Seminars in Cancer Biology 95 (2023) 13–24

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Accelerating antibody discovery and design with artificial intelligence: Recent advances and prospects

Ganggang Bai [a,1], Chuance Sun [a,1], Ziang Guo [b,1], Yangjing Wang [a], Xincheng Zeng [a], Yuhong Su [a], Qi Zhao [b,c,*], Buyong Ma [a,d,**]

[a] Engineering Research Center of Cell & Therapeutic Antibody (MOE), School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Cancer Center, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Taipa, Macao Special Administrative Region of China
[c] MoE Frontiers Science Center for Precision Oncology, University of Macau, Taipa, Macao Special Administrative Region of China
[d] Shanghai Digiwiser BioTechnolgy, Limited, Shanghai 201203, China

ARTICLE INFO

Keywords:
Antibody
Artificial intelligence
Antibody design
Machine learning , therapeutic

ABSTRACT

Therapeutic antibodies are the largest class of biotherapeutics and have been successful in treating human diseases. However, the design and discovery of antibody drugs remains challenging and time-consuming. Recently, artificial intelligence technology has had an incredible impact on antibody design and discovery, resulting in significant advances in antibody discovery, optimization, and developability. This review summarizes major machine learning (ML) methods and their applications for computational predictors of antibody structure and antigen interface/interaction, as well as the evaluation of antibody developability. Additionally, this review addresses the current status of ML-based therapeutic antibodies under preclinical and clinical phases. While many challenges remain, ML may offer a new therapeutic option for the future direction of fully computational antibody design.

1. Introduction

Monoclonal antibodies (mAbs) have become the leading products in the biopharmaceutical market over the past twenty years [1]. The mAbs have widely been used in treating human cancer, autoimmune diseases and infectious diseases [2–4]. The successful examples are anti-programmed death-1 (PD-1) and anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) mAbs that have significantly improved the survival of advanced non-small cell lung cancer (NSCLC) [5,6]. Additionally, multiple antibody-based therapeutic modalities, such as bispecific antibodies [7,8], antibody-drug conjugates [9,10], chimeric antigen receptors (CARs) [11,12], have shown potential in cancer treatment. However, there are many challenges in designing effective antibody drugs, including antibody discovery, optimization, developability, and antibody delivery to certain organs [13,14]. Traditional computational strategies have been used to optimize antibody affinity and bioactivity through energy calculation in combination with display library selection [15–17], but these approaches are often limited by the low reliability of free energy estimation and high experimental time-cost. Therefore, there is a need to develop novel in silico techniques for antibody discovery and engineering.

Artificial intelligence (AI) techniques have already been widely used in various directions in biomedical developments[18–21]. Traditional machine learning (ML) and deep learning (DL) have played decisive roles in aiding to evaluate the efficacy of anti-PD-1 antibody immunotherapy for lymphoma [22], and 3D computational modeling and ML were used to provide accurate predictions for optimizing drug treatments of different anti-angiogenic drugs on the treatment of solid tumors [23].

One of the major interests of AI in the biological pharmaceutical industry is the antibody design in the discovery, optimization, and developability evaluation, where the recent developments in ML and DL have opened up much broader avenues for therapeutic antibody design with implications for fully computational antibody design [24]. ML

* Corresponding author at: Cancer Center, Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Taipa, Macao Special Administrative Region of China.
** Corresponding author at: Engineering Research Center of Cell & Therapeutic Antibody (MOE), School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail addresses: qizhao@um.edu.mo (Q. Zhao), mabuyong@sjtu.edu.cn (B. Ma).
[1] Ganggang Bai, Chuance Sun, and Ziang Guo are co-first authors

https://doi.org/10.1016/j.semcancer.2023.06.005
Received 23 March 2023; Received in revised form 9 June 2023; Accepted 18 June 2023
Available online 22 June 2023

models can predict antibody affinity and non-specific binding using the dataset of antibody libraries for high and low levels of affinity and non-specific binding, enabling the co-optimization of therapeutic antibody affinity and specificity to accelerate the development of highly potent antibody drugs[25].

This article introduces AI approaches commonly used in antibody design and critically reviews the major applications of AI in antibody design for discovery, optimization, and developability evaluation. In antibody design, important design parameters include paratope, epitope, affinity, and developability, all of which should be included and optimized simultaneously. In the following reviews, we analyze each parameter in individual sections to provide readers with a better understanding.

2. Overview of machine learning methods for antibody design

After decades of vigorous development, the meaning of AI has continued to expand, and it has become a general term encompassing artificial neural networks, ML, DL and other technologies [26–28].

Traditional ML primarily focuses on how to learn a prediction model. Data is typically represented as a set of features, which can take the form of continuous numerical values, discrete symbols, or other formats. These features are then fed into the prediction model, which outputs the prediction results. This type of ML can be categorized as shallow learning, as it does not involve feature learning, and its features are mainly extracted using artificial experience or feature conversion methods. The ML method can be roughly divided into three basic elements: model, learning criteria, and optimization algorithm. There are generally three main types of ML: supervised learning, semi-supervised learning, and unsupervised learning. Supervised learning generates a function through some existing input data and output data to map the input to the appropriate output. Unsupervised learning divides unlabeled sample sets into several categories by using the internal relations between data. Semi-supervised learning uses both labeled and unlabeled data to generate appropriate functions. Support Vector Machine (SVM) [29] and Random Forest (RF) [30] are commonly used ML models. SVM is a generalized linear classifier that performs binary classification of data using supervised learning. Its decision boundary is a maximum-margin hyperplane for learning samples. For example, Daberdaku and Ferrari developed SVM for antibody interface prediction [31]. Their method outperformed other antigen-binding interface prediction software including Paratome[32], Antibody i-Patch[33] and Parapred[34]. RF is an integrated algorithm (ensemble learning) that combines multiple weak classifiers to produce a final result that is voted or averaged. This approach results in high accuracy and generalization performance for the overall model. For example, Griffiths et al. used random forest predicting antibody properties[35]. The Random Forest Tree showed the best performance as compared to Logistic regression, K-Nearest Neighbor (KNN), SVM and Naïve Bayesian for membrane and soluble proteins, respectively.

DL has been widely used in many scientific fields over the last decade. In 2020, there was an unprecedented breakthrough in DL due to the AlphaFold2 artificial intelligence system developed by Google's DeepMind team, which achieved remarkable accuracy in the International Protein Structure Prediction Competition (CASP)[36]. Most of the prediction models produced were highly consistent with experimentally measured protein structure models, which has drawn worldwide attention and made AI technology highly anticipated by scientific researchers.

Several common DL frameworks have been developed to facilitate the construction of DL models, such as Caffe [37], TensorFlow [38], Pytorch [39], Keras, MXNet [40], CNTK [41] (Table 1).

Various deep neural network algorithms have been developed to process protein-related AI models (Fig. 1), and new algorithms are being proposed continuously. Common DL models include fully connected (FC) network structure, convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory network (LSTM), autoencoder (AE), and graph neural network (GNN), among others. Each neural network has its own strengths and is suitable for different tasks.

Table1
Comparison of different deep learning frameworks.

Framework	Release time and Developers	Git star	Underlying language	Interface language
Caffe	2014 BVLC	32900 +	C+ +	C+ +/Python/ Matlab
TensorFlow	2015 Google	168000 +	C+ +/Python	C+ +/Python/ Java
Pytorch	2017 Facebook	59400 +	C/C+ +/ Python	Python
Keras	2015 Google	56300 +	Python	Python
MXNet	2016 DMLC	20100 +	C+ +	C+ +/Python/R
CNTK	2016 Microsoft	17200 +	C+ +	C+ +/Python/C/ Java

CNN is a multi-layer feedforward network that excels in image processing and recognition, especially for large images [42]. It consists of a convolutional layer, a pooling layer, and a fully connected layer. The convolution layer and pooling layer work together to form multiple convolution groups, extracting features layer by layer, and finally completing classification through several fully connected layers. The advantage of CNN is that it can extract dominant features from large dimension. For example , in order to get information from both sequential and spatial neighbors to understand more about the local environment of target amino acid residue in antibody-antigen complex, Lu et al. combined CNN and graph convolution networks (GCNs, see next paragraph) which describe special connections among contacting residues. Here CNN may provide global features hidden in local environment. [43].

RNN is designed to describe the relationship between current output and previous input information. Each neuron performs the same task and is often used in machine translation, speech recognition, and text similarity tasks [44]. This feature fits amino acid sequence based prediction well, as did in Parapred, a sequence-based probabilistic machine learning algorithm for paratope prediction [34]. Parapred used a recurrent neural network architecture to leverage features from both local residue neighbourhoods and across the entire sequence.

LSTM is a temporal recurrent neural network that is suitable for processing and predicting important events with long intervals and delays in time series, such as predicting diseases, click-through rates, and stocks [45]. Toshiaki et al. proposed a deep learning method based on long short-term memory with an attention mechanism to consider the characteristics of a whole antigen protein in addition to the target sequence. The proposed method achieves better accuracy compared with the conventional method in the experimental prediction of epitope regions using the data from the immune epitope database [46].

AE is mainly used for data compression by capturing the most critical factors that can represent the input information. It approximates the input content with the output content, preserving the essential characteristics of the input [47]. Friedens et al. employed a deep learning approach utilizing variational autoencoders (VAEs) to model the underlying process of B cell receptor (BCR) recombination and assume that the data generation follows a Gaussian mixture model (GMM) in latent space. This provides both a latent embedding and cluster labels that group similar sequences, thus enabling the discovery of a multitude of convergent, antigen-associated sequence patterns. Their work highlights to which extent convergence in antibody repertoires can occur and shows how deep learning can be applied for immunodiagnostics and

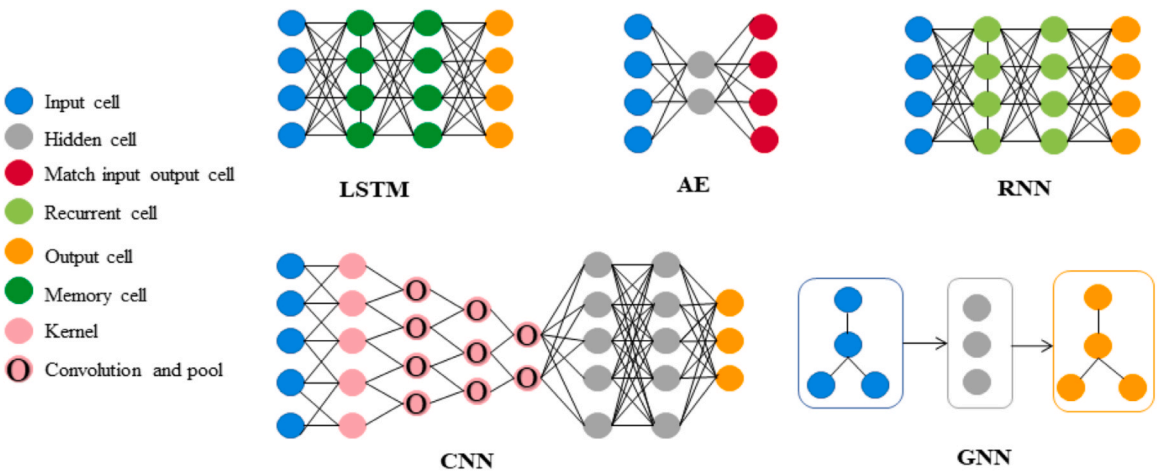

Fig. 1. Schematic diagram showing various neural network architectures.

antibody discovery and engineering [48].

GNN is a framework for learning graph structure data. By formulating certain strategies on the nodes and edges of the graph, GNN transforms graph structure data into a standard representation, which can be input into a variety of different neural networks for training [49]. It achieves excellent results in tasks such as node classification, edge information dissemination, and graph clustering. GNN is widely used in structure based protein interaction predictions, including antibody-antigen interaction[43]. It is natural to see GNN has been used in antibody sequence and structure co-design [50].

3. Antibody structure prediction and design

Traditionally, homology modeling and energy-based methods have been the popular solutions for predicting protein structures. In protein structure design, one typically samples from known conformations or protein fragments, generates structures based on experimental databases (such as PDB), mutates some residues to obtain a large number of candidate antibodies, and finally uses a scoring function to select the most promising protein or antibody structure. The protein folding principles learned from structures in the PDB database still guide the design process, and are useful for generating new protein structures by assembling continuous or discontinuous three-dimensional elements, as well as for the development and optimization of design energy functions used to rank candidate designs [51,52].

Recent advances in DL have led to the development of new methods for de novo protein design [53,54]. In 2022, Liu's group combined SCUBA-driven stochastic dynamics (SD) simulations with the data-driven fixed-backbone amino acid sequence selection program ABACUS2 to design de novo proteins with new topological architectures and sequences that meet various design specifications [55]. In this model, a new statistical learning strategy is used, and the computational neural network is trained on more than 12,000 non-redundant natural PDB structures to guide the optimization of the main chain structure. The resulting high-dimensional correlation in the designable main chain structure can be described with high fidelity. Experimental validation of the crystal structures of nine de novo proteins demonstrated that four of them had novel, non-natural overall architectures. These results suggest that SCUBA+ABACUS2, alongside the state-of-the-art RosettaDesign method [56], is a useful method for de novo protein design. These approaches enable substantial expansion of the structural diversity and functionalities accessible to de novo protein design. More details on the development and typical design of de novo protein design can be found in a series of reviews by Baker's group [57], DeGrado's group[58], Woolfson's group[59] and Kortemme's group[60].

Antibody structure prediction is a special sub-area of protein structure prediction. Antibodies belong to the immunoglobulin protein superfamily (Igs). A basic immunoglobulin unit (Ig) monomer (150 kDa) has a similar "Y" shape, consisting of two light and two heavy chains joined by disulfide connections. Each light chain comprises two regions, each of which is made up of one variable region (VL) and one constant region (CL), while each heavy chain has one variable domain (VH) and three constant domains (CH1–3) [61]. The two antigen-binding fragments (Fabs) are responsible for binding to the specific molecular target with high avidity, while the Fc region connects to immune receptors to activate effector actions. The variable sequences locate in the N-terminal half of the Fab arms and vary between antibodies to give them unique specificities. Each chain has three complementarity-determining region (CDR) loops that contain hypervariable sequences near the antigen-binding interface [62]. The length of these CDR loops has a significant impact on the nature of antigen binding. Among these six CDR regions, the CDR3 of the heavy chain has the largest diversity, with lengths varying from 4 to more than 20 residues. However, a survey of 137 clinical-stage antibody therapeutics (CSTs) has shown that CST CDRH3 loops had a shorter median length compared to that of human VdH Ig sequences [63] (Fig. 2).

Computational antibody design relies on accurate structural models of both the antibody and target antigen [64] [65]. However, predicting the structure of antibody CDRs, especially CDR-H3, remains challenging due to their highly diverse conformations. Several ab initio protocols, such as OptCDR[66], OptMAVEn[67], RosettaAntibodyDesign (RAbD) [68] and AbDesign[69], have been developed to design novel paratopes through four sequential steps: CDR generation, structure modelling, antibody-antigen docking, and binding energy evaluation. The first step is always the redesign of CDRs to enhance antibody stability and affinity by optimizing conformational and free energy changes in specific residues. RosettaAntibodyDesign (RAbD) can perform de novo antibody design from a nonbinding antibody and affinity maturation of an existing antibody [68]. It classifies the antibody into regions, including framework, the five canonical loops, and the HCDR3 loop, following the methodology in RAbD [70]. Recent progress using libraries of antibody loop conformations from the PDB has shown promise in creating stable antibodies with defined loop structure. Baran et al. developed a set of algorithms based on the skeleton structure and conservative sequence information of natural antibodies, which can design antibody sequences different from natural antibody sequences but with similar affinity and stability when combined with the Rosetta platform and experimental methods [71].

The exceptional performance of several leading AI algorithms such as AlphaFold2 or RoseTTAFold in predicting protein structures provides new approaches for generating protein structures [72–75]. However, accurately modeling the structure of antibody variable regions,

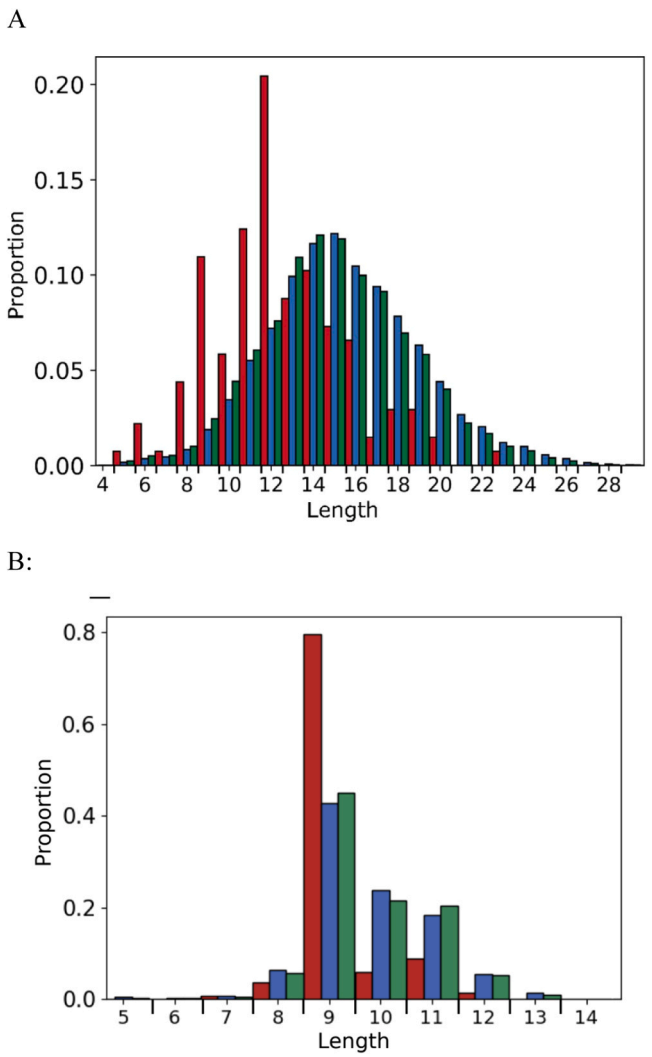

Fig. 2. Comparison of the length distributions of CDRH3 (A) and CDRL3 (B) in 137 clinical-stage antibody therapeutics (CSTs in red), 105,458 non-redundant human VdH Ig-seq CDR3s (in blue), and 551,193 non-redundant human VdH Ig-seq heavy chains (in green)
Figures are reproduced from reference [63] , Copyright (2019) National Academy of Sciences.

particularly CDR-H3, remains a challenge due to highly diverged sequences and a lack of evolution pattern. Recent progress has proposed some antibody design protocols to address these challenges, including DeepH3[76], DeepAb[77], ABlooper [78] and Nanonet[79], demonstrating the power of ML techniques in antibody structural predictions.

DeepH3 is a deep residual neural network trained to identify near-native CDR-H3 loops [76]. The training dataset for DeepH3 was extracted from SAbDab [80]. The output of DeepH3 was converted to geometric potentials and used to score poses generated by RAbD [81]. Compared to the standard Rosetta energy function, DeepH3 improved the average RMSD of prediction by 32.1% (1.4 Å). This work also revealed that inter-residue orientations are more effective discriminators for scoring CDR-H3 structures than distances.

Although DeepH3 can predict CDR-H3 with high accuracy , rebuilding the entire region of the antibody variable is the ultimate goal. DeepSCAb (deep side-chain antibody) is a deep neural network that focuses on predicting both the backbone and side-chain conformations of an antibody variable fragment simultaneously and employs Rosetta to predict the full Fv structure [82]. The network comprises two modules: an inter-residue module for predicting backbone geometries like DeepH3 and a rotamer module for predicting side-chain dihedrals. The network required the antibody sequence as the input. After the input passes through two ResNet modules, the resulting tensors are converted to pairwise probability distributions over Cβ distance, d, the orientation dihedrals ω and θ, and the planar angle ϕ. The rotamer module obtains features from the inter-residue module, updates the d, ω, θ, and ϕ outputs, and adds them back into the inter-residue module. These rotamer probability distributions generated by DeepSCAb were utilized with Rosetta for structure realization and side-chain packing to predict full Fv structure. This method proves to be robust when the backbone is perturbed or deviates from the crystal structure.

As mentioned above, DeepSCAb, and DeepH3 use inter-residue geometric potentials and adopt the energy minimization method to produce the final structure. In contrast, AbLooper utilizes end-to-end Equivariant Graph Neural Networks to predict CDR (all loops) structures [83]. Antibodies extracted from SAbDab were encoded into 41-dimensional vectors, including amino acid type, atom type, and attribution of the variable region. AbLooper employed L1-loss, which comprises five terms, distances about the true and predicted inter-atom to encourage the conservation of distances between neighboring atoms in the backbone chain. The output from the five networks was then averaged to obtain a final prediction. The advantage of AbLooper is its ability to produce antibody models with high accuracy and less time-consuming.

Pre-trained language model was demonstrated to be very effective for inference of full atomic-level protein structure[84]. Therefore, several antibody structure prediction methods have been developed in combination with antibody pre-trained language models. DeepAb is a method for reconstructing the entire antibody variable region using antibody pre-trained language model with recurrent neural network [77]. The input immunoglobulin sequences are represented using an RNN encoder-decoder model consisting of two LSTMs trained on a set of 118,386 paired heavy and light chain sequences from the observed antibody space (OAS) [85]. The general structure prediction architecture of DeepAb was based on DeepH3 with the addition of interpretable attention layers. On benchmarks of challenging, therapeutically important targets, DeepAb consistently generated more precise structures compared to alternatives based on grafting. Additionally, this approach offers an explanation for predictions of a CDR H3 loop and reveals several interactions with adjacent residues considered crucial for the structure. Inspired by the success of the DeepAb, Ruffolo developed IgFold for predicting antibody structures, using a pre-trained language model trained on 558 M natural antibody sequences followed by graph networks that directly predict backbone atom coordinates [86]. The method incorporates AntiBERTy, a transformer language model that generates embeddings for structure prediction [87]. Using these sequence embeddings, IgFold employs a series of transformer layers to directly predict atomic coordinates for protein backbone atoms. For each residue, IgFold also provides an estimate of prediction quality. The refinement of predictions and addition of side chains is done by Rosetta. This method stands out for its high speed, accuracy, and support for nanobody modeling.

Nanobodies (Nb) have gained increasing attention as an important and popular therapeutic format [88]. However, due to the lack of a second immunoglobulin chain, the prediction of nanobodies differs slightly from paired antibodies. NanoNet is designed to directly produce the 3D coordinates of the backbone and Cβ atoms of the entire VH domain of nanobodies from a given sequence [89]. This method employs a CNN consisting of two 1D Residual Neural Networks. The variable domain of the training set, which consists of ~ 2,000 heavy chains of mAbs and nanobodies structures, is aligned using the MultiProt algorithm with order-dependence and a distance threshold of 1.4 Å to enable the network to directly learn the VH domain 3D structure [90]. Since the structure prediction is accomplished within a single CNN network without structure relaxation, NanoNet can produce an Nb structure in just a few milliseconds, enabling accurate modeling of entire antibody

repertoires from Next-generation sequencing experiments.

Brennan developed ImmuneBuilder[91] to predict the structures of immune proteins with deep-learning method. ImmuneBuilder contained three deep learning models: ABodyBuilder2, NanoBodyBuilder2 and TCRBuilder2 which were trained to predict the structure of antibodies, nanobodies and T-Cell receptors. The ABodyBuilder2 was developed based on AlphaFold-Multimer and composed of four deep-learning models trained independently to predict antibody structures. The final prediction was the closest one to the average of the four predicted structures and utilized OpenMM to obtain the refined antibody structure. On a benchmark of 34 recently solved antibodies, ABodyBuilder2 predicted CDR-H3 loops with lower RMSD than ABlooper and IgFold.

In recent years, there has been growing interest in co-designs of antibody sequence and structure. Generative ML has emerged as a major driver in the computational design of antigen-specific mAbs. For instance, a lattice-based antibody-antigen binding simulation framework incorporates various physiological antibody-binding parameters. This deep generative model, trained exclusively on antibody sequence data, can be used to design three-dimensional conformational epitope-specific antibodies that match or even surpass the training dataset in terms of affinity and developability parameter value variety [92].

4. Antibody-antigen interface/interaction prediction

Predicting the antibody-antigen binding interface and binding affinity are crucial challenges in ML for antibody-antigen interactions. These tasks are also essential in antibody design, as they allow for the prediction of the paratope, epitope, and paratope-epitope interactions [93].

4.1. Antigen epitope prediction

Antigen epitope prediction can be divided into two types: those done with the presence of antibodies and those done without. The former aims to identify the most probable epitopes of the antigen, while the latter aims to identify the epitope that a known antibody will bind to. Early epitope prediction methods used propensity scales like BcePred [94], ABCPred[95], and iBCE-EL[96] to infer residues of contiguous epitopes based on a few hundred linear epitopes. iBCE-EL an ensemble method that combined extremely randomized tree and gradient boosting algorithms, used a combination of amino acid composition and physicochemical properties as input features to predict the class and probability values of a given peptide. SEMA [97], an AI model, used a transfer learning approach to predict epitopes based on the primary antigen sequence and tertiary structure. It fine-tuned a pretrained protein language model, ESM-1v, and an inverse folding model, ESM-IF1, to quantitatively predict antibody-antigen interaction features and distinguish between epitope and non-epitope residues.

Linear epitopes account for about 10% of B-cell epitopes, while the remaining 90% are non-contiguous sequences and conformational [98]. As a result, some prediction methods, such as DiscoTope[99], SEPPA [100], and PEPITO[101], are trained on antibody-antigen structures and then identify the antigen structures using traditional geometric features like the number of neighbors according to different distance thresholds. BCEs, a DL prediction model [102], employed two parallel modules to extract features (local and global) from the antigen and identify B-cell epitopes. The local features of the target residue were captured using Graph Convolutional Networks (GCNs), while global information from the entire antigen sequence was extracted using Attention-Based Bidirectional Long Short-Term Memory (Att-BLSTM) networks. This method declares incorporating the global features leads to improved prediction of BCEs.

It is now widely accepted that epitopes of antigens can be located in nearly any surface accessible region recognized by an antibody [103]. Thus, an epitope can become functional or vice versa depending on the context of the antibody (a paratope). EpiPred [104] uses a combination of conformational matching of antibody-antigen structures and a specific antibody-antigen score to identify the epitope region. DLAB-Re [105] employs CNNs to improve the ranking of docks from the ZDock docking algorithm and, in combination with docking scores generated by ZDock, to predict antibody-antigen binding.

4.2. Antibody Paratope Prediction

The paratope, which is the region of the antibody that binds to the antigen, can improve the accuracy of antibody-antigen structure prediction and increase the antibody affinity by targeting mutations on paratope residues. Most paratope residues are located within the CDR loops and are spatially close to each other. Unlike epitopes, each CDR's paratope residue has its own preferential amino acid usage [106]. Parapred is a DL model that incorporates both local residue neighborhood information and overall sequence information [34]. It requires only a stretch of the amino acid sequence corresponding to a hypervariable region as input, without any information regarding the antigen. Paramatome identifies consensus antigen-binding regions through structural alignments [32]. The server applies structural consensus regions from multiple structure alignments of a reference set of antibody-antigen complexes to identify antibody binding regions. AG-Fast-Parapred addresses Parapred's limitations using self-attention convolutions [107]. This method significantly reduces computation time and moderately improves accuracy (AUC = 0.90) compared to Parapred (AUC = 0.88). PECAN is a unified DL-based framework that predicts binding interfaces on both antibodies and antigens [108]. It employs graph convolutions to aggregate properties across local regions in a protein, transfer learning to leverage this data as a prior for the specific case of antibody-antigen interactions, and an attention layer to explicitly encode the context of the partner. It demonstrated better performance in predicting epitopes by the paratope prediction networks compared to networks trained solely for epitope prediction. EPMP is a highly asymmetrical neural network for paratope (Para-EPMP) and epitope (Epi-EPMP) predictors [109]. Para-EPMP was designed to exploit paratope sequence as input features and followed by a graph structure to predict the paratope. Epi-EPMP was purely structural, and used distinct neural message passing architectures that are geared towards the specific aspects of paratope and epitope prediction and obtains significant improvements on both tasks. This method adopts separate neural message passing architectures specifically designed for paratope and epitope prediction and demonstrates improvements in both tasks.

4.3. Prediction of antibody-antigen interaction and binding affinity

AI techniques, such as CNNs and GNNs, have found broad applications in various fields such as protein structure prediction, protein function prediction, genetic engineering, systems biology, and drug design. In this context, we will focus on the application of AI in protein-protein docking and de novo protein design. Protein docking methods aim to predict the overall quaternary structure of a protein complex from the tertiary structure of individual chains. Despite considerable progress in ab initio protein docking, selecting near-native models out of a large number of decoys remains a challenging task [110]. Reau et al. have developed DeepRank, which is a user-friendly, open-source, and configurable DL framework implemented in Python3 [111] (Fig. 3). DeepRank utilizes a customizable 3D CNN pipeline to learn interaction patterns specific to protein-protein interactions (PPIs) by mapping atomic and residue-level features to 3D grids. However, CNNs have limitations in processing data that are not in Euclidean space with a very regular structure, which is the case with PPIs. DeepRank-GNN, which utilizes graph representations and graph convolutions, was proposed by Reau et al. as a solution to this issue [112]. DeepRank-GNN has been shown to outperform five different software programs used to assess docking conformations, including DeepRank, DOVE from CNN, iScore, and the classical scoring function HADDOCK, with an AUC value of 0.68,

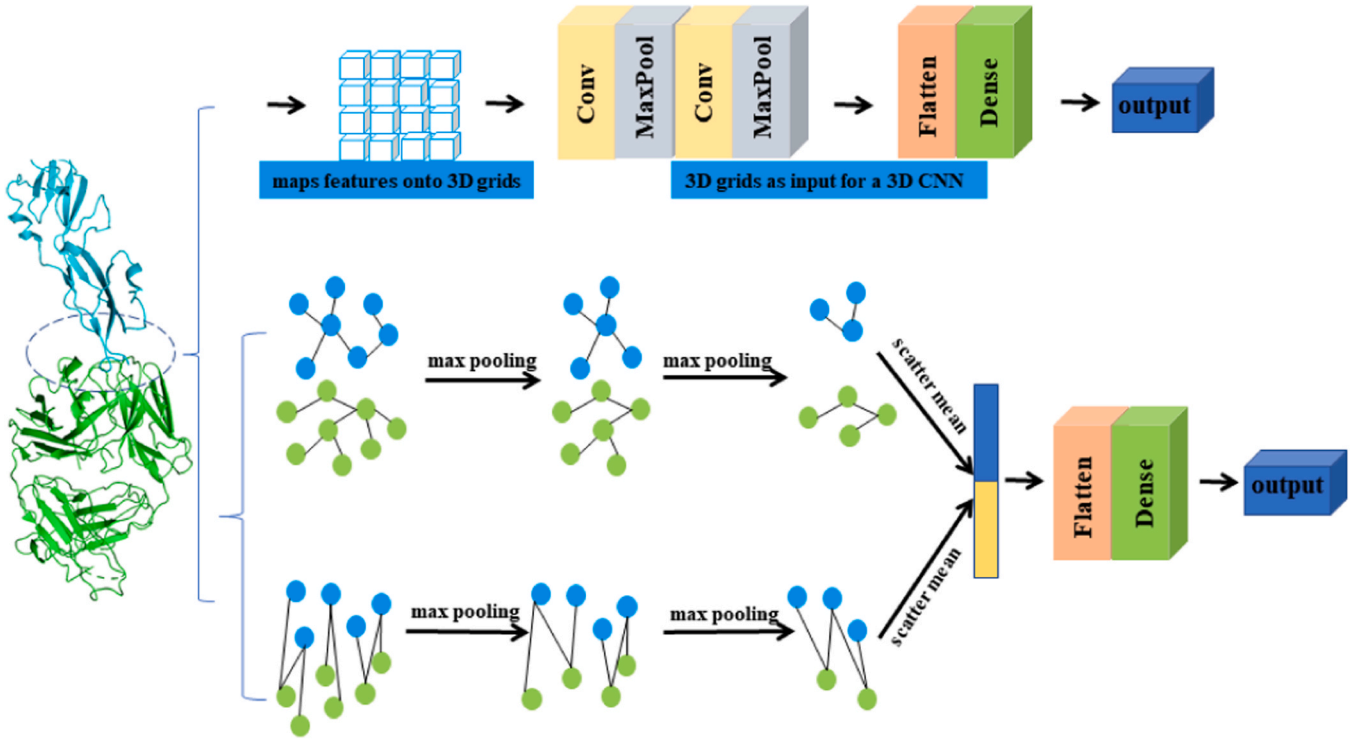

Fig. 3. Overview of the DeepRank-CNN(top) and DeepRank-GNN(bottom) frameworks.

which is the highest accuracy achieved [110].

The strength of interactions between an antibody and an antigen is referred to as the antibody binding affinity, which is an important factor in antibody research. In the past, traditional in-silico methods have been used to predict the binding affinity of antibody-antigen interactions based on the atom's chemical properties, such as polarity and charge [113]. However, with the advent of DL, new approaches to predicting antibody binding affinity have been developed and successfully applied to antibody screening. High-throughput sequencing and display technologies have enabled the generation of sufficient sequence-only data to train DL models. With these antibody sequences, DL models can explore a wide sequence space and generate novel and improved antibody sequences.

Ens-Grad is an ML method developed to design complementarity determining regions of antibodies with superior target affinities compared to candidates obtained from phage display panning experiments [114]. This two-stage method models antibody affinity with an ensemble of neural networks trained with target phage display sequences. It then optimizes the sequence using gradient-based optimization. This approach allows Ens-Grad to generate sequences with greater enrichment than the training dataset, suggesting that the neural network model can provide efficient strategies for exploring promising subsets of sequence space for antibody design. In another study, Koichiro Saka et al. employed a long short-term memory network (LSTM) to generate and prioritize antibodies for efficiently discovering antibody sequences with higher affinity with kynurenine [115]. They enriched the dataset of kynurenine binding sequences through phage display panning using a kynurenine-binding oriented human synthetic Fab library. The generated sequences exhibited over 1800-fold higher affinity than that of the parental clone.

On the other hand, Kang and colleagues proposed a different approach for antibody maturation using GNNs and sequence data from the AB-Bind dataset, in contrast to methods utilizing high-throughput sequencing data [116,117]. The GNN-based modeling strategies investigated the contributions of interacting contacts (ICs) and non-interacting surfaces (NIS), as well as antibody-based features in predicting affinity. Different modeling designs were employed using Hag-Net GNN, resulting in satisfactory classification and regression performance in the AB-Bind dataset.

5. Antibody developability evaluation

Antibody drugs or candidates used for treatment must endure various challenges across production, transportation, storage, and administration. Consequently, comprehensive evaluation is a vital aspect of antibody development. The assessment of antibody developability generally follows a linear process that varies depending on the stage of antibody discovery. It encompasses numerous factors such as antibody solubility and aggregation, thermal unfolding, non-specific protein-protein interactions, charge heterogeneity, immunogenicity, pharmacokinetics, and toxicity [118]. In comparison to experimental methods, evaluation approaches powered by AI and machine learning offer the advantages of speed and cost-effectiveness. Here are several examples of developability evaluation driven by AI and ML.

One study by Hashemi et al. utilized ML modeling to predict the solubility of a recombinant antibody fragment in four different E. coli strains. The study optimized the soluble production of a single-chain variable fragment (scFv) antibody by modeling it as a function of post-induction temperature, post-induction time, cell density of induction time, and inducer concentration. The predicted values obtained using an artificial neural network (ANN) were closer to the experimental results than those obtained using response surface methodology (RSM) [119].

In another study, molecular dynamics simulation and ML were combined to predict antibody aggregation and viscosity. The study used features obtained from molecular dynamics simulations of the full-length antibody and sequences for ML model construction. The classification model's accuracy and the area under precision-recall curve in validation tests were 0.86 and 0.70, respectively. The best model was a logistic regression model with two features: the number of hydrophobic residues on the light chain variable region and net charges on the light chain variable region [120]. An earlier study also suggested that high viscosity is correlated with more hydrophilic and fewer hydrophobic

residues in the Fv region [121].

Current AI-driven immunogenicity predictions often involve predicting antibody binding to B cell or T cell epitopes. Timothy P. Riley et al. developed a peptide/MHC structure simulation program and a peptide database containing immunogenic and non-immunogenic peptides. They constructed an artificial neural network that utilizes this information to predict the immunogenicity of HLA-A2 binding [122]. This neural network model takes the three-dimensional model's structure and energy characteristics as inputs and generates a score between 0 and 1, indicating the confidence level in immunogenicity. Similarly, Karina Winterling et al. proposed a method to identify highly immunogenic T cell epitopes by simulating changes in the amino acid sequence through computer modeling. This approach enables the prediction and avoidance of recombinant FVIII (rFVIII) immunogenicity [123]. Alan M. Luu et al. demonstrated a convolutional neural network model that incorporates deep metric learning and multimodal learning. This model can identify the T cell receptor (TCR) associated with a given epitope from the TCR library and visualize the binding epitope [124]. Currently, the available dataset on TCR-ligands is limited. Dan Hudson aims to address the challenge of predicting TCR antigen specificity by fostering interdisciplinary collaborations to combine experimental data with AI and machine learning techniques [125]. The establishment of a reliable map linking TCRs to their homologous antigens through such efforts will undoubtedly greatly enhance computer-based immunogenicity prediction and advance the field.

When it comes to pharmacokinetic evaluation, ML offers evident advantages in analyzing diverse data types and large datasets [126]. By integrating traditional experimental models with machine learning methods, standardized datasets can be established, leading to reduced uncertainty, failure rates, and experimental costs. Moreover, this approach supports the ongoing efforts to minimize, enhance, and replace animal experiments. Zhou et al. demonstrated the feasibility of ML in constructing pharmacokinetic/pharmacodynamic (PK/PD) models [127]. They developed a PK/PD GRNN model with 6 input neurons, 23 hidden layer neurons, and one output neuron for Zingiberis Rhizoma. This model successfully investigated the impact of the concentration of active components from Zingiberis Rhizoma and Zingiberis Rhizoma Carbonisata on the pharmacodynamics. Hop et al. conducted a study showing that the GCN model exhibited excellent predictive performance for properties such as pKa, clearance, and plasma protein binding [128]. Bies et al. employed the genetic algorithm (GA) method to construct a PK model and compared the modeling performance of GA with manual step-by-step selection using PK data from seven drugs. The results demonstrated that GA provided nearly equivalent or superior model fitting for pharmacokinetic data [129]. This study also represents an exploration of machine learning techniques in the field of population pharmacokinetics (PPK).

6. Therapeutic antibodies discovery and optimization

AI methods have been used to learn the coding and genetic sequences of all known proteins, and generate different structures and functions for proteins with specified therapeutic functions. However, creating functional de novo proteins/antibodies that can be customized to achieve the desired function requires challenging protein design approaches. Such "purposefully-functional" de novo protein design requires a high level of control over the final design's shape, size, and function, as well as prior knowledge or assumptions about the structure of the targeted functional site. While the de novo design of functional proteins remains an elusive landmark in the field, the combination of structural design, ML, and experiments has achieved several successes in therapeutic antibody discovery and optimization. During this process, mutation libraries, flow cytometry, and deep sequencing are usually used to generate data for training neural networks. In this section, we review AI-driven antibodies that have been described in peer-reviewed articles and conference abstracts. The details of these antibodies are described below and summarized in Table 2.

Cetuximab is a mAb approved for treating patients with colorectal cancer (mCRC). However, acquired resistance to cetuximab is mediated by mutations in the cetuximab epitope in the epidermal growth factor receptor (EGFR) ectodomain, hindering its clinical application. Zhuang et al. [128] successfully applied a structure-guided and phage-assisted evolution (SGAPAE) approach based on the Rosetta DL platform to direct the evolution of cetuximab [130]. They reversed acquired drug resistance conferred by EGFR ECD mutations in colorectal cancer.

Mason et al. used DL to predict antigen specificity and identified optimized antibody variants from a massively diverse space of antibody sequences. In this study, deep-sequenced libraries of the therapeutic antibody trastuzumab (consisting of approximately 1×10^4 variants) were screened for specificity to HER2, and used as a training set to optimize a neural network. This network was then used to screen a computational library of approximately 1×10^8 trastuzumab variants and predict the HER2-specific subset (consisting of approximately 1×10^6 variants). Experimental testing of 30 randomly selected variants verified the model's accuracy [25].

Additionally, Sharrol Bachas et al. developed the Activity-specific Cell-Enrichment (ACE) assay, which generates high-throughput measurements of trastuzumab binding affinities with HER2 for the training of deep contextual language models [131]. These trained models can accurately predict binding affinities of unseen antibody variants, enabling virtual screenings and augmenting the accessible sequence space by orders of magnitude. To improve developability and immunogenicity properties, a genetic algorithm was developed to efficiently identify sequences with both strong binding affinity and high naturalness.

Makowski and colleagues developed an ML model to optimize the affinity and specificity of antibodies. Their model employed a projector network, which generated a one-dimensional prediction for each protein sequence, and a final prediction layer used to predict the affinity and specificity of the antibody [132]. The projector network used Linear Discriminant Analysis function, and the network was trained using the Sparse Categorical Cross-entropy loss function and the ADAM optimizer. The dataset used to train the network was generated by sorting and deep sequencing a mutation antibody library that the researchers constructed.

The researchers optimized the affinity and specificity of Emibetuzumab, an antibody that inhibits the MET signal pathway activation of both HGF-dependent and HGF-independent tumors by blocking the binding of MET and HGF [133]. Emibetuzumab is currently in phase II clinical trials for non-small cell lung cancer treatment [134].

In a similar vein, Jonathan Parkinson built an AI model called RESP to identify antibodies with high affinity and increased the K_D value of Atezolizumab (Tecentriq), an antibody that targets PD -L1, by 17 times through RESPs [135]. The RESP model consists of four parts: an autoencoder model trained on over 3 million human B-cell receptor (BCR) sequences to simulate trends in fluorescence-activated cell sorting (FACS) data more accurately; a yeast display library of Atezolizumab mutants to provide affinity data and protein sequences; a Bayesian network using ordinal regression to predict the likelihood of a given sequence having a low K_{off} value; and a modified simulated annealing algorithm that developed a directed evolution algorithm, allowing the RESP model to explore sequences with high affinity that did not exist in the mutation library.

AU-007, developed by Biolojic, is likely the first computationally designed human antibody to enter clinical development. It binds to the CD25-binding portion on IL-2, blocking the Treg expansion auto-inhibitory loop and preventing IL-2 binding to the trimeric receptor. This indicates its unique therapeutic profile and potential as a novel cancer treatment [136]. The Biolojic platform utilizes ML algorithms trained by antigen-antibody pairs and billions of human somatic hypermutations to find a good template and predict mutation sites for improving affinity, humanness, stability, and developability.

Table 2
Summary for AI-driven therapeutic antibodies.

	Description	Method	Status	References
Antibody discovery platform	Bamlanivimab, an antibody neutralizes SARS-Cov-2	ML based paired-chain antibody sequences prediction	Approved by FDA with emergency use authorization (EUA) on 9 November 2020, and revoked by FDA on April 16, 2021	[138]
	IBIO-101, an antibody binds CD25 on Tregs cells without blocking the IL-2 signaling pathway associated with Teffs cells, thereby depleting immunosuppressive Treg cells and stimulating anti-tumor immunity	Meso-scale engineered molecules (MEMs) AI model	Pre-clinical studies	U.S. Patent No. 11,545,238
	Several antibodies bind to SARS-CoV-2 spike proteins	Antibody-GAN, trained by 4 million light- and heavy-chain human antibody sequence	Pre-clinical studies	[139]
	AU-007, an antibody which binds the CD25-binding portion on IL-2, preventing the binding of IL-2 to the trimeric receptor while simultaneously blocking the Treg expansion autoinhibitory loop	AI platform trained with antigen-antibody pairs and billions of human somatic hypermutations	Phase 1/2	[136],[140]
	Generate variable length CDR3 sequences that resemble real sequences	The adversarial network models trained by the data which generated by microfluidics, yeast display, and deep sequencing	Pre-clinical studies	[137]
Antibody optimization platform	Optimize Cetuximab to avoid the drug resistance conferred by EGFR ECD mutations	Combined Rosetta deep learning platform with structure-guided and phage-assisted evolution (SGAPAE)	Pre-clinical studies	[130]
	Optimize Emibetuzumab with the affinity and specificity	Mechanical learning neural network based on mutation library		[132]
	Optimize Atezolizumab with 17 times KD increasement	RESP model		[135]
	Identify optimized trastuzumab variants	Neural network model trained by mutation library data		[25]
	Seeking for sequences with high affinity and naturalness of trastuzumab variants by screening in silico	Combined a deep contextual language model trained by Activity-specific Cell-Enrichment (ACE) assay data with a genetic algorithm		[131]
	Optimize affinity of Omburtamab, an anti-CD276 antibody	Docking simulations by ZDOCK and homology modeling by MODELLER		[17,141]
	Optimize affinity of Naxitamab, an anti-GD2 antibody	Docking simulations by ZDOCK and homology modeling by MODELLER	Granted as Orphan Drug Designation (ODD) and Rare Pediatric Disease Designation (RPDD).	[15]

Lim, Adler, and Johnson utilized microfluidics, yeast display, and deep sequencing to generate a panel of binder and non-binder antibody sequences to cancer immunotherapy targets PD-1 and CTLA-4. Then they trained convolutional neural network models to classify binders and non-binders using encoded antibody images. Their generative adversarial network models generated variable length CDR3 sequences resembling real sequences [137].

AbCellera Biologics Inc. and Eli Lilly and Company developed Bamlanivimab, a SARS-Cov-2 neutralizing antibody helped by ML based paired-chain antibody sequences prediction to select antibody gene from Next-generation sequencing (NGS) of B-cell, leading to finding of 440 paired antibodies. [138].

IBIO-101 is an antibody developed by IBIO Inc using its AI model meso-scale engineered molecules (MEMs) and ML algorithms combined with empirical analysis of protein structure. MEMs use an engineered scaffold to preserve natural conformation and motion present in the target protein, leading to the discovery of epitope-selective antibodies. IBIO-101 has been shown to bind CD25 on Tregs cells without blocking the IL-2 signaling pathway associated with Teffs cells, thereby depleting immunosuppressive Treg cells and stimulating anti-tumor immunity. This antibody is currently in preclinical studies.

Antibody-GAN, developed by Tileli Amimeur et al., is a platform for designing new antibodies using modified Wassertein-GANs trained with over 400,000 human antibody sequences to generate single-chain or double-chain antibody sequences [139]. Transfer learning can be used to encode key attributes of interest into the library, which can be used to find antibodies with interesting characteristics. This model was validated by the discovery of multiple candidate antibodies that specifically bind to the SARS-CoV-2 spike protein.

Omburtamab, a monoclonal antibody known as 8H9, is currently under development as a therapeutic agent targeting B7-H3-expressing cells. It is intended for the treatment of embryonal tumors, carcinomas, sarcomas, and brain tumors. Naxitamab, a humanized anti-GD2 monoclonal antibody referred to as 3F8, is also in development. It is being explored as a potential treatment for neuroblastoma, osteosarcoma, and other GD2-positive cancers. Cheung and Zhao et al. introduced a platform that utilizes in silico modeling for affinity maturation and epitope mapping of both 8H9 and 3F8 [15,17]. The affinity-matured humanized versions of 3F8 and 8H9 exhibited improved in vitro antibody-dependent cell-mediated cytotoxicity (ADCC) and demonstrated high tumor uptake in mouse xenograft models when tested in vivo.

7. Challenges and future perspectives

Antibody therapy has emerged as an effective approach for cancer treatment, with the discovery of therapeutic antibodies relying on animal immunization, in vitro display methodologies (such as phage library display), and helped by AI [142]. The AI-driven antibody discovery model offers numerous advantages over traditional methods. Firstly, it allows for rapid identification of antibody sequences that bind to the target protein. Secondly, this approach has lower requirements for experimental instruments, making it more convenient to build or expand. Thirdly, high-throughput screening can be performed more efficiently and cost-effectively.

Despite the benefits of AI-driven antibody discovery, there are still

challenges to overcome. To train accurate and effective models, a large amount of high-quality data is required, which can only be obtained through biological, physical, or chemical experiments [13]. However, data standardization has not been established, and collating and cleaning the published data for AI training can be time-consuming. Moreover, with the increasing demand for accurate protein simulation or molecular docking, the hash rate of computers may become a core challenge in the future.

Currently, efforts are underway to standardize data and establish open-source databases, which could propel the development of antibody drugs to a new level. In the future, precision medicine AI models for patients with drug resistance gene mutations may become a reality. An efficient AI-driven antibody discovery model could also serve as a potent weapon against large-scale epidemic diseases, allowing for rapid drug discovery and efficient control of such diseases.

Recently, OpenAI introduced multimodal large language model (LLM) GPT-4 which promoted us to look at the future of application of Natural Language Processing (NLP) methods in antibody design. Embedding antibody sequences in textual representation into vector space (vector representation) allows an implicit account for intrinsic biophysical properties[143]. Deep learning shows great potential in learning unobserved patterns from amino acid sequences that are relevant to their structure and function. currently, the application of specified NLP models like ESMfold [84]. AntiBERTy [87] has already shown good potential in antibody related models. However, it is still need to see if the GPT-4 like LLM or antibody specified models are better suitable for actual antibody design.

8. Conclusion

Although the application of AI in antibody development is still in its early stages, there have been significant successes in general protein modeling, and the increasing number of antibody models indicates a strong trend towards using AI to improve the developability of biotherapeutics, reduce costs, and expand access [144]. In addition to tools with specific structural predictions, such as DeepH3 and DeepAb, more general platforms for antibody design, humanization, and humanness evaluation based on DL have emerged [145]. In the near future, AI-powered antibody prediction and design tools will have even more applications in cancer therapy and will continue to advance the field.

Funding

B. Ma thanks support from Natural Science Foundation of China (Grant No. 32171246) and Shanghai Municipal Government Science Innovation grant 21JC1403700. Y. Wang thanks support from the grants from the National Natural Science Foundation of China (No.32200531), the Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2022QN114 and YG2022QN082), and Startup Fund for Young Faculty at SJTU (SFYF at SJTU). Q. Zhao has been supported by the National Key R&D Program of China (2019YFA0904400), the Natural Science Foundation of Guangdong Province (2023A1515010549), the Shenzhen Science and Technology Project (SGDX2020110309280301), the Science and Technology Development Fund of Macau (FDCT/0043/2021/A1 and FDCT/0002/2021/AKP), the University of Macau (MYRG2022-00143-FHS), Dr. Stanley Ho Medical Development Foundation (SHMDF-VSEP/2022/002), and Zhongnanshan Medical Foundation of Guangdong Province (ZNSA-2021016).

CRediT authorship contribution statement

Q. Z and B.M. designed the study. All authors co-wrote the manuscript and approved this version of the article.

Declaration of Competing Interest

None.

Data Availability

No data was used for the research described in the article.

Acknowledgments

The computational simulations are supported by the Center for High Performance Computing at Shanghai Jiao Tong University.

References

[1] Z. Elgundi, M. Reslan, E. Cruz, V. Sifniotis, V. Kayser, The state-of-play and future of antibody therapeutics, Adv. Drug Deliv. Rev. 122 (2017) 2–19.
[2] G. Zhou, Q. Zhao, Perspectives on therapeutic neutralizing antibodies against the Novel Coronavirus SARS-CoV-2, Int J. Biol. Sci. 16 (10) (2020) 1718–1723.
[3] Q. Zhao, Bispecific antibodies for autoimmune and inflammatory diseases: clinical progress to date, BioDrugs 34 (2) (2020) 111–119.
[4] Q. Zhao, M. Ahmed, D.V. Tassev, A. Hasan, T.Y. Kuo, H.F. Guo, R.J. O'Reilly, N. K. Cheung, Affinity maturation of T-cell receptor-like antibodies for Wilms tumor 1 peptide greatly enhances therapeutic potential, Leukemia 29 (11) (2015) 2238–2247.
[5] Y. Zhang, Y. Goto, S. Yagishita, Y. Shinno, K. Mizuno, N. Watanabe, Y. Yamamoto, N. Ota, T. Ochiya, Y. Fujita, Machine learning-based exceptional response prediction of nivolumab monotherapy with circulating microRNAs in non-small cell lung cancer, Lung Cancer 173 (2022) 107–115.
[6] S.R. Punekar, E. Shum, C.M. Grello, S.C. Lau, V. Velcheti, Immunotherapy in non-small cell lung cancer: Past, present, and future directions, Front Oncol 12 (2022) 877594.
[7] J. Liu, S. Yang, B. Cao, G. Zhou, F. Zhang, Y. Wang, R. Wang, L. Zhu, Y. Meng, C. Hu, H. Liang, X. Lin, K. Zhu, G. Chen, K.Q. Luo, L. Di, Q. Zhao, Targeting B7-H3 via chimeric antigen receptor T cells and bispecific killer cell engagers augments antitumor response of cytotoxic lymphocytes, J. Hematol. Oncol. 14 (1) (2021) 21.
[8] W. Chen, Y. Feng, Q. Zhao, Z. Zhu, D.S. Dimitrov, Human monoclonal antibodies targeting nonoverlapping epitopes on insulin-like growth factor II as a novel type of candidate cancer therapeutics, Mol. Cancer Ther. 11 (7) (2012) 1400–1410.
[9] D.Z. Li, B.N. Han, R. Wei, G.Y. Yao, Z. Chen, J. Liu, T.C.W. Poon, W. Su, Z. Zhu, D. S. Dimitrov, Q. Zhao, N-terminal alpha-amino group modification of antibodies using a site-selective click chemistry method, MAbs 10 (5) (2018) 712–719.
[10] L. Zhu, J. Liu, G. Zhou, H.M. Ng, I.L. Ang, G. Ma, Y. Liu, S. Yang, F. Zhang, K. Miao, T.C.W. Poon, X. Zhang, Z. Yuan, C.X. Deng, Q. Zhao, Targeting immune checkpoint B7-H3 antibody-chlorin e6 bioconjugates for spectroscopic photoacoustic imaging and photodynamic therapy, Chem. Commun. 55 (95) (2019) 14255–14258.
[11] B. Cao, M. Liu, L. Wang, K. Zhu, M. Cai, X. Chen, Y. Feng, S. Yang, S. Fu, C. Zhi, X. Ye, J. Zhang, Z. Zhang, X. Yang, M. Zhao, Q. Wu, L. Xu, L. Yang, H. Lian, Q. Zhao, Z. Zhang, Remodelling of tumour microenvironment by microwave ablation potentiates immunotherapy of AXL-specific CAR T cells against non-small cell lung cancer, Nat. Commun. 13 (1) (2022) 6203.
[12] J. Liu, F. Zhang, J. Yu, Q. Zhao, Programmed death-ligand 1 expression on CD22-specific chimeric antigen receptor-modified T cells weakens antitumor potential, MedComm (2020), e140.
[13] R. Akbar, H. Bashour, P. Rawat, P.A. Robert, E. Smorodina, T.S. Cotet, K. Flem-Karlsen, R. Frank, B.B. Mehta, M.H. Vu, T. Zengin, J. Gutierrez-Marcos, F. Lund-Johansen, J.T. Andersen, V. Greiff, Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies, MAbs 14 (1) (2022) 2008790.
[14] J. Kim, M. McFee, Q. Fang, O. Abdin, P.M. Kim, Computational and artificial intelligence-based methods for antibody development, Trends Pharm. Sci. 44 (3) (2023) 175–189.
[15] Q. Zhao, M. Ahmed, H.F. Guo, I.Y. Cheung, N.K. Cheung, Alteration of electrostatic surface potential enhances affinity and tumor killing properties of anti-ganglioside GD2 monoclonal antibody hu3F8, J. Biol. Chem. 290 (21) (2015) 13017–13027.
[16] D. Li, R. Gong, J. Zheng, X. Chen, D.S. Dimitrov, Q. Zhao, Engineered antibody CH2 domains binding to nucleolin: Isolation, characterization and improvement of aggregation, Biochem. Biophys. Res. Commun. 485 (2) (2017) 446–453.
[17] M. Ahmed, M. Cheng, Q. Zhao, Y. Goldgur, S.M. Cheal, H.F. Guo, S.M. Larson, N. K. Cheung, Humanized affinity-matured monoclonal antibody 8H9 has potent antitumor activity and binds to FG loop of tumor antigen B7-H3, J. Biol. Chem. 290 (50) (2015) 30018–30029.
[18] S. Huang, J. Yang, N. Shen, Q. Xu, Q. Zhao, Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective, Semin Cancer Biol. 89 (2023) 30–37.
[19] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives, Int. J. Biol. Sci. 17 (6) (2021) 1581–1587.

[20] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges, Cancer Lett. 471 (2020) 61–71.
[21] J. Yang, S. Fong, H. Wang, Q.Y. Hu, C. Lin, S.G. Huang, J. Shi, K. Lan, R. Tang, Y. Y. Wu, Q. Zhao, Artificial intelligence in ophthalmopathy and ultra-wide field image: a survey, Expert Syst. Appl. 182 (2021).
[22] J. Fang, Z. Chen, Evaluation of short-term efficacy of PD-1 monoclonal antibody immunotherapy for lymphoma by positron emission tomography/computed tomography imaging with convolutional neural network image registration algorithm, Contrast Media Mol. Imaging 2022 (2022) 1388517.
[23] M. Mousavi, M.D. Manshadi, M. Soltani, F.M. Kashkooli, A. Rahmim, A. Mosavi, M. Kvasnica, P.M. Atkinson, L. Kovacs, A. Koltay, N. Kiss, H. Adeli, Modeling the efficacy of different anti-angiogenic drugs on treatment of solid tumors using 3D computational modeling and machine learning, Comput. Biol. Med 146 (2022), 105511.
[24] W. Wilman, S. Wrobel, W. Bielska, P. Deszynski, P. Dudzic, I. Jaszczyszyn, J. Kaniewski, J. Mlokosiewicz, A. Rouyan, T. Satlawa, S. Kumar, V. Greiff, K. Krawczyk, Machine-designed biotherapeutics: opportunities, feasibility and advantages of deep learning in computational antibody discovery, Brief Bioinform. 23 (4) (2022).
[25] D.M. Mason, S. Friedensohn, C.R. Weber, C. Jordi, B. Wagner, S.M. Meng, R. A. Ehling, L. Bonati, J. Dahinden, P. Gainza, B.E. Correia, S.T. Reddy, Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning, Nat. Biomed. Eng. 5 (6) (2021) 600–612.
[26] C.-M. Lo, U. Iqbal, Y.-C. Li, Cancer quantification from data mining to artificial intelligence, Comput. Methods Prog. Biomed. 145 (2017) A1.
[27] J. Abbasi, Artificial intelligence tools for sepsis and cancer, JAMA 320 (22) (2018), 2303-2303.
[28] W.L. Bi, A. Hosny, M.B. Schabath, M.L. Giger, N.J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout, C. Abbosh, I.F. Dunn, R.H. Mak, R.M. Tamimi, C. M. Tempany, C. Swanton, U. Hoffmann, L.H. Schwartz, R.J. Gillies, R.Y. Huang, H. Aerts, Artificial intelligence in cancer imaging: Clinical challenges and applications, CA Cancer J. Clin. 69 (2) (2019) 127–157.
[29] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[30] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[31] S. Daberdaku, C. Ferrari, Antibody interface prediction with 3D Zernike descriptors and SVM, Bioinformatics 35 (11) (2019) 1870–1876.
[32] V. Kunik, S. Ashkenazi, Y. Ofran, Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure, Nucleic Acids Res. (2012) W521–W524.
[33] K. Krawczyk, T. Baker, J. Shi, C.M. Deane, Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking, Protein Eng. Des. Sel. 26 (10) (2013) 621–629.
[34] E. Liberis, P. Velickovic, P. Sormanni, M. Vendruscolo, P. Lio, Parapred: antibody paratope prediction using convolutional and recurrent neural networks, Bioinformatics 34 (17) (2018) 2944–2950.
[35] O.A. Egaji, S. Ballard-smith, I. Asghar, M. Griffiths, A machine learning approach for predicting antibody properties. Proceedings of the 10th International Conference on Information Communication and Management, Association for Computing Machinery, Paris, France, 2020, pp. 20–24.
[36] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature (2021) 583–589.
[37] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe, Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 675–678.
[38] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: a system for large-scale machine learning. Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, USENIX Association, Savannah, GA, USA, 2016, pp. 265–283.
[39] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in PyTorch, 2017.
[40] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, Z. Zhang, MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems ArXiv abs/1512 01274 2015.
[41] F. Seide, A. Agarwal, Cntk, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 2135–2135.
[42] J. Teuwen, N. Moriakov, Chapter 20 - Convolutional neural networks, in: S. K. Zhou, D. Rueckert, G. Fichtinger (Eds.), Handbook of Medical Image Computing and Computer Assisted Intervention, Academic Press, 2020, pp. 481–501.
[43] S. Lu, Y. Li, F. Wang, X. Nan, S. Zhang, Leveraging sequential and spatial neighbors information by using CNNs linked with GCNs for paratope prediction, IEEE/ACM Trans. Comput. Biol. Bioinforma. 19 (1) (2022) 68–74.
[44] G.R. Kanagachidambaresan, A. Ruwali, D. Banerjee, K.B. Prakash, Recurrent neural network, in: K.B. Prakash, G.R. Kanagachidambaresan (Eds.), Programming with TensorFlow: Solution for Edge Computing Applications, Springer International Publishing, Cham, 2021, pp. 53–61.
[45] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[46] N. Toshiaki, I. Seiichi, F. Haruka, S. Kugatsu, S. Makoto, T. Akiko, N. Hironori , Epitope Prediction of Antigen Protein using Attention-Based LSTM Network bioRxiv 2020.2020.07.27.224121.
[47] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.
[48] F. Simon, N. Daniel, A.K. Tarik, C. Lucia, P. Cristina, R.Gd.V. Arthur, E. Lena, M. M. Derek, T.R. Sai, , Convergent selection in antibody repertoires is revealed by deep learning, bioRxiv (2020), 2020.02.25.965673.
[49] J. Bruna, W. Zaremba, A. Szlam, Y. Lecun, Spectral networks and locally connected networks on graphs, 2014.
[50] W. Jin, J. Wohlwend, R. Barzilay, T. Jaakkola, Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-design, arXiv e-prints (2021).
[51] S.J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution, Cell 149 (2) (2012) 262–273.
[52] X. Pan, T. Kortemme, Recent advances in de novo protein design: Principles, methods, and applications, J. Biol. Chem. 296 (2021), 100558.
[53] I. Anishchenko, S.J. Pellock, T.M. Chidyausiku, T.A. Ramelot, S. Ovchinnikov, J. Z. Hao, K. Bafna, C. Norn, A. Kang, A.K. Bera, F. DiMaio, L. Carter, C.M. Chow, G. T. Montelione, D. Baker, De novo protein design by deep network hallucination, Nature 600 (7889) (2021) 547.
[54] N. Ferruz, S. Schmidt, B. Hocker, ProtGPT2 is a deep unsupervised language model for protein design, Nat. Commun. 13 (1) (2022).
[55] B. Huang, Y. Xu, X. Hu, Y. Liu, S. Liao, J. Zhang, C. Huang, J. Hong, Q. Chen, H. Liu, A backbone-centred energy function of neural networks for protein design, Nature 602 (7897) (2022) 523–528.
[56] R.F. Alford, A. Leaver-Fay, J.R. Jeliazkov, M.J. O'Meara, F.P. DiMaio, H. Park, M. V. Shapovalov, P.D. Renfrew, V.K. Mulligan, K. Kappel, J.W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R.L. Dunbrack, Jr, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J.J. Gray, The rosetta all-atom energy function for macromolecular modeling and design, J. Chem. Theory Comput. 13 (6) (2017) 3031–3048.
[57] P.S. Huang, S.E. Boyken, D. Baker, The coming of age of de novo protein design, Nature 537 (7620) (2016) 320–327.
[58] I.V. Korendovych, W.F. DeGrado, De novo protein design, a retrospective, Q Rev. Biophys. 53 (2020), e3.
[59] D.N. Woolfson, A. Brief, History of De Novo protein design: minimal, rational, and computational, J. Mol. Biol. 433 (20) (2021), 167160.
[60] B.A. Meinen, C.D. Bahl, Breakthroughs in computational design methods open up new frontiers for de novo protein engineering, Protein Eng. Des. Sel. 34 (2021).
[61] M.L. Chiu, G.L. Gilliland, Engineering antibody therapeutics, Curr. Opin. Struct. Biol. 38 (2016) 163–173.
[62] D.R. Goulet, W.M. Atkins, Considerations for the design of antibody-based therapeutics, J. Pharm. Sci. 109 (1) (2020) 74–103.
[63] M.I.J. Raybould, C. Marks, K. Krawczyk, B. Taddese, J. Nowak, A.P. Lewis, A. Bujotzek, J. Shi, C.M. Deane, Five computational developability guidelines for therapeutic antibody profiling, Proc. Natl. Acad. Sci. USA 116 (10) (2019) 4025–4030.
[64] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Zidek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G.J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S.A. A. Kohl, A. Potapenko, A.J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A.W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Highly accurate protein structure prediction for the human proteome, Nature 596 (7873) (2021) 590–596.
[65] A. Kovaltsuk, K. Krawczyk, J.D. Galson, D.F. Kelly, C.M. Deane, J. Truck, How B-cell receptor repertoire sequencing can be enriched with structural antibody data, Front. Immunol. 8 (2017).
[66] R.J. Pantazes, C.D. Maranas, OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding, Protein Eng. Des. Sel. 23 (11) (2010) 849–858.
[67] R. Chowdhury, M.F. Allan, C.D. Maranas, OptMAVEn-2.0: de novo design of variable antibody regions against targeted antigen epitopes, Antibodies 7 (3) (2018).
[68] J. Adolf-Bryfogle, O. Kalyuzhniy, M. Kubitz, B.D. Weitzner, X. Hu, Y. Adachi, W. R. Schief, R.L. Dunbrack Jr., RosettaAntibodyDesign (RAbD): a general framework for computational antibody design, PLoS Comput. Biol. 14 (4) (2018), e1006112.
[69] R. Lipsh-Sokolik, D. Listov, S.J. Fleishman, The AbDesign computational pipeline for modular backbone assembly and design of binders and enzymes, Protein Sci. 30 (1) (2021) 151–159.
[70] A. Sircar, E.T. Kim, J.J. Gray, RosettaAntibody: antibody variable region homology modeling server, Nucleic Acids Res. (2009) W474–W479.
[71] D. Baran, M.G. Pszolla, G.D. Lapidoth, C. Norn, O. Dym, T. Unger, S. Albeck, M. D. Tyka, S.J. Fleishman, Principles for computational design of binding antibodies, Proc. Natl. Acad. Sci. USA 114 (41) (2017) 10900–10905.
[72] Z.Y. Du, H. Su, W.K. Wang, L.S. Ye, H. Wei, Z.L. Peng, I. Anishchenko, D. Baker, J. Y. Yang, The trRosetta server for fast and accurate protein structure prediction, Nat. Protoc. 16 (12) (2021) 5634–5651.
[73] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, Critical assessment of methods of protein structure prediction (CASP)-Round XIV, Proteins-Struct. Funct. Bioinforma. 89 (12) (2021) 1607–1617.
[74] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior,

K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature (2021) 583–589.
[75] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R.D. Schaeffer, C. Millan, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. van Dijk, A. C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K.C. Garcia, N.V. Grishin, P. D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, Science 373 (6557) (2021) 871.
[76] J.A. Ruffolo, C. Guerra, S.P. Mahajan, J. Sulam, J.J. Gray, Geometric potentials from deep learning improve prediction of CDR H3 loop structures, Bioinformatics 36 (2020) 268–275.
[77] J.A. Ruffolo, J. Sulam, J.J. Gray, Antibody structure prediction using interpretable deep learning, Patterns 3 (2) (2022).
[78] B. Abanades, G. Georges, A. Bujotzek, C.M. Deane, ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation, Bioinformatics (2022).
[79] T. Cohen, M. Halfon, D. Schneidman-Duhovny, NanoNet: Rapid and accurate end-to-end nanobody modeling by deep learning, Front Immunol. 13 (2022), 958584.
[80] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J.Y. Shi, C. M. Deane, SAbDab: the structural antibody database, Nucleic Acids Res. 42 (D1) (2014) D1140–D1146.
[81] B.D. Weitzner, D. Kuroda, N. Marze, J.Q. Xu, J.J. Gray, Blind prediction performance of RosettaAntibody 3.0: Grafting, relaxation, kinematic loop modeling, and full CDR optimization, Proteins-Struct. Funct. Bioinforma. 82 (8) (2014) 1611–1623.
[82] D. Akpinaroglu, J.A. Ruffolo, S.P. Mahajan, J.J. Gray, Simultaneous prediction of antibody backbone and side-chain conformations with deep learning, PLoS One 17 (6) (2022), e0258173.
[83] B. Abanades, G. Georges, A. Bujotzek, C.M. Deane, ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation, Bioinformatics 38 (7) (2022) 1877–1880.
[84] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (6637) (2023) 1123–1130.
[85] A. Kovaltsuk, J. Leem, S. Kelm, J. Snowden, C.M. Deane, K. Krawczyk, Observed antibody space: a resource for data mining next-generation sequencing of antibody repertoires, J. Immunol. 201 (8) (2018) 2502–2509.
[86] J.A. Ruffolo, L.-S. Chu, S.P. Mahajan, J.J. Gray, Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies, Nat. Commun. 14 (1) (2023) 2389.
[87] J.A. Ruffolo, J.J. Gray, J. Sulam, Deciphering antibody affinity maturation with language models and weakly supervised learning, arXiv Prepr. 2112 (2021) 07782.
[88] F. Van Bockstaele, J.B. Holz, H. Revets, The development of nanobodies for therapeutic applications, Curr. Opin. Investig. Drugs 10 (11) (2009) 1212–1224.
[89] T. Cohen, M. Halfon, D. Schneidman-Duhovny, NanoNet: rapid and accurate end-to-end nanobody modeling by deep learning, Front. Immunol. 13 (2022).
[90] M. Shatsky, R. Nussinov, H.J. Wolfson, A method for simultaneous alignment of multiple protein structures, Proteins-Struct. Funct. Bioinform. 56 (1) (2004) 143–156.
[91] B. Abanades, W.K. Wong, F. Boyles, G. Georges, A. Bujotzek, C.M. Deane, ImmuneBuilder: deep-Learning models for predicting the structures of immune proteins, Commun. Biol. 6 (1) (2023) 575.
[92] R. Akbar, P.A. Robert, C.R. Weber, M. Widrich, R. Frank, M. Pavlovic, L. Scheffer, M. Chernigovskaya, I. Snapkov, A. Slabodkin, B.B. Mehta, E. Miho, F. Lund-Johansen, J.T. Andersen, S. Hochreiter, I. Hobaek Haff, G. Klambauer, G. K. Sandve, V. Greiff, In silico proof of principle of machine learning-based antibody design at unconstrained scale, MAbs 14 (1) (2022) 2031482.
[93] P.A. Robert, R. Akbar, R. Frank, M. Pavlović, M. Widrich, I. Snapkov, A. Slabodkin, M. Chernigovskaya, L. Scheffer, E. Smorodina, P. Rawat, B. B. Mehta, M.H. Vu, I.F. Mathisen, A. Prósz, K. Abram, A. Olar, E. Miho, D.T. T. Haug, F. Lund-Johansen, S. Hochreiter, I.H. Haff, G. Klambauer, G.K. Sandve, V. Greiff, Unconstrained generation of synthetic antibody-antigen structures to guide machine learning methodology for real-world antibody specificity prediction, bioRxiv (2022) s2021.07.06.451258.
[94] S. Saha, G.P.S. Raghava, BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties, Lect. Notes Comput. Sc. 3239 (2004) 197–204.
[95] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, Proteins-Struct. Funct. Bioinform. 65 (1) (2006) 40–48.
[96] B. Manavalan, R.G. Govindaraj, T.H. Shin, M.O. Kim, G. Lee, iBCE-EL: a new ensemble learning framework for improved linear B-cell epitope prediction, Front. Immunol. 9 (2018).
[97] T.I. Shashkova, D. Umerenkov, M. Salnikov, P.V. Strashnov, A.V. Konstantinova, I. Lebed, D.N. Shcherbinin, M.N. Asatryan, O.L. Kardymon, N.V. Ivanisenko, SEMA: antigen B-cell conformational epitope prediction using deep transfer learning, Front Immunol. 13 (2022), 960985.
[98] S.E.C. Caoili, Hybrid methods for B-cell epitope prediction, in: R.K. De, N. Tomar (Eds.), Immunoinformatics, Springer, New York, New York, NY, 2014, pp. 245–283.
[99] P.H. Andersen, M. Nielsen, O. Lund, Prediction of residues in discontinuous B-cell epitopes using protein 3D structures, Protein Sci. 15 (11) (2006) 2558–2567.
[100] J. Sun, D. Wu, T. Xu, X. Wang, X. Xu, L. Tao, Y.X. Li, Z.W. Cao, Nucleic Acids Res 37(Web Server issue) (2009 a computational server for spatial epitope prediction of protein antigens, SEPPA W612 W616.
[101] M.J. Sweredoski, P. Baldi, PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure, Bioinformatics 24 (12) (2008) 1459–1460.
[102] S. Lu, Y.G. Li, Q. Ma, X.F. Nan, S.T. Zhang, A. Structure-Based, B-cell epitope prediction model through combing local and global features, Front. Immunol. 13 (2022).
[103] I. Sela-Culang, Y. Ofran, B. Peters, Antibody specific epitope prediction - emergence of a new paradigm, Curr. Opin. Virol. 11 (2015) 98–102.
[104] K. Krawczyk, X.F. Liu, T. Baker, J. Shi, C.M. Deane, Improving B-cell epitope prediction and its application to global antibody-antigen docking, Bioinformatics 30 (16) (2014) 2288–2294.
[105] C. Schneider, A. Buchanan, B. Taddese, C.M. Deane, DLAB: deep learning methods for structure-based virtual screening of antibodies, Bioinformatics 38 (2) (2022) 377–383.
[106] V. Kunik, Y. Ofran, The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops, Protein Eng. Des. Sel. 26 (10) (2013) 599–609.
[107] A. Deac, P. Velickovic, P. Sormanni, Attentive cross-modal paratope prediction, J. Comput. Biol. 26 (6) (2019) 536–545.
[108] S. Pittala, C. Bailey-Kellogg, Learning context-aware structural representations to predict antigen and antibody binding interfaces, Bioinformatics 36 (13) (2020) 3996–4003.
[109] A. Del Vecchio, A. Deac, P. Liò, P. Veličković, Neural message passing for joint paratope-epitope prediction, arXiv Prepr. 2106 (2021) 00757.
[110] X. Wang, S.T. Flannery, D. Kihara, Protein docking model evaluation by graph neural networks, Front. Mol. Biosci. 8 (2021), 647915.
[111] N. Renaud, C. Geng, S. Georgievska, F. Ambrosetti, L. Ridder, D.F. Marzella, M. F. Réau, A.M.J.J. Bonvin, L.C. Xue, DeepRank: a deep learning framework for data mining 3D protein-protein interfaces, Nat. Commun. 12 (1) (2021) 7068.
[112] M. Réau, N. Renaud, L.C. Xue, A.M.J.J. Bonvin, DeepRank-GNN: a graph neural network framework to learn patterns in protein-protein interfaces, bioRxiv (2021), 2021.12.08.471762.
[113] L.C. Xue, J.P. Rodrigues, P.L. Kastritis, A.M. Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein-protein complexes, Bioinformatics 32 (23) (2016) 3676–3678.
[114] G. Liu, H. Zeng, J. Mueller, B. Carter, Z. Wang, J. Schilz, G. Horny, M. E. Birnbaum, S. Ewert, D.K. Gifford, Antibody complementarity determining region design using high-capacity machine learning, Bioinformatics 36 (7) (2019) 2126–2133.
[115] K. Saka, T. Kakuzaki, S. Metsugi, D. Kashiwagi, K. Yoshida, M. Wada, H. Tsunoda, R. Teramoto, Antibody design using LSTM based deep generative model from phage display library for affinity maturation, Sci. Rep. 11 (1) (2021).
[116] Y. Kang, D. Leng, J. Guo, L. Pan, Sequence-based deep learning antibody design for in silico antibody affinity maturation, 2021, p. arXiv:2103.03724.
[117] S. Sirin, J.R. Apgar, E.M. Bennett, A.E. Keating, AB-Bind: Antibody binding mutational database for computational affinity predictions, Protein Sci. 25 (2) (2016) 393–409.
[118] W. Zhang, H. Wang, N. Feng, Y. Li, J. Gu, Z. Wang, Developability assessment at early-stage discovery to enable development of antibody-derived therapeutics, Antib. Ther. 6 (1) (2023) 13–29.
[119] A. Hashemi, M. Basafa, A. Behravan, Machine learning modeling for solubility prediction of recombinant antibody fragment in four different E. coli strains, Sci. Rep. 12 (1) (2022) 5463.
[120] P.K. Lai, A. Gallegos, N. Mody, H.A. Sathish, B.L. Trout, Machine learning prediction of antibody aggregation and viscosity for high concentration formulation development of protein therapeutics, MAbs 14 (1) (2022) 2026208.
[121] P.K. Lai, A. Fernando, T.K. Cloutier, Y. Gokarn, J. Zhang, W. Schwenger, R. Chari, C. Calero-Rubio, B.L. Trout, Machine learning applied to determine the molecular descriptors responsible for the viscosity behavior of concentrated therapeutic antibodies, Mol. Pharm. 18 (3) (2021) 1167–1175.
[122] T.P. Riley, G.L.J. Keller, A.R. Smith, L.M. Davancaze, A.G. Arbuiso, J.R. Devlin, B. M. Baker, Structure based prediction of neoantigen immunogenicity, Front Immunol. 10 (2019) 2047.
[123] K. Winterling, W.D. Martin, A.S. De Groot, J. Daufenbach, S. Kistner, J. Schuttrumpf, Development of a novel fully functional coagulation factor VIII with reduced immunogenicity utilizing an in silico prediction and deimmunization approach, J. Thromb. Haemost. 19 (9) (2021) 2161–2170.
[124] A.M. Luu, J.R. Leistico, T. Miller, S. Kim, J.S. Song, Predicting TCR-epitope binding specificity using deep metric learning and multimodal learning, Genes 12 (4) (2021).
[125] D. Hudson, R.A. Fernandes, M. Basham, G. Ogg, H. Koohy, Can we predict T cell specificity with digital biology and machine learning? Nat. Rev. Immunol. (2023) 1–11.
[126] N. Maharao, V. Antontsev, M. Wright, J. Varshney, Entering the era of computationally driven drug development, Drug Metab. Rev. 52 (2) (2020) 283–298.
[127] S.J. Zhou, J. Meng, B. Liu, Investigation into the pharmacokinetic-pharmacodynamic model of Zingiberis Rhizoma/Zingiberis Rhizoma Carbonisata and contribution to their therapeutic material basis using artificial neural networks, RSC Adv. 7 (41) (2017) 25488–25496.
[128] P. Hop, B. Allgood, J. Yu, Geometric deep learning autonomously learns chemical features that outperform those engineered by domain experts, Mol. Pharm. 15 (10) (2018) 4371–4377.

[129] R.R. Bies, M.F. Muldoon, B.G. Pollock, S. Manuck, G. Smith, M.E. Sale, A genetic algorithm-based, hybrid machine learning approach to model selection, J. Pharmackinet. Pharmacodyn. 33 (2) (2006) 195–221.
[130] X. Zhuang, Z. Wang, J. Fan, X. Bai, Y. Xu, J.J. Chou, T. Hou, S. Chen, L. Pan, Structure-guided and phage-assisted evolution of a therapeutic anti-EGFR antibody to reverse acquired resistance, Nat. Commun. 13 (1) (2022) 4431.
[131] S. Bachas, G. Rakocevic, D. Spencer, A.V. Sastry, R. Haile, J.M. Sutton, G. Kasun, A. Stachyra, J.M. Gutierrez, E. Yassine, B. Medjo, V. Blay, C. Kohnert, J.T. Stanton, A. Brown, N. Tijanic, C. McCloskey, R. Viazzo, R. Consbruck, H. Carter, S. Levine, S. Abdulhaqq, J. Shaul, A.B. Ventura, R.S. Olson, E. Yapici, J. Meier, S. McClain, M. Weinstock, G. Hannum, A. Schwartz, M. Gander, R. Spreafico, Antibody optimization enabled by artificial intelligence predictions of binding affinity and naturalness, bioRxiv (2022) 2022.08.16.504181.
[132] E.K. Makowski, P.C. Kinnunen, J. Huang, L. Wu, M.D. Smith, T. Wang, A.A. Desai, C.N. Streu, Y. Zhang, J.M. Zupancic, J.S. Schardt, J.J. Linderman, P.M. Tessier, Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space, Nat. Commun. 13 (1) (2022) 3788.
[133] L. Liu, W. Zeng, M.A. Wortinger, S.B. Yan, P. Cornwell, V.L. Peek, J.R. Stephens, J. W. Tetreault, J. Xia, J.R. Manro, K.M. Credille, D.W. Ballard, P. Brown-Augsburger, V. Wacheck, C.K. Chow, L. Huang, Y. Wang, I. Denning, J. Davies, Y. Tang, P. Vaillancourt, J. Lu, LY2875358, a neutralizing and internalizing anti-MET bivalent antibody, inhibits HGF-dependent and HGF-independent MET activation and tumor growth, Clin. Cancer Res. 20 (23) (2014) 6059–6070.
[134] D.R. Camidge, T. Moran, I. Demedts, H. Grosch, K. Mileham, J. Molina, O. Juan-Vidal, G. Bepler, J.W. Goldman, K. Park, J. Wallin, S.R. Wijayawardana, X. A. Wang, V. Wacheck, E. Smit, A randomized, open-label phase II study evaluating emibetuzumab plus erlotinib and emibetuzumab monotherapy in MET immunohistochemistry positive NSCLC patients with acquired resistance to erlotinib, Clin. Lung Cancer 23 (4) (2022) 300–310.
[135] J. Parkinson, R. Hard, W. Wang, The RESP AI model accelerates the identification of tight-binding antibodies, Nat. Commun. 14 (1) (2023) 454.
[136] I. Amit, I. Levin, T. Wyant, N. Levitin, R. Barak, M. Ben-Mayor, O. Bluvshtein, N. Grossman, Y. Sasson, G. Nimrod, M. Zehnin, S. Fischman, M. Strajbl, L. Danielpur, A. Knickerbocker, J. Vasselli, Y. Ofran, The computationally designed human antibody, Au-007, mediates human immune activation by endogenous Il-2, while uniquely breaking the Il-2 auto-inhibitory loop and preventing treg expansion, J. Immunother. Cancer 9 (2021) A732–A734.
[137] Y.W. Lim, A.S. Adler, D.S. Johnson, Predicting antibody binders and generating synthetic antibodies using deep learning, MAbs 14 (1) (2022) 2069075.
[138] B.E. Jones, P.L. Brown-Augsburger, K.S. Corbett, K. Westendorf, J. Davies, T. P. Cujec, C.M. Wiethoff, J.L. Blackbourne, B.A. Heinz, D. Foster, R.E. Higgs, D. Balasubramaniam, L. Wang, Y. Zhang, E.S. Yang, R. Bidshahri, L. Kraft, Y. Hwang, S. Zentelis, K.R. Jepson, R. Goya, M.A. Smith, D.W. Collins, S. J. Hinshaw, S.A. Tycho, D. Pellacani, P. Xiang, K. Muthuraman, S. Sobhanifar, M. H. Piper, F.J. Triana, J. Hendle, A. Pustilnik, A.C. Adams, S.J. Berens, R.S. Baric, D.R. Martinez, R.W. Cross, T.W. Geisbert, V. Borisevich, O. Abiona, H.M. Belli, M. de Vries, A. Mohamed, M. Dittmann, M.I. Samanovic, M.J. Mulligan, J. A. Goldsmith, C.L. Hsieh, N.V. Johnson, D. Wrapp, J.S. McLellan, B.C. Barnhart, B.S. Graham, J.R. Mascola, C.L. Hansen, E. Falconer, The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates, Sci. Transl. Med. 13 (593) (2021).
[139] T. Amimeur, J.M. Shaver, R.R. Ketchem, J.A. Taylor, R.H. Clark, J. Smith, D. Van Citters, C.C. Siska, P. Smidt, M. Sprague, B.A. Kerwin, D. Pettit, Designing feature-controlled humanoid antibody discovery libraries using generative adversarial networks, bioRxiv (2020), 2020.04.12.024844.
[140] I. Amit, I. Levin, T. Wyant, N. Levitin, R. Barak, M. Ben-Mayor, O. Bluvshtein, N. Grossman, Y. Sasson, G. Nimrod, M. Zehnin, S. Fischman, M. Štrajbl, L. Danielpur, A. Knickerbocker, J. Vasselli, Y. Ofran, 704The computationally designed human antibody, AU-007, mediates human immune activation by endogenous IL-2, while uniquely breaking the IL-2 auto-inhibitory loop and preventing Treg expansion, J. Immunother. Cancer 9 (2) (2021) A732–A734.
[141] J.W.K. Ng, E.H.Q. Ong, L. Tucker-Kellogg, G. Tucker-Kellogg, Deep learning for de-convolution of Smad2 versus Smad3 binding sites, BMC Genom. 23 (Suppl 1) (2022) 525.
[142] A.H. Laustsen, V. Greiff, A. Karatt-Vellatt, S. Muyldermans, T.P. Jenkins, Animal immunization, in vitro display technologies, and machine learning for antibody discovery, Trends Biotechnol. 39 (12) (2021) 1263–1273.
[143] E. Asgari, M.R. Mofrad, Continuous distributed representation of biological sequences for deep proteomics and genomics, PLoS One 10 (11) (2015), e0141287.
[144] J.M. Shaver, J. Smith, T. Amimeur, Deep learning in therapeutic antibody development, Methods Mol. Biol. 2390 (2022) 433–445.
[145] D. Prihoda, J. Maamary, A. Waight, V. Juan, L. Fayadat-Dilman, D. Svozil, D. A. Bitton, BioPhi: a platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning, MAbs 14 (1) (2022) 2020203.