Compound Design using Reaction Networks

James Wallace

All reactions quoted sourced from the Med. Chem. literature
Project Aims

• Find new pathways and new molecules in molecular design space based on known reactions

• Applications include
  • Improved *de novo* design
  • Alternative routes to known compounds
  • Exploration of SAR
Overview of Talk

- Brief review of reaction vectors (RVs) and their application to de novo design
- Introduction to reaction sequence vectors (RSVs)
- Generation of reaction network from database of reactions
- Representation of paths in the network as RVs and RSVs
- Application of RSVs and RV to explore SAR
Given a Molecule (A Hit) what can I do with it?

A typical SAR strategy

Medicinal Chemistry Strategy

Molecules from J. Med. Chem
Medicinal Chemistry Strategy

• Need to apply known transformations to hit compound, to find worthwhile analogues.

• Use transformation ‘rule set’ to keep generated results chemically feasible.
Reaction Vectors

\[
\begin{array}{c}
\text{Br(1,0,0)-2(1)-C(2,0,0)} \\
\downarrow \\
\text{Br\textsuperscript{-}CH\textsubscript{3}} \\
\Rightarrow \\
\text{Cl\textsuperscript{-}CH\textsubscript{3}} \\
\text{Cl(1,0,0)-2(1)-C(2,0,0)}
\end{array}
\]

Atom Pair removed during reaction

Atom Pair added during reaction


De Novo Design using RVs

Check starting material against RV database

Remove negative atom pairs from molecule (where applicable)

Add positive atom pairs to molecule wherever environments match (multiple locations if necessary)

RVs in Multi-Step *De Novo* Design

- Scoring functions assume a smooth progression from starting material to product, however, synthetic chemistry does not conform to this pattern

- Complex intermediates may score poorly compared to both starting material and product
Reaction Sequence Vectors

Molecules from J. Med. Chem
Reaction Sequence Vectors

Pairs removed over sequence

- C(2,2,1)-2(4)-C(2,2,1)
- I(1,0,0)-2(1)-C(3,2,1)
- C(2,2,1)-3-C(2,2,1)
- C(2,2,1)-3-C(2,2,1)
- I(1,0,0)-3-C(2,2,1)
- I(1,0,0)-3-C(2,2,1)

Pairs added over sequence

- C(3,1,0)-2(1)-C(1,0,0)
- C(3,2,1)-2(1)-C(3,1,0)
- C(3,2,1)-2(4)-C(3,2,1)
- O(1,0,0)-2(1)-C(3,1,0)
- O(1,1,0)-2(2)-C(3,1,0)
- O(1,1,0)-2(2)-C(3,1,0)
- O(2,0,0)-2(1)-C(3,1,0)
- O(2,0,0)-2(1)-C(3,2,1)
- C(3,1,0)-3-C(2,2,1)
- C(3,2,1)-3-C(2,2,1)
- C(3,2,1)-3-C(2,2,1)
- O(2,0,0)-3-C(2,2,1)
- O(2,0,0)-3-C(2,2,1)
- O(2,0,0)-3-O(1,1,0)
Sequence Generation

Collect database of reactions

Connect reactions where product of one reaction is reactant of another

Build network of reaction sequences

Molecules from J. Med. Chem
Network of reactions

Molecules from J. Med. Chem
Alternative synthetic routes
Reaction Network

- Reaction network can be generated from any source of Reaction SMILES, RDFile data.

- For example:
  - In house collections (ELNs, etc)
  - Literature abstraction (4.4M reactions in SPRESI)
  - NextMove collection
    - 1M+ reactions from US Patent Database
Reaction Network
A Med. Chem. Reaction Network

- 26K reactions produce 266K sequences
Reaction Sequence Vectors

Molecules from J. Med. Chem

RSV Structure

Generation

Structure

Generation

RSV Structure

Generation

Molecules from J. Med. Chem

26/06/2014 © The University of Sheffield
RSV Method Validation

- 88.4% success rate in reproducing known product from starting material and RSV
Generation of New Products

- 500 molecules < 500 Da used as starting materials
Applications

• Extend RSV approach to proof of concept applications

• Investigate and explore known drug sequences in SAR space
SAR exploration 1: Cilomast

In-Silico Products (Tanimoto 0.8–1)

Synthetic routes to Cilomast and NN

SAR exploration 2: Hydroxamates

PCA of property space & near neighbours

Legend

- **Known products**
- **Near neighbours** (Tanimoto 1.0-0.8)

PCA of property space
- all results

Legend

- Known products
- Near neighbours (Tanimoto 1.0 -0.8)
- All other products

SAR exploration 2: Hydroxamates

- Expand network around synthetic routes to NNs

Find sequences used to make products, and collect the individual reaction vectors

Pool all starting materials and reagents for these reactions, as well those used in the original *de novo* process

Apply all reaction vectors to all starting materials iteratively

Generate reaction network as before

---

26/06/2014 © The University of Sheffield
SAR exploration 2: Hydroxamates

SAR exploration 3: Carboxamide synthesis

SAR exploration 3: Carboxamide synthesis

Legend
- **Known products**
- **Near neighbours** (Tanimoto 1.0 - 0.8)
- **All other products**

SAR exploration 3: Carboxamide synthesis

Analysis of intermediates shows alternative route to all compounds from different starting material.

Code Release

• Regent Court Chemoinformatics:
  http://tech.knime.org/cheminformatics-extensions

• Reaction Vector Database Reader/Writer
• De Novo Structure Generator
• Multi-Objective Looping
• Pareto/Desirability Ranking

• More to come in near future
Conclusions

• Created algorithms to generate reaction networks and exploit them for *de novo* design.
• Demonstrated approaches to SAR exploration around products of interest, with visual feedback for medicinal chemists.
• Releasing early code as open source.
• Need to investigate how sensitive RSV method is to differing reaction, sequence input, etc.
Acknowledgements

Supervisors:

• Prof. Val Gillet (Information School)
• Dr. Beining Chen (Chemistry)
• Dr. Michael Bodkin (Industrial Supervisor)
• Dr. David Thorner (Industrial Supervisor)

Other Contributors:

• Hina Patel
• Ben Allen
• Dimitar Hristozov

Funding:

• Eli Lilly UK
• EPSRC/CASE Award