Transfer Learning for Drug-Target Interaction Prediction

A.Dalkıran[1], Ahmet Atakan[1], A.S. Rifaioğlu[2], M.J. Martin[3], R. Çetin Atalay[4], A. Acar[1], T. Doğan[5] and V. Atalay[1]

[1]Middle East Technical University, Turkey [2]University of Heidelberg, Germany [3]EMBL-European Bioinformatics Institute, UK, [4]University of Chicago, USA, [5]Hacettepe University, Turkey

Abstract

Utilizing AI-driven approaches for DTI prediction require large volumes of training data which are not available for the majority of target proteins. In this study, we investigate the use of deep transfer learning for the prediction of interactions between drug candidate compounds and understudied target proteins with scarce training data. The idea here is to first train a deep neural network classifier with a generalized source training dataset of large size and then reuse this pre-trained neural network as an initial configuration for re-training/fine-tuning purposes with a small-sized specialized target training dataset. To explore this idea, we selected six protein families that have critical importance in biomedicine: *kinases*, *G-protein-coupled receptors* (*GPCRs*), *ion channels*, *nuclear receptors*, *proteases*, and *transporters*. The protein families of *transporters* and *nuclear receptors* were individually set as the target datasets, while the other five families were used as the source datasets. Several size-based target family training datasets were formed in a controlled manner. Here we present a disciplined evaluation by pre-training a feed-forward neural network with source training datasets and applying different modes of transfer learning from the pre-trained source network to a target dataset. The performance of deep transfer learning is evaluated and compared with that of training the same deep neural network from scratch. We found that when the training dataset is smaller than 100 compounds, transfer learning yields significantly better performance compared to training the system from scratch, suggesting an advantage to using transfer learning to predict binders to under-studied targets.

1 INTRODUCTION

Drugs are chemical compounds that are used to treat diseases or to increase the quality of life. A drug is intended to interact with a target biomolecule (e.g., a single protein, several proteins or a protein complex) by regulating or correcting cellular functions in pathological conditions. Although drug discovery is traditionally a long, laborious and costly process, recently, there have been innovative and promising computational solutions based on machine learning and deep learning. Virtual screening of compounds against a target cell or protein is used widely during the initial steps of the drug discovery process. Lately, deep learning-based models for virtual screening and drug-target interaction (DTI) prediction have yielded highly promising results (Bagherian *et al.*, 2020; Baskin, 2020; Chen *et al.*, 2018; Du *et al.*, 2022; Elbadawi *et al.*, 2021; Ezzat *et al.*, 2018; Jing *et al.*, 2018; Kim *et al.*, 2021; Lo *et al.*, 2018; Pan *et al.*, 2022; Réda *et al.*, 2020; Rifaioglu *et al.*, 2019; Vamathevan *et al.*, 2019; Wang and Kurgan, 2018; Zhang *et al.*, 2022). However, the majority of deep learning models developed thus far require a large volume of training data. Such a large volume of data is not available for many of the target proteins or protein families, and therefore, no prediction models are available for these classes of proteins.

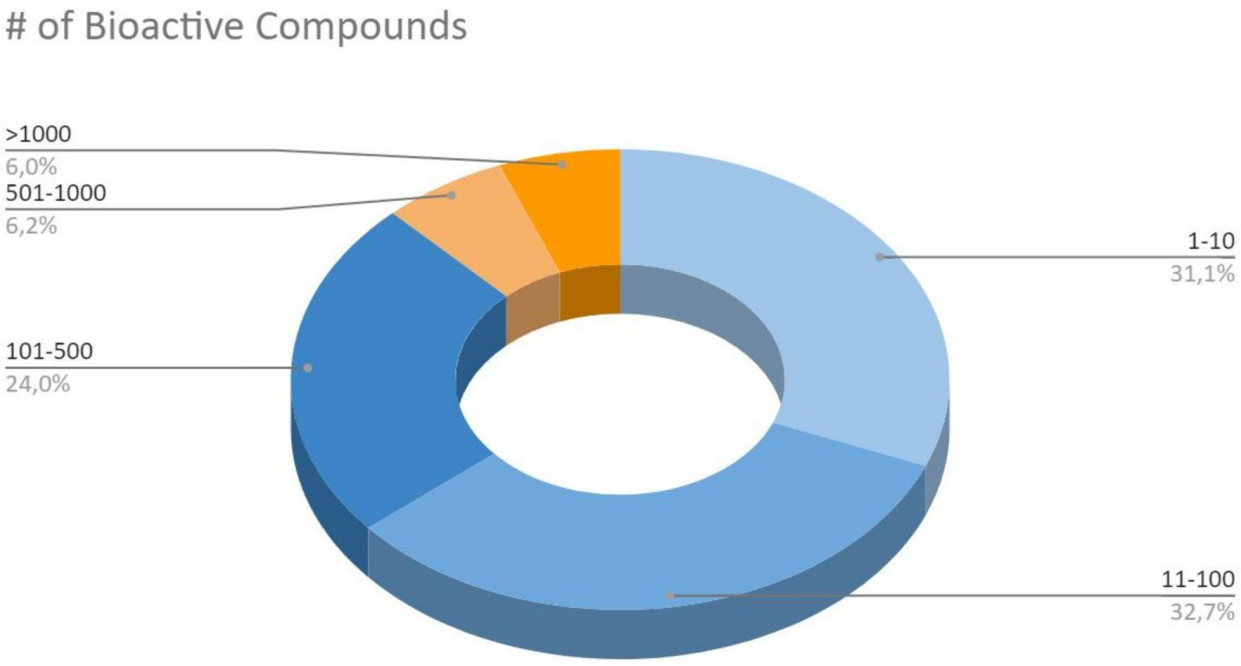


Figure 1. The distribution of the number of compounds per target protein in the bioactivity data from the ChEMBL_29 database.

The distribution of compounds for target proteins over the bioactivity data in the ChEMBL_29 database indeed reveals the problem of limited training data precisely. Figure 1 shows this distribution of compounds with associated bioactivity data for target proteins in ChEMBL_29. In this case, we ask the following question: is drug-target interaction prediction possible when a limited amount of bioactivity data is available for a target protein? The crucial issue is to find a solution when a protein (or a protein family) has a low amount of bioactivity data, particularly in the case where there is the risk of overfitting for the prediction model since the input feature vector is generally of high dimensionality and deep learning models are prone to "memorizing" rather than learning without sufficient training data. When transfer learning or few- or zero-shot learning is incorporated, it becomes possible to learn from such a low amount of data. *Transfer learning* is a machine learning approach where a model is trained for a *source* task, and this pre-trained source model is then reused as an initial configuration to build (train) a model (target model) for a different but related *target* task. (Yosinski *et al.*, 2014) explain the basic principles and methods of transfer learning in deep artificial neural networks, while (Tan *et al.*, 2018) have compiled studies using deep transfer learning.

Deep transfer learning has not been extensively exploited so far in the area of drug-target interaction prediction (Cai *et al.*, 2020; Kao *et al.*, 2021; Lee *et al.*, 2019; Li *et al.*, 2021; Playe and Stoven, 2020). To this end, we investigate the use of deep transfer learning for the prediction of interactions between drugs/compounds and understudied target proteins that have scarce training data and we present a disciplined evaluation for this aim. We formulated drug-target interaction prediction as a ligand-based binary classification problem. For this, a feed-forward neural network (FNN) with two hidden layers is used in which the input is learned representations from Chemprop (Yang *et al.*, 2019). Chemprop employs a directed message-passing deep neural network model that transforms the

graphical representation of molecules into continuous vectors via a directed connection-based message-passing approach. In terms of data, we have selected six of the main protein families: *G-protein-coupled receptors (GPCRs), ion channels, kinases, nuclear receptors, proteases, and transporters*. *Transporters* and *nuclear receptors* were each family was separately used as the target dataset, while the other five families were set as the source datasets. Deep transfer learning was carried out by training the FNN with the experimental bioactivity measurements of a source dataset generated from one of the five protein families and applying the three modes of transfer learning on the small-sized target family (transporter or nuclear receptor) training datasets. The small-sized target family training datasets are generated in a controlled manner to pursue a disciplined evaluation approach. We then compared the performance of this deep transfer learning approach with the case where the FNN was trained from scratch. We also compared it against a shallow classifier.
In Section 2, we give background information on deep transfer learning. This is followed by a discussion of the related research in Section 3. The data (Section 4) and method (Section 5) are then presented in addition to the experimental evaluation (Section 6). Section 7 finally presents a discussion and conclusions.

2 BACKGROUND INFORMATION-DEEP TRANSFER LEARNING

A machine learning problem involves a domain, D, and a task, T. Given a source problem and a target problem, the source domain is D_s, and the target domain is called D_t, while the source task is T_s, and the target task is T_t. Transfer learning aims to learn D_t and improve the performance of T_t with the help of D_s and T_s. In practice, a domain is represented by a dataset. During the initial steps of drug discovery, the task is typically the prediction of the interaction or bioactivity of the drug with the target protein(s) or the prediction of the absorption, distribution, metabolism, elimination, and toxicity (ADMET) of the drug. The domain is typically the set of molecules described by features such as chemical descriptors. In our case, the task remains the same, and the transfer is between domains, i.e., between different molecular (compound) datasets.

Deep transfer learning is applying transfer learning on deep neural networks. The training phase of deep transfer learning is composed of two stages.

Stage I: A *source model* is obtained by training the network with a sufficient number of source training data. This is also referred to as the pre-trained source model.
Stage II: The pre-trained source model is used as an initial configuration and re-trained using target training data (which is typically small) to obtain a *target model*.

Techniques for Stage II are grouped under three modes. Note that the architecture of a deep neural network can be functionally decomposed into roughly two parts: the bottom layer(s) where feature extraction is performed and the upper layer(s) where prediction is performed. Mode 2 and Mode 3 make use of this functional decomposition of the network.

Mode 1 – Full Fine-tuning: The most common deep transfer learning technique is fine-tuning, which is in fact parameter-based transfer learning. Based on the assumption that the learned parameter values (weights) contain useful knowledge obtained from the source domain, we seek to achieve better performance by moving these parameter values (weights) to the target model. The parameter values acquired from the source model form the initial values of the parameters of the target model. In this way, the weights of the target model do not start with random values but with the converged values of the weights of the pre-trained source model, and the target model is re-trained with a small number of target training data and converges faster as well with a reduced number of training epochs.

Mode 2 – Feature Transformer: The source model is in fact used to form a latent feature space that is common to both source data and target data. This is indeed feature-based transfer learning. The feature transformer can be obtained by freezing the bottom layers (which are used for feature extraction) of the pre-trained source model during Stage II; that is, the weights of the nodes at the bottom layers are not updated during re-training with the target training data. Only the weights of the nodes at the output layer (i.e., the predictor) are modified with the limited number of target training data.

Mode 3 – Shallow Classifier: In Stage II, the output layer (predictor) of the source model is replaced with a shallow classifier. Hence, only the shallow classifier is trained with the target data and the feature vectors for the target data are extracted by the frozen bottom layers of the source model. Mode 3 is similar to Mode 2, except that a shallow classifier is trained instead of the output layer (predictor part) of the model.

3 RELATED RESEARCH

A comprehensive literature review on transfer learning in drug discovery is given by (Cai *et al.*, 2020). In the drug discovery field, most deep transfer learning studies have been carried out for the prediction of compound properties, generation of molecules, and structure-based virtual screening. Here, we focus on deep transfer learning studies related to ligand-based and feature-based chemogenomic drug-target interaction prediction methods. In (Lee *et al.*, 2019), a pairwise-input neural network model (chemogenomic drug-target interaction predictor) was trained for the classification of compound-target protein interaction, and toxicity (activity) was chosen as the target task in transfer learning. When compared, transfer learning was more successful than training from scratch. (Li *et al.*, 2021) employed graph neural networks with the aim of compound representation learning and by using this GNN as a feature transformer, a chemogenomic drug-target interaction predictor was assessed. The study was not necessarily carried out on limited data or small datasets. (Kao *et al.*, 2021) made an extensive analysis and examined how much data a network (chemogenomic drug-target interaction regressor) needs to achieve an acceptable drug-target interaction prediction performance via transfer learning with full fine-tuning through several datasets, including KIBA, Davis, and some others extracted from ChEMBL. The main critical point of this study is that most of these datasets are from the same protein family, i.e., enzymes. Another comprehensive study on transfer learning in drug-target interaction prediction (chemogenomic drug-target interaction binary classifier) is by (Playe and Stoven, 2020). They reported that transfer learning by full fine-tuning technique might improve the prediction performance if the source task is highly similar to the target task. (Dey *et al.*, 2022) used instance-based and feature-based transfer learning in contrast to the popular parameter-based transfer learning, such as pre-training.

4 DATA

Training datasets and test datasets were generated from the ChEMBL database version 29 by applying the data filtering protocol developed in our previous study (Rifaioglu *et al.*, 2020). pChEMBL value=7.0 (100 nM) was used to separate the active and inactive compounds of each target. We clustered compounds and selected representative member compounds from each cluster to avoid chemical series bias during training and the evaluation of the model. The statistics for the datasets after the filtering steps are given in Table 1. To construct source training datasets, target training datasets and test datasets, we have selected six main protein families: G-protein-coupled receptors (GPCR), ion channel, kinase, nuclear receptor, protease, and transporter. The *transporter and nuclear receptor* families were selected as the target datasets (separately), while the other five families were selected as the source datasets.

Table 1. Numbers of active and inactive compounds, training dataset size, and test dataset size for all protein families.

protein family	*active*	*inactive*	*training dataset*	*test dataset*
GPCR	36,924	31,085	56,675	11,334
ion channel	5,996	14,167	16,803	3,360
kinase	35,531	30,778	55,259	11,050
nuclear receptor	5,099	6,668	9,807	1,960
protease	15,718	19,518	29,364	5,872
transporter	3,666	5,898	7,970	1,594

To generate training datasets containing lower numbers of drug-target interaction data points in a controlled manner, we randomly selected compounds from the original *transporter* training dataset and *nuclear receptor* training dataset. Eight smaller and balanced (containing the same number of active and inactive compounds) target training datasets were constructed where the numbers of bioactivities are 2, 6, 12, 48, 96, 400, 1,000 and 4,000. Tests for all eight smaller target training datasets were carried out with the *transporter* family test dataset (containing 1,594 bioactivity data points in the test dataset) and the *nuclear receptor* family test dataset (containing 1,960 bioactivity data points in the test dataset).

5 METHOD

We formulated drug-target interaction prediction as a ligand-based binary classification problem. Therefore, we considered deep neural networks having compound features at the input which perform binary classification (i.e., having binary output).

The training phase is sketched in Figure 2. At the same time as the training phase, for comparison purposes, we trained, from scratch, an FNN having exactly the same configuration (reference model) as well as a shallow classifier (base model), using this same target training dataset. During the test phase, all three models trained with the same target training dataset are tested with an independent target test dataset and the performance is evaluated and comparisons are made. All datasets are generated using the learned representations of Chemprop (Yang *et al.*, 2019).

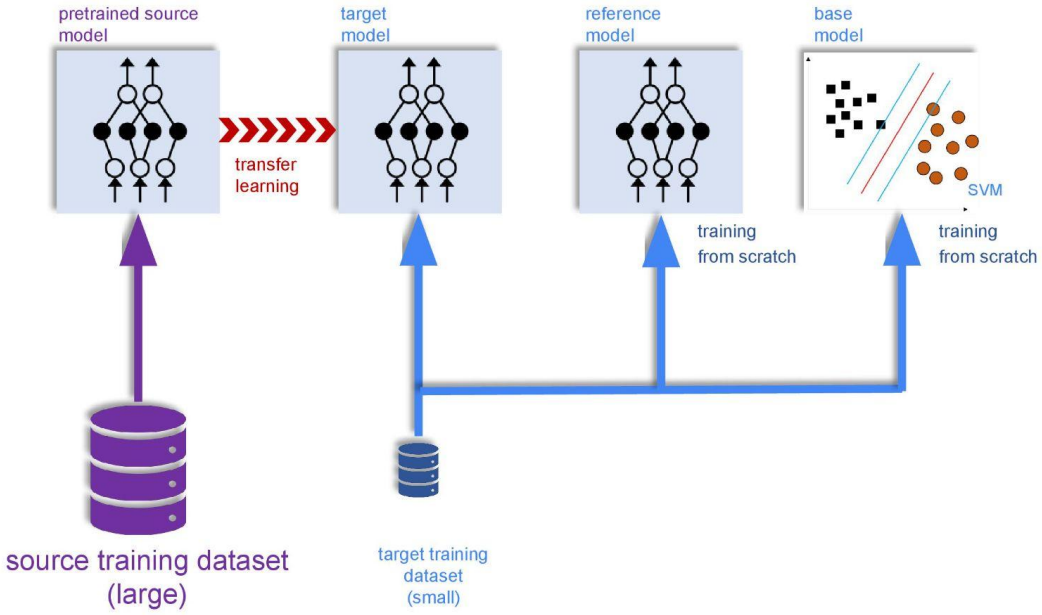
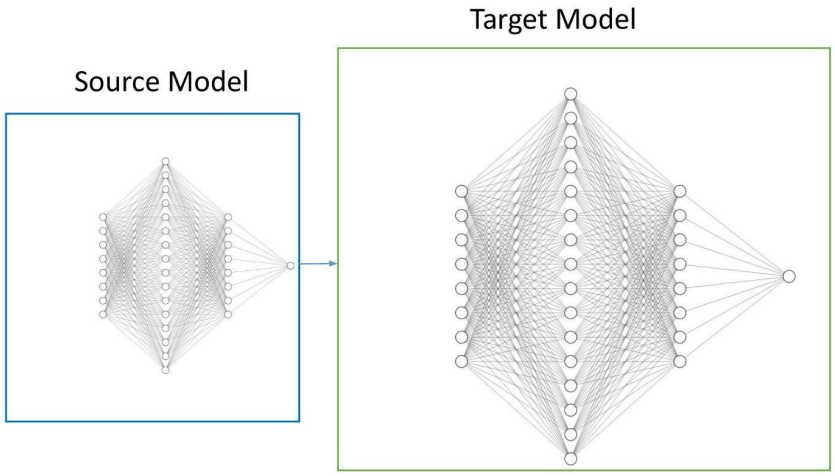
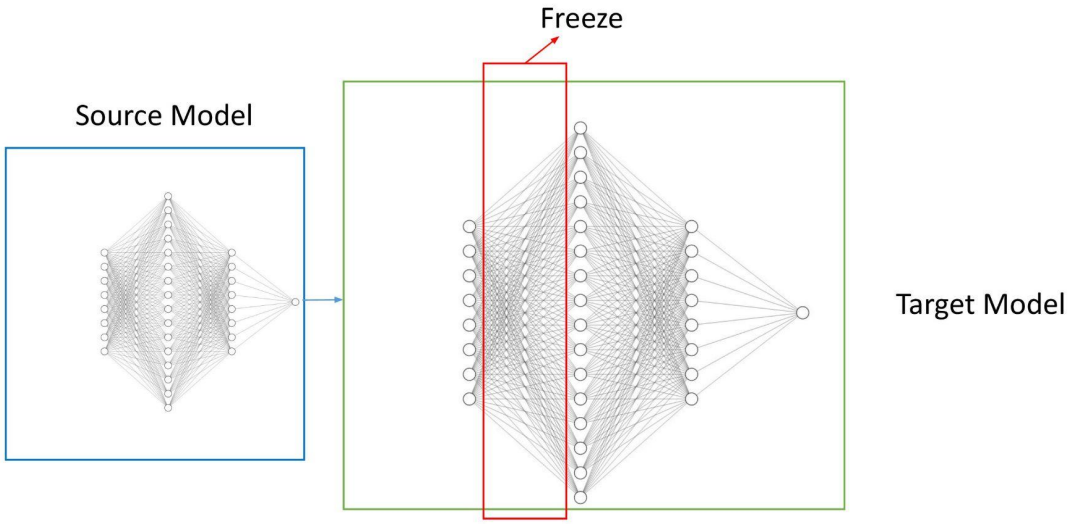


Figure 2. Sketch of the training phase. During the training phase, we first trained a source neural network model with a training dataset of a source family (Stage I). This pre-trained source model is then used for transfer learning to retrain it with a small-sized target training dataset (Stage II). We also trained, from scratch, an FNN having exactly the same configuration (reference model) as well as a shallow classifier (base model), using this same target training dataset.

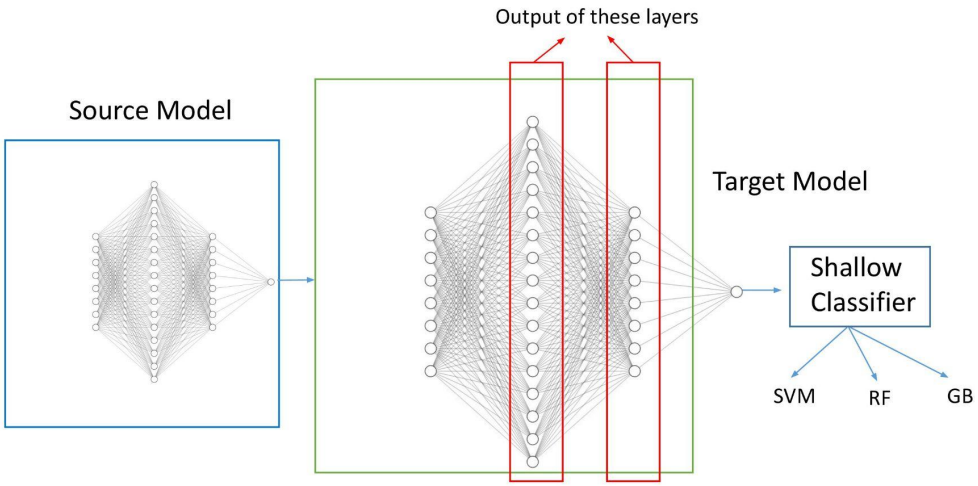
To choose a neural network architecture and determine its configuration, we compared the performances of several different architectures such as feed-forward neural network (FNN) with various numbers of hidden layers, one-dimensional convolutional neural network (1D-CNN), two-dimensional CNN (2D-CNN) with various input compound representations. The best performance is obtained by the model composed of an FNN with two hidden layers where Chemprop learned representations (FNN-2-Chemprop) are used as input. FNN-2-Chemprop performs binary classification (as active or inactive) using compound features at the input level. In FNN-2-Chemprop, a compound is represented by a numerical vector of length 300 which is obtained by using the learned representations of Chemprop. Training and test split was employed to tune the hyperparameter values. The final values of hyperparameters used in FNN-2-Chemprop are as follows: number of hidden layers = 2; hidden layer sizes = 1200 and 300; learning rate = 0.0001; number of training epochs = 100; batch size = 256. Visual representations of three modes of transfer learning (described in Section II) on FNN-2-Chemprop are demonstrated in Figure 3.


Target Model
Source Model

a)


Freeze
Source Model
Target Model

b)



c)

Figure 3. Visual representations of three modes of transfer learning (described in Section II) on FNN-2-Chemprop a) Mode 1 Full Fine-tuning, b) Mode 2 Feature Transformer, c) Mode 3 Shallow Classifier.

Representations (features) of compounds are learned by using Chemprop (Yang *et al.*, 2019). Chemprop is a graph convolutional neural network model consisting of two parts: a Directed Message Passing Neural Network (DMPNN) and a Feed-forward Neural Network (FFN). Message Passing Neural Network (MPNN) is a model that works on an undirected graph with node properties and edge properties. In Chemprop, training data for each compound includes the SMILES string and a target value for the task. In this study, we trained Chemprop to perform the relevant task as a binary classifier (e.g., discriminating between active and inactive compounds in the training dataset of the kinase protein family). We then removed the final classifier FNN layer and we used the values of 300 nodes in the last layer of the DMPNN as the representative (feature vector) of the compounds both in the training dataset and test dataset of the protein families. Thus, representations were learned for a specific task.

6 EXPERIMENTAL EVALUATION

For comparison purposes, we selected the FNN-2-Chemprop that was trained from scratch with the whole target training dataset (without any transfer learning involved) as the *reference model* and a Support Vector Machine (SVM) that was trained from scratch again with the whole target training dataset (without any transfer learning involved) as the *base model*. An SVM was used as the shallow classifier in Mode 2. Matthews correlation coefficient (MCC) was chosen as the evaluation metric to measure performance. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch), the base (shallow) model (an SVM trained from scratch) and the three transfer learning modes are shown in Figure 4, for four test dataset sizes. MCC values are the averages of several repeated experiments. Small datasets were created randomly for every experiment. In each case, the *transporter* is the target family and one of the other five families is used as the source family. A similar evaluation is given in Figure 5 where the *nuclear receptor* is the target family and one of the other five families is used as the source family.

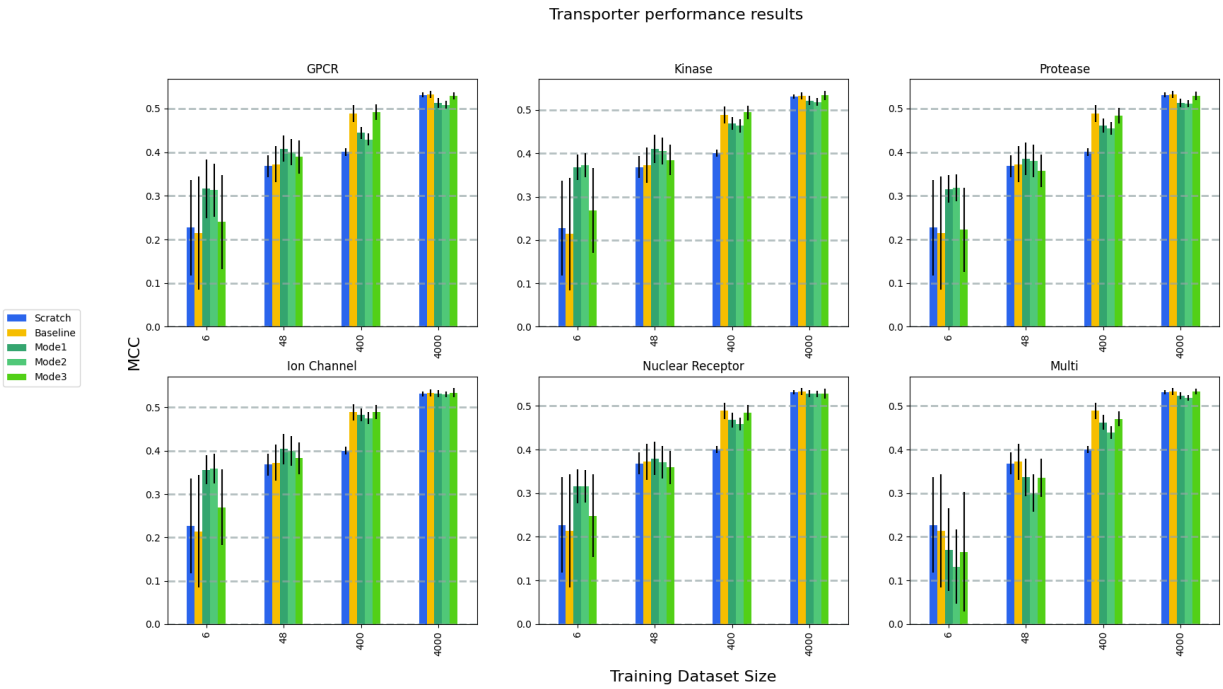


Figure 4. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch), the base model (an SVM trained from scratch) and the three transfer learning modes. The results are given for four cases and in each case, the transporter is the target family and one of the other five families is used as the source family.

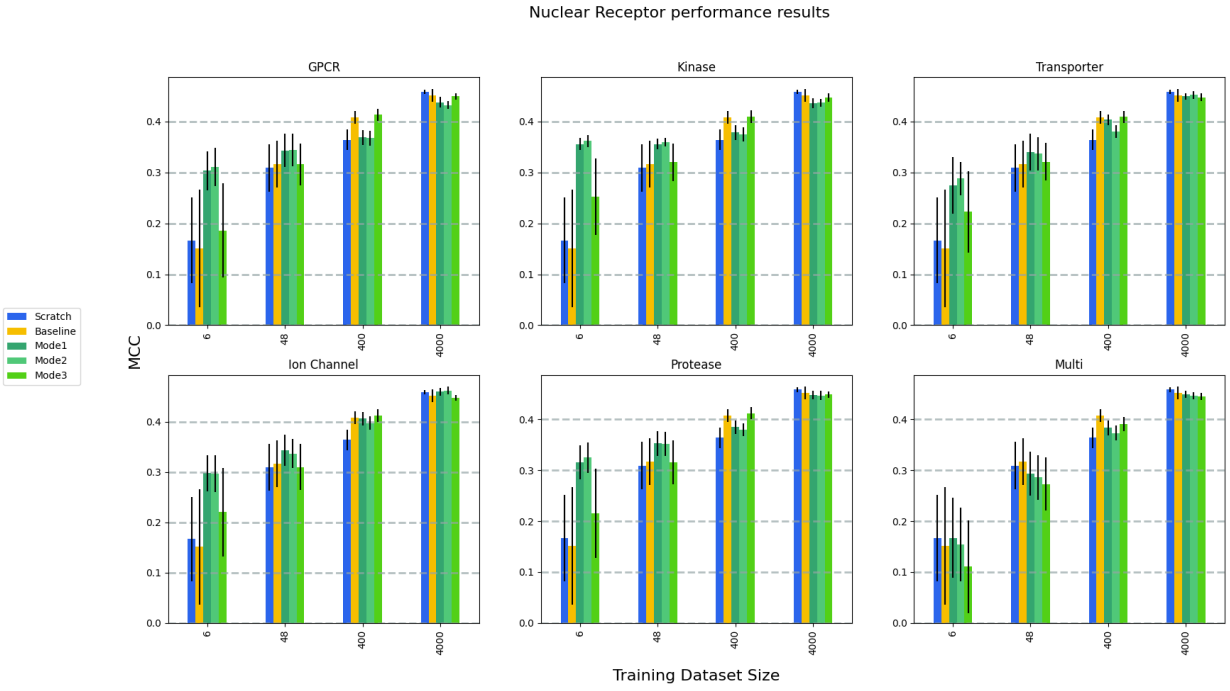


Figure 5. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch), the base model (an SVM trained from scratch) and the three transfer learning modes. The results are given for eight cases and in each case, the nuclear receptor is the target family and one of the other five families is used as the source family.

One of the subplots of Figure 4, where the source protein family is kinase and the target protein family is the transporter, is given in detail in Figure 6. The effect of learning by transfer is better understood in this plot. A similar subplot is given in Figure 7, where the source protein family is the nuclear receptor. Similar trends occur when plots are drawn for other source protein families.

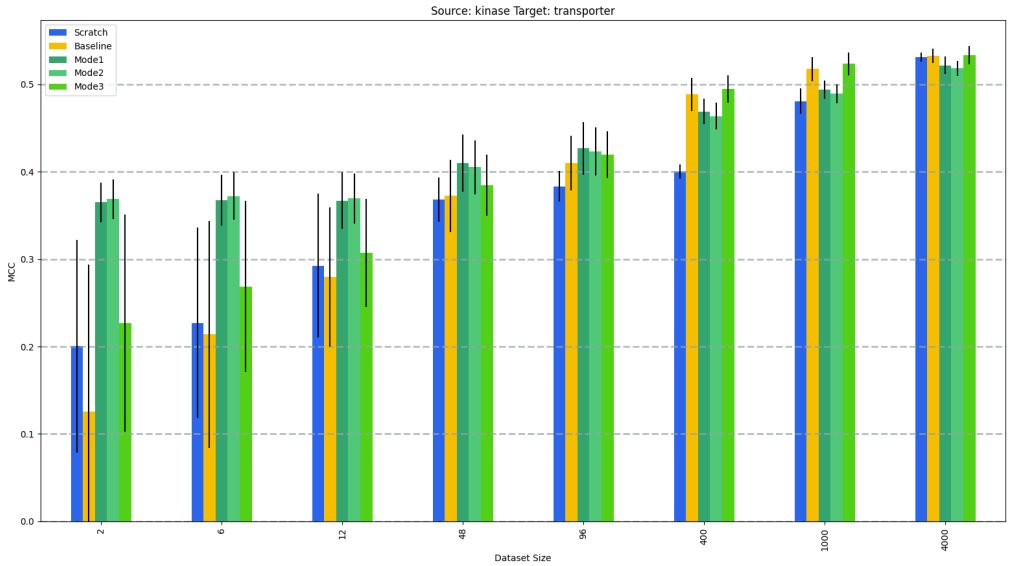


Figure 6. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch), the base model (an SVM trained from scratch) and the three transfer learning modes for all of the eight cases when kinase is the source family and transporter is the target family.

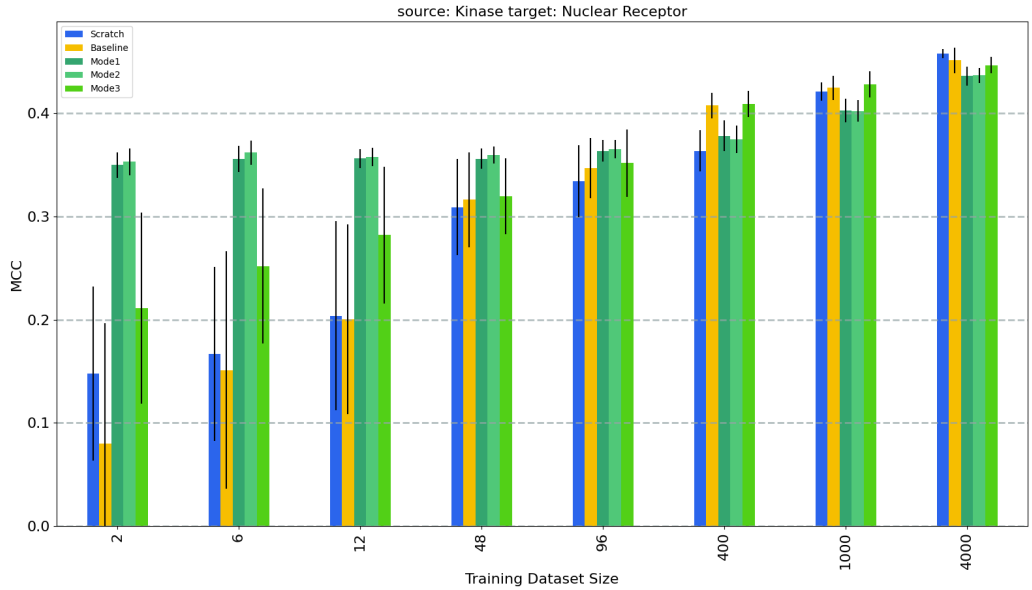


Figure 7. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch), the base model (an SVM trained from scratch) and the three transfer learning modes for all of the eight cases when kinase is the source family and nuclear receptor is the target family.

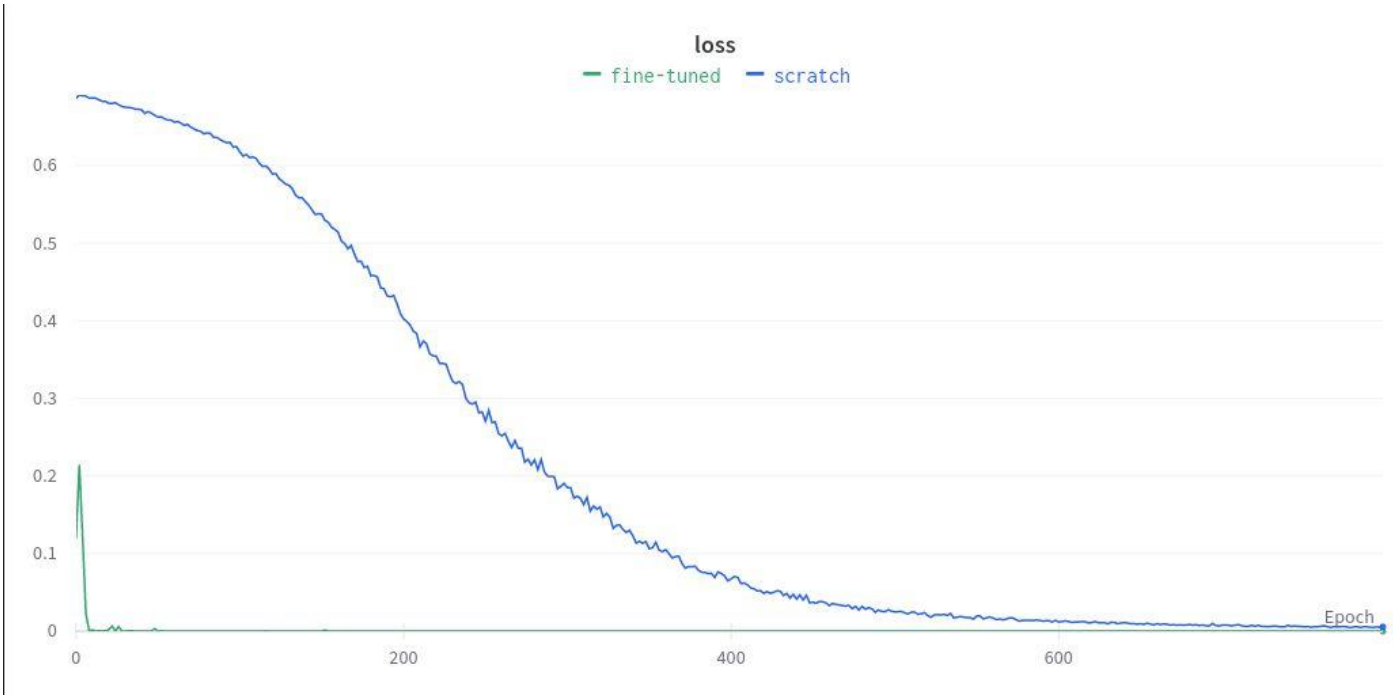
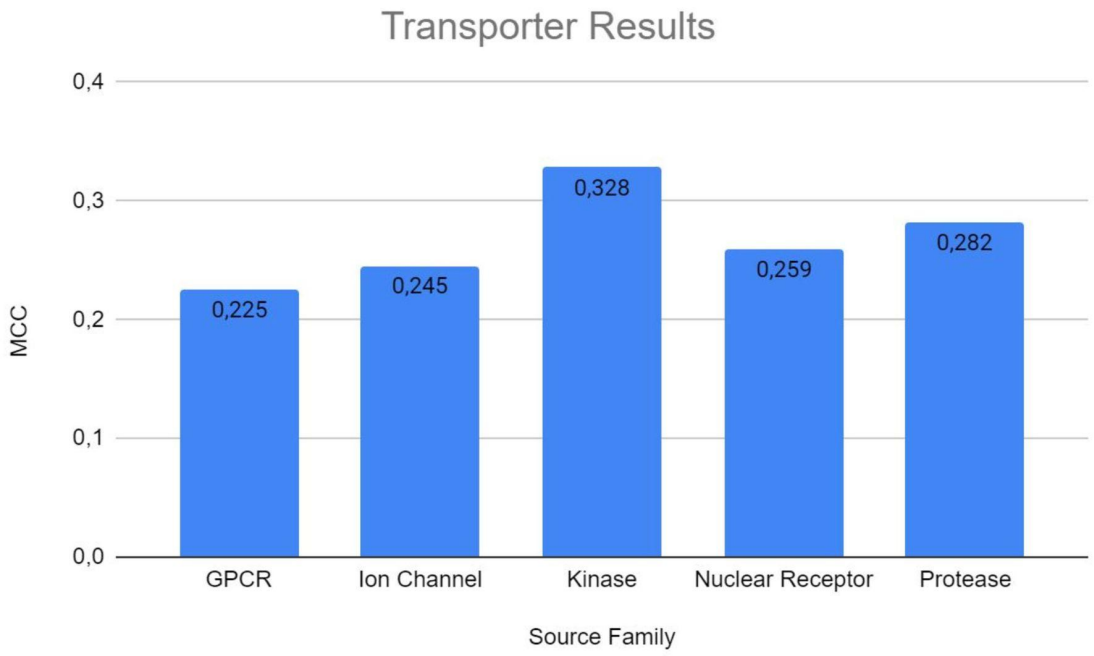
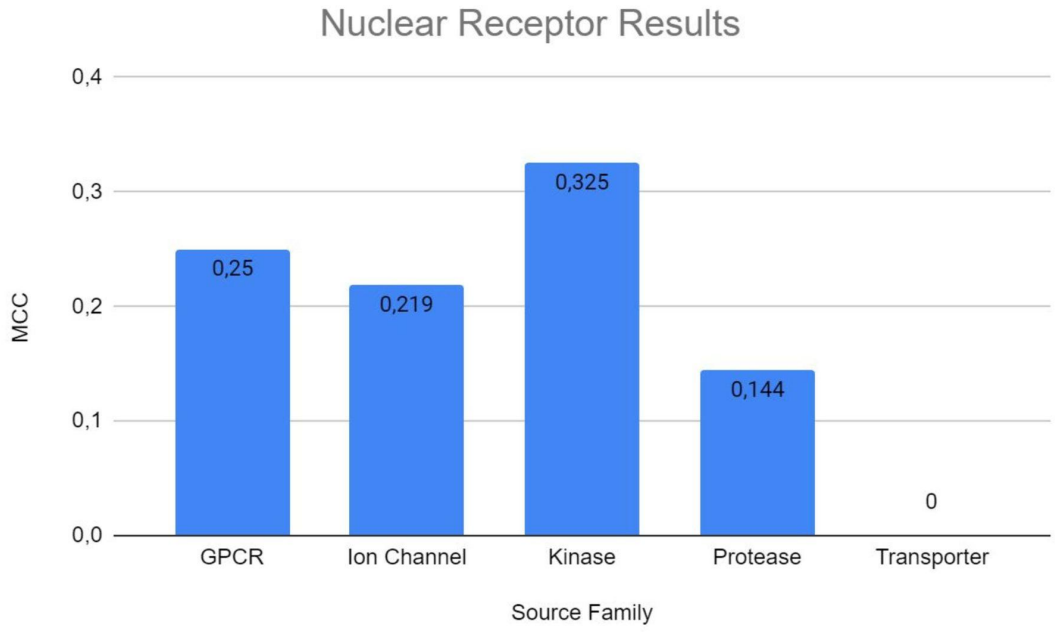


Figure 8. When transfer learning is used, the target model (fine-tuned) starts from lower loss values compared to the reference model (scratch).

In general, when the training dataset size is less than 100, transfer learning has better performance than training from scratch (i.e., compared to the reference model and the base model). When the size of the target training dataset is greater than 100, transfer learning performance is very close to that of training the network model from scratch. Transfer learning should still be preferred since it requires a smaller number of training epochs. In all of the cases for which the training dataset size is less than 100, transfer learning methods performed better than the reference model and base model. Furthermore, when transfer learning is used, target models start from lower loss values when compared to the reference model (see Figure 8). Therefore, a lower number of epochs are generally sufficient for training, significantly reducing training time.We have also evaluated the performance when the source models are directly used for the tests of target data; that is, no transfer learning (no re-training is applied on the pre-trained source model). Figure 9 shows the average test MCC values of the reference model (FNN-2-Chemprop trained from scratch) when a) transporter and b) nuclear receptor are the target families, respectively. Although these results may still be acceptable, it is easily observed that even a very small amount of training via transfer learning boosts performance. For example, when ion channel is the source family and transporter is the target family, 0.333 MCC is obtained if the pre-trained source model is trained with only 2 samples while 0.245 MCC is obtained when there is no training data.



a)



b)

Figure 9. Average test MCC values of the reference model (FNN-2-Chemprop trained from scratch) when a) transporter and b) nuclear receptor are the target families, respectively, when the source models are directly used for the tests of target data; that is, no transfer learning (no further training is applied on the pre-trained source model).

7 DISCUSSION and CONCLUSIONS

We presented a systematic evaluation of deep transfer learning for drug-target interaction prediction when a limited amount of bioactivity data is available. With this approach, learning is still possible even when there is a very low amount of data, as low as two compounds (i.e, one positive interaction data point and one data point corresponding to an inactive compound), are available. Although fine-tuning is the most popular transfer learning technique, we show that the other transfer learning techniques (Feature Transformer and Shallow Classifier) described in this paper deserve attention as well. Furthermore, deep transfer learning is effective even in the general case, even where there is sufficient data to train from scratch; since convergence becomes faster. When the source models are directly used for the tests of target data; that is, no transfer learning (zero-shot learning) is also possible when a pre-trained model is directly used for the predictions of a target protein family other than that used for pre-training. Although the performance is still acceptable, even a very small amount (even two target samples) of training via transfer learning boosts the performance. We intend to use these models as a basis for developing target-specific models. Last but not least, transfer learning is not limited to DTI; the methodology presented here can be applied to other machine learning applications in bioinformatics, such as protein function prediction.

AVAILABILITY

All the source code and data are available on the GitHub page github.com/cansyl/TransferLearning4DTI The website is available at tl4dti.kansil.org.

FUNDING

This work was supported by TUBITAK project number: 121E208.

REFERENCES

Bagherian,M. *et al.* (2020) Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief. Bioinform.*, **22**, 247–269.

Baskin,I.I. (2020) The power of deep learning to ligand-based novel drug discovery. *Expert Opin. Drug Discov.*, **15**, 755–764.

Cai,C. *et al.* (2020) Transfer Learning for Drug Discovery. *J. Med. Chem.*, **63**, 8683–8694.

Chen,H. *et al.* (2018) The rise of deep learning in drug discovery. *Drug Discov. Today*, **23**, 1241–1250.

Dey,V. *et al.* (2022) Improving Compound Activity Classification via Deep Transfer and Representation Learning. *ACS Omega*, **7**, 9465–9483.

Du,B.-X. *et al.* (2022) Compound–protein interaction prediction by deep learning: Databases, descriptors and models. *Drug Discov. Today*, **27**, 1350–1366.

Elbadawi,M. *et al.* (2021) Advanced machine-learning techniques in drug discovery. *Drug Discov. Today*, **26**, 769–777.

Ezzat,A. *et al.* (2018) Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief. Bioinform.*, **20**, 1337–1357.

Jing,Y. *et al.* (2018) Correction to: Deep Learning for Drug Design: an Artificial Intelligence Paradigm for Drug Discovery in the Big Data Era. *AAPS J.*, **20**, 79.

Kao,P.-Y. *et al.* (2021) Toward Drug-Target Interaction Prediction via Ensemble Modeling and Transfer Learning. In, *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. ieeexplore.ieee.org, pp. 2384–2391.

Kim,J. *et al.* (2021) Comprehensive Survey of Recent Drug Discovery Using Deep Learning. *Int. J. Mol. Sci.*, **22**.

Lee,M. *et al.* (2019) Multi-channel PINN: investigating scalable and transferable neural networks for drug discovery. *J. Cheminform.*, **11**, 46.

Li,P. *et al.* (2021) An effective self-supervised framework for learning expressive molecular global representations to drug

discovery. *Brief. Bioinform.*, **22**.
Lo,Y.-C. *et al.* (2018) Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today*, **23**, 1538–1546.
Pan,X. *et al.* (2022) Deep learning for drug repurposing: Methods, databases, and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **12**.
Playe,B. and Stoven,V. (2020) Evaluation of deep and shallow learning methods in chemogenomics for the prediction of drugs specificity. *J. Cheminform.*, **12**, 11.
Réda,C. *et al.* (2020) Machine learning applications in drug development. *Comput. Struct. Biotechnol. J.*, **18**, 241–252.
Rifaioglu,A.S. *et al.* (2020) DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.*, **11**, 2531–2557.
Rifaioglu,A.S. *et al.* (2019) Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief. Bioinform.*, **20**, 1878–1912.
Tan,C. *et al.* (2018) A Survey on Deep Transfer Learning. In, *Artificial Neural Networks and Machine Learning – ICANN 2018*. Springer International Publishing, pp. 270–279.
Vamathevan,J. *et al.* (2019) Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.*, **18**, 463–477.
Wang,C. and Kurgan,L. (2018) Review and comparative assessment of similarity-based methods for prediction of drug–protein interactions in the druggable human proteome. *Brief. Bioinform.*, **20**, 2066–2087.
Yang,K. *et al.* (2019) Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.*, **59**, 3370–3388.
Yosinski,J. *et al.* (2014) How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.*, **27**.
Zhang,Z. *et al.* (2022) Graph neural network approaches for drug-target interactions. *Curr. Opin. Struct. Biol.*, **73**, 102327.