Exploring Chemical Space

Two (still immature) ways of traversing in the molecule space

Nico Domschke

Bioinformatik Uni Leipzig

February 17, 2023

Chemical space in numbers

- New innovative molecules are needed for drug discovery.
- Only 12267 approved drugs[1]
- How big is the chemical space? 10^{60} drug like molecules[4]
- Huge untapped potential

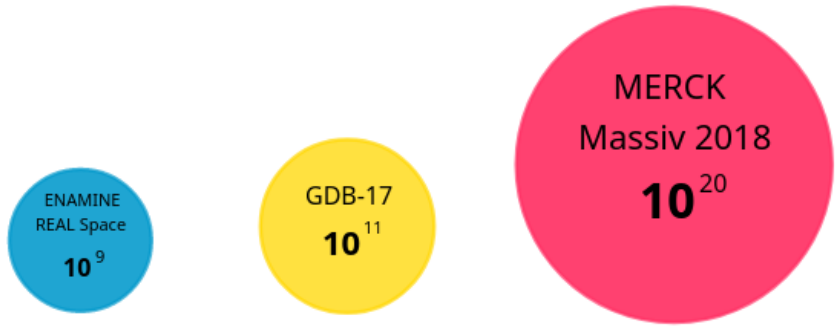


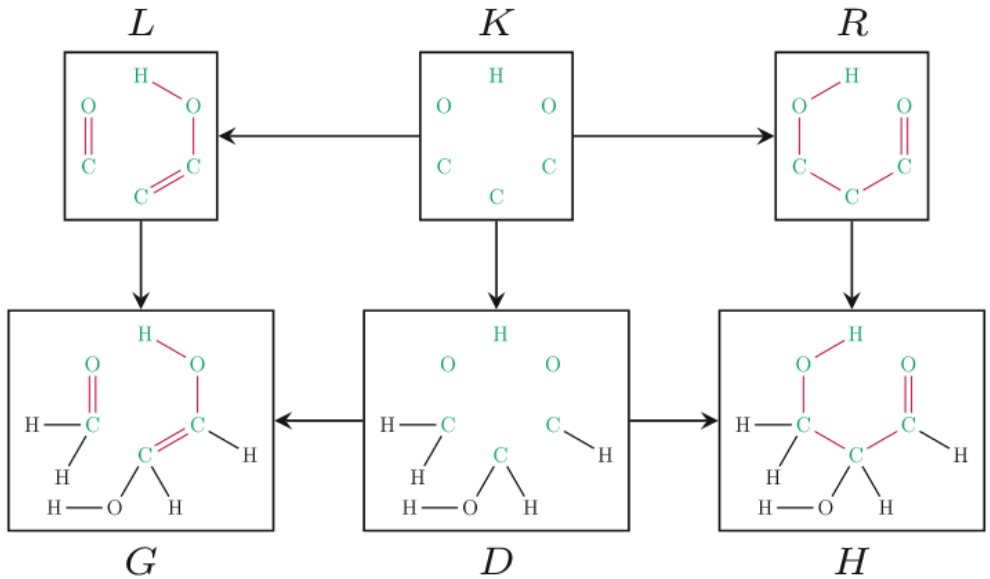
Figure: Chemical Space in Numbers.[2]

[1] on DrugBank 11.02.23

[2] Hoffmann, Torsten. The next level in chemical space navigation: Going far beyond enumerable compound libraries. Drug Discovery Today, 24(5):1148–1156, 2019

A rule based approach

- Using MØD as the underlying framework [1]
- Category theory embedded subgraph matching on molecule graphs
- A restriction on actual chemistry produces synthesizable molecules [5]
- Curated collection of 54 organic reactions by Hartenfeller et al.[3]

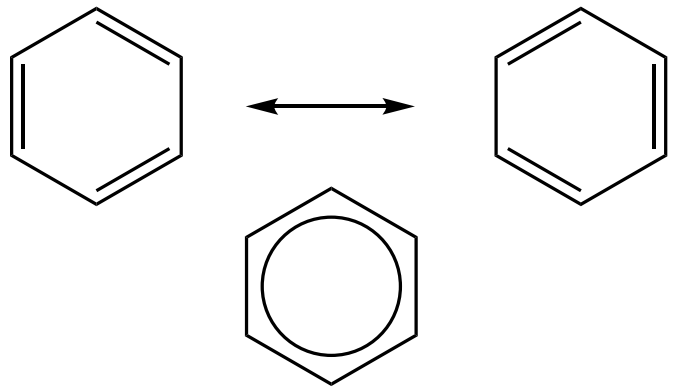


Term-Rewriting

- Chemical context needs to be explicity defined
- `ConstrainLabelAny` gives a lot of flexibility
- A liberal definition should be the sweet spot
- Much Needed: Negative Lists and Subgraph Matches

```
(...)
node [ id 1 label "_U" ]
edge [ source u target v label "_A"]
constrainLabelAny [
    label "Meso(_A,_B,_C,_D,_E,_F)"
    labels [label "Meso(-,=,-,=,-,=)" label "Meso(=,-,=,-,=,-)"]
    ]
constrainLabelAny [
    label "Halo(_U)"
    labels [label "Halo(Cl)" label "Halo(Br)" label "Halo(I)"]
    ]
]
```

Term-Rewriting

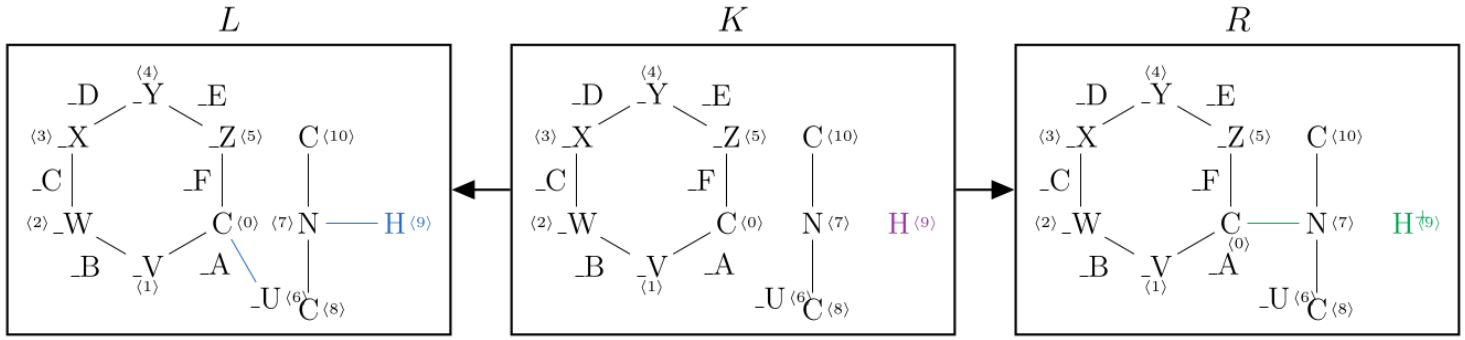


```
1   (...)
2   node [ id 1 label "_U" ]
3   edge [ source u target v label "_A"]
4   constrainLabelAny [
5       label "Meso(_A,_B,_C,_D,_E,_F)"
6       labels [label "Meso(-,=,-,=,-,=)" label "Meso(=,-,=,-,=,-)"]
7       ]
8   constrainLabelAny [
9       label "Halo(_U)"
10      labels [label "Halo(Cl)" label "Halo(Br)" label "Halo(I)"]
11      ]
12  ]
```

Term-Rewriting

- Chemical context needs to be explicity defined
- `ConstrainLabelAny` gives a lot of flexibility
- A liberal definition should be the sweet spot
- Much Needed: Negative Lists and Subgraph-Terms

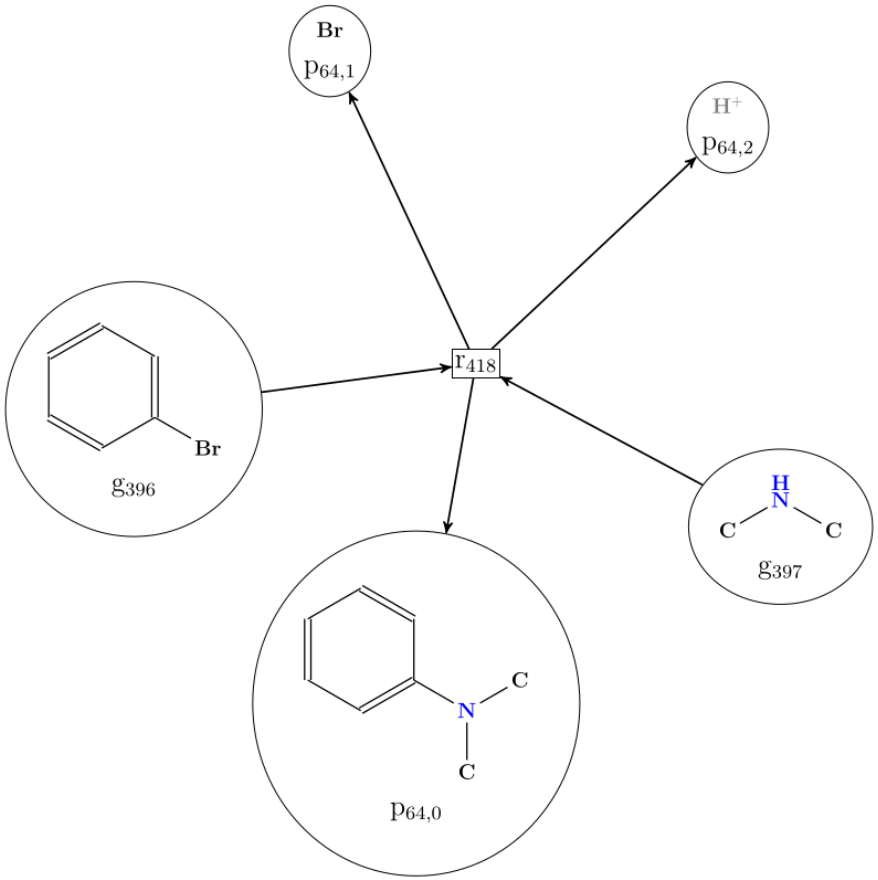
65.1.1 Buchwald-Hartwig-amine



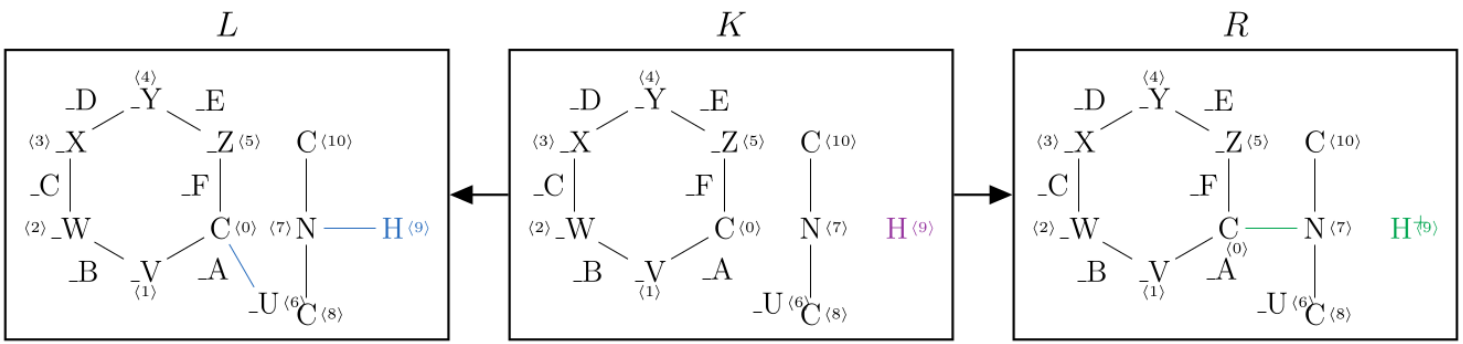
Files: out/819_r_418_10300010_{L, K, R}

`Permut(_V,_W,_X,_Y,_Z)` ∈ {'`Permut(C,C,C,C,C)`', '`Permut(N,C,C,C,C)`', '`Permut(C,C,C,C,N)`'
`Meso(_A,_B,_C,_D,_E,_F)` ∈ {'`Meso(-,=,-,=,-,=)`', '`Meso(=,-,=,-,=,-)`'}
`Halo(_U)` ∈ {'`Halo(Cl)`', '`Halo(Br)`', '`Halo(I)`'}

Term-Rewriting



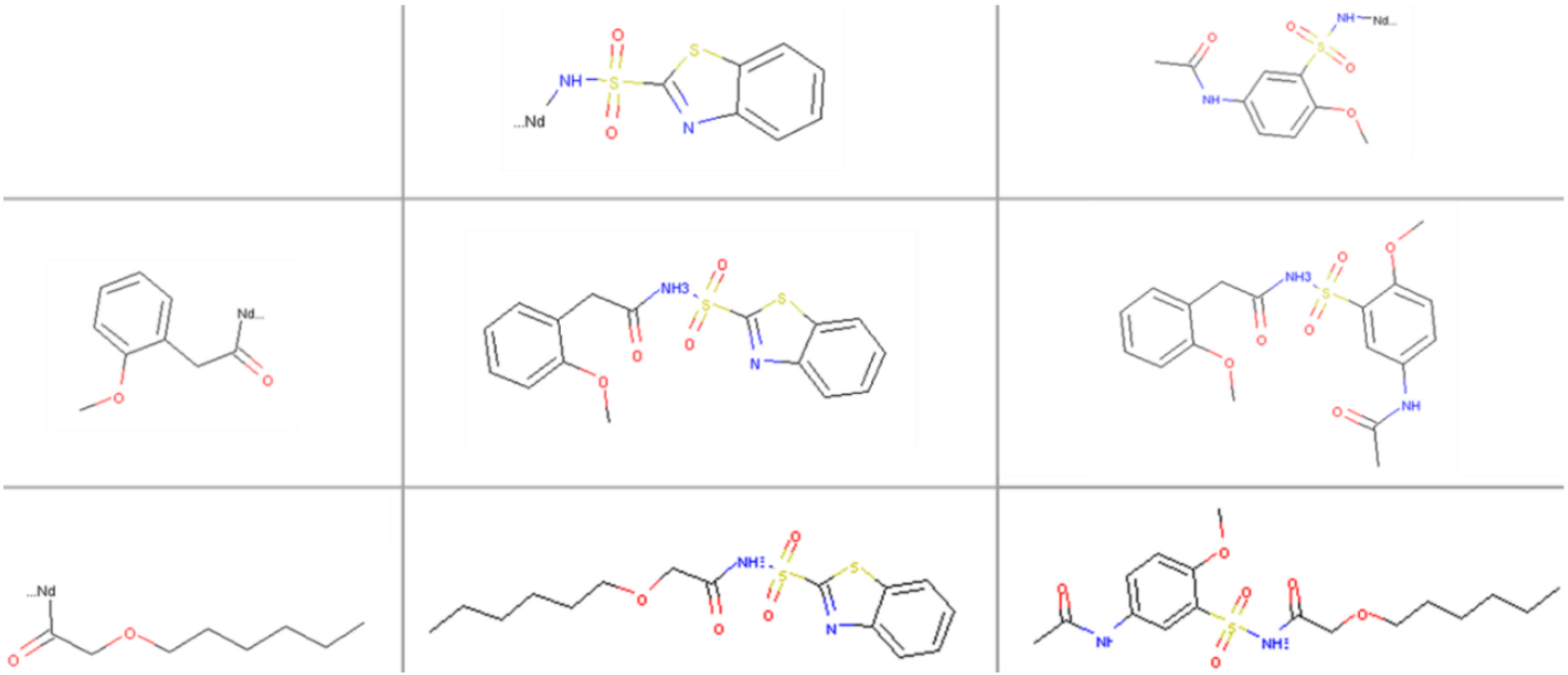
65.1.1 Buchwald-Hartwig-amine



Files: out/819_r_418_10300010_{L, K, R}

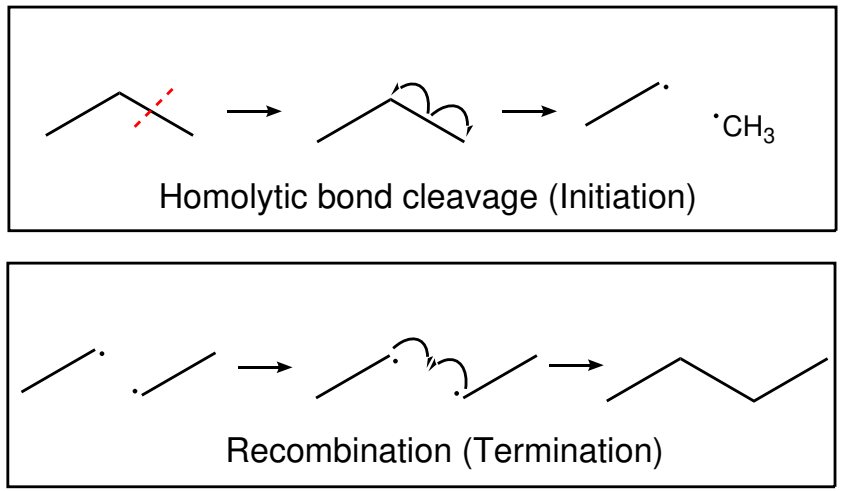
A rule based approach

- A first test is *planned* using the ENAMINE synthesize-on-demand library



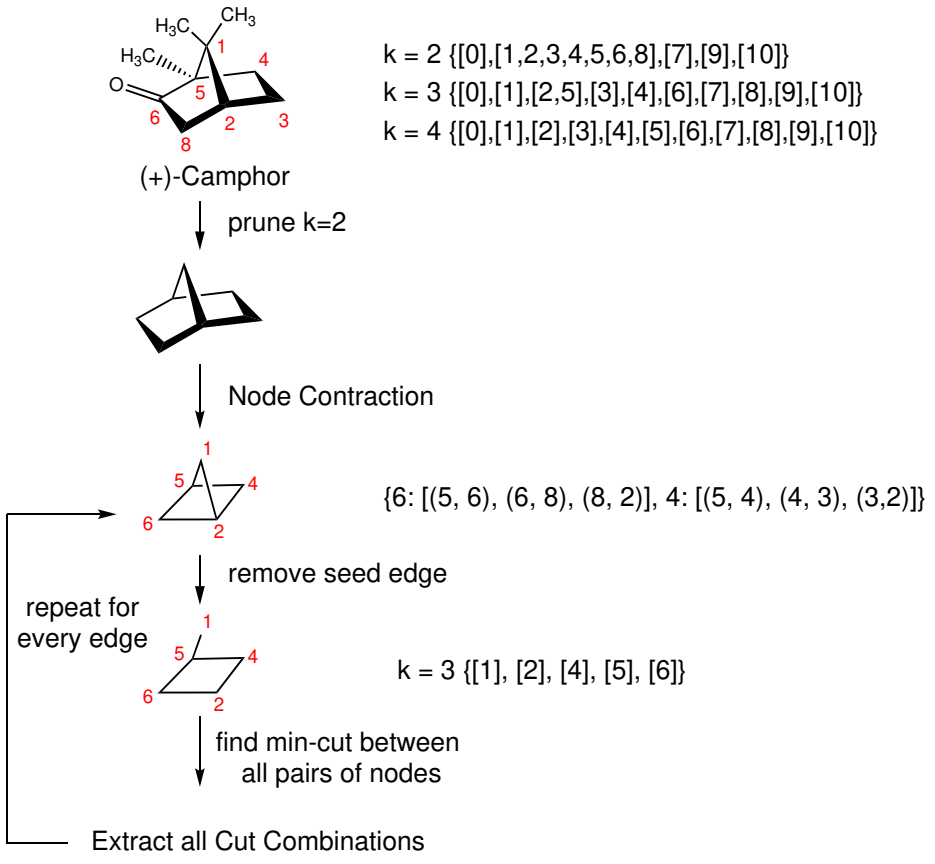
Crossover mutations

- Two parent molecules each cut into two fragments
- Corresponds to the formation of radicals in chemistry
- No restriction by relying on reactions
- exhaustive graph bipartitioning



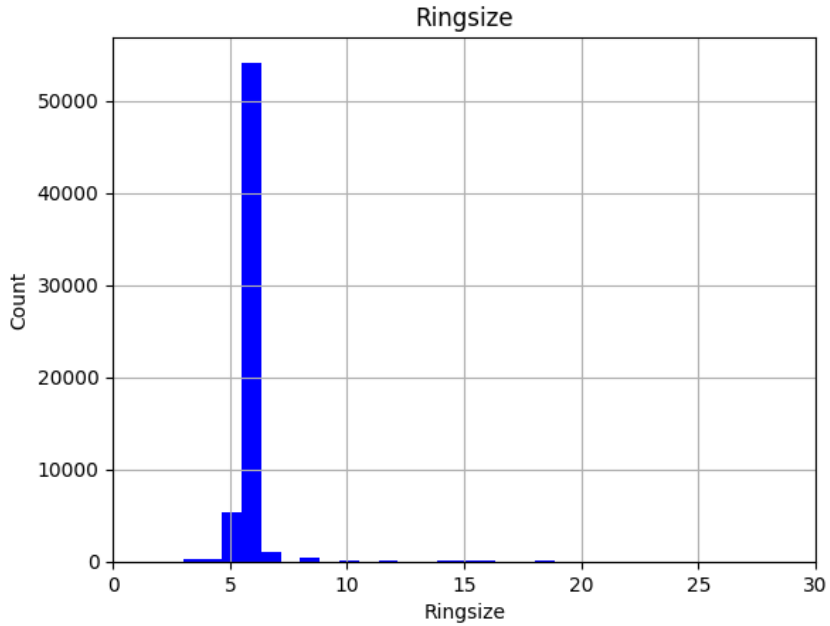
Generating all possible cuts

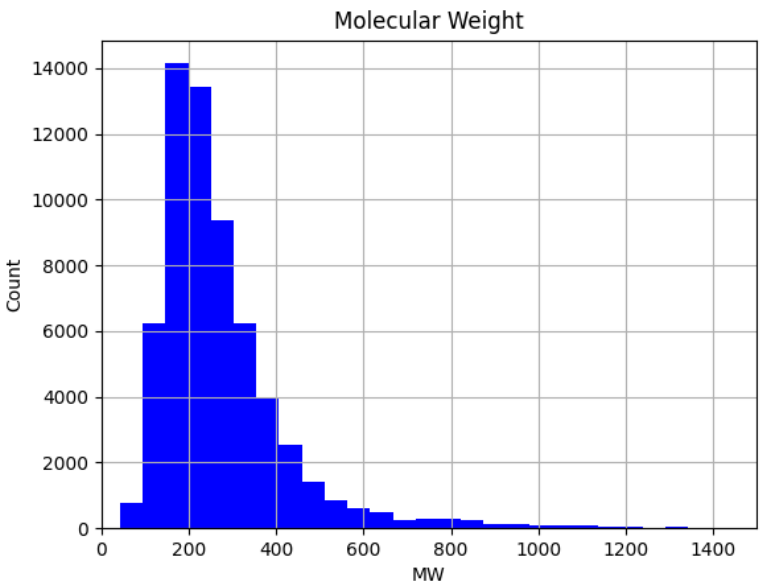
- k-connected-components for k = 2,3,4 [6]
- remove sidechains (k=2)
- contract nodes
- remove random seed-edge of graph
- check k-connected-components after removal
- get minimum-cut of every pair of atoms in these connected components

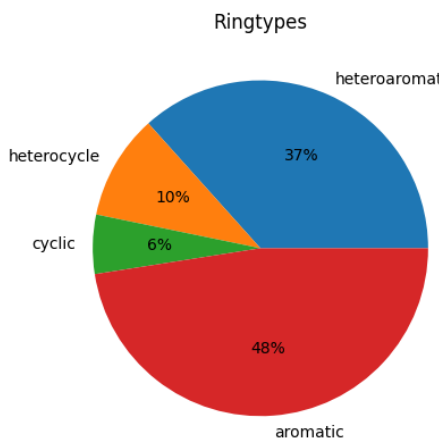


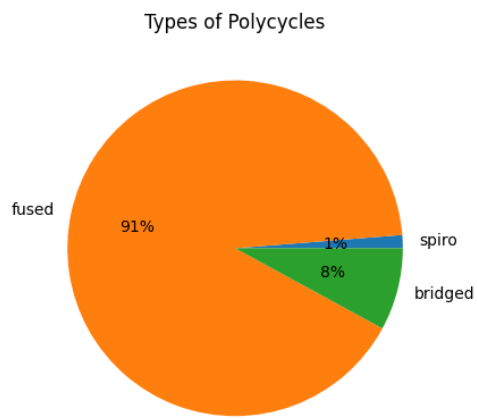
A superficial look at Sigma Aldrich catalog

- Atleast 1 Ring,
  1 Carbon-Atom,
  no disconnected
  components
- 117k $\rightarrow$ 62k compounds
- High diversity



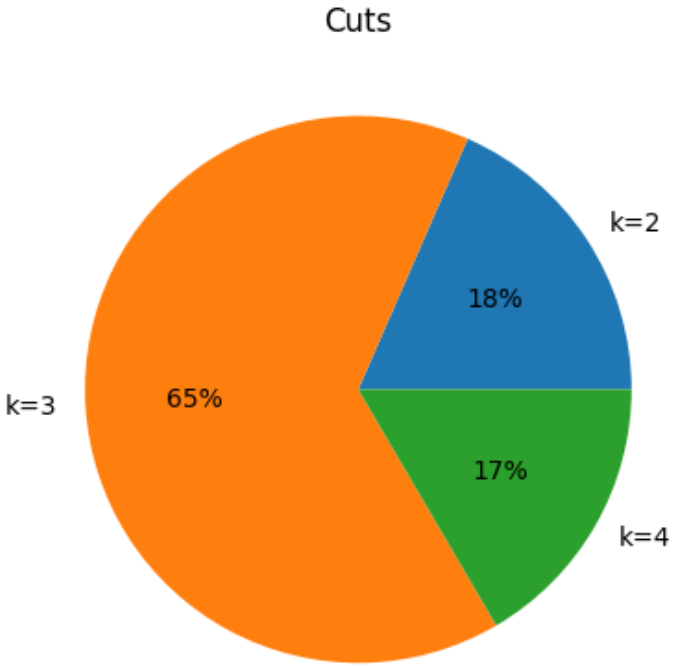


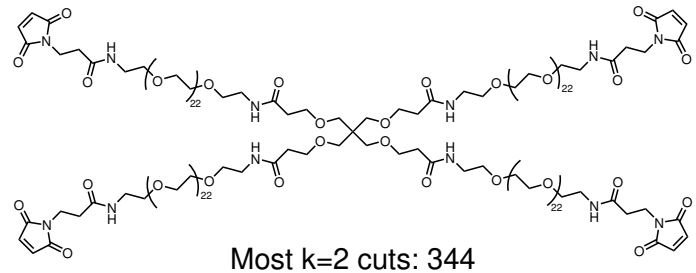




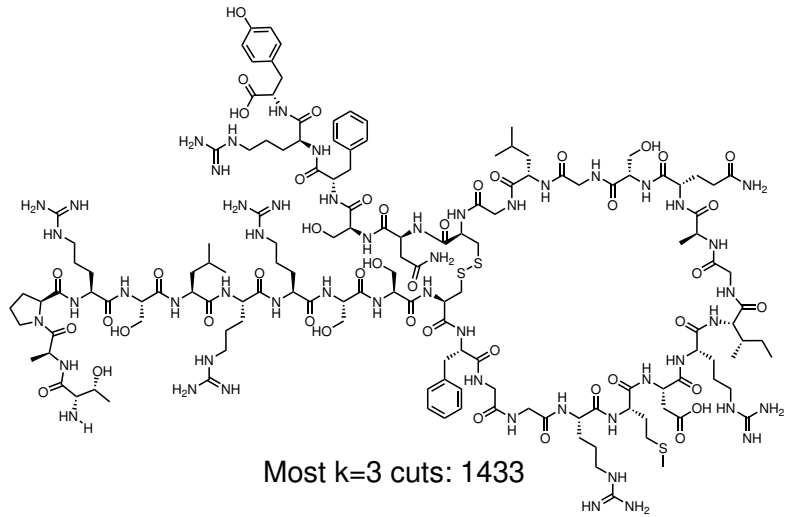
A superficial look at Sigma Aldrich catalog

- Found 3.1 M cut combinations

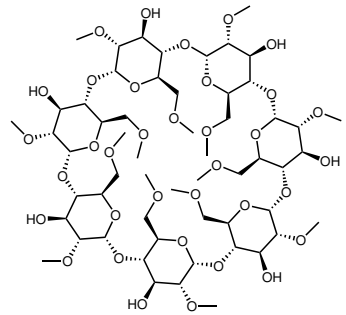




Most k=2 cuts: 344



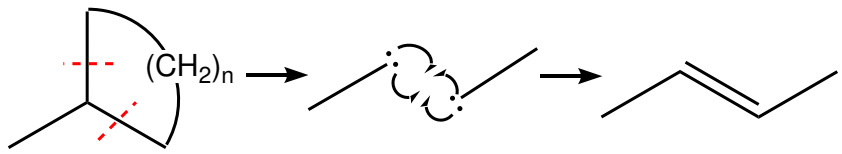
Most k=3 cuts: 1433



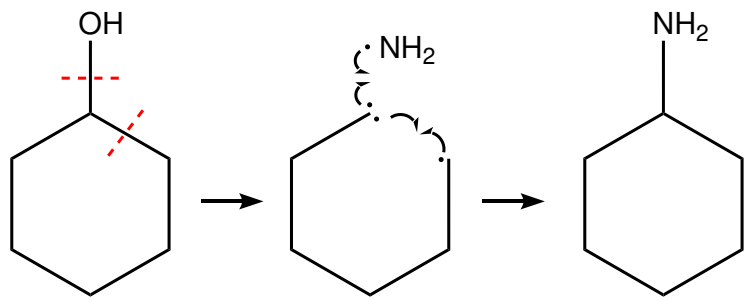
Most k=4 cuts: 630

Open Questions of the fragment recombination

- Allow Termination of multiple radicals at the same atoms?
- How to proceed with fragments having a missmatched number of radicals?
- Allow Recombination to previous bond pattern?
- Matching of single-node-fragments?
- Essentially point mutations, takes a lot of space ($\sim$ 32%) in the cut space



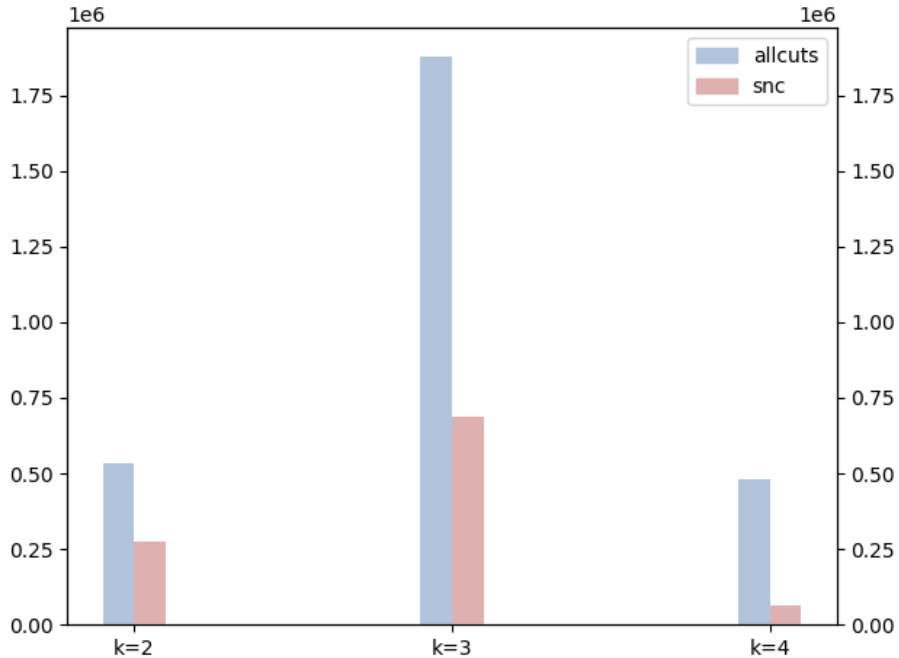
Changing of bondorder



Recombination into "old" bonds

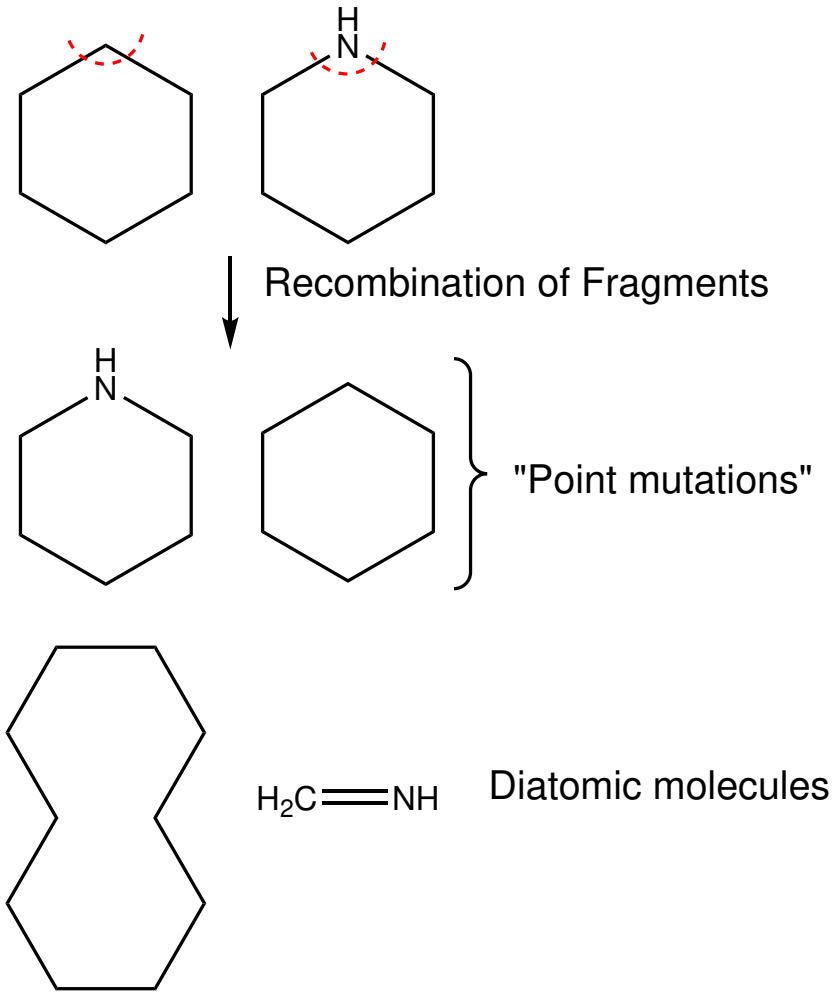
Open Questions of the fragment recombination

- Allow Termination of multiple radicals at the same atoms?
- Matching of single-node-fragments?
- How to proceed with fragments having a missmatched number of radicals?
- Allow Recombination to previous bond pattern?
- A lot of point mutations, taking lot of space ($\sim$ 32%) in the cut space



Open Questions of the fragment recombination

- Allow Termination of multiple radicals at the same atoms?
- Matching of single-node-fragments?
- How to proceed with fragments having a missmatched number of radicals?
- Allow Recombination to previous bond patterns?
- Essentially point mutations, takes a lot of space ($\sim$ 32%) in the cut space



How to ensure chemical meaningfulness?

- Possible Restrictions and filters:

- Remove unstable functional groups (violating Erlenmeyer-Rule etc.)
- Account for ring strain (violating Bredts-Rule, etc.)
- Most lead compounds are in range $100 < \text{MW} < 350$ Da
- Empiric MedChem rules: Lipinski Ro5, Lily Medchem [2]

Thanks to my Supervisors Peter and xtof

Special Thanks to Bruno!

Thank You!

Nico Domschke
Bioinformatik Uni Leipzig
dnico@bioinf.uni-leipzig.de
www.bioinf.uni-leipzig.de

Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler.

A Software Package for Chemically Inspired Graph Transformation.
In Rachid Echahed and Mark Minas, editors, *Graph Transformation*, volume 9761, pages 73–88.
Springer International Publishing, Cham, 2016.
Series Title: Lecture Notes in Computer Science.

Robert F. Bruns and Ian A. Watson.

Rules for Identifying Potentially Reactive or Promiscuous Compounds.
Journal of Medicinal Chemistry, 55(22):9763–9772, November 2012.

Markus Hartenfeller, Martin Eberle, Peter Meier, Cristina Nieto-Oberhuber, Karl-Heinz Altmann, Gisbert Schneider, Edgar Jacoby, and Steffen Renner.

A Collection of Robust Organic Synthesis Reactions for *In Silico* Molecule Design.
Journal of Chemical Information and Modeling, 51(12):3093–3098, December 2011.

Asher Mullard.

The drug-maker's guide to the galaxy.
Nature, 549(7673):445–447, September 2017.

Lars Ruddigkeit, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond.

Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17.
Journal of Chemical Information and Modeling, 52(11):2864–2875, November 2012.

Tianhao Wang, Yong Zhang, Francis Y. L. Chin, Hing-Fung Ting, Yung H. Tsin, and Sheung-Hung Poon.

A Simple Algorithm for Finding All k-Edge-Connected Components.
PLOS ONE, 10(9):e0136264, September 2015.