InDeep: 3D fully convolutional neural networks to assist in silico drug design on protein–protein interactions

Vincent Mallet, Luis Checa Ruano, Alexandra Moine Franel, Michael Nilges, Karen Druart, Guillaume Bouvier, Olivier Sperandio

Bioinformatics, 38(5), 2022, 1261–1268
https://doi.org/10.1093/bioinformatics/btab849
Advance Access Publication Date: 15 December 2021
Original Paper

OXFORD

Structural bioinformatics

InDeep: 3D fully convolutional neural networks to assist *in silico* drug design on protein–protein interactions

Vincent Mallet[1,2], Luis Checa Ruano[1,3], Alexandra Moine Franel[1,3], Michael Nilges[1], Karen Druart[1], Guillaume Bouvier [1,†] and Olivier Sperandio [1,*,†]

[1]Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, Institut Pasteur, Université de Paris, CNRS UMR3528, Paris F-75015, France, [2]Center for Computational Biology, Mines ParisTech, Paris-Sciences-et-Lettres Research University, Paris 75272, France and [3]Collège Doctoral, Sorbonne Université, Paris F-75005, France

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

Associate Editor: Pier Luigi Martelli

Received on October 1, 2021; revised on November 15, 2021; editorial decision on December 7, 2021; accepted on December 13, 2021

Abstract

Motivation: Protein–protein interactions (PPIs) are key elements in numerous biological pathways and the subject of a growing number of drug discovery projects including against infectious diseases. Designing drugs on PPI targets remains a difficult task and requires extensive efforts to qualify a given interaction as an eligible target. To this end, besides the evident need to determine the role of PPIs in disease-associated pathways and their experimental characterization as therapeutics targets, prediction of their capacity to be bound by other protein partners or modulated by future drugs is of primary importance.

Results: We present InDeep, a tool for predicting functional binding sites within proteins that could either host protein epitopes or future drugs. Leveraging deep learning on a curated dataset of PPIs, this tool can proceed to enhanced functional binding site predictions either on experimental structures or along molecular dynamics trajectories. The benchmark of InDeep demonstrates that our tool outperforms state-of-the-art ligandable binding sites predictors when assessing PPI targets but also conventional targets. This offers new opportunities to assist drug design projects on PPIs by identifying pertinent binding pockets at or in the vicinity of PPI interfaces.

Availability and implementation: The tool is available on GitLab at https://gitlab.pasteur.fr/InDeep/InDeep.

Contact: olivier.sperandio@pasteur.fr

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

1 Introduction

1.1 Protein–protein interactions as therapeutic targets

Protein–protein interactions (PPIs) are central elements in numerous biological pathways. They represent increasing interests as therapeutic targets, with a growing number of published studies describing the successful modulation of PPIs using small molecules (Torchet *et al.*, 2021). Yet, identifying chemical probes or drugs on PPIs remains a difficult task. As opposed to more conventional drug discovery targets, such as G-protein coupled receptors (GPCRs) or enzymes and more recently protein kinases, PPIs have not evolved to bind small molecules. Therefore, the proof of their ligandability has to be made on a case by case scenario (Sperandio *et al.*, 2010). Indeed, the design of small molecules binding orthosterically at the interface to prevent protein interactions is not achievable for all PPIs (Lu *et al.*, 2020). Given their number and heterogeneity of structures, it is therefore of primary importance to have powerful tools to efficiently evaluate the feasibility of considering PPIs as targets in complement of the unavoidable biological evaluations.

In the situation of designing orthosteric inhibitors of PPIs (iPPIs) using small molecules, strategies like epitope mimetics can be envisaged (Ashkenazi *et al.*, 2017). This was successfully made against the B-cell lymphoma-2 (Bcl-2) family to combat chronic lymphocytic leukaemia (D'Aguanno and Del Bufalo, 2020). This led to the development of Venetoclax which was approved by the FDA in 2016 as the first orthosteric PPI drug. The design of iPPIs implies to evaluate two complementary features within the interface: (i) the knowledge of an epitope binding at the interface and the presence of hotspot residues that carry out most of the binding energy of interaction (Clackson and Wells, 1995) and (ii) the existence of a ligandable binding site around these hotspots that could host a small molecule.

1.2 Predicting and profiling epitope binding sites

Several *in silico* tools can predict hot spot residues within PPIs (Krüger and Gohlke, 2010; Tuncbag *et al.*, 2010) but they necessitate the structure of a complex and the fore knowledge of an identified protein partner. To predict protein interactions, some tools

directly use the sequence information (Murakami and Mizuguchi, 2010) or evolutionary data (Cong *et al.*, 2019). However, leveraging the structure of the protein has shown to drastically increase performance of prediction of interface regions. Moreover, structural motifs and local arrangements of atoms can be highly conserved even across different secondary structures and different global protein folding. These local motifs are hypothesized to be the key element of partner binding. This has motivated using convolutional strategies to encode such local information about binding sites. Some of these methods focus on predicting the interaction patch on the protein: that is determine which residues are involved in an interaction (Gainza *et al.*, 2020; Dai and Bailey-Kellogg, 2021). Some methods also take as input the partner to predict the interaction patch (Dai and Bailey-Kellogg, 2021), deemed as partner-specific predictions. In that case, the prediction can be more fine-grained and also provide contact prediction: which residue interacts with which other (Sanchez-Garcia *et al.*, 2019; Townshend *et al.*, 2019). All of these tools annotate the sequence by predicting residue-level information. However, knowing not only which residues are involved in the binding (sequence or surface derived info) but which types of partner residues and where they bind in the vicinity of the protein surface is highly desirable to understand the mechanisms of epitope binding or the design of future drugs mimicking these epitopes.

1.3 Predicting ligandable binding sites within PPIs

A plethora of tools is now available to predict binding sites and binding site ligandability. One can cite historical and efficient geometric-based methods such as Fpocket (Guilloux *et al.*, 2009), VolSite (Da Silva *et al.*, 2018) and mkgridXf (Monet *et al.*, 2019), fragment-based methods like FTMap (Kozakov *et al.*, 2015) and more recent and powerful methods using deep learning such as DeepSite (Jiménez *et al.*, 2017), P2rank (Krivák and Hoksza, 2018), Kalasanty (Stepniewska-Dziubinska *et al.*, 2020), DeepSurf (Mylonas *et al.*, 2021) or OctSurf (Liu *et al.*, 2021), with Kalasanty being a reference in the field. These methods have demonstrated their predictive capacity to identify ligandable binding sites on conventional drug targets. Although FTMap was the first method to rise the question of PPI ligandability without really providing a ligandability score, none of the methods cited above is specific to the ligandability of PPIs. Nonetheless, interfaces of PPIs are historically described as rather flat, large and devoid from deep binding pockets (Arkin *et al.*, 2014). It is therefore most legitimate to anticipate a specific form of ligandability in the case of PPIs. There are numerous examples of co-crystallized orthosteric iPPIs (Torchet *et al.*, 2021). Using machine learning and PPI-specific datasets, we can expect to address the specificity of PPIs ligandability.

1.4 Capturing holo-likeness along MD trajectories

PPIs are known to undertake important conformational changes depending on their binding state: apo, holo with ligand or a protein partner. These conformational changes affect the shape and binding capacity of interface binding pockets (Johnson and Karanicolas, 2013). It is therefore of primary importance to take these conformational changes into account when profiling epitope and ligandable binding sites (Kozakov *et al.*, 2015) as those will condition binding to partners. This represents a major challenge when attempting to identify chemical probes, using *in silico* methods such as virtual screening or designing epitope mimetics, in the absence of the partner bound. Indeed, it is for example key for such methods to sample and identify so-called holo-like conformations prior to virtual screening in the context of ensemble docking (Ivetac and McCammon, 2012; Amaro *et al.*, 2018). Previous works have addressed the holo-like sampling challenge using MD simulations with methanol solvent. The use of a less polar solvent than water favors the opening of transient hydrophobic pockets, resulting in an improvement of docking results (Eyrisch *et al.*, 2012). Other methods have already been developed to monitor ligandability along molecular dynamics (MD) trajectories using geometric (Guilloux *et al.*, 2009) or deep learning approaches (Kozlovskii and Popov, 2020), although none of these are specific to PPIs. Moreover, no method is available to monitor interactability patches and epitope binding sites along MD trajectories.

1.5 Contribution

Our work builds upon our last release of iPPI-DB (Torchet *et al.*, 2021) and of its new target-centric mode, and aims to facilitate the identification of iPPIs. Our tool InDeep has capitalized on iPPI-DB structural data to train predictive models relying on neural networks with a 3D fully convolutional U-Net architecture. It is a unified multi-tasking prediction tool that uses the 3D structure of proteins to predict ligandable binding for iPPIs and so-called interactability patches for epitope binding. We show that InDeep outperforms the state of the art of binding pockets detection methods and that our tool is especially efficient to detect iPPI and epitope binding sites. While remaining competitive on annotating the protein sequence with interactability, it also predicts the spatial location of its putative partner. Our tool also enables tracking of these druggability and interactability scores for a given detected pocket along MD trajectories. It is integrated in a PyMol (DeLano *et al.*, 2002) plugin (see Supplementary Section S7) for easy visualization of the predictions, making it a real toolbox for iPPI drug design. It is freely available at https://gitlab.pasteur.fr/InDeep/InDeep. Finally, the results of InDeep predictions (before post-processing) can be consulted on the iPPI-DB website for every heterodimer and iPPI-bound protein in the database at https://ippidb.pasteur.fr/targetcentric/.

2 Materials and methods

2.1 Data curation, splitting and representation

For training and assessing the models, we have used the dataset available in the iPPI-DB (Torchet *et al.*, 2021). The dataset relies on two subsets: one contains hetero–dimeric complexes (HD interactions) and the other iPPI-bound protein complexes (PL interactions) where ligands bind one of the two partners at the interface within a HD complex (orthosteric inhibitors). We have then split this dataset based on CATH (Sillitoe *et al.*, 2021) folds to avoid any structural overlap between our train, validation and test datasets.

Equipped with these sets of co-crystallized proteins with either a protein or ligand partner, we wish to represent them in a way a neural network can learn on. We follow the volumetric CNN framework used, for instance, in Jiménez *et al.* (2017) and Stepniewska-Dziubinska *et al.* (2020). This framework consists in treating the interaction sites as 3D images, whose color channel is functional atom types. We introduce five functional atom types: α-carbon (Cα), donor and receptor of h-bonds and positively and negatively charged atoms and hydrophobic/aromatic atoms. We then put a Gaussian function around each atom centre based on its type and interpolate its values on a regular 3D grid with 1A spacing. The details and results of these procedures are described in Supplementary Section S1.

2.2 Model architecture and learning

We now want to build a model that takes a protein structure as input and predicts iPPI ligandability (PL interaction) as well as protein interactability (HD interaction). These two machine learning tasks can benefit from the concept of multitasking, a well-documented phenomenon that means that one hybrid machine learning model that solves two tasks usually performs better than two separate ones (Goodfellow *et al.*, 2016). Indeed, in the multitasking setting, each task benefits from the representations learned using the other task's supervision. After several shared layers, our network is split into a PL and an HD branch. These branches are two independent sequences of layers with a sigmoid and softmax activations for PL and HD, respectively. We use a U-Net (Ronneberger *et al.*, 2015) architecture for our prediction. Our model does not use fixed-size linear layers (fully convolutionnal network), which enables it to take any grid size as input. A visual representation of this branching scheme is

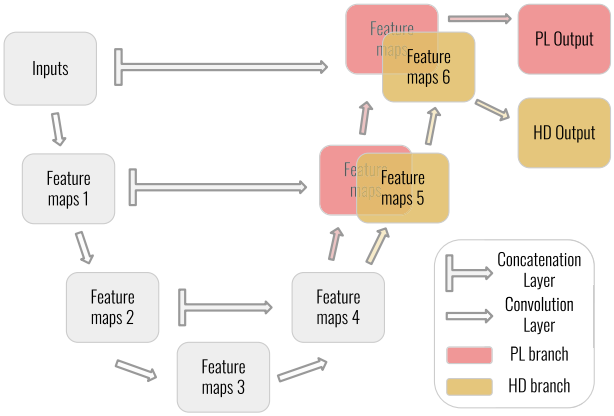

Fig. 1. Visual representation of InDeep's architecture

available in Figure 1 and a detailed description of the model is available in Supplementary Section S2.

We then train the resulting models with batches containing a mix of PL and HD data points. For memory limitations issues, we used the accumulated gradient trick to use batches of size greater than 1. Since we encode the 3D structure into voxels, the resulting representation is sparse. Moreover, the key part of the prediction lies in the space surrounding the protein surface and especially around the position of the true ligand, while having a hard zero inside the protein of very far from its surface is less relevant. To account for these two points, we use the weighted versions of cross entropy (CE) and binary CE as the loss ℓ to train our network. All voxels receive a small weight $\mathbf{w}^{\text{background}}$ of 0.05, then the voxels closer than 6 Å from the surface receive an additional weight $\mathbf{w}^{\text{surface}}$ of 0.35 and finally the voxels corresponding to the target voxels get an additional value $\mathbf{w}^{\text{partner}}$ of 1. We found this weighting scheme to stabilize the learning and optimized our network with an Adam (Kingma and Ba, 2014) optimizer

$$\ell(\hat{y}, y) = \sum_{(i,j,k)\in \text{Grid}} (\mathbf{w}^{\text{background}}_{i,j,k} + \mathbf{w}^{\text{surface}}_{i,j,k} + \mathbf{w}^{\text{partner}}_{i,j,k}) \text{CE}_{i,j,k}(\hat{y}, y).$$

2.3 Post-processing and optimization

Once equipped with scalar fields prediction, we need to segment them in contiguous regions of high values. Other approaches have simply used a mean-shift algorithm (Jiménez *et al.*, 2017; Stepniewska-Dziubinska *et al.*, 2020), but such algorithms can split the prediction or discard important neighbouring parts of the prediction. To address this segmenting problem, we have relied on the watershed (Beucher, 1979) algorithm. The watershed algorithm finds all basins around local minima. We build a graph whose nodes are these basins and whose edges contain the Euclidean distance as well as a normalized value of the lowest saddle point joining two neighbouring basins. Then, we merge the neighbouring nodes in a greedy manner, by prioritizing the ones with the smallest edge. We stop the merging process when the merged nodes exceed a geometric distance threshold of 15 and 20 Å for PL and HD, respectively. Each of the resulting group of basins is denoted as a predicted pocket and scored based on the mean values of its best 150 voxels. Finally, we filter these predicted pockets to remove the smallest or least high-scoring ones, yielding the final list of predictions.

To choose the optimal values for the hyperparameters of the network and the post-processing (number of layers, number of neurons per layer, thresholds), we conducted a hyperparameter optimization (HPO). The HPO metric and optimization procedure is described in detail in the Supplementary Section S3. We then ran the HPO for approximately 100 experiments on the validation set (Supplementary Fig. S2), which gave us our final predictive model.

2.4 MD trajectory analysis

To distinguish the conformations with higher ligandability (PL) or interactability (HD) propensity, the developed model can be used on an ensemble of protein conformations generated by MD. To do so, each snapshot of a MD simulation is treated as an input for InDeep. Therefore, one can specify some residues to InDeep along which to monitor the ligandability or interactability, resulting in a reduced grid compared with an inference on the whole protein. Then, the post-processing is simplified because the prediction size is reduced: a spatial anchor is chosen either by the user or as the point of the grid in the solvent closest to the grid centre. Finally, we simply grow a volume around this anchor following a greedy nearest neighbour policy. The average value of the voxels in this volume represents a ligandability or interactability score that can be easily tracked along the MD time steps. We chose a volume of 150 $Å^3$, close to the cutoff of 100 $Å^3$ used in the study by Gao and Skolnick (2013) which considers that 80% of pockets occupied by ligands are encompassed by this cutoff.

3 Results

3.1 Prediction of ligandable binding sites

3.1.1 Benchmark of InDeep on conventional target binding sites

Rosell and Fernández-Recio (2020) have introduced a method to detect iPPI binding sites that use FPocket (Guilloux *et al.*, 2009) along a MD trajectory to detect transient binding sites. These binding sites are then selected to be nearby potential interfaces based on protein–protein docking results. However, the docking step requires the structure of the partner, while our method just uses the structure of the protein meant to be bound. Moreover, these steps require substantial computing time, making it less suitable for the investigation of several structures. There are several tools that aim to predict small molecule binding sites, among which Kalasanty is a reference in the field. In the absence of other iPPI dedicated tool for ligandability, we benchmarked InDeep against Kalasanty. The other major difference between our tool and Kalasanty is that Kalasanty was trained on VolSite predicted cavities in ligand locations, whereas our model has been trained on ligand position directly. For fairness, we compare the ability of our tool to predict VolSite predicted cavities as well as the ligand location with Kalasanty.

We use the same dataset as the authors of Kalasanty did for validation: a distinct dataset, made by Chen *et al.* (2011). The original test set is composed of 111 protein–ligand holo structures and 104 corresponding apo structures. We have filtered out a few systems that were too similar to our training set according to the TM-score metric (Zhang and Skolnick, 2004) and end up with 187 and 196 pockets evaluation for apo and holo structures, respectively. Ligands coordinates were extracted from the holo structures of the Chen benchmark and VolSite was used to describe cavities for each ligand, as shown in Supplementary Figure S3. Following Kalasanty, the number of retained predicted pockets is the number of small molecules present in the deposited PDB system.

We then used the metrics used by Kalasanty on their dataset: DCC and DVO. The DCC metric computes the distance in Angstrom between the centre of mass of the predicted pocket and the one of the ground truth. We denote a prediction as successful when its DCC is below a distance threshold and we plot the success rate at different thresholds of the different methods. We compare predictions of InDeep compared with the ones of Kalasanty, for the bound conformation (holo) as well as the unbound one (apo). The DCC values were computed in three conditions: between the computational predictions and (i) the VolSite cavities, (ii) the ligand positions and (iii) the ligand positions that have a VolSite cavity associated with them (top of Fig. 2). The DVO metric is only computed on successful prediction at 6 Å and consists in the volume of the overlap over the volume of the union. This procedure is illustrated in Supplementary Figure S3 and the DVO results are presented in the top of Supplementary Figure S4.

First of all, we reproduce the results claimed by Kalasanty. In this benchmark dataset, InDeep outperforms Kalasanty on all

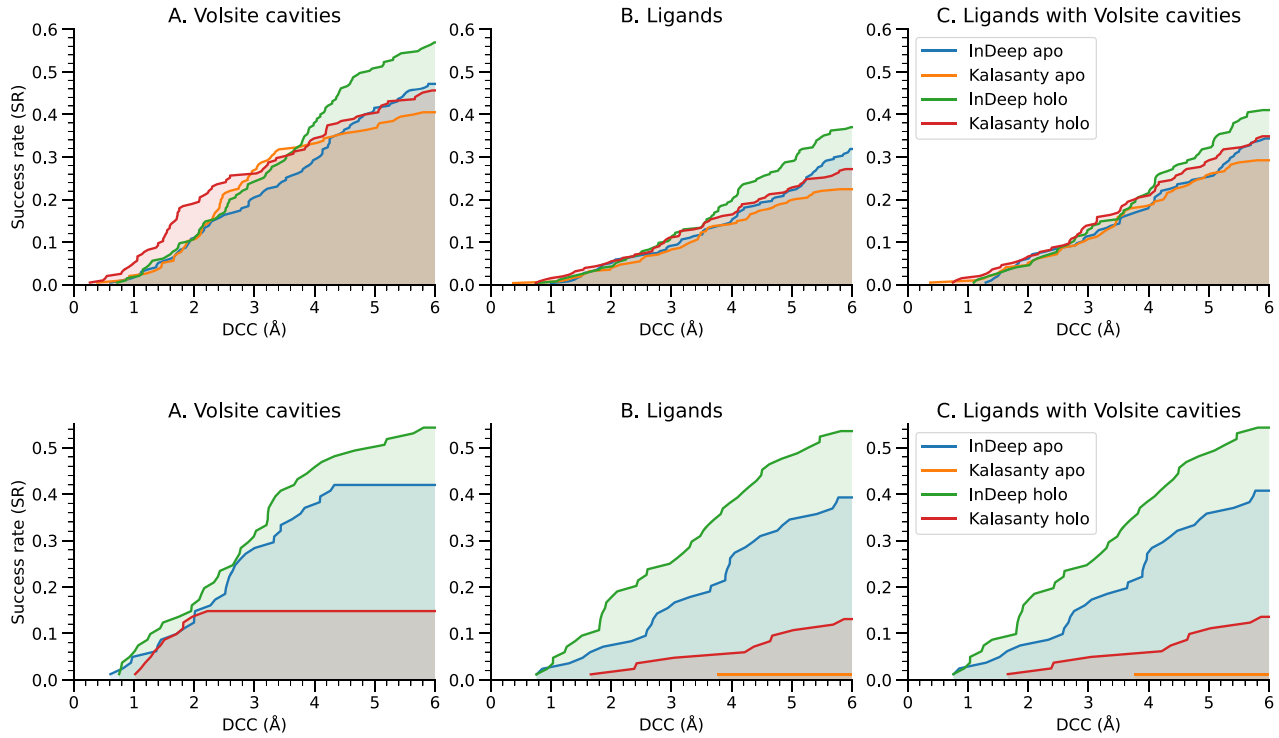

Fig. 2. (Top) DCC evaluation for InDeep and Kalasanty on the Chen benchmark. We plot the success rate (SR), the fraction of systems for which we have a DCC value below a threshold, for different thresholds. (Bottom) Performance on the test set filtered by TM-score. The plots are produced following the same procedure as the ones above on this new dataset. The metrics are computed with the VolSite cavities associated with the ligand position given by the PDB (**A**), the ligand position itself (**B**) and the ligand position having a cavity detected by VolSite (**C**)

settings except in the very low DCC range and for VolSite cavities only. At a 6-Å threshold, we have an average relative performance boost of 27%. We see that this difference is most important in the ligand setting that resembles our training procedure, but that the performance remains stable on the cavities setting. We also see that this boost of performance is observed across holo and apo predictions. We detect approximately 40% of the binding sites in the best ranked predictions. We then compute the DVO values for the different methods among the successful predictions at 6 Å. We obtain comparable values of DVO overall. Because Volsite cavities tend to be deeper into the pocket, Kalasanty predictions tend to be at the bottom of binding sites. This explains why the average DVO value for cavities is better for Kalasanty and worse for InDeep (0.612 versus 0.503) and vice-versa for ligands (0.257 versus 0.320). Moreover, one should note that the total populations are not the same because the DVO is computed only on the 'successful predictions' with low DCC values. Therefore, the better performance in terms of DCC can hinder the DVO distribution performance.

Finally, we further tested InDeep on the complete sc-PDB, an annotated database of druggable binding sites from the protein databank. We computed the DCC values to compare InDeep predictions with the Volsite cavities stored in the sc-PDB and the position of the ligand. At 6 Å, we found a success rate value of 0.71 and 0.68 for the Volsite cavities and the ligand positions, respectively. The full results are available in Supplementary Figure S5. Overall, InDeep yields state-of-the-art performance on the binding site prediction for general ligands, even in the unfavourable setting of predicting cavities, as opposed to actual ligand localization.

3.1.2 Benchmark of InDeep on iPPI binding sites

We have then repeated the same data extraction pipeline as used for the Chen dataset and described in Supplementary Figure S3 on our test set. We have applied the same TM-score filtering between the train set of Kalasanty and our test set to avoid data leakage. This represents a more suitable application of InDeep, as it was designed to identify PPI-specific binding pockets. We also note that this procedure results in few (81) systems, because of the large overlap with the training set of Kalasanty. We present the DCC results in Figure 2 (bottom) and those for DVO in Supplementary Figure S4 (bottom).

Despite the limited size of this filtered dataset, we see that InDeep clearly outperforms Kalasanty. The SR is increased 5-fold and the DVO values remain reasonable. Moreover, we retain 80% of the performance when predicting on the apo form of the protein, an important feature for these binding sites that are known to be hard to detect. This shows that InDeep is not only a good predictor for conventional target binding sites, but that it is much more efficient than existing methods for iPPI binding site detection.

3.2 Predicting and profiling epitope binding sites

3.2.1 Benchmark of InDeep on PPI datasets

As for PL, several tools exist that predict which region of a protein interacts with another. Once again, we choose to compare against the state of the art and reproduce the results of PInet (Dai and Bailey-Kellogg, 2021). We use two benchmarking datasets they propose: DBD5 (Guest *et al.*, 2021) and EpiPred (Krawczyk *et al.*, 2014). DBD5 is a protein docking benchmark that offers several pairs of structures of interacting proteins. This dataset is split into a train and test set by PInet. EpiPred is a dataset centred around interactions between antigens and antibodies. For fairness, we used their protocol to annotate the data as in PInet and have rerun their method. Finally, we have used their tool in the partner-specific setting (giving the partner as input) and in a blind setting that is closer to our use case.

We then had to slightly adapt our validation pipeline. Indeed, we have found no study trying to predict the actual location of the partner in the vicinity of the protein surface. The tools we compare against always project the predicted interactability onto the surface residues of the protein or the sequence. To project our 3D prediction onto the

sequence, we use a convolution with Gaussian kernel between our 3D prediction and the coordinates of the atoms of the protein.

We compute the area under the precision–recall curve for the DBD5 test set and EpiPred. On the DBD5 test set, InDeep is widely outperformed by PInet. However, PInet was trained on a different split of the same dataset, so we turned to EpiPred, that was not used for training either methods. PInet gets a value of 0.235 and 0.217 in the native and blind setting, respectively. On this dataset, we get a value of 0.232, close to the performance of the partner-specific setting. We achieve a state-of-the-art performance in our blind setting. Moreover, it should be noted that InDeep performance suffers from the extra step of sequence projection. Overall, this shows that our tool is able to accurately predict the interaction sites of a protein.

3.2.2 Localization of interacting partner

We complement this comparison to other tools with a validation of InDeep with metrics closer to the PL validation. We compute these metrics on our test set as well as on DBD5 and EpiPred. We note that to our knowledge, no tool exist that output 3D prediction of the volume occupied by a putative protein partner. However, this prediction is of a great use to assess if the 3D prediction for a small molecule binding would collide into its corresponding protein partner, opening new doors for therapeutic design of iPPIs. Since for PPIs, the number of observed partners is just 1, we have computed DCC and DVO values for one, three and all predicted binding sites. Because the interfaces are bigger, we also present our results up to a distance threshold of 10 A. The results are presented in Figure 3.

At the 10-Å threshold, we have success rates of 42%, 24% and 11% with only the first prediction, and of 78%, 72% and 69% using all of them for our test set, DBD5 and EpiPred, respectively. This means that our method finds the correct binding site in about 70% of the cases, but that a significant amount of times, the correct predicted volume is not ranked as the first one. This can be partially imputed to the fact that a given protein can have several partners, so the first prediction might actually be a correct one that does not correspond to the partner at hand. However, InDeep's performance on the top-3 falls in between the performance of the top-1 and keeping all pockets, which indicates that the correct binding sites is often among the best scored position, proving once again the relevance of the tool for epitope binding site prediction.

3.2.3 Epitope binding site prediction: atom-typed channel validation

We now turn to the channel evaluation. We have used five channels to encode our protein environment and our prediction: α-carbon ('CA'), donor and receptor of h-bonds ('HAD'), positively and negatively charged atoms ('POS' and 'NEG') and hydrophobic/aromatic atoms ('COB').

This means that beyond prediction of the presence or absence of a protein partner, the model also predicts which protein atom type should be at a given voxel. This is close to the idea developed by LigVoxel (Skalic *et al.*, 2019) but for the interactability model. However, finding a quantitative metric to describe the quality of these channels is not easy. Indeed, the target now contains several little volumes (each atom type environment) that can be split across the protein partner interface. Therefore, we cannot use the DCC metric easily, because the centres of mass of split volumes do not represent our objects accurately. Moreover, we cannot easily interpret the DVO values, as previous experiments only plotted the DVO for successful DCCs, which we do not have anymore. This is even more true if we consider the large size of the interface, which explains why we cannot use the same validation procedure as LigVoxel.

We have turned to a more direct method for assessing the performance. At each voxel of our prediction, we have a distribution of probability for each channel. We can aggregate these voxel distributions for all voxels around an atom of the ligand to obtain a mean distribution of channels probabilities. We also compute an atom-type specific distribution by aggregating only the voxels around atoms of each specific channel. We plot a heatmap representing the Z-scores of the observed channels distributions compared with the overall ones. We expect to see enhanced values on the diagonal and decreased ones off the diagonal. The results are presented in Supplementary Figure S6.

We see that the hydrophobic channel (COB) performs well at localizing hydrophobic patches of protein partners (Z-score = 1.6). It is an important result as transient PPIs, that represents most of the known PPI targets, are often mediated by hydrophobic patches at the interface whose seclusion from the solvent upon binding helps to regulate protein association. This COB channel can therefore be used as a way to suggest point mutations at the interface when dealing with hydrophobic interaction, or in the context of epitope mimicking or peptide design. The backbone channel (CA) has a more modest performance than COB, although it displays a partial enrichment. In this case, the perspective of depicting the backbone of a putative partner for a given interactability patch is also very pertinent. Indeed, for example the spatial arrangements of Cα within a α-helix are fitting very nicely within a cylinder that can be clearly identified within some PPIs mediated by such secondary structure at the interface (see case study about Bcl-2). Nevertheless, the other channels are clearly non-specific and shall be the subject of improvements in the future.

3.3 Case study: Bcl-2 as therapeutic target

The Bcl-2 protein is the eponymous protein of the Bcl-2 family, which is central to the regulation of apoptosis and vital for proper tissue development and cellular homeostasis (Bajwa *et al.*, 2012). Upon interaction with pro-apoptotic BH3 domains containing

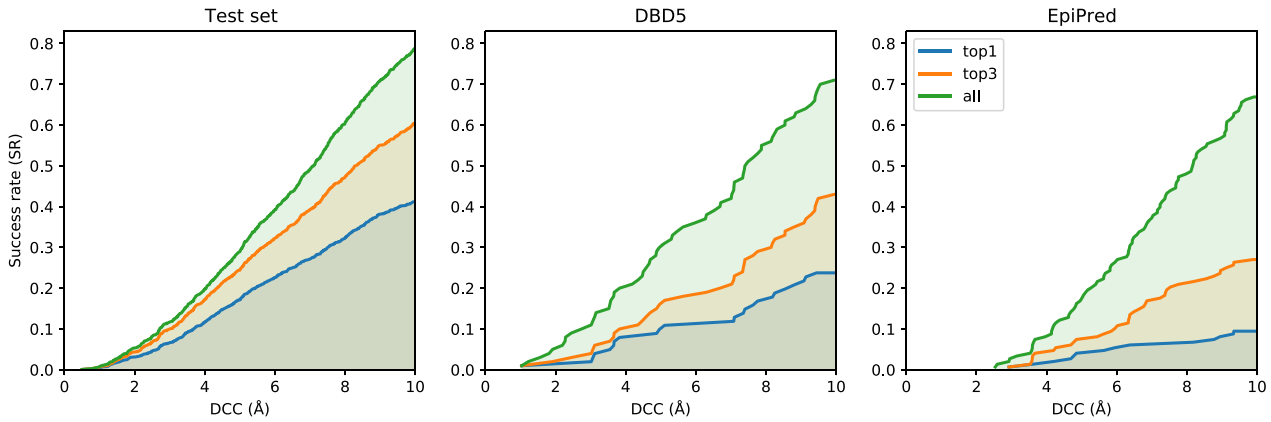

Fig. 3. DCC values for InDeep on the test set (Left), DBD5 (Centre) and EpiPred (Right) datasets, when considering the best, top-3 or all pockets predicted

proteins, Bcl-2 inhibits cell death. In the last two decades, small molecules that disrupt this interaction by binding, to Bcl-2 and other anti-apoptotic proteins of this family, have been successfully designed and clinically approved to induce apoptosis of cancer cells (Vogler *et al.*, 2009). Pro-apoptotic partners of Bcl-2 possess a 20 Å-long α-helical-containing BH3 domains that interact at Bcl-2 surface through an extended hydrophobic groove. A recent review underpinned that the successful development of drugs, such as Venetoclax, against this target was mainly due to the fact they manage to mimic two of the hotspot residues within the BH3 domain binding this groove (Ashkenazi *et al.*, 2017).

This protein family is therefore an excellent case study to retrospectively evaluate the pertinence of using a tool like InDeep. The fact that this tool can predict ligandable/druggable binding pockets, interactability patches including hydrophobic and backbone atom-typed channels and also monitor such predictions along MD trajectories allows a retrospective analysis of feasibility of designing ligands binding to the BH3 groove of Bcl-2. Although Bcl-2 complexes were present in our training set, it is worth noting that the different InDeep predictions below have been made exclusively on the sole structure of Bcl-2 without any consideration for Bax or known co-crystallized ligands.

We can first use InDeep to predict interactability patches at the surface of Bcl-2. As can be seen in Figure 4 (left panel), InDeep correctly predicts (first ranked patch), within the BH3 groove, the location of the interactability patch with the α-helix of its protein partner Bax. Inspecting more specifically the Cα- and hydrophobic-atom-typed channels within the interactability patch Supplementary Figure S7 (Top panel), one can observe, respectively, (i) a faithful depiction of the α-helix shape of the Bax epitope binding the BH3 groove (as a cylinder shape) and (ii) a proper localization of the hydrophobic hotspots known for this system (Ashkenazi *et al.*, 2017).

If we now use the ligandability prediction of InDeep on the same structure of Bcl-2 as co-crystallized with Bax (pdb: 2xa0), one can see in Supplementary Figure S7 (Bottom panel) that InDeep correctly highlights the known aforementioned hotspots as the most ligandable regions of the BH3 groove of Bcl-2. In the end, one can notice in Figure 4 (right panel) that the same regions that were successfully targeted by Venetoclax analogs (ex with pdb: 4lvt) and correctly highlighted by InDeep.

NMR and X-ray crystallography have highlighted important backbone rearrangements within Bcl-2 upon peptide and small molecule binding to the BH3 groove (Liu *et al.*, 2003). Early virtual screening approaches failed to identify validated small molecules inhibitors, as they did not consider protein flexibility (Scott *et al.*, 2016). It is therefore essential to properly sample the flexibility of the system and be able to monitor both ligandable and interactability patches on representative Bcl-2 structure conformations. To do so, a 1-μs-long MD simulation (see MD parameters in Supplementary Section S5), starting from the apo form of Bcl-2 (pdb 1gjh), was run to monitor the ligandability and the interactability of the BH3 groove known to bind Venetoclax and other chemical analogs. The InDeep prediction has been focused on this region of Bcl-2, as the goal of this approach is to detect holo-like conformations, that is favourable conformations for ligand binding and not to detect the binding region on the whole protein surface. Figure 5 (left panel) shows the values of the InDeep ligandability score of each frame (red line) and the minimal binding site root mean square deviation (RMSD) value (blue line) with respect to 16 ligand-bound PL structures (see binding site definition and the list of PL structures in Supplementary Section S6). It can be noted from the fluctuation profiles of the RMSD and the ligandabilty scores that local minima in the RMSD value (highlighted as black points) correspond to ligandability peaks. Therefore, holo-like conformations (with low RMSD against a PL) tend to be correctly predicted as ligandable by InDeep. A similar trend was observed by other groups on conventional targets (Kozlovskii and Popov, 2020). Likewise, the interactability was monitored along the same simulation and the RMSD of the binding site residues was computed against a holo HD structure of Bcl-2 bound to Bax (pdb 2xa0) (Fig. 5, right panel). Similarly to the ligandability model, holo-like conformations have a higher score of InDeep interactability than conformations distant from the HD conformations.

These predictions collectively show that a proper usage of InDeep early in the drug discovery initiative against Bcl-2/Bax would have highlighted the hydrophobic hotspots residues within the BH3 groove of Bcl-2 and the most ligandable binding regions of these spots to assist the design of selective BH3 mimetics. Moreover, InDeep would have been efficient at profiling holo-like conformations prior to virtual screening campaigns even when starting from apo structures, and in the absence of Bax bound to Bcl-2, within a MD trajectory. Finally, as holo-like conformations are not always accessible with regular water-solvated MD simulations (Eyrisch and Helms, 2009), InDeep can be easily combined to other sampling protocols in order to profile the resulting pocket conformations.

4 Discussion

We have introduced InDeep, a unified prediction tool for structure-based drug design targeting protein interfaces. We show that this tool is competitive in detecting the residues that interact with a protein partner. We go beyond this sequence prediction by predicting the localization in space of a putative partner using atom-typed channels signal that helps understanding how such protein interaction can take place. Seventy per cent of the observed binding site are present in one of our prediction and 35% in the top three ones.

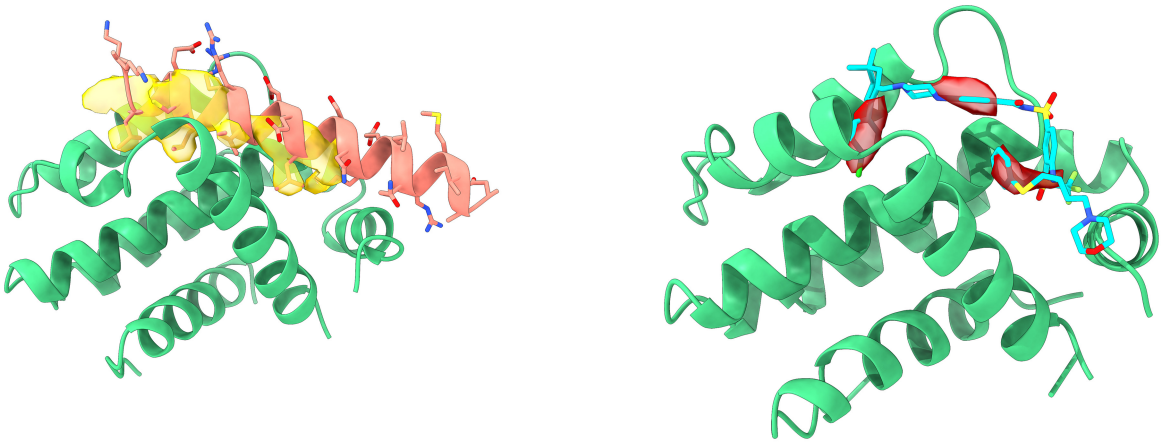

Fig. 4. (Left) InDeep interactability patch prediction on Bcl-2 (pdb 2xa0). (Right) Ligandability prediction (red surface) performed on Bcl-2 (pdb 4lvt) surface. The red surface patches of InDeep ligandability are localized around the known hot spots of the Bcl-2/Bax interaction that are mimicked by some of the ligand atoms

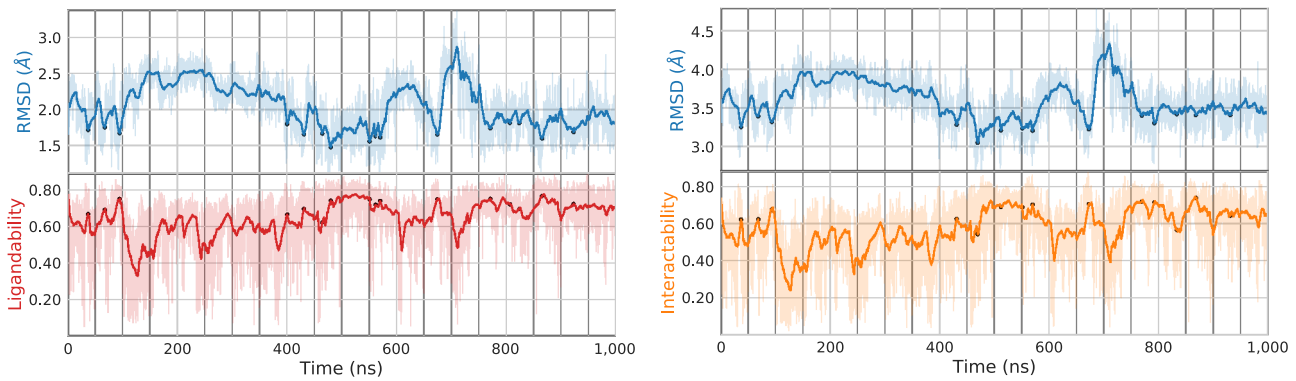

Fig. 5. InDeep predictions (red and orange) along MD trajectory of Bcl-2. Moving average (exponential) on 200 frames are represented with solid lines. (Left) InDeep ligandability score evolution (red) compared with the minimal binding site RMSD (blue) with respect to 16 PL structures. (Right) InDeep interactability score evolution (orange) compared with the RMSD (blue) with respect to the reference HD structure. Local minima on the RMSD curve and their corresponding InDeep predictions are highlighted as black points

Moreover, InDeep clearly outperforms the state of the art of binding pockets detection for iPPI binding sites but even for conventional targets. Combining those two predictions for a newly solved structure, one can investigate binding sites for ligands that would potentially disrupt a given PPI. Given such a detected binding site, our tool also enables tracking ligandability or interactability scores along a MD trajectory, which opens the door for a refined ligandability assessment as well as conformation selection for virtual screening or unravel epitope binding-prone conformations. Finally, we illustrate these functionalities in a retrospective drug discovery use case on Bcl-2. InDeep is integrated in a PyMol (DeLano *et al.*, 2002) plugin for easy visualization of the predictions (see Section 7).

Despite several promising results, the therapeutic use of iPPIs remains a minority. We hope this dedicated tool can help enhance their use as well as spark a development of other methods following this line of work. InDeep uses 3D-FC U-Net because the grid-like prediction enables actual localization and profiling of protein partners or future drugs. However, this method, as many others, is sensitive to rotation, a possible extension would be to use equivariant networks (Weiler *et al.*, 2018) to bake the rotation invariance in the network.

Future developments also include actual drug discovery project use of the tool as well as the implementation of a web-server for easier access to the predictions.

Acknowledgements

We thank IBM for their sponsorship for the optimization of the hyperparameters. In particular, we thank Jean Armand Broyelle, Maxime Deloche and Xavier Vasquez for their expert support. We are grateful to the Institut Pasteur and the CNRS for their continued support for our research. We also thank Arnaud Blondel for feedback and discussions.

Financial Support: V.M. is recipient of a doctoral fellowship from the INCEPTION project [PIA/ANR-16-CONV-0005] and benefits from support from the CRI through 'Ecole Doctorale FIRE—Programme Bettencourt'.

Conflict of Interest: none declared.

References

Amaro,R.E. *et al.* (2018) Ensemble docking in drug discovery. *Biophys. J.*, **114**, 2271–2278. ISSN 00063495. doi: 10.1016/j.bpj.2018.02.038.

Arkin,M.R. *et al.* (2014) Small-molecule inhibitors of protein–protein interactions: Progressing toward the reality. *Chem. Biol.*, **21**, 1102–1114. doi: 10.1016/j.chembiol.2014.09.001.

Ashkenazi,A. *et al.* (2017) From basic apoptosis discoveries to advanced selective bcl-2 family inhibitors. *Nat. Rev. Drug Discov.*, **16**, 273–284.

Bajwa,N. *et al.* (2012) Inhibitors of the anti-apoptotic Bcl-2 proteins: A patent review. *Exp. Opin. Ther. Pat.*, **22**, 37–55. doi: 10.1517/13543776.2012.644274.

Beucher,S. (1979) Use of watersheds in contour detection. In *Proceedings of the International Workshop on Image Processing*, CCETT.

Chen,K. *et al.* (2011) A critical comparative assessment of predictions of protein-binding sites for biologically relevant organic compounds. *Structure*, **19**, 613–621.

Clackson,T. and Wells,J.A. (1995) A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386. doi: 10.1126/science.7529940.

Cong,Q. *et al.* (2019) Protein interaction networks revealed by proteome coevolution. *Science*, **365**, 185–189.

Silva,F.D. *et al.* (2018) IChem: A versatile toolkit for detecting, comparing, and predicting protein–ligand interactions. *ChemMedChem*, **13**, 507–510.

D'Aguanno,S., and Bufalo,D.D. (2020) Inhibition of anti-apoptotic Bcl-2 proteins in preclinical and clinical studies: Current overview in cancer. *Cells*, **9**, 1287.doi: 10.3390/cells905

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

DeLano,W.L. *et al.* (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newslett. Prot. Crystallogr.*, **40**, 82–92.

Eyrisch,S. and Helms,V. (2009) What induces pocket openings on protein surface patches involved in protein–protein interactions? *J. Comput. Aided Mol. Des.*, **23**, 73–86.

Eyrisch,S. *et al.* (2012) Transient pockets on xiap-bir2: Toward the characterization of putative binding sites of small-molecule xiap inhibitors. *J. Mol. Model.*, **18**, 2031–2042.

Gainza,P. *et al.* (2020) Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods*, **17**, 184–192.

Gao,M. and Skolnick,J. (2013) A comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.*, **9**, e1003302.ISSN 1553734X. doi: 10.1371/journal.pcbi.1003302.

Goodfellow,I. *et al.* (2016) *Deep Learning*. MIT Press.

Guest,J.D. *et al.* (2021) An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure*, **29**, 606–621.

Guilloux,V.L. *et al.* (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.*, **10**, 168–111. doi: 10.1186/1471-2105-10-168.

Ivetac,A. and McCammon,J.A. (2012) A molecular dynamics ensemble-based approach for the mapping of druggable binding sites. *Methods Mol. Biol.*, **819**, 3–12. doi: 10.1007/978-1-61779-465-0{_}1.

Jiménez,J. *et al.* (2017) Deepsite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics*, **33**, 3036–3042.

Johnson,D.K. and Karanicolas,J. (2013) Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Comput. Biol.*, **9**, e1002951.

Kingma,D.P. and Ba,J. (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kozakov,D. *et al.* (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.*, **10**, 733–755. doi: 10.1038/nprot.2015.043.

Kozlovskii,I. and Popov,P. (2020) Spatiotemporal identification of druggable binding sites using deep learning. *Commun. Biol.*, **3**, 1–12. doi: 10.1038/s42003-020-01350-0.

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

Krivák,R. and Hoksza,D. (2018) P2Rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminform.*, **10**, 1–12. doi: 10.1186/S13321-018-0285-8.

KrügerD.M. and Gohlke,H. (2010) DrugScorePPI webserver: Fast and accurate in silico alanine scanning for scoring protein–protein interactions. *Nucleic Acids Res.*, **38**, W480–7. doi: 10.1093/nar/gkq471.

Liu,Q. *et al.* (2021) OctSurf: Efficient hierarchical voxel-based molecular surface representation for protein–ligand affinity prediction. *J. Mol. Graph. Model.*, **105**, 107865.doi: 10.1016/J.JMGM.2021.107865.

Liu,X. *et al.* (2003) The structure of a Bcl-xL/Bim fragment complex: Implications for bim function. *Immunity*, **19**, 341–352.

Lu,H. *et al.* (2020) Recent advances in the development of protein–protein interactions modulators: Mechanisms and clinical trials. *Signal Transduct. Target. Ther.*, **5**, 1–23.

Monet,D. *et al.* (2019) mkgridXf: Consistent identification of plausible binding sites despite the elusive nature of cavities and grooves in protein dynamics. *J. Chem. Inform. Model.*, **59**, 3506–3518.

Murakami,Y. and Mizuguchi,K. (2010) Applying the naïve Bayes classifier with kernel density estimation to the prediction of protein–protein interaction sites. *Bioinformatics*, **26**, 1841–1848.

Mylonas,S.K. *et al.* (2021) DeepSurf: A surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*, **37**, 1681–1690. doi: 10.1093/BIOINFORMATICS/BTAB009.

Ronneberger,O. *et al.* (2015) U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.

Rosell,M. and Fernández-Recio,J. (2020) Docking-based identification of small-molecule binding sites at protein–protein interfaces. *Comput. Struct. Biotechnol. J.*, **18**, 3750–3761.

Sanchez-Garcia,R. *et al.* (2019) BIPSPI: a method for the prediction of partner-specific protein–protein interfaces. *Bioinformatics*, **35**, 470–477.

Scott,D.E. *et al.* (2016) Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. Nature Reviews Drug Discovery, 15, 533–550.

Sillitoe,I. *et al.* (2021) Cath: Increased structural coverage of functional space. *Nucleic Acids Res.*, **49**, D266–D273.

Skalic,M. *et al.* (2019) Ligvoxel: Inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics*, **35**, 243–250.

Sperandio,O. *et al.* (2010) Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discov. Today*, **15**, 220–229.

Stepniewska-Dziubinska,M.M. *et al.* (2020) Improving detection of protein–ligand binding sites with 3D segmentation. *Sci. Rep.*, **10**, 1–9.

Torchet,R. *et al.* (2021) The iPPI-DB initiative: A community-centered database of protein–protein interaction modulators. *Bioinformatics*, **37**, 89–96. doi: 10.1093/bioinformatics/btaa1091.

Townshend,R. *et al.* (2019) End-to-end learning on 3D protein structure for interface prediction. *Adv. Neural Inform. Process. Syst.*, **32**, 15642–15651.

Tuncbag,N. *et al.* (2010) HotPoint: Hot spot prediction server for protein interfaces. *Nucleic Acids Res.*, **38**, W402–W405. doi: 10.1093/NAR/GKQ323.

Vogler,M. *et al.* (2009) Bcl-2 inhibitors: Small molecules with a big impact on cancer therapy. *Cell Death Differ.*, **16**, 360–367.

Weiler,M. *et al.* (2018) 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Adv. Neural Inform. Process. Syst.*, 10381–10392.

Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Prot. Struct. Funct. Bioinform.*, **57**, 702–710.