PEPSI: Post-docking Evaluation with Protein-Small Molecules Interaction

Andrea Karlova [*1,2] **Wim Dehaen** [*3] **Andrei Penciu** [*2] **Richard Dallaway** [4,2] **Suran Goonatilake** [1,2]

Abstract

Successful leads in drug discovery are determined by sufficient knowledge of how the drug-like molecule candidate interacts and potentially forms a complex with the protein target. Protein-ligand docking aims to predict the probable experimental orientation as well as the binding affinity needed to form a stable complex structure and proves to be highly useful in applications such as virtual high-throughput screening of candidate drugs or binding target detection. Nevertheless, the traditional docking algorithms are computationally costly. Recently, Stärk et al. (2022) explored an SE(3)-equivariant geometric deep learning model with promising speed-ups in order of magnitudes. We present a methodology for post-docking evaluation of the protein-ligand complexes based on an assessment of the intermolecular interactions between the target and the candidate molecule and use the interaction fingerprints as a tool to investigate the deep learning docking algorithms. We also demonstrate the advantages of incorporating interaction features in the prediction of bioactivity.

1. Introduction

Protein-ligand docking explores possible binding poses of the ligand to a given molecular target (Friesner et al., 2004; Wang et al., 2016). With the good knowledge of the ligand binding mode we can better predict either bioactivity or determine the optimal binding affinity. Correctly predicting the bio-activity of the molecule across a large spectrum of protein targets increases the probability of successful hits. Alternatively, reverse docking can be used to screen a drug candidate or existing drug against a large set of potential targets with applications such as side-effects prediction (Sachdev & Gupta, 2020) or drug re-purposing (Kumar & Kumar, 2019).

Classical docking algorithms either sample various conformations of multiple ligand binding poses or explores the ligand and protein surface complementarity. The quality of the binding pose is evaluated by a scoring function. Due to potentially burdensome computational cost, molecular docking does not scale well to the screening of large chemical libraries. It creates a computational bottleneck when used in machine learning optimisation, making it infeasible as a part of reward in the novel drug generation. The use of a rigid receptor, which represents no induced fit or receptor flexibility, and limitations of the scoring function and force-field (Eberhardt et al., 2021) can negatively influence the performance. An appropriate selection of the location and size of the boundary box within which the docking is performed is also important (Coleman et al., 2013). The blind docking without a boundary box is significantly harder because it necessitates binding site detection on the whole protein target.

Representations of the intermolecular interactions as depicted in Figure 1 allows for more realistic modelling of binding affinities (Da & Kireev, 2014). These fingerprints have been used to compare docked ligands to a reference structure, usually a protein-ligand complex structure determined by X-ray crystallography. In particular a protein-ligand interaction fingerprint encodes the presence or absence of specific non-covalent interactions between a ligand and the protein target at the binding pocket. Information such as atom types on the ligand and protein end, interaction type such as electrostatic interactions, Van Der Waals interactions, hydrogen bonds, interacting residue and more can be used to distinguish between interactions (Chupakhin et al., 2014). The presence and strength of interactions is extracted from the complex structure by measuring the distance and orientation of complementary chemical groups such as ring centroids and lone electron pairs which must be within a certain predefined interval.

The empirical post-docking filters make use of various types of structureactivity information e.g. quantify how the known ligands differ against the background decoys, the negative

[*]Equal contribution [1]Department of Computer Science, University College London, London, United Kingdom [2]April19 Discovery Inc., Delaware, USA [3]University of Chemistry and Technology, Prague, Czech Republic [4]School of Engineering and Informatics, University of Sussex, Brighton, United Kingdom. Correspondence to: Andrea Karlova <akarlova@cs.ucl.ac.uk>, Wim Dehaen <dehaeni@vscht.cz>.

The 2022 ICML Workshop on Computational Biology. Baltimore, Maryland, USA, 2022.

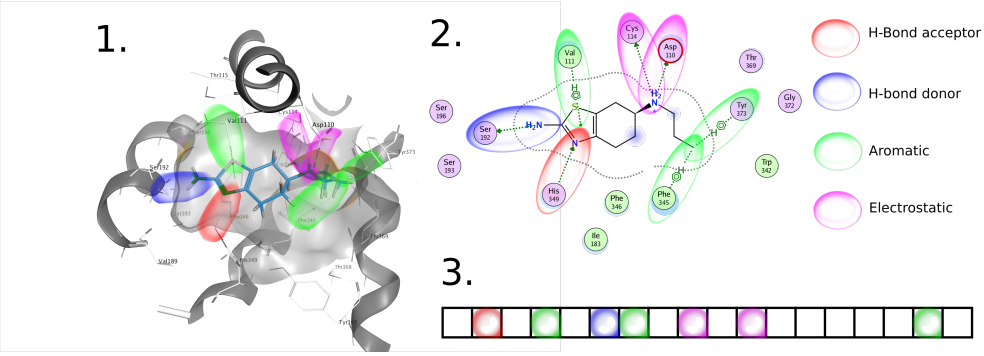

Figure 1. Interaction Fingerprints: 1. Structure of Pamipexole (Cyan) bound to the pocket of the D3 subtype Dopamine receptor. Interactions are highlighted in color in all figures (legend of interaction types on right of figure). 2. 2D representation of Pamipexole's interaction with the receptor 3. presence of interactions distributed to bins of fingerprint, which usually incorporates amino acid, interaction and ligand information.

control molecules which do not bind with the target of interest (Mysinger et al., 2012). To successfully predict the molecule binding affinity and its strength, we need to first determine whether the candidate molecule even binds to the target as for decoys the assay values such as pKi are experimentally non-measurable.

Deep learning has been extensively used in docking algorithms, either in improving the scoring function directly or learning on the generated poses and thus tackling the issues with re-ranking or ex-post assessment. Most recently, deep learning models are used to generate optimal docking poses directly. The models are usually trained in supervised learning manner on experimentally measured ligand-protein complexes and heavily rely on the co-crystalised complexes as a source of the training information. These needs to be obtained experimentally and does not provide the complementary information about the decoys. Training models to the docked poses or utilising the docking scores in the machine learning algorithm introduces a new source of errors and uncertainty.

We present a methodology for post-docking evaluation of the docked complexes based on protein-ligand atom connectivity interaction representations such as the one used in ECIF. We predict the docking poses of the complexes using Equibind and use the predicted conformations as a part of feature pre-processing for modelling bioactivity patterns. We also demonstrate the efficiency of our approach in prediction of the bioactivity.

2. Methods

To investigate the quality of the docked poses we define a metric based on the "Extend Circular Interaction Features" (ECIFs) (Sánchez-Cruz et al., 2020). These are constructed such that each atom in the molecule we encode according the following criteria: atom symbol; explicit valence; number of attached heavy atoms; number of attached hydrogens; aromaticity; and ring membership. The list of 22 protein encoded atom types is combined with 70 ligand types based on PDBBindv2016 and forms a bag of the most relevant 1540 atom type interactions. For a given radius, the atom pair of the atom from ligand and the atom from protein are considered as interacting whenever the L2 norm of their 3D coordinates lies within the cutoff radius. For keeping fixed length fingerprint we count all interactions from the bag of 1540 and drop those outside the bag. In our investigation we take ECIFs for 10 different radii 1.5A(Amstrong), 2A, 2.5A, 3A, 3.5A, 4A, 4.5A, 5A, 5.5A and 6A. The value 1.5A corresponds to nearly atomic bond while value 6A contains various distant interactions. The interaction fingerprints are stacked into a matrix with rows corresponding to radius cutoffs. This allows to captures the changes in the interactions as the atom neighbourhood expands. We refer to these as 2D ECIFs. Figure 2 demonstrates that the specific amino-acid sequence of the target characterises the representation of the interactions.

ECIFs were design for a post-docking evaluation and require a conformer of the optimal structure. We explore the docking approach introduced by (Stärk et al., 2022; Ganea et al., 2021) based on the approximation of the binding pockets and predicting the docking poses while using keypoint matching and aligning through optimal transport and a differentiable Kabsch algorithm. We use Equibind to predict 3D conformations of complexes from DUD-E (Mysinger et al., 2012). DUD-E dataset contains mostly drug-like molecules with molecular weight less then 500 and less

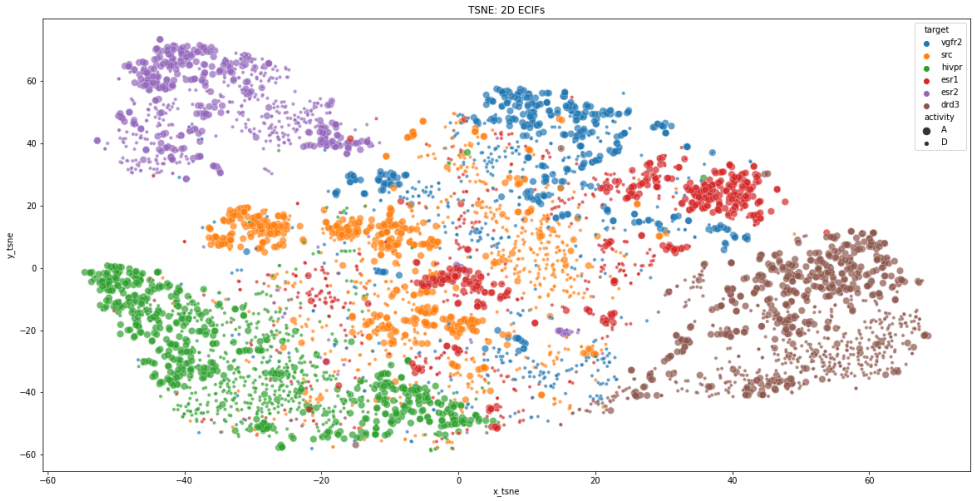

Figure 2. 2D ECIF: for each complex we compute interaction fingerprint with the radius cutoff taking values in 1.5A, 2A, 2.5A, 3A, 3.5A, 4A, 4.5A, 5A, 5.5A and 6A. The value 1.5A corresponds to nearly atomic bond while value 6A contains various distant interactions. The interaction fingerprints are stacked into a matrix with rows corresponding to radius cutoffs. We visualise 2D ECIFs using t-SNE for 6 protein targets: dopamine receptor DRD3, estrogen receptors ESR1 and ESR2, HIV type 1 protase receptor HIVPR, tyrosine-protein kinase SRC and vascular endothelial growth factor receptor VGFR2 together with corresponding ligands and decoys. The visible clustering indicated that the specific amino-acid sequence of the target characterises the representation of the interactions.

then 20 rotatable bonds. We construct a metrics based on 2D ECIFs and investigate the relation between the interaction metrics and root mean square errror between the ligand conformation predicted by EquiBind and experimentally measured one condition by the molecular weight, number of rotatable bonds, drug-likeness and NHOH counts. This motivates our post-docking evaluation with protein-small molecules interactions (PEPSI): we take the Frobenius norm of the difference of 2D ECIFs between two complexes and regress them on the L2 norm of the 3D conformations of the complexes. That way we capture the spatial aspect of the error and inter-atomic interaction. We then order the values according the desired chemical property, such as number of rotatable bonds, logP, molecular weight or any relevant chemical property of interest.

We further investigate whether the ECIFs are useful for detection of bioactive molecules and decoys. Multi-task deep learning is good fit for drug discovery problems as a natural progression of quantitative structure-activity relationship (QSAR) models (Geppert et al., 2010). The multi-task architecture tends to better leverage the shared input representations while it is co-jointly trained to multiple tasks and when trained on sufficiently large dataset it overperforms the single-task methods(Ramsundar et al., 2015; 2017).

3. Experiments

The PDBbind database (Liu et al., 2017) is a benchmark dataset of systematically annotated protein–ligand complexes in the Protein Data Bank (PDB) with experimental binding data. It has been updated annually since its first public release in 2004. The latest release provides binding data for 19k biomolecular complexes in PDB.

DUD-E (Mysinger et al., 2012), Database of Useful Decoys - Enhanced, is a docking benchmark dataset. It consists of a curated subset of ChEMBL(Gaulton et al., 2016), 102 protein targets with associated crystal structure and a total of 20.000 ligands with known, measured activity at one of these targets. Each of these ligands have 50 associated decoys - molecules with matched physiochemical properties such as molecular weight and lipophilicity that are topologically distinct and expected to be inactive. These compounds are selected from the ZINC database. The number of ligands varies per different target. For our experiments we sampled roughly the same number of decoys per each target to construct a balanced dataset.

To assess the actives-decoys recognition task, we constructed the dataset of approximately 50 actives from CHEMBL and we sampled a set of approximately 50 decoys from DUD-E per each target.

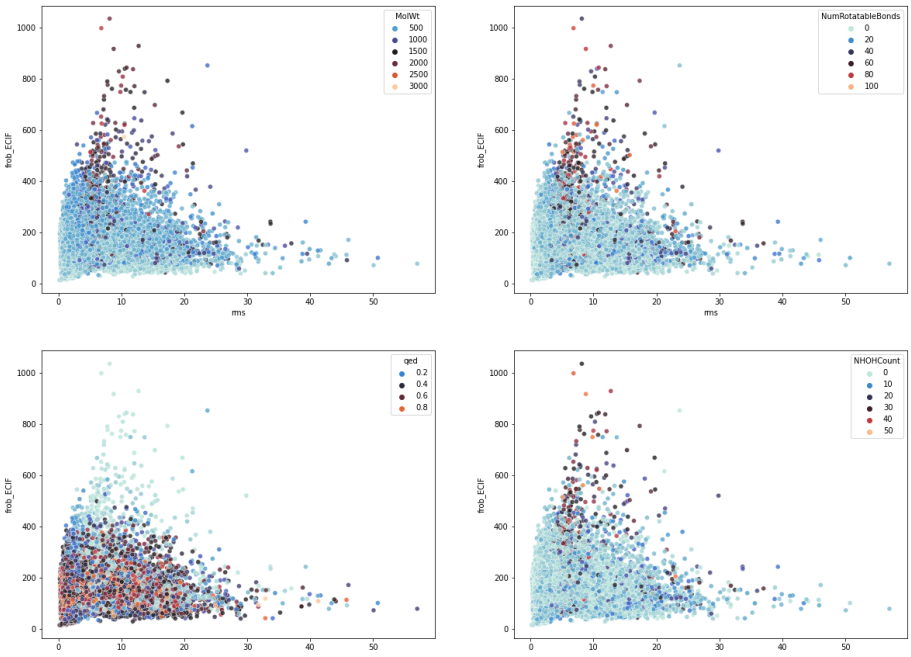

Figure 3. PEPSI: RMSE (x-axis) plotted against Frobenius norm of difference between 2D ECIFs measured on complexes predicted by Equibind and complexes measured experimentally (PDBBindv2020). Large values of Frobenius norm of ECIFs together with low RMSE indicates that EquiBind fails for largely flexible molecules with low QED(drug-likeness) score. Sharp increase of Frobenius norm of 2D ECIFs for small molecules and close to 0 RMSE indicates that Equibind introduces error into the prediction of the inter-molecular activity of the complex.

To obtain the ECIFs we first use Equibind to obtained the optimal docked poses for the complexes. Each ligand and protein is processed with OpenBabel and all missing hydrogen atoms are added to the ligand. Additionally, each protein goes through a hydrogen correction procedure using REDUCE[1]. From the protein, we keep only the connected components which within a 10A radius of any ligand atom. We compute ECIFs for radius: 1.5A, 2A, 2.5A, 3A, 3.5A, 4A, 4.5A, 5A, 5.5A and 6A and fit 102 task multitask neural network classifier for each single radius with the MLP architecture: [2048, 512, 256, 128, 32, 8], ReLu activation, and dropout 0.2 on all layers. As a benchmark model for a comparison we use the Random Forest Classifier (RF). The results are reported in Table 1. Despite the fact the RF lightly outperforms the multi-task NN on the unseen CHEMBL data, we assume that this is due to the size and drug-like character of the DUD-E dataset.

4. Conclusions and Discussion

We investigated the performance and the usability of the Equibind model. The efficiency of the use outperforms traditional docking tools. As Equibind aims to only alter torsion angles of rotatable bonds while keeping bond angles and lengths fixed, the rotatable bonds constrain together with the alignment of the structure creates issues for larger and flexible molecules. Most of these molecules may most likely not be the potential drug-like hits. As measured by PEPSI, the Equibind also tends to fit the structures closer to the protein with significantly larger number of interactions. This introduces uncertainty which needs to handled when used in the model pipeline. However there is degree of uncertainty in the experimental data for particular atomic distances either.

[1]https://github.com/rlabduke/reduce

		Metrics			
Cutoff	Model	Accuracy	Precision	Recall	AUC
1.5A	MNN	0.623	0.586	0.651	0.625
	RF	0.632	0.595	0.657	0.687
2.5A	MNN	0.728	0.719	0.684	0.726
	RF	0.740	0.735	0.692	0.860
3.5A	MNN	0.806	0.826	0.740	0.802
	RF	0.813	0.840	0.742	0.893
4.5A	MNN	0.842	0.871	0.776	0.838
	RF	0.847	0.890	0.766	0.925
5A	MNN	0.854	0.883	0.793	0.857
	RF	0.856	0.903	0.775	0.934
6A	MNN	0.864	0.9	0.796	0.860
	RF	0.868	0.92	0.784	0.94

Table 1. Actives-Decoys Recognition Task on CHEMBL Testset Performance: we provide comparison of the multi-task NN classifier with 102 task and Random Forest Classifier used as a benchmark. The features for the model are 1540x1 ECIFs counting the interaction between the target and the ligand within the cutoff radius reported in the first column.

References

Chupakhin, V., Marcou, G., Gaspar, H., and Varnek, A. Simple ligand–receptor interaction descriptor (silirid) for alignment-free binding site comparison. *Computational and Structural Biotechnology Journal*, 10(16):33–37, 2014. ISSN 2001-0370. doi: https://doi.org/10.1016/j.csbj.2014.05.004. URL https://www.sciencedirect.com/science/article/pii/S2001037014000051.

Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J., and Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS ONE*, 8(10):e75992, oct 2013. doi: 10.1371/journal.pone.0075992.

Da, C. and Kireev, D. Structural protein–ligand interaction fingerprints (splif) for structure-based virtual screening: Method and benchmark study. *Journal of Chemical Information and Modeling*, 54(9):2555–2561, 2014. doi: 10.1021/ci500319f.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. AutoDock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, jul 2021. doi: 10.1021/acs.jcim.1c00203.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47 (7):1739–1749, feb 2004. doi: 10.1021/jm0306430.

Ganea, O.-E., Pattanaik, L., Coley, C. W., Barzilay, R., Jensen, K. F., Green, W. H., and Jaakkola, T. S. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. arXiv, 2021. doi: https://doi.org/10.48550/arXiv.2106.07802.

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M. P., Overington, J. P., Papadatos, G., Smit, I., and Leach, A. R. The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1):D945–D954, 11 2016. doi: 10.1093/nar/gkw1074.

Geppert, H., Vogt, M., and Bajorath, J. Current trends in ligand-based virtual screening: Molecular representations, data mining methods, new application areas, and performance evaluation. *Journal of Chemical Information and Modeling*, 50(2):205–216, jan 2010. doi: 10.1021/ci900419k.

Kumar, S. and Kumar, S. Chapter 6 - molecular docking: A structure-based approach for drug repurposing. In Roy, K. (ed.), *In Silico Drug Design*, pp. 161–189. Academic Press, 2019. ISBN 978-0-12-816125-8. doi: https://doi.org/10.1016/B978-0-12-816125-8.00006-7. URL https://www.sciencedirect.com/science/article/pii/B9780128161258000067.

Liu, Z., Su, M., Han, L., Liu, J., Yang, Q., Li, Y., and Wang, R. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of Chemical Research*, 50(2):302–309, 2017. doi: 10.1021/acs.accounts.6b00491.

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. Directory of useful decoys, enhanced (dud-e): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14):6582–6594, 2012. doi: 10.1021/jm300687e. URL https://doi.org/10.1021/jm300687e.

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., and Pande, V. Massively multitask networks for drug discovery, 2015. URL https://arxiv.org/abs/1502.02072.

Ramsundar, B., Liu, B., Wu, Z., Verras, A., Tudor, M., Sheridan, R. P., and Pande, V. Is multitask deep learning practical for pharma? *Journal of Chemical Information and Modeling*, 57(8):2068–2076, 2017. doi: 10.1021/acs.jcim.7b00146.

Sachdev, K. and Gupta, M. K. A comprehensive review of computational techniques for the prediction of drug side effects. *Drug Development Research*, 81(6):650–670, 2020. doi: https://doi.org/10.1002/ddr.21669.

Stärk, H., Ganea, O.-E., Pattanaik, L., Barzilay, R., and Jaakkola, T. Equibind: Geometric deep learning for drug binding structure prediction. 2022. doi: 10.48550/ARXIV.2202.05146. URL https://arxiv.org/abs/2202.05146.

Sánchez-Cruz, N., Medina-Franco, J. L., Mestres, J., and Barril, X. Extended connectivity interaction features: improving binding affinity prediction through chemical description. *Bioinformatics*, 37(10):1376–1382, 12 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa982. URL https://doi.org/10.1093/bioinformatics/btaa982.

Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., and Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics*, 18(18): 12964–12975, 2016. doi: 10.1039/c6cp01555g.