Emerging Big Data: Chemoinformatics-Driven View of Kinase Drug Discovery

Jürgen Bajorath
Life Science Informatics
University of Bonn
Protein Kinases

- **Kinases** and G protein coupled receptors currently are the most intensely investigated drug targets.

- **Kinases** play a key role in native and aberrant signaling pathways and are prime targets in oncology, immunology/inflammation etc.
Inhibitor Coverage of the Kinome

- Human kinome comprises 518 kinases
- Inhibitors with high-confidence data are available for 286 kinases

**Ki measurements**

**IC$_{50}$**

![Kinome Tree](image)
Kinase Inhibitors

- **Scaffolds** of structurally diverse and potent inhibitors

<table>
<thead>
<tr>
<th></th>
<th>$K_i$ subset</th>
<th>$IC_{50}$ subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 compounds</td>
<td>PI3K alpha</td>
<td>9.57 – 10.40</td>
</tr>
<tr>
<td>6 compounds</td>
<td>PI3K alpha</td>
<td>8.72 – 9.33</td>
</tr>
<tr>
<td>6 compounds</td>
<td>MAP kinase p38 alpha</td>
<td>9.22 – 9.70</td>
</tr>
<tr>
<td>5 compounds</td>
<td>Ser/Thr protein kinase PIM3</td>
<td>9.00 – 9.52</td>
</tr>
<tr>
<td>10 compounds</td>
<td>HGF receptor</td>
<td>8.05 – 9.00</td>
</tr>
<tr>
<td>9 compounds</td>
<td>GSK-3 beta</td>
<td>8.10 – 8.70</td>
</tr>
<tr>
<td>6 compounds</td>
<td>MAP kinase p38 alpha</td>
<td>8.14 – 8.60</td>
</tr>
<tr>
<td>6 compounds</td>
<td>CSF1R</td>
<td>8.60 – 9.52</td>
</tr>
</tbody>
</table>
Big Compound Data

- Compound activity data from biological screening and medicinal chemistry increasingly meet Big Data criteria

*The ‘7 Vs’*

<table>
<thead>
<tr>
<th>Volume</th>
<th>Heterogeneity (across databases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>Complexity (multi-layered data structures)</td>
</tr>
<tr>
<td>Variety</td>
<td>Confidence (experimental stringency)</td>
</tr>
<tr>
<td>Veracity</td>
<td></td>
</tr>
<tr>
<td>Variability</td>
<td>(Hu &amp; Bajorath, 2017)</td>
</tr>
<tr>
<td>Visualization</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td></td>
</tr>
</tbody>
</table>

(Van Rijmenam, 2013)

- Kinase inhibitors are a representative example
Activity Data Confidence Levels

Low-confidence set
activity against human targets

All available activity annotations

Medium-confidence set
+ highest assay relevance

Activity data with:
- highest assay relevance & reliability

High-confidence set
+ highest measurement reliability

Activity data with:
- highest assay relevance & reliability
- highest measurement reliability
Data Growth

- Kinase inhibitors with **high-confidence** activity data

<table>
<thead>
<tr>
<th>ChEMBL 23</th>
<th>ChEMBL 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>2015</td>
</tr>
<tr>
<td>• 45,728 kinase inhibitors</td>
<td>• 18,951 kinase inhibitors</td>
</tr>
<tr>
<td>• 286 kinases</td>
<td>• 266 kinases</td>
</tr>
<tr>
<td>• 12 kinase groups</td>
<td>• 10 kinase groups</td>
</tr>
</tbody>
</table>
Kinase Inhibitors

- More than 95% of current kinase inhibitors target the largely conserved ATP (cofactor) binding site (type I)

Gavrin & Saiah
Kinase Inhibitors

- Type I inhibitors are expected to be **promiscuous** (active against multiple kinases)

Gavrin & Saiah
Kinase Inhibitors

- **Promiscuity degrees** (high-confidence data)

<table>
<thead>
<tr>
<th>ChEMBL 23 2017</th>
<th>ChEMBL 18 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 45,728 kinase inhibitors</td>
<td>• 18,951 kinase inhibitors</td>
</tr>
<tr>
<td>• 286 kinases</td>
<td>• 266 kinases</td>
</tr>
<tr>
<td>• 12 kinase groups</td>
<td>• 10 kinase groups</td>
</tr>
<tr>
<td>• 33,157 (76.5%) single-kinase inhibitors</td>
<td>• 14,892 (78.6%) single-kinase inhibitors</td>
</tr>
<tr>
<td>• 504 (1.2%) with activity against ≥ 5 kinases</td>
<td>• 353 (1.9%) with activity against ≥ 5 kinases</td>
</tr>
</tbody>
</table>
Kinase Inhibitors

- Low detectable promiscuity due to **data sparseness?**

<table>
<thead>
<tr>
<th></th>
<th>ChEMBL 23 2017</th>
<th>ChEMBL 18 2015</th>
</tr>
</thead>
</table>
| Number of kinase inhibitors | • **45,728** kinase inhibitors  
                          | • 286 kinases  
                          | • 12 kinase groups                                                                 |
| Single-kinase inhibitors | • **33,157 (76.5%)** single-kinase inhibitors  
                          | • 504 (1.2%) with activity against ≥ 5 kinases                                |
| Activity against ≥ 5 kinases | • **14,892 (78.6%)** single-kinase inhibitors  
                          | • 353 (1.9%) with activity against ≥ 5 kinases                                |
Iterative removal of confidence criteria leads to increasing data volumes and kinome coverage.
Extensively Assayed Compounds

- 437,257 screening compounds assembled from PubChem BioAssays tested in both primary and confirmatory assays (>800 targets)

Mean: 411
Median: 437
Kinase Inhibitors - Promiscuity

- Subset of **437,257 extensively assayed screening** compounds contains **28,172 inhibitors** of a total of **43 human kinases**

<table>
<thead>
<tr>
<th><strong>Extensively assayed screening compounds</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># Kinase inhibitors</td>
<td>28,172</td>
</tr>
<tr>
<td># Kinase targets</td>
<td>43</td>
</tr>
<tr>
<td>Promiscuity degree (PD)</td>
<td>1 – 11</td>
</tr>
<tr>
<td>PD [mean]</td>
<td>1.1</td>
</tr>
<tr>
<td># Kinases [tested]</td>
<td>1 – 23 (mean 14)</td>
</tr>
</tbody>
</table>
Kinase Inhibitors - Promiscuity

PD [mean] 1.1
Type I inhibitors are expected to have limited **selectivity** for individual kinases – **promiscuous** inhibitors should be **least selective**

Gavrin & Saiah
Kinase Inhibitors - Selectivity

- **Promiscuous** human kinase inhibitors with high-confidence data

ChEMBL 23

Inhibitors active against 2 or more human kinases

10,060

Active against 266 kinases

Formation of **target pairs that share ≥ 10 inhibitors**

\[
\text{Kinase}_1 - \text{Kinase}_2: \geq 10 \text{ shared inhibitors} \\
\text{Kinase}_2 - \text{Kinase}_3: \geq 10 \text{ shared inhibitors} \\
\text{...}
\]

596 pairs 141 kinases
Kinase Pair Categories

- 596 protein kinase pairs sharing at least 10 inhibitors (median value of 18 compounds per pair)

- 3 different pair categories according to their phylogenetic distance

1-3

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>same family</td>
<td>132 pairs</td>
</tr>
<tr>
<td>(ii)</td>
<td>different families</td>
<td>262 pairs</td>
</tr>
<tr>
<td>(iii)</td>
<td>different groups</td>
<td>202 pairs</td>
</tr>
</tbody>
</table>

Selectivity Analysis

- By kinase pair-based analysis, all possible selectivity relationships for kinases and inhibitors were quantified.

- For each kinase pair, potency differences ($\Delta pIC_{50}$) between shared inhibitors were calculated as a measure of selectivity.
Kinase Inhibitors - Selectivity

- Small global potency differences (median $\Delta pIC_{50}$ value 0.64)
- Different result for most selective inhibitors per pair (median 2.37)

<table>
<thead>
<tr>
<th>Potency difference ($\Delta pIC_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

All compounds 23,719 pairs

Most selective compounds 596 pairs

> 100-fold difference
Kinase Inhibitors - Selectivity

- Similar differences for pairs with increasing phylogenetic distances

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>median ΔpIC$_{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>same family</td>
<td></td>
</tr>
<tr>
<td>(ii)</td>
<td>different families</td>
<td></td>
</tr>
<tr>
<td>(iii)</td>
<td>different groups</td>
<td></td>
</tr>
</tbody>
</table>

Graph showing the median ΔpIC$_{50}$ for compounds grouped by similarity: 
- All compounds
- Most selective compounds
Selectivity Analysis

- **Pair-based selectivity profiles**: recording inhibitors with largest potency differences per kinase pair

- **Compound-based selectivity profiles**: ordering available inhibitors according to increasing potency differences between paired kinases
Category-Based Selectivity Profiles

- More than 50% of kinase pairs from each category have inhibitors with more than 100-fold potency differences

<table>
<thead>
<tr>
<th>Same family</th>
<th>Target Pairs</th>
<th>Potency difference ($\Delta IC_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Different families</th>
<th>Target Pairs</th>
<th>Potency difference ($\Delta IC_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Different groups</th>
<th>Target Pairs</th>
<th>Potency difference ($\Delta IC_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>
Kinase Inhibitors - Selectivity

- **Unexpected selectivity** among **promiscuous** kinase inhibitors

<table>
<thead>
<tr>
<th>Same family</th>
<th>Target Pairs</th>
<th>Potency difference ($\Delta pIC_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different families</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Different groups</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>
Compound-Based Selectivity Profiles

Kinases from the same family

### Protein kinase pair | Gatekeeper
---|---
PKCh | PKCt | M|M

<table>
<thead>
<tr>
<th>Least selective</th>
<th>Most selective</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Least selective structure" /></td>
<td><img src="image2" alt="Most selective structure" /></td>
</tr>
</tbody>
</table>
Compound-Based Selectivity Profiles

Kinases from different families

<table>
<thead>
<tr>
<th>Protein kinase pair</th>
<th>Gatekeeper</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMS</td>
<td>LCK</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
</tbody>
</table>

Potency

Least selective

Most selective
Compound-Based Selectivity Profiles

Kinases from different groups

Protein kinase pair | Gatekeeper
---|---
PDK1 | AurA | L|L

Potency

Least selective

Most selective
Profiling Clinical Kinase Inhibitors

- Most comprehensive kinase inhibitor profiling study currently available\(^1\)
- Cell-based profiling of 243 clinical kinase inhibitors
- “Kinobeads” assays, followed by qMS
- Cellular interactions with 253 human kinases detected

Chemoinformatics perspective:
Reconciling observed promiscuity vs. selectivity trends on the basis of compound activity data from medicinal chemistry

\(^1\)Klaeger et al. Science 2017, 358, eaan4368.
Activity Data

202 inhibitors with human kinase annotations

Activity data
- No activity threshold
- Confidence level 1
- Confidence level 2
- Activity < 10 μM
  - Confidence level 1
  - Confidence level 2

ChEMBL23

Klaeger et al.

243 kinase inhibitors

Analysis

1. Data confidence levels and promiscuity degrees
2. Most and least selective inhibitors
3. Inhibitors with different binding modes
4. Inhibitors designated as chemical probes
Promiscuity Degrees (PDs)

Highest Assay Confidence + Highest Measurement Confidence

No activity threshold
Confidence level 1: medium
Confidence level 2: high

Activity < 10 μM
Confidence level 1
Confidence level 2

166 inhibitors
Confidence Level 2

185 inhibitors
Confidence Level 1

164 inhibitors
Activity <10 μM

172 inhibitors
Activity <10 μM
Promiscuity Degrees (PDs)

Strong influence of data confidence criteria
Small median PD values

No activity threshold
Confidence level 1: medium
Confidence level 2: high

Activity < 10 μM
Confidence level 1
Confidence level 2

166 inhibitors
185 inhibitors
164 inhibitors
172 inhibitors
**Kinome Coverage**

- Cell-based profiling: 253 of 363 expressed human kinases
- 122 kinases (< 10 μM: 122)
- 394 kinases (< 10 μM: 379)

**Confidence levels**
- Confidence level 1: medium
- Confidence level 2: high

**Activity threshold**
- Activity < 10 μM

**Inhibitors**
- Confidence Level 2: 166 inhibitors
- Confidence Level 1: 185 inhibitors
- Activity < 10 μM: 164 inhibitors
- Activity < 10 μM: 172 inhibitors

**Compounds**
- Confidence Level 2: 3 compounds
- Confidence Level 1: 7 compounds
- Activity < 10 μM: 3 compounds
- Activity < 10 μM: 4 compounds
Kinase Pair-Based Selectivity Analysis

Klaeger et al.

Kinobeads assays
Apparent dissociation constants
\[ K_d \Rightarrow pK_d \]

Formation of target pairs that share \( \geq 10 \) inhibitors

- \( \text{Kinase}_1 - \text{Kinase}_2: \geq 10 \) shared inhibitors
- \( \text{Kinase}_2 - \text{Kinase}_3: \geq 10 \) shared inhibitors
- ...

2369 pairs
190 inhibitors
137 kinases

216 inhibitors
225 kinases
Kinase Inhibitors - Selectivity

- Small global potency differences:
  \( \text{median } \Delta pK_d \text{ value } 0.67 \) (med. chem. data: \( \text{median } \Delta pIC_{50} 0.69 \))

- Different picture for most selective inhibitors:
  \( \text{median } \Delta pK_d \text{ value } 2.24 \) (med. chem. data: \( \text{median } \Delta pIC_{50} 2.37 \))
Compound-Based Selectivity Profiles

- Uni-directional selectivity profiles revealed inhibitors with exclusive selectivity for one kinase over the other

Compounds are ordered according to increasing potency differences for each kinase from the left to right and vice versa.
**Compound-Based Selectivity Profiles**

- **Bi-directional** profiles uncovered inhibitors with inverted selectivity for paired kinases

### Kinase Profiles

**Same family**
- Kinase A: PRKAA1, SIK2
- Kinase B: PF-3758309 (ΔpK_d = 2.32)

**Different families**
- Kinase A: PDGFRB, YES1
- Kinase B: AZD-7762 (ΔpK_d = 2.05)

**Different groups**
- Kinase A: ABL1, TGFBR1
- Kinase B: Sunitinib (ΔpK_d = 2.32), TAK-901 (ΔpK_d = 2.31), Dasatinib (ΔpK_d = 2.56), CP-547632 (ΔpK_d = 2.64)
Conclusions

- Advent of the ‘big data‘ era in medicinal chemistry
- Kinase inhibitors as a representative example
- Most kinase inhibitors target the largely conserved ATP binding site
- Promiscuity and lack of selectivity are anticipated
- Systematic activity data analysis is strongly influenced by data confidence criteria
- Varying confidence criteria and inclusion of screening data put the data sparseness issue into perspective
Conclusions

- Data analysis reveals that ATP site-directed kinase inhibitors are less promiscuous and more specific than often assumed.

- In part surprising agreement of computational selectivity analysis on the basis of activity data from medicinal chemistry and cell-based profiling.