



HAL
open science

Artificial Intelligence Methods and Models for Retro-Biosynthesis

Guillaume Gricourt, Philippe Meyer, Thomas Duigou, Jean-Loup Faulon

► **To cite this version:**

Guillaume Gricourt, Philippe Meyer, Thomas Duigou, Jean-Loup Faulon. Artificial Intelligence Methods and Models for Retro-Biosynthesis. AI Methods and Models for (bio)catalysis and Synthetic Biology, Jun 2024, Montpellier, France. hal-04603192

HAL Id: hal-04603192

<https://hal.science/hal-04603192>

Submitted on 6 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Artificial Intelligence Methods and Models for Retro-Biosynthesis¹

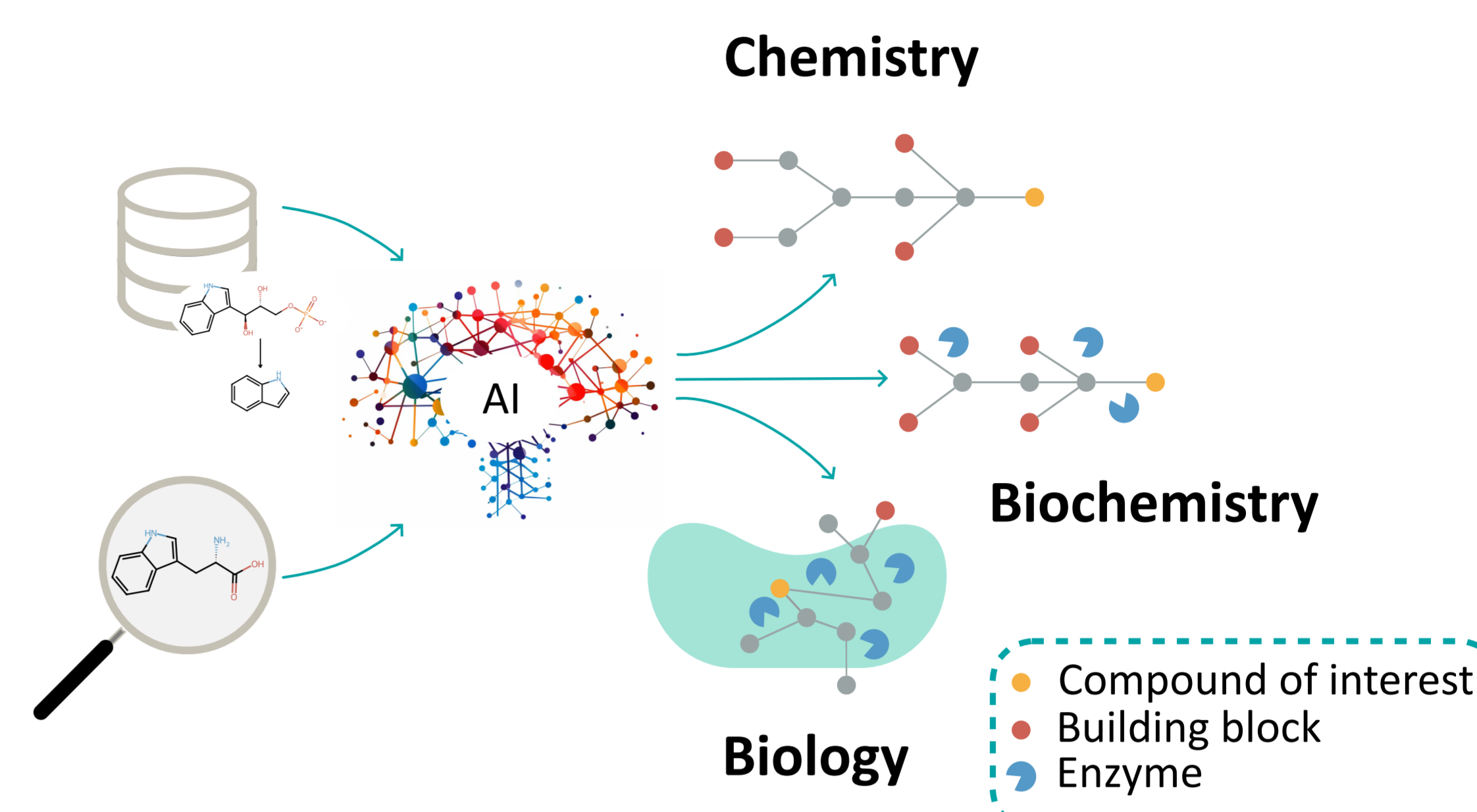
G. Gricourt[†], P. Meyer[†], T. Duigout[†], and J.-L. Faulon^{†,‡}

[†] Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78352 Jouy-en-Josas, France

[‡] Manchester Institute of Biotechnology, SYNBIOCHEM center, School of Chemistry, University of Manchester, Manchester M1 7DN, UK

1 — Context

- **Retrosynthesis**: process of **planning routes** for **chemical synthesis** by iteratively searching for substrates/precursors of reactions starting from the final product.
- **Retro-biosynthesis**: extension of retrosynthesis to other fields, **biocatalysis** and **synthetic biology**, aligned with **green chemistry**.
- Retro-(bio)synthesis main stages consist of (i) **single-step** executions where substrates are predicted from products and (ii) a **multi-step** strategy to explore as efficiently as possible the tree of possibilities for reaching best synthesis plannings.
- **AI methods** are nowadays employed within both single-step and multi-step stages.



2 — Single-step methods

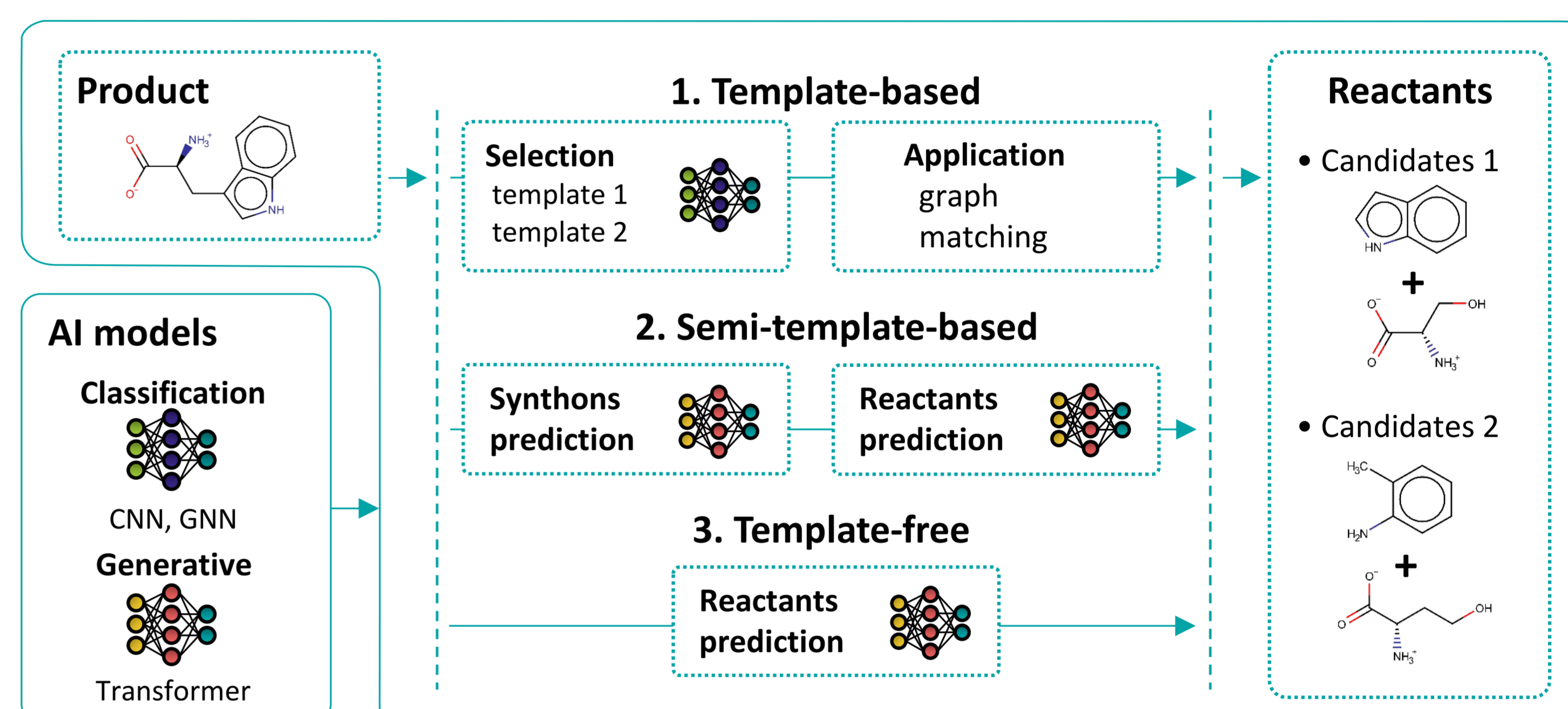
Identify reactants given a product using **template-based** (1), **semi-template-based** (2), or **template-free** (3) methods.

Key considerations

- Chemical reaction database: Brenda, Rhea, ...
- Molecular representation: SMILES, Fingerprint, ...
- Model evaluation: MRR, MaxFrag, ...

Representative implementations

EHReact², RetrosimEnz³, ...



3 — Multi-step approaches

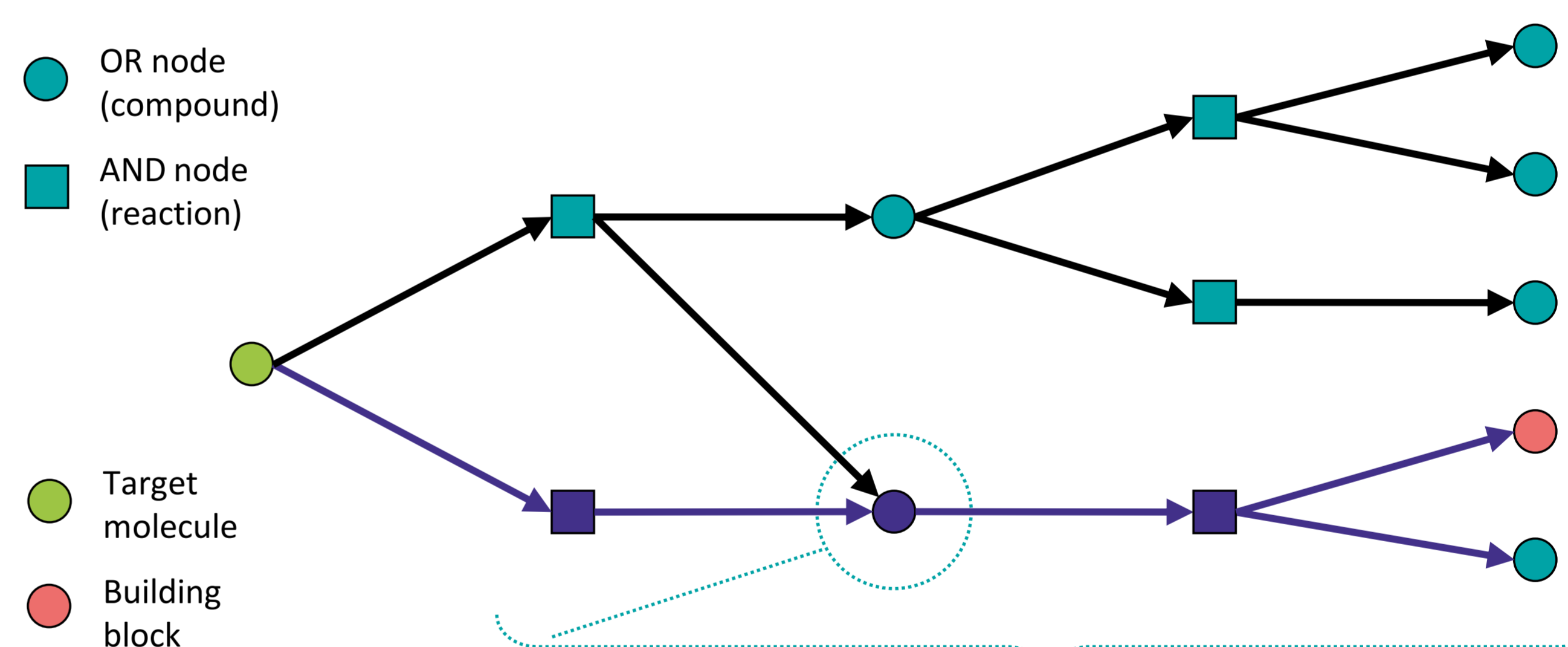
Planning the **selection** of single-step **reactions** from target molecule to building blocks.

Key considerations

- Planning algorithms: Beam Search, MCTS, A*, Reinforcement learning, ...
- Building block molecules: buyable chemicals, chassis metabolites, ...
- Scoring functions, assessing predicted chemicals, reactions and routes: DFRScore, ...

Representative implementations

Retropath2.0⁴, BioNavi-NP⁵, ...

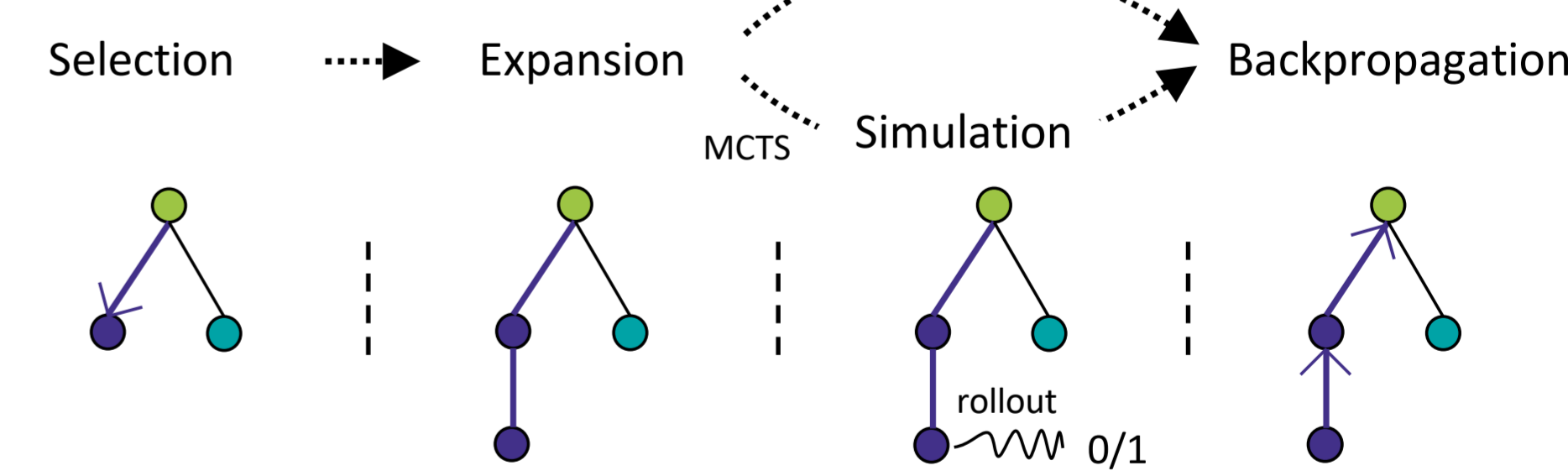


Scoring function

- Chemical evaluations
- Reaction conditions
- Pathway related

Planning algorithms

A* search & Monte Carlo tree search

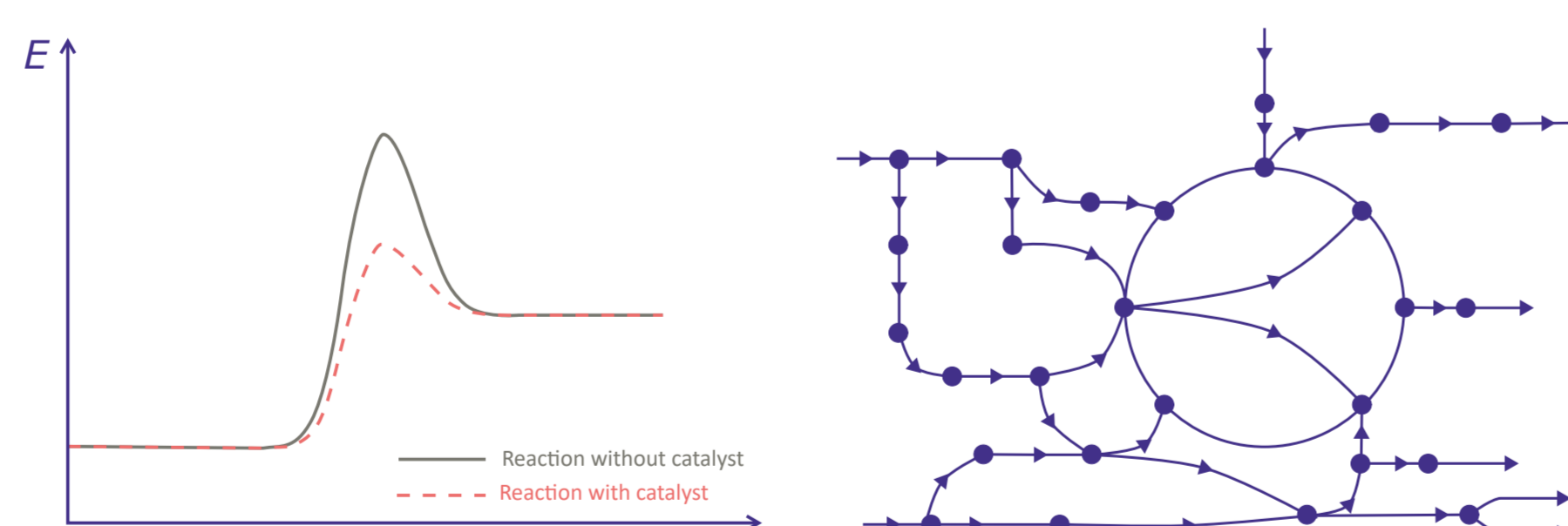


4 — The specifics of retro-biosynthesis

Retro-biosynthesis involves **enzymes** as catalyst and specific reaction conditions, eventually taking place within a **chassis**.

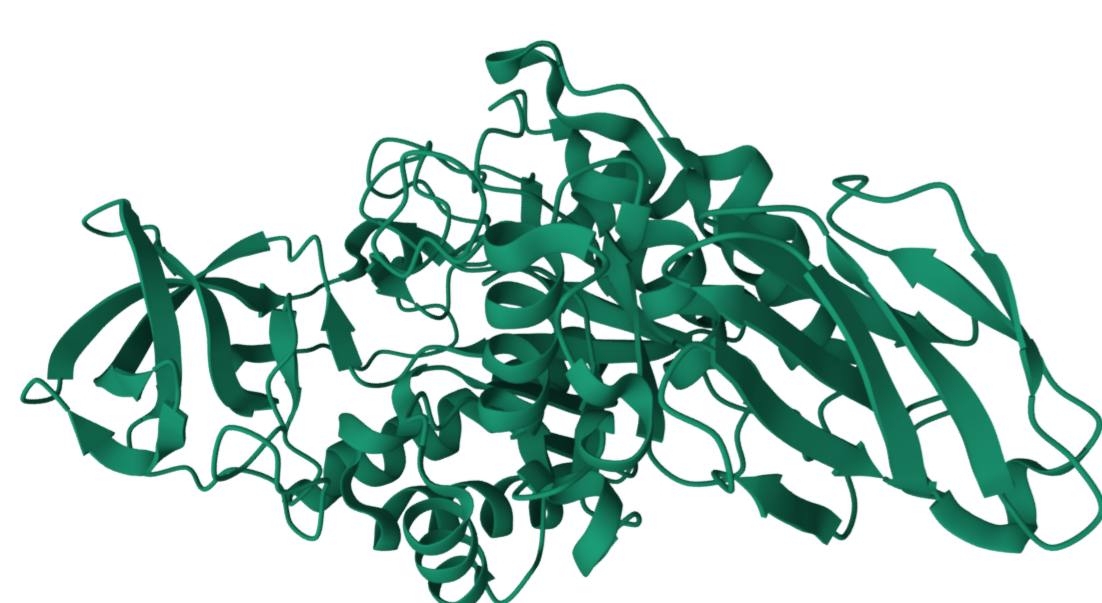
Assert pathway feasibility

- Reaction stoichiometry
- Thermodynamic
- Kinetic



Enzyme constraints

- Identify an enzyme: DeepRFC, ...
- Reaction conditions, such as solvent, pH, temperature, and pressure: InFlow, ExtractionScore, ...



Bibliography

1. Gricourt G.; Meyer, P.; Duigou, T.; Faulon, J.L.. Artificial Intelligence Methods and Models for Retro-Biosynthesis: a scoping review. ACS Synthetic Biology. in press.
2. Heid, E.; Goldman, S.; Sankaranarayanan, K.; Coley, C. W.; Flamm, C.; Green, W. H. EHreact: Extended Hasse Diagrams for the Extraction and Scoring of Enzymatic Reaction Templates. J. Chem. Inf. Model. 2021, 61 (10), 4949–4961. <https://doi.org/10.1021/acs.jcim.1c00921>.
3. Sankaranarayanan, K.; Heid, E.; Coley, C. W.; Verma, D.; Green, W. H.; Jensen, K. F. Similarity Based Enzymatic Retrosynthesis. Chem. Sci. 2022, 13 (20), 6039–6053. <https://doi.org/10.1039/D2SC01588A>.
4. Delépine, B.; Duigou, T.; Carbonell, P.; Faulon, J.-L. RetroPath2.0: A Retrosynthesis Workflow for Metabolic Engineers. Metab. Eng. 2018, 45, 158–170. <https://doi.org/10.1016/j.ymben.2017.12.002>.
5. Zheng, S.; Zeng, T.; Li, C.; Chen, B.; Coley, C. W.; Yang, Y.; Wu, R. Deep Learning Driven Biosynthetic Pathways Navigation for Natural Products with BioNavi-NP. Nat. Commun. 2022, 13 (1), 3342. <https://doi.org/10.1038/s41467-022-30970-9>.