, utilizing DL architectures for Ab design [12] might present advantages over traditional computational methods [37]. Despite this, we concur that DL-based PPI modeling methods may not always be the optimal choice, particularly due to the insufficient availability of training data or when training is not robust. For instance, Dequeker et al. demonstrated that their comprehensive cross-docking approach surpassed a sequence-based DL method [38].

### *1.2. Sequence-Based and Structure-Based Approaches with Implications for Antibody Design*

Ab design methodologies can be classified into sequence-based or structure-based methods, contingent upon the types of input and output data. In this review, we delve into each of these categories with relevant examples and evaluations. In this subsection, we provide a summary of each category. The development of Ab-based therapeutics typically spans several years, and enhancing therapeutic efficacy can result in substantial cost savings. In order to assess therapeutic efficacy at an early stage, a number of prediction methods based on DL have been developed to predict Ab affinity [13,15,39]. As we analyzed, it is clear that most of these DL-assisted Ab modeling approaches only require protein sequence data as input. This is an ideal approach because it allows for quick implementation of the model [40], and, most importantly, Ab sequence data can be produced on a much larger scale, significantly lowering costs compared to generating structural information about Abs [41]. Similar to the situation present in general protein data, there is a huge amount of sequence data available (for example, metagenomic sequences surpass 1.6 billion [42], while the Protein Data Bank (PDB) [43] only has 214,000 structures). Nonetheless, creating sequences without corresponding structures may result in suboptimal outcomes [34]. Alternatively, structure-based models can offer details of the models with structural features [12], like structural paratope and epitope information, along with additional interpretable physicochemical properties [6]. Furthermore, it is generally assumed that most antigenic determinants exhibit some degree of conformational (i.e., discontinuous) structure [32]. Hence, including Ab structural information, either as an input or output, can aid in identifying potential modes of action. Nevertheless, the required Ab structure is often not available. For instance, there are only around 2000 to 5000 unique Ab–Ag complex structures [44,45], while the number of Ab sequences is over  $10^{13}$ . Therefore, a hybrid model that utilizes both sequence and structural information is considered optimal for Ab design [34].

## **2. Antibody Design**

### *2.1. Role of Antibodies in Mediating Protein–Protein Interactions*

PPIs play a crucial role in various cellular responses and functions, making them essential targets for the development of biomarkers and pharmaceuticals. To save experimental resources, a variety of machine learning (ML) methods have been developed, primarily focused on predicting PPI sites or residues, and these methods have been ex-

tensively reviewed [46]. Among many PPI cases, Ab–Ag interactions are distinct because most cross-interface hydrogen bonds are generated between sidechains rather than between backbones [47]. Moreover, interfaces between Ab and Ag are likely to exhibit fewer hydrophobic interactions than those observed in typical PPIs [5]. The quantity of AAs involved in the design process also differs. For instance, processes such as VDJ recombination and somatic hypermutation expand immunoglobulin (Ig) diversity to an experimentally confirmed extent greater than  $10^{13}$  and could theoretically surpass  $10^{26}$  [48,49]. On the other hand, the diversity of non-immune protein ranges from  $10^5$  to  $10^6$  [41].

Many researchers proposed that the fundamental principles governing PPI prediction can be adapted to predict Ab–Ag interactions as well [50]. An example of this is the DL-based model for the one-sided design of general PPI interfaces, which was trained with general peptide ligands and their binding complexes, then applied to Ab–Ag interfaces [33]. The highly successful attention DL-based *Binding-ddg-predictor* [15] (Supplementary Note S2) redesigned the CDR to enhance Ab affinity (toward multiple virus variants) and was validated with *SKEMPI* (Structural database of Kinetics and Energetics of Mutant Protein Interactions) version 2 [51]. This kind of approach (training with PPI, then application to Abs) is theoretically amenable with DL, as training with general 20 million protein sequences can be fine-tuned on target sequences to be optimized [52]. Additionally, general hot-spot prediction methods [53] may be applicable to immunogenic regions [54].

## 2.2. Generative Modeling for Antibody Sequences

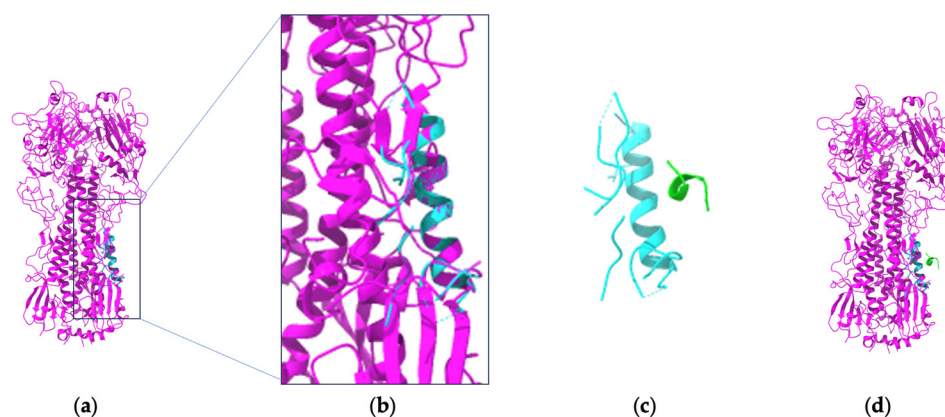
In this section, we discuss how generative modeling techniques are used to predict and optimize Ab sequences. Gated recurrent units (GRUs), long short-term memory (LSTM) networks, variational autoencoders (VAEs), generative adversarial networks (GANs) and language models (LMs) have been utilized as deep generative models in molecular design [55,56] and computational protein design [57–60]. Language model-based sequence generation can involve the use of either an autoregressive model (e.g., *ProtGPT* [61], *ProGen* [62]) or non-autoregressive language (e.g., *AntiBERTy*, *ESM*) [63]. Consequently, it is expected that these generative DL methods have also been applied to generate or design novel Ab sequences [64]. The designs of Ab sequences have primarily focused on the CDR region, as it determines binding specificity [34]. The design of protein (including Ab) sequences refers to either sidechain design with a fixed backbone [65] or concomitant design of both the backbone and sidechain [66]. We consider generative methods for Ab sequences to contain both cases, as Jin et al. asserted [34]. Generative methods for Ab sequences include the identification of tight-binding Abs using variational Bayesian neural networks [67]. Furthermore, an ensemble approach utilizing five CNNs, along with additional methods, such as a VAE, a genetic algorithm-augmented KNN, and a genetic algorithm-augmented CNN, was employed to generate novel CDR-H3 sequences [68].

### 2.2.1. Introduction to Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) Models

Both GRUs and LSTM models are variations of RNNs designed to tackle the vanishing and exploding gradient challenges inherent in traditional RNNs. For the generation of CDR-H3 sequences and synthetic Ab–Ag structures, Robert and Akbar utilized a GRU-based model [54]. Their methodology was predicated on the energetically optimal binding structures located within a 3D lattice, providing a copious volume of training data. It is worth noting that LSTM models have gained substantial popularity over GRUs within the broader DL community. A parallel trend is discernible in the field of Ab–Ag studies as well [2]. For example, LSTM memory models were used to generate novel Ab sequences for affinity maturation [39] and the classification of Ag specificity (e.g., predicting the probability of each sequence variant as either Ag binders or non-binders) [9]. Since these approaches require a massive amount of sequence data for prioritization, they utilized deep-sequenced libraries of therapeutic Abs. An LSTM was also used to generate Ag-specific CDR-H3



sequences with respect to various developability parameters [69]. Humanization of Ab sequences has been an important goal to minimize the immunogenicity of mAbs derived from xenogeneic sources and to improve effectiveness in the human immune system [70]. Therefore, LSTM models have been used to distinguish natural human Ab sequences from those originating from other species [71]. Another LSTM application of protein sequence design is *iNNterfaceDesign* [33,72]. Specifically, it designs one-sided PPI interfaces based on features of the protein receptor. Here, we show an example that we could run (Figure 2, Supplementary Note S3). *iNNterfaceDesign*-validated benchmarks include Ab–Ag complexes. Using an LSTM model with attention, it designs binder sequences (from polyglycines) with two steps (e.g., construction of binding sites centered at anchor residues, extraction of features of the binding sites, and prediction of AA sequences). Just as *DeepAb* [12] does, *iNNterfaceDesign* [33] uses the *PyRosetta* package [73]. Like other *Rosetta*-based design research [74], *iNNterfaceDesign* uses a native sequence recovery rate as the success metric. It is very encouraging to observe that *iNNterfaceDesign* achieves better sequence recovery rates than *FastDesign* [75] (*RosettaDesign* [66] during *FastRelax* [24]). Given the long history of PPI design with *Rosetta* [76], it was an expected direction that *Rosetta*-based protein interface design would incorporate DL-based modeling.



**Figure 2.** An example of *iNNterfaceDesign* running with human H3 Influenza hemagglutinin. The *iNNterfaceDesign* treats the structures of protein as the 3D object to be captured. (a) Target (3ztj\_ABCDEF); (b) pocket in target; (c) pocket + binder; (d) target + binder. Target: pink; pocket: cyan; binder: green.

### 2.2.2. Introduction to Variational Autoencoders (VAEs)

VAE models have been widely employed in various unsupervised DL applications. Unlike GANs, VAEs enable visualization of the latent space, allowing for the representation of similar clusters together. For Ab modeling, VAEs have been utilized to model B-cell receptor (BCR) recombination. For example, Friedensohn et al. identified sequence patterns that are predictive of antigenic exposure by VAE [77]. Later, they experimentally confirmed their binding specificity to target Ags. Another example is to learn the rules of VDJ recombination [78]. A VAE is used also to directly generate the 3D coordinates of immunoglobulins with torsion and distance awareness [79]. As shown for cryo-EM-based 3D volume generation per class [80], VAEs also prove useful in latent space sampling in Ab design.

### 2.2.3. Application of Generative Adversarial Networks (GANs)

GANs have been used to model various properties, including images [81]. For Ab sequence design, GANs have been used to design mAbs, which retain typical human repertoire characteristics such as diversity and immunogenicity while biasing the libraries to achieve other biotherapeutic features. In particular, Just-Evotec Biologics used a Wasserstein-GAN (WGAN) with gradient penalty for this purpose [82]. To bias their GAN toward molecules with developability properties of interest, they utilized transfer learning. Like *DeepAb* and *BioPhi* [83], they used the Observed Antibody Space (OAS) database [84]

for training and testing. This database contains more than five hundred million human sequences from more than five hundred human subjects. For Abs, it contains more than 118,386 paired heavy- and light-chain sequences and unpaired sequences.

#### 2.2.4. Introduction to Autoregressive Method

Autoregressive methods in the DL field refers to models that use previous output as input, often for sequential data. One autoregressive generative method for Abs includes the co-design of sequences and 3D structures for CDRs [34], as we described in the Ab structural modeling section, since it iterates design along with updated sequence and structural information. Other examples include causal CNNs and transformers. Specifically, using a residual causal dilated CNN, Shin et al. generated millions of novel Nb sequences [60]. Recently, *BioPhi*, a platform for Ab humanization, was released [83]. It is constituted with *Sapiens*, a transformer-based Ab sequence humanization model, and *OASis*, a humanness evaluation program based on a 9-mer peptide search in the *OAS* database [84]. We found that *BioPhi* is a very user-friendly application. For example, it provides an easy-to-use website interface. Figure 3 represents a possible use case. Among more than 26 DL-based Ab modeling programs, this is one of the few cases that provides such a function. As we tried this program on our own hardware, the provided instructions for installation and execution were easy to follow as well. The runtime is very fast (i.e., fully completing within 1 min).

Name	OASis identity		OASis percentile		Germline content		Germlines		Humanizing mutations		Actions
	Before	After	Before	After	Before	After	VH	VL	VH	VL	
Antibody1	37%	74%	0%	35%	66%	79%	IGHV1-46*01	IGKV3-11*01	15	16	Detail

**Figure 3.** An example *BioPhi* humanization result. The *OASis* identity can be used as a threshold to separate human, humanized, chimeric, and murine Abs. The *OASis* percentile represents the percentile of *OASis* identity among therapeutic Abs (Supplementary Note S4).

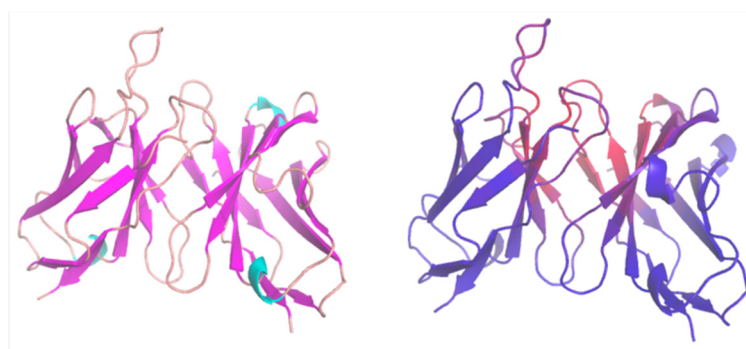
Most LMs use autoregressive methods, and there are several LM-based Ab sequence generation tools. For example, the Gray group shared the Ig Language Model (*IgLM*), which is trained with 558M Ab heavy- and light-chain variable sequences [85]. The *IgLM* generates full-length Ab sequences based on chain type and species of origin. By diversifying loops within an Ab, it creates high-quality synthetic libraries that exhibit biophysical properties consistent with natural Ab sequences. These synthetic libraries also demonstrate lower immunogenicity and greater resemblance to human Abs compared to baseline models. Another LM-based Ab sequence generation method is the ESM-1b transformer-based, ML-guided Antigenic Evolution Prediction (MLAEP) model [86]. *ReprogBERT* generates diverse Ab (CDR) sequences (more than two-fold increase) without losing structural integrity and naturalness [87].

### 3. Antibody Structural Modeling

#### 3.1. Fragment Variable Structure and Predicting the Impacts of Mutations on the Structure and Function

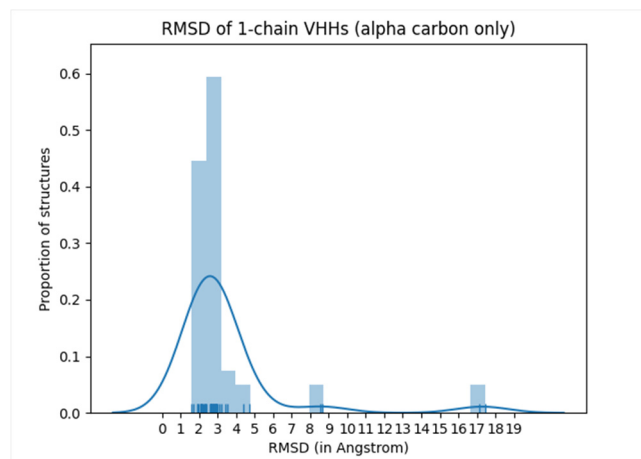
We include this subsection on fragment variables (Fvs), as they comprise the VH and VL domains of Abs, representing the smallest segment retaining the complete binding capacity of the intact Ab. The CDR region, which many computational modeling tools aim

to predict, is also located within the Fv. Therefore, predicting binding based on mutations in the Fv region has become a prevalent objective. *DeepAb* is one of the recent programs addressing this pressing requirement. *DeepAb* details the alterations in categorical cross entropy of designed Ab sequences. Specifically, *DeepAb* offers predictions on the effects of Ab mutations on binding against a target, such as lysozyme. Furthermore, *DeepAb* predicts the structures of both Fvs and Nbs. Its Fv structure prediction accuracy was the best among four benchmark methods across all loops. Fv structure is predicted with two stages (i.e., identifying residue relationships and structure refinement). The first stage uses ResNet, which predicts relative distances and orientations between pairs of residues, similar to *trRosetta* [88]. Interestingly, *DeepAb* builds initial protein structures through multi-dimensional scaling (MDS) to bypass expensive sampling for much of the Ab structure. This approach becomes possible due to the high conservation of the framework structural regions of Abs. This MDS approach is different than most computational protein structure predictions [88,89] that sample protein backbone torsion angles (phi and psi) explicitly. Of course, further refinement is needed to remove clashes and non-ideal geometries right after MDS-based initial structure generation. As transformer attention visualization [90], *DeepAb* uses crisscross attention [91] to represent which residues attend more with each other (Figure 4). The second stage is a *Rosetta*-based protocol for structure realization/optimization using quasi-Newton minimization that has been used traditionally [92]. At this stage, explicit values of protein backbone dihedral angles and  $d_{CA}$  are used. As a recent dilated CNN-based embedding cluster protein function, *DeepAb* further projected sequence-averaged LSTM embedding by species and loop structures. When we tested this program, installation and execution were easy to follow (Supplementary Note S5).



**Figure 4.** *DeepAb*-predicted Fv structures. **Left:** colored by secondary structures (i.e., loops in salmon and beta sheets in pink); **right:** colored by attention scores (red residues have high attention values, while blue residues have low attention values).

Other Ab structure prediction methods using DL include *IgFold*, which is a pre-trained LM [93]. Despite not employing VAE, *IgFold* resembles *IG-VAE*, as it produces 3D coordinates of full-atom Abs directly. This direct reconstruction process is carried out by graph networks for the backbone atom coordinates and by *Rosetta* for the sidechain. As with *IgLM*, it is trained on 558M natural Ab sequences. *IgFold* predicts Ab structure faster than *DeepAb* and *AlphaFold* with comparable or slightly better accuracy. The rise of these LM-based Ab models has been expected due to the development of various LM-based protein designs [61,94]. Here, we share our independent benchmark result with *IgFold* (Figure 5, Supplementary Note S6, see Methods section for detail). Even without *PyRosetta*-based structure refinement, most predictions are accurate and produced with fast execution speed (i.e., a few seconds per ~100 AA sequence, even without a GPU).



**Figure 5.** Small benchmark result of *IgFold* with Nb structures. All predictions that result in a 8~17 Å root mean squared deviation (RMSD) between experiment structure and predicted structure have floppy long terminal (either N-terminal or C-terminal) regions in experimental structures. Therefore, it is quite likely that either the experiments themselves were incomplete or *IgFold* prefers well-folded Nb structures.

Another recent attention-based DL model of immune protein structure prediction is ImmuneBuilder [95]. It predicts the structure of Abs (ABodyBuilder2), Nbs (NanoBodyBuilder2), and T-cell receptors (TCRBuilder2). A notable improvement of ABodyBuilder2 is that it ran over a hundred times faster than AlphaFold-Multimer [29], while it predicted the CDR-H3 loop structures with marginally better accuracy in Abs and Nbs. This is interesting, considering that ABodyBuilder2 is an Ab-specific version of the structure module in AlphaFold-Multimer with several tweaks. Similar results were reported for NanoBodyBuilder2 and TCRBuilder2 as well. Errors are estimated for every residue with an ensemble of structures.

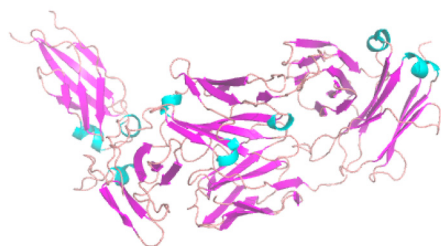
### 3.2. Methods and Techniques in Screening for Binding Antibodies

Numerous programs have been generated to address Ab–Ag interactions. Given the vast number of Ab sequences to assess, these quick in silico screening techniques have been eagerly anticipated. *DLAB* (Deep Learning approach for AntiBody screening) [96] is one such example. *DLAB* is one of the structure-based Ab-DL methods, along with the topology-based approach [50] and the geometric representation of the surface patches [97]. It uses scores of the docking poses of Ab–Ag pairings. However, it adopts three-dimensional gridding with Conv3D. While the depiction of the protein structure/interface within a voxel grid is common in other protein DL models [98,99] and relatively easy to comprehend, it is acknowledged that this approach is computationally demanding, particularly in terms of memory usage [100]. The input size of the network restricts the voxel space to a single size [5]. In fact, most traditional multilayer perceptron (MLP) and CNN architectures do not robustly deal with rotation-invariant features [101,102]. Therefore, efforts toward SE(3) representation such as data augmentation [103], extraction of spatially invariant features as seen in spatial-VAE [104], and point cloud data [105] are desired. Trained with Ab structures (modeled with *ABodyBuilder* [106] and *ZDOCK* [107]), *DLAB* predicts Ab–Ag binding for Ags with no known Ab binders. This trait is similar to *DeepAb* [12], which shows moderate predictability of mutational tolerability, even without explicit knowledge of the Ag. Therefore, *DLAB* and *DeepAb* can be useful for early-stage development of Ab therapeutics (i.e., virtual screening).

In an alternative screening approach, Kim’s team employed LM-based affinity maturation techniques [108]. They examined fewer than 20 variants for each Ab, aiming to improve binding affinities. This effort also included Abs known for their excellent thermostability and their ability to neutralize Ebola and SARS-CoV-2 pseudoviruses. A



noteworthy takeaway is that more than half of the mutations proposed in this LM pertain to framework regions, areas traditionally deemed less relevant than the CDR [109]. Other LM (*ProtTrans* [110])-based prediction methods of the binding affinity of Abs (BCR) and Ags include *DeepAIR* [111]. The *Binding-ddg-predictor* [15] focuses on the design of Ab–Ag interfaces, employing an attention-based geometric neural network for analysis. Specifically, the geometric part of the model learns an embedding for each AA based on the proximity of its neighboring atoms. The attention network, reflecting the learned geometric embeddings, recognizes key residue pairs near the protein interface that have an impact on binding affinity. As a result, it generates libraries of CDR mutations, ranking each mutation according to its influence on binding affinity and structural stability. After this cyclical optimization process, this DL approach improved an Ab, which displayed significantly increased and more effective virus-neutralizing activity compared to the original Ab. We present an example that can be leveraged using the *Binding-ddG-predictor* (Figure 6). In this structure, even a single mutation (among 628 total AAs) was predicted to alter the ddG by 0.3.



**Figure 6.** One of the *Binding-ddG-predictor*-applicable structures (i.e., beta sandwich protein). *Binding-ddG-predictor* is a semi-generative method (i.e., it generates an in silico mutation library of CDRs). Specifically, it ranks sequences by trained geometric neural network in such a way that they should improve Ab binding to the SARS-CoV-2 Delta variant RBD).

### 3.3. Strategies in Designing Both Sequence and Structure in Optimizing Antibody Efficacy

A recent trend in Ab modeling includes the co-design of sequence and structure to overcome previous approaches. For example, Jin et al. employed an autoregressive/generative graph neural network (GNN) to co-design the sequence and 3D structure of CDRs [34]. Essentially, this approach is similar to *RosettaAntibodyDesign* [112], as it focuses solely on predicting CDRs while leaving the framework region unaltered. However, instead of a physics-based score function as used in *RosettaAntibodyDesign*, they used graph-based DL. Here, a graph is a sequence–structure pair, and it models the conditional relationship between a CDR and its environments. This unique capability is possible by modeling protein backbone angles on top of AA identity and a joint graph representation that connects between CDR residues and between the CDR and framework (rather than single-residue prediction with CDR only). Since framework+CDR has more than 100 residues, graph convolution over the whole graph (>100 nodes) is challenging. Therefore, they clustered framework residue into K-mers to minimize the graph size. Unlike previous autoregressive models that never update residue distances, even when a new residue is added, their model updates residue distances whenever a new residue is introduced. Consequently, this method was very effective with a rotation/translation-invariant loss function and outperformed all other methods (e.g., *RosettaAntibodyDesign*, LSTM, and AutoRegression) in terms of speed, sequence recovery rate (<29% vs. 34%), and structure prediction. As a multi-objective optimization, this model designs CDR-H3 sequences that have higher neutralization probability as well.

### 3.4. Application of Diffusion Methods

Another co-design of CDR sequence and structure involves a diffusion probabilistic model, which demonstrated competitive binding affinities according to *Rosetta* energy functions, other protein design metrics, and in vitro experimental validation [113]. *AbDiffuser*,

in particular, utilizes the Aligned Protein Mixer (APMixer), an innovative neural network designed for the processing of proteins from an aligned protein family [113]. APMixer is more memory-efficient and operates more rapidly than GNNs and transformer-based models. Additional applications of diffusion models in Ab applications encompass *diffusionN Optimized Sampling (NOS)* [63] and *EvoDiff* [114]. *NOS* enables direct execution of designs within the sequence space, effectively bypassing substantial constraints associated with structure-based methods (i.e., limited data and complex inverse design challenges). Remarkably, *NOS* shows better sequence recovery by infilling than *IgLM* [85], *RFdiffusion* [115], and *DiffAb* [116]. *EvoDiff* utilizes evolutionary data to generate proteins beyond the reach of structure-based modeling techniques. It improves Fréchet ProtT5 distance (FPD) even better than *RosettaFoldDiffusion (RFdiffusion)* [115]. The denoising diffusion probabilistic model has recently been employed in numerous general protein structure prediction and design models [115,117–123] and in generating end-to-end protein–ligand complexes [124]. Hence, the application of the SE(3) diffusion model specifically to Abs was anticipated. A drawback of the diffusion-based Ab model is its need for an Ab framework attached to the target Ag.

### 3.5. Graph-Based Supervised Learning for Biophysical Property Prediction

In this section, we discuss the application of graph-based learning in predicting biophysical properties of Abs. GNNs have been widely used across all scientific domains [17,125]. For applications in property prediction, message-passing neural networks (MPNNs) have been used to predict IC<sub>50</sub> values with anti-SARS database and molecular property [125]. For PPI research, GNNs have been used to predict various features, such as the 3D structure of a protein–protein complex [126], synergy scores of drug combinations [127], and effects of mutations on protein–protein binding affinity [128], PPI link [129], PPI site [130,131] and patterns in protein–protein interfaces [132]. For Ab modeling specifically, GNNs have been used to co-design the sequence and 3D structure of CDRs and affinity maturation [34,133].

To consider physiological impact of novel therapeutics, DL-based Ab developability filtering methods [9,82] can save a huge amounts of resources. These Ab developability filtering methods aim to predict various biophysical properties. These physicochemical/biophysical properties include thermal and colloidal stability, aggregation, fragmentation, hydrophobic patches/surfaces, solubility, post-translation modification (PTM), and half-life (pharmacokinetics), as reviewed in [1]. Most of these properties can be trained on either numerical or categorical values with DL methods. However, when Ab-related biophysical predictions use generative methods, the copy problem (i.e., generative modeling may reproduce the training data too closely) [134] should be avoided to generate reasonably novel Ab sequences. Overall, computational constraints that govern the developability of therapeutic Abs are summarized as evenly distributed hydrophobic residues across the surface, avoiding glycosylation motifs and CDR residues with reasonable levels of charges [135]. Various non-DL-based in silico methods for Ab developability parameter computation were summarized by Akbar et al. [41]. Other aspects of Ab developability, such as avoiding unusual CDR sequences that are not explicitly explained by biophysical properties, can be examined through the perplexity calculated from an ensemble of LMs [136]. A supervised graph-based approach shows considerable potential in predicting biophysical properties during Ab design. Even when non-DL ML methods such as random forest, Gaussian processes, and nearest neighbors or a simple MLP were employed, graph-based signatures demonstrated their effectiveness in capturing the interaction interfaces between Ab and Ag and in predicting binding affinity [137].

### 3.6. Curation of Sequence and Structural Datasets to Develop Unsupervised Machine Learning Methods

To achieve success with unsupervised ML techniques, it is essential to have high-quality datasets comprising sequences and structures. With abundant data, these methods

can excel in identifying patterns or tendencies and grouping them within the latent space, broadening their domain of application. Moreover, it is crucial to tackle data bias in the training set to guarantee precise extrapolation. However, the existing databases of Ab and PPI sequences and structures (Tables 1 and 2), which were used for development of various DL-driven Ab modeling programs (Table 3), do not offer comprehensive coverage of all known antibody sequences. Additionally, many Ab sequences and structures are redundant. This obvious limitation of the datasets is evident, given the fact that in addition to the core regions of Abs (i.e., the framework) but the CDR region alone requires a  $20^{60}$  possible combinatorial search space of various sequences [34]. Therefore, to minimize data bias, reasonable in silico generation methods (such as *Absolut!* [54]) can be considered to reduce the gap. Additionally, interpretable models would be useful to assess data completeness [138]. Effective embedding can also minimize the data bias issue by better intra/extrapolation. For example, recently, a dilated CNN-based embedding was employed to model a protein function [18]. Specifically, the findings demonstrate that using contextualized word-embedding representation for protein sequences [139] eliminates the need to incorporate explicit structural information, which, in turn, effectively simplifies the modeling process [140]. These computational efforts will eventually allow for effective Ab sequence modeling. A similar analogy can be found in computational protein structure prediction. Due to recent advancements in multiple sequence alignment (MSA)-based structure prediction [141], the sampling process has become significantly more efficient. This enhancement significantly addresses the vast potential inherent in protein folding, a concept previously referred to as Levinthal's paradox [142]. Therefore, as the Akbar group has contributed a substantial amount of Ab structure data for training with GRUs [54], scientists have widely adopted transformer based MSA programs like *AlphaFold* [141] and *RoseTTAFold* [143] to generate synthetic Ab structures. Nonetheless, it is crucial to acknowledge the superiority of LMs, such as *IgFold* [93] and *OmegaFold* [144], in terms of the overall prediction accuracy for Ab structures. This holds true particularly for the complex structure of CDR3. Furthermore, these LMs exhibit a notable advantage in terms of speed when compared to *AlphaFold*, representing a significant improvement in the field.

**Table 1.** Ab sequence and structure databases. We present mostly large-scale databases. Other databases were reviewed by Akbar et al. [41] and Wilman et al. [2].

Data Source	Description	Number of Entries
AbDb [45]	Expert-curated Ab structure database	~2 k full structures
Absolut! [54]	In silico generated Ab–Ag bindings	159 antigens times 6.9 million CDR-H3 murine sequences
AntiBodies Chemically Defined Database (ABCD) [145]	Manually curated depository of sequenced Abs	23 k sequenced Abs against 4 k Ags
CoV-AbDab (in SAbDab) [146]	Coronavirus-binding Ab sequences and structures	4 k homology models and 500 PDB structures
iReceptor [147]	Ab/B-cell and T-cell receptor repertoire data	>5 B
Observed Antibody Space (OAS) [84]	Paired and unpaired (VH/VL) Ab sequences	>1 B
SAbDab [44]	Ab structures available in PDB	>5 k

**Table 2.** PPI sequence and structure databases. We present mostly large-scale protein interaction databases. Other databases were reviewed by Akbar et al. [41]. SKEMPI V2.0 is for general PPIs, yet it is also the largest Ab–Ag binding affinity database.

Data Source	Description	Number of Entries
IntAct [148]	Binary interactions from the literature and user submissions	>1 M
MINT (in IntAct)	Protein interaction information disseminated in the literature	>130 k
SKEMPI V2.0	Structural Kinetic and Energetic database of Mutant Protein Interactions	7 k
STRING [149]	Direct (physical) and indirect (functional) PPIs	>20 B

**Table 3.** Some of the prominent antibody modeling programs. A full list of programs is presented in Supplementary Note S7.

Model	Goal	Input Type	Output	Architecture	Metrics	Note
Binding-ddg-predictor	Redesign the CDR to enhance Ab affinity (targeting multiple virus variants)	Sequence	Predicted binding affinity	Attention-based geometric neural network	kD (dissociation constant)	Through an iterative optimization procedure, this DL method found that the optimized Ab exhibited broader and much more potent neutralizing activity compared to the original Ab
BioPhi	Humanize the sequence and evaluate the humanness of the sequence	Sequence	Sequence and predicted humanization	Transformer	Accuracy (%), ROC, AUC, and R <sup>2</sup>	Different methods were more successful in different cases, further encouraging the assembly of a diverse arsenal of humanization methods
DeepAb	Predict the Ab mutation effect on binding	Sequence	Structure and predicted affinity	RNN for sequence representation and ResNet to predict six distances and angles	Oriental coordinate distance and AUC	Provides an attention layer to interpret the features contributing to its predictions
IgFold	Predict Ab (Fv) structure	Sequence	Predicted Ab structures	Pre-trained language model followed by graph networks that directly predict backbone atom coordinates	Oriental coordinate distance and RMSD	Representations from IgFold may be useful as features for ML models
IG-VAE	Directly generate 3D coordinates of full-atom Abs	Known IG structures	Diversified IG structures	VAE	Distance matrix reconstruction and torsion angle inference	Intended for use with existing protein design suites such as Rosetta
iNNterfaceDesign	One-sided design of protein–protein interfaces	Both sequence and structure (features of protein receptors)	Redesigned protein interface sequence and structures	LSTM with attention	Recovery rates of the native sequence and hot spot	First neural network model for prediction of amino acid sequences for peptides involved into interchain interactions
RefineGNN	Co-design of the sequence and 3D structure of CDRs as graph	Both sequence and structure	Both sequence and structure	Autoregressive/generative graph neural network	Perplexity of sequences and the RMSD	Co-designs the sequence and 3D structure of CDRs as a graph



## 4. The Role of Antibodies and Deep Learning in the Fight against SARS-CoV-2

### 4.1. Understanding How SARS-CoV-2 Interacts with Host Cells

To more effectively illustrate the application of DL in Ab research, we discuss SARS-CoV-2 in this section. The human death toll of SARS-CoV-2 is estimated to be 10~20 million so far [150]. Additionally, this disease has impacted economics negatively. Therefore, many structural studies about the disease have been reported, such as those investigating Nb [151] or Fab [152] binding to the receptor-binding domain (RBD). Often, structural studies of the target protein suggest plausible routes of development, such as the suggestion of an anti-neuraminidase Ab as a starting point for the design of an anti-SARS Ab [153]. The authors of the abovementioned NB and Fab binding studies attempted to inhibit angiotensin-converting enzyme 2 (ACE2) binding either directly or indirectly. This ACE2-RBD interaction can be blocked by fusion proteins (such as extracellular portions of ACE2 or RBD fused to the Fc portion of human IgG1) as well [154]. Fusion glycoproteins lie on the surface of enveloped viruses and are important for the cell entry of viruses [155]. This fusion protein-targeting approach has been gaining traction for other viruses, such as respiratory syncytial virus. For example, the King group computationally designed immunogens that induced potent neutralizing Ab responses [156]. All three classes of fusion proteins (i.e., I, II, and III) have different structures, mechanisms (triggering molecule/pH or reversibility between pre-fusion and post-fusion), and applicable viral families. For SARS-CoV-2, which belongs to the Coronaviridae family, class I fusion protein acts as the fusion machinery. As a result, class I fusion proteins have been studied most extensively. To develop ML models to combat SARS-CoV-2, it is necessary to have a specific dataset for class I fusion proteins. Well-known cryo-EM structures show that class I fusion proteins have a high proportion of alpha helices in their post-fusion conformation with coiled coils [152]. The structural stability of the pre-fusion and pre-hairpin states, which are intermediate stages between pre-fusion and post-fusion, is lower compared to that of the post-fusion state [141]. This implies that the examination of pre-states should incorporate both kinetic and thermodynamic properties, leveraging tools like molecular dynamics or Monte Carlo simulations, as well as quantum mechanical calculations. In contrast, the post-fusion state can primarily be evaluated based on its thermodynamic properties. Stabilization of the pre-fusion and intermediate states prior to reaching the post-fusion stage can aid in averting viral infections. Alternatively, as suggested earlier, direct inhibition of the binding interaction between ACE2 and the receptor-binding domain (RBD) is also a viable strategy.

### 4.2. Overview of Experimental Datasets in Studying SARS-CoV-2

We present various experimental datasets that were used as training sets for DL related to SARS-CoV-2. Neutralizing antibodies (nAbs) are effective for the prevention and treatment of SARS-CoV-2-related infections. Therefore, phage-display immune libraries were employed to isolate effective nAbs against SARS-CoV-2 from pooled peripheral blood mononuclear cells (PBMCs) of COVID-19 convalescent patients [157]. This phage-display screening identified a neutralizing IgG that attaches to an epitope located on the N-terminal domain of SARS-CoV-2 [158]. mAbs that neutralize and obstruct the binding of the SARS-CoV-2 spike protein to ACE2 can also be discovered through target-ligand blocking methods and BCR sequencing. This involves linking the BCR to Ag specificity via sequencing [159]. For such campaigns or screenings, the dissociation constant ( $K_D$ ) and  $IC_{50}$  are crucial measurements. The binding kinetics ( $K_D$ ) of the mAbs to the target molecule are typically assessed using surface plasmon resonance (SPR) [1]. To quickly assess developability, protein thermal unfolding temperatures, such as the midpoint temperature ( $T_m$ ) and onset temperature ( $T_{onset}$ ), can be gauged using differential scanning fluorimetry. Of course, these experimental screenings can be lessened or supplemented by virtual screenings of Abs [9,39,96].

#### 4.3. How Deep Learning Is Advancing Research on SARS-CoV-2

In this section, we present a range of DL techniques used to address SARS-CoV-2 challenges on top of a well-summarized review [160]. For instance, CNNs have been employed to pinpoint the representative genomic sequence of SARS-CoV-2 among various viral genome strains [161]. Furthermore, DL methods have been utilized to repurpose existing drugs for COVID-19 through network-based approaches [162] and to analyze COVID-19 computed tomography imaging using UNET [163]. Thus, the impact of DL methods in the fight against SARS-CoV-2 is significant, with an ever-growing list of specific contributions. However, DL's role in SARS-CoV-2 Ab design has largely been confined to predicting binding affinity related to general PPIs [164]. This limitation stems from the challenge of fine tuning existing general PPI-based DL models for SARS-CoV-2-targeted Ab design due to a scarcity of SARS-CoV-2-specific datasets. For instance, only around 500 experimentally determined SARS-CoV-2-specific structures exist (Table 1). Nevertheless, given the recent successful applications of DL for the optimization of specific SARS-CoV-2 variants [15], it is expected that the development of DL-based COVID-19 Ab designs will become more prevalent. DL methods have expedited the development of therapeutics by identifying epitopes, offering a time-efficient alternative to experimental screening [165,166]. Many of these methods have been employed to combat SARS-CoV-2 or hold the potential to be applied for the same. Specifically, T-cell epitope prediction utilizing artificial neural networks includes RNN-based prediction of peptide–human leukocyte antigen (HLA) class II binding [167], along with sparse encoding and BLOSUM [168] encoding-based prediction of HLA-DR binding, an MHC class II cell surface receptor encoded by the HLA [169]. Conversely, B-cell epitope prediction employing artificial neural networks includes RNN-based prediction of linear or continuous B-cell epitopes of an antigen [170].

There is substantial room for enhancements in DL methodologies used in Ab design that can be universally applied, not only for SARS-CoV-2. First, incorporating evolutionarily conserved sequence information (MSA) can enhance DL-based Ag design, as its effectiveness has already been demonstrated in general protein structure prediction [141,143] and design [74]. Aside from the CDR-H3 region, Ab sequences exhibit high similarity, which supports the feasibility of using MSA. Secondly, diversifying Ab repertoires reflecting *in vivo* insertion and deletion of AAs into the V region, post-translational modifications, and the use of non-protein cofactors [171] have not been fully realized in DL approaches [41]. Nonetheless, this limitation is being addressed; for instance, glycan information in proteins has been converted to lattice representations to generate a wealth of DL training data [54]. DL has not advanced as much in certain areas, including germ lines, Ab formats (Fc-fusion, scFv, and Fab), specific sequence liabilities (deamidation and glycosylation sites), and clearance likelihood [4]. Lastly, many generative programs for Ab sequence design lack appropriate code/document sharing. To make a greater impact on the community, more collaborative approaches should be encouraged.

Establishing a foundation for ML and structural modeling for SARS-CoV-2 requires careful consideration. For instance, smaller monomer–monomer interactions (such as those between the RBD and Nbs) can be investigated via docking [29,172]. However, the entirety of the class I fusion protein often exceeds 2000 AA, making *AlphaFold-Multimer* studies impractical due to memory limitations. Attempts to circumvent this by removing the C-terminal region to fit into the *AlphaFold-Multimer*'s memory (such as on Google Colab Pro+) often results in orientations between chains that significantly deviate from experimental findings. As such, a threading method with individual structure is recommended. Vaccine design strategies using non-DL ML methods encompass combinatorial ML approaches such as support vector machine, k-nearest neighbors, logistic regression, random forest, and extreme gradient boosting. These methods underpin reverse vaccinology, which starts by predicting the optimal vaccine candidate through bioinformatics analysis of the Ag genome [173].

## 5. Conclusions and Future Directions

We have highlighted several DL-based Ab modeling programs [12,15,33,79] that utilize *Rosetta* [174] for ddG calculation, structure idealization, sequence design, sidechain optimization, and visualization. Given that the *Rosetta* group is one of the forefront groups of computational protein design [66], it has been expected that they will continue incorporating DL methods to enhance their current capabilities in Ab design [175,176] and Ab and protein stability improvement [177,178]. This will aid in addressing challenges such as Ab thermal stability [179] and Ab structure prediction [180].

Reliable Ab–Ag interaction prediction, particularly due to the diverse conformations of the H3 loop in CDR, has remained an elusive goal [12,68]. For example, *AlphaFold-Multimer* has not demonstrated reliable Ab–Ag binding prediction [29]. The limited amount of data, varying biological platforms [4], and structural alterations upon Ag interaction [181] also present challenges when applying DL to Ab research. Consequently, it has been a typical approach to employ databases that encompass both Ab-specific and broader PPI data [182], as we previously mentioned.

However, recent advances in DL approaches for Ab sequence design and classification have accelerated Ab development, enabling the exploration of a much larger protein sequence space than display libraries can offer [9]. For instance, generating a vast number of simulated 3D Ab–Ag complex structures that represent various biological complexities [54] may provide a valuable foundation for the enhancement of DL techniques in Ab sequence design. Furthermore, LM-based Ab structure prediction methods such as *IgFold* and *OmegaFold*, which are faster and offer similar or better quality predictions compared to *AlphaFold*, are starting to emerge. LMs hold considerable promise and are highly generalizable. For instance, Hie and colleagues have shown that their approach is not limited to Abs but can be effectively extended to other proteins as well [109]. In every test instance, it was observed that the general-purpose protein LM yielded better results than methods focused solely on Abs.

Here, we share opinions that are applicable both to general PPIs and Ab sequence design. First, other than two cases [35,83], most current sequence-based DL models for Ab sequence design and modeling represent protein sequences with one-hot encoding. This traditional method often used in RNNs is fine for most cases and sometimes ideal for ease of training on the available limited data [12]. However, transformer models can deal with long protein sequence information more easily [19]. Therefore, it better captures relationships between sequences with an attention model and can visualize these relationships intuitively as well [31].

Secondly, it is worthwhile to focus on updating models originally designed for individual monomer prediction. A prime example of this methodology is *AlphaFold-Multimer* [29]. It employs the same *Evoformer*, a transformer with multiple sequence alignment (MSA) representation and pair-wise information, as the original *AlphaFold* [141]. However, it adopts a longer-distance cutoff for the frame-aligned point error (FAPE) loss. This loss refers to the distances between the actual and predicted atoms in the local reference frame of each residue. This adjustment is implemented to facilitate the training of interchain pairs. Notably, this method has demonstrated higher accuracy compared to previously prominent techniques, including *AlphaFold*-refined *ClusPro* docking. Thirdly, the protein–protein docking score is better in a standardized form between the DockQ score and TRScore [183] for consistent comparison and reporting.

Finally, as highlighted by Shaver et al. [4], the use of pre-training techniques like masked language modeling (MLM) embodied in models such as Bidirectional Encoder Representations from Transformers (BERT) [184] and the Generative Pre-trained Transformer (GPT) [185] could greatly benefit the Ab sequence research community. The MLM technique involves obscuring certain residues in a protein sequence and training the model to predict the hidden AAs based on the rest of the sequence. This approach has already demonstrated potential in protein sequence applications [19,186]. One of its key benefits is its swift deployment time. Even though the training process and final structural refine-

ment can be time-consuming, the actual implementation of pre-trained models in both the pharmaceutical industry and academic research is quite manageable. As a result, we can anticipate the development of more LM-based Ig modeling methods, such as *IgLM* and *IgFold*.

## 6. Methods

We used *ChimeraX* [187] to generate Figure 1. The examples that we ran and the analysis scripts that we used for Ab programs are presented in Supplementary Notes S2–S6. For the *IgFold* benchmark (Figure 3, Supplementary Note S6), we downloaded the entire 1252 VHH (i.e., Nb) PDB structures from SAbDab (as of February of 2023). Then, according to the Chothia naming convention, we chose all one-chain PDB structures and removed entries with duplicated structures, leaving 51 unique representative Nb structures. We converted these into fasta files and ran *IgFold*. RMSD values were calculated by *BioPython* [188] after superimposing comparing structures. The RMSD values calculated in this way match to those calculated by *ChimeraX*.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bioengineering11020185/s1>, References [189–192] are cited in the Supplementary Materials, Supplementary Note S1. Running and interpretation of *AlphaFold-Multimer* result; Supplementary Note S2. Example *binding-ddg-predictor* script; Supplementary Note S3. Example *iNNterfaceDesign* script; Supplementary Note S4. Example *BioPhi* script; Supplementary Note S5. Running *DeepAb*; Supplementary Note S6. Example *IgFold* script; Supplementary Note S7. Detailed list of Ab modeling programs.

**Author Contributions:** Conceptualization, D.N.K. and N.K.; methodology, D.N.K. and A.D.M.; software, D.N.K. and A.D.M.; validation, D.N.K.; formal analysis, D.N.K.; investigation, D.N.K.; resources, N.K.; data curation, D.N.K.; writing—original draft preparation, D.N.K.; writing—review and editing, D.N.K., A.D.M. and N.K.; visualization, D.N.K.; supervision, N.K.; project administration, N.K.; funding acquisition, N.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Laboratory Directed Research and Development Program at the Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** For the benchmark input and output, refer to 10.5281/zenodo.10632946. We obtained 1252 VHH (also known as Nb) PDB structures from SAbDab in 2023. We did not generate any new experimental data.

**Acknowledgments:** This research used computational resources provided by Research Computing at the PNNL. We appreciate Kurt Glaesemann for PNNL's research computing support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bailly, M.; Mieczkowski, C.; Juan, V.; Metwally, E.; Tomazela, D.; Baker, J.; Uchida, M.; Kofman, E.; Raoufi, F.; Motlagh, S.; et al. Predicting Antibody Developability Profiles through Early Stage Discovery Screening. *mAbs* **2020**, *12*, 1743053. [[CrossRef](#)] [[PubMed](#)]
2. Wilman, W.; Wróbel, S.; Bielska, W.; Deszynski, P.; Dudzic, P.; Jaszczyszyn, I.; Kaniewski, J.; Młokosiewicz, J.; Rouyan, A.; Satława, T.; et al. Machine-designed biotherapeutics: Opportunities, feasibility and advantages of deep learning in computational antibody discovery. *Brief. Bioinform.* **2022**, *23*, bbac267. [[CrossRef](#)] [[PubMed](#)]
3. Lu, R.-M.; Hwang, Y.-C.; Liu, I.-J.; Lee, C.-C.; Tsia, H.-Z.; Li, H.-J.; Wu, H.-C. Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci.* **2020**, *27*, 1. [[CrossRef](#)] [[PubMed](#)]
4. Shaver, J.M.; Smith, J.; Amimeur, T. Deep Learning in Therapeutic Antibody Development. *Methods Mol. Biol.* **2022**, *2390*, 433–445. [[PubMed](#)]



5. 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. [[CrossRef](#)] [[PubMed](#)]
6. Laustsen, A.H.; Greiff, V.; Karatt-Vellatt, A.; Muyldermans, S.; Jenkins, T.P. Animal Immunization, in vitro Display Technologies, and Machine Learning for Antibody Discovery. *Trends Biotechnol.* **2021**, *39*, 1263–1273. [[CrossRef](#)]
7. Greiff, V.; Yaari, G.; Cowell, L.G. Mining adaptive immune receptor repertoires for biological and clinical information using machine learning. *Curr. Opin. Syst. Biol.* **2020**, *24*, 109–119. [[CrossRef](#)]
8. 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. [[CrossRef](#)]
9. Mason, D.M.; Friedensohn, S.; Weber, C.R.; Jordi, C.; Wagner, B.; Meng, S.M.; Ehling, R.A.; Bonati, L.; Dahinden, J.; Gainza, P.; et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat. Biomed. Eng.* **2021**, *5*, 600–612. [[CrossRef](#)]
10. Deac, A.; Veličković, P.; Sormanni, P. Attentive Cross-Modal Paratope Prediction. *J. Comput. Biol.* **2018**, *26*, 536–545. [[CrossRef](#)]
11. 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. [[CrossRef](#)]
12. Ruffolo, J.A.; Sulam, J.; Gray, J.J. Antibody structure prediction using interpretable deep learning. *Patterns* **2022**, *3*, 100406. [[CrossRef](#)]
13. Warszawski, S.; Katz, A.B.; Lipsh, R.; Khmel'nitsky, L.; Nissan, G.B.; Javitt, G.; Dym, O.; Unger, T.; Knop, O.; Albeck, S.; et al. Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLoS Comput. Biol.* **2019**, *15*, e1007207. [[CrossRef](#)]
14. Koehler Leman, J.; 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.; et al. Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol.* **2020**, *16*, e1007507. [[CrossRef](#)] [[PubMed](#)]
15. Shan, S.; Luo, S.; Yang, Z.; Hong, J.; Su, Y.; Ding, F.; Fu, L.; Li, C.; Chen, P.; Ma, J.; et al. Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2122954119. [[CrossRef](#)]
16. Huang, L.; Jiao, S.; Yang, S.; Zhang, S.; Zhu, X.; Guo, R.; Wang, Y. LGFC-CNN: Prediction of lncRNA-Protein Interactions by Using Multiple Types of Features through Deep Learning. *Genes* **2021**, *12*, 1689. [[CrossRef](#)] [[PubMed](#)]
17. Knutson, C.; Bontha, M.; Bilbrey, J.A.; Kumar, N. Decoding the protein–ligand interactions using parallel graph neural networks. *Sci. Rep.* **2022**, *12*, 7624. [[CrossRef](#)]
18. Bileschi, M.L.; Belanger, D.; Bryant, D.H.; Sanderson, T.; Carter, B.; Sculley, D.; Bateman, A.; DePristo, M.A.; Colwell, L.J. Using deep learning to annotate the protein universe. *Nat. Biotechnol.* **2022**, *40*, 932–937. [[CrossRef](#)]
19. Brandes, N.; Ofer, D.; Peleg, Y.; Rappoport, N.; Linial, M. ProteinBERT: A universal deep-learning model of protein sequence and function. *Bioinformatics* **2022**, *38*, 2102–2110. [[CrossRef](#)] [[PubMed](#)]
20. Joshi, R.P.; Gebauer, N.W.A.; Bontha, M.; Khazaieli, M.; James, R.M.; Brown, J.B.; Kumar, N. 3D-Scaffold: A Deep Learning Framework to Generate 3D Coordinates of Drug-like Molecules with Desired Scaffolds. *J. Phys. Chem. B* **2021**, *125*, 12166–12176. [[CrossRef](#)]
21. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)] [[PubMed](#)]
22. Yu, X.; Conyne, M.; Lake, M.R.; Walter, K.A.; Min, J. In silico high throughput mutagenesis and screening of signal peptides to mitigate N-terminal heterogeneity of recombinant monoclonal antibodies. *mAbs* **2022**, *14*, 2044977. [[CrossRef](#)] [[PubMed](#)]
23. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [[CrossRef](#)] [[PubMed](#)]
24. Tyka, M.D.; Keedy, D.A.; André, I.; Dimairo, F.; Song, Y.; Richardson, D.C.; Richardson, J.S.; Baker, D. Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *J. Mol. Biol.* **2011**, *405*, 607–618. [[CrossRef](#)] [[PubMed](#)]
25. McPartlon, M.; Xu, J. An end-to-end deep learning method for protein side-chain packing and inverse folding. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2216438120. [[CrossRef](#)] [[PubMed](#)]
26. Misiura, M.; Shroff, R.; Thyer, R.; Kolomeisky, A.B. DLpucker: Deep learning for prediction of amino acid side chain conformations in proteins. *Proteins* **2022**, *90*, 1278–1290. [[CrossRef](#)] [[PubMed](#)]
27. Gao, M.; Nakajima An, D.; Parks, J.M.; Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat. Commun.* **2022**, *13*, 1744. [[CrossRef](#)]
28. Basu, S.; Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS ONE* **2016**, *11*, e0161879. [[CrossRef](#)]
29. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **2022**. [[CrossRef](#)]
30. Jin, W.; Barzilay, R.; Jaakkola, T. Antibody-Antigen Docking and Design via Hierarchical Structure Refinement. *Proc. Mach. Learn. Res.* **2022**, *162*, 10217.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762v7. [[CrossRef](#)]

32. Barlow, D.J.; Edwards, M.S.; Thornton, J.M. Continuous and discontinuous protein antigenic determinants. *Nature* **1986**, *322*, 747–748. [[CrossRef](#)]
33. Syrlybaeva, R.; Strauch, E.-M. Deep learning of Protein Sequence Design of Protein-protein Interactions. *bioRxiv* **2022**. [[CrossRef](#)] [[PubMed](#)]
34. Jin, W.; Wohlwend, J.; Barzilay, R.; Jaakkola, T. Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-design. *arXiv* **2021**, arXiv:2110.04624.
35. Widrich, M.; Schäfl, B.; Ramsauer, H.; Pavlović, M.; Gruber, L.; Holzleitner, M.; Brandstetter, J.; Sandve, G.K.; Greiff, V.; Hochreiter, S.; et al. Modern Hopfield Networks and Attention for Immune Repertoire Classification. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18832–18845.
36. Goldenzweig, A.; Fleishman, S.J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129. [[CrossRef](#)]
37. Barlow, K.A.; Conchúir, S.Ó.; Thompson, S.; Suresh, P.; Lucas, J.E.; Heinonen, M.; Kortemme, T. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J. Phys. Chemistry. B* **2018**, *122*, 5389–5399. [[CrossRef](#)] [[PubMed](#)]
38. Dequeker, C.; Mohseni Behbahani, Y.; David, L.; Laine, E.; Carbone, A. From complete cross-docking to partners identification and binding sites predictions. *PLoS Comput. Biol.* **2022**, *18*, e1009825. [[CrossRef](#)] [[PubMed](#)]
39. Saka, K.; Kakuzaki, T.; Metsugi, S.; Kashiwagi, D.; Yoshida, K.; Wada, M.; Tsunoda, H.; Teramoto, R. Antibody design using LSTM based deep generative model from phage display library for affinity maturation. *Sci. Rep.* **2021**, *11*, 5852. [[CrossRef](#)]
40. Sher, G.; Zhi, D.; Zhang, S. DRREP: Deep ridge regressed epitope predictor. *BMC Genom.* **2017**, *18*, 676. [[CrossRef](#)]
41. Akbar, R.; Bashour, H.; Rawat, P.; Robert, P.A.; Smorodina, E.; Cotet, T.S.; Flem-Karsen, K.; Frank, R.; Mehta, B.B.; Vu, M.H.; et al. Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. *mAbs* **2022**, *14*, 2008790. [[CrossRef](#)] [[PubMed](#)]
42. Steinegger, M.; Söding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **2018**, *9*, 2542. [[CrossRef](#)] [[PubMed](#)]
43. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Christie, C.H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J.M.; et al. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **2021**, *49*, D437–D451. [[CrossRef](#)]
44. Schneider, C.; Raybould, M.I.J.; Deane, C.M. SAbDab in the age of biotherapeutics: Updates including SAbDab-nano, the nanobody structure tracker. *Nucleic Acids Res.* **2022**, *50*, D1368–D1372. [[CrossRef](#)]
45. Ferdous, S.; Martin, A.C.R. AbDb: Antibody structure database—A database of PDB-derived antibody structures. *Database J. Biol. Databases Curation* **2018**, *2018*, bay040. [[CrossRef](#)]
46. Sarkar, D.; Saha, S. Machine-learning techniques for the prediction of protein-protein interactions. *J. Biosci.* **2019**, *44*, 104. [[CrossRef](#)]
47. Kuroda, D.; Gray, J.J. Shape complementarity and hydrogen bond preferences in protein-protein interfaces: Implications for antibody modeling and protein-protein docking. *Bioinformatics* **2016**, *32*, 2451–2456. [[CrossRef](#)] [[PubMed](#)]
48. Greiff, V.; Menzel, U.; Miho, E.; Weber, C.; Riedel, R.; Cook, S.; Valai, A.; Lopes, T.; Radbruch, A.; Winkler, T.H.; et al. Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Rep.* **2017**, *19*, 1467–1478. [[CrossRef](#)]
49. Elhanati, Y.; Sethna, Z.; Marcou, Q.; Callan, C.G., Jr.; Mora, T.; Walczak, A.M. Inferring processes underlying B-cell repertoire diversity. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **2015**, *370*, 20140243. [[CrossRef](#)]
50. 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. [[CrossRef](#)]
51. Jankauskaitė, J.; Jiménez-García, B.; Dapkūnas, J.; Fernández-Recio, J.; Moal, I.H. SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **2019**, *35*, 462–469. [[CrossRef](#)]
52. Biswas, S.; Khimulya, G.; Alley, E.C.; Esvelt, K.M.; Church, G.M. Low-N protein engineering with data-efficient deep learning. *Nat. Methods* **2021**, *18*, 389–396. [[CrossRef](#)]
53. Fleishman, S.J.; Whitehead, T.A.; Ekiert, D.C.; Dreyfus, C.; Corn, J.E.; Strauch, E.M.; Wilson, I.A.; Baker, D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **2011**, *332*, 816–821. [[CrossRef](#)] [[PubMed](#)]
54. Robert, P.A.; Akbar, R.; Frank, R.; Pavlović, M.; Widrich, M.; Snapkov, I.; Chernigovskaya, M.; Scheffer, L.; Slabodkin, A.; Mehta, B.B.; et al. One billion synthetic 3D-antibody-antigen complexes enable unconstrained machine-learning formalized investigation of antibody specificity prediction. *bioRxiv* **2021**. [[CrossRef](#)]
55. Joshi, R.P.; Kumar, N. Artificial intelligence for autonomous molecular design: A perspective. *Molecules* **2021**, *26*, 6761. [[CrossRef](#)] [[PubMed](#)]
56. Xu, M.; Ran, T.; Chen, H. De Novo Molecule Design through the Molecular Generative Model Conditioned by 3D Information of Protein Binding Sites. *J. Chem. Inf. Model.* **2021**, *61*, 3240–3254. [[CrossRef](#)] [[PubMed](#)]
57. Ovchinnikov, S.; Huang, P.-S. Structure-based protein design with deep learning. *Curr. Opin. Chem. Biol.* **2021**, *65*, 136–144. [[CrossRef](#)]

58. Wu, Z.; Johnston, K.E.; Arnold, F.H.; Yang, K.K. Protein sequence design with deep generative models. *Curr. Opin. Chem. Biol.* **2021**, *65*, 18–27. [[CrossRef](#)] [[PubMed](#)]
59. Defresne, M.; Barbe, S.; Schiex, T. Protein Design with Deep Learning. *Int. J. Mol. Sci.* **2021**, *22*, 11741. [[CrossRef](#)]
60. Shin, J.-E.; Riesselman, A.J.; Kollasch, A.W.; McMahon, C.; Simon, E.; Sander, C.; Manglik, A.; Kruse, A.C.; Marks, D.S. Protein design and variant prediction using autoregressive generative models. *Nat. Commun.* **2021**, *12*, 2403. [[CrossRef](#)]
61. Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **2022**, *13*, 4348. [[CrossRef](#)]
62. Madani, A.; Krause, B.; Greene, E.R.; Subramanian, S.; Mohr, B.P.; Holton, J.M.; Olmos, J.L., Jr.; Xiong, C.; Sun, Z.Z.; Socher, R.; et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, *41*, 1099–1106. [[CrossRef](#)]
63. Gruver, N.; Stanton, S.; Frey, N.C.; Rudner, T.G.J.; Hotzel, I.; Lafrance-Vanasse, J.; Rajpal, A.; Cho, K.; Wilson, A.G. Protein Design with Guided Discrete Diffusion. *arXiv* **2023**, arXiv:2305.20009. [[CrossRef](#)]
64. Shanehazzadeh, A.; Bachas, S.; Kasun, G.; Sutton, J.M.; Steiger, A.K.; Shuai, R.; Kohnert, C.; Morehead, A.; Brown, A.; Chung, C.; et al. Unlocking de novo antibody design with generative artificial intelligence. *bioRxiv* **2023**. [[CrossRef](#)]
65. Murphy, G.S.; Sathyamoorthy, B.; Der, B.S.; Machius, M.C.; Pulavarti, S.V.; Szyperki, T.; Kuhlman, B. Computational de novo design of a four-helix bundle protein—DND\_4HB. *Protein Sci. A Publ. Protein Soc.* **2015**, *24*, 434–445. [[CrossRef](#)] [[PubMed](#)]
66. Kuhlman, B.; Dantas, G.; Ireton, G.C.; Varani, G.; Stoddard, B.L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364–1368. [[CrossRef](#)] [[PubMed](#)]
67. Parkinson, J.; Hard, R.; Wang, W. The RESP AI model accelerates the identification of tight-binding antibodies. *Nat. Commun.* **2023**, *14*, 454. [[CrossRef](#)] [[PubMed](#)]
68. Liu, G.; Zeng, H.; Mueller, J.; Carter, B.; Wang, Z.; Schilz, J.; Horny, G.; Birnbaum, M.E.; Ewert, S.; Gifford, D.K. Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics* **2019**, *36*, 2126–2133. [[CrossRef](#)] [[PubMed](#)]
69. Akbar, R.; Robert, P.A.; Weber, C.R.; Widrich, M.; Frank, R.; Pavlović, M.; Scheffer, L.; Chernigovskaya, M.; Snapkov, I.; Slabodkin, A.; et al. In silico proof of principle of machine learning-based antibody design at unconstrained scale. *bioRxiv* **2021**. [[CrossRef](#)] [[PubMed](#)]
70. Choi, Y.; Hua, C.; Sentman, C.L.; Ackerman, M.E.; Bailey-Kellogg, C. Antibody humanization by structure-based computational protein design. *mAbs* **2015**, *7*, 1045–1057. [[CrossRef](#)] [[PubMed](#)]
71. Wollacott, A.M.; Xue, C.; Qin, Q.; Hua, J.; Bohnuud, T.; Viswanathan, K.; Kolachalama, V.B. Quantifying the nativeness of antibody sequences using long short-term memory networks. *Protein Eng. Des. Sel. PEDS* **2019**, *32*, 347–354. [[CrossRef](#)]
72. Syrlybaeva, R.; Strauch, E.-M. One-sided design of protein-protein interaction motifs using deep learning. *bioRxiv* **2022**. [[CrossRef](#)]
73. Chaudhury, S.; Lyskov, S.; Gray, J.J. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26*, 689–691. [[CrossRef](#)] [[PubMed](#)]
74. Schmitz, S.; Ertelt, M.; Merkl, R.; Meiler, J. Rosetta design with co-evolutionary information retains protein function. *PLoS Comput. Biol.* **2021**, *17*, e1008568. [[CrossRef](#)] [[PubMed](#)]
75. Maguire, J.B.; Haddox, H.K.; Strickland, D.; Halabiya, S.F.; Coventry, B.; Griffin, J.R.; Pulavarti, S.V.S.R.K.; Cummins, M.; Thieker, D.F.; Klavins, E.; et al. Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **2021**, *89*, 436–449. [[CrossRef](#)] [[PubMed](#)]
76. Stranges, P.B.; Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **2013**, *22*, 74–82. [[CrossRef](#)] [[PubMed](#)]
77. Friedensohn, S.; Neumeier, D.; Khan, T.A.; Csepregi, L.; Parola, C.; de Vries, A.R.G.; Erlach, L.; Mason, D.M.; Reddy, S.T. Convergent selection in antibody repertoires is revealed by deep learning. *bioRxiv* **2020**. [[CrossRef](#)]
78. Davidsen, K.; Olson, B.J.; DeWitt, W.S., 3rd; Feng, J.; Harkins, E.; Bradley, P.; Matsen, F.A., 4th. Deep generative models for T cell receptor protein sequences. *eLife* **2019**, *8*, e46935. [[CrossRef](#)]
79. Eguchi, R.R.; Anand, N.; Choe, C.A.; Huang, P.-S. IG-VAE: Generative Modeling of Immunoglobulin Proteins by Direct 3D Coordinate Generation. *bioRxiv* **2020**. [[CrossRef](#)]
80. Zhong, E.D.; Bepler, T.; Berger, B.; Davis, J.H. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **2021**, *18*, 176–185. [[CrossRef](#)]
81. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv* **2018**, arXiv:1809.11096.
82. Amimeur, T.; Shaver, J.M.; Ketchum, R.R.; Taylor, J.A.; Clark, R.H.; Smith, J.; Van Citters, D.; Siska, C.C.; Smidt, P.; Sprague, M.; et al. Designing Feature-Controlled Humanoid Antibody Discovery Libraries Using Generative Adversarial Networks. *bioRxiv* **2020**. [[CrossRef](#)]
83. Prihoda, D.; Maamary, J.; Waight, A.; Juan, V.; Fayadat-Dilman, L.; Svozil, D.; Bitton, D.A. BioPhi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *mAbs* **2022**, *14*, 2020203. [[CrossRef](#)] [[PubMed](#)]
84. Olsen, T.H.; Boyles, F.; Deane, C.M. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci. A Publ. Protein Soc.* **2022**, *31*, 141–146. [[CrossRef](#)] [[PubMed](#)]



85. Shuai, R.W.; Ruffolo, J.A.; Gray, J.J. Generative Language Modeling for Antibody Design. *bioRxiv* **2021**. [CrossRef]
86. Han, W.; Chen, N.; Xu, X.; Sahil, A.; Zhou, J.; Li, Z.; Zhong, H.; Gao, E.; Zhang, R.; Wang, Y.; et al. Predicting the antigenic evolution of SARS-COV-2 with deep learning. *Nat. Commun.* **2023**, *14*, 3478. [CrossRef] [PubMed]
87. Melnyk, I.; Chenthamarakshan, V.; Chen, P.-Y.; Das, P.; Dhurandhar, A.; Padhi, I.; Das, D. Reprogramming Pretrained Language Models for Antibody Sequence Infilling. *arXiv* **2022**, arXiv:2210.07144. [CrossRef]
88. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496–1503. [CrossRef]
89. Xu, J.; Mcpartlon, M.; Li, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat. Mach. Intell.* **2021**, *3*, 601–609. [CrossRef]
90. Vig, J. Visualizing Attention in Transformer-Based Language Representation Models. *arXiv* **2019**, arXiv:1904.02679v2. [CrossRef]
91. Huang, Z.; Wang, X.; Wei, Y.; Huang, L.; Shi, H.; Liu, W.; Huang, T.S. CCNet: Criss-Cross Attention for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *45*, 6896–6908. [CrossRef] [PubMed]
92. Lemn, 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.; et al. Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* **2020**, *17*, 665–680. [CrossRef] [PubMed]
93. Ruffolo, J.A.; Chu, L.-S.; Mahajan, S.P.; Gray, J.J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nat. Commun.* **2023**, *14*, 2389. [CrossRef] [PubMed]
94. Ferruz, N.; Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **2022**, *4*, 521–532. [CrossRef]
95. Abanades, B.; Wong, W.K.; Boyles, F.; Georges, G.; Bujotzek, A.; Deane, C.M. ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. *Commun. Biol.* **2023**, *6*, 575. [CrossRef] [PubMed]
96. Schneider, C.; Buchanan, A.; Taddese, B.; Deane, C.M. DLAB: Deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics* **2021**, *38*, 377–383. [CrossRef]
97. Jespersen, M.C.; Mahajan, S.; Peters, B.; Nielsen, M.; Marcantili, P. Antibody Specific B-Cell Epitope Predictions: Leveraging Information from Antibody-Antigen Protein Complexes. *Front. Immunol.* **2019**, *10*, 298. [CrossRef]
98. Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D.R. Protein–Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957. [CrossRef]
99. Imrie, F.; Bradley, A.R.; van der Schaar, M.; Deane, C.M. Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data. *J. Chem. Inf. Model.* **2018**, *58*, 2319–2330. [CrossRef]
100. Li, N.; Kaehler, O.; Pfeifer, N. A Comparison of Deep Learning Methods for Airborne Lidar Point Clouds Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6467–6486. [CrossRef]
101. Rosebrock, A. Are CNNs Invariant to Translation, Rotation, and Scaling? 2021. Available online: <https://www.pyimagesearch.com/2021/05/14/are-cnns-invariant-to-translation-rotation-and-scaling/> (accessed on 5 February 2024).
102. Balci, A.T.; Gumeli, C.; Hakouz, A.; Yuret, D.; Keskin, O.; Gursoy, A. DeepInterface: Protein-protein interface validation using 3D Convolutional Neural Networks. *bioRxiv* **2019**. [CrossRef]
103. Si, D.; Moritz, S.A.; Pfab, J.; Hou, J.; Cao, R.; Wang, L.; Wu, T.; Cheng, J. Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM Density Maps. *Sci. Rep.* **2020**, *10*, 4282. [CrossRef]
104. Bepler, T.; Zhong, E.D.; Kelley, K.; Brignole, E.; Berger, B.; Wallach, H. Explicitly disentangling image content from translation and rotation with spatial-VAE. *arXiv* **2019**, arXiv:1909.11663.
105. Zhou, Q.-Y.; Park, J.; Koltun, V. Open3D: A Modern Library for 3D Data Processing. *arXiv* **2018**, arXiv:1801.09847. [CrossRef]
106. 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. [CrossRef] [PubMed]
107. Pierce, B.G.; Hourai, Y.; Weng, Z. Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS ONE* **2011**, *6*, e24657. [CrossRef]
108. Hie, B.L.; Shanker, V.R.; Xu, D.; Bruun, T.U.J.; Weidenbacher, P.A.; Tang, S.; Wu, W.; Pak, J.E.; Kim, P.S. Efficient evolution of human antibodies from general protein language models. *Nat. Biotechnol.* **2023**. [CrossRef] [PubMed]
109. Outeiral, C.; Deane, C.M. Perfecting antibodies with language models. *Nat. Biotechnol.* **2023**. [CrossRef] [PubMed]
110. Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. ProfTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7112–7127. [CrossRef] [PubMed]
111. Zhao, Y.; He, B.; Xu, F.; Li, C.; Xu, Z.; Su, X.; He, H.; Huang, Y.; Rossjohn, J.; Song, J.; et al. DeepAIR: A deep learning framework for effective integration of sequence and 3D structure to enable adaptive immune receptor analysis. *Sci. Adv.* **2023**, *9*, eabo5128. [CrossRef]
112. Adolf-Bryfogle, J.; Kalyuzhnyi, O.; Kubitz, M.; Weitzner, B.D.; Hu, X.; Adachi, Y.; Schief, W.R.; Dunbrack, R.L., Jr. RosettaAntibodyDesign (RABD): A general framework for computational antibody design. *PLoS Comput. Biol.* **2018**, *14*, e1006112. [CrossRef] [PubMed]
113. Martinkus, K.; Ludwiczak, J.; Cho, K.; Liang, W.-C.; Lafrance-Vanasse, J.; Hotzel, I.; Rajpal, A.; Wu, Y.; Bonneau, R.; Gligorijevic, V.; et al. AbDiffuser: Full-Atom Generation of In-Vitro Functioning Antibodies. *arXiv* **2023**, arXiv:2308.05027.



114. Alamdari, S.; Thakkar, N.; van den Berg, R.; Lu, A.X.; Fusi, N.; Amini, A.P.; Yang, K.K. Protein generation with evolutionary diffusion: Sequence is all you need. *bioRxiv* **2023**. [CrossRef]
115. 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.; et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *bioRxiv* **2022**. [CrossRef]
116. Luo, S.; Su, Y.; Peng, X.; Wang, S.; Peng, J.; Ma, J. Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures. *bioRxiv* **2022**. [CrossRef]
117. Chu, A.E.; Cheng, L.; El Nesr, G.; Xu, M.; Huang, P.S. An all-atom protein generative model. *bioRxiv* **2023**. [CrossRef]
118. Lee, J.S.; Kim, J.; Kim, P.M. Score-based generative modeling for de novo protein design. *Nat. Comput. Sci.* **2023**, *3*, 382–392. [CrossRef]
119. Yim, J.; Trippe, B.L.; De Bortoli, V.; Mathieu, E.; Doucet, A.; Barzilay, R.; Jaakkola, T. SE(3) diffusion model with application to protein backbone generation. *arXiv* **2023**, arXiv:2302.02277. [CrossRef]
120. Ingraham, J.B.; Baranov, M.; Costello, Z.; Barber, K.W.; Wang, W.; Ismail, A.; Frappier, V.; Lord, D.M.; Ng-Thow-Hing, C.; Van Vlack, E.R.; et al. Illuminating protein space with a programmable generative model. *bioRxiv* **2022**. [CrossRef] [PubMed]
121. Ni, B.; Kaplan, D.L.; Buehler, M.J. Generative design of de novo proteins based on secondary-structure constraints using an attention-based diffusion model. *Chem* **2023**, *9*, 1828–1849. [CrossRef] [PubMed]
122. Anand, N.; Achim, T. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models. *arXiv* **2022**, arXiv:2205.15019. [CrossRef]
123. Lisanza, S.L.; JGershon, J.M.; Tipps, S.; Arnoldt, L.; Hendel, S.; Sims, J.N.; Li, X.; Baker, D. Joint Generation of Protein Sequence and Structure with RoseTTAFold Sequence Space Diffusion. *bioRxiv* **2023**. [CrossRef]
124. Nakata, S.; Mori, Y.; Tanaka, S. End-to-end protein–ligand complex structure generation with diffusion-based generative models. *BMC Bioinform.* **2023**, *24*, 233. [CrossRef]
125. Bilbrey, J.; Ward, L.; Choudhury, S.; Kumar, N.; Sivaraman, G. Evening the Score: Targeting SARS-CoV-2 Protease Inhibition in Graph Generative Models for Therapeutic Candidates. *arXiv* **2021**, arXiv:2105.10489.
126. Ganea, O.-E.; Huang, X.; Bunne, C.; Bian, Y.; Barzilay, R.; Jaakkola, T.; Krause, A. Independent {SE}(3)-Equivariant Models for End-to-End Rigid Protein Docking. *arXiv* **2022**, arXiv:2111.07786.
127. Wang, X.; Zhu, H.; Jiang, Y.; Li, Y.; Tang, C.; Chen, X.; Li, Y.; Liu, Q.; Liu, Q. PRODeepSyn: Predicting anticancer synergistic drug combinations by embedding cell lines with protein–protein interaction network. *Brief. Bioinform.* **2022**, *23*, bbab587. [CrossRef] [PubMed]
128. Liu, X.; Luo, Y.; Li, P.; Song, S.; Peng, J. Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. *PLoS Comput. Biol.* **2021**, *17*, e1009284. [CrossRef] [PubMed]
129. Xiang, Z.; Gong, W.; Li, Z.; Yang, X.; Wang, J.; Wang, H. Predicting Protein–Protein Interactions via Gated Graph Attention Signed Network. *Biomolecules* **2021**, *11*, 799. [CrossRef] [PubMed]
130. Mahbub, S.; Bayzid, M.S. EGRET: Edge aggregated graph attention networks and transfer learning improve protein–protein interaction site prediction. *Brief. Bioinform.* **2022**, *23*, bbab578. [CrossRef]
131. Yuan, Q.; Chen, J.; Zhao, H.; Zhou, Y.; Yang, Y. Structure-aware protein-protein interaction site prediction using deep graph convolutional network. *Bioinformatics* **2021**, *38*, 125–132. [CrossRef]
132. Réau, M.; Renaud, N.; Xue, L.C.; Bonvin, A.M.J.J. DeepRank-GNN: A Graph Neural Network Framework to Learn Patterns in Protein-Protein Interfaces. *bioRxiv* **2021**. [CrossRef]
133. Kang, Y.; Leng, D.; Guo, J.; Pan, L. Sequence-based deep learning antibody design for in silico antibody affinity maturation. *arXiv* **2021**, arXiv:2103.03724.
134. Renz, P.; Van Rompaey, D.; Wegner, J.K.; Hochreiter, S.; Klambauer, G. On failure modes in molecule generation and optimization. *Drug Discov. Today. Technol.* **2019**, *32–33*, 55–63. [CrossRef] [PubMed]
135. Raybould, M.I.; Marks, C.; Krawczyk, K.; Taddese, B.; Nowak, J.; Lewis, A.P.; Bujotzek, A.; Shi, J.; Deane, C.M. Five computational developability guidelines for therapeutic antibody profiling. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4025–4030. [CrossRef]
136. Jin, W. Structured Refinement Network for Antibody Design. 2022. Available online: [https://www.youtube.com/watch?v=uDTcdbg\\_Ai4&list=PL27Hzl3ugX\\_okAYK-HmUJ8wHEVS1n\\_5u&index=1&t=1035s&ab\\_channel=ValenceDiscovery](https://www.youtube.com/watch?v=uDTcdbg_Ai4&list=PL27Hzl3ugX_okAYK-HmUJ8wHEVS1n_5u&index=1&t=1035s&ab_channel=ValenceDiscovery) (accessed on 5 February 2024).
137. Myung, Y.; Pires, D.E.V.; Ascher, D.B. CSM-AB: Graph-based antibody-antigen binding affinity prediction and docking scoring function. *Bioinformatics* **2021**, *38*, 1141–1143. [CrossRef]
138. Julie Josse, N.P.; Scornet, E.; Varoquaux, G. On the consistency of supervised learning with missing values. *arXiv* **2020**, arXiv:1902.06931.
139. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. *arXiv* **2018**, arXiv:1802.05365.
140. Villegas-Morcillo, A.; Makrodimitris, S.; van Ham, R.C.; Gomez, A.M.; Sanchez, V.; Reinders, M.J. Unsupervised protein embeddings outperform hand-crafted sequence and structure features at predicting molecular function. *Bioinformatics* **2021**, *37*, 162–170. [CrossRef]

141. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
142. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **1968**, *65*, 44–45. [[CrossRef](#)]
143. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]
144. Wu, R.; Ding, F.; Wang, R.; Shen, R.; Zhang, X.; Luo, S.; Su, C.; Wu, Z.; Xie, Q.; Berger, B.; et al. High-resolution de novo structure prediction from primary sequence. *bioRxiv* **2022**. [[CrossRef](#)]
145. Lima, W.C.; Gasteiger, E.; Marcatili, P.; Duek, P.; Bairoch, A.; Cosson, P. The ABCD database: A repository for chemically defined antibodies. *Nucleic Acids Res.* **2020**, *48*, D261–D264. [[CrossRef](#)] [[PubMed](#)]
146. Raybould, M.I.J.; Kovaltsuk, A.; Marks, C.; Deane, C.M. CoV-AbDab: The coronavirus antibody database. *Bioinformatics* **2021**, *37*, 734–735. [[CrossRef](#)] [[PubMed](#)]
147. Corrie, B.D.; Marthandan, N.; Zimonja, B.; Jaglale, J.; Zhou, Y.; Barr, E.; Knoetze, N.; Breden, F.M.; Christley, S.; Scott, J.K.; et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **2018**, *284*, 24–41. [[CrossRef](#)] [[PubMed](#)]
148. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; Del-Toro, N.; et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [[CrossRef](#)]
149. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [[CrossRef](#)]
150. Adam, D. The pandemic's true death toll: Millions more than official counts. *Nature* **2022**, *601*, 312–315. [[CrossRef](#)]
151. Shi, Z.; Li, X.; Wang, L.; Sun, Z.; Zhang, H.; Chen, X.; Cui, Q.; Qiao, H.; Lan, Z.; Zhang, X.; et al. Structural basis of nanobodies neutralizing SARS-CoV-2 variants. *Structure* **2022**, *30*, 707–720.e5. [[CrossRef](#)]
152. Yin, W.; Xu, Y.; Xu, P.; Cao, X.; Wu, C.; Gu, C.; He, X.; Wang, X.; Huang, S.; Yuan, Q.; et al. Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. *Science* **2022**, *375*, 1048–1053. [[CrossRef](#)]
153. Zhang, X.W.; Yap, Y.L. The 3D structure analysis of SARS-CoV S1 protein reveals a link to influenza virus neuraminidase and implications for drug and antibody discovery. *Theochem* **2004**, *681*, 137–141. [[CrossRef](#)]
154. Chaouat, A.E.; Achdout, H.; Kol, I.; Berhani, O.; Roi, G.; Vitner, E.B.; Melamed, S.; Politi, B.; Zahavy, E.; Brizic, I.; et al. SARS-CoV-2 receptor binding domain fusion protein efficiently neutralizes virus infection. *PLoS Pathog.* **2021**, *17*, e1010175. [[CrossRef](#)] [[PubMed](#)]
155. Narkhede, Y.B.; Gonzalez, K.J.; Strauch, E.-M. Targeting Viral Surface Proteins through Structure-Based Design. *Viruses* **2021**, *13*, 1320. [[CrossRef](#)] [[PubMed](#)]
156. Marcandalli, J.; Fiala, B.; Ols, S.; Perotti, M.; de van der Schueren, W.; Snijder, J.; Hodge, E.; Benhaim, M.; Ravichandran, R.; Carter, L.; et al. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **2019**, *176*, 1420–1431.e17. [[CrossRef](#)] [[PubMed](#)]
157. Pan, Y.; Du, J.; Liu, J.; Wu, H.; Gui, F.; Zhang, N.; Deng, X.; Song, G.; Li, Y.; Lu, J.; et al. Screening of potent neutralizing antibodies against SARS-CoV-2 using convalescent patients-derived phage-display libraries. *Cell Discov.* **2021**, *7*, 57. [[CrossRef](#)] [[PubMed](#)]
158. Yuan, T.Z.; Garg, P.; Wang, L.; Willis, J.R.; Kwan, E.; Hernandez, A.G.L.; Tuscano, E.; Sever, E.N.; Keane, E.; Soto, C.; et al. Rapid discovery of diverse neutralizing SARS-CoV-2 antibodies from large-scale synthetic phage libraries. *mAbs* **2022**, *14*, 2002236. [[CrossRef](#)] [[PubMed](#)]
159. Shiakolas, A.R.; Kramer, K.J.; Johnson, N.V.; Wall, S.C.; Suryadevara, N.; Wrapp, D.; Periasamy, S.; Pilewski, K.A.; Raju, N.; Nargi, R.; et al. Efficient discovery of SARS-CoV-2-neutralizing antibodies via B cell receptor sequencing and ligand blocking. *Nat. Biotechnol.* **2022**, *40*, 1270–1275. [[CrossRef](#)] [[PubMed](#)]
160. Abubaker Bagabir, S.; Ibrahim, N.K.; Abubaker Bagabir, H.; Hashem Ateeq, R. COVID-19 and Artificial Intelligence: Genome sequencing, drug development and vaccine discovery. *J. Infect. Public Health* **2022**, *15*, 289–296. [[CrossRef](#)] [[PubMed](#)]
161. Lopez-Rincon, A.; Tonda, A.; Mendoza-Maldonado, L.; Mulders, D.G.; Molenkamp, R.; Perez-Romero, C.A.; Claassen, E.; Garssen, J.; Kraneveld, A.D. Classification and specific primer design for accurate detection of SARS-CoV-2 using deep learning. *Sci. Rep.* **2021**, *11*, 947. [[CrossRef](#)]
162. Zeng, X.; Song, X.; Ma, T.; Pan, X.; Zhou, Y.; Hou, Y.; Zhang, Z.; Li, K.; Karypis, G.; Cheng, F. Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning. *J. Proteome Res.* **2020**, *19*, 4624–4636. [[CrossRef](#)]
163. Wang, B.; Jin, S.; Yan, Q.; Xu, H.; Luo, C.; Wei, L.; Zhao, W.; Hou, X.; Ma, W.; Xu, Z.; et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system. *Appl. Soft Comput.* **2021**, *98*, 106897. [[CrossRef](#)] [[PubMed](#)]
164. Chen, J.; Gao, K.; Wang, R.; Nguyen, D.D.; Wei, G.-W. Review of COVID-19 Antibody Therapies. *Annu. Rev. Biophys.* **2021**, *50*, 1–30. [[CrossRef](#)] [[PubMed](#)]
165. Darmawan, J.T.; Leu, J.-S.; Avian, C.; Ratnasari, N.R.P. MITNet: A fusion transformer and convolutional neural network architecture approach for T-cell epitope prediction. *Brief. Bioinform.* **2023**, *24*, bbad202. [[CrossRef](#)] [[PubMed](#)]

166. Bukhari, S.N.H.; Jain, A.; Haq, E.; Mehbodniya, A.; Webber, J. Machine Learning Techniques for the Prediction of B-Cell and T-Cell Epitopes as Potential Vaccine Targets with a Specific Focus on SARS-CoV-2 Pathogen: A Review. *Pathogens* **2022**, *11*, 146. [CrossRef] [PubMed]
167. Liu, Z.; Jin, J.; Cui, Y.; Xiong, Z.; Nasiri, A.; Zhao, Y.; Hu, J. DeepSeqPanII: An interpretable recurrent neural network model with attention mechanism for peptide-HLA class II binding prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 2188–2196. [CrossRef]
168. Hess, M.; Keul, F.; Goesele, M.; Hamacher, K. Addressing inaccuracies in BLOSUM computation improves homology search performance. *BMC Bioinform.* **2016**, *17*, 189. [CrossRef]
169. Nielsen, M.; Lundegaard, C.; Blicher, T.; Peters, B.; Sette, A.; Justesen, S.; Buus, S.; Lund, O. Quantitative Predictions of Peptide Binding to Any HLA-DR Molecule of Known Sequence: NetMHCIIpan. *PLoS Comput. Biol.* **2008**, *4*, e1000107. [CrossRef]
170. Saha, S.; Raghava, G.P.S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **2006**, *65*, 40–48. [CrossRef]
171. Kanyavuz, A.; Marey-Jarossay, A.; Lacroix-Desmazes, S.; Dimitrov, J.D. Breaking the law: Unconventional strategies for antibody diversification. *Nature reviews. Immunology* **2019**, *19*, 355–368.
172. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [CrossRef] [PubMed]
173. 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. [CrossRef] [PubMed]
174. Leaver-Fay, A.; Tyka, M.; Lewis, S.M.; Lange, O.F.; Thompson, J.; Jacak, R.; Kaufman, K.W.; Renfrew, P.D.; Smith, C.A.; Sheffler, W.; et al. Chapter nineteen—Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Computer Methods, Part C*; Johnson, M.L., Brand, L., Eds.; Academic Press: Hoboken, NJ, USA, 2011; Volume 487, pp. 545–574.
175. Leaver-Fay, A.; Froning, K.J.; Atwell, S.; Aldaz, H.; Pustilnik, A.; Lu, F.; Huang, F.; Yuan, R.; Hassanali, S.; Chamberlain, A.K.; et al. Computationally Designed Bispecific Antibodies using Negative State Repertoires. *Structure* **2016**, *24*, 641–651. [CrossRef] [PubMed]
176. Lewis, S.M.; Wu, X.; Pustilnik, A.; Sereno, A.; Huang, F.; Rick, H.L.; Guntas, G.; Leaver-Fay, A.; Smith, E.M.; Ho, C.; et al. Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface. *Nat. Biotechnol.* **2014**, *32*, 191–198. [CrossRef]
177. Miklos, A.E.; Kluwe, C.; Der, B.S.; Pai, S.; Sircar, A.; Hughes, R.A.; Berrondo, M.; Xu, J.; Codrea, V.; Buckley, P.E.; et al. Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* **2012**, *19*, 449–455. [CrossRef] [PubMed]
178. Kim, D.N.; Jacobs, T.M.; Kuhlman, B. Boosting protein stability with the computational design of  $\beta$ -sheet surfaces. *Protein Sci.* **2016**, *25*, 702–710. [CrossRef]
179. Harmalkar, A.; Rao, R.; Richard Xie, Y.; Honer, J.; Deisting, W.; Anlahr, J.; Hoenig, A.; Czwikla, J.; Sienz-Widmann, E.; Rau, D.; et al. Toward generalizable prediction of antibody thermostability using machine learning on sequence and structure features. *mAbs* **2023**, *15*, 2163584. [CrossRef]
180. Liang, T.; Jiang, C.; Yuan, J.; Othman, Y.; Xie, X.Q.; Feng, Z. Differential performance of RoseTTAFold in antibody modeling. *Brief. Bioinform.* **2022**, *23*, bbac152. [CrossRef]
181. Fernández-Quintero, M.L.; Kraml, J.; Georges, G.; Liedl, K.R. CDR-H3 loop ensemble in solution—Conformational selection upon antibody binding. *mAbs* **2019**, *11*, 1077–1088. [CrossRef]
182. Gainza, P.; Sverrisson, F.; Monti, F.; Rodola, E.; Boscaini, D.; Bronstein, M.M.; Correia, B.E. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **2020**, *17*, 184–192. [CrossRef]
183. Guo, L.; He, J.; Lin, P.; Huang, S.-Y.; Wang, J. TRScore: A three-dimensional RepVGG-based scoring method for ranking protein docking models. *Bioinformatics* **2022**, *38*, 2444–2451. [CrossRef]
184. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
185. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165.
186. Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C.L.; Ma, J.; et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2016239118. [CrossRef] [PubMed]
187. Goddard, T.D.; Huang, C.C.; Meng, E.C.; Pettersen, E.F.; Couch, G.S.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27*, 14–25. [CrossRef] [PubMed]
188. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [CrossRef] [PubMed]
189. Temitope Sobodu. How to Deploy and Interpret AlphaFold2 with Minimal Compute. 2023. Available online: <https://towardsdatascience.com/how-to-deploy-and-interpret-alphafold2-with-minimal-compute-9bf75942c6d7> (accessed on 5 February 2024).

190. Yin, R.; Feng, B.Y.; Varshney, A.; Pierce, B.G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **2022**, *31*, e4379. [[CrossRef](#)] [[PubMed](#)]
191. O'Reilly, F.J.; Graziadei, A.; Forbrig, C.; Bremenkamp, R.; Charles, K.; Lenz, S.; Elfmann, C.; Fischer, L.; Stülke, J.; Rappsilber, J. Protein complexes in cells by AI-assisted structural proteomics. *Mol. Syst. Biol.* **2023**, *19*, e11544. [[CrossRef](#)]
192. *The PyMOL Molecular Graphics System, Version 1.8*; Schrödinger, LLC: New York, NY, USA, 2016.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.