Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

GRAPH NEURAL NETWORKS FOR KINASE–INHIBITOR AFFINITY PREDICTION USING DOCKING SCORES AND BINDING-SITE FEATURES

Michael Turner[1], Sophia Nguyen[2]*, David Clark[1], Emma Wilson[3]

1. ***Department of Computational Pharmaceutical Sciences, Faculty of Pharmacy, School of Biomedical Sciences, University of Glasgow, Glasgow, United Kingdom.***
2. ***Department of Artificial Intelligence in Drug Discovery, Faculty of Pharmaceutical Sciences, National University of Singapore, Singapore.***
3. ***Department of Pharmaceutical Data Analytics, Faculty of Medicine and Health, University of Sydney, Sydney, Australia.***

ARTICLE INFO

Received:
09 September 2024
Received in revised form:
19 November 2024
Accepted:
23 November 2024
Available online:
28 December 2024

Keywords: Graph neural networks, Kinase inhibitors, Binding affinity prediction, Molecular docking, Binding-site features, Virtual screening

ABSTRACT

Kinases are major therapeutic targets, yet accurately predicting true inhibitor binding affinity remains a significant challenge. While molecular docking is useful for generating binding poses, docking scores alone often fail to capture the full complexity of kinase–inhibitor recognition. Current predictive models frequently overlook the spatial graph of the protein–ligand interface and the physicochemical characteristics of the binding site, limiting their effectiveness in compound ranking and reasoning about kinase selectivity. To address these limitations, this article proposes a graph neural network model that integrates kinase–inhibitor complex structures, docking-derived scores, and binding-site descriptors, aiming to provide an interpretable framework for affinity prediction that conceptually surpasses docking-only and ligand-only approaches. The model represents ligand atoms and kinase binding-site residues as graph nodes, with spatial contacts and docking-derived interaction terms encoded as edges, while pocket descriptors are incorporated as residue-level or graph-level features to convey local chemical context. This design enables the model to learn interaction patterns beyond what a single docking score can capture and to identify residues and ligand substructures that contribute most strongly to predicted affinity. By combining structure-based scoring with interpretable deep learning, a docking-informed graph neural network has the potential to advance kinase drug discovery, provided it is rigorously evaluated using kinase-specific benchmarks and prospective validation strategies.

To Cite This Article: Turner M, Nguyen S, Clark D, Wilson E. Graph Neural Networks for Kinase–Inhibitor Affinity Prediction Using Docking Scores and Binding-Site Features. Pharmacophore. 2024;15(6):15-23. https://doi.org/10.51847/VepyukPdnR

Introduction

Protein kinases are central therapeutic targets because they regulate signaling pathways involved in cancer, inflammation, metabolism, and cell-cycle control. Predicting kinase–inhibitor affinity is difficult because many inhibitors compete for the conserved ATP-binding pocket, where small structural differences can strongly affect potency and selectivity. Deep learning approaches for kinase binding prediction have therefore begun to move beyond ligand-only similarity and toward target-aware models that represent kinase-specific binding behavior [1]. Kinome-wide profiling models also show why affinity prediction should consider selectivity across related kinases rather than treating each target as isolated.

Molecular docking remains a practical tool for virtual screening because it generates binding poses and approximate scores for large compound libraries. However, docking scores are simplified estimates and may fail when kinase flexibility, induced-fit effects, water networks, or pocket-state changes dominate binding. Deep learning scoring functions such as structure-based convolutional models were introduced partly to address these limitations by learning from protein–ligand complexes rather than relying only on fixed scoring terms [2, 3]. GNINA further illustrates how docking and deep learning can be combined so that pose generation and learned scoring support each other rather than functioning as separate steps [4].

Graph neural networks provide a natural framework for kinase–inhibitor modeling because both molecules and binding sites can be represented as structured graphs. PotentialNet demonstrated how molecular graph learning can encode chemically

Corresponding Author: Sophia Nguyen; Department of Artificial Intelligence in Drug Discovery, Faculty of Pharmaceutical Sciences, National University of Singapore, Singapore. E-mail: sophia.nguyen@gmail.com.

meaningful interactions for property prediction [5], while graph-based drug–target models extended this idea to compound–protein interaction prediction [6]. GraphDTA showed that representing compounds as graphs can improve drug–target affinity modeling compared with sequence-only or fingerprint-only representations [7]. The remaining opportunity is to combine graph-based structure learning with docking-derived signals and explicit kinase binding-site features in one model.

The central thesis of this MDL article is that a graph neural network integrating docking scores and binding-site descriptors could provide a more informative kinase–inhibitor affinity predictor. In this framework, docking-derived values would act as structured prior knowledge rather than final affinity estimates, while pocket descriptors would supply local kinase context. InteractionGraphNet supports the value of representing protein–ligand contacts through graph-based interaction learning [8], and structure-aware interactive graph models show how complex-level graph representations can support affinity prediction [9]. Such a model would also be useful for medicinal chemistry if its attention or attribution maps identified interpretable kinase residues and ligand substructures [10].

Background

Kinase Drug Discovery and Molecular Docking

Kinase drug discovery is complicated by the structural conservation of the ATP-binding site and the diversity of active, inactive, and intermediate kinase conformations. Docking can propose plausible inhibitor poses in these pockets, but its scoring functions may not reliably distinguish strong from weak binders when selectivity depends on subtle residue-level differences. Machine learning scoring functions have been reviewed as a way to learn target-sensitive corrections to docking scores instead of accepting docking ranks directly [11]. Kinase-specific deep learning models reinforce this need because selectivity prediction depends on both inhibitor chemistry and target-pocket context [1].

Graph Representations of Protein–Ligand Complexes

Protein–ligand complexes can be encoded as graphs in which ligand atoms, protein atoms, or binding-site residues are nodes, and covalent bonds or spatial contacts are edges. This is useful for kinase inhibitors because hinge contacts, gatekeeper-region interactions, hydrophobic-pocket occupancy, and solvent-front exposure can be represented locally. Torng and Altman used graph convolutional neural networks to model drug–target interactions from graph-structured representations [6], while Tsubaki, Tomii and Sese connected compound graphs with protein sequence representations for interaction prediction [12]. Complex-level graph representations go further by encoding the geometry of the bound structure directly [8, 9].

Graph Neural Networks for Binding Affinity Prediction

Graph neural networks learn node embeddings through message passing, allowing local chemical and structural information to propagate across the graph. In affinity prediction, this is important because binding is determined by distributed interaction patterns rather than isolated ligand properties. GraphDTA used molecular graphs for drug–target binding affinity prediction [7], and MONN modeled compound–protein interactions using a multi-objective neural framework [13]. Later graph and interaction-aware affinity models further developed this direction by combining protein–ligand structural information with learned graph representations [14, 15].

Binding-Site Descriptors and Pocket Features

Binding-site descriptors encode information such as pocket volume, residue composition, hydrophobicity, electrostatic character, solvent accessibility, and shape complementarity. These features are especially relevant for kinases because selectivity may depend on small changes in the hinge region, gatekeeper position, back pocket, or activation-loop state. P2Rank demonstrated that machine learning can identify ligand-binding pockets from protein structure using local geometric and physicochemical features [16]. DeepSite similarly showed that three-dimensional deep learning can detect protein-binding sites from structural environments, supporting the use of pocket-aware descriptors in downstream affinity models [17].

Integrating Docking Scores into Deep Learning

Docking scores can provide useful but incomplete information about estimated interaction strength, steric complementarity, and pose plausibility. Instead of using a docking score as the final prediction, a deep model can treat it as one feature among many and learn when it should be trusted or corrected. KDEEP used three-dimensional convolutional learning for protein–ligand affinity prediction from complex structures [2], and GNINA demonstrated that deep learning can be embedded directly into docking workflows [4]. Reviews of structure-based deep learning scoring functions also emphasize that learned models can complement classical docking by capturing empirical interaction patterns [18, 19].

Model Development Overview

High-Level Pipeline

The proposed pipeline begins with a kinase–inhibitor three-dimensional complex obtained from crystallography, structural modeling, or docking. A graph construction module then converts the complex into ligand nodes, binding-site residue nodes, intramolecular edges, and protein–ligand contact edges. The GNN processes this graph and outputs a predicted affinity value, while optional attention or attribution modules produce residue-level and ligand-substructure explanations. Similar structure-

aware pipelines have been explored in protein–ligand scoring models such as Pafnucy [20], KDEEP [2], and graph-based interaction frameworks [8].

Figure 1 presents the proposed docking-informed graph neural network architecture for integrating kinase–inhibitor complex structures, docking-derived interaction signals, binding-site descriptors, affinity prediction, and interpretable medicinal chemistry outputs.

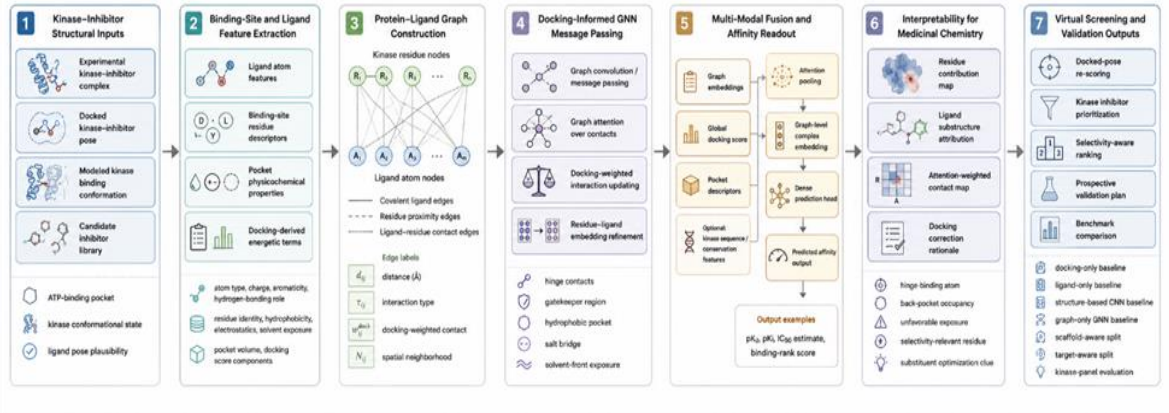

Figure 1. Docking-Informed Graph Neural Network Architecture for Kinase–Inhibitor Affinity Prediction.

Core Input Features

Ligand atom nodes would encode atom type, aromaticity, hybridization, charge, hydrogen-bonding role, and local topology. Binding-site residue nodes would encode amino-acid identity, physicochemical class, solvent exposure, conservation, secondary-structure context, and pocket membership. Edge features would include covalent connectivity, ligand–residue distances, interaction categories, and docking-derived energetic terms. This design follows the broader logic of graph-based molecular learning in PotentialNet [5] and protein–ligand interaction graph modeling in InteractionGraphNet [8].

Design Principles

The model should be end-to-end learnable while preserving chemically meaningful structure-based inputs. It should work with both experimental complexes and docked poses because prospective virtual screening usually depends on docked structures, whereas training data may include crystallographic complexes. PIGNet illustrates the value of incorporating physics-informed interaction ideas into deep learning for drug–target prediction [15], while PIGNet2 extends this concept toward broader binding affinity scoring and virtual screening use [21]. Interpretability should also be designed into the architecture so that model explanations can be connected to medicinal chemistry hypotheses rather than remaining abstract.

Data Sources and Graph Construction

Training and Benchmark Datasets

Training data would conceptually include general protein–ligand affinity resources and kinase-specific affinity benchmarks. General datasets are useful for learning broad interaction patterns, while kinase-focused subsets are necessary for evaluating selectivity and ATP-pocket recognition. DeepDTA and WideDTA provide examples of benchmark-driven drug–target affinity modeling using curated compound–protein affinity data [22, 23]. DeepDTAF further shows how affinity models can combine compound and protein information in a deep learning framework that should be evaluated carefully on target-aware splits [24].

Graph Construction from 3D Complexes

Graph construction would define the kinase binding site as residues within a spatial neighborhood of the ligand, with optional inclusion of known ATP-pocket selectivity regions. Ligand atoms would be connected by covalent bonds, residue nodes by spatial or sequence proximity, and ligand–residue pairs by interaction-distance edges. OnionNet used contact shells to represent protein–ligand interactions in affinity prediction [25], and OnionNet-2 refined this idea using residue–atom contacting shells [26]. These contact-based approaches support the use of explicit local interaction structure when constructing graphs for kinase–inhibitor complexes.

Table 1 defines how ligand atoms, kinase binding-site residues, spatial contacts, docking-derived terms, and pocket descriptors can be encoded as complementary graph components in the proposed affinity prediction framework.

Table 1. Graph Construction and Feature Encoding Strategy for Docking-Informed Kinase–Inhibitor Affinity Prediction

Model component	Structural unit represented	Primary encoded features	Docking-informed information	Kinase-specific relevance	Expected modeling contribution
Ligand atom nodes	Individual inhibitor atoms	Atom type, formal charge, aromaticity, hybridization, hydrogen-bond	Local docking contribution	Captures substituent-level effects that influence potency and selectivity	Enables the model to learn which ligand substructures drive

		donor/acceptor role, local topology	summaries assigned to ligand atoms		affinity or reduce compatibility
Binding-site residue nodes	Kinase residues within the ligand-contact neighborhood	Amino-acid identity, physicochemical class, secondary-structure context, conservation, solvent exposure, pocket membership	Residue-level docking interaction summaries and contact-weighted energetic descriptors	Represents hinge residues, gatekeeper position, back pocket, solvent front, activation-loop region, and selectivity-relevant pocket variation	Allows target-aware learning beyond ligand-only similarity
Ligand covalent edges	Bonds between ligand atoms	Bond type, conjugation, ring membership, local chemical topology	Not usually docking-derived, but linked to pose-dependent ligand orientation	Preserves chemical structure of inhibitor scaffolds and substituent connectivity	Supports propagation of chemical information across the ligand graph
Residue proximity edges	Spatial or sequence-neighbor relationships among kinase binding-site residues	Residue distance, sequence adjacency, structural neighborhood, pocket-region identity	Optional docking-weighted residue proximity when contacts influence pose stability	Captures local pocket geometry and conformational arrangement of kinase binding regions	Helps encode pocket shape, residue clustering, and spatial context
Ligand–residue contact edges	Noncovalent contacts between ligand atoms and kinase residues	Interatomic distance, contact type, hydrogen bond, hydrophobic contact, electrostatic interaction, π-stacking, steric clash	Edge-level docking terms, interaction scores, pose confidence, contact-specific energy proxies	Directly models ATP-pocket recognition, hinge binding, hydrophobic-pocket occupancy, and solvent-front exposure	Provides the main interface through which the GNN learns structure–affinity relationships
Pocket-level graph descriptors	Global kinase binding-site environment	Pocket volume, hydrophobicity, electrostatics, residue composition, solvent accessibility, shape complementarity	Global docking score, docking rank, pose plausibility, estimated binding energy	Distinguishes kinase pockets with similar docking scores but different local chemical environments	Supplies contextual information for affinity readout and selectivity-aware prediction
Optional kinase sequence/conservation features	Kinase family or residue-position annotations	Kinase family identity, conserved motifs, variable pocket residues, alignment-derived features	Not docking-derived, but can contextualize docking contacts across related kinases	Supports kinome-wide selectivity reasoning and transfer across kinase families	Improves generalization when structural data are limited or kinase families are related
Graph-level complex embedding	Full kinase–inhibitor complex	Learned ligand, residue, contact, docking, and pocket representations	Integrated docking-informed representation after message passing and pooling	Summarizes the complete kinase–inhibitor binding event	Feeds the final affinity predictor and supports compound prioritization

Docking Score Integration as Edge/Node Features

Docking scores would be incorporated into the graph as structured features rather than appended only as a single pose-level scalar. Ligand atom nodes could receive local docking contribution summaries, residue nodes could receive interaction-weighted docking descriptors, and ligand–residue edges could encode contact-specific scoring terms. Hybrid models that combine molecular docking with graph neural networks and physics-based scoring functions support the idea that docking information can guide learned protein–ligand prediction [27]. AK-score also illustrates the broader value of learned structure-based scoring for affinity estimation from protein–ligand complexes [28].

Graph Neural Network Architecture

Graph Convolution / Attention Layers

The proposed architecture would use graph convolutional, graph attention, or message-passing layers to update ligand and binding-site embeddings. These layers would allow information from kinase residues, ligand atoms, and docking-weighted contacts to interact over multiple message-passing steps. A graph attention mechanism would be particularly useful because the model could learn to focus on hinge interactions, hydrophobic-pocket occupancy, salt bridges, or solvent-front contacts depending on the inhibitor and kinase conformation. Graph attention affinity models and nested interpretable GNNs demonstrate why attention-based interaction learning is relevant for drug–target prediction [10, 29].

Global Pooling and Affinity Readout

After message passing, the model would aggregate node embeddings into a graph-level complex representation using attention-based or interaction-aware pooling. This pooled representation would then be combined with global docking scores and pocket-level descriptors before passing through dense layers to output predicted affinity. DeepBindGCN provides an example of integrating molecular representation with graph convolutional learning for protein–ligand interaction prediction [30]. Fusion-

based models that combine graph neural networks with three-dimensional protein–ligand interaction features also support this readout strategy [31].

Loss Function and Training Protocol
The training objective would conceptually use affinity regression and could be extended with ranking or selectivity-aware losses for kinase panels. Data augmentation could include alternative docked poses, randomized rotations, or conformational variants, but such strategies should be evaluated rather than assumed to improve performance. Structure-based scoring studies caution that validation must avoid overestimating generalization when closely related ligands or targets appear across training and test partitions [11, 18]. The model should therefore be assessed with scaffold-aware, target-aware, and time-aware splits when used for kinase inhibitor discovery.

Integrating Docking Scores and Binding-Site Features
Docking Scores as Injected Knowledge
Docking scores would be treated as injected structural knowledge rather than as final affinity predictions. In the proposed model, docking-derived terms could guide attention toward plausible ligand–residue contacts while still allowing the GNN to correct scoring errors caused by simplified energy functions. GNINA demonstrates that docking workflows can benefit from learned scoring components [4], while hybrid docking, graph neural network, and physics-based scoring approaches show how docking information can be combined with learned interaction representations [27]. Reviews of structure-based deep learning scoring further support the view that docking features are most useful when embedded within models that learn empirical corrections from protein–ligand structures [11, 18, 19].

Pocket Descriptors as Global Context
Binding-site descriptors would provide global and residue-level context for the kinase pocket, helping the model distinguish complexes that may have similar docking scores but different local environments. Pocket volume, hydrophobicity, electrostatic character, solvent exposure, and residue composition could be concatenated with the graph-level embedding or assigned directly to binding-site residue nodes. P2Rank supports the relevance of local geometric and physicochemical features for ligand-binding pocket characterization [16], and DeepSite shows that three-dimensional structural environments can be learned for binding-site recognition [17]. Sequence-search and representation tools such as MMseqs2 also support conservation-aware residue annotation, which could help identify kinase pocket positions that are preserved or variable across related targets [32].

Multi-Modal Fusion and Pre-Training
A multi-modal fusion strategy would combine ligand graph embeddings, residue graph embeddings, docking-derived edge features, pocket descriptors, and optional protein sequence representations. Pre-trained sequence or protein-family embeddings could initialize kinase residue features, while the structure-based GNN would refine them using local ligand contacts and docking-informed interactions. This approach is conceptually aligned with drug–target affinity models that combine compound and protein information, such as DeepDTA, WideDTA, and DeepDTAF [22–24]. It also complements graph and sequence interaction approaches in which compound graphs and protein representations are learned jointly for compound–protein interaction prediction [12, 33].
Table 2 shows the components of the proposed multi-modal fusion strategy for integrating ligand, protein, and docking-informed representations in drug–target affinity prediction frameworks.

Table 2. Multi-modal fusion components for ligand–protein interaction modeling and docking-informed representation learning

Modality	Representation / Input	Role in Fusion	Example Methods / Notes
Ligand graph embeddings	Molecular graph (atoms as nodes, bonds as edges)	Encodes chemical structure and substructure-level features of the ligand	GNNs such as GCN, GAT, MPNN
Residue graph embeddings	Protein residue-level graph (residues as nodes, spatial edges)	Captures 3D structural context of binding pocket residues	Structure-based GNNs, contact maps
Docking-derived edge features	Docking poses, interaction scores, distance/contact constraints	Injects spatial and energetic priors into ligand–protein interaction modeling	Docking scores, interaction fingerprints (IFPs)
Pocket descriptors	Physicochemical pocket features (hydrophobicity, charge, shape)	Provides global binding site characterization	Geometric deep learning, descriptor-based encodings
Protein sequence embeddings	Amino acid sequence representations	Provides sequence-level biological context when structure is incomplete	CNN/RNN/Transformer-based encoders
Pre-trained protein-family embeddings	Evolutionary or family-aware embeddings (e.g., kinase-specific features)	Initializes protein representations with biologically informed priors	ESM, ProtBert, or kinase-specific pretraining models
Fusion mechanism	Cross-attention / concatenation / gated fusion	Integrates heterogeneous modalities into a unified interaction representation	Multi-head attention, multimodal GNN fusion layers

Model Interpretability for Medicinal Chemistry

Attention-Based Interaction Mapping

Attention-based interaction mapping would allow the model to identify kinase residues and ligand substructures that contribute most strongly to the predicted affinity. In a kinase context, such maps could highlight hinge-binding atoms, gatekeeper-adjacent interactions, back-pocket occupancy, salt bridges, or solvent-front contacts that medicinal chemists already interpret during lead optimization. Graph attention affinity models provide a natural mechanism for assigning higher weights to important contacts [29], while interpretable nested GNN approaches show how drug–target predictions can be linked to explanatory graph components [10]. InteractionGraphNet and structure-aware interactive graph models further support the idea that learned protein–ligand interaction graphs can produce mechanistically meaningful signals rather than only graph-level scores [8, 9].

From Prediction to Decision Support

The model's explanations would be most useful if they connected affinity predictions to medicinal chemistry decisions, such as whether a substituent improves hinge engagement, fills a hydrophobic pocket, or introduces an unfavorable solvent-exposed contact. Instead of ranking compounds only by predicted affinity, the workflow could present each compound with a predicted binding rationale based on residue-level and ligand-substructure contributions. Kinase-focused deep learning models motivate this decision-support role because selectivity depends on target-specific structural differences as much as on ligand potency [1]. Kinome-wide profiling and selectivity prediction platforms further suggest that interpretable outputs could help chemists reason across kinase panels rather than optimizing a single target in isolation.

Deployment and Integration into Virtual Screening Workflows

Inference Pipeline for Docked Pose Re-Scoring

In deployment, the trained GNN would operate as a re-scoring model for docked kinase–inhibitor poses generated by standard virtual screening workflows. Docking would provide candidate poses and preliminary scores, while the GNN would re-evaluate each complex using ligand topology, kinase pocket context, docking-derived edge features, and learned protein–ligand interaction patterns. GNINA illustrates how learned scoring can be integrated into docking-centered workflows [4], and structure-based deep learning reviews emphasize that learned scoring functions are best evaluated as complements to docking rather than as isolated predictors [18, 19]. A kinase-specific implementation would therefore be expected to support prioritization of compounds for follow-up testing, while requiring prospective validation before any claim of screening improvement.

High-Throughput Screening Mode

For high-throughput screening, the full structure-aware model could be approximated through staged inference in which inexpensive ligand- or docking-based filters precede the GNN re-scoring step. A lighter surrogate model could also be trained to imitate selected components of the full GNN when speed is more important than interpretability, while the complete model remains available for final prioritization and explanation. Deep affinity models such as DeepDTA, WideDTA, and DeepDTAF illustrate how lower-cost compound–protein representations can support broad screening before more detailed structural analysis [22–24]. Learned structure-based scoring models, including AK-score and docking-aware neural frameworks, provide complementary examples of how complex-based predictors may fit into virtual screening pipelines [4, 28].

Evaluation Strategy

Affinity Prediction Metrics

Affinity prediction should be evaluated using regression and rank-correlation metrics, but these metrics should be reported only after careful separation of related ligands, targets, and time periods. Scaffold-aware and target-aware splits would be particularly important for kinase inhibitors because close analogues can otherwise make a model appear more generalizable than it actually is. Pafnucy, KDEEP, and protein–ligand CNN scoring studies provide examples of structure-based affinity evaluation on protein–ligand complexes [2, 3, 20]. GraphDTA, MONN, and DeepDTAF further show why affinity models should be assessed under validation settings that reflect compound–target generalization rather than only interpolation within familiar chemical series [7, 13, 24].

Ranking and Selectivity Assessment

Ranking evaluation should examine whether the model can prioritize more plausible kinase inhibitors within docked libraries and distinguish selective from broadly active compounds across kinase panels. Selectivity assessment would require comparisons across related kinases, because an inhibitor that binds one ATP pocket strongly may still be undesirable if it is predicted to engage many off-target kinases. Calibrated geometric deep learning for kinase–drug binding prediction supports the importance of kinase-aware evaluation [1], and large-scale comparisons of machine learning methods for kinase inhibitor profiling show why panel-level prediction is central to selectivity modeling. KinomePro-DL further illustrates the need for models that estimate small-molecule kinome selectivity profiles rather than single-target affinity alone.

Comparison with Docking and Other Deep Learning Baselines

The proposed model should be compared conceptually against docking-only scores, ligand-only affinity predictors, protein–sequence drug–target models, three-dimensional convolutional scoring functions, and graph-only GNNs that do not include docking or pocket descriptors. Such comparisons would clarify whether docking-derived edge features and binding-site descriptors add value beyond learned ligand or complex representations. KDEEP and AK-score provide relevant three-dimensional structure-based baselines [2, 28], while OnionNet and OnionNet-2 offer contact-shell alternatives for protein–ligand affinity modeling [25, 26]. Graph-based and interaction-aware baselines such as PotentialNet, GraphDTA, InteractionGraphNet, and structure-aware interactive graph networks would also be important comparators for isolating the contribution of the proposed docking-informed kinase pocket design [5, 7–9].

Table 3 consolidates the validation, ablation, interpretability, and deployment criteria needed to determine whether the proposed docking-informed GNN is genuinely useful for kinase inhibitor discovery.

Table 3. Evaluation and Deployment Readiness Framework for Docking-Informed GNN Affinity Prediction

Evaluation domain	Key question addressed	Recommended assessment strategy	Essential comparators	Main risk controlled	Practical interpretation for kinase drug discovery
Affinity regression accuracy	Does the model predict continuous kinase–inhibitor affinity values reliably?	Report RMSE, MAE, Pearson correlation, and Spearman correlation on kinase-relevant test sets	Docking-only scores; ligand-only models; sequence-based drug–target models; graph-only GNNs	Overclaiming predictive accuracy from simple interpolation	Determines whether the model improves affinity estimation beyond classical docking and non-structural predictors
Scaffold-aware generalization	Does the model generalize to chemically distinct inhibitor scaffolds?	Split compounds by chemical scaffold before training and testing	Ligand fingerprint models; GraphDTA-style compound graph models; docking score baselines	Inflated performance caused by close analogue leakage	Tests whether the model can support discovery of new inhibitor chemotypes
Target-aware generalization	Does the model transfer across related and less familiar kinases?	Hold out kinase targets or kinase families during validation	Protein-sequence affinity models; kinase-specific profiling models; graph-only complex models	Inflated performance caused by target overlap	Evaluates whether the model can reason across kinase families rather than memorizing individual targets
Pose-dependence analysis	How sensitive are predictions to docking pose quality?	Compare crystallographic poses, top-ranked docked poses, alternative docked poses, and perturbed poses	Docking-only ranking; GNINA-like learned scoring; structure-based CNN scoring	Misleading predictions from incorrect binding modes	Clarifies whether the GNN improves scoring only when poses are structurally plausible
Docking-feature ablation	Do docking-derived edge and node features add value?	Train variants with and without global docking scores, local docking terms, and docking-weighted contact edges	Same GNN architecture without docking features	Unclear contribution of docking information	Determines whether docking acts as useful structured prior knowledge rather than redundant input
Pocket-descriptor ablation	Do binding-site descriptors improve kinase specificity?	Remove pocket volume, hydrophobicity, electrostatics, solvent exposure, and residue-composition descriptors in controlled experiments	Graph-only protein–ligand models; ligand-only models	Failure to isolate value of kinase pocket context	Shows whether binding-site features help distinguish similar inhibitors across related kinases
Selectivity-aware ranking	Can the model prioritize compounds across kinase panels?	Evaluate ranking performance across multiple kinases using enrichment metrics, panel-level selectivity profiles, and off-target prediction	Docking-only panel ranking; kinome-wide profiling models; single-target affinity predictors	Optimizing potency without accounting for off-target kinase binding	Tests usefulness for medicinal chemistry programs where selectivity is as important as potency
Interpretability validity	Are residue and ligand attribution maps chemically meaningful?	Compare attention or attribution outputs with known hinge contacts, gatekeeper interactions, hydrophobic-pocket occupancy, and structure–activity relationships	Attention-free GNNs; black-box neural scoring functions; docking interaction diagrams	Producing explanations that are visually plausible but chemically weak	Determines whether model explanations can guide substituent optimization and lead prioritization
Prospective screening value	Does the model improve real virtual screening outcomes?	Re-score a held-out or prospective docked compound library and measure experimental hit enrichment	Standard docking workflow; learned docking score; ligand-based virtual screening	Benchmark success that does not translate to laboratory prioritization	Establishes whether the framework has practical value for kinase inhibitor discovery
Reproducibility and reporting	Can other researchers evaluate and	Release code, data splits, preprocessing rules, graph-construction settings, model	Published structure-based scoring and GNN benchmarks	Irreproducible model claims and hidden benchmark leakage	Supports transparent adoption and comparison across kinase families and inhibitor chemotypes

extend the framework?	hyperparameters, and trained checkpoints

Limitations

Pose-Dependence and Docking Quality

The model's predictions would depend strongly on the quality of the input kinase–inhibitor pose. If docking fails to generate a realistic binding mode, the GNN may learn from misleading ligand–residue contacts even if its architecture is expressive. This limitation is especially relevant for kinases because activation-loop movement, DFG-state variation, side-chain rearrangement, and water-mediated interactions can alter the pocket geometry seen by the inhibitor. Docking-aware neural scoring methods and reviews of structure-based deep learning emphasize that learned models can improve scoring workflows, but they do not remove the need for reliable structural inputs and pose validation [4, 11, 18, 19].

Data Scarcity for Atypical Kinases

Many kinases and inhibitor classes have limited experimentally measured affinity data, which could restrict the model's ability to generalize to atypical pockets, rare conformational states, or under-studied kinase families. Transfer learning, multi-task learning, and data augmentation could help, but these strategies should be evaluated carefully rather than assumed to solve data imbalance. Compound–protein interaction models that combine graph, sequence, and affinity information suggest possible routes for learning from broader datasets before kinase-specific fine-tuning [12, 22–24, 33]. Kinase-focused models and profiling platforms also show that selectivity prediction depends on broad and reliable kinase-panel information, which may not be uniformly available across the kinome [1].

Conclusion

A graph neural network that integrates docking scores and binding-site features offers a conceptually strong framework for kinase–inhibitor affinity prediction. By representing ligand atoms, kinase residues, spatial contacts, and docking-derived interaction information in a unified graph, the model could learn richer structure–affinity relationships than docking scores or ligand fingerprints alone.

The main strength of the proposed approach is its structure-aware and interpretable design. It could support virtual screening by re-scoring docked poses, while also helping medicinal chemists understand which residues and ligand substructures drive the predicted affinity.

Important challenges remain before such a model could be used confidently in drug discovery campaigns. Pose quality, kinase conformational diversity, water-mediated interactions, data imbalance, and benchmark leakage would all need careful treatment through rigorous validation and transparent reporting.

Open-source implementation would help the community test, reproduce, and extend the proposed framework across kinase families and inhibitor chemotypes. Integration into standard docking workflows could make the model more useful for practical lead discovery, provided that prospective validation confirms its value in real screening and optimization settings.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Luo Y, Liu Y, Peng J. Calibrated geometric deep learning improves kinase–drug binding predictions. Nat Mach Intell. 2023;5(12):1390-401.
2. Jiménez J, Skalic M, Martinez-Rosell G, De Fabritiis G. K deep: protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model. 2018;58(2):287-96.
3. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein–ligand scoring with convolutional neural networks. J Chem Inf Model. 2017;57(4):942-57.
4. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. J Cheminform. 2021;13(1):43.
5. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y, et al. PotentialNet for molecular property prediction. ACS Cent Sci. 2018;4(11):1520-30.
6. Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions. J Chem Inf Model. 2019;59(10):4131-49.

7. Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140-7.
8. Jiang D, Hsieh CY, Wu Z, Kang Y, Wang J, Wang E, et al. Interactiongraphnet: a novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. J Med Chem. 2021;64(24):18209-32.
9. Li S, Zhou J, Xu T, Huang L, Wang F, Xiong H, et al. Structure-aware interactive graph neural networks for the prediction of protein–ligand binding affinity. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21). New York: Association for Computing Machinery; 2021. p. 975-85.
10. Sun Y, Li YY, Leung CK, Hu P. iNGNN-DTI: prediction of drug–target interaction with interpretable nested graph neural network and pretrained molecule models. Bioinformatics. 2024;40(3):btae135.
11. Pellicani F, Dal Ben D, Perali A, Pilati S. Machine learning scoring functions for drug discovery from experimental and computer-generated protein–ligand structures: towards per-target scoring functions. Molecules. 2023;28(4):1661.
12. Tsubaki M, Tomii K, Sese J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2019;35(2):309-18.
13. Li S, Wan F, Shu H, Jiang T, Zhao D, Zeng J. MONN: a multi-objective neural network for predicting compound-protein interactions and affinities. Cell Syst. 2020;10(4):308-22.
14. Son J, Kim D. Development of a graph convolutional neural network model for efficient prediction of protein-ligand binding affinities. PLoS One. 2021;16(4):e0249404.
15. Moon S, Zhung W, Yang S, Lim J, Kim WY. PIGNet: a physics-informed deep learning model toward generalized drug–target interaction predictions. Chem Sci. 2022;13(13):3661-73.
16. Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J Cheminform. 2018;10(1):39.
17. Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics. 2017;33(19):3036-42.
18. Meli R, Morris GM, Biggin PC. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: a review. Front Bioinform. 2022;2:885983.
19. Wang Z, Wang S, Li Y, Guo J, Wei Y, Mu Y, et al. A new paradigm for applying deep learning to protein–ligand interaction prediction. Brief Bioinform. 2024;25(3):bbae145.
20. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. Bioinformatics. 2018;34(21):3666-74.
21. Moon S, Hwang SY, Lim J, Kim WY. PIGNet2: a versatile deep learning-based protein–ligand interaction prediction model for binding affinity scoring and virtual screening. Digit Discov. 2024;3(2):287-99.
22. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics. 2018;34(17):i821-9.
23. Öztürk H, Ozkirimli E, Özgür A. WideDTA: prediction of drug-target binding affinity. arXiv. 2019:arXiv:1902.04166.
24. Wang K, Zhou R, Li Y, Li M. DeepDTAF: a deep learning method to predict protein–ligand binding affinity. Brief Bioinform. 2021;22(5):bbab072.
25. Zheng L, Fan J, Mu Y. Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction. ACS Omega. 2019;4(14):15956-65.
26. Wang Z, Zheng L, Liu Y, Qu Y, Li YQ, Zhao M, et al. OnionNet-2: a convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells. Front Chem. 2021;9:753002.
27. Hong Y, Ha J, Sim J, Lim CJ, Oh KS, Chandrasekaran R, et al. Accurate prediction of protein–ligand interactions by combining physical energy functions and graph-neural networks. J Cheminform. 2024;16(1):121.
28. Kwon Y, Shin WH, Ko J, Lee J. AK-score: accurate protein-ligand binding affinity prediction using an ensemble of 3D-convolutional neural networks. Int J Mol Sci. 2020;21(22):8424.
29. Yuan H, Huang J, Li J. Protein-ligand binding affinity prediction model based on graph attention network. Math Biosci Eng. 2021;18(6):9148-62.
30. Zhang H, Saravanan KM, Zhang JZ. DeepBindGCN: integrating molecular vector representation with graph convolutional neural networks for protein–ligand interaction prediction. Molecules. 2023;28(12):4691.
31. Dong L, Shi S, Qu X, Luo D, Wang B. Ligand binding affinity prediction with fusion of graph neural networks and 3D structure-based complex graph. Phys Chem Chem Phys. 2023;25(35):24110-20.
32. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026-8.
33. Ahmed A, Mam B, Sowdhamini R. DEELIG: a deep learning approach to predict protein-ligand binding affinity. Bioinform Biol Insights. 2021;15:11779322211030364.