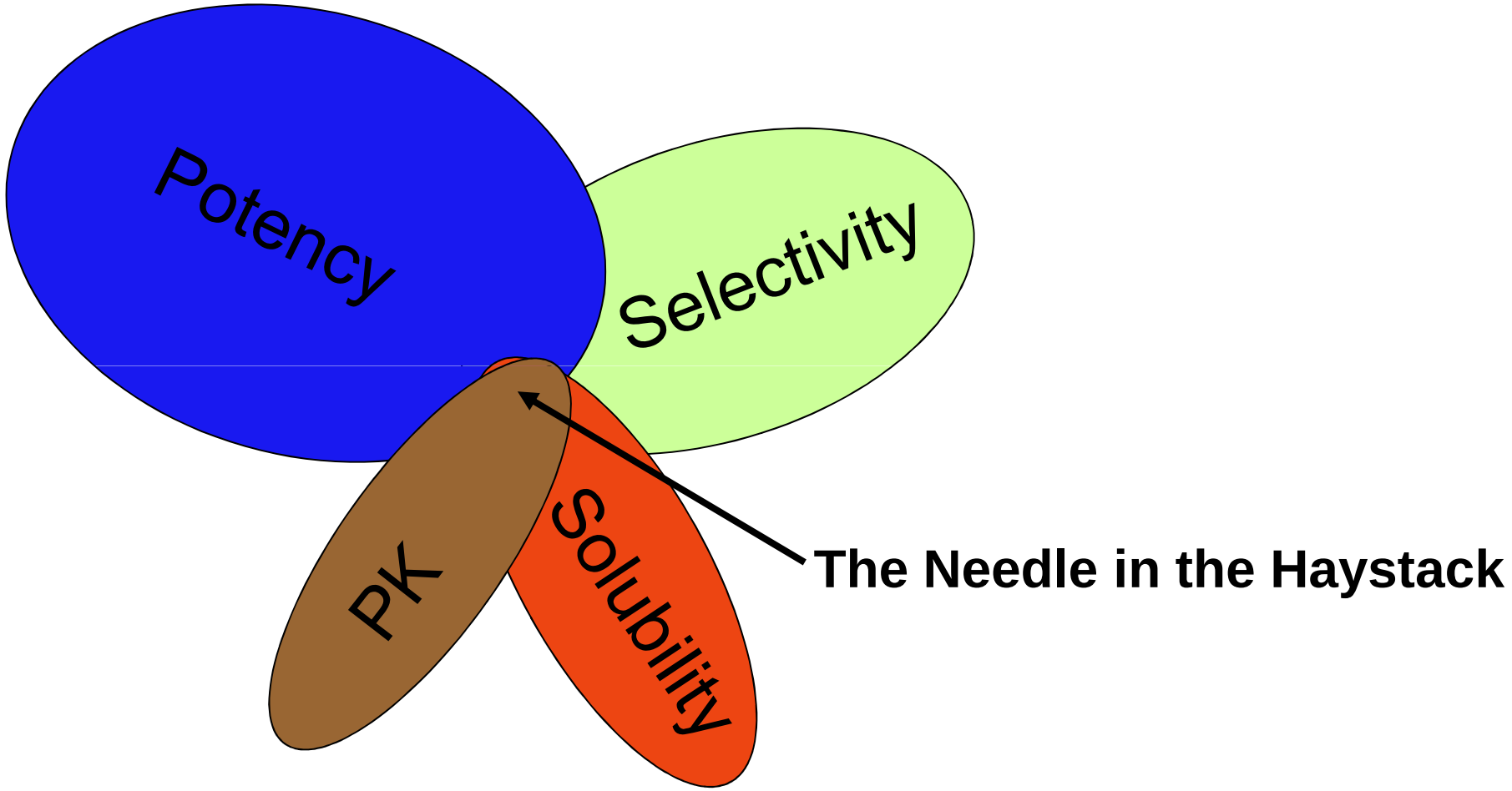
# Improving Drug Discovery Efficiency via In Silico Calculation of Properties

D. Ortwine

# Outline

- Background:  Why Calculate Properties?
- Calculable properties
- Modeling Methods and Molecule Descriptors
- Reporting Results From Calculations
- Available Commercial Software
- Strategies for Implementation
- A Real Project Example
- The Future
- Conclusions
- References

# Lead Optimization in Drug Discovery



# Why Calculate Properties?

## They can be related to the developability of drugs!

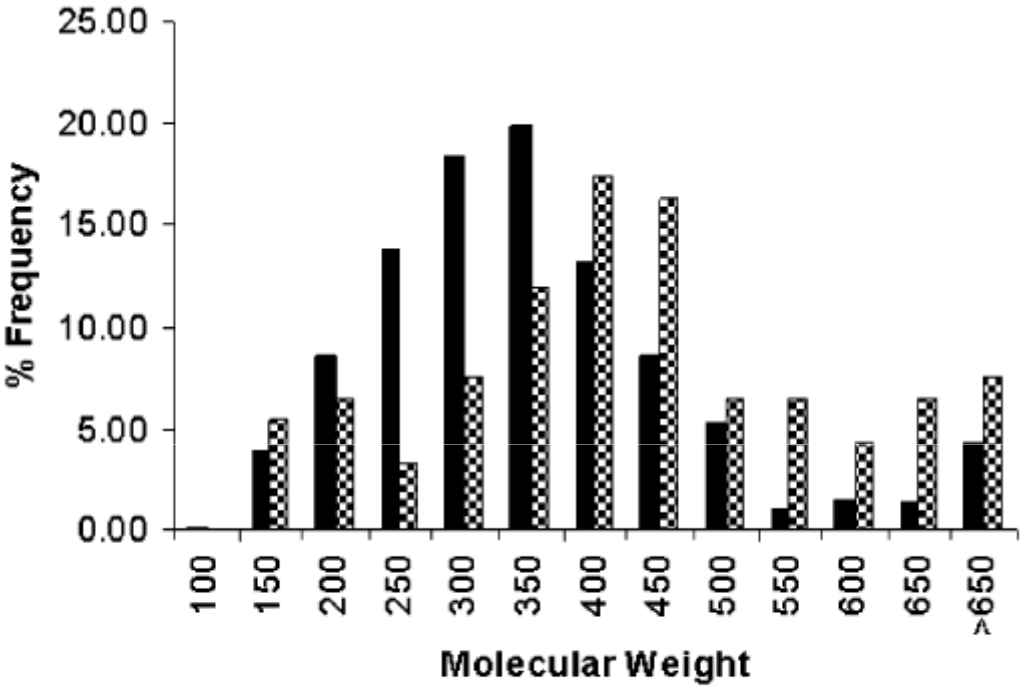



**Figure 2.** Distributions of molecular weight for marketed oral drugs (black) and development phase I oral drugs (checkered).

Paul D. Leeson and Brian Springthorpe, "The Influence of Drug-Like Concepts on Decision-Making in Medicinal Chemistry", Nature Reviews Drug Discovery, vol. 6, pp. 881-890, 2007.

Mark C. Wenlock, et.al, "A Comparison of Physiochemical Property Profiles of Development and Marketed Oral Drugs", J. Med. Chem, 2003, 46, 1250-1256.

# Why Calculate Properties?

## They can also be related to the ADMET Profile

**Table 3.** Indication of How Changes in Key Molecular Properties will Affect a Range of ADMET Parameters[a]

| neutral molecules | MWT < 400 and clogP < 4 | MWT > 400 and/or clogP > 4 |
|---|---|---|
| solubility | average | lower |
| permeability* | higher | average/higher |
| bioavailability | average | lower |
| volume of Dist.** | average | average |
| plasma protein binding | average | higher |
| CNS penetration*** | higher/average | average/lower |
| brain tissue binding | lower | higher |
| P-gp efflux | average | higher/average |
| in-vivo clearance | average | average |
| hERG Inhibition | lower | lower |
| P450 inhibition**** | lower 2C9, 2C19, 2D6 & 3A4 inhibition | higher 2C9, 2C19 & 3A4 inhibition |
| P450 inhibition**** | higher 1A2 inhibition | lower 1A2 inhibition |
| P450 inhibition**** | | average 2D6 inhibition |

(a)

M. Paul Gleeson. Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. J. Med. Chem. (2008), 51(4), 817-834.

# Why Calculate Properties?

- Prioritize synthesis
  -> Generate virtual individual molecules or combinatorial libraries, calculate properties, map back to R groups
- Build an understanding of SAR
- Combine with docking scores in a multiparameter optimization paradigm
- Assist HTS triage
- Replace measurements
- Guide the growth of the compound collection
- Guide the subsetting of the compound collection


Formulate This

# Calculable Properties

**Descriptive**

N+O, Donors, Rings,
TPSA, Size, Similarity,
Connectivity

**Binding, 3D Shape**

Docking scores,
Fit to a pharmacophore,
Shape overlap

**Physiochemical**

MW, Log P, Log D, pKa,
Solubility, Polarizability,
Critical packing (crystallinity)

**Composite**

Ligand Efficiency,
Ligand Lipophilic Efficiency,
Cellular Efficiency

**DMPK**

LM, Hep Stability,
PPB, Permeability,
Vdiss, Cyp inhibition,
Metabolic 'hotspots'

| Accuracy | Calculation Difficulty |
|---|---|
| High | Easy |
| Moderate | Medium |
| Low | Hard |

# Property Definitions

TPSA Topological polar surface area

<120 desirable; <80 for CNS drugs

LE Ligand Efficiency = -1.4 logKi / # of heavy atoms[1]

0.3 is a good hit; 0.35-0.5 is a good clinical candidate

LLE Ligand Lipophilic Efficiency = -log(Ki) – logD[2]

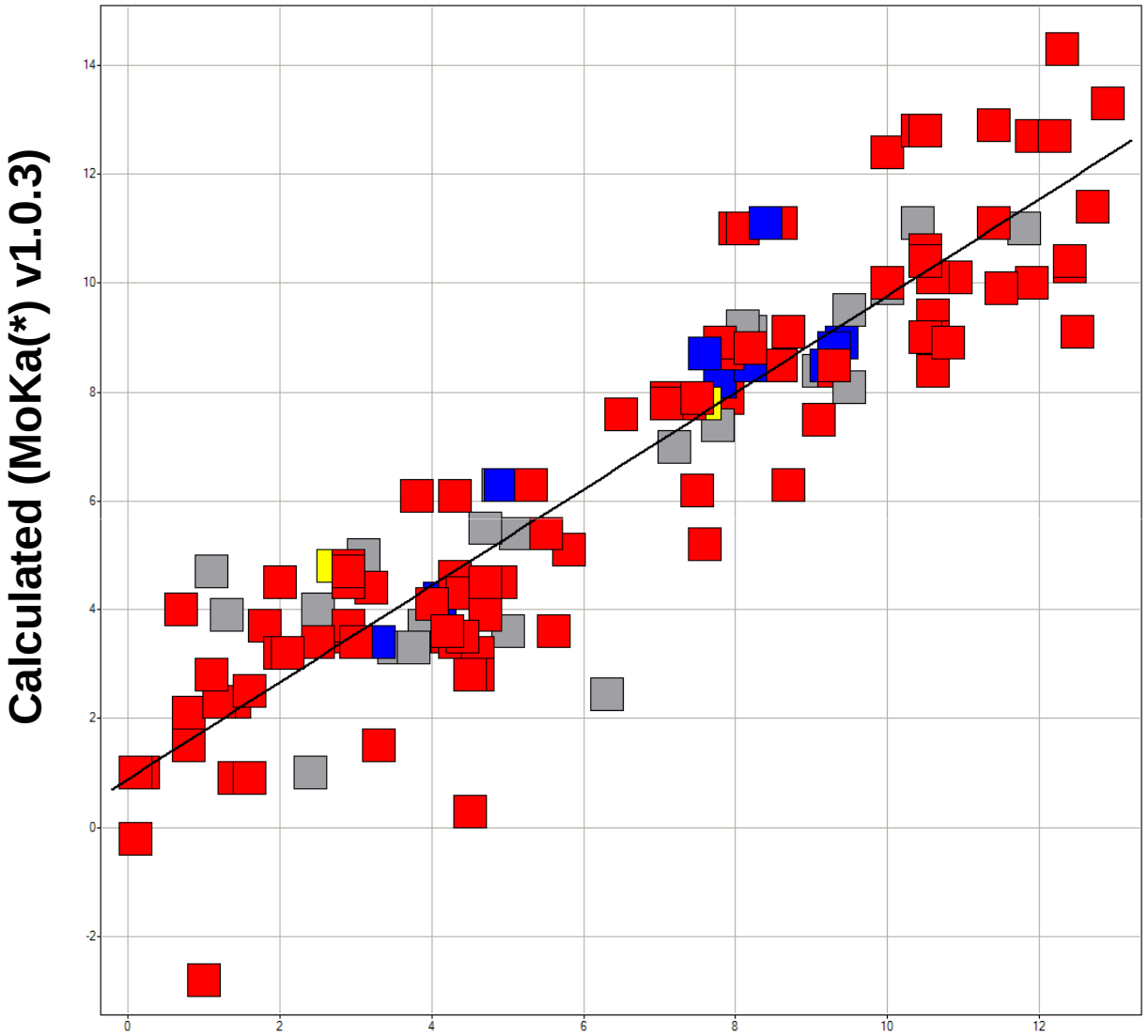
7-9 is a good clinical candidate

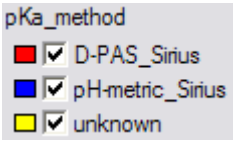
CellE Cellular Efficiency = -log(*in vitro* Ki) – log(cellular EC50)

0 is goal, <1.5 is acceptable

[1]Andrew Hopkins, et.al., Drug Discovery Today, 2004, 9, 430-431.
[2]Paul Leeson and Brian Springthorpe, Nature Reviews Drug Discovery, 2007, 6, 881-890.

# Calculated vs Measured pKa




**$R^2$ = 0.82**
**n = 133**
**s = 1.5**
**(75 compounds)**
**(intercept not signif.)**

pKa_for_plots = 0.2963984 + 0.9271127*c_pKa_MoKa_for_plots

**Summary of Fit**

| | |
|---|---|
| RSquare | 0.822564 |
| RSquare Adj | 0.821209 |
| Root Mean Square Error | 1.496272 |
| Mean of Response | 6.318797 |
| Observations (or Sum Wgts) | 133 |

**Lack Of Fit**

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 1 | 1359.6259 | 1359.63 | 607.2931 |
| Error | 131 | 293.2867 | 2.24 | Prob > F |
| C. Total | 132 | 1652.9126 | | <.0001* |

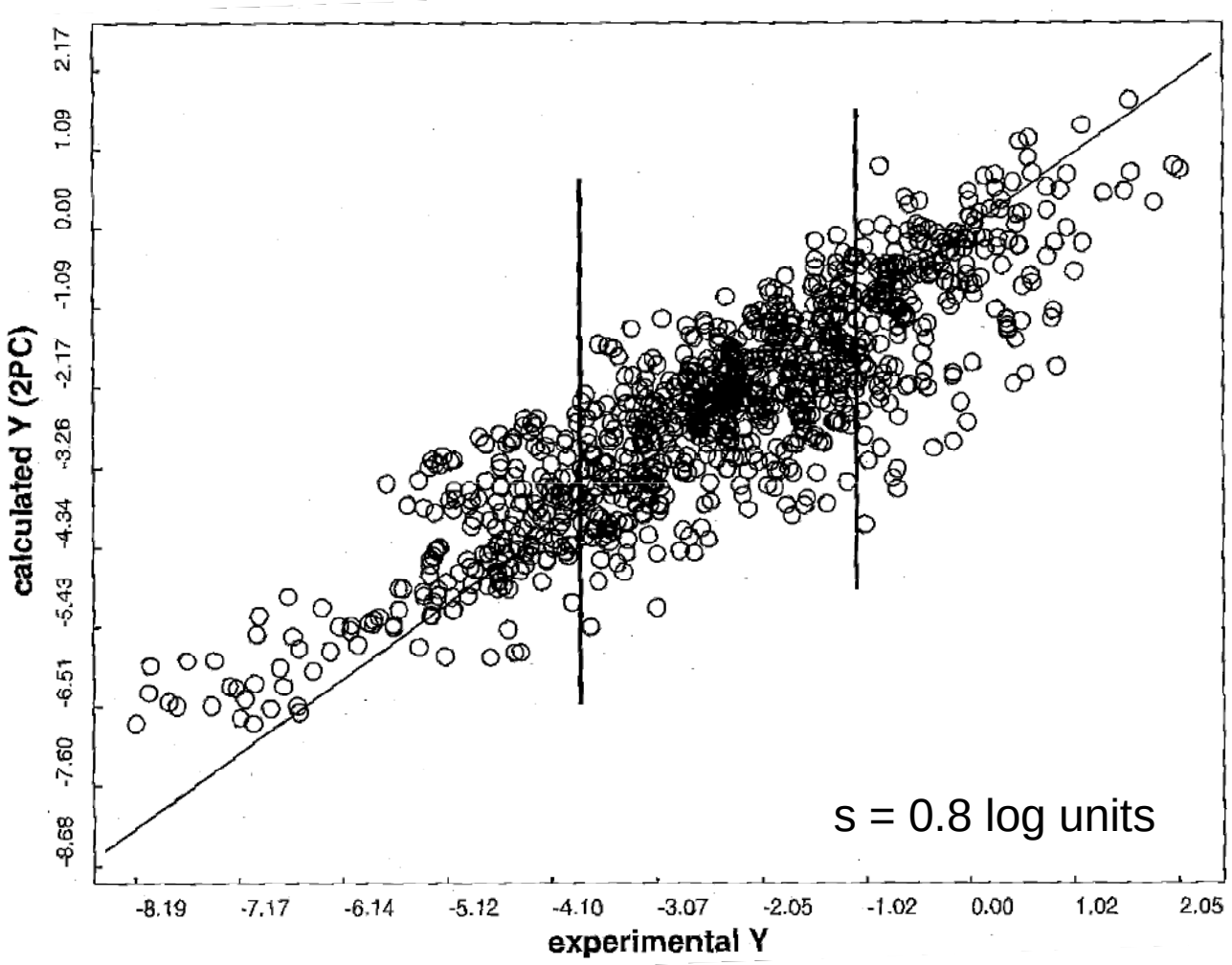
**Parameter Estimates**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 0.2963984 | 0.276688 | 1.07 | 0.2860 |
| c_pKa_MoKa_for_plots | 0.9271127 | 0.037621 | 24.64 | <.0001* |

(*) http://www.moldiscovery.com

# Calcd. (Volsurf+) vs Expt'l. Thermodynamic Sol'y.

(970 compounds)



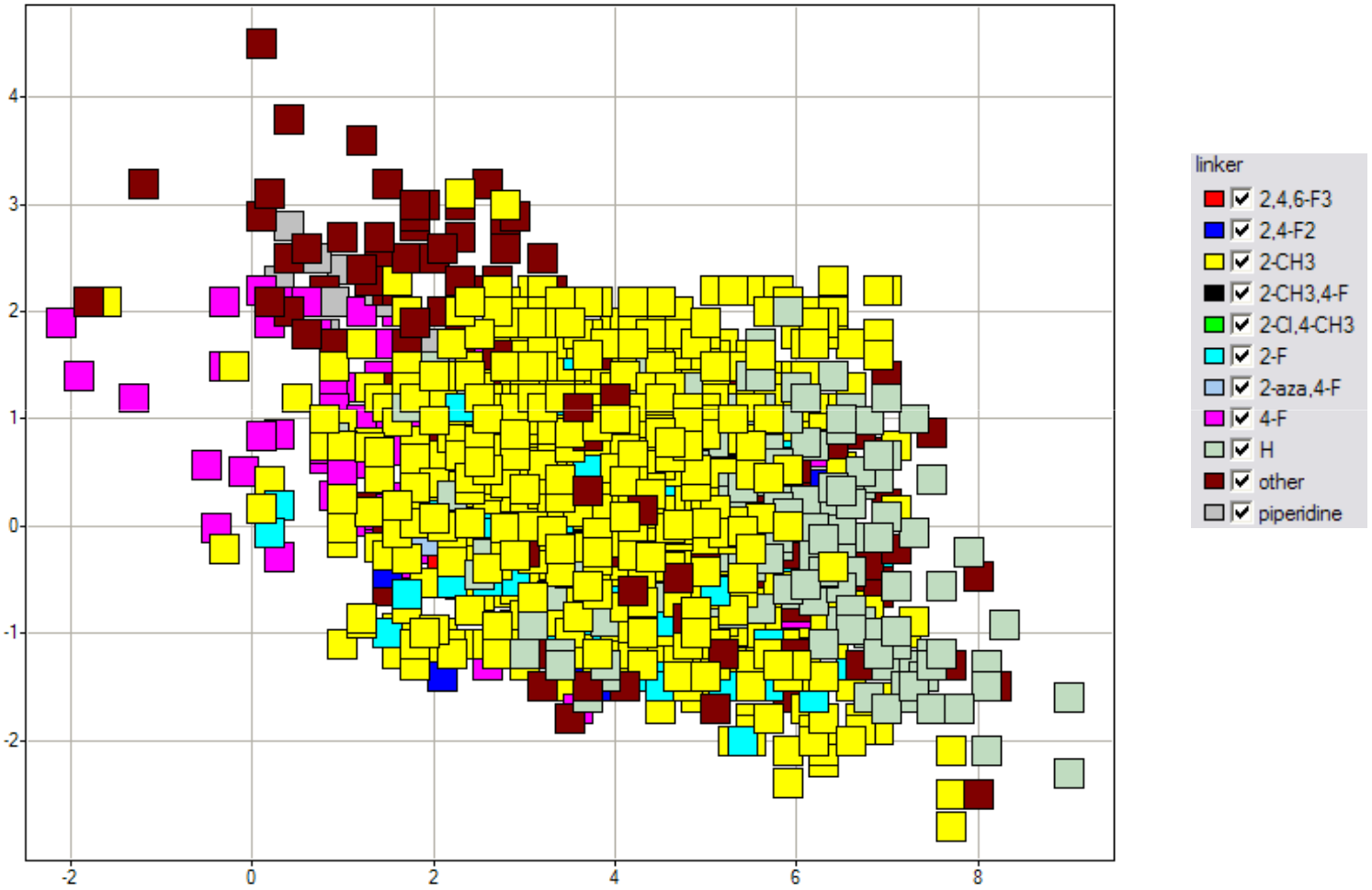

G. Cruciani, P. Crivori, P.-A. Carrupt, B. Testa. Journal of Molecular Structure: THEOCHEM 503, 17-30, 2000. Volsurf+ manual (http://www.moldiscovery.com).

# It’s Not Just About Lipophilicity

## Calculated Solubility vs. cLogD7.4

**cLogS, µM**

**(Stardrop (*) software)**




**cLogD7.4**

(*) http://www.optibrium.com/

# Methods for Deriving DMPK Models

- Regression
- Partial Least Squares
- Neural Networks
- Discriminant Analysis (ADAPT, SIMCA, Support Vector Machines)
- Decision Trees (Random Forest)
- Baysean Methods (probabilistic approaches)
- Use of 3D Structure of Target (CYPs, Transporters, Efflux Pumps,...)

Models are derived using a subset of compounds, then the property is predicted for the held-out compounds (prediction or validation set)

# Molecule Descriptors Used to Derive DMPK Models

- Molecular fingerprints (Pipeline Pilot, MOE, Unity)
- General molecular descriptors (MDL keys, OEchem)
- Calculated properties (AlogP, ClogP, TPSA, ...)
- Connectivity descriptors (e-state keys from Molconn-Z)
- Geometrically derived from 3D structure of target (pharmacophores, correlograms in MetaSite)

# Reporting Results

- Predicted value
- Confidence in the prediction (std. error, probability, quality of prediction)
- The nearest neighbor in the training set (Tc* is typically used)
- The number of near neighbors in the training set (above a Tc threshhold)
- Geometric fit score (docking, pharmacophore overlap, shape similarity)
- Details (model version, date)

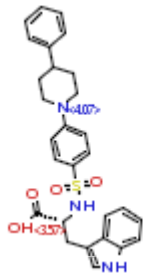
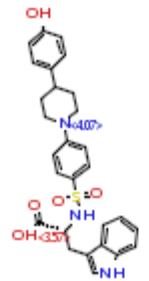
*Tc = Tanimoto coefficient = difference in binary fingerprints between two compounds = $T(A,B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B}$

# Reporting Results: Example (*)

**cLM_HRM = Calculated liver microsomal stability in Human, Rat, and Mouse**

Calculated Properties

A1 | Structures

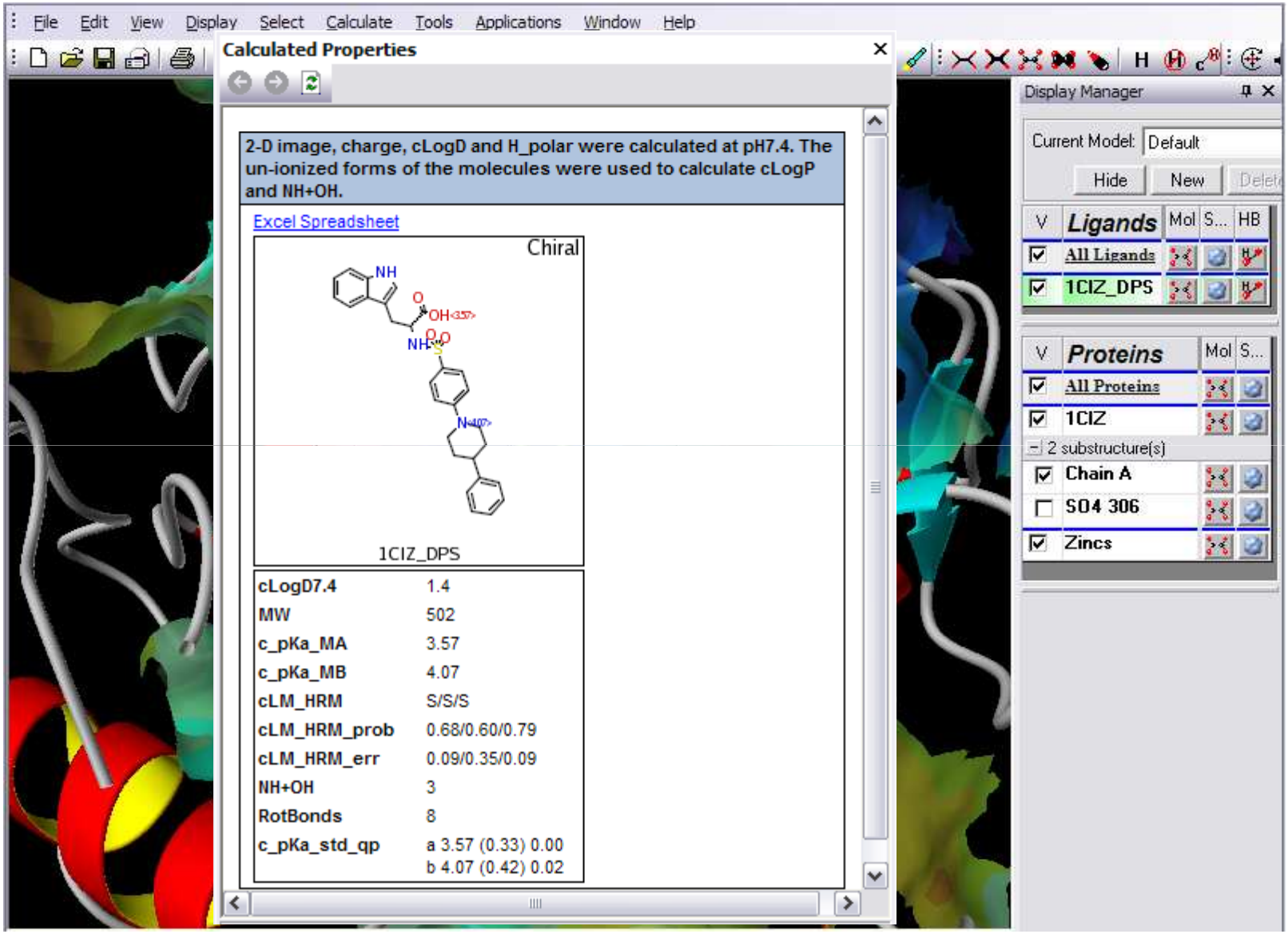
| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Structures | Name | cLogD7.4 | MW | c_pKa_MA | c_pKa_MB | cLM_HRM | cLM_HRM_prob | cLM_HRM_err | cHLM |
| 2 | Chiral  1CIZ_DPS | 1CIZ_DPS | 1.4 | 502 | 3.57 | 4.07 | S/S/S | 0.68/0.60/0.79 | 0.09/0.35/0.09 | S |
| 3 | Chiral  1CIZ_DPS_4-OH | 1CIZ_DPS_4-OH | 0.7 | 518 | 3.57 | 4.07 | S/L/S | 0.75/0.49/0.70 | 0.09/0.40/0.16 | S |
| 4 | | | | | | | | | | |

**Result, probability, and error are reported**

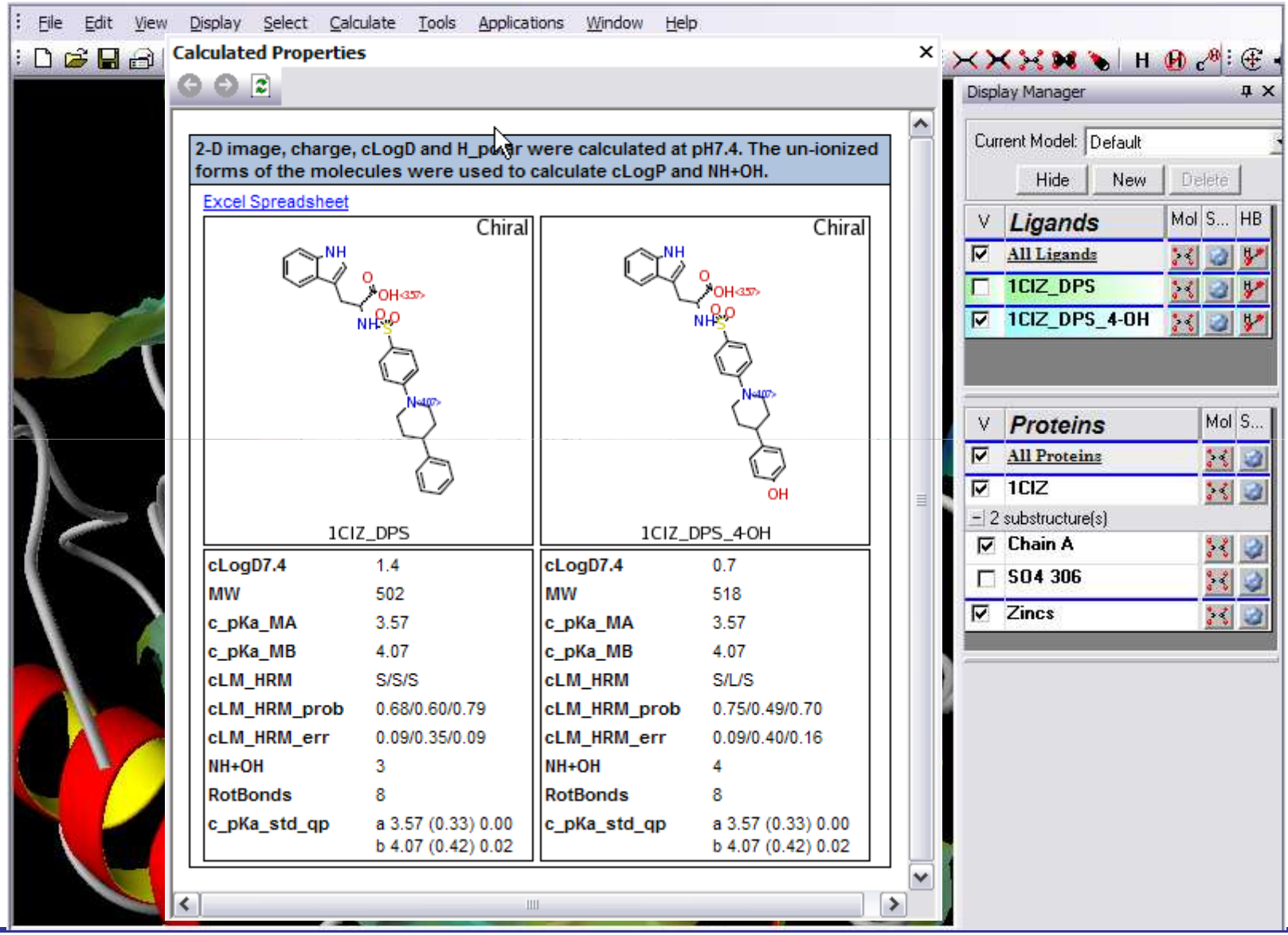
horizontal_view | for_plotting

(*) Output from a property calculator used at Genentech

# Combining 2D and 3D Worlds



# Comparing Multiple Ligands



# Commercial Software

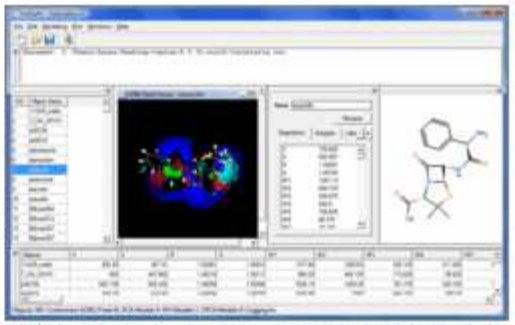
| | |
|---|---|
| • Pipeline Pilot v7.5 | Accelrys |
| • Stardrop | Optibrium |
| • MoKa, MetaSite, Volsurf+ | Molecular Discovery |
| • ACDlabs | Molecular Discovery Ltd. |
| • ADME Boxes | PharmaAlgorithms |
| • ADMET predictor | Simulations Plus (Now merged with Molecular Discovery Ltd.) |
| • SARchitect | Strand Life Sciences |
| • Metabolizer | ChemAxon |
| • Spotfire (visualization) | Tibco |
| • Vortex (visualization) | Dotmatics |

# From Molecular Discovery Ltd.

2009-02-16 18:23

**New Release! VolSurf+ 1.0**

Use predictive ADME more effectively in lead identification and optimization. Calculate over 100 GRID-based ADME relevant descriptors to prioritize hits, create and explore models, and interactively optimize compounds in ADME space using created or provided libraries.



VolSurf+ is a completely re-architected solution based on the popular VolSurf 4, with improved usability and data handling, as well as new descriptors and analyses.

Multi-platform support enables computational and medicinal chemists to work together more effectively, and three task-based interfaces are now provided to help support this: VolSurf+ Selector enables the virtual screening of compounds using ADME relevant descriptors, VolSurf+ Modeller enables the detailed modelling of physicochemical properties through a range of statistical analyses and graphs, and VolSurf+ Designer allows the interactive design of compounds with simultaneous projection in multiple models.

VolSurf+ creates 128 molecular descriptors from 3D Molecular Interaction Fields (MIFs) produced by our software GRID, which are particularly relevant to ADME prediction and are also simple to interpret. One example would be the interaction energy moment descriptor between hydrophobic and hydrophilic regions, which is important for membrane permeability prediction. These can then be used with provided chemometric tools to build statistical models.

VolSurf+ also comes with a number of models that we have developed using both public and pharmaceutical data, including passive intestinal absorption, blood-brain barrier permeation, solubility, protein binding, volume of distribution, and metabolic stability.

# Strategies For Implementation

Obtain commercial software or develop your own

-> A full-featured ‘chemically aware’ graphing package is a must

Try generating global models on DMPK endpoints first

->If this fails, try project- or chemotype- specific models

Report probabilities and errors along with the calculated values

Track calculated vs. measured values on a regular basis

-> Continually check the models’ performance (predicted vs. measured)

-> Update the model regularly with new data

-> Regularly discuss results with chemists

Add calculated properties to your compound database

-> Facilitates searching, subsetting, and rank ordering of compounds

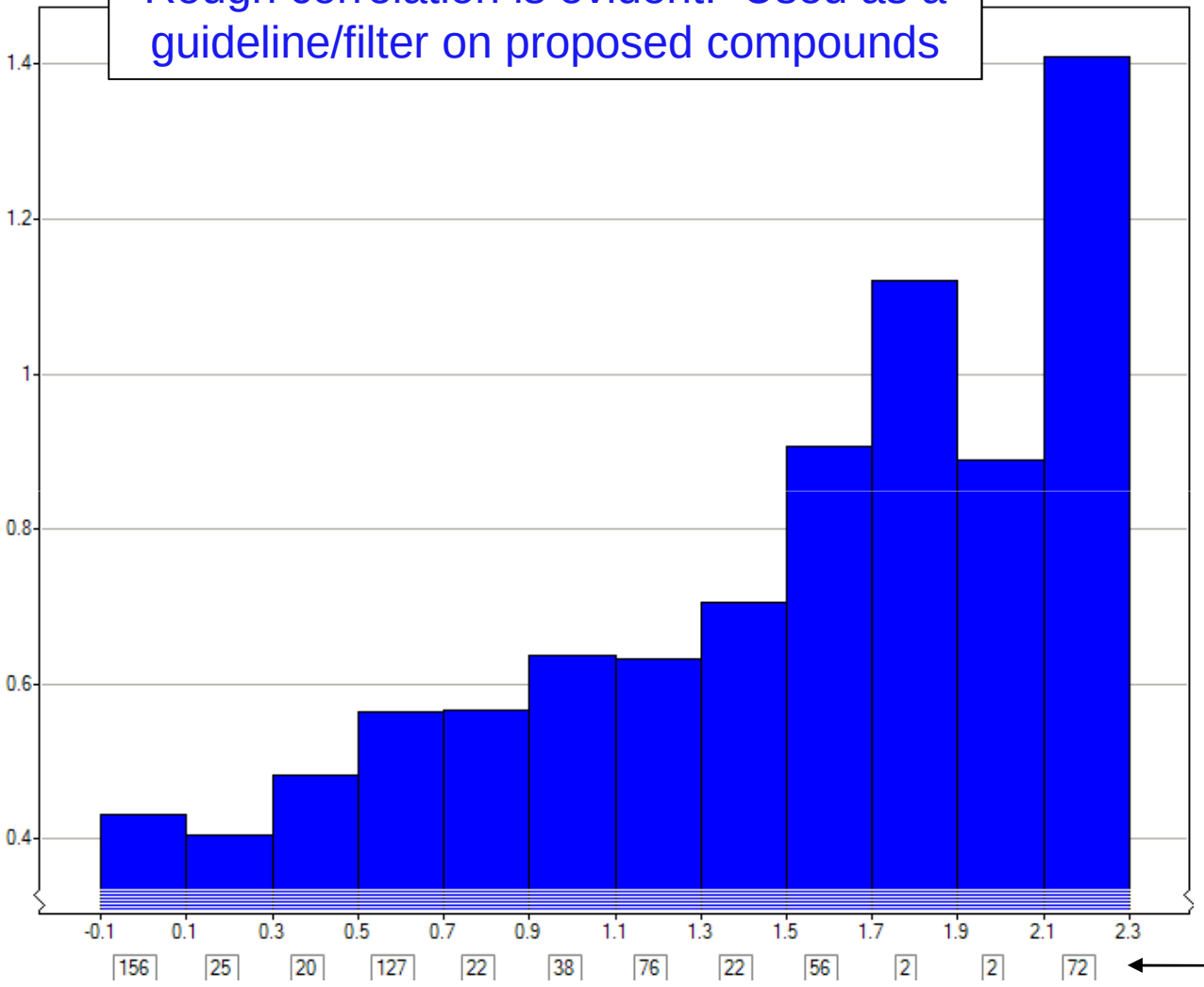
# “Real Project” Example

# Solubility: Calculated vs Measured (Kinetic)

Rough correlation is evident: Used as a guideline/filter on proposed compounds

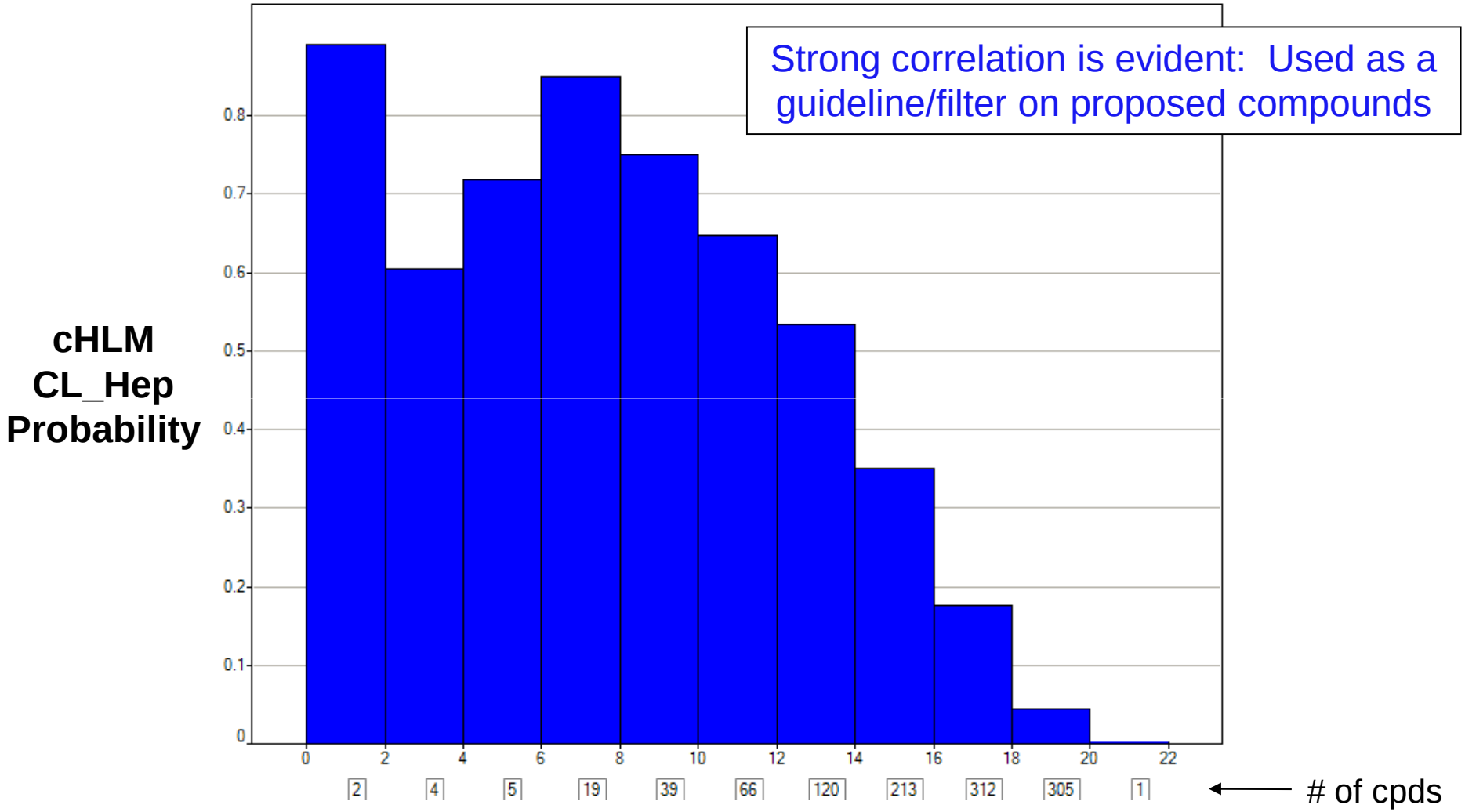
**Average cLogS, µM**

**(from Stardrop software)**

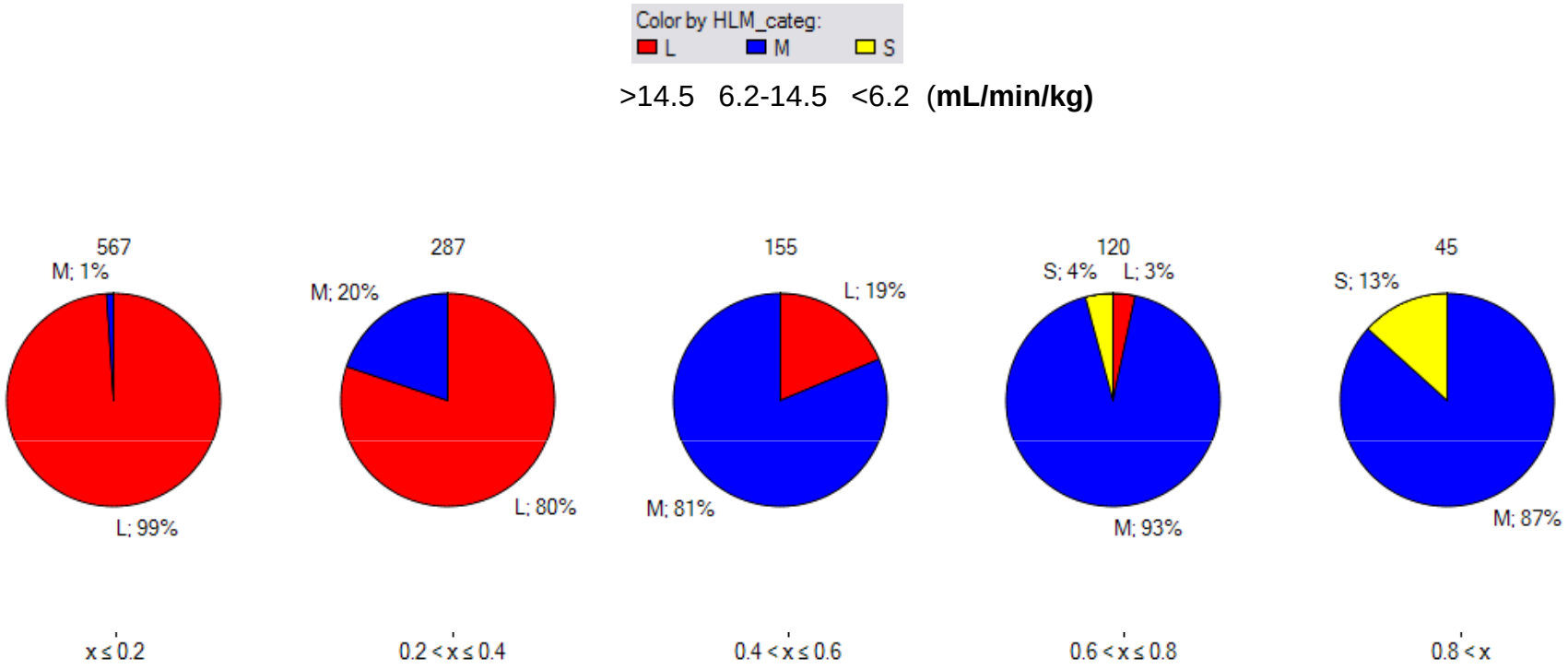



**Measured LogS (kinetic)**

# Performance of Human Liver Microsome (HLM) Stability Model



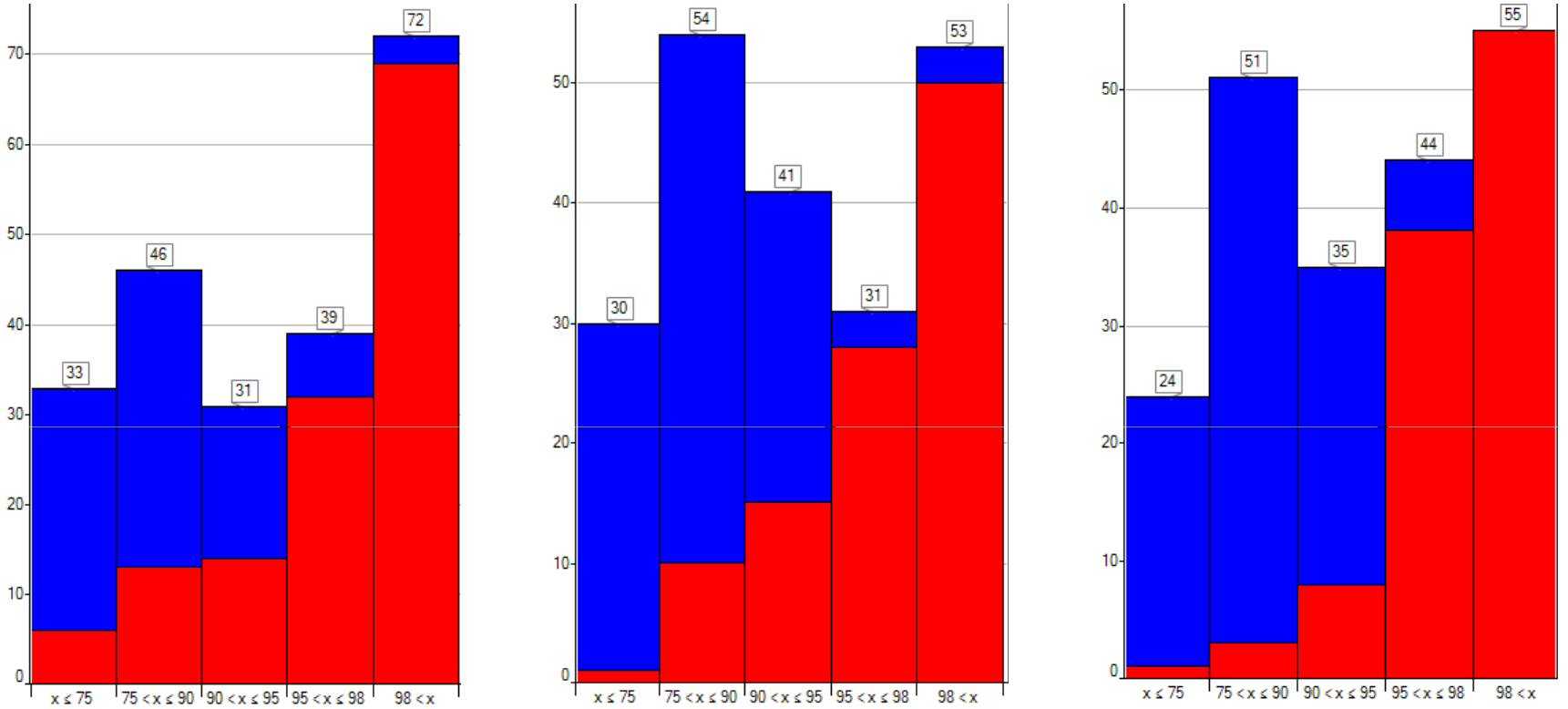
# Predictivity of HLM Stability Model




## Calculated HLM Probability

# PPB Model Validation: Human, Rat, Mouse

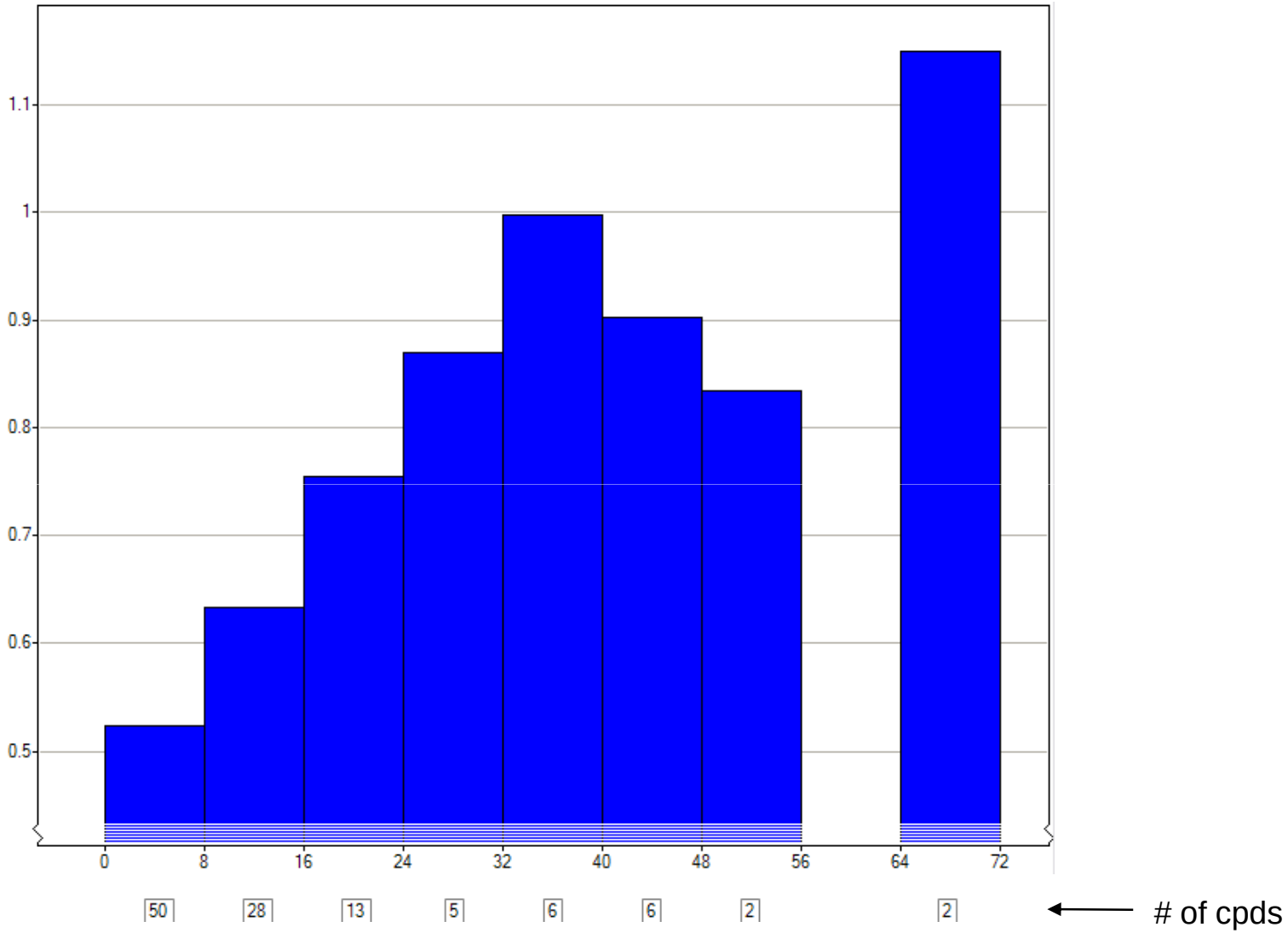
**(Training set, 750 compounds; Test set, ~250 compounds)**



# Calculated Solubility vs PK Rat F%

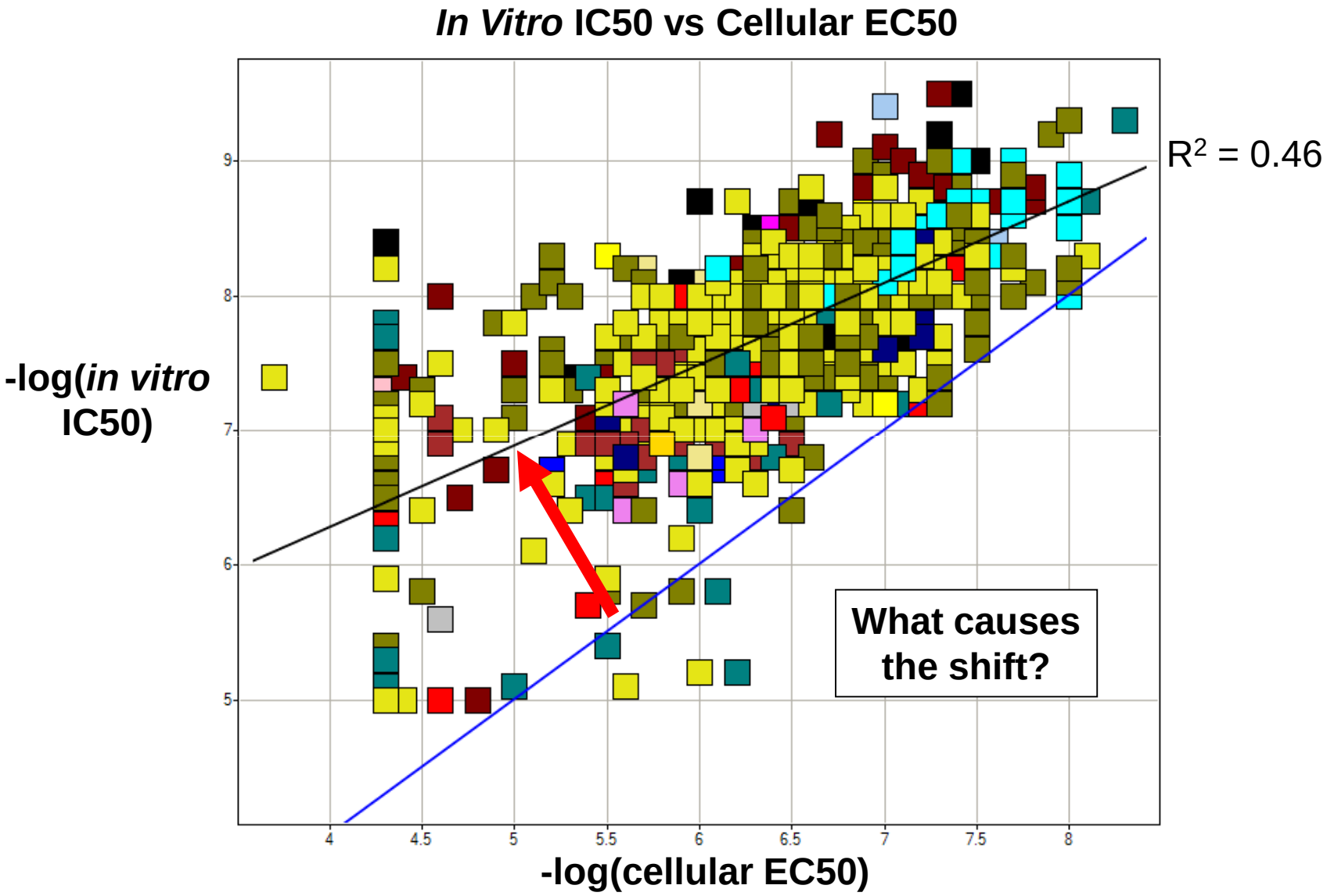
**Average cLogS, µM**

**(from Stardrop software)**



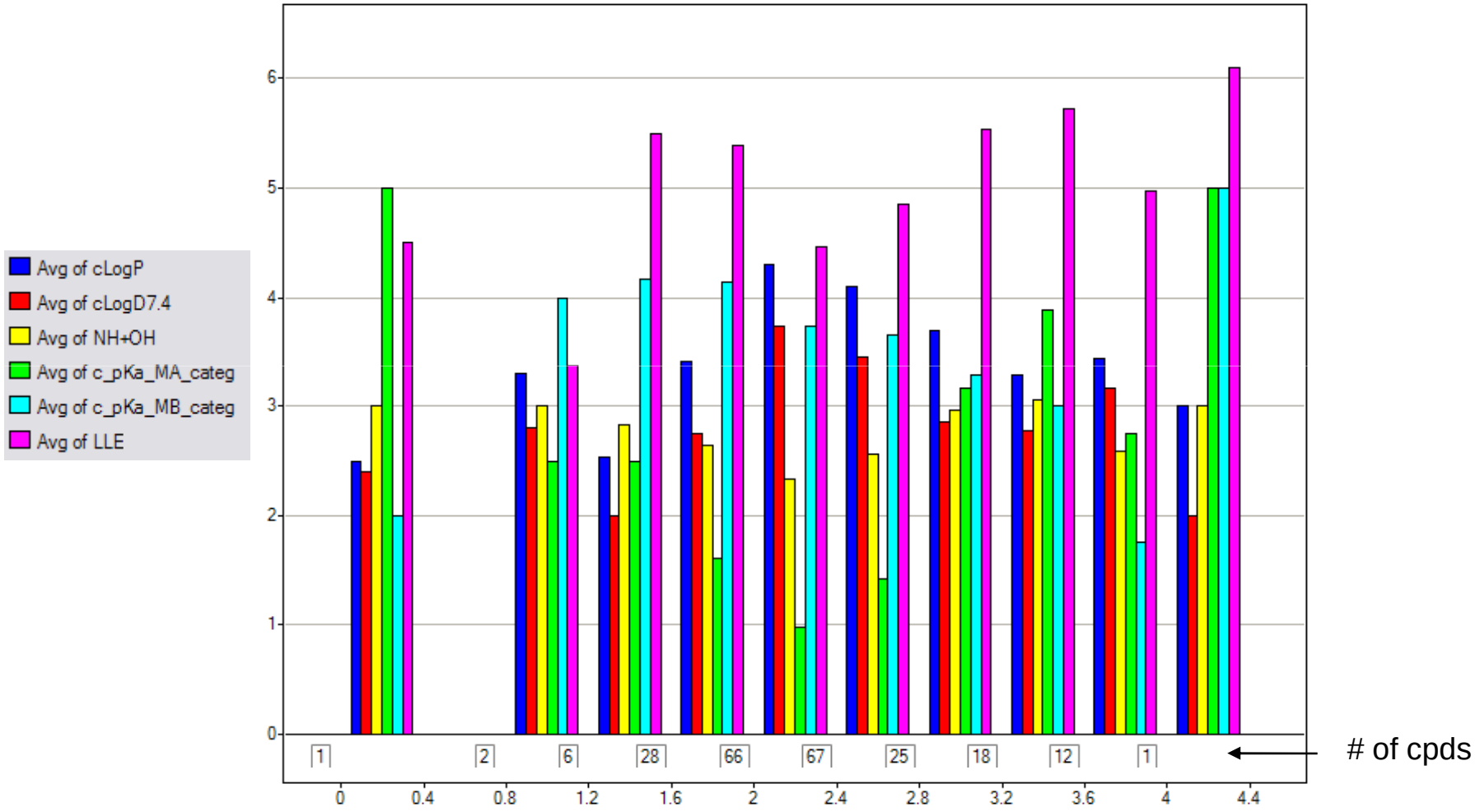

**PK Rat F%**

# In Vitro vs. Cellular Potency Disconnects



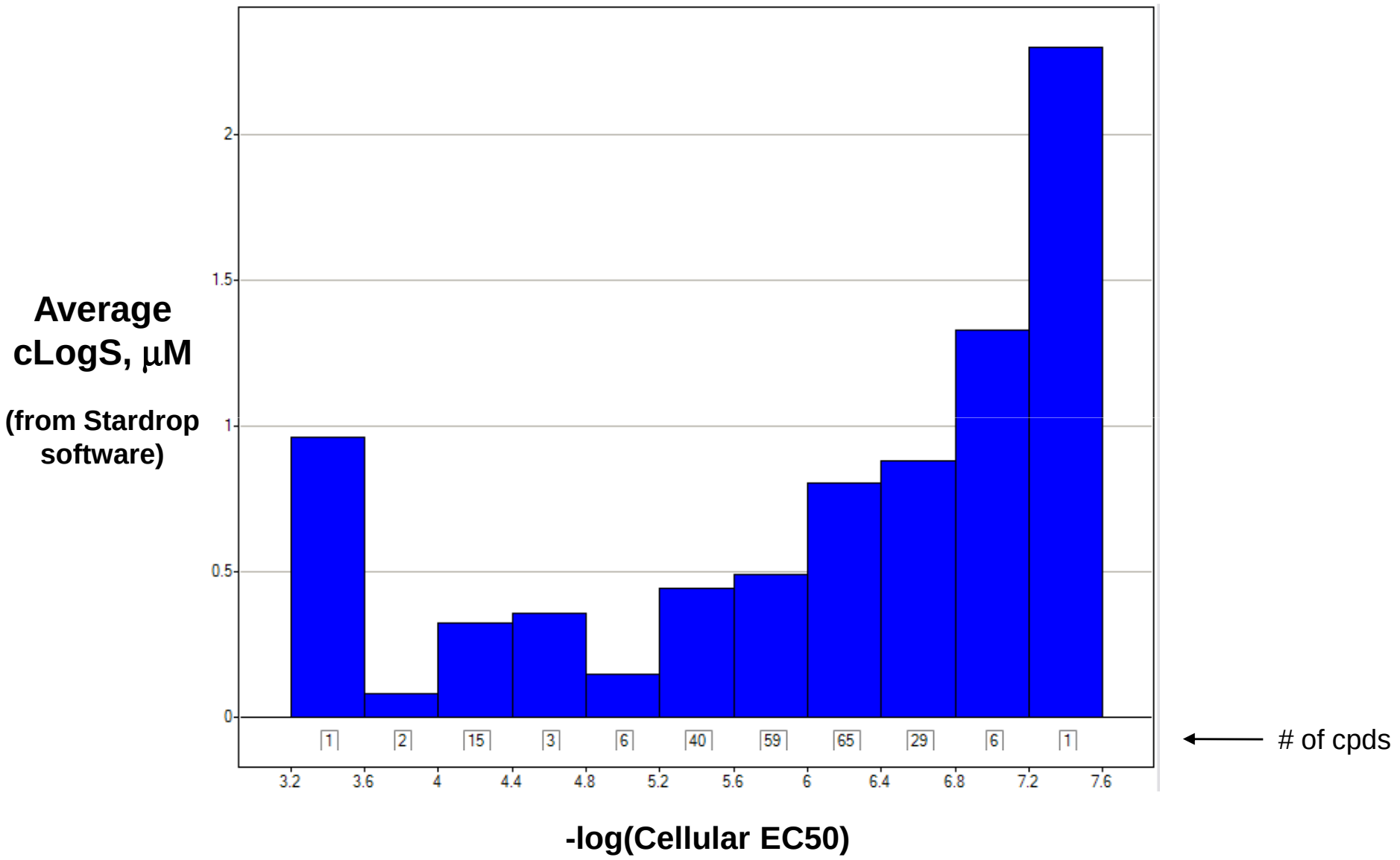
# In Vitro vs. Cellular Potency Disconnects

## Looking for Relationships With Calculated Properties




**CellE In Vitro vs Cellular Potency**

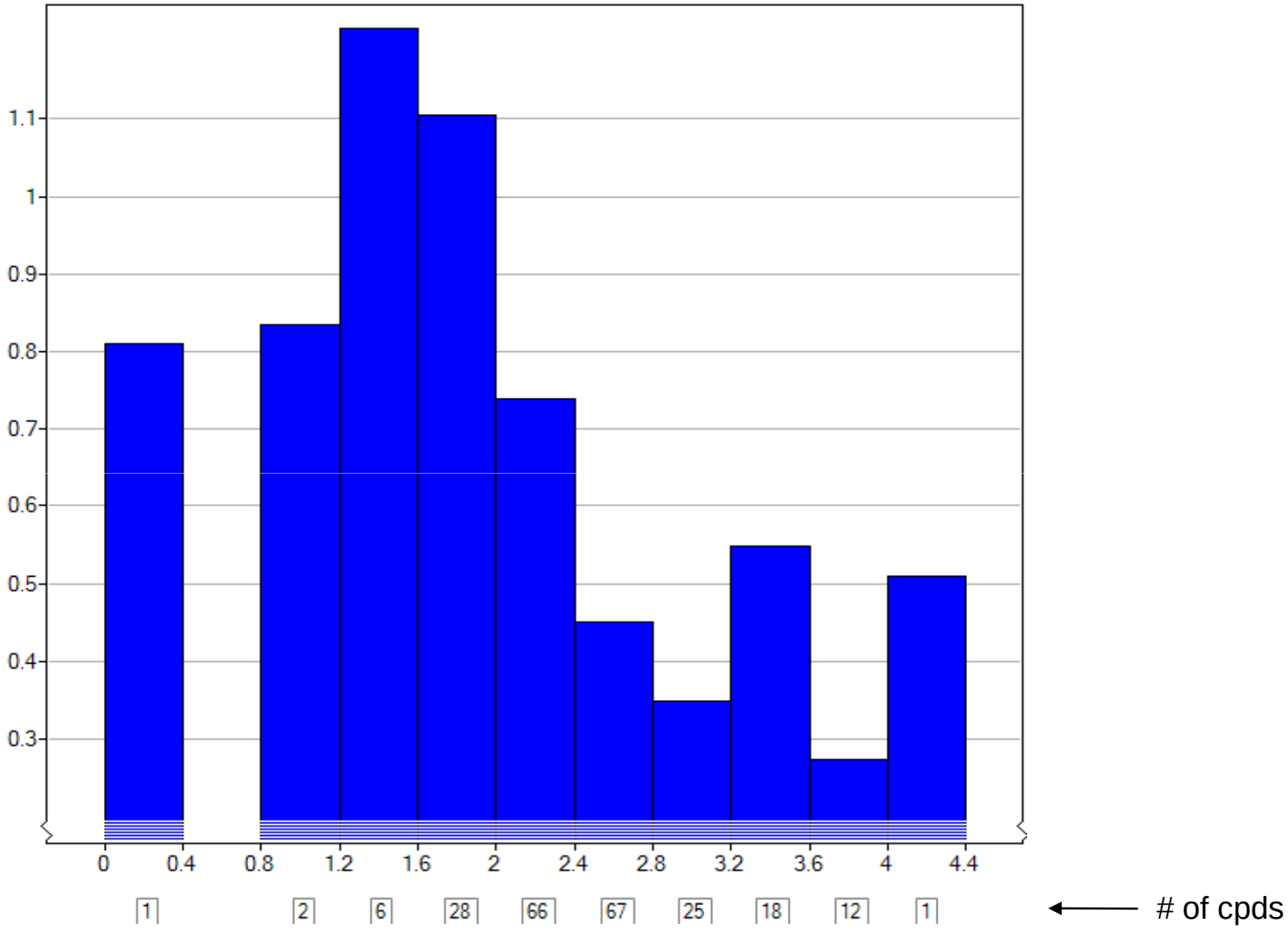
# Calcd. Solubility vs. Cellular Potency



# Increased Sol'y. Reduces In Vitro/Cellular Pcy. Disconnect

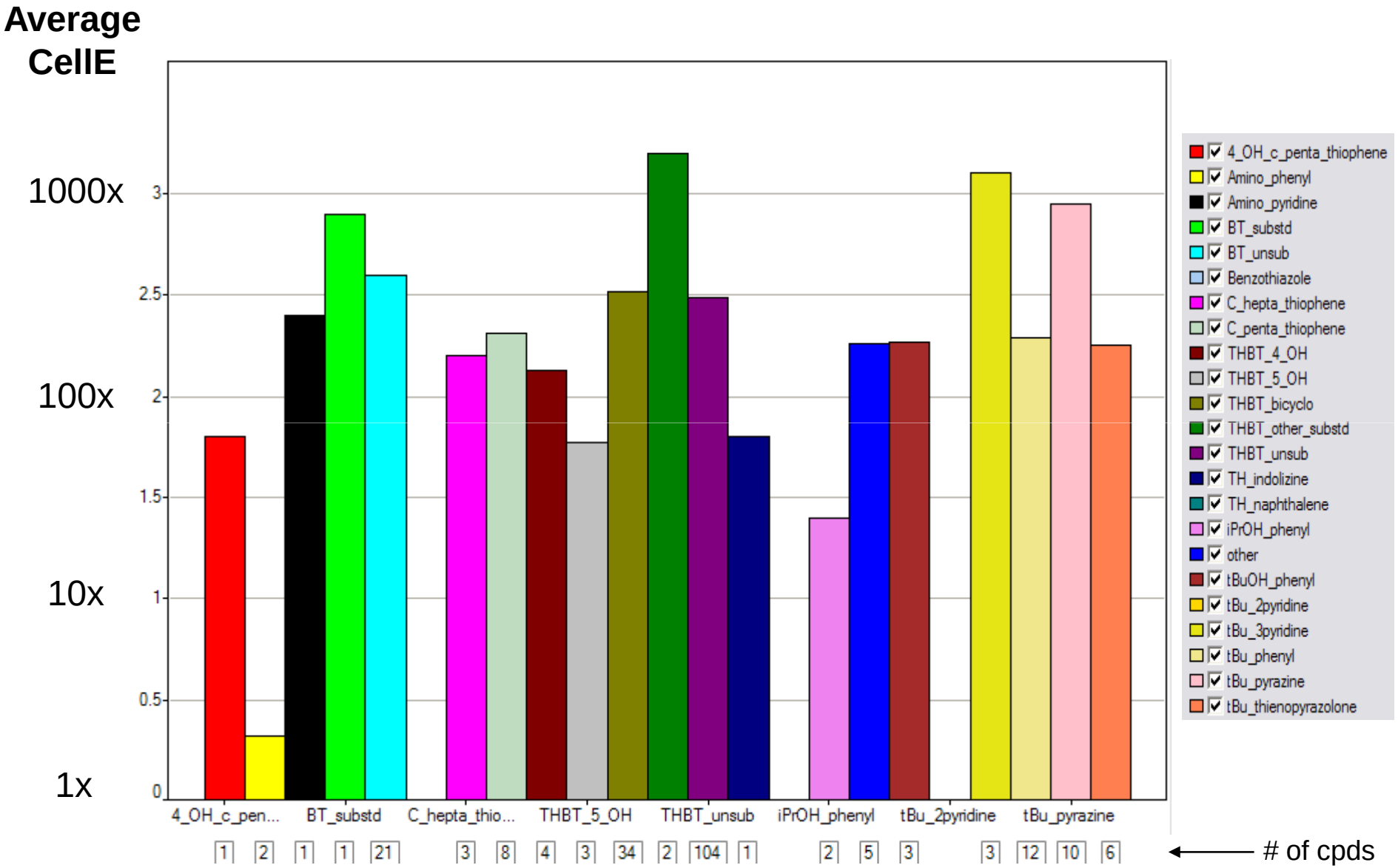
**Average cLogS, μM**

**(from Stardrop software)**

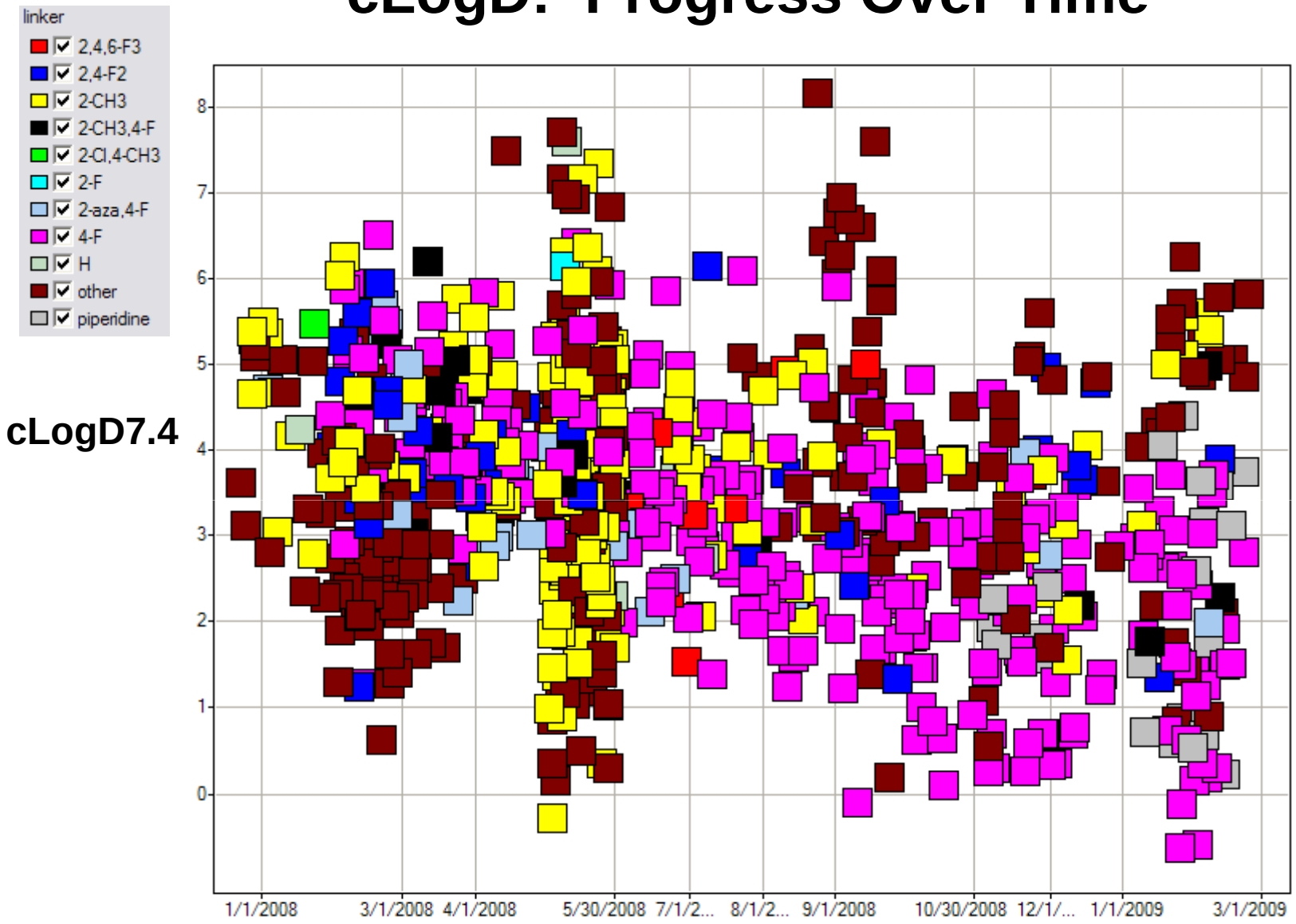



**CellE (in vitro/cellular potency)**

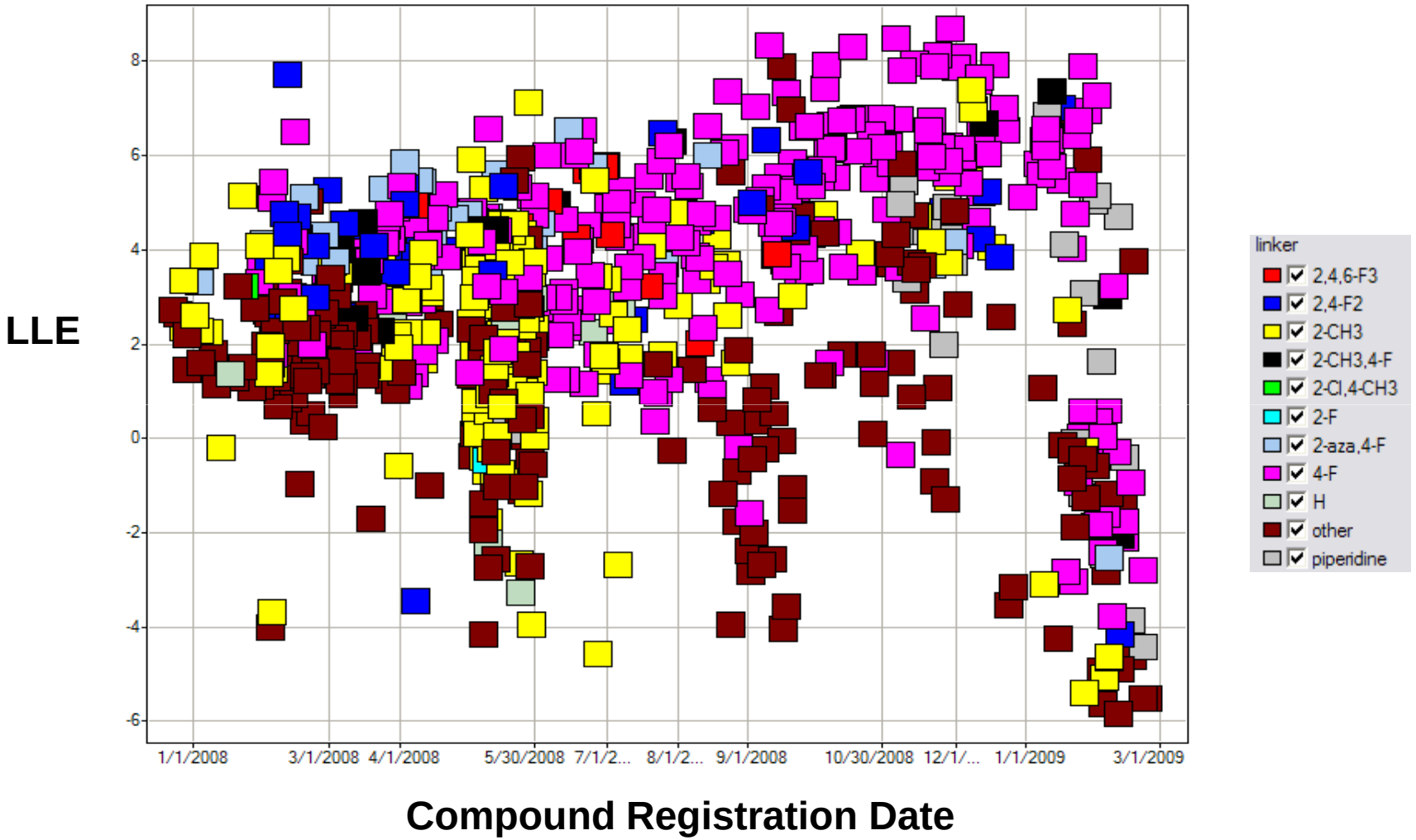
# CellE In Vitro vs Cellular potency by R Group



# cLogD: Progress Over Time



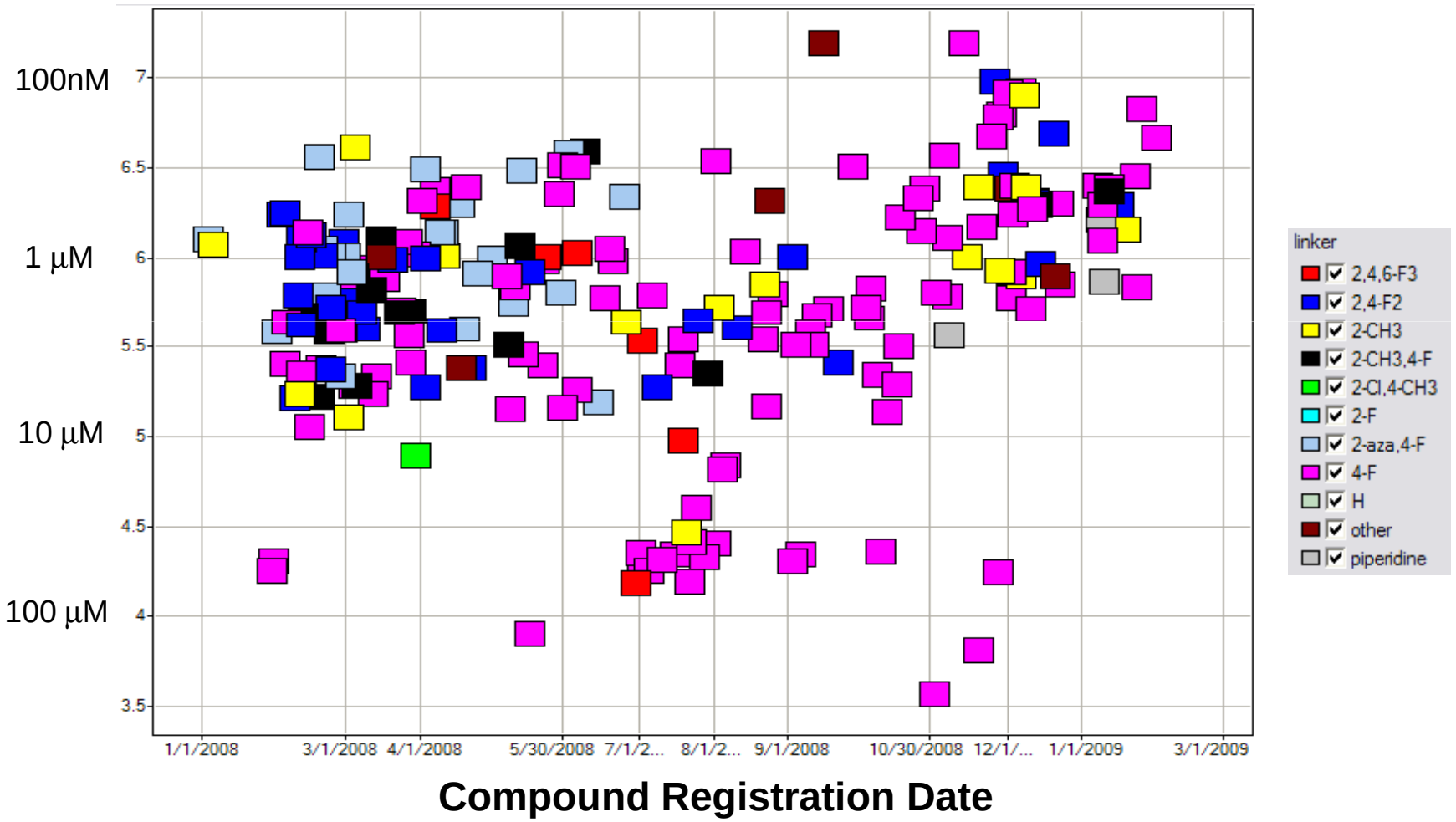
# Ligand Lipophilic Efficiency(*): Progress Over Time




(*) LLE = pIC50 – cLogD7.4 [Values of 7-9 Desirable]

# Whole Blood Potency:  Progress Over Time

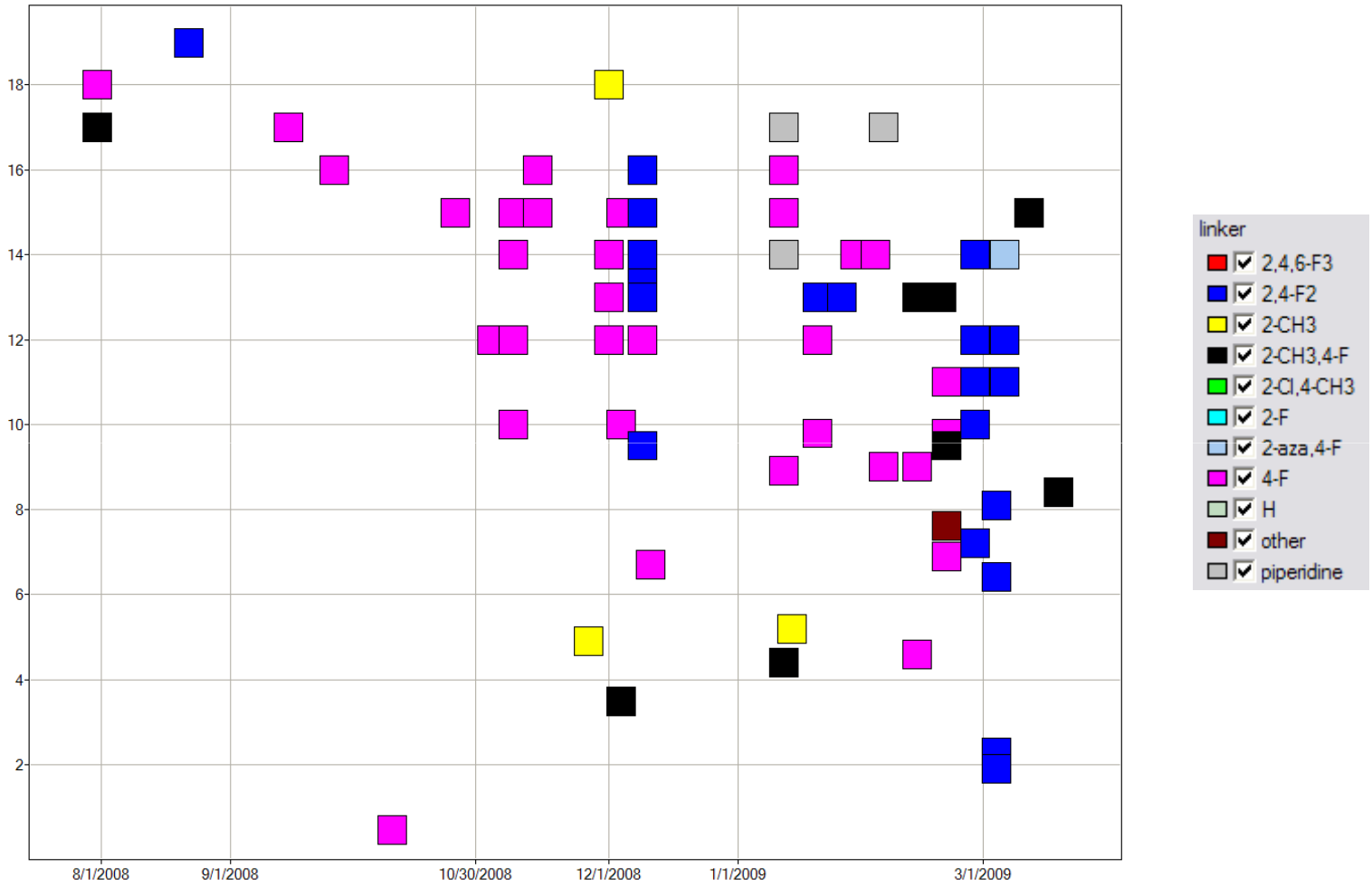
**-log(EC50)**




**Compound Registration Date**

# Stability in Hepatocytes: Progress Over Time

**CLhep, mL/min/kg**




**Compound Registration Date**

# Med Chem Prioritizes Based on Filters

7041 compound targets

- 87% cHLMprob >0.6

884

67% of these have TPSA >120

- 72% clogS >0.8

241

- 58% TPSA <130

100

- 19% clogD<3; pKa>6

81

SAR sense synthesis

Attrition rate = 99%

Survivors = high quality targets

Are these the right filters?

**Output limited only by input**

# The Future of *In Silico* Property Calcns

- Wider availability of in silico methods, models, databases
- Improved predictions of solubility, crystallinity

  -> Avantium
- Improved prediction of *in vivo* endpoints
- Combination of 2D and 3D models
- Models that suggest molecules to make
- Application to exhaustive chemical databases

  -> eMolecules, ChemUniverse
- Toxicity modeling

  -> Pharmatrope

# Conclusions

- Marketed drugs exhibit defined property profiles
- Calculating properties in advance helps avoid unproductive compounds
  - -> Use calculated properties where it makes sense
  - -> You can get there faster!
- Projects benefit by calculating properties on proposed cpds.
- Not all models will work for all projects
  - -> “The important thing is not to stop questioning”
- Calculations are meant to be guidelines.....
  - -> If there are compelling reasons to make the compound, do so!

# Conclusions

- Commercial software is getting better, but ‘built-in’ DMPK models remain approximate
  - -> Usually better to derive your own models if data are available!
- “Global” models are preferable
  - -> Many more and varied molecules used -- more robust predictions
  - -> In many cases, more approximate predictions result
- If Global models don’t work, develop “Local” models on data from just one project
  - -> Quite accurate predictions inside compound space possible
  - -> Often, limited prediction accuracy outside compound space
- Delivering models to bench scientists facilitates their use/uptake
- Delivering results from approximate DMPK models as probabilities is preferable to delivering the actual prediction

# References

**Drug-like Properties: Concepts, Structure Design and Methods.** Kerns, Edward H and Di, Li. UK. (2008), 526 pp. Publisher: (Elsevier Ltd., Oxford, UK)

*-> Basic textbook that contains sections on in silico calculations DMPK properties. A good place to start!*

**Comprehensive Medicinal Chemistry II, Vol 5: ADME-Tox Approaches.** Taylor, John B.; Triggle, David J.; Editors. UK. (2006), 1152 pp. Publisher: (Elsevier Ltd., Oxford, UK)

*-> An in-depth treatment of in silico tools to predict DMPK properties, written by experts in the field.*

**Molecular Drug Properties. [In: Methods and Principles in Medicinal Chemistry, 2008; 27].** Mannhold, Raimund, Editor. Germany. (2008), 471 pp. Publisher: (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany)

*-> Focuses on molecular descriptors and their calculation.*

**Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability. [In: Methods and Principles in Medicinal Chemistry, 2003; 18].** Van de Waterbeemd, Han; Lennernas, Hans; Artursson, Per. Germany. (2003), 579 pp. Publisher: (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany)

*-> Somewhat dated, but contains useful practical advice on the application of in silico models.*

# References

**Defining optimum lipophilicity and molecular weight ranges for drug candidates-Molecular weight dependent lower log D limits based on permeability.** Michael J. Waring. Bioorganic & Medicinal Chemistry Letters (2009), 19(10), 2844-2851.

*-> An example of a body of literature that report relationships between DMPK properties (permeability in this paper) to optimum property ranges (molecular weight and log D in this paper).*

**Physicochemical drug properties associated with in vivo toxicological outcomes: a review.**
David A. Price, Julian Blagg, Lyn Jones, Nigel Greene, Travis Wager. Expert Opinion in Drug Metab. Toxicol. (2009) vol. 5 (8) pp. 921-931.

*-> A forward-looking review of in silico calculation of in vivo tox endpoints by experts in the field.*