Algorithms, Evolution and Network-Based Approaches in Molecular Discovery
Drug Discovery

The State of the Art

Chemistry of Life

- Schizophrenia
- Anxiety
- Happiness

- Depression
- Love
- Fight or Flight

- Dopamine
- Serotonin
- Oxytocin
- Norepinephrine
- Epinephrine

A man doesn't know what happiness is until he's married, by then it's too late! (Frank Skinner)
Antipsychotic drugs and their Poly-Pharmacology

The chart shows the known affinity (Ki) values of antipsychotic drugs for a panel of receptors.

How can we go about discovering a novel antipsychotic?.
What do I make next?

Medicinal Chemistry Optimisation
Mapping Computational Drug Discovery

The toolbox

- Target Identification
- Ligand Identification
- Ligand Optimisation

- Protein – Protein Interaction Networks
- ADMET in-Silico profile
- Library Design
- Molecular Simulation
- SAR Networks
- SAR Analysis
- SAR Visualisation

- Acquire the tools
- Evidence of SAR
- Explore the SAR Envelope

- Target Deconvolution
- Phenotypic Deconvolution
- Sequence Alignment
- Homology Modelling Threading
- Enzyme Function Inference
- Pharmacophores
- 3D Structure

- Molecular Field Screening
- Docking
- Quantum Mechanics
- Activity Prediction

- Conformational Analysis
- Molecular Overlays
- Matched Pairs

- Target Druggability
- Multi-objective Optimisation
- Multi-objective Compound Design

- De novo Design Tools
- 2D/3D QSARs
- Change in Structure vs Change in Activity

- Scaffold Hopping

- Bioinformatics
ChEMBL

ChEMBLSpace

10.6k Targets
1.4m Compounds
12.8m Activities

ChEMBLSpace – a graphical explorer of the chemogenomic space covered by ChEMBL

Bioinformatics (2013) 29 (4): 523-524

https://www.ebi.ac.uk/chembldb/
ChEMBLSpace search: D2 & $\alpha_1$BA

Dopamine D2

$\alpha_1$B-Adrenergic
ChEMBLSpace: D2, $\alpha_1$BA, H1

Available for download search: ChEMBLSpace@sourceforge.net
Similarity Ensemble Approach (SEA)

Keiser et al.
Phenotypic Deconvolution

- Combination of:
  - Target Prediction
  - BioSAR - Laplacian-modified Naïve Bayes algorithm
  - Information gain prefiltering
  - Decision Trees (C4.5) for Classification

- Yields 70% accurate,

- Interpretable model for sleep outcome.

IF
ACTIVITY_ON D(2) Dopamine receptor
AND
ACTIVITY_ON Histamine H1 receptor
AND
ACTIVITY_ON 5-hydroxytryptamine receptor 2A
THEN
“Good Sleep”
Accessing Chemical Space

A virtual enumeration of chemical space up to 17 heavy atoms generated 166,443,860,262 molecules.

A Pharma screening collection up to 17 heavy atoms is typically 100–500K molecules, which is equal to 0.000003% of accessible space.
Strategies in Automated Molecule Design

Grow
A → Core X → B → C

Replace Scaffold
A → Core X → B → C

Merge Molecules
A → Core Y → B → C

Constraints
- 2D or 3D Generation
- Protein Cavity
- Volume Based
- Feature Based

Synthetic Feasibility
- Reaction Based
- Reagent Enumeration
- Transformation Based
- Complexity Score
- User Assessment
Motivation

Using reaction databases

- Using reaction databases
- Public reaction databases
- Commercial reaction databases
- Adding say ~250,000 reactions per year, strong medicinal chemistry bias
- Wealth of reaction data:
  - Extract the knowledge hidden in these data
  - Use this knowledge to assist the medicinal chemist
  - Suggest new, synthetically feasible molecules with desired bio profile
Reaction Vectors

\[ \text{reactant vector, } R = (R1 + R2) \]

\[ \text{product vector, } P \]

\[ \text{reaction vector, } D = P - R \]
Reaction Vectors in Structure Generation

- The reaction vector, $D$, equals the difference between the product vector, $P$, and the reactant vector, $R$

$$D = P - R$$

Given a reaction vector, $D$, and a reactant vector, $R$, the product vector, $P$, can be obtained

$$P = D + R$$

Given a product vector, $P$, can we reconstruct the product molecule(s)?

<table>
<thead>
<tr>
<th>I</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond</td>
<td>C-C</td>
<td>C=O</td>
<td>C-OH</td>
<td>C-OR</td>
</tr>
<tr>
<td>#</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

better descriptor is required
Modified Atom Pairs

Atom Pairs 2 (AP2) : $X_1(n, p, r) - 2(BO) - X_2(n, p, r)$

- $X$: element type
- $n$: number of bonds to heavy atoms
- $p$: number of $\pi$ bonds
- $r$: number of ring memberships
- $BO$: bond order

Atom Pairs 3 (AP3): $X_1(n, p, r) - 3 - X_2(n, p, r)$

- Extending the bond distance in atom pairs encodes more of the environment of the reaction centre
Beckmann Rearrangement

Reaction vector

<table>
<thead>
<tr>
<th>Negative APs</th>
<th>Positive APs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- C(3,2,1)-2(1)-C(3,1,0)</td>
<td>1+ C(3,2,1)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>2- C(3,1,0)-2(2)-N(2,1,0)</td>
<td>2+ C(3,1,0)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>3- N(2,1,0)-2(1)-O(1,0,0)</td>
<td>3+ C(3,1,0)-2(2)-O(1,1,0)</td>
</tr>
<tr>
<td>a- C(3,2,1)-3-N(2,1,0)</td>
<td>a+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>b- C(3,2,1)-3-C(1,0,0)</td>
<td>b+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>c- C(3,1,0)-3-C(2,2,1)</td>
<td>c+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>d- C(3,1,0)-3-C(2,2,1)</td>
<td>d+ N(2,0,0)-3-C(1,0,0)</td>
</tr>
<tr>
<td>e- C(3,1,0)-3-O(1,0,0)</td>
<td>e+ N(2,0,0)-3-O(1,1,0)</td>
</tr>
<tr>
<td>f- N(2,1,0)-3-C(1,0,0)</td>
<td>f+ O(1,1,0)-3-C(1,0,0)</td>
</tr>
</tbody>
</table>

X element type  
n number of bonds to heavy atoms  
p number of $\pi$ bonds  
r number of ring memberships  
BO bond order
Applying a RV to a reactant to generate a Product

1. Removing the negative atom pairs from the reactant

<table>
<thead>
<tr>
<th>Negative APs</th>
<th>Positive APs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- C(3,2,1)-2(1)-C(3,1,0)</td>
<td>1+ C(3,2,1)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>2- C(3,1,0)-2(2)-N(2,1,0)</td>
<td>2+ C(3,1,0)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>3- N(2,1,0)-2(1)-O(1,0,0)</td>
<td>3+ C(3,1,0)-2(2)-O(1,1,0)</td>
</tr>
<tr>
<td>a- C(3,2,1)-3-N(2,1,0)</td>
<td>a+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>b- C(3,2,1)-3-C(1,0,0)</td>
<td>b+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>c- C(3,1,0)-3-C(2,2,1)</td>
<td>c+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>d- C(3,1,0)-3-C(2,2,1)</td>
<td>d+ N(2,0,0)-3-C(1,0,0)</td>
</tr>
<tr>
<td>e- C(3,1,0)-3-O(1,0,0)</td>
<td>e+ N(2,0,0)-3-O(1,1,0)</td>
</tr>
<tr>
<td>f- N(2,1,0)-3-C(1,0,0)</td>
<td>f+ O(1,1,0)-3-C(1,0,0)</td>
</tr>
</tbody>
</table>
Applying a RV to a reactant to generate a Product

2. Adding positive atom pairs to the fragment

<table>
<thead>
<tr>
<th>Negative APs</th>
<th>Positive APs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- C(3,2,1)-2(1)-C(3,1,0)</td>
<td>1+ C(3,2,1)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>2- C(3,1,0)-2(2)-N(2,1,0)</td>
<td>2+ C(3,1,0)-2(1)-N(2,0,0)</td>
</tr>
<tr>
<td>3- N(2,1,0)-2(1)-O(1,0,0)</td>
<td>3+ C(3,1,0)-2(2)-O(1,1,0)</td>
</tr>
<tr>
<td>a- C(3,2,1)-3-N(2,1,0)</td>
<td>a+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>b- C(3,2,1)-3-C(1,0,0)</td>
<td>b+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>c- C(3,1,0)-3-C(2,2,1)</td>
<td>c+ C(2,2,1)-3-N(2,0,0)</td>
</tr>
<tr>
<td>d- C(3,1,0)-3-C(2,2,1)</td>
<td>d+ N(2,0,0)-3-C(1,0,0)</td>
</tr>
<tr>
<td>e- C(3,1,0)-3-O(1,0,0)</td>
<td>e+ N(2,0,0)-3-O(1,1,0)</td>
</tr>
<tr>
<td>f- N(2,1,0)-3-C(1,0,0)</td>
<td>f+ O(1,1,0)-3-C(1,0,0)</td>
</tr>
</tbody>
</table>

Atom Pairs 2 (AP2) : X1(n, p, r)-2(BO)-X2(n, p, r)

- X: element type
- n: number of bonds to heavy atoms
- p: number of π bonds
- r: number of ring memberships
- BO: bond order

No AP2s left in the reaction vector that match atom 11

Final Solution
Duplicate solution
How well does it work?

Organic Chemistry Database

<table>
<thead>
<tr>
<th>Reaction Type</th>
<th>Number of Reactions</th>
<th>Correctly Reproduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epoxide reduction</td>
<td>450</td>
<td>449 (99.8%)</td>
</tr>
<tr>
<td>Epoxide formation</td>
<td>450</td>
<td>444 (98.7%)</td>
</tr>
<tr>
<td>Ester to amide</td>
<td>172</td>
<td>172 (100.0%)</td>
</tr>
<tr>
<td>Alcohol dehydration</td>
<td>171</td>
<td>169 (98.8%)</td>
</tr>
<tr>
<td>Claisen rearrangement</td>
<td>61</td>
<td>54 (88.5%)</td>
</tr>
<tr>
<td>Beckmann rearrangement</td>
<td>123</td>
<td>123 (100.0%)</td>
</tr>
<tr>
<td>Friedel Crafts acylation</td>
<td>113</td>
<td>113 (100.0%)</td>
</tr>
<tr>
<td>Olefin metathesis</td>
<td>9</td>
<td>7 (77.8%)</td>
</tr>
<tr>
<td>Dieckmann condensation</td>
<td>98</td>
<td>91 (92.9%)</td>
</tr>
<tr>
<td>Nitro reduction</td>
<td>231</td>
<td>230 (99.6%)</td>
</tr>
<tr>
<td>Alkene oxidation</td>
<td>272</td>
<td>272 (100.0%)</td>
</tr>
<tr>
<td>Cope rearrangement</td>
<td>453</td>
<td>306 (67.5%)</td>
</tr>
<tr>
<td>Aldol condensation</td>
<td>134</td>
<td>134 (100.0%)</td>
</tr>
<tr>
<td>Alcohol amination</td>
<td>97</td>
<td>97 (100.0%)</td>
</tr>
<tr>
<td>Amide reduction</td>
<td>51</td>
<td>51 (100.0%)</td>
</tr>
<tr>
<td>Diels-Alder hetero</td>
<td>441</td>
<td>320 (72.6%)</td>
</tr>
<tr>
<td>Ether halogenation</td>
<td>58</td>
<td>58 (100.0%)</td>
</tr>
<tr>
<td>Ozonolysis</td>
<td>132</td>
<td>125 (94.7%)</td>
</tr>
<tr>
<td>Claisen condensation</td>
<td>98</td>
<td>77 (78.6%)</td>
</tr>
<tr>
<td>Carboxylic acids to aldehydes</td>
<td>194</td>
<td>194 (100.0%)</td>
</tr>
<tr>
<td>Nitrile reduction</td>
<td>102</td>
<td>102 (100.0%)</td>
</tr>
<tr>
<td>Diels-Alder cycloaddition</td>
<td>106</td>
<td>65 (61.3%)</td>
</tr>
<tr>
<td>Fischer indole</td>
<td>230</td>
<td>94 (40.9%)</td>
</tr>
<tr>
<td>Alkene halogenation</td>
<td>310</td>
<td>281 (90.6%)</td>
</tr>
<tr>
<td>Nitrile hydrolysis</td>
<td>460</td>
<td>460 (100.0%)</td>
</tr>
<tr>
<td>Olefination</td>
<td>455</td>
<td>427 (93.8%)</td>
</tr>
<tr>
<td>Wittig-Horner</td>
<td>211</td>
<td>190 (90.0%)</td>
</tr>
<tr>
<td>Robinson annulation</td>
<td>13</td>
<td>10 (76.9%)</td>
</tr>
</tbody>
</table>

Total: 5,695 reactions, 5,115 correctly reproduced (89.8%)

- Products generated for 5,115 reactions (~90% of the 5,695)

- ~3 seconds per reaction average, 0.015 seconds median run time
Evolutionary Design

Load Molecule → Generate new molecule → Mutate → Score → Rank → Best new molecules

Options:
- Fragment preservation
- How variables
- Memory Policy

Input:
- Loading Molecules
- Column containing parent molecules
- Column containing reaction pathway
- Column containing molecule scores
- Reagents Column

Settings:
- Use reagents from the knowledge base
- Include input structures in output table
- Number of results to return before terminating: 20
- Time out for de novo algorithm (seconds): 15
- Minimum number of reactions to be permitted: 5

Structure generation:
- Generate a maximum of 100 structures per input molecule
- Use tournament selection to generate new structures
- Tournament size: 4
- Allow use of parent with a probability of: 0.1

Duplicates:
- Include the most similar duplicates

Selected Files:

evotec

Open Source KNIME Contributions

http://tech.knime.org/community

Community Nodes
- CDK
- EMBL-EBI
- Erl Wood Cheminformatics
- Groovy Scripting
- KNIME Tools
- Indigo
- KNIME Image Processing
- Matlab Scripting
- NGS
- Palladian
- Python Scripting
- R Scripting
- RDKit
- REST

Regent Court Chemoinformatics
- De Novo Generation
  - De Novo Reaction Vectors Database Reader
  - De Novo Reaction Vectors Database Writer
  - De Novo Structure Generator
- Multi-Objective Molecule Evolution
  - Desirability
  - Multi-Objective Loop End
  - Multi-Objective Loop Start
  - Pareto Ranking
- Vernalis

Erl Wood Cheminformatics
- Activity CIs
  - Activity CIs Viewer
  - Similarity network viewer
- Converters
  - Column Merger
  - Fingerprint Exporter
  - Fingerprin Similarity
  - Virtual Screening Matrics

Docking
- Docking Job Loader
- Docking Job Retriever
- Docking Job Submitter

3D
- Chemical Reactions File Reader
- Test Input

Multi-objective
- Desirability
- Multi-Objective Loop End
- Multi-Objective Loop Start
- Pareto Ranking

Group Analysis
- MCS Distance
- MCS Matrix
- Matched Pairs Detector
- Matched Pairs Finder
- Group Efficiency

Reaction Generation
- Reaction Generator
- Reaction Vectors Database Reader
- Reaction Vectors Database Writer

Viewers
- 2D/3D Scatterplot
- Jmol Docking Pose Viewer
- Jmol Viewer
- Similarity Viewer
- Video Viewer


2D/3D viewer
“Score” = Similarity + D2\text{predAct} + \alpha 1\text{BApredAct} + H1\text{predAct}

\begin{align*}
\text{Q}^2 &= 0.40 \\
\text{Actual} &\quad \text{Predicted} \\
\text{Q}^2 &= 0.68 \\
\text{Actual} &\quad \text{Predicted} \\
\text{Q}^2 &= 0.76 \\
\text{Actual} &\quad \text{Predicted}
\end{align*}

Union of Descriptor

Haloperidol + Ziprasidone

\begin{align*}
\text{Haloperidol} &\quad \text{Ziprasidone} \\
\text{Cl} &\quad \text{H}
\end{align*}
Results: Piperidine

26K Reactions, 93K Reagents
It looks great but ...

1. The algorithm only knows about transformation types that are in the Db!
2. The AP2/3’s cover 1 and 2 bonds. Remote functionality isn’t considered.
3. A reaction path is not a drug “optimisation”!

But how do we get from here to here?
Reaction Sequence Vectors

Tools for molecular design

Sequence Vectors

Molecules Nodes // RVs Edges
Validating Sequence vectors

J. MedChem 2009-2011: 26K reactions

26K => 266K Sequences

Reproducing a => x
SAR Exploration

Succinyl Hydroxamates

Succinyl hydroxamates as potent and selective non-peptidic inhibitors of procollagen C-proteinase: Design, synthesis, and evaluation as topically applied, dermal anti-scarring agents

Simon Bailey a,b,*, Paul V. Fish a,*, Stephane Billotte a, Jon Bordner a, Doris Greiling b, Kim James a, Andrew McElroy a, James E. Mills c, Charlotte Reed c, Robert Webster d

a Department of Discovery Chemistry, Pfizer Global Research and Development, Sandwich Laboratories, Ramgate Road, Sandwich, Kent CT13 9NJ, UK
b Department of Discovery Biology, Pfizer Global Research and Development, Sandwich Laboratories, Ramgate Road, Sandwich, Kent CT13 9NJ, UK
c Department of Pharmacological Sciences, Pfizer Global Research and Development, Sandwich Laboratories, Ramgate Road, Sandwich, Kent CT13 9NJ, UK
d Department of Pharmacokinetics, Dynamics & Metabolism, Pfizer Global Research and Development, Sandwich Laboratories, Ramgate Road, Sandwich, Kent CT13 9NJ, UK

ABSTRACT

Succinyl hydroxamates 1 and 2 are disclosed as novel series of potent and selective inhibitors of procollagen C-proteinase (PCP) which may have potential as anti-fibrotic agents. Carboxamide 7 demonstrated good PCP inhibition and had excellent selectivity over MMPs involved in wound healing. In addition, 7 was effective in a cell-based model of collagen deposition (fibroplasia model) and was very effective at penetrating human skin in vitro. Compound 7 (UK-383,367) was selected as a candidate for evaluation in clinical studies as a topically applied, dermal anti-scarring agent.

![Chemical Structures and Table]

<table>
<thead>
<tr>
<th>Compound</th>
<th>NR1R2</th>
<th>PCP IC50 (nM)</th>
<th>MMP-2 IC50 (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>NH2</td>
<td>44</td>
<td>&gt;50,000</td>
</tr>
<tr>
<td>(S)-7</td>
<td>NH2</td>
<td>&gt;2000</td>
<td>NT</td>
</tr>
<tr>
<td>8</td>
<td>NHMe</td>
<td>28</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>9</td>
<td>NMe2</td>
<td>10</td>
<td>&lt;100,000</td>
</tr>
<tr>
<td>10</td>
<td>NH-pPr</td>
<td>32</td>
<td>21,700</td>
</tr>
<tr>
<td>11</td>
<td>NH-iPr</td>
<td>21</td>
<td>&lt;100,000</td>
</tr>
<tr>
<td>12</td>
<td>NHCH3, cPr</td>
<td>33</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>13</td>
<td>NHCH2, Ph</td>
<td>57</td>
<td>NT</td>
</tr>
<tr>
<td>14</td>
<td>NHCH2, (2-py)</td>
<td>37</td>
<td>&gt;30,000</td>
</tr>
<tr>
<td>15</td>
<td>NHCH2COH</td>
<td>21</td>
<td>&gt;100,000</td>
</tr>
<tr>
<td>16</td>
<td>Pyrrolidine</td>
<td>43</td>
<td>NT</td>
</tr>
<tr>
<td>17</td>
<td>Piperidine</td>
<td>26</td>
<td>74,000</td>
</tr>
<tr>
<td>18</td>
<td>Morpholine</td>
<td>25</td>
<td>NT</td>
</tr>
<tr>
<td>19</td>
<td>4-Methylpiperazine</td>
<td>64</td>
<td>NT</td>
</tr>
<tr>
<td>20</td>
<td>NMeCH3, Ph</td>
<td>65</td>
<td>81,800</td>
</tr>
<tr>
<td>21</td>
<td>NMeCH2(2-py)</td>
<td>40</td>
<td>50,500</td>
</tr>
</tbody>
</table>
Novel SAR

Succinyl Hydroxamates
Principle Components Analysis of Property Space

Succinyl Hydroxamates

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

Legend
- Known products
- Near neighbours (Tanimoto 1.0-0.8)
- All other products
Mapping Discovery Space

Sequence vector network

1M+ reactions from US Patent Database: Nextmove
The PGC GWAS
Genome Wide Association Studies
GPCRs associated with Schizophrenia GWAS genes

2 step shortest paths network

- Schizophrenia GWAS genes
- GPCR receptor subtypes

Eg, From: Dopamine receptor subtypes
To: Proteins defined by PGC2 schizophrenia GWAS genes

- Dopamine receptor network
- Adenosine receptor network
- 5HT receptor network
- Histamine receptor network
- Adrenergic receptor network
- Opioid receptor network
- Metabotropic glutamate receptor network
Hubs in the GPCR schizophrenia network

Proteins identified by schizophrenia GWAS
- GPCRs and GPCR signalling proteins

Hubs involved in GPCR signalling in schizophrenia

Remove proteins degree <10
Alzheimer’s Disease Neuropathology

- Brain Atrophy
  - Aβ peptide
  - Extracellular deposits

- β-Amyloid plaques
  - Aβ peptide
  - Extracellular deposits

- Neurofibrillary tangles
  - Tau protein
  - Intraneuronal filamentous inclusions
Alzheimer’s Disease Network

Canonical network
Network Based Design

1. Disease Hypothesis

2. Identify cmpds with known/predicted selectivity


4. Relate chemistry to gene Bioprint.

5. Network Enrichment. Novel target identification

6. Off-Target Hypothesis
In-Silico Network Based Design

Connecting Gene Expression Data from Connectivity Map and In Silico Target Predictions For Small Molecule Mechanism-of-Action Analysis

Bender et al., Molecular BioSystems 2014 – accepted
The Drug Discovery Challenge

Summary

Given a disease signature how do we best sample appropriate chemistry space?
Acknowledgments

Val Gillet
Beining Chen
Hina Patel
Ben Allen
James Wallace

Dimitar Hristozov
David Thorner
David Evans
Nik Fechner
George Papadatos
Suzanne Brewerton

Andreas Bender
Georgios Drakakis

John Liebeschuetz
Jason Cole
Your contact:
Mike Bodkin
VP Computational Chemistry & Cheminformatics
114 Innovation Drive,
Milton Park, Abingdon
Oxfordshire OX14 4RZ, UK
T: +44 (0)1235 44 1207
Mike.Bodkin@evotec.com