Iterative Medium Throughput Screening for Kinase Lead Discovery

Problem: Experimental high-throughput screening of a large (1.5 million) compound collection typically takes 6 months and costs $1,000,000.

Alternative: Iterative MTS screening using structural, IC50, EC50 data driven, adaptive, modeling methods that treat each new kinase as a member of a family, rather than an idiosyncratic new target.

Eric Martin*, David Sullivan and Prasenjit Mukherjee
Novartis Institutes for BioMolecular Design, Emeryville, CA


**Hypothetical Experimental “profile-QSAR” Model for IC₅₀ Prediction**

Extra Info from data on many cpds across many kinases

\[ aK_1 + bK_2 - cK_3 + \ldots + zzK_{70} = K_{71}, K_{72}, \ldots, K_{518} \]
Computational 2D Profile-QSAR Models for a New Kinase

1) Can’t do experimental regression model because 90% missing values,
2) Dumb to measure 70 activities to predict 1?
so:
1) Build individual Bayesian QSARs for each of 70 kinases,
2) Use predicted activity probability profile as chemical descriptor for a profile QSAR.

Resulting profile-QSAR models predict activity much better than individual models!
2D Profile-QSAR Models: Combine data across Kinase Family

A PLS model combining Bayesian models across 70 enzymes correlates with activity much better than a single Bayes Model alone.

"Profile" and simple QSAR models

R2 (25% held-out)

cRAF, PAR1bMARK2, bRafmutV599E, TAK1, FLT3, GSK3b, Tie2, PAK4, bRAF, CDC7, p13Kb, EphA2, PIM1, CK1g2, PKCe, CTK

BayesProfileR2, BayesR2
Cellular Activity Prediction with profile-QSAR models

- 2D Profile QSAR model
  - PLS using kinase enzyme and kinase cellular activity bayesian predictions as descriptors
  - Including physical properties did not help: MW, ClogP, PSA, Rot_bnds
- Model validation
  - Leave 20% out Cross Validation
  - 25%External test set
Profile QSAR for 32 Cellular assays
Kinase Ensemble Surrogate “AUTOSHIM” Docking

2D profile-QSAR models rely on topological similarity. 3-D Docking can make larger “scaffold hops”

Docking suffers from 3 serious limitations:
1) It requires a protein structure.
2) It is slow.
3) It cannot predict affinity.
3-D Docking: *All-Purpose* scoring functions can’t predict affinity.

**GlaxoSmithKline Docking Study**

“there is no statistically significant correlation between measured affinity and any of the [37] scoring functions [from 10 docking programs] evaluated across all eight protein targets examined (Table 7). An extremely modest positive correlation was observed for Chk1 \( r^2 = 0.32 \).”

Can we use available IC\(_{50}\) data to make target specific *predictive* scoring functions?

Let \( r^2 = 0.32 \) be our criterion for success.

**Magnet:** An expert system for target-tailored scoring functions

Replace visual review with “HTI” (High Throughput Intuition)

- Specific H-bonds
- Atom counts filling critical pockets
- Substructures (SMARTS) in pockets
- % of buried ligand surface area
- Hydrophobic/Hydrophilic contact area
- Large contact area

Hanneke Jansen wrote original prototype in *Sybyl*. Partner with Metaphorics. Scott Dixon wrote SEA interaction def. language.

E.g. CSF1R features
**PLS/Magnet:** Parameterized Magnet with *Pose Selection*

Assay Data *(CSF1R: 2,149 compounds & IC50 values)*

Dock *(ensemble of 2 CSF1R X-rays)*

102,610 poses – Minimize & score in Flo+ *(R²=0.27)*

Docking

Extract Magnet features

Train Set

- 75% (1,612 cpds)
- (76,719 poses)

Test Set

- 25% (537 cpds)
- (25,891 poses)

Score and keep 1 best pose / cpd

Score of best pose is activity prediction

Pose selected on best score, *not* best Q²!

**Docking**

**QSAR**

Train PLS model on best pose & IC₅₀ data

*calc. predictive R²*

CSF-1R: Predictive R² = 0.50
(bar was R² = 0.32)
"AutoShim": Automated "Shim" generation
Adjustable Pharmacophore-like Shims in an NMR Magnet

**PDK1 Positive Shims**

- **HBD, HBA, Aromatic, Non-polar**

**Recursive Partitioning (RP): Multi-points and thresholds**

- HBA at location 1?
- !Polar at 2?
- HBD at 1?
- Flo score > 4.32?
- !Polar at 3?

### Additional AUTOSHIM Models

<table>
<thead>
<tr>
<th>Kinase</th>
<th>Shims</th>
<th>Pred. $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDK1 (8X)</td>
<td>1PT</td>
<td>0.61</td>
</tr>
<tr>
<td>FGFR</td>
<td>1PT</td>
<td>0.55</td>
</tr>
<tr>
<td>Chk1</td>
<td>1PT</td>
<td>0.40</td>
</tr>
<tr>
<td>Pim1</td>
<td>1PT</td>
<td>0.17</td>
</tr>
<tr>
<td>Chk1</td>
<td>RP</td>
<td>0.58</td>
</tr>
<tr>
<td>CSF1R (2X,50K)</td>
<td>RP</td>
<td>0.55</td>
</tr>
<tr>
<td>PDK1 (1X, 50K)</td>
<td>RP</td>
<td>0.52</td>
</tr>
<tr>
<td>AurA</td>
<td>RP</td>
<td>0.42</td>
</tr>
<tr>
<td>GSK3b</td>
<td>RP</td>
<td>0.41</td>
</tr>
</tbody>
</table>

**CSF1R: 2 X-ray Ensemble**

<table>
<thead>
<tr>
<th>Method</th>
<th>Pred. $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Docking (Flo+)</td>
<td>0.27</td>
</tr>
<tr>
<td>PLS on Hand Magnet:</td>
<td>0.50</td>
</tr>
<tr>
<td>AutoShim (1pt Shims)</td>
<td>0.50</td>
</tr>
<tr>
<td>AutoShim (RP Shims)</td>
<td><strong>0.58</strong></td>
</tr>
</tbody>
</table>
**“Surrogate AutoShim”: Docking with no X-ray or Homology Model**

Docking of 42 Kinases into 5 Surrogate X-ray structures

<table>
<thead>
<tr>
<th>Crystal Structure</th>
<th>Test set R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSF1R</td>
<td>0.56</td>
</tr>
<tr>
<td>CHK1</td>
<td>0.56</td>
</tr>
<tr>
<td>PDK1</td>
<td>0.49</td>
</tr>
<tr>
<td>AurA</td>
<td>0.49</td>
</tr>
<tr>
<td>GSK3</td>
<td>0.50</td>
</tr>
<tr>
<td>PI3K</td>
<td>0.47</td>
</tr>
<tr>
<td>Tie2</td>
<td>0.35</td>
</tr>
<tr>
<td>PIM1</td>
<td>0.42</td>
</tr>
<tr>
<td>16X Ensemble</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Prediction Quality: limited by training data set

Model quality vs. Dynamic Range: $R^2=0.57$

Red => more pts
Additional AUTOSHIM Profile boost: almost as predictive as 2D

A chemo-kinometric ensemble of 51 AUTOSHIM models:
9 w/ X-ray: median=0.58, max=0.68,
42 w/o X-ray: median=0.52, max=0.75
**Archive Pre-Dock:** 1.5 million cpds into 16 kinase structures

1) Superpose 16 diverse kinase active sites (universal kinase model)
2) Dock 1.5M cpds x 16 X-ray x 100 poses (dockit, 8 CPU yrs)
   a) Filter out conformations lacking hinge H-bond (MAGNET)
3) Minimize 200e+6 poses (flo+, 40 CPU yrs ... 35 on grep?)
4) Extract Magnet features for each minimized pose
5) 5 months elapsed time!

---

### Tight on hinge

![Tight on hinge](http://www.daylight.com/download/contrib/autoshim.html)

### Xtal ligands

![Xtal ligands](http://sourceforge.net/projects/autoshim/)

---
MTS/Virtual Screening: Iterative Kinase MTS before HTS

1) MTS 8 pt. IC$_{50}$s of 10,000 Core Set.
2) Build 2D-profile & Surrogate AutoShim models from IC$_{50}$s.
3) If OK, predict full collection.
4) Prioritize predicted hits.
5) Order cpds and measure IC50s.
   a) Good hits? Get started!
   b) Tool cpd., X-ray, etc.
   c) Re-train models and repeat?
6) Full HTS to follow?
   a) C.f. to model to prioritize marginal hits & ID false negatives
>30X enrich. on PDK1 “KSP” 50K (4 pt. HTS IC₅₀s)

New screen vs. 2 old screens

At 3uM Recovered 962 of 1236 hits (78%) with 283 FPs (HR=77%)

“80/20 solution”

FP:283  R²=0.5  TP:962

TN:47643  FN:264
C.f. 2D Profile-QSAR to 3D AutoShim on KSP 50K at 1uM

<table>
<thead>
<tr>
<th>Metd</th>
<th>Pred. Rate</th>
<th>Hit Rate</th>
<th>Recov Rate</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>72%</td>
<td>62%</td>
<td>39X</td>
<td></td>
</tr>
<tr>
<td>2D</td>
<td>71%</td>
<td>62%</td>
<td>39X</td>
<td></td>
</tr>
<tr>
<td>AND</td>
<td>74%</td>
<td>49%</td>
<td>40X</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>70%</td>
<td>76%</td>
<td>38X</td>
<td></td>
</tr>
</tbody>
</table>

KSP 50K cpds at 1 uM

All=49,145

Expt=903

2D=794

777=3D

120 443

153

61

77

219
Based on 2D and 3D predictions, MW, Lipinski, chemical novelty and PDK1 team input, 342 cpds were ordered and tested. Hit rate = 24%. Histogram shows how fraction active increases with similarity to known actives.

J. Chem. Inf. Model; 2008; 48(4); 861-872 & 873-881
Hit rate (5 uM) by similarity for 342 ordered cpds and EMV screen

Histogram shows how fraction active decreases with dissimilarity to known actives.

Overall 5 uM Hit Rates
- 342 = 19.0%
- HTS = 0.7%
- Enrichment = 27X

Blue=Avalon
Red=342
Similar hit-rates for 3D AutoShim and 2D profile-QSAR methods.
Kinase “L” preliminary results

8900 compounds

Molwt <= 450, PSA <= 100

894 compounds

Clustered and compared against Familiar scaffolds

80 compounds

52 tested so far
35 < 10 uM (67% hit rate)
28 < 1 uM (54% hit rate)
Most potent 5 nM

Acitivity Distribution
Kinase “L” hit-rates vs. Similarity to Known Actives

![Graph showing hit-rates at 10 uM and 1 uM against similarity to nearest known active.](image)
Kinase “R” Iterative screening

Model Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>R²</th>
<th>Q²</th>
<th>R²_ext</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoshim</td>
<td>2685</td>
<td>0.57</td>
<td>0.47</td>
<td>0.40</td>
</tr>
<tr>
<td>ProfileQSAR</td>
<td>3055</td>
<td>0.60</td>
<td>0.53</td>
<td>0.53</td>
</tr>
</tbody>
</table>

- Build Autoshim and PLSB models
- Carry out enrichment analysis
- Rank pre-docked database (1.5 million)

High ligand eff.
- Top 2000 high LE
- Heavy atoms ≤ 21

High pred. activ.
- Top 500 FIT
- 20 (top pred) + 56(diverse)/480

SAR sets
- Known kinase inhibitor scaffold
- MW ≤ 500
  - FIT ≥ 5.5
- Scaffold based SAR sets
- MW ≤ 400
  - Rot bonds ≤ 4
  - FIT ≥ 5.1

Novel selection
- No Known Kinase inhibitor scaffold
- MW ≤ 400
- 344 molecules

1100 molecules (Total) ~900 ordered for testing

- Selection threshold
- % Retrieved vs % Database
- P_active
- P_inactive
- Random
- 77 act
- 2 inact
- 38 : 1 selectivity ratio

238th National ACS meeting | E. Martin| March 2009|Adaptive Scoring Functions
Protein “R” Hit-rates by Category

- **25 uM**
  - 37% active

- **10 uM**
  - 28% active
“R” Hit-rate vs. Similarity to Nearest Known Active

Hitrates Vs Similarity Bins

Similarity

- Yellow: Hit-rate at 25 uM
- Red: Hit-rate at 10 uM
- Blue: Hit-rate at 1 uM
Minimal Kinase Ensemble Receptor (MKER) of 7-8 structures

SFFS on 20 of 74 models (max 2000 data points)
Conclusion: Iterative Kinase MTS

Experimental high-throughput screening of a large (1.5 million) compound collection typically takes 6 months and costs $1,000,000.

2D profile-QSAR models combine vast data for accurate predictions, including cellular activity.

Ensemble Surrogate AutoShim solves 3 serious limitations of virtual screening by docking:
1) Does not require a (new) protein structure.
2) It as fast as 2D QSAR.
3) Highly predictive.

2D+3D union scoring finds majority of active compounds very efficiently.

Has been successfully applied to 6/6 active Novartis projects. \[Q^2 = 0.35 \text{ – } 0.7, \text{ enrichments of } 20x \text{ to } 40x\]

7-8 kinase structure ensemble in a minimal spanning set.
Acknowledgements

David Sullivan (now at Anacor): Original Method Development, PDK1
Prasenjit Mukherjee: Recent applications, Cells, MKER
Hanneke Jansen: Programmed and refined 1st MAGNET prototype
Scott Dixon: Developed SEA and commercial MAGNET
Jason Kondracki & Joe Ringgenberger: Kinase data warehouse
Mika Lindvall: proteins “R”, “K1”, and “K2”
Paulette Greenidge: protein “L”
Henrik Moebitz: protein “C”
Bob Warne, Paul Feucht, Nichole Robison: Iterative MTS Assays