Enabling Interactive Multi-Objective Optimization for Drug Discovery Scientists

Matthew Segall – matt.segall@optibrium.com
Lorentz Workshop – Optimizing Drug Design

Optibrium™, STarDrop™, Auto-Modeler™ and Glowing Molecule™ are trademarks of Optibrium Ltd.
Overview

• Multi-objective optimisation in drug discovery
  – Issues and analogies

• Guiding decisions in drug discovery
  – Balancing properties
  – Selecting compounds, diversity vs. quality
  – Feedback to guide design

• Helping the decision-makers
  – Challenges to delivering technological solutions

• Challenging the decision-making process
  – Case study

© 2009 Optibrium Ltd. 2
Multi-objective optimisation: Challenges and analogies
The Objectives

• Identify chemistries with an optimal **balance** of properties

• Quickly identify situations when such a balance is not possible
  – Fail fast, fail cheap
  – Only when **confident**
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?**

<table>
<thead>
<tr>
<th>Airplane</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Cost</td>
<td>• Potency</td>
</tr>
<tr>
<td>• Efficiency</td>
<td>• Selectivity</td>
</tr>
<tr>
<td>• Range</td>
<td>• Absorption</td>
</tr>
<tr>
<td>• Capacity</td>
<td>• Metabolic Stability</td>
</tr>
<tr>
<td>• Safety</td>
<td>• Safety</td>
</tr>
<tr>
<td>• ...</td>
<td>• ....</td>
</tr>
</tbody>
</table>
Why Does this Analogy Break Down?
Precision of Models

Airplane

\[
\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0
\]
\[
\rho \frac{D\mathbf{v}}{Dt} = -\nabla \rho + \nabla \cdot \mathbf{T} + \mathbf{f}
\]
\[
\rho \frac{De}{Dt} = -\rho \nabla \cdot \mathbf{v} + \Phi - \nabla \cdot \mathbf{q} + \rho \mathbf{r}
\]

Drug

Caco2 vs. Human Intestinal Absorption*

* R^2 = 0.81, RMSE = 0.8 log units

An Alternative Analogy
Card Counting in Blackjack*

• Uniquely among casino games, the outcome of a Blackjack hand is, to some degree, predictable

• The cards that have been dealt and discarded define the probabilities of drawing cards in the future

• High cards (10 through Ace) favour the player over the dealer

• Card counters use this information to *bias the odds* in their favour

• N.B. This is not a recommendation of card counting, it may be illegal in some jurisdictions.

* Bringing Down the House, Ben Mezrich
An Alternative Analogy
Card Counting in Blackjack

Discarded

Remaining

$$ \text{Discarded: } +11 $$

$$ \text{Remaining: } $$$ $$

$$ \text{Discarded: } -10 $$

$$ \text{Remaining: } $ $$
An Alternative Analogy
Card Counting in Blackjack
How does this apply to drug discovery?
Guiding decisions in drug discovery
Challenges of Decision-Making in Drug Discovery

• Importance of multiple, sometimes conflicting, criteria to the success of a potential drug molecule
  – Different degrees of importance of properties

• Lots of data
  – Potentially large numbers of compounds
  – Multiple properties

• Uncertain information
  – Variability/error in data or predictions
  – *in silico* predictions, *in vitro* assays and *in vivo* models are only approximations of the ultimate human target

• Objectives
  – Find the ‘hot’ decks
  – Deciding when to ‘kill’ a project/chemical series – no ‘hot’ decks
Applying Data to Making Decisions

Value created by good decisions, not data

Data → Prioritise → Selection

- In silico
- In vitro
- In vivo

Importance
Uncertainty
Quality
Diversity
‘Manual’
Balancing Properties
Defining your objectives

Profile: Diabetes

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>logIC50</td>
<td>≤ -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 → 3.5</td>
<td></td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>≤ 5</td>
<td></td>
</tr>
<tr>
<td>BBB log([brain]:[blood])</td>
<td>-0.20 → 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>≤ 6</td>
<td></td>
</tr>
<tr>
<td>PPB category</td>
<td>low</td>
<td></td>
</tr>
</tbody>
</table>
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. Data at bottom 50% may be rejected with confidence, as error bars overlap.
Bringing all data together
‘Kill’ with confidence!
Bringing all data together
Launched drugs scored for oral, non-CNS

Most drugs are predicted to have high chance of success.

Note: Set includes non-oral drugs
Objective: Select 200 compounds from scored library of 13,000 compounds
Selecting Compounds
Balancing Quality and Diversity

Balanced Diverse Sample Top 200 ranked compounds Balance Diversity: Rank = 80:20
Assessing Quality vs. Diversity

Selecting 20 compounds from library of ~270.

Full library

Score: Diversity = 60:40
Feedback for (Re)Design

Data → Prioritise → Selection

Redesign

© 2009 Optibrium Ltd.
Interpreting Models

Motivation

Desc. 1

Desc. 2

Property
Interactive Redesign: The ‘Glowing Molecule’

• Provides visual interpretation of structural influences on predicted properties
  – “Why is a property value predicted?”
  – “Where can I change this property?”
  – Interpret SAR
  – Guide efficient redesign of molecules

• Applies to linear and non-linear
  – No-more ‘black box’ models!
  – Individual properties or scores
Helping the decision makers
Who Are the Decision Makers?

• Multi-disciplinary project teams
  – Medicinal chemistry
  – Biology/pharmacology
  – DM/PK
  – Computational chemistry/cheminformatics
  – Bioinformatics
  – Toxicology

• Implications
  – Speak different scientific ‘languages’
Who Are the Decision Makers?

- Multi-disciplinary project teams
  - Medicinal chemistry
  - Biology/pharmacology
  - DM/PK
  - Computational chemistry/cheminformatics
  - Bioinformatics
  - Toxicology

- Implications
  - Speak different scientific ‘languages’
  - Most are non-computational scientists
  - Uncomfortable with uncertainty
  - Scientists want to understand
Bridging the Language Gap

Link the realms of biology and chemistry

Biology

Chemistry

© 2009 Optibrium Ltd.
Bridging the Language Gap

Link the realms of biology and chemistry
Bridging the Language Gap

Link the realms of biology and chemistry
Accessibility

What?

100% value

'Nobody' used it!

Ahah!

Experts can access 100%
No Black Boxes

Score = 0.18

Why?

Why?
Make Uncertainty Explicit
Interactivity

- Project timelines often mean that significant delays before making decisions can’t be tolerated
  - Analysis by experts in a particular domain can often lag behind by an iteration
  - Resources are often wasted making/testing compounds that analysis would have discarded if available in time

- Scientists don’t like to wait!
  - Will often make a decision ‘right now’ rather than wait for analysis by someone else

- Rapid feedback helps with learning
Automatic Model Generation

• Splitting data into training, validation and test sets
  – cluster analysis by structural similarity at certain Tanimoto level
  – Y based
  – random
  – manual split

• Descriptor calculation and filtering
  – 2D SMARTS, logP, TPSA, charge...
  – user defined SMARTS
  – imported descriptor values, e.g. experimental data
Automatic Model Generation

- Modelling techniques applied to training set
  - Partial Least Squares (PLS)
  - Radial Basis Functions with Genetic Algorithm
  - Gaussian Processes
  - Decision Trees (category models)

- Selection of the best model by performance on the validation set
  - $R^2$, RMSE for continuous models
  - Accuracy, kappa-statistic for category models

- Test set is an independent set
Automatic Model Generation

Data set

trn  val  test

Build models

PLS  RBF  GPs

Best model

Evaluate multiple models

Test the best model

New compounds

Prediction and uncertainty

© 2009 Optibrium Ltd.
Integration
Don’t work in isolation

StarDrop™
Balanced compounds from better decisions

Compound Database
SQL via Python

StarDrop Model Server
Plug-in

In-house Models

Web services

Other informatics platforms, e.g. Pipeline Pilot

© 2009 Optibrium Ltd.
Challenging the Decision-Making Process
Challenging the Decision-Making Process

- A robust scoring algorithm allows objective prioritisation of compounds against a profile of required properties

- However, we can also reverse the process to test the decision making process

- Assess the impact of different processes/criteria on the decisions we may make
  - Do different processes give rise to different outcomes?
  - If so, we need to carefully consider the process and criteria
  - If not, we can proceed with confidence
Illustrative Example

• *In vitro* data being generated
  – Potency
  – Selectivity
  – Solubility
  – Microsomal stability (rat and human: RLM and HLM)

• Original process focused on potency and selectivity, filtering compounds that did not meet requirements

• Results
  – Low but prolonged activity after IP dosing
  – No correlation between *in vitro* and *in vivo* potency
  – Problems with solubility and metabolic stability
### Historical Process - Filtering

#### No uncertainty

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>log Selectivity</td>
<td>$&gt; 0.9$</td>
<td></td>
</tr>
<tr>
<td>pIC50</td>
<td>$&gt; 6$</td>
<td></td>
</tr>
</tbody>
</table>

Selectivity 11 fold  
Potency 0.21 uM

<table>
<thead>
<tr>
<th>Name</th>
<th>Profile 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX322</td>
<td>1</td>
</tr>
<tr>
<td>XXX326</td>
<td>2</td>
</tr>
<tr>
<td>XXX137</td>
<td>3</td>
</tr>
<tr>
<td>XXX292</td>
<td>4</td>
</tr>
<tr>
<td>XXX104</td>
<td>5</td>
</tr>
<tr>
<td>XXX318</td>
<td>6</td>
</tr>
<tr>
<td>XXX540</td>
<td>7</td>
</tr>
<tr>
<td>XXX313</td>
<td>8</td>
</tr>
<tr>
<td>XXX160</td>
<td>9</td>
</tr>
<tr>
<td>XXX289</td>
<td>10</td>
</tr>
<tr>
<td>XXX280</td>
<td>11</td>
</tr>
<tr>
<td>XXX502</td>
<td>12</td>
</tr>
<tr>
<td>XXX582</td>
<td>13</td>
</tr>
<tr>
<td>XXX572</td>
<td>14</td>
</tr>
<tr>
<td>XXX316</td>
<td>15</td>
</tr>
</tbody>
</table>
Taking Uncertainty into Account

With uncertainty

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>log Selectivity</td>
<td>&gt; 0.9</td>
<td></td>
</tr>
<tr>
<td>pIC50</td>
<td>&gt; 6</td>
<td></td>
</tr>
</tbody>
</table>

Selectivity 7 fold
Potency 0.12 uM

<table>
<thead>
<tr>
<th>Name</th>
<th>Profile 2</th>
<th>Profile 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX326</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>XXX137</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>XXX322</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>XXX313</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>XXX540</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>XXX160</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>XXX572</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>XXX104</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>XXX292</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>XXX541</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>XXX318</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>XXX561</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>XXX280</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>XXX537</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>XXX502</td>
<td>15</td>
<td>12</td>
</tr>
</tbody>
</table>
Comparing With and Without Uncertainty

![Graph showing IC50 and Selectivity with and without uncertainty.](image)
## Looking for Balance

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>log Selectivity</td>
<td>&gt; 0.9</td>
<td></td>
</tr>
<tr>
<td>Expt. Sol.</td>
<td>&gt; 100</td>
<td></td>
</tr>
<tr>
<td>pIC50</td>
<td>&gt; 6</td>
<td></td>
</tr>
<tr>
<td>Expt. HLM</td>
<td>≤ 60</td>
<td></td>
</tr>
<tr>
<td>Expt. RLM</td>
<td>≤ 60</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Profile 3</th>
<th>Profile 2</th>
<th>Profile 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX572</td>
<td>1</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>XXX518</td>
<td>2</td>
<td>52</td>
<td>36</td>
</tr>
<tr>
<td>XXX582</td>
<td>3</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>XXX295</td>
<td>4</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>XXX321</td>
<td>5</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>XXX025</td>
<td>6</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>XXX502</td>
<td>7</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>XXX274</td>
<td>8</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>XXX292</td>
<td>9</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>XXX316</td>
<td>10</td>
<td>44</td>
<td>15</td>
</tr>
<tr>
<td>XXX280</td>
<td>11</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>XXX278</td>
<td>12</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td>XXX319</td>
<td>13</td>
<td>51</td>
<td>44</td>
</tr>
<tr>
<td>XXX294</td>
<td>14</td>
<td>17</td>
<td>24</td>
</tr>
<tr>
<td>XXX282</td>
<td>15</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

- Selectivity 11 fold
- Potency 0.21 uM
- Solubility 136 uM
- HLM 36%
- RLM 86%

- Selectivity 5 fold
- Potency 1.67 uM
- Solubility 138 uM
- HLM 4%
- RLM 38%
$IC_{50}$ and Potency vs. All Data
Conclusion to Example

• Balancing the key properties in a drug molecule is important from the outset

• Probabilistic scoring takes uncertainty into account and reduces bias in decision making
  – More appropriate compounds brought forward
  – Broader view of SAR taken
  – Improved in vivo profile achieved

• An objective scoring framework allows the impact of different criteria to be rigorously compared
  – When are your decisions sensitive to the criteria you set?
  – Prospective application – future research
Conclusions

• A robust approach to guiding decisions on compound selection and design must take into account:
  – Multiple criteria with different degrees of importance
  – Uncertainty in the underlying data
  – A balance of diversity and ‘quality’

• These methods can be placed directly in the hands of the key decision makers
  – ‘Human’ factors are critical to adoption

• Knowing when we have the ‘right’ criteria can be tricky
  – Test criteria and heuristics and understand limitations
Acknowledgements and References

• Acknowledgements – StarDrop group past and present
  - Ed Champness
  - Olga Obrezanova
  - Chris Leeding
  - Alan Beresford
  - Dan Hawksley
  - Joelle Gola
  - Brett Saunders
  - Simon Lister
  - Mike Tarbit
References

• General Interest

• Scoring, Glowing Molecule and Chemical Space

• Gaussian Processes and Automatic QSAR modelling

www.optibrium.com
Backup Slides
Importance of Uncertainty

Desired value > Threshold

Property Y

© 2009 Optibrium Ltd.
Importance of Uncertainty

Desired value > Threshold

Property Z

© 2009 Optibrium Ltd.
• A chemical space allows you to visualise trends across your data set
• Each point represents one compound
• The closer two points are the greater their similarity
  – Structure
  – Properties
  – Mixed
• A space is defined by a single data set...
• ...but other data sets can be plotted in that space at the same time
Chemical Space
Zooms to appropriate resolution

Space defined by diverse drug-like chemistry (black)

Space defined by virtual library (blue)

Virtual library (blue) and selection (red)
Automatic Models versus Manual

• Model built ‘manually’ by computational chemists

• Model from the automatic model generation process (apply to all compounds in a dataset)

• Compare ‘automatic’ and ‘manual’ models by testing on external data set

• Two data sets:
  - Blood-brain barrier penetration (151 compounds with logBB values)
  - Intrinsic aqueous solubility (3313 compounds with logS values from PHYSPROP database)
Blood-Brain Barrier Penetration

• ‘Manual’ model
  – 2D SMARTS descriptors reduced by FVS, various modelling techniques (PLS, RBF, MLR) – performance supervised on test set
  – Final model is built by RBF on 7 descriptors (logP, flexibility, charge, hydrogen bonding...)

• ‘Automatic’ model built by Gaussian Processes with nested sampling on 162 descriptors

• External set: 143 compounds from ‘Abraham’ set
  – Abraham et al. J.Pharm. Sci., 2006, 95

---

<table>
<thead>
<tr>
<th>Test set</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.73</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.36</td>
</tr>
</tbody>
</table>

manual

<table>
<thead>
<tr>
<th>Val+Test set</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ val</td>
<td>0.72</td>
</tr>
<tr>
<td>$R^2$ test</td>
<td>0.66</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.44</td>
</tr>
</tbody>
</table>

automatic
Blood-Brain Barrier Penetration
Performance on external ‘Abraham’ test set

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE pred</th>
<th>% pred within ±0.4 log unit</th>
<th>% pred within ±0.8 log unit</th>
<th>R²</th>
<th>r² corr</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>manual</td>
<td>0.36</td>
<td>62.9</td>
<td>93.0</td>
<td>0.39</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>automatic</td>
<td>0.44</td>
<td>63.6</td>
<td>90.9</td>
<td>0.27</td>
<td>0.36</td>
<td>0.49</td>
</tr>
</tbody>
</table>

![Manual scatter plot](image1.png)

![Automatic scatter plot](image2.png)
Aqueous Solubility

• Manual model is built by RBF method on ~ 100 descriptors

• Automatic model is produced by Gaussian Processes with 2D search

• External test data – 564 compounds from ‘Huuskonen’ set

<table>
<thead>
<tr>
<th>Test set</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Val+Test set</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R² val</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>R² test</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.69</td>
<td></td>
</tr>
</tbody>
</table>

manual

automatic
Aqueous Solubility
Performance on external ‘Huuskonen’ test set

<table>
<thead>
<tr>
<th>Model</th>
<th>Desc</th>
<th>% pred within ±0.7 log unit</th>
<th>% pred within ±1.4 log unit</th>
<th>R²</th>
<th>r² corr</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>manual</td>
<td>108</td>
<td>39.9</td>
<td>70.9</td>
<td>0.68</td>
<td>0.80</td>
<td>1.28</td>
</tr>
<tr>
<td>automatic</td>
<td>166</td>
<td>54.1</td>
<td>85.9</td>
<td>0.82</td>
<td>0.86</td>
<td>0.96</td>
</tr>
</tbody>
</table>

© 2009 Optibrium Ltd.
Data Doesn’t Distinguish Best with Confidence!

Snake Plot for Expt

![Snake Plot](image-url)