Can We Really do Computer-Aided Drug Design?

ACS Spring National Meeting, March 29th 2012
Matt Segall
Overview

• Design vs. Discovery

• Accuracy of predictive models in drug discovery

• How accurate do models need to be?

• Adding value with predictive models

• Moving toward drug design

• Conclusion
Design vs. Discovery
Design vs. Discovery

Design

Discovery
An Analogy of Drug Design
The Boeing 777*

- Designed entirely on computer
- Full-scale prototype built
- Successfully flown first time
- Compared with the “crash test” paradigm of drug discovery

Why Does this Analogy Break Down?
Complexity of Design Goals?

Airplane

• Cost
• Efficiency
• Range
• Capacity
• Safety
• ...

Drug

• Potency
• Selectivity
• Absorption
• Metabolic Stability
• Safety
• .....
Why Does this Analogy Break Down?
Precision of Models

Airplane

\[
\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0
\]

\[
\rho \frac{D\mathbf{v}}{Dt} = -\nabla \rho + \nabla \cdot \mathbf{T}^* + \rho \mathbf{f}
\]

\[
\rho \frac{D\mathbf{e}}{Dt} = -\rho \nabla \cdot \mathbf{v} + \Phi - \nabla \cdot \mathbf{q} + \rho \mathbf{r}
\]

Drug

Caco2 vs Human Intestinal Absorption*

\[ R^2 = 0.81, \text{RMSE} = 0.8 \log \text{units} \]

How Accurate are Predictive Models?
2D QSAR Models of Target Potency*
Root Mean Square Error

- Average RMSE on validation set = 0.76 log units (factor of 5.8)
- Average RMSE on test set = 0.8 log units (factor of 6.3)

* Segall et al. ACS Spring National Meeting, 2012 COMP Thursday 2pm, Room 28E
Other Methods

Some examples

• 2D and 3D Similarity
  – Hit-rate of 20-30% among most similar compounds*

• Docking
  – Similar hit-rate, 20-30% †

• Structure-property relationships
  – Solubility models found to have RMSE of between 0.47 to 1.96 log units on 122 drugs‡

† Kroemer RT. (2007) Curr. Protein Pept. Sci. 8:312-328
How Accurate Do Models Need to Be?
How Well Does this Model Help Us to Identify Active Compounds?

• In your screening deck, you expect to have a hit-rate of 0.1% against a target

• You choose to use a predictive model to classify active compounds to prioritise for screening
  - The model is 90% accurate (90% specific and 90% sensitive)

• What proportion of compounds that are predicted to be active actually are?
  - a) about 0.1%
  - b) about 1%
  - c) about 10%
  - d) about 50%
  - e) about 90%

• Answer: b)
  - E.g. Of 10,000 compounds 9990 x 0.1 + 10 x 0.9 = 1008 would be predicted as active, of which only 9 really are.
What Prior Probability Do We Need for a 90% Accurate Model to be Useful?

• Depends on what we mean by useful!
  – E.g. 1 in 10 compounds predicted to be active would be expected to be confirmed

• Answer: 1.2%

• Required accuracy depends on the prior probability
  – Until we know this, we don’t know the accuracy we require
Sequential Filtering
Compounding errors

Potency

Absorption

Metabolic Stability

CAUTION
Adding Value With Predictive Models
Probabilistic Scoring

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>logIC50</td>
<td>$\leq -0.3$</td>
<td></td>
</tr>
<tr>
<td>logS</td>
<td>$&gt; 1$</td>
<td></td>
</tr>
<tr>
<td>HIA category</td>
<td>$+$</td>
<td></td>
</tr>
<tr>
<td>logP</td>
<td>$0.0 -&gt; 3.5$</td>
<td></td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>$\leq 5$</td>
<td></td>
</tr>
<tr>
<td>BBB log([brain]:[blood])</td>
<td>$-0.20 -&gt; 1.00$</td>
<td></td>
</tr>
<tr>
<td>2D6 affinity category</td>
<td>low medium</td>
<td></td>
</tr>
<tr>
<td>BBB category</td>
<td>$+$</td>
<td></td>
</tr>
<tr>
<td>P-gp category</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>2C9 pKi</td>
<td>$\leq 6$</td>
<td></td>
</tr>
<tr>
<td>PPB category</td>
<td>low</td>
<td></td>
</tr>
</tbody>
</table>

Importance of Uncertainty

Desired value > Threshold

Property Y

100
10
1
0.1

UNDESIRABLE

DESIRABLE

A B C

X X X X

© 2012 Optibrium Ltd.
StarDrop Prioritisation
Probabilistic Scoring

- Property data
  - Experimental or predicted
- Criteria for success
  - Relative importance
- Uncertainties in data
  - Experimental or statistical

- Score (Likelihood of Success)
- Confidence in score

Error bars show confidence in overall score. Bottom 50% may be rejected with confidence as error bars overlap.

Data do not separate these, as error bars overlap.

Score

Compounds ordered by score

Best

Worst

© 2012 Optibrium Ltd.
Visualising ‘Chemical Space’
Exploring trends in chemical diversity

‘Hot spot’ of good compounds
Balance Quality Against Diversity
Mitigating risk

Key
Bad
Good

Inset graph:
- X-axis: Compounds ordered by score
- Y-axis: Score
- Red line represents higher scores
- Blue line represents lower scores
Moving Towards Drug Design
Improve Accuracy of Prediction

• Better modelling algorithms?
  – Advanced machine learning, e.g. random forests, Gaussian processes, support-vector machines...

• Better data?
  – Always welcome! But, lots more than data is available than ever before, e.g. PubChem, PDB, Chemble, Bindingdb...

• Better descriptors?
Structural Descriptors
Better Description of Physics/Chemistry
E.g. Fields
• Quantum mechanics captures electronic properties and energetics with a high degree of accuracy
  – Slow
  – But, becoming more accessible on a routine basis

• Examples:
  – Hydrogen bonding acidity
  – Lability to metabolism
  – Binding energies
  – Classical MD parameterised using DFT
Conclusions

• Predictive models are not yet accurate enough to enable a true drug design paradigm

• However, models provide value by helping to reduce wasted effort and focus efforts on chemistries with the best chance of success

• QM approaches may offer one way to move towards true drug design
  – Still some way to go before these methods can be routinely used

• Of course, modelling also adds value by helping to understand and interpret SAR