

Machine learning approaches for interpretable antibody property prediction using structural data

Kevin Michalewicz, Mauricio Barahona, and Barbara Bravi

Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

{k.michalewicz22,m.barahona,b.bravi}@imperial.ac.uk

Abstract. Understanding the relationship between antibody sequence, structure and function is essential for the design of antibody-based therapeutics and research tools. Recently, machine learning (ML) models mostly based on the application of large language models to sequence information have been developed to predict antibody properties. Yet there are open directions to incorporate structural information, not only to enhance prediction but also to offer insights into the underlying molecular mechanisms. This chapter provides an overview of these approaches and describes two ML frameworks that integrate structural data (via graph representations) with neural networks to predict properties of antibodies: ANTIPASTI predicts binding affinity (a global property) whereas INFUSSE predicts residue flexibility (a local property). We survey the principles underpinning these models; the ways in which they encode structural knowledge; and the strategies that can be used to extract biologically relevant statistical signals that can help discover and disentangle molecular determinants of the properties of interest.

Keywords: antibody, structural biology, molecular design, interpretability, machine learning, binding affinity, flexibility, graph representations

1 Introduction

Antibodies are key actors in the adaptive immune system. Antibodies are proteins secreted by B cells that can bind with high affinity and specificity to cognate targets (*antigens*) at specific sites (the *epitope*). As such, antibodies are central to triggering the immune response and clearing infection, and are sought-after design targets for research and therapeutic applications.

Antibodies are Y-shaped proteins, where each tip of the Y is formed by a heavy and a light chain, each split into constant and variable regions (Figure 1). The variable domains carry the molecular determinants of antigen recognition, and each of its chains is organized into three highly variable complementarity-determining regions (CDRs) and four surrounding framework regions (FRs) that stabilize the structural scaffold [1, 2]. CDRs constitute the majority of the antigen-binding site (the *paratope*); among them, the CDR3 of the heavy chain

(CDR-H3), often the main contributor to antigen binding [3], stands out for its sequence diversity, which is actively generated within single individuals by the VDJ recombination process [4].

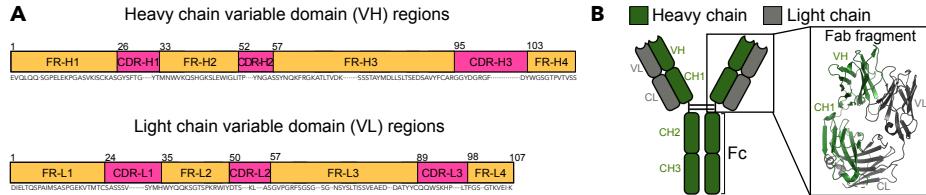


Fig. 1. Antibody sequence and structure. (A) Each variable domain of the heavy (VH) and light (VL) chains contains 3 CDRs and 4 FRs, defined via the Chothia position numbering scheme [5]. (B) The structure of an antibody is Y-shaped, with two identical antigen binding fragments (Fab) and one *crystallizable* fragment (Fc). Each Fab consists of the variable (VH and VL) and first constant domains (CH1 and CL), while the Fc comprises the CH2 and CH3 domains.

Recently, there has been a surge in the development of computational antibody design tools aimed at selecting and optimizing candidate antibodies *in silico* in terms of key properties such as binding affinity, binding specificity, solubility and stability [6–8]. These methods are critical to filter large candidate pools to prioritize variants most likely to succeed in downstream experimental validation. Machine Learning (ML) methods hold great promise in this setting, thanks to their success in protein-related prediction tasks [9–11].

Among the properties that are a focus of computational modeling, binding affinity is essential to antibody design, as it governs the strength of antigen engagement, antigen/antibody dose requirements, as well as antibody developability trade-offs in multiple trait optimization protocols [12–14].

Early antibody property prediction methods were primarily sequence-based, benefiting from large-scale datasets available at the sequence level [15, 16]. More recently, several ML models have incorporated both amino acid sequence and structural information for antibody property prediction [17, 18], including binding affinity and specificity [19, 20]. Yet, despite the steadily expanding amount of structural data, including AlphaFold models [21, 22], the potential of structure-informed prediction remains underexploited. This is partly because training models on fully spatially resolved structures can be computationally intensive and data-hungry. Network models and geometric graphs, built from relative positions, distances and interactions within proteins, provide natural abstractions to account for structural information in a parsimonious yet effective way for protein modeling within ML pipelines [23, 24].

The importance of interpretability in ML for scientific inference is broadly appreciated [25]. In the context of antibody discovery, models that are able to highlight the specific regions or interactions explaining their predictions offer

valuable mechanistic insight, which in turn can strengthen model trustworthiness. The present chapter builds on this view, discussing ML principles and practices that encode protein structural knowledge via network or geometric graph constructions, while enabling biophysical interpretation in the context of antibody modeling.

Protein property prediction tasks can be divided into *global* or *local*. Global properties describe a characteristic of the entire molecule, such as its stability or solubility, while local properties are defined at the level of individual structural units, such as individual atoms, residues or groups of residues. For example, single-residue properties include features like solvent accessibility [26], conservation score [27], secondary structure or local flexibility. Interestingly, local flexibility, which can be quantified through crystallographic B-factors, has been suggested to be linked to global properties such as binding affinity [28], as it shapes conformational selection and interface complementarity in antibody-antigen complexes [29–31].

We exemplify these two prediction categories through two ML structure-based methods that we have recently developed: ANTIPASTI [32] and INFUSSE [33]. ANTIPASTI (ANTIBody Predictor of Affinity from STructural Information) exemplifies this philosophy for the prediction of a global property, binding affinity. It does so by creating representations of antibody-antigen complexes based on a dynamical interpretation of Elastic Network Models (ENMs) and processing them through a compact Convolutional Neural Network (CNN) that is made interpretable through model-agnostic and model-dependent strategies. INFUSSE (Integrated Network Framework Unifying Structure and Sequence Embeddings) provides a complementary case study at the single-residue level by integrating sequence-level embeddings with a diffusive Graph Convolutional Network (diff-GCN) defined on geometric graphs to predict B-factors in antibody-antigen complexes.

The chapter is organized as follows. Section 2 introduces protein structure and sequence representations that serve as model inputs, including network abstractions, Normal Mode Analysis, and data-driven embeddings. Section 3 briefly presents ANTIPASTI and INFUSSE model architectures and details how they leverage structural and sequence information for antibody property prediction. Section 4 discusses biophysical interpretability strategies, covering both model-agnostic and model-dependent approaches. Finally, Section 5 concludes with a summary and discussion.

2 Protein Structure and Sequence Representations

2.1 Network Models of Protein Structures

The vastness of protein conformational space and associated computational cost make it challenging to model protein structures with full 3D spatial resolution at the atomistic level [23, 34, 35]. Instead of an all-atom representation, it is possible to work with coarse-grained protein structure representations, focusing only on

the heavy atoms or, most commonly, just one atom per amino acid, *i.e.*, the α -carbon atoms. Such coarse-grained, residue-level protein models have proven effective in various applications, including protein folding [36–38], aggregation studies [39], protein-protein docking [40], and protein flexibility prediction [41].

Normal Mode Analysis. Well-established coarse-grained models of protein structures are formulated in terms of Elastic Network Models (ENMs) [42, 43], *i.e.*, harmonic networks of residues. A prime example is the Anisotropic Network Model (ANM) [44], which employs a simplified central potential but preserves the representation of anisotropic motions of the particles. Normal Mode Analysis (NMA) is then applied to ENMs to capture dynamical fluctuations around equilibrium through a set of generalized coordinates (Normal Modes, NM) [45]. In our biomolecular modeling, NMs are useful to describe small-scale molecular deformations in relation to biological function [46–48]. Although early applications of NMA were performed at the atomistic level, where each atom was treated as a particle [49], residue-level NMA on coarse-grained ENMs offers a computationally efficient alternative that captures collective motions relevant to protein function [44, 50, 51].

Let $\mathbf{r} = [\mathbf{r}_1 \dots \mathbf{r}_N]^\top \in \mathbb{R}^{3N}$ be the vector containing the 3D coordinates of the α -carbons of the N residues of a protein, and let \mathbf{r}^* denote that vector at equilibrium (Figure 2A). The coarse-grained protein structure can then be viewed as a system of N interacting particles under a potential energy dependent on the residue positions, $U(\mathbf{r})$. Under small spatial displacements from equilibrium, the potential energy can be written (to second order) as:

$$U(\mathbf{r}) \simeq U(\mathbf{r}^*) + (\mathbf{r} - \mathbf{r}^*)^\top \nabla U(\mathbf{r})|_{\mathbf{r}^*} + \frac{1}{2}(\mathbf{r} - \mathbf{r}^*)^\top \nabla^2 U(\mathbf{r})|_{\mathbf{r}^*} (\mathbf{r} - \mathbf{r}^*) . \quad (1)$$

The potential energy of the equilibrium configuration, $U(\mathbf{r}^*)$, can be shifted to zero without loss of generality [52], and the gradient term $\nabla U(\mathbf{r})|_{\mathbf{r}^*}$ also cancels since \mathbf{r}^* is a local minimum of the potential energy. Therefore, only the second order term survives, and the potential energy close to equilibrium can be approximated as a harmonic potential, *i.e.*, a quadratic form of the Hessian matrix $\nabla^2 U(\mathbf{r}) \in \mathbb{R}^{3N \times 3N}$ evaluated at \mathbf{r}^* .

The solutions to the equations of motion for small displacements of the residues under the quadratic potential (equation 1) take the form:

$$\boldsymbol{\delta}(t) := \mathbf{r}(t) - \mathbf{r}^* \propto \sum_{l=0}^{3N-1} C^{(l)} \cos(\omega^{(l)} t + \phi^{(l)}) \mathbf{a}^{(l)} . \quad (2)$$

In other words, the small displacements (or fluctuations) of the residues $\boldsymbol{\delta}(t)$ can be expressed as linear combinations of NMs indexed by l . Here, $C^{(l)}$ and $\phi^{(l)}$ are, respectively, the amplitude and phase of the l^{th} NM, $\mathbf{a}^{(l)} \in \mathbb{R}^{3N}$ is the l^{th} eigenvector of the Hessian matrix $\nabla^2 U(\mathbf{r})|_{\mathbf{r}^*}$ and $\omega^{(l)}$ is its associated eigenvalue, setting the natural frequency of the NM [53]. It is important to note that the first six NMs are trivial, as they are associated with rotational and translational invariances [54].

Ref. [55] proposed a particular form of the potential $U(\mathbf{r})$:

$$U(\mathbf{r}) = \sum_{\substack{i,j=1 \\ r_{ij}^* < \epsilon}}^N k(r_{ij}^*)(r_{ij} - r_{ij}^*)^2, \quad (3)$$

where $r_{ij} := \|\mathbf{r}_{ij}\|_2 = \|\mathbf{r}_i - \mathbf{r}_j\|_2$ is the Euclidean distance between particles i and j , with r_{ij}^* the distance at equilibrium; $\epsilon = 7\text{\AA}$ is a cut-off radius; and $k(r_{ij}^*)$ is the strength of the pairwise elastic force given by an α -carbon force field derived by fitting to the Amber94 potential [56]:

$$k(r) = \begin{cases} 8.6 \times 10^2 r - 2.39 \times 10^3 & \text{for } r < 4\text{\AA} \\ 128 \times 10^4 r^{-6} & \text{otherwise} \end{cases} \quad (4)$$

Using the NMIs, the *correlation of the fluctuations* of residues i and j can be computed as:

$$X_{ij} := \frac{\langle \boldsymbol{\delta}_i \cdot \boldsymbol{\delta}_j \rangle}{\langle \boldsymbol{\delta}_i^2 \rangle^{1/2} \langle \boldsymbol{\delta}_j^2 \rangle^{1/2}} \in [-1, 1], \quad \text{with} \quad \langle \boldsymbol{\delta}_i \cdot \boldsymbol{\delta}_j \rangle = k_B T \sum_{l=0}^{3N-1} \frac{\mathbf{a}_i^{(l)} \cdot \mathbf{a}_j^{(l)}}{\omega^{(l)}}, \quad (5)$$

where \cdot is the dot product in \mathbb{R}^3 ; the angular brackets indicate an ensemble average [57]; k_B is the Boltzmann constant; T is the absolute temperature; and $\mathbf{a}_i^{(l)}$ contains the three coordinates of the l^{th} eigenvector of the Hessian matrix $\nabla^2 U(\mathbf{r})|_{\mathbf{r}^*}$ corresponding to residue i (and similarly for $\mathbf{a}_j^{(l)}$). In this way, the correlations X_{ij} provide a global representation of the effective linkage between residues mediated by the small-displacement dynamics of the protein.

Graph-based representations. Protein structures at the residue level can also be represented more abstractly through an undirected graph \mathcal{G} (Figure 2A), where each of the N nodes represents an amino acid and e_{ij} , the edge connecting the nodes i and j , has weight $w(e_{ij})$. For unweighted graphs, $w(e_{ij}) = 1$ if there is an edge e_{ij} and $w(e_{ij}) = 0$ otherwise. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ of \mathcal{G} is symmetric with entries $A_{ij} = w(e_{ij})$. Another matrix associated with \mathcal{G} is the $N \times N$ Laplacian matrix $\mathbf{L} := \mathbf{D} - \mathbf{A}$, where $\mathbf{D} = \text{diag}(\mathbf{A} \mathbf{1}) \in \mathbb{R}^{N \times N}$ and $\mathbf{1}$ is the $N \times 1$ vector of ones (*i.e.*, \mathbf{D} contains the node degrees on the diagonal). For undirected graphs, \mathbf{L} is symmetric and positive semi-definite [58]. The graph Laplacian matrix is a central concept in spectral graph theory and is directly linked to diffusive processes on graphs, as its eigenvalues determine how fast such processes converge to equilibrium [59, 60]. It has also been shown that, although simplified, such protein graph representations capture information about geometric neighborhoods and weighted local energetic interactions, which can deliver important functional and structural insights [23, 34].

The simplest graph construction for protein structures is the geometric ϵ -ball graph [61], whereby residues within an ϵ distance of each other are assigned an

unweighted edge. This graph has the following adjacency matrix:

$$A_{ij} = \begin{cases} 1 & \text{if } r_{ij} < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The cutoff is typically chosen in the range $\epsilon \in [8\text{\AA}, 10\text{\AA}]$, following Ref. [62].

An alternative geometric construction is a simple *weighted* Gaussian graph with (full) weighted adjacency matrix given by:

$$A_{ij} = \exp(-r_{ij}^2/\eta^2), \quad (7)$$

with length scale $\eta = 8\text{\AA}$ as in Ref. [63].

More sophisticated approaches construct atomistic graphs with energy-weighted edges (reflecting different chemical bonds and physicochemical interactions), which can be subsequently coarse-grained to the residue level [23, 64].

Machine-learned structure embeddings. Recently, ML models trained on a very large portion of the Protein Data Bank (PDB, [65]) have been leveraged to encode structures into fixed-size embeddings, producing high-dimensional residue-level representations that are enriched with spatial information [66]. Such structure-aware embeddings have been applied to tasks including structure similarity assessment [67], structure searching [68], property prediction [69, 70], and domain classification [66]. Transformer-based models for protein modeling like Refs. [21, 22, 71, 72] learn *state* and *pair* embeddings, which correspond to per-residue and residue-residue properties of the protein structure, respectively (see Figure 2B). Structure embeddings of this type can be integrated into graph-based frameworks, as they encode geometric and interaction context analogous to biophysical descriptors, yet learned directly from data [73].

2.2 Sequence Encodings

Protein sequences can be represented in several ways to serve as input for ML models (see Figure 2C for a schematic summary). The most basic representation, which does not consider relationships between residues, is *one-hot encoding*, a binary vector representation in which each amino acid is given a unique index in a fixed dictionary of size F_{aa} [74] (typically $F_{aa} = 20$, corresponding to the standard amino acid types). Given a sequence \mathbf{x}_s of length N , $\Phi(\mathbf{x}_s) \in \{0, 1\}^{N \times F_{aa}}$ is called the one-hot encoded version of \mathbf{x}_s .

The biochemical information carried by sequences is also routinely described by considering a set of physicochemical properties for each amino acid, such as aromaticity [75], hydrophobicity [76, 77] or composite measures like Kidera factors [78]. Beyond commonly used scales, hundreds of additional descriptors can be generated, covering properties such as mutability, free energy and steric bulk. These have been classified into several categories: conformation, energy, composition, volume, structure-activity, shape and polarity [79]. The drawback of hand-picked feature approaches is that the chosen properties may not be

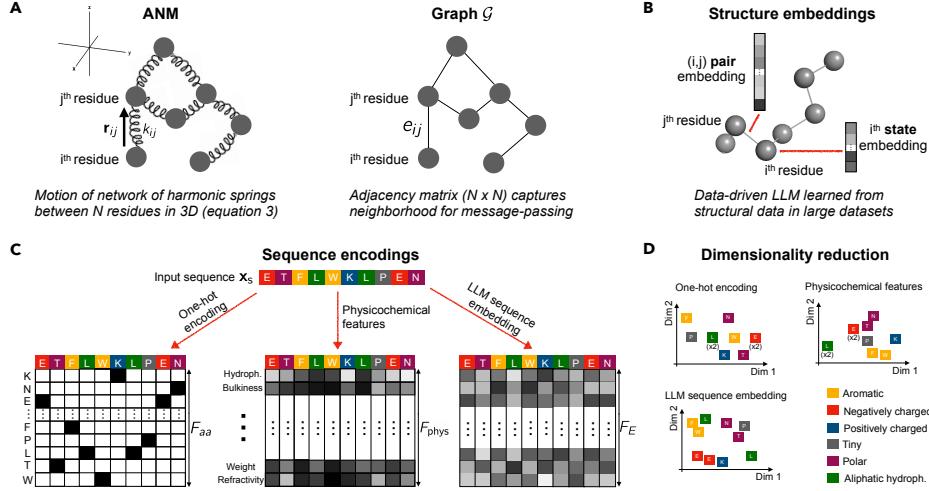


Fig. 2. Protein representations. (A) *Left*: In ENMs, a protein structure is modeled as a network in which the nodes (typically residues) are connected by springs that represent the elastic force acting between them. For two connected nodes i and j , the spring constant k_{ij} quantifies the elastic interaction strength. The ANM is a particular case where $k_{ij} = k(r_{ij}^*)$, with r_{ij}^* denoting the equilibrium distance between nodes i and j . *Right*: Graph representation of a protein structure as a network of N nodes (typically residues) connected by edges e_{ij} with weight $w(e_{ij})$ with adjacency matrix $A_{ij} = w(e_{ij})$. The diffusive dynamics and message-passing behavior of this system is governed by the associated graph Laplacian \mathbf{L} . (B) Structure embeddings (data-driven and derived by LLMs) capture both per-residue (*state*) and residue-residue (*pair*) properties of the protein structure. (C) A protein sequence x_s can be represented using one-hot encoding, physicochemical feature sets, or context-rich embeddings learned by an LLM. (D) Dimensionality reduction applied to the representations in C facilitate inspection through visualization in a lower-dimensional space. One-hot encodings do not capture similarity in physicochemical properties (denoted by colors) between residues. The F_{phys} features quantifying physicochemical properties exhibit consistent groupings. The LLM-learned embeddings capture both biochemical similarity and context-dependence, so that the representation of a given residue is influenced by its physicochemical properties and by the sequence to which it belongs (context).

the most relevant or informative for the specific task under consideration [80]. Another option is to represent sequences at the *atomic level* via SMILES [81], a line notation that encodes molecules as strings that specify atom types and bond connectivity.

Although the above representations (*embeddings*) are interpretable, they are semantically limited: one-hot encodings ignore residue-residue relationships and lack biochemical context, while physicochemical feature sets depend heavily on the quality and appropriateness of the features chosen for the representation. To overcome these limitations, there has been a recent surge in the use of advanced deep learning methods based on Large Language Models (LLMs) to create data-driven *sequence embeddings* that capture complex statistical dependencies between residues along the sequence obtained from training on very large datasets. Indeed, LLMs specifically tailored to proteins and trained in a self-supervised way [82] can learn, through their intermediate layers, high-dimensional sequence embeddings that are shown to recapitulate evolutionary, structural and functional signals (see Ref. [83] and Figure 2C). Examples include ProtBERT [84], ESM [85], and ProGen [86]. For a protein sequence \mathbf{x}_s of length N , we denote the embedding given by a hidden layer of a pre-trained LLM as $\mathbf{E}_s \in \mathbb{R}^{N \times F_E}$ with F_E being the embedding dimension (*e.g.*, $F_E = 1024$ for ProtBERT [84]). Unlike the interpretable but semantically limited representations discussed above, LLM-derived embeddings are not directly interpretable in physicochemical terms, yet they encode much richer information learned from large-scale sequence statistics, enabling meaningful comparisons across sequences. Each residue in the sequence is represented by a context-dependent embedding, meaning that the same amino acid type acquires a different representation depending on its surrounding sequence environment. This allows such embeddings to capture functional, evolutionary, and sequence-context information beyond the simple amino acid identity, as it typically becomes clear when visualized in reduced dimensions (Figure 2D). These high-dimensional, semantically rich representations thus provide residue-level features that can be directly used when modeling proteins as graphs.

However, although such sequence models have shown their utility and power for many tasks, they tend to perform accurately only when there are well-defined and statistically strong signals of residue covariation that are informative about the structural and functional properties modeled. Antibody interfaces, on the other hand, involve mainly the CDRs, whose behavior is hard to infer from sequence alone due to the fact that they are regions with high amino acid composition variability and conformational flexibility acting as binding interfaces in contact with external structural motifs (epitope). All these properties together make it difficult to rely on sequence-only methods and motivate the need for including and encoding structural information into our descriptions, *e.g.*, as graphs or network protein representations.

In ANTIPASTI, inputs are *dynamics-aware* residue-residue correlation maps derived from normal modes of an ANM (equation 5). In INFUSSE, the graph is *geometry-driven*, defined either by an ϵ -ball graph (equation 6) or a smoothly decaying weighted-Gaussian adjacency (equation 7). Both networks of residues

represent different types of structural signals: ANTIPASTI’s input maps capture patterns of correlated fluctuations, which involve global dynamical properties of the network, whereas INFUSSE is based on spatial proximity (representing packing [87] or biophysical interactions [64]) as the basis for propagating node features into the graph neighborhood via message-passing.

3 Integrating Structural Information for Predictive Modeling of Antibody Properties

We now introduce the architectures of ANTIPASTI and INFUSSE as two examples of ML pipelines that integrate structural information for the prediction of, respectively, global and local protein properties. While exemplified here with *regression* tasks, *i.e.*, prediction of continuous properties, the techniques can be adapted to classification tasks, *i.e.*, prediction of discrete class labels.

3.1 ANTIPASTI for Binding Affinity Prediction: An Example of Global Property Prediction

Consider the following general regression problem. Given the input-output pairs $(\mathbf{X}^{(i)}, y^{(i)})_{i=1}^{N_s} \subseteq \mathcal{X} \times \mathbb{R}$, the aim is to learn a function $f \in \mathcal{F}$ such that it captures the input-output relationship, *i.e.*, $f(\mathbf{X}^{(i)}) \approx y^{(i)}$, where N_s is the number of available data entries and \mathcal{F} is the model class, *e.g.*, neural networks. For 2D image data, for instance, $\mathcal{X} = \mathbb{R}^{N \times N}$, where N^2 is the number of pixels.

ANTIPASTI trains a CNN in a supervised fashion (see Figure 3A,C) on input pairs given by: (i) NM correlation maps $\mathbf{X} \in \mathbb{R}^{N \times N}$ (equation 5), computed from antibody-antigen structures, and (ii) experimentally measured values of the binding dissociation constant (K_D). The model comprises a bank of n_f $k \times k$ convolutional filters, followed by a ReLU activation function and a $p \times p$ max pooling layer with stride s . The latter is followed by a bias-free fully-connected layer from which the ANTIPASTI’s global prediction, the value of $\log_{10}(K_D)$, is computed as follows.

Let $\mathbf{z} \in \mathbb{R}^M$ be the flattened input of the ANTIPASTI fully-connected layer, produced by the network’s convolutional layer (with n_f filters and the non-linear activation) and the max pooling layer applied to a NM correlation map \mathbf{X} . Let $\mathbf{w} \in \mathbb{R}^M$ be the weights of the fully-connected layer. Then, since there is no bias, the predicted value of K_D (denoted by \hat{K}_D) is obtained simply from:

$$\log_{10}(\hat{K}_D) = \mathbf{w}^\top \mathbf{z} = \sum_{i=0}^{M-1} w_i z_i, \quad \text{with } M = n_f \left\lfloor \frac{N - k + 1 - p}{s} + 1 \right\rfloor^2. \quad (8)$$

The value of M arises because every convolution operation yields a matrix of shape $(N - k + 1) \times (N - k + 1)$ which is downsized through max pooling with kernel shape $p \times p$ and stride s .

Convolutions are powerful when dealing with protein structures due to their robustness to detect features regardless of their exact position. This makes it

possible to find common structural motifs relevant to protein function [88, 89]. Further, given the multiscale nature of proteins, the hierarchical architecture of CNNs is ideal for capturing features at different levels of resolution [23, 90].

Antibody-antigen binding typically follows an *induced fit* mechanism, characterized by small, localized arrangements such as CDR loop adjustments or slight domain reorientations [91–93]. As NMA captures small-amplitude displacements around the equilibrium position, it is well-suited for binding affinity prediction, a relationship also proposed in Refs. [52, 94]. The induced fit model inherently assumes small structural adjustments and is thus embedded in the choice of the harmonic ANM potential energy (equation 3), which serves as a biophysical prior for the ML model.

After learning the CNN model, we tested the accuracy of ANTIPASTI for prediction of the binding affinity constant K_D on curated, unseen test datasets from SAbDab [95, 96]. Figure 3B shows the values \hat{K}_D predicted by the CNN with pooling on an unseen test set for five training/test splits compared to the true values K_D , achieving a Pearson correlation of $R = 0.86$ when combining data across splits. The random splits were prepared robustly to ensure there is less than 90% antibody sequence identity and that the antigens are different between the training and test data.

For comparison, a linear regression approach to predicting K_D from NMs achieves a mean $R = 0.64$ (see Ref. [32]), substantially lower than ANTIPASTI, indicating that it is crucial to include non-linearities and layer types with robust feature detection properties, such as convolutional layers.

3.2 INFUSSE for B-factor Prediction: An Example of Single-Residue Property Prediction

In contrast to the prediction of a global protein property achieved by ANTIPASTI, we now consider the prediction of a residue-level property. This can be formulated as a graph node regression task, where each node of the graph is associated with a residue.

For a protein with N residues, let $\mathbf{X} \in \mathbb{R}^{N \times F}$ be a matrix that contains F features for each residue, *e.g.*, a one-hot encoding $\Phi(\mathbf{x}_s)$ or a residue-level embedding obtained from learned representations \mathbf{E}_s of the protein sequence \mathbf{x}_s (Figure 2C). Let $\mathbf{L} \in \mathbb{R}^{N \times N}$ be the Laplacian matrix of the graph \mathcal{G} , derived from the atomic coordinates of the residues, *e.g.*, the Cartesian coordinates of the N α -carbons $\mathbf{r} \in \mathbb{R}^{3N}$, which can be found in PDB files. Finally, let $\mathbf{B} \in \mathbb{R}^N$ be a vector containing the target values that we aim to predict for each node, *i.e.*, at the local residue level.

The INFUSSE architecture consists of a *sequence block* and a *graph block* that are learned sequentially and combined to produce the predictions:

$$\hat{\mathbf{B}} := \text{INFUSSE}(\mathbf{x}_s, \mathbf{r}) := S_{\text{block}}(\mathbf{x}_s, \mathbf{E}_s) + G_{\text{block}}(\mathbf{r}, \mathbf{X}(\mathbf{x}_s, \mathbf{E}_s)). \quad (9)$$

The sequence block S_{block} integrates one-hot encoded residue information and LLM embeddings through a series of learnable transformations that produce an enriched residue-level representation fine-tuned for the prediction task.

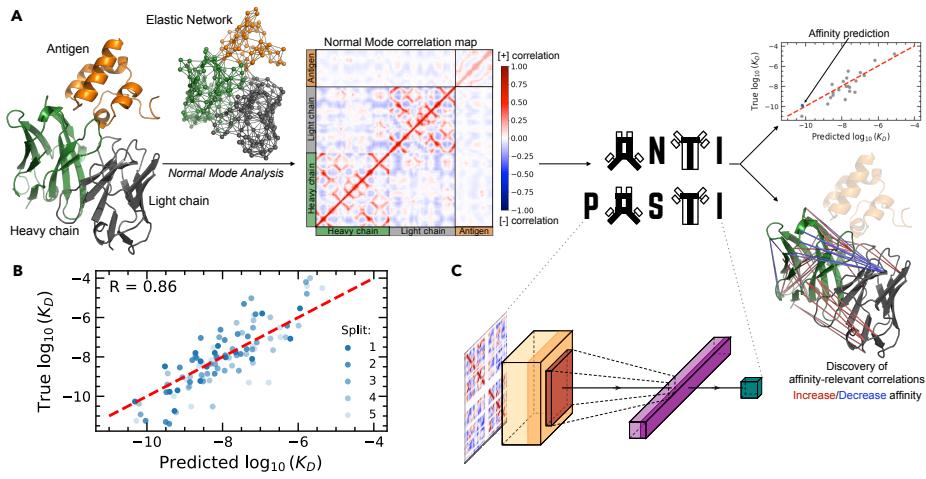


Fig. 3. Overview of ANTIPASTI. (A) From the PDB structure of the antibody-antigen complex, an elastic network model representation is created and a map of correlations between residues is calculated from normal modes, which carry the antigen's imprint on the antibody residues. Gaps are then added in place of absent residues through an alignment, and the antigen residues are removed from the correlation map to produce the input image to the ANTIPASTI's CNN architecture (see C). ANTIPASTI processes this input image and yields a binding affinity prediction and a map of affinity-relevant correlations. (B) ANTIPASTI predictions on the test set are compared to the ground-truth affinities for five training/test splits. (C) ML architecture of ANTIPASTI. Figure adapted from Ref. [32].

The graph block corresponds to a two-layer diff-GCN [97]: $G_{\text{block}}(\mathbf{r}, \mathbf{X}) := \text{diff-GCN}_t(\mathbf{L}(\mathbf{r}), \mathbf{X}) := e^{-t\mathbf{L}} \text{ReLU}(e^{-t\mathbf{L}} \mathbf{X} \mathbf{W}^{(0)}) \mathbf{W}^{(1)}$, where $\mathbf{W}^{(l)} \in \mathbb{R}^{F_{l+1} \times F_l}$ is the matrix of learnable weights for the l^{th} layer of the diff-GCN. Note that the matrix $e^{-t\mathbf{L}}$ is the diffusion transition associated with the graph Laplacian \mathbf{L} , and $t \in \mathbb{R}_{>0}$ is a learnable scale parameter.

GCNs differ fundamentally from CNNs. Instead of applying spatially localized filters, GCNs update node features via a message-passing scheme derived from a first-order approximation of spectral graph convolutions [98, 99], such that each node aggregates information from its neighbors. In the diff-GCN variant, this propagation is further modulated by the graph diffusion operator $e^{-t\mathbf{L}}$, where the parameter t controls the effective range of information flow.

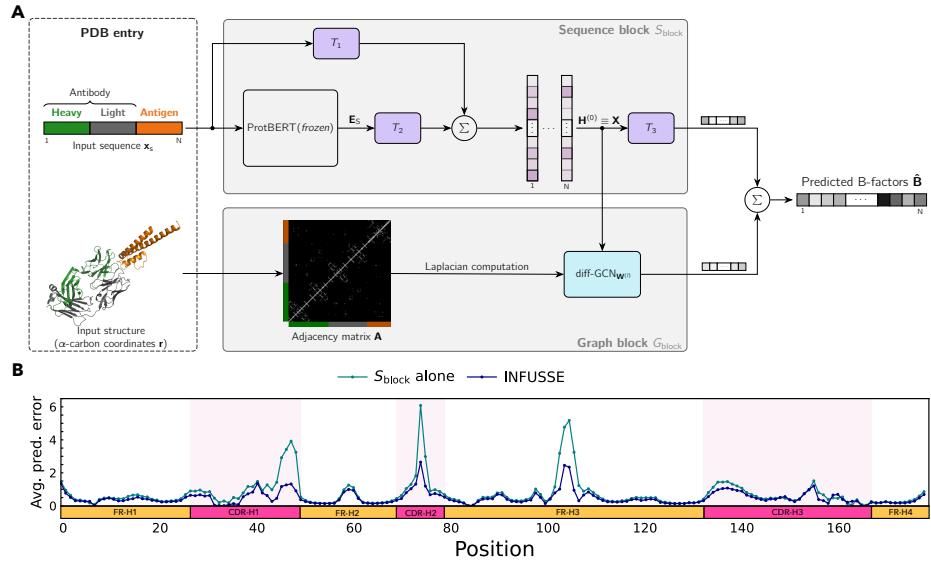


Fig. 4. Overview of INFUSSE. (A) In the INFUSSE architecture, a frozen LLM (ProtBERT) and a diffusive Graph Convolutional Network (diff-GCN) are combined to predict B-factors for antibody-antigen complexes. In the sequence block (S_{block}), input sequences are encoded by the frozen ProtBERT model and passed through a learnable non-linear layer T_2 , then summed with their one-hot encoded version transformed by another learnable non-linear layer T_1 , producing enriched sequence embeddings \mathbf{X} that are further transformed by a learnable non-linear layer T_3 . The resulting representations are summed with the output of the graph block G_{block} , *i.e.*, of the diff-GCN with learnable parameters t , $\mathbf{W}^{(0)}$ and $\mathbf{W}^{(1)}$ that takes \mathbf{X} from S_{block} as input node features and leverages the Laplacian of a geometric graph constructed from the antibody-antigen structure. (B) Prediction errors of INFUSSE, $\varepsilon_{\text{INFUSSE},n}^{(q)}$, and S_{block} alone, $\varepsilon_{S_{\text{block}},n}^{(q)}$, for each position n of the heavy chain antibody variable region averaged over the samples $q \in \mathcal{Q}$. Figure adapted from [33].

We have applied the INFUSSE model (Figure 4A) to the prediction of experimental standardized B-factors, which can also be extracted directly from PDB entries [33]. To train INFUSSE, we retrieved antibody-antigen complexes and unbound antibodies with B-factor annotation. INFUSSE is trained in a supervised way by minimizing the mean squared error of B-factor prediction in a *two-step* manner. In a first step, we train the sequence block S_{block} by learning the parameters of T_1 , T_2 and T_3 to predict B-factors from sequence alone, *i.e.*, from the sequence information \mathbf{x}_s and the LLM (ProtBERT) embedding \mathbf{E}_s . In a second step, we incorporate the graph block G_{block} and we optimize the parameters of the diff-GCN (t , $\mathbf{W}^{(l)}$) *jointly* with the ones of T_1, T_2, T_3 (initialized to the values obtained from the first step). Note that ProtBERT is kept frozen, and is only used to compute informative sequence embeddings \mathbf{E}_s used as input.

After training, the model was evaluated on the test set by computing the Pearson correlation coefficient R between the ground truth \mathbf{B} and predicted standardized B-factors $\hat{\mathbf{B}}$, repeating training and testing for 10 independent training/test data splits. INFUSSE achieved $R = 0.71 \pm 0.01$, averaged over the 10 training/test splits, for the prediction of standardized B-factors (see Table 1). This performance was achieved with a weighted Gaussian graph (equation 7 with $\eta = 8\text{\AA}$), but other similar setups performed similarly, confirming the robustness of the INFUSSE architecture.

Table 1 shows the results of two other models that contain only *part* of the full INFUSSE architecture. The ‘ S_{block} alone’ model is our ‘step 1’ model trained only on sequence and ProtBERT embeddings, *i.e.*, $S_{\text{block}}(\mathbf{x}_s, \mathbf{E}_s)$, and it achieved an average $R = 0.64$. The ‘ S_{block} (no LLM embeddings) + G_{block} ’ model corresponds to $S_{\text{block}}(\mathbf{x}_s, \mathbf{0}) + G_{\text{block}}(\mathbf{r}, \mathbf{X}(\mathbf{x}_s, \mathbf{0}))$, *i.e.*, one-hot encodings are used in place of ProtBERT embeddings, and it achieved an average $R = 0.55$. Altogether, these results underscore the importance of including both sequence embeddings and structural graphs within INFUSSE for the prediction of single-residue properties that are influenced by local residue neighborhoods.

In Table 1, we also compared INFUSSE with two models from the literature that perform well on B-factor prediction for generic proteins. Although the state-of-the-art (SOTA) Long Short-Term Memory (LSTM) model has been shown to achieve R between 0.6 – 0.8 on generic protein benchmarks [100, 101], it only produced an average $R = 0.48$ when applied to our antibody-antigen dataset. This reflects the difficulty of predicting B-factors for antibodies due to their high sequence variability and the presence of unstructured regions. Similarly, a baseline graph model that relies on the pseudoinverse of the graph Laplacian with no learning [102] produced $R = 0.01$ for both weighted Gaussian and ϵ -ball graphs on our antibody dataset, in contrast to R between 0.65 and 0.8 on generic proteins [103]. Again, this observation underscores the challenges of B-factor prediction in antibody-antigen complexes, and the need to combine and leverage jointly enriched sequence representations and structural graph information, as in our INFUSSE model.

Table 1. Performance in residue B-factor prediction for antibody-antigen complexes. The measure of performance reported for different methods is the Pearson correlation coefficient R between experimental and predicted B-factors of the test set, averaged over 10 training/test splits, with standard deviation across splits. Table adapted from [33].

Method	Sequence representation	Structure representation	R
INFUSSE ($S_{\text{block}} + G_{\text{block}}$)	One-hot encoding & LLM embeddings	Weighted Gaussian graph	0.71 ± 0.01
S_{block} alone (no structure)	One-hot encoding & LLM embeddings	—	0.64 ± 0.02
S_{block} (no LLM embeddings) + G_{block}	One-hot encoding	Weighted Gaussian graph	0.55 ± 0.04
LSTM [100] (SOTA for general proteins)	One-hot encoding	Raw coordinates \mathbf{r} , secondary structure, and chain breaks	0.48 ± 0.06
Laplacian pseudoinverse (no learning)	—	Weighted Gaussian graph	0.01 ± 0.04

4 Biophysical Interpretability

Interpretability in ML seeks to explain model predictions, both in connection to its internal workings and to the modeled system, *e.g.*, by relating variations in inputs and learned representations to variations in outputs in terms of intelligible, meaningful features or variables. Ref. [104] spells out the desiderata of interpretability approaches, emphasizing fidelity to the original input-output mapping, clarity, and parsimony of the explanation. In practice, a simple model such as linear regression is intrinsically explainable insofar as the model’s prediction depends in a clear and transparent way on a weighted sum of features, with each regression coefficient reflecting the importance of the corresponding feature. The prediction by a decision tree can be rationalized in terms of a set of decisions in the feature space, and a parsimonious summary of the predictive contribution of each feature can be obtained via metrics of feature importance [105].

On the other hand, deep neural networks typically trade algorithmic transparency for performance, requiring *post hoc* interpretability methods to ascribe performance to particular variables or, alternatively, by constructing architectures that are interpretable by design [106].

4.1 Model-Agnostic Interpretability Strategies: Feature Importance

Interpretability approaches can be broadly classified as either *model-agnostic* or *model-dependent*. Within the first group, different approaches are typically employed to compute concise, *post hoc* metrics of feature importance through *importance factors*, which measure how strongly an input variable (feature) influences the predictive performance of a model [107]. LIME (Local Interpretable Model-agnostic Explanations) fits a simple linear regression model around a prediction and uses the regression coefficients to quantify feature importance [108], whereas SHAP (SHapley Additive exPlanations) assigns each feature a Shapley value that reflects its expected marginal contribution, thereby providing an

information-theoretic decomposition of the model’s output into contributions from each input variable [109].

These methods have been applied in the context of immune proteins. Examples include the prediction of epitope-specific T-cell receptors with random forest models, where Gini-based importance factors rank discriminatory β -chain CDR3 positional and physicochemical features [110], and a *post hoc* pipeline based on the model-agnostic interpretability strategy of “anchors” [111] to extract human-readable, specificity-determining binding motifs in T-cell receptors [112]. Additional examples of ML interpretability strategies applied to molecular immunology are discussed in Ref. [113].

Antibody region importance for binding affinity. Model-agnostic interpretability strategies have been used within ANTIPASTI to introduce a region-level importance factor that quantifies the change in the prediction error when a given region is excluded from the final affinity estimation. The importance I for a region s^* is calculated as:

$$I(s^*) = 100 \frac{\bar{N}(s_{\text{Best}})}{\bar{N}(s^*)} \frac{\left| \text{MSE}_{s \neq s^*} - \text{MSE} \right|}{\left| \text{MSE}_{s \neq s_{\text{Best}}} - \text{MSE} \right|}, \quad (10)$$

where $\bar{N}(s^*)$ denotes the mean number of residues across the entire dataset. In this way, the importance $I(s^*)$ is given as a percentage of s_{Best} , the region that, when excluded, causes the MSE to deviate the most, *i.e.*, it contributes the most to an accurate prediction of K_D . Since targets of different types (*e.g.*, proteins or peptides) appeared to be associated to different binding configurations, we computed antibody region importance $I(s^*)$ for each target type t . Thus, the mean squared error in equation 10 was given by $\text{MSE} = \frac{1}{A_t} \sum_{i=0}^{A_t-1} (\log_{10}(K_D^{(i)}) - f_{\theta^*}(\mathbf{X}^{(i)}))^2$, where A_t is the number of antibodies in the training set that bind to target type t , and f_{θ^*} denotes the binding affinity prediction by the final ANTIPASTI architecture with learned parameters θ^* .

We computed the region importance $I(s^*)$ for the FRs and CDRs across the heavy and light chains. For example, for antibodies binding to protein targets (Figure 5A), the three CDR loops of the heavy chain rank among the top four regions, consistent with their harboring most epitope-binding residues [114, 115]. The CDR loops of the heavy chain are surpassed only by the second framework of the light chain (FR-L2). Previous experimental work has shown that mutations in FR-L2 can lead to improved thermostability, a pre-requisite for effective binding [114]. Further, residues within FR-L2 are part of the heavy-light chain interface involved in highly stabilizing interactions with the CDR-H3, thus modulating its rigidity [116, 117].

Importance of structural information for B-factor prediction. Within INFUSSE we perform a different type of ablation analysis that assesses the importance of structural information for B-factor prediction. Specifically, INFUSSE

quantifies the effect of incorporating structural information through graph representations by comparing per-residue squared errors before and after adding the graph block G_{block} :

$$\varepsilon_{S_{block},j}^{(q)} := \left(\hat{B}_{S_{block},j}^{(q)} - B_j^{(q)} \right)^2 = \left([S_{block}(\mathbf{x}_s^{(q)}, \mathbf{E}_s^{(q)})]_j - B_j^{(q)} \right)^2, \quad (11)$$

$$\varepsilon_{\text{INFUSSE},j}^{(q)} := \left(\hat{B}_j^{(q)} - B_j^{(q)} \right)^2 = \left([\text{INFUSSE}(\mathbf{x}_s^{(q)}, \mathbf{r}^{(q)})]_j - B_j^{(q)} \right)^2, \quad (12)$$

where $j = 1, \dots, N_q$ indexes the residues of the PDB entries $q \in \mathcal{Q}$ in the test set. We then collect the differences in prediction errors into the set:

$$\Delta_{\text{graph}} = \left\{ \Delta_{\text{graph},j}^{(q)} \right\} \quad \text{where} \quad \Delta_{\text{graph},j}^{(q)} := \varepsilon_{S_{block},j}^{(q)} - \varepsilon_{\text{INFUSSE},j}^{(q)}. \quad (13)$$

The distribution of Δ_{graph} across residues enables us to assess whether specific antibody regions or structural motifs are associated to statistically significant performance gains when the structural graph information is added. For instance, we analyzed different secondary structure motifs (α -helices, β -strands and loops) across antigens, and found that graph information brings an improvement for α -helices and loops, but no significant change, compared to sequence-based prediction alone, for β -strands (see Figure 5B). We found a significantly higher Δ_{graph} mean value for α -helices compared to loops (difference of means $\Delta\mu = 0.08$ with p-value 0.05), while the interquartile range (IQR) for loops is wider than that of helices ($\Delta\text{IQR} = 0.25$ with p-value 10^{-6}), suggesting more heterogeneity of graph-induced effects on performance across positions in loops. This aligns with our intuition: although graph information is beneficial to predictions in *unstructured* regions, their intrinsic flexibility and conformational variability make the mapping between sequence, structure and B-factors less consistent in the training dataset, leading to more heterogeneous performance. On the other hand, the narrow, near-zero distribution of Δ_{graph} for β -strands indicates that the sequence block is already able to capture most of their geometry and conformational flexibility, which are encoded to a large extent in their amino acid composition. β -strands are indeed particularly regular structural motifs whose stability is enabled by specific patterns of hydrophobic residues and hydrogen bonds [118].

4.2 Model-Dependent Interpretability: Using the Learned Representations

Model-dependent interpretability examines and probes the internal mechanisms of a model, such as parameter values or latent representations, resulting in explanations that are tied to the model architecture itself rather than being model-agnostic. Visualizing such internal parameters and representations, *e.g.*, by projecting onto lower dimensions (see Figure 2D), provides complementary insight into the information that is encoded and leveraged for prediction at different layers [119].

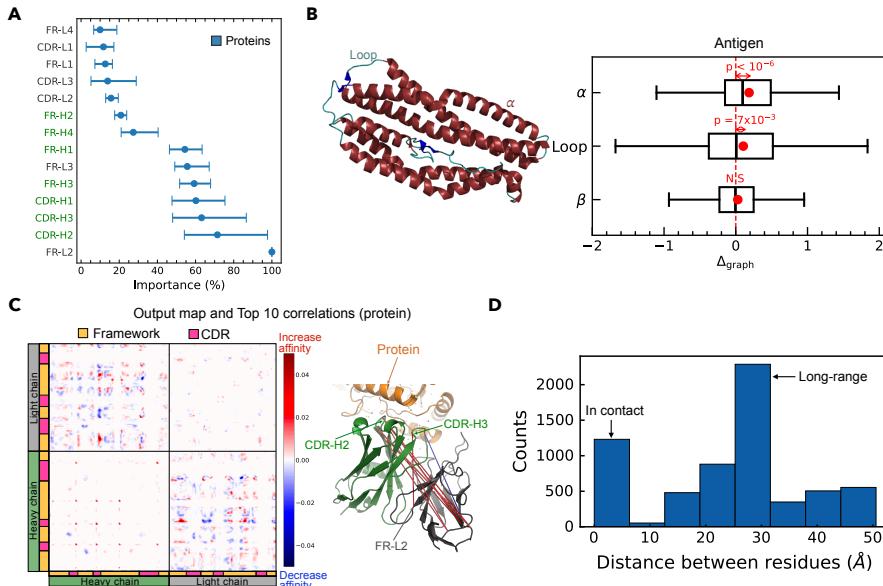


Fig. 5. Biophysical interpretation in ML models. (A) Ranking of ANTIPASTI importance factors for different antibody regions for protein targets. The importance factor (equation 10) is expressed as a percentage of that of the best region, and we show the average (dot) and extreme values (error bars) over 5 training/test splits. (B) Boxplot of INFUSSE’s Δ_{graph} stratified by secondary structure types (α -helix, β -strand or loop) for the antigens in the test set. (C) ANTIPASTI’s affinity-relevant correlations for an antibody with a protein target (Mntc with Mab 305-78-7 complex, PDB entry: 5hdq). (D) Distribution of α -carbon pairwise distances involved in the top 10 ANTIPASTI’s affinity-relevant correlations across all antibody structures. Panel B adapted from Ref. [33]; panels A, C and D adapted from Ref. [32].

Beyond visualization, gradient-based perturbation methods use the value of gradients upon small perturbations of the inputs and their propagation to probe the predictive contribution of single neural network components (like hidden unit activations) or input features [120, 121]. DeepLIFT addresses key design limitations of seminal gradient-based perturbation methods by propagating difference-from-reference contributions to quantify input feature importance [122].

Within specific architectures, attention weights are another prototypical example, as the attention mechanism is designed to highlight which positions a model focuses on during prediction, and in protein language models these attention patterns have been shown to correlate with functional correlations and structural contacts [123], including epitope-paratope ones [124]. Attention scores can also be recast through the lens of pairwise statistical mechanics models of proteins (Potts models), thereby linking learned dependencies to physical interactions between residues [125]. Finally, Transfer Learning applied learning network weights within the architecture of Restricted Boltzmann Machines can flag up distinctive amino acid patterns underlying molecular binding specificity in T-cell mediated immune responses [126].

Within our methods, model-dependent interpretability has been employed in ANTIPASTI. Since the output layer of ANTIPASTI is designed such that the predicted binding affinity is a weighted sum of M representational elements (equation 8), these can be arranged into an interpretable, 2D map \mathbf{F} given by:

$$F_{jk} := - \sum_{\ell=0}^{n_f-1} w_i z_i = F_{kj}, \quad \text{with } i = j + \sqrt{\frac{M}{n_f}} k + \ell \frac{M}{n_f}, \quad (14)$$

where $j, k = 0, \dots, \sqrt{M/n_f} - 1$ and n_f is the number of filters in the convolutional layer. Through this reshaping, each element F_{jk} can be interpreted as a representation of the NM-mediated correlation between amino acids j and k learned for the binding affinity prediction task.

Given that equation 8 can be rewritten as $\log_{10}(\hat{K}_D) = - \sum_{j,k} F_{jk}$, the positive (respectively, negative) elements of \mathbf{F} can be interpreted as the residue-residue correlations responsible for its high (respectively, low) binding affinity. We call these F_{jk} pairs “affinity-relevant correlations”. For instance, Figure 5C shows, among the ANTIPASTI’s affinity-relevant correlations that are highest in magnitude, correlations linking two *distant* regions: CDR-H3 and FR-L2. To further validate this insight that affinity-relevant correlations can span large distances in the structure, we considered the 10 correlations highest in magnitude for each structure in the dataset. We found that most of the important correlations are long-range: 80% of the top correlations are between residues more than 10Å apart. Yet the distribution also has a peak at short distances, indicating that the model detects important short-range, contact-related physical interactions as well (Figure 5D). An open question amenable to future investigations is what biophysical processes underpin the long-range affinity-relevant correlations. In our preliminary inspection of a case where a 10-fold increase in binding affinity was observed upon a mutation on the CDR-H3 [127], we found that the largest

changes in the affinity-relevant correlations concerning the mutated residue connected it to two β -sheets that underwent substantial conformational shifts, as the mutation improved the steric fit of the antigen within the antibody’s binding pocket (see Ref. [32]).

In both our models, the performance comparison to simpler approaches revealed that putative molecular determinants of the predicted properties entail structural and residue-residue interaction patterns rather than isolated amino acids. INFUSSE allows for the diffusion of information over local neighborhoods on geometry-derived graphs, capturing short- to mid-range motifs (*e.g.*, loops, helices, or interfacial patches). ANTIPASTI emphasizes global dynamical couplings, as its affinity-relevant correlations are frequently inter-region and long-range. This is consistent with the well-acknowledged view that cooperative interactions and collective motions underlie many aspects of protein function [128], more than any single residue in isolation.

5 Conclusion

In this chapter, we have examined how sequence and structure protein representations, ML architectures and interpretability strategies can be tailored to antibody modeling and its specific challenges posed, for instance, by the functional constraints of antigen-binding specificity, by the scarcity of functionally annotated data, and by antibodies’ sequence variability and conformational flexibility. We have argued that structure-derived representations, such as graphs, are particularly valuable in this context, as they can capture both local interactions and chemical neighborhoods, as well as long-range cooperative effects.

We have discussed two ML methods for antibody property prediction, ANTIPASTI and INFUSSE, both of which are underpinned by graph representations of protein structure, and which also illustrate the broader principle that ML predictive performance can be combined with interpretability for biological insight.

ANTIPASTI provides an example of how binding affinity predictions (a *global* property) can be linked to, and elucidated by, the direct inspection of learned representations as well as interpretable region-level importance factors. ANTIPASTI reveals that long-range cooperative correlations between regions, including contributions from the second framework of the light chain (FR-L2), can often be crucial to binding affinity.

Our application of INFUSSE, on the other hand, focuses on B-factor predictions (a *local*, residue-level property) and its modular architecture (one block trained on sequences, one block incorporating the graph representation) allows one to systematically assess the effect of incorporating structural information for improved accuracy of residue-level predictions. Systematic benchmarking against sequence-only and structure-only ablations confirms that coupling context-rich sequence embeddings with graphs improves accuracy, and pinpoints where graph information confers the greatest enhancement in prediction, notably within CDRs and at paratopes and epitopes. Due to its model-agnostic nature,

the importance factor strategy used for ANTIPASTI could be extended to the local predictions by INFUSSE. All such quantitative analyses of region or residue importance to determining a given property could inform targeted antibody mutagenesis and ultimately accelerate antibody discovery.

Future work will be needed to embed such predictions into broader pipelines for computational antibody design by combining them with predictors of additional properties such as solubility, stability, and humanness [129–131]. This integration would allow antibody candidates to be assessed more holistically *in silico*, with different predictors capturing orthogonal aspects of their molecular behavior. It remains however an open challenge to understand how to navigate the inevitable trade-offs between these properties in a systematic rather than an *ad hoc* manner.

For design and more general purposes, a key step will be to create interpretability frameworks that can merge outputs across multiple models, connecting global and local predictions in a coherent way. This step would enable the development of ML models that not only score antibody candidates but also provide insight into the biophysical properties underlying their expected behavior, thereby enhancing model reliability and generalizability. Another avenue for future work is the quantification of predictive uncertainty, in such a way that practitioners can gauge the extent to which they can rely on computational predictions ahead of experimental validation.

References

1. Janeway, C. A., Travers, P., Walport, M., Shlomchik, M.: The structure of a typical antibody molecule. In: Immunobiology: The Immune System in Health and Disease (5th edition), pp. 116–118. Garland Science, New York (2001)
2. Li, L., Chen, S., Miao, Z., Liu, Y., Liu, X., Xiao, Z.-X., Cao, Y.: AbRSA: A robust tool for antibody numbering. *Protein Science* 28, 1524–1531 (2019)
3. Chiu, M. L., Goulet, D. R., Teplyakov, A., Gilliland, G. L.: Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies* 8(4), 55 (2019)
4. Mora, T., Walczak, A.: Quantifying lymphocyte receptor diversity. arXiv:1604.00487 (2016)
5. Chothia, C., Lesk, A. M.: Canonical structures for the hypervariable regions of immunoglobulins. *Journal of molecular biology* 196(4), 901–917 (1987)
6. Seeliger, D., Schulz, P., Litzenburger, T., Spitz, J., Hoerer, S., Blech, M., Enenkel, B., Studts, J. M., Garidel, P., Karow, A. R.: Boosting antibody developability through rational sequence optimization. *mAbs* 7(3), 505–515 (2015)
7. Jain, T., Sun, T., Durand, S., Hall, A., Houston, N. R., Nett, J. H., Sharkey, B., Bobrowicz, B., Caffry, I., Yu, Y., Cao, Y., Lynaugh, H., Brown, M., Baruah, H., Gray, L. T., Krauland, E. M., Xu, Y., Vásquez, M., Wittrup, K. D.: Biophysical properties of the clinical-stage antibody landscape. *Proceedings of the National Academy of Sciences of the United States of America* 114(5), 944–949 (2017)
8. Sormanni, P., Aprile, F. A., Vendruscolo, M.: Third generation antibody discovery methods: *in silico* rational design. *Chemical Society reviews* 47(24), 9137–9157 (2018)

9. Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M., Church, G. M.: Unified rational protein engineering with sequence-based deep representation learning. *Nature methods* 16(12), 1315–1322 (2019)
10. Yang, K. K., Wu, Z., Arnold, F. H.: Machine-learning-guided directed evolution for protein engineering. *Nature methods* 16(8), 687–694 (2019)
11. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., Hassabis, D.: Improved protein structure prediction using potentials from deep learning. *Nature* 577(7792), 706–710 (2020)
12. Tabasinezhad, M., Talebkhan, Y., Wenzel, W., Rahimi, H., Omidinia, E., Mahboudi, F.: Trends in therapeutic antibody affinity maturation: From in-vitro towards next-generation sequencing approaches. *Immunology letters* 212, 106–113 (2019)
13. Wang, B., Gallolu Kankamalage, S., Dong, J., Liu, Y.: Optimization of therapeutic antibodies. *Antibody therapeutics* 4(1), 45–54 (2021)
14. Makowski, E. K., Kinnunen, P. C., Huang, J., Wu, L., Smith, M. D., Wang, T., Desai, A. A., Streu, C. N., Zhang, Y., Zupancic, J. M., Schardt, J. S., Linderman, J. J., Tessier, P. M.: Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nature communications* 13(1), 3788 (2022)
15. Sirin, S., Apgar, J. R., Bennett, E. M., Keating, A. E.: AB-Bind: Antibody binding mutational database for computational affinity predictions. *Protein Science* 25(2), 393–409 (2016)
16. Kang, Y., Leng, D., Guo, J., Pan, L.: Sequence-based deep learning antibody design for in silico antibody affinity maturation. *arXiv:2103.03724* (2021)
17. Sormanni, P., Aprile, F. A., Vendruscolo, M.: The CamSol method of rational design of protein mutants with enhanced solubility. *Journal of molecular biology* 427(2), 478–490 (2015)
18. Wang, H., Hao, X., He, Y., Fan, L.: AbImmPred: An immunogenicity prediction method for therapeutic antibodies using AntiBERTy-based sequence features. *PloS one* 19(2), e0296737 (2024)
19. Kurumida, Y., Saito, Y., Kameda, T.: Predicting antibody affinity changes upon mutations by combining multiple predictors. *Scientific reports* 10(1), 19533 (2020)
20. Yang, Y. X., Huang, J. Y., Wang, P., Zhu, B. T.: AREA-AFFINITY: A Web Server for Machine Learning-Based Prediction of Protein-Protein and Antibody-Protein Antigen Binding Affinities. *Journal of chemical information and modeling* 63(11), 3230–3237 (2023)
21. Jumper, J. M., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., ..., Hassabis, D.: Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873), 583–589 (2021)
22. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., ..., Jumper, J. M.: Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630(8016), 493–500 (2024)
23. Delmotte, A., Tate, E. W., Yaliraki, S. N., Barahona, M.: Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin-myosin light chain interaction. *Physical biology* 8(5), 055010 (2011)

24. Qian, Y., Expert, P., Panzarasa, P., Barahona, M.: Geometric graphs from data to aid classification tasks with Graph Convolutional Networks. *Patterns* 2(4), 100237 (2021)
25. Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., Zhong, C.: Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *arXiv:2103.11251* (2021)
26. Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R., Meiler, J.: Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *Journal of molecular modeling* 15(9), 1093–1108 (2009)
27. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., Ben-Tal, N.: ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic acids research* 33(Web Server issue), W299–W302 (2005)
28. Ovchinnikov, V., Louveau, J. E., Barton, J. P., Karplus, M., Chakraborty, A. K.: Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. *eLife* 7, e33038 (2018)
29. Teilum, K., Olsen, J. G., Kragelund, B. B.: Functional aspects of protein flexibility. *Cellular and molecular life sciences* 66(14), 2231–2247 (2009)
30. Schneider, B., Gelly, J. C., de Brevern, A. G., Černý, J.: Local dynamics of proteins and DNA evaluated from crystallographic B factors. *Acta Crystallographica Section D, Biological crystallography* 70(9), 2413–2419 (2014)
31. Carugo, O.: How large B-factors can be in protein crystal structures. *BMC bioinformatics* 19(1), 61 (2018)
32. Michalewicz, K., Barahona, M., Bravi, B.: ANTIPASTI: Interpretable prediction of antibody binding affinity exploiting normal modes and deep learning. *Structure* 32(12), 2422–2434 (2024)
33. Michalewicz, K., Barahona, M., Bravi, B.: Integrating protein sequence embeddings with structure via graph-based deep learning for the prediction of single-residue properties. *arXiv:2502.17294* (2025)
34. Amor, B. R., Schaub, M. T., Yaliraki, S. N., Barahona, M.: Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature communications* 7, 12477 (2016)
35. Kmiecik, S., Gront, D., Kolinski, M., Witeska, L., Dawid, A. E., Kolinski, A.: Coarse-Grained Protein Models and Their Applications. *Chemical reviews* 116(14), 7898–7936 (2016)
36. Liwo, A., Khalili, M., Scheraga, H. A.: Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America* 102(7), 2362–2367 (2005)
37. Kmiecik, S., Gront, D., Kouza, M., Kolinski, A.: From coarse-grained to atomic-level characterization of protein dynamics: transition state for the folding of B domain of protein A. *The journal of physical chemistry. B* 116(23), 7026–7032 (2012)
38. Maisuradze, G. G., Senet, P., Czaplewski, C., Liwo, A., Scheraga, H. A.: Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *The journal of physical chemistry. A* 114(13), 4471–4485 (2010)
39. Nasica-Labouze, J., Nguyen, P. H., Sterpone, F., Berthoumieu, O., Buchete, N. V., Coté, S., De Simone, A., Doig, A. J., Faller, P., Garcia, A., Laio, A., Li, M. S., Melchionna, S., Mousseau, N., Mu, Y., Paravastu, A., Pasquali, S., Rosenman, D. J., Strodel, B., Tarus, B., ... Derreumaux, P.: Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chemical reviews* 115(9), 3518–3563 (2015)

40. Fleishman, S. J., Corn, J. E., Strauch, E. M., Whitehead, T. A., Andre, I., Thompson, J., Havranek, J. J., Das, R., Bradley, P., Baker, D.: Rosetta in CAPRI rounds 13–19. *Proteins* 78(15), 3212–3218 (2010)
41. Frezza, E., Lavery, R.: Internal Normal Mode Analysis (iNMA) Applied to Protein Conformational Flexibility. *Journal of chemical theory and computation* 11(11), 5503–5512 (2015)
42. Tirion, M. M.: Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters* 77(9), 1905–1908 (1996)
43. Bahar, I., Atilgan, A. R., Erman, B.: Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding & design* 2(3), 173–181 (1997)
44. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., Bahar, I.: Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical journal* 80(1), 505–515 (2001)
45. Dykeman, E. C., Sankey, O. T.: Normal mode analysis and applications in biological physics. *Journal of Physics: Condensed Matter* 22(42), 423202 (2010)
46. Levitt, M., Sander, C., Stern, P. S.: Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of molecular biology* 181(3), 423–447 (1985)
47. Brooks, B., Karplus, M.: Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* 82(15), 4995–4999 (1985)
48. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., Kern, D.: A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450(7171), 913–916 (2007)
49. Go, N., Noguti, T., Nishikawa, T.: Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences of the United States of America* 80(12), 3696–3700 (1983)
50. Haliloglu, T., Bahar, I., Erman, B.: Gaussian Dynamics of Folded Proteins. *Physical Review Letters* 79(16), 3090–3093 (1997)
51. Hinsen, K.: Analysis of domain motions by approximate normal mode calculations. *Proteins* 33(3), 417–429 (1998)
52. Bauer, J. A., Pavlović, J., Bauerová-Hlinková, V.: Normal mode analysis as a routine part of a structural investigation. *Molecules* 24(18), 3293 (2019)
53. Tama, F., Sanejouand, Y.-H.: Conformational change of proteins arising from normal mode calculations. *Protein Engineering* 14(1), 1–6 (2001)
54. Dubanevics, I., McLeish, T. C.: Optimising elastic network models for protein dynamics and allostery: Spatial and modal cut-offs and backbone stiffness. *Journal of Molecular Biology* 434(17), 167696 (2022)
55. Hinsen, K.: Normal mode theory and harmonic potential approximations. In: *Normal Mode Analysis. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, pp. 1–16. Chapman & Hall / CRC (2005)
56. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A.: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* 117, 5179–5197 (1995)
57. Ichijo, T., Karplus, M.: Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 11(3), 205–217 (1991)
58. Stehlík, P., Vaněk, P.: Sign-changing diagonal perturbations of Laplacian matrices of graphs. *Linear Algebra and its Applications* 531, 64–82 (2017)

59. Chung, F. R. K.: Spectral Graph Theory. American Mathematical Society, Providence (1997)
60. Lambiotte, R., Delvenne, J.-C., Barahona, M.: Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. *IEEE Transactions on Network Science and Engineering* 1(2), 76–90 (2014)
61. Liu, Z., Barahona, M.: Graph-based data clustering via multiscale community detection. *Applied Network Science* 5, 3 (2020)
62. Baldi, P., Pollastri, G.: The Principled Design of Large-Scale Recursive Neural Network Architectures—DAG-RNNs and the Protein Structure Prediction Problem. *Journal of Machine Learning Research* 4, 575–602 (2003)
63. Opron, K., Xia, K., Burton, Z., Wei, G.-W.: Flexibility–rigidity index for protein–nucleic acid flexibility and fluctuation analysis. *Journal of Computational Chemistry* 37(14), 1283–1295 (2016)
64. Song, F., Barahona, M., Yaliraki, S. N.: BagPype: A Python package for the construction of atomistic, energy-weighted graphs from biomolecular structures. Preprint <https://doi.org/10.6084/m9.figshare.14039723.v1> (2021)
65. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E.: The Protein Data Bank. *Nucleic acids research* 28(1), 235–242 (2000)
66. Lau, A. M., Bordin, N., Kandathil, S. M., Sillitoe, I., Waman, V. P., Wells, J., Orengo, C. A., Jones, D. T.: Exploring structural diversity across the protein universe with The Encyclopedia of Domains. *Science* 386(6721), eadq4946 (2024)
67. Kandathil, S. M., Lau, A. M., Buchan, D. W. A., Jones, D. T.: Foldclass and Merizo-search: scalable structural similarity search for single- and multi-domain proteins using geometric learning. *Bioinformatics* 41(5), btaf277 (2025)
68. Greener, J. G., Jamali, K.: Fast protein structure searching using structure graph embeddings. *Bioinformatics Advances* 5(1), vbaf042 (2024)
69. Blaabjerg, L. M., Jonsson, N., Boomsma, W., Stein, A., Lindorff-Larsen, K.: SSEmb: A joint embedding of protein sequence and structure enables robust variant effect predictions. *Nature communications* 15(1), 9646 (2024)
70. Danner, M., Begemann, M., Elbracht, M., Kurth, I., Krause, J.: Utilizing protein structure graph embeddings to predict the pathogenicity of missense variants. *NAR genomics and bioinformatics* 7(3), lqaf097 (2025)
71. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., ..., Baker, D.: Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557), 871–876 (2021)
72. Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., Lauko, A., ..., Baker, D.: De novo design of protein structure and function with RFdiffusion. *Nature* 620(7976), 1089–1100 (2023)
73. Jing, B., Eismann, S., Soni, P. N., Dror, R. O.: Equivariant Graph Neural Networks for 3D Macromolecular Structure. arXiv:2106.03843 (2021)
74. ElAbd, H., Bromberg, Y., Hoarfrost, A., Lenz, T., Franke, A., Wendorff, M.: Amino acid encoding for deep learning applications. *BMC Bioinformatics* 21, 235 (2020)
75. Ofer, D., Linial, M.: ProFET: Feature engineering captures high-level protein functions. *Bioinformatics* 31(21), 3429–3436 (2015)

76. Harding-Larsen, D., Funk, J., Madsen, N. G., Gharabli, H., Acevedo-Rocha, C. G., Mazurenko, S., Welner, D. H.: Protein representations: Encoding biological information for machine learning in biocatalysis. *Biotechnology advances* 77, 108459 (2024)
77. Li, D., Pucci, F., Rooman, M.: Prediction of Paratope–Epitope Pairs Using Convolutional Neural Networks. *International Journal of Molecular Sciences* 25(10), 5434 (2024)
78. Rackovsky S.: Global characteristics of protein sequences and their implications. *Proceedings of the National Academy of Sciences of the United States of America* 107(19), 8623–8626 (2010)
79. Breimann, S., Kamp, F., Steiner, H., Frishman, D.: AAontology: An Ontology of Amino Acid Scales for Interpretable Machine Learning. *Journal of molecular biology* 436(19), 168717 (2024)
80. Yang, K. K., Wu, Z., Bedbrook, C. N., Arnold, F. H.: Learned protein embeddings for machine learning. *Bioinformatics* 34(15), 2642–2648 (2018)
81. Weininger, D.: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28(1), 31–36 (1988)
82. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A. Y.: Self-taught learning: transfer learning from unlabeled data. In: *Proceedings of the 24th International Conference on Machine Learning*, pp. 759–766. Association for Computing Machinery, New York (2007)
83. Brandes, N., Ofer, D., Peleg, Y., Rappoport, N., Linial, M.: ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics* 38(8), 2102–2110 (2022)
84. Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., Bhowmik, D., Rost, B.: ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE transactions on pattern analysis and machine intelligence* 44(10), 7112–7127 (2022)
85. Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., Fergus, R.: Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences of the United States of America* 118(15), e2016239118 (2021)
86. Madani, A., Krause, B., Greene, E. R., Subramanian, S., Mohr, B. P., Holton, J. M., Olmos, J. L., Jr, Xiong, C., Sun, Z. Z., Socher, R., Fraser, J. S., Naik, N.: Large language models generate functional protein sequences across diverse families. *Nature biotechnology* 41(8), 1099–1106 (2023)
87. Halle, B.: Flexibility and packing in proteins. *Proceedings of the National Academy of Sciences of the United States of America* 99(3), 1274–1279 (2002)
88. Singh, R., Saha, M.: Identifying structural motifs in proteins. In: *Pacific Symposium on Biocomputing*, pp. 228–239. Accelrys Inc, San Diego (2003)
89. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big data* 8(1), 53 (2021)
90. Soleymani, S., Dabouei, A., Kazemi, H., Dawson, J. M., Nasrabadi, N. M.: Multi-Level Feature Abstraction from Convolutional Neural Networks for Multimodal Biometric Identification. In: *24th International Conference on Pattern Recognition (ICPR)*, pp. 3469–3476. IEEE Press, Beijing (2018)

91. Kourentzi, K., Srinivasan, M., Smith-Gill, S. J., Willson, R. C.: Conformational flexibility and kinetic complexity in antibody-antigen interactions. *Journal of Molecular Recognition* 21(2), 114–121 (2008)
92. Sela-Culang, I., Alon, S., Ofran, Y.: A systematic comparison of free and bound antibodies reveals binding-related conformational changes. *Journal of immunology* 189(10), 4890–4899 (2012)
93. Barozet, A., Bianciotto, M., Siméon, T., Minoux, H., Cortés, J.: Conformational changes in antibody Fab fragments upon binding and their consequences on the performance of docking algorithms. *Immunology letters* 200, 5–15 (2018)
94. Lee, M. S., Olson, M. A.: Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophysical journal* 90(3), 864–877 (2006)
95. Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., Deane, C. M.: SAbDab: the structural antibody database. *Nucleic acids research* 42(Database issue), D1140–D1146 (2014)
96. Schneider, C., Raybould, M. I. J., Deane, C. M.: SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. *Nucleic acids research* 50(D1), D1368–D1372 (2022)
97. Peach, R. L., Arnaudon, A., Barahona, M.: Semi-supervised classification on graphs using explicit diffusion dynamics. *Foundations of Data Science* 2(1), 19–33 (2020)
98. Kipf, T. N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. In: 5th International Conference on Learning Representations (2017)
99. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., Dahl, G.E.: Neural Message Passing for Quantum Chemistry. In: Proceedings of the 34th International Conference on Machine Learning, pp. 1263–1272 (2017)
100. Pandey, A., Liu, E., Graham, J., Chen, W., Keten, S.: B-factor prediction in proteins using a sequence-based deep learning model. *Patterns* 4(9), 100805 (2023)
101. Bramer, D., Wei, G. W.: Blind prediction of protein B-factor and flexibility. *The Journal of chemical physics* 149(13), 134107 (2018)
102. Bahar, I., Lezon, T. R., Yang, L. W., Eyal, E.: Global dynamics of proteins: bridging between structure and function. *Annual review of biophysics* 39, 23–42 (2010)
103. Rader, A., Chennubhotla, C., Yang, L.-W., Bahar, I.: The Gaussian Network Model: Theory and Applications. In: Normal Mode Analysis. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems, pp. 41–64. Chapman & Hall / CRC (2005)
104. Ras, G., van Gerven, M., Haselager, P.: Explanation Methods in Deep Learning: Users, Values, Concerns and Challenges. In: Escalante, H., et al. Explainable and Interpretable Models in Computer Vision and Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham (2018)
105. Breiman, L.: Random Forests. *Machine Learning* 45, 5–32 (2001)
106. Abdullah, T. A. A., Zahid, M. S. M., Ali, W.: A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions. *Symmetry* 13(12), 2439 (2021)
107. Schossler, R. T., Ojo, S., Jiang, Z., Hu, J., Yu, X.: A novel interpretable machine learning model approach for the prediction of TiO₂ photocatalytic degradation of air contaminants. *Scientific reports* 14, 13070 (2024)
108. Ribeiro, M. T., Singh, S., Guestrin, C.: “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144. Association for Computing Machinery, New York (2016)

109. Lundberg, S. M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777. Curran Associates Inc., Red Hook, New York (2017)
110. De Neuter, N., Bittremieux, W., Beirnaert, C., Cuypers, B., Mrzic, A., Moris, P., Suls, A., Van Tendeloo, V., Ogunjimi, B., Laukens, K., Meysman, P.: On the feasibility of mining CD8+ T cell receptor patterns underlying immunogenic peptide recognition. *Immunogenetics* 70(3), 159–168 (2018)
111. Ribeiro, M. T., Singh, S., Guestrin, C.: Anchors: High-Precision Model-Agnostic Explanations. In: AAAI Conference on Artificial Intelligence (2018)
112. Papadopoulou, I., Nguyen, A. P., Weber, A., Martínez, M. R.: DECODE: a computational pipeline to discover T cell receptor binding rules. *Bioinformatics* 38(Suppl 1), i246–i254 (2022)
113. Bravi B.: Development and use of machine learning algorithms in vaccine target selection. *NPJ vaccines* 9(1), 15 (2024)
114. Koenig, P., Lee, C. V., Walters, B. T., Janakiraman, V., Stinson, J., Patapoff, T. W., Fuh, G.: Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proceedings of the National Academy of Sciences of the United States of America* 114(4), E486–E495 (2017)
115. Akbar, R., Robert, P. A., Weber, C. R., Widrich, M., Frank, R., Pavlović, M., Scheffer, L., Chernigovskaya, M., Snapkov, I., Slabodkin, A., Mehta, B. B., Miho, E., Lund-Johansen, F., Andersen, J. T., Hochreiter, S., Hobæk Haff, I., Klambauer, G., Sandve, G. K., Greiff, V.: In silico proof of principle of machine learning-based antibody design at unconstrained scale. *mAbs* 14(1), 2031482 (2022)
116. Xu, H., Schmidt, A. G., O'Donnell, T., Therkelsen, M. D., Kepler, T. B., Moody, M. A., Haynes, B. F., Liao, H. X., Harrison, S. C., Shaw, D. E.: Key mutations stabilize antigen-binding conformation during affinity maturation of a broadly neutralizing influenza antibody lineage. *Proteins* 83(4), 771–780 (2015)
117. Phillips, A. M., Maurer, D. P., Brooks, C., Dupic, T., Schmidt, A. G., Desai, M. M.: Hierarchical sequence-affinity landscapes shape the evolution of breadth in an anti-influenza receptor binding site antibody. *eLife* 12, e83628 (2023)
118. Wouters, M. A., Curni, P. M.: An analysis of side chain interactions and pair correlations within antiparallel beta-sheets: the differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 22(2), 119–131 (1995)
119. Lipton, Z. C.: The Mythos of Model Interpretability. *Communications of the ACM* 61, 36–43 (2016)
120. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In: Workshop at International Conference on Learning Representations (2014)
121. Alipanahi, B., Delong, A., Weirauch, M. T., Frey, B. J.: Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature biotechnology* 33(8), 831–838 (2015)
122. Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences. In: Proceedings of the 34th International Conference on Machine Learning, pp. 3145–3153 (2017)
123. Vig, J., Madani, A., Varshney, L.R., Xiong, C., Socher, R., Rajani, N.: BERTology Meets Biology: Interpreting Attention in Protein Language Models. In: International Conference on Learning Representations (2021)

124. Leem, J., Mitchell, L. S., Farmery, J. H. R., Barton, J., Galson, J. D.: Deciphering the language of antibodies using self-supervised learning. *Patterns* 3(7), 100513 (2022)
125. Rende, R., Gerace, F., Laio, A., Goldt, S.: Mapping of attention mechanisms to a generalized Potts model. *Physical Review Research* 6, 023057 (2024)
126. Bravi, B., Di Gioacchino, A., Fernandez-de-Cossio-Diaz, J., Walczak, A. M., Mora, T., Cocco, S., Monasson, R.: A transfer-learning approach to predict antigen immunogenicity and T-cell receptor specificity. *eLife* 12, e85126 (2023)
127. Schiele, F., van Ryn, J., Litzenburger, T., Ritter, M., Seeliger, D., Nar, H.: Structure-guided residence time optimization of a dabigatran reversal agent. *mAbs* 7(5), 871–880 (2015)
128. Eyal, E., Dutta, A., Bahar, I.: Cooperative dynamics of proteins unraveled by network models. *Wiley interdisciplinary reviews. Computational molecular science* 1(3), 426–439 (2011)
129. Rosace, A., Bennett, A., Oeller, M., Mortensen, M. M., Sakhnini, L., Lorenzen, N., Poulsen, C., Sormanni, P.: Automated optimisation of solubility and conformational stability of antibodies and proteins. *Nature communications* 14(1), 1937 (2023)
130. Ramon, A., Ali, M., Atkinson, M., Saturnino, A., Didi, K., Visentin, C., Ricagno, S., Xu, X., Greenig, M., Sormanni, P.: Assessing antibody and nanobody nativeness for hit selection and humanization with AbNatiV. *Nature Machine Intelligence* 6, 74–91 (2024)
131. Ali, M., Greenig, M., Oeller, M., Atkinson, M., Xu, X., Sormanni, P.: Automated optimization of the solubility of a hyper-stable α -amylase. *Open biology* 14(5), 240014 (2024)