Computers in Biology and Medicine 147 (2022) 105772

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

DTITR: End-to-end drug–target binding affinity prediction with transformers

Nelson R.C. Monteiro [a,*], José L. Oliveira [b], Joel P. Arrais [a]

[a] Univ Coimbra, Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Coimbra, Portugal
[b] IEETA, Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal

ARTICLE INFO

Keywords:
Drug–target interaction
Binding affinity
Deep learning
Transformer
Attention

ABSTRACT

The accurate identification of Drug–Target Interactions (DTIs) remains a critical turning point in drug discovery and understanding of the binding process. Despite recent advances in computational solutions to overcome the challenges of *in vitro* and *in vivo* experiments, most of the proposed *in silico*-based methods still focus on binary classification, overlooking the importance of characterizing DTIs with unbiased binding strength values to properly distinguish primary interactions from those with off-targets. Moreover, several of these methods usually simplify the entire interaction mechanism, neglecting the joint contribution of the individual units of each binding component and the interacting substructures involved, and have yet to focus on more explainable and interpretable architectures. In this study, we propose an end-to-end Transformer-based architecture for predicting drug–target binding affinity (DTA) using 1D raw sequential and structural data to represent the proteins and compounds. This architecture exploits self-attention layers to capture the biological and chemical context of the proteins and compounds, respectively, and cross-attention layers to exchange information and capture the pharmacological context of the DTIs. The results show that the proposed architecture is effective in predicting DTA, achieving superior performance in both correctly predicting the value of interaction strength and being able to correctly discriminate the rank order of binding strength compared to state-of-the-art baselines. The combination of multiple Transformer-Encoders was found to result in robust and discriminative aggregate representations of the proteins and compounds for binding affinity prediction, in which the addition of a Cross-Attention Transformer-Encoder was identified as an important block for improving the discriminative power of these representations. Overall, this research study validates the applicability of an end-to-end Transformer-based architecture in the context of drug discovery, capable of self-providing different levels of potential DTI and prediction understanding due to the nature of the attention blocks. The data and source code used in this study are available at: https://github.com/larngroup/DTITR.

1. Introduction

The therapeutic effects of active compounds are determined through the observation of DTIs, where the role enforced by the drug (pharmacological activity) regulates the target's biological process. Therefore, identifying new molecules with relevant binding activity against targets with biological interest is crucial in the early drug discovery stages, considering that the ability of a drug to bind plays an important role in the execution of its intrinsic activity [1]. However, conducting low or high-throughput bioassays for the screening of potential leads is time-consuming, labor-intensive, and unfeasible for the vast compound and protein space, compromising the effectiveness of these approaches [2].

In recent years, *in silico* DTI prediction has attracted increasing attention and holds broad interest to address several challenges, including target fishing, drug repositioning, and polypharmacology studies. These computational methods through the scanning of large amounts of pharmacogenomic data in shorter periods of time and leveraging of the knowledge available to characterize the proteins and/or compounds have been determinant in the discovery of new drugs, new findings for existing drugs, and improving the overall understanding of the biological, chemical and pharmacological processes involved in the DTIs [3].

In spite of the encouraging results and performances obtained by numerous computational studies proposed to solve the DTI prediction challenge, most of these methodologies rely on shallow binary associations to characterize the interaction and conduct the experiments [4]. On that account, the importance of DTA, which considers all the comprehensive processes involved in the interaction, i.e., reflects the magnitude and rank order of the pair association, is usually overlooked, especially given that predicting DTA is substantially more challenging. Hence, the quality of the predictions is usually compromised or at least limited, particularly in the identification of primary interactions.

* Corresponding author.
E-mail addresses: nelsonrcm@dei.uc.pt (N.R.C. Monteiro), jlo@ua.pt (J.L. Oliveira), jpa@dei.uc.pt (J.P. Arrais).

https://doi.org/10.1016/j.compbiomed.2022.105772
Received 7 April 2022; Received in revised form 7 June 2022; Accepted 19 June 2022
Available online 21 June 2022

The interaction between compounds and proteins results from the recognition and complementarity of certain groups (binding regions) and it is supported by the joint action of other individual substructures scattered across the protein and compound. However, most DTI prediction models simplify the interaction mechanism and do not take simultaneously into consideration the magnitude of certain local regions of each binding component and the interacting substructures involved. Furthermore, several studies neglect that the interactions are substructural and characterize DTIs with global features, limiting the inferring process and introducing noise in the predictions [5].

On account of the progressive advances in computing and the growth of available data to train complex models, deep learning algorithms have been successfully employed in several fields of interest, including critical contexts such as bioinformatics, cheminformatics, and medical image analysis [6]. The higher modular capability of these architectures to estimate nonlinear mapping between data input and output, and discover appropriate representations from structured or unstructured raw data, has led to interesting findings in the DTI domain [7]. However, these methods progressively transform the input in order to increase the selectivity and invariance of the representations, resulting in abstract learned features, which are essentially non-human interpretable. Furthermore, these representations do not provide a tractable path to the input domain, leading to inadequate explanations about the context that is responsible for a specific decision [8].

Based on these reported characteristics and drawbacks, we propose an end-to-end Transformer-based architecture for predicting DTA measured in terms of the dissociation constant (K_d), where 1D sequential and structural data, specifically protein sequences and SMILES (Simplified Molecular Input Line Entry System) strings, are used to represent the targets and compounds, respectively. We employ three Transformer-Encoder blocks, particularly a protein encoder, a compound encoder, and a protein-compound encoder, and concatenate the resulting aggregate representations to feed into a Fully-Connected Feed-Forward Network (FCNN). This architecture, drug–target Interaction TRansformer (DTITR), leverages the use of self-attention layers to learn the short and long-term biological and chemical context dependencies between the sequential and structural units of the proteins and compounds, respectively, and cross-attention layers to exchange information and make the interaction between the proteomic and chemical domains (pharmacological space). Overall, the proposed model's emphasis is not only on the predictive performance, where the results were better than or on a par with state-of-the-art baselines, but also on the self-capability of the architecture to provide three different levels of potential DTI and prediction understanding due to the nature of the attention blocks, which give information about the overall importance of the input components and their associations to the model. Fig. 1 illustrates the proposed computational framework to solve the challenge of predicting DTA based on Transformer-Encoders and 1D raw sequential and structural data to represent the proteins and compounds.

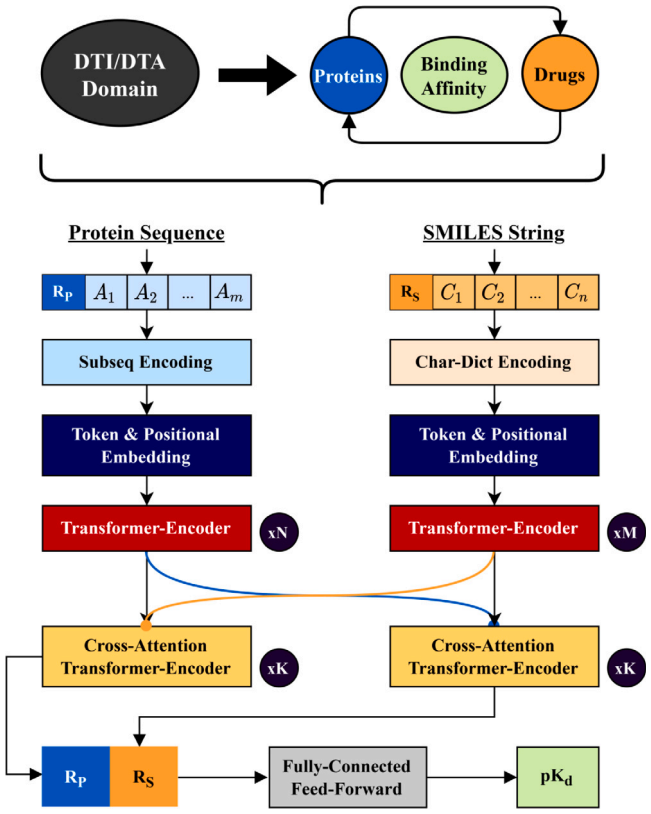

Fig. 1. DTA prediction computational framework based on three Transformer-Encoders, particularly a protein encoder, a compound encoder, and a protein-compound encoder, where 1D raw sequential and structural data is used to represent the proteins and compounds.

2. Related work

Different perspectives and approaches have been proposed over the past years to solve the computational challenge of identifying new DTIs. Structure-based methods, commonly known as docking simulation, simulate and score the interaction according to the intermolecular energy and individual contributions of the receptor and ligand, in which the 3D coordinates of the ligand and receptor are used to predict the coordinates of the resulting complex [9]. Docking methods essentially differ from each other in terms of the different degrees of molecular flexibility considered, the direction of the docking process, and the scoring function employed [10–12]. Despite interesting findings in terms of discovering potential leads, e.g., the work of Gowthaman et al. (2016) [13] identified relevant inhibitors (biochemical assay validated) based on binding pocket topography mapping, most of the results are usually limited or unreliable due to the complexity and number of possible conformations. Following the docking methods, ligand-based approaches have also been explored, which pair multiple (similar) compounds with a specific bioactivity of interest, building prediction models to determine the correlation between chemical structures and biological activity (quantitative structure–activity relationship). Machine learning methods, including Random Forest (RF) or Support Vector Machine (SVM), and deep learning architectures, e.g., Feed-Forward Neural Network (FFN), have been considered as the prediction models, where different descriptors either related to molecular or 3D structural geometric properties have been used to characterize the ligands [14–17]. These methods, however, are heavily dependent on the amount of known and available ligands (or knowledge available about known interactions), performing poorly when this number is scarce.

The abundance of useful biological and chemical data and the growth of available computational power has motivated new predictive solutions in the DTI domain, leading to the chemogenomic approaches, which integrate the genomic, chemical, and/or pharmacological spaces for the inferring process. On that account, some studies have explored similarity-based methodologies, where target and compound associations of similar compounds and targets, respectively, are shared to make new assumptions. For instance, the work of Yamanashi et al. (2008) [18] uses a kernel-driven regression method to infer new interactions based on bipartite graph learning. Peng et al. (2017) [19] proposed a semi-supervised framework, NormMulInf, based on collaborative filtering theory, where similarities among the samples and local correlations among the labels are incorporated into a robust Principal Component Analysis (PCA) model. Additionally, other approaches have explored the use of matrix factorization to decompose the interaction matrix as a product of latent variables that express each drug/target

and determine the missing interactions that are likely to exist [20,21]. Given that similarity-based approaches have shown to present lower performances for some protein classes and that the protein sequence similarity is not always a good indicator due to the conformation complexity, feature-based methods have gained special interest over time. Feature-based DTI studies have been exploring different properties and representations to characterize the proteins, including CTD (Composition, Transition, and Distribution) descriptors and evolutionary profiles, and compounds, e.g., fingerprints, which represent the presence or absence of certain substructures, and molecular descriptors. These features, comprising different attributes of the proteins and compounds, are usually combined to characterize the DTI pair and used as input for machine learning models, e.g., RF and SVM, and also deep learning architectures, including FFN, Deep Belief Neural Networks, or Long Short-Term Memory Neural Networks [22–27]. To improve the performance of feature-based models and overcome the limits of using global descriptors, some studies have been focusing on the use of raw sequential and structural data to characterize the proteins and compounds, specifically amino acid sequences and SMILES strings, respectively, combined with deep learning architectures such as Convolutional Neural Networks (CNNs) [28–31]. Inspired by the remarkable success of the Transformers in different domains, such as natural language processing and image analysis [32–34], the recent study by Huang et al. (2021) [35] proposed a deep learning architecture based on Transformer-Encoders and CNNs, where the Transformers are used to encode and extract an augmented contextual representation from the protein sequences and SMILES strings, and the CNN to model the higher-order interaction.

Despite the interesting results obtained in the field of DTI prediction, the use of binary associations to perform the experiments limits the quality of the results, leading to an increased number of false negatives and a lack of target selectivity. The expansion of certain databases of interactions with known binding affinity or activity metrics, such as ChEMBL [36] or BindingDB [37], has been instrumental in shifting computational drug discovery towards DTA prediction. Given the limitations of some of the original score metrics used in structure-based virtual screening, DTA prediction methods have initially focused on improving and incorporating more information, e.g., additional energetic terms, into these functions. Machine learning methods, including RF, and deep learning architectures, such as FFN, have been proposed as replacements for the scoring functions, predicting the putative strengths of protein–ligand complexes based on different features associated with the 3D structures [38–43]. Additionally, given the remarkable ability of 3D CNNs to capture spatial context, recent studies have explored employing these architectures in combination with 3D single instance learning to predict the binding strength [44–47]. To circumvent the limitations of 3D single instance learning and the confined space of proteins and ligands with known/determined 3D structure, some research studies have pursued more realistic and reproducible methodologies to predict DTA, making use of the abundant chemogenomic data and lower structural information. Apart from algorithms such as RF and SVM [48], Kronecker-Regularized Least Squares [49] or Gradient Boosting Regression Trees [50], several recent studies have been exploring the use of 1D CNNs, 2D CNNs or Graph CNNs in combination with different representations of the proteins and compounds, including 1D structures, 2D similarity matrices, feature vectors or even graph representations [51–57]. The existing DTA prediction methodologies, however, still rely on the use of biased binding affinity metrics, i.e., dependent on the experimental conditions, mechanism of inhibition, and concentrations. Furthermore, the majority of these models, especially those based on deep learning, have yet to consider including interpretability in the inner structure of the architectures or providing potential explainability to the predictions, thus, limiting the results.

Table 1
Original and pre-processed Davis dataset: unique proteins, compounds, and DTIs.

Davis Kinase Dataset					
	Proteins	Compounds	DTI	pKd = 5	pKd > 5
Original	442	72	31824	22400	9424
Pre-Processed	423	69	29187	20479	8708

3. Material and methods

3.1. Binding affinity dataset

We evaluated our proposed model on the Davis et al. (2011) [58] research study dataset, which contains a total of 31 824 interactions between 72 kinase inhibitors (compounds) and 442 kinases (proteins). This dataset covers a large percentage of the human catalytic protein kinome, and the binding strength of the DTI pairs is measured in terms of a quantitative dissociation constant (K_d), which expresses a direct measurement (unbiased) of the equilibrium between the receptor–ligand complex and dissociation components, in which lower values are associated with strong interactions.

The protein sequences of the Davis dataset were extracted from the UniProt [59] database based on the corresponding accession numbers (identifiers). In order to avoid increased noise due to excessive padding, or loss of relevant sequential information potentially related to binding regions, we have selected only proteins with a length between 264 and 1400 residues, which corresponds to 95.7% of the information present in the dataset.

Davis compound SMILES strings were collected in their canonical notation from the PubChem [60] database based on their compound identifiers (CIDs). Even though the canonical notation is unique, where the atoms are consistently numbered, there are some differences in the representation across different data sources. On that account, we have also applied the canonical transformation from the RDKit [61] package in order to guarantee a consistent notation to represent the chemical structure of the compounds and increase the overall reproducibility. Similar to the protein sequences, we have selected only SMILES strings with a length between 38 and 72 chemical characters, which corresponds to 95.8% of the information.

The Davis binding strength distribution ranges from low values (strong interactions) to high values (weak interactions), in which the majority of the DTI pairs are characterized by a binding affinity equal to 10 000 nM. Hence, in order to reduce the effects of the high variance of this distribution on the learning loss, we have applied a normalization strategy (Eq. (1)) to the Kd values, transforming them into the logarithmic space (pK_d). The distribution of the pK_d values ranges from 5 (10 000 nM) to approximately 11.

$$pK_d = -log_{10}(\frac{K_d}{10^9}) \quad (1)$$

Table 1 summarizes the statistics of the original and pre-processed Davis Dataset.

3.2. Input representation

We used an integer-based encoding to represent the structural characters of the SMILES strings, where we scanned the different SMILES in the Davis dataset and extracted 26 categories (unique characters). This 26-character dictionary is used to encode each character with the corresponding integer. SMILES strings shorter than the maximum length threshold of 72 characters were padded. Fig. 2 illustrates the integer-based encoding applied to the SMILES string associated with the Dasatinib compound.

In the case of protein sequences, it is not reasonable to apply the same encoding method of the SMILES strings given the computational complexity of the self-attention layer of $O(n^2)$ with respect to the

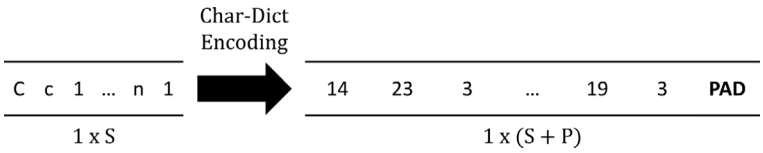

Fig. 2. Integer-based encoding applied to the Dasatinib SMILES string, where each character is encoded into the corresponding integer. S is the length of the SMILES string and P is the number of padding tokens (zeros).

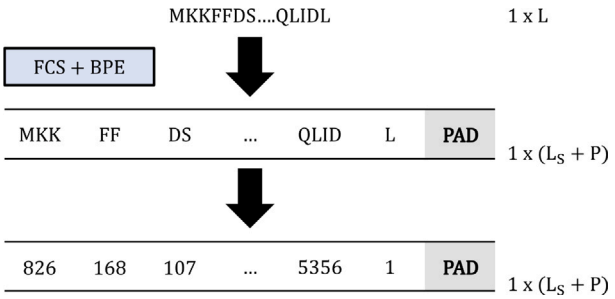

Fig. 3. FCS and BPE encoding applied to the AAK1 kinase amino acid sequence, where the sequence is decomposed into an order of discovered frequent subsequences followed by integer encoding. L is the length of the amino acid sequence, L_S is the length of sequence decomposed into subsequences, and P is the number of padding tokens (zeros).

sequence length. On that account, we used the same approach proposed in the research study by Huang et al. (2021) [35], which combines a Frequent Consecutive Subsequence (FCS) mining method with the Byte Pair Encoding (BPE) algorithm. The FCS method examines large amounts of unlabeled data to discover frequent substructures and create a set of recurring subsequences (subwords). On the other hand, BPE decomposes the sequence into an order of discovered frequent subsequences, where each subsequence must be exclusive and must not overlap, and the aggregation of all subsequences must recover the original sequence. The hierarchy set of frequent subsequences contains a total of 16 693 different subwords, which results in a maximum length of 556 subwords for the protein sequences present in the Davis dataset. Similar to the SMILES strings, protein sequences shorter than this maximum length were padded. Fig. 3 depicts the FCS and BPE encoding approach applied to the AAK1 kinase.

3.3. DTITR framework

The DTITR framework learns to predict the binding strength of DTIs, where 1D sequential and structural information, protein sequences and SMILES strings, respectively, are used as input. This architecture makes use of two parallel Transformer-Encoders [32] to compute a contextual embedding of the protein sequences and SMILES strings. The outputs are then fed into a Cross-Attention Transformer-Encoder block, which comprises cross-attention and self-attention layers, to exchange information and model the interaction space. The resulting aggregate representations, which correspond to the final hidden states of the start tokens added to the protein sequences and SMILES strings, are concatenated and used as input for an FCNN. The final layer, which is composed of a single neuron, outputs the binding affinity measured in terms of pK_d.

3.3.1. Embedding block

The protein sequences and SMILES strings are initially processed based on their length (Section 3.1) and then encoded according to the approaches mentioned in Section 3.2. Similar to the BERT architecture [33], we have added a special token of regression R_P and R_S to the beginning of every protein sequence and SMILES string, respectively. We have assigned an embedding layer to the protein sequences and SMILES strings, which generates a learned embedding to every token with a fixed size of d^P_{model} and d^S_{model}, respectively, via a learnable dictionary matrix. Following the embedding layers, we have also multiplied the embedding values with $\sqrt{d^P_{model}}$ and $\sqrt{d^S_{model}}$ to initially rescale their value.

Considering that the Transformer-Encoder is permutation invariant, it is necessary to add additional information about the relative or absolute position of the tokens in the sequence. We have used the same approach applied in the study by Vaswani et al. (2017) [32] and added a positional encoding based on sine and cosine functions of different frequencies, which outputs a unique encoding for each position. The final embeddings for the ith and jth input tokens of the protein sequence ($E^{P_k}_i$) and SMILES string ($E^{S_k}_j$), respectively, associated with the kth DTI pair are given by the sum of the token embedding and the positional embedding:

$$\begin{aligned} E^{P_k}_i &= E^{P_k}_{token_i} + E^{P_k}_{pos_i} \\ E^{S_k}_j &= E^{S_k}_{token_j} + E^{S_k}_{pos_j}, \end{aligned} \tag{2}$$

where $E^{P_k}_{token_i} \in R^{d^P_{model}}$ and $E^{S_k}_{token_j} \in R^{d^S_{model}}$, and $E^{P_k}_{pos_i} \in R^{d^P_{model}}$ and $E^{S_k}_{pos_j} \in R^{d^S_{model}}$ are the token embeddings and the positional embeddings for the ith and jth inputs tokens of the protein sequence P_k and SMILES string S_k, respectively.

Following the sum of the two types of embedding, we have added a dropout layer.

3.3.2. Transformer-Encoder

In order to capture the biological and chemical context information present in the protein sequences and SMILES strings, respectively, we propose the use of two Transformer-Encoders in parallel. The Transformer-Encoder architecture is composed of a stack of identical blocks, where each block contains a Multi-Head Self-Attention layer (MSA) with an FFN. Residual connections are applied after every block followed by Layer Normalization (LN), and dropout is applied after each MSA layer and after each Dense layer of the FFN. Considering B^1 the output of the first subunit and B^2 the output of the second subunit, the output of the kth block can be expressed as:

$$\begin{aligned} B^1_k &= \mathbf{LN}(B^2_{k-1} + \mathbf{dropout}(\mathbf{MSA}(B^2_{k-1}))) \\ B^2_k &= \mathbf{LN}(B^1_k + \mathbf{FFN}(B^1_k)), \end{aligned} \tag{3}$$

where $B^1_k, B^2_k \in R^{N_P \times d^P_{model}}$ in the case of the protein sequences (N_P is the number of protein subwords), and $B^1_k, B^2_k \in R^{N_S \times d^S_{model}}$ in the case of the SMILES strings (N_S is the number of SMILES characters).

The MSA layer takes its input in the form of three parameters, specifically Query, Key, and Value, which are generated from the same input sequence. This layer applies self-attention, i.e., the input sequence attends to itself multiple times in parallel, where the queries, keys, and values are linearly projected and split across different heads of attention. Each head (self-attention layer) maps a query and a set of key–value pairs to an output, which is computed as a weighted sum of the values. The attention weights assigned to each value are obtained by applying a softmax to the scaled (divided by $\sqrt{d^{P/S}_{model}}$) dot-product between the queries and keys, i.e., each element of the query attends to all elements of the key–value pair. The outputs of each head of attention

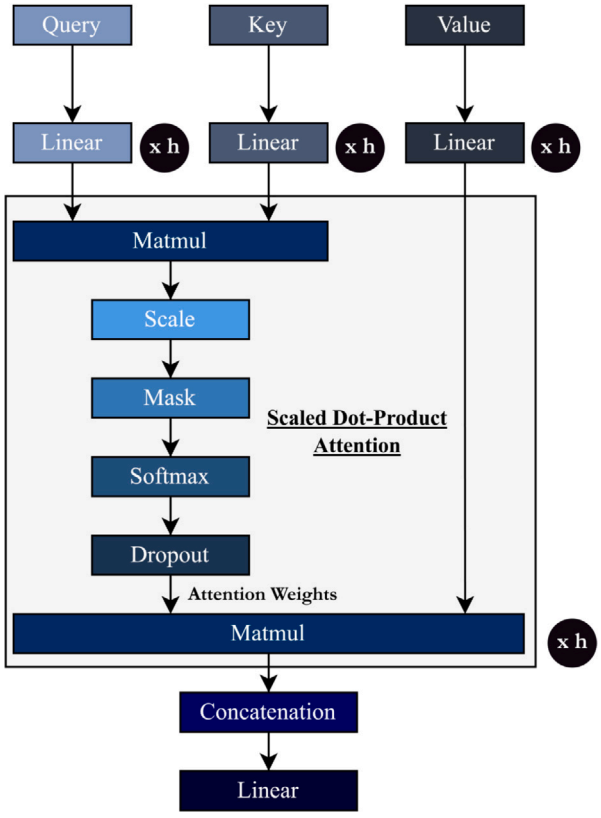

Fig. 4. Multi-head attention architecture, where each head of attention maps a query and set of key–value pairs to an output, which is computed as a weighted sum of the values. h is the number of heads of attention and *mask* corresponds to the masking of the PAD tokens.

are concatenated and linearly projected, where the size of the final output is the same as the query sentence.

$$\begin{aligned}\mathbf{attn}(Q,K,V) &= \mathbf{softmax}(\frac{QK^T}{\sqrt{d_K}})V \\ \mathbf{MSA}(Q,K,V) &= [\mathbf{attn}(QW_1^Q, KW_1^K, VW_1^V); ...; \\ &\quad \mathbf{attn}(QW_h^Q, KW_h^K, VW_h^V)]W^O,\end{aligned} \tag{4}$$

where $Q \in R^{N_{P/S}\ x\ d_{model}^{P/S}}$ is the Query, $K \in R^{N_{P/S}\ x\ d_{model}^{P/S}}$ is the Key, $V \in R^{N_{P/S}\ x\ d_{model}^{P/S}}$ is the Value, $W_{1,...,h}^Q \in R^{d_{model}^{P/S}\ x\ d_Q}$ are the Query projection matrices, $W_{1,...,h}^K \in R^{d_{model}^{P/S}\ x\ d_K}$ are the Key projection matrices, $W_{1,...,h}^V \in R^{d_{model}^{P/S}\ x\ d_V}$ are the Value projection matrices, $W^O \in R^{h\ x\ d_V\ x\ d_{model}^{P/S}}$ is the output projection matrix, h is the number of heads of attention, $[;]$ denotes concatenation, $d_Q = d_K = d_V = \frac{d_{model}^{P/S}}{h}$, and P or S in the case of the protein sequences or the SMILES strings, respectively. Fig. 4 illustrates the architecture of a multi-head attention layer with h heads of attention in parallel.

Regarding the FFN, it is composed of two Dense layers applied to the last dimension (position-wise), where dropout is added after each one of these layers. This block is used to project the attention outputs in order to potentially give them an individually more robust representation. On that account, the first dense layer initially projects the attention outputs to a higher dimension with a certain expansion ratio, and the second dense layer projects it back to the initial last dimension. Thus, this block is usually compared to two 1×1 convolutions layers. Moreover, the FFN improves the learning capacity of the architecture.

Overall, these two stacked Transformer-Encoders in parallel compute a contextual embedding for the protein sequences and SMILES strings, in which the self-attention mechanisms condition the weight given to input elements by learning the short and long-term context dependencies between the individual units.

3.3.3. Cross-attention Transformer-Encoder

Apart from attending individually and learning the context dependencies between the individual units of each element of the DTI pair (Section 3.3.2), it is crucial for the compounds and proteins to attend mutually to each other, i.e., to exchange information, especially when considering that DTIs are primarily substructural, where the complementarity of certain regions is key for the binding process. Hence, we propose a Cross-Attention Transformer-Encoder block to learn the pharmacological context information associated with the interaction space. The Cross-Attention Transformer-Encoder architecture is composed of a stack of two parallel identical blocks, where each block contains a Multi-Head Cross-Attention layer (MCA), an MSA, and an FFN. Similar to the Transformer-Encoder, residual connections are applied after every block followed by LN, and dropout is applied after each MCA and MSA layers and each Dense layer of the FFN.

The two MCA layers are responsible for the exchange of information between the proteins and compounds, and to model the substructural space of the interaction. Instead of employing a full attention approach, i.e., the whole protein and compound attending to the whole compound and protein, respectively, which is computationally expensive and complex, and also redundant since the two attention matrices would have to satisfy the condition $W_{P-S} = W_{S-P}^T$, we use the R_P and R_S tokens for the exchange of context information [62]. These tokens previously learn (Section 3.3.2) the overall biological and chemical context information amongst the individual units of the protein sequence and SMILES string, respectively, and therefore are considered as an aggregate representation. On that account, these can be efficiently used as the attending agents (Query) in a Multi-Head Attention Layer, where each one of these tokens attends to the information present in the corresponding interaction component, i.e., the R_P token attends to the tokens of the SMILES string and the R_S token attends to the tokens of the protein sequence. Hence, these tokens interact and learn the context information present in the corresponding binding component, which further enriches their representation. The MCA layers work similarly to the MSA layer (Eq. (4)), but instead of the input attending to itself, i.e., the Query, Key, and Value being generated from the same input sequence, the Query will correspond to R_P or R_S token, and the Key and Value to the concatenation of the R_P or R_S token with the corresponding interaction component tokens. Considering X_P and X_S the representation of the protein sequence and SMILES string, respectively, the outputs for the two MCA subunits associated with the kth Cross-Attention Transformer-Encoder block (X_P^k and X_S^k) can be expressed as:

$$\begin{aligned} X_P^{k-1} &= [R_P^{k-1} \parallel T_P^{k-1}]\ ,\ X_S^{k-1} = [R_S^{k-1} \parallel T_S^{k-1}] \\ \mathbf{Q_P} &= R_P^{k-1}\ ,\ \mathbf{Q_S} = R_S^{k-1} \\ \mathbf{K_P/V_P} &= [R_P^{k-1} \parallel T_S^{k-1}]\ ,\ \mathbf{K_S/V_S} = [R_S^{k-1} \parallel T_P^{k-1}] \\ R_P^k &= \mathbf{LN}(R_P^{k-1} + \mathbf{dropout}(\mathbf{MCA}(Q_P, K_P, V_P))) \\ R_S^k &= \mathbf{LN}(R_S^{k-1} + \mathbf{dropout}(\mathbf{MCA}(Q_S, K_S, V_S))) \\ X_P^K &= [R_P^k \parallel T_P^{k-1}]\ ,\ X_S^K = [R_S^K \parallel T_S^{k-1}], \end{aligned} \tag{5}$$

where X_P^{k-1}, $X_P^k \in R^{N_P\ x\ d_{model}^P}$; R_P^{k-1}, $R_P^k \in R^{d_{model}^P}$; T_P^{k-1}, $T_P^k \in R^{(N_P-1)\ x\ d_{model}^P}$; X_S^{k-1}, $X_S^k \in R^{N_S\ x\ d_{model}^S}$; R_S^{k-1}, $R_S^k \in R^{d_{model}^S}$; and T_S^{k-1}, $T_S^k \in R^{(N_S-1)\ x\ d_{model}^S}$.

Following each one of these MCA layers, we apply an MSA layer in order to improve the internal connections between the individual units and enhance the representation of each token based on the learnt cross-attention context information. Similar to the Transformer-Encoder, an FFN is added and applied to the output of each MSA layer. Considering B^1 the output of the first subunit, B^2 the output of the second subunit, and B^3 the output of the third subunit, the outputs of

the kth Cross-Attention Transformer-Encoder block can be expressed as:

$$\begin{aligned} B^1_{k_P-1} &\xrightarrow{\text{Eq. (5)}} B^1_{k_P}, B^1_{k_S-1} \xrightarrow{\text{Eq. (5)}} B^1_{k_S} \\ B^2_{k_P} &= \mathbf{LN}(B^1_{k_P} + \mathbf{dropout}(\mathbf{MSA}(B^1_{k_P}))) \\ B^2_{k_S} &= \mathbf{LN}(B^1_{k_S} + \mathbf{dropout}(\mathbf{MSA}(B^1_{k_S}))) \\ B^3_{k_P} &= \mathbf{LN}(B^2_{k_P} + \mathbf{FFN}(B^2_{k_P})) \\ B^3_{k_S} &= \mathbf{LN}(B^2_{k_S} + \mathbf{FFN}(B^2_{k_S})), \end{aligned} \tag{6}$$

where $B^1_{k_P-1}$, $B^1_{k_P}$, $B^2_{k_P}$, $B^3_{k_P} \in R^{N_P \times d^P_{model}}$, and $B^1_{k_S-1}$, $B^1_{k_S}$, $B^2_{k_S}$, $B^3_{k_S} \in R^{N_S \times d^S_{model}}$.

3.3.4. Fully-connected feed-forward

The final hidden states of the aggregated representations, which correspond to the start tokens R_P and R_S added to the protein sequences and SMILES strings, respectively, are concatenated and used as input for an FCNN, which is essentially a Multilayer Perceptron (MLP). After each Dense layer of this block, we have added a dropout layer. Following the FCNN, a Dense layer with a single neuron is applied to predict the binding affinity of the DTI pair measured in terms of the logarithmic-transformed dissociation constant (pK_d).

Fig. 5 illustrates the proposed DTITR architecture.

3.4. Hyperparameter optimization approach

The most common approach to determine the model's best architecture and set of parameters is grid search with cross-validation, in which the dataset is split across different folds under different conditions depending on the methodology used, e.g., stratified K-fold splits the dataset into different folds taking into consideration the distribution of the classes. However, in the context of the problem, traditional cross-validation approaches are usually not satisfactory or representative, especially when considering that the Davis dataset is extremely imbalanced towards the pK_d values distribution and that 1D raw sequential and structural data is used to characterize the proteins and compounds. On that account, the DTI representability of each fold is determinant in the learning process of the architecture.

We used the Chemogenomic Representative K-Fold method to split the dataset into representative folds and determine the hyperparameters. This method takes into consideration the pK_d values distribution, the protein sequences similarity, and the SMILES strings similarity during the splitting process. It initially distributes the DTI pairs with a pK_d value greater than 5 (relevant interactions) across the different K folds based on the lowest similarity score. This metric corresponds to the weighted mean between the median value across all the protein sequences similarity scores and the median value across all the SMILES strings similarity scores, which are calculated between the sample and each entry in the corresponding set. Additionally, this method also guarantees that every set is equally sized, thus, only sets that had not previously been assigned a sample are considered at each step (until it is reset). Following the pairs with a pK_d value greater than 5, this process is repeated for the DTIs with a pK_d value equal to 5 (weak interactions).

Considering the improved representability of each fold obtained by this splitting methodology, it is also possible to extract an independent testing set in order to estimate the model's performance in the context and chemogenomic domain of the problem and evaluate the generalization capacity.

Fig. 6 illustrates the Chemogenomic Representative K-Fold method.

4. Experimental setup

The hyperparameters for the DTITR architecture were determined by the chemogenomic K-fold cross-validation method (Section 3.4). The protein sequences similarity matrix was obtained using the Smith–Waterman local alignment algorithm, which was implemented using the Biostrings R Package [63]. The substitution matrix selected was the BLOSUM62, and the gap penalty for opening and extension was fixed at 10 and 0.5, respectively. The final alignment scores were normalized to a [0,1] range [18]:

$$SW_{Normalized}(p_1, p_2) = \frac{SW(p_1, p_2)}{\sqrt{SW(p_1, p_1)} * \sqrt{SW(p_2, p_2)}}, \tag{7}$$

where p_1 and p_2 are the two proteins of a certain pair (p_1, p_2). On the other hand, the SMILES similarity matrix was obtained by computing the Tanimoto Coefficient (Eq. (8)), where the SMILES strings were initially converted to the Morgan circular fingerprints with a radius of 3 using the RDKit Python package [61].

$$d(i, j) = \frac{i \cdot j}{|i|^2 + |j|^2 - i \cdot j}, \tag{8}$$

where i and j are the vector (fingerprint) representations of two different compounds, respectively.

The dataset was split into six different folds, in which one of the folds was selected to evaluate the generalization capacity of the model (independent test set) and the remaining folds to determine the hyperparameters of the architecture. We have hyperoptimized several parameters: number of protein transformer-encoders, number of SMILES transformer-encoders, number of cross-attention transformer-encoder blocks, number of heads for the self-attention and cross-attention layers, embedding dimension for the protein sequences and the SMILES strings, FFN hidden neurons, FCNN number of layers, FCNN hidden neurons, dropout rate, optimizer learning rate, and optimizer weight decay. We initially considered a wide range of values for each hyperparameter and then narrowed the search range around the best performing parameter values.

The Gaussian Error Linear Unit (GELU) [64] was selected as the activation function for every layer, with the exception of the final output dense layer which uses a linear activation. The GELU function weights its input by their value rather than gating the input depending upon its sign, thus, it can be seen as a smoother ReLU. Moreover, this activation function avoids the *dead neurons* problem and is able to more easily approximate complicated functions due to the increased curvature and non-monotonicity.

$$\begin{aligned} GELU(x) &= xP(X \le x) = x\Phi(x) \\ &\approx 0.5x(1 + tanh[\sqrt{2/\pi}(x + 0.044715x^3)]), \end{aligned} \tag{9}$$

where $\Phi(x)$ is the cumulative distribution function for the standard normal distribution (Gaussian) and $P(X) \sim N(0, 1)$.

Considering that the context of the problem focuses on a regression task, the loss function selected was the mean squared error (MSE), which measures the average squared difference between the predicted values and the real values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{10}$$

where n is the number of samples, y_i the real value and $\hat{y}_i$ the predicted value.

Regarding the optimizer function, Rectified Adaptive Moment Estimation (RAdam) [65] was used to update the network weights in each iteration of the training process. This function is an improved version of the Adam optimizer and it dynamically adjusts the adaptive learning rate based on the underlying divergence of the variance. Thus, it avoids the need to use a warmup heuristic, which is usually required for

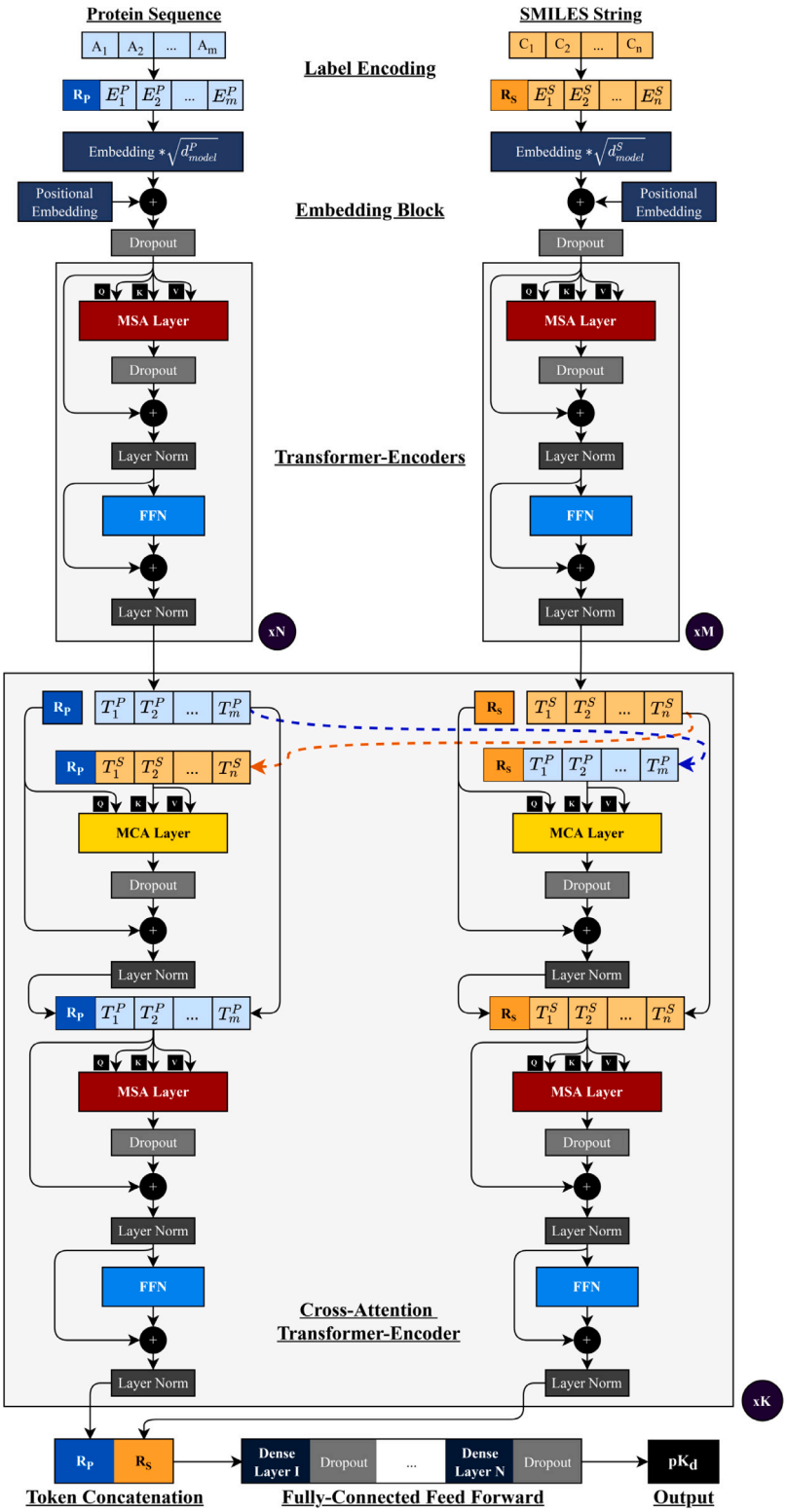

Fig. 5. DTITR: End-to-End Transformer-based architecture. Two parallel Transformer-Encoders compute a contextual embedding of the protein sequences and SMILES strings, and a Cross-Attention Transformer-Encoder models the interaction space and learns the pharmacological context of the interaction. The resulting aggregate representations of the proteins (R_P) and compounds (R_S) are concatenated and used as input for an FFN. The final dense layer outputs the binding affinity measured in terms of pK_d.

adaptive learning rate optimizers due to the excessive variance of the initial training steps.

In order to avoid potential overfitting, two callbacks were considered during the training process, specifically early stopping with a patience of 30 and model checkpoint. The hyperparameter combination that provided the best average MSE score over the validation sets was selected to establish the optimized model and evaluate the generalization capacity on the independent test set. Table 2 summarizes the parameters settings for the DTITR architecture.

In order to validate and assess the prediction efficiency of the proposed DTITR architecture, we have evaluated and compared the performance with different state-of-the-art binding affinity regression

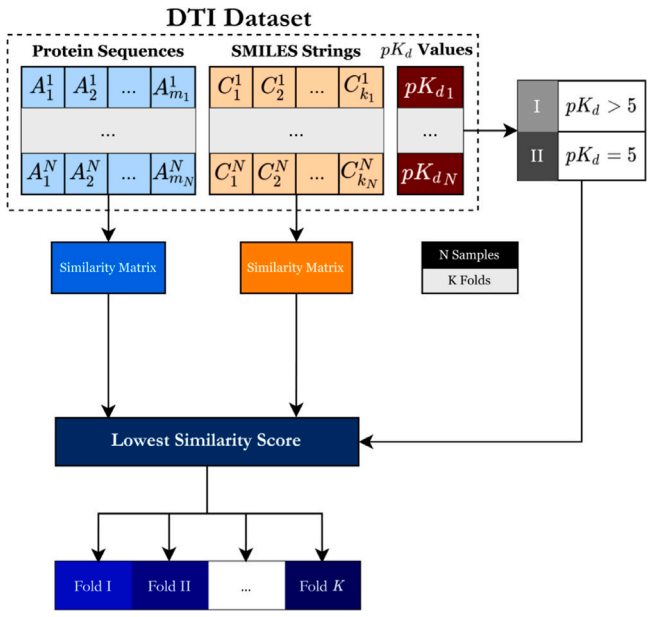

Fig. 6. Chemogenomic representative K-fold, where DTI pairs are distributed based on the pK_d value, protein sequence similarity, and SMILES string similarity. The DTI pairs with a $pK_d > 5$ are initially assigned to the K set with the lowest similarity score followed by the DTI pairs with a $pK_d = 5$. The similarity score corresponds to the weighted mean between the median value across all the protein sequences similarity scores and the median value across all the SMILES strings similarity scores, which are computed between the sample and each entry in the corresponding set.

Table 2
DTITR architecture parameter settings.

Parameter	Value
Protein Transformer-Encoders	3
SMILES Transformer-Encoders	3
Cross-Attention Transformer-Encoders	1
Protein Self-Attention Heads	4
SMILES Self-Attention Heads	4
Cross-Attention Heads	4
Protein Embedding Dim	128
SMILES Embedding Dim	128
Protein FFN Hidden Neurons	512
SMILES FFN Hidden Neurons	512
Activation Function	GELU
Activation Function (Output)	Linear
Dropout Rate	0.1
FCNN Dense Layers	3
FCNN Hidden Neurons	[512,512,512]
Loss Function	Mean Squared Error
Optimizer Function	RAdam
Optimizer Learning Rate	1e−04
Optimizer Beta 1	0.9
Optimizer Beta 2	0.999
Optimizer Epsilon	1e−08
Optimizer Weight Decay	1e−05
Batch Size	32
Epochs[a]	500

[a]Initial number of epochs to allow convergence of the model, where early stopping and model checkpoint were applied to avoid overfitting.

baselines: KronRLS [49], SimBoost [50], Sim-CNN-DTA [55], DeepDTA [51], DeepCDA [54], and all the different formulations of the GraphDTA [53]. The same folds obtained from the chemogenomic K-fold cross-validation methodology were considered to train these models and the testing fold to evaluate their performance. Additionally, we have applied the same encoding approach to the protein sequences (Section 3.2) in the research work where the proteins are represented by their 1D amino acid sequence in order to ensure fairness in the comparisons.

Apart from evaluating the prediction efficiency of the proposed architecture, different alternatives for the DTITR model were also explored, where we have evaluated the efficacy of the Cross-Attention Transformer-Encoder block (Section 3.3.3) by applying and training the model with and without this module, the differences in the prediction efficiency of the architecture by employing the FCS and BPE encoding approach (Section 3.2) to the SMILES strings instead of the character-dictionary integer-based method, and the increasing learning capacity of the model due to the FCNN block (Section 3.3.4) by applying and training the model with and without this module.

We used Python 3.9.6 and Tensorflow 2.6.0 to develop the model and the experiments were run on AMD Ryzen 9 3900X and GeForce RTX 3070 8 GB.

4.1. Evaluation metrics

There are many metrics used to evaluate the performance and capacity of the models as predictors. However, the choice of which ones to use greatly depends on the context of the problem. Therefore, in order to evaluate the performance of the proposed architecture (DTITR), which outputs a continuous value, we have chosen the MSE, root mean squared error (RMSE), concordance index (CI), coefficient of determination (r^2) and Spearman rank correlation (ρ).

- **RMSE**: measures the square root of the average squared difference between the predicted values and the real values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{11}$$

where n is the number of samples, y_i the real value, and $\hat{y}_i$ the predicted value.

- **CI**: measures the probability of non-equal pairs being correctly predicted in terms of order.

$$CI = \frac{1}{Z}\sum_{y_i > y_j} h(p_i - p_j),\ h(p) = \begin{cases} 1, p > 0 \\ 0.5, p = 0 \\ 0, p < 0 \end{cases}, \tag{12}$$

where Z corresponds to the number of non-equal pairs, p_i to the predicted value for the larger affinity y_i, and p_j to the predicted value for the smaller affinity y_j.

- r^2: measures the ratio between the total variance explained by the model and the total variance.

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{13}$$

where $\hat{y}_i$ is the predicted value, y_i the real value, and $\bar{y}$ the mean of the real values.

- **Spearman**: measures the strength and direction of association between two ranked variables (non-parametric).

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}(R(y_i) - \overline{R(y)}) \cdot (R(\hat{y}_i) - \overline{R(\hat{y})})}{\sqrt{(\frac{1}{n}\sum_{i=1}^{n}(R(y_i) - \overline{R(y)})^2) \cdot (\frac{1}{n}\sum_{i=1}^{n}(R(\hat{y}_i) - \overline{R(\hat{y})})^2)}}, \tag{14}$$

where $R(\hat{y}_i)$ is the predicted value rank, $R(y_i)$ the real value rank, $\overline{R(\hat{y})}$ the mean of the predicted values ranks, and $\overline{R(y)}$ the mean of the real values ranks.

5. Results and discussion

5.1. Predictive performance evaluation

In the context of drug discovery and drug repositioning, it is crucial to accurately predict the binding strength of DTI pairs to properly identify and distinguish main interactions from those with secondary

Table 3
Binding affinity prediction results over the Davis independent testing set.

Method	Protein Rep.	Compound Rep.	↓ MSE	↓ RMSE	↑ CI	↑ r^2	↑ Spearman
Baseline Methods							
KronRLS [49]	Smith–Waterman	PubChem-Sim	0.443	0.665	0.847	0.473	0.624
GraphDTA-GCNNet [53]	1D-Subseq	Graph	0.311	0.558	0.883	0.630	0.681
GraphDTA-GATNet [53]	1D-Subseq	Graph	0.286	0.535	0.881	0.660	0.688
SimBoost [50]	Smith–Waterman	PubChem-Sim	0.277	0.526	0.891	0.670	0.694
GraphDTA-GAT-GCN [53]	1D-Subseq	Graph	0.269	0.518	0.874	0.680	0.670
Sim-CNN-DTA [55]	Smith–Waterman	PubChem-Sim	0.266	0.516	0.884	0.683	0.674
GraphDTA-GINConvNet [53]	1D-Subseq	Graph	0.238	0.488	0.899	0.717	**0.741**
DeepDTA [54]	1D-Subseq	1D	0.215	0.464	0.891	0.743	0.691
DeepCDA [54]	1D-Subseq	1D	0.208	0.457	0.895	0.752	0.689
Proposed Method							
DTITR	1D-Subseq	1D	**0.192**	**0.438**	**0.907**	**0.771**	0.712

Table 4
Binding affinity prediction results over the Davis independent testing set for the different alternatives of the DTITR model.

Method	Protein Rep.	Compound Rep.	↓ MSE	↓ RMSE	↑ CI	↑ r^2	↑ Spearman
DTITR - Without FCNN Block	1D-Subseq	1D	0.232	0.481	0.906	0.724	0.712
DTITR - Both Subseq	1D-Subseq	1D-Subseq	0.205	0.453	0.905	0.756	0.712
DTITR - Without Cross-Block	1D-Subseq	1D	0.196	0.443	0.899	0.766	0.703
DTITR	1D-Subseq	1D	**0.192**	**0.438**	**0.907**	**0.771**	**0.712**

targets (off-targets). In order to validate the performance of the proposed DTITR architecture, we have evaluated and compared the prediction efficiency with different state-of-the-art binding affinity regression models. Table 3 reports the binding affinity prediction results over the Davis independent testing set in terms of five different metrics: MSE, RMSE, CI, r^2 and Spearman rank correlation.

The proposed DTITR architecture achieved superior performance across almost all metrics, specifically MSE (0.192), RMSE (0.438), CI (0.907), and r^2 (0.771), when compared to the state-of-the-art baselines. The lower MSE and RMSE scores demonstrate the capacity of the model to correctly predict the binding strength values, and the higher CI score indicates the ability of the architecture to correctly distinguish the binding strength rank order across DTI pairs, which is not only crucial in the drug discovery context to differentiate primary from secondary or weak interactions, but also of special interest given the imbalance nature of the pK_d values distribution of the Davis dataset.

Contrarily to the majority of the baseline methods, where either only individual representations of the proteins and the compounds are being learnt by the model or only the mutual interaction space is being considered during the inferring process, the DTITR architecture takes simultaneously into consideration the magnitude of certain local regions of each binding component (and their intra-associations) and the involving interaction substructures, resulting in robust representations of the protein sequences and SMILES strings. On that account, the results demonstrate that the DTITR model is properly learning the biological, chemical, and pharmacological context information of the proteins, compounds, and protein-compounds interactions, respectively, considering that the final aggregate representations are robust and discriminative for the prediction of binding affinity.

Fig. 7 illustrates the predictions from the DTITR model against the actual (true) binding affinity values for the Davis testing set, where it is possible to observe a significant density around the *predicted = true value* reference line (perfect model).

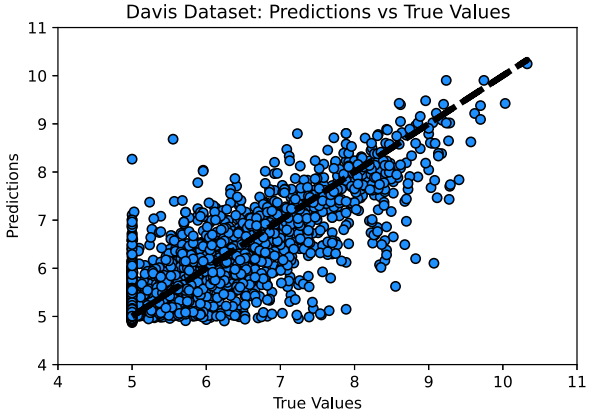

Fig. 7. DTITR predictions against the true values for the Davis testing set, where the diagonal line is the reference line (*predicted = true value*).

5.2. Ablation study

In order to further validate the DTITR architecture, we have explored three different alternatives for the DTITR model, specifically (i) DTITR architecture without the Cross-Attention Transformer-Encoder block, (ii) DTITR architecture without the FCNN block, and (iii) FCS and BPE encoding applied to the SMILES strings instead of the integer-based character-dictionary method. Table 4 reports the binding affinity prediction results over the Davis independent testing set in terms of the five different metrics for the different alternatives of the DTITR model.

To properly assess the efficacy of the Cross-Attention Transformer-Encoder block, which is responsible for the exchange of context information between proteins and compounds (pharmacological space), we have evaluated the model prediction efficiency with and without this module. The DTITR architecture with the Cross-Attention Transformer-Encoder block resulted in overall better performance in terms of the MSE (0.192), RMSE (0.438), CI (0.907), r^2 (0.771) and Spearman (0.712) scores when compared to the DTITR architecture without the Cross-Block (MSE - 0.196, RMSE - 0.443, CI - 0.899, r^2 - 0.766 and Spearman - 0.703). These results demonstrate that using the Cross-Attention Transformer-Encoder block to learn the pharmacological context information associated with the interaction space improves the discriminative power of the final aggregated representation hidden states for the prediction of binding affinity. Moreover, it indicates that the use of only the individual biological and chemical contextual information of the protein sequences and SMILES strings, respectively, leads to worse performance when compared to combining the biological, chemical, and pharmacological contexts, which is in agreement with the fact that DTIs result from the recognition and complementarity of certain substructures (pharmacological space) but are supported by the joint action of other individual substructures scattered across the proteins (biological space) and compounds (chemical space).

Regarding the prediction efficiency of the DTITR model without the FCNN block, the performance obtained over the independent testing set

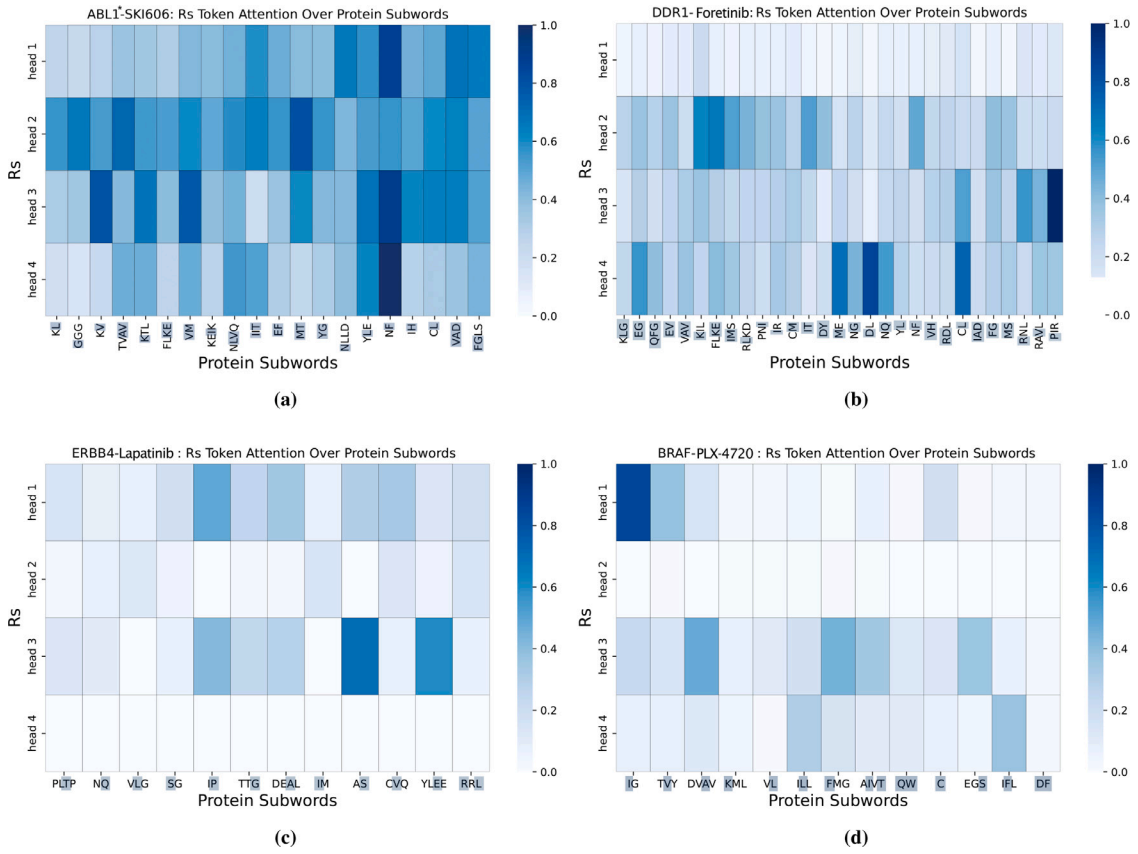

Fig. 8. Attention maps for the attention of the R_S token over the protein substructures, where the interacting residues within the protein subwords are highlighted in gray. (a) ABL1(E255K)-phosphorylated - SKI-606; (b) DDR1 - Foretinib; (c) ERBB4 - Lapatinib; (d) BRAF - PLX-4720.

is worse in terms of the MSE (0.231), RMSE (0.481), and r^2 (0.724) scores when compared to the DTITR architecture with this block (MSE - 0.192, RMSE - 0.438, and r^2 - 0.771). These results demonstrate that the use of the FCNN increases the learning capacity of the architecture and aids in the generalization from the concatenated aggregated representations space, which describes the DTI, to the output space.

Additionally, we have evaluated the differences in the prediction performance of the model by applying the same encoding approach of the protein sequences to the SMILEs strings instead of using the character-dictionary encoding method mentioned in Section 3.2. The performance achieved is substantially worse (MSE - 0.205, RMSE - 0.453 and r^2 - 0.756), except for the CI (0.905) and Spearman (0.712) scores, when compared to using the proposed integer-based encoding method. These results suggest that employing the FCS and BPE algorithms to represent the SMILES strings reduces the learning capacity of the DTITR model, which might be a consequence of the restrictive representation of the SMILES strings since this encoding method results in a maximum length of 15 for the SMILES strings in the Davis dataset.

Overall, the use of an end-to-end Transformer-based architecture for predicting binding affinity demonstrates the ability to use Transformer-Encoders to learn robust and discriminative aggregated representations of the protein sequences and SMILES strings. Moreover, it shows the capacity of the self-attention layers to learn the context dependencies between the sequential and structural units of the proteins and compounds, respectively, and the cross-attention layers to exchange information and model the interaction space.

5.3. Attention maps

DTIs are primarily substructural, where the recognition and complementarity of certain substructures are crucial for the interaction, but the support of the joint action of other individual substructures scattered across the protein and compound also plays a key role in the overall binding process. On that account, visualizing the overall importance of the input components and their associations to the model may potentially lead not only to understanding the model prediction but also to significant findings in the DTI domain. The DTITR architecture contains three different levels of attention: (i) self-attention over the individual units of the protein sequences and SMILES strings; (ii) cross-attention between the protein sequences and SMILES strings; and (iii) self-attention over the individual units of the protein sequences and SMILES strings after the cross-attention (interaction). The first level of attention provides information about the overall importance of the individual units (substructures) and intra-associations of protein sequences and SMILES strings prior to the interaction, i.e., the individual importance of the biological space and chemical space. On the other hand, the second level provides clues about which protein and compound substructures lead to the interaction, in particular, which compound substructures the protein attends to and vice versa. The third level of attention provides information about how the individual biological and chemical importance shifts after the interaction, i.e., how the pharmacological information affects the overall importance of the individual units and intra-associations of protein sequences and SMILES strings.

In order to visualize the attention levels, we have generated heat maps for the second level of attention, specifically for the attention of the R_S token over the protein substructures (subwords). We have selected four different DTI pairs, particularly ABL1(E255K)-phosphorylated - SKI-606, DDR1 - Foretinib, ERBB4 - Lapatinib, and BRAF - PLX-4720, where only subwords associated with interaction residues were considered for visualization. In the case of the ERBB4 - Lapatinib and BRAF - PLX-4720 DTI pairs, the binding positions were collected from the sc-PDB [66] database, which is a specialized structure database focused on ligand binding site in ligandable proteins,

i.e., contains some experimental 3D interaction complexes with the binding regions known/available. On the other hand, the BL1(E255K)-phosphorylated - SKI-606 and DDR1 - Foretinib DTI pairs do not have experimental 3D interaction complexes available/known, thus, we have explored the 3D interaction space using docking approaches. On that account, we have selected potential binding positions (≤ 5 Å) from the resulting 3D receptor–ligand complexes, which were obtained by using guided docking (AutoDock Vina [67]) based on the highest scoring binding pocket from the DoGSiteScorer [68] platform. Fig. 8 illustrates the attention heat maps for ABL1(E255K)-phosphorylated - SKI-606, DDR1 - Foretinib, ERBB4 - Lapatinib, and BRAF - PLX-4720, where the attention weights were normalized across all the positions for each head of attention.

These visual results show that the R_S token, which is an aggregate representation of the compound, is attending, i.e., giving weight, to substructures of the protein sequences associated with binding residues. For each one of these DTI pairs, there are binding-related substructures with a high percentage of significance (weight) in almost every head of attention, e.g., head 4 - motif NF, head 3 - motif PIR, head 3 - motif AS, and head 1 - motif IG for the ABL1(E255K)-phosphorylated - SKI-606, DDR1 - Foretinib, ERBB4 - Lapatinib and BRAF - PLX-4720 interaction pairs, respectively. Moreover, in the particular case of the ABL1(E255K)-phosphorylated - SKI-606 interaction pair, all heads of attention highly attend to almost every substructure. Overall, these findings demonstrate that the Cross-Attention Transformer-Encoder block is learning the pharmacological context of the DTIs, indicating that the DTITR architecture is capable of providing reasonable evidence for understanding the model prediction and potentially leading to new knowledge about DTIs.

6. Conclusion

In this research study, we propose an end-to-end Transformer-based architecture (DTITR) for predicting the logarithmic-transformed quantitative dissociation constant (pK_d) of DTI pairs, where self-attention layers are exploited to learn the short and long-term biological and chemical context dependencies between the sequential and structural units of the protein sequences and compound SMILES strings, respectively, and cross-attention layers to exchange information and learn the pharmacological context associated with the interaction space. The architecture makes use of two parallel Transformer-Encoders to compute a contextual embedding of the protein sequences and SMILES strings, and a Cross-Attention Transformer-Encoder block to model the interaction, where the resulting aggregate representations are concatenated and used as input for an FCNN. We perform our experiments on the Davis kinase binding affinity dataset and compare the performance of the proposed model with different state-of-the-art binding affinity regression baselines.

The proposed model yielded better results than state-of-the-art baselines. It obtained lower MSE and RMSE values and a higher CI score, demonstrating the model's ability to correctly predict the value of the binding strength and correctly distinguish the rank order of binding strength between the DTI pairs, respectively. In addition, the DTITR architecture is shown to efficiently learn the biological, chemical, and pharmacological context of the proteins, compounds, and protein-compound interactions, respectively, given the robustness and discriminative power of the resulting aggregate representations of the protein sequences and SMILES strings.

We have examined various formulations of the DTITR architecture. It was found that the Cross-Attention Transformer-Encoder, which is responsible for the exchange of information between the protein sequences and SMILES strings and learning the pharmacological context, leads to better performance than when only the two initial parallel Transformer-Encoders are used. These results show that combining the biological, chemical, and pharmacological contexts improves the robustness and discriminative power of the aggregate representations compared to using only the individual biological and chemical context information of the protein sequences and SMILES strings. In addition, the FCNN block was found to improve the learning capacity of the architecture as it can improve the generalization from the concatenated aggregate representations space to the output space.

Considering the nature of the attention layers, which give information about the overall importance of the input components and their associations to the model, the DTITR architecture provides three different levels of potential DTI and prediction understanding. We have visualized the attention maps for the second level of attention (Cross-Attention Transformer-Encoder block), specifically for the attention of the aggregate representation of the compounds over the protein sequences substructures. The results show that the compounds are attending to subwords of the protein sequences associated with binding residues, confirming the ability of this block to properly learn the pharmacological context of the DTIs. It also demonstrates that the DTITR architecture is capable of providing reasonable model understanding and potentially leading to new insights in the DTI field.

The major contribution of this study is an efficient and novel end-to-end Transformer-based deep learning architecture for predicting binding affinity that simultaneously considers the magnitude of certain local regions of each binding component (biological and chemical context) and the interacting substructures involved (pharmacological context). Moreover, this architecture provides three different levels of potential DTI and prediction understanding, which is critical in the context of drug discovery.

Deep learning-based architectures perform significantly better when the dataset becomes larger. Therefore, in future work, we will focus on building a larger and more valid DTI dataset measured in terms of the K_d constant, which is one of few unbiased binding affinity/activity metrics. Furthermore, we will focus on extending this work to include additional information associated with the binding sites during the training phase, which is still a major challenge when trying to realistically modulate DTIs and understand the interaction process.

Funding

This research was funded by the Portuguese Research Agency Fundação para Ciência e Tecnologia (FCT) through the Ph.D. scholarship 2020.04741.BD, and by FCT under the project D4 - Deep Drug Discovery and Deployment (CENTRO-01-0145-FEDER-029266).

CRediT authorship contribution statement

Nelson R.C. Monteiro: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft. **José L. Oliveira:** Conceptualization, Writing – review & editing, Project administration. **Joel P. Arrais:** Conceptualization, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.P. Hughes, S. Rees, S.B. Kalindjian, K.L. Philpott, Principles of early drug discovery, Br. J. Pharmacol. 162 (6) (2011) 1239–1249, http://dx.doi.org/10.1111/j.1476-5381.2010.01127.x.

[2] S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, C.C. Persinger, B.H. Munos, S.R. Lindborg, A.L. Schacht, How to improve R & D productivity: the pharmaceutical industry's grand challenge, Nat. Rev. Drug Discov. 9 (3) (2010) 203–214, http://dx.doi.org/10.1038/nrd3078.

[3] P. Schneider, W.P. Walters, A.T. Plowright, N. Sieroka, J. Listgarten, R.A. Goodnow, J. Fisher, J.M. Jansen, J.S. Duca, T.S. Rush, M. Zentgraf, J.E. Hill, E. Krutoholow, M. Kohler, J. Blaney, K. Funatsu, C. Luebkemann, G. Schneider, Rethinking drug design in the artificial intelligence era, Nat. Rev. Drug Discov. 19 (5) (2020) 353–364, http://dx.doi.org/10.1038/s41573-019-0050-3.

[4] S. D'Souza, K.V. Prema, S. Balaji, Machine learning models for drug–target interactions: current knowledge and future directions, Drug Discov. Today 25 (4) (2020) 748–756, http://dx.doi.org/10.1016/j.drudis.2020.03.003.

[5] F.E. Agamah, G.K. Mazandu, R. Hassan, C.D. Bope, N.E. Thomford, A. Ghansah, E.R. Chimusa, Computational/in silico methods in drug target and lead prediction, Brief. Bioinform. 21 (5) (2019) 1663–1675, http://dx.doi.org/10.1093/bib/bbz103.

[6] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE J. Biomed. Health Inf. 21 (1) (2017) 4–21, http://dx.doi.org/10.1109/JBHI.2016.2636665.

[7] A.S. Rifaioglu, H. Atas, M.J. Martin, R. Cetin-Atalay, V. Atalay, T. Doğan, Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases, Brief. Bioinform. 20 (5) (2018) 1878–1912, http://dx.doi.org/10.1093/bib/bby061.

[8] A.J. London, Artificial intelligence and black-box medical decisions: Accuracy versus explainability, Hastings Cent. Rep. 49 (1) (2019) 15–21, http://dx.doi.org/10.1002/hast.973.

[9] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: Current status and future challenges, Proteins Struct. Funct. Bioinform. 65 (1) (2006) 15–26, http://dx.doi.org/10.1002/prot.21082.

[10] J.-C. Wang, P.-Y. Chu, C.-M. Chen, J.-H. Lin, idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach, Nucleic Acids Res. 40 (W1) (2012) W393–W399, http://dx.doi.org/10.1093/nar/gks496.

[11] F. Wang, F.-X. Wu, C.-Z. Li, C.-Y. Jia, S.-W. Su, G.-F. Hao, G.-F. Yang, ACID: a free tool for drug repurposing using consensus inverse docking strategy, J. Cheminformatics 11 (1) (2019) 73, http://dx.doi.org/10.1186/s13321-019-0394-z.

[12] W. Zhang, E.W. Bell, M. Yin, Y. Zhang, EDock: blind protein–ligand docking by replica-exchange monte carlo simulation, J. Cheminformatics 12 (1) (2020) 37, http://dx.doi.org/10.1186/s13321-020-00440-9.

[13] R. Gowthaman, S.A. Miller, S. Rogers, J. Khowsathit, L. Lan, N. Bai, D.K. Johnson, C. Liu, L. Xu, A. Anbanandam, J. Aubé, A. Roy, J. Karanicolas, DARC: Mapping surface topography by ray-casting for effective virtual screening at protein interaction sites, J. Med. Chem. 59 (9) (2016) 4152–4170, http://dx.doi.org/10.1021/acs.jmedchem.5b00150.

[14] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, Nature Biotechnol. 25 (2) (2007) 197–206, http://dx.doi.org/10.1038/nbt1284.

[15] M. Luo, X.S. Wang, B.L. Roth, A. Golbraikh, A. Tropsha, Application of quantitative structure–activity relationship models of 5-HT1A receptor binding to virtual screening identifies novel and potent 5-HT1A Ligands, J. Chem. Inf. Model. 54 (2) (2014) 634–647, http://dx.doi.org/10.1021/ci400460q.

[16] J. Ma, R.P. Sheridan, A. Liaw, G.E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure–activity relationships, J. Chem. Inf. Model. 55 (2) (2015) 263–274, http://dx.doi.org/10.1021/ci500747n.

[17] B.J. Neves, R.F. Dantas, M.R. Senger, C.C. Melo-Filho, W.C.G. Valente, A.C.M. de Almeida, J.M. Rezende-Neto, E.F.C. Lima, R. Paveley, N. Furnham, E. Muratov, L. Kamentsky, A.E. Carpenter, R.C. Braga, F.P. Silva-Junior, C.H. Andrade, Discovery of new anti-schistosomal hits by integration of QSAR-based virtual screening and high content screening, J. Med. Chem. 59 (15) (2016) 7075–7088, http://dx.doi.org/10.1021/acs.jmedchem.5b02038.

[18] A. Gutteridge, M. Araki, M. Kanehisa, W. Honda, Y. Yamanishi, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces, Bioinformatics 24 (13) (2008) i232–i240, http://dx.doi.org/10.1093/bioinformatics/btn162.

[19] L. Peng, B. Liao, W. Zhu, Z. Li, K. Li, Predicting drug–target interactions with multi-information fusion, IEEE J. Biomed. Health Inf. 21 (2) (2017) 561–572, http://dx.doi.org/10.1109/JBHI.2015.2513200.

[20] X. Zheng, H. Ding, H. Mamitsuka, S. Zhu, Collaborative matrix factorization with multiple similarities for predicting drug-target interactions, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 1025–1033, http://dx.doi.org/10.1145/2487575.2487670.

[21] A. Ezzat, P. Zhao, M. Wu, X. Li, C. Kwoh, Drug-target interaction prediction with graph regularized matrix factorization, IEEE/ACM Trans. Comput. Biol. Bioinform. 14 (3) (2017) 646–656, http://dx.doi.org/10.1109/TCBB.2016.2530062.

[22] H. Yu, J. Chen, X. Xu, Y. Li, H. Zhao, Y. Fang, X. Li, W. Zhou, W. Wang, Y. Wang, A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data, PLoS One 7 (5) (2012) e37608, http://dx.doi.org/10.1371/journal.pone.0037608.

[23] E.D. Coelho, J.P. Arrais, J.L. Oliveira, Computational discovery of putative leads for drug repositioning through drug-target interaction prediction, PLoS Comput. Biol. 12 (11) (2016) e1005219, http://dx.doi.org/10.1371/journal.pcbi.1005219.

[24] L. Peng, W. Zhu, B. Liao, Y. Duan, M. Chen, Y. Chen, J. Yang, Screening drug-target interactions with positive-unlabeled learning, Sci. Rep. 7 (1) (2017) 8087, http://dx.doi.org/10.1038/s41598-017-08079-7.

[25] K. Tian, M. Shao, Y. Wang, J. Guan, S. Zhou, Boosting compound-protein interaction prediction by deep learning, Methods 110 (2016) 64–72, http://dx.doi.org/10.1016/j.ymeth.2016.06.024.

[26] P.-W. Hu, K.C.C. Chan, Z.-H. You, Large-scale prediction of drug-target interactions from deep representations, in: 2016 International Joint Conference on Neural Networks, IJCNN, 2016, pp. 1236–1243, http://dx.doi.org/10.1109/IJCNN.2016.7727339.

[27] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, K. Zheng, A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network, BMC Med. Inf. Decis. Mak. 20 (2) (2020) 49, http://dx.doi.org/10.1186/s12911-020-1052-0.

[28] M. Tsubaki, K. Tomii, J. Sese, Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, Bioinformatics 35 (2) (2019) 309–318, http://dx.doi.org/10.1093/bioinformatics/bty535.

[29] I. Lee, J. Keum, H. Nam, Deepconv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences, PLoS Comput. Biol. 15 (6) (2019) e1007129, http://dx.doi.org/10.1371/journal.pcbi.1007129.

[30] N.R.C. Monteiro, B. Ribeiro, J.P. Arrais, Drug-target interaction prediction: End-to-end deep learning approach, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (6) (2021) 2364–2374, http://dx.doi.org/10.1109/TCBB.2020.2977335.

[31] Q. Zhao, H. Zhao, K. Zheng, J. Wang, HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism, Bioinformatics 38 (3) (2022) 655–662, http://dx.doi.org/10.1093/bioinformatics/btab715.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017, arXiv:1706.03762v5.

[33] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2019, arXiv:1810.04805v2.

[34] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, 2021, arXiv:2010.11929v2.

[35] K. Huang, C. Xiao, L.M. Glass, J. Sun, MolTrans: Molecular interaction transformer for drug–target interaction prediction, Bioinformatics 37 (6) (2021) 830–836, http://dx.doi.org/10.1093/bioinformatics/btaa880.

[36] A. Gaulton, A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magariños, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucleic Acids Res. 45 (D1) (2017) D945–D954, http://dx.doi.org/10.1093/nar/gkw1074.

[37] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 44 (D1) (2016) D1045–D1053, http://dx.doi.org/10.1093/nar/gkv1072.

[38] P.J. Ballester, J.B.O. Mitchell, A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking, Bioinformatics 26 (9) (2010) 1169–1175, http://dx.doi.org/10.1093/bioinformatics/btq112.

[39] P.J. Ballester, A. Schreyer, T.L. Blundell, Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity? J. Chem. Inf. Model. 54 (3) (2014) 944–955, http://dx.doi.org/10.1021/ci500091r.

[40] J.D. Durrant, J.A. McCammon, NNScore: A neural-network-based scoring function for the characterization of protein-ligand complexes, J. Chem. Inf. Model. 50 (10) (2010) 1865–1871, http://dx.doi.org/10.1021/ci100244v.

[41] J.D. Durrant, J.A. McCammon, NNScore 2.0: A neural-network receptor-ligand scoring function, J. Chem. Inf. Model. 51 (11) (2011) 2897–2903, http://dx.doi.org/10.1021/ci2003889.

[42] S. Kumar, M.-h. Kim, SMPLIP-Score: predicting ligand binding affinity from simple and interpretable on-the-fly interaction fingerprint pattern descriptors, J. Cheminformatics 13 (1) (2021) 28, http://dx.doi.org/10.1186/s13321-021-00507-1.

[43] R. Meli, A. Anighoro, M.J. Bodkin, G.M. Morris, P.C. Biggin, Learning protein-ligand binding affinity with atomic environment vectors, J. Cheminformatics 13 (1) (2021) 59, http://dx.doi.org/10.1186/s13321-021-00536-w.

[44] I. Wallach, M. Dzamba, A. Heifets, AtomNet: A Deep convolutional neural network for bioactivity prediction in structure-based drug discovery, 2015, arXiv:arXiv:1510.02855.

[45] M.M. Stepniewska-Dziubinska, P. Zielenkiewicz, P. Siedlecki, Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, Bioinformatics 34 (21) (2018) 3666–3674, http://dx.doi.org/10.1093/bioinformatics/bty374.

[46] J. Jiménez, M. Škalič, G. Martínez-Rosell, G. De Fabritiis, KDEEP: PRotein-ligand absolute binding affinity prediction via 3D-convolutional neural networks, journal=j. chem. inf. model., 58 (2) (2018) 287–296, http://dx.doi.org/10.1021/acs.jcim.7b00650.

[47] D. Jones, H. Kim, X. Zhang, A. Zemla, G. Stevenson, W.F.D. Bennett, D. Kirshner, S.E. Wong, F.C. Lightstone, J.E. Allen, Improved protein–ligand binding affinity prediction with structure-based deep fusion inference, journal=j. chem. inf. model., 61 (4) (2021) 1583–1592, http://dx.doi.org/10.1021/acs.jcim.0c01306.

[48] P.A. Shar, W. Tao, S. Gao, C. Huang, B. Li, W. Zhang, M. Shahen, C. Zheng, Y. Bai, Y. Wang, Pred-binding: large-scale protein–ligand binding affinity prediction, J. Enzyme Inhib. Med. Chem. 31 (6) (2016) 1443–1450, http://dx.doi.org/10.3109/14756366.2016.1144594.

[49] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Szwajda, J. Tang, T. Aittokallio, Toward more realistic drug-target interaction predictions, Brief. Bioinform. 16 (2) (2014) 325–337, http://dx.doi.org/10.1093/bib/bbu010.

[50] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, SimBoost: A read-across approach for predicting drug–target binding affinities using gradient boosting machines, J. Cheminformatics 9 (1) (2017) 24, http://dx.doi.org/10.1186/s13321-017-0209-z.

[51] H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug-target binding affinity prediction, Bioinformatics 34 (17) (2018) i821–i829, http://dx.doi.org/10.1093/bioinformatics/bty593.

[52] Q. Feng, E. Dueva, A. Cherkasov, M. Ester, PADME: A Deep learning-based framework for drug-target interaction prediction, 2019, arXiv:arXiv:1807.09741v4.

[53] T. Nguyen, H. Le, T.P. Quinn, T. Nguyen, T.D. Le, S. Venkatesh, GraphDTA: Predicting drug-target binding affinity with graph neural networks, Bioinformatics (2020) http://dx.doi.org/10.1093/bioinformatics/btaa921.

[54] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J.B. Ghasemi, A. Masoudi-Nejad, DeepCDA: deep cross-domain compound-protein affinity prediction through LSTM and convolutional neural networks, Bioinformatics 36 (17) (2020) 4633–4642, http://dx.doi.org/10.1093/bioinformatics/btaa544.

[55] J. Shim, Z.-Y. Hong, I. Sohn, C. Hwang, Prediction of drug-target binding affinity using similarity-based convolutional neural network, Sci. Rep. 11 (1) (2021) 4416, http://dx.doi.org/10.1038/s41598-021-83679-y.

[56] K. Wang, R. Zhou, Y. Li, M. Li, DeepDTAF: a deep learning method to predict protein–ligand binding affinity, Brief. Bioinform. 22 (5) (2021) http://dx.doi.org/10.1093/bib/bbab072, bbab072.

[57] A.S. Rifaioglu, R. Cetin Atalay, D. Cansen Kahraman, T. Doğan, M. Martin, V. Atalay, MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery, Bioinformatics 37 (5) (2021) 693–704, http://dx.doi.org/10.1093/bioinformatics/btaa858.

[58] M.I. Davis, J.P. Hunt, S. Herrgard, P. Ciceri, L.M. Wodicka, G. Pallares, M. Hocker, D.K. Treiber, P.P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, Nature Biotechnol. 29 (11) (2011) 1046–1051, http://dx.doi.org/10.1038/nbt.1990.

[59] T.U. Consortium, UniProt: The universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (D1) (2020) D480–D489, http://dx.doi.org/10.1093/nar/gkaa1100.

[60] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem In 2021: new data content and improved web interfaces, Nucleic Acids Res. 49 (D1) (2020) D1388–D1395, http://dx.doi.org/10.1093/nar/gkaa971.

[61] G. Landrum, RDKit: Open-source cheminformatics, 2021, URL http://www.rdkit.org.

[62] C.-F. Chen, Q. Fan, R. Panda, CrossViT: CRoss-attention multi-scale vision transformer for image classification, 2021, arXiv:2103.14899v2.

[63] H. Pagés, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings, 2022, R package version 2.64.0, URL https://bioconductor.org/packages/Biostrings.

[64] D. Hendrycks, K. Gimpel, Gaussian Error linear units (GELUs), 2020, arXiv:1606.08415v4.

[65] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, J. Han, On the variance of the adaptive learning rate and beyond, 2021, arXiv:1908.03265v4.

[66] J. Desaphy, G. Bret, D. Rognan, E. Kellenberger, sc-PDB: A 3D-database of ligandable binding sites—10 years on, Nucleic Acids Res. 43 (D1) (2015) D399–D404, http://dx.doi.org/10.1093/nar/gku928.

[67] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings, J. Chem. Inf. Model. 61 (8) (2021) 3891–3898, http://dx.doi.org/10.1021/acs.jcim.1c00203.

[68] A. Volkamer, D. Kuhn, T. Grombacher, F. Rippmann, M. Rarey, Combining global and local measures for structure-based druggability predictions, J. Chem. Inf. Model. 52 (2) (2012) 360–372, http://dx.doi.org/10.1021/ci200454v.