Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Statistical and machine learning approaches to predicting protein–ligand interactions

Lucy J Colwell

Data driven computational approaches to predicting protein–ligand binding are currently achieving unprecedented levels of accuracy on held-out test datasets. Up until now, however, this has not led to corresponding breakthroughs in our ability to design novel ligands for protein targets of interest. This review summarizes the current state of the art in this field, emphasizing the recent development of deep neural networks for predicting protein–ligand binding. We explain the major technical challenges that have caused difficulty with predicting novel ligands, including the problems of sampling noise and the challenge of using benchmark datasets that are sufficiently unbiased that they allow the model to extrapolate to new regimes.

**Address**
Department of Chemistry, Cambridge University, Cambridge, UK

Corresponding author: Colwell, Lucy J (ljc37@cam.ac.uk)

**Current Opinion in Structural Biology** 2018, **49**:123–128

This review comes from a themed issue on **Theory and simulation**

Edited by **Robert Best** and **Kresten Lindorff-Larsen**

For a complete overview see the Issue and the Editorial

Available online 20th February 2018

**https://doi.org/10.1016/j.sbi.2018.01.006**



## Introduction

Molecular recognition is a fundamental requirement of biological systems. The interactions between proteins and small molecules are central to biology, allowing cells to sense their surroundings and respond appropriately. Estimates place the number of small molecules that can be synthesized at $\approx 10^{60}$, yet just a small fraction of potential protein–ligand interactions have been explored. Finding novel interactions is of great importance to drug discovery and basic biology. Given the enormity of the search space, computational approaches can narrow down the possibilities. However, despite three decades of computational effort, biochemical experiments are still essential to determine the efficacy of ligand binding to a protein target [1,2]. The results of computational analysis have been decidedly mixed: it is challenging to use even experimentally well-characterized ligand–protein interactions to computationally design novel interactions [1,2], much less explore the vast space of possibilities.

There are three highly demanding tasks in protein–ligand binding prediction: *virtual screening* predicts whether a ligand binds to a given target; *affinity prediction* predicts the binding affinity; and *pose prediction* identifies the molecular interactions causing binding to occur. In this review we focus on the first; the others have been reviewed elsewhere [3,4]. Approaches to virtual screening can be categorized as physical or statistical. The idea of using first principles physical models to describe protein–ligand interactions is attractive, however timescale and computational resource constraints mean that simplified descriptions of features such as protein flexibility, and solvent are necessary. Docking algorithms are an important example of a coarse-grained physical model, however, even the most sophisticated versions cannot accurately reproduce large numbers of known interactions, much less predict new ones. The scoring functions used in such approaches can be empirical [5–9] or knowledge-based [10–14], and significant expertise is required to encode physico-chemical interactions through the use of hand-tuned features and parameters. Moreover, the results can be highly specific to the system that they are designed for [15].

Recently, the use of high throughput methods to screen large libraries of proteins and small molecules and quantify their interactions has made it possible to correlate activity with representations of proteins and small molecules, to infer predictive models. Techniques from machine learning and artificial intelligence have been introduced, allowing both the parameters and the model to be learned from the wealth of experimental data available in databases such as Chembl [16–19]. Increasingly publications are demonstrating that data driven approaches have the potential to make significant contributions to these problems [20••,21•,22•,23•,24–27].

## Machine learning — potential and limitations

The aim of any machine learning or statistical approach is to identify patterns among training examples that can be used to make predictions outside the training set. An algorithm achieves this by mapping the training set representation to a space in which active and inactive ligands segregate — this mapping can be guided by physical models [13,23•] but is more often learned directly from the data without addition of extensive physico-chemical knowledge [21•,22•,26]. Once this mapping

has been learned, it is hoped that the location of new examples in this space — clustered with training set actives or inactives — will accurately predict their activity. To compare the performance of different algorithms, a test set of ligands with known activity is held-out during the training process. An algorithm that can make accurate predictions for this unseen data is presumed to extrapolate well.

In particular, approaches that use deep neural networks (DNNs) have been shown to make predictions on held-out test sets with eye-opening levels of accuracy, exceeding 0.95 AUC (Area under the Receiver Operator Characteristic) on benchmark datasets [22•,28]. However, the extent to which such results extrapolate is not yet clear — are they overfit to the training data [1,29,30••]? A number of studies posit that the test/train protocol is not as exacting as it might appear due to the non-uniformity of the distribution of ligands in chemical space. If training set actives are closer to test set actives than to training set inactives, by some metric that is not fully predictive of protein–ligand binding activity such as molecular weight, or number of ring systems, then the algorithm can appear to make accurate predictions for test set molecules without being able to extrapolate this predictive ability [31,19,32]. Machine learning performs best when abundant data drawn uniformly from the space of interest is available, but in this setting human chemists choose which molecules to work with, often based on clear similarities to known success stories [33,34].

Definitively showing that these approaches generalize is perhaps the outstanding challenge facing this field today. In the search for novel pharmaceuticals, the ability to predict the binding of ligands that are chemically distinct from those in the training data is highly valuable, but much more challenging for algorithms that are expert at identifying patterns among training set ligands. The goal of this article is to review statistical approaches to molecular recognition in the context of protein–ligand binding, focusing on recent results that exploit DNNs (see Figure 1). Here we briefly outline the basic steps of machine learning algorithm that predicts protein–ligand binding.

## Molecular representation

There are almost as many choices for representation of the input data as there are for the machine learning algorithm employed [35,36]. The simplest involve counting the numbers of different heavy atoms present in a ligand, together with other features such as hydrogen bond donors/acceptors, chiral centres and ring systems [19,30••]. Some information about the chemical structure is retained by descriptors such as atom pairs or donor–acceptor pairs [37,38] where each element has the form (atom type *i*) — (distance in bonds) — (atom type *j*). More information is encoded by chemical fingerprints, for example MACCS keys [39] and ECFP fingerprints [40]; fixed length binary descriptors which can be generated by the package RDKit [41]. Here, each non-hydrogen atom is used as a centre from which fragments are generated by extending radially from the centre along bonds to neighbouring atoms; the maximum radius considered *N* is encoded in the name as ECFP2*N*. A unique identifier is assigned to each fragment, and the set of identifiers for molecules is mapped to a fixed length bit vector to yield the molecular fingerprint.

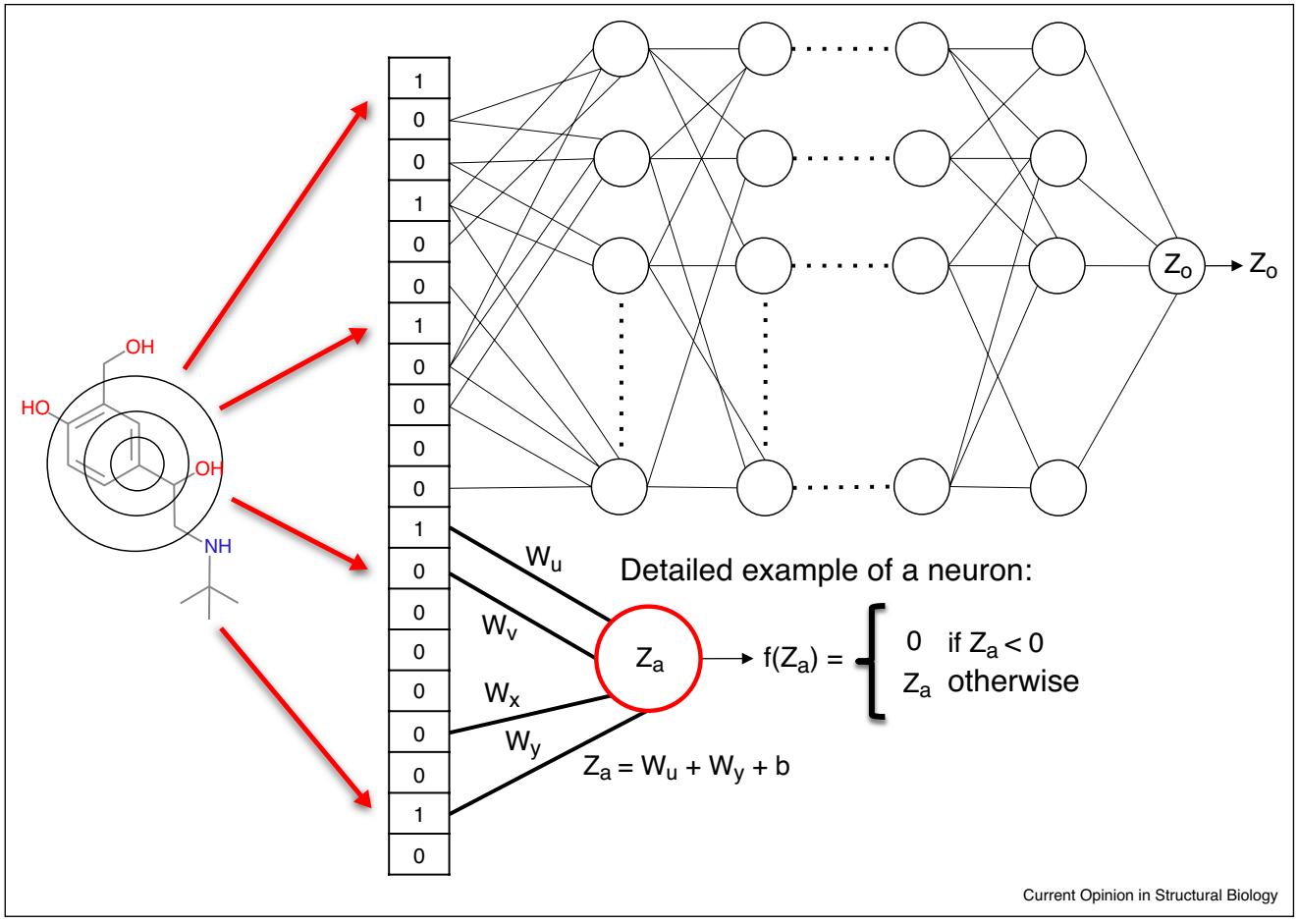
This abundance raises the question of which representation is most useful for different prediction tasks. Recently the suggestion has been made that it may be more effective to also learn the molecular representation itself, alongside the metric and corresponding embedding space used to distinguish active from inactive ligands [21•,22•,27,42]. However, counter-intuitively it has also been reported that the use of more complex molecular descriptors can result in little gain of predictive ability [30••,43].

## Representation and sampling noise

One rational for the finding that more complex representations can result in little improvement is noise due to finite sampling. The basic premise of any predictive algorithm is that similarities among known interaction partners can reveal the requirements of the binding site, and thus predict novel interactions. A straightforward approach is to compile the set of ligands known to bind to a protein receptor of interest, and identify those features that show statistically significant enrichment among this set [44]. However, because there are only finitely many samples (i.e. known ligand binders), some features will be enriched purely by chance. This chance similarity increases with the number of variables, so representations that have more variables will lead to greater random similarity between features. For DNNs in particular, representations with thousands or even millions of features have recently been employed [24,21•,22•]; although these algorithms have the ability to share information between targets it is still important that the level of chance similarity between small molecules is quantified and accounted for.

This phenomena has been carefully studied in the field of random matrix theory, which provides a null distribution that describes the similarity between samples (ligands) that can be expected by chance due to finite sampling as a function of the number of samples available, and the number of variables present in the ligand descriptor [45,46•]. A simpler method for generating this null distribution involves computing the covariance matrices of multiple sets of *n* random ligands, where *n* is the sample size, using the same ligand descriptor for each set, to obtain the distribution of the largest entries that occur due to finite sampling noise. A similar approach can be taken for any measure of molecular similarity, defining a

**Figure 1**




Schematic of a deep neural network for predicting properties of small molecule ligands. In this schematic, the input molecule is first encoded in a binary molecular fingerprint of fixed length. Convolutional architectures are a popular choice among practitioners and typically require that a fixed length representation be either used or learned from the molecular graph. The neural network simply consists of a very large number of simple functions (or neurons) of the form indicated in the schematic. Deep neural networks have large numbers of layers of neurons, resulting in a huge number of parameters that are learnt during training from the data, which is typically split into distinct training, cross-validation and test datasets.

statistical null distribution with which to compare putative active molecules.

## Benchmark datasets

The application of machine learning algorithms to problems such as protein–ligand binding is facilitated by the availability of large experimental datasets. For both algorithms that require 3D complexes and those that work with 2D ligand structures, a number of benchmark datasets have emerged that allow researchers to compare the performance of different algorithms. The PDBbind benchmark [47] comprises 1300 diverse protein–ligand complexes with high quality structural and binding data. Another popular choice is the Directory of Useful Decoys (DUD) [48] or Directory of Useful Decoys Enhanced (DUDE) benchmark [18]. These datasets have been carefully constructed to attempt to control for biases that could enable a machine learning algorithm to learn features distinguishing actives from inactives without providing information about the relevant molecular interactions. The construction of benchmark datasets that are unbiased, and so provide an accurate indication of how well an algorithm can extrapolate is an important research direction that requires attention [30••,49•,50].

## Choice of algorithm

Early linear regression methods were soon super-seeded by nonlinear models constructed using early versions of neural networks, among other techniques. Algorithms such as support vector machines [51,52], Gaussian Processes [53,54] and Random Forests (RF) [55,15,56,43] have all received significant attention within the community. One advantage of these approaches is that many allow some level of interpretability, enabling the medicinal chemist to rationalize and evaluate the resulting model. This is more challenging for DNNs (see Figure 1),

which have recently been found to be highly successful on a variety of different tasks including computer vision [57] and speech recognition [58]. In the last few years there has been a burst of activity involving the construction of DNNs to predict protein–ligand binding [20••,21•,22•,24–26,28,59–62].

A key breakthrough came in 2012 when Merck sponsored a Kaggle contest, in which the winning entry submitted by G. Dahl improved the mean squared Pearson correlation coefficient ($R^2$) between predicted and observed binding activities over 15 proprietary Merck datasets from 0.42 (attained using RFs) to 0.49 [20••]. Although a combination of DNNs [63•], gradient boosting machine [64] and Gaussian process regression [53] were used by the winning entry, the authors suggest that the DNN was the main cause of the performance increase. A careful comparison of DNNs and RFs at protein-binding prediction over 15 datasets found that the best $R^2$ values over all parameter sets are obtained using DNNs for five datasets, and using RFs for one, with mixed results for the remaining datasets. For refined parameter ranges, this increased to nine datasets for DNNs compared to one for RFs, leading the authors to conclude that DNNs have the potential to outperform RFs generally [20••].

A crucial issue with DNNs is the set of algorithmic parameters affecting predictive performance that must be decided by the researcher, such as network size, choice of activation functions and use of dropout [65], which are distinct from parameters that are learned from the data. In general it is not computationally feasible to search the space of parameter values, in particular because different parameters are not independent of each other. Some analysis has been presented by varying one parameter at a time, and recording the $R^2$ values of the resulting DNNs [20••].

## Conclusion

Advances in technology have led to enormous opportunity for using machine learning algorithms to learn models of protein–ligand interactions. Machine learning algorithms are inherently flexible and with a sufficient amount of carefully curated data, the algorithms have the potential to find patterns that humans cannot find on their own. However, besides the issues already discussed with this review, other challenges must be overcome: the parameters learned from the data can lead to overfitting — the unintended specialization of the trained model to the dataset used for training. For example, if the training and test datasets were identical, the algorithm would simply memorize all details of the training data, and achieve perfect performance on the test dataset. However, it would likely be unable to make accurate predictions for additional molecules that it had not previously seen before.

For chemical data, a key problem is to define what constitutes a distinct partitioning between test and training data. Simply requiring that the molecules are different from one another is not enough [19,30••,31,32]. Machine learning algorithms are extremely effective at learning any feature that distinguishes samples from different classes (such as binding and non-binding). This means that it is necessary to go further, and require that active molecules do not share any nuisance similarity or property that distinguishes them from inactive molecules, while not being the crucial molecular feature that enables binding. The situation is complicated by the fact that in most cases the crucial feature that enables binding is not known — indeed identifying these feature(s) is the ultimate goal of the exercise. However relatively simple devices such as synthetic datasets, in which binding is said to occur if particular logical combinations of features are present, or bias measures that quantify simple molecular similarities can be surprisingly effective.

Despite these challenges, this is an exciting time for building models of protein–ligand interactions. One could hope that a review written five years hence would be filled more with victories than warnings.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest


1. Peón A, Naulaerts S, Ballester PJ: **Predicting the reliability of drug–target interaction predictions with maximum coverage of target space**. *Sci Rep* 2017, **7**.
2. Rathi PC, Ludlow RF, Hall RJ, Murray CW, Mortenson PN, Verdonk ML: **Predicting "hot" and "warm" spots for fragment binding**. *J Med Chem* 2017, **60**:4036-4046.
3. Baron R, McCammon JA: **Molecular recognition and ligand association**. *Annu Rev Phys Chem* 2013, **64**:151-175.
4. Durrant JD, McCammon JA: **Molecular dynamics simulations and drug discovery**. *BMC Biol* 2011, **9**:71.
5. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: **Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes**. *J Comput Aided Mol Des* 1997, **11**:425-445.
6. Böhm H-J: **The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure**. *J Comput Aided Mol Des* 1994, **8**:243-256.
7. Wang R, Lai L, Wang S: **Further development and validation of empirical scoring functions for structure-based binding affinity prediction**. *J Comput Aided Mol Des* 2002, **16**: 11-26.
8. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK *et al.*: **Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy**. *J Med Chem* 2004, **47**:1739-1749.

9. 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**:455-461.

10. Muegge I, Martin YC: **A general and fast scoring function for protein–ligand interactions: a simplified potential approach**. *J Med Chem* 1999, **42**:791-804.

11. Gohlke H, Hendlich M, Klebe G: **Knowledge-based scoring function to predict protein–ligand interactions**. *J Mol Biol* 2000, **295**:337-356.

12. Zhou H, Skolnick J: **Goap: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction**. *Biophys J* 2011, **101**:2043-2052.

13. Verdonk ML, Ludlow RF, Giangreco I, Rathi PC: **Protein–ligand informatics force field (PLiff): toward a fully knowledge driven "force field" for biomolecular interactions**. *J Med Chem* 2016, **59**:6891-6902.

14. Sliwoski G, Kothiwale S, Meiler J, Lowe EW: **Computational methods in drug discovery**. *Pharmacol Rev* 2014, **66**: 334-395.

15. Ballester PJ, Mitchell JB: **A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking**. *Bioinformatics* 2010, **26**:1169-1175.

16. 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* 2016, **45**: D945-D954.

17. 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**:D1045-D1053.

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**:6582-6594.

19. Rohrer SG, Baumann K: **Maximum unbiased validation (MUV) data sets for virtual screening based on pubchem bioactivity data**. *J Chem Inf Model* 2009, **49**:169-184.

20. •• Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V: **Deep neural nets as a method for quantitative structure–activity relationships**. *J Chem Inf Model* 2015, **55**:263-274.
Beyond the interesting results reported, that use deep neural nets (DNNs) to learn quantitative structure–activity relationships, this clearly written paper also uses careful analysis to shed light on the impact of DNN hyperparameters on predictive performance.

21. • Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP: **Convolutional networks on graphs for learning molecular fingerprints**. *Advances in Neural Information Processing Systems*. 2015:2224-2232.
Introduces an approach in which molecular descriptors are learned directly from the molecular graph. The authors report that these data-driven features are more interpretable, and display better performance on a number of prediction tasks.

22. • Kearnes S, McCloskey K, Berndl M, Pande V, Riley P: **Molecular graph convolutions: moving beyond fingerprints**. *J Comput Aided Mol Des* 2016, **30**:595-608.
The process of learning a molecular representation from the data is explained in this paper, which also explores different network architectures, and compares the resulting performance.

23. • Bartok AP, De S, Poelking C, Bernstein N, Kermode J, Csanyi G, Ceriotti M: *Machine Learning Unifies the Modelling of Materials and Molecules*. 2017arXiv:1706.00179.
The authors introduce a novel structure-based machine learning model that uses Gaussian process regression together with a local description of chemical environments to develop a unified framework that makes accurate predictions for problems that span from protein–ligand binding to the modelling of hard matter.

24. Unterthiner T, Mayr A, Klambauer G, Steijaert M, Wegner JK, Ceulemans H, Hochreiter S: **Deep learning as an opportunity in virtual screening**. In *Proceedings of the Deep Learning Workshop at NIPS*. 2014.

25. Wallach I, Dzamba M, Heifets A: *Atomnet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-Based Drug Discovery*. 2015arXiv:1510.02855.

26. Goh GB, Siegel C, Vishnu A, Hodas NO, Baker N: *Chemception: A Deep Neural Network with Minimal Chemistry Knowledge Matches the Performance of Expert-Developed QSAR/QSPR Models*. 2017arXiv:1706.06689.

27. Altae-Tran H, Ramsundar B, Pappu AS, Pande V: **Low data drug discovery with one-shot learning**. *ACS Cent Sci* 2017, **3**: 283-293.

28. Ramsundar B, Kearnes S, Riley P, Webster D, Konerding D, Pande V: *Massively Multitask Networks for Drug Discovery*. 2015arXiv:1502.02072.

29. Gabel J, Desaphy J, Rognan D: **Beware of machine learning-based scoring functions on the danger of developing black boxes**. *J Chem Inf Model* 2014, **54**:2807-2815.

30. •• Wallach I, Heifets A: *Most Ligand-Based Benchmarks Measure Overfitting Rather than Accuracy*. 2017arXiv:1706.06619.
Using a novel metric the authors demonstrate that many benchmark datasets are biased. This implies that machine learning algorithms trained on these datasets may not generalize beyond the domain of the training data as well as might be expected, given the reported performance on held out test data.

31. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P: **Virtual screening using protein–ligand docking: avoiding artificial enrichment**. *J Chem Inf Comput Sci* 2004, **44**:793-806.

32. Ripphausen P, Wassermann AM, Bajorath J: **REPROVIS-DB: a benchmark system for ligand-based virtual screening derived from reproducible prospective applications**. *J Chem Inf Model* 2011, **51**:2467-2473.

33. Cleves AE, Jain AN: **Effects of inductive bias on computational evaluations of ligand-based modeling and on drug discovery**. *J Comput Aided Mol Des* 2008, **22**:147-159.

34. Jain AN, Cleves AE: **Does your model weigh the same as a duck?** *J Comput Aided Mol Des* 2012, **26**:57-67.

35. Maggiora G, Vogt M, Stumpfe D, Bajorath J: **Molecular similarity in medicinal chemistry: miniperspective**. *J Med Chem* 2013, **57**:3186-3204.

36. Cereto-Massagué A, Ojeda MJ, Valls C, Mulero M, Garcia-Vallvé S, Pujadas G: **Molecular fingerprint similarity search in virtual screening**. *Methods* 2015, **71**:58-63.

37. Carhart RE, Smith DH, Venkataraghavan R: **Atom pairs as molecular features in structure–activity studies: definition and applications**. *J Chem Inf Comput Sci* 1985, **25**:64-73.

38. Kearsley SK, Sallamack S, Fluder EM, Andose JD, Mosley RT, Sheridan RP: **Chemical similarity using physiochemical property descriptors**. *J Chem Inf Comput Sci* 1996, **36**: 118-127.

39. Durant JL, Leland BA, Henry DR, Nourse JG: **Reoptimization of MDL keys for use in drug discovery**. *J Chem Inf Comput Sci* 2002, **42**:1273-1280.

40. Rogers D, Hahn M: **Extended-connectivity fingerprints**. *J Chem Inf Model* 2010, **50**:742-754.

41. Landrum G: *Rdkit: Open-Source Cheminformatics*. 2006:2012 http://www.rdkit.org. (accessed 03 (04), online).

42. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE: *Neural Message Passing for Quantum Chemistry*. 2017arXiv:1704.01212.

43. Ballester PJ, Schreyer A, Blundell TL: **Does a more precise chemical description of protein–ligand complexes lead to more accurate prediction of binding affinity?** *J Chem Inf Model* 2014, **54**:944-955.

44. Todeschini R, Consonni V, Xiang H, Holliday J, Buscema M, Willett P: **Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets**. *J Chem Inf Model* 2012, **52**:2884-2901.

45. Edelman A, Wang Y: **Random matrix theory and its innovative applications**. *Advances in Applied Mathematics, Modeling, and Computational Science*. Springer; 2013:91-116.

46. • Lee AA, Brenner MP, Colwell LJ: **Predicting protein–ligand affinity with a random matrix framework**. *Proc Natl Acad Sci U S A* 2016, **113**:13564-13569.
This paper applies random matrix theory to develop a novel approach that identifies the spurious enrichment of features among active or inactive ligands that can be caused by finite sampling effects.

47. 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* 2014, **31**:405-412.

48. Huang N, Shoichet BK, Irwin JJ: **Benchmarking sets for molecular docking**. *J Med Chem* 2006, **49**:6789-6801.

49. • Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V: *Moleculenet: A Benchmark for Molecular Machine Learning*. 2017arXiv:1703.00564.
Provides a comprehensive overview of datasets that can be used to train and test machine learning algorithms.

50. Lagarde N, Zagury J-F, Montes M: **Benchmarking data sets for the evaluation of virtual ligand screening methods: review and perspectives**. *J Chem Inf Model* 2015, **55**:1297-1307.

51. Burbidge R, Trotter M, Buxton B, Holden S: **Drug design by machine learning: support vector machines for pharmaceutical data analysis**. *Comput Chem* 2001, **26**:5-14.

52. Jorissen RN, Gilson MK: **Virtual screening of molecular databases using a support vector machine**. *J Chem Inf Model* 2005, **45**:549-561.

53. Burden FR: **Quantitative structure–activity relationship studies using Gaussian processes**. *J Chem Inf Comput Sci* 2001, **41**:830-835.

54. Obrezanova O, Csányi G, Gola JM, Segall MD: **Gaussian processes: a method for automatic QSAR modeling of ADME properties**. *J Chem Inf Model* 2007, **47**:1847-1857.

55. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP: **Random forest: a classification and regression tool for compound classification and QSAR modeling**. *J Chem Inf Comput Sci* 2003, **43**:1947-1958.

56. Zilian D, Sotriffer CA: **Sfcscore rf: a random forest-based scoring function for improved affinity prediction of protein–ligand complexes**. *J Chem Inf Model* 2013, **53**:1923-1933.

57. Krizhevsky A, Sutskever I, Hinton GE: **Imagenet classification with deep convolutional neural networks**. *Advances in Neural Information Processing Systems*. 2012:1097-1105.

58. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A-r, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN *et al.*: **Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups**. *IEEE Signal Process Mag* 2012, **29**:82-97.

59. Kearnes S, Goldman B, Pande V: *Modeling Industrial ADMET Data with Multitask Networks*. 2016arXiv:1606.08793.

60. Gonczarek A, Tomczak JM, Zareba S, Kaczmar J, Dabrowski P, Walczak MJ: *Learning Deep Architectures for Interaction Prediction in Structure-Based Virtual Screening*. 2016arXiv:1610.07187.

61. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR: **Protein–ligand scoring with convolutional neural networks**. *J Chem Inf Model* 2017, **57**:942-957.

62. Gomes J, Ramsundar B, Feinberg EN, Pande VS: *Atomic Convolutional Networks for Predicting Protein–Ligand Binding Affinity*. 2017arXiv:1703.10603.

63. • Goodfellow I, Bengio Y, Courville A: *Deep Learning*. MIT Press; 2016.
A comprehensive and clearly written introduction to machine learning algorithms and techniques.

64. Svetnik V, Wang T, Tong C, Liaw A, Sheridan RP, Song Q: **Boosting: an ensemble learning tool for compound classification and QSAR modeling**. *J Chem Inf Model* 2005, **45**:786-799.

65. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R: **Dropout: a simple way to prevent neural networks from overfitting**. *J Mach Learn Res* 2014, **15**:1929-1958.