Machine learning methods in catalyst selection and reactivity and enantioselectivity prediction in asymmetric catalysis with magnesium complexes

mgr Paulina Baczewska

promotor: prof. dr hab. Jacek Mlynarski

The pace at which chemical data are being generated has long exceeded the limits of manual analysis. The knowledge contained in published reactions is scattered across hundreds of articles and supporting information files, and relationships that become visible only at the level of a larger dataset can easily remain unnoticed.

This is where I see the role of machine learning. Not as a value in itself, but as a tool that makes it possible to extract from existing results relationships that are useful for asymmetric catalysis, which remains a field full of challenges.

This dissertation is based on the assumption that the value of a model depends on whether the question asked of it matches the information actually contained in the data and in their representation. For example, one cannot ask about the effect of solvent on the reaction outcome if the dataset does not contain examples of the same reaction performed in different solvents. I formulated these research questions in the area of asymmetric catalysis with magnesium complexes.

The first project shows how one can work with a large and heterogeneous dataset, for which the process of standardization itself was a major challenge. I transformed the scattered literature on magnesium catalysis into a consistent dataset that also included low-enantiomeric-excess results and unfavorable examples manually recovered from supporting information. The model built on this dataset indicated ligands that, in combination with an appropriate magnesium salt, had the highest chance of inducing enantioselectivity in a given reaction. I confirmed its value experimentally. The recommended catalytic systems improved the outcomes of known α-ketoester reductions, enabled access to products for which no magnesium-complex-catalyzed reduction had previously been reported, and in some reactions made it possible to replace a labile organomagnesium reagent with a more convenient and less expensive magnesium salt. The results were published in *Angewandte Chemie*.[1]

The second project concerned the opposite situation. The starting point was a set of small but internally consistent datasets for two reactions previously developed in our laboratory, namely chalcone epoxidation and thia-Michael addition. These datasets were supplemented with rarely published negative results. In this setting, we studied when a quantitative question can be asked. In chalcone epoxidation, the enantioselectivity values were close to one another, and the absence of product could not be treated as 0% ee, so the most meaningful question concerned reactivity. In the thia-Michael addition, which covered a broader substrate space and a wider range of outcomes, it was possible to predict enantiomeric excess and capture non-obvious trends. I tested both types of predictions prospectively on substrates that had not been used to build the models. The results were published in *ACS Catalysis*.[2]

Both studies lead to the same conclusion. The usefulness of a model is determined not by the number of data points alone, nor by the complexity of the algorithm, but by whether the question being asked falls within the limits of the information contained in the dataset. Understood in this way, the model does not replace experiment. It helps indicate what should be tested first and where predictions

---

[1] Baczewska, P.; Kulczykowski, M.; Zambroń, B.; Jaszczewska-Adamczak, J.; Pakulski, Z.; Roszak, R.; Grzybowski, B. A.; Mlynarski, J. Machine Learning Algorithm Guides Catalyst Choices for Magnesium-Catalyzed Asymmetric Reactions. *Angew. Chemie Int. Ed.* **2024**, *63* (37), e202318487. https://doi.org/https://doi.org/10.1002/anie.202318487

[2] Baczewska, P.; Nowak, D.; Jaszczewska-Adamczak, J.; Bachorz, R. A.; Hoffmann, M.; Mlynarski, J. Predicting Reactivity and Enantioselectivity from Small Experimental Datasets: A Case Study in Asymmetric Magnesium Catalysis. *ACS Catal.* **2026**. https://doi.org/10.1021/acscatal.6c05090.

should be treated with caution. For this reason, I treated every recommendation as a hypothesis requiring experimental verification. In this way, machine learning can turn accumulated literature knowledge into informed starting points for catalyst discovery.