Library Compound Design Methods for Custom Library Synthesis
Offers
Usage

REALITY TV
Library design by ChemAxon

Fragmentation
R-group decomposition
Fragmentation
Reagent clipping

Enumeration
Fuse fragments
R-group composition
Reaction enumeration
Markush enumeration

Databases
Reactions
Molecules
Markush structures
Queries

Compound selection
Similarity searches
Substructure searches

Library analysis
Clustering
2D similarity screen
3D Shape similarity screen
## Library design by ChemAxon

<table>
<thead>
<tr>
<th>Feature</th>
<th>ChemAxon Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical data storage</td>
<td>JChemBase</td>
</tr>
<tr>
<td></td>
<td>JChem Cartridge for Oracle</td>
</tr>
<tr>
<td>Chemical data search</td>
<td>JChem search technology</td>
</tr>
<tr>
<td>Chemical data visualization</td>
<td>JChem for Excel – Marvin</td>
</tr>
<tr>
<td></td>
<td>Instant Jchem - Marvin</td>
</tr>
<tr>
<td>Chemical data characterization</td>
<td>Calculator plugins – logP, pKa ...</td>
</tr>
<tr>
<td>Enumeration</td>
<td>Reactor – reaction enumeration</td>
</tr>
<tr>
<td></td>
<td>Markush enumeration</td>
</tr>
<tr>
<td></td>
<td>R-group composition</td>
</tr>
<tr>
<td></td>
<td>Fragment fusion</td>
</tr>
<tr>
<td>Fragmentation</td>
<td>Fragmenter</td>
</tr>
<tr>
<td></td>
<td>R-group decomposition</td>
</tr>
<tr>
<td>Analysis</td>
<td>JKlustor</td>
</tr>
<tr>
<td></td>
<td>Screen 3D - Screen 2D</td>
</tr>
</tbody>
</table>
Databases: displaying content on your desktop

Instant JChem

JChem for Excel
Databases: displaying content: JSP application

ONLINE TRYOUT
https://www.chemaxon.com

- Search technology
- Descriptors
- Alignments
- Chemical Terms filter
- Import / Export / Edit
- AJAX in JChem Webservices
Building blocks for library enumeration

Instant JChem
- Fragmentation

JChem for Excel
- R-group decomposition

Command line
- Fragmentation
- R-group decomposition
Which fragments?

Optimization of Similarity search metrics:
ECFP/FPCP/Chemical FP/Pharmacophore FP

regular Tanimoto: 0.47
optimized Tanimoto: 0.57
regular Tanimoto: 0.55

0.20
0.28
0.06
Similarity searching statistics

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Metric</th>
<th>Enrichment</th>
<th>Selectivity Effectiveness</th>
<th>Active Hit Distribution</th>
<th>Threshold</th>
<th>Test Hits</th>
<th>Target Hits</th>
<th>Before actives</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>Tanimoto</td>
<td>NaN</td>
<td>0.5</td>
<td>0.15</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>623 33, 14, 49, 102, 89, 248, 35, 53</td>
</tr>
<tr>
<td>PP</td>
<td>Euclidean</td>
<td>NaN</td>
<td>0.5</td>
<td>0.29</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>669 21, 23, 40, 35, 176, 70, 22, 282</td>
</tr>
<tr>
<td>PP</td>
<td>TanT</td>
<td>1.5</td>
<td>0.56</td>
<td>0.15</td>
<td>0.82</td>
<td>3</td>
<td>174</td>
<td>623 33, 14, 49, 102, 89, 248, 35, 53</td>
</tr>
<tr>
<td>PP</td>
<td>TanA</td>
<td>1.45</td>
<td>0.56</td>
<td>0.15</td>
<td>0.7</td>
<td>3</td>
<td>180</td>
<td>623 33, 14, 49, 102, 89, 248, 35, 53</td>
</tr>
<tr>
<td>PP</td>
<td>TanS</td>
<td>1.05</td>
<td>0.52</td>
<td>0.19</td>
<td>0.87</td>
<td>5</td>
<td>415</td>
<td>626 47, 18, 17, 55, 74, 308, 68, 39</td>
</tr>
</tbody>
</table>

Number of queries 1
Number of test set elements 8
Number of targets 700

Graph: Number of Hits vs. Number of Active Hits

- **Tanimoto**
- **Euclidean**
- **Optimized**
- **Ideal**
## Enumeration

<table>
<thead>
<tr>
<th>Output files</th>
<th>ChemAxon technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragments</td>
<td>Fragment fusion</td>
</tr>
<tr>
<td>Markush structures</td>
<td>Markush enumeration (search without enumeration)</td>
</tr>
<tr>
<td>Reactants – generic reactions</td>
<td>Reaction enumeration</td>
</tr>
<tr>
<td>R-tables</td>
<td>R-group composition</td>
</tr>
</tbody>
</table>
### Enumeration

#### R-table

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><img src="image1" alt="Image" /></td>
<td>R1</td>
<td>R1</td>
<td>R2</td>
</tr>
<tr>
<td><img src="image2" alt="Image" /></td>
<td>H2N</td>
<td>N2H</td>
<td>Br</td>
</tr>
<tr>
<td><img src="image3" alt="Image" /></td>
<td>H3C</td>
<td>CH3</td>
<td></td>
</tr>
<tr>
<td><img src="image4" alt="Image" /></td>
<td>Cl</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td><img src="image5" alt="Image" /></td>
<td>H</td>
<td>CH3</td>
<td></td>
</tr>
<tr>
<td><img src="image6" alt="Image" /></td>
<td>H</td>
<td>OH</td>
<td>COOH</td>
</tr>
<tr>
<td><img src="image7" alt="Image" /></td>
<td>H</td>
<td>F</td>
<td>Cl</td>
</tr>
</tbody>
</table>

#### Markush structure

- **R1** = Cl, NO
- **R2** = H, CH3
- **R3** = H, OH, COOH
- **R4** = H, F, Cl
ChemAxon in Knime

MultipleMolImporter ➔ LibraryMCS ➔ MarvinTable ➔ MarvinTable ➔ MolExporter

Node 1 ➔ Node 2 ➔ Node 3 ➔ Node 8 ➔ Node 12

Row Splitter ➔ R-group Composition ➔ Markush Enumeration

selection of t... ➔ Node 10 ➔ Node 11

MultipleMolImporter ➔ MarvinTable ➔ RxnSimilarity

Node 4 ➔ Node 9 ➔ Node 13 ➔ Node 15
Reaction Enumeration

EXCLUDE:

\[
\text{match(reactant(1), } "[\text{Cl, Br, I}]C([=O,S])C=C"	ext{) or}
\]
\[
\text{match(reactant(0), } "[\text{H}][O,S]C=[O,S]"	ext{) or}
\]
\[
\text{match(reactant(0), } "[\text{P}][\text{H}]"	ext{) or}
\]
\[
(\text{max(pka(reactant(0), filter(reactant(0),}
\]
\[
"\text{match('[O,S;H1]')", "acidic"}}) > 14.5) or
\]
\[
(\text{max(pka(reactant(0), filter(reactant(0),}
\]
\[
"\text{match('[#7:1][H]', 1)"}, "basic"}})) > 0)\]
ChemAxon in Knime
• Characterisation of library:
  – Fragments - Fragmenter
  – Molecular descriptors – Calculator plugins
Library analysis – 3D shape similarity search

Test on DUD

1% Enrichment

Library analysis

Wide range of methods
- Unsupervised, agglomerative clustering
- Hierarchical and non-hierarchical methods
- Similarity based and structure based techniques

Flexible search options
- Tanimoto and Euclidean metrics, weighting
- Maximum common substructure identification
- Chemical property matching including atom type, bond type, hybridization, charge
Synthetically Feasible Virtual Database

- Diverse, virtual library to be used in virtual screening, lead hopping and lead optimization for drug discovery
  - Sampling of chemically accessible space
- Synthetically accessible products
  - Well defined reaction scheme
  - Readily available starting materials
- Tools
  - ChemAxon Reactor
    - Enumerate synthetically feasible virtual products
  - Pipeline Pilot
    - Process the enumerated reactions
Use cases

Reactor in Pipeline Pilot

- Supports "smart" reactions
  - generic reaction equations combined with reaction rules generating chemically feasible products
- Reaction library of 145 reactions
- Synthesis code generation
  - Unique identifier for product
- Reaction Rules
  - Reactivity and Exclude
  - Selectivity
  - Tolerance

New Upgrades

Parameters

<table>
<thead>
<tr>
<th>String</th>
<th>File</th>
<th>Reactant Combination</th>
<th>Output Type</th>
<th>Mapping Style</th>
<th>Included Products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Combinatorial</td>
<td>Product</td>
<td>none</td>
<td>Product 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Synthesys Code Generation
- Generate Synthesis Code
  - True
- Reaction Synthesis Code
  - R1

Advanced

- Unambiguous only
  - False
- Ignore reaction rules
  - Reactivity and Exclude
    - False
  - Selectivity
    - False
  - Tolerance
    - False

Tagging

- Product Number Tag
  - Product_Number
- Reaction Number Tag
  - Reaction_Number
**Use cases**

**Reactor Library Reactions: Bischler-Mohlau Indole Synthesis**

- **Reactivity** *(select appropriate reagents)*
  - First reactant is aryl-methyl-ketone with alpha halogen, amino or activated hydroxy derivative (tosyl, mesyl…)
  - Second reactant is aniline not acylated

- **Selectivity** *(finds most reactive sites)*
  - Aromatic electrophilic substitution goes faster on the most negative carbon

- **Exclude** *(compounds that may yield side reactions)*
  - Exclude compounds which may react with anilines
Reactor Library Reactions: Suzuki Coupling

(A,a),1 → B 3 + [O,S,Cl,Br,I]4 L → (A,a),2 → (A,a),1 → (A,a),2

Organo-boron Halide or Pseudo-halide C-C Coupled

- **Reactivity (select appropriate reagents)**
  - First reactant is boronic acid, ester or anhydride, borane or trifluoroborate salt
  - Second reactant may be halogen or organo-sulfonate or methyl sulfide (behave as pseudo-halogen)

- **Selectivity (finds most reactive sites)**
  - The most negative carbon atom reacts in oxidative addition from the second molecule

- **Exclude (compounds that may yield side reactions)**
  - Second reactant may not contain such functional groups which compete with the boron containing group in the transmetallation step of the catalytic cycle
CombiChem Example Using Reactor Library

Use cases

**Synthesis Code Generation:**

R1(A1,B1): 1/1

1. Bischler-Mohlau Indole Synthesis

Bischler-Mohlau Indole (CC13105, TL00240): 1/1

2. Suzuki Coupling

Suzuki_coupling (CC04212, Bischler-Mohlau Indole (CC13105, TL00240): 1/1): 1/1
### Reaction Rules in Reactor

<table>
<thead>
<tr>
<th>Reaction Rules</th>
<th>ON</th>
<th>OFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bischler-Mohlau</td>
<td>51</td>
<td>310</td>
</tr>
<tr>
<td>(10 Ketones x 20 Amines)</td>
<td>(200)</td>
<td>(200)</td>
</tr>
<tr>
<td>2. Suzuki</td>
<td>30</td>
<td>3380</td>
</tr>
<tr>
<td>(Indoles x 10 Boron)</td>
<td>(510)</td>
<td>(3100)</td>
</tr>
</tbody>
</table>

- **Reactions Rules: ON**
  - May yield fewer results than theoretical since REACTOR eliminates starting materials that are synthetically unfeasible

- **Reactions Rules: OFF**
  - Could use for debugging purposes
  - May yield more results than theoretical since the reactions are interpreted generically
Use cases

Reaction Rules OFF – Caution!

Reactivity: O is organo-sulfonate ("pseudo-halogen")

Suzuki Coupling

(Generic interpretation of reaction!)

(A,a),1 + [O,S,Cl,Br,I]4 \rightarrow (A,a),2

Bischler-Mohlau Indole(F1011-0534, F0074-0075):2/1

Synthetically Unfeasible Product

Incorrect Product

Suzuki_coupling(CC04212, Bischler-Mohlau Indole(F1011-0534, F0074-0075):2/1):1/1
Suzuki_coupling(CC04212, Bischler-Mohlau Indole(F1011-0534, F0074-0075):2/1):2/1
Protocol Summary

1. Search for applicable starting materials.
2. Run ChemAxon Reactor to do combinatorial chemistry.
3. Filter and Cluster output to obtain sampling of reaction.
Pipeline Pilot Protocol for CombiChem Example
Target-focused libraries: rapid selection of potential PDE inhibitors from multi-million compounds’ repositories

Why do we need rapid selection of target-focused libraries?

Design inputs

2D similarity searching strategy

Property-based filtering

Seed/chemotype representation (diversity)

Conclusion/Proposals
Target-focused libraries via Virtual Screening

TargetEx Ltd.,
György Dormán

Source Compounds
Commercial Samples
Combinatorial Libraries/Historical collections
De Novo Compounds

Filtering
ADMET
Lead-likeliness

Docking

Target structure

Known Active Compounds

2D Substructure-Similarity Searching
Partitioning
Data Fusion
Clustering
Kernels
SVM

2D fingerprint

3D Pharmacophore
Shape Similarity
3D/4D-QSAR

Final Visual Inspection
Acquisition
Plating

Biological testing

Focused library

H-bond Acceptor
Cation
Aromatic
H-bond donor
2D similarity selection

5 M compounds (10 vendors)

2D similarity search

Commercial Compound Libraries

Seed compounds
Example of known XXX antagonists

2D fingerprint (bitstrings)

Analogue library

Property filtering

Diversity selection

Visual inspection

8,655 compounds

2,009 compounds

Not available

Under progress

compounds to be purchased
Similarity searching strategy: execution

- Setting the starting similarity level (dependent on the fingerprint S/W, T= 60-75 % for ChemAxon)
- Iteration based on the results (scenarios):
  - the number of virtual hits are between 50 and 500, OK
  - the number of virtual hits are <50 or >500
    - if <50 lower the similarity treshold with 5 %
    - if >500 increase the similarity treshold with 5 %
    - This can be continued until the optimal range achieved
    - If 5 % decrease results in >500 compounds the search can be refined by 2% (alternatively a diversity selection would be needed, but that is not available)
    - Duplications can be removed when merged the resulting DBs
How to reduce the number of the hits?

Normally screening companies would like to buy 100-1000 compounds

• Since from the various vendor DBs we can obtain 2000-10,000 virtual hits their number can be reduced

• 1. Applying the reference property space (Lipinski and Veber rules) (IJchem OK)

• 2. There are overrepresented seeds thus virtual hits coming from those seeds can be reduced (IJchem OK)

• 3. Applying an optimal distribution of the resulting chemotypes (removing the overrepresented compounds) (Limited with Jklustor)

• 4. Simple diversity analysis (JKlustor)
1. Applying the reference property space:
Structural determinants: H-bond donor/acceptor, hydrophobic interactions
(property space determination)

Pharmacophore fingerprints requires more computation and time consuming.
In simple similarity search pharmacophore features can only be considered as statistical features (not connected to structures).
The similarity search results can be filtered based on the physico-chemical parameter space of the seed compounds (+10/-10 % range applied).
Results and further reduction

• Similarity search results: 8655
• After property filtering: 2009
• 2. There are overrepresented seeds thus virtual hits coming from those seeds can be reduced
• When combining the similarity search the contribution of the seeds can be controlled (or set the number of analogues derived from certain seeds)
2. Overrepresented seeds

Seeds leading to highest number of similar hits

#4 (Sildenafil)
238 analogues
(60 % similarity or above)

#13
328 analogues
(60 % similarity or above)

#18 (desantafil)
4494 analogues
(60-80 % similarity)
2. Overrepresented seeds

*Seeds leading to highest number of similar hits*

#27
237 analogues
(60 % similarity or above)

#28
272 analogues
(60 % similarity or above)

#30
466 analogues
(60 % similarity or above)

#44
2726 analogues
(60 % similarity or above)
Recurring structural motifs in the seed structures
Recurring structural motifs in the similarity search results
3. Applying an optimal distribution of the resulting chemotypes
Proposed application of JKlustor/LibMCS

- Taking into consideration of the substructure where the maximum number of connection (bond) is found
  - it can be an option
  - Maybe difficult to define

- Using such option the „real” core structure can be found easier
A brief history of time
Life of Evotec cheminformatics software

- 1992: Oxford Asymmetry International created
- 1999: CCD1 (MDL)
- 2000: CCD2 Viewer (Daylight)
- 2001: Evotec Supplier Database (ESD) (Daylight)
- 2002: ESMA EVOseeq (Daylight)
- 2003: Start working with ChemAxon
- 2004: ELN Viewer (ChemAxon)
- 2005: ESMA EVOseeq (ChemAxon)
- 2006: CCD3 (ChemAxon)
- 2007: EVOsource (ChemAxon)
- 2008: PFA (ChemAxon)
- 2009: Instant JChem (ChemAxon)
- 2010: Reaxys (ChemAxon)
- 2011: JC4XL (ChemAxon)
- ...
Evotec Library Profiler

• Aim is to be able to select from a large virtual library either:
  – A combinatorial subset
    • Typically small focussed libraries
  – A non-combinatorial subset
    • Medicinal chemistry projects

• Desirable to allow access to all scientists
  – Creativity
  – Share ideas
  – Security aspect

• Interactive

• Subsets need to satisfy multiple criteria
Workflow using the Library Profiler

- Enumerate Virtual Library
- Select properties to calculate
- Filtering / analysis
- Export to file
- Further analysis
<table>
<thead>
<tr>
<th>1 (Reagent 2) Structure</th>
<th>logP for Step 1 (Reagent 2)</th>
<th>Hydrogen bond acceptors for Step 1</th>
<th>Molecular weight for Step 1 (Reagent 2)</th>
<th>Hydrogen bond donors for Step 1 (Reagent 2)</th>
<th>Product</th>
<th>logP for Product</th>
<th>BBB for Product</th>
<th>Hydrogen bond acceptors for Product</th>
<th>HSA for Product</th>
<th>Molecular weight for Product</th>
<th>TPSA for Product</th>
<th>AVERAGE Energy (kcal/mol) for Product</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Structure 1" /></td>
<td>2.61</td>
<td>4</td>
<td>59.42</td>
<td>233.19527</td>
<td><img src="image2" alt="Product 1" /></td>
<td>2.92</td>
<td>-0.3120915</td>
<td>5</td>
<td>0.27364637</td>
<td>316.32697</td>
<td>60.45</td>
<td></td>
</tr>
<tr>
<td><img src="image3" alt="Structure 2" /></td>
<td>2.332</td>
<td>6</td>
<td>77.88</td>
<td>259.2143</td>
<td><img src="image4" alt="Product 2" /></td>
<td>2.342</td>
<td>-1.201831</td>
<td>7</td>
<td>-0.06353126</td>
<td>342.546</td>
<td>78.91</td>
<td></td>
</tr>
<tr>
<td><img src="image5" alt="Structure 3" /></td>
<td>1.92</td>
<td>4</td>
<td>68.01</td>
<td>209.21326</td>
<td><img src="image6" alt="Product 3" /></td>
<td>1.66</td>
<td>-0.7910311</td>
<td>5</td>
<td>0.11273974</td>
<td>289.34808</td>
<td>69.04</td>
<td></td>
</tr>
<tr>
<td><img src="image7" alt="Structure 4" /></td>
<td>3.4</td>
<td>3</td>
<td>55.12</td>
<td>272.22924</td>
<td><img src="image8" alt="Product 4" /></td>
<td>3.237</td>
<td>0.10484234</td>
<td>4</td>
<td>0.3357661</td>
<td>385.38492</td>
<td>56.15</td>
<td></td>
</tr>
<tr>
<td><img src="image9" alt="Structure 5" /></td>
<td>1.386</td>
<td>4</td>
<td>66.4</td>
<td>209.17386</td>
<td><img src="image10" alt="Product 5" /></td>
<td>1.655</td>
<td>-0.7354481</td>
<td>5</td>
<td>0.1051364</td>
<td>292.30557</td>
<td>67.43</td>
<td></td>
</tr>
<tr>
<td><img src="image11" alt="Structure 6" /></td>
<td>0.742</td>
<td>3</td>
<td>50.19</td>
<td>149.14617</td>
<td><img src="image12" alt="Product 6" /></td>
<td>1.722</td>
<td>-0.42710893</td>
<td>3</td>
<td>0.021787751</td>
<td>244.31326</td>
<td>41.99</td>
<td></td>
</tr>
<tr>
<td><img src="image13" alt="Structure 7" /></td>
<td>3.383</td>
<td>3</td>
<td>46.53</td>
<td>214.2167</td>
<td><img src="image14" alt="Product 7" /></td>
<td>4.573</td>
<td>-0.15614685</td>
<td>3</td>
<td>0.6135623</td>
<td>309.38324</td>
<td>36.30</td>
<td></td>
</tr>
<tr>
<td><img src="image15" alt="Structure 8" /></td>
<td>3.903</td>
<td>3</td>
<td>46.53</td>
<td>214.2167</td>
<td><img src="image16" alt="Product 8" /></td>
<td>4.573</td>
<td>-0.17267993</td>
<td>3</td>
<td>0.614996</td>
<td>309.38324</td>
<td>36.30</td>
<td></td>
</tr>
</tbody>
</table>
Charting – Scatter plot
Pivot View - Properties
Using ChemAxon tools

• High usage of Marvin View and Sketch
  – Easy to integrate

• JChem cartridge for filtering
  – Experience in using cartridge

• JChem tools for many of the property calculations
  – HBD, HBA, ROT, AMW, TPSA, Veber Bioavailability, BBB distribution, undesirable functional groups, Andrews AVERAGE energy, Bioavailability score, Ligand binding efficiency, PGP Substrate prediction, pKa, protonated atom count, non-H atom count
Focused and diverse library generation by ChemAxon technology

WORKSHOP AT 14:00
HANDS-ON SESSION
Visit other technical presentations

www.chemaxon.com

**Visualization**
- Marvin
  - Structure, query & reaction editor, viewer & visualization

**Property Prediction**
- Calculator Plugins
  - Structure property prediction & calculation

  - pKa, Major microspecies
  - logP, logD
  - Charge
  - Tautomerization
  - Stereoisomer
  - Conformation and 3D alignment
  - Topology/Analysis
  - Molecular Surface Area
  - Markush Enumeration
  - Hydrogen bond donor/acceptor
  - Structural Frameworks
  - Structure to Name
  - etc...

**Chemical DB – toolkit**
- JChem Cartridge
  - JChem/Oracle integration

- JChem Base
  - Structure searching & db access

- Standardizer
  - Chemical business rules processing

- Structure checker
  - Batch structure file validation and correction

**Chemical DB – desktop**
- Instant JChem
  - Structure db management, search & prediction

- JChem for Excel
  - Enabling chemistry in Excel

**Nomenclature**
- Name to Structure
  - Import & search chemical names

**Enumeration**
- Reactor
  - Enumeration via reaction modelling

**Library analysis**
- JKlustor
  - Clustering & diversity analysis

- Fragmenter
  - Decomposition to fragments and R-groups

- Screen
  - HT pharmacophore screening

**Add-ons**
- Markush Search
  - Store & search Markush structures

- JChem Webservices
  - Web services integration interface