Marcus Gastreich
Sally Hindle

Searching the Impossible
FTrees in Fragment Space
1. Overview of Descriptors and FTrees
2. Fragment Spaces:
   - Generation
   - FTrees Fragment Spaces
   - Assembly and searching
   - Application
3. Outlook
Molecular Similarity Descriptors: Bitstrings

typical 2D and 3D descriptors:

00101110101010110001
Molecular Similarity Descriptors: FTrees

2½D FTrees descriptor:
Feature Tree

- Molecule represented as a tree
  - no bit string
  - no 3D conformers
- Nodes store chemistry / physics
- Topology preserved

Molecular Similarity Descriptors: FTrees

Advantages of FTrees:

- delivers much more than just a similarity score
- optimum similarity
- fast like 2D, retains topology
- retains topology, remains fuzzy

100 000 comparisons → 1 minute

**FTrees “Intuition”**

**2D Fingerprints (Unity)**

<table>
<thead>
<tr>
<th>Compound</th>
<th>2D Fingerprint</th>
<th>FTrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.203</td>
<td>0.764</td>
<td></td>
</tr>
<tr>
<td>0.203</td>
<td>0.655</td>
<td></td>
</tr>
<tr>
<td>0.203</td>
<td>0.550</td>
<td></td>
</tr>
</tbody>
</table>
FTrees Scaffold Hopping: Dopamine D4 Antagonists

Cpds from: M.A. Sanner, Exp. Opin. Ther. Patents, 8 (1998)
957 compounds from MDDR
383 cover 5 activity classes

experiment:
- 10 nearest neighbors for each active
- hit rate = number from correct class X 10
- average the hit rate over 383

FTrees: Comparison with Other Descriptors

query
PAF antagonist

hit
PAF antagonist

2D: 729 / 957
FTrees: 5 / 957
Fragment Spaces

- The basic idea:
  - *de novo* assembly of fragments to generate novel molecules

- Problems:
  - Where do the fragments come from?
  - How do we ensure high probability of synthesizability?

- Answer:
  - Take a combi lib OR take a compound set and “shred” via…

...a retrosynthetic approach: RECAP

Fragment Spaces: Generation

RECAP RULES

Fragment Space
Fragment Spaces: Generation

corporate collection

RECAP RULES

Fragment Space
FTrees Fragment Spaces: Generation

RECAP RULES

Fragment Space
FTrees Fragment Spaces: Generation

RECAP RULES
→ SMARTS MATCHER

H₂N-CH₂-CH₂-OH

*C(=O)O >> *.C(-[OH])O

H₂N-CH₂-CH₃ + CO₂H
FTrees Fragment Spaces: Generation

cSLN
(input combi libs without shredding)
FTrees Fragment Spaces: Assembly

- Convert fragments
- Make linkage with special linker nodes
- Combined tree loses linkers
- Back to molecule
FTrees Fragment Spaces: Searching

query

fragment space

dynamic programming *


* deterministic:
  - same result for every run
  - always identify best mol
Application: WDI Fragment Space

RECAP RULES

1. Cleave at 11 default bond cleavage types

2. Don't cleave small groups (e.g., Me, MeO, H…)

3. Don't cleave rings (nodes contain complete rings)

Application: WDI Fragment Space

~35 000 subset of WDI

RECAP RULES

Fragment Space

16 780 fragments

Impossible?...

<= 5 fragments → searching in $10^{18}$ compounds!

1 search → a couple of minutes…

### Application: WDI & Histamine H1 Antagonists

- **Target similarity level = 0.9**

<table>
<thead>
<tr>
<th>Query</th>
<th>Generated structure</th>
<th>Known Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimethindene</td>
<td><img src="image1.png" alt="Generated structure" /></td>
<td><img src="image2.png" alt="Known Active" /></td>
</tr>
<tr>
<td>Mianserin</td>
<td><img src="image3.png" alt="Generated structure" /></td>
<td><img src="image4.png" alt="Known Active" /></td>
</tr>
</tbody>
</table>
**Application: WDI & Angiotensin2 Inhibitors**

- **Target similarity level = 0.9**

<table>
<thead>
<tr>
<th>Query</th>
<th>Generated structure</th>
<th>Known Active</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Query Structure" /></td>
<td><img src="image" alt="Generated Structure" /></td>
<td><img src="image" alt="Known Active" /></td>
</tr>
</tbody>
</table>
Application: WDI & COX2 inhibitors

- Hop between structural classes:

1. **Search**
   - ![Structure 1](image1.png)

2. **Similar to**
   - ![Structure 2](image2.png)

3. **Known active**
   - ![Structure 3](image3.png)

4. **Search**
   - ![Structure 4](image4.png)
Application: Drug Discovery Workflow

corporate database

shredding & converting

known actives

FTrees-FS

structure based design

superposition

data analysis module
Application: Drug Discovery Workflow

BS Python modules:
PyFTrees
PyFlexS
PyFlexX

Excel
Oracle
OELib
PyMOL
PyMOL
Daylight
Chemdraw
...

Python

ACS Philadelphia 2004
Outlook

- Up-to-date FTrees evaluations
- FTrees GUI
- Fragment space creation:
  - Shredding: SMARTS matcher
  - cSLN input
- New fragment spaces and fragment space analyses
- Integrated modular python workflows
Acknowledgements

- Matthias Rarey, ZBH Hamburg, Germany
- Martin Stahl, Roche, Switzerland
- Thomas Lengauer, MPII Saarbrücken, Germany
- Marc Zimmermann, FhG SCAI, Germany
- Colleagues at BioSolve IT