Universiteit
Leiden
The Netherlands

Proteins in harmony: Tuning selectivity in early drug discovery

Burggraaff, L.

Citation

Burggraaff, L. (2020, October 21). *Proteins in harmony: Tuning selectivity in early drug discovery*. Retrieved from https://hdl.handle.net/1887/136965

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/136965

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/136965 holds various files of this Leiden University dissertation.

Author: Burggraaff, L.
Title: Proteins in harmony: Tuning selectivity in early drug discovery
Issue date: 2020-10-21

Summary

Summary

Computational modeling plays an important role in early drug discovery. Using computers, time spend on medicine development can be decreased significantly. Moreover, since fewer resources are needed to detect potential new medicines, the drug discovery process becomes more cost-effective. This thesis describes how computational models can be applied to enhance the drug discovery process by the prediction of selectivity of compounds, and therewith implicitly toxicity.

The subject of computational medicine development and selectivity is touched upon in **Chapter 1**. Here, general drug discovery concepts are outlined and the role of computers in medicine development is highlighted. Moreover, it is made clear why one would want to obtain selectivity.

Chapter 2 goes into depth on computational methods that can be used in drug discovery. The roles of both ligand- and protein-based methods are discussed in the light of their applicability in target prediction studies. Furthermore, the methods' requirements and restrictions are discussed, which gives the reader a good understanding of the applicability of these tools in drug discovery in general.

A selection of the computational tools that are described in **Chapter 2** are applied in case studies in the subsequent **Chapters 3**, **4**, and **5**. **Chapter 3** is limited to the use of ligand-based methods due to the lack of structure-based protein information. In contrast, **Chapter 4** and **5** describe the use of both ligand- and structure-based methods. With the machine learning method applied in **Chapter 3**, chemically novel SGLT1 inhibitors could be detected. The key to success in this chapter was to create a chemically diverse training set in order to expand the applicability domain.

Ligand-based machine learning was also put to the test in **Chapter 4**, where it was used to directly model the difference of selectivity between two proteins – the A_1 and A_{2A} adenosine receptors. Where traditionally machine learning in drug discovery is used to model bioactivity (as in **Chapter 3**), it was now applied to predict the bioactivity difference. Since this method only scales new compounds based on selectivity, inactive compounds cannot be detected with this model. Therefore, it was combined with structure-based docking models to filter for inactive compounds.

In **Chapter 5**, a different combination of ligand-based and structure-based techniques was applied. A subsequent workflow was developed that utilized machine learning, structure-based docking, and binding pose metadynamics, to detect kinase inhibitors. Here, multiple machine learning models were designed to filter and score compounds. In contrast to **Chapter 4** where docking was used to exclude inactive compounds, in this workflow docking was used to rank compounds. Together with pose metadynamics, a more accurate binding affinity score could be obtained than with docking by itself. The final combined ligand- and structure-based score was used to select novel RET kinase inhibitors.

Chapter 6 addresses a subject that is worth considering when modeling selectivity: the binding location of the ligand on the protein. Some proteins contain multiple binding sites, which can bind chemically different compounds. This can influence the predictive capability of models when not considered. **Chapter 6** displays the effect on model predictive performance when

binding locations on proteins are included in the model. Moreover, it shows that the binding location can be predicted when training data is sufficient.

The final chapter, **Chapter 7**, lists the most important conclusions that can be drawn from this thesis. Attention is given to the applicability of selectivity modeling in different situations. Moreover, both the limitations and opportunities of computational models in a broader perspective are described with an emphasis on available data in terms of quantity and quality (diversity).