# Global or local QSAR: is there a way out?

**Miklos Feher**
Campbell Family Institute for Breast Cancer Research,
University Health Network, Toronto, ON, Canada

**Todd Ewing**
Plexxikon Inc., Berkeley, CA, USA

*** All experimental work was performed at Neurocrine Biosciences, San Diego, CA, USA**

September 22, 2008

# Many seemingly worrying recent articles

***In silico* ADME/Tox: why models fail**
Terry R. Stouch et al.
J Comput Aided Mol Des **17**, 83, 2003

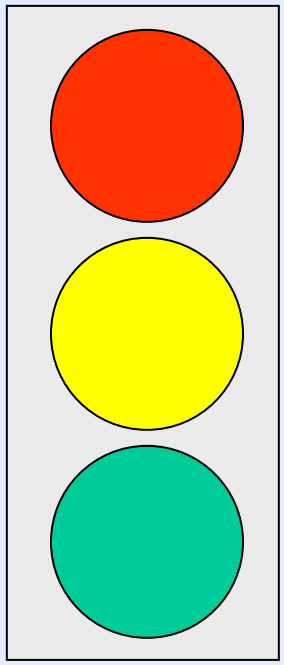
**The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy)**
Stephen R. Johnson
*J. Chem. Inf. Model.*, **48**, 25, 2008.

**QSAR: dead or alive?**
Arthur M. Doweyko
J. Comp.-Aided Mol. Des. **22**, 81, 2008

**On Outliers and Activity Cliffs- Why QSAR Often Disappoints**
Gerald M. Maggiora
*J. Chem. Inf. Model.*, **46**, 1535, 2006

# Similar issues observed in our research

## Some examples (based on proprietary data):



2D6 inhibition (Accelrys, GeneGo, Inpharmatica), 3A4 inhibition (GeneGo), BBB penetration (Schrodinger, Rational discovery), HERG binding (Schrodinger), various published HERG models, PGP substrate (Inpharmatica)

BBB (Accelrys, Simulations Plus, Inpharmatica), HIA (Accelrys), PGP substrate (Pharma Algorithms), Solubility (ACD, Simulations Plus, MOE, Inpharmatica), logP (Accelrys)

logP (ACD), logD (ACD), pKa (ACD)

…but even the best ones have rather poor statistics and the generate the occasional huge misprediction…

# Potential issues with commercial ADME predictors

Sources of the problem - in broad strokes:

- Data issues (e.g. data often from incompatible sources)
- Training issues (e.g. overtraining, bias, chance correlation)
- Prediction issues (e.g. domain applicability, global vs. local models)

# Global vs. local models

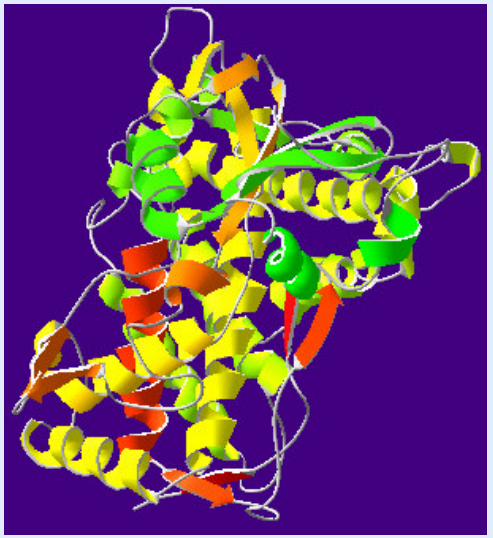
| | Global model | Local model |
|---|---|---|
| **source of compounds** | (ideally) large and diverse dataset | narrow range of targets or chemotypes |
| **generality** | (ideally) wide domain of applicability | narrow domain of applicability, small changes in chemistry may render model useless |
| **statistics** | good statistics for a diverse set<br>sometimes poor results for highly similar molecules | good statistics only for ‘similar’ compounds |
| **servicing** | minimal (generate and use) | model needs to be continually updated |

**We developed a global and local fusion model…**

# CYP Predictions

**Cytochrome P450s**

- major enzymes involved in drug metabolism, accounting for ~75% of the total metabolism
- many drugs may increase or decrease the activity of various CYP isozymes (enzyme induction and inhibition) ⇒ major source of adverse drug interactions (changes in CYP activity may affect the metabolism and clearance of other drugs)
- humans have 57 genes and divided among 18 families and 43 subfamilies of cytochrome P450's



**We concentrated on two CYP isozymes: 3A4 , 2D6:**
**- these were most relevant for discovery programs**
**- thousands of data points were available and data was continually generated**

# Methodology of model generation and testing

**Forward prediction:**

- initial training set: available data
- predictions, statistics for next group
- this group is then added to training set, predictions for the subsequent group
- weekly cycle (except for global model)

More realistic and rigorous than other validation methods:

- assesses both interpolation and extrapolation performance
- as chemistry keeps changing with time, performance is expected to be worse than with other methods

**Descriptors**: small number of ‘reasonable’ physicochemical descriptors

# Scenario 1: Global model

**Training set**

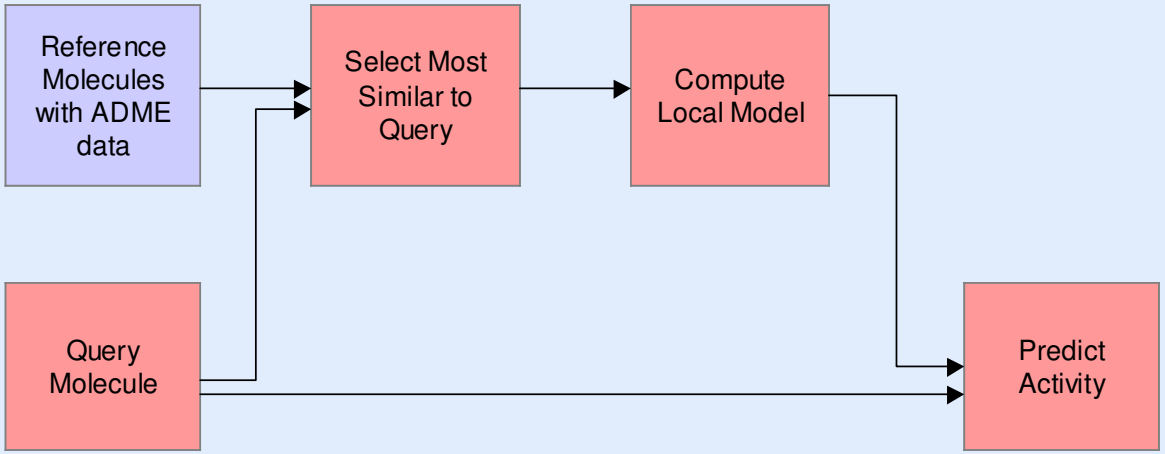
| | CYP2D6 Global Model | | CYP3A4 Global Model | |
|---|---|---|---|---|
| **Property [a]** | **Coefficient** | **Importance** | **Coefficient** | **Importance** |
| constant | 5.5 | | 2.9 | |
| logD (7.4) | 0.083 | 0.15 | 0.16 | 0.35 |
| MW | N.D. | | 0.0027 | 0.29 |
| PSA | -0.016 | 0.44 | N.D. | |
| pKa (acid) | 0.33 | 0.08 | 0.75 | 0.23 |
| pKa (base) | 0.46 | 0.20 | 0.27 | 0.13 |
| heme chelator | 0.43 | 0.09 | 1.8 | 0.37 |
| **n** | **2798** | | **4310** | |
| **Weighted $r^2$** | **0.35** | | **0.44** | |
| **Weighted S.E.** | **0.95** | | **0.79** | |

**Prediction set**

| | CYP2D6 N = 2787 | | CYP3A4 N = 4228 | |
|---|---|---|---|---|
| **Method** | **$r^2$** | **s.e.** | **$r^2$** | **s.e.** |
| Global QSAR | 0.26 | 0.89 | 0.29 | 0.74 |

**Performance is insufficient for practical use**

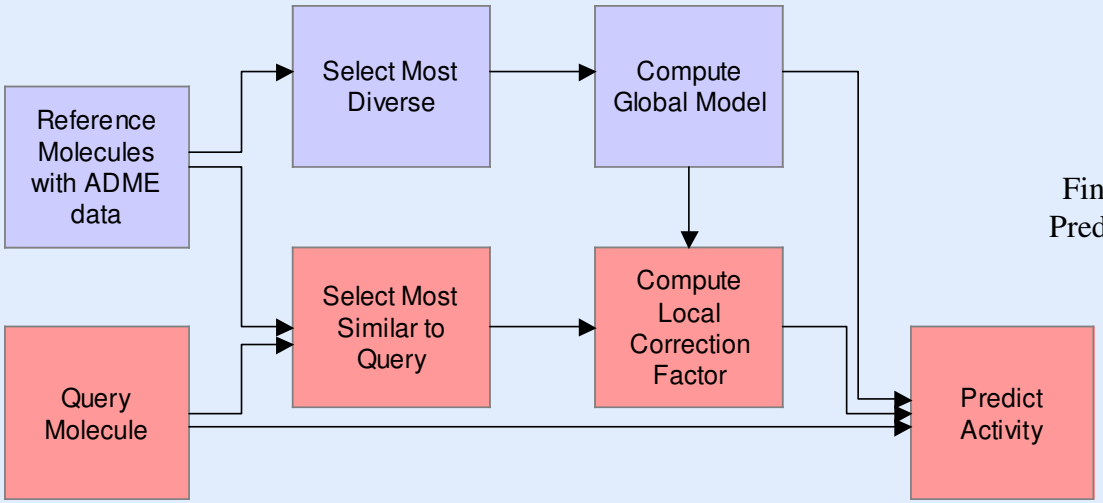
# Scenario 2: Automatically generated local model

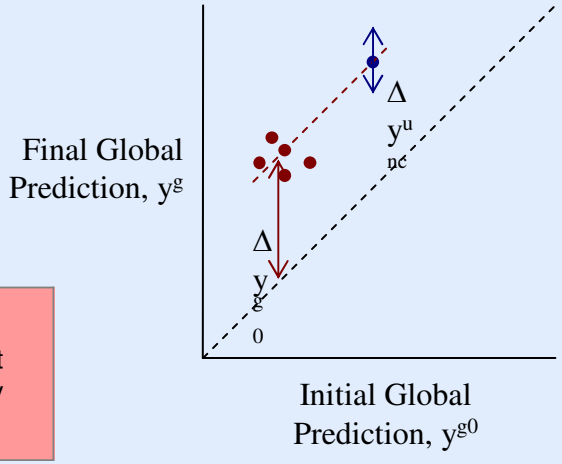



| Forward prediction | CYP2D6 Predictions, N = 2787 | | CYP3A4 Predictions, N = 4228 | |
|---|---|---|---|---|
| Method | $r^2$ | s.e. | $r^2$ | s.e. |
| Global QSAR | 0.26 | 0.89 | 0.29 | 0.74 |
| NN/Local QSAR | 0.43 | 0.76 | 0.41 | 0.69 |

**Issue: 'do not see forest from the trees' (i.e. important contributions that arise in all examples are not selected)**

# Scenario 3: Global-local fusion model



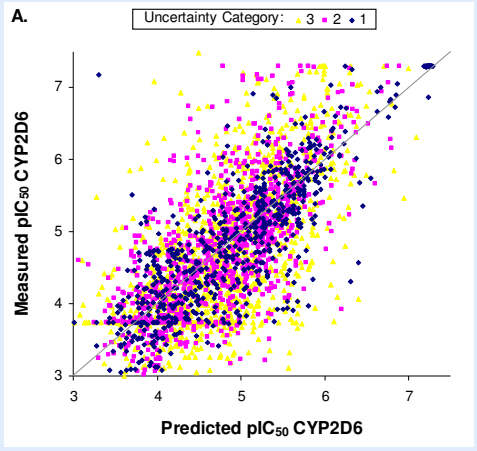



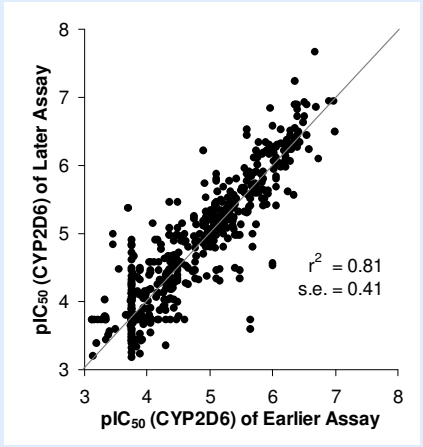

| Forward prediction | CYP2D6 Predictions, N = 2787 | | CYP3A4 Predictions, N = 4228 | |
|---|---|---|---|---|
| **Method** | **$r^2$** | **s.e.** | **$r^2$** | **s.e.** |
| Global QSAR | 0.26 | 0.89 | 0.29 | 0.74 |
| NN/Local QSAR | 0.43 | 0.76 | 0.41 | 0.69 |
| NN/Global QSAR + NN/Local QSAR | 0.47 | 0.73 | 0.47 | 0.64 |

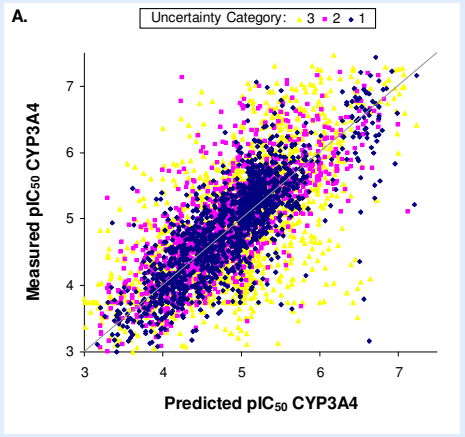
**However, fingerprint similarity also allows reliability estimates…**

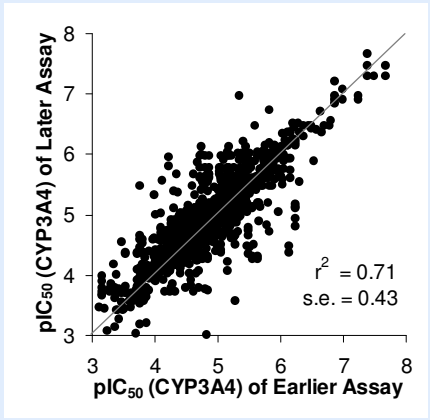
# Classifying predictions according to estimated standard error

| Predicted Standard Error | CYP2D6 Validation Set N | CYP2D6 Validation Set $r^2$ | CYP2D6 Validation Set S.E. | CYP3A4 Validation Set N | CYP3A4 Validation Set $r^2$ | CYP3A4 Validation Set S.E. |
|---|---|---|---|---|---|---|
| PSE < 0.6 | 1287 | 0.66 | 0.47 | 1193 | 0.69 | 0.41 |
| 0.6 ≤ PSE < 0.8 | 1120 | 0.46 | 0.63 | 807 | 0.39 | 0.59 |
| PSE ≥ 0.8 | 1011 | 0.28 | 0.78 | 1369 | 0.26 | 0.68 |




**Reasonable predictive performance…**

# Conclusions

## Global-local fusion model

- advantages of local models (e.g. good predictivity within a narrow range of chemical diversity, always developed using the most suitable training set)
- advantages of global models (no need to keep servicing models)
- indicates if available data is unsuitable for predictions
- was used at Neurocrine in a weekly cycle (predictions put in a database prior to the availability of experimental data)
- method applicable to other properties (e.g. similar model was generated for predicting PGP efflux)

# Acknowledgements

**Neurocrine Biosciences**

- John Saunders and Warren Wade (Med.Chem.)

- Med. Chem. groups for new compound synthesis

- Preclinical Development Group for generating CYP450 $IC_{50}$ data