Activity Landscapes

Jürgen Bajorath
Department of Life Science Informatics
University of Bonn
Activity Landscapes

Chemoinformatics methods for landscape design

Navigation and Interpretation

Medicinal chemistry applications
‘Activity landscape’ - generally defined as any graphical representation that integrates molecular similarity and potency relationships between active compounds.
Concept of Activity Landscapes

- Often idealized as a 2D projection of chemical space with a compound potency surface added in the third dimension
Idealized Activity Landscapes and SARs

**Continuous SAR**
- gradual changes in structure result in moderate changes in activity
- “rolling hills”

**Discontinuous SAR**
- small changes in structure have dramatic effects on activity
- “activity cliffs”
Calculated 3D Activity Landscapes

Squalene synthase inhibitors

- Reference space 2D projection
  - MACCS Euclidean distances
  - Multidimensional scaling
3D Surface Generation

1. Add potency as 3rd dimension

2. Interpolate on a regular grid

3. Generate 3D surface

Squalene synthase inhibitors
Influence of Molecular Representations

MACCS

TGT

Molprint2D

1. 0.02 nM
2. 590 nM
3. 0.4 nM
4. 32 μM
Molecular Network-Based Landscapes

Network-like Similarity Graph (NSG)
Annotated graph of similarity relationships in compound data sets

Designed to explore **global** and **local** SAR features in data sets

**Global scores**
- continuity: 0.79
- discontinuity: 0.99
- SAR Index: 0.40
NSG Information Layers

Edges

\[ T_c > x \]

\[ T_c < x \]

Calculated fingerprint Tanimoto similarity, e.g. 75 %

1. similarity relationships
2. potency distribution
3. compound discontinuity scores
NSG Information Layers

Node color

1. similarity relationships
2. potency distribution
3. compound discontinuity scores
NSG Information Layers

Node size

- high compound score
- low compound score

1. similarity relationships
2. potency distribution
3. compound discontinuity scores
NSG Information Layers

Graph layout: Fruchterman-Reingold

1. similarity relationships
2. potency distribution
3. compound discontinuity scores
Globally Discontinuous SAR

Thrombin inhibitors

Densely connected clusters with large-magnitude activity cliffs

Global scores:
- continuity: 0.081
- discontinuity: 0.665
- SAR Index: 0.208
Heterogeneous SAR

Squalene synthase inhibitors

activity cliff

continuous local SAR

discontinuous local SAR

Global scores
continuity 0.79
discontinuity 0.99
SAR Index 0.40
3D vs. 2D Activity Landscapes

Squalene synthase inhibitors

Activity cliff region

Interpolated area

Smooth region

NSG cluster discontinuity: 1.00

NSG cluster discontinuity: 0.01
SAR Landscapes of Evolving Data Sets
Adenosine A2 Receptor Ligands: 2001

SAR discontinuity

SAR continuity
Adenosine A2 Receptor Ligands: 2003

increasingly potent compounds

flat SAR
Adenosine A2 Receptor Ligands: 2005
Adenosine A2 Receptor Ligands: 2007

new chemotypes explored?
NSG Extensions: Selectivity Landscapes

- **Target-pair selectivity:**
  - potency ratios
  - logarithmic potency differences
  \[ S_{A/B}(i) = P_A(i) - P_B(i) \]

- **Selectivity NSGs**

- **From activity cliffs to selectivity cliffs**
  - pIC50 = 9
  - pIC50 = 7
  - \( S_{L/B} = 2 \)
Activity Landscape

Potency-based NSG

Potency:

10.4  3.0

Compound discontinuity score:

1  activity cliff markers
0

Cluster discontinuity score

1  “rough” SAR
0  “smooth” SAR
Activity Landscape Comparison

cathepsin L

0.48

0.05

cathepsin B

0.10

0.27
Selectivity Landscape

Selectivity-based NSG

Selectivity:

3.2 (L) – 3.2 (B)

Compound discontinuity score:

- 1 selectivity cliff markers
- 0

Cluster discontinuity score

- 1 “rough” SSR
- 0 “smooth” SSR

Structure-Selectivity Relationship (SSR)
Local Environments

cathepsin L / cathepsin B

0.73

discontinuous SSR

0.72
Activity Cliffs vs. Selectivity Cliffs

L: 3.6 μM
B: 102 μM
L/B: 1.5

L: 26 μM
B: 5.3 μM
L/B: -0.7

selectivity cliff markers
Selectivity Determinants

- Selectivity determinants are often found in selectivity cliff environments
- Selectivity rules can be derived

Halogen groups with increasing bulkiness and decreasing electronegativity shift selectivity toward cat L.
Molecular Mechanism-Based NSG (M-NSG)

- Color code:
  - Mechanism
    - Antagonist (94)
    - Partial agonist (54)
    - Agonist (107)
    - Inverse agonist (52)

Adenosine A1 receptor ligands
Molecular Mechanism-Based NSG (M-NSG)
M-NSG ‘Mechanism Hopping’ Regions
M-NSG ‘Mechanism Hopping’ Regions

Exemplary ‘Mechanism Hops’
Structural Neighborhoods in Data Sets

**Structural neighborhood**
all compounds that are more similar to a *reference compound* than a similarity threshold $T$
Similarity–Potency Trees (SPTs)

Compound-centric view of an activity landscape

SPTs systematically organize structural neighborhoods
SPT Interpretation

reference compound

nearest neighbor relationship

vertical series of compounds spanning multiple nearest neighbor relationships

(Factor Xa inhibitors)

(pKᵢ)

11.4

4.5

decreasing similarity to root compound

horizontal different nearest neighbors of the same compound
NSG-SPT Analysis

potency (node color)  discontinuity (node size)

high
low

vertical pattern

horizontal pattern

decreasing similarity to root compound

nearest neighbor relationship

Rank 1

Rank 15
Screening Data Analysis

Data

- Anti-malarial screening hits (GSK)
- ~13,500 active compounds
- phenotypic assay (parasite growth inhibition)
- estimated EC$_{50}$ values (based on percent inhibition)
- tested and made publicly available by GlaxoSmithKline
- Gamo et al., Nature 465, 305, 2010
Most prominent regions were selected from the NSG and subjected to SPT analysis.
SPT Analysis of Series 1

Horizontal potency patterns

“Clustering” of compounds with similar potency levels

Well-defined SAR patterns: all series selected for the initial evaluation consisted of previously known anti-malarial compounds
Removal of Known Active Compounds

Known active compounds* and similar structures were removed (2914 compounds representing 1186 scaffolds)

(*from BindingDB/ChEMBL)
SPT Analysis of Series 1

- Absence of highly potent compounds
- Horizontal and vertical patterns detectable
- Clustering of similarly potent compounds

Novel active compounds - potency distribution more characteristic of screening hits
Summary

Activity landscapes
- data-driven analysis
- focus on SAR visualization
- grid- and graph-based designs

No pre-defined SAR models
- structure and activity are independent parameters
- SAR patterns are an emergent property
Summary

Advanced SAR analysis
- large data sets: global and local SAR views
- SAR monitoring of data sets evolving over time
- selectivity landscapes
- mode-of-action analysis and mechanism hopping
- compound-centric SAR environments