[Artificial Intelligence in the Life Sciences 2 (2022) 100037](https://doi.org/10.1016/j.ailsci.2022.100037)

Contents lists available at [ScienceDirect](http://www.ScienceDirect.com/)

Artificial Intelligence in the Life Sciences

journal homepage: [www.elsevier.com/locate/ailsci](http://www.elsevier.com/locate/ailsci)

Viewpoint

Deep learning of protein–ligand interactions—Remembering the actors

Jürgen Bajorath

*Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 5/6, Bonn D-53115, Germany*

One of the intensely investigated applications of deep learning in drug design is the prediction of compound potency (aﬃnity) based upon three-dimensional structures of protein–ligand complexes. Consistently accurate ligand binding aﬃnity predictions would represent a milestone event for the field and put structure-based ligand design on a new level. For this purpose, convolutional neural networks (CNNs) [[1]](#_bookmark1) with voxel representations of ligand binding sites as well as graph neural networks (GNNs) [[2]](#_bookmark2) including message passing neural networks (MPNNs) [[3]](#_bookmark3) are applied. GNNs/MPNNs learn directly from molecular graphs. In general, MPNNs are becoming increasingly popular for representation learning in chemistry. For aﬃnity predictions using GNNs/MPNNs, protein–ligand complex structures are translated into interaction graphs.

Employing these neural network architectures, a variety of aﬃn- ity prediction models based on protein–ligand complex structures have been reported in recent years (for an up-to-date summary see, for ex- ample, [[4](#_bookmark4)]). Graph-based models often achieve high correlation (at the 80% level) between predicted and experimentally observed ligand bind- ing aﬃnities and prediction accuracy within or close to an order of magnitude (10-fold). These observations have triggered recurrent as- sumptions or claims that deep neural networks are capable of learn- ing specific protein–ligand interactions. However, this would also im- ply that the resulting models would capture, in one way or another, the physico-chemical foundations of these interactions. Might this be conceivable by learning from molecular interaction graphs? Regardless of such principal considerations, several observations suggest that the promising results reported for various protein–ligand aﬃnity prediction models should be considered with caution.

An essential resource for protein–ligand complex structures with available experimental aﬃnity measurements is the PDBbind database

[[5]](#_bookmark5) that provides the basis for many investigations. However, the num- ber of high-resolution structures with high-quality aﬃnity measure- ments is limited and the composition of PDBbind is biased towards pre- ferred crystallographic targets [[6](#_bookmark6),[7](#_bookmark7)], which naturally limits the gener- alization ability of predictive models derived from these data. This is consistent with the finding that different training and test data parti- tions can significantly influence model performance [[6]](#_bookmark6). On the other hand, training sets of varying size often yield similarly accurate protein– ligand interaction models [[4](#_bookmark4)], which is counterintuitive for deep learn- ing. It has also been observed that CNN models trained only on protein or ligand representations can approach or meet the accuracy of models trained on protein–ligand interaction data [[6–9]](#_bookmark6). Taken together, these

*E-mail address:* [bajorath@bit.uni-bonn.de](mailto:bajorath@bit.uni-bonn.de)

findings indicate that CNN and GNN aﬃnity prediction models might primarily memorize training data information from complex structures, rather than learn specific protein–ligand interactions. Similar memo- rization effects have been observed for deep compound classification models displaying limited generalization ability [[10]](#_bookmark8).

The most recent scientifically rigorous investigation of protein– ligand interaction models has employed an MPNN architecture to learn from different graph representations of protein–ligand complexes from PDBbind [[4](#_bookmark4)]. For a given complex structure, Rognan and colleagues gen- erated graph representations of the ligand (L), protein (P), and protein– ligand interactions (I) using differently defined nodes and edges, as illus- trated in [Fig. 1](#_bookmark0) [[4](#_bookmark4)]. For example, protein binding site residues were rep- resented as nodes annotated with interaction-relevant chemical prop- erty information. Furthermore, in interaction graphs, different nodes representing interaction sites formed by ligand atoms and binding site residues, respectively, were combined and edges accounted for non- covalent short-range interactions (annotated with their distances).

From the three graph representations, seven training constellations were generated including the individual graphs (L, P, I), three pair- wise combinations (PL, PI, LI), and the triplet (PLI) combining all three graphs. This framework was used to derive different MPNN models, which were then applied to predict the ligand aﬃnities of complex struc- tures from different test sets. Predictive performance was quantified by calculating Pearson’s correlation coeﬃcient (RP) for predicted and ex- perimental aﬃnities as well as the root mean square error (RMSE).

above 0.6 and maximally ∼0.8. Importantly, models based only on the The models were generally found to be predictive, with RP values of

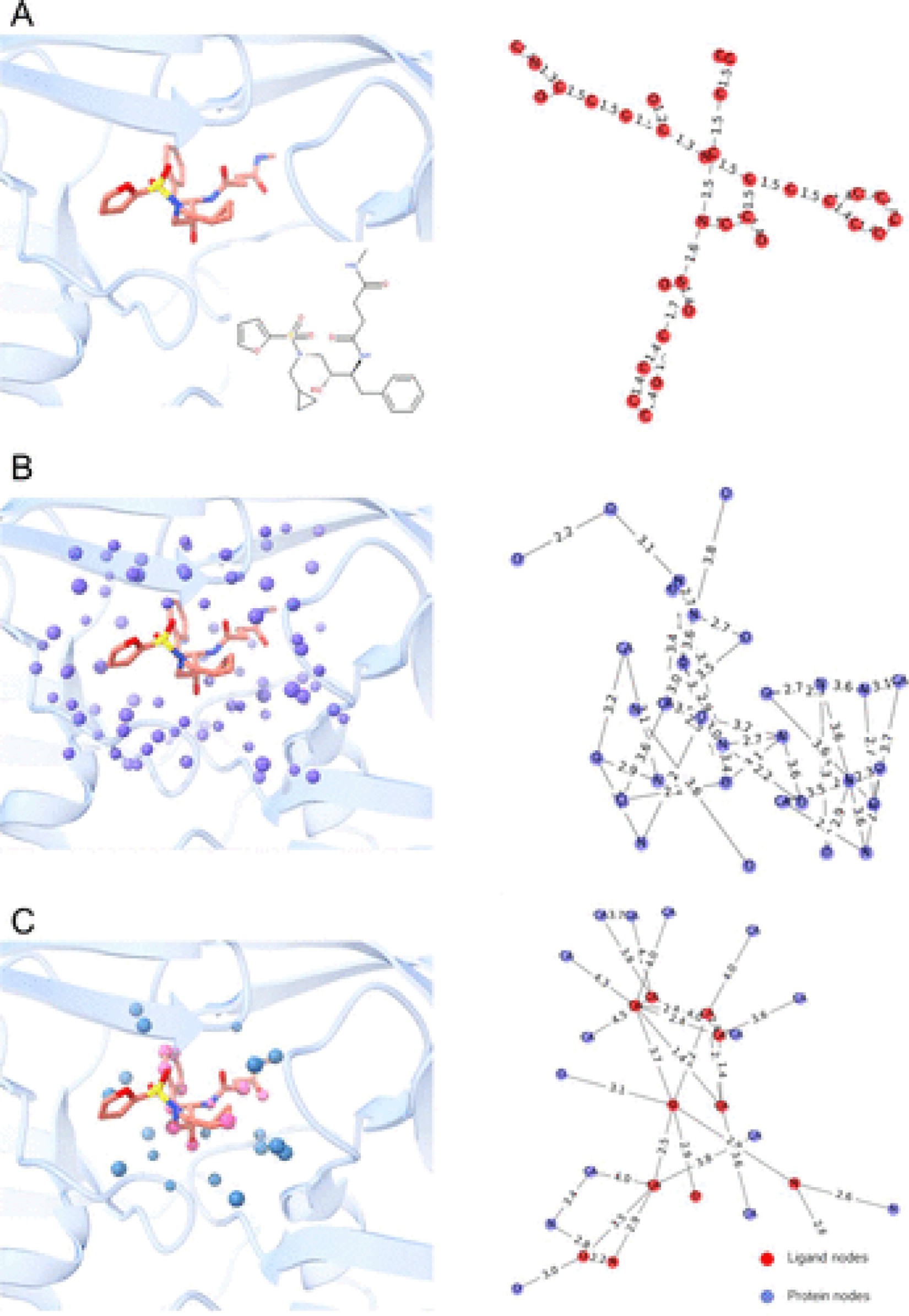
L or P graphs were more accurate than the model trained on the inter- action graph (I model), with the ligand-based model performing best. The accuracy was further increased by the PL model, the performance of which was very similar to the PLI model trained on all three graphs. These findings conclusively demonstrated that aﬃnity predictions using MPNNs did not depend on learning specific protein–ligand interactions, let alone the underlying physics.

The authors further extended their analysis. For example, the MPNN models were found to display limited generalization potential, but were insensitive to reduction of training set size. In addition, simple “memo- rization baseline” models were generated predicting the aﬃnity of com- plexes based on averages for the most similar ligands or proteins from the training set. Interestingly, the ligand baseline model nearly reached the accuracy of the I model. Taken together, these results showed that

<https://doi.org/10.1016/j.ailsci.2022.100037> Received 25 May 2022; Accepted 25 May 2022

Available online 26 May 2022

2667-3185/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



**Fig. 1. Graph representations of proteins, ligands, and their interactions**. Based on the structure of a protein–ligand complex, graph representations encode the

(**A**) ligand (red nodes, edges), (**B**) protein’s ligand binding site (blue nodes, edges), and (**C**) protein–ligand interactions (blue/red nodes, edges). For further details, see Volkov et al. [[4](#_bookmark4)]. The figure was reprinted with permission from Volkov, M.; Turk, J.A.; Drizard, N.; Martin, N.; Hoffmann, B.; Gaston-Mathé; Rognan, D. On the frustration to predict binding aﬃnities from protein-ligand structures with deep neural networks. J Med Chem 2022. doi: 10.1021/acs.jmedchem.2c00487. in press. Copyright 2022 American Chemical Society.

the aﬃnity predictions were mostly driven by memorizing patterns from training data, with ligand similarity relationships playing a major role. The authors also showed that increasing the complexity of interaction graphs by including more interactions over longer distances further in- creased the prediction accuracy of the I model. This observation led to the conclusion that an interaction model deprived of additional pro- tein or ligand context information should provide a reasonable basis for further exploring the ability of deep neural networks to learn protein– ligand interactions. Like others before, the authors also emphasized that the current sparsity of high-quality and diverse protein–ligand com- plex data represents a major limitation for the further development of protein–ligand aﬃnity models.

For compound potency predictions and structure-based drug design, insights provided by careful studies like the one by Rognan and col- leagues are of fundamental relevance, putting putative methodological advances into scientific perspective, raising awareness of potential over- interpretation, and balancing expectations. Moreover, for deep learning across the life sciences, investigations demonstrating limitations of cur- rent approaches and potential caveats or misinterpretation of results are as important for the further development of the field as methodological breakthroughs.

# Declaration of Competing Interest

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

# References

1. [He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Pro- ceedings of the IEEE conference on computer vision and pattern recognition; 2016.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0001)

[p. 770–8.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0001)

1. [Xiong J, Xiong Z, Chen K, Jiang H, Zheng M. Graph neural networks for automated de novo drug design. Drug Discov Today 2021;26:1382–93.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0002)
2. [Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proceedings of the 34th international conference on machine learning, 70; 2017. p. 1263–72.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0003)
3. Volkov M., Turk J.A., Drizard N., Martin N., Hoffmann B., Gaston M., Rognan D., On the frustration to predict binding aﬃnities from protein-ligand structures with deep neural networks. J Med Chem 2022. doi: 10.1021/acs.jmedchem.2c00487, in press.
4. [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;47:2977–80.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0005)
5. [Yang J, Shen C, Huang N. Predicting or pretending: artificial intelligence for pro- tein-ligand interactions lack of suﬃciently large and unbiased datasets. Front Phar- macol 2020;11:69.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0006)
6. [Wang J, Dokholyan NVYuel. Improving the generalizability of structure-free com- pound-protein interaction prediction. J Chem Inf Model 2022;62:463–71.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0007)
7. [Chen L, Cruz A, Ramsey S, Dickson CJ, Duca JS, Hornak V, Koes DR, Kurtzman T. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. PLoS One 2019;14:e0220113.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0008)
8. [Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: pre- dicting drug-target binding aﬃnity with graph neural networks. Bioinformatics 2021;37:1140–7.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0009)
9. [Wallach I, Heifets A. Most ligand-based classification benchmarks reward memo- rization rather than generalization. J Chem Inf Model 2018;58:916–32.](http://refhub.elsevier.com/S2667-3185(22)00008-3/sbref0010)