**In silico drug discovery: identifying potential lead structure for PTP1B inhibition**

**Authors: Thomas M., Mania, H, Timo D.**

**Introduction**

Diabetes is a disorder to regulate homeostatic glucose levels. A total of 56.5 million European adults were estimated to have diabetes, both diagnosed and undiagnosed in 2019. Left untreated diabetes has some serious implications, including blindness, kidney failure and limb amputations.1 Diabetes can be grouped into two types, diabetes type 1 and diabetes type 2. Type 1 is characterised by the immune system attacking insulin-producing cells while type 2 is the disability of the body to produce insulin or the disability of cells to react to insulin. Of these two types, diabetes type 2 is more common.2 Risk factors include of genetic background, age and obesity.3 Under physiological conditions, insulin binds to the insulin receptor kinase (IRK) which autophosphorylates and activates downstream signalling.4 The IRK signalling is antagonized by protein tyrosine phosphatases (PTP), which includes PTP1B. PTP1B dephospharylates the IRK and thereby deactivates insulin signalling.5 In order to prolong the effects of IRK signalling the antagonistic effects of PTP1B should be inhibited. A limited amount of inhibitors has entered clinical trials, some compounds include Ertiprotafib, Trodusquemine and JTT-551.6 The limited availability of inhibiting compound emphasizes the need for discovery of potential small molecule inhibitors. The discovery of potential drugs usually takes a few years. However, the use of computational approaches significantly speeds up this process. Nowadays, a considerable amount of data is being generated. The ever increasing pile of data is often too hard too analyse, this is where computational approaches come to play.7 This work aims to discover potential new antagonistic ligands for PTP1b by using computational approaches.

**Methods**

Data gathering

The ChEMBL database was used to query for bioactivity data of the PTP1B protein and filtered to only include the mean of the same compound-entries. The data was further filtered to only include bioactivity data measured in nM. Next, the compound structures were retrieved in canonical SMILES format and lastly were filtered to comply with the Lipinski rule of 5 which resulted in a final dataset of 1992 compounds.

The IC50 values were converted to their pIC50 counterparts by 9*-log10(nM)* and compounds displaying a pIC50 higher than 7 were labelled active while pIC50 below 7 were labelled as inactive. Furthermore, the SMILES were converted to MACCS keys.

Activity prediction model

The labelled dataset was split into a training set 70% and a validation set 30% and was trained on a neural network consisting of two dense hidden layers with 32 nodes each and a output layers of 1 dense node with a sigmoidal activation to reflect the probability of a compound having a pIC50 higher than 7.

De Novo Drug generation (DrugEx)

To generate new drugs which fall within the chemical space of the original bio-activity tested compounds of the PTP1B protein, the pipeline of Martin Sicho was used termed DrugEx.8

De Novo dataset filter

The generated dataset was filtered on Lipinski rule of 5 and compounds having unwanted substructures9 or PAINS were removed from the dataset. 391 of the total 1000 compounds were left. These 391 compounds were fed into the activity prediction model and predictions which resulted in a confidence higher than 50% of higher of compounds being active were selected.

Docking

For the docking of the de novo ligands the pipeline of Willem Jespers was used.10 First PDB files of the protein and co-crystalized ligand were retrieved after which Vina was used to dock the ligands.

**Results**

Activity prediction model

The model was trained on the Lipinski dataset for 100 epochs and resulted in a train accuracy of 99% and a validation accuracy of 97%. Evaluating the model on the entire dataset resulted in 59.2% correct predictions for active compounds and 99.7% correct predictions for inactive compounds. (Table 1).

|  |  |  |
| --- | --- | --- |
| **Table 1** | Active | Inactive |
| Predicted Active | 42 | 5 |
| Predicted Inactive | 29 | 1916 |

De novo potential ligands

The generated 1000 compounds of the DrugEx pipeline were filtered on Lipinski rules and unwanted substructures, resulting in 391 compounds. These 391 models were evaluated by the activity prediction model and compound displaying a confidence higher than 0.5 were selected to be active. It resulted in 4 ligands displaying a confidence higher than 0.5 (Figure 1).

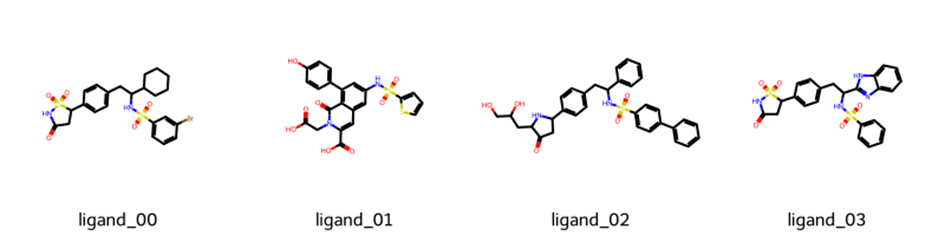


Figure 1. The filtered ligands which are probably active. Ligand 00 displayed the lowest confidence (60%) while the other ligands were more confident (80%).

Docking

So far we have only looked into the 2D structure of the ligand and haven’t considered the position of the ligands within the binding site. Therefore, we’ll use Vina docking to assess the interactions of the ligands on the binding site. The predicted pIC50 values were also given by Vina which were all lower than 7 (Table 2).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2** | Kcal/mol | Model Predicted Activity | pIC50 |
| Ligand\_00 | -7.799 | 60.1% | 5.72 |
| Ligand\_01 | -6.676 | 81.9% | 4.9 |
| Ligand\_02 | -7.865 | 90.6% | 5.77 |
| Ligand\_03 | -7.466 | 85.8% | 5.46 |

The positions of the ligand within the binding site were also evaluated (Figure 2). All ligands fit quite well in the binding pocket. Interestingly, the 3 ligands displaying the lowest energy binding poses do share a similar common structures (Figure 3) and all the ligands share a sulfur dioxide bond which is attached to a nitrogen. Lastly, looking at ligand 02, which is the only ligand which seems to interact with a residue outside the binding pocket, in an encapsulating manner. This pose might be beneficial for staying inside the binding pocket. The structure of ligand 02 resembles a scissor which enables the ligand to only go inside the binding pocket and also just a few ways to go out.

Afbeelding met verschillend, groente

Automatisch gegenereerde beschrijving

Figure 2. Docking positions of the ligands. Illustrations gathered from RBS protein bank <https://www.rcsb.org/3d-view>.

Afbeelding met tekst, horloge

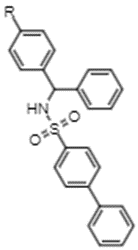
Automatisch gegenereerde beschrijving

Figure 3. Common structure of the lowest energy binding poses.

**Discussion**

As we have seen in the docking poses of the ligands, ligand 02 seems to be interesting in occupying the binding pocket for a longer period of time. Therefore, the structure of ligand 02 will be incorporated into the lead compound. To ensure better IC50 values the part which binds within the pocket could be altered to maximize interactions. Which finally results in the final lead compound (Figure 4). The single benzene ring which is part of the scissor structure could also be replaced by another bulky group since this part interacts with the elongated part of the binding pocket. The other ligands do occupy this elongated part of the binding pocket. Especially ligan 03 seems to maximise the space within this elongated part. So this single benzene group could be replaced by a purine. This work may have not find all potential ligand within the de novo dataset due to the model’s inability to predict all active compounds (59.2%). However, due to the similar bond types within the ligands such as the sulfur dioxide-nitrogen bond. A study which discovered several ligands identified 3 potential ligands which also include this sulfur dioxide-nitrogen bond, implying the importance of such bond in inhibiting PTP1B.6 Our discovered ligand 02 has not yet been tested in vitro and may show very different results compared to it’s in silico predictions. Therefore we suggest to test this structure in vitro and to keep on improving our model to distinguish the existing bias to correctly predicting inactive compounds but not predicting the active compounds accurately. Lastly, with this work we haven’t included protein-ligand interactions so it would be more beneficial to look into these interaction to gain more insight into beneficial substructures for maximizing interactions.

Figure 4. Suggested lead structure for inhibiting PTP1B.



Scissor structure

Binding Pocket

**References**

1. OECD/European Union (2020), Health at a Glance: Europe 2020: State of Health in the EU Cycle, OECD Publishing, Paris, <https://doi.org/10.1787/82129230-en>.
2. NHS website. (2022, June 28). Diabetes. nhs.uk. <https://www.nhs.uk/conditions/diabetes/>
3. Fletcher, B., Gulanick, M., & Lamendola, C. (2002). Risk Factors for Type 2 Diabetes Mellitus. The Journal of Cardiovascular Nursing, 16(2), 17–23. <https://doi.org/10.1097/00005082-200201000-00003>
4. Posner, B. I. (2017). Insulin Signalling: The Inside Story. Canadian Journal of Diabetes, 41(1), 108–113. <https://doi.org/10.1016/j.jcjd.2016.07.002>
5. Tiganis, T. (2012). PTP1B and TCPTP - nonredundant phosphatases in insulin signaling and glucose homeostasis. FEBS Journal, 280(2), 445–458. <https://doi.org/10.1111/j.1742-4658.2012.08563.x>
6. Sharma, B., Xie, L., Yang, F., Wang, W., Zhou, Q., Xiang, M., Zhou, S., Lv, W., Jia, Y., Pokhrel, L., Shen, J., Xiao, Q., Gao, L., & Deng, W. (2020). Recent advance on PTP1B inhibitors and their biomedical applications. European Journal of Medicinal Chemistry, 199, 112376. <https://doi.org/10.1016/j.ejmech.2020.112376>
7. Wooller, S., Benstead-Hume, G., Chen, X., Ali, Y., & Pearl, F. (2017). Bioinformatics in translational drug discovery. Bioscience Reports, 37(4). <https://doi.org/10.1042/bsr20160180>
8. GitHub - martin-sicho/drugex-demo. (n.d.). GitHub. Retrieved November 6, 2022, from <https://github.com/martin-sicho/drugex-demo>
9. Brenk, R., Schipani, A., James, D., Krasowski, A., Gilbert, I., Frearson, J., & Wyatt, P. (2008). Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. ChemMedChem, 3(3), 435–444. <https://doi.org/10.1002/cmdc.200700139>
10. GitHub - jesperswillem/CBR\_teaching. (n.d.). GitHub. Retrieved November 6, 2022, from <https://github.com/jesperswillem/CBR_teaching>