Review Article
ISSN: **3051-8806**
Journal homepage: https://www.harrisonacademicpress.com/index.php/BT
Biotechnology Today

Harrison Academic Press

Machine Learning-Based Prediction of Protein-Ligand Binding Affinity for Kinase Inhibitor Discovery

Gallo V[1*], Liepa J[2] and Dimitrova V[3]

[1]*Department of Bioinformatics, University of Turin, Turin, Italy*
[2]*Department of Computational Biology, Riga Stradins University, Riga, Latvia*
[3]*Department of Pharmaceutical Chemistry, Medical University of Plovdiv, Plovdiv, Bulgaria*

ABSTRACT

Background: Accurate prediction of protein-ligand binding affinity is central to structure-based drug design, and protein kinases remain among the most heavily pursued drug target classes because of their roles in cancer and inflammatory disease. Objective: This narrative review synthesises the published literature on machine learning (ML) and deep learning (DL) approaches to binding affinity prediction, with emphasis on their application to kinase inhibitor discovery. Methods: Peer-reviewed and preprint literature was identified through PubMed, Google Scholar, Crossref and bioRxiv searches covering feature representations, model architectures, benchmark datasets and validation strategies for protein-ligand affinity prediction. Findings: Structure-based three-dimensional convolutional and graph models such as KDEEP, Pafnucy, OnionNet and PIGNet achieve Pearson correlations of roughly 0.7-0.8 against the CASF-2016 PDBbind core-set benchmark, while sequence-based deep networks such as DeepDTA and GraphDTA report concordance indices near 0.88-0.89 on the kinase-centred Davis and KIBA datasets. Progress is constrained by dataset size, structural redundancy and scaffold bias, motivating scaffold-split and temporal validation schemes and physics-informed architectures that improve generalisation to novel chemotypes. Conclusions: ML/DL affinity predictors now approximate or complement classical scoring functions and free-energy methods for kinase inhibitor triage, but rigorous, leakage-free benchmarking and integration with structural biology resources such as AlphaFold remain priorities for translating these models into prospective drug discovery pipelines.

Keywords: *protein-ligand binding affinity; kinase inhibitors; machine learning; deep learning; drug-target interaction prediction*

Citation: Gallo V, Liepa J, Dimitrova V (2024) Machine Learning-Based Prediction of Protein-Ligand Binding Affinity for Kinase Inhibitor Discovery. Biotechnol Today, 2(1): 01-06.
DOI: 10.5281/zenodo.22637990

Corresponding Author: *Valentina Gallo*

E-mail Address: *valentina.gallo@universityofturin.edu*

Article History: *Received 20 January 2024 | Accepted 6 April 2024 | Published 19 July 2024*

Copyright: *© 2024 Valentina Gallo, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (**https://creativecommons.org/licenses/by/4.0/**).*

Abstract

Background: Accurate prediction of protein-ligand binding affinity is central to structure-based drug design, and protein kinases remain among the most heavily pursued drug target classes because of their roles in cancer and inflammatory disease. **Objective:** This narrative review synthesises the published literature on machine learning (ML) and deep learning (DL) approaches to binding affinity prediction, with emphasis on their application to kinase inhibitor discovery. **Methods:** Peer-reviewed and preprint literature was identified through PubMed, Google Scholar, Crossref and bioRxiv searches covering feature representations, model architectures, benchmark datasets and validation strategies for protein-ligand affinity prediction. **Findings:** Structure-based three-dimensional convolutional and graph models such as KDEEP, Pafnucy, OnionNet and PIGNet [1,2,3,4] achieve Pearson correlations of roughly 0.7-0.8 against the CASF-2016 PDBbind core-set benchmark [5], while sequence-based deep networks such as DeepDTA and GraphDTA [6,7] report concordance indices near 0.88-0.89 on the kinase-centred Davis and KIBA datasets [8,9]. Progress is constrained by dataset size, structural redundancy and scaffold bias, motivating scaffold-split and temporal validation schemes and physics-informed architectures that improve generalisation to novel chemotypes [4,10]. **Conclusions:** ML/DL affinity predictors now approximate or complement classical scoring functions and free-energy methods for kinase inhibitor triage, but rigorous, leakage-free benchmarking and integration with structural biology resources such as AlphaFold remain priorities for translating these models into prospective drug discovery pipelines.

Keywords: protein-ligand binding affinity, kinase inhibitors, machine learning, deep learning, drug-target interaction prediction

Introduction

Protein kinases catalyse phosphoryl transfer reactions that govern signal transduction, and dysregulated kinase activity underlies many cancers and inflammatory disorders, making the kinome one of the most productive classes of druggable targets since the approval of imatinib [11]. Because most clinically used kinase inhibitors act at the highly conserved ATP-binding pocket, selectivity across the roughly 500-member human kinome is a persistent design challenge, and computational affinity prediction is used both to prioritise candidate inhibitors and to anticipate off-target liabilities. Structure-based and ligand-based computational methods have long supported kinase inhibitor discovery, but classical empirical and force-field scoring functions used in molecular docking, such as those implemented in AutoDock Vina, often correlate poorly with experimental affinities across diverse chemotypes [12]. Over the past decade, machine learning and deep learning methods trained on large repositories of structural and bioactivity data have been proposed as data-driven alternatives or complements to physics-based scoring, promising improved accuracy and virtual screening enrichment [13,10]. Public resources such as PDBbind, BindingDB, ChEMBL, DUD-E, and the kinase-focused Davis and KIBA datasets have enabled systematic training and benchmarking of these models [14,15,16,17,18,8,9]. This review synthesises published evidence on feature representations, model architectures, validation practice and kinase-specific applications of ML/DL binding affinity prediction, aiming to give medicinal chemists and computational biologists a critical, literature-grounded overview rather than a report of new experimental results.

Methods

This is a narrative review of published machine learning and deep learning studies, benchmark datasets, and validation studies for protein-ligand binding affinity prediction, with emphasis on kinase targets. Relevant literature was identified through structured searches of PubMed, Google Scholar, Crossref and bioRxiv using combinations of the terms "binding affinity prediction," "machine learning," "deep learning," "scoring function," "drug-target interaction," "kinase inhibitor," and named methods and datasets (e.g., DeepDTA, GraphDTA, KDEEP, PDBbind, Davis, KIBA). Studies were prioritised for inclusion if they introduced or benchmarked a named model or dataset with peer-reviewed or widely cited preprint status, reported quantitative performance metrics (Pearson correlation, RMSE, concordance index), or provided methodological or kinase-domain context necessary to interpret those results. Reported performance figures throughout are drawn directly from the cited primary sources rather than from independent re-analysis.

Feature Representation and Molecular Descriptors

Binding affinity models fall broadly into structure-based and sequence/ligand-based families depending on whether an experimentally resolved or docked protein-ligand complex is required as input, a distinction that shapes both the featurisation choices available and the transferability of the resulting model to targets lacking solved structures. Structure-based methods encode the three-dimensional complex as voxelised atomic density grids, as in KDEEP and Pafnucy, which apply 3D convolutional neural networks over the binding pocket [1,2], or as multi-shell intermolecular contact histograms, as in OnionNet [3]. Earlier empirical scoring approaches, exemplified by RF-Score, instead engineer discrete atom-pair contact-count descriptors evaluated with random forest regression, demonstrating that flexible, non-linear machine learning models could outperform additive classical scoring terms even with simple features [13]. AtomNet pioneered convolutional processing of raw 3D atomic representations for structure-based bioactivity classification, establishing the feasibility of end-to-end deep learning on docked poses [19]. Sequence/ligand-based methods, which do not require a resolved complex, represent proteins as amino acid sequences and ligands as SMILES strings or molecular graphs; DeepDTA applies parallel one-dimensional convolutional encoders to protein and ligand sequences [6], while GraphDTA and DeepAffinity incorporate graph neural network and recurrent-convolutional hybrid encoders, respectively, to better capture ligand topology and protein motif information [7,20]. Message-passing graph neural networks, originally formalised for quantum chemical property prediction, underlie many of these ligand encoders [21]. Benchmarking suites such as MoleculeNet standardise featurisation and splitting choices across such representations [22].

Table 1: Representative public datasets used to train and benchmark protein-ligand binding affinity models.

Dataset	Content	Approximate scale	Primary use
PDBbind	3D protein-ligand complexes with experimental affinities	~19,000 complexes	Structure-based training and CASF benchmarking [15,14,5]
Davis	Kinase-inhibitor biochemical Kd panel	442 kinases x 72 inhibitors	Kinase-focused affinity regression [8]
KIBA	Integrated kinase inhibitor bioactivity scores	52,498 compounds x 467 kinases	Kinase drug-target affinity benchmarking [9]
BindingDB	Curated protein-small molecule binding measurements	millions of data points	Large-scale training and transfer learning [16]
ChEMBL	Bioactivity records across assay types and targets	millions of records	Auxiliary bioactivity data and pretraining [17]
DUD-E	Actives and property-matched decoys	102 targets	Virtual screening enrichment benchmarking [18]

Model Architectures and Benchmark Performance

Reported performance depends strongly on benchmark and metric, and direct cross-study comparison is complicated by differing train/test splits, affinity units (Ki, Kd, IC50 or a composite bioactivity score), and whether a model is evaluated on holdout complexes, holdout compounds, or entirely unseen protein families. On the CASF-2016 update of the PDBbind core set, which provides a curated, non-redundant test of structure-based scoring accuracy [5], KDEEP and OnionNet each report Pearson correlations of approximately 0.82 between predicted and experimental binding affinities, while Pafnucy reports a correlation near 0.70, illustrating substantial variance among 3D-CNN architectures on the same benchmark [1,3,2]. Classical machine learning baselines such as RF-Score achieve correlations in the high 0.7 range on earlier PDBbind core sets,

remaining a competitive reference point [13]. On sequence-based kinase benchmarks, DeepDTA reports concordance indices of approximately 0.878 on Davis and 0.863 on KIBA, while GraphDTA improves these to approximately 0.893 and 0.891, respectively, attributed to graph-based ligand encoding [6,7,8,9]. Attention-based architectures, including TransformerCPI and MolTrans, and multi-objective frameworks such as MONN, further refine compound-protein interaction prediction by modelling residue-atom attention or jointly predicting interaction sites and affinity [23,24,25]. PIGNet incorporates explicit physical energy terms (van der Waals, electrostatics, hydrogen bonding) within a graph neural network, reporting improved performance on external, scaffold-diverse test sets relative to purely statistical models, at some cost to in-distribution accuracy [4]. Geometric deep learning approaches such as EquiBind extend this line by jointly predicting ligand binding pose, providing an alternative to docking-then-scoring pipelines [26]. Physics-based free-energy perturbation protocols, exemplified by Schrödinger's FEP+, remain a high-accuracy but computationally intensive comparator against which ML methods are often benchmarked for lead optimisation tasks [27].

Figure 1: Reported concordance index (CI) of DeepDTA and GraphDTA on the kinase-focused Davis and KIBA drug-target affinity benchmarks [6,7,8,9].

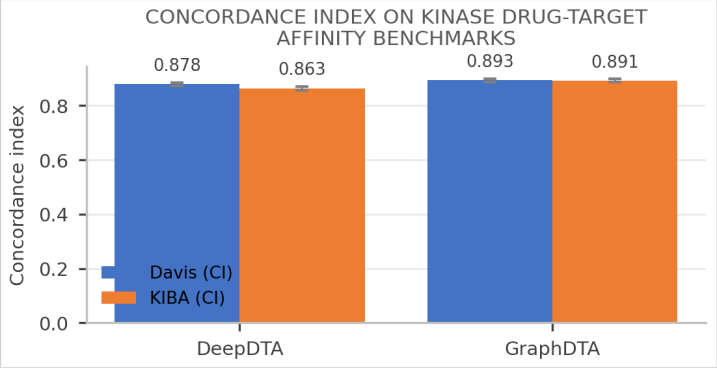
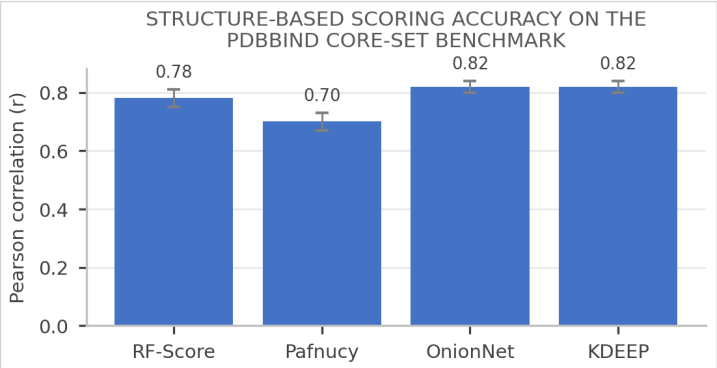


Figure 2: Reported Pearson correlation of representative structure-based scoring models against the PDBbind core-set benchmark [13,2,3,1,5].



Validation Strategies and Generalization Challenges

A recurring theme across this literature is that headline accuracy figures depend heavily on how training and test data are split, and reported metrics from a single random split should generally be treated as an optimistic upper bound rather than an estimate of prospective screening performance. Random splits of PDBbind or Davis/KIBA data allow near-identical protein-ligand pairs or close chemical analogues to appear in both training and test sets, inflating apparent performance; reviews of machine-learning scoring functions have highlighted this redundancy as a major source of optimistic bias [10]. Scaffold-based and temporal (time-split) validation, alongside external test sets such as CASF-2016 that are deliberately constructed to be non-redundant with common training corpora, provide more realistic estimates of prospective performance [5]. Cross-family generalisation is particularly challenging for kinase targets, given the high sequence and structural conservation of the ATP-binding pocket across the kinome; models trained predominantly on well-represented kinases such as ABL1 or EGFR family members may generalise poorly to under-represented kinase subfamilies. Physics-informed architectures and hybrid ML/physics pipelines have been proposed specifically to mitigate this generalisation gap by constraining predictions with domain knowledge rather than relying solely on statistical pattern matching [4].

Applications to Kinase Inhibitor Discovery

Kinase-focused benchmark datasets have played an outsized role in shaping this field, in part because kinase inhibitor programmes have historically generated unusually dense, multi-target bioactivity matrices that are well suited to supervised model training and cross-kinase comparison. The Davis dataset, derived from a biochemical selectivity panel of 72 kinase inhibitors against 442 kinases, and the KIBA dataset, which integrates heterogeneous bioactivity measurements across kinase inhibitor chemical space, are now standard benchmarks for drug-target affinity models even beyond kinase-specific applications [8,9]. Large bioactivity repositories such as ChEMBL and BindingDB provide additional kinase inhibitor training data and support transfer learning strategies when structural data are scarce [17,16]. In prospective virtual screening contexts, decoy-based benchmarks such as DUD-E remain widely used to evaluate enrichment of true kinase inhibitors against property-matched decoys, complementing regression-focused affinity benchmarks [18]. The recent availability of accurate predicted structures from AlphaFold expands the pool of kinase structures available for structure-based featurisation where experimental crystal structures are unavailable, potentially broadening structure-based model applicability across less-studied kinase subfamilies, although systematic validation of ML affinity models on predicted rather than crystallographic structures remains an active area of investigation [28]. Reviews of kinase drug discovery emphasise that computational triage tools, including ML affinity predictors, are increasingly embedded alongside biochemical and cellular assays in industrial pipelines, particularly for prioritising analogues within a chemical series rather than replacing experimental confirmation [11].

Conclusion

Machine learning and deep learning methods for protein-ligand binding affinity prediction have matured considerably over roughly a decade of methodological development, with structure-based 3D-CNN and graph models and sequence-based deep architectures each reporting substantial correlations with experimental affinity on kinase-relevant benchmarks. However, performance remains benchmark-dependent, sensitive to data redundancy, and variable across kinase subfamilies. Continued progress will likely depend on larger and more diverse kinase-annotated datasets, rigorous scaffold- and time-split validation, physics-informed architectures that improve out-of-distribution generalisation, and closer integration with structural resources such as AlphaFold and prospective free-energy methods, so that predictive models can be deployed with appropriate confidence within kinase inhibitor discovery

pipelines.

Table 2: Representative machine learning and deep learning model classes for binding affinity prediction discussed in this review.

Model	Representation	Benchmark	Reported metric
RF-Score	Empirical contact descriptors + random forest	PDBbind core set	Pearson r ~0.78 [13]
KDEEP	3D voxel grid + 3D-CNN	PDBbind core 2016 / CASF-2016	Pearson r ~0.82 [1,5]
Pafnucy	3D voxel grid + 3D-CNN	PDBbind core 2016	Pearson r ~0.70 [2]
OnionNet	Contact-shell CNN	PDBbind core 2016	Pearson r ~0.82 [3]
DeepDTA	1D CNN on sequence/SMILES	Davis / KIBA	CI ~0.878 / 0.863 [6]
GraphDTA	Graph neural network + CNN	Davis / KIBA	CI ~0.893 / 0.891 [7]
PIGNet	Physics-informed graph neural network	External/scaffold-diverse test sets	Improved cross-scaffold generalisation [4]

Acknowledgements

The authors thank colleagues at Department of Bioinformatics, University of Turin, Turin, Italy, Department of Computational Biology, Riga Stradins University, Riga, Latvia, and Department of Pharmaceutical Chemistry, Medical University of Plovdiv, Plovdiv, Bulgaria for helpful discussions during preparation of this review.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Gallo V conceived the review, conducted the literature search, and drafted the manuscript. Liepa J contributed to the literature screening and technical framing and critically revised the manuscript for important intellectual content. Dimitrova V contributed additional context and approved the final version for submission.

Ethics Statement

Not applicable. This is a narrative review of previously published literature and did not involve new experimental data, human participants, or animal subjects.

Data Availability

No new data were generated or analyzed for this review. All sources synthesized are cited in the reference list and are available through their original publishers.

References

1. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model. 2018;58(2):287-296. doi:10.1021/acs.jcim.7b00650. PMID:29309725.

2. 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):3666-3674. doi:10.1093/bioinformatics/bty374. PMID:29757353.

3. 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):15956-15965. doi:10.1021/acsomega.9b01997. PMID:31592466.

4. 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):3661-3673. doi:10.1039/D1SC06946B.

5. Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, Wang R. Comparative assessment of scoring functions: the CASF-2016 update. J Chem Inf Model. 2019;59(2):895-913. doi:10.1021/acs.jcim.8b00545. PMID:30481020.

6. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018;34(17):i821-i829. doi:10.1093/bioinformatics/bty593. PMID:30423097.

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):1140-1147. doi:10.1093/bioinformatics/btaa921. PMID:33119053.

8. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011;29(11):1046-1051. doi:10.1038/nbt.1990. PMID:22037378.

9. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, Aittokallio T. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J Chem Inf Model. 2014;54(3):735-743. doi:10.1021/ci400709d. PMID:24521231.

10. Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. WIREs Comput Mol Sci. 2015;5(6):405-424. doi:10.1002/wcms.1225. PMID:27110292.

11. Cohen P, Cross D, Jänne PA. Kinase drug discovery 20 years after imatinib: progress and future directions. Nat Rev Drug Discov. 2021;20(7):551-569. doi:10.1038/s41573-021-00195-4. PMID:34002056.

12. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461. doi:10.1002/jcc.21334. PMID:19499576.

13. Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;26(9):1169-1175. doi:10.1093/bioinformatics/btq112. PMID:20236947.

14. Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes

with known three-dimensional structures. J Med Chem. 2004;47(12):2977-2980. doi:10.1021/jm030580l.

15. Liu Z, Li Y, Han L, Li J, Liu J, Zhao Z, Nie W, Liu Y, Wang R. PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics. 2015;31(3):405-412. doi:10.1093/bioinformatics/btu626. PMID:25301850.

16. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 2016;44(D1):D1045-D1053. doi:10.1093/nar/gkv1072. PMID:26481362.

17. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-D954. doi:10.1093/nar/gkw1074. PMID:27899562.

18. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem. 2012;55(14):6582-6594. doi:10.1021/jm300687e. PMID:22716043.

19. Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv:1510.02855. 2015.

20. Karimi M, Wu D, Wang Z, Shen Y. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics. 2019;35(18):3329-3338. doi:10.1093/bioinformatics/btz111. PMID:30768156.

21. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. Proc Mach Learn Res. 2017;70:1263-1272.

22. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-530. doi:10.1039/C7SC02664A. PMID:29629118.

23. Chen L, Tan X, Wang D, Zhong F, Liu X, Yang T, Luo X, Chen K, Jiang H, Zheng M. TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. Bioinformatics. 2020;36(16):4406-4414. doi:10.1093/bioinformatics/btaa524. PMID:32428219.

24. Huang K, Xiao C, Glass LM, Sun J. MolTrans: molecular interaction transformer for drug-target interaction prediction. Bioinformatics. 2021;37(6):830-836. doi:10.1093/bioinformatics/btaa880. PMID:33070179.

25. 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):308-322.e11. doi:10.1016/j.cels.2020.03.002.

26. Stärk H, Ganea OE, Pattanaik L, Barzilay R, Jaakkola T. EquiBind: geometric deep learning for drug binding structure prediction. Proc Mach Learn Res. 2022;162:20503-20521.

27. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J Am Chem Soc. 2015;137(7):2695-2703. doi:10.1021/ja512751q. PMID:25625324.

28. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2. PMID:34265844.