Expert Systems With Applications 306 (2026) 130985

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Molecular mechanics-aware feature fusion framework for predicting protein-ligand binding affinity

Xing Lv [a,b,c], Xinhui Tu [a,b,c], Tingting He [a,b,c], Weizhong Zhao [a,b,c,*]

[a] Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, Wuhan, Hubei postcode, China
[b] School of Computer, Central China Normal University, Wuhan, Hubei postcode, China
[c] National Language Resources Monitoring & Research Center for Network Media, Central China Normal University, Wuhan, Hubei postcode, China

ARTICLE INFO

Keywords:
Protein-ligand binding affinity
Molecular mechanism
Graphtransform
Highway network

ABSTRACT

Drug-target affinity (DTA) prediction is a key indicator in identifying drug-target interactions (DTIs), which is the most crucial step of drug discovery. Traditional experiment-based methods can accurately measure affinity but are typically labor-intensive and time-consuming. The intrinsic intermolecular forces, such as hydrogen bonds and electrostatic interactions, are crucial for determining binding affinity. Molecular mechanics utilizes potential functions based on atomic coordinates and associated parameters to describe the energy of a system, capturing these essential forces. Existing deep learning approaches primarily focus on sequence and structural information of protein-ligand complexes while overlooking the essential role of molecular mechanics. This limitation results in reduced generalization capability. To address this gap, we propose MMA-DTA (Molecular Mechanics-aware Drug-Target Affinity prediction), a novel framework that explicitly integrates molecular mechanics theory into deep learning-based representation learning. Our model consists of three components: (1) the input module computes covalent features (i.e., bond stretching, angle bending, and dihedral torsion) and non-covalent features (i.e., electrostatic interactions, van der Waals forces, and hydrogen bonds) using molecular mechanics functions; (2) the feature fusion module processes covalent features via Highway Networks with multi-head attention and non-covalent features via GraphTransform with edge attention mechanisms; and (3) the prediction module employs a gating mechanism to dynamically integrate both obtained features. Comprehensive experiments on PDBBind datasets demonstrate that MMA-DTA performs better than SOTA baselines on the task of DTA prediction according to extensive evaluation metrics. Moreover, although the absolute performance gain over the best baseline MMGA is narrow (e.g., 1.102 vs. 1.104 on RMSE), the results of t-test confirm that the performance improvement is statistically significant and robust. By effectively combining covalent and non-covalent interactions, the proposed model is able to learn meaningful representations that enable accurate protein-ligand binding affinity prediction.

1. Introduction

Pharmaceutical compounds are diverse and numerous, with estimates ranging from 10^{23} to 10^{60} (Polishchuk et al., 2013). Efficiently screening these molecules is therefore a critical challenge. Drug-target affinity (DTA) is a key indicator of drug-target interactions (DTIs) (Pahikkala et al., 2015), commonly expressed in units of $-logK_d$. Traditional experimental methods, such as Isothermal Titration Calorimetry (ITC) (Velázquez-Campoy et al., 2004), Surface Plasmon Resonance (SPR) (Yamamoto, 2002), and Enzyme-Linked Immunosorbent Assay (ELISA) (Lequin, 2005), can accurately measure binding affinity scores and screen drug molecules accordingly. However, they require substantial human resources and time, making it difficult to process large-scale datasets.

As a result, methods that use simulation-assisted predictions of DTIs have emerged. Molecular docking (Fan et al., 2019) and molecular dynamics simulations (Hollingsworth & Dror, 2018) can accurately predict protein-ligand affinity values, but these methods also require significant computational resources, making them difficult to handle large-scale data. In contrast, machine learning and deep learning have shown excellent performance in DTA prediction, as they can predict a large number of unrecognized samples with relatively small amounts of data. Deep learning methods, in particular, are better at extracting meaningful features from protein-ligand interactions.

* Corresponding author.
E-mail addresses: chichu1116@mails.ccnu.edu.cn (X. Lv), tuxinhui@ccnu.edu.cn (X. Tu), tthe@mail.ccnu.edu.cn (T. He), wzzhao@ccnu.edu.cn (W. Zhao).

https://doi.org/10.1016/j.eswa.2025.130985
Received 22 August 2025; Received in revised form 13 December 2025; Accepted 23 December 2025
Available online 27 December 2025

Generally, deep learning-based methods can be categorized into sequence-based (Chen et al., 2024; Hua et al., 2023; Jin et al., 2023; Tang et al., 2024b; Wang et al., 2021) and structure-based approaches (Li et al., 2023; Nguyen et al., 2021; Stepniewska-Dziubinska et al., 2018; Tang et al., 2024a; Wang et al., 2023; Xia et al., 2023; Yi et al., 2024). Sequence-based models require only the amino acid sequences of proteins and the SMILES strings of ligands as input, without relying on the 3D structural information of proteins. However, this lack of structural context limits their ability to directly capture the 3D spatial interactions between proteins and ligands, such as hydrogen bonds, hydrophobic interactions, and van der Waals forces, and often results in reduced interpretability. Structural information has been shown to be beneficial for binding affinity prediction (Callaway, 2020). Structure-based models address this by integrating global sequence features with local 3D structural characteristics, often employing architectures like 3D Convolutional Neural Networks (3D-CNNs) or Graph Neural Networks (GNNs). These models can simultaneously capture long-range indirect interactions and short-range direct interactions between proteins and ligands. However, these approaches typically focus solely on structural information, potentially overlooking the critical role of molecular mechanics in protein-ligand binding.

Molecular mechanics involves the use of potential functions based on atomic coordinates and associated parameters to describe the energy of a system (Demtröder, 2008). It forms the foundation of molecular dynamics simulations, enabling the rapid and precise calculation of molecular interactions. Molecular mechanics has proven effective across various domains (Ahmed et al., 2018; Cornell et al., 1995; Lechelon et al., 2022) and can also be applied to predict protein-ligand affinities. Its advantage lies in its ability to model fundamental intermolecular forces, such as hydrogen bonds and electrostatic interactions, which are crucial to binding affinity. Existing studies have also confirmed this point. Janbaz-amirani et al. (2025) revealed the relationship between the thermodynamic characteristics (e.g., hydrogen bonds, van der Waals forces, electrostatic interactions, etc.) of the β-Lg-EA complex and its spontaneous binding and stability. Yang et al. (2025) elucidated the interaction mechanism between peanut protein and resveratrol, indicating that the binding occurs primarily through non-covalent interactions. For the protein TLR8, biochemical mutation experiments and cellular functional assays directly demonstrated that disrupting its interaction with the imidazoquinoline ligand (Resiquimod) significantly impairs the protein's ligand-binding capacity and biological function (Cai et al., 2024).

A critical consideration in applying molecular mechanics to protein-ligand binding affinity prediction is the distinctive characteristics of covalent and non-covalent interactions, which require their separate treatments in prediction models. Covalent interactions define the internal geometry and conformational stability of individual molecules, operating at short spatial scales with relatively high energy magnitudes. These interactions are intramolecular in nature and determine the intrinsic structural properties of proteins and ligands independently (Du et al., 2016). In contrast, non-covalent interactions govern the intermolecular recognition and binding between proteins and ligands, operating at longer spatial scales with lower energy magnitudes (Nilofer et al., 2017). These interactions are highly dependent on the relative spatial arrangement and chemical complementarity between binding partners (Leckband & Israelachvili, 2001). From a physical perspective, this fundamental distinction suggests that covalent features primarily capture the "readiness" of individual molecules to participate in binding (i.e., their conformational states and flexibility) (Amaral et al., 2017; Celej et al., 2003), while non-covalent features capture the binding state capability through intermolecular recognition. Treating these two types of interactions separately will be helpful to learn specialized representations for enhancing generalization ability of prediction models.

Based on the above discussion, we propose a novel model for DTA prediction called MMA-DTA, which applies a molecular mechanics-aware information fusion mechanism for drug-target representation learning. MMA-DTA utilizes a Highway Network (Srivastava et al., 2015) combined with CNN (Yamashita et al., 2018) to extract both global and local features of covalent interactions for proteins and ligands. This architectural choice is critical because simple dense layers often suffer from signal attenuation when processing high-magnitude physical constraints (e.g., rigid bond lengths and angles) (He et al., 2016). By employing a gating mechanism that facilitates direct information flow, the Highway Network ensures that these intrinsic geometric properties are preserved and not diluted during deep non-linear feature extraction (Srivastava et al., 2015). A multi-head attention mechanism is employed to better learn the features within these interactions. Additionally, a GraphTransform (Yun et al., 2019) architecture is used to process the protein-ligand graph structure. Three types of non-covalent interactions are used as attention weight inputs to the GraphTransform to learn potential feature representations. Finally, the three sets of features are effectively fused to jointly predict the binding affinity score. In summary, our main contributions are listed as follows:

- We propose a novel framework for predicting protein-ligand binding affinity, in which the fundamental theory of molecular mechanics is explicitly utilized for representation learning.
- Considering the different mechanisms characterizing covalent interactions and non-covalent interactions, different models are employed for modeling covalent features and non-covalent features, respectively.
- Comprehensive experiments are conducted and the results demonstrate the effectiveness of the proposed model.

2. Related work

In this section, we briefly review existing work related to our model MMA-DTA, including sequence-based models and structure-based models.

2.1. Sequence-based models

Sequence-based models for predicting protein-ligand binding affinity primarily leverage the linear sequences of proteins and ligands as input. These models often integrate multiple features of sequences, including physicochemical properties of amino acids, secondary structure elements, and molecular fingerprints. In DeepDTAF (Wang et al., 2021) and TC-DTA (Tang et al., 2024b), the physicochemical properties of amino acids and secondary structure features are employed as input representations, and then CNN is used to merge the linear sequences of proteins and ligand molecular features for binding affinity prediction. The model called CAPLA (Jin et al., 2023) incorporates multi-dimensional protein-binding pocket features including amino acid types, secondary structure elements (SSEs), and residue physicochemical properties, and predicts the binding affinity by using a cross-attention mechanism. In MFR-DTA (Hua et al., 2023), it combines the chemical features of molecular fingerprints (FCFP) with topological features by using Graph Convolutional Networks (GCN), deriving the binding affinity prediction accordingly. In addition, DEAttentionDTA (Chen et al., 2024) leverages dynamic embedding techniques to capture the ternary interactions among amino acid sequences, active pocket sequences, and small molecule sequences, on which the prediction of binding affinity is conducted.

Generally, the limitation of methods mentioned above lies in ignoring completely the three-dimensional structural information. Specifically, they fail to effectively integrate key physicochemical features such as atomic distances and molecular interactions. This limitation makes it challenging to accurately simulate the spatial structure-activity relationship of protein-ligand complexes.

2.2. Structure-based model

Structure-based models for predicting protein-ligand binding affinity focus on explicitly utilizing the three-dimensional structures of the

molecules involved. These approaches take into account the detailed structural features of both the protein and ligand, employing advanced techniques such as 3D Convolutional Neural Networks (3D-CNN) and graph-based models. By integrating spatial and physicochemical properties, these methods aim to capture the complex dynamics of molecular interactions. The Pafnucy (Stepniewska-Dziubinska et al., 2018) framework was the first to use 3D-CNN, with its four-dimensional tensor representation integrating the spatial and physicochemical descriptors of the binding pocket. DeepAtom (Li et al., 2023) expanded on this approach by employing voxelized 3D grids to automatically extract stereochemical and electronic interaction patterns through hierarchical feature learning in CNNs. In addition, the advancements in geometric deep learning have made it possible to directly process molecular graphs. In ENGA (Xia et al., 2023), it proposes an empirical graph neural network with an empirical interaction representation layer to represent the interactions between the protein and ligand, facilitating information exchange. In GraphscoreDTA (Wang et al., 2023), it introduces three types of graph structures (i.e., protein graph, ligand graph, and interaction graph) to extract features, incorporating a dual-transmission information mechanism to facilitate bidirectional information transfer between proteins and ligands. The model called ELGN (Yi et al., 2024) constructs an Interaction Graph to represent the bonding topology network and employs a dual-modal embedding strategy to jointly represent atomic-level spatial coordinates and bonding features, thereby achieving high-resolution modeling of molecular 3D geometric configurations. EHIGN (Yang et al., 2024) implements a heterogeneous graph module, in which four types of interactions including within-ligand, within-protein, ligand→protein, and protein→ligand are processed through dedicated graph convolutional operators. da Silva et al. (2024) proposed a bilinear optimization-based technique that precisely finds the global minimum of the protein energy function to predict the protein's native structure.

Moreover, some models integrate the characteristics of both sequence-based and structure-based methods, combining sequence and structural information to predict protein-ligand binding affinity. GraphDTA (Nguyen et al., 2021) has validated the effectiveness of integrating protein feature information with various graph neural networks and 1D-CNN in the field of binding affinity prediction. KC-DTA (Jiang et al., 2023) utilizes k-mer analysis and Cartesian product to convert sequence information into 2D and 3D matrices, and it applies GNN to the drug molecule graphs for obtaining structural features. In MFF-DTA (Tang et al., 2024a), features of drugs such as SMILES sequences, Morgan and Avalon fingerprints are combined by using Graph Convolutional Networks (GCN) and skip connections; while for proteins, 1D-CNN and self-attention mechanisms are used to extract both local and global semantics from the amino acid sequence. PocketDTA (Zhao et al., 2024) proposes a multimodal contrastive learning architecture that integrates molecular fingerprint features with protein pocket structures. However, it only considers dihedral angle potential energy and fails to comprehensively account for the role of molecular mechanics. Lv et al. proposed a model called MMGA (Lv et al., 2024), in which it simply uses a heterogeneous graph neural network to integrate the molecular mechanics-based features for predicting the binding affinity. More specifically, in MMGA, covalent and non-covalent interactions are distinguished simply by different edge types in the graph topology, which may overlook the complex interactions between intramolecular stability and intermolecular binding. In contrast, the proposed model MMA-DTA decouples the processing flow for covalent interactions (via Highway Network) and non-covalent interactions (via GraphTransform), and integrates them through an adaptive gating mechanism. This unique feature fusion strategy allows the model to dynamically weigh the contributions of conformational constraints and binding interactions, thereby capturing subtle differences that are ignored by fixed edge types in heterogeneous graphs.

Although many methods have been proposed for the task of binding affinity prediction, the common issue lies in the failure to make full use of the sequence and structural information contained in ligands and proteins, leaving the room to improve the performance of prediction models.

3. Methods

In this section, we describe the detailed implementation of the MMA-DTA model, which consists of three parts, i.e., the input module, the molecular mechanics-aware feature fusion module, and the prediction module. The schematic representation of MMA-DTA is shown in Fig. 1. The input module processes the protein and ligand data to obtain the covalent and non-covalent feature matrices for the protein, ligand and protein-ligand complex, respectively. The molecular mechanics-aware feature fusion module utilizes the Highway Networks and GraphTransform networks to make full use of the covalent and non-covalent features. The prediction module uses a gate mechanism-based fusion of obtained features to complete the binding affinity score prediction.

3.1. The input module

In this module, we employ molecular mechanics functions to simulate and obtain the molecular mechanics features of the protein, ligand, and protein-ligand complex. Specifically, we first process the protein and ligand files to obtain their atomic node types and spatial data (i.e., the coordinate information of proteins and ligands). For atomic types, one-hot encoding is used as the initial graph node features, while the spatial data is used to determine the existence of edges between the pair of protein and ligand, which is described in detail in the Graph Construction section. We use P_i and L_j to represent a protein and its corresponding ligand, respectively. For a given protein P_i, the total number of protein atoms is denoted as M_i, with each atom indexed by m ($1 \le m \le M_i$). Similarly, for a ligand L_j, the total number of ligand atoms is denoted as N_j, with each atom indexed by n ($1 \le n \le N_j$).

3.1.1. Protein / ligand features

For proteins or ligands, we simulate the molecular mechanics mechanism and calculate specific values based on interactions between atomic pairs. Since the processing procedure is the same for proteins and ligands, we take a protein as an example to describe the processing procedure for extracting molecular mechanics features.

The bond stretching energy quantifies the energy required to elongate or compress a chemical bond between atoms within a molecule, which is associated with the relative positional changes of bonded atomic pairs. Bond angle bending energy denotes the energy penalty arising from deviations of bond angles between adjacent atoms from their ideal equilibrium values, reflecting the inherent geometric constraints imposed by atomic orbital hybridization. Given each atom p_m of protein P_i, harmonic potential functions (Katznelson, 2004) are implemented to formally describe bond stretching and angle bending interactions as follows,

$$E^{\text{bond}}(p_m) = \frac{1}{2} k\,(r - r_0)^2 \tag{1}$$

where k is the bond force constant, r is the interatomic distance, and r_0 is the equilibrium bond length. In this study, k and r_0 are treated as fixed empirical constants to standardize the input. Specifically, these values are determined based on atom-pair types (r_0 is set to 1.54 Å for C-C bonds and 1.33 Å for C-N bonds) consistent with standard parameters in the MMFF94 force field (Halgren, 1996).

$$E^{\text{angle}}(p_m) = \frac{1}{2} k_\theta\,(\theta - \theta_0)^2 \tag{2}$$

where k_θ is the angular force constant, θ is the measured angle, and θ_0 is the ideal bond angle. Similarly, k_θ and θ_0 are set as fixed generalized constants (specifically, θ_0 is set to 109.5 degrees to represent ideal tetrahedral geometry) to capture geometric deviations uniformly across the molecular structure.

The dihedral torsion characterizes the energy variation associated with rotation around a chemical bond that connects four sequentially

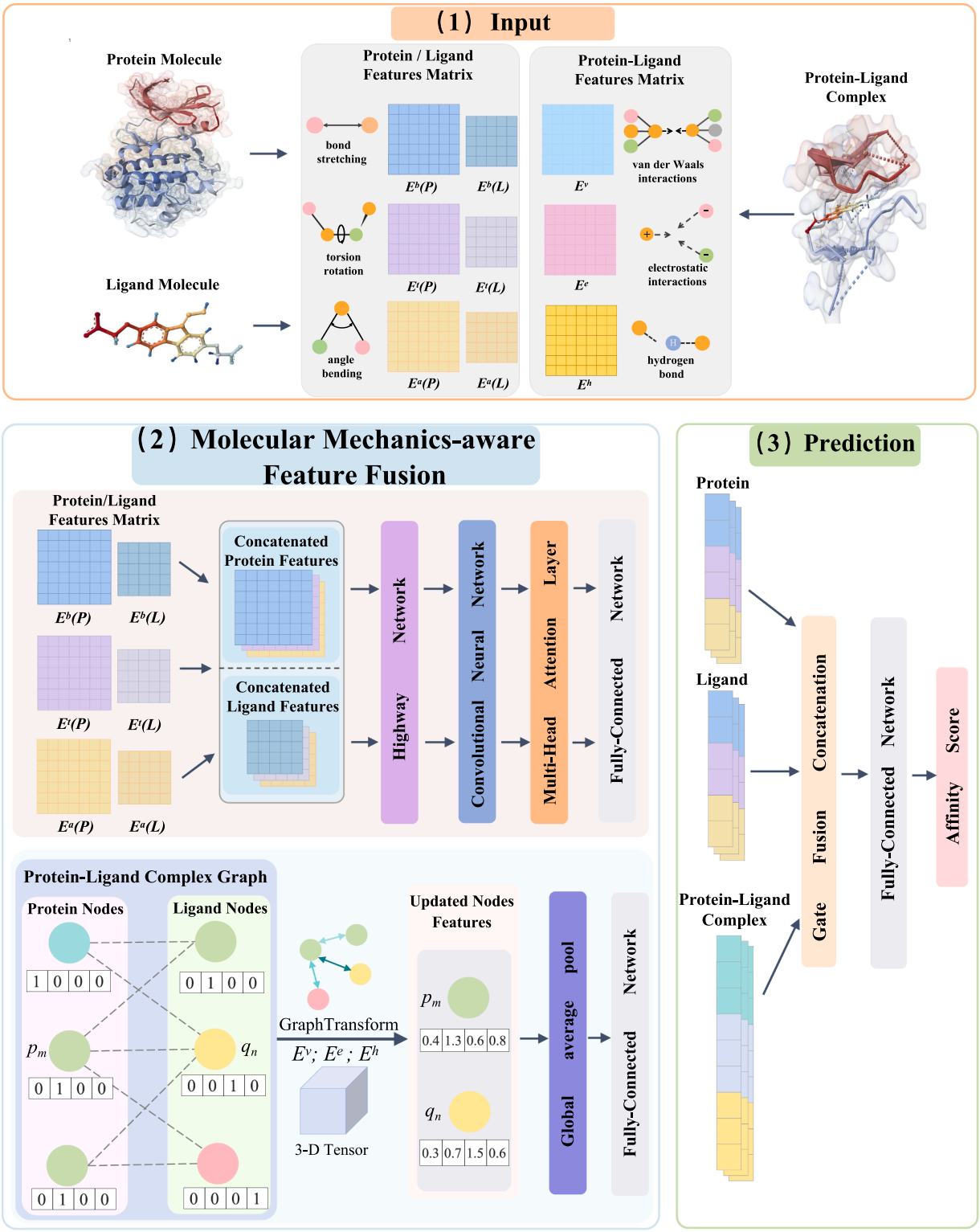

Fig. 1. The overview of MMA-DTA model. The model consists of three modules: the input, the molecular mechanics-aware feature fusion, and the prediction. First, the protein and ligand are processed separately to obtain their covalent features, while the complex is processed to derive its non-covalent features. Next, the obtained feature matrices are processed in the molecular mechanics-aware feature fusion module to derive the high-quality representations. Finally, the derived features are integrated via a gating mechanism to predict the binding affinity score.

bonded atoms, playing a critical role in governing conformational transitions between distinct molecular states and directly influencing chemical stability, dynamic behavior, and functional properties. Formally, given each atom p_m of protein P_i, the dihedral torsion energy is defined as follows,

$$E^{\text{torsion}}(p_m) = \sum_{t \in \mathcal{T}_m} V_t \left[1 + \cos\left(n_t \phi_t - \gamma_t\right)\right] \tag{3}$$

where $\mathcal{T}_m$ denotes the set of torsion angles involving atom p_m. For the tth torsion angle in this set, V_t is the force constant, n_t is the periodicity factor, ϕ_t is the torsion angle, and γ_t is the phase offset. The parameters V_t, n_t, and γ_t are pre-defined coefficients which are derived from general force field specifications.

Based on the formulations above, we obtain three types of covalent molecular mechanics features for individual atoms within both proteins and ligands. Specifically, for a given protein P_i, the molecular mechanics features of all constituent atoms are aggregated to construct feature matrices, denoted as $E^b(P_i)$, $E^a(P_i)$, and $E^t(P_i)$, corresponding to bond stretching, angle bending, and dihedral torsion interactions, respectively. Similarly, for each ligand L_j, the corresponding three types of covalent feature matrices are denoted as $E^b(L_j)$, $E^a(L_j)$, and $E^t(L_j)$.

3.1.2. Protein-ligand complex features

In order to obtain the non-covalent features of the given pair of protein and ligand, we calculate three types of non-covalent interactions, including electrostatic interactions, van der Waals forces, and hydrogen bonds. Electrostatic interaction is one of the core types of non-covalent forces occurring between or within molecules. Electrostatic interactions arise from the differing abilities of elements to attract electrons, leading to an uneven distribution of charges within the molecule and playing a dominant role in molecular recognition, structural stability, and energy transfer. We simulate and calculate the electrostatic interactions by using the point charge formula. For each pair of protein P_i and ligand L_j, the electrostatic interactions feature is denoted by a $M_i \times N_j$ matrix $E_{ij}^{\text{electrostatic}}$, which is defined as follows,

$$E_{ij}^{\text{electrostatic}} = \begin{bmatrix} \frac{K\, C_{p_1}\, C_{q_1}}{\epsilon\, r_{11}} & \cdots & \frac{K\, C_{p_1}\, C_{q_n}}{\epsilon\, r_{1n}} & \cdots & \frac{K\, C_{p_1}\, C_{q_{N_j}}}{\epsilon\, r_{1N_j}} \\ \vdots & & \vdots & & \vdots \\ \frac{K\, C_{p_m}\, C_{q_1}}{\epsilon\, r_{m1}} & \cdots & \frac{K\, C_{p_m}\, C_{q_n}}{\epsilon\, r_{mn}} & \cdots & \frac{K\, C_{p_m}\, C_{q_{N_j}}}{\epsilon\, r_{mN_j}} \\ \vdots & & \vdots & & \vdots \\ \frac{K\, C_{p_{M_i}}\, C_{q_1}}{\epsilon\, r_{M_i 1}} & \cdots & \frac{K\, C_{p_{M_i}}\, C_{q_n}}{\epsilon\, r_{M_i n}} & \cdots & \frac{K\, C_{p_{M_i}}\, C_{q_{N_j}}}{\epsilon\, r_{M_i N_j}} \end{bmatrix} \tag{4}$$

where K is Coulomb's constant ($\approx 8.98755 \times 10^9\, \text{N}\,\text{m}^2/\text{C}^2$), ϵ is the dielectric constant, and r_{mn} is the distance between the mth atom p_m of protein P_i and the nth atom q_n of ligand L_j, C_{p_m} and C_{q_n} represent the charges of the atom p_m and the atom q_n, respectively. In our implementation, these parameters are assigned based on element-specific empirical values (e.g., distinct typical charge ranges for C, N, and O atoms) to approximate the electrostatic potential. These values are kept unchanged to preserve chemical consistency.

Since electrostatic interactions can not fully encompass all non-bonded interactions within a system, van der Waals interactions are calculated as well, which primarily arise from three sources including induction, dispersion, and orientation forces. In this study, the Lennard-Jones 6–12 function (Ferguson & Kollman, 1991), which is a commonly used method for computing van der Waals interactions, is used to compute the van der Waals interaction feature. More specifically, for each pair of protein P_i and ligand L_j, the van der Waals interaction feature is denoted by a $M_i \times N_j$ matrix E_{ij}^{vdW}, and the entry of mth row and nth column denotes the van der Waals interaction between the mth atom p_m of protein P_i and the nth atom q_n of ligand L_j, which is defined by the following equation,

$$E_{ij}^{\text{vdW}}(m,n) = 4\,\epsilon \left[\left(\frac{\sigma}{r_{mn}}\right)^{12} - \left(\frac{\sigma}{r_{mn}}\right)^{6} \right] \tag{5}$$

where ϵ is the potential well depth, σ is a distance parameter, and r_{mn} is the interatomic distance between the atom p_m and the atom q_n. Note that the coefficients ϵ and σ are set to fixed empirical constants (e.g., $\sigma = 3.5$Å) for all atom pairs in this study.

Hydrogen bond is a special type of interaction formed by a hydrogen atom, which is covalently bonded to a highly electronegative atom (such as O, N), interacting with another highly electronegative atom through directional electrostatic attraction. Given each pair of protein P_i and ligand L_j, the hydrogen bond feature is represented by a $M_i \times N_j$ matrix E_{ij}^{hbond}, in which the entry of mth row and nth column denotes the hydrogen bond interaction between the mth atom p_m of protein P_i and the nth atom q_n of ligand L_j, which is defined as follows,

$$E_{ij}^{\text{hbond}}(m,n) = E_{\text{scale}} \cdot \left(1 - \frac{r_{mn}}{r_{\text{cutoff}}}\right) \tag{6}$$

where E_{scale} is a scaling factor, r_{cutoff} is the cutoff value of distance, and r_{mn} is the interatomic distance between the atom p_m and the atom q_n. E_{scale} and r_{cutoff} are fixed geometric constraints used to explicitly quantify the interaction feasibility based on spatial proximity. By fixing these coefficients, we construct a deterministic interaction landscape that serves as a robust inductive bias for the deep learning module.

For the three non-covalent molecular mechanics features, we use three matrices E_{ij}^{v}, E_{ij}^{e}, E_{ij}^{h}, respectively in the following descriptions.

3.2. The molecular mechanics-aware feature fusion module

In this module, we employ different deep neural networks to respectively process covalent features and non-covalent features to effectively extract information from the obtained molecular mechanics features.

3.2.1. Covalent feature processing

For proteins and ligands, the covalent bond information is processed following the same process. In this section, we use the processing of proteins as an example to introduce how to process the covalent features.

For protein P_i, three covalent bond information matrices $E^b(P_i)$, $E^a(P_i)$, and $E^t(P_i)$ are concatenated and the resultant matrix is denoted by $E(P_i)$. Then, a Highway convolutional structure is conducted on $E(P_i)$ for hidden representation learning. The Highway convolutional structure is a neural network with a gating mechanism that establishes a shortcut directly between the input and output to avoid the gradient exploding or vanishing. Specifically, the Highway convolutional structure consists of two parts: a conventional transformation branch and a gating branch. The conventional transformation branch employs convolution operations along with the ReLU activation function to extract feature representations $Q(P_i)$, while the gating branch generates weights $T(P_i)$ via the Sigmoid function to dynamically control the information flow between $Q(P_i)$ and the original input $E(P_i)$. Finally, $X(P_i)$ is computed as a weighted sum of $Q(P_i)$ and $E(P_i)$ based on the gating weights, thereby preserving the original information while introducing a more expressive nonlinear transformation in a learnable manner. This design effectively mitigates the gradient vanishing issue of deep neural networks and has demonstrated strong generalization capabilities across multiple tasks. The formal descriptions are presented as follows,

$$Q(P_i) = \text{ReLU}\left(\text{Conv}_{3\times3}(E(P_i))\right) \tag{7}$$

$$T(P_i) = \sigma\left(\text{Conv}_{1\times1}(E(P_i))\right) \tag{8}$$

$$X(P_i) = Q(P_i) \odot T(P_i) \;+\; E(P_i) \odot \left(1 - T(P_i)\right) \tag{9}$$

where $Conv_{3\times3}$ and $Conv_{1\times1}$ respectively denote the conventional convolution branch and the gating convolution branch; $\odot$ denotes the element-wise multiplication, performing a weighted fusion of the output feature $Q(P_i)$ and the original input $E(P_i)$ according to the gating mechanism $T(P_i)$.

To further process $X(P_i)$, we employ the convolution and pooling layers to extract spatial features, and the obtained feature vector is denoted as $X'(P_i)$ in Eq. (10).

$$X'(P_i) = MAXPool(ReLU(Conv(X(P_i)))) \tag{10}$$

Then, $X'(P_i)$ is flattened and fed into a multi-head attention module to derive more meaningful representation, which is defined formally as in Eq. (11).

$$X''(P_i) = Multi - Attention(X'(P_i)) \tag{11}$$

Finally, $X''(P_i)$ is fed into a fully connected layer with a non-linear activation $sigmoid$ to generate the final representation $H(P_i)$ defined in Eq. (12).

$$H(P_i) = \sigma(W^p X''(P_i) + b^p) \tag{12}$$

Through the above operations, we obtain the protein's covalent feature $H(P_i)$. Similarly, the ligand's covalent feature is obtained by the same procedure, which is denoted by $H(L_j)$.

3.2.2. Non-covalent feature processing

Given each pair of protein and ligand, we first construct a graph based on the spatial data of the complex. By applying GraphTransform (Yun et al., 2019), we then make full use of the semantic information contained in non-covalent bonds to obtain more enriched representations for each atom in the graph. Finally, all atomic representations are integrated to derive a graph-level representation, which serves as the final representation of the protein-ligand complex. Each step will be described in detail as follows.

Graph construction. Given a pair of protein P_i and ligand L_j, we construct a bipartite graph based on non-covalent features, in which nodes are the atoms of the protein and ligand, edges are non-covalent bonds connecting each atom in the protein with each atom in the ligand. For each node, the one-hot encoded vector is utilized as its initial feature. Formally, for the atom p_m in P_i or the atom q_n in L_j, the initial feature is denoted as $f(p_m)$ or $f(q_n)$. For the non-covalent bond information, the spatial data is obtained by processing the protein-ligand complex with RDKit (Landrum, 2013), and an edge is formed between a protein atom p_m and a ligand atom q_n if the distance between them is less than or equal to 20 Å.

Graph message passing. Based on the constructed graph, we apply the GraphTransform framework (Yun et al., 2019) to capture the semantic information contained in the non-covalent bonds in the protein-ligand complex. More specifically, given a pair of protein P_i and ligand L_j, we first concatenate the three types of non-covalent molecular force matrices to obtain a 3D tensor, which is denoted as E_{ij}. Then, three independent projection modules are conducted on E_{ij} to derive the query (Q), key (K), and value (V) for multi-head attention mechanism. In each projection module, a two-layer fully connected network with Gaussian Error Linear Unit (GELU) activation is employed to derive the feature matrices.

$$\begin{aligned} Q_{ij} &= \mathrm{GELU}(W_2^q \cdot \mathrm{GELU}(W_1^q \cdot E_{ij})) \\ K_{ij} &= \mathrm{GELU}(W_2^k \cdot \mathrm{GELU}(W_1^k \cdot E_{ij})) \\ V_{ij} &= \mathrm{GELU}(W_2^v \cdot \mathrm{GELU}(W_1^v \cdot E_{ij})) \end{aligned} \tag{13}$$

In addition, we incorporate the spatial data of P_i and L_j as additional features, denoted as R_{ij}. Then, R_{ij} is processed through a two-layer linear network to obtain the hidden representation $\phi_{\mathrm{spatial}}(R_{ij})$, which is defined formally in Eq. (14).

$$\phi_{\mathrm{spatial}}(R_{ij}) = W_2^s \cdot \mathrm{ReLU}(W_1^s R_{ij}) \tag{14}$$

For each node, the updated representation is obtained by aggregating the information from its neighbors with the multi-head attention mechanism. Specifically, for each atom p_m in protein P_i, the node set of neighbors in the constructed graph is denoted by $\mathcal{N}(m)$. For each neighbor node $q_n \in \mathcal{N}(m)$, the attention score is computed as follows,

$$\alpha_{mn} = \mathrm{softmax}\left(\frac{Q_{ij}(m,n) \cdot K_{ij}(m,n)}{\sqrt{d_{head}}} + \phi_{\mathrm{spatial}}(R_{ij}(m,n))\right) \tag{15}$$

where d_{head} represents the dimension of attention head. $Q_{ij}(m,n)$, $K_{ij}(m,n)$, and $R_{ij}(m,n)$ correspond to the entry of the mth row and the nth column in Q_{ij}, K_{ij}, and R_{ij}, respectively.

By multiplying attention weight α_{mn} with the corresponding vector V_{mn} and using the aggregation of message passing to accumulate information from each edge, we obtain the updated representation $f'(p_m)$ for node p_m, which is formally defined as follows,

$$f'(p_m) = f(p_m) + \frac{1}{|\mathcal{N}(m)|} \sum_{n \in \mathcal{N}(m)} \alpha_{mn} \cdot V_{mn} \tag{16}$$

in which $\mathcal{N}(m)$ denotes the set of nodes connecting with the node p_m.

Through the same process of graph message passing, the enriched representations of all nodes in the constructed graph are derived accordingly, which are denoted by F_{ij} for the given pair of P_i and L_j.

Graph-level feature aggregation. Given each pair of protein P_i and ligand L_j, the graph-level representation is obtained by a global average pooling operation to integrate the representations of all nodes in the constructed graph. Formally, the graph-level representation H_{ij} is defined as follows,

$$H_{ij} = \sigma(W^{pl} F_{ij} + b^{pl}) \tag{17}$$

where a fully connected layer with the activation function $sigmoid$ is applied to obtain the final protein-ligand complex feature.

3.3. The prediction module

Through the above steps, we have obtained three types of features: protein feature $H(P_i)$, ligand feature $H(L_j)$, and protein-ligand complex feature H_{ij}. In this module, we employ a gated mechanism-based feature fusion method to fully leverage the complementary effect between covalent bond information (contained in $H(P_i)$ and $H(L_j)$) and non-covalent bond information (contained in H_{ij}). Specifically, a gated fusion mechanism is used to generate the discriminative weights of covalent bond features (i.e., $H(P_i)$ and $H(L_j)$) and non-covalent bond feature (i.e., H_{ij}). Formally, the adaptive gating weight of non-covalent bond feature is calculated as shown in Eq. (18).

$$W^{\mathrm{gate}} = \sigma\big(\mathrm{Gate}\big(\mathrm{concat}(H_{ij}, [H(P_i) : H(L_j)])\big)\big) \tag{18}$$

Based on the obtained gating weight, the final representation H^{fused} is obtained by integrating both covalent bond features (i.e., $H(P_i)$ and $H(L_j)$) and non-covalent bond feature (i.e., H_{ij}), which is computed as follows,

$$H^{\mathrm{fused}} = \sigma(W^{\mathrm{gate}} \cdot H_{ij} + (1 - W^{\mathrm{gate}}) \cdot [H(P_i) : H(L_j)]) \tag{19}$$

in which the $sigmoid$ is used as a non-linear activation.

The final representation is fed into a fully connected network to derive the final binding affinity score (defined by Eq. (20)), completing the prediction task.

$$\mathrm{Affinity\_Score} = \mathrm{W}^{\mathrm{fused}} \cdot \mathrm{H}^{\mathrm{fused}} + \mathrm{b}^{\mathrm{fused}} \tag{20}$$

Note that since the range of ground-truth affinity scores is [0, 16], the binding affinity score prediction task is a typical regression problem.

4. Experiments and results

In this section, we conduct comprehensive experiments to evaluate the proposed model, including performance comparison with baseline models, ablation study, cold-start evaluation, and case study.

4.1. Experimental setup

4.1.1. Dataset

In experiments, we use PDBBind+ (Wang et al., 2004) as the training dataset for optimizing our model. PDBBind+[1] consists of numerous

[1] The original dataset is publicly available at https://www.pdbbind-plus.org.cn/download

Table 1
Summary of key hyper-parameter settings for model training.

Hyper-parameter	Value
Model Architecture	
Covalent Feature Dimension	128
Non-covalent Feature Dimension	256
Optimizer & Scheduler	
Optimizer	AdamW
Maximum Learning Rate	3×10^{-4}
Weight Decay (Backbone)	0.01
Weight Decay (Classifier)	0.05
Training Process	
Loss Function	SmoothL1Loss
Training Epochs	20
Batch Size	32

annotated protein-ligand pairs with experimentally measured binding affinity scores. We first perform a filtering process to exclude the samples that could not be read by RDKit (Landrum, 2013). Finally, 14,035 protein-ligand pairs are obtained for training the proposed model. For testing, two commonly used datasets (which are parts of PDBBind+ as well), the 2013Core dataset and 2016Core dataset, are utilized to evaluate our model and baseline models.

4.1.2. Hyper-parameter setting

In this study, the dimensions of the hidden representations for covalent and non-covalent features are set to 128 and 256, respectively. For training our model, we employed a grouped optimization strategy, assigning different learning rates and weight decay values to the backbone networks and the classifier of the model. We selected the AdamW optimizer for its effectiveness in training deep neural networks. The learning rate was dynamically managed using the OneCycleLR scheduler (Smith & Topin, 2019), with a maximum learning rate set to 3×10^{-4} over a training cycle of 20 epochs. This approach is designed to enhance both training efficiency and model performance. For the loss function in our regression task, we utilized SmoothL1Loss (Liu et al., 2021) to ensure a stable optimization process. Furthermore, mixed-precision training, facilitated by GradScaler, was implemented to accelerate computation while maintaining numerical stability. The detailed hyper-parameter settings for the training process are summarized in Table 1.

4.1.3. Baselines and evaluation metrics

For performance comparison, several representative sequence-based and structure-based models are selected as baselines. More specifically, the selected sequence-based baselines include DeepDTAF (Wang et al., 2021), CAPLA (Jin et al., 2023), MFR-DTA (Hua et al., 2023), and DEAttentionDTA (Chen et al., 2024), while the selected six structure-based baselines include GraphDTA (GCN, GAT, GIN, GAT + GCN) (Nguyen et al., 2021), Pafnucy (Stepniewska-Dziubinska et al., 2018), GraphscoreDTA (Wang et al., 2023), ELGN (Yi et al., 2024), EHIGN (Yang et al., 2024), and MMGA (Lv et al., 2024). All baseline models were implemented following the hyper-parameter settings from the original papers.

In experiments, we employed a comprehensive suite of evaluation metrics to rigorously assess the performance of all models. Following the treatment of previous work (de Azevedo Takara et al., 2024; de T. et al., 2025), we employed the following evaluation metrics in our experiments.

The Root Mean Squared Error (RMSE) is utilized to quantify the quadratic mean of the differences between predicted values and the ground-truth values. A lower RMSE indicates a better fit of the model to the data, which is defined formally as follows,

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

where y_i represents the ground-truth value of the ith sample, $\hat{y}_i$ is the predicted value for the ith sample, and n is the total number of samples.

The Mean Absolute Error (MAE) provides a direct measure of the model's average prediction bias by calculating the mean of the absolute differences between predicted and ground-truth values, which is calculated as follows,

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|.$$

Note that unlike RMSE, MAE is less sensitive to large outliers.

To evaluate the linear relationship between the predicted and ground-truth values, we compute the Pearson Correlation Coefficient (PCC) defined as follows,

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the means of the ground-truth and predicted values, respectively.

Finally, the Standard Deviation (SD) of the prediction errors $(y_i - \hat{y}_i)$ is calculated as well to measure the degree of dispersion or variability in the prediction results. A lower SD suggests that the model's predictions are more stable.

4.2. Results and analysis

4.2.1. Overall performance comparison

Table 2 presents the overall performance comparison of our model with baselines on the 2013 core and 2016 core datasets.

Based on the results on the 2013 core dataset, among the sequence-based methods, DeepDTAF shows the worst performance according to all evaluation measures. Although other sequence-based methods like CAPLA, MFR-DTA, and DEAttentionDTA perform reasonably well, their predictive accuracy is generally lower than that of graph neural network-based models, indicating that only using the sequence information can not effectively capture the complex semantics in protein-ligand interactions. Although GraphDTA is a structure-based model, all GraphDTA-type baselines derive undesirable results, with GraphDTA (GCN) performing even worse than DeepDTAF. The possible reason lies in that the features extracted by GraphDTA are insufficient due to the failure of making full use of biological knowledge in the information aggregation of GraphDTA. Other structure-based baselines like ELGN, EHIGN, and MMGA derive good prediction results. Generally, the proposed model MMA-DTA, as a multi-modal fusion model, achieves the best performance according to RMSE (1.198), MAE (1.023), SD (1.231) and PCC (0.894), demonstrating a significant performance improvement over other baseline models. The improvements are attributed to the fact that the covalent features and non-covalent features can capture the fine-grained semantics among protein-ligand interactions, from which more meaningful representations can be obtained for predicting binding affinity scores. In addition, the designed gated mechanism for feature fusion can further advance the representation learning.

According to results on the 2016 core dataset, we can find the similar observations. Specifically, our model MMA-DTA derives the best performance on RMSE of 1.102 and PCC of 0.903. Although the obtained MAE (1.001) and SD (1.221) values are not the best among all models, they are still comparable with the best MAE (0.996) and best SD (1.163), respectively. It is worth noting that while quantitative improvement on certain metrics is narrow (e.g., the improvement over the best baseline MMGA is only 0.2% on RMSE), it is crucial to interpret these results through the lens of statistical robustness. The statistical significance is

Table 2
Results of overall comparison on PDBbind 2013 core set and 2016 core set.

Methods	2013 Core				2016 Core			
	RMSE ↓	MAE ↓	SD ↓	PCC ↑	RMSE ↓	MAE ↓	SD ↓	PCC ↑
DeepDTAF	1.791	1.514	1.780	0.758	1.356	1.075	1.337	0.789
CAPLA	1.446	1.155	1.436	0.770	1.200	**0.996**	1.170	0.843
MFR-DTA	1.412	1.139	1.421	0.774	1.200	1.028	**1.163**	0.832
DEAttentionDTA	1.359	1.102	1.292	0.819	1.224	1.003	1.166	0.845
GraphDTA(GCN)	1.894	1.593	1.868	0.716	1.735	1.343	1.719	0.613
GraphDTA(GAT)	1.745	1.407	1.762	0.739	1.765	1.354	1.740	0.601
GraphDTA(GIN)	1.639	1.456	1.656	0.763	1.640	1.261	1.621	0.667
GraphDTA(GAT-GCN)	1.692	1.365	1.602	0.782	1.562	1.191	1.558	0.697
Pafnucy	1.623	1.318	1.613	0.697	1.423	1.135	1.378	0.774
GraphscoreDTA	1.542	1.236	1.479	0.758	1.349	1.053	1.281	0.810
ELGN	1.398	1.216	1.307	0.723	1.285	1.013	1.263	0.810
EHIGN	1.297	1.103	1.316	0.841	1.150	1.053	1.281	0.854
MMGA	1.214	1.037	1.235	0.887	1.104	1.013	1.266	0.898
MMA-DTA	**1.198**	**1.023**	**1.231**	**0.894**	**1.102**	1.001	1.221	**0.903**

Table 3
Results of t-test on PDBbind 2016 Core Set.

Method	2016 Core (Mean ± Std)			
	RMSE ↓	MAE ↓	SD ↓	PCC ↑
MMGA	1.119 ± 0.012	1.023 ± 0.009	1.280 ± 0.015	0.896 ± 0.005
MMA-DTA	**1.112** ± 0.010	1.009 ± 0.007	1.232 ± 0.012	0.902 ± 0.004

verified by the subsequent t-test, the detailed analysis of which will be presented in the following paragraph.

Since MMGA (Lv et al., 2024) is the best baseline (as shown in Table 2) which also utilizes molecular mechanics-based features, we investigate further the statistical significance of MMA-DTA's performance improvement over the baseline MMGA. Specifically, we conducted 10 independent training runs with different random seeds on the PDBbind 2016 Core set. As shown in Table 3, MMA-DTA consistently outperforms MMGA across all runs according to both RMSE and PCC metrics. Note that the paired t-test provides the statistical validation (with $p < 0.05$) for the results, confirming that the performance advantage–despite being narrow in absolute evaluation values–is statistically significant and not an artifact of random variation. In the real applications of drug discovery, consistency (lower standard deviation and statistical significance) is as valuable as raw accuracy.

The paired t-test on RMSE values yielded a highly significant p-value of less than 0.05 (with $t = 5.29$ and $df = 9$), confirming that the reduction in prediction error is not due to random chance. Similarly, Fisher's z-transformation test on PCC values also showed highly significant improvement with p-value < 0.05 (with $z = 3.62$). These results demonstrate that MMA-DTA's molecular mechanics-aware feature fusion provides statistically significant and robust enhancements in binding affinity prediction accuracy.

4.2.2. Ablation study

To investigate the impact of three different components in our model on the performance of binding affinity prediction, we conduct the ablation study, and the results are shown in Table 4. Note that in Table 4, "w/o non-covalent" denotes the variant without using the non-covalent feature matrix, "w/o covalent" denotes the variant without using the covalent feature matrix, and "w/o gate fusion" denotes the variant without using the gated fusion mechanism but simply concatenating the covalent and non-covalent features for prediction.

Generally, experimental results in Table 4 show that models lacking key components perform poorly across all evaluation metrics, highlighting the contributions of the main components to the performance improvement. Specifically, after removing the non-covalent features, the model's prediction performance declines drastically, with RMSE of 1.675, MAE of 1.326, SD of 1.654, and PCC of 0.647. This observation

Table 4
Results of ablation study.

Method	2016 Core			
	RMSE ↓	MAE ↓	SD ↓	PCC ↑
w/o non-covalent	1.632	1.379	1.638	0.664
w/o covalent	1.197	1.073	1.345	0.848
w/o gate fusion	1.169	1.032	1.276	0.882
MMA-DTA	**1.102**	**1.001**	**1.221**	**0.903**

Table 5
Results of cold start evaluation.

Training Size	2016 Core			
	RMSE ↓	MAE ↓	SD ↓	PCC ↑
10% train-dataset	1.675	1.326	1.654	0.647
30% train-dataset	1.498	1.246	1.463	0.684
50% train-dataset	**1.361**	**1.197**	**1.324**	**0.732**

indicates that the non-covalent features play a critical role in capturing the complex semantics among protein-ligand interactions. In addition, by comparing the performance of "w/o covalent" and "w/o gate fusion" with that of our model MMA-DTA (with all components), we can find that both the covalent features and the gated fusion mechanism are beneficial for learning meaningful representations in our model. The above observations show that the three components (i.e., non-covalent features, covalent features, and the gated fusion mechanism) are all essential for the proposed model MMA-DTA.

4.2.3. Cold-start evaluation

Cold − start is the most challenging setting in drug discovery (Vefghi et al., 2025), in which the associated drugs or targets in the testing set do not exist in the training set. In this section, we test the performance of our model under the cold-start scenario to evaluate the robustness and generalization ability of our model. Specifically, we conduct experiments on the PDBbind2020+ refined dataset, mimicking the strategy used for KIBA (Tang et al., 2014). Moreover, we evaluate our model on three different ratios (i.e., 10%, 30%, and 50%) of training set to investigate the influence of training set size on the model's prediction performance. The results are shown in Table 5.

From results in Table 5, we can find that the proposed model performs well in the "Cold-start" scenario. For example, even training with only 10% of the training data, our model still maintains predictive stability, evidenced by the Standard Deviation (SD) of 1.654 and the RMSE of 1.675. And the obtained performance is comparable with the performance obtained by GraphDTA variants trained on the whole training data (as shown in Table 2). The results demonstrate that even in data-sparse scenarios, the proposed molecular mechanics-aware feature

Table 6
Binding analysis of EAI045 with three EGFR variants.

EGFR Variant	Predicted Affinity	van der Waals		Electrostatic	
		total energy	strong interactions	charge complementarity	strong interactions
Wild-type	3.8	−1,362,998	3498	0.508	7145
L858R mutant	7.3	−1,641,162	5769	0.674	7769
Triple mutant	**8.1**	**-4,405,564**	**24,899**	**0.689**	**8726**

fusion mechanism can still effectively capture key information related to protein-ligand binding. Moreover, another observation is that MMA-DTA's prediction performance increases consistently as the size of training set increases. For example, when the ratio of training set increases from 10% to 50%, the RMSE and MAE respectively drop to 1.361 and 1.197, suggesting a clearly improved prediction performance. Note that, results according to the other two metrics (SD and PCC) demonstrate a similar pattern. Overall, MMA-DTA provides a feasible solution to applications in the "Cold-start" scenario, and its performance can be advanced with increasing size of training data.

4.2.4. Case study

In this section, we conduct the case study to validate the effectiveness of our model from two distinct perspectives: the ligand-centric viewpoint and the target-centric viewpoint. In both cases, the binding affinity of protein-ligand complexes serves as the primary evaluation metric.

From viewpoint of ligand-centric. EAI045 (Wang et al., 2017) is an allosteric inhibitor of the epidermal growth factor receptor (EGFR). Note that EAI045 is specifically designed to target mutants of drug-resistant EGFR (Jiang et al., 2023), while it exhibits minimal activity against wild EGFR. EAI045 performs in the non-covalent manner, forming highly specific conformational interactions at a conserved allosteric site to inhibit EGFR kinase activity. In this study, we test the binding affinity of EAI045 with three protein variants obtained from the RCSB Protein Data Bank (Rose et al., 2016), including the wild EGFR, the L858R mutant EGFR, and the L858R/T790M/C797S triple mutant EGFR (triple mutant for short), and the results are shown in Table 6.

Results in Table 6 show that the binding affinity of the ligand EAI045 with two EGFR variants is drastically greater than that with the wild EGFR, and the binding affinity of EAI045 with the triple mutants variant EGFR is the greatest among all proteins. This observation is consistent with results of previous experiments (Wang et al., 2017), indicating the effectiveness of our model for binding affinity prediction. To investigate further the details of binding process, we calculate the molecular mechanics interaction energy, encompassing both van der Waals and electrostatic components. As shown in Table 6, the triple mutant complex demonstrates the most favorable interaction profile, exhibiting the lowest van der Waals energy (-4,405,564), the greatest number of strong van der Waals interactions (24,899), superior electrostatic complementarity (0.689), and the greatest number of strong electrostatic interactions (8,726). These detailed factors explain the significantly enhanced binding strength of EAI045 to the triple mutants EGFR, which is consistent with experimental binding assays. The above analysis validates the contribution of non-covalent interactions in the binding affinity prediction, suggesting further the predictive power of our model accordingly.

From viewpoint of target-centric. Beta-secretase (BACE1) is a transmembrane aspartyl protease that initiates the generation of $A\beta$-amyloid ($A\beta$) by cleaving the amyloid precursor protein (APP) (Sathya et al., 2012). And the aberrant accumulation of $A\beta$ constitutes a core pathological hallmark of Alzheimer's disease (AD) (Azargoonjahromi, 2024). Since its discovery, BACE1 has been recognized as a rate-limiting enzyme in $A\beta$ production, and thus serves as a critical therapeutic target for AD (Yan & Vassar, 2014). Current research efforts focus on the development of highly selective BACE1 inhibitors, aiming to achieve precise intervention during the early stages of AD (Xia et al., 2025). In this study, we test the binding affinity of BACE1 with different ligands. More specifically, we select five positive ligands which are effective BACE1 inhibitors, including Cyclohexyl-QA carboxamide, Imatinib derivatives (denoted by $5HA$), LY2886721, LY2817228, AZD3839 (Rose et al., 2016). Note that all five ligands are not included in the training set. Additionally, EAI045 (Wang et al., 2017) is employed as a negative control for comparison. The results of predicted binding affinity are presented in Table 7.

Table 7
Predicted binding affinity of selected ligands with BACE1.

Ligand	True Affinity	Predicted Affinity
Cyclohexyl-QA carboxamide	7.52	7.89
5HA	7.82	8.03
LY2886721	7.70	8.09
LY2817228	8.01	8.48
AZD3839	8.00	8.36
EAI045	–	3.12

From Table 7, we can find that the predicted binding affinities (shown in the 3rd column) of the five positive ligands with BACE1 are really close to the corresponding true values (shown in the 2nd column), demonstrating our model's strong generalization capability to unseen pairs of ligands and targets. Furthermore, as a negative control ligand, the predicted binding affinity of EAI045 with the same target BACE1 is only 3.12, which is significantly lower and consistent with its inability to effectively bind the target BACE1.

In addition, we take the positive ligand $5HA$ as an example to further investigate the binding with BACE1. More specifically, we use PyMOL (DeLano et al., 2002) to extract the binding sites of BACE1 and 5HA, and the visualization is shown in Fig. 2. Moreover, we calculate the average van der Waals and electrostatic interaction energies for the entire protein-ligand complex as well as for the binding sites. For each binding site or residue, the ratio of its interaction strength to the overall average strength of non-covalent interactions is used to quantify the non-covalent interaction contribution. The results are shown in Fig. 3. Note that the density of color bar denotes the relative energy ratio which is calculated by normalizing the residue-specific interaction energy against the average interaction energy of the complex.

From Fig. 3, we can find that the non-covalent interaction energy ratios of the binding sites in the BACE1-5HA complex exhibit considerable variation compared to the overall protein-ligand interaction energy. The ratios of van der Waals (vdW) energy reveal that residues such as A/VAL332, A/ILE226, and A/TYR198 contribute prominently to the type of non-covalent interactions, with energy ratios approaching the maximum scale of 10. This suggests that these residues play a pivotal role in maintaining the binding stability through strong hydrophobic interactions, facilitating a tight spatial fit within the binding pocket. As for the electrostatic energy, the results show that the binding residues A/ILE226, A/SER229, and A/TYR198 exhibit significant contributions with electrostatic ratios exceeding 2.0. These findings indicate their critical involvement in charge complementarity and polar interactions at the binding interface. Note that the significant contributions of the binding residues mentioned above can also be observed in Fig. 2. This detailed analysis verifies the model's ability to learn non-covalent interactions between protein-ligand complexes and provides support for accurate binding affinity prediction.

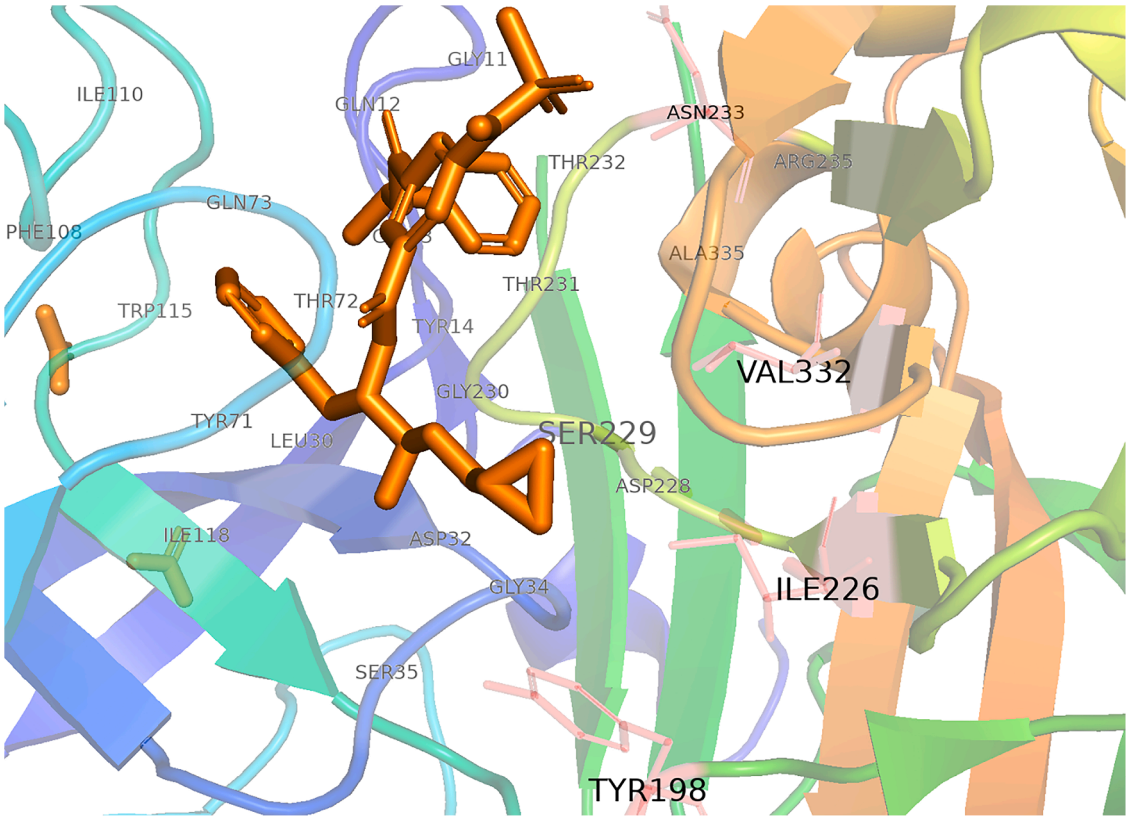

Fig. 2. Visualization of the binding sites between BACE1 and 5HA.

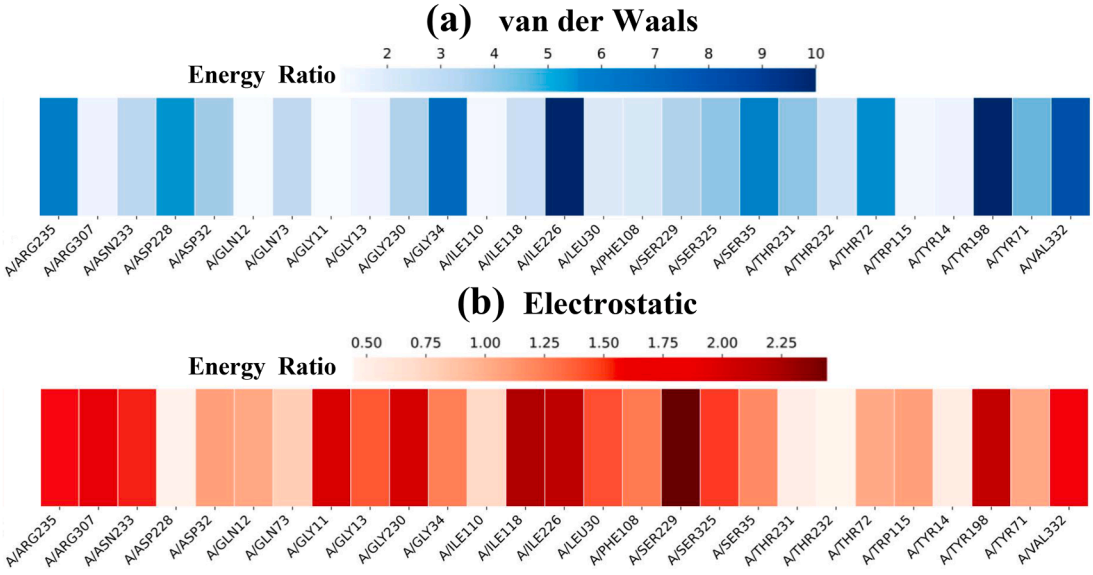

Fig. 3. The ratios of the non-covalent interaction energy of the binding sites of the BACE1-5HA pair: (1) van der Waals, (2) Electrostatic.

4.3. Limitations

Although MMA-DTA derives desirable performance in this study, it still has the following limitations.

Firstly, MMA-DTA has limited capability to capture sequence information, which requires further improvement to obtain more meaningful representations.

Secondly, constrained by the accuracy of force field-based molecular mechanics simulations, we obtained a relatively coarse molecular mechanics numerical matrix, which might influence the performance of the down-stream binding affinity prediction. Although quantum chemistry could be used to advance the accuracy of molecular mechanics information between protein and ligand, it is extremely resource-intensive and time-intensive (Ginex et al., 2024; Molani & Cho, 2024), which is the main reason we did not adopt it in this study.

Thirdly, MMA-DTA's predictive performance is sensitive to the quality of the input structure. The reason lies in that the calculation of molecular mechanics energies (Eqs. (4)–(6)) is intrinsically influenced by the

resolution of the input crystal structures (PDB files). For non-covalent interaction features, even minor coordinate inaccuracies in low-resolution structures can lead to disproportionate energy fluctuations.

Fourthly, although the results of the case study demonstrated the effectiveness of the proposed model, the experiments are conducted in a computational or simulated manner. The real binding affinity of ligand-target pairs still needs to be evaluated by wet-lab experiments, which needs to be conducted by collaborating with biochemical or biomedical labs.

5. Conclusion and future work

In this study, we have proposed a multimodal molecular mechanics-aware feature fusion model, MMA-DTA, for predicting protein-ligand binding affinity. By leveraging multiple molecular mechanics-aware features, MMA-DTA is able to effectively capture the complex interactions between proteins and ligands. More specifically, the covalent features are processed by using Highway Network and CNN, while the non-covalent features are processed by using GraphTransform. Finally, both types of features are dynamically fused through a gating mechanism, and the prediction of binding affinity is completed accordingly. The experimental results on PDBbind show that MMA-DTA outperforms several selected benchmark models according to multiple evaluation measures, achieving the best performance on both 2013 core dataset (with RMSE 1.198, MAE 1.023, SD 1.231 and PCC 0.894) and 2016 core dataset (with RMSE 1.102 and PCC 0.903). The results of ablation study further verify the effectiveness of the main components in the proposed model, especially the non-covalent features for capturing the complex semantics among protein-ligand interactions. The results of cold-start evaluation demonstrate that our model has the potential for addressing the challenge of “Cold-start” applications. In case study, MMA-DTA is able to accurately quantify the interaction energies of mutant complexes, which provides a foundation for interpreting drug resistance mechanisms and guiding the design of novel ligands. Therefore, MMA-DTA can play a significant role in real-world virtual screening workflows, in which it can be utilized to increase the hit rate of lead compounds, accelerating the drug discovery iteration cycle.

In future work, we plan to incorporate sequence information as additional features to derive more meaningful representations of proteins and ligands. In addition, we aim to generalize our framework to other real-world applications, such as drug-target interaction (DTIs) prediction and protein-protein interaction (PPI) prediction.

CRediT authorship contribution statement

Xing Lv: Investigation, Writing – original draft; **Xinhui Tu:** Visualization, Writing – review & editing; **Tingting He:** Visualization, Writing – review & editing; **Weizhong Zhao:** Conceptualization, Methodology, Investigation, Writing – original draft.

Data availability

Data will be made available on request.

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.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (62472192) and (62372205), the Self-determined Research Funds of CCNU from the Colleges’ Basic Research and Operation of MOE (CCNU25JC008). Authors are grateful to the anonymous reviewers for helpful comments.

References

Ahmed, W. W., Fodor, É., Almonacid, M., Bussonnier, M., Verlhac, M.-H., Gov, N., Visco, P., van Wijland, F., & Betz, T. (2018). Active mechanics reveal molecular-scale force kinetics in living oocytes. *Biophysical Journal*, *114*(7), 1667–1679.

Amaral, M., Kokh, D. B., Bomke, J., Wegener, A., Buchstaller, H. P., Eggenweiler, H. M., Matias, P., Sirrenberg, C., Wade, R. C., & Frech, M. (2017). Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature Communications*, *8*(1), 2276.

Azargoonjahromi, A. (2024). The duality of amyloid-β: Its role in normal and alzheimer’s disease states. *Molecular Brain*, *17*(1), 44.

Cai, W., Liu, Y., Lin, X., Li, Z., Zhang, Y., & Newth, D. (2024). Nonlinear country-heterogenous impact of the indian ocean dipole on global economies. *Nature Communications*, *15*(1), 5009.

Callaway, E. (2020). ’It will change everything’: Deepmind’s AI makes gigantic leap in solving protein structures. *Nature*, *588*(7837), 203–205.

Celej, M. S., Montich, G. G., & Fidelio, G. D. (2003). Protein stability induced by ligand binding correlates with changes in protein flexibility. *Protein Science*, *12*(7), 1496–1506.

Chen, X., Huang, J., Shen, T., Zhang, H., Xu, L., Yang, M., Xie, X., Yan, Y., & Yan, J. (2024). Deattentiondta: Protein–ligand binding affinity prediction based on dynamic embedding and self-attention. *Bioinformatics*, *40*(6), btae319.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, *117*(19), 5179–5197.

da Silva, L. S. A., Seman, L. O., Camponogara, E., Mariani, V. C., & dos Santos Coelho, L. (2024). Bilinear optimization of protein structure prediction: An exact approach via AB off-lattice model. *Computers in Biology and Medicine*, *176*, 108558.

de Azevedo Takara, L., Teixeira, A. C., Yazdanpanah, H., Mariani, V. C., & dos Santos Coelho, L. (2024). Optimizing multi-step wind power forecasting: integrating advanced deep neural networks with stacking-based probabilistic learning. *Applied Energy*, *369*, 123487.

de, T. G. R., Yamaguchi, C. K., Stefenon, S. F., dos, S. C. L., & Mariani, V. C. (2025). Deep learning approach for automatic heartbeat classification. *Sensors*, *25*(5), 1400.

DeLano, W. L. et al. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter Protein Crystallography*, *40*(1), 82–92.

Demtröder, W. (2008). Molecular physics: Theoretical principles and experimental methods. John Wiley & Sons.

Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., & Liu, S.-Q. (2016). Insights into protein–ligand interactions: Mechanisms, models, and methods. *International Journal of Molecular Sciences*, *17*(2), 144.

Fan, J., Fu, A., & Zhang, L. (2019). Progress in molecular docking. *Quantitative Biology*, *7*, 83–89.

Ferguson, D. M., & Kollman, P. A. (1991). Can the Lennard–Jones 6–12 function replace the 10–12 form in molecular mechanics calculations? *Journal of Computational Chemistry*, *12*(5), 620–626.

Ginex, T., Vázquez, J., Estarellas, C., & Luque, F. J. (2024). Quantum mechanical-based strategies in drug discovery: Finding the pace to new challenges in drug design. *Current Opinion in Structural Biology*, *87*, 102870.

Halgren, T. A. (1996). Merck molecular force field. i. basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, *17*(5-6), 490–519.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, *99*(6), 1129–1143.

Hua, Y., Song, X., Feng, Z., & Wu, X. (2023). MFR-DTA: A multi-functional and robust model for predicting drug–target binding affinity and region. *Bioinformatics*, *39*(2), btad056.

Janbaz-amirani, A., Majrashi, T. A., Alshehri, S. A., Farhadian, S., Shareghi, B., Alghamdi, S. S., Atiya, A., & Gholizadeh, M. (2025). Fatty acid-whey protein interactions: A multi-spectroscopic and computational study. *International Dairy Journal*, *166*, 106250.

Jiang, M., Shao, Y., Zhang, Y., Zhou, W., & Pang, S. (2023). A deep learning method for drug-target affinity prediction based on sequence interaction information mining. *PeerJ*, *11*, e16625.

Jin, Z., Wu, T., Chen, T., Pan, D., Wang, X., Xie, J., Quan, L., & Lyu, Q. (2023). Capla: Improved prediction of protein–ligand binding affinity by a deep learning approach based on a cross-attention mechanism. *Bioinformatics*, *39*(2), btad049.

Katznelson, Y. (2004). An introduction to harmonic analysis. Cambridge University Press.

Landrum, G. (2013). Rdkit documentation. *Release*, *1*(1–79), 4.

Lechelon, M., Meriguet, Y., Gori, M., Ruffenach, S., Nardecchia, I., Floriani, E., Coquillat, D., Teppe, F., Mailfert, S., Marguet, D. et al. (2022). Experimental evidence for long-distance electrodynamic intermolecular forces. *Science Advances*, *8*(7), eabl5855.

Leckband, D., & Israelachvili, J. (2001). Intermolecular forces in biology. *Quarterly Reviews of Biophysics*, *34*(2), 105–267.

Lequin, R. M. (2005). Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical Chemistry*, *51*(12), 2415–2418.

Li, S., Zhou, J., Xu, T., Huang, L., Wang, F., Xiong, H., Huang, W., Dou, D., & Xiong, H. (2023). Giant: Protein-ligand binding affinity prediction via geometry-aware interactive graph neural network. *IEEE Transactions on Knowledge and Data Engineering*, *36*(5), 1991–2008.

Liu, C., Yu, S., Yu, M., Wei, B., Li, B., Li, G., & Huang, W. (2021). Adaptive smooth l1 loss: A better way to regress scene texts with extreme aspect ratios. In *2021 IEEE Symposium on computers and communications (ISCC)* (pp. 1–7). IEEE.

Lv, X., Zhao, W., Tu, X., & He, T. (2024). Predicting protein-ligand binding affinity via molecular mechanics-guided graph aggregation. In *2024 IEEE International conference on bioinformatics and biomedicine (BIBM)* (pp. 1072–1077). IEEE.

Molani, F., & Cho, A. E. (2024). Accurate protein-ligand binding free energy estimation using QM/MM on multi-conformers predicted from classical mining minima. *Communications Chemistry*, *7*(1), 247.

Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, *37*(8), 1140–1147.

Nilofer, C., Sukhwal, A., Mohanapriya, A., & Kangueane, P. (2017). Protein-protein interfaces are vdw dominant with selective h-bonds and (or) electrostatics towards broad functional specificity. *Bioinformation*, *13*(6), 164.

Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., & Aittokallio, T. (2015). Toward more realistic drug–target interaction predictions. *Briefings in Bioinformatics*, *16*(2), 325–337.

Polishchuk, P. G., Madzhidov, T. I., & Varnek, A. (2013). Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of Computer-Aided Molecular design*, *27*, 675–679.

Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., Costanzo, L. D., Duarte, J. M., Dutta, S., Feng, Z. et al. (2016). The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Research*, (p. gkw1000).

Sathya, M., Premkumar, P., Karthick, C., Moorthi, P., Jayachandran, K. S., & Anusuyadevi, M. (2012). Bace1 in Alzheimer's disease. *Clinica Chimica Acta*, *414*, 171–178.

Smith, L. N., & Topin, N. (2019). Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications* (pp. 369–386). SPIE (*vol. 11006*).

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. arXiv preprint arXiv:1505.00387, .

Stepniewska-Dziubinska, M. M., Zielenkiewicz, P., & Siedlecki, P. (2018). Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics*, *34*(21), 3666–3674.

Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K., & Aittokallio, T. (2014). Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *Journal of Chemical Information and Modeling*, *54*(3), 735–743.

Tang, X., Ma, W., Yang, M., & Li, W. (2024a). MFF-DTA: Multi-scale feature fusion for drug-target affinity prediction. *Methods*, *231*, 1–7.

Tang, X., Zhou, Y., Yang, M., & Li, W. (2024b). TC-DTA: Predicting drug–target binding affinity with transformer and convolutional neural networks. *IEEE Transactions on NanoBioscience*, *23*, 572–578.

Vefghi, A., Rahmati, Z., & Akbari, M. (2025). Drug-target interaction/affinity prediction: Deep learning models and advances review. arXiv preprint arXiv:2502.15346.

Velázquez-Campoy, A., Ohtaka, H., Nezami, A., Muzammil, S., & Freire, E. (2004). Isothermal titration calorimetry. *Current Protocols in Cell Biology*, *23*(1), 8–17.

Wang, K., Zhou, R., Li, Y., & Li, M. (2021). DeepDTAF: A deep learning method to predict protein–ligand binding affinity. *Briefings in Bioinformatics 22*(5), bbab072.

Wang, K., Zhou, R., Tang, J., & Li, M. (2023). GraphscoreDTA: Optimized graph neural network for protein–ligand binding affinity prediction. *Bioinformatics*, *39*(6), btad340.

Wang, R., Fang, X., Lu, Y., & Wang, S. (2004). The PDBbind database: Collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, *47*(12), 2977–2980.

Wang, S., Song, Y., & Liu, D. (2017). Eai045: The fourth-generation EGFR inhibitor overcoming t790m and c797s resistance. *Cancer Letters*, *385*, 51–54.

Xia, C., Feng, S.-H., Xia, Y., Pan, X., & Shen, H.-B. (2023). Leveraging scaffold information to predict protein–ligand binding affinity with an empirical graph neural network. *Briefings in Bioinformatics*, *24*(1), bbac603.

Xia, L., Li, J., Pang, Y., Dai, C., Xu, M., Du, Y., Tian, Q., Yi, L., Wu, B., Chen, M. et al. (2025). Disruption of BAG3-mediated BACE1 stabilization alleviates neuropathology and memory deficits in a mouse model of alzheimer's disease. *Science Advances 11*(21), eadt7981.

Yamamoto, M. (2002). Surface plasmon resonance (SPR) theory: Tutorial. *Review of Polarography*, *48*(3), 209–237.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, *9*, 611–629.

Yan, R., & Vassar, R. (2014). Targeting the β secretase BACE1 for alzheimer's disease therapy. *The Lancet Neurology*, *13*(3), 319–329.

Yang, Y.-q., Gao, Q., Yue, S.-q., Peng, X., Wang, N., Xin, J.-l., Yu, M., Rao, J.-j., & Xue, Y.-l. (2025). Investigating the interaction mechanisms between arachin and resveratrol: Utilizing multi-spectroscopy and computational chemistry. *Food Chemistry*, *463*, 141435.

Yang, Z., Zhong, W., Lv, Q., Dong, T., Chen, G., & Chen, C. Y.-C. (2024). Interaction-based inductive bias in graph neural networks: Enhancing protein-ligand binding affinity predictions from 3d structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*, 8191–8208.

Yi, Y., Wan, X., Zhao, K., Ou-Yang, L., & Zhao, P. (2024). Equivariant line graph neural network for protein-ligand binding affinity prediction. *IEEE Journal of Biomedical and Health Informatics*, *28*, 4336–4347.

Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph transformer networks. *Advances in Neural Information Processing Systems*, *32*, 11983-11993.

Zhao, L., Wang, H., & Shi, S. (2024). PocketDTA: An advanced multimodal architecture for enhanced prediction of drug- target affinity from 3D structural data of target binding pockets. *Bioinformatics*, *40*(10), btae594.