Analyzing Building Blocks Diversity for DNA Encoded Library Design

Cresset User Group Meeting
Nik Stiefl & Finton Sirockin, Novartis
2016.06.16
Outline

- DNA Encoded Libraries (DEL)
- Building blocks selection process
- Capping group selection
  - 2D fingerprints
  - Cresset fields
- Caveats
- Conclusion
DEL technology uses DNA oligonucleotides to record the combinatorial synthesis of organic molecules...

- **Dimer library, W x X compounds**
  - Pos 1: W building blocks, A₁ to Aₓ
  - Pos 2: X building blocks, B₁ to Bₓ

  ![Diagram of dimer library](image)

- **Trimer library, W x X x Y compounds**

  ![Diagram of trimer library](image)

- **Tetramer library, W x X x Y x Z compounds**

  ![Diagram of tetramer library](image)
Building Blocks selection process

General process

Internal Vendors

Diversity

Price

Chemical intuition

# Building Blocks Selection Process

*What makes DNA Encoded Libraries different?*

<table>
<thead>
<tr>
<th></th>
<th>Project libraries</th>
<th>DNA Encoded Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td># of building blocks</td>
<td>10-50</td>
<td>100s – 1000s</td>
</tr>
<tr>
<td>Design principle</td>
<td>Target knowledge</td>
<td>Diversity &amp; Density</td>
</tr>
<tr>
<td>Priority on</td>
<td>Scaffold</td>
<td>Combinator’ics’</td>
</tr>
<tr>
<td>Property space</td>
<td>Target driven and limited</td>
<td>Simultaneous SAR and physico-chemical exploration</td>
</tr>
</tbody>
</table>
2D descriptions

Example: Atom Pairs\(^{[1]}\)

- Atom properties (atomic number, degree, number pi electrons) and inter-atom distances encode molecules

![Diagram of molecules with atom pairs AP: 0.58](image)

2D descriptions

Influence of reacting group on descriptors

- Modifying atom invariants and Rooted Fingerprints \[1\]

\[\text{AP : 0.58}\]

\[\text{AP : 1.0}\]

\[\text{APMI : 0.43}\]

\[\text{APRooted : 0.29}\]

\[\text{APMI : Atom Pairs Modified Invariant}\]

Capping groups specific fingerprints

Clockwise: AP vs. APMI; APRooted vs. APMI; APRooted vs. AP.

Similarity plots of pairs of 1832 aldehydes
From 2D to 3D

*Brainstorming in Novartis*
The Cauliflower™ (Frédéric Berst - Novartis)

**Distributing pharmacophores (Cresset fields) in space**

Diverse (2D) BBs

Distributing different PH4 (fields) types as evenly as possible in space

Hydrophobic
- Polar +
- Polar –
- Space around derivatisation point
Le «Caulitree»
Building Blocks clustering with Cresset technology

*Workflow developed by Paolo Tosco (Cresset)*

**BBs preparation**
- Cleaning
- Reacting group labelling

**BBs conformations**
- 3D coordinates generation and protonation
- Conformations generation

**BBs Cresset processing**
- Alignment on bond vector
- Cresset similarity matrix calculation
- Clustering

**BBs selection**
- Cluster visualisation
- BBs selection by clusters using multiple parameters (price, reactivity profiling, density of in-house availability, etc)
Pharmacophore distribution clustering

*Help chemists choosing BBs to maximise pharmacophore coverage and BB attractiveness*

*Example clusters from the processing of 1832 aldehydes (~ 12 hours on a cluster)*
Handling the reactive handle

*Functional group vs. isotope labelling*

- Minor influence on 2D description depending on BB size and functionality

- Major influence on Cresset fields (transform with respect to resulting product)
Handling the reactive handle

*Replace vs. add (“CC” vs. “CCC”)*

![Diagram showing the reactive handle comparison between “CC” and “CCC”](image-url)
Pharmacophore spaces

*Identifying missing spots*

1300 NAS reagents aligned on the derivatisation vector. Top and side views.

Disc-like electropositive density with obvious hole on top and bottom.
2D vs 3D - does it make a difference?

*Cresset vs. APRooted*

Cresset: 0.92 APRooted: 0.33

Cresset: 0.90 APRooted: 0.40
2D vs 3D - does it make a difference?

*Cresset vs. APRooted*

- **Cresset:** 0.57
  **APRooted:** 0.88

- **Cresset:** 0.66
  **APRooted:** 1.00

- **Cresset:** 0.39
  **APRooted:** 0.85
Caveats

- Reactive group treatment

- Tautomers

- Capping group only (for the moment)
Outlook

Core scaffolds – exit vector distribution analysis

Identification of unexplored exit vector space (branches in the caulitree)

To be combined with spatial distribution of cresset fields
Conclusions

- Library design tool to distribute pharmacophoric constraints evenly around linking vector
- Used to aid chemists visually select BBs from pharmacophoric clusters
- Modified 2D descriptors should be used in conjunction
- Applied in-house to large BB datasets for prioritization
Acknowledgments

F. Berst  V. Acker  H. Hemmerle
J. Ottl  P. Drueckes  J. Duca
C. Freslon  K. Twesten  G. Landrum (now @ Knime)
R. Martinez  S. Haddad
P. Fille  C. Dumelin  P. Tosco
A. Marzinzik  M. Huebscher  T. Cheeseright
X. Pelle  J. Giovannoni  M. Mackey
Y. Ruff  D. Ouzineb  R. Chauhan
S. Liverneau  N. Schneider
S. Pickett  C. Gaul
P. Lehr  P. Ertl
M.-H. Bellance  R. Lewis
BACKUPS
Large discrepancies between AP and APRooted
A DELibrary is **not limited to a single central scaffold**

**Elements of design**

- **Scaffold fragment set**
  - Derivatization vectors
  - Pharmacophoric elements

- **Derivatisation fragment set**
  - Pharmacophoric elements