AK-score: Accurate protein-ligand binding affinity prediction using the ensemble of 3D-convolutional neural network

Yongbeom Kwon¹, Woong-Hee Shin²*, Junsu Ko³*, and Juyong Lee¹*

1. Department of Chemistry, Kangwon National University, Gangwon-do, Chuncheon, 24341, Republic of Korea

2. Department of Chemical Science Education, Sunchon National University, Jeollanam-do, Suncheon, 57922, Republic of Korea

3. Arontier, 241 Gangnam-daero, Seocho-gu, Seoul, 06735, Republic of Korea

KEYWORDS: Protein-ligand binding affinity, Convolutional Neural Network, ResNext, Deep-learning, binding affinity prediction, docking score

ABSTRACT

Accurate prediction of the binding affinity of a protein-ligand complex is essential for efficient and successful rational drug design. In this work, a new neural network model that predicts the binding affinity of a protein-ligand complex structure is developed. Our new model predicts the binding affinity of a complex using the ensemble of multiple independently trained networks that consist of multiple channels of 3D convolutional neural network layers. Our model was trained using the 3740 protein-ligand complexes from the refined set of the PDBbind database and tested
using the 270 complexes from the core set. The benchmark results show that the correlation coefficient between the predicted binding affinities by our model and the experimental data is higher than 0.72, which is comparable with the state-of-the-art binding affinity prediction methods. In addition, our method also ranks the relative binding affinities of possible multiple binders of a protein quite accurately. Last, we measured which structural information is critical for predicting binding affinity.

**Introduction**

Predicting the binding affinity of a protein-ligand complex plays a central role in drug design and discovery. In order for a molecule to be a lead molecule for drug discovery, generally, it is required to bind with a target protein tightly. However, the experimental measurement of protein-ligand binding affinity is difficult and time-consuming, which is one of the major bottlenecks of the drug discovery process. If one can predict the affinity of a specific ligand to a target protein quickly and accurately, the efficiency of virtual screening would be significantly improved. Thus, to accelerate the drug discovery process, many computational binding affinity prediction methods have been developed.

Generally, computational methods for binding affinity prediction are classified into three categories: 1) physics-based, 2) empirical, and 3) knowledge-based methods. The first approach mainly uses theoretically rigorous binding free calculations based on forcefield models. The strongest advantage of physics-based methods is that, with the help of the state-of-the-art forcefield models, they predict the binding free energies of arbitrary small-molecule ligands to a protein with an average error of about 1 kcal/mol\(^1\text{–}^4\). However, rigorous free energy calculations require a significant amount of computational resources. With state-of-the-art molecular dynamics (MD)
code running on a graphics processing unit, binding free energies of only one or two ligands can be calculated depending on the sizes of a ligand and a protein. Such computational burden hinders the use of MD and free energy calculations for high-throughput screening of drug-like molecules.

The empirical scoring functions have been extensively used in many protein-ligand docking programs and virtual screening processes\textsuperscript{5,6}. They approximate protein-ligand interactions using equations consisting of several physics-based terms, mimicking van der Waals interaction, solvation free energy, electrostatic interactions, and etc. The parameters of the physics-based terms are generally fitted with experimental data to reproduce measured binding affinity values. Because of the simplicity of calculation and their close relationship with physics-based interactions, empirical scoring functions are still actively developed.

Recently, due to the recent emergence of deep-learning methods, more accurate data-driven predictions have become possible in various scientific disciplines\textsuperscript{7,8}. For protein-ligand binding affinity prediction, many deep-learning-based methods have been suggested\textsuperscript{9–12}. Ragoza et al. suggested a small network consisting of three sequential layers of 3D convolutional neural network (3D-CNN) with pooling layers\textsuperscript{11}. Similarly, Stepniewska-Dziubinska et al. developed a binding affinity prediction model consisting of three consecutive 3D-CNN layers followed by three dense layers\textsuperscript{12}. Jiménez et al.\textsuperscript{9} developed a binding affinity prediction model, $K_{\text{deep}}$, based on the SqueezeNet architecture\textsuperscript{13}, which was originally designed for image classification. The $K_{\text{deep}}$ model consists of multiple 3D-CNN with about 1.3 million parameters. Zhang et al.\textsuperscript{14} developed the DeepBindRG model, which uses the 2D representation of a protein-ligand interface and the ResNet architecture\textsuperscript{15}. Similarly, Zheng et al.\textsuperscript{10} also converted a protein-ligand binding structure into a 2D tensor with a single channel and processed it through three 2D-CNN layers and four dense layers.
In this paper, we present a new protein-ligand binding affinity prediction model inspired by the ResNext architecture\textsuperscript{16}, which was originally developed for accurate image classification. In addition to using a novel network architecture compared to previous models, we demonstrate that an ensemble-based approach, using an average of multiple predictors instead of a single predictor, significantly improves prediction quality. The advantage of an ensemble approach is that it does not require further modification of network architectures and can be readily applicable to most existing models. The benchmark results using the CASF-2016 dataset show that the performance of our model is comparable to the best existing scoring functions. We also analyzed relative feature importance to gain insights on which physical properties are most essential in determining binding affinity.

**Methods**

**Data set**

The protein-ligand binding affinity data for training and testing the network was adopted from the PDBBind database\textsuperscript{17,18}. As of August 2018, the refined set of the database contains the experimental binding affinities of 3767 protein-ligand complexes. Among them, the manually curated core set containing 290 high-resolution data was used for testing. The remaining 4055 complexes grouped as the refined binding affinities were used as a training set.

**Convolutional Neural Network**

To utilize the power of the convolutional neural networks, the structures of protein-ligand complexes were represented as three-dimensional (3D) grids. For each protein-ligand complex, the center of mass of the complex was set to the origin and the neighboring atomic environment
was embedded into a 3D-grid whose edge length is 30 Å. Along the X, Y, and Z-axes, 30 grid boxes were generated with a spacing of 1.0 Å. To capture the pattern of protein-ligand interactions, for each grid box, atomic number density was calculated using the following density function.

\[ n(r) = 1 - \exp \left[ 1 - \left( \frac{r_{vdw}}{r} \right)^{12} \right], \]

where \( r_{vdw} \) is the van der Waals radius of an atom and \( r \) is a distance from an atom to the center of a grid.

Atoms were classified into 8 classes, and they were represented as the different channels of the input data. We treated atoms from proteins and ligands separately, which leads to 16 real-valued channels representing the aggregated number density of each protein-ligand complex. The description of the atom-types used in this study is listed in Table 1.

**Table 1.** Atom types used to classify atoms forming protein-ligand binding sites.

| Atom type               | Definition                                                                 |
|-------------------------|-----------------------------------------------------------------------------|
| Hydrophobic             | Aliphatic or aromatic C                                                     |
| Aromatic                | Aromatic C                                                                  |
| Hydrogen bond donor     | Donor 1 H-bond or Donor S Spherical H with NA, NS, OA, OS, SA              |
| Hydrogen bond acceptor  | Acceptor 1 H-bond or S Spherical N; Acceptor 2 H-bonds or S Spherical O;   |
|                         | Acceptor 2 H-bonds S                                                       |
| Positive                | ionizable Gasteiger positive charge                                         |
| Negative                | ionizable Gasteiger negative charge                                         |
Metallic | Mg, Zn, Mn, Ca, or Fe
---|---
Excluded Volume | All atom types

To reduce the orientation dependency of a complex structure, we augmented the number of data set by rotating a grid with all 24 possible rotational operations.

Network architecture

![Network architecture diagram](image)

**Figure 1.** (a) The overall architecture of the network. The complete network mainly consists of 14 stacked layers of an ensemble-based residual block. (b) The structure of each residual block is illustrated. A residual block consists of three stacks of convolutional layers combined with batch normalization and ReLU activation layers. In the middle of the block, each channel is distributed to 16 3D-convolutional layers and processed in a parallel way. Conv3D: 3D convolutional neural network layer, BN: Batch Normalization layer, RL: Residual Layer.
The main component of our network is an ensemble-based residual network, which was used in the RexNext model for image recognition. The overall structure of the network is illustrated in Figure 1 and the structure of each residual block is shown in Figure 1-(b). At each residual block, each channel is distributed to multiple convolutional layers and processed in a parallel way. The number of parallel residual networks is also called cardinality. In this study, we used 16 Conv3D layers for each residual block. We call our network architecture, AK-score (Arontier-Kangwon docking scoring function).

The ReLU activation function was used. All weight parameters were initialized with the He_normal initialization scheme. The model loss was calculated with mean absolute error (MAE) between the experimental and predicted binding affinities in a kcal/mol unit. For parameter optimization, the Adam optimizer was used with the following parameters, beta-1 = 0.99 and beta-2=0.999. The model was trained with multiple learning rates to test the effect of the learning rate on the final prediction quality. Learning rates of 0.0001, 0.0005, 0.0007, and 0.0010 were tested. While training the network, the whole dataset was randomly permuted.

**Ensemble prediction**

To enhance prediction accuracy, we employed an ensemble prediction scheme, obtaining the final prediction value from the average of multiple independently trained models. In many machine-learning tasks, the parameters of each prediction model are optimized from initial random values. When the number of parameters is large, the final parameter set does not converge generally. To reduce possible such biases, we trained multiple networks independently and checked whether the average of multiple predictions yields better predictions.

**Performance Assessment**
To assess the performance of our model, we compared our model with the previously suggested 3D-CNN based binding affinity prediction model. We implemented the K_deep model, which is based on SqueezeNet architecture, which was used for image classification. The model was trained with the same parameters reported in the reference. All models reported in this study are implemented in Keras-2.2.4 with Tensorflow-1.13.1 backend\textsuperscript{19}.

We compared the performance of models using multiple criteria. The CASF-2016 benchmark set has been used as a common ground for a comparison of various docking models\textsuperscript{18}. The accuracy of protein-ligand docking prediction is generally measured in three aspects: 1) scoring, how well predicted binding affinity values are correlated with experimental values? 2) ranking, how well relative binding propensities are predicted? 3) docking, whether correct docking poses are correctly identified to have lower energy than decoys? To assess the scoring power of a model, the Pearson correlation coefficient is calculated. To assess the ranking power, Spearman, Kendall tau, and Predictive index (PI) values are used. For the assessment of docking power, whether the ligand with the true binding ligand is included in the top1, top2 and top3 percent predictions of possible decoys.

**Results & Discussion**

**Binding affinity prediction accuracy**

We performed three different types of experiments based on AK-score architecture: AK-score-single, AK-score-small, and AK-score-ensemble. AK-score-single uses a single prediction network as shown in Figure 1. AK-score-small uses only 10 features, which exclude relatively sparse features than the others: positive, negative, and metallic (Table 1). AK-score-ensemble uses an average of 20 independently trained networks as the final prediction value. The mean absolute
error (MAE) and root mean squared error (RMSE) between predicted and experimental values of various models trained with different learning rates are listed in Table 2. For comparison, the results of the other 3D-CNN-based deep-learning model, the K_{deep} model, implemented by our group are also provided.

The benchmark results show that the AK-score-ensemble model yields the most accurate prediction results. Among the tested models, AK-score-ensemble has the lowest accuracy metric values with an MAE of 1.01 kcal/mol and an RMSE of 1.29 kcal/mol. Compared with a single model, the average errors of the ensemble model are lower by about 0.1 kcal/mol. In addition, when compared with the K_{deep} model, AK-score-ensemble has a lower average error by 0.2 kcal/mol. Our results also show that choosing the best learning rate improves MAE by about 0.05 kcal/mol and RMSE by 0.10 kcal/mol.

Table 2. Assessment of prediction accuracy of ResNext-ensemble, ResNext, and K_{deep} using the PDBbind-2016 dataset and mean absolute error and root mean square error metrics.

| Model            | Learning Rate | MAE     | RMSE    |
|------------------|---------------|---------|---------|
| K_{deep}         | 0.0001        | 1.131   | 1.462   |
|                  | 0.0005        | 1.200   | 1.519   |
|                  | 0.0006        | 1.164   | 1.534   |
|                  | 0.0010        | 1.219   | 1.536   |
| AK-score single  | 0.0001        | 1.159   | 1.511   |
|                  | 0.0005        | 1.101   | 1.415   |
|                  | 0.0007        | 1.130   | 1.425   |
|                  | 0.0010        | 1.110   | 1.406   |
We also assess the performance of our models based on three criteria used in the CASF-2016 dataset, scoring, ranking, and docking power (Table 2). In all three criteria, the AK-score-ensemble model shows the best performance. Scoring power is indicated by the Pearson correlation coefficient between predicted and experimental values. Overall, AK-score models result in higher correlation values than the $K_{\text{deep}}$ model. Among all the tested models, only AK-score-ensemble yields a correlation coefficient value higher than 0.8. In terms of ranking power, the AK-score models outperform the $K_{\text{deep}}$ model on average. Among them, the AK-score-ensemble model results in the highest rank correlation coefficients. For docking power, the difference in prediction performance of the $K_{\text{deep}}$ and the AK-score-single model is not as prominent as in the other criteria. However, the prediction results of the AK-score-ensemble model are clearly better than those of the $K_{\text{deep}}$ model.

**Table 3.** A comparison of prediction accuracy with the CASF-2016 dataset.

| Model          | Scoring | Ranking | Docking |
|----------------|---------|---------|---------|
|                | learning rate | Pearson (R) | Spearman (SP) | Kendall (tau) | Predictive (PI) | Top1 (%) | Top2 (%) | Top3 (%) |
| $K_{\text{deep}}$ | 0.0001  | 0.738   | 0.539   | 0.435   | 0.559   | 24.8      | 38.5      | 52.2      |
|                | 0.0005  | 0.709   | 0.486   | 0.389   | 0.535   | 29.1      | 39.9      | 49.6      |
|                | 0.0006  | 0.701   | 0.528   | 0.439   | 0.558   | 29.1      | 39.9      | 49.6      |
|                | 0.0010  | 0.715   | 0.479   | 0.400   | 0.492   | 24.8      | 36.3      | 44.6      |
| AK-score single | 0.0001  | 0.719   | 0.572   | 0.456   | 0.600   | 34.9      | 48.6      | 56.1      |
An ensemble of networks improves the quality of prediction

We investigated the change of prediction quality by the number of networks. The results show that overall the prediction accuracy increases as more networks are used. In all three quality metrics, scoring, ranking and docking power, the accuracy rapidly increases from a single network to five networks. In terms of the scoring power, the Pearson R correlation coefficient between experiments and predictions increases until the number of networks reaches 25 (Figure 3-(a)). When a single network is used, the correlation coefficient is 0.74. However, when the average of five networks is used, the value becomes higher than 0.80. After 10 networks, the improvement becomes modest but is kept until 25 networks are used. Similarly, the ranking power keeps improving until 25 networks are used. All three ranking measures, SP, tau and PI, are improved consistently when the ensemble average of networks (Figure 3-(b)). These results clearly show that using the ensemble of prediction networks significantly improves prediction quality, which is a simple and straightforward way to improve prediction accuracy without further exploration of various network architectures.
Figure 2. The change of prediction quality by the size of the ensemble of networks. (a) Scoring power is measured by the Pearson correlation coefficient between the experimental and predicted binding affinities. (b) Ranking power is measured by three rank correlation coefficients, Spearman (SP), Kendall tau (tau), and Predictive Index (PI).

Comparison with existing scoring functions

The benchmarking result of AK-score shows that its prediction accuracy is comparable with the best existing scoring functions based on the CASF-2016 dataset. The CASF-2016 dataset provides the pre-calculated prediction results of known scoring functions, which allows a fair comparison of our model with existing scoring functions using the exact same test set. In terms of scoring power, the highest correlation coefficient obtained with AK-score is 0.828, which is a little higher than the best value obtained with vina-RF20, 0.816. In addition, the ranking power of AK-score, 0.736, is slightly worse than the best value by vina-RF20, 0.761, which corresponds to the second place among the tested scoring functions.
Figure 3. Benchmark results of AK-score with existing protein-ligand binding affinity scoring functions. (a) AK-score shows the highest scoring power than existing scoring functions. (b) AK-score is ranked right after the known best performing function vina-RF20.

Comparisons between the experimental binding affinities and predicted values obtained with AK-score-ensemble, X-score, and Autodock vina are displayed in Figure 4. Overall, the results of the AK-score-ensemble model correlates well with experimental values compared to all tested complexes with a correlation coefficient of 0.827. On the other hand, the X-score and Autodock results are showing biases clearly. On average, the X-score significantly underestimates absolute binding affinities, which is indicated by a small slope coefficient of a regression line. The Autodock-vina results are showing a better correlation than X-score, but they are also underestimated than the experimental values to some degree. In summary, it is clear that AK-score-ensemble outperforms the widely used empirical scoring functions in absolute binding
Affinity prediction.

Figure 4. The scatter plots of the experimental binding affinities and prediction results obtained with (a) AK-score-ensemble, (b) Autodock vina, and (c) X-score are depicted.

Assessment of feature importance

To obtain chemical and biological insights from the trained networks, it is necessary to identify which atomic features play important roles in determining the binding affinity of a protein-ligand complex. To achieve this goal, we performed additional experiments that perform predictions by 1) making the values of a specific channel zero or 2) randomly shuffling the values of a channel. The rationale of the second experiment is based on the conjecture that making all values of a channel zero may be too drastic loss of information and conserving the average and variance of values of a channel may be important for making reasonable predictions.

Overall, from both experiments, the excluded volumes of a ligand and a binding-site are identified to be the most important features in determining the binding affinity of a protein-ligand complex (Figure 4). In other words, the shape complementarity between a binding-site and a ligand is most important in determining the binding affinity of a complex. When excluded volume information of a ligand is missing, the average binding affinity prediction accuracy deteriorates by 1.4 kcal/mol (Figure 4-(a)). Following the excluded volume information, the hydrophobic atom
The information of a ligand and a binding-site is identified to be the second important factor. For a binding-site, the hydrogen acceptor atoms of a binding-site play the third important role. Interestingly, for ligands, aromatic atoms play the third important role. In the shuffling experiment, the overall trend is similar to that of the zeroing experiment, but the average decrease in prediction accuracy is smaller (Figure 4-(b)). The most prominent difference is that, for a binding-site, the relative importance of the hydrogen bond acceptor atoms becomes larger than that of hydrophobic atoms.

**Figure 5.** Feature importance calculation results measured by the loss of prediction accuracy, the increase of mean absolute error of predictions in kcal/mol (Y-axis). (a) A channel corresponding to a feature (X-axis) is filled with zero. (b) The values of a channel corresponding to a feature are randomly shuffled.

**Conclusion**
We developed a new binding affinity prediction model, AK-score, by combining a multi-branched deep-learning network architecture and an ensemble predictor approach. Our model predicts the binding affinity of a protein-ligand complex with a high accuracy, which is comparable to the best existing scoring functions in terms of scoring and ranking power. Our results suggest that an ensemble-based approach, using the average of multiple independently trained models, is a straightforward but powerful approach. A similar approach may be applicable to existing machine-learning-based models. In addition, our study gives an insight into the relative importance of atoms based on their chemical properties. The feature importance tests show that, for a ligand, the excluded volume of atoms, the spatial distribution hydrophobic and aromatic atoms are critical in determining binding affinity. For a protein, the excluded volume of atoms, and the distribution of hydrophobic atoms and hydrogen bond acceptors are identified to be important factors. We believe that our results provide useful guidelines for the development of next-generation deep-learning-based protein-ligand scoring functions.

AUTHOR INFORMATION

Corresponding Author

Junsu Ko, e-mail: junsuko@arontier.co

Woonghee Shin, e-mail: whshin@scnu.ac.kr

Juyong Lee, e-mail: juyong.lee@kangwon.ac.kr

Author Contributions
The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding Sources**

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2019M3E5D4066897, and 2019M3E5D4066898) This research was also partially supported by Arontier co.

**Notes**

The code is available upon request.

**ACKNOWLEDGMENT**

**ABBREVIATIONS**

Conv3D, 3-dimensional convolutional neural network layer; ReLU, rectified linear unit activation function layer; BN, batch normalization.

**REFERENCES**

(1) Lee, J.; Miller, B. T.; Brooks, B. R. Computational Scheme for pH-Dependent Binding Free Energy Calculation with Explicit Solvent. *Protein Science*. 2016, pp 231–243. https://doi.org/10.1002/pro.2755.

(2) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; 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.

(3) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; et al. Overview of the SAMPL6 Host-Guest Binding Affinity Prediction Challenge. *J. Comput. Aided Mol. Des.* 2018, 32 (10), 937–963.

(4) Nishikawa, N.; Han, K.; Wu, X.; Tofoleanu, F.; Brooks, B. R. Comparison of the Umbrella Sampling and the Double Decoupling Method in Binding Free Energy Predictions for SAMPL6 Octa-Acid Host-Guest Challenges. *J. Comput. Aided Mol. Des.* 2018, 32 (10),
1075–1086.

(5) Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.* **2003**, *46* (12), 2287–2303.

(6) Perryman, A. L.; Santiago, D. N.; Forli, S.; Santos-Martins, D.; Olson, A. J. Virtual Screening with AutoDock Vina and the Common Pharmacophore Engine of a Low Diversity Library of Fragments and Hits against the Three Allosteric Sites of HIV Integrase: Participation in the SAMPL4 Protein–ligand Binding Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 429–441.

(7) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature*. 2015, pp 436–444. https://doi.org/10.1038/nature14539.

(8) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710.

(9) Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *Journal of Chemical Information and Modeling*. 2018, pp 287–296. https://doi.org/10.1021/acs.jcim.7b00650.

(10) 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.

(11) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model*. **2017**, *57* (4), 942–957.

(12) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein–ligand Binding Affinity Prediction. *Bioinformatics*. 2018, pp 3666–3674. https://doi.org/10.1093/bioinformatics/bty374.

(13) Iandola, F. N.; Han, S.; Moskewicz, M. W.; Ashraf, K.; Dally, W. J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. *arXiv [cs.CV]*. 2016.

(14) Zhang, H.; Liao, L.; Saravanan, K. M.; Yin, P.; Wei, Y. DeepBindRG: A Deep Learning Based Method for Estimating Effective Protein-Ligand Affinity. *PeerJ* **2019**, *7*, e7362.

(15) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv [cs.CV]*. 2015.

(16) Sharma, A.; Muttoo, S. K. Spatial Image Steganalysis Based on ResNeXt. *2018 IEEE 18th International Conference on Communication Technology (ICCT)*. 2018. https://doi.org/10.1109/icct.2018.8600132.

(17) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48* (12), 4111–4119.

(18) 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.

(19) Martin Abadi; Ashish Agarwal; Paul Barham; Eugene Brevdo; Zhifeng Chen; Craig Citro; Greg S. Corrado; Andy Davis; Jeffrey Dean; Matthieu Devin; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015.