Decision-Making with Predictive ADME Data in Context of Experimental Variability

Brian E. Mattioni, Ph.D., Principal Scientist
Computational/Preclinical ADME
Department of Pharmacokinetics, Pharmacodynamics, and Drug Metabolism
Merck Research Laboratories
Mindset shift towards synergistic in silico↔in vitro↔in vivo design cycles and focus on property space enrichment
  - re: *in silico*, progress beyond calculated physchem properties (when appropriate)

Experimental variability and an argument on why we need to reset expectations of *in silico* (& *in vitro*) ADME models (especially with N=1 screening data)

Examples of real-time, prospective decision-making with Predictive ADME data at the design stage
In silico ↔ in vitro ↔ in vivo design cycles
~ as early as possible, at the chemotype level ~

Harness information from *in silico* and *in vitro* and *in vivo* assays to maximize information content from each design cycle

Consider more often
How can I trust and proactively use *in silico* ADME models for virtual & experimental design?

- **Iterative design/learning cycles are key**
  - Builds trust in the *prospective* ability of models for contemporary SAR space
  - Helps understand model strengths/limitations
  - If global model validation has been robust throughout time, a certain degree of trust exists in using proactively as hypothesis “generators” earlier rather than later (typical of models for *in vitro* endpoints)
  - For models built with *in vivo* PK data, frequent learning cycles are a must

**Capitalize on where the models work and exploit strengths**
(spend less time figuring out why they shouldn’t be used)
Setting realistic expectations for *in silico* ADME models

When performing “predicted vs. measured” analyses...acknowledge experimental variability

\[
\text{Error}_{\text{total}} = \text{Error}_{\text{model}} + \text{Error}_{\text{experimental}}
\]

This is not zero!

*In vitro/in vivo* methods still only provide an estimate of “truth”

As most screening data stand today, building those elusive models with \(R^2=0.8\) or \(\text{RMSE}=0.1\) is simply not realistic
$P_{\text{app}}$ \textit{in silico} model prospective validation

Quantitative view

Is this model “predictive”?

\[ R^2 \approx 0.60 \]

$N = 7518$ compounds

Raw-scale ($\times 10^{-6}$ cm/s)
**P_{\text{app}}** data reproducibility/variability

- plot of run1 vs run2 for repeat measurements -

\[ R^2 \sim 0.60 \]

N = 352 compounds

Raw-scale (x10^{-6} cm/s)
Large majority of repeat measurements were within 2-fold of the original indicating the assay is robust, reproducible, and enables decision-making (just not as quantitatively delineating as most scientists think or want)
P_{app} \textit{in silico} model prospective validation

Quantitative view

\[ R^2 \sim 0.60 \]
N = 7518 compounds
Raw-scale (x10^{-6} \text{ cm/s})
P<sub>app</sub> in silico model prospective validation

Quantitative view

\[ R^2 \approx 0.60 \]
N = 7518 compounds
Raw-scale (x10^-6 cm/s)
**Qualitative view (same data as last slide)**

**Predicted vs measured $P_{app}$ categories for the last 7,518 compounds tested with control BA/AB ratios between {0.5-2.0}**

**Design Guidelines:**

*in silico* $P_{app} \leq 5$ {deprioritize}

5 < *in silico* $P_{app} \leq 20$ {indeterminate}

*in silico* $P_{app} > 20$ {prioritize}
Rat CL_{plasma} \textit{in silico} model demonstrates robust categorical enrichment.

**Predicted Rat \textit{in vivo} CL_{plasma} category**

- **Represents \sim 1 years worth of PK data**

<table>
<thead>
<tr>
<th>Category</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x \leq 21.00)</td>
<td>15.0%</td>
</tr>
<tr>
<td>(21.00 &lt; x \leq 63.00)</td>
<td>47.7%</td>
</tr>
<tr>
<td>(63.00 &lt; x)</td>
<td>74.2%</td>
</tr>
</tbody>
</table>

**Design Guidelines:**

- \textit{in silico} Rat \textit{in vivo} CL_{plasma} \leq 21 \{prioritize\}
- \(21 < \textit{in silico} \textit{in vivo} CL_{plasma} \leq 63\} \{indeterminate\}
- \textit{in silico} Rat \textit{in vivo} CL_{plasma} > 63 \{deprioritize\}
Are better decisions made when *in silico* data are included in the thought process?

- This should be the question asked of any model
- Specifically with respect to *in silico* ADME models:
  - Shift focus to **property space enrichment** and rank-ordering
  - Absolute accuracy (*e.g.*, RMSE) or correlation (*e.g.*, \( R^2 \)) is less important … **since we’re not replacing measured data**
  - If *in cerebro* (+ *in silico*) enriches the decision-making process at the design stage … apply!

**Models prolong the shelf-life of data & each is a summary of our “experience” with a given biological endpoint**

*Why not use them to our advantage?*
Progressing beyond calculated physchem properties…QSAR (often) provides higher granularity

**Rat P-gp substrate recognition**

- **N = 355**
  - $PPV = \sim 96\%$

- **N = 1911**
  - $NPV = \sim 84\%$

- **N = 1110**

**Predicted rat P-gp category**

- **N = 1533**
  - $PPV = \sim 62\%$

- **N = 564**
  - $NPV = \sim 74\%$

- **N = 1279**

---

**Merck**

Be well
A QSAR model (often) offers higher granularity when compared to a calculated physchem property

~ Trade-off between interpretability & accuracy acknowledged ~

**Solution:** use both to your advantage at design
• **Mindset shift** towards synergistic *in silico*↔*in vitro*↔*in vivo* design cycles and focus on property space enrichment
  – re: *in silico*, progress beyond calculated physchem properties (when appropriate)

• **Experimental variability** and an argument on why we need to reset expectations of *in silico* (& *in vitro*) ADME models (especially with N=1 screening data)

• **Examples** of real-time, prospective decision-making with *Predictive ADME* data at the design stage
• **Premise**, can we exploit the strengths (i.e., predicting inactivity) of our *in silico* ADME models to advance compounds faster down the flow-scheme?

• Get to the more informative experiments faster

• Teams have full autonomy to decide how much risk (if any) is taken based on risk-reward cutoffs


DOI: 10.1021/ci500666m
Future: Adaptive & Probabilistic Screening Funnels

- Adapt compound flow through assays to de-risk early, but be thoughtful to maximize information gain (take risk where appropriate)
- Bring a more proactive & holistic approach to design aspects (instead of reactive)
- Cheminformatics tools needed for data management and handling logistics

Linear & binary view of the funnel

Adaptive & probabilistic

obtain crucial data sooner

accelerate promising candidates

don't make compounds with low POS
The Z-score is the ratio of the difference between the prediction and the threshold divided by the uncertainty of the prediction:

\[ Z\text{-score} = \frac{(\text{Prediction} - \text{threshold})}{\text{Prediction RMSE}} \]

Quantifying the likelihood of activity: Prediction + confidence → “Z-score”
Prospective validation of the probabilistic confidence metric

- The probability of being **active** (IC50 < 10 µM) vs. **inactive** is a function of the Z-score metric

![Graphs showing the probability of CaV1.2 and CYP 3A4 inhibition vs. Z-score.](image)
Prediction and confidence can be converted into risk/reward profiles

- By testing only compounds with very low probability of activity, few false negatives will result, but benefit will be modest also.

- As we increase the probability threshold, the benefits increase along with the false negative rate.

Risk/reward profiles can differ for each assay, per project, per chemical series.

Confidence cutoff of 0.0: 60% benefit, but ~13% false negative rate.

Confidence cutoff of 2.0: <10% benefit.

Confidence cutoff of Z-score = 0.5: Good compromise?
Workflow for *in silico* counterscreening on a Discovery project

1. Mine predicted versus observed data
2. Determine risk/reward profile
3. Set rules to be applied to each counterscreen assay

Revisit as needed

Repeat weekly

- **Compound predicted active?** Screen
- **Compound predicted inactive but with insufficient confidence?** Screen
- **Compound of high interest?** Screen
- **Compound randomly selected (to assure model quality and monitor predictivity)** Screen
- **Large discrepancy between pred and exp n=1?** (Re-) Screen
- ...otherwise Defer screen

Request additional compounds/assays as needed
eCS Workbench

- Automates the process of examining chemical series for predictivity and setting appropriate rules

**eCS rule-setting:**
Which models are working on my project, and what are the decision rules for testing/deferring?

**eCS submission:**
View, adjust, and submit weekly testing choices
Two project-specific case studies that demonstrate...

- Establishing & utilizing in silico ↔ in vitro ↔ in vivo connectivity at the molecular & experimental design stages

- Exploiting the strengths of both in silico and in vitro models to prioritize synthesis and testing

- Prioritizing chemical synthesis using both calculated physchem props and QSAR predictions for CL_{int}/P-gp substrate recognition

- Probabilistic approach to “design”
Project-specific case study #1

circa 2013-14, IVIVC established & utilized routinely

circa 2014-15, intentional drive to bring a holistic ISIVIVC perspective to design
Usable IVIVC established for both Lead series

- Usable IVIVC in rat for both lead series.
- Robust correlation between human LM and rat LM \( \text{in vitro} \) \( \text{CL}_{\text{int,u}} \)

88% of compounds with \( \text{in vitro} \) \( \text{CL}_{\text{int,u}} \) > 400 have an \( \text{in vivo} \) \( \text{CL}_{\text{int}} \) > 1000

Design guidance:
Rat mic. \( \text{in vitro} \) \( \text{CL}_{\text{int}} \) > ~400 mL/min/kg {deprioritize}
Microsomal $\text{CL}_{\text{int}}$ Prospective Validation
~ robust enrichment demonstrated by \textit{in silico} models ~

**Design guidance:**
- Rat mic. \textit{in silico} $\text{CL}_{\text{int}} < \sim 200 \text{ mL/min/kg}$
- Human mic. \textit{in silico} $\text{CL}_{\text{int}} < \sim 100 \text{ mL/min/kg}$

- For a QD drug, target human \textit{in vitro} $\text{CL}_{\text{int}} < 100 \text{ mL/min/kg}$
P-gp Sub. Recognition Prospective Validation
~ robust enrichment demonstrated by *in silico* models ~

**Design guidance:**
- Rat P-gp ER < 1.5 (prioritize)
- Rat P-gp ER > 4 (deprioritize)

**Design guidance:**
- Human P-gp ER < 1.5 (prioritize)

- Models showed robust enrichment on the extremes of the BA/AB ratio scale (*i.e.*, *bonafide* P-gp substrates/nonsubstances)
Multiparameter Optimization (MPO) scoring function used at molecular design

Strategy: use a 10-pt MPO scoring approach to triage synthetic targets

- 6-pt CNS MPO*:
  - cLogP \( \leq 3 \), \{3-5\}, >5
  - cLogD \( \leq 2 \), \{2-4\}, >4
  - MW \( \leq 360 \), \{360-500\}, >500
  - PSA \( \leq 20 \), \{20-40\}, \{40-90\}, \{90-120\}, >120
  - HBD \( \leq 0.5 \), \{0.5-3.5\}, >3.5
  - \( pK_{a_{\text{basic}}} \) \( \leq 8 \), \{8-10\}, >10

- Combined with internal models for human/rat microsomal intrinsic clearance and P-gp substrate recognition

Standing the test of time...prospective (& automated) spot-checks are useful

Since January 1, 2015

- Rat
  - ~89% PPV* (30 of 44 predictions confirm experimentally)
  - ~68% NPV* (30 of 44 predictions confirm experimentally)

- Human
  - ~88% PPV* (108 of 123 predictions confirm experimentally)
  - ~68% NPV* (30 of 44 predictions confirm experimentally)

*PPV – positive predicted value (given positive prediction, % confirm experimentally)
*NPV – negative predicted value (given negative prediction, % confirm experimentally)
Project-specific case study #2

Team established & used *in silico*↔*in vivo* connectivity

**IVIVC disconnected**
Early on, in vivo unbound clearance (CL\textsubscript{u}) was trending high for the Lead series.

- In vitro data tended to under-predict in vivo CL\textsubscript{u}.

**Initial IVIVC snapshot**

- RAT \textit{in vitro} hepatocyte CL\textsubscript{int,u}
- RAT in vivo CL\textsubscript{u}

**Initial ISIVC snapshot**

- Predicted RAT in vivo CL\textsubscript{u}
- RAT in vivo CL\textsubscript{u}
In an effort to improve both IVIVC and ISIVC, a virtual library was designed where...

...in silico models for rat in vivo CL_{u} and in vitro CL_{int} were used to prioritize chemical synthesis (pink)
Tracking progress... circa May, 2015
ISIVC solidified & expanded, IVIVC strengthened

Initial ISIVC snapshot

Initial IVIVC snapshot

May, 2015 ISIVC snapshot

May, 2015 IVIVC snapshot
Take home

- **Find synergy** between *in silico*, *in vitro*, and *in vivo* data & exploit where *in silico*/*in vitro* models work

- **Experimental variability** should be taken into account when building/validating *in silico* ADME models (especially with N=1 screening data)

- Think **enrichment**!

- *Predictive ADME* data can be used to **augment real-time, prospective decisions** at the design stage
Acknowledgements

**Structural Chemistry**
Michael Altman  
Chris Culberson  
Hakan Gunaydin  
Prabha Karnachi  
Daniel McMasters  
John Sanders  
Bob Sheridan  
Deping Wang  
Li Xiao

**PPDM (DMPK)**
Tjerk Bueters  
Kerry Fillgrove  
Chris Gibson  
Iain Martin  
Karsten Menzel  
Sue Hill

**Discovery Chemistry**
Doug Beshore  
Chris Burgey  
Mike Ellis  
Iyassu Sebhat  
Abbas Walji