Predictive Modeling of On-And-Off Target Bioactivities

Presented By  Tim Hoctor, Vice President Professional Services
Date  Sept 16, 2016

KNIME Fall Summit 2016 - San Francisco
Today’s Presentation

• Elsevier is a partner and vendor to 100% of the world’s top pharma
• Providing leading data solutions in chemistry, biology, drug safety and biomedical review
• Workflow solutions are based on clearly defined use cases

BUT:
• As with many data-driven companies, our data is provided for specific use

Our Challenge:

• How to externalize data for:
  • Complementary data integration
  • Utilization for additional use cases and workflows
Research Data Needs in Life Sciences

External Collaborations
CRO - across whole process

In-licensing – Repurpose drugs

Clinician – assessing disease Progression (incl Biomarkers)

Data associated to Targets/Diseases/Drug

Patient – phenotype his/her disease

Drug Safety/Efficacy

PK/PD – Pharmacol Effects

Understand the biological system

Make/obtain compound series (chemicals and biologicals)

Compound/Agent with optimal pharmacological properties Chemicals and biologicals

Metabolizing Enzymes – Drug Metabolism routes
The Challenges Pharma Faces

“We cannot fail for reasons we could have predicted. We should fail only for reasons we could not predict”

- Dr Moncef Saloui, Head of Global R&D, GSK

“Usability & Re-Usability of data is critical to our success. We are experts at capturing data for a purpose, and poor at making that data available for any other purpose”

- Dr Mark Ramsey, Chief Data Officer, GSK
Developing an API strategy

Data processing engine (mostly Knime or PipelinePilot) queries and retrieves data

Elsevier data cloud

“Server”

Analysis and visualization

“Client”

By using Knime, we can help our customers by providing “out of the box” Reaxys API components

Modeling

QSAR/QSPR
Challenge: Make Best Decisions Based on Data

Issue

• Need to make best possible decisions to rank/decide on drug candidates
• Use all available information to help make decisions about potential therapeutics

Method

• Build hundreds of models for on and off-target activities
• Create a model building engine in KNIME that can make models for large lists of targets using data from Reaxys Medicinal Chemistry
  • Selectivity panels
  • Toxicology pathways (from pathway analysis)
  • PK/ADME models
Random Selection to ‘Flatten’ Activity Distribution

Raw reports – may be thousands for a popular target. Activities and structure distribution may skew statistics.

Randomly selected ‘N’ compounds per decade

More diverse set of structures and more uniform distribution of activities.
Classification Using Z Value for Confidence Scoring

Active/inactive classification can be adjusted by error of each individual model as measured by test set

- When using model for an activity cutoff one uses a single-tailed test: “is value greater than this value”
- Individual model $\sigma$ is used to adjust the predicted value to make an activity classification
High-Level Workflow for Creating Models

- Uses a list of UNIPROT target names, and NCBI synonyms
- Attempts to create a model for each target
  - Rejects models for insufficient data, or low quality
Model Engine

- Creates series of Reaxys queries for each decade of activity
- Uses PLS/R to create and evaluate models
- Creates spreadsheet with RMS error of prediction and plots
Spreadsheet created for each model

- Training set, test set
- Statistics for training and test sets.
Prediction of Activities

- Reads in all models
- Reads in a set of molecules in SDF format
- Predicts activity for each model
A tale of two JAK2 inhibitors - fedratinib and ruxolitinib

Fedratinib – development stopped due to brain injury
Ruxolitinib – a successful drug

Could the toxicity of fedratinib have been anticipated?
- Analysis of kinase selectivity
- Pathway analysis for kinases related to brain edema/toxicity
- Identification of targets related to toxicities → IRAK3
- Use of predictive models for many targets to anticipate toxicities based on predicted/measured bioactivities and pathways
Find Most Relevant Regulators of Cerebral Edema

Scan Literature For Possible Adverse Events Related to Target Modulation

1. Map all key regulators Encephalopathy and related brain swelling diseases in Pathway Studio
2. Only retain regulators with strong evidence of involvement in disease pathology (at least 10 references)
3. For sake of presentation both positive and negative regulators were retained but were removed from further analysis for clarity sake
4. Only negative regulators will potentially lead to brain swelling when inhibited.
Prediction of Panel of Kinase Activities

Predicted Activities for Comparison; with Selectivity Metric

Target associated with efficacy highlighted
Result of Analysis: IRAK3 activity may be marker for cerebral edema toxicity - models can provide a way to screen in silico for this issue

<table>
<thead>
<tr>
<th>Kinase inhibitor</th>
<th>JAK2 activity</th>
<th>IRAK3 activity</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>staurosporine</td>
<td>10</td>
<td>7.8</td>
<td>multi-toxic</td>
</tr>
<tr>
<td>lestaurtinib</td>
<td>9</td>
<td>7.8</td>
<td>patient had cerebral event</td>
</tr>
<tr>
<td>fedratinib</td>
<td>8.9</td>
<td>6.9</td>
<td>development stopped</td>
</tr>
<tr>
<td>gefitinib</td>
<td>5.8</td>
<td>5.8</td>
<td>approved</td>
</tr>
<tr>
<td>vandetanib</td>
<td>5.6</td>
<td>1</td>
<td>approved</td>
</tr>
<tr>
<td>ruxolitinib</td>
<td>10.4</td>
<td>1</td>
<td>approved</td>
</tr>
</tbody>
</table>
**Conclusions**

- One can make a large set of models from the Reaxys Medicinal Chemistry data set
  - Data from journals, and patents

- These models can then be used for in-silico screening and prioritization of compounds for
  - Selectivity
  - Toxicity related to specific targets

- Pathway analysis can direct attention to models/targets that may be predictive of specific toxicities
More importantly

Using this approach and understanding, Elsevier:

• Provided a set of ‘out of the box’ KNIME nodes at no additional cost to their product subscription will help our customers maximize value of their data
• Built internal competencies in our Professional Services team to extend KNIME nodes and apply to custom solutions
• Strengthened our API strategy
• Gained insights into researcher needs
• Better served our customers
Acknowledgements

- Dr Matthew Clark - Elsevier
- Dr Jim Rinker - WuXi/Elsevier
Thanks and questions
Appendix

The following slides detail the full presentation given by Dr. Matthew Clark on developing and predictive models using the Elsevier Reaxys Medicinal Chemistry data set and KNIME. For more detail, please contact t.hoctor@elsevier.com.

For further information on modeling and data services, please see https://www.elsevier.com/solutions/professional-services
Conversion of % inhibition

- If Concentration of the compound is not Available  \( p[X] \) is not calculated
- IF CONCENTRATION Available as a range
  - \( p[X] \) is not calculated
- If concentration is available as a single value
  - \( p[X] \) is calculated

- If % inhibition > 25
  \[
  IC_{50} = \frac{100 \times [C]}{\text{% inhibition} - [C]}
  \]
- If % inhibition < 25
  \( p[X] = 1 \)
Qualitative results

- **Reported as Inactive**
  - \( p[x] = 1 \)

- **Reported as Active**
  - Concentration reported
    - \( p[X] = - \log_{10}(\text{Concentration}) \)
  - Concentration is not reported
    - \( p[X] \) not calculated
Process for Computing pX

1. Data
   - Conc Unit?
     - concentration
     - Not concentration
       - Conversion to concentration
         - Not convertible
           - pX not calculated
         - Low quality
           - Acceptable quality
             - Range processing
               - Yes
                 - Range
                   - Log unit?
                     - No
                       - pX not calculated
                     - Yes
                       - pX
     - Quality metric?
       - Yes
         - Range
           - Log unit?
             - No
               - pX not calculated
             - Yes
               - -log_{10}(affinity)
             - No
               - Log unit?
                 - Yes
                   - pX
                 - No
                   - pX not calculated
Elements of the Model Making Workflow

- **Issue:** some targets have thousands of measurements
- **Solution:** randomly select N examples from each decade of pX activity
  - Selecting from each decade provides a more balanced data set across the activity range
  - In this study only IC\textsubscript{50} measurements are used, not % inhibition or K\textsubscript{d}
  - N is from 50 to 200 in these examples.

- **Compounds are selected so that training and test sets have no overlap**
  - Use random selection of both

- **For each compound the median pX value is chosen when multiple measurements are reported**
  - pX = -\log(\text{molar activity})

- **Partial-least-squares correlation is used, selecting the optimal number of descriptor components**

- **Model quality is assessed**
  - I like RMS error of prediction because it most directly answers the question “\textit{how accurate do I expect the prediction to be?”}
  - \( r^2, F \) and other statistics are also computed.
Huuskonen Molecule Sets – *Predicted with Model Created from Reaxys Data Set*

- Same molecule sets – **Model Trained with Reaxys Training Set**
- Standard error 0.98 log units – not bad
Random Selection to ‘Flatten’ Activity Distribution

Raw reports – may be thousands for a popular target. Activities and structure distribution may skew statistics

Randomly selected ‘N’ compounds per decade

More diverse set of structures and more uniform distribution of activities
**Input:**
- target name,
- number of compounds to use from each decade of log(Activity)
- activity type, e.g. IC50

**Output:**
- model from training set
- predictions of test set
The Importance of the Test Set

- Powerful statistical methods can select variables to create a model that looks great on the training set

  - *However the real test is how well the model can predict activity for a compound not in the training set.*

- In these examples the Training and Test sets are not overlapping; no test compound is in the training set.

- In this study the best metric for the model is the RMS Error of Prediction – the errors seen when predicting activities for compounds not in the training set.
  
  - This is useful to provide the “error bar” for the predictions.
  - The required accuracy depends on what you are using the prediction for.
  - The Pearson $r^2$ provides a metric of “fit” but is dependent on the range of the data and does not directly tell you how much error to expect.

  - Other metrics for prediction might include how similar the compound is any compounds in the training set. One might expect the predictions to be less reliable if the compound is very different.
Descriptors

- The descriptors are properties computed for the compounds that are related to the activities.
- They encode elements of the structure and other properties of the compounds.
- In this study we combined 2 sets of computed descriptors:
  - RD Kit
  - CDK Toolkit
R Script for Partial Least Squares

Create model with all possible components

```r
library(pls)
model <- plsr(pX ~ ., data = knime.in)
# compute RMS error of prediction for each number of components
components <- RMSEP(model)
optimalComps = length(components$comps) - 1

# find optimal number of components in the model based on decreasing rms error of prediction for the models.
for (i in 1:optimalComps) {
  if (components$val[i] - components$val[i+1] <= 0.0001) {
    optimalComps = i
    break
  }
}
```

Find model with optimal # of components

```r
# make final model with optimal components
model <- plsr(pX ~ ., data = knime.in, ncomp = optimalComps)
```

Make final model with optimal # of components

```r
sprintf("number optimal components: %d   RMS error of prediction: %2.3f", optimalComps, components$val[optimalComps] )
```
This model is moderately useful to separate best from worst molecules.

The test set statistics are influenced by 2 outliers with absurdly high predicted binding. Without those, the rms error is about 0.9.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>0.71</td>
</tr>
<tr>
<td>Test Set</td>
<td>1.3</td>
</tr>
</tbody>
</table>
p38a (Q16539)

- p38a has 2 binding sites ('normal' and allosteric) which make a simple QSAR more difficult if the compounds are not separated
- Test set RMSEP suggests predictive value is limited

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>1.15</td>
</tr>
<tr>
<td>Test Set</td>
<td>1.45</td>
</tr>
</tbody>
</table>
ABL Kinase (P00519)

- Test set suggests reasonable predictive power to distinguish active from inactive compounds in-silico

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>0.63</td>
</tr>
<tr>
<td>Test Set</td>
<td>0.96</td>
</tr>
</tbody>
</table>
ErbB1

• Prediction error of 1 log unit is useful model for in-silico screening

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>0.89</td>
</tr>
<tr>
<td>Test Set</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Q: Since random sets are selected, what happens if you repeat trials?
A: The results are nearly identical in a statistical sense

- Results are similar for this ABL example.
- For ‘final’ models larger selections or all data points can be used to increase robustness
- The automation aspect makes repeated trials easy to do
Q: What if you use a higher level class instead of a single protein?
A: In this case at least, the model isn’t as good.

**Training Set**

\[ y = 0.9495x \]

\[ R^2 = 0.3565 \]

**Test Set**

\[ y = 0.9526x \]

\[ R^2 = 0.0029 \]

VEGFR* (200,000 bioactivities reported)
Includes VEGFR 1, 2, 3, FLT1, FLT3 etc. etc.
There is probably too much variation in the sub-target structures that makes the model ineffective
Model is not useful.
Your mileage may vary with class

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>1.47</td>
</tr>
<tr>
<td>Test Set</td>
<td>1.99</td>
</tr>
</tbody>
</table>
EGFR (P00533)

- EGFR* (200,000 bioactivities reported)
  - Includes VEGFR, FLT1, FLT3 etc. etc.
- There is probably too much variation in the sub-target structures that makes the model ineffective
- Model is not useful.
- Your mileage may vary with class

<table>
<thead>
<tr>
<th>Model</th>
<th>RMS Error of Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>1.24</td>
</tr>
<tr>
<td>Test Set</td>
<td>1.59</td>
</tr>
</tbody>
</table>

\[
y = 0.9721x \\
R^2 = 0.711
\]

\[
y = 0.9514x \\
R^2 = 0.4955
\]
Models for Sub-Variants of VEGFR – Some of the variants contribute more to the model than others

- **VEGFR**
  - FLT1
  - FLT4
  - FLK-1 (VEGFR2)
  - VEGFR [Human]

<table>
<thead>
<tr>
<th>Target</th>
<th>RMS EP Test Set</th>
<th>RMS EP Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGFR*</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td>FLT1</td>
<td>1.4</td>
<td>1.8</td>
</tr>
<tr>
<td>FLT4</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>FLK-1</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>VEGFR [Human]</td>
<td>1.0</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Conclusions

• This tool makes it very easy to do answer many many questions with minimal time and effort, as shown.
  • One can add ability to search for specific scaffolds and combine with target query.

• It provides an unprecedented tool to explore QSAR by target, class and other criteria.

• One can assess the likelihood that there is sufficient data to make a model for a given target
  • The automatic evaluation using a test set provides nearly instant feedback on the prospects for making a predictive model.

• Because it uses the KNIME system, nearly any descriptor set can be used for making the models.
  • There are additional open-source and commercial descriptor tools available.
  • One can probably get better results with better descriptors.