Bidirectional Mass Spectrometry Modeling Aided by Auxiliary Intermediate Steps From Graph Transformation Tools

Philipp Reiser

TBI Uni Vienna

2026-02-11

What is my thesis about?

- Formulated mathematically as a function:

$$f : \text{molecular structure} \rightarrow \text{mass spectrum}$$

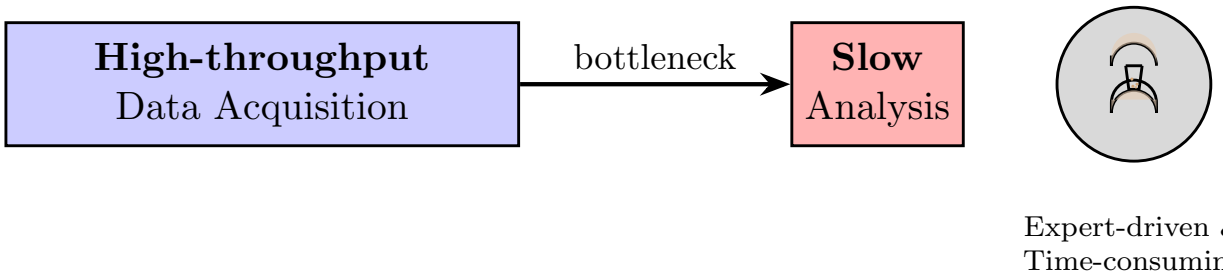
- Investigation of the **inverse problem**:

$$f^{-1} : \text{mass spectrum} \rightarrow \text{molecular structures}$$

- Integration of **chemically studied reaction rules** modeled in **MØD**
- Use of fragmentation trees to support traditional machine-learning-based MS models

Motivation

MS is widely used, but analysis is the challenge

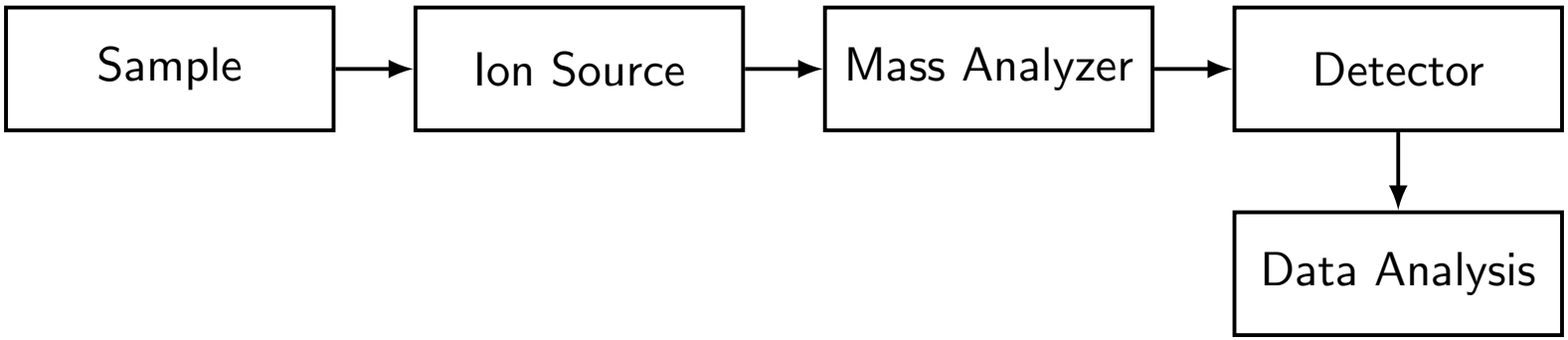


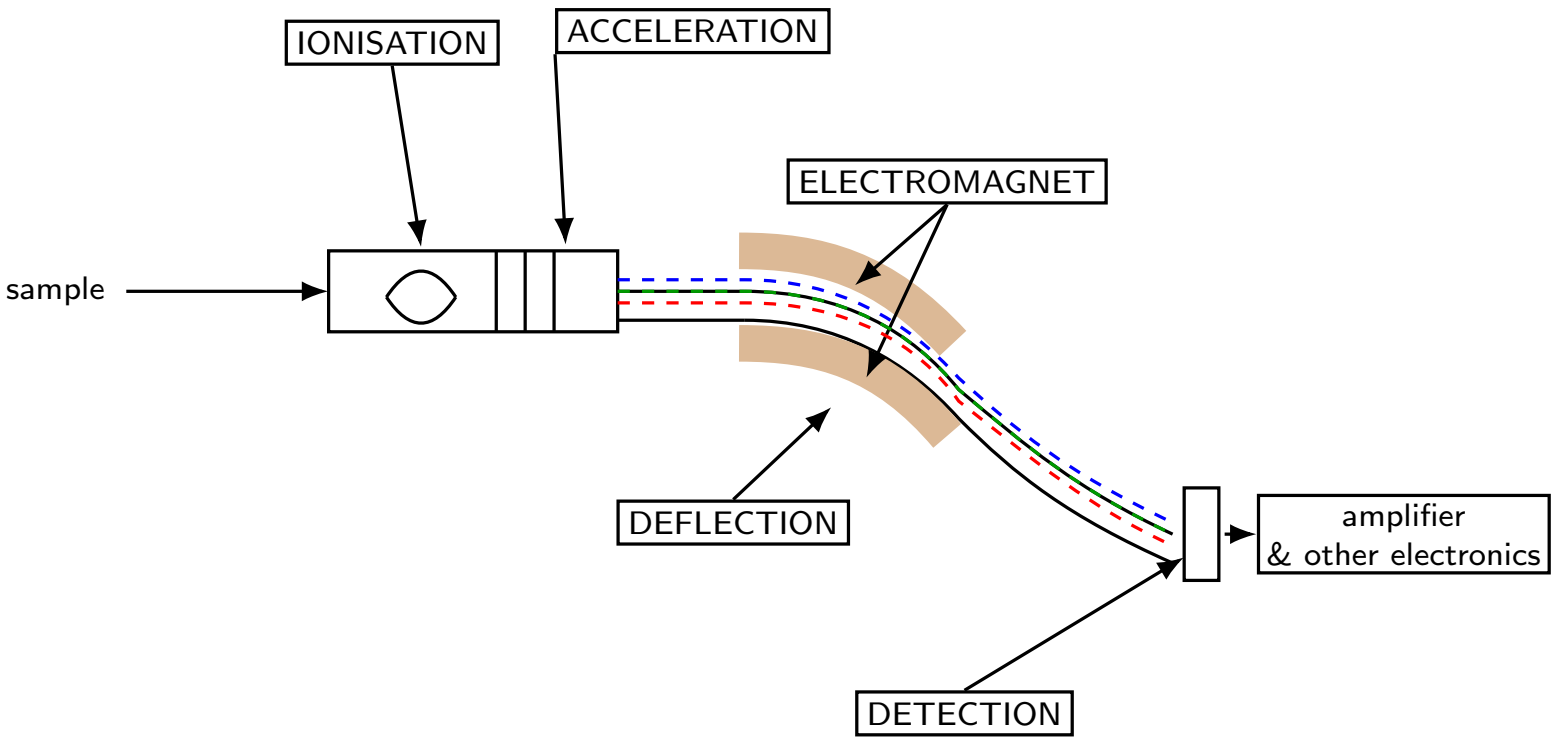
Pure ML
No structure
× interpretability
× generalization

VS

ML + Structure
Chemical rules
✓ interpretable
✓ generalizable

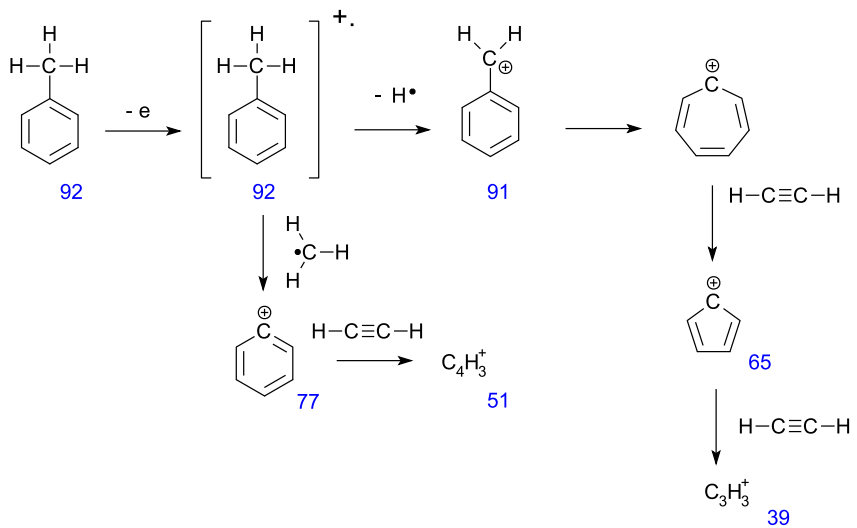
Mass Spectrometer

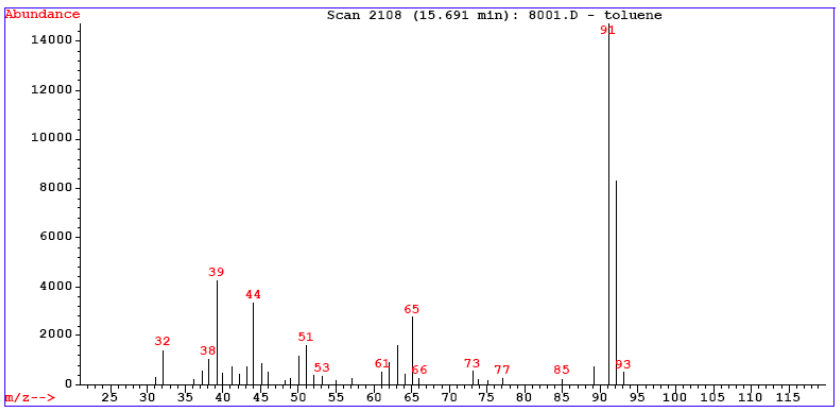




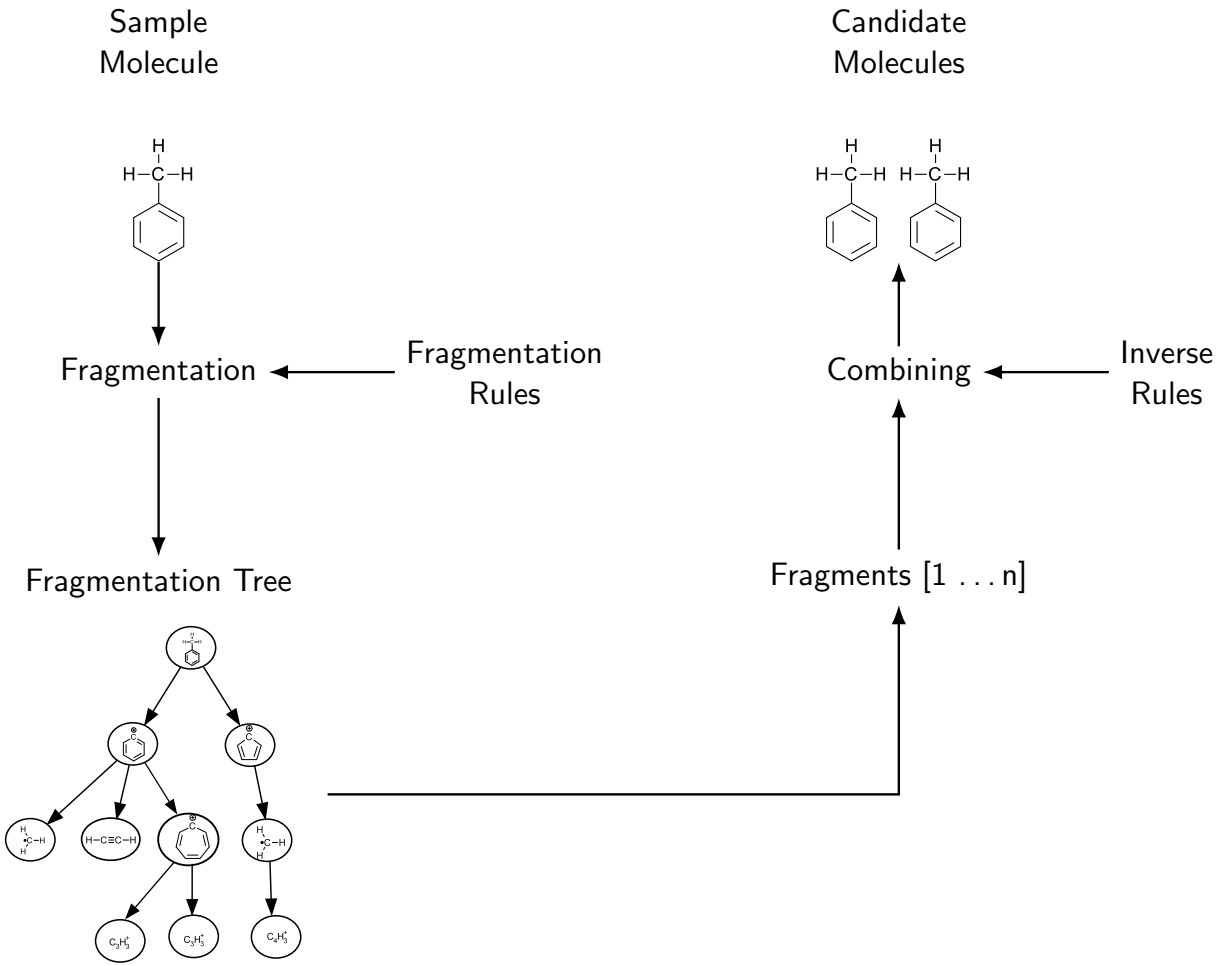
Data Sources

- Mass spectrometry data from NIST → SMILES strings & spectra
- Curated EI fragmentation reactions → **MØD** rules

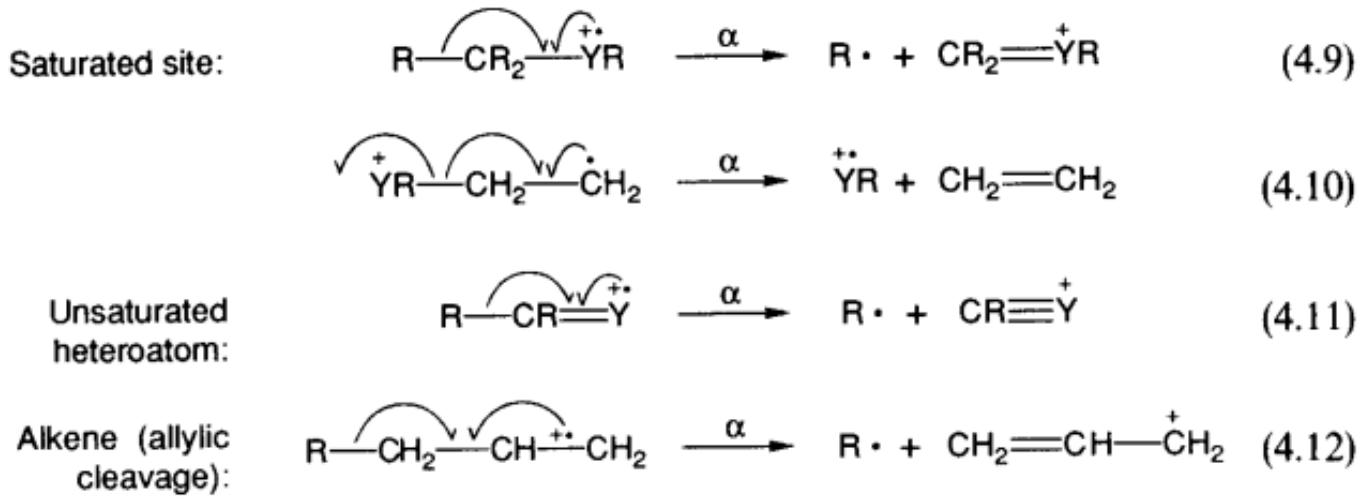




Data Generation Overview

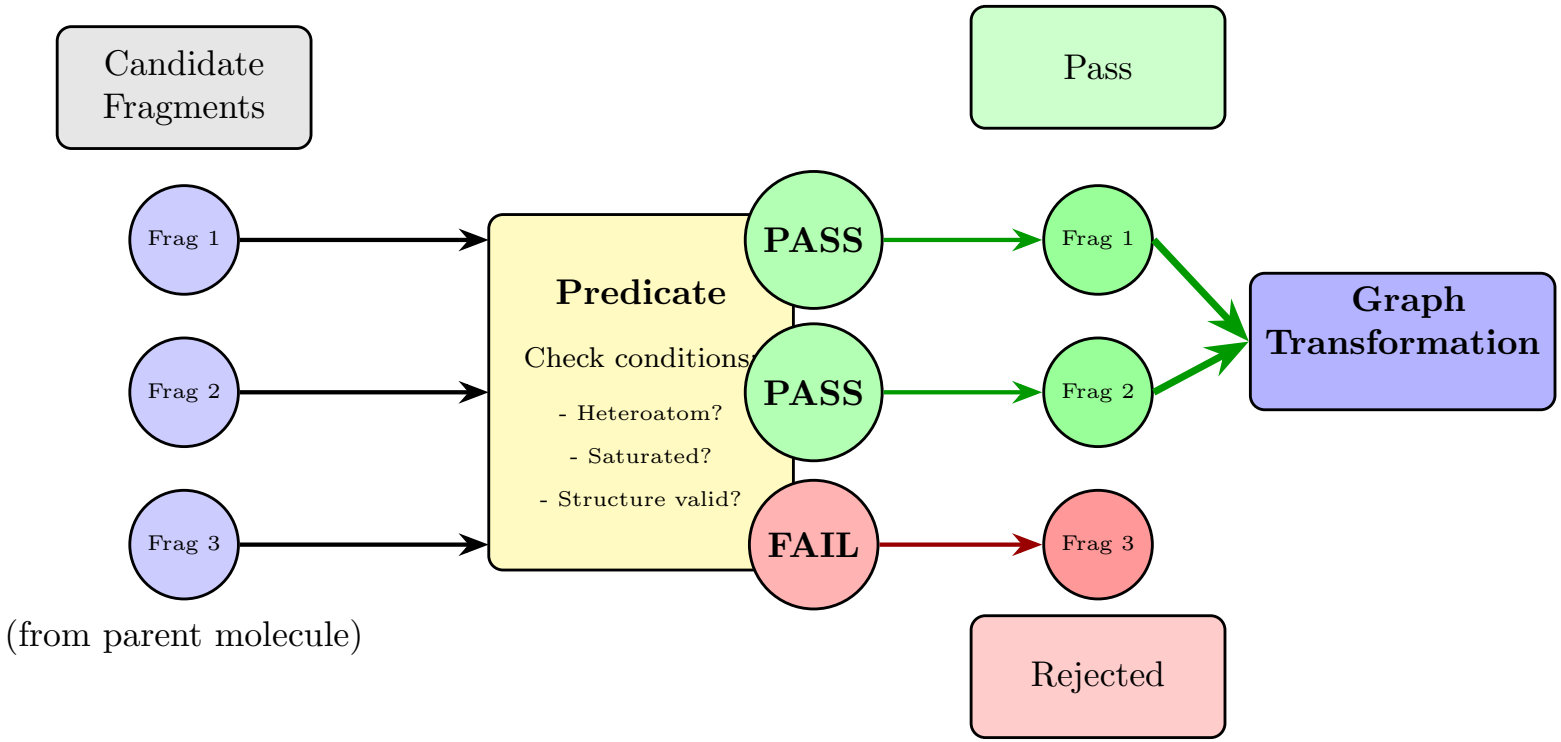


Sample Fragmentation Rules

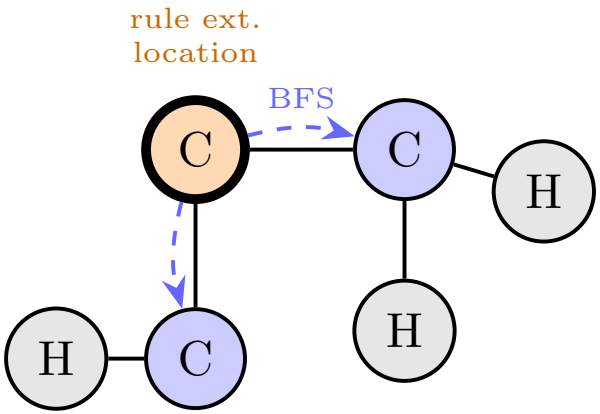



```
IMS_4_9 = mod.Rule.fromDFS(
    s =
    "[C]1[C]2[C]3([C]4)[_A+.]5"
    ">>"
    "[C.]1" "." "[C]2[C]3([C]4){=}[_A+.]5",
    name =
    "radical induced (alpha-)cleavage for a saturated site"
    " 4.9"
    " !R1R3R4Y5"
)
```

MØD - Predicate

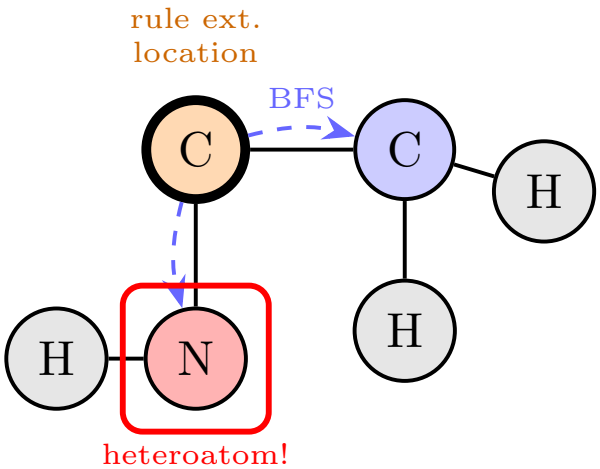


MØD - Predicate - Alkyl Group



VALID: Alkyl Group

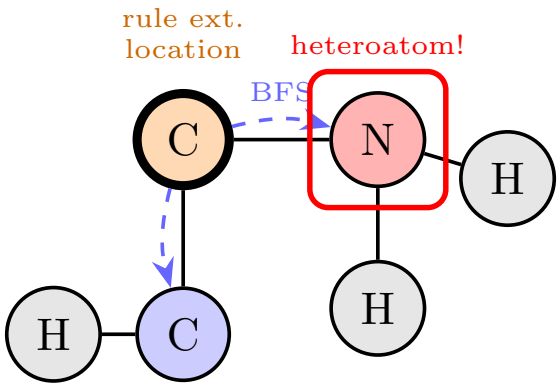
BFS neighbors $\subseteq \{H, C\}$ ✓



INVALID: Not Alkyl

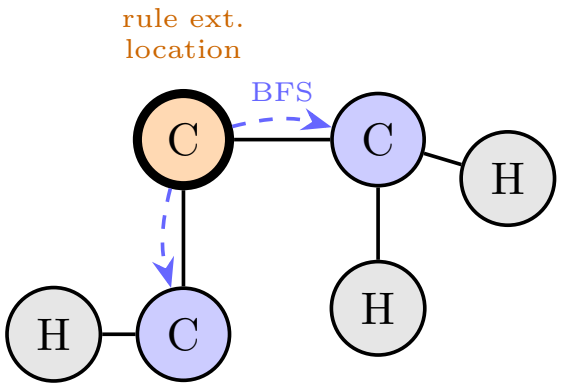
BFS neighbors $\not\subseteq \{H, C\}$ ×

MØD - Predicate - Heteroatom Group



VALID: Heteroatom Group

BFS \ {H, C} ≠ ∅ ✓



INVALID: No Heteroatom

BFS \ {H, C} = ∅ ×

MØD - Predicate - Saturation

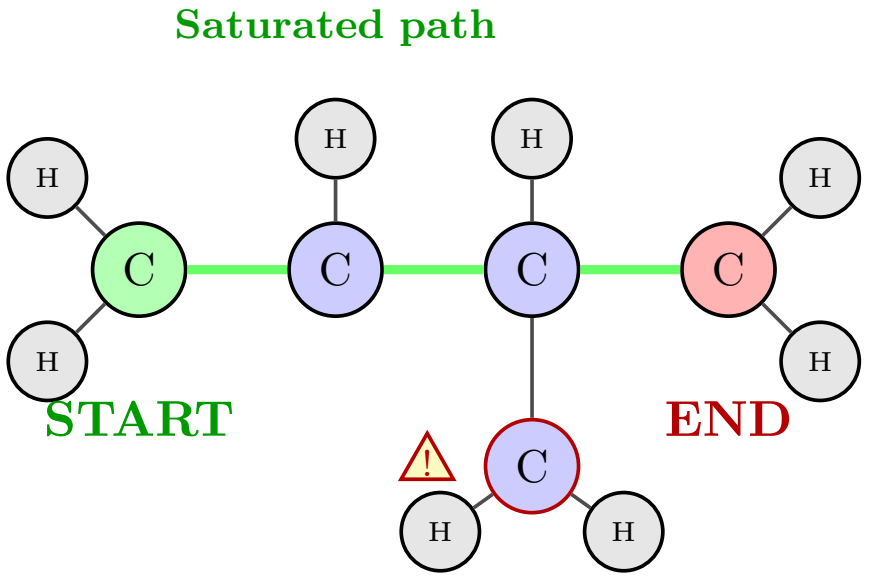
Goal: Find saturated path

Requirements:

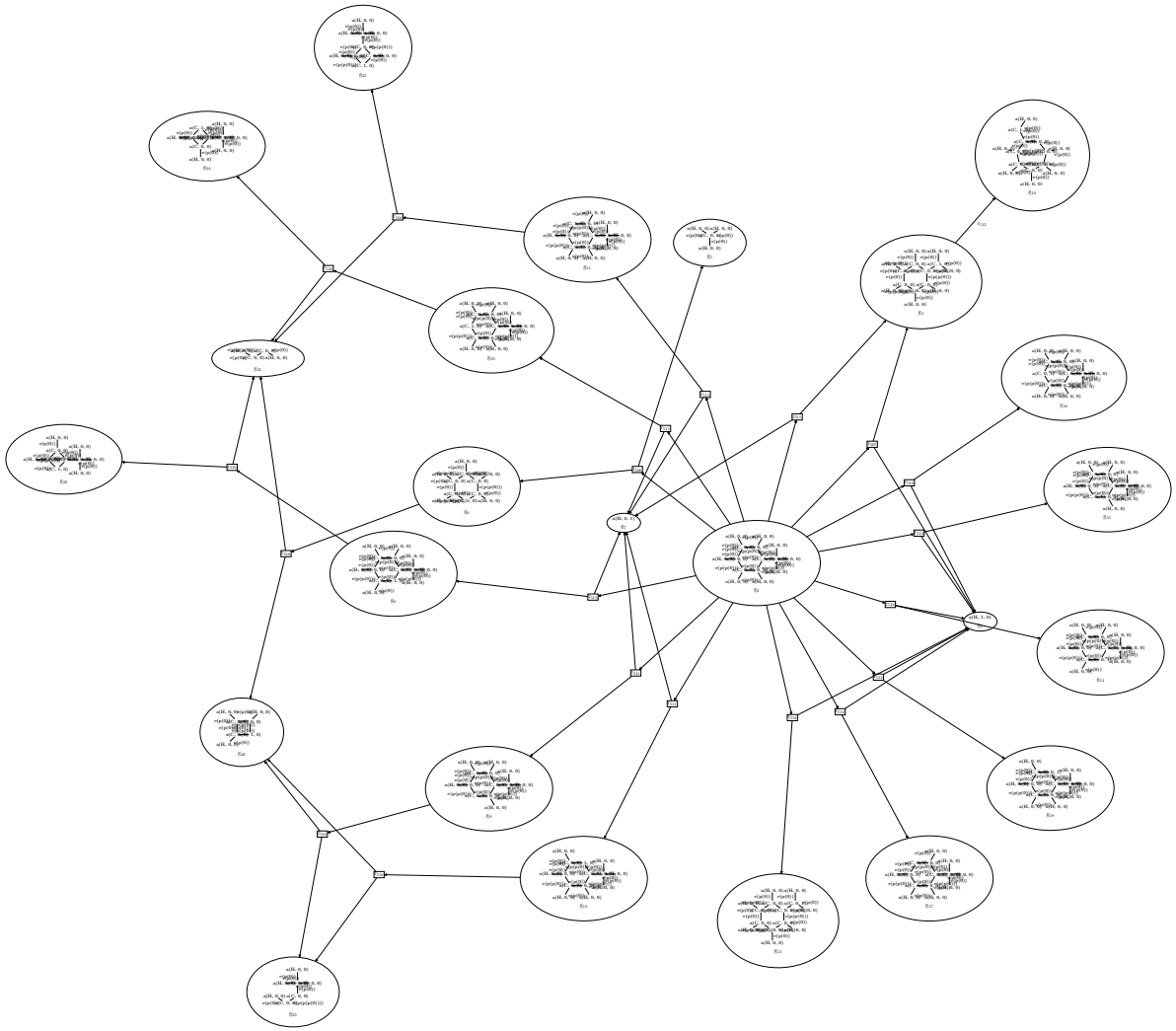
✓ Only carbon backbone
✓ Single bonds only (C—C)
✓ Only H side branches

Method:

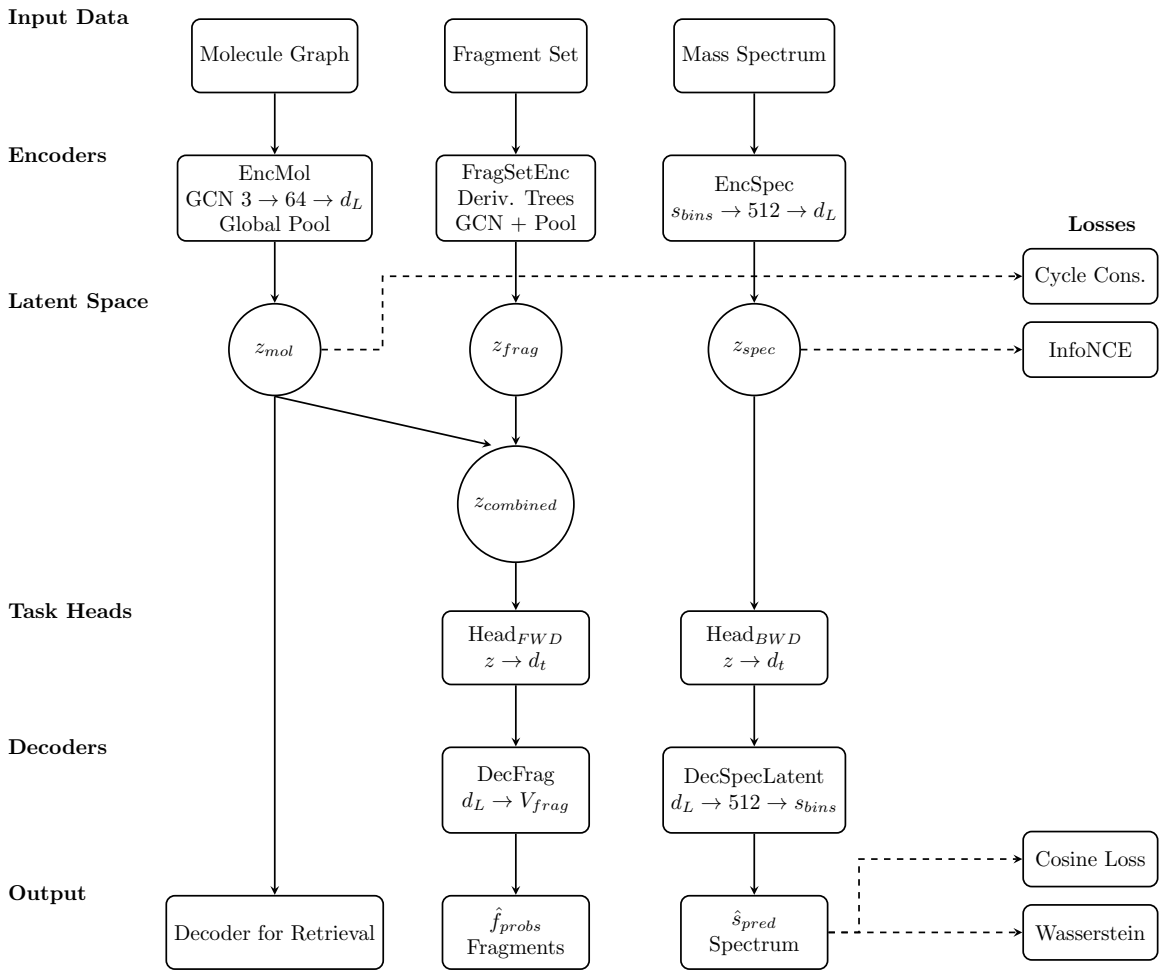
Search from start to end
following C-C single bonds



Sample Derivation Graph



Architecture Diagram



Results

Diverse Model scores:

Best Recall@1:	0.000
Best Recall@5:	0.111
Best Recall@10:	0.259

References

[1] V8rik. *Toluene Fragmentation*. 2008. URL: https://upload.wikimedia.org/wikipedia/commons/6/6c/TolueneFragmentation.svg (visited on 01/20/2026).

[2] Paginazero, Ronhjones. *Toluene Spectra*. 2017. URL: https://upload.wikimedia.org/wikipedia/commons/2/2c/Spettro_di_massa_del_toluene.PNG (visited on 01/20/2026).

[3] Fred W McLafferty and František Tureček. *Interpretation von Massenspektren*. ger. Spektrum Lehrbuch. Berlin [u.a.]: Spektrum Akad. Verl., 1995. ISBN: 9783642398483.

Discussion

Questions?