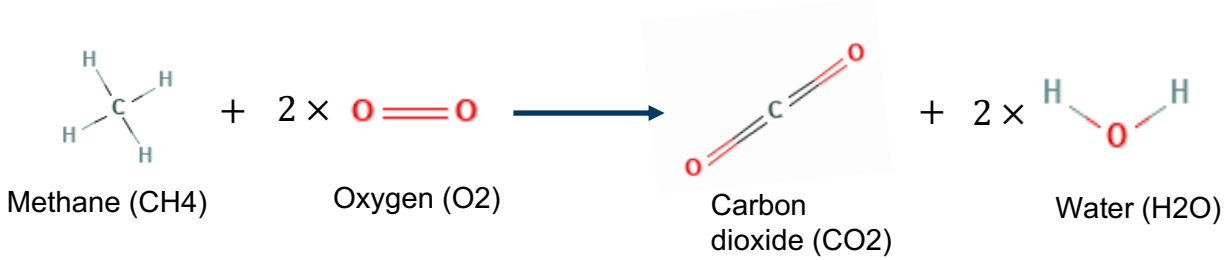
Heinrich Heine
Universität Düsseldorf

Representing chemical reactions

What is a chemical reaction?

- A chemical reaction is a process where substances (reactants/substrates) interact to form different substances (products)
- This is achieved by:
  - Break bonds
  - Forming bonds
  - Rearranging bonds



- Law of conservation of mass: Matter cannot be created or destroyed in a chemical reaction
  - The number and type of atoms must remain constant throughout the reaction
- Different initiation or catalysis mechanisms:
  - Catalytic reactions
  - Enzyme-catalyzed reactions
  - Thermal reactions
  - …

Using SMILES strings to represent reactions

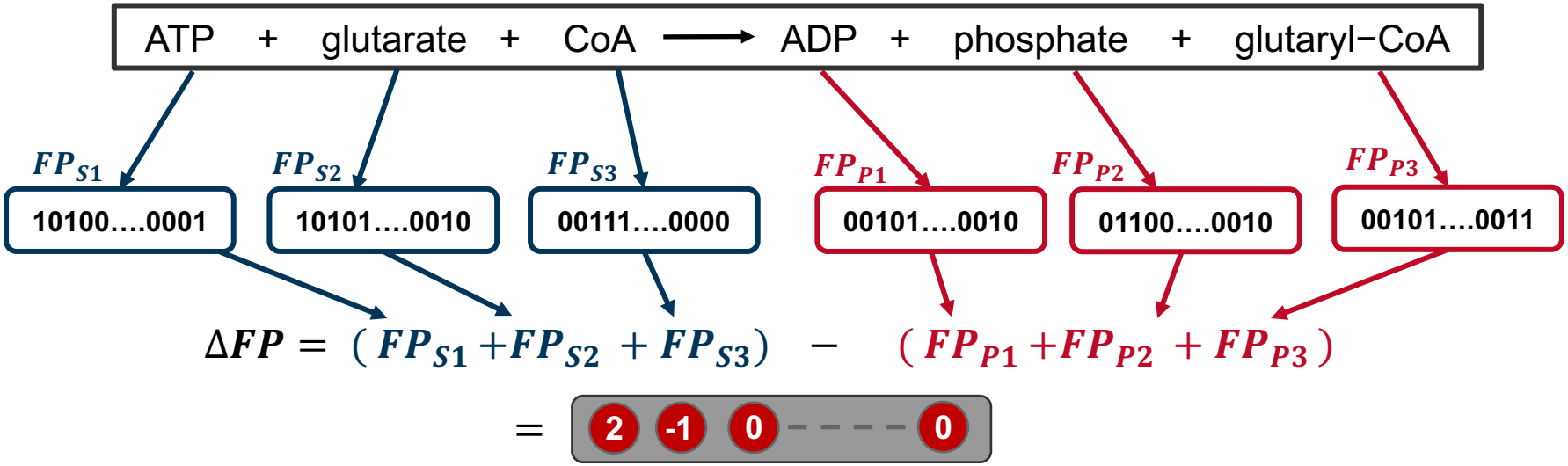
- We can use SMILES strings to represent the reactants and products
- We can combine SMILES strings to write down the whole reaction equations
  - “.” for separating the reactants and products
  - ”>>” for separating reactants from products
- Example:

Methane (CH4) | Oxygen (O2) | Carbon dioxide (CO2) | Water (H2O)

SMILES: C | O=O | O=C=O | O

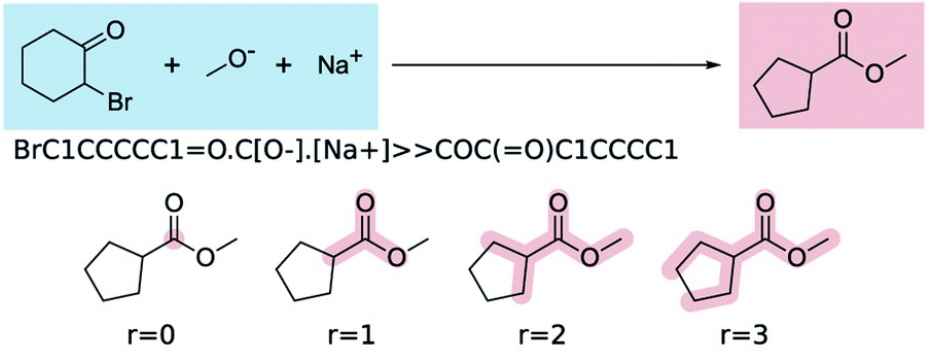
Reaction SMILES: C.O=O.O=O >> O=C=O.O.O

Traditional numerical representations – difference reaction fingerprints



Difference reaction fingerprint: Hu, Q.-N. et al. (2012), *PLoS ONE, 7*, 1–6.

Traditional numerical representations – DRFP



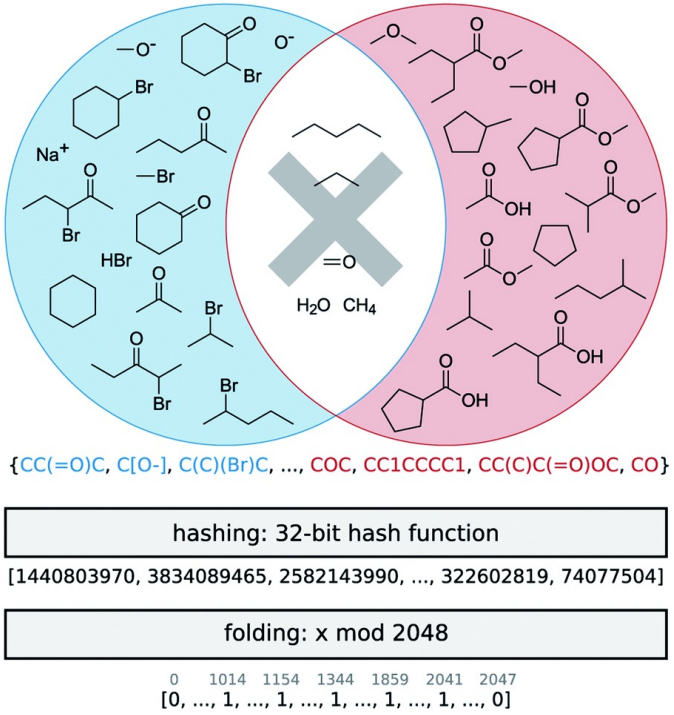
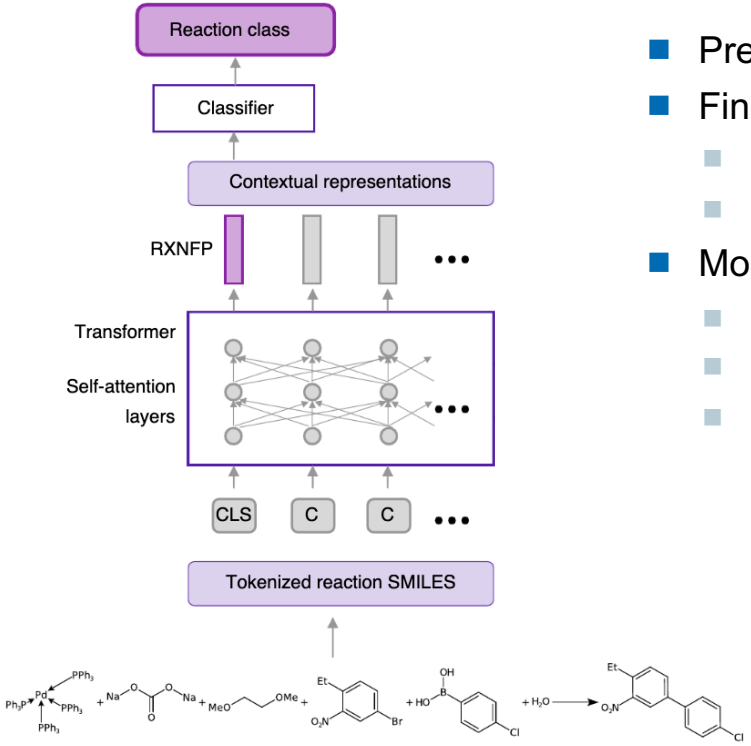


Figure extracted from: Probst, D. *et al.* (2022). Reaction classification and yield prediction using the differential reaction fingerprint DRFP. *Digital discovery*, 1(2), 91-97.

Reaction Transformer Network (rxnfp)



- Pre-training: Masked Language Modelling (MLM)
- Fine-tuning: Predicting reaction classes
  - 2.6 million reaction equations
  - ~1000 reaction classes
- Model Architecture
  - Encoder-only model (BERT)
  - Hidden dimension $d = 256$
  - Number of layers encoder layer:

Figure extracted from: Schwaller, Philippe, et al. "Mapping the space of chemical reactions using attention-based neural networks." Nature machine intelligence 3.2 (2021)