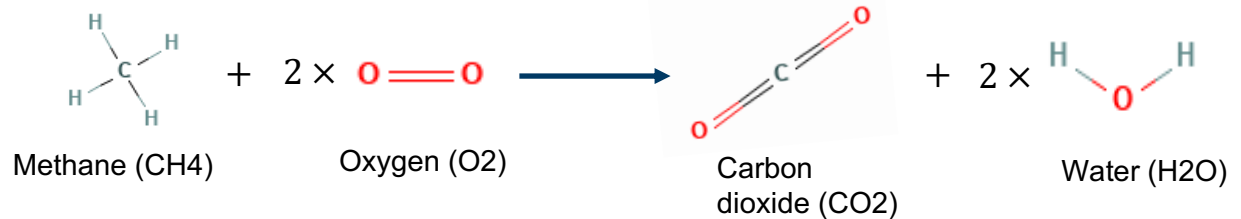# Representing chemical reactions

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

Methane (CH4) Oxygen (O2) → Carbon dioxide (CO2) + 2 × Water (H2O)

- 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
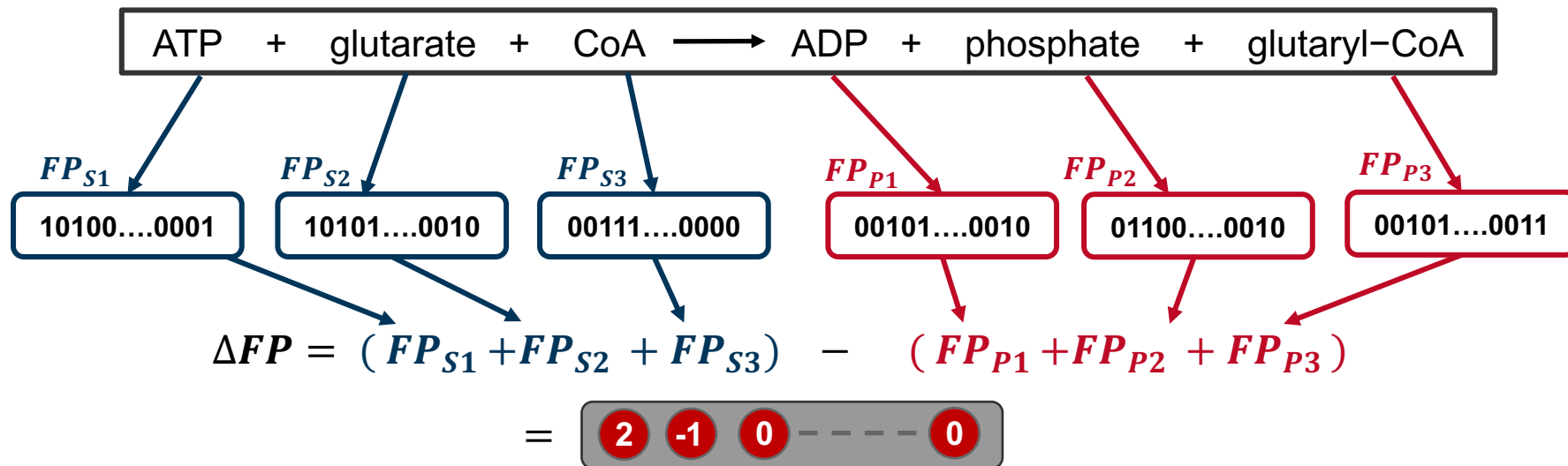 - …

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)
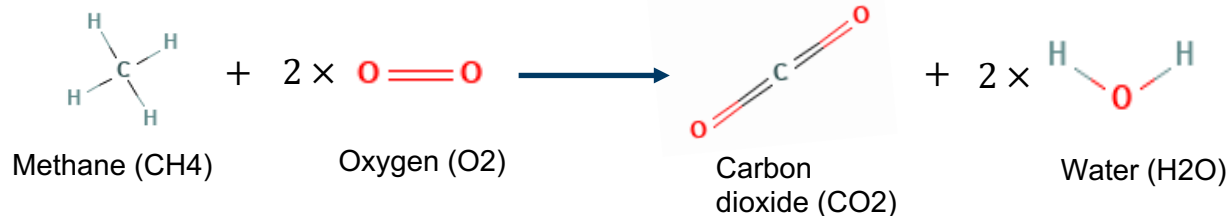
SMILES: C   O=O   O=C=O   O

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

Traditional numerical representations – difference reaction fingerprints
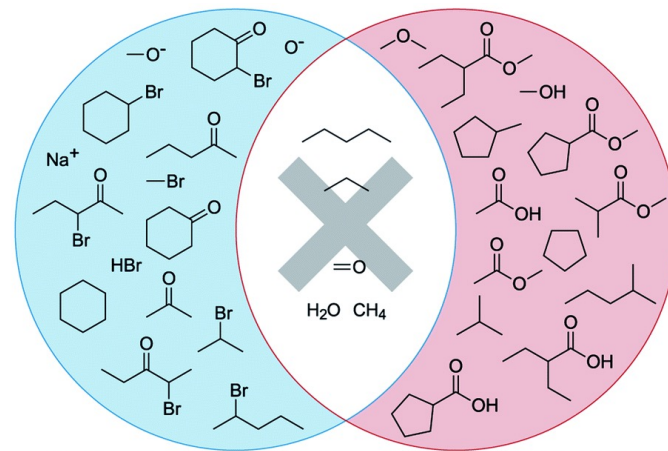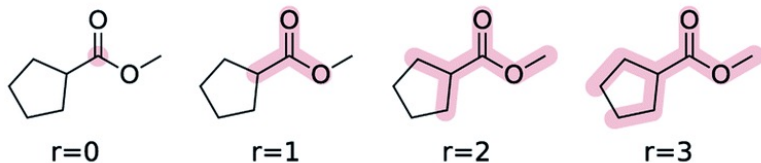
$$ATP \;+\; glutarate \;+\; CoA \longrightarrow ADP \;+\; phosphate \;+\; glutaryl-CoA$$

FP_{S1} 10100….0001

FP_{S2} 10101….0010

FP_{S3} 00111….0000

FP_{P1} 00101….0010

FP_{P2} 01100….0010

FP_{P3} 00101….0011

$$\Delta FP = (FP_{S1} + FP_{S2} + FP_{S3}) - (FP_{P1} + FP_{P2} + FP_{P3})$$

$$= \quad \boxed{2 \quad -1 \quad 0 \quad - - - - \quad 0}$$

hhu.de

Reaction Transformer Network (rxnfp)
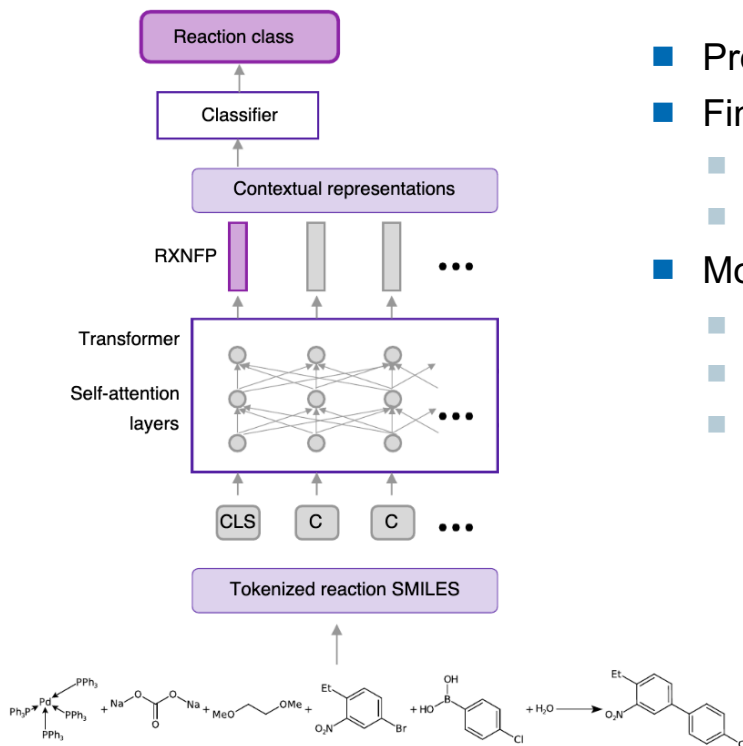
- Pre-training: Masked Language Modelling (MLM)
- Fine-tuning: Predicting reaction classes
  - 2.6 million reaction equations
  - ~1000 reaction classes
- Model Architecture
  - Encoder-only model (BERT)
  - Hidden dimension $d = 256$
  - Number of layers encoder layer: 12

hhu.de