Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances in computational structure-based antibody design

Alissa M. Hummer[a], Brennan Abanades[a] and Charlotte M. Deane

Abstract

Antibodies are currently the most important class of biotherapeutics and are used to treat numerous diseases. Recent advances in computational methods are ushering in a new era of antibody design, driven in part by accurate structure prediction. Previously, structure-based antibody design has been limited to a relatively small number of cases where accurate structures or models of both the target antigen and antibody were available. As we move towards a time where it is possible to accurately model most antibodies and antigens, and to reliably predict their binding site, there is vast potential for true computational antibody design. In this review, we describe the latest methods that promise to launch a paradigm shift towards entirely *in silico* structure-based antibody design.

Addresses

Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, UK

Corresponding author: Deane, Charlotte M (deane@stats.ox.ac.uk)
(Hummer A.M.), (Abanades B.), (Deane C.M.)
[a] These authors contributed equally.

Current Opinion in Structural Biology 2022, **74**:102379

This review comes from a themed issue on **Engineering and Design**

Edited by **Sarel J. Fleishman** and **Roy A. Mariuzza**

For complete overview of the section, please refer the article collection - Engineering and Design

Available online 28 April 2022

https://doi.org/10.1016/j.sbi.2022.102379

Introduction

Computational structure-based antibody design is on the cusp of a revolution in which it should be possible to rapidly design binders for a desired target — whether to combat a new disease, or to facilitate molecular biology research — purely *in silico*. While this future may seem distant, the required data and computational tools are becoming readily available.

Antibodies are proteins produced by the immune system as part of the defence mechanism against foreign molecules and pathogens. A key property of antibodies, their ability to bind strongly and specifically to a target (antigen), has made them an important class of therapeutics [1], consistently topping the list of best-selling drugs [2]. Therapeutic antibodies have been developed for the treatment of a wide range of diseases, from viruses (including HIV and SARS-CoV-2) to cancers. Antibodies are also valuable tools in molecular biology research with applications including structural biology, functional assays and imaging [3].

However, traditional methods for antibody development, such as deriving antibodies from the hybridomas of inoculated animals or from library assembly followed by display techniques, are not only costly and time-consuming but also are not necessarily able to produce antibodies that bind to the desired site (epitope) on an antigen [4]. Computational antibody design methods, which have been part of therapeutic discovery for decades (for comprehensive overviews see Refs. [5,6]), offer a way to overcome these limitations. However, structure-based antibody design has been held back by the lack of accurate antibody and antigen structures.

Recent breakthroughs in computational structure prediction [7] have ushered in a new era where it is reasonable to assume that accurate structures will be available for most proteins [8]. This, in combination with the rapidly increasing numbers of available antibody sequences [9], will allow for large-scale structural antibody virtual screening, as is already done in small molecule drug development [10,11]. In this review, we describe how advances in protein structure prediction and other areas are bringing us closer to being able to entirely computationally design antibodies that bind strongly to a defined antigen epitope.

Structure prediction

Deep learning has revolutionised the field of protein structure prediction, evolving from adapting computer vision methods to designing complex task-specific machine learning algorithms [12–18]. For a review see Ref. [19]. Given a sequence of amino acids, methods like AlphaFold2 [7] are now capable of accurately predicting the structure of most proteins. For computational antibody design, this means that a structural

model of the target antigen will almost always be available [8]. However, methods like AlphaFold2 do have limitations in predicting physiologically relevant components or states of proteins. For example, they cannot yet model disordered or post-translationally modified (e.g. glycosylated) proteins [20,21].

For antibodies, the major remaining challenge in terms of structure prediction lies in accurately predicting the structure of the complementarity determining regions (CDRs), in particular CDR-H3 (Figure 1). The CDR loops tend to form most of the binding contacts with the antigen [22] and are thus of particular importance to model accurately. While it has long been possible to predict structural models of the framework region with sub-angstrom accuracy using homology modelling [23], the diverse set of possible conformations of the CDR loops makes these less amenable to prediction with homology-based methods [24].

Deep learning-based methods developed specifically to predict antibody structures [25–27] predict CDR loops more accurately than those trained for general structure prediction. By leveraging knowledge specific to antibody structures, these methods can also be substantially faster, enabling the rapid generation of large numbers of antibody structures. For example, while DeepAb [27] takes around 10 min per modelled Fv, and ABlooper [25] only 10 s to model all CDRs, AlphaFold2 [7] takes over an hour to model a single chain on the same device. With the arrival of large antibody sequence databases [9,28], fast methods like ABlooper [25] allow for accurate structural studies at previously unseen scales.

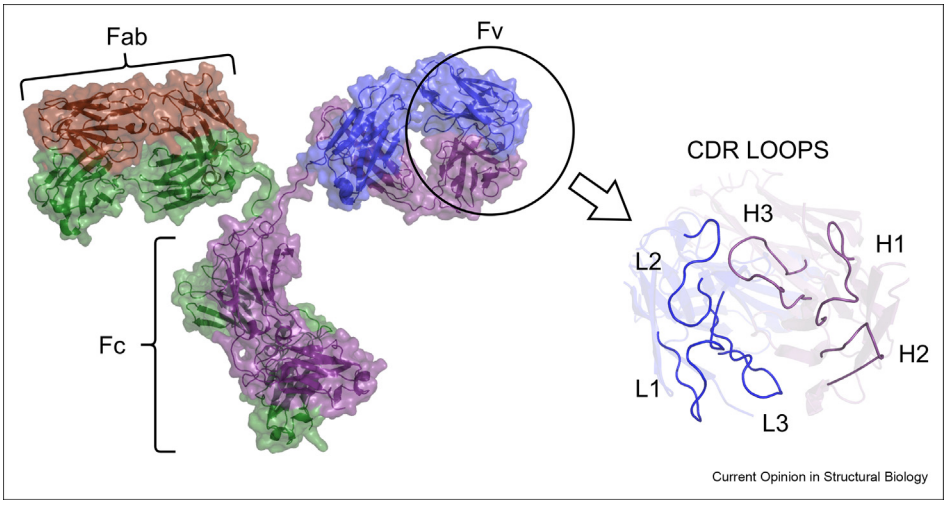
Structure-driven antibody design

Once structures have been obtained for the antibody and antigen, these can be used to assess binding potential (Figure 2). The residues involved in binding on the antibody are called the paratope and those on the antigen the epitope. Accurate prediction of the paratope and/or epitope is often one of the first steps in designing antibodies.

Paratope and epitope residues can be identified by solving the structure of a bound antibody-antigen pair. With the rapidly increasing size of antibody sequence datasets [9,28], the ability to predict paratopes in a high throughput manner will be essential for large-scale studies. Current methods for paratope prediction offer reasonable accuracy (e.g. Parapred [29]). Epitope prediction has proven to be more challenging, with a structure-based antibody-agnostic method like Epitope3D [30] only achieving a Matthews correlation coefficient of 0.45.

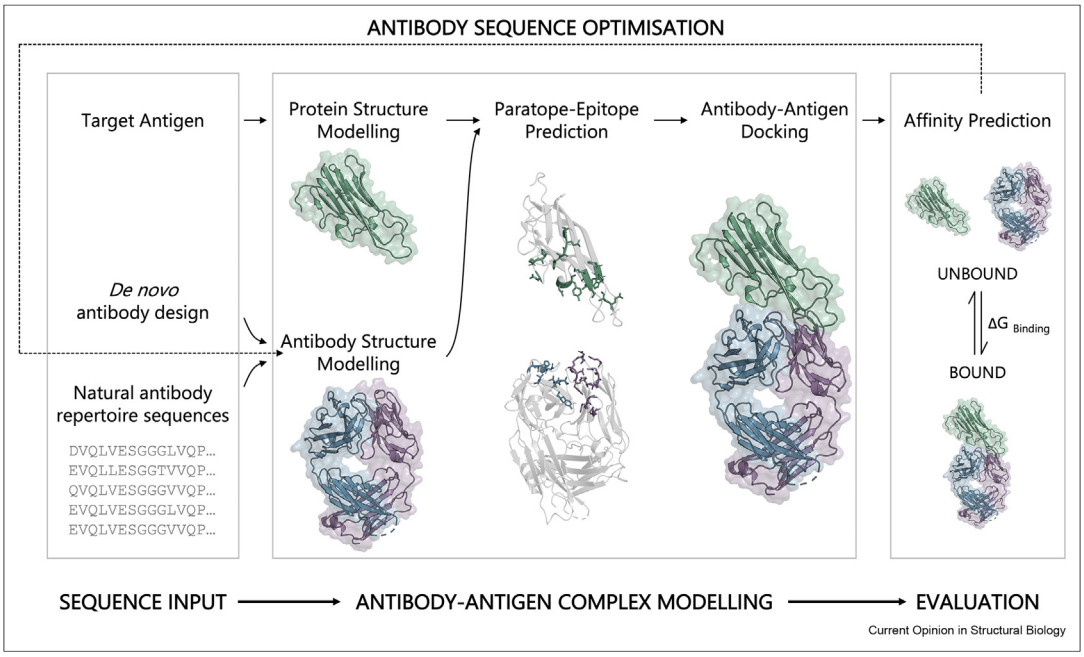
PECAN [31] is a deep learning-based method for paratope and epitope prediction. It outperformed all previously available methods for paratope prediction, achieving an Area Under the Curve-Precision Recall (AUC-PR) of 0.70 for the paratope and 0.21 for the epitope, by considering the structure of both the antibody and antigen. The authors of EPMP [32] take a different approach, claiming that while the position of the epitope is dependent on the structure of both the antibody and antigen, the position of the paratope is antigen-independent. Following this hypothesis, they developed two distinct architectures for paratope and

Figure 1



Schematic of a standard antibody, showing its two heavy chains (purple and green) and light chains (brown and blue). Antibodies contain two antigen binding regions (Fab) and one crystallisable domain (Fc). The area responsible for binding in the Fab is called the variable region (Fv), with most of the variability required for binding concentrated in six loops known as the complementarity determining regions (CDRs). The part of the Fv that does not belong to the CDRs is called the framework. On the right side of the figure, the structure of the Fv region is shown, with the CDR loops highlighted and labelled. PDB IDs: 1IGT (left), 15C8 (right).

Figure 2



Outline of a potential *in silico* structure-based design pipeline. Starting antibody sequences are obtained via *de novo* design or from natural antibody repertoire sequences. The antibody-antigen complex is then modelled. This could consist of separate steps of modelling the antibody and antigen structures, followed by paratope-epitope prediction and antibody-antigen docking, as visualised here. Alternatively, future work employing generative models may enable complex modelling in a single generative step. The resulting antibody constructs are then evaluated on the basis of binding affinity, as well as other desirable properties. The construct can be iteratively optimised by altering residues based on information obtained in the design procedure. Following mutagenesis, the modelled structure, predicted paratope and docked complex are updated and the binding affinity is re-evaluated. This process is repeated until a suitable antibody construct has been designed. PDB ID: 3WD5.

epitope prediction, achieving an AUC-PR of 0.75 for the paratope and 0.28 for the epitope. The current state-of-the-art method for epitope prediction is PInet [33], which achieves an AUC-PR of 0.37 using an antibody and antigen-aware geometric deep neural network.

However, even a perfect paratope-epitope predictor does not tell us the relative orientations of the antibody and antigen in the bound complex. The binding mode and relative positions of the antibody-antigen pair can be predicted via docking. Docking is computationally expensive, especially for large and flexible molecules such as antibodies and their protein targets [34]. Paratope and epitope prediction can, in addition to providing useful information for designing experiments, be used to limit the docking search space as they predict the binding interfaces [5].

Protein docking methods can be classified into two main categories, flexible and rigid body. There is a trade-off between the two approaches, with the latter being much faster — important for high-throughput methods — but less accurate. Docking is widely used in computational small molecule drug design but has been of more limited use in biotherapeutic discovery due to challenges in terms of accuracy [35,36]. Recently, deep learning has been applied to enhance antibody-antigen docking in a way which mimics advances to small molecule docking [37]. DLAB employs a convolutional neural network to improve docking pose ranking and identify antibody-antigen pairs which are more likely to be docked accurately [37]. This approach is promising for enhancing the virtual screening of antibodies against desired antigen targets. In separate approaches, AlphaFold2 has been used as a basis for protein multimer prediction [38,39]. While this has achieved success in general protein—protein docking, antibody-antigen complexes have proven much more challenging to model accurately [38,39].

Epitope and paratope (defined as CDR loops) information has been shown to be valuable for docking [37,40]. However, the accuracy of protein docking is still limited by algorithms and structures. Existing algorithms for docking do not yet perform sufficiently well for entirely *in silico* structure-based design [35,36]. Additionally, any error in the antigen or antibody structure — for example in the CDR-H3 loop — or conformational change on binding will hinder docking, particularly rigid body docking [36,41].

While structure modelling, paratope-epitope prediction and docking approaches are not yet perfect, foundations have been laid to identify binders for a given site through virtual screening [7,37,42]. However, as the methodology currently stands, the antibody initially identified is unlikely to be a strong binder. Weakly-binding antibodies can be *in silico* affinity maturated to strengthen their binding affinity for an antigen. Recent computational affinity maturation methods have been built from machine learning models, which predict the effect a mutation would have on binding affinity [43–45]. These models can then be used to evaluate mutations and identify those which increase binding affinity.

The primary components of antibody-antigen complex modelling – structure prediction, paratope-epitope prediction and docking (Figure 2) – could potentially be achieved simultaneously, in a single step, with a generative modelling approach. A recent study has suggested a method for co-designing antibody sequence and structure, with 3D coordinates updated in response to sequence changes [46]. However, there is need for further development, including the explicit consideration of the binding antigen structure and clearer guidelines for initialising structures, to apply this method for *de novo* design. Additionally, a thorough comparison of the accuracies and speeds of the sequential vs. generative approaches will be required.

Future directions

The overarching, long-term aim of structure-based antibody design is the ability to generate antibodies that will bind strongly to a desired target at a desired site, entirely computationally. There have been studies proposing ideas towards reaching *de novo* design [47–51]. For example, Liu and colleagues developed a hotspot grafting method, where a motif at the interface of a protein–protein interaction is selected for grafting onto a library of antibody scaffolds [47]. However, this approach will remain limited to epitopes for which an existing interaction interface with a binding partner is known. The approach suggested by Nimrod et al., in which antibody scaffolds with preliminary complementarity to the desired epitope are selected and docked, is more general [48]. Nevertheless, this strategy still draws from, and is thus limited by, existing antibody-antigen complex structures, in addition to relying on experimental yeast surface display for generating a binder from a starting scaffold [48]. There are a range of further proposed design methods, for example involving antibody assembly from CDR and framework fragments followed by sequence optimisation [49,52] or using deep learning models. In deep learning, there are a variety of ways in which antibody/protein space can be represented, and subsequently sampled from, both structurally (e.g. low-dimensional latent space capturing antibody structural fluctuations [51]; distance and orientation maps for optimisation and network 'hallucination' [53]; graphs (e.g. Refs. [7,25,31,33])) and sequence-based (e.g. sequence embeddings [54–57]). Additionally, first attempts have been made to investigate whether the antibody-antigen interaction is learnable and to work towards generalisations [58]. All of these studies have advanced *de novo* design, although future work is still needed for the computational, end-to-end design of high-affinity binders.

Virtual screening of natural antibody sequences presents another avenue for structure-based design. The availability of paired antibody sequence data is growing in both quantity and quality, thanks to advances in next generation sequencing techniques, although there still remains far more unpaired than paired sequence data available (e.g. Refs. [9,28]). A few methods have tried to overcome this by computationally pairing the data (e.g. Ref. [42]). Binders are already being identified from B-cell receptor sequences from convalescent patients (e.g. Refs. [42,59–61]). Furthermore, a comparison of therapeutic and natural repertoire sequences demonstrated that therapeutics in many cases exhibit high similarity to natural sequences [62]. These results suggest that using the vast available sequence resources as a virtual screening library is a promising way to begin the computational design process (Figure 2). The selected antibody sequences could then be modelled, docked and evaluated by predicting the binding affinity to the target antigen (Figure 2).

Several challenges still remain for true computational structure-based antibody design. While there has been great progress in protein structure prediction, current methods are not yet able to accurately predict the position of side chain atoms or structural changes on binding. For antibodies, accurately modelling the CDR-H3 loop remains a major obstacle. Additionally, improvements in paratope and epitope prediction, both in terms of accuracy and specificity (predicting the types of binding interactions for residues), will be needed to help improve docking for high-throughput virtual screening.

It will also be essential to have suitable metrics for evaluating designed antibodies. Tools for predicting antibody-antigen binding affinity could fill such a role, however both machine learning [63] and physics-based [64] methods currently struggle to achieve the required levels of accuracy. A hybrid method capable of combining the strengths of statistical and physics-based approaches (as done in Refs. [7,25,27] for structure prediction or in Ref. [65] to predict ligand binding affinity) may fill such a role. Advances in binding affinity prediction would enable discrimination between designed antibodies based on binding strength. As such, constructs carried forward for experimentation and further development could be

rationally selected. In this review we have focussed on the prediction of binding but the full design of any therapeutic would involve optimising for multiple characteristics, for example developability (e.g. Refs. [66,67]) and humanness (e.g. Refs. [68,69]).

An additional consideration is the shift in the biologics therapeutics field towards alternative antibody formats, such as single-chain nanobodies and asymmetric, multi-specific antibodies [70]. These formats may pose further difficulties, particularly given the even more limited availability of data (for example, SAbDab contains 4543 antibody Fv structures but only 923 nanobody (VHH) structures, as of February 2022 [71,72]). However, we expect the fundamentals of the binder design process should remain the same across formats, given the common goal of designing CDR-mediated interaction interfaces.

In order to facilitate AlphaFold2-level breakthroughs for the remaining steps in the design pipeline (Figure 2), a two-pronged approach of designing experiments to obtain high quality data — for example on interface residues and complex binding affinity — and concurrent advances in computational techniques will be required. While there are still areas to improve upon, research, data collection and computational method development over the past decades have laid a strong foundation for eventually achieving *in silico* antibody design.

Conflict of interest statement

Nothing declared.

Acknowledgements

Funding: This work was supported by the Medical Research Council [grant number: MR/N013468/1 awarded to A.M.H] and the Engineering and Physical Sciences Research Council [grant number: EP/S024093/1 awarded to B.A.].

References

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

* of special interest
** of outstanding interest

1. Chiu ML, Goulet DR, Teplyakov A, Gilliland GL: **Antibody structure and function: the basis for engineering therapeutics**. *Antibodies* 2019, **8**:55.

2. Lu RM, Hwang YC, Liu IJ, Lee CC, Tsai HZ, Li HJ, Wu HC: **Development of therapeutic antibodies for the treatment of diseases**. *J Biomed Sci* 2020, **27**:1–30.

3. Basu K, Green EM, Cheng Y, Craik CS: **Why recombinant antibodies — benefits and applications**. *Curr Opin Biotechnol* 2019, **60**:153–158.

4. Sormanni P, Aprile FA, Vendruscolo M: **Third generation antibody discovery methods:: in silico rational design**. *Chem Soc Rev* 2018, **47**:9137–9157.

5. * Norman RA, Ambrosetti F, Bonvin AMJJ, Colwell LJ, Kelm S, Kumar S, Krawczyk K: **Computational approaches to therapeutic antibody design: established methods and emerging trends**. *Briefings Bioinf* 2020, **21**:1549–1567.
This review gives a comprehensive overview of existing computational methods for therapeutic antibody design, as well as insights into emerging trends.

6. Roy A, Nair S, Sen N, Soni N, Madhusudhan MS: **In silico methods for design of biological therapeutics**. *Methods* 2017, **131**:33–65.

7. ** Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.
This study presents a deep learning method that has revolutionized the field of protein structure prediction, achieving high levels of accuracy.

8. ** Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, Bridgland A, Cowie A, Meyer C, Laydon A, *et al.*: **Highly accurate protein structure prediction for the human proteome**. *Nature* 2021, **596**:590–596.
The authors apply the state-of-the-art AlphaFold2 protein structure prediction method ([7]**) to the human proteome. They provide a residue-level metric for model confidence, which they make available alongside all modelled structures. The resulting database allows researchers to access predicted structures for nearly any protein without needing to run a modelling algorithm.

9. Olsen TH, Boyles F, Deane CM: **Observed Antibody Space: a diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences**. *Protein Sci* 2021, **31**:141–146.

10. Batool M, Ahmad B, Choi S: **A structure-based drug discovery paradigm**. *Int J Mol Sci* 2019, **20**:2783.

11. Sliwoski G, Kothiwale S, Meiler J, Lowe EW: **Computational methods in drug discovery**. *Pharmacol Rev* 2014, **66**:334–395.

12. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T: **Identification of direct residue contacts in protein-protein interaction by message passing**. *Proc Natl Acad Sci U S A* 2009, **106**:67–72.

13. Wang S, Sun S, Li Z, Zhang R, Xu J: **Accurate de novo prediction of protein contact map by ultra-deep learning model**. *PLoS Comput Biol* 2017, **13**, e1005324.

14. Xu J: **Distance-based protein folding powered by deep learning**. *Proc Natl Acad Sci U S A* 2019, **116**:16856–16865.

15. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Žídek A, Nelson AWR, Bridgland A, *et al.*: **Improved protein structure prediction using potentials from deep learning**. *Nature* 2020, **577**:706–710.

16. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D: **Improved protein structure prediction using predicted interresidue orientations**. *Proc Natl Acad Sci U S A* 2020, **117**:1496–1503.

17. Ingraham J, Riesselman A, Sander C, Marks D: **Learning protein structure with a differentiable simulator**. In *7th Int Conf Learn Represent ICLR* 2019; 2019.

18. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Dustin Schaeffer R, *et al.*: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science (80-)* 2021, **373**:871–876.

19. AlQuraishi M: **Machine learning in protein structure prediction**. *Curr Opin Chem Biol* 2021, **65**:1–8.

20. Perrakis A, Sixma TK: **AI revolutions in biology: the joys and perils of AlphaFold**. *EMBO Rep* 2021, **22**, e54046.

21. Jones DT, Thornton JM: **The impact of AlphaFold2 one year on**. *Nat Methods* 2022, **19**:11–26.

22. Sela-Culang I, Kunik V, Ofran Y: **The structural basis of antibody-antigen recognition**. *Front Immunol* 2013, **4**:302.

23. Leem J, Dunbar J, Georges G, Shi J, Deane CM: **ABodyBuilder: automated antibody structure prediction with data–driven accuracy estimation**. *mAbs* 2016, **8**:1259–1268.

24. Teplyakov A, Luo J, Obmolova G, Malia TJ, Sweet R, Stanfield RL, Kodangattil S, Almagro JC, Gilliland GL: **Antibody modeling assessment II. Structures and models**. *Proteins Struct Funct Bioinforma* 2014, **82**:1563–1582.

25. * Abanades B, Georges G, Bujotzek A, Deane CM: **ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation**. *Bioinformatics* 2022:btac016.

By using equivariant graph neural networks, the authors accurately predict the structure of CDR loops at previously unseen timescales.

26. Ruffolo JA, Guerra C, Mahajan SP, Sulam J, Gray JJ: **Geometric potentials from deep learning improve prediction of CDR H3 loop structures**. *Bioinformatics* 2020, **36**:I268–I275.

27. * Ruffolo JA, Sulam J, Gray JJ: **Antibody structure prediction using interpretable deep learning**. *Patterns* 2022, **3**:100406.

This method is the current state-of-the-art for antibody structure prediction.

28. Kovaltsuk A, Leem J, Kelm S, Snowden J, Deane CM, Krawczyk K: **Observed antibody space: a resource for data mining next-generation sequencing of antibody repertoires**. *J Immunol* 2018, **201**:2502–2509.

29. Liberis E, Velickovic P, Sormanni P, Vendruscolo M, Lio P: **Parapred: antibody paratope prediction using convolutional and recurrent neural networks**. *Bioinformatics* 2018, **34**:2944–2950.

30. da Silva BM, Myung Y, Ascher DB, Pires DEV: **epitope3D: a machine learning method for conformational B-cell epitope prediction**. *Briefings Bioinf* 2021, **23**:bbab423.

31. * Pittala S, Bailey-Kellogg C: **Learning context-aware structural representations to predict antigen and antibody binding interfaces**. *Bioinformatics* 2020, **36**:3996–4003.

By incorporating structural knowledge of both the antibody and antigen, the authors outperform all previous paratope and epitope prediction methods.

32. Del Vecchio A, Deac A, Liò P, Veličković P: **Neural message passing for joint paratope-epitope prediction**. *arXiv* 2021.

33. * Dai B, Bailey-Kellogg C: **Protein interaction interface region prediction by geometric deep learning**. *Bioinformatics* 2021, **37**:2580–2588.

This paper uses a geometric deep learning method to predict the interface between two proteins with a high level of accuracy.

34. Dauzhenka T, Kundrotas PJ, Vakser IA: **Computational feasibility of an exhaustive search of side-chain conformations in protein-protein docking**. *J Comput Chem* 2018, **39**:2012–2021.

35. Lensink MF, Brysbaert G, Mauri T, Nadzirin N, Velankar S, Chaleil RAG, Clarence T, Bates PA, Kong R, Liu B, *et al.*: **Prediction of protein assemblies, the next frontier: the CASP14-CAPRI experiment**. *Proteins Struct Funct Bioinforma* 2021, **89**:1800–1823.

36. * Guest JD, Vreven T, Zhou J, Moal I, Jeliazkov JR, Gray JJ, Weng Z, Pierce BG: **An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants**. *Structure* 2021, **29**:606–621. e5.

The authors assemble an expanded dataset of antibody-antigen complexes, for which unbound and bound structures are available, for benchmarking docking and affinity prediction. Additionally, they present an analysis into how conformational changes upon binding influence docking and affinity prediction.

37. * Schneider C, Buchanan A, Taddese B, Deane CM: **DLAB: deep learning methods for structure-based virtual screening of antibodies**. *Bioinformatics* 2021, **38**:377–383.

This paper applies a deep learning convolutional neural network framework for structure-based antibody-antigen docking, improving docked pose ranking and identifying antibody-antigen complexes for which docking is more accurate.

38. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, *et al.*: **Protein complex prediction with AlphaFold-Multimer**. *bioRxiv* 2021.

39. Yin R, Feng BY, Varshney A, Pierce BG: **Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants**. *bioRxiv* 2021.

40. Ambrosetti F, Jiménez-García B, Roel-Touris J, Bonvin AMJJ: **Modeling antibody-antigen complexes by information-driven docking**. *Structure* 2020, **28**:119–129. e2.

41. Zacharias M: **Accounting for conformational changes during protein-protein docking**. *Curr Opin Struct Biol* 2010, **20**:180–186.

42. Raybould MIJ, Marks C, Kovaltsuk A, Lewis AP, Shi J, Deane CM: **Public Baseline and shared response structures support the theory of antibody repertoire functional commonality**. *PLoS Comput Biol* 2021, **17**:1–23.

43. Myung Y, Rodrigues CHM, Ascher DB, Pires DEV: **mCSM-AB2: guiding rational antibody design using graph-based signatures**. *Bioinformatics* 2020, **36**:1453–1459.

44. Myung Y, Pires DEV, Ascher DB: **mmCSM-AB: Guiding rational antibody engineering through multiple point mutations**. *Nucleic Acids Res* 2021, **48**:W125–W131.

45. * Wang M, Cang Z, Wei G: **A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation**. *Nat Mach Intell* 2020, **2**:116–123.

This study presents a convolutional neural network-based model for predicting changes in binding affinity in response to mutation. The model is trained on topological barcodes obtained using a persistent homology approach to represent structural changes.

46. Jin W, Wohlwend J, Barzilay R, Jaakkola T: **Iterative refinement graph neural network for antibody sequence-structure Co-design**. *arXiv* 2021.

47. Liu X, Taylor RD, Griffin L, Coker S-F, Adams R, Ceska T, Shi J, Lawson ADG, Baker T: **Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping**. *Sci Rep* 2017, **7**:41306.

48. Nimrod G, Fischman S, Austin M, Herman A, Keyes F, Leiderman O, Hargreaves D, Strajbl M, Breed J, Klompus S, *et al.*: **Computational design of epitope-specific functional antibodies**. *Cell Rep* 2018, **25**:2121–2131.e5.

49. Aguilar Rangel M, Bedwell A, Costanzi E, Ricagno S, Frydman J, Vendruscolo M, Sormanni P: **Fragment-based computational design of antibodies targeting structured epitopes**. *bioRxiv* 2021.

50. Adolf-Bryfogle J, Kalyuzhniy O, Kubitz M, Weitzner BD, Hu X, Adachi Y, Schief WR, Dunbrack RL: **RosettaAntibodyDesign (RAbD): a general framework for computational antibody design**. *PLoS Comput Biol* 2018, **14**, e1006112.

51. Eguchi RR, Anand N, Choe CA, Huang P-S: **IG-VAE: generative modeling of immunoglobulin proteins by direct 3D coordinate generation**. *bioRxiv* 2020.

52. Baran D, Pszolla MG, Lapidoth GD, Norn C, Dym O, Unger T, Albeck S, Tyka MD, Fleishman SJ: **Principles for computational design of binding antibodies**. *Proc Natl Acad Sci U S A* 2017, **114**:10900–10905.

53. Anishchenko I, Pellock SJ, Chidyausiku TM, Ramelot TA, Ovchinnikov S, Hao J, Bafna K, Norn C, Kang A, Bera AK, *et al.*: **De novo protein design by deep network hallucination**. *Nature* 2021, **600**:547–552.

54. Olsen TH, Moal IH, Deane CM: **AbLang: an antibody language model for completing antibody sequences**. *bioRxiv* 2022.

55. Leem J, Mitchell LS, Farmery JHR, Barton J, Galson JD: **Deciphering the language of antibodies using self-supervised learning**. *bioRxiv* 2021.

56. Shuai RW, Ruffolo JA, Gray JJ: **Generative language modeling for antibody design**. *bioRxiv* 2021.

57. Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM: **Unified rational protein engineering with sequence-based deep representation learning**. *Nat Methods* 2019, **16**:1315–1322.

58. Akbar R, Robert PA, Pavlović M, Jeliazkov JR, Snapkov I, Slabodkin A, Weber CR, Scheffer L, Miho E, Haff IH, *et al.*: **A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding**. *Cell Rep* 2021, **34**:108856.

59. Wu Y, Wang F, Shen C, Peng W, Li D, Zhao C, Li Z, Li S, Bi Y, Yang Y, *et al.*: **A noncompeting pair of human neutralizing**

antibodies block COVID-19 virus binding to its receptor ACE2. *Science* 2020, **368**:1274–1278.

60. Bullen G, Galson JD, Hall G, Villar P, Moreels L, Ledsgaard L, Mattiuzzo G, Bentley EM, Masters EW, Tang D, *et al.*: **Cross-reactive SARS-CoV-2 neutralizing antibodies from deep mining of early patient responses**. *Front Immunol* 2021, **12**: 678570.

61. Raybould MIJ, Rees AR, Deane CM: **Current strategies for detecting functional convergence across B-cell receptor repertoires**. *mAbs* 2021, **13**:1996732.

62. Krawczyk K, Raybould MIJ, Kovaltsuk A, Deane CM: **Looking for therapeutic antibodies in next-generation sequencing repositories**. *mAbs* 2019, **11**:1197–1205.

63. Myung Y, Pires DEV, Ascher DB: **CSM-AB : graph-based antibody–antigen binding affinity prediction and docking scoring function**. *Bioinformatics* 2021, **38**:1141–1143.

64. Lees WD, Stejskal L, Moss DS, Shepherd AJ: **Investigating substitutions in antibody-antigen complexes using molecular dynamics: a case study with broad-spectrum, influenza A antibodies**. *Front Immunol* 2017, **8**:143.

65. Cleves AE, Johnson SR, Jain AN: **Synergy and complementarity between focused machine learning and physics-based simulation in affinity prediction**. *J Chem Inf Model* 2021, **61**: 5948–5966.

66. Raybould MIJ, Marks C, Krawczyk K, Taddese B, Nowak J, Lewis AP, Bujotzek A, Shi J, Deane CM: **Five computational developability guidelines for therapeutic antibody profiling**. *Proc Natl Acad Sci U S A* 2019, **116**:4025–4030.

67. Sormanni P, Aprile FA, Vendruscolo M: **The CamSol method of rational design of protein mutants with enhanced solubility**. *J Mol Biol* 2015, **427**:478–490.

68. Marks C, Hummer AM, Chin M, Deane CM: **Humanization of antibodies using a machine learning approach on large-scale repertoire data**. *Bioinformatics* 2021, **37**:4041–4047.

69. Prihoda D, Maamary J, Waight A, Juan V, Fayadat-Dilman L, Svozil D, Bitton DA, BioPhi: **A platform for antibody design, humanization and humanness evaluation based on natural antibody repertoires and deep learning**. *bioRxiv* 2021.

70. Raybould MIJ, Marks C, Lewis AP, Shi J, Taddese B, Deane CM, Bujotzek A: **Thera-SAbDab : the therapeutic structural antibody database**. *Nucleic Acids Res* 2020, **48**:383–388.

71. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, Shi J, Deane CM: **SAbDab: the structural antibody database**. *Nucleic Acids Res* 2014, **42**:1140–1146.

72. Schneider C, Raybould MIJ, Deane CM: **SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker**. *Nucleic Acids Res* 2022, **50**: D1368–D1372.