Abstract:

Machine learning to predict Michaelis constants for aldehyde oxidase enzymes

Co-supervisor 1: Dr. Ariane Nunes Alves, TU Berlin (UniSysCat, BIMoS)

Co-supervisor 2: Prof. Dr. Silke Leimkühler, U Potsdam (UniSysCat)

Mammalian aldehyde oxidases (AOXs) are important drug and xenobiotic metabolizing enzymes which are present in many tissues in various mammalian species, including humans and rodents. Different species contain a different number of AOX isoforms. The reasons why mammals other than humans express a multiplicity of tissue-specific AOX enzymes is unknown. The physiological function of AOX for mammals remains enigmatic to date. To get more insights into the physiological role of AOX and the isoforms, the problem can be approached by defining their substrate specificities. In a previous study the four mouse AOX isoforms were shown to have different catalytic rates and Michaelis constants for the same substrates. The aims of this project are to build regression models to predict Michaelis constants for the binding of diverse substrates to mouse and human AOX enzymes and later test these models prospectively, characterizing Michaelis constants for other candidate substrates experimentally. Regression models will be done in the group of Dr. Ariane Nunes Alves, and experiments will be performed in the group of Prof. Dr. Silke Leimkühler. The results will enhance our understanding of factors regulating substrate specificity in enzymes.

Extended description version of the project:

Machine learning to predict Michaelis constants for aldehyde oxidase enzymes

Co-supervisor 1: Dr. Ariane Nunes Alves, TU Berlin (UniSysCat, BIMoS)

Co-supervisor 2: Prof. Dr. Silke Leimkühler, U Potsdam (UniSysCat)

1. Overall goal of the project

The aim of this project is to understand the factors modulating substrate selectivity in mouse and human aldehyde oxidase (AOX) isoenzymes. In order to achieve this goal, regression models will be built to predict Michaelis constants (K_M) for AOX enzymes. Different regression methods (partial least squares regression, random forest) and features (ligand chemical descriptors and protein-ligand interactions) will be tested, and the models with highest predictive power will be selected for further analysis. Later, these models will be tested prospectively, by experimental characterization of K_M values for other candidate substrates.

2. State of the art

Mammalian aldehyde oxidases (AOXs) are multifunctional molybdo-flavoenzymes. The importance of human AOX is related to its prominent role as a drug and xenobiotic metabolizing enzyme, being responsible for unexpected drug metabolism, for promoting drug interactions, for affecting drug's efficacy and ultimately, for leading to failures in clinical trials. AOXs in general accept a wide range of unrelated substrates, and are able to perform surprisingly diverse reactions such as oxidations (aldehydes and N-heterocycles), amide hydrolysis, and reductions (e.g. N-oxides, sulfoxides, organic nitrites). The challenges related to the role of AOX enzymes in drug metabolism also reside on the fact that AOXs are species-specific and, while humans have only one active gene, preclinical animal models like mice have four genes encoding four different isoenzymes with different substrate specificities. The other enigma on AOXs relates to its true physiological function that is still unknown. It is highly unlikely that its primary function is xenobiotic metabolism. There are some hypotheses based on known endogenous substrates and the fact that human AOX1 is an important source of reactive oxygen/nitrogen species (ROS/RNS) suggests a role in signaling and cellular redox stress. Additionally, there are in vitro data supporting the role of human AOX1 in cancer that may also be related to the generation of ROS upon substrate oxidation (1). To get more insights into the enigma why mammals other than humans express a multiplicity of tissue-specific AOX enzymes, we want to characterize the substrate specificities of mouse and human isoenzymes in more detail. Despite the amino acid similarity of 60% among the four mouse AOXs, they can display K_M values that differ by orders of magnitude for the same substrate (2). Moreover, ref. 2 also showed that AOX4 displays different substrate specificity as compared to the other three AOX enzymes. By exchanging amino acids in the active site and substrate-binding funnel using site-directed mutagenesis, the authors concluded that one of the reasons for this different specificity are the amino acids in the substrate-binding funnel of AOX4. Computational and data-driven methods can also be used to further understand substrate specificity in AOX enzymes.

Data-driven methods, especially regression, can be used not only to predict K_M values and other constants, but also to reveal mechanistic insights of substrate binding. A recent study has built models using partial least squares regression and ligand-protein interaction energies to predict dissociation rate constants (k_{off}) for ligands bound to a kinase (3). The models from this study revealed protein-ligand interactions which modulate k_{off} values for ligand-kinase complexes. Recent studies employed neural networks to predict K_M values for β-glucosidases (4) and for different enzymes (5), achieving a coefficient of correlation (R) of 0.6 and a coefficient of determination (R^2) of 0.42 for the test sets, respectively. Considering that the maximum value that R and R^2 can achieve is 1, such results show that there is room for improving regression models to predict K_M values.

Our studies will be particularly important to define the substrate specificity of AOX isoforms in mice and will help to understand the differences in the substrate selectivity of the human enzyme and why there is only one isoform present.

3. Specific aims and how they may be reached

Work package 1 (WP1): building models to predict K_M values for mouse AOX3

The aim of this WP is to build models to predict K_M values for different substrates binding to mouse AOX3. AOX3 will be tested first, because this is the only mouse AOX isoenzyme with a crystal structure available. K_M values available in the literature for AOX3 will be used to train and test the models. Different regression methods, such as partial least squares regression and random forest, will be tested. Different features will be tested to build the models, such as substrate-enzyme interaction energies, and chemical descriptors of the substrates and of the enzyme catalytic site. Docking will be used to build substrate-enzyme complexes. Molecular mechanics will be used to obtain substrate-enzyme interaction energies. The most predictive models will be used in WP2.

WP2: building models to predict K_M values for mouse AOX1-4 and human AOX1

The aim of this WP is to expand the models obtained in WP1 to mouse AOX1-4 and human AOX1. K_M values available in the literature for mouse AOX1-4 and human AOX1 will be used to train and test the models. Crystal structures for human AOX1 are available, and homology modeling will be used to build the structures of mouse AOX1, AOX2 and AOX4. Docking will be used to build substrate-enzyme complexes. Molecular mechanics will be used to obtain substrate-enzyme interaction energies. The most predictive models will be used in WP3.

WP3: prospective tests of the models

In WP1 and WP2, the models will be tested retrospectively (the experimental results are available before the predictions). In this WP, the models will be tested prospectively (the experimental results will be known after the predictions). The best models from WP2 will be used to predict K_M values for a library made of molecules with high similarity to one or more of the substrates used in WP1 and WP2. A subset of these molecules will be selected and have their K_M values determined experimentally, to allow comparison with the machine learning predictions. Further, we will define amino acids that are important for the substrate selectivity. These amino acids will be exchanged to other amino acids by site-directed mutagenesis to validate the predicted interactions from the models. We established codon-optimized heterologous expression systems for the AOX isoforms 1-4 from mouse in *Escherichia coli*. These expression systems will enable us to obtain sufficient amounts of active protein for kinetic characterizations and set the basis for site-directed mutagenesis and kinetic studies of the variant enzymes.

This project will provide models to predict K_M values for human and mouse AOXs and mechanistic insights about the enzyme-substrate interactions and molecular features modulating substrate specificity.

References:

1. M. Terao *et al.* Evolution, expression, and substrate specificities of aldehyde oxidase enzymes in eukaryotes. J. Biol. Chem., 295: 5377-5389, 2020.
2. G. Kücükgöze and S. Leimkühler. Direct comparison of the four aldehyde oxidase enzymes present in mouse gives insight into their substrate specificities. PLoS One, 13: e0191819, 2018.
3. A. Nunes-Alves *et al.* Prediction of the drug-target binding kinetics for flexible proteins by Comparative Binding Energy analysis. J. Chem. Inf. Model., 61: 3708-3721, 2021.
4. S.-M. Yan *et al.* Predicting Km values of beta-glucosidases using cellobiose as substrate. Interdiscip. Sci. Comput. Life Sci. 4: 46-53, 2012.
5. A Kroll *et al.* Prediction of Michaelis constants from structural features using deep learning. BioRxiv 2020.12.01.405928.