Bioinformatics, 36(13), 2020, 3996–4003
doi: 10.1093/bioinformatics/btaa263
Advance Access Publication Date: 22 April 2020
Original Paper

OXFORD

Structural bioinformatics

Learning context-aware structural representations to predict antigen and antibody binding interfaces

Srivamshi Pittala and Chris Bailey-Kellogg *

Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA

*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on October 20, 2019; revised on April 10, 2020; editorial decision on April 13, 2020; accepted on April 15, 2020

Abstract

Motivation: Understanding how antibodies specifically interact with their antigens can enable better drug and vaccine design, as well as provide insights into natural immunity. Experimental structural characterization can detail the 'ground truth' of antibody–antigen interactions, but computational methods are required to efficiently scale to large-scale studies. To increase prediction accuracy as well as to provide a means to gain new biological insights into these interactions, we have developed a unified deep learning-based framework to predict binding interfaces on both antibodies and antigens.

Results: Our framework leverages three key aspects of antibody–antigen interactions to learn predictive structural representations: (i) since interfaces are formed from multiple residues in spatial proximity, we employ graph convolutions to aggregate properties across local regions in a protein; (ii) since interactions are specific between antibody–antigen pairs, we employ an attention layer to explicitly encode the context of the partner; (iii) since more data are available for general protein–protein interactions, we employ transfer learning to leverage this data as a prior for the specific case of antibody–antigen interactions. We show that this single framework achieves state-of-the-art performance at predicting binding interfaces on both antibodies and antigens, and that each of its three aspects drives additional improvement in the performance. We further show that the attention layer not only improves performance, but also provides a biologically interpretable perspective into the mode of interaction.

Availability and implementation: The source code is freely available on github at https://github.com/vamships/PECAN.git.

Contact: cbk@cs.dartmouth.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

As one of its mechanisms to combat disease, the immune system develops B cells that secrete antibodies to specifically recognize and either neutralize or help drive functional responses against a pathogen. An antibody recognizes a particular region, called its epitope, on a particular part of the pathogen, called its antigen; the region of the antibody directly involved in the recognition is called its paratope. The interface between an epitope and paratope is crucial to the affinity and specificity of an antibody–antigen interaction, and thus the antibody's function. Characterizing antibody–antigen interactions at the epitope–paratope resolution can thus reveal mechanisms of immune recognition, and, over a set of antibodies, can even provide insights into the development of the immune response. For example, recent studies have revealed new insights into antibody evolution (Mishra and Mariuzza, 2018; Sok *et al.*, 2013), and have shown that those insights could be used for guiding the affinity maturation process using appropriate immunogens (Briney *et al.*, 2016). Such characterization can also benefit the development of therapeutics and vaccines. For example, therapeutic antibodies are being used to treat many different diseases (Carter, 2006; Holliger and Hudson, 2005), and early development processes typically yield large arrays of candidate antibodies from which to select. Understanding their different recognition mechanisms can aid selection and subsequent development. Similarly, subunit vaccines are being developed to train the immune system against a pathogen by mimicking an important part but without causing actual infection (Briney *et al.*, 2016; Delany *et al.*, 2014; Doria-Rose and Joyce, 2015). Understanding the recognition processes driving beneficial responses, as well as those that are not helpful, can guide the development of these vaccines so as to ensure the desired immune targeting.

Experimental structure determination methods, namely X-ray crystallography, nuclear magnetic resonance spectroscopy and cryoelectron microscopy, provide the gold standard for characterizing antibody–antigen binding modes (Bai *et al.*, 2015; Lee *et al.*, 2015). Unfortunately, they remain expensive and time consuming, and

cannot feasibly keep up with the exploding amount of antibody sequence data for which it is desirable to understand antigen recognition, e.g. the millions of sequences obtained from analysis of an immune repertoire (Miho *et al.*, 2018; Truck *et al.*, 2015; Zhu *et al.*, 2013). Alternative experimental methods like H–D exchange mass spectrometry (Gallagher and Hudgens, 2016) and alanine scanning (Weiss *et al.*, 2000) are faster and cheaper, and of lower resolution/confidence, but still require substantial experimental effort per target. Higher-throughput methods such as multiplexed surface plasmon resonance can characterize many interactions simultaneously but do not provide direct localization information (Brooks *et al.*, 2014; Safsten, 2009). Computational methods thus have the most promise to scale to characterization of large numbers of possible epitope–paratope interactions, but it is necessary to ensure that predictions provide sufficient grounds to support further investigations, in terms of overall accuracy as well as the underlying reasoning for a prediction.

Prediction of antibody–antigen binding interfaces can be seen as a special case of predicting protein–protein binding interfaces. However, as discussed above these particular interactions are of significant importance, and the interfaces have their own special characteristics (Esmaielbeiki *et al.*, 2016; Kunik and Ofran, 2013) (as does each different class of protein–protein interaction), specific methods have been developed for epitope prediction and others for paratope prediction. Many methods make predictions based on amino acid sequence alone, e.g. predicting epitopes based on neural networks (Saha and Raghava, 2006), support vector machines (SVMs) (El-Manzalawy *et al.*, 2008; Singh *et al.*, 2013), hidden Markov models (Zhao *et al.*, 2011) and random forests (Jespersen *et al.*, 2017), and paratopes using long short-term memory networks (Deac *et al.*, 2019; Liberis *et al.*, 2018) and random forests (Olimpieri *et al.*, 2013). Though sequence-based methods can perform well on paratope prediction, most sequence-based epitope predictions are limited to the special case of a sequentially contiguous epitope (Yao *et al.*, 2013), whereas in contrast, most epitopes are found to be conformational (distal in sequence, but close in 3D structure) (Regenmortel, 1996; Walter, 1986). Thus we and others focus on structure-based methods that leverage geometric information in making predictions. Fortunately, while the complex structure is not known, in most common scenarios, the structure of the antigen by itself is available, and antibody structure prediction techniques enable confident prediction of most of the antibody's structure (Sircar, 2012; Sircar *et al.*, 2009). We thus briefly review this body of most closely related work on structure-based prediction of epitopes and paratopes.

Docking: Many structure-based methods for epitope and paratope predictions rely on computational docking techniques, which estimate the most likely conformations of a complex based on complementarity (geometric, chemical, energetic) between the individual proteins in many possible poses (Chen *et al.*, 2003; Schneidman-Duhovny *et al.*, 2005; Sircar and Gray, 2010). The resulting docking models may be ranked using a scoring function incorporating many different geometric and physicochemical parameters; defining a good scoring function is a challenging task that typically relies on domain expertise (Pedotti *et al.*, 2011), and has been specialized for antibody–antigen docking (Brenke *et al.*, 2012). From the top-ranked conformations, regions on one protein that are close to the partner protein can be identified as binding interfaces. Thus antibody–antigen docking can simultaneously predict epitopes and paratopes. While docking achieves a fairly high recall rate (fraction of true interface residues that are computationally identified) if enough docking models are considered, the corresponding precision rate (fraction of computationally identified residues that are truly in the interface) is then fairly low. This has prompted the development of methods that design targeted mutagenesis experiments so as to evaluate predicted docking models and improve precision (Hua *et al.*, 2017).

Epitope prediction: Some approaches, e.g. PEPITO (Sweredoski and Baldi, 2008), ElliPro (Ponomarenko *et al.*, 2008), EPSVR (Liang *et al.*, 2010) and DiscoTope (Kringelum *et al.*, 2012), apply machine learning methods to structural features of the antigen's residues. These methods can be considered antibody-agnostic as they do not use information from the partner antibody, and thus just reveal parts of the antigen generally amenable to antibody binding (Sela-Culang *et al.*, 2015). For prediction of an epitope targeted by a particular antibody, the context of which residues are likely to be involved in the interaction can improve the prediction performance, as well as distinguish specificity differences among different antibodies. This aspect of antibody–antigen interactions was leveraged by the antibody-specific prediction method EpiPred (Krawczyk *et al.*, 2014) to achieve state-of-the-art performance. EpiPred first performs geometric matching of patches (e.g. based on docking models) and then scores residues on the antigen with a customized binding potential specific for antibodies and antigens.

Paratope prediction: Many paratope predictors focus on special regions on antibodies called complementarity determining regions (CDRs), as they are well-defined from sequence and constitute the majority of the paratope and the majority of the differences among antibodies driving antigen-specific recognition. In the nonparametric method Paratome (Kunik *et al.*, 2012), the query antibody's structure and sequence are compared against a nonredundant dataset of antibodies, and paratopes are predicted based on resemblance to those on the closest matching antibody. Antibody i-Patch (Krawczyk *et al.*, 2013) uses a scoring function derived from an analysis of antibody–antigen interactions in a nonredundant training set. Recently, Daberdaku and Ferrari (2019) achieved state-of-the-art performance with a method that applies SVMs to classify patches extracted from the surface of the antibody, based on rototranslationally invariant shape descriptors and other physicochemical properties representing the patches.

A common drawback of current structure-based methods for epitope and paratope predictions is the use of fixed representations, which can be limited by the extent of available domain knowledge. Furthermore, epitope and paratope predictions are treated as two separate tasks, leading to the use of different representations and prediction methods for antigens and antibodies. Sequence-based methods, Parapred (Liberis *et al.*, 2018) and AG-Parapred (Deac *et al.*, 2019), demonstrate the utility of learning representations for better paratope prediction. However, there are currently no methods to learn structural representations for either epitope or paratope prediction tasks. Recently, a spatial graph convolution network was proposed to learn structural representations of proteins for interface prediction in general protein–protein interactions (Fout *et al.*, 2017). While graph convolution networks can encode structural representations of residues with information from their spatial neighborhood, they do not encode the context of the target protein. As shown by current methods, embedding the correct context of the target protein can improve the prediction performance (Krawczyk *et al.*, 2013, 2014). Therefore, there is a need to develop methods for learning context-aware structural representations for epitope and paratope predictions.

In this work, we present a unified deep learning-based framework for learning context-aware structural representations of antigens and antibodies to predict their binding interfaces. Our framework consists of a novel combination of graph convolution networks, attention and transfer learning to capture several desired aspects of antibody–antigen interactions. We show that the models trained on our framework can overcome the limitations of current computational methods and achieve state-of-the-art performance on both epitope and paratope prediction tasks. We also show that the prediction performance of these networks is only marginally affected when using homology models of antibodies compared to that when using their crystallized structures. Using the attention layer, we demonstrate the ability of our framework to reveal the mode of interaction between antigens and antibodies, enabling a deeper study of the biological factors driving their interactions. Therefore, our framework improves prediction accuracy and provides interpretable results to expedite the process of large-scale antibody–antigen characterization.

2 Materials and methods

We propose a novel deep learning framework, PECAN, Paratope and Epitope AQ5prediction with graph Convolution Attention Network (Fig. 1), to learn the structural representations of antigens and antibodies to predict their binding interfaces (i.e. antigen epitopes and antibody paratopes). Our framework comprises three components to leverage biological insights: (i) graph convolutions to capture the spatial relationships of the interfaces, (ii) an attention layer to enable each protein's interface predictions to account for the potential binding context provided by its partner and (iii) transfer learning to leverage the larger set of data available for general protein–protein interactions to provide a baseline model to be fine-tuned with antibody–antigen data.

We use this general framework to train two separate networks for the two prediction tasks: (i) an epitope prediction network in which the antigen is the primary protein on which we want to predict the interface (epitope) and the antibody is the secondary protein providing the context for a suitable interface; (ii) a paratope prediction network with the antibody as primary for interface prediction (paratope) and the antigen secondary providing the context. We note that in both tasks, the interface labels of the secondary protein are hidden during training and prediction phases, forcing the attention layer to learn the correct context of the secondary protein in an unsupervised fashion.

Problem statement: The objective is to assign a label, either *interface* (positive class) or *non-interface* (negative class), to each residue of the primary protein.

A. Input representation: Each protein structure is represented as a graph, with nodes for the amino acid residues and edges between residues with Cβ–Cβ distance less than 10 Å. For Gly residues, which lack a β carbon, Cα was used. Associated with each node is a 62-dimension feature vector encoding important sequence and structural properties as used in Fout *et al.* (2017): (i) a one-hot encoding of the amino acid type ($d=20$); (ii) a conservation profile for that position across a set of homologous proteins returned by PSI-BLAST (Altschul *et al.*, 1997) ($d=20$); (iii) the absolute and relative solvent accessible surface area of the residue as computed by STRIDE (Heinig and Frishman, 2004) ($d=2$); (iv) a local amino acid profile indicating the number of times each amino acid type appears within 8 Å of the residue ($d=20$). The structure-based features (iii and iv) were calculated for each protein in isolation from its partner.

As antibody CDRs drive antigen-specific recognition and the rest of the antibody framework is quite similar across all antibodies, nodes in the antibody graph are limited to 'CDR clouds' as follows: (i) identify the six CDRs using the IMGT annotation tool (Lefranc *et al.*, 2003); (ii) for each CDR, consider two sequentially adjacent residues; (iii) further extend these sets to include all residues within 6 Å in the structure (the maximum of the minimum Cβ–Cβ distance between any two CDR residues in the training sets).

B. Neural network: The neural network consists of graph convolution, attention and fully connected layers. Given two input graphs, primary $P=\{p_i\}_{i=1}^{N}$ and secondary $S=\{s_i\}_{i=1}^{M}$, the network assigns to each node $p_i \in P$ a probability of belonging to the positive class (i.e. binding interface).

B.1 Graph convolutions: Graph convolution (Fout *et al.*, 2017) enables order-independent aggregation of properties over a neighborhood of residues that together contribute to the formation of a binding interface. For a node x_i and its receptive field consisting of K spatial neighbors $G_i=\{g_j\}_{j=1}^{K}$ from the input graph, the convolution operation results in a vector $\hat{z}_i \in \mathbb{R}^{v}$, where v is a specified number of filters for the layer (Eq. 1). The parameters of this operation represent the aggregation weight matrix $\mathbf{W^c}$ for the center node, the aggregation weight matrix $\mathbf{W^g}$ for the neighboring nodes, and the bias vector $\hat{b}$. Thus, the convolution operation for a node x_i results in a spatial vector representation $\hat{x}_i'$ in the latent space $\mathbb{R}^{v}$.

$$\hat{z}_i = \mathrm{ReLU}\left(\mathbf{W^c}\hat{x}_i + \frac{1}{|G_i|}\sum_{j=1}^{K}\mathbf{W^g}\hat{g}_j + \hat{b}\right) \quad (1)$$

Multiple layers can be stacked to produce high-level representations for each node. Each convolution layer has two weight-shared graph convolution modules, one for the primary graph and one for the secondary graph.

B.2 Attention: An attention layer encodes the context of the secondary graph in the residue-level representations of the primary graph, providing information for each primary residue about secondary residues that are likely to interact with it. An attention score a_{ij} is computed between all node pairs $p_i \in P$ and $s_j \in S$ after projecting them into a latent space via an attention weight matrix $\mathbf{W^a}$ (Eq. 2). The dimensions of $\mathbf{W^a}$ are determined by the number of neurons in the final convolution layer and the desired dimension of the latent space. This dot product style of computing attention scores is

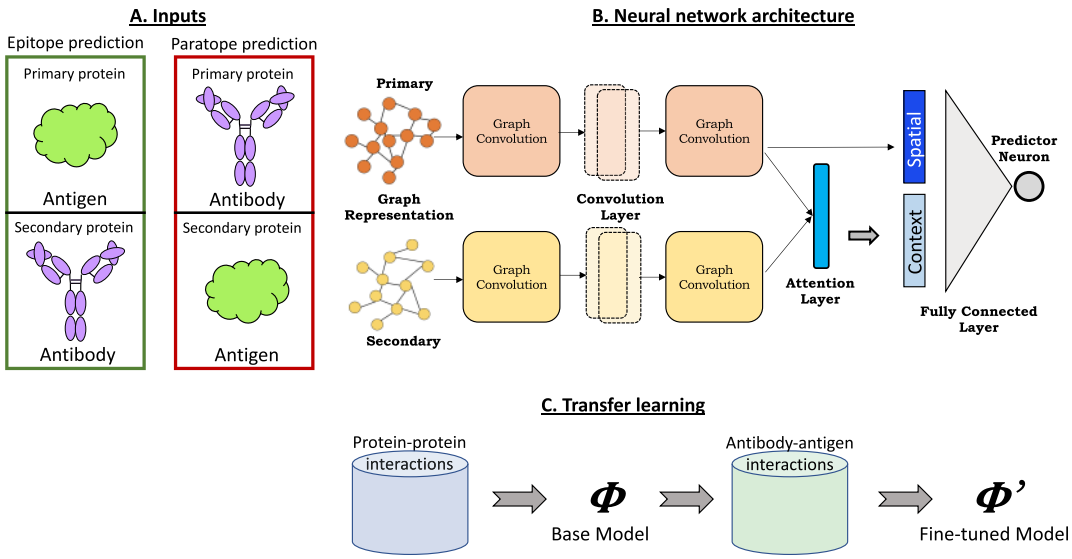

Fig. 1. Schematic overview of PECAN, Paratope and Epitope prediction with graph Convolution Attention Network. (**A**) The antibody and antigen protein structures are ordered as primary and secondary according to the task. (**B**) Graph representations of the two proteins are input to the neural network consisting of graph convolution, attention and fully connected layers. (**C**) For transfer learning, a base network trained on general proteins is used as initialization for training epitope and paratope prediction networks

used to directly estimate complementarity between hidden representations as in Luong *et al.* (2015) and Deac *et al.* (2019). The context vector $\hat{c}_i \in \mathbb{R}^v$ for node p_i is then computed by aggregating the node-level representations of S using normalized attention scores (Eq. 3). The normalization is performed to calibrate the score between each pair with respect to scores across all possible pairs. The normalized score α_{ij} can therefore be interpreted as a pairwise interaction potential between p_i and s_j.

$$a_{ij} = \mathrm{ReLU}(h_i^{\mathrm{T}} h_j), \quad \hat{h}_i = \mathbf{W}^{\mathbf{a}} \hat{p}'_i, \quad \hat{h}_j = \mathbf{W}^{\mathbf{a}} \hat{s}'_j. \tag{2}$$

$$\hat{c}_i = \sum_{j=1}^{M} \alpha_{ij} \hat{s}'_j, \quad \text{where } \alpha_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i,j=1}^{i=N,j=M} a_{ij}^2}} \tag{3}$$

B.3 Node classification: A final fully connected layer performs classification for each primary node p_i based on its spatial vector $\hat{p}'_i$ and context vector $\hat{c}_i$ (Eq. 4). An inverse logit function transforms each node's output y_i to indicate the probability of belonging to the positive class.

$$y_i = \hat{W}^{\mathrm{T}}(c_i || p'_i) + b, \quad \text{where} || \text{ indicates concatenation} \tag{4}$$

C. Transfer learning: A base network ϕ for interface prediction is learned for a relatively larger set of general protein–protein interactions. The learned weights from the base model are then used to initialize weights for training the two task-specific networks, essentially fine-tuning the general base network for epitope and paratope predictions using antibody–antigen data.

2.1 Implementation details

The framework was implemented in TensorFlow (Abadi *et al.*, 2015). Validation sets were used to find the optimal set of network training parameters for final evaluation. A grid search was performed over the following parameters: (i) Optimizer: Stochastic gradient descent, Momentum (Sutskever *et al.*, 2013), Adagrad (Duchi *et al.*, 2011) or Adam (Kingma and Ba, 2015); (ii) learning rates: 0.0001, 0.001, 0.005, 0.01, 0.05 or 0.1; (iii) batch size: 32, 64 or 128 and (iv) dropout: 0.5 or 0.8. For each combination, networks were trained until the performance on the validation set stopped improving or for a maximum of 250 epochs. For both epitope and paratope predictions, the best validation set performance was achieved when training till 120 epochs using the Momentum optimizer with Nesterov accelerated gradients (Su *et al.*, 2016) at a learning rate of 0.001, with batch size of 32 and 50% dropout rate. Training was carried out by minimizing the weighted cross-entropy loss function as in Fout *et al.* (2017). The same network settings were used for training on general protein–protein complexes, but the fine-tuning was carried out on antibody–antigen complexes for half the original time (i.e. 60 epochs). The graph convolution layers were set to have 32 filters and the latent space dimension for attention was also set to 32. All weight matrices were initialized as in He *et al.* (2015) and biases set to zero. For graph convolution, the receptive field (spatial neighborhood) for each node was set to include the 15 nearest nodes in the graph.

2.2 Datasets

For benchmarking purposes, we use the datasets provided with the state-of-the-art epitope predictor, EpiPred (Krawczyk *et al.*, 2014) and the state-of-the-art paratope predictor, by Daberdaku and Ferrari (2019), and compare the performance against these predictors. We note that the datasets were already curated to avoid overlap (too much similarity in antibodies or antigens) between testing and training.

Epitope prediction: The dataset from EpiPred (Krawczyk *et al.*, 2014) consists of 148 antibody–antigen complexes, 118 for training and 30 for testing, filtered to ensure that antigens shared no more than 90% pairwise sequence identity. As a separate validation set was not used, we constructed one from the antibody–antigen complexes in the Docking Benchmarking Dataset (DBD) v5 (Vreven *et al.*, 2015), ensuring that antigens in the validation set had no more than 25% pairwise sequence identity to those in the test set. This yielded 103 complexes for training, 29 for validation and 30 for testing. To test on homology models of the antibodies, in addition to the crystal structures provided in the EpiPred comparison, the PIGS server (Marcatili *et al.*, 2014) was used to generate homology models. In homology modeling, we ensured that structures of the same or closely related antibodies were not used as templates for modeling by considering only those antibody structures with less than 95% sequence identity to the query sequence (on heavy and light chains). The definition of the true epitope was still taken from the cocrystal structures. In addition, we compiled from SabDab (Krawczyk *et al.*, 2014) a new, distinct epitope prediction test set of 53 Ab:Ag complexes (Supplementary Table S1). In this new dataset, we ensured that no antigen had more than 50% sequence similarity with any antigen in this set and in the Epipred dataset (which used a cutoff of 80% sequence identity).

Paratope prediction: The dataset from Daberdaku and Ferrari (2019) consists of 471 antibody–antigen complexes, with 213 complexes for training, 106 for validation and 152 for testing, filtered to ensure that antibodies shared no more than 95% pairwise sequence identity. As our framework accepts only proteins, we discarded complexes with nonprotein antigens (e.g. DNA), resulting in 205 complexes for training, 103 for validation and 152 for testing. For evaluation on homology modeled structures, we used the same antibody models provided in the dataset, which had been built using the ABodyBuilder server (Leem *et al.*, 2016).

Transfer learning: To facilitate unbiased transfer learning, the DBD v5 dataset (Vreven *et al.*, 2015) was processed to discard complexes that were categorized as antibody–antigen, resulting in a dataset of 191 protein–protein complexes.

Following the previous studies (Daberdaku and Ferrari, 2019; Krawczyk *et al.*, 2014), residues were labeled as part of the interface if they had any non-hydrogen atoms within 4.5 Å of any non-hydrogen atoms of residues on the other protein.

Table 1 summarizes the percentage of residues that form the interface in training, validation and test sets for epitope and paratope prediction tasks.

3 Results

We evaluate our approach in head-to-head benchmark comparisons against state-of-the-art epitope and paratope predictors, showing that our unified framework outperforms approaches specifically targeted to each. Furthermore, we elaborate general precision and recall trends in the full architecture, as well as versions enabling characterization of the contributions of convolution, attention and transfer learning. Finally, we explore the ability of the attention layer to provide insights into the basis for the model's predictions.

Table 1. Summary of datasets used for training, validation and testing

Epitope prediction dataset		
Data split	# complexes	% epitopes
Training	103	8.9%
Validation	29	8.7%
Test	30	7.8%
Paratope prediction dataset		
Data split	# complexes	% paratopes
Training	205	8.8%
Validation	103	8.9%
Test	152	9.4%

3.1 Evaluation

Epitope and paratope prediction networks were trained using the validation-set optimized hyperparameters from above. Per-protein prediction performance was measured on the test sets by comparing predicted scores against ground truth labels.

While our networks output a probability for each residue, to enable a direct comparison to other epitope prediction methods we computed precision and recall by predicting as interface residues those with probability above 0.5. To further elaborate precision-recall trade-offs, we considered all such classification thresholds and computed the area under the precision recall curve (AUC-PR). Though some previous methods have used the area under the receiver operating characteristics curve (AUC-ROC) metric, AUC-PR is more suitable here as the emphasis is on predicting binding interfaces (positive class) and the negative class constitutes roughly 90% of the samples. To summarize the performance, AUC-PR was averaged over all proteins in the test set. To provide robust estimates of performance, the training and testing procedures were repeated five times, and the mean and standard error reported. As results from the previous studies were reported as single numbers, from either one evaluation or an average of multiple evaluations, the statistical significance of the difference between our performance and those single values was evaluated by way of a tail probability, the fraction of times our networks performed worse than the reported values.

Our evaluations included two learning schema: task-specific learning (i.e. just using antibody–antigen data) and transfer learning (i.e. fine tuning from a model trained with general protein–protein data). For each schema, five networks were evaluated: one network with a single fully connected layer (no convolution), one with a single graph convolution layer (Conv1-layer) and one with two (Conv2-layer), and likewise one network with the attention layer following a single graph convolution layer (Conv1-layer+Attn) and one following two convolution layers (Conv2-layer+Attn). Two-sided Wilcoxon-Mann-Whitney tests were performed to estimate the statistical significance of the performance differences between different models.

3.2 Epitope prediction

Figure 2 summarizes the epitope test set prediction performance for our different neural network implementations along with the state-of-the-art Epipred (Krawczyk *et al.*, 2014) and DiscoTope (Kringelum *et al.*, 2012). Our networks all perform better than Epipred and DiscoTope in terms of precision and recall (substantially stronger on recall alone) at the 0.5 cutoff (Fig. 2A) (tail probability $P < 0.001$). Elaborating performance for a range of cut-offs via AUC-PR in Figure 2B enables further comparison among our architecture implementations (these numbers are not available for the other methods). The network with an attention layer after two convolution layers achieves the best performance, confirming the utility of embedding the context of the target antibody into the representation of antigen's residues in addition to information from their spatial neighbors. The models predicting epitopes without the context of the antibody can be considered as predicting general hotspots on the antigens, which succeeds to a certain extent, as is the basis for generic (not antibody-specific) 'B cell epitope prediction' methods. The Conv2-layer+Attn network achieves similar prediction performance on the additional epitope prediction test set (Supplementary Fig. S7) as on the Epipred dataset, thus confirming our results on this completely separate dataset.

Furthermore, the improvements in performance of all models after transfer learning illustrates the benefits of leveraging data from general protein–protein interactions to establish a base model that can be fine-tuned with antibody–antigen data. The networks trained on general protein–protein interactions, however, show poor prediction performance compared to the task-specific and transfer learning networks (Supplementary Fig. S1A), demonstrating that these base models only served as good initialization points to networks that were trained for predicting epitopes.

To identify most importance features for prediction, networks were trained independently on each of the four feature types. Feature-specific networks did not perform as well as those networks trained on all features together (Supplementary Fig. S2A), thus showing that features extracted from both sequence and structure are necessary for better epitope prediction. We also show that we avoid overfitting to the training dataset by limiting our training runs to 120 epochs, after which the prediction performance on the validation set starts to worsen (Supplementary Fig. S8).

Similar prediction performance can be observed when homology modeled structures of antibodies, instead of their crystal structures,

A

	Task-specific learning		Transfer learning	
Method	**Precision**	**Recall**	**Precision**	**Recall**
No Convolution	0.142	0.637	0.146	0.570
Conv 1-Layer	0.148	0.727	0.155	0.568
Conv 1-Layer+Attn	0.147	0.688	0.159	0.625
Conv 2-Layer	0.152	0.692	0.157	0.730
Conv 2-Layer+Attn	0.154	0.672	0.158	0.628
EpiPred	0.136	0.436	-	-
DiscoTope	0.214	0.110	-	-

B

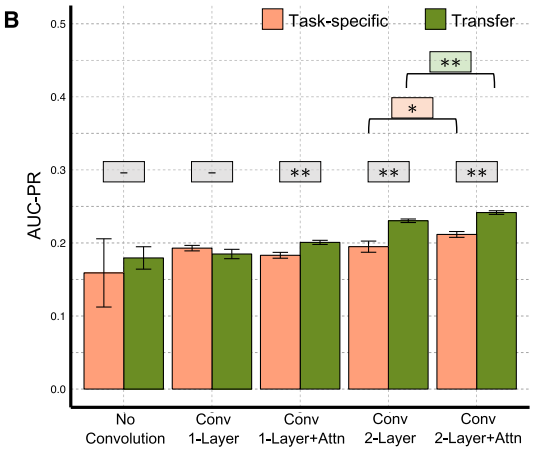

Fig. 2. Epitope prediction networks performance summary. (**A**) Precision and recall metrics obtained at decision boundary of 0.5. The measures for EpiPred (Krawczyk *et al.*, 2014) and DiscoTope (Kringelum *et al.*, 2012) were taken from Krawczyk *et al.* (2014). (**B**) AUC-PR measures obtained over a range of cutoffs. (Wilcoxon–Mann–Whitney *P*-values: $**P < 0.01$, $*P < 0.05$, -not significant)

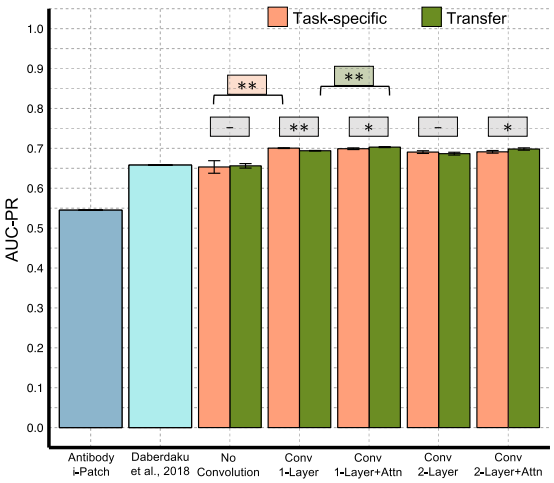

Fig. 3. Paratope prediction networks performance summary. Performance measures of Antibody i-Patch (Krawczyk *et al.*, 2013) and Daberdaku and Ferrari (2019) were taken from Daberdaku and Ferrari (2019). (Wilcoxon–Mann–Whitney *P*-values: $**P < 0.01$, $*P < 0.05$, -not significant)

are used to predict epitopes on corresponding antigens (Supplementary Fig. S3A). Interestingly, networks performed slightly better when using homology models compared to the crystal structures, which may be due to some relaxation of the structure during the modeling process. However, we found no significant correlation between the change in performance and the root mean squared distance between the models and crystal structures of CDR-H3 loops (Supplementary Fig. S3C).

3.3 Paratope prediction

Figure 3 summarizes the paratope test set prediction performance of our different neural networks, and state-of-the-art structure-based methods (Daberdaku and Ferrari 2019) and antibody i-Patch (Krawczyk *et al.*, 2013). To enable a direct comparison to previous studies, we predict for the entire structure of the antibody Fv region instead of just the CDR clouds, as described in our methods. Our networks perform better than the other methods on both AUC-PR and AUC-ROC (Supplementary Fig. S4) (tail probability $P < 0.001$), establishing the superior performance of learned features over predefined features as used by Daberdaku and Ferrari (2019). The no-convolution network performs equivalently to Daberdaku and Ferrari (2019) (tail probability $P = 0.4$), and the addition of convolution adds a statistically significant boost. The network with a single layer of convolution and attention achieves the best performance, but the attention layer provides only a small performance improvement over convolution. We hypothesize that since paratopes are mostly localized to regions around the CDRs, the context of the antigen may not provide much more information regarding exact paratope location than the structural properties already captured by convolution. Nonetheless, as we show in the next section, the attention layer offers the benefit of making the network interpretable, which can be a difficult task for convolution layers alone.

As was done in the case of epitope prediction, feature-specific networks were trained to identify most important feature types for paratope prediction. Interestingly, networks trained on conservation profile alone performed nearly as well as networks trained with all features (Supplementary Fig. S2B), thus showing that features extracted from sequence played a significant role in paratope prediction.

While we observe a minor drop in prediction performance when predicting on homology models compared to that obtained using crystal structures, our networks perform better than Daberdaku and Ferrari (2019) (Supplementary Fig. S5).

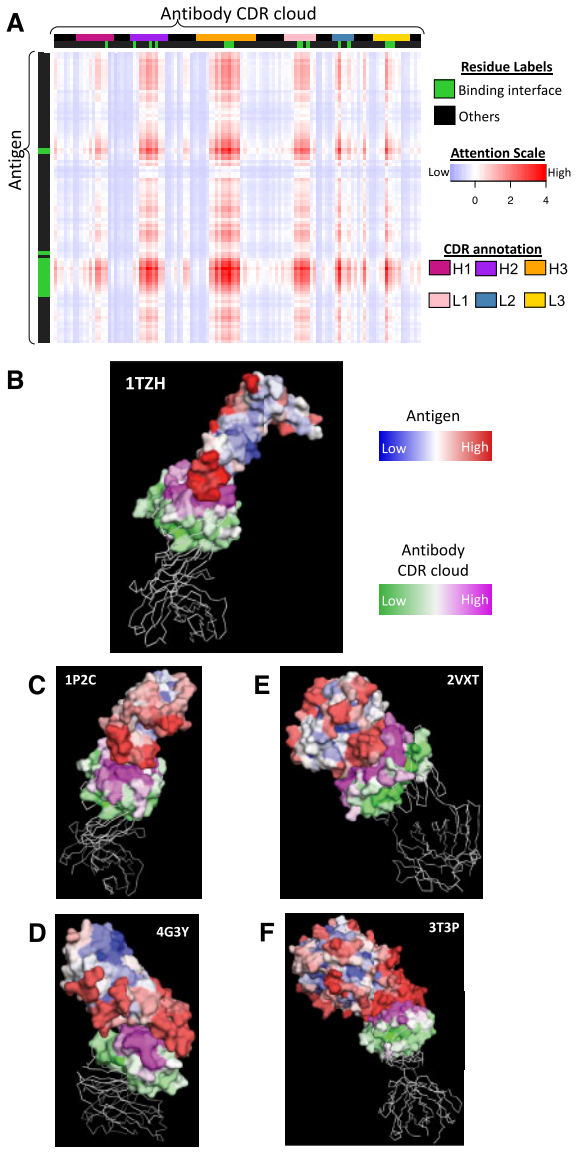

Fig. 4. Attention visualization (**A**) Heatmap of the attention score matrix for an antibody in complex with vascular endothelial growth factor A (VEGFA) (PDB ID 1TZH). The scores were normalized to have zero mean and unit variance, and truncated to the range [-6,6]. (**B–E**) Projection of max-pooled attention scores onto structures of example antigen and antibody complexes in the test set. Structural visualizations made using PyMOL (Schrödinger, 2015)

3.4 Assessing the contributions of attention

The attention layer provides the opportunity to study the mode of interaction by revealing the learned context of the target protein without requiring additional inference techniques. The attention score between every pair of residues can be visualized as a matrix, as Figure 4A illustrates for the complex on which our epitope prediction network performed best. In this heatmap, epitopes have a substantially distinct attention profile compared to other residues on the antigen, which results in improved epitope prediction (AUC-PR: 0.8) compared to convolution alone (AUC-PR: 0.48). These attention scores can further be projected onto the structures (Fig. 4B) by taking for each residue the maximum of its scores with partner residues. This projection shows that attention is high between residues in and around the actual interface region, suggesting that the attention layer encodes the correct context (i.e. paratopes) of the antibody for epitopes. The same pattern of high attention scores near the interface regions was also observed for other antibody–antigen complexes (Fig. 4C–F illustrate the next four top performers).

Intrigued by the attention layer's ability to localize the appropriate context during epitope prediction, we hypothesized that the same ability could benefit paratope prediction. We thus performed a 'cross-task evaluation', in which a network was trained to predict

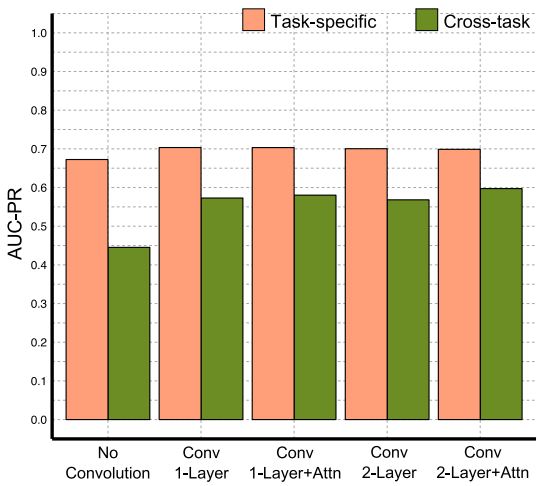

Fig. 5. Performance summary from 'cross-task evaluation', predicting paratopes using networks trained only for predicting epitopes

epitopes using the antigens in the epitope prediction training set. This epitope prediction network was then evaluated for its performance at also predicting *paratopes* for the antibodies in the paratope prediction test set—the reciprocal task to that for which it was trained. For reference, paratope prediction networks were trained on the antibodies from the epitope prediction training set and applied (as normal) to the CDR clouds of antibodies in the paratope prediction test set. Figure 5 summarizes the prediction performance of our different neural networks on the cross-task evaluation. As expected, the paratope prediction networks perform significantly better at predicting epitopes than do the networks trained to predict epitopes. However, the results from cross-task evaluation show that even though none of the networks were trained to predict paratopes, those with an attention layer perform better than convolution-only networks. This suggests that the attention layer is indeed able to better capture the specificity of antibody–antigen interactions, thereby also benefiting paratope prediction.

4 Conclusion

We have presented a unified deep learning framework for predicting binding interfaces on antibodies and antigens. Our results demonstrate that the networks learn structural representations that capture many desired aspects of antibody–antigen interactions and simultaneously achieve state-of-the-art performance on both epitope and paratope prediction tasks. We also show that the attention layer successfully encodes the context of partner proteins, improving prediction performance and providing an interpretable view of the mode of interaction. Future work includes including additional residue features while imposing sparsity constraints on the attention matrix, applying the same framework to other large protein families with specific recognition modes, and using predictions to focus docking as well as experimental evaluation.

Acknowledgements

The authors thank Margie Ackerman, Karl Griswold and members of the Bailey-Kellogg lab for helpful comments, and Casey Hua for sharing antibody art.

Funding

This research was supported in part by National Institutes of Health grant 2R01GM098977, with computational support from the anthill and discovery clusters at Dartmouth.

Conflict of Interest: none declared.

References

Abadi,M. *et al.* (2015) TensorFlow: large-scale machine learning on heterogeneous systems. www.tensorflow.org.

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Bai,X.C. *et al.* (2015) How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.*, **40**, 49–57.

Brenke,R. *et al.* (2012) Application of asymmetric statistical potentials to antibody–protein docking. *Bioinformatics*, **28**, 2608–2614.

Briney,B. *et al.* (2016) Tailored immunogens direct affinity maturation toward HIV neutralizing antibodies. *Cell*, **166**, 1459–1470.

Brooks,B.D. *et al.* (2014) High-throughput epitope binning of therapeutic monoclonal antibodies: why you need to bin the fridge. *Drug Discov. Today*, **19**, 1040–1044.

Carter,P.J. (2006) Potent antibody therapeutics by design. *Nat. Rev. Immunol.*, **6**, 343–357.

Chen,R. *et al.* (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, **52**, 80–87.

Daberdaku,S. and Ferrari,C. (2019) Antibody interface prediction with 3D Zernike descriptors and SVM. *Bioinformatics*, **35**, 1870–1876.

Deac,A. *et al.* (2019) Attentive cross-modal paratope prediction. *J. Comput. Biol.*, **26**, 536–545.

Delany,I. *et al.* (2014) Vaccines for the 21st century. *EMBO Mol. Med.*, **6**, 708–720.

Doria-Rose,N.A. and Joyce,M.G. (2015) Strategies to guide the antibody affinity maturation process. *Curr. Opin. Virol.*, **11**, 137–147.

Duchi,J. *et al.* (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, **12**, 2121–2159.

El-Manzalawy,Y. *et al.* (2008) Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.*, **21**, 243–255.

Esmaielbeiki,R. *et al.* (2016) Progress and challenges in predicting protein interfaces. *Brief. Bioinf.*, **17**, 117–131.

Fout,A. *et al.* (2017) Protein interface prediction using graph convolutional networks. In: Guyon,I. *et al.* (eds.) *Advances in Neural Information Processing Systems*. **Vol. 30**, Curran Associates, Inc., Red Hook, NY, pp. 6530–6539.

Gallagher,E.S. and Hudgens,J.W. (2016) Mapping protein–ligand interactions with proteolytic fragmentation, hydrogen/deuterium exchange-mass spectrometry. *Methods Enzymol.*, **566**, 357–404.

He,K. *et al.* (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *ICCV'15: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile*. IEEE, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.

Heinig,M. and Frishman,D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–502.

Holliger,P. and Hudson,P.J. (2005) Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.*, **23**, 1126–1136.

Hua,C.K. *et al.* (2017) Computationally-driven identification of antibody epitopes. *Elife*, **6**, e29023.

Jespersen,M.C. *et al.* (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.*, **45**, W24–W29.

Kingma,D.P. and Ba,J. (2015) Adam: A Method for Stochastic Optimization. In: *Published as a Conference Paper at the 3rd International Conference for Learning Representations, San Diego*.

Krawczyk,K. *et al.* (2013) Antibody i-Patch prediction of the antibody binding site improves rigid local antibody–antigen docking. *Protein Eng. Des. Select.*, **26**, 621–629.

Krawczyk,K. *et al.* (2014) Improving B-cell epitope prediction and its application to global antibody–antigen docking. *Bioinformatics*, **30**, 2288–2294.

Kringelum,J.V. *et al.* (2012) Reliable b cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput. Biol.*, **8**, 1–10.

Kunik,V. and Ofran,Y. (2013) 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**, 599–609.

Kunik,V. *et al.* (2012) Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. *Nucleic Acids Res.*, **40**, W521–W524.

Lee,H. *et al.* (2015) A cryo-electron microscopy study identifies the complete H16.V5 epitope and reveals global conformational changes initiated by binding of the neutralizing antibody fragment. *J. Virol.*, **89**, 1428–1438.

Leem,J. *et al.* (2016) ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. *MAbs*, **8**, 1259–1268.

Lefranc,M.-P. *et al.* (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and IG superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.

Liang,S. *et al.* (2010) EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics*, **11**, 381.

Liberis,E. *et al.* (2018) Parapred: antibody paratope prediction using convolutional and recurrent neural networks. *Bioinformatics*, **34**, 2944–2950.

Luong,M.-T. *et al.* (2015) Effective approaches to attention-based neural machine translation. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.

Marcatili,P. *et al.* (2014) Antibody modeling using the prediction of immunoglobulin structure (PIGS) web server [corrected]. *Nat. Protoc.*, **9**, 2771–2783.

Miho,E. *et al.* (2018) Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front. Immunol.*, **9**, 224.

Mishra,A.K. and Mariuzza,R.A. (2018) Insights into the structural basis of antibody affinity maturation from next-generation sequencing. *Front. Immunol.*, **9**, 117.

Olimpieri,P.P. *et al.* (2013) Prediction of site-specific interactions in antibody–antigen complexes: the proABC method and server. *Bioinformatics*, **29**, 2285–2291.

Pedotti,M. *et al*. (2011) Computational docking of antibody–antigen complexes, opportunities and pitfalls illustrated by influenza hemagglutinin. *Int. J. Mol. Sci*., **12**, 226–251.

Ponomarenko,J. *et al*. (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, **9**, 514.

Regenmortel,M.H.V. (1996) Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods*, **9**, 465–472.

Safsten,P. (2009) Epitope mapping by surface plasmon resonance. *Methods Mol. Biol*., **524**, 67–76.

Saha,S. and Raghava,G.P. (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, **65**, 40–48.

Schneidman-Duhovny,D. *et al*. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*., **33**, W363–367.

Schrödinger,L.L.C. (2015) The PyMOL molecular graphics system, version 1.8.

Sela-Culang,I. *et al*. (2015) Antibody specific epitope prediction-emergence of a new paradigm. *Curr. Opin. Virol*., **11**, 98–102.

Singh,H. *et al*. (2013) Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One*, **8**, e62216–8.

Sircar,A. *et al*. (2009) RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Res*., **37**, W474–W479.

Sircar,A. (2012) Methods for the homology modeling of antibody variable regions. *Methods Mol. Biol*., **857**, 301–311.

Sircar,A. and Gray,J.J. (2010) SnugDock: paratope structural optimization during antibody–antigen docking compensates for errors in antibody homology models. *PLoS Comput. Biol*., **6**, e1000644–13.

Sok,D. *et al*. (2013) The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. *PLoS Pathog*., **9**, e1003754.

Su,W. *et al*. (2016) A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *J. Mach. Learn. Res*., **17**, 1–43.

Sutskever,I. *et al*. (2013) On the importance of initialization and momentum in deep learning. In: Dasgupta,S. and McAllester,D. (eds.) *Proceedings of the 30th International Conference on Machine Learning, Volume 28 of Proceedings of Machine Learning Research*, PMLR, Atlanta, Georgia, USA, pp. 1139–1147.

Sweredoski,M.J. and Baldi,P. (2008) PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics*, **24**, 1459–1460.

Truck,J. *et al*. (2015) Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *J. Immunol*., **194**, 252–261.

Vreven,T. *et al*. (2015) Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *J. Mol. Biol*., **427**, 3031–3041.

Walter,G. (1986) Production and use of antibodies against synthetic peptides. *J. Immunol. Methods*, **88**, 149–161.

Weiss,G.A. *et al*. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA*, **97**, 8950–8954.

Yao,B. *et al*. (2013) Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. *PLoS One*, **8**, 1–4.

Zhao,L. *et al*. (2011) Antibody-specified B-cell epitope prediction in line with the principle of context-awareness. *IEEE/ACM Trans. Comput. Biol. Bioinf*., **8**, 1483–1494.

Zhu,J., *et al*.; NISC Comparative Sequencing Program. (2013) Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc. Natl. Acad. Sci. USA*, **110**, 6470–6475.