ChemAxon

Dora Barna
Tim Parrott

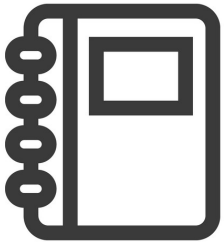
How to reach the hidden world of chemistry in patents

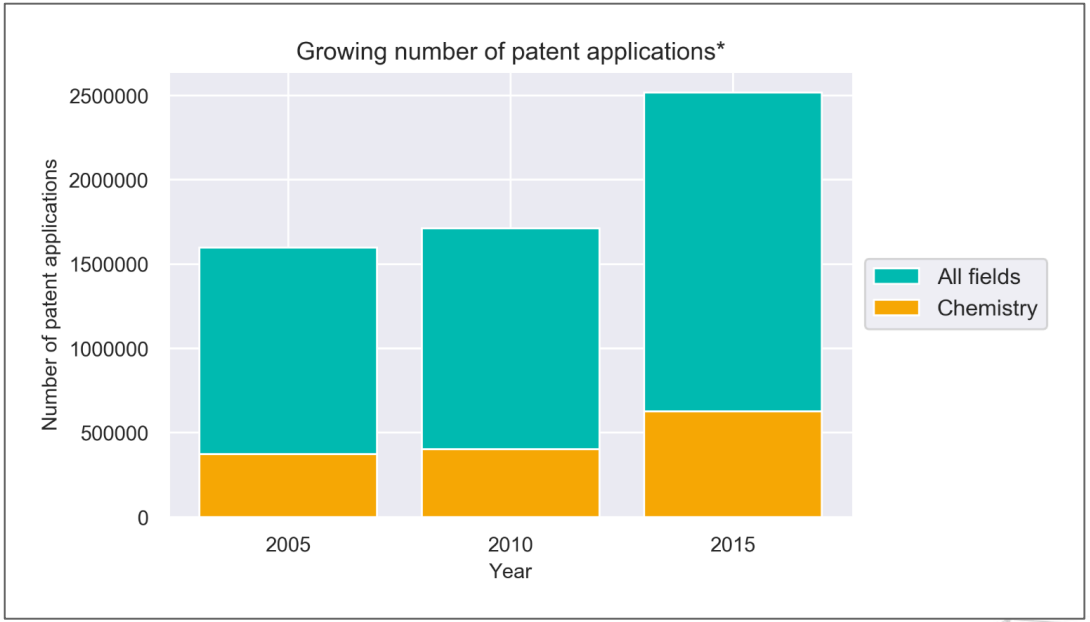
ChemAxon

Dora Barna
Tim Parrott

How to reach the hidden world of chemistry in documents

THE PROBLEM



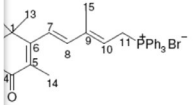


*World Intellectual Property Indicators 2017 Patents,
https://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2017-chapter2.pdf

ChemAxon

4–Oxo-β-ionylidene-ethyl-triphenylphosphonium bromide (38):

To a solution of PPh_3HBr (10.3 g, 28.6 mmol) in anhydrous methanol (100 mL) was added **36** (6.7 g, 28.6 mmol) in methanol (100 mL). After stirring for 75 h at room temp., the solvent was evaporated and **38** (10.57 g, 18.9 mmol, 66 %, lit.[30c] yield: 84 %) was obtained as a yellow solid. (m.p. 50 °C). The NMR spectra correspond to the



38

M⁺-Br].

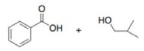
cm^{-1} (w), 1655 (s, C=O), 1588 (w),1480 (s), 1107 (s), 756

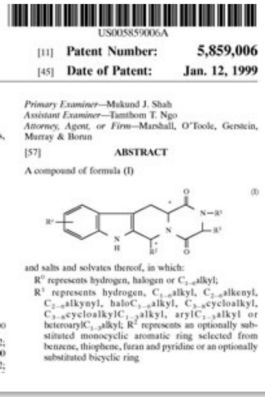
List of Schemes

1 Background

Our research project requires a target ester, whose starting material requires 15 synthetic steps. We decided to optimize the conditions of this Fischer esterification by preparing a closely related analogue **3**. The starting materials in this model reaction is the readily available benzoic acid (**1**) and *i*-PrOH (**2**).

f





US005859006A

United States Patent [19]

[11] **Patent Number: 5,859,006**

[45] **Date of Patent: Jan. 12, 1999**

Primary Examiner—Mukund J. Shah
Assistant Examiner—Tamthom T. Ngo
Attorney, Agent, or Firm—Marshall, O'Toole, Gerstein, Murray & Borun

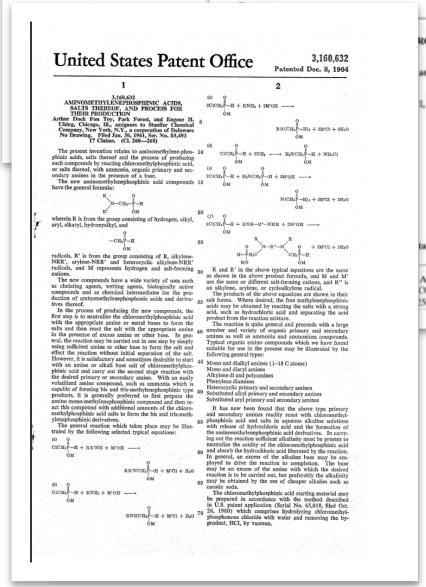
[57] **ABSTRACT**

A compound of formula (I)

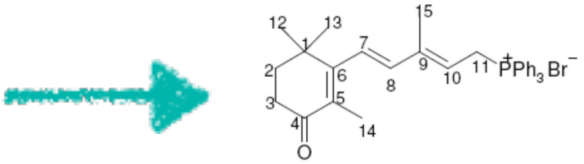
and salts and solvates thereof, in which:

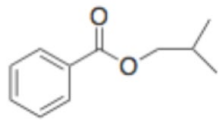
R^0 represents hydrogen, halogen or C_{1-4}alkyl;

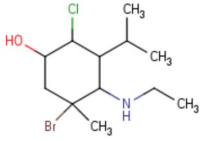
R^1 represents hydrogen, C_{1-6}alkyl, C_{2-6}alkenyl, C_{2-6}alkynyl, haloC_{1-6}alkyl, C_{1-6}cycloalkyl, C_{3-6}cycloalkylC_{1-3}alkyl, arylC_{1-3}alkyl or heteroarylC_{1-3}alkyl; R^2 represents an optionally substituted monocyclic aromatic ring selected from benzene, thiophene, furan and pyridine or an optionally substituted bicyclic ring

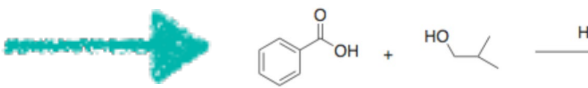


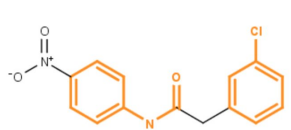
United States Patent Office 3,160,632 Patented Dec. 8, 1964

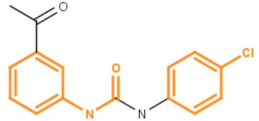


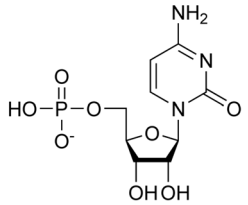












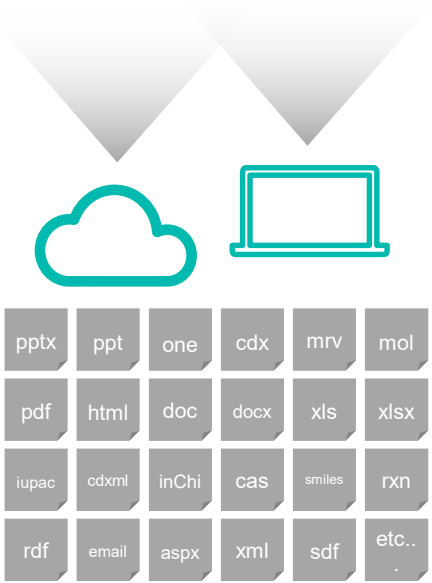
ChemAxon

UNSTRUCTURED → STRUCTURED

ChemLocator

ChemLocator

Aspirin, also known as acetylsalicylic acid (ASA), is a medication used to treat pain, fever, or inflammation.[4] Specific inflammatory conditions in which aspirin is used include Kawasaki disease, pericarditis, and rheumatic fever.[4] Aspirin given …

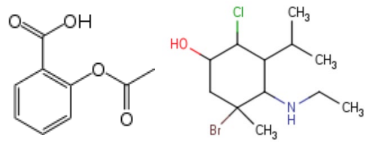


Free text indexing

ChemAxon Naming +

Semantic indexing

Chemical NER*
OCR
OSR
Metadata extraction
Tagging
Keywords
Biological NER
etc.



***NER:** Named Entity Recognition

ChemAxon

An Experiment

24,000 patent documents

- From the **USPTO** patent database
- Medicinal chemistry patents from 2015

771,000 unique molecules

- exact molecules: 740,000
- fragments: 23,000
- generic structures: 8,800

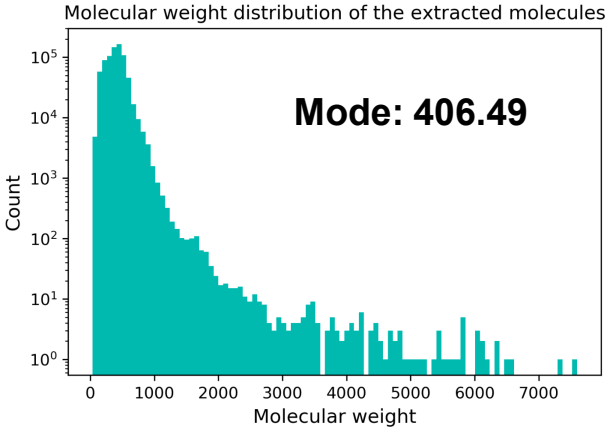
3.1 million molecules in total

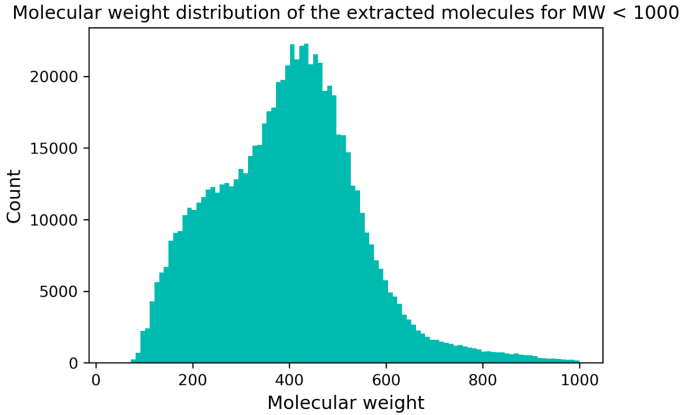
Processing

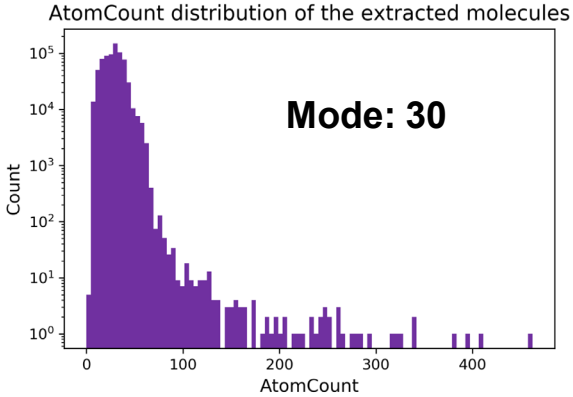
- Indexing time: 7.08 days
- Chemical database: 13.5 GB
- Elastic database: 25 GB

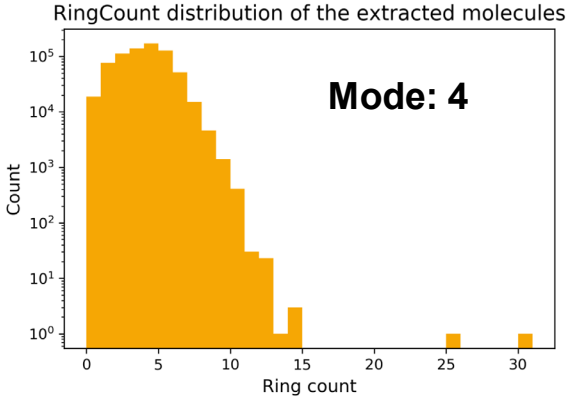
ChemAxon

Chemical space - Basic properties



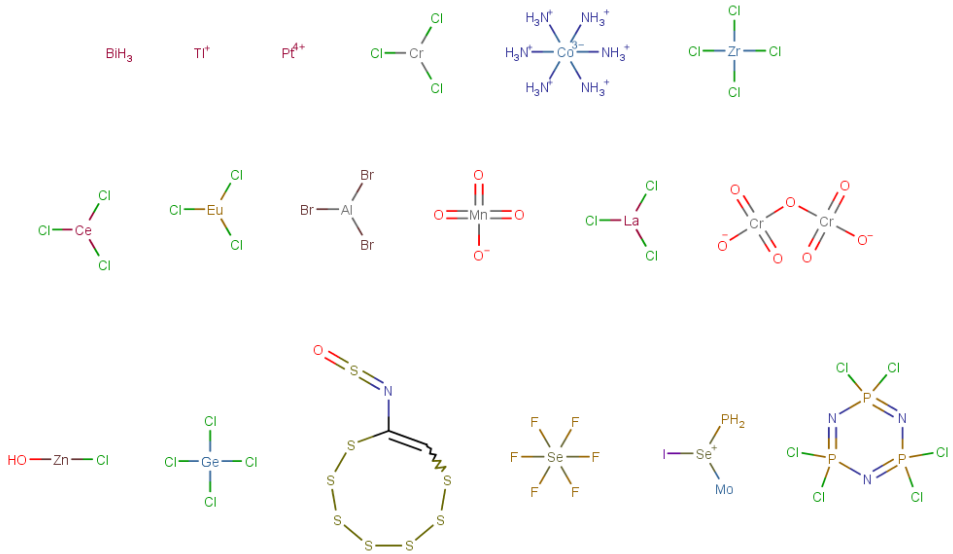


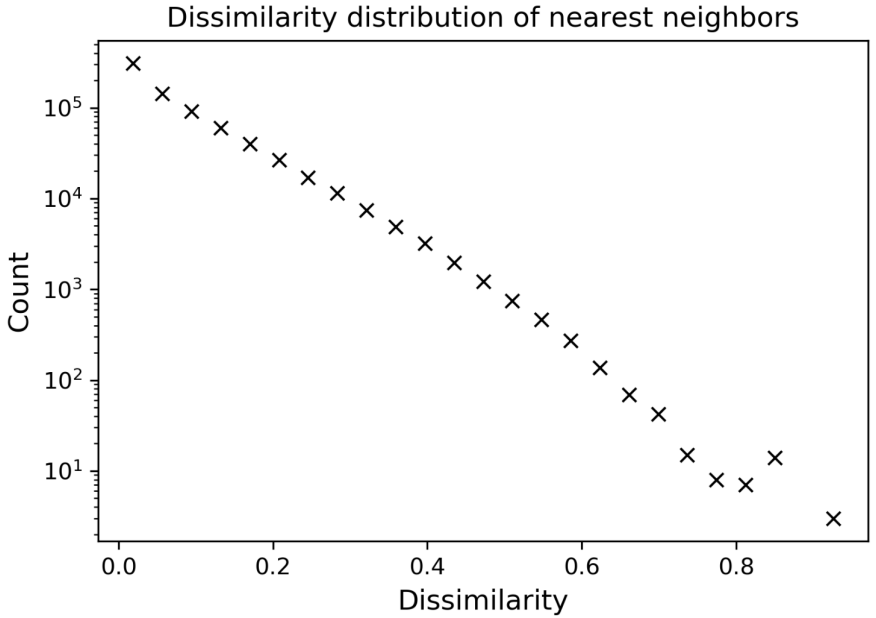




Chemical space - Structural similarity

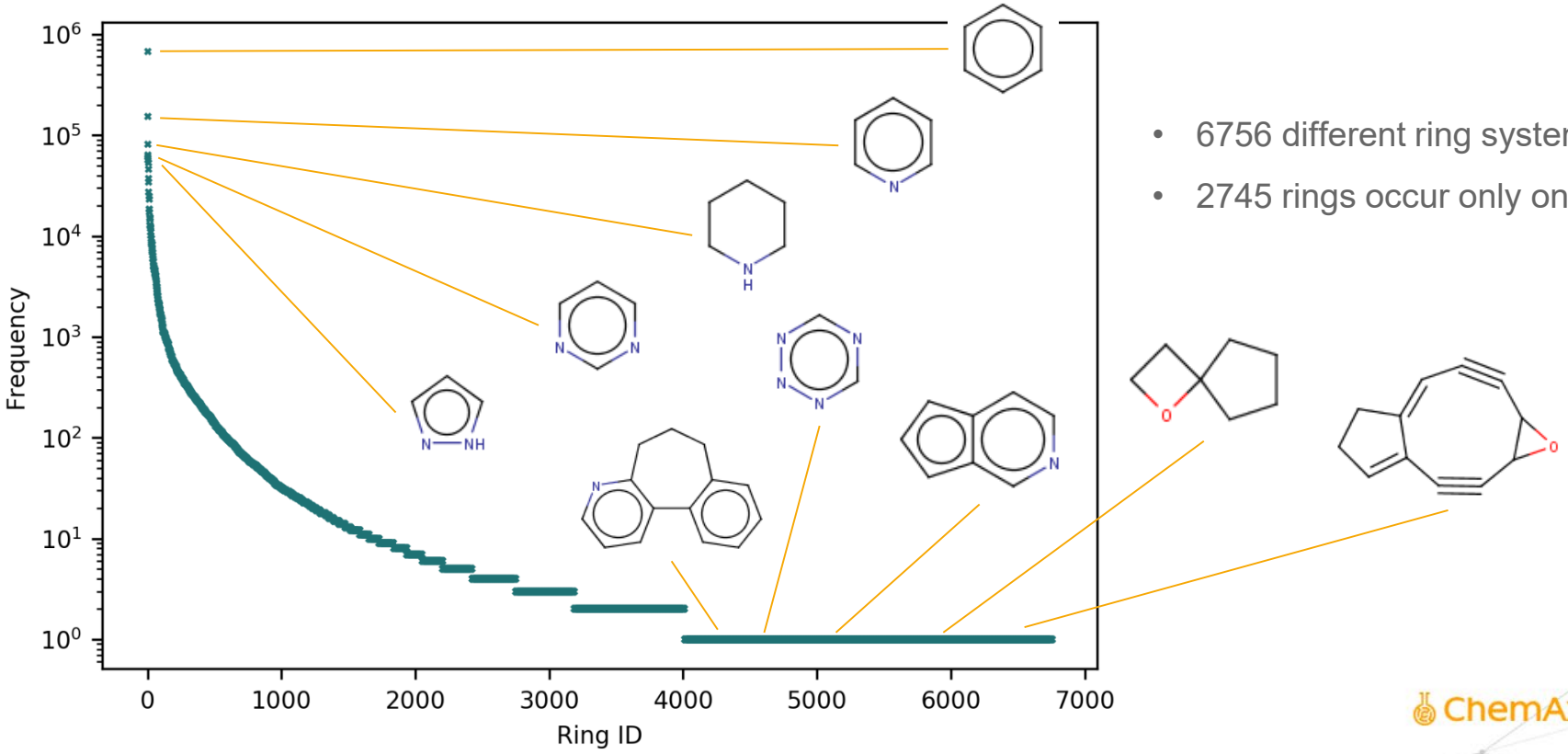
- 50% of the molecules had its most similar pair within 0.05 dissimilarity;
- Molecules that are very different from their nearest neighbors:





ChemAxon

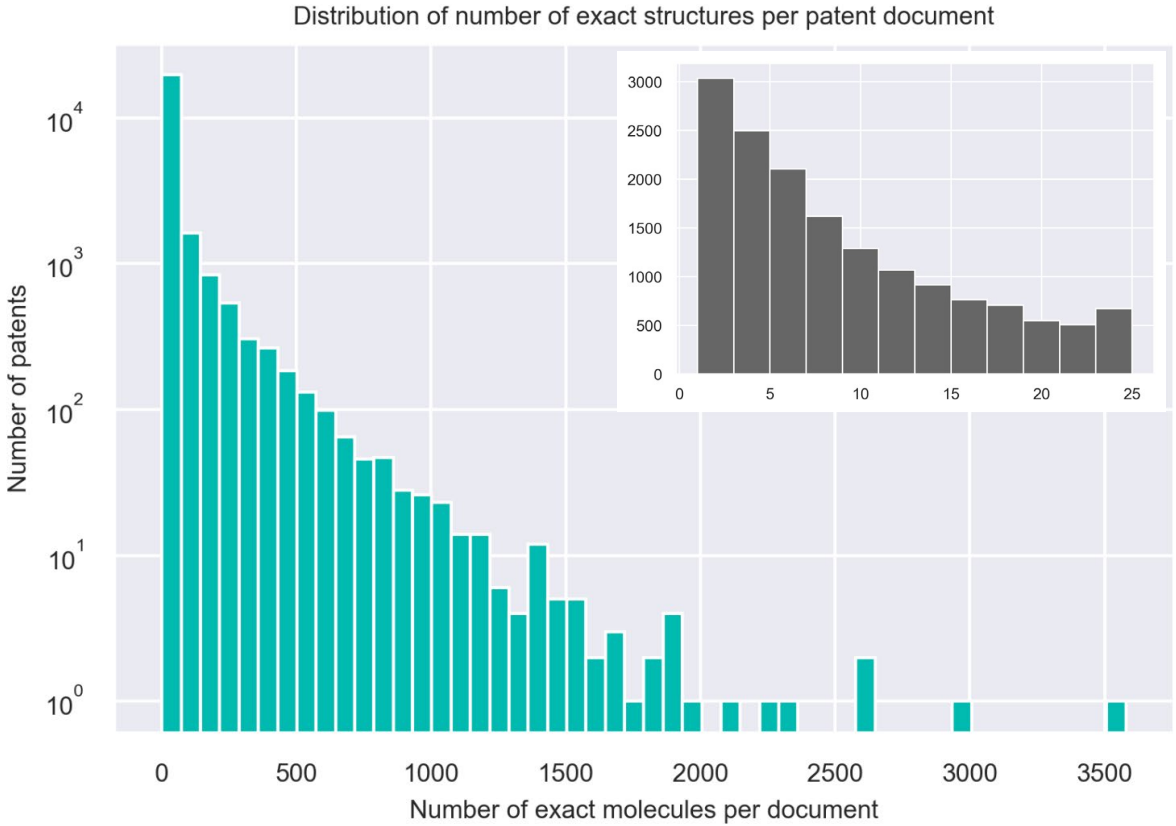
Chemical space - Typical ring systems



- 6756 different ring systems;
- 2745 rings occur only once;

ChemAxon

Chemical structures per document

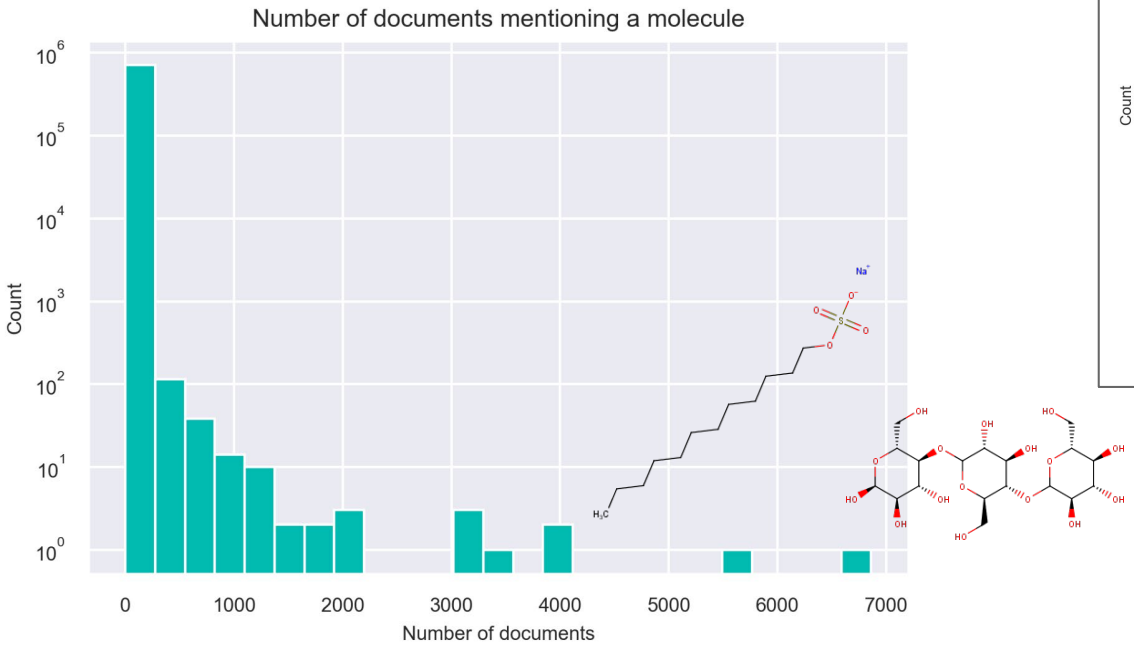


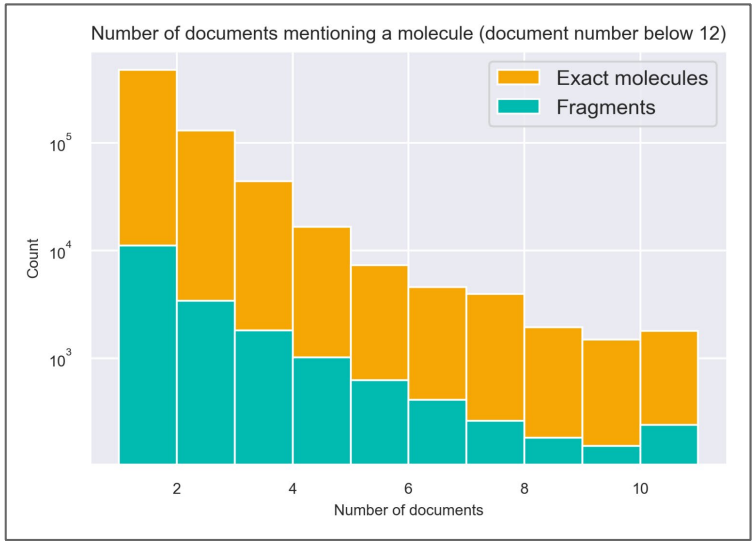
- 1548 patents with a single exact molecule;
- Topics:
  - pharmaceutical preparation
  - pharmaceutical composition
  - method of treatment

- 25 documents with more than 1500 exact molecules;
- kinase inhibitors, H4 receptor inhibitors, glucokinase activators...

ChemAxon

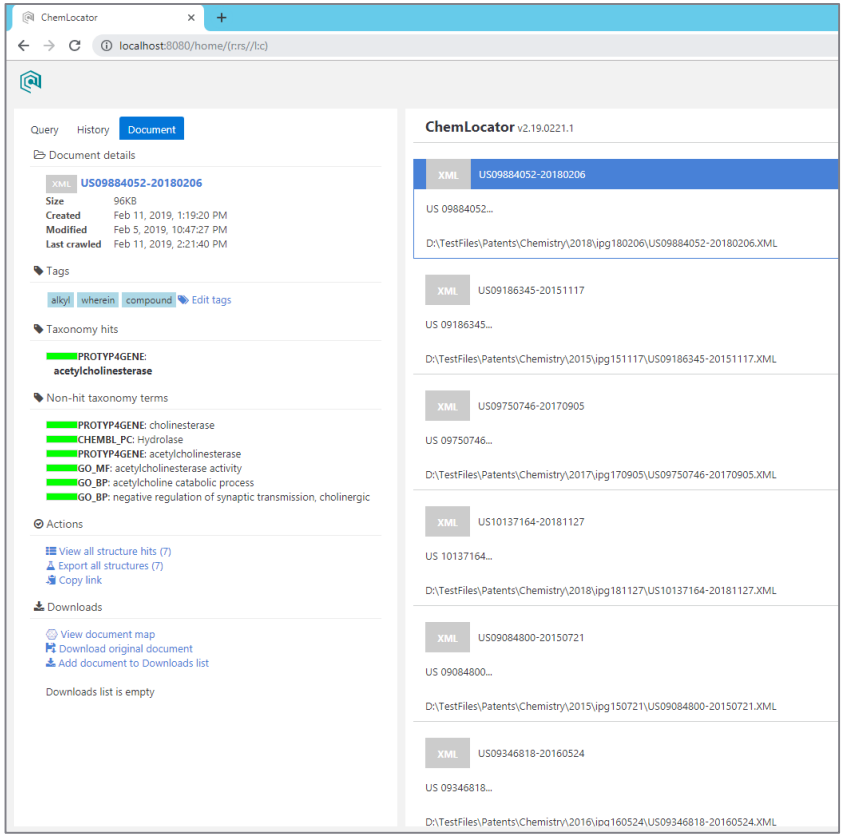
Frequency of chemical structures

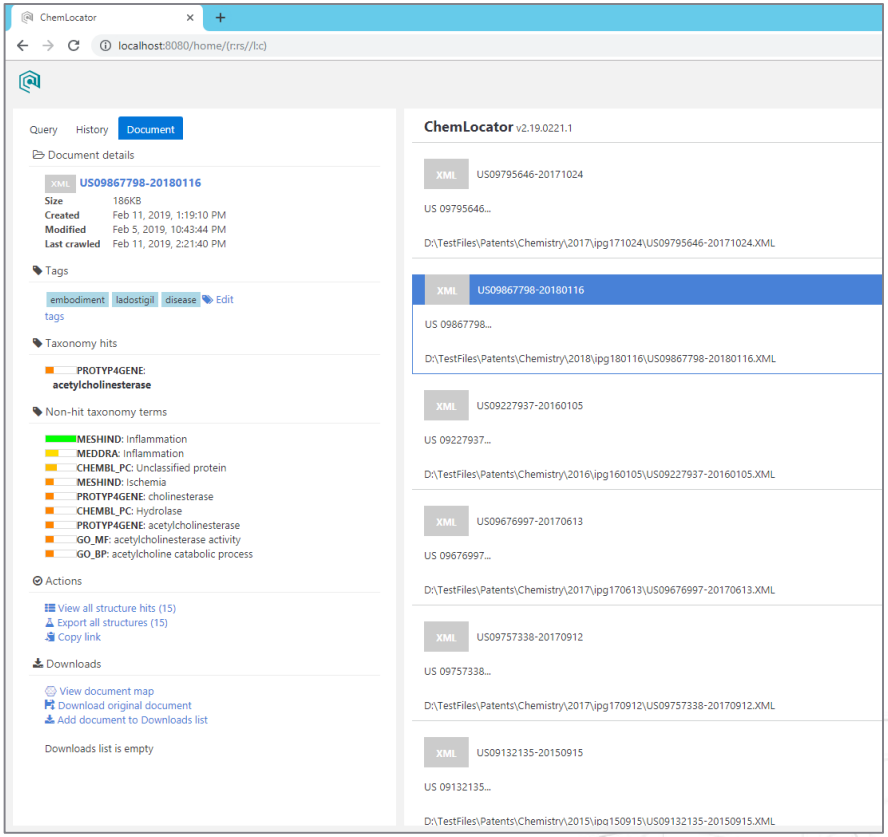




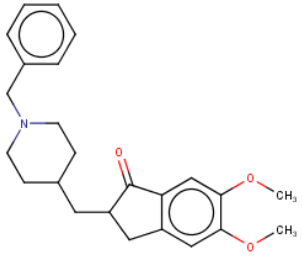
ChemAxon

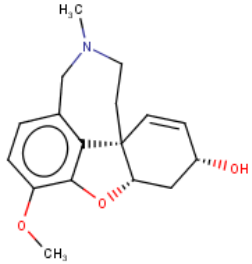
Chemical space overlap of two targets 1.

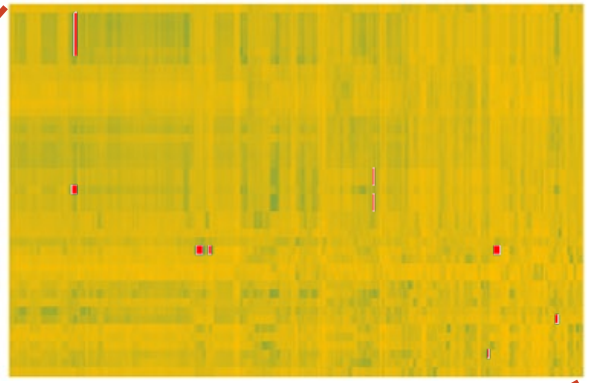


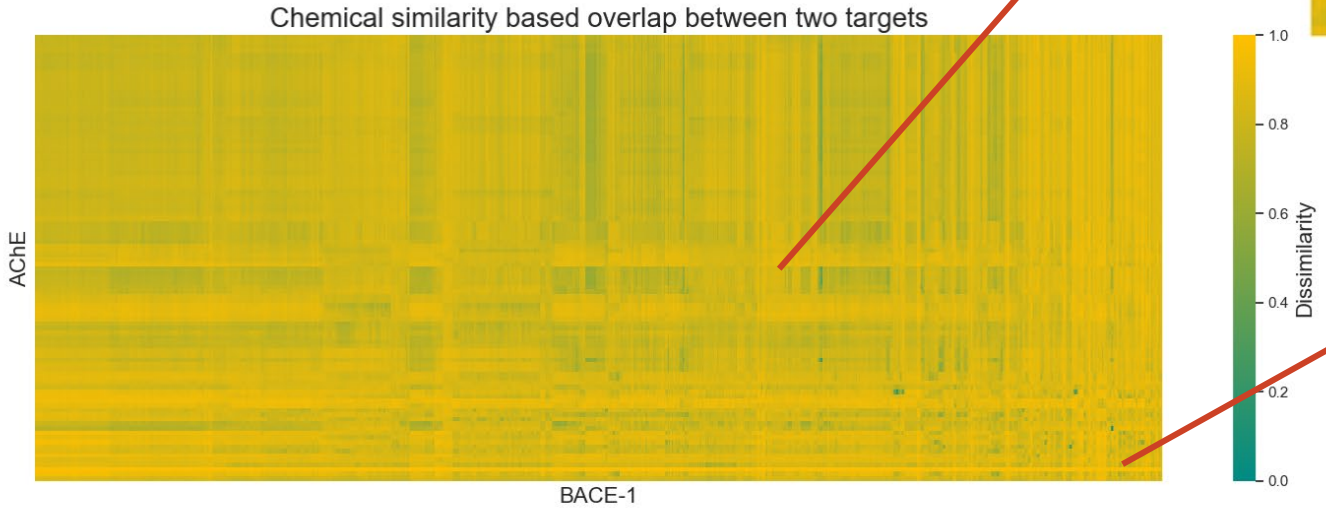


Chemical space overlap of two targets 2.



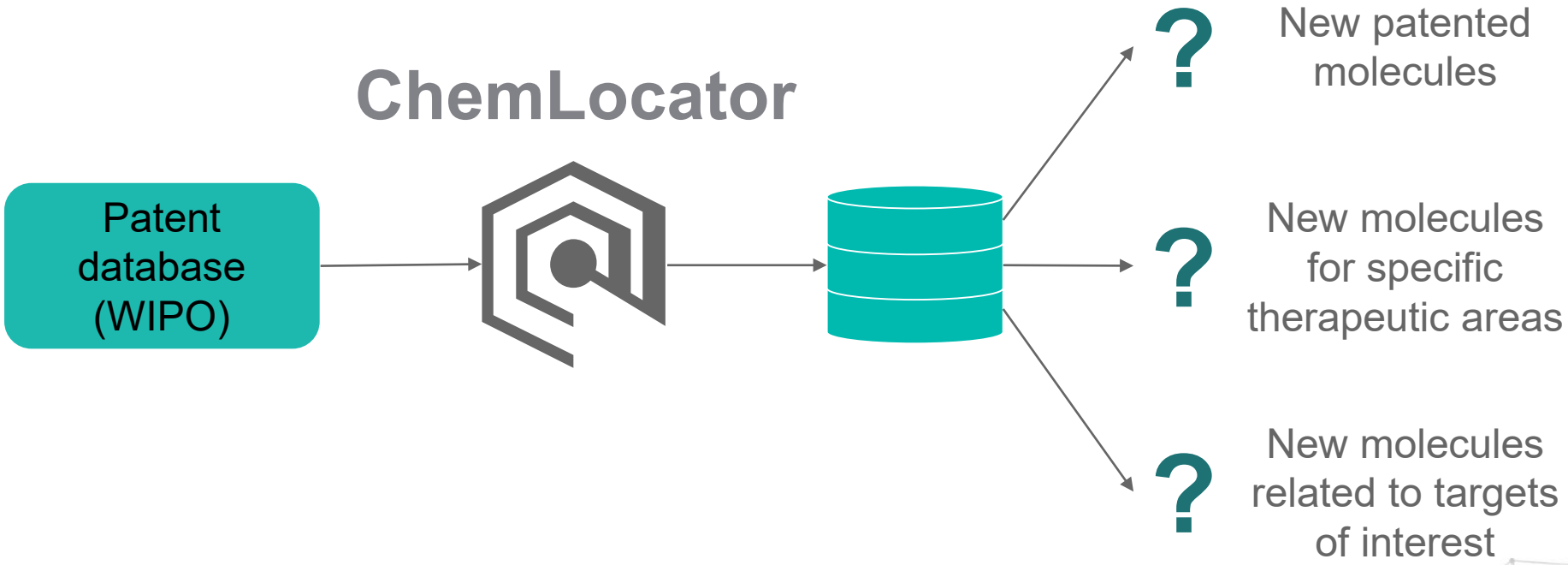






ChemAxon

Up-to-date information automatically



ChemAxon

THANK YOU!

ChemAxon