Guiding Compound Design in 2, 3… N Dimensions

Streamlining Drug Discovery and Development – Boston, April 11th 2016
Matthew Segall - matt.segall@optibrium.com
Overview

• 2-Dimensions
  – Qualitative SAR: Activity cliffs, matched molecular pair analysis...
  – Quantitative SAR: Predictive models

• 3-Dimensions
  – Structure-based design
  – Scoring/affinity prediction
  – Understanding 3D SAR

• N-Dimensions
  – Multi-parameter optimisation

• Linking 2D, 3D,... ‘N’D SAR to guide design

• Conclusions
2-Dimensions
Many methods are routinely used for analysis of data to reveal patterns and trends to guide compound optimisation, e.g.

- **Clustering**
  - Group ‘similar’ compounds to identify series with interesting SAR

- **Activity cliff detection**
  - Small changes in structure that cause a large change in activity

- **Matched molecular pair analysis**
  - Pairs of compounds that are identical except for one small change at the same position
• Freedom from the constraints of ‘chemical spreadsheets’
  – Represent compound relationships

• Work the way you think
  – Cards: Display key compound data
  – Stacks: Summarise and compare data for groups of compounds
  – Links: Highlight compound relationships

• Intuitive visualisation of analyses
  – Clustering, activity cliffs, matched molecular pairs...

• Quickly identify optimisation strategies
Activity Neighbourhood
Activity Cliff Visualisation

Matched Molecular Pair Analysis

Quantitative Structure-Activity Relationships

Principles

\[ y = f(x_1, x_2, x_3, \ldots) \pm \varepsilon \]

- Data
  - Quality data is essential
  - Public data needs very careful curation (and may not be good enough)

- Descriptors, e.g.
  - Whole molecule properties, e.g. logP, MW, PSA...
  - Structural descriptors, SMARTS, fingerprints...

- Statistical fitting or machine learning method, e.g.
  - Partial least squares, artificial neural networks, support vector machines, random forest, Gaussian processes...

- Widely applied to prediction of ADME and physicochemical properties
Interactive Redesign

- QSAR models provide estimates of compounds’ properties

- Instant feedback on how properties are likely to change
  - Explore strategies for redesign

- But, important questions
  - “Why is a property value predicted?”
  - “Where can I change this property?”

- Glowing Molecule™:
  - Visual indication of structural influences on predicted properties

3-Dimensions
Visual Understanding of 3D Affinity Data
HYDE: A Different View @ Energetics*

- Physics only
- No calibration to complexes
- Relates to real Free Energies

Please see Carsten Detering’s talk for more details
HYDE Scoring Function – Concept

\[ \Delta G_{\text{HYDE}}^i = \sum_{\text{atom } i} \Delta G_{\text{dehydration}}^i + \Delta G_{\text{H-bond}}^i \]

*Reulecke et al., ChemMedChem’08*
Hyde - Visual Affinities

HYDE color code:

- **+ΔG contribution**
- **-ΔG contribution**
- **no ΔG contribution**

**HYDE Color Code Values**

- **receptor carbonyl oxygen**: 8.2 kJ/mol
- **ligand aromatic oxygen**: 2.4 kJ/mol
- **total desolvation cost**: 10.6 kJ/mol

- **receptor aromatic carbons**: -5.2 kJ/mol
- **ligand aromatic carbon**: -2.0 kJ/mol
- **total desolvation gain**: -7.2 kJ/mol

- **receptor amide N dehydrat**: 6.3 kJ/mol
- **interaction energy**: -7.4 kJ/mol
- **ligand aromatic N dehydrat**: 6.4 kJ/mol
- **interaction energy**: -7.5 kJ/mol
- **total H-bond energy**: -2.2 kJ/mol
Visual Understanding of 3D Affinity Data

PDB: 1GKC
N-Dimensions
The Objectives of Drug Discovery
Multi-parameter optimisation

- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible
  - Fail fast, fail cheap
  - Only when **confident**

Multi-Parameter Optimisation
Probabilistic Scoring

Integrated assessment of data against project criteria
Accounts for the uncertainties in all compound-related data (experimental or calculated)

Project specific scoring profile

Histories for quick visual guide to compound properties

Multi-Parameter Optimisation
Probabilistic Scoring

- Property data
  - Experimental or predicted
- Criteria for success
  - Relative importance
- Uncertainties in data
  - Experimental or statistical

- Score (Likelihood of Success)
- Confidence in score

Data do not separate these as error bars overlap

Error bars show confidence in overall score

Bottom 50% may be rejected with confidence

Probabilistic Scoring
Guide redesign to improve chance of success

Linking 2D and 3D SAR to Guide Design
Understanding Activity Cliffs in 3D
PPARγ PDB 4EMA
Understanding Activity Cliffs in 3D

PPARγ PDB 4EMA
Understanding Activity Cliffs in 3D

PPARγ PDB 4EMA
Understanding Activity Cliffs in 3D

PPARγ, PDB 4EMA
Understanding Activity Cliffs in 3D
PPARγ PDB 4EMA
Exploration of Virtual Screening Results
HSP90

- Crystal structure PDB ref. 2XJX

- Virtual library generated using STORM workflow in KNIME
  - Amide substitution using Schotten Baumann reaction on beta resorcylic core
  - Building blocks from vendor catalogues
  - ‘Tail’ of molecule not contributing to affinity

- Resulting library docked with FlexX

- Scored using SeeSAR and HYDE to estimate $pK_i$
Matched Molecular Pair Analysis
Matched Molecular Pair Analysis
Matched Molecular Pair Analysis
Matched Molecular Pair Analysis
HYDE Analysis in SeeSAR

Row 731

Row 208

Hyde pKi e...: 8.142

Hyde pKi e...: 4.138
Combine with 2D QSAR Predictions
Multi-parameter optimisation

![Graphical User Interface](image-url)

**Profile:** Hyde pKi + Oral Non CNS Scoring Profile

- **Hyde pKi estimate:** > 7
- **logS:** > 1
- **HI4 category:** +
- **logP:** 0 -> 3.5
- **HERG pIC50:** % 5
- **2Daffinity category:** low medium
- **2C9 pKi:** % 6
- **P-gp category:** no
- **PBE category:** low
- **BBB category:** -
- **BBB log[brain/blood]:** % -0.5

**Available Properties:**
- HERG pIC50
- HI4 category
- BBB log[brain/blood]
- NIF

**Scoring Profiles:**
- Oral Non CNS Scoring Profile
- Oral CNS Scoring Profile
- Lipinski Rule of Five
- Intravenous Non CNS Scoring Profile
- Intravenous CNS Scoring Profile
- Hyde pKi + Oral Non CNS Scoring ...

**Data Set:**
- HSP90 ligands

**Columns:**
- Hydrogen bond
- Oral CNS
- Hydrogen bond
- Herceptin
- BBB log[brain/blood]
- pKi
3D View... Optimisation opportunities
HYDE Analysis in SeeSAR
Optimisation Idea
Modify pKa of Nitrogen by amide substitution
Optimisation Idea
Add polar group to phenyl ring
Repose in SeeSAR
Conclusions

- Both 2D and 3D information are important to interpret SAR and guide design
- A seamless combination between these two views of the chemical world maximises the benefits that they bring
  - Understanding SAR from experimental data
  - Analysis of virtual screening/docking results
- Multi-parameter optimisation of potency, physicochemical and ADMET properties
  - Quickly target high quality compounds
Acknowledgements

• BioSolve IT
  – Christian Lemmen
  – Markus Gastreich
  – Carsten Detering

• Optibrium
  – Peter Hunt
  – James Chisholm
  – Ed Champness
  – Chris Leeding
  – Nick Foster
  – Tamsin Mansley
  – Alex Elliot
  – Sam Dowling
  – Fayzan Ahmed
  – Many more…. 