

# Talk Chemistry to Me

## LLMs die Sprache der Reaktionen lehren

Inola Bauer, Marco Grünewald, Paul Schubert, Brian Smith, Marius Verdaasdonk, Burkhard Wagler, Maike Mücke, Prof.Dr. Ricardo A. Mata

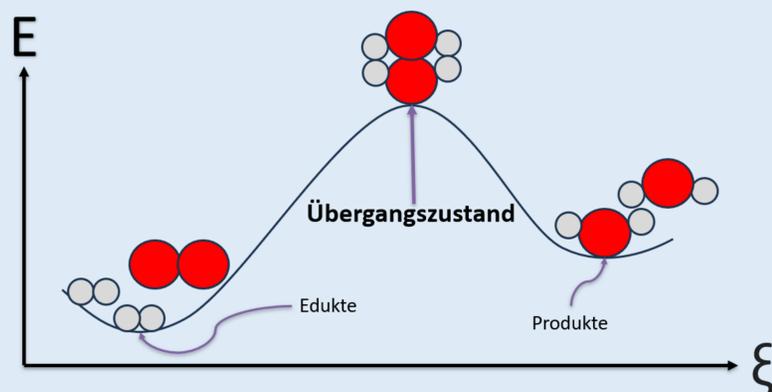
### Einführung

**Übergangszustände** sind eine Schlüsselfigur für die Katalysechemie und die Aufklärung von Reaktionsmechanismen, die aufgrund ihrer extremen Kurzlebigkeit nicht isolierbar sind und mühselig berechnet werden müssen. Um herauszufinden, wie gut künstliche Intelligenz (KI) dazu in der Lage ist, Übergangszustände vorherzusagen, haben wir uns zum Ziel gesetzt, das chemische Verständnis verschiedener **Large Language Models (LLM)** zu testen.

### Probleme

- LLMs sind textbasiert, Moleküle nicht
- Große Datensätze müssen automatisiert verarbeitet werden können
- Datensätze müssen vor der Nutzung auf fehlerhafte oder problematische Einträge überprüft werden

### Übergangszustände



### Moleküldarstellung

- Automatisierung aller Schritte mit **Python** (Hauptsächlich RDKit)
- Umwandlung von Strukturformeln in **text- und matrixbasierte Darstellungsmethoden**

	1-Cyclopropylethen-1-ol
Darstellung	Moleküldarstellung
Summenformel	C <sub>5</sub> H <sub>8</sub> O
InChi	InChi=1S/C5H8O/c1-4(6)5-2-3-5/h5-6H,1-3H2
Smiles	C=C(O)C1CC1
Smarts	[#6](=[#6](-[#8]-[H]))-[#6]1(-[#6](-[#6]-1(-[H])-[H])(-[H])-[H])-[H])(-[H])-[H]
Selfies	[C]=[C][Branch1][C][O][C][C][C][Ring1][Ring1]

### LLMs

- Einigung auf drei verschiedene LLMs:
  - Llama**: US-Modell von Meta, bekanntestes der drei
  - Mistral**: vielversprechendes, französisches Modell
  - Qwen**: stark zensiertes, chinesisches Modell, erkennt matrixbasierte Darstellungen nicht
- Jede KI soll mit jeder Darstellungsmethode getestet werden
- Formulierung eines umfassenden **System Prompts**
- Lückentext-artige **Templates** zur schnellen Kombination von neuen Prompts und Fragestellungen
- Fokus hauptsächlich auf leicht beantwortbare und auszuwertende single shot prompts
- Statistische Auswertung der erhaltenen Antworten

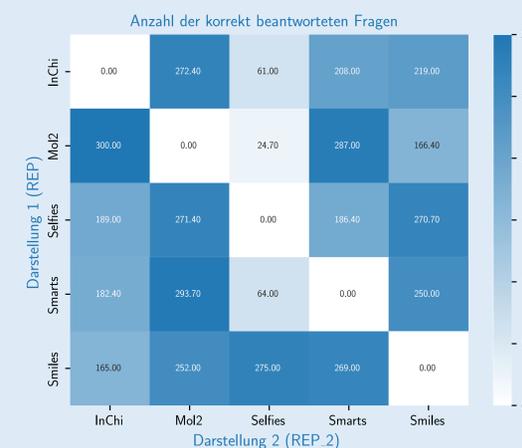
### Ergebnisse

Es konnten nicht alle Prompts getestet werden. Die getesteten liefern folgende Ergebnisse:

**Prompt 3:** Of which molecule is REP a correct METHOD representation?



**Prompt 5 mit dem LLM Mistral:** Does the METHOD representation REP refer to the same molecule as the METHOD\_2 representation REP\_2?



**Prompt 7:** Of which method is REP a valid representation?



### Abschließendes

- Alle LLMs performen vergleichbar
  - InChis werden zuverlässig gelesen
  - Mol2 mit größeren Hürden versehen
- Wir sind der Ansicht, dass LLMs in naher Zukunft Übergangszustände erkennen werden und sind auf Fortschritte in dem Forschungsfeld gespannt!