Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Boehringer Ingelheim

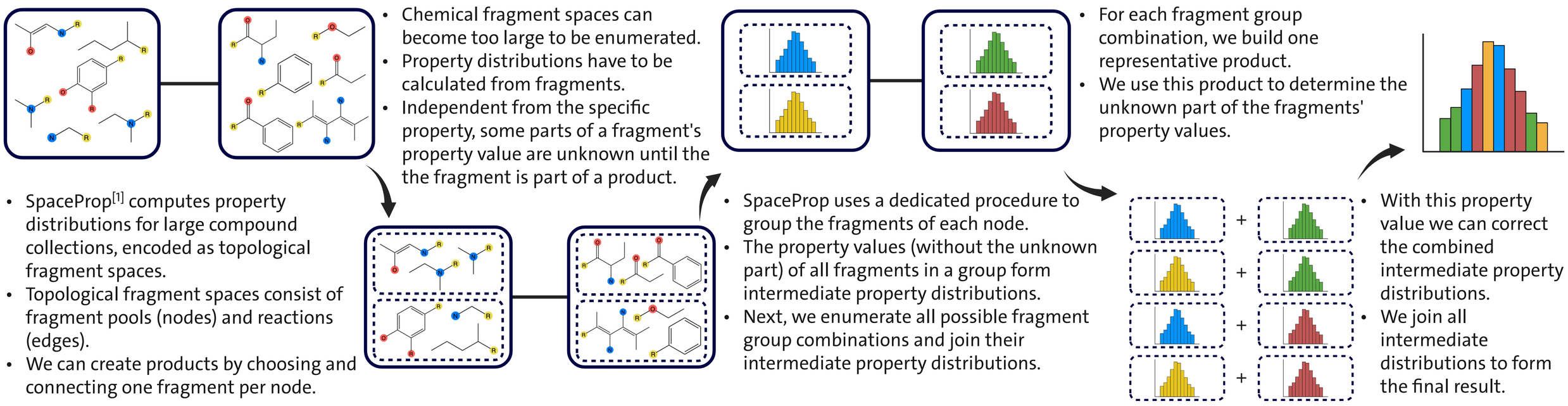
MIN FACULTY
CENTER FOR BIOINFORMATICS

Calculating More Property Distributions of Chemical Fragment Spaces

JUSTIN LÜBBERS | UTA LESSEL | LOUIS BELLMANN | MATTHIAS RAREY

The recent development of large chemical fragment spaces has significantly expanded the range of synthetically accessible compound collections. With sizes ranging between 10^8 and 10^{26} products, these catalogs often cannot be enumerated due to their immense size, necessitating new algorithms for cheminformatics tasks. The SpaceProp algorithm was developed to calculate exact physicochemical property distributions of chemical fragment spaces. The original work included the properties molecular weight, the number of hydrogen bond acceptors and donors as well as the clogP. We present recent extensions of SpaceProp that allow for the consideration of additional molecular properties, namely the topological polar surface area, the number of rotatable bonds and the occurrence of structural patterns in the form of SMARTS expressions. Our expanded algorithm offers greater insight into the composition of non-enumerable compound collections, aiding in the analysis of these vast libraries and paving the way for the construction of optimized chemical fragment spaces.

SpaceProp Algorithm



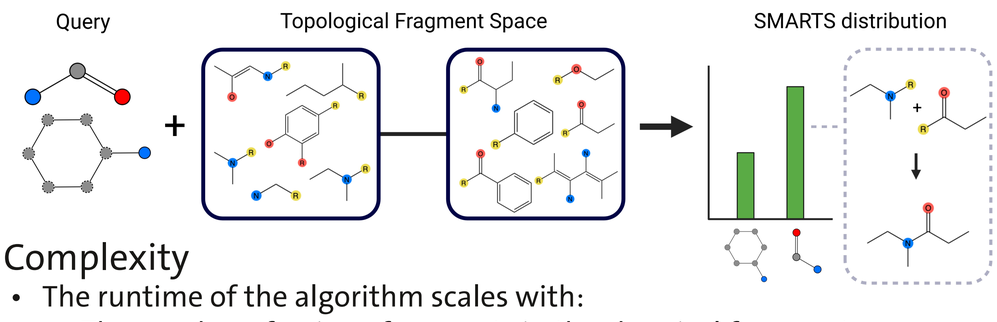
Methods

New SpaceProp Features

- We added three additional properties to SpaceProp:
  - The topological polar surface area (TPSA)[2]
  - The number of rotatable bonds
  - The occurrence of molecular structures based on SMARTS
- We enabled the generation of example molecules for each entry in the property distributions.

Motivation

- The number of rotatable bonds and the TPSA are commonly used molecular descriptors that influence oral bioavailability[3].
- SMARTS distributions offer insight into the structural composition
  - Many drug design projects are associated with specific structural motifs of interest
  - How many products in a chemical fragment space contain how many of these structures?
  - Which specific structures are present and how often do they occur?
- Example molecules increase explainability of the results.



Complexity

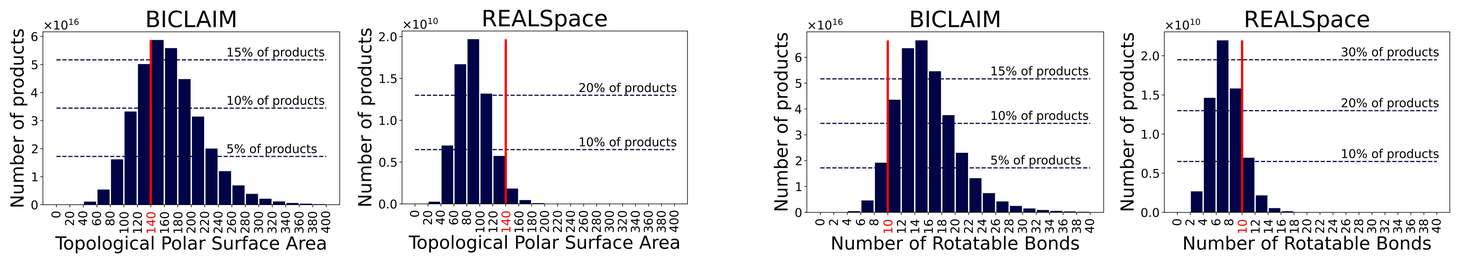
- The runtime of the algorithm scales with:
  - The number of unique fragments in the chemical fragment space
  - The degree to which fragments can be grouped
  - The value range of the given property
- The runtime does not scale with the number of products.
- Average runtimes range from a few minutes to several hours.

Fragment Spaces

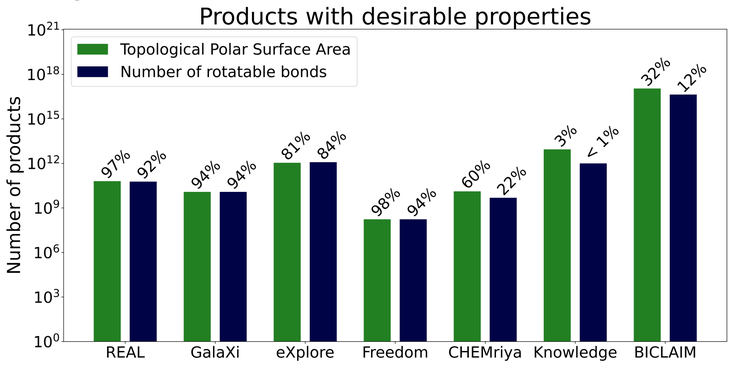
- We applied SpaceProp to the following chemical fragment spaces[4]:
  - REAL Space by Enamine Ltd. — 3.6×10^{10} products
  - GalaXi by WuXi LabNetwork — 1.2×10^{10} products
  - eXplore by eMolecules — 7.0×10^{12} products
  - Freedom Space by Chemspace — 1.8×10^{8} products
  - CHEMriya by OTAVA — 1.2×10^{10} products
  - KnowledgeSpace — 2.9×10^{14} products
  - BICLAIM by Boehringer Ingelheim[5] — 3.4×10^{17} products

Results

Example Histograms



Analysis



- Veber's Rule[3] for high oral bioavailability:
  - 10 or fewer rotatable bonds
  - A TPSA of 140Å or less
- Chemical fragment spaces can contain numerous compounds with undesired properties.
- SpaceProp reveals such imbalances.
- Example molecules show which reactions and fragment combinations are responsible.
- This opens up opportunities for improvement.

SMARTS Distribution Application Example

- Question: Is any of the chemical fragment spaces suitable for covalent drug discovery?
- Query: 135 SMARTS patterns of electrophilic warheads, taken from the WHdb[6]

- Although neither of the chemical fragment spaces was designed for covalent drug discovery, the results indicate that all contain a significant amount of potential compounds for this purpose.
- SpaceProp can be used to further investigate which query structures occur most frequently, which structures do not occur at all, and which reactions produce the structure matches.

Conclusion

- The properties of chemical fragment spaces are difficult to anticipate without detailed analysis.
- Our new software tool SpaceProp2 can:
  - Analyze and compare the contents of these otherwise intransparent compound libraries
  - Reveal imbalances in the composition of chemical fragment spaces and identify their causes
  - Aid in the optimization of non-enumerable, combinatorial compound collections with regard to general physico-chemical properties as well as target specific requirements

Justin Lübbers* - justin.luebbers@uni-hamburg.de
Dr. Uta Lessel** - uta.lessel@boehringer-ingelheim.com
Prof. Matthias Rarey* - matthias.rarey@uni-hamburg.de
* Center for Bioinformatics, Bundesstr. 43, 20146 Hamburg, Germany
** Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach an der Riss, Germany

NAOMI ChemBio Suite
https://uhh.de/naomi

Chemical fragment spaces provided by BioSolveIT GmbH
https://www.biosolveit.de/

[1] Bellmann *et al.*, *J. Chem. Inf. Model.*, 2022. 10.1021/acs.jcim.2c00334
[2] Ertl *et al.*, *J. Med. Inf.*, 2000. 10.1021/jm000942e
[3] Veber *et al.*, *J. Med. Chem.* 2002. 10.1021/jm020017n
[4] *BioSolveIT Chemical Spaces.* https://www.biosolveit.de/products/infinisee/ (accessed 11/06/2023)
[5] Lessel *et al.*, *J. Chem. Inf. Model.*, 2009. 10.1021/ci800272a (updated version from 2020)
[6] Péczka *et al.*, *Expert Opin Drug Discov.*, 2022. 10.1080/17460441.2022.2034783
Figures created with BioRender.com