Target Validation and Druggability

SMR
Hot Topics in Drug Discovery:
Finding the Next Lead
11 November 2009

James Mills
and
Stephen Campbell, Anna Gaulton, Lee Harland,
Andrew Hopkins, Jerry Lanfear, Gaia Paolini,
Anne Phelan
Overview

• The target universe
  – The “druggable genome”/proteome/targetome

• Target identification
  – Genetic approaches
  – Data-mining approaches
  – *in vitro* approaches

• Target validation
  – Biological approaches
  – Computational assessments of “druggability”
The Target Universe: Genes

- 30,000 genes from human genome
  - Initial estimate: 600–1500 are “druggable”
    - Capable of being affected by orally-available drug
    - ~300 so far have yielded products

Rich Targets?
Drug targets
~600-1500

Poor Targets?
Disease modifying
~3,000

Druggable
~3,000

Human genome
~30,000 genes

- No consideration of
  - non-human targets
    - Pathogens
  - non-oral routes
    - iv/inhaled
  - non-small molecule
    - Vaccines
    - Antibodies
    - Proteins
    - siRNA

The Target Universe: Proteins

• One gene can give rise to many targets

Gene

mRNA

Protein

Splice variants

- Multiple complexes
  - e.g. ligand-gated ion channels
- Multiple states
  - e.g. GPCR agonist/antagonists
- Post-translational modifications
  - e.g. phosphorylated kinases

• Some targets are combinations of proteins e.g. Gleevec
• Can be target/mode/location e.g. AChE in eye/CNS/NMJ

Target Identification

- Genetic evidence: identification of genes that
  - cause disorder if mutated
  - or increase risk of disorder if mutated
  - protect against disorder if mutated
  - show changes in expression in disease states
  - occur in pathogens but have no human homologue
- Data-driven evidence: identification of proteins that
  - play an important role in disease-associated pathway
- Experimental evidence: identification of targets by
  - chemogenomic screening of small-molecule tools
Mutations Causing Disease

• Leptin and obesity

• Mutation leading to no leptin production
  – Highly obese phenotype

• Developed leptin analogue
  – Worked in patients with this mutation
    • but not in the majority of remaining obese patients
Mutations Protecting Against Disease

• CCR5 and HIV infection
  – 32 base pair deletion in CCR5 gene
  – Receptor does not express on T cell surface
  – No deleterious phenotype
  – Remarkably resistant to HIV infection
  • CCR5 is a co-receptor for viral cell penetration
Changes in Gene Expression

• Compare normal and disease tissue
  – Levels of
    • protein: harder (antibodies)
    • mRNA: easier (hybridisation)
  – Changes imply a target/pathway involved in disease
    • or involved in response to disease
    • e.g. wound healing
      – histone methylases (PcGs, Eed, Suz12, Exh2) ↓
      – histone demethylases (Jmjd3, Utx) ↑

Shaw & Martin (2009) EMBO Rep 10 881
# Targets from Automated Data-Mining

- **e.g. STAT3 involvement in COPD**

<table>
<thead>
<tr>
<th>Loose association</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGF, a potent angiogenesis factor, likely contributes to airway remodeling in asthma.</td>
</tr>
<tr>
<td>The STAT3 inhibitor piceatannol decreased both OSM-induced VEGF release...</td>
</tr>
<tr>
<td>The authors review the progress in understanding how STAT3 and SOC3 regulate the lung inflammatory response.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Another loose association</th>
</tr>
</thead>
<tbody>
<tr>
<td>In both cancer and COPD, the STAT3 gene was up-regulated</td>
</tr>
<tr>
<td>Several STAT3 down-regulated genes also showed differential expression patterns in carcinoma and COPD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Direct association</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3 could either be a target or a biomarker for COPD</td>
</tr>
</tbody>
</table>

---

---
Targets from Small Molecules

- Define target by means of chemical tools
  - Launched products
    - e.g. D2 role in schizophrenia
    - D2 potency vs efficacy
  - Natural products
    - e.g. role of capsaicin in pain
      - Identified TrpV1 as target
  - Designed small-molecule subsets
    - Selective probes to ascertain protein roles
    - e.g. histone modification enzymes

<table>
<thead>
<tr>
<th>Dose</th>
<th>D2 IC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promazine</td>
<td>Clozapine</td>
</tr>
<tr>
<td>Chlorpromazine</td>
<td>Haloperidol</td>
</tr>
<tr>
<td>Fluphenazine</td>
<td>Pimozide</td>
</tr>
<tr>
<td>Benperidol</td>
<td>Spiroperidol</td>
</tr>
</tbody>
</table>

Seeman et al. (1976) Nature 261 717
Caterina et al. (1997) Nature 389 816
Target Validation

• Does target knock-out have the desired effects?
  – Disease-related phenotype, no deleterious effects
• Where in the body is the gene expressed?
  – Does expression vary with e.g. age/gender?
• Are there alternative pathways available?
• Can we prosecute a drug-discovery program?
  – Is the target druggable?
  – Is selectivity necessary/achievable?
  – Is there a suitable animal model system?
    • All of the above for animal model
Knock-ins/downs/outs

- Reduce expression of protein to assess involvement/toxicity
  - Knockout: whole organism
  - Knock-in: gene replaced with non-functional mutant
  - Knock-down e.g. siRNA: organ specific
- Can provide go/no-go decisions but
  - embryonic lethality can be uninterpretable
    - Knock-out can affect development
      - may not replicate effects of reversible antagonism
      - organism can compensate for knock-out
Understand the System: Enzymes

- Rate-determining step
- Functional degeneracy e.g. COX-1/2
- Pathway consequences e.g. COX/5-LO

Diagram:
- Arachidonic acid
- COX-1/2
- 5-LO
- PGH2
- COX-1/2 5-LO
- NSAIDs
- Pain/inflammation
- Gut protection
- Clotting
- LTA4
- LTB4
- LTC4
- LTD4
- LTE4
- Bronchoconstriction
- PGD2
- PGE2
- PGI2
- TXA2
- HPETE
- 5 s⁻¹
- 100 s⁻¹
Understand the System: Receptors

- Concentration and potency of endogenous ligands
- Peptidic GPCRs are notoriously difficult targets (NK1)
  - Endogenous ligand usually sub-nM
  - Occupies small fraction of receptors to give full response
  - Released in huge concentrations at synapses
  - Located in CNS
- What can we do?
  - Agonist projects
  - Systems with low concentrations of circulating hormone
    - $\text{AT}_1$, $\text{ET}_A$, $\text{CCR}_5$
  - Non-competitive and/or slow-offset modulation
Druggability

• A good pocket tends to be
  – the right size: accommodate drug-sized molecule
  – buried: increases interaction surface area
  – not too polar: allow drug-like properties in ligands
Practical approaches

• Screen using representation of chemical space
• More active compounds = more druggable target
  – High throughput screen
    • Thorough
    • Expensive, with false positives and negatives
  – Screen fragment-based library
    • Cover chemical space more effectively
    • Need high-sensitivity assays/biophysical methods

Hajduk et al. (2005) J Med Chem 48 2518
Analytical approaches

- Define pocket
- Compute non-polar SA
- Compute curvature
- Potency estimate

Druggable:
- Factor Xa
- Thrombin
- ACE
- HIV-RT (nuc)
- Neuraminidase
- Cathepsin K
- PTP1b
- ICE1
- HIV-integrase

Undruggable:
- NNRTI
- cAbl
- PDE5
- PDE4D
- HIV-protease
- Prodrug/transporter

Calc pKᵢ

Hajduk et al. (2005) DDT 1675
Cheng et al. (2007) Nature Biotech 25, 71
Selectivity: Site vs Whole Sequence

• Antifungals: broad spectrum and selectivity over human

<table>
<thead>
<tr>
<th>Identity</th>
<th>Human</th>
<th>C. albicans</th>
<th>A. fumigatis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Site</td>
<td>25%</td>
<td>38%</td>
<td>35%</td>
</tr>
<tr>
<td>Site</td>
<td>15%</td>
<td>72%</td>
<td>78%</td>
</tr>
</tbody>
</table>
Caveats in Cavity Analysis

- Irreversible inhibitors do not need as much SA/burial
- Allosteric pockets induced by ligands
  - Cannot predict pocket properties from apo structures

HIV reverse transcriptase

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Volume (Å³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo structure</td>
<td>293</td>
</tr>
<tr>
<td>Efavirenz</td>
<td>577</td>
</tr>
<tr>
<td>Delavirdine</td>
<td>823</td>
</tr>
</tbody>
</table>
Empirical Approaches

• Homology-based
  – Sequence similarity to known druggable proteins

• Non-homology based
  – If structure available, calculate pocket properties
    • Volume, depth, curvature, accessibility, PSA…
    • Apply model based on 400 druggable pockets
  – Calculate sequence-based properties
    • #helices, size of protein…
    • Apply model based on 1400 known targets

• Combine all this information into one score

Agüero et al. (2007) Nature Drug Disc 7 900
TDR Targets Database

- Tropical Disease Pathogen genome database

The TDR Targets Database
Identification and ranking of targets against neglected tropical diseases

Search results for query: #1 (untitled query)

91 records found | Showing page 1 of 4 (records 1-25) | Number of records to display 25

Go to page: next | last

<table>
<thead>
<tr>
<th>Organism</th>
<th>Name</th>
<th>Ortholog group</th>
<th>Product</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. falciparum</em></td>
<td>MAL10P.1.46</td>
<td>OG12_492</td>
<td>AMP deaminase, putative</td>
<td>PlasmoDB 5.0</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>MAL10P.1.166</td>
<td>OG12_98</td>
<td>helicase, putative</td>
<td>PlasmoDB 5.0</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>MAL10P.1.279</td>
<td>OG12_302</td>
<td>cell division control protein 2 homolog</td>
<td>PlasmoDB 5.0</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>MAL10P.1.56</td>
<td>OG12_5262</td>
<td>m1-family aminopeptidase</td>
<td>PlasmoDB 5.0</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>MAL8P.1.156</td>
<td>OG12_963</td>
<td>hypothetical protein</td>
<td>PlasmoDB 5.0</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>PF00_9108</td>
<td>OG12_2639</td>
<td>pepsinogen, putative</td>
<td>PlasmoDB 5.0</td>
</tr>
</tbody>
</table>

http://www.tdrtargets.org
Confidence in Druggability

**Sequence-based Druggability**

- Predicted druggability based on primary sequence features?
- Protein structure contains drug-like pockets?
- Does this protein bind any endogenous drug-like small molecules?

**Predicted pocket**

- Sequence Analysis
- PDB Protein-Structure Mining

**Ligands**

- ChEBI / Metabo-cards

**Compounds**

- Are there any known chemical tools for this protein?
- ChEMBL, Internal Data

**Clinical**

- Drugs for this target in clinical trials for any indication?
- Industry Databases

**Clinical**

- Is the protein an established small-molecule drug target for any indication?
- Industry Databases, DrugBank

**Novel**

- Precedented
Target Opportunity Universe

Confidence in druggability vs confidence in mechanism

High-risk but testable
New indications for cpds
Done

Campbell et al. (2009) DDT in press
Summary

• 2-3 targets per year yield launched drugs
• Many ways to associate a target with a disorder
• Many ways to further explore this link
  – Target confidence building rather than validation
• Reasons for optimism?
  – Technology is moving on
    • Molecular biology
    • Data mining
  – Seeds of change in culture
    • Pre-competitive research
Genes Unique to Pathogens

- Genomes of pathogens tend to be smaller
  - Less degeneracy: more proteins essential
- Some pathways not found in humans
  - *e.g.* cell wall synthesis, folate synthesis
- Easier to prioritise genome experimentally

**Yeast genome triage**

**YPD data**
- Lethal (777)
- Unknown (2646)
- Viable (2722)

**PDB neighbours**
- Beautiful
- Ugly
- GotNoLigand

- 166 LB targets
- 76 targets