Deep neural network affinity model for BACE inhibitors in D3R Grand Challenge 4

Bo Wang · Ho-Leung Ng

Abstract

Drug Design Data Resource (D3R) Grand Challenge 4 (GC4) offered a unique opportunity for designing and testing novel methodology for accurate docking and affinity prediction of ligands in an open and blinded manner. We participated in the beta-secretase 1 (BACE) Subchallenge which is comprised of cross-docking and redocking of 20 macrocyclic ligands to BACE and predicting binding affinity for 154 macrocyclic ligands. For this challenge, we developed machine learning models trained specifically on BACE. We developed a deep neural network (DNN) model that used a combination of both structure and ligand-based features that outperformed simpler machine learning models. According to the results released by D3R, we achieved a Spearman's rank correlation coefficient of 0.43(7) for predicting the affinity of 154 ligands. We describe the formulation of our machine learning strategy in detail. We compared the performance of DNN with linear regression, random forest, and support vector machines using ligand-based, structure-based, and combining both ligand and structure-based features. We compared different structures for our DNN and found that performance was highly dependent on fine optimization of the regularization hyperparameter, alpha. We also developed a novel metric of ligand three-dimensional similarity inspired by crystallographic difference density maps to match ligands without crystal structures to similar ligands with known crystal structures. This report demonstrates the detailed parameterization and careful data training and implementation necessary to obtain strong performance with more complex machine learning methods. Our DNN approach tied for fourth in predicting BACE-ligand binding affinities.

Bo Wang
Biochemistry & Molecular Biophysics, 141 Chalmers Hall, Kansas State University, 1711 Claflin Rd., Manhattan, 66506, KS, U.S.A.
Tel.: 785-532-6121, Fax: 785-532-7278
E-mail: wangbo@ksu.edu

Ho-Leung Ng
Biochemistry & Molecular Biophysics, 141 Chalmers Hall, Kansas State University, 1711 Claflin Rd., Manhattan, 66506, KS, U.S.A.
Tel.: 785-532-2518, Fax: 785-532-7278
E-mail: hng@ksu.edu

Keywords D3R, Docking, Binding affinity, Structure-based drug design, Machine learning, Deep neural network, BACE, beta-secretase 1

Introduction

The Drug Design Data Resource (D3R) has organized four Grand Challenges (GC) for docking, affinity, and free energy predictions for protein-ligand complexes.[1–3] By providing high quality, blinded protein-ligand crystal structures and affinity datasets, D3RGC has attracted extensive attention from computational drug design researchers. Assessment of results from blinded competition has provided unbiased insights into the most effective strategies as well as the many shortcomings of the state of the art. This platform has sparked numerous novel computer-aided drug designing methods. In this article, we present our machine learning approach in GC4 for predicting ligand binding affinities for beta-secretase 1 (BACE), a key protein for the formation of amyloid β-peptide and a leading drug target for Alzheimer's disease.[4]

Computer-aided drug design facilitates the whole process of drug development, such as virtual screening, lead optimization, structure-activity relationships (SAR) analysis, and ADMET modeling.[5] The designing and modeling of drug molecules can be classified into structure-based drug design (SBDD) and ligand-based drug design (LBDD), depending on whether three-dimensional structural information is used.[6] To design a novel drug molecule, an accurate prediction of the binding affinity between a ligand and its target protein is helpful. However, due to the complex nature of intermolecular interactions, protein flexibility, solvation, and the entropic effect, the prediction of docked structures and affinities of protein-ligands is very challenging. Classically, a predicted docked pose needs to be generated first, then the affinity is calculated by force-field-based, knowledge-based, or an empirical scoring function. Popular docking programs and scoring functions such as the AutoDock family[7–9], Glide[10], GOLD[11], and X-Score[12] have demonstrated their advantages in predicting docked poses and protein-ligand affinities.

Machine learning models have shown great potential in advancing current computer-aided drug designing methodology.[13] The advance of machine learning platforms and tools for chemistry, such as TensorFlow[14] and scikit-learn[15], and the public availability of high-quality protein-drug datasets, such as the PDB[16] and PDBbind[17], greatly enables the application of machine learning to drug design. Novel machine learning-based scoring methods, such as RF-Score[18] and K_{DEEP}[19], or machine learning-optimized software, such as Vinardo[20], smina[21], RF-Score-v3[22], have demonstrated superior performance in addition to their generalizability and accessibility. Machine learning methods have been shown to perform better than conventional scoring methods in benchmark studies.[23] For example, K_{DEEP}, RF-Score, X-Score, and Cyscore have Pearson's correlation coefficients (R) of 0.82, 0.80, 0.66, 0.65, respectively, for 290 protein-ligand affinities in the PDBbind v.2016 core set.[19] Surprisingly, these four scoring methods are very poor at predicting the affinities of ligands to BACE, as they yield Pearson's correlation coefficients (R) of -0.06, -0.14, -0.12, and 0.2, respectively for 36 BACE ligands. It is still an unresolved question why some protein targets are more difficult than others for different algorithms and scoring functions.

We are especially interested in answering the question of whether a machine learning model trained on a specific target can improve the affinity prediction performance for ligands to this target. Since D3R GC4 provided a high-quality and blinded BACE affinity dataset, we took this opportunity to explore the performance of a target-trained machine learning model. We also demonstrate how the combination of structure-based and ligand-based features benefit machine learning performance.

Methods

Affinity model overview

To build and test the affinity prediction performance for BACE-specific trained machine learning models, three essential elements: training BACE input features (X training), training BACE experimental affinities (y training), and BACE test input features (X test), see **Fig. 1**. To generate input features of ligands, we compared using structure-based features and ligand-based features, thus the method of obtaining accurate docked is important.

The compilation of training dataset and test input features are discussed in the "**Test dataset compilation for BACE-ligand affinity modeling**" section. Before that, our workflow of generating predicted docked pose is introduced in "**Semi-automated ligand pose generation and docking**" section.

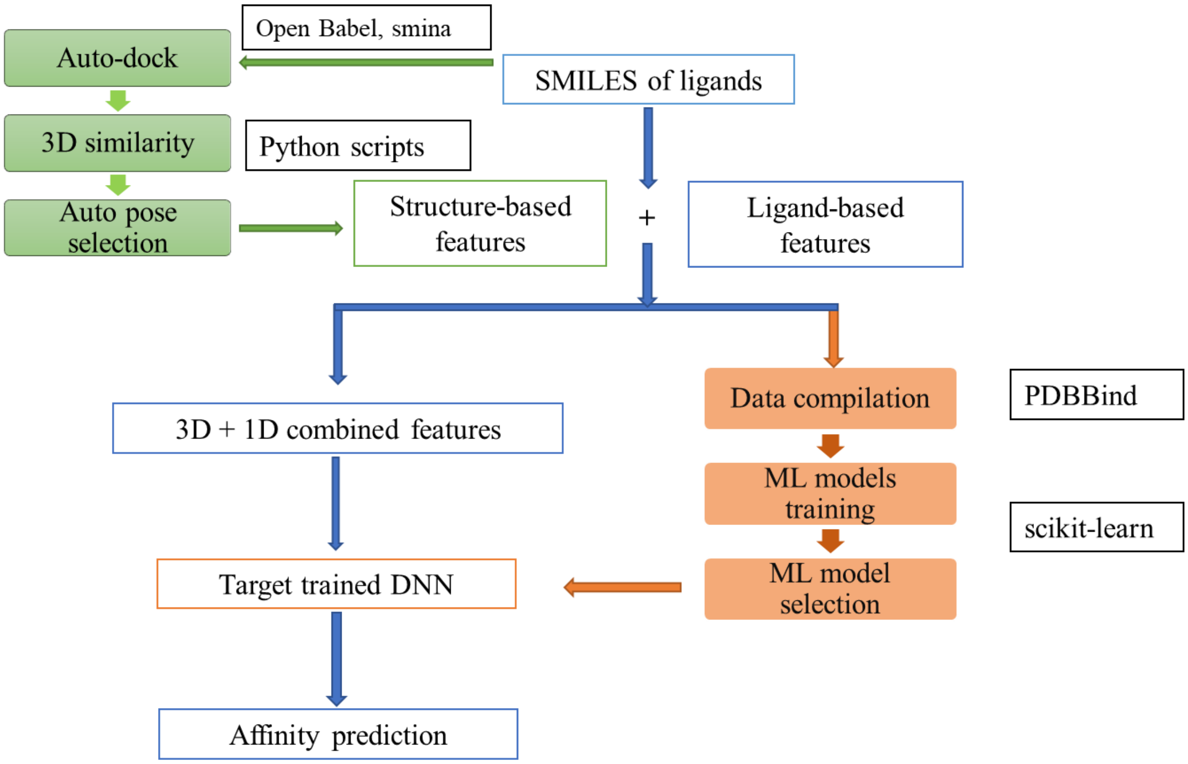

Fig. 1 Workflow for generating machine learning models for affinity prediction. Green boxes indicate the steps in the auto-pose generator for the 154 ligands in D3R GC4 BACE Stage 2 affinity prediction challenge. Orange boxes represent procedures in the BACE-specific machine learning models of affinity prediction.

Semi-automated ligand pose generation and docking

Our first objective was to develop a docking workflow for the 154 ligands without crystal structures. It was shown in D3R GC3 that docking to the appropriate receptor structure is important for success [3]. In GC4 Stage 2, D3R provided SMILES codes of the 154 macrocyclic ligands for the affinity prediction test; 20 co-crystal structures were released earlier in D3R GC4 Stage 1B. We looked for chemical similarity between the 154 ligands (BACE_1 to BACE_158) for the affinity prediction test and the 20 ligands (BACE_1 to BACE_20) with crystal structures using the FragFp method in DataWarrior 4.7.2,[24] see supplemental **Table S1** for the full similarity list. The FragFp descriptor uses chemical fingerprints based on substructure fragment matching.

We cross-docked each of the 154 test ligands to the BACE receptor bound to the ligand with the highest similarity out of BACE_1 to BACE_20. An automated cross-docking python script using

Open Babel[25] and smina[21] was written to generate the docked poses for the 154 test ligands for Stage2. Smina was used for the cross-docking procedure with the Vinardo scoring function [20]. The docking box was centered on the most similar ligand with a reduced buffer padding of 2.0 Å (default 4.0 Å) to confine docking poses in a narrower space. The reduced docking volume helped eliminate false positive poses.

Widely used simply root-mean-square deviation (RMSD) metrics for measuring the 3D similarity for different molecules are not applicable, since atom pairs for different compounds are not identical. Given multiple docked poses for one ligand, a 3D structural similarity score (R_{sim}) inspired by crystallographic electron density difference maps was calculated for each docked pose to the most chemical similar ligand with an available cocrystal structure:

$$R_{sim} = \frac{\sum|EG1-EG2|}{\sum EG1} \qquad \text{(Eq. 1)}$$

where EG denotes the electron grid of a molecule. The electron grid of a molecule is calculated by setting a molecule centered in a box with a user-defined padding distance in each dimension (5 Å used here). The box was treated as a 3D electron grid. To calculate the grid density, each atom of the molecule is treated as a spherical Gaussian distribution of electrons, with the integration of electron density being the atomic number, and with the standard deviation being the van der Waals radii divided by 2.3548.[26] For example, consider two conformers (A and B) of the same compound. If A and B overlap with each other perfectly, R_{sim} equals 0 since $EG_A=EG_B$. If A and B do not overlap at all spatially, R_{sim} would be 2. An example is shown in **Fig. 2**. The Stage 2 affinity test ligand BACE_73 is most similar to BACE_10 which has a known co-crystal structure. Using smina with the Vinardo scoring function, multiple docked poses were generated to the receptor structure from the co-crystal structure of BACE_10. The second-best pose (Vinardo affinity score: -11.2 kcal/mol) and the fourth-best pose (Vinardo affinity score: -8.8 kcal/mol) are shown in **Fig. 2.** Using the R_{sim} method, the 3D-similarity of the BACE_73 second pose was shown to be more 3D-similar to BACE_10 than the BACE_73 fourth pose, as the R_{sim} is lower for the 2nd pose (0.72 versus 1.36).

This 3D similarity algorithm works well when comparing docked poses to cocrystal structures. For all Stage 2 ligands, the docked poses were scored with this method. Docked poses with the lowest R_{sim} value (highest 3D similarity) were picked for further analysis for affinity estimation.

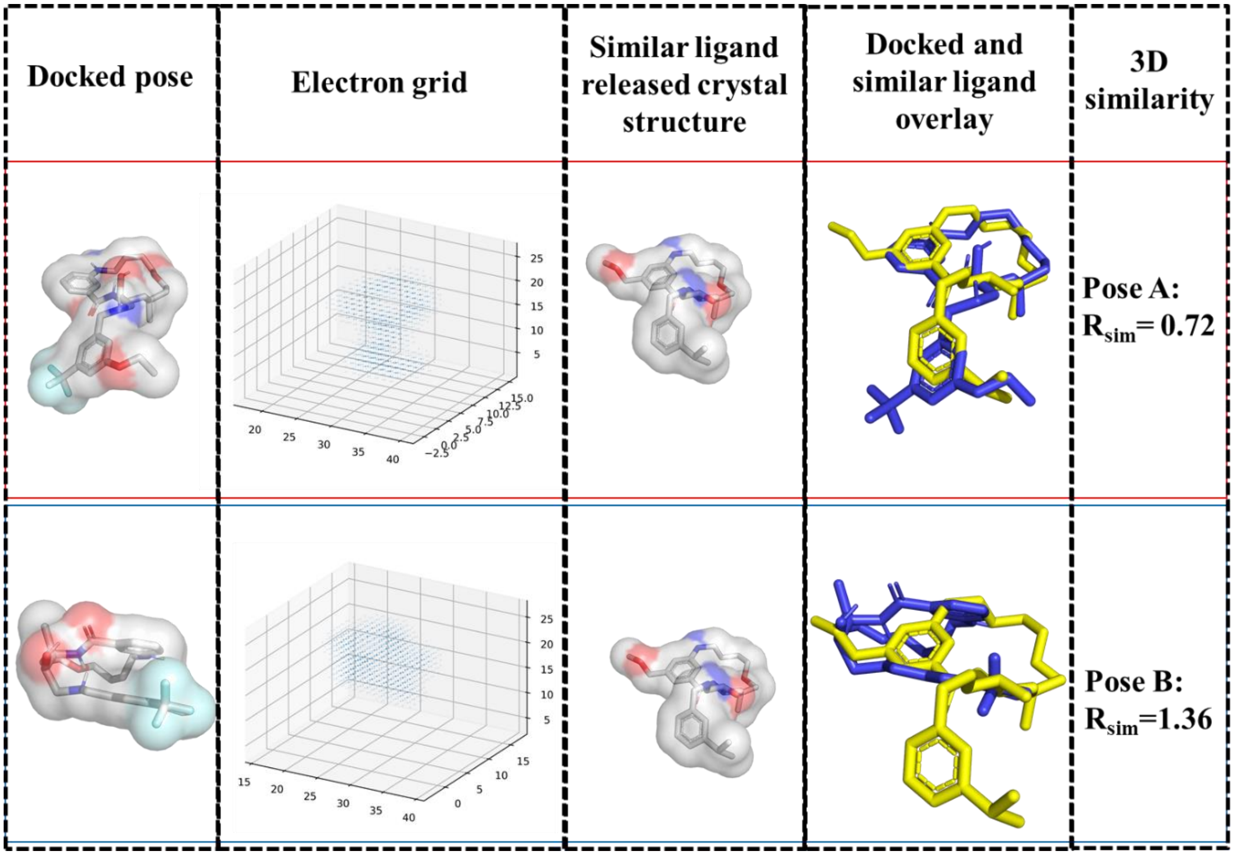
Docked pose
Electron grid
Similar ligand released crystal structure
Docked and similar ligand overlay
3D similarity
Pose A: R_sim= 0.72
Pose B: R_sim=1.36

Fig. 2 3D similarity with the R_{sim} method for D3R GC4 BACE ligands. Using the R_{sim} method, the 3D-similarity of the BACE_73 second pose (top) was shown to be more 3D-similar to BACE_10 than the BACE_73 fourth pose (bottom), as the R_{sim} is lower for the second pose. PyMOL was used for generating part of this figure.[27]

Training dataset compilation for affinity modeling of BACE-ligand

To build a BACE-specific affinity machine learning model, a dataset comprised of 222 published BACE ligand affinities (label value $y_{training}$, dimension: 222×1) was extracted from PDBbind v.2017 which contains 14,761 protein-ligand complexes affinity data, see Table S2.[17]

In this work, structure-based and/or ligand-based features were used as input features (X) for our machine learning models. To generate the structure-based features used for machine learning model training, ten AutoDock Vina-like scoring terms were generated for the 222 BACE-ligands using smina, (**Table 1)**. To obtain the ten scoring values, “scoring_only” functionality was used in smina, without conformation search and docking. The scores from smina were compiled as the training set ($X_{training\_structure\text{-}based}$, dimension: 222×10). To generate the ligand-based chemoinformatic features, DataWarrior 4.7.2 was used to produce twenty-four ligand-based descriptors ($X_{training\_ligand\text{-}based}$, dimension: 222×24) for the 222 training ligands (**Table 2)**. When structure-based and ligand-based features were used together, a full training set ($X_{training}$, dimension: 222×34) was obtained by addition of the two datasets ($X_{training\_structure\text{-}based} + X_{training\_ligand\text{-}based}$).

Table 1 Ten structure-based AutoDock Vina terms generated by smina. Fine parameters o, w, c, g, b, s, i, j, ^ are used in the generation of the terms. Briefly, o is offset between atom pairs, w is the width of the Gaussian function, c is distance cutoff, g is a good distance, b is a bad distance, s is a smoothing, i and j are Lennard-Jones exponents, ^ is the cap.

Term description	**Fine parameters**
Gaussian	o=0, w=0.5, c=8
Gaussian	o=3, w=2, c=8
repulsion	o=0, c=8
hydrophobic	g=0.5, b=1.5, c=8
van der Waals	i=6, j=12, s=1, ^=100, c=8
non_dir_h_bond	g=-0.7, b=0, c=8
non_dir_h_bond_lj	o= -0.7, ^=100, c=8
non_dir_anti_h_bond_quadratic	o=0, c=8
acceptor_acceptor_quadratic	o=0, c=8
donor_donor_quadratic	o=0, c=8

Table 2 Twenty-four ligand-based features generated by DataWarrior.

Ligand-based features	**Ligand-based features**
Molecular weight	Non-H atoms
cLogP	Non-C/H atoms
cLogS	Electronegative atoms
H-acceptors	Stereo-centers
H-donors	Rotatable bonds
Total surface area	Ring closures
Relative PSA	Small rings
Polar surface area	Aromatic rings

Drug-likeness	Aromatic atoms
Shape index	sp^3-Atoms
Molecular flexibility	Symmetric atoms
Molecular complexity	Amides

Test dataset compilation for BACE-ligand affinity modeling

D3R GC4 BACE Stage 2 requires the affinity prediction of 154 ligands. To utilize our BACE specifically trained machine learning model for affinity prediction, the same ten structure-based features ($X_{test\_structure\text{-}based}$, dimension: 154 × 10) and twenty-four ligand-based features ($X_{test\_ligand\text{-}based,}$ dimension: 154 × 24) for the test ligands were generated. To calculate structure-based features for the 154 test ligands, the docked poses for the ligands was generated using the method described previously. Then, the ten AutoDock Vina terms (**Table 1**) was calculated using smina. The ligand-based features for this test set were generated in the same way as the training set, using the twenty-four molecular descriptors (**Table 2**) from DataWarrior 4.7.2.

Construction, training, and tuning of machine learning models for BACE-ligands affinity

After we obtained a training dataset ($X_{training}$) and training affinities ($y_{training}$), five common machine learning models (**Table 3**) were constructed, refined, and compared, using structure-based and/or ligand-based features, in Python using scikit-learn.

To validate and compare machine learning models, the training dataset was first linearly rescaled to facilitate model convergence, then treated with 10-fold cross-validation with the dataset shuffled ("random state" defined as one, for reproducibility). To refine each model, fine-tuning of machine learning models was scrutinized and adjusted. For evaluation, the coefficient of determination (R^2) of 10-fold validations of each model was compared.

Table 3 Machine learning models for BACE-ligand affinity investigated in this study.

Number	Model Type	Fine-tuning parameters adjusted
1	Linear regression	N/A

2	Support vector machine regressor	Kernel functions, regularization (C)
3	Random forest regressor	Number of estimators, max depth
4	Regularized linear regression	Regularization (λ)
5	Deep neural network	Hidden layer size, regularization (λ)

Result and Discussion

Pose generator and docking performance evaluation

In the D3R GC4 BACE Subchallenge, we also submitted pose predictions in Stages 1A (cross-docking for BACE_1 through BACE_20, with no receptor coordinates of the 20 ligands provided) and Stage 1B (redocking for BACE_1 through BACE_20, with receptor coordinates of the 20 ligands provided). We decided on a strategy based on optimizing the AutoDock Vina algorithm and scoring method which has been shown to perform well for pose ranking[7], where docked poses are scored based on a linear combination of five terms:

$E_{inter} = w_1 \times E_{gauss1} + w_2 \times E_{gauss2} + w_3 \times E_{gauss3} + w_4 \times E_{hydrophobic} + w_5 \times E_{hydrogenbond}$ [7, 21] (Eq. 2)

To achieve an improved cross-docking pose prediction performance for BACE specifically, the weights (w_n) in the Vina scoring function (**Equation 2**) were optimized and customized on a training dataset of 24 BACE ligands and test dataset of 229 BACE ligands deposited in the PDB. The weights of the five Vina terms were refined via partial gradient descent of each weight until the overall RMSD reached a local minimum (**Table 4**).

Table 4 Cross-docking performance evaluation for BACE-ligands. The weights (w_n) are the scaling factors in Equation (2). The mean and standard error of RMSD for each method were evaluated on 229 BACE-ligand cocrystal structures deposited in the PDB using smina (receptor PDB used: 4L7G).

	w_1	w_2	w_3	w_4	w_5	Mean(SE) of RMSD (Å)
Vina[7]	-0.035579	-0.005156	0.840245	-0.035069	-0.587439	2.42(10)
Vinardo[a][20]	-0.045	0.000	0.800	-0.035	-0.600	2.55(10)
BACE custom	-0.02558	-0.00516	1.19025	-0.05507	-0.83744	2.33(10)

[a]Vinardo also has modified term functions besides weight.

Although a statistically improved cross-docking performance was achieved for 229 BACE ligands deposited in PDB, we discovered that the Vinardo scoring function yielded the best re-docking performance (lowest RMSD) for the macrocyclic BACE ligands that were the subject of D3R GC4 (**Table 5**). Thus, Vinardo scoring was applied to D3R GC4 BACE Stage 1B for BACE_1 to BACE_20 redocking. The re-docking performance of our method was released by D3R, this method yielded a mean (standard deviation) RMSD of 1.97 (1.55) Å for the 20 BACE macrocyclic ligands, ranking 48th place among 70 submissions. A RMSD below 2.0 Å has traditionally been considered a good result for docking. However, other research teams demonstrated superior cross-docking and re-docking performance during this challenge. We speculate that an additional relaxation and optimization step after docking would improve our docking performance.

Table 5. Redocking performance evaluation for BACE-macrocyclic ligands deposited in PDB. Docking was conducted with smina with Vina/Vinardo/custom scoring methods [20]. Test cocrystal structures were obtained from the PDB. RMSD was calculated and compared using Pymol.

PDB ID	Vina	Vinardo	BACE custom[a]
4KE1	0.185	0.178	0.266
4KE0	0.154	0.136	0.122
4K8S	0.164	0.319	0.374
3K5C	0.913	2.165	2.145
3DV5	0.571	0.988	0.578
3DV1	0.290	0.618	2.608

2F3E	3.844	4.031	2.352
1XS7	2.708	3.410	2.787
Mean	**1.481**	**1.104**	**1.404**
Median	**0.803**	**0.431**	**1.362**
STD	**1.537**	**1.399**	**1.164**

[a]The BACE custom scoring function is shown in Table 4.

After D3R GC4 Stage1B, twenty cocrystal structures of BACE_01:BACE_20 macrocyclic ligands were released. When overlaying the D3R ligands with macrocyclic ligands deposited in PDB, it was found that these ligands shared a similar structural motif in the main site: the macrocycle occupying an empty cavity with a large substituent extended to the other side of the main cavity via a linker that usually contains hydrogen-binding, electronegative functional groups, as shown in **Fig 3**.

Since the Vinardo scoring function was shown to be effective in generating docked poses for macrocyclic BACE ligands (**Table 5**), we used an automated pose selection method to select the best pose from the multiple docked poses generated by smina/Vinardo.

We hypothesized that a chemically similar BACE macrocyclic ligands should share a similar docked pose (154 BACE chemical similarity pairs are provided in **Table S1**). The semi-automated docking and pose selection workflow was described above. Given multiple cross-docked poses of the 154 ligands without crystal structures, the pose with the lowest R_{sim} (defined in **Equation 1**) was selected for structure-based scoring. Two examples are shown in **Fig 4**. BACE_26 (test) is chemically similar to BACE_3 (crystal structure known). A docked pose generated by smina/Vinardo produced a lower R_{sim} = 0.553 and was used for affinity prediction. BACE_137 is similar to BACE_12. A docked pose with R_{sim} = 0.564 was used for further calculation. Using this semi-automated workflow, 154 predicted docked structures were generated for BACE-ligand affinity prediction. We expect that further optimization of the selected pose will improve pose accuracy.

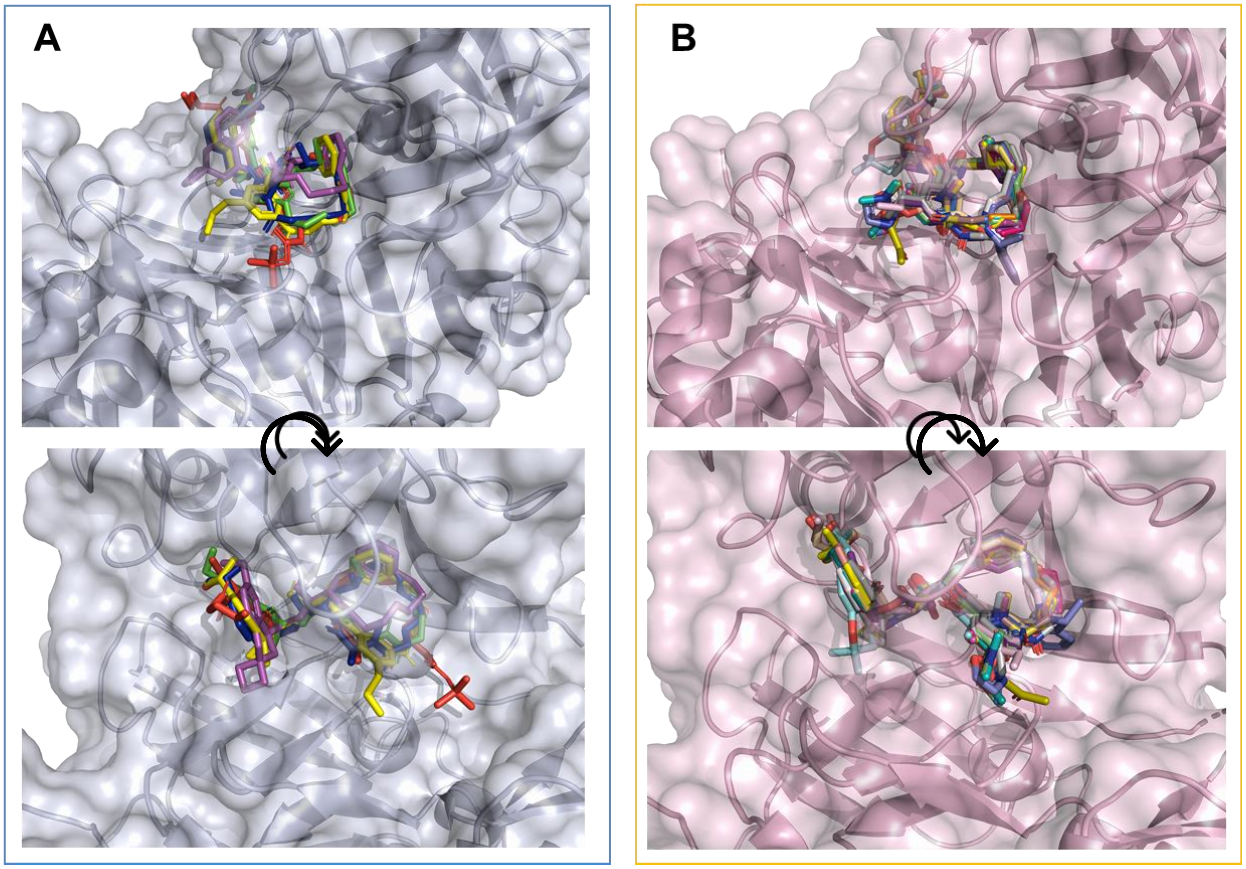

Fig. 3 Structure overlay of macrocyclic ligands of BACE. A, five published ligands structures deposited in the PDB: 2F3E-AXQ(red), 3DV1-AR9(green), 3DV5-BAV(blue), 3K5C-OBI(yellow), 4KE0-1R8(magenta), receptor PDB used: 4KE0.[16] B, twenty D3R GC4 released macrocyclic ligands (BACE_1 to BACE_20, receptor used BACE_BA01. Top and bottom figures show rotated views of the same structure. PyMOL was used to generate this figure.

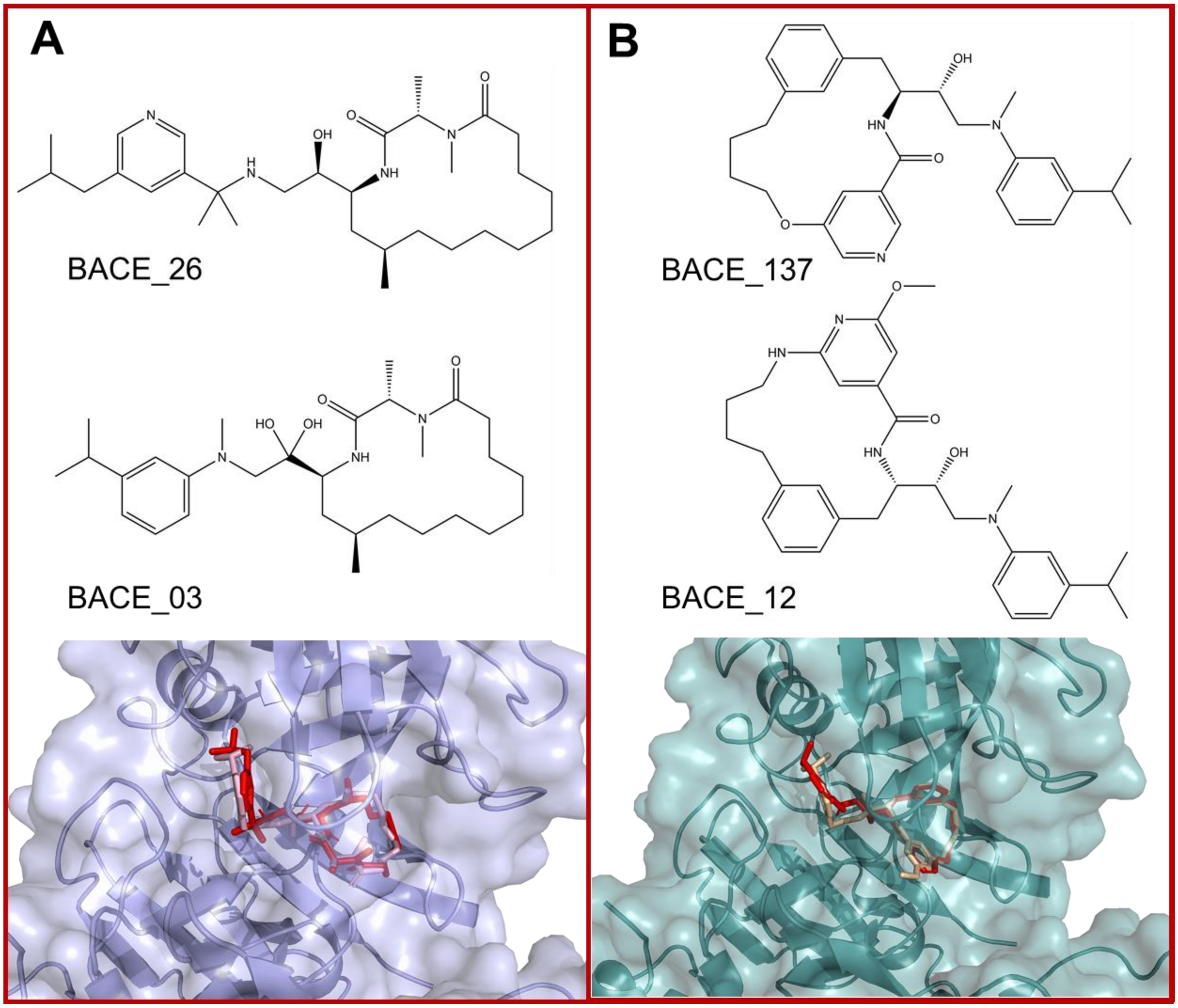

Fig. 4 Chemical structures and structural overlays of two sets of docked macrocyclic ligands in D3R GC4 BACE Stage 2. A, BACE_26 (red) was found to be most chemically similar to BACE_3 (light pink), that has a released cocrystal structure, via the FragFp method in DataWarrior. Between two docked poses of BACE_26 generated by Vinardo scoring using smina, the pose with the lowest R_{sim} = 0.553 was used for structure-based affinity calculations. B, similarly, ligand BACE_137 (red) was docked to mimic the 3D structure of the most chemically similar ligand BACE_12 (light orange); the pose with the lowest R_{sim} = 0.564 was used for affinity modeling. ChemBioDraw and PyMOL were used to generate this figure.

Comparisons between machine learning models of BACE-ligand affinities

We obtained a dataset ($X_{training}$) of 222 BACE-ligands deposited in PDB with their affinities ($y_{training}$) extracted from PBDBind v2017[17]. We investigated three aspects of applying machine learning in BACE-ligand affinity prediction: input feature selection, machine learning model selection, and machine learning model regularization.

First, for feature selection, we compared the model performance based on structure-based features only **(Table 1**), ligand-based features only (**Table 2**), and a combination of both using five machine learning models (**Table 3**), linear regression, regularized linear regression, random forest regressor (RF), support vector machine regressor (SVM), and deep neural network/multilayer perceptron (DNN/MLP).

When only structure-based terms were used for affinity modeling (**Fig. 5A**), linear regression (LR) and regularized LR yielded poor performance, with low coefficients of determination (R^2) at 0.3(7), 0.28(8) for the ten-fold cross-validation set ($X_{training}$). This indicates the limitations of linear customization of Vina terms to predict ligand affinity. We point out that multilinear regression is still the most commonly used method for developing docking and affinity scoring functions. Using random forest regressor (RF, number of estimators: 100, maximum depth: 1), support vector machine regressor (SVM, kernel: Gaussian, regularization: C = 2), and multiple layer perceptron (MLP, layer structure: 10 × 8 × 8, activation function: ReLU, L2 regularization: alpha = 10) with structure-based only terms yielded reasonable performance after fine-tuning the hyperparameters. The SVM regressor model performed well and achieved an R^2 of 0.65(11) for the 10-fold cross-validation set, and 0.77(1) for the same 10-fold cross-training set.

The performance of only using ligand-based terms was also evaluated (**Fig. 5B)**. Interestingly, all five machine learning models yielded comparable results, suggesting that linear regression adequately utilizes most of the information in the input features. The SVM performed the best among the models, with an R^2 of 0.62(13) for the 10-fold cross-validation set and 0.836(9) for the cross-training set.

When a combination of structure-based terms and ligand-based terms was utilized in machine learning models, slightly improved performance was obtained across the different modeling methods (**Fig. 5C**). LR and regularized LR yielded equivalent performance, with R^2 = 0.59(13) for the 10-fold cross-validation set. RF (number of estimators: 200, maximum depth: 8) yielded an R^2 of 0.61(19) for the cross-validation set. SVM regressor (kernel: Gaussian, regularization: C = 2.5) had a R^2 = 0.64(17) for the validation set. We compared DNN/MLP architecture with 10 × 8 × 8, 8 × 10 × 10, 8 × 8, and 10 (**Table 6)**. Only four networks were explored given limits on our time. It was found that MLP model performance is slightly affected by the choice of the number of layers and numbers of neurons in layers, as R^2 of 0.65 were achieved in all four architectures. Careful hyperfine tuning is necessary to obtain an effective model. For example, in the 10 × 8 × 8 DNN, an alpha of 0.01 overfit the training set and yielded a poor R^2 (0.3±0.4) of the ten-fold cross-validation sets; when alpha of 50 was used, the model was underfitted, with poor R^2 (0.54±0.15).

For the D3R GC4 BACE Stage 2 affinity predictions, we selected the refined 10 × 8 × 8 DNN model due to its superior performance over other architectures. The hyperfine tuning of this model is shown in **Fig. 6A**. When the regularization parameter alpha was 10, the model achieves optimal performance for the ten-fold cross-validation, with Pearson's correlation (R) = 0.82. Using this model, the predicted versus experimental affinities for the whole training dataset are shown in **Fig. 6B** (the last cross-validated model in the 10-fold cross-validation dataset). This model exhibited very good correlation metrics for BACE affinity prediction. It greatly outperforms the published performance of K_{DEEP}, RF-Score, X-Score, and Cyscore, with their R equal to -0.06, -0.14, -0.12, and 0.2, respectively to 36 BACE ligands[19]. The architecture and mapping matrix (W_n) are represented in **Fig. 6C**. Every neuron in the MLP take a linear combination of earlier neurons as input, and output after ReLU (rectified linear unit) activation.

D3R released the performance of GC4 BACE Stage 2. Our results tied for the fourth best performance with a Kendall's $\tau = 0.30(5)$ and Spearman's $\rho = 0.43(7)$.

Table 6 Evaluation of different layer structures of MLP models for the 10-fold cross-validation set for the 222 BACE-ligands dataset ($X_{training}$). Activation function: ReLU, L2 regularization parameter was adjusted.

Regularization (alpha)	**MLP (10 × 8 × 8)**	**MLP (8 × 8 × 10)**	**MLP (8 × 8)**	**MLP (10)**
0.01	0.3(4)	0.3(4)	0(2)	0.3(5)
0.1	0.4(2)	0.3(3)	0(1)	0.4(5)
1	0.4(4)	0.5(4)	0.6(2)	0.6(3)
10	0.67(13)	0.66(14)	0.66(16)	0.65(16)
20	0.61(17)	0.61(17)	0.62(17)	0.61(19)
50	0.54(15)	0.54(15)	0.57(15)	0.59(17)

Numpy and scikit-learn were used to generate the values in this table.

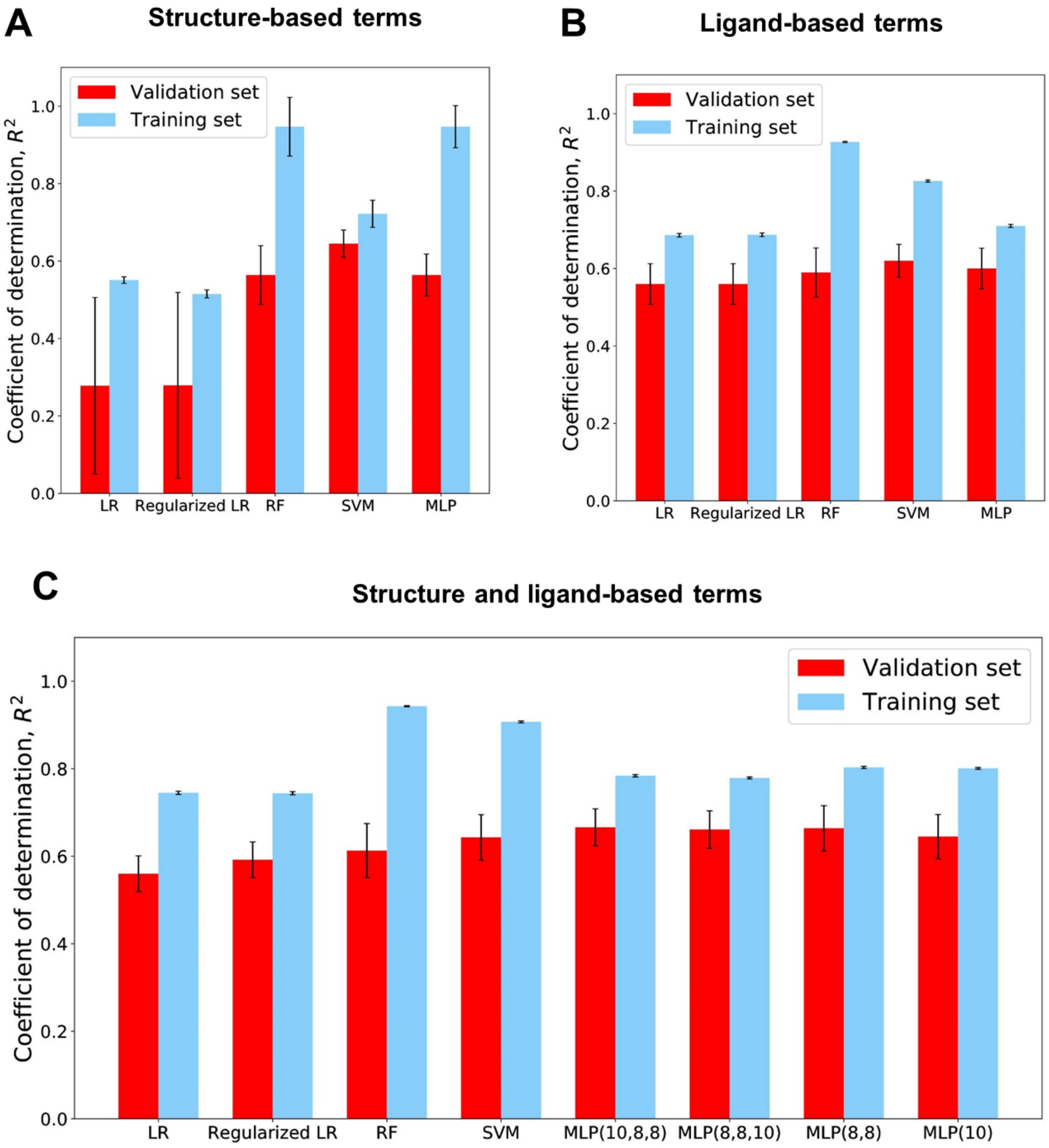

Fig. 5 Performance of machine learning models of BACE affinity. Red boxes indicate the mean of R^2 for the 10-fold cross-validations, cyan boxed indicate the mean of R^2 for the training set. The error bars represent the standard error (SE) for the 10-fold cross-validation metric evaluation. A, only the ten structure-based Vina terms were used. B, only the twenty-four ligand-based terms (Table 2) were used. C, both the structure-based and ligand-based terms were applied. Numpy, Matplotlib, and scikit-learn were used to generate this figure.

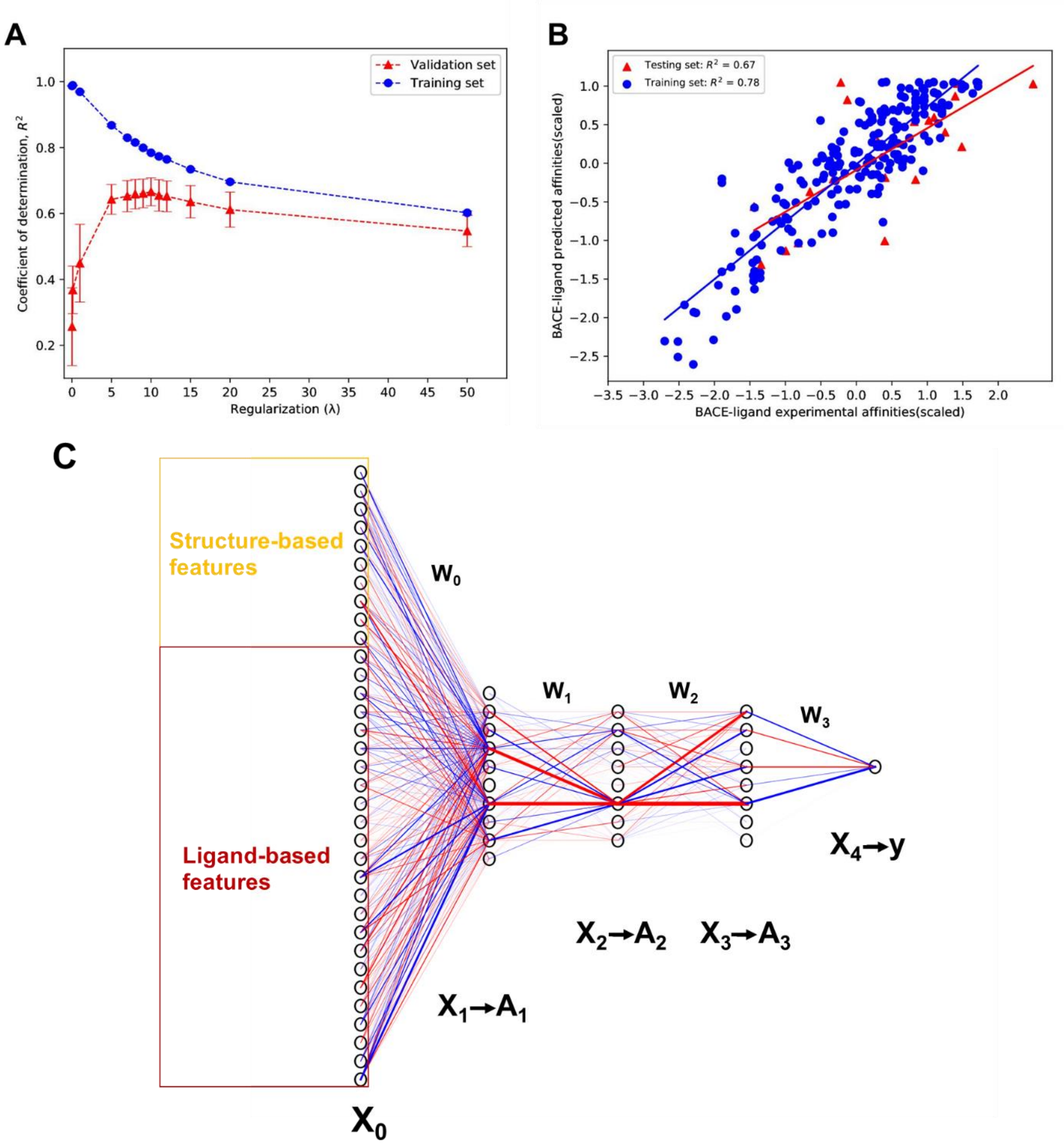

Fig. 6 The regularization, fitting, and representation of our DNN/MLP model submitted for D3R GC4 BACE Stage 2. Model parameters: layer structure: 10 × 8 × 8, activation function: ReLU, regularization: alpha = 10. A, MLP model performance (R^2) over variations of a range of L2 regularization alpha (0.01-50.0). B, BACE affinity fitting ability of this MLP model. Blue dots are for the last training set in the 10-fold cross-training, and red spots are for the last validation set in 10-fold cross-validation. C, the architect and mapping coefficient of this MLP. Redlines indicate positive coefficients and blue lines indicate negative coefficients, and the thickness indicates the absolute value of coefficients connecting neurons. Numpy, Matplotlib, and scikit-learn were used to generate this figure.

Conclusions

Through participating in the D3R GC4 BACE Subchallenge, we investigated optimizing the AutoDock Vina docking scoring function and designed and compared machine learning models using a combination of structure-based and ligand-based terms. To generate docked poses for BACE macrocyclic ligands, a 3D similarity pose automated selection script was shown to be effective in generating accurate docked poses. Five different machine learning models were explored ranging in complexity from linear regression to a deep neural network. From tuning different models, we found that hyperparameter tuning greatly affects the accuracy of protein-ligand affinity prediction.

This work shows that machine learning models are highly effective for protein-ligand affinity prediction if high-quality training datasets are available for the target protein. We found that the Vinardo scoring function, developed from a broad set of ligands, performed best for docking macrocyclic ligands to BACE. We expect docking performance can be further improved by careful choice and optimization of receptor structures. In contrast, a deep neural network trained specifically on BACE ligands performed best for affinity prediction. Affinity prediction can probably be improved by training on larger datasets, training on ligand/target-specific datasets, using deeper neural networks and adopting advanced neural networks such as convolutional neural networks, automated tuning of hyperparameters, and carefully selecting a larger set of informative input features.

Acknowledgments

This work was funded by NSF CAREER MCB 1833181.

References

1. Gathiaka S, Liu S, Chiu M, et al (2016) D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. J Comput Aided Mol Des 30:651–668. https://doi.org/10.1007/s10822-016-9946-8

2. Gaieb Z, Liu S, Gathiaka S, et al (2018) D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. J Comput Aided Mol Des 32:1–20. https://doi.org/10.1007/s10822-017-0088-4

3. Gaieb Z, Parks CD, Chiu M, et al (2019) D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings. J Comput Aided Mol Des 33:1–18. https://doi.org/10.1007/s10822-018-0180-4

4. Vassar R, Bennett BD, Babu-Khan S, et al (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 286:735–741

5. Bajorath J (2015) Computer-aided drug discovery. F1000Research 4:. https://doi.org/10.12688/f1000research.6653.1

6. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD (2015) Molecular Docking and Structure-Based Drug Design Strategies. Molecules 20:13384–13421. https://doi.org/10.3390/molecules200713384

7. Trott O, Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem NA-NA. https://doi.org/10.1002/jcc.21334

8. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256

9. Ravindranath PA, Forli S, Goodsell DS, et al (2015) AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. PLoS Comput Biol 11:1–28. https://doi.org/10.1371/journal.pcbi.1004586

10. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430

11. Taylor R, Cole J, Cosgrove D, et al (2012) Development and validation of an improved algorithm for overlaying flexible molecules. J Comput Aided Mol Des 26:451–472. https://doi.org/10.1007/s10822-012-9573-y

12. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 16:11–26

13. Khamis MA, Khamis, Mohamed A. M (20150301) Machine learning in computational docking. Artif Intell Med 63:135–152

14. Abadi M, Agarwal A, Barham P, et al (2016) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv160304467 Cs

15. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825−2830

16. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

17. Liu Z, Su M, Han L, et al (2017) Forging the Basis for Developing Protein–Ligand Interaction Scoring Functions. Acc Chem Res 50:302–309. https://doi.org/10.1021/acs.accounts.6b00491

18. Ballester PJ, Mitchell JBO (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinforma Oxf Engl 26:1169–1175. https://doi.org/10.1093/bioinformatics/btq112

19. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G (2018) KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. J Chem Inf Model 58:287–296. https://doi.org/10.1021/acs.jcim.7b00650

20. Quiroga R, Villarreal MA (2016) Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. PLoS ONE 11:1–18. https://doi.org/10.1371/journal.pone.0155183

21. Koes DR, Baumgartner MP, Camacho CJ (2013) Lessons Learned in Empirical Scoring with smina from the CSAR 2011 Benchmarking Exercise. J Chem Inf Model 53:1893–1904. https://doi.org/10.1021/ci300604z

22. Li H, Leung K-S, Wong M-H, Ballester PJ (2015) Improving AutoDock Vina Using Random Forest: The Growing Accuracy of Binding Affinity Prediction by the Effective Exploitation of Larger Data Sets. Mol Inform 34:115–126. https://doi.org/10.1002/minf.201400132

23. Ashtawy HM, Mahapatra NR (2012) A Comparative Assessment of Ranking Accuracies of Conventional and Machine-Learning-Based Scoring Functions for Protein-Ligand Binding Affinity Prediction. IEEE/ACM Trans Comput Biol Bioinform 9:1301–1313. https://doi.org/10.1109/TCBB.2012.36

24. Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. J Chem Inf Model 55:460–473. https://doi.org/10.1021/ci500588j

25. O’Boyle NM, Banck M, James CA, et al (2011) Open Babel: An open chemical toolbox. J Cheminformatics 3:33. https://doi.org/10.1186/1758-2946-3-33

26. Alvarez S (2013) A cartography of the van der Waals territories. Dalton Trans 42:8617–8636. https://doi.org/10.1039/C3DT50599E

27. Schrodinger, LLC. PYMOL, The PyMOL Molecular Graphics System, Version 2.0