Deep Learning of Reaction Properties: Role of Reaction Representation and Model Architecture

Esther Heid[1]

[1]*Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, A-1060 Vienna, Austria*

Machine learning for predicting and designing molecular and reaction properties nowadays enable the discovery of complex structure-activity relationships directly from data. In contrast to molecules, however, chemical reactions are inherently more complex, involving changes from one or multiple molecules and spanning different levels of structural and mechanistic detail. This makes the choice of reaction representation and model architecture a central challenge for accurate and data-efficient learning.

In this talk, I will discuss how different reaction representations and architectures affect the performance and generalization of predictive and generative deep learning models for reaction property prediction. I will contrast graph-based approaches like encoding reactant and product graphs separately or as a condensed graph of reaction (CGR) [2, 3], to sequence-based representations like reaction SMILES and different CGR SMILES variants [4, 5], and encodings of the reaction path as Cartesian coordinates [6], as well as introduce a new reaction representation inspired by protein modeling. Graph neural networks, graph transformers, transformer-based language models and equivariant neural networks in Euclidian space will be discussed as backbone architectures [7] processing these inputs, combined with recent work on generative models to impute missing three-dimensional information [8]. Overall, this work aims to clarify the design principles underlying effective machine learning models for chemical reactions and to advance their practical applicability in sustainable chemistry.

Bibliography :

[1] J. De Landsheere, M. P.-P. Kovar, K. Mark, L. Ganser, L. Galustian, J. Karwounopoulos, C. Gerhaher, E. Heid. Submitted (2026), preprint: 10.26434/chemrxiv-2026-np10c
[2] E. Heid, W. H. Green. J. Chem. Inf. Model. 62 (2022) 2101.
[3] Fujita. J. Chem. Inf. Comput. Sci. 26 (1986) 205.
[4] G. Sulpizio, C. Gerhaher, E. Heid, K. Jorner. Submitted (2026), preprint: 10.26434/chemrxiv.15000926
[5] W. Bort, I. I. Baskin, T. Gimadiev, A. Mukanow, R. Nugmanov, P. Sidorov, G. Marcou, D. Horvath, O. Klimchuk, T. Madzidov, A. Varnek. Sci. Rep. 11 (2021), 3178.
[6] J. Karwounopoulos, J. De Landsheere, L. Galustian, T. Jechtl, E. Heid. Dig. Discov. 4 (2025) 3208
[7] J. De Landsheere, A. Zamyatin, J. Karwounopoulos, E. Heid. J. Chem. Inf. Model. 66 (2026) 2434.
[8] L. Galustian, K. Mark, J. Karwounopoulos, M. P.-P. Kovar, E. Heid. Dig. Discov. 4 (2025) 3492