BigChem ESR7 PhD Project

Mid-term Progress Report

Exploration of uncharted regions of chemical space by reaction-driven *de novo* design

October 23, 2017

Xuejin Zhang

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Overview

1. Introduction
2. Methodology
3. Enumeration Software
4. Preliminary Results
5. Conclusion and Outlook
6. Education

1. Introduction

张雪瑾 *Xuejin Zhang, or JJ,* **中国人**

25 years old, Bcs & Msc China Pharmaceutical University; worked at WuXi AppTec Co. 06.2015-07.2016.

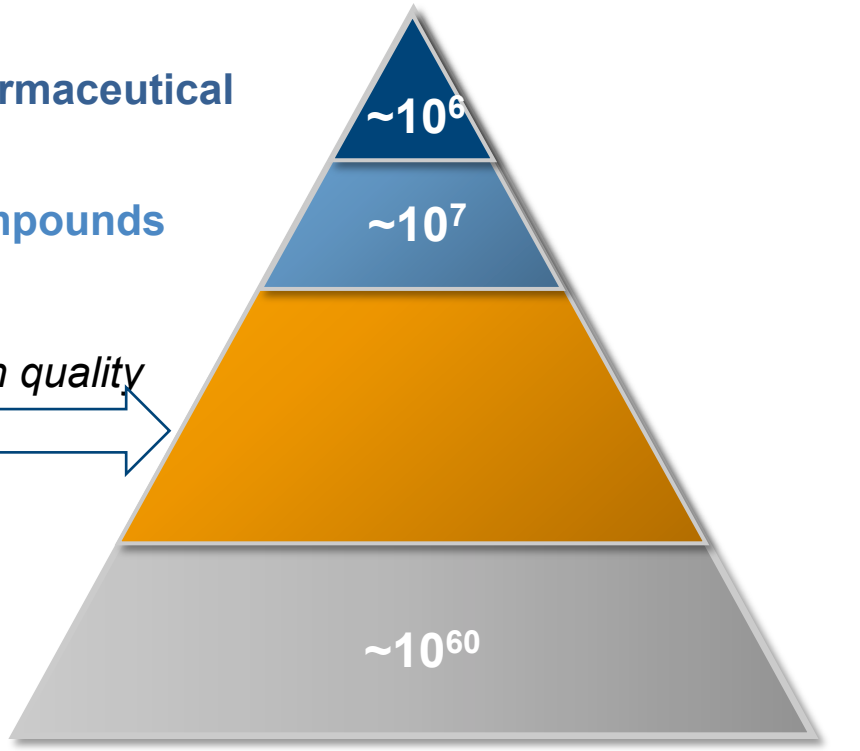
Currently is registered at ETHZ & working at BI (secondment).

Screening collection of pharmaceutical companies

Commercially available compounds

Synthetically accessible & high quality chemical space

Theoretical total chemical space



1. Introduction

Objectives of ESR7:

Bigger is not better:

- Implement an algorithm to translate multi-step organic synthesis protocols into a **high-quality & novel** chemical space of **synthetically accessible** molecules.

Clever Mining:

- Perform extensive data analysis to inspect the chemical space properties
- Quantitative comparison of this chemical space with available chemical libraries

Exploit future:

- Employ *de novo* design of compounds with a desired profile.

Proof-of-Concept:

- Synthesize & test.

2. Methodology

Open source software:

- RDKit: toolkit for cheminformatics
- Scikit learn: data mining, data analysis and machine learning in Python

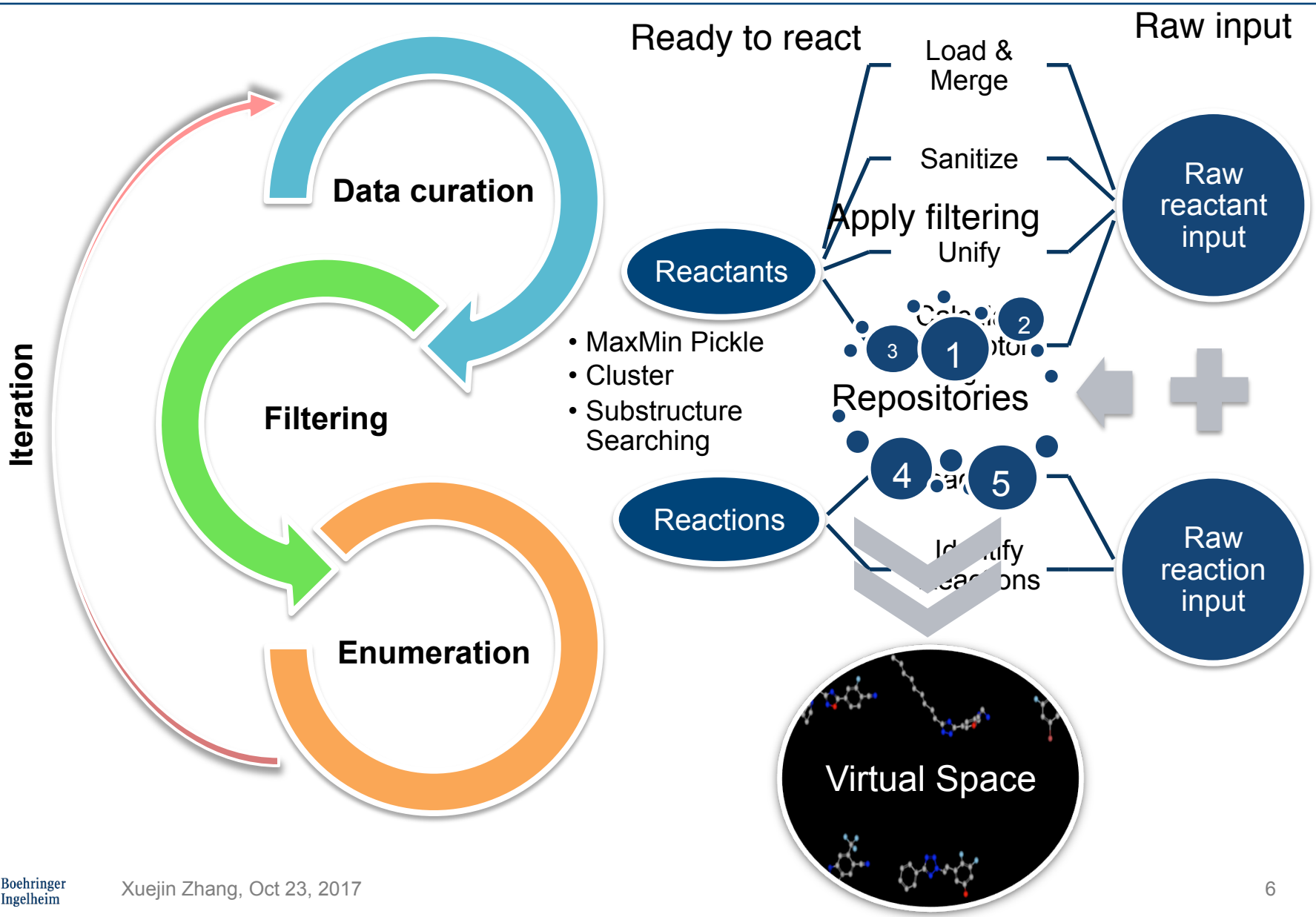
Chemical reactions:

- 209 reactions from ChemAxon reaction library
- *literature collection of most commonly used organic chemical reactions*

Reagents input source:

- *11045 reagents BIOVIA Available Chemicals Directory (ACD, MW <= 250 Da)*

3. Enumeration Software Cycle



3. Enumeration Software Features

- User friendly
  - Automatically enumeration after uploading input
  - Automatically calculation of molecular properties
  - Automatically generated graphic reports after enumeration

- Full flexibility
  - Predefined quality control rules & customized constraints on space generation

- Diverse methodologies
  - MaxMin diversity, clustering, random, MW, number heavy atoms etc.

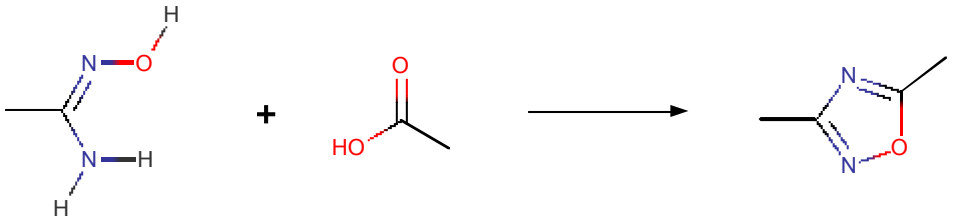
- Multiprocessing jobs

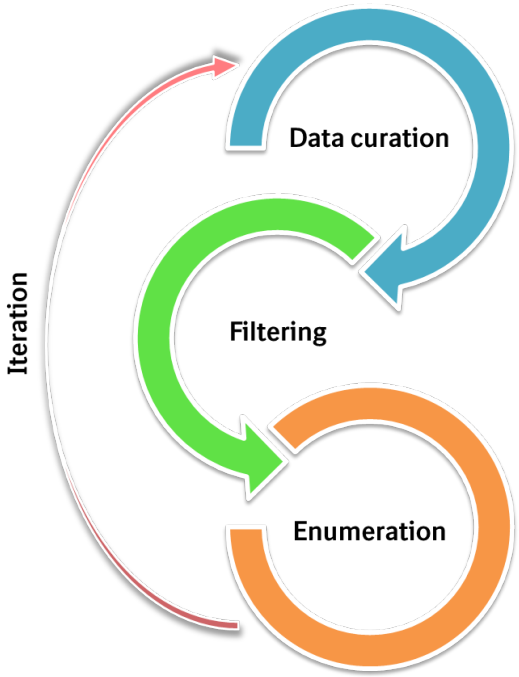
- Complete open source

4. Preliminary Results

In-depth analysis on five chemical reactions

1) 1-2-4-Oxadiazole formation

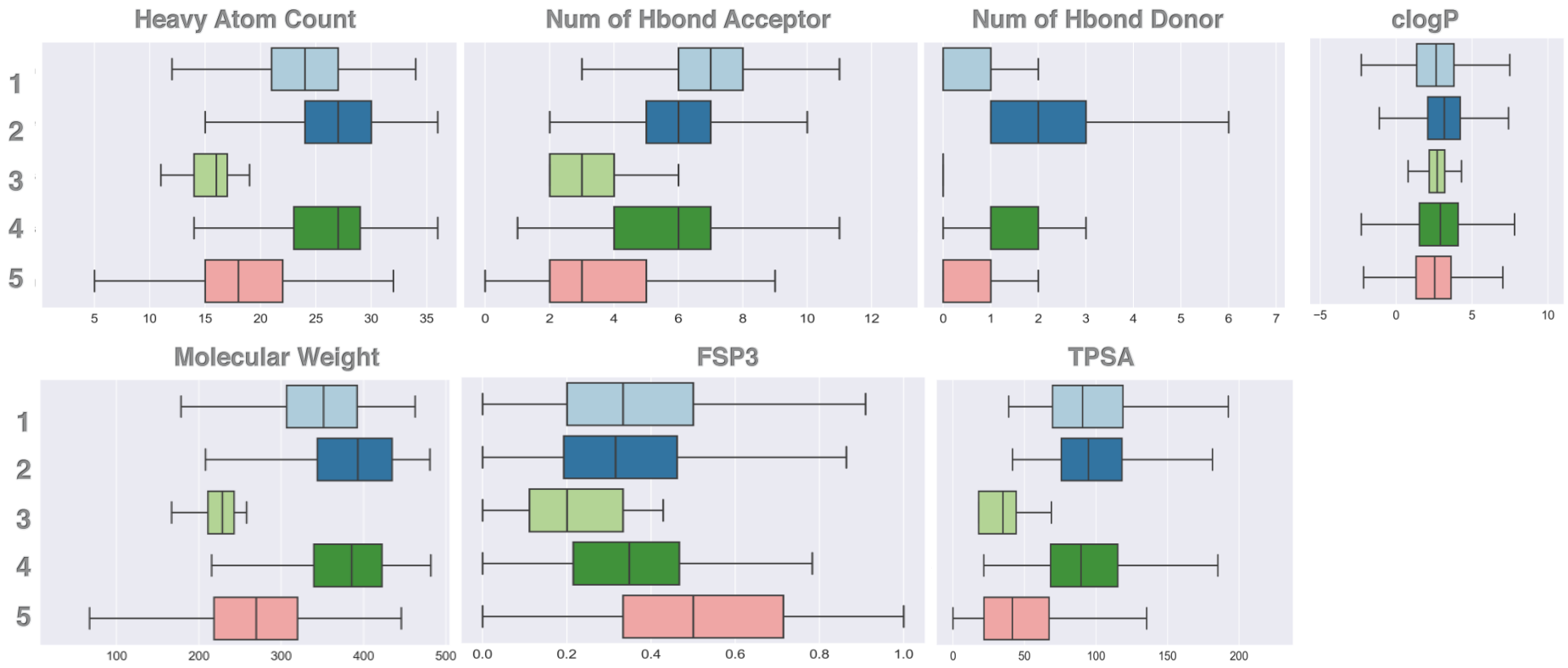




2) 1-2-4-Triazole formation from nitrile and hydrazide
3) Benzimidazole formation from 1-2-phenyldiamine
4) Friedel-Crafts acylation of heteroaromatics
5) Suzuki coupling

4. Preliminary Results

A glance at enumerated space: 100*100 reactants chosen by MaxMin diversity:



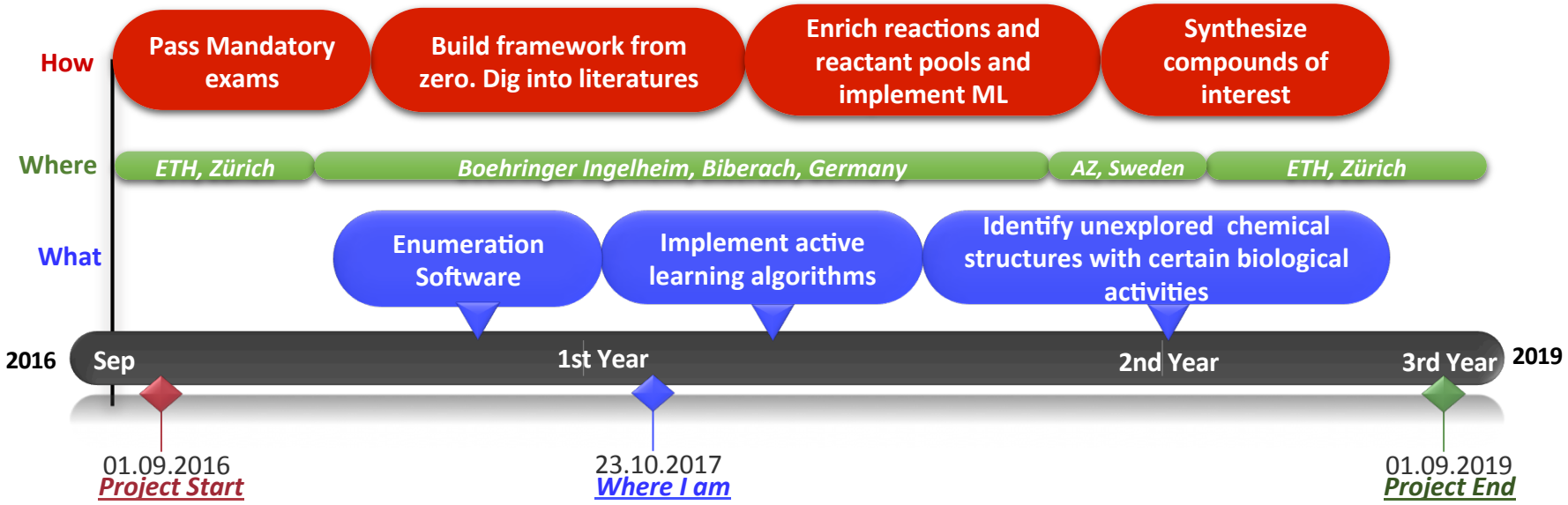
5. Conclusion and Outlook

Done:

- ***Reaction-driven enumeration software established***

Ongoing activities:

- Implementation of additional reaction-specific substructure filter
- Enrichment of reaction- and reactant-repositories
- Comparison of enumerated space with available compound collections: *e.g. DrugBank, ChEMBL*



6. Education

Research training

- BigChem weekly online courses
- 2016.09 to 2016.12, at ETH, Mandatory exams:
  - ✓ Medizinische Chemie I,
  - ✓ Therapeutic Proteins,
  - ✓ Biotransformation of Drugs and Xenobiotics.
- 2016.09 to 2016.12, at ETH, other course:
  - ✓ Computer-Assisted Drug Design
- Language course:

  two German lessons per week, since 01.2017 at Boehringer Ingelheim.
- Programming:

  two Python courses at Boehringer Ingelheim

- ETH, Lab open day, April 2017

Networking and transfer of knowledge

- GCC meeting in Mainz 05.11.-07.11.2017
- A visit & Poster Presentation at the Institute of Molecular Biology in Mainz 13.11.2017
- Poster Presentation & Discussion with other PhD & Post-Doc students on BI Science Day in Biberach (22.11.2017)

Acknowledgement

Many thanks to my supervisor, Dr. Alexander Weber from Boehringer Ingelheim;
Prof. Dr. Gisbert Schneider and Dr. Jan A. Hiss from ETH, Zürich;
Dr. Daniel Seeliger, Dr. Oliver Hucke, Dr. Bernd Wellenzohn,
Dr. Peter Haebel, Dr. Jan Kriegl from Boehringer Ingelheim.

Boehringer Ingelheim