DATA, DATA, DATA: is there something else in in silico ADME? May be their quality and their sources!

F. Lombardo
Modeling, Computation and Molecular Properties (MCMP) Group

NEDMDG, June 11 2015
ADME optimization.....early!
This is why WE are in this business

Slide courtesy of Professor G. Cruciani, U of Perugia
What is the meaning of ‘drug like’?

Not much of a structural resemblance, right? Then what?

One definition could be “pharmacokinetics friendly” as we have to deliver these drugs to humans...therefore we need to reason about pharmacokinetics. (Zhao, DDT, \textbf{2011}, 64, 158-163).
The pillars of ADMET

Due to increasing complexity in drug discovery/development, larger focus on computational data interpretation needed

*Slide courtesy of Dr. H. Fischer, Roche*
A very complex interplay of factors... unraveled?

Physicochemical properties are
• key to understand ADME
• the link between chemical structure and PK properties... via **DATA INTEGRATION**

Um...What if we ignored them?
Good PK → Competitive Advantage...how?

- "Good" Pharmacokinetics
- Better Efficacy
- Convenient Dosing Regimens
- Higher Patient Compliance
- Competitive Advantage
How it was and how it is now

**Monodimensional**
- High risk
- Low chance of success
- Time consuming
  
  ~ 1960

**Consecutive Multidimensional**
- Increased chance of success
  
  ~ 1990

**Parallel Multidimensional (MDO)**
- Reduced time
- Increased productivity
- Increased chance of success
  
  since 1999 - implemented 2005

*Slide courtesy of Dr. H. Fischer, Roche*
Are wee agreeing so far?

Let us recap:

• **Yes** we need to worry sooner rather than later about ADME.
• The Sirens of potency **can be “deadly”**
• ADME parameters are **complex**.
• We **DO** wish to use parallel multidimensional optimization schemes.
• There are **competitive advantages** for a PK-ADME “job” well done.
The Human Factor
(and its impact on data and modeling efforts)

- There is generally no formal training in isADME.
- Some of us come from a computational background and have learned experimental techniques and caveats with data.
- Some of us come from an experimental ADME/PK background and have learned computational techniques.

- Do you prefer to build local or global models?
- Let us not start an argument. Both of them may have a place and time plus data for! *Personal bias: global models. But I can be educated!"
The Human Factor_2
(and its impact on data and modeling efforts)

- Is a multi-Company cooperative effort with a large and curated data set, a good thing to do?
  - Are we keeping this work too close to our chests?
  - Are we involving the “right” people? Do we ask end-users?
  - Some of these aspects being considered by the IQ isADME Discussion Group

- Organizational considerations*
  - Pressures on productivity. Explore “this” chemistry..easier to do.
  - Pragmatic chemistry as a result?
  - Is your organization over-relying on potency? (remember the Sirens)
  - Application of computational model
    - To do that you need good data. (They said it, honest!)

*(Leeson and St-Gallay, Nat. Rev. Drug. Disc. 2011, 10, 749-765)
Is there a trend in data generation and use in the industry?

- Yes, many pharmaceutical companies, large or small, use CROs (contract research organization) for a variety of screens and determination. **Single determination or replicates?**
  

- Some do not have the capability to run screens in house at all (they outsource physchem and ADME) while some will run each compound made through, say, a logD or solubility screen in house.

- Data use and access: not a trivial consideration, and seemingly dependent on IT work not necessarily size. **Very difficult to get to a steady state of efficient data retrieval.**
  
  - If we do not have the visualization tools or build them is like not having the data.
  - Sometimes data may just be end up being ignored..and it is quite an expense!
  - Importantly, **is somebody QC’ing the data?** Blind upload and blind retrieval? No curation? A dangerous practice. Data quality may (or may not) depend on stage of the project.

> Very, very, important to maintain expertise in data generation and target property quality…otherwise…you may be sold an egg without yolk and still be told it is an egg!

(and then wonder about “Why models fail” Stouch et al *JCAMD*, **2003**, *17*, 83-92)
The Three Pillars of Modeling

Descriptors and statistical approaches will add to the error.
Sooo..you say you wish to use literature data to generate in silico PK models, right?

• Search started in....2001 at Pfizer (Obach and Lombardo). First VD_{ss} publication with 64 compounds (in vitro data with some computational tests J. Med. Chem. 2002, 45, 2867-2876)

• As of June 2015 ~ 1,200 compounds all with iv data in human. (Clearance data for 1003 compounds available as SI for J. Med. Chem. 2014, 57, 4397–4405).

• Highly screened plasma data. We looked at analytical procedures, (e.g. no total radioactivity data), weighted-average across studies and doses (based on number of subjects), many plots digitized if data not available, total clearance and VDss (no VD_{β}), data mostly from healthy subjects.
And then branching out to animal to human scaling?

• From the above human data, a search for data in rat, dog and monkey was launched (2 years!). Resulted in 400 compounds with data for at least one species: 331 in rat, 250 in dog and 132 in monkey. No QSPkR in animal attempted yet.

• Some iv data in rat were generated in house (if dog and monkey were available). 170 ppb ($f_u$) data points also generated in house.

• Would an accurate in silico QSPkR animal model be useful?

Available Databases
(seen through industry glasses and not comprehensive...some sites offer isADME models)

Table 1. Summary Statistics for the Public and Commercial Chemistry Databases above or near Half a Million Structures (at the Time of Writing), Most of Which Include Linkages to Bioactivity and Biological Data

<table>
<thead>
<tr>
<th>name</th>
<th>total (million)</th>
<th>URL</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDB13</td>
<td>977</td>
<td><a href="http://www.gdb.unibe.ch/gdb/home.html">http://www.gdb.unibe.ch/gdb/home.html</a></td>
<td>Virtual compounds, no bioactivity data</td>
</tr>
<tr>
<td>SciFinder</td>
<td>89</td>
<td><a href="http://www.cas.org/products/scifinder">http://www.cas.org/products/scifinder</a></td>
<td>Includes 28 million vendor compounds</td>
</tr>
<tr>
<td>UniChem</td>
<td>71</td>
<td><a href="https://www.ebi.ac.uk/unchem/">https://www.ebi.ac.uk/unchem/</a></td>
<td>Includes 15 million SureChEMBL from patents</td>
</tr>
<tr>
<td>CSLS</td>
<td>46</td>
<td><a href="http://caict.ni.nih.gov/cgi-bin/lookup/search">http://caict.ni.nih.gov/cgi-bin/lookup/search</a></td>
<td>Update status unclear</td>
</tr>
<tr>
<td>ChemSpider</td>
<td>32</td>
<td><a href="http://www.chemspider.com/">http://www.chemspider.com/</a></td>
<td>Includes 12 million vendor compounds</td>
</tr>
<tr>
<td>Reaxys</td>
<td>25</td>
<td><a href="http://www.elsevier.com/online-tools/reaxys">http://www.elsevier.com/online-tools/reaxys</a></td>
<td>5.1 million medicinal chemistry data</td>
</tr>
<tr>
<td>ZINC</td>
<td>23</td>
<td><a href="http://zinc.docking.org/">http://zinc.docking.org/</a></td>
<td>All vendor compounds, 8.1 million in PubChem</td>
</tr>
<tr>
<td>GOSTAR</td>
<td>6.3</td>
<td><a href="https://gostardb.com/gostar/">https://gostardb.com/gostar/</a></td>
<td>Activity linked</td>
</tr>
<tr>
<td>Liceptor</td>
<td>3.2</td>
<td><a href="http://www.evlolution.com/products/databases/liceptordatabase.html">http://www.evlolution.com/products/databases/liceptordatabase.html</a></td>
<td></td>
</tr>
<tr>
<td>ChEMBL</td>
<td>1.4</td>
<td><a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a></td>
<td>0.94 million inside PubChem</td>
</tr>
<tr>
<td>BindingDB</td>
<td>0.48</td>
<td><a href="http://www.bindingdb.org/bind/index.jsp">http://www.bindingdb.org/bind/index.jsp</a></td>
<td></td>
</tr>
</tbody>
</table>

*Note that apart from the three sources that have update cycles within PubChem (Thomson Pharma, ChEMBL, and BindingDB) all the others are likely to have at least a proportion of unique content (e.g., extractions from different journal articles).*


OCHEM: [https://ochem.eu/home/show.do](https://ochem.eu/home/show.do). 1.3 M data points. Chemical and biological data
What happens when you look in house?

*True Value vs. sample mean considering repeat measurements 2.95, 3.20, 3.30*

Definition of a good QSA(P)R model

Is this a good model for decision-making?

If experimental error ~0.2

YES

(0.2^2 + 0.2^2)^{0.5} \sim 0.28
RMSE: 0.29

Predicted

Observed
Definition of a good QSA(P,Pk)R model

Contextualized against inherent experimental error

• Experimentally determined response variables in the training and test sets should be the true value

• No physical quantity can be measured with absolute certainty; it is unlikely that the value of a single measurement will be equal to the true value for a molecule

• Any experimental error in a model’s training set will be propagated through into a similar prediction error (assume no modelling error)

• What is a reasonable experimental error?
  • \( \leq 0.2 \) (log units)
  • Subsequent RMSE for a good model related to the propagated error in the x and y data:
    • \((0.2^2 + 0.2^2)^{0.5} \approx 0.28\)
Experimental errors
(Do we download or check too?)

Gross
• Serious deviations from a validated procedure, more so than would be expected from random variability (i.e. extreme outliers)
  • Remove prior to modelling

Systematic
• Measurements biased either negatively or positively
  • Remove prior to modelling

Random
• Measurements tending to fall randomly either side of an average value
  • Evaluate prior to modelling to understand the model’s predictivity.

### Table 1. Overview of Data Sets for the Eight DMPK Response Variables

<table>
<thead>
<tr>
<th>response variable</th>
<th>number of molecules</th>
<th>number of results</th>
<th>number of molecules to consider</th>
<th>percentage of data set with a single measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>human hep ( CL_{int} )</td>
<td>10668</td>
<td>22588</td>
<td>9819</td>
<td>40</td>
</tr>
<tr>
<td>human mic ( CL_{int} )</td>
<td>32492</td>
<td>47566</td>
<td>31215</td>
<td>74</td>
</tr>
<tr>
<td>human PPB</td>
<td>61356</td>
<td>80725</td>
<td>59852</td>
<td>89</td>
</tr>
<tr>
<td>log ( D_{7,4} )</td>
<td>115441</td>
<td>140662</td>
<td>113339</td>
<td>93</td>
</tr>
<tr>
<td>rat hep ( CL_{int} )</td>
<td>39112</td>
<td>55969</td>
<td>36807</td>
<td>77</td>
</tr>
<tr>
<td>rat PPB</td>
<td>16476</td>
<td>23738</td>
<td>16037</td>
<td>85</td>
</tr>
<tr>
<td>solubility (dried DMSO)</td>
<td>44256</td>
<td>49043</td>
<td>42821</td>
<td>95</td>
</tr>
<tr>
<td>solubility (solid)</td>
<td>38722</td>
<td>42736</td>
<td>36256</td>
<td>95</td>
</tr>
</tbody>
</table>

---

## Experimental errors

<table>
<thead>
<tr>
<th>response variable</th>
<th>no. molecules with $\geq 3$ repeat measurements</th>
<th>range in observed stdev</th>
<th>typical stdev</th>
<th>lower 95% confidence limit for stdev</th>
<th>upper 95% confidence limit for stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human hep $\text{CL}_{\text{int}}$</td>
<td>540</td>
<td>0.01 to 0.61</td>
<td>0.11</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>Human mic $\text{CL}_{\text{int}}$</td>
<td>830</td>
<td>0.01 to 0.67</td>
<td>0.12</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>Human PPB</td>
<td>1696</td>
<td>0.01 to 1.56</td>
<td>0.16</td>
<td>0.11</td>
<td>0.21</td>
</tr>
<tr>
<td>Log $D_{7.4}$</td>
<td>1445</td>
<td>0.01 to 2.12</td>
<td>0.19</td>
<td>0.11</td>
<td>0.27</td>
</tr>
<tr>
<td>Rat hep $\text{CL}_{\text{int}}$</td>
<td>919</td>
<td>0.01 to 0.92</td>
<td>0.16</td>
<td>0.1</td>
<td>0.22</td>
</tr>
<tr>
<td>Rat PPB</td>
<td>668</td>
<td>0.01 to 1.25</td>
<td>0.16</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>Solubility (dried DMSO)</td>
<td>363</td>
<td>0.01 to 1.78</td>
<td>0.25</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Solubility (solid)</td>
<td>466</td>
<td>0.01 to 1.60</td>
<td>0.28</td>
<td>0.1</td>
<td>0.46</td>
</tr>
</tbody>
</table>
**HQ: n ≥ 2, stddev = x; LQ: n = 1**

<table>
<thead>
<tr>
<th>response variable</th>
<th>stddev criterion</th>
<th>RMSE training set HQ</th>
<th>RMSEP test set HQ</th>
<th>RMSE training set LQ</th>
<th>RMSEP test set LQ</th>
<th>∆ RMSEP</th>
<th>size of training set HQ</th>
<th>size of test set HQ</th>
<th>size of training set LQ</th>
<th>size of test set LQ</th>
<th>size of training set LQ / size of training set HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human mic CL(_{int})</td>
<td>0.1</td>
<td>0.16</td>
<td>0.37</td>
<td>0.15</td>
<td>0.44</td>
<td>-0.07</td>
<td>5774</td>
<td>1372</td>
<td>22,132</td>
<td></td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.16</td>
<td>0.38</td>
<td>0.15</td>
<td>0.44</td>
<td>-0.06</td>
<td>6451</td>
<td>1537</td>
<td>22,132</td>
<td></td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.16</td>
<td>0.37</td>
<td>0.15</td>
<td>0.44</td>
<td>-0.07</td>
<td>6646</td>
<td>1591</td>
<td>22,132</td>
<td></td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.16</td>
<td>0.38</td>
<td>0.15</td>
<td>0.44</td>
<td>-0.06</td>
<td>6707</td>
<td>1605</td>
<td>22,132</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>Human PPB</td>
<td>0.1</td>
<td>0.19</td>
<td>0.45</td>
<td>0.17</td>
<td>0.43</td>
<td>0.02</td>
<td>3060</td>
<td>750</td>
<td>34,972</td>
<td></td>
<td>11.43</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.18</td>
<td>0.42</td>
<td>0.17</td>
<td>0.45</td>
<td>-0.02</td>
<td>4592</td>
<td>1084</td>
<td>34,972</td>
<td></td>
<td>7.62</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.18</td>
<td>0.43</td>
<td>0.17</td>
<td>0.46</td>
<td>-0.04</td>
<td>5255</td>
<td>1240</td>
<td>34,972</td>
<td></td>
<td>6.65</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.18</td>
<td>0.44</td>
<td>0.17</td>
<td>0.47</td>
<td>-0.03</td>
<td>5531</td>
<td>1303</td>
<td>34,972</td>
<td></td>
<td>6.32</td>
</tr>
<tr>
<td>response variable</td>
<td>RMSE training set HQ</td>
<td>RMSEP test set HQ</td>
<td>RMSE training set LQ</td>
<td>RMSEP test set LQ</td>
<td>size of training sets (average)</td>
<td>size of test set HQ</td>
<td>$\Delta$ RMSEP</td>
<td>% improvement in RMSEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>--------------------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human hep CL$_{int}$</td>
<td>0.17</td>
<td>0.42</td>
<td>0.1</td>
<td>0.55</td>
<td>147, 143</td>
<td>5037</td>
<td>-0.13</td>
<td>23.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human mic CL$_{int}$</td>
<td>0.19</td>
<td>0.46</td>
<td>0.14</td>
<td>0.6</td>
<td>448, 443</td>
<td>6698</td>
<td>-0.14</td>
<td>23.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human PPB</td>
<td>0.2</td>
<td>0.47</td>
<td>0.22</td>
<td>0.5</td>
<td>1790, 1780</td>
<td>2020</td>
<td>-0.03</td>
<td>5.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log $D_{7.4}$</td>
<td>0.27</td>
<td>0.66</td>
<td>0.3</td>
<td>0.68</td>
<td>1817, 1800</td>
<td>3489</td>
<td>-0.02</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat hep CL$_{int}$</td>
<td>0.16</td>
<td>0.39</td>
<td>0.15</td>
<td>0.54</td>
<td>651, 642</td>
<td>6663</td>
<td>-0.15</td>
<td>27.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat PPB</td>
<td>0.21</td>
<td>0.51</td>
<td>0.23</td>
<td>0.54</td>
<td>572, 563</td>
<td>815</td>
<td>-0.02</td>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solubility (dried DMSO)</td>
<td>0.29</td>
<td>0.7</td>
<td>0.29</td>
<td>0.91</td>
<td>774, 766</td>
<td>648</td>
<td>-0.21</td>
<td>23.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solubility (solid)</td>
<td>0.3</td>
<td>0.74</td>
<td>0.29</td>
<td>0.77</td>
<td>750, 745</td>
<td>196</td>
<td>-0.03</td>
<td>4.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HQ data model improvement over LQ data model relative to training set size

\[ y = 23.323e^{-0.063x} \]

16% improvement with 1/6 HQ data!
Conclusions

• Remove gross and systematic errors

• Training data sets
  • Assuming Gaussian PDFs
  • Random errors should be defined by repeat measurements that pass a stddev criterion.
  • Models built on training sets with stddev ≤ 0.1 better than those built on sets with stddev ≥ 0.20.
    • For 8 DMPK assays, range: 3.3–27.5% improvements

Conclusions

• Do not strive to measure every new molecule in a assay
• Instead:
  • Ensure assays give rise to experimental data with a stdev $\leq 0.20$ on 99% of the occasions from at least triplicate repeat measurements.
  • Measure a representative 1/6th of all possible molecules..i.e. the ones that you would have measured anyway with N=1
• Benefits
  • Reduce experimental resources potentially by 50%
  • Significant improvements in QSA(P,Pk)R models as the training sets will have less uncertainty in the estimates of the response variable true value for molecules and thus less error is propagated through into test set predictions

In silico models: a shared compound choice scheme (whether or not the algorithm is trainable)

1. Compounds are made
   - Known scaffold
   - Novel scaffold

2. Model as filter
   - Chemist’s ideas, given scaffold, modifications sought
   - Update Quarterly?

3. Subsetting algorithm
   - Screen 30-40%?

4. Calculation and analysis of data
   - Screen 5%?
   - Reasonable agreement with experiment (e.g. AAE < 0.5)
   - Disagreement with experiment (e.g. AAE > 0.5)
Some take home messages – Data and modeling

- Need we run each and every compound through physicochemical and ADME screens? Probably NOT... Especially if we use trainable models (time-series QSA(P)R).

- Single measurements may be fine in some areas/cases if some discrimination is all that it is wanted. But for modeling we need high quality.

- Importantly, there are no panaceas: what question(s) are you trying to answer?

- Consensus among various approaches: in vitro, in vivo, in silico. A possible path?

- Do not lose the “know how” of the data in house however you generate them (in house or outsourcing).

- We can get quite far with in silico ADME models but it is important to manage expectations.
Acknowledgments

• Prof. G. Cruciani (slide)
• Dr. H. Fischer (slides)
• Dr. M. Wenlock (slides and discussions)
• The great many number of colleagues who helped build the PK dataset and discussed these aspects along the years. In particular, R. S. Obach (Pfizer) and G. Berellini (Novartis).
For your attention