Exercising receptor-site similarity:

From Off-Target Identification to Scaffold Hopping

Steven Muskal, Ph.D.
Chief Executive Officer
Eidogen-Sertanty, Inc.
smuskal@eidogen-sertanty.com
Protein Structure Growth is Accelerating

> 50K Structures/co-complexes (Aug-2008)
> 600 deposits per month → >150/week!

PDB Growth
source: rcsb.org
## Drugs developed using SBDD

<table>
<thead>
<tr>
<th>Inhibitor/Drug</th>
<th>Disease</th>
<th>Company(s)</th>
<th>Protein targeted</th>
<th>Enzyme Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>STI-571/Gleevec</td>
<td>Chronic Myeloid Leukemia</td>
<td>Novartis</td>
<td>c-Abl kinase</td>
<td>Tyrosine kinase</td>
</tr>
<tr>
<td>Fluoroquinolone/Ciprofloxacin</td>
<td>Bacterial infection</td>
<td>Bayer</td>
<td>Gyrase</td>
<td>ATP Hydrolase</td>
</tr>
<tr>
<td>Saquinavir/Invirase, Ritonavir/Norvir, Indinavir/ Crixivan, Nelfinavir/Viracept, Amprenavir/Agenerase, Fosamprenavir/Lexiva,</td>
<td>AIDS</td>
<td>Roche, Abbott, Agouron, Merck, Vertex</td>
<td>HIV-1 Protease</td>
<td>Aspartylprotease</td>
</tr>
<tr>
<td>Trusopt</td>
<td>Glaucoma</td>
<td>Merck</td>
<td>Carbonic Anhydrase</td>
<td>Lyase</td>
</tr>
<tr>
<td>Thymitaq</td>
<td>Cancer</td>
<td>Agouron</td>
<td>Thymidylate synthase</td>
<td>Methyl transferase</td>
</tr>
<tr>
<td>Celecoxib/Celebrex, Rofecoxib/Vioxx</td>
<td>Inflammation, rheumatoid arthritis</td>
<td>Searle, Merck</td>
<td>Cox-2</td>
<td>Oxidoreductase</td>
</tr>
<tr>
<td>AG3340/Prinomastat</td>
<td>Cancer</td>
<td>Agouron</td>
<td>Matrix metalloprotease</td>
<td>Metalloprotease</td>
</tr>
<tr>
<td>Oseltamivir phosphate/Tamiflu, Zanamivir/Relenza</td>
<td>Influenza</td>
<td>Roche</td>
<td>Neuraminidase</td>
<td>Glycosidase</td>
</tr>
</tbody>
</table>

Industrializing an Information Rich Craft

• Pharma cost reductions (reductions in jobs, spending, etc.)
  → Fewer IT specialists with less resource, supporting more people

• One-at-a-time computational efforts are bottlenecks
  → Many proteomic riches remain untapped

Non-specialist driven workflow-apps have become necessary!
Bringing Proteomic Riches to Non-Specialists

Automated Modeling and Proteomic Structural Mining

- **Sequence-to-Structure Calculation Cascade**
- **Search-by:** KeyWord, Sequence, Ligand, Protein Structure, Receptor-Sites, etc.
- **Exploit Structural fold and receptor-site conservation**
  - **Off-Target Identification (opportunities v. liabilities)**

- **Borrowing Matter Ideas from co-complexes and protein structures**
### About Eidogen-Sertanty

- **Knowledge-Driven Discovery Solutions Provider**
  - Formed in March 2005 when Sertanty (Libraria → Sertanty 2003) acquired Eidogen (Bionomix 2000)
  - >$20M Invested in Technology Development
  - 12 FTE’s
  - Worldwide Customerbase
  - Cash-Positive

- **Chemogenomic Databases & Analysis Software**
  - *TIP™* - Structural Informatics Platform
  - *KKB™* - Kinase SAR and Chemistry Knowledgebase
  - *CHIP™* - Chemical Intelligence Platform

- **DirectDesign™ Discovery Collaborations**
  - In Silico Target Screening ("Target Fishing" and Repurposing)
  - Target and compound prioritization services
  - Fast Follower Design: Novel, Patentable Leads
TIP Algorithm Engine

- > 400K Sequences
- > 158K Chains & Models
- > 388K Sites

> 33M Sequence Similarities
> 69M Structure Similarities
> 62M Site Similarities
**STRUCTFAST™**

Structure Realization Utilizing Cogent Tips From Aligned Structural Templates

**Basic Principle:** Gaps known to exist should not be strongly penalized.

Leverages experimental structure and structural alignment data to create better alignments


2) STRUCTFAST: Protein Sequence Remote Homology Detection and Alignment Using Novel Dynamic Programming and Profile-Profile Scoring
Proteins. 2006 64:960-967
SiteSeeker™

Geometric Site-Finding Algorithms Find Many Pockets
But they don’t know which pockets are important!

Evolutionary Trace Approach
Can’t clearly define site boundary
Not all conserved residues are functionally relevant

SiteSeeker combines both methods

Reliability & Confidence
We use proteins with apo- & co-crystal structures in the PDB to test the accuracy & reliability of method

Allows us to map SiteSeeker score to predict confidence!
(e.g. At this SiteSeeker score, 80% are “real” co-crystal sites)
→ Sites with <60% confidence are not stored in TIP
**Weighted Clique Detection Algorithm**

Importance of Points Related To Conservation In Multiple Sequence Alignment

Surface Atoms Assigned One of 5 Different Chemical Characters

Matching points increase the *SiteSorter* similarity score
TIP Content

> 75,000 Human Sequences

> 116,000 Total PDB chains (~50K PDBs)
> 42,000 Homology Models

> 194,000 PDB co-crystal sites
> 190,000 Predicted Sites (on PDBs & Models)

> 33M Sequence Similarities

> 69M Structural Similarities

> 62M Site Similarities

Updated monthly with new PDBs and models:

**e.g. March 2006:**
- 661 new PDBs added
- 447 new models built
  - 153 had no previous structure in TIP
  - 294 had “better” models built

**e.g. July 2008:**
- 576 new PDBs added
- 1045 new models built

Automatically updated with new models as the PDB grows
Off-Target Opportunities

**Intra-Family Opportunities**

- **B-RAF**
  - BAY 43-9006
  - Sequence ID = 30%
  - Site ID = 60%
  - Top 10 SiteSorter rank

- **C-KIT**
  - B-RAF inhibitor BAY 43-9006 also inhibits C-KIT

**Inter-family Opportunities**

- **HIV protease**
  - Viracept
  - Key contacts conserved

- **Cathepsin D**
  - Cathepsin D is inhibited by HIV protease inhibitors
PXR – Promiscuous Ligand-Binding Site

Example High-ranking similar sites:

- Bile Acid Receptor FXR
- PPAR-γ receptor
- ACE2
- Thyroid Receptor
- Caspase-3
- HMG-CoA Reductase (statin target)

Pregnane X-receptor – PXR (“sensor”) → CYP3A4 (“executioner”)
PXR Binds > 50% drugs
Including some bile acids, statins, herbal components, a selection of HIV protease inhibitors, calcium channel modulators, numerous steroids, plasticizers and monomers, organochlorine pesticides, a peroxisome proliferator-activated receptor-α antagonist, xenobiotics and endobiotics…
LIMK1 – ATP binding site comparison

The ATP site of LIMK1 shares a high level of homology with several well-studied kinases.

Eidogen-Sertanty KKB SAR Data Point Distribution

Kinase SAR Knowledgebase – Hot Targets

>362,000 SAR data points curated from
>4,270 journal articles and patents
>130 Bayesian QSAR Models
Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

> 4278 journal articles/patents

- **KKB Content Summary (Q2-2008):**
  - # of kinase targets: **>390**
  - # of SAR Data points: **> 362,000**
  - # of unique kinase molecules with SAR data: **>120,000**
  - # of annotated assay protocols: **>16,000**
  - # of annotated chemical reactions: **>2,300**
  - # of unique kinase inhibitors: **>465,000** (~340K enumerated from patent chemistries)

- **KKB Growth Rate:**
  - Average **15-20K** SAR data points added per quarter
  - Average **20-30K** unique structures added per quarter
Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from >4100 journal articles/patents

Kinase Validation Set

Three sizable datasets freely available to the research community

http://www.eidogen-sertanty.com/kinasednld.php

Average 20-30K unique structures added per quarter
Lead Discovery: Knowledge-Based Design

LigandCross Workflow

- TIP Co-Crystals
- Proprietary Co-Crystals
- Docked Actives

EVE-2D
Site Overlays
LigandCross

New Molecules (.sdf)

EVE-3D
Pose Visualization

New Molecules via LigandCross
Novel Ligands via Ligand Crossover

Starting ligands

Hybridized product ligands
From Ligand Query to Sites to New Ligand Ideas
### Step 1: Find Co-complexes and Sites from **Ligand-Structure-Search**

<table>
<thead>
<tr>
<th>Molecule</th>
<th>ligname</th>
<th>similarity</th>
<th>pdbcode</th>
<th>siteid</th>
<th>FourCode</th>
<th>pdbIdx</th>
<th>number</th>
<th>proteinid</th>
<th>title</th>
<th>classification</th>
<th>source</th>
<th>compound</th>
<th>releaseDate</th>
<th>journalTitle</th>
<th>journalReference</th>
<th>exptype</th>
</tr>
</thead>
<tbody>
<tr>
<td>STI</td>
<td>1</td>
<td>2iDA</td>
<td>1305707</td>
<td>2i0</td>
<td>2i0</td>
<td>1305799</td>
<td>42526</td>
<td>L0X bound to MATINB</td>
<td>TRANSFERASE</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
<td>HUMAN; GENE: L0X; EXPRESSION_SYSTEM: SPEROPTERA; FRUITEMPERA</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
</tr>
<tr>
<td>STI</td>
<td>1</td>
<td>2iQX</td>
<td>1148614</td>
<td>2i0q</td>
<td>2i0q</td>
<td>1126100</td>
<td>26318</td>
<td>STRUCTURE OF CHICKEN C-SRC KINASE DOMAIN IN COMPLEX WITH THE CANCER DRUG MATINB</td>
<td>TRANSFERASE</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
<td>HUMAN; GENE: ABL; EXPRESSION_SYSTEM: SPEROPTERA; FRUITEMPERA</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
</tr>
<tr>
<td>STI</td>
<td>1</td>
<td>2iyyA</td>
<td>518007</td>
<td>2iyyA</td>
<td>2iyyA</td>
<td>504013</td>
<td>16081</td>
<td>HUMAN ABL KINASE DOMAIN IN COMPLEX WITH MATINB (STRIKING OLIVE)</td>
<td>TRANSFERASE</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
<td>HUMAN; GENE: ABL; EXPRESSION_SYSTEM: SPEROPTERA; FRUITEMPERA</td>
<td>MOL_ID:</td>
<td>MOL_ID:</td>
<td>1: ORGANISM:SCIENTIFIC; HOMO SAPIENS; ORGANISM:COMMON</td>
</tr>
</tbody>
</table>
Step 2: Find Other Receptor Sites from Site-Similarity Search

Site: 1309707
### Example Site Similarity Results (Query: s1309707)

<table>
<thead>
<tr>
<th>Site</th>
<th>SiteLigand</th>
<th>SiteProtein</th>
<th>SiteScore</th>
<th>ContactScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1309707</td>
<td>STI</td>
<td>2pl0A</td>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>1420904</td>
<td>C92</td>
<td>3cpbB</td>
<td>110.906</td>
<td>0.7</td>
</tr>
<tr>
<td>1384893</td>
<td>900</td>
<td>3b8qB</td>
<td>121.051</td>
<td>0.67</td>
</tr>
<tr>
<td>1322334</td>
<td>276</td>
<td>2qu5A</td>
<td>117.866</td>
<td>0.66</td>
</tr>
<tr>
<td>1284638</td>
<td>406</td>
<td>2e2bA</td>
<td>119.18</td>
<td>0.64</td>
</tr>
<tr>
<td>1396160</td>
<td>2RL</td>
<td>2rl5A</td>
<td>121.208</td>
<td>0.63</td>
</tr>
<tr>
<td>1400124</td>
<td>NIL</td>
<td>3cs9D</td>
<td>111.198</td>
<td>0.62</td>
</tr>
<tr>
<td>867405</td>
<td>7MP</td>
<td>2hiwA</td>
<td>101.948</td>
<td>0.61</td>
</tr>
<tr>
<td>916548</td>
<td>242</td>
<td>2ofvB</td>
<td>109.214</td>
<td>0.6</td>
</tr>
<tr>
<td>1147514</td>
<td>MUH</td>
<td>2oscA</td>
<td>104.115</td>
<td>0.6</td>
</tr>
<tr>
<td>776230</td>
<td>WBT</td>
<td>1wbtA</td>
<td>101.635</td>
<td>0.6</td>
</tr>
<tr>
<td>916805</td>
<td>1N8</td>
<td>2og8A</td>
<td>116.819</td>
<td>0.59</td>
</tr>
<tr>
<td>394066</td>
<td>PRC</td>
<td>1fpuB</td>
<td>107.297</td>
<td>0.57</td>
</tr>
<tr>
<td>1415780</td>
<td>C19</td>
<td>3cp9A</td>
<td>104.078</td>
<td>0.56</td>
</tr>
<tr>
<td>911671</td>
<td>KIN</td>
<td>2hz0A</td>
<td>106.08</td>
<td>0.56</td>
</tr>
<tr>
<td>1148488</td>
<td>608</td>
<td>2p2iB</td>
<td>109.41</td>
<td>0.55</td>
</tr>
<tr>
<td>1300447</td>
<td>GIG</td>
<td>2oh4A</td>
<td>110.471</td>
<td>0.53</td>
</tr>
<tr>
<td>1320735</td>
<td>857</td>
<td>2qu6B</td>
<td>116.424</td>
<td>0.52</td>
</tr>
<tr>
<td>437653</td>
<td>B96</td>
<td>1kv2A</td>
<td>107.323</td>
<td>0.52</td>
</tr>
<tr>
<td>691631</td>
<td>L11</td>
<td>1w83A</td>
<td>101.268</td>
<td>0.52</td>
</tr>
<tr>
<td>1147212</td>
<td>RAJ</td>
<td>2oo8X</td>
<td>104.058</td>
<td>0.52</td>
</tr>
<tr>
<td>910098</td>
<td>GIN</td>
<td>2hz0B</td>
<td>108.713</td>
<td>0.51</td>
</tr>
<tr>
<td>1396708</td>
<td>P38</td>
<td>3bv2A</td>
<td>124.962</td>
<td>0.51</td>
</tr>
<tr>
<td>436174</td>
<td>BMU</td>
<td>1kv1A</td>
<td>88.568</td>
<td>0.5</td>
</tr>
<tr>
<td>1412158</td>
<td>G2G</td>
<td>2puuA</td>
<td>118.296</td>
<td>0.5</td>
</tr>
<tr>
<td>775147</td>
<td>LI3</td>
<td>1wbvA</td>
<td>85.135</td>
<td>0.5</td>
</tr>
<tr>
<td>1415688</td>
<td>CS2</td>
<td>3cpcB</td>
<td>102.25</td>
<td>0.48</td>
</tr>
<tr>
<td>1431710</td>
<td>GK6</td>
<td>3d83A</td>
<td>104.164</td>
<td>0.48</td>
</tr>
</tbody>
</table>
### Example Ligands Extracted from Similar Sites

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Ligand</th>
<th>Ligand</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Ligand 1" /></td>
<td><img src="image2.png" alt="Ligand 2" /></td>
<td><img src="image3.png" alt="Ligand 3" /></td>
<td><img src="image4.png" alt="Ligand 4" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Ligand 5" /></td>
<td><img src="image6.png" alt="Ligand 6" /></td>
<td><img src="image7.png" alt="Ligand 7" /></td>
<td><img src="image8.png" alt="Ligand 8" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Ligand 9" /></td>
<td><img src="image10.png" alt="Ligand 10" /></td>
<td><img src="image11.png" alt="Ligand 11" /></td>
<td><img src="image12.png" alt="Ligand 12" /></td>
</tr>
<tr>
<td><img src="image13.png" alt="Ligand 13" /></td>
<td><img src="image14.png" alt="Ligand 14" /></td>
<td><img src="image15.png" alt="Ligand 15" /></td>
<td><img src="image16.png" alt="Ligand 16" /></td>
</tr>
</tbody>
</table>
Step 3: LigandCross – Mixing Ligand Features from Aligned Sites
Example LigandCross Results
Step 4: LigandCross Ligands reDocked into s1309707
# LigandCross Ligands with Reported Biological Activity

<table>
<thead>
<tr>
<th>Kinase Knowledgebase (pIC50)</th>
<th>Bayesian Model Predictions (PP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LC4D</strong></td>
<td><strong>ABL</strong></td>
</tr>
<tr>
<td>G26-STI_12</td>
<td>6.7</td>
</tr>
<tr>
<td>900_STI_1</td>
<td>6.1</td>
</tr>
<tr>
<td>7MP_1N8_4</td>
<td>7.8</td>
</tr>
<tr>
<td>7MP_1N8_2</td>
<td>6.8</td>
</tr>
<tr>
<td>7MP_RAJ_3</td>
<td>8.4</td>
</tr>
<tr>
<td>7MP_GIN_4</td>
<td>7.5</td>
</tr>
<tr>
<td>242_C52_2</td>
<td>7.9</td>
</tr>
<tr>
<td>L13_L11_1</td>
<td>7.2</td>
</tr>
<tr>
<td>608_GIG_7</td>
<td>6.1</td>
</tr>
<tr>
<td>KIN_BMU_4</td>
<td>6.1</td>
</tr>
<tr>
<td>G26_KIN_3</td>
<td>6.1</td>
</tr>
</tbody>
</table>

1: G26-STI_12  | 2: 900_STI_1  | 3: 7MP_1N8_4  | 4: 7MP_1N8_2  |
5: 7MP_RAJ_3  | 6: 7MP_GIN_4  | 7: 242_C52_2  | 8: L13_L11_1  |
9: 608_GIG_7  | 10: KIN_BMU_4 | 11: G26_KIN_3
Conclusions

- Significant receptor-site similarities exist within and across target families

- The structurally resolved and modelable proteome is a very rich source for new matter ideas

- LigandCross can be an effective strategy to generate novel, bioactive molecules from co-complex information.
Acknowledgements

- Stephan Schürer
- Kevin Hambly
- Joe Danzer
- Brian Palmer
- Derek Debe
- Aleksandar Poleksic

- Accelrys/Scitegic - Shikha Varma-O'Brien/Ton van Daelen

- ChIP: National Institute of Standards and Technology (NIST) – ATP program: ‘Chemical Intelligence Platform for Rapid Discovery of DrugLeads’

Contact
Steven Muskal
Chief Executive Officer
smuskal@eidogen-sertanty.com
Chemical Intelligence Platform (ChIP™)
ChIP Reaction Transforms | Building Blocks Databases

Development funded by a $2.5M NIST-ATP Grant

ChIP: Navigating Accessible Synthetic Space

Building Block Sources May Be Changed, Enabling Diversity Oriented or Focused Synthesis
ChIP™ mixes and matches reaction methods so that novel scaffolds are generated with their synthetic road-maps.

**ChIP: Protocol Shuffling Chemistries**

- **S1**: Nucleophilic Aromatic Substitution Reactions
- **S2**: Pd-Catalyzed Aromatic Substitutions
- **S3**: Functional Group Transformations
- **S4**: Amine Acylation Reactions (Amides/Carbamates)
- **S5**: Amine Acylation Reactions (Ureas/Thioureas)
- **S6**: Formation of Diverse Heterocyclic Systems
- **S7**: Formation of Thioimidazoles
- **S8**: Standard Deprotection Steps

![Diagram of protocol shuffling chemistries](image-url)
Ligand pharmacophoric potential

Journal of Chemical Information and Computer Sciences, 1999, 39 (3) : 569-74
Pharmacophoric Feature

7 pharmacophore types:
- H-bond acceptor (A) & donor (D)
- negative (N) & positive charges (P)
- aromatic (R), hydrophobic (H)
- other (X)

6 distance ranges:
2-4.5, 4.5-7, 7-10, 10-14, 14-19, 19-24 Å

Enumerate 3-point pharmacophores
-> 10,549
Target specific ligands share pharmacophoric feature.

- NMDA Receptor Antag
- Leukotriene Antag
- PAF Antag
- Angiotensin II Blocker
- Ca Channel Blocker
- K Channel Activator
- Substance P Antag
- ACAT Inhibitor
- Cyclooxygenase Inhib
- Lipoxygenase Inhib
Non-obvious Me-Too’s

RANITIDINE (Zantac)
“Antiulcer”

MWT: 314.4; LogP: 0.27; pKa [2.30, 8.20]
Target: Histamine H2-antagonist.
Oral Avail.: 52% (±11)
Urinary Excretion: 69% (±6)
Plasma Bound: 15% (±3)
Clearance: 730 mL/min (±80)
Half-Life: 2.1 hr (±0.2)
Effective Conc.: 100 ng/mL

CIMETIDINE (Tagamet)
“Antiulcer”

MWT: 252.3; LogP: 0.40; pKa: [6.80]
Target: Histamine H2-antagonist.
Oral Avail.: 62% (±6)
Urinary Excretion: 62% (±20)
Plasma Bound: 19%
Clearance: 540 mL/min (±130)
Half-Life: 1.9 hr (±0.3)
Effective Conc.: 800 ng/mL

MDLSim: 46.8/100.0
DaySim: 0.32/1.0
PFPSim: 0.88/1.0
ChIP-ing Towards Me-Too's Known PDE-IV Inhibitors

Example ChIP Generated Synthetic Road-Maps
e.g. Non-specialist Proteomic Mining

Goal: Obtain ligand/scaffold ideas from co-complex examples

Workflow:
1) Ligand SSS/Sim/3D search in co-complexes
2) Issue Target-side SiteSearches across proteome
3) Pull Ligand examples from similar sites
4) Launch EVE

Steps:
1. Company HTTP Proxy required?
2. Clean-up from previous runs
3. ES Authenticate
4. SD Reader to CTAB
   - SOAP - 2D/3D TIP/Ligand Search
   - SOAP - TIP/Peptide
   - SOAP - TIP/Data Search
5. Parse SiteSearch Results
   - Normalized siteSim
   - Sort Data by Normalized percentSim
   - Reduce Site Redundancies
   - Reduce Site Ligand Redundancies
6. Launch EVE
   - SOAP - TIP/SiteSim Search
   - SOAP - TIP/Ligand Search
   - SQL Construct
   - Top N Filter
   - Save Parsed SiteSearch Results
Various dynamic programming seeding methods are used in order to utilize as much information as is available.

Dynamic programming scores are fit to an EVD to assess alignment significance.

1) PDB sequences clustered at 90% identity and 95% coverage.

2) N-by-N comparison of one representative chain from each cluster
   (All other chains are only compared to the representative's significant hits)

Allows structural alignment database to be computed in 1.5 months instead of 2.5 years.

StructSorter Example: Rhinovirus protease

StructSorter computes and stores alignments between Rhinovirus Protease and other mammalian proteases in TIP, despite very low overall sequence and structural similarity.
SiteSeeker Example

Fructose-1,6-bisphosphatase FBPase

SiteSeeker

Pass

Allosteric Site
(Discovered By Pfizer)

>80%

Rank 3

>85%

Rank 4

>60%

Rank 7

F6P Site

AMP Site

Color Key

Missed By Prediction
Correctly Predicted
Falsely Predicted

EidogenSctancy
SiteSeeker Example: PTP1B Allosteric Site

All structures in TIP are annotated with known and predicted binding sites, along with confidence levels for each annotation.
Virtual Target Screening Example: COX-2

Example: Identifying potential “off-targets” for COX-2 inhibitors

Use site to query TIP and rank similar binding sites

Prioritize “off-targets” based on Site-Ligand Contact analysis and/or biological relevance

Rank #1: PPAR-Gamma (Diabetes target)

- Confirmed “Off-Target” for COX-2 inhibitors

Rank #10: Estrogen Sulfotransferase (Estrogen Metabolism pathway)

- Possible “Off-Target” for COX-2 inhibitors
Virtual Target Screening Example: Statins

Example 1: Searching for off-targets to explain pleiotropic effects of statins

Simvastatin (Zocor)

HMG-CoA Reductase Active Site

Perform Site Similarity Search in TIP

Overlay of simvastatin in Human Arginase site:

Identify binding site

Visualize Site alignments in EVE

Supporting Evidence

1) Statins have anti-atherosclerosis effects independent of HMG-CoA Reductase inhibition (Circulation. 2003;108:1368)

2) Statins increase NO synthesis and bioavailability (J Am Soc Nephrol 15:1098-1100, 2004)

3) Arginase is a novel therapeutic target for atherosclerosis, as it decreases the bioavailability of nitric oxide (NO), leading to endothelial dysfunction (Curr Hypertens Rep. 2006 Apr ;8:54-9)

Identify binding site

Similar Surface patches

Dissimilar Surface patches
Virtual Target Screening Example: Statins

Example 2: Searching for off-targets to explain adverse effects of statins

**Supporting Evidence**

1) Statins induce apoptosis in muscle tissue (myotoxicity) via unknown mechanism
   *(Pharmacol Exp Ther. 2005 Sep;314(3):1032)*

2) Statins activate mitochondrial pathway of apoptosis, and this is mediated by the Mitochondrial Permeability Transition Pore (MPTP)

3) ANT1 is an essential component of the MPTP, and is activated by carboxyatractyloside to permeabilize the mitochondria and induce apoptosis

**Overlay of fluvastatin in human ANT1 site**

- **Highly similarly charged pockets** (conserved Arg/Lys/Asp residues) and similar H-bond contacts

**Carboxyatractyloside-binding site of human mitochondrial Adenosine Nucleotide Translocator 1 (ANT1)**
Supplemental Slides
EidoSert Products & Services

Data

- Protein Sequences
- Experimental Protein Structures
- Ligand-binding sites and binding modes

Knowledge-Based Solution

- Proteome-wide homology modeling and structural alignment calculations
- Correlating binding site and binding mode similarities for selectivity & cross-reactivity prediction

- Kinase database and infrastructure for managing SAR information
- Guided in-silico focused library design using programmed reaction transformations

DirectDesign™ Collaborations

TIP™
EVE™

KKB™
ARK™

ChIP™
Target Informatics Platform (TIP™)

Comparative Visualizer (EVE™)
In Silico Target Screening ("Target Fishing")

Example Reference: Interrogating the Druggable Genome with Structural Informatics
Kevin Hambly*, Joseph Danzer, Steven Muskal, and Derek A. Debe.
Molecular Diversity, 2006.
TIP Druggable Genome Coverage

- 5,715 Sequences From 157 Druggable PFAM Domains
- 4,679 Druggable Targets with one or more structures (81.9% Coverage)
- 1,263 with PDB’s
- 3,416 with STRUCTFAST Models

**Major Membrane Protein Targets**
- 983 with TIP structure (65% coverage) → 2.7% PDBs → 97.3% Models

**Major Enzyme Targets**
- 1,767 with TIP structure (95% coverage) → 23.5% PDBs → 76.5% Models

**KEY**

<table>
<thead>
<tr>
<th>Description</th>
<th>%Structural Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Kinases</td>
<td>98.0%</td>
</tr>
<tr>
<td>Trypsin-like Proteases</td>
<td>98.0%</td>
</tr>
<tr>
<td>Metalloproteases</td>
<td>91.0%</td>
</tr>
<tr>
<td>ATPases</td>
<td>80.1%</td>
</tr>
<tr>
<td>Short chain dehydrogenases</td>
<td>95.6%</td>
</tr>
<tr>
<td>P450s</td>
<td>96.9%</td>
</tr>
<tr>
<td>Phospholipases</td>
<td>82.1%</td>
</tr>
<tr>
<td>Cyclophilin-type isomerase</td>
<td>100.0%</td>
</tr>
<tr>
<td>Protein Tyr. Phosphatases</td>
<td>100.0%</td>
</tr>
<tr>
<td>Other Enzymes</td>
<td>88.7%</td>
</tr>
<tr>
<td>GPCRs, Family 1</td>
<td>78.2%</td>
</tr>
<tr>
<td>Ion Pumps</td>
<td>33.9%</td>
</tr>
<tr>
<td>Ion Channels</td>
<td>69.1%</td>
</tr>
<tr>
<td>GPCRs, Family 2</td>
<td>5.3%</td>
</tr>
<tr>
<td>Mitochondrial transporters</td>
<td>69.7%</td>
</tr>
<tr>
<td>Other Cell Surf. Receptors</td>
<td>62.1%</td>
</tr>
<tr>
<td>Other Membrane Transporters</td>
<td>4.5%</td>
</tr>
<tr>
<td>Nuclear Hormone Receptors</td>
<td>100.0%</td>
</tr>
<tr>
<td>Other Drug Target Domains</td>
<td>78.9%</td>
</tr>
</tbody>
</table>
Animal Model Suitability

Cathepsin S Inhibition by JNJ 10329670

<table>
<thead>
<tr>
<th>Species</th>
<th>IC50 (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>34</td>
</tr>
<tr>
<td>Dog</td>
<td>124</td>
</tr>
<tr>
<td>Monkey</td>
<td>266</td>
</tr>
<tr>
<td>Bovine</td>
<td>411</td>
</tr>
<tr>
<td>Mouse</td>
<td>2364</td>
</tr>
</tbody>
</table>

Anti-Infective Spectrum

Comparison of fabH from Several Pathogens

For broad spectrum inhibition, avoid interactions with non-conserved regions in C. pneumonia fabH.
Druggability and Selectivity Analysis

MMP Substrate Site Similarity

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Locus</th>
<th>Description</th>
<th>% Cont</th>
<th>SiteSorter</th>
<th>Similarity (%)</th>
<th>SiteResidue</th>
<th>Conservation</th>
<th>SimilarityDendrogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>pdb1mms46</td>
<td>MMP03</td>
<td>HIN: N (R.C..)</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model25148</td>
<td>MMP12</td>
<td>HTA: N(O-L..)</td>
<td>100</td>
<td>94.98</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model11126</td>
<td>MMP15</td>
<td>HTA: N(O-L..)</td>
<td>100</td>
<td>98.78</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1mms46</td>
<td>MMP07</td>
<td>RVE: N(H-L..)</td>
<td>100</td>
<td>101.42</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model28047</td>
<td>MMP20</td>
<td>HTA: N(O-L..)</td>
<td>100</td>
<td>94.03</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb2kwe402</td>
<td>MMP01</td>
<td>RO:O[T1-N..]</td>
<td>100</td>
<td>96.38</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1mms46</td>
<td>MMP08</td>
<td>BAT: 4(N.H..)</td>
<td>100</td>
<td>98.82</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model15220</td>
<td>MMP24</td>
<td>BAT: 4(N-H..)</td>
<td>100</td>
<td>98.21</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1mms46</td>
<td>MMP16</td>
<td>BAT: 4(N-H..)</td>
<td>100</td>
<td>97.33</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1q3b48</td>
<td>MMP10</td>
<td>NGH: N-Iso..</td>
<td>100</td>
<td>93.85</td>
<td>88</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1f1m1e36</td>
<td>MMP13</td>
<td>WAY: N-Hyd..</td>
<td>100</td>
<td>87.43</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pdb1h0w441</td>
<td>MMP02</td>
<td>HS: N-[4{(1..)]</td>
<td>100</td>
<td>101.51</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model14014</td>
<td>MMP14</td>
<td>Predicted Site</td>
<td>81</td>
<td>69.38</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>model0578</td>
<td>MMP12</td>
<td>Predicted Site</td>
<td>81</td>
<td>67.42</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Druggability

- Depth
- H-bonding
- Hydrophobicity

Selectivity

- MMP-13
- MMP-3
- MMP-14
- MMP-1
Allosteric Site Opportunities

Allosteric site on p38, behind ATP site

Other kinases with same allosteric site

Predicted site

DATABASE SEARCH
SiteSorter Example: ATP Sites

Overlay of ATP binding sites from completely different folds

- PHOSPHOENOLPYRUVATE CARBOXYKINASE 1AYL
- DETHIOBIOTIN SYNTHETASE 1A82
- RUVB HOLLIDAY JUNCTION DNA HELICASE 1J7K
- THYMIDYLATE KINASE 1E2Q
EVE: TIP’s Fully Integrated Analysis Tool

EVE-2D

EVE-3D

TIP Project

Data
EVE Comparative Visualizer Layout

- Interactive Structure/Site Alignment Window
- Multiple Structure/Site viewing
- Similarity Clustering View
- Customizable Selection & Analysis Toolbar
Selectivity Opportunities

Finds the basis for selectivity in RSK’s (p90 ribosomal S6 Kinases)

cSLiC Binding Mode Analysis

Similar to pSIFT approach developed by Jus Singh’s Group at Biogen

cSLiC: Composite Site-Ligand Contacts
cSLiC Binding Mode Analysis

Docking Analysis & Rescoring

EVE
Structure & cSLiC Analysis

TIP Structure DB

Small Molecule DB

Spreadsheet Score Analysis & Plotting

Docking

Docked Poses (.sdf)

EVE Pose Visualization

Structure Family

Structure & Docking Site (.pdb)

cSLiC & Docking Scores

Docking Analysis & Rescoring

EidogenSertanty
Dramatically enhances docking-based screening

**cSLiC: CDK2 Enrichment**

![Graph showing the comparison of different methods in terms of database screening and active compounds recovery.](image)

- **Ideal**
- **Co-Crystal cSLiC + Docking**
- **Co-Crystal cSLiC**
- **Docking**
- **Random**

**Legend:**
- % of Actives Recovered vs % of Database Screened
TIP and EVE Flexibility

**TIP Database**
- Upload new sequences (or genomes)
- Define and query custom binding sites
- Build STRUCTFAST models from multiple templates

**EVE Comparative Visualizer**
- Import your own docked ligands
- Import your own homology models
- Import your own crystal and co-crystal structures

Public data is made available to all users.
Your data and calculation results stay *private*.

© EidogenS Certainty
<table>
<thead>
<tr>
<th>LC-ID</th>
<th>KKB-ID</th>
<th>Structures</th>
<th>ABL</th>
<th>PDGFR</th>
<th>PDGFRB</th>
<th>JAK3</th>
<th>KDR</th>
<th>LCK</th>
<th>MAPK14</th>
<th>TEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2G_STI_12</td>
<td>2082</td>
<td><img src="image1" alt="Structure" /></td>
<td>6.7</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>900_STI_1</td>
<td>2083</td>
<td><img src="image2" alt="Structure" /></td>
<td>6.1</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7MP_1N8_4</td>
<td>4336542</td>
<td><img src="image3" alt="Structure" /></td>
<td></td>
<td></td>
<td></td>
<td>7.8</td>
<td>9</td>
<td>9.5</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>7MP_1N8_2</td>
<td>4336547</td>
<td><img src="image4" alt="Structure" /></td>
<td></td>
<td></td>
<td></td>
<td>6.8</td>
<td>8.3</td>
<td>9.5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>7MP_RAJ_3</td>
<td>4307626</td>
<td><img src="image5" alt="Structure" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.4</td>
</tr>
</tbody>
</table>

The table above lists various compounds with their corresponding KKB-ID, structures, and activities against different kinases.
<table>
<thead>
<tr>
<th>LC-ID</th>
<th>KKB-ID</th>
<th>Structures</th>
<th>ABL</th>
<th>PDGFR</th>
<th>PDGFRB</th>
<th>JAK3</th>
<th>KDR</th>
<th>LCK</th>
<th>MAPK14</th>
<th>RAF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>7MP_GIN_4</td>
<td>4344221</td>
<td><img src="image1" alt="Structure 1" /></td>
<td></td>
<td></td>
<td></td>
<td>7.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>242_C52_2</td>
<td>4360910</td>
<td><img src="image2" alt="Structure 2" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI3_L11_1</td>
<td>4198240</td>
<td><img src="image3" alt="Structure 3" /></td>
<td></td>
<td></td>
<td></td>
<td>7.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>608_GIG_7</td>
<td>4360457</td>
<td><img src="image4" alt="Structure 4" /></td>
<td></td>
<td></td>
<td></td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIN_BMU_4</td>
<td>4360452</td>
<td><img src="image5" alt="Structure 5" /></td>
<td></td>
<td></td>
<td></td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Kinase Knowledgebase™
(KKB™)
eScreen QSAR Models

3D-QSAR models built from data annotated in KKB

eScreen Training Set Criteria (all data mined from KKB)

- **Modes of ligand interaction with biological target?**
  - PKC: ATP site, phosphatidyl serine site, diacyl glycerol site, substrate site
  - SRC: ATP site, substrate site, SH2-binding domain

- **Substrate for the enzyme?**
  - Peptide, Carbohydrate, Pyruvate, ADP, Creatine

- **What kind of assay was run?**
  - Ligands only grouped from purified enzyme assays

- **What is the size and shape of the ligands in the assay protocol?**
  - Peptides, Small molecules, Heterocycles, unnatural amino acids, peptidomimetics

- **Experimental assay conditions (ATP concentration 5-20 uM)**

- **Is the information suspicious in the paper?**
eScreen Enrichment of Bio-Activity Space

SAR-Based Activity Matrix
eScreen Enrichment of Bio-Activity Space

eScreen Amplified Activity Matrix
Example Enrichment Study - WyethABL

ABL enrichment

%Wyeth-ABL actives found

% Screened

- WyethABL (N:14) + SCREEN-DDRA (N: 26K)
Enrichment analysis for 14 "new" ABL active compounds published by Wyeth (Diane H. Boschelli,* Yanong D. Wang, Steve Johnson, Biqi Wu, Fei Ye, Ana Carolina Barrios Sosa, Jennifer M. Golas, and Frank Boschelli "7-Alkoxy-4-phenylamino-3-quinolinecarbonitriles as Dual Inhibitors of Src and Abl Kinases" J. Med. Chem. 2004, 47, 1599-1601). None used in eABL eScreen development.

Enrichment greater than 7-fold. 13 of the 14 had Observed pIC50's >= 6.0 and were termed "WyethABL active" and thrown into the SCREEN-DDRA pot of compounds previously described. 1 compound overlap (molid: 13783, SCREEN-DDRAId: 301966) between the 13 WyethABL compounds and the 26K SCREEN-DDRA set (actually denoted as a Src protein kinase inhibitor).

Several known kinase actives in SCREEN-DDRA, many of which seem to bubble to the top in a general eABL ranking. In the rank-ordered compounds, for example, several of the top compounds are noted PDGF, FGF, and SRC inhibitors. Also, no less than 45 of the top-300 in this set are classified as Antineoplastic.

10 of 13 are identified when screening through just 10% of the set and all 13 are identified before screening through 40% of the set.
eScreen Example

- Prioritized 51,000 compound library using a collection of eScreens
- Compounds screened in lung cancer cell-based assay*

~5-fold enrichment after screening 10% of the library.

*Screening data courtesy of Hakim Djaballah, Memorial Sloan-Kettering Cancer Center
3,169 of the 51,000 compounds were “active” in at least one of 5 cell-based assays.
ChIP Smart Library Generation

**AMP Li F IC A T I O N**

- Chemical transformation types:
- Available building blocks:
- Potential synthetic strategies:

**K N O W L E D G E**

- Corporate synthetic expertise
- Reaction knowledgebase

**T I P + A R K**

- Activity and Relevancy Assessment
- Selectivity and Suitability Assessment

**N o v e l , T r a c t a b l e C o m p o u n d L i b r a r i e s**

Guided chemistry evolution directed by family, target, prior-art
Consistent with accessible synthesis methods and commercially available starting materials.
• ~1,300 journal articles from 80 journals and ~80 patents
  ➢ Publications cover parallel, solid-phase and solution-phase methods

• ~15,000 generic reactions → ~1.8 million products
  ➢ Reactions with their starting materials correspond to ~2.8 million specific reactions
  ➢ Reaction exports available in RDFfile/SMIRK and building-blocks/product molecules in SDFfile/SMI formats

• Hierarchically organized chemistry content covers a variety of synthetic methodology
  ➢ Solution phase reactions
  ➢ Solid-phase reactions
  ➢ Polymer-supported solution-phase reactions
Enhanced Enumeration-Ready (“ChIP-able”) Reaction Content

• Reaction transforms – generic reactions in SMIRKS format
  ➢ Associated incompatibility SMARTS filters (“required” and “exclude”)
  ➢ Introspective reactivity filters are included in the SMIRKS representation

• Example reaction types
  Nucleophilic Aromatic Substitution Reactions
  Pd-Catalyzed Aromatic Substitutions
  Functional Group Transformations
  Amine Acylation Reactions (Amides / Carbamates)
  Amine Acylation Reactions (Ureas / Thioureas)
  Formation of Diverse Heterocyclic Systems
  Michael addition, cyclo-condenzation (formation of 4-quinolinones)
  Formation of Thioimidazoles
  Standard Deprotection Steps
  Robinson Annulation
  Diels-Alder Reaction (26 representations)
  Fisher-Indole Synthesis

• Filters of reactive functionalities/undesired motifs in SMARTS
ChIP-able Nucleophilic Aromatic Substitutions

- aliphatic primary or secondary amine
- aromatic primary amine
- 2-chloropyrimidine, but not 4-chloropyrimidine
- aliphatic primary or secondary amine
- aromatic primary amine
- 2-chloropyrimidine, but not 4-chloropyrimidine
- aromatic primary amine
- aliphatic primary or secondary amine
- aromatic primary amine
- aliphatic primary or secondary amine
- aromatic primary amine
- aromatic primary amine
- amine aliphatic primary or secondary
ChIP-able Diverse Heterocycles

any 2-amino arylamide not in a ring; any carboxylic acid; excludes acylators, alkylators, nucleophiles, other acids, activated aryl halides, aryl bromides/iodides

any 2-amino arylamide not in a ring; carboxylic acid chloride, no chloroformate, carbamoyl chloride, etc.; excludes other acylators, alkylators, nucleophiles, acids, activated aryl halides, aryl bromides/iodides

any 2-amino arylamide not in a ring; carboxylic acid methyl ester, does not allow any other ester, which is too restrictive; no differentiation between different ester reactivity; excludes acylators, alkylators, nucleophiles, acids, activated aryl halides, aryl bromides/iodides

any 2-ureido aryl methyl carboxylate not in a ring; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides

any 2-ureido aryl carboxylic acid not in a ring; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides

any 2-hydroxyaryl-(NH)-amide, thioamide, urea, thiourea, etc.; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides

any 2-mercaptoaryl-(NH)-amide, thioamide, urea, thiourea, etc.; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides

any 2-aminoaryl-(NH)-amide, thioamide, urea, thiourea, etc.; not amido aryl amide; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromide/iodide

1,2-diamino aryl (amino-2-arylamine); carboxylic acid methyl ester, does not allow any other ester, which is too restrictive; no differentiation between different ester reactivity; exclude, nucleophiles, acids, acylators, alkylators, activated aryl halides, aryl bromide/iodide, aldehydes, other esters

1,2-diamino aryl; aldehyde; exclude, nucleophiles, acids, acylators, alkylators, activated aryl halides, aryl bromide/iodide
Leveraging Building Block Complexity

Diels-Alder

Amide Formation

Benzimidazole Formation

A landslide of interesting pharmacophores, all just one Rxn step away
Substantial One-Step Diversity
Good Filtering Is Critical...

Each reaction transform is encoded with the information necessary to prevent undesirable products.

Side reactions are avoided using...

*Incompatibility Filters* to analyze available starting materials to exclude multiple reactive functional groups

Synthetic infeasibility is avoided using...

*Introspective filters* to analyze the reaction center environment to exclude reactants with groups that prevent successful synthesis, e.g. nucleophilic amine, activated aromatic chloride,…

Reactive product molecules are removed using...

*Global "bad-frag" filters* to avoid reactive functionality or undesired motifs, e.g. acylators, undesired elements/isotopes…
Filtering Example

**How The Computer Thinks**

RXN Smirk with “introspective filter”

Required
1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

Not Allowed
nucleophiles, acids, acylators, alkylators, activated aryhalides, aryl bromide/iodide, aldehydes, any other esters (no differentiation between reactivities of different esters)

**How The Chemist Thinks**

1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

Required
nucleophiles, acids, acylators, alkylators, activated aryhalides, aryl bromide/iodide, aldehydes, any other esters (no differentiation between reactivities of different esters)

Not Allowed
1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

1) Filtering Example

How The Chemist Thinks

Required
1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

Not Allowed
nucleophiles, acids, acylators, alkylators, activated aryhalides, aryl bromide/iodide, aldehydes, any other esters (no differentiation between reactivities of different esters)

How The Computer Thinks

RXN Smirk with “introspective filter”

Required
1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

Not Allowed
nucleophiles, acids, acylators, alkylators, activated aryhalides, aryl bromide/iodide, aldehydes, any other esters (no differentiation between reactivities of different esters)
De-"Know"-vo Drug Design

**Chemical Intelligence Platform (ChIPTM)**

*Synthetically Focused, Active Guided Lead Design*

Active Seeds From SAR Database

**Kinase Knowledgebase Target Informatics Platform (KKB™/TIPTM)**

**Design Candidates**

**Synthetic, Activity, and Structural Knowledge Leveraged Simultaneously**

Kinase Knowledgebase

Target Informatics Platform

Ligand and Target Analyses

Active Seeds from SAR Database

Ki = 27 nM
Other ChIP Generated PDE-IV “Me-Too’s”
ChIP – “Diversity” Example Simulation

- Start with ~ 40 generic reactions with “introspective” filters
- Generate virtual protocols by graph traversal algorithm based on compatibility of generic representations
- Enumerate virtual protocols using commercially available starting materials as input
- Eliminate structures that overlap with database TDTFile (e.g. KKB)
- Apply Lipinski and structural filters (> 90)
  - MWT, HBA/HBD, ClogP, TPSA, rotBond
  - reactive functionalities like alkylators / acylators, electrophiles, nucleophiles, etc.
  - Undesired motifs (non-standard elements, >2 halo or >1 nitro per aryl, thioesters / ureas, un-branched chains, etc.
- Excerpted set: 28K novel compounds
Example Simulation Results

Compound Novelty (ChIP sim-3 excerpt)

Maximal compound similarity

Frequency

KB "BioMol" (23K)
MDDR (v2202.2) (65K)
ACD (v2003.1) (45K)
KB "BioMol+commercial" (850K)
Excerpted ePotency & eSelectivity (sim3)

FlexiChem Excerpt (sim3)

<table>
<thead>
<tr>
<th>Excerpted ePotency &amp; eSelectivity (sim3)</th>
<th>MoleculeID</th>
<th>eABL</th>
<th>eSRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2270249</td>
<td>7.96; 6.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2269216</td>
<td>7.75; 6.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2357124</td>
<td>7.51; 6.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2280014</td>
<td>7.36; 7.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2267424</td>
<td>7.04; 4.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Activity Screen: MAPK14/p38alpha Pharmacophore Model
High Scoring Reaction Products: 4-Aminopyridopyrimidinones

P38 ATP-Site Directed Simulation
De-"Know"-vo Design Summary

- Diversity and/or focused library evolution through tractable chemistry with corresponding available building blocks
- Flexible evolution direction and scoring: e.g. PFPSim, (Q)SAR, Target Structure-directed, etc.
- Project-based collaboration engagements and/or several components individually licensable (e.g. ARK, Reaction Content, EVE, KKB, TIP, command line enumeration, etc.)