

Construction of QSAR models for passive permeability from a Parallel Artificial Membrane Permeability Assay (PAMPA) dataset

UK QSAR and Chemoinformatics meeting,
Spring 2004

**Matt Wood, Physical Chemistry Group, Mereside, AstraZeneca, Alderley Park,
Macclesfield, Cheshire, SK10 4TG**

**Patrick Barton and Roger Fessey, Department of Physical and Metabolic
Science, AstraZeneca R&D, Charnwood, Bakewell Road,
Loughborough, LE11 5RH**

Overview

- Permeability in discovery
- Experimental set-up
- PAMPA QSAR
 - Literature studies
 - AZ screening:
 - Dataset characterisation
 - Correlation with $\log D_{7.4}$
 - Predictive models

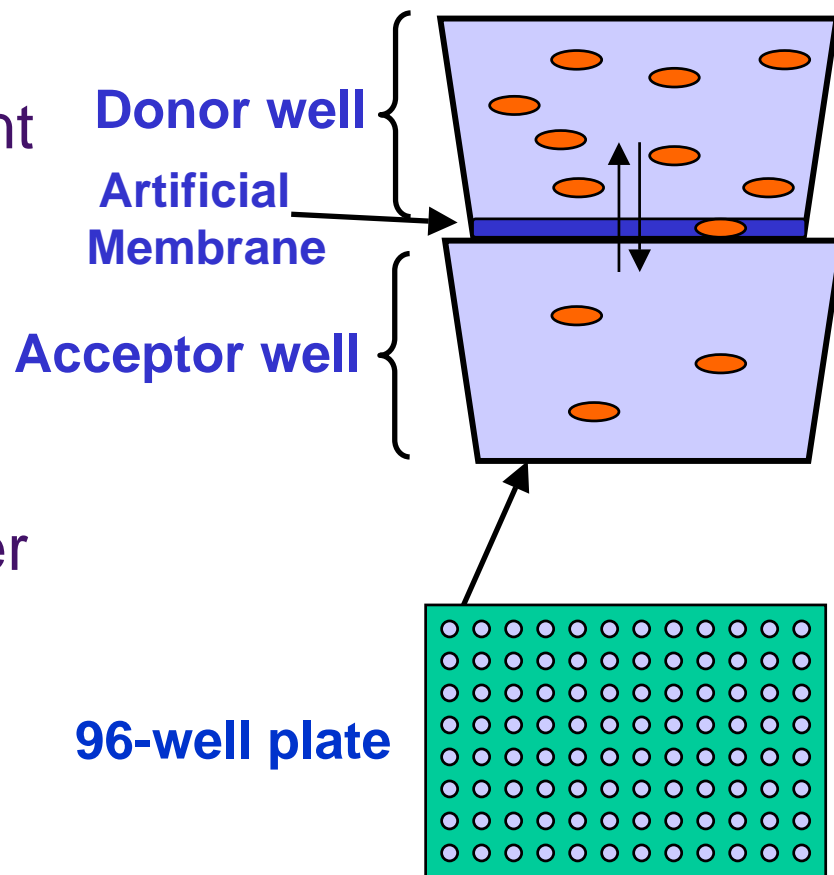
Permeability in discovery

- Need to predict compound permeability early in drug discovery to reduce attrition.
- Widely used Caco-2 screen is a relatively low throughput and expensive model of *in-vivo* situation.
- The screening or prediction of drug candidates for passive permeability provides essential ADME information.
- PAMPA (Parallel Artificial Membrane Permeability Assay) provides a rapid method for the measurement of a compound's ability to cross an artificial membrane.
- Allows passive permeability to be determined directly.
- Valuable front line screen, combination with Caco-2 would enable mechanistic studies.

Experimental PAMPA

PAMPA experimental set-up

- Sample is allowed to permeate for 4-18 hours.
- Fast or slow flux, do not want equilibrium as cannot then quantify permeability.
- Plate 'sandwich' is then separated and each well analysed *via* UV plate reader or LC-MS.
- Apparent permeability constant, P_{app} , calculated from relative concentrations.

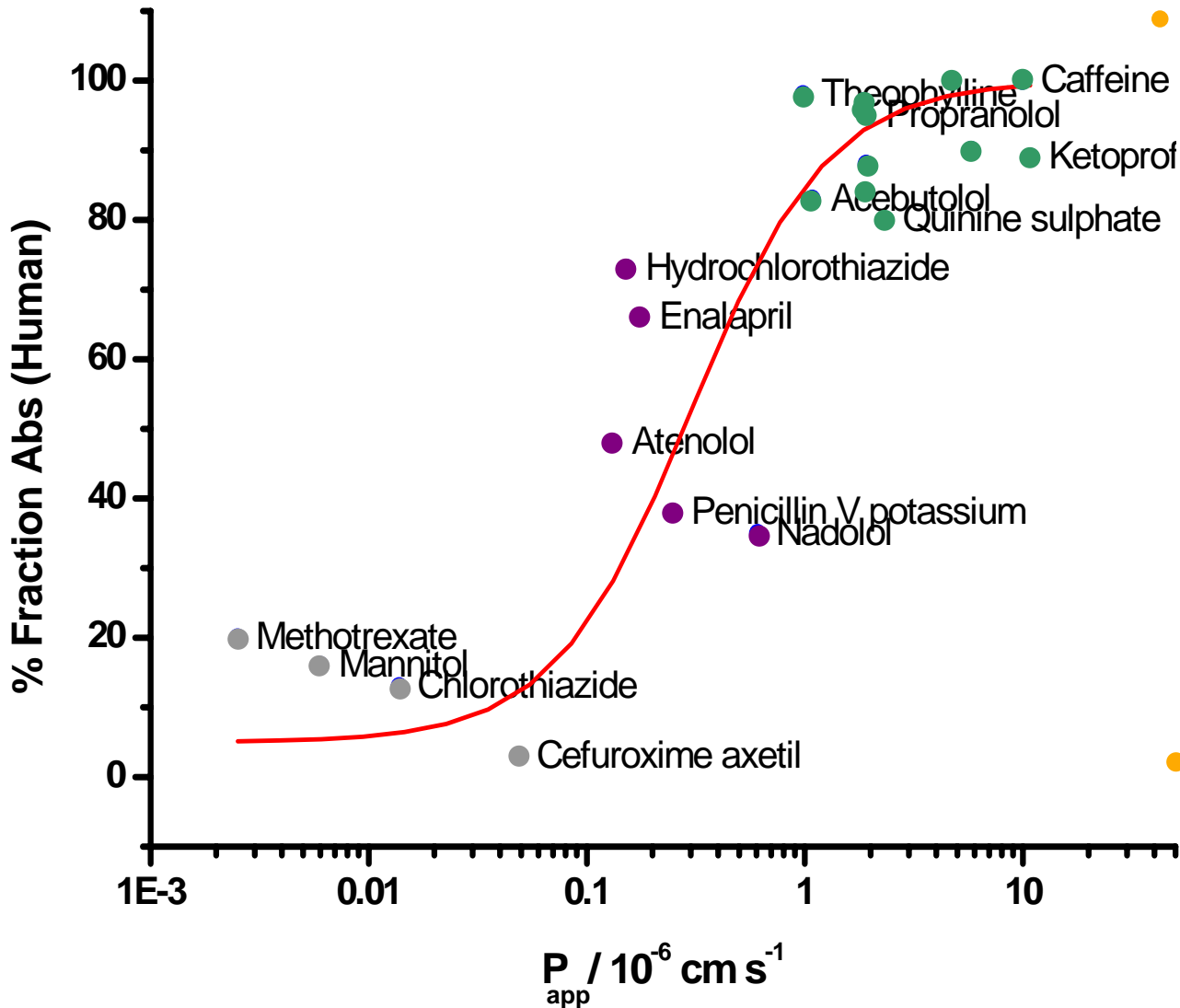


2-way flux equation

$$C_{A(t)} = \left(\frac{n - n_{mem}}{V_D - V_A} \right) + \left(C_{A(0)} - \left(\frac{n - n_{mem}}{V_D - V_A} \right) \right) e^{-P_{app} A (1/V_D + 1/V_A) t}$$

- Solution to Fick's First Law of Diffusion. Adapted from Palm *et al. J. Pharm. Exp. Ther.*, 1999, 291, 435.
- Equation determines apparent permeabilities, P_{app} .
- Does not assume sink conditions.
- Accounts for mass of compound lost to membrane, %Membrane retention.
- Accounts for back-flux if membrane partition coefficients are equal in acceptor and donor (iso-pH).

Correlation with %Fa (Human)



- Distinct groups:

- $P_{app} \geq 1.0 =$
Fa ~80-100 %

- $P_{app} \geq 0.1 \leq 1.0 =$
Fa ~20-80 %

- $P_{app} \leq 0.1 =$
Fa < 20 %

- $P_{app} \text{ RSD} <$
20 %

QSAR of PAMPA permeability

Literature work

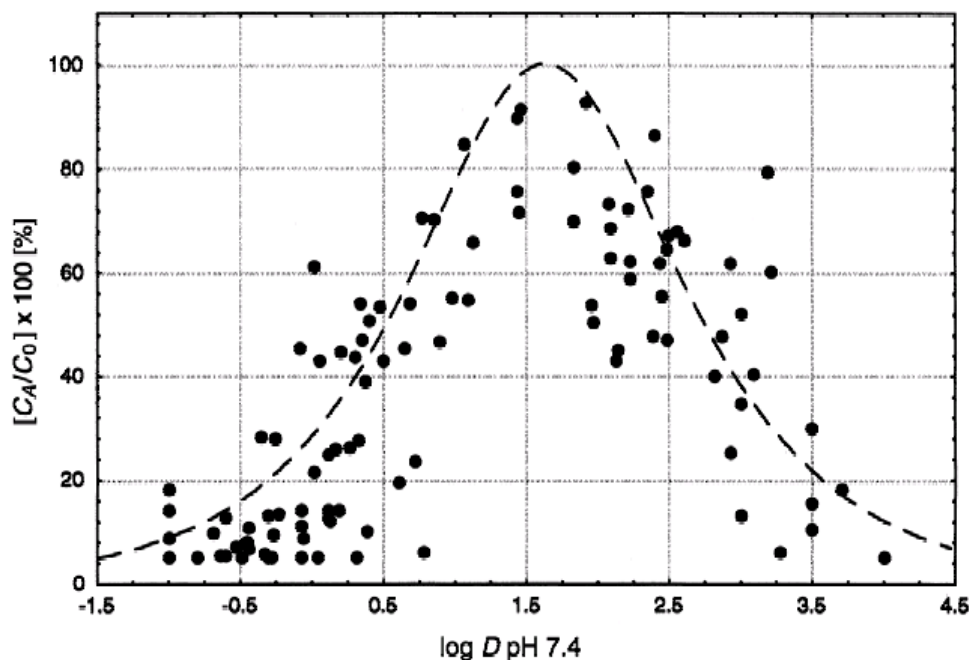
- Most studies show correlations with permeability measurements from more complex systems:
 - %Fa (Human) Kansy (*J. Med. Chem.*, 1998, 41, 1007)
Sugano (*Int. J. Pharm.*, 2001, 228, 181)
Faller (*J. Med. Chem.*, 2001, 44, 923)
Zhu (*J. Pharm. Pharmaceut. Sci.*, 2003, 6, 315)
 - Caco-2 P_{app} / %Fa Zhu (*Eur. J. Med. Chem.*, 2002, 37, 399)
 - Caco-2 P_{app} / P_{eff} (rat) Avdeef (*Eur. J. Pharm. Sci.*, 2004, 21, 429)
 - P_{eff} (Human) Avdeef (*Drug Bioavailability*, 2002, Wiley-VCH)
 - CNS +/- Kerns (*Eur. J. Med. Chem.*, 2003, 38, 223)
- Sugano has added functions for paracellular and UWL permeability to give combined model for %Fa. (*Int. J. Pharm.*, 2002, 241, 241, *Int. J. Pharm.*, 2003, 257, 245)

Literature PAMPA QSAR

- Work on permeability of peptide and indole derivatives (Ano *et al.* *Bioorg. Med. Chem.*, 2004, 12, 257).
- Mixture of charge types, logP -0.48 to 4.93, pH 7.3.
- Correlation with Caco-2 reveals transport mechanism.
- Analysis of PAMPA data:
 - “Classical” QSAR:
 - $\log P_{\text{app}} = 0.56 \log P_{\text{oct}} - 0.27(\text{pK}_a - \text{pH}) - 2.34 \text{SA}_{\text{HA}} - 5.14$,
 $n = 22$, $s = 0.28$, $R^2 = 0.8$
 - Volsurf PLS analysis:
 - 4 components, $n = 22$, $s_{\text{calc}} = 0.26$, $R^2 = 0.87$, $s_{\text{pred}} = 0.43$, $Q^2 = 0.65$.
 - Hydrophobic regions and size / shape increase P_{app} .
 - Hydrophilic regions and hydrogen bonding decrease P_{app} .

LogP 2000 Conference proceedings

- Work by Kansy *et al.* has highlighted role of $\log D_{7.4}$ in PAMPA permeation.
- Builds on work by Hansch and Kubinyi on bilinear correlations with $\log P$.



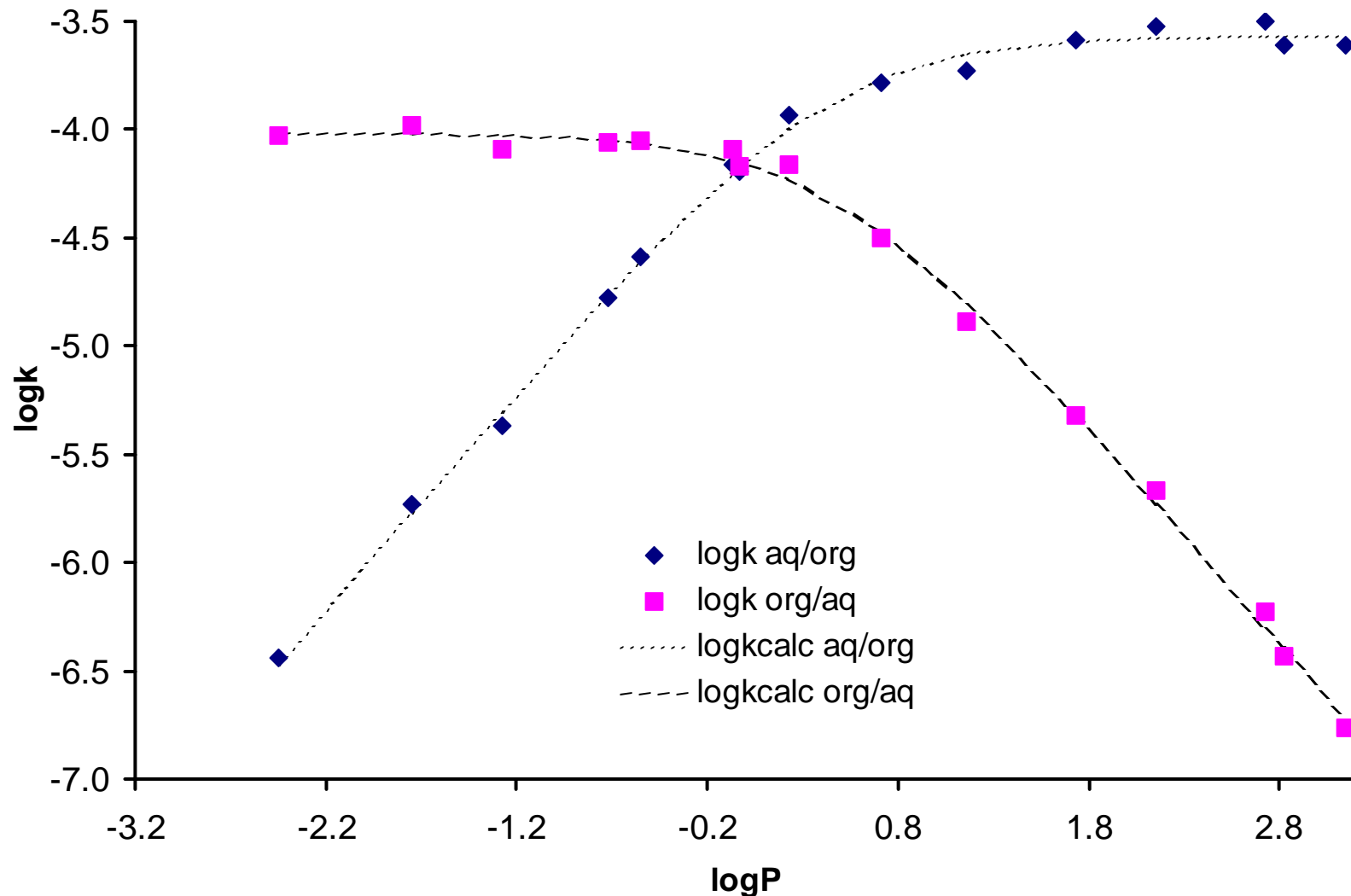
- Transport is balance of partitioning in and out of membranes.
- %Flux does not account for %Mem or plastic binding.
- Optimum $\log D_{7.4}$ of 1.7 (100% Flux).

Literature precedent

- Early work by Hansch, followed up by Kubinyi and van de Waterbeemd.
- Numerous *in-vivo* and *in-vitro* examples.
- Kinetic transport rates across the aq:org and org:aq interfaces are different and depend upon logP.
- Changing logP results in changing RLS and bilinear correlation.
- From diverse chemical series:

$$\log k_{aq/org}^{obs} = \log P - \log \left(\frac{k_{org}}{k_{aq}} P + 1 \right) + \log k_{org}$$

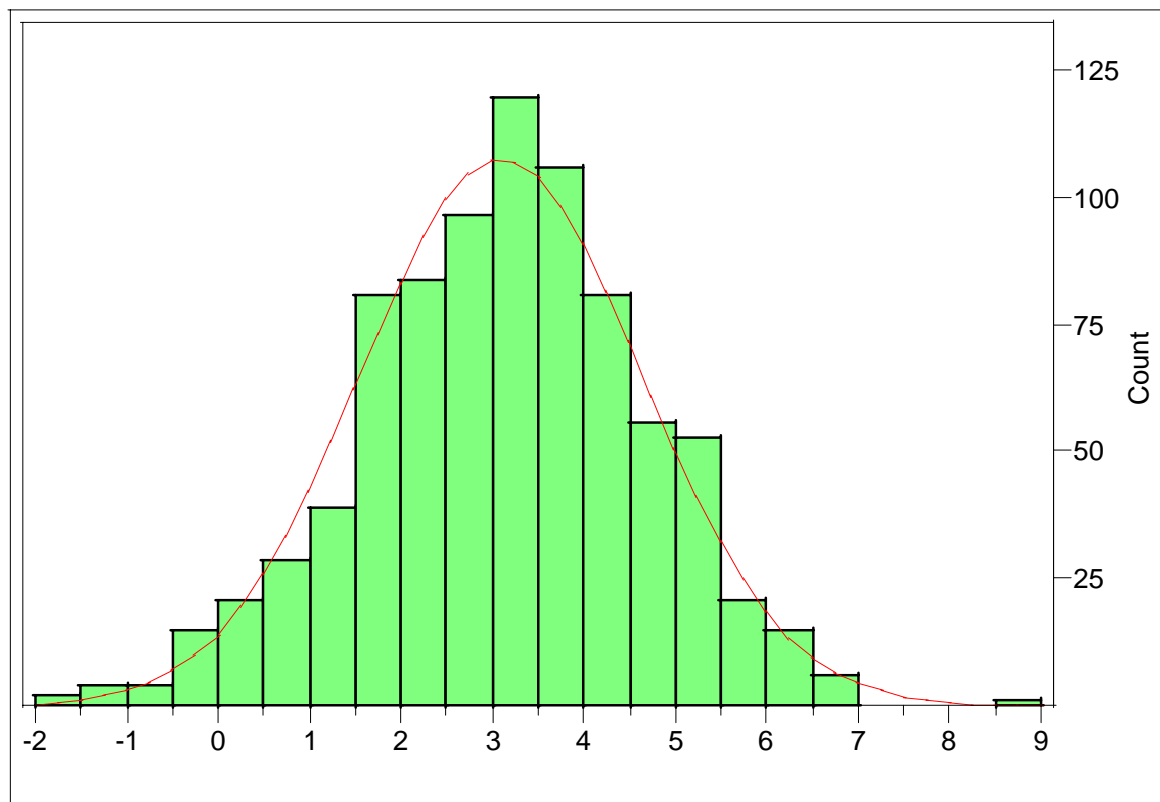
Change in rate with logP



J. Pharm. Sci., 1981, 70, 9, 1081

AZ screening

Distribution of clogP



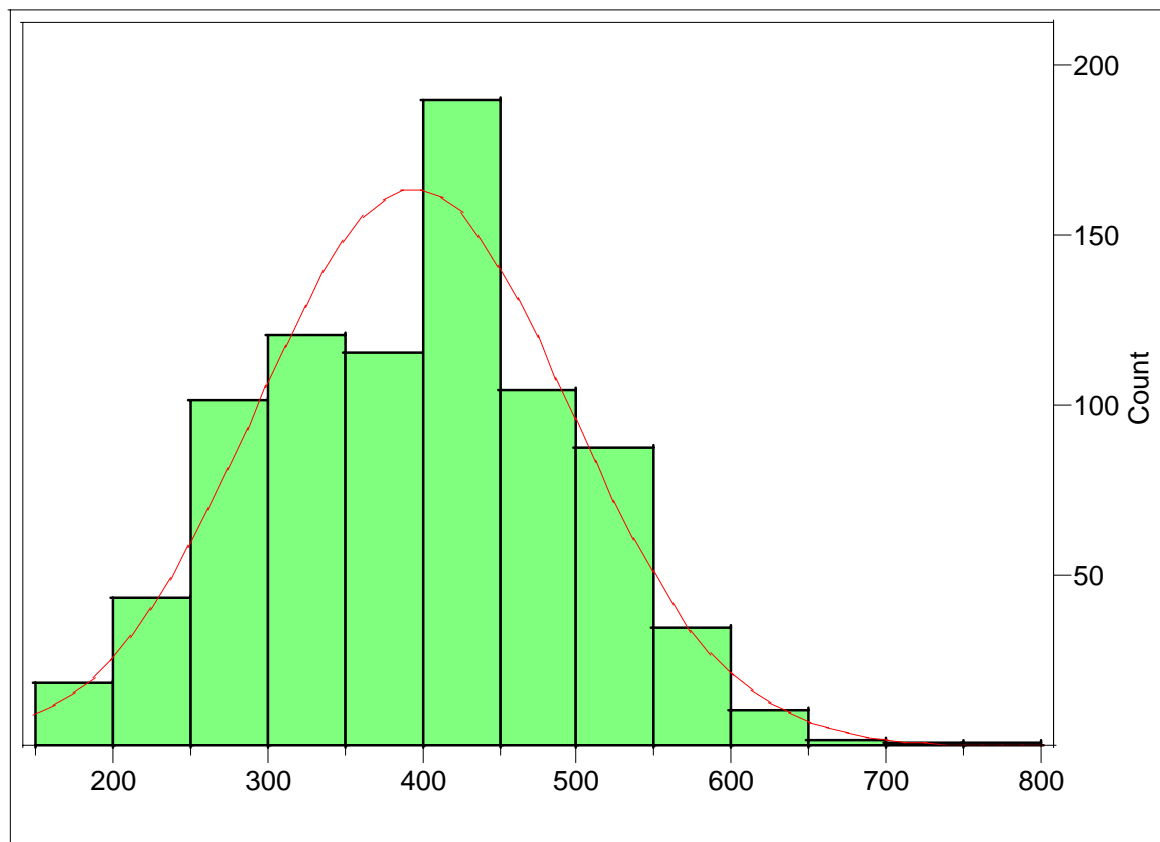
Quantiles

97.5%		6.093
90.0%		5.100
75.0%	quartile	4.136
50.0%	median	3.156
25.0%	quartile	2.092
10.0%		1.099
2.5%		-0.093

Moments

Mean	3.114012
Std Dev	1.546431
Std Err Mean	0.0535165

Distribution of MW



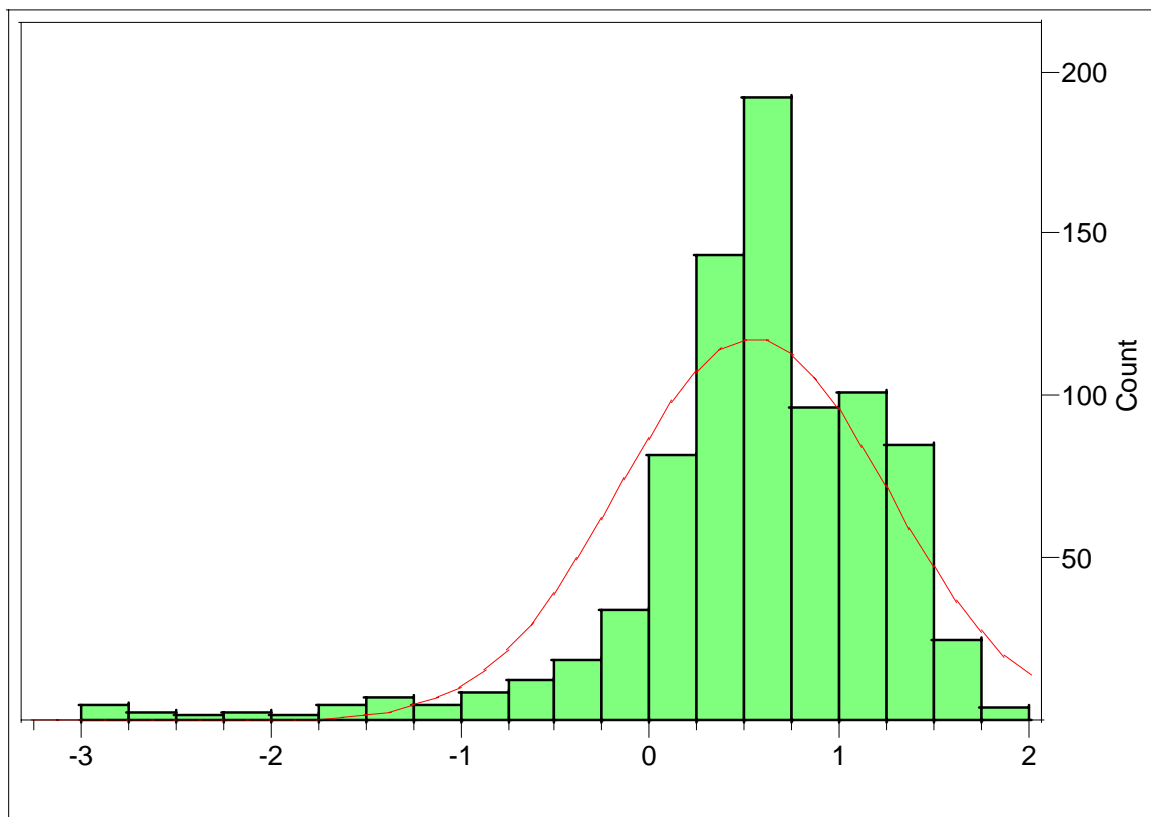
Quantiles

97.5%		590.99
90.0%		526.96
75.0%	quartile	466.86
50.0%	median	403.86
25.0%	quartile	320.39
10.0%		262.97
2.5%		202.12

Moments

Mean	395.50718
Std Dev	101.59578
Std Err Mean	3.5158668

Distribution of PAMPA $\log P_{app}$



Quantiles

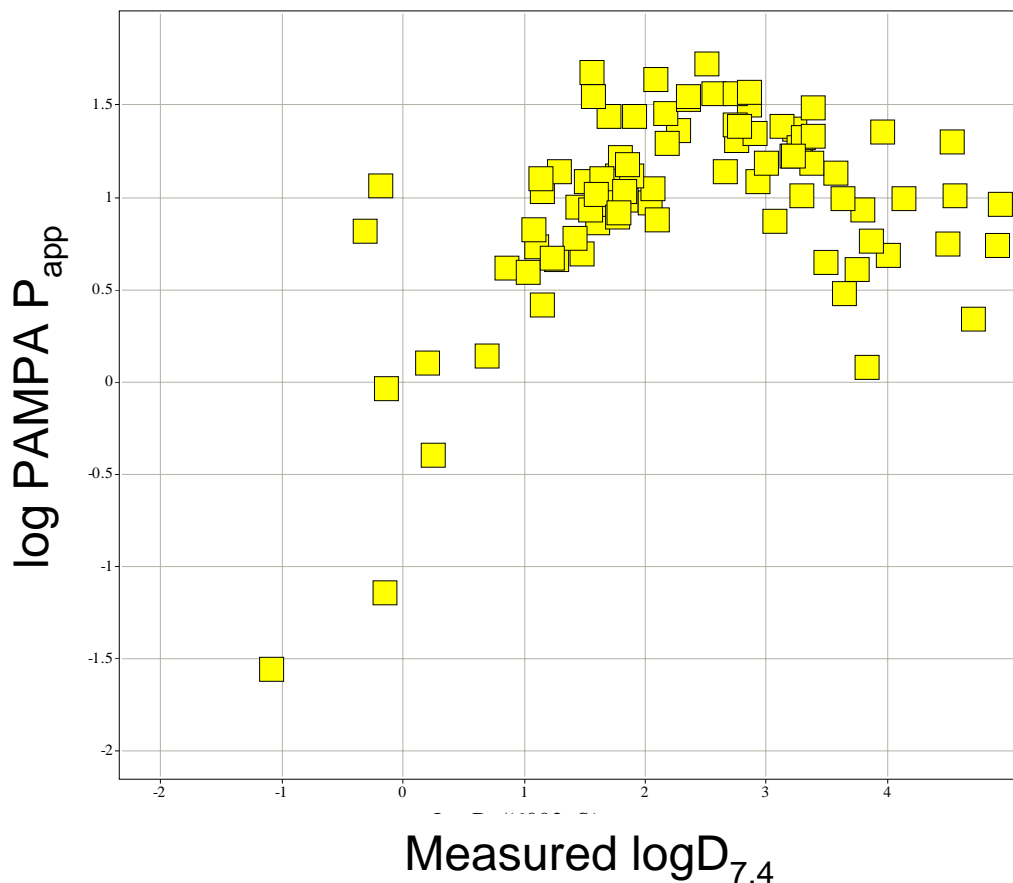
97.5%		1.544
90.0%		1.329
75.0%	quartile	1.019
50.0%	median	0.603
25.0%	quartile	0.309
10.0%		-0.122
2.5%		-1.371

Moments

Mean	0.551985
Std Dev	0.7086258
Std Err Mean	0.0244937

Correlation with $\log D_{7.4}$

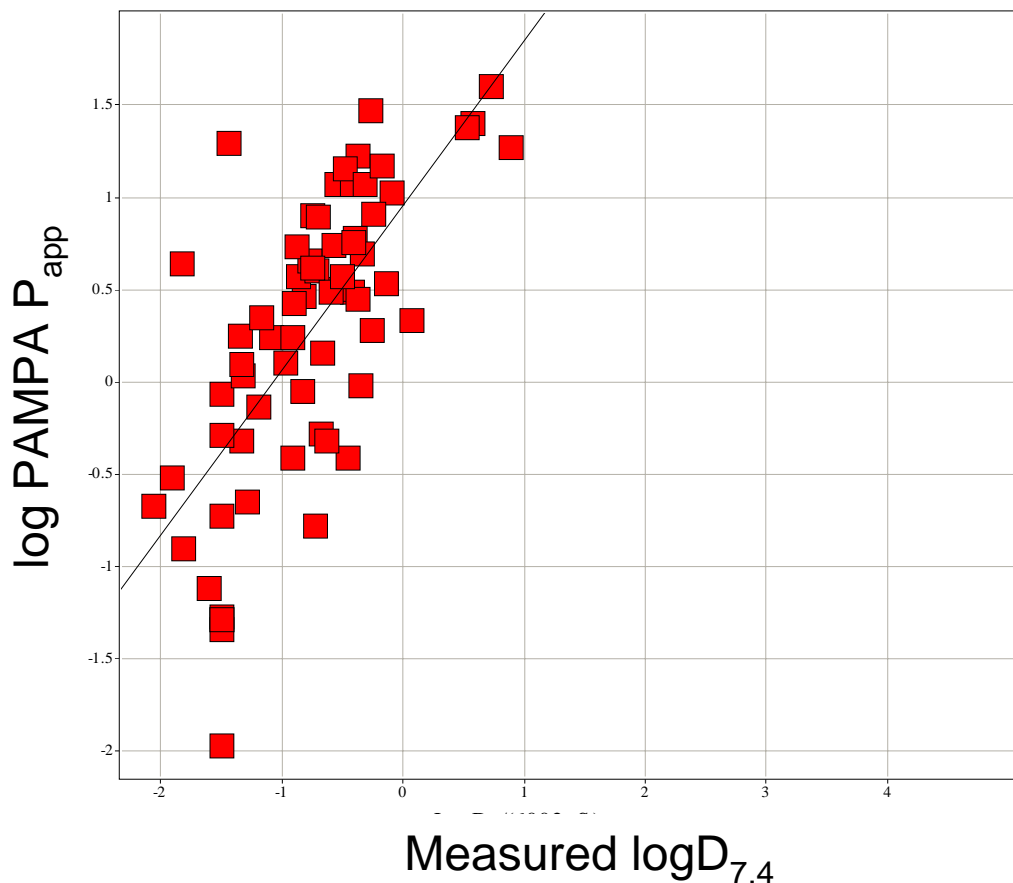
- PAMPA permeability exhibits a non-linear dependence upon measured $\log D_{7.4}$.



- For neutrals ■.
- Significant parabolic relationship.
- $\log P_{app} = 0.92 + 0.14^* \log D - 0.16^*(\log D - 2.32)^2$
- $R^2 = 0.57$, $RMSE = 0.35$, $F = 58$, $p < 0.0001$

Correlation with $\log D_{7.4}$

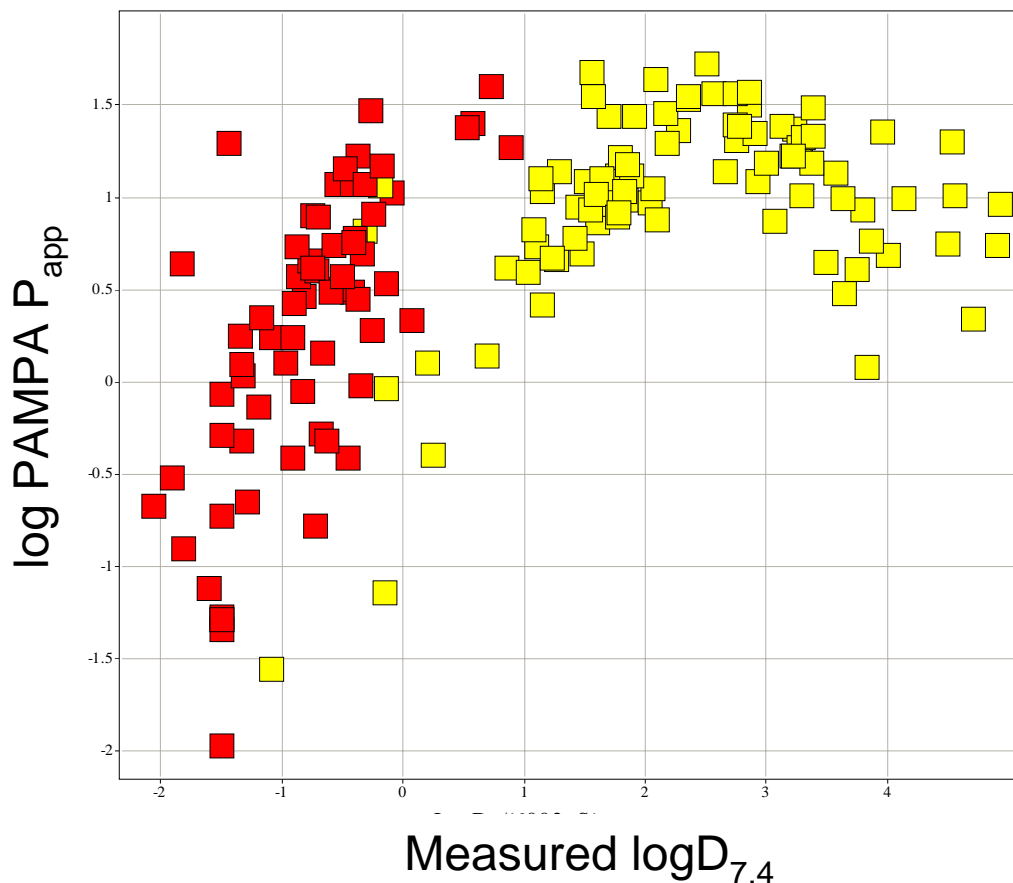
- PAMPA permeability exhibits a non-linear dependence upon measured $\log D_{7.4}$.



- For acids ■.
- Significant linear relationship.
- $\log P_{app} = 0.96 + 0.89 * \log D$
- $R^2 = 0.50$, $RMSE = 0.57$, $F = 62$, $p < 0.0001$

Correlation with $\log D_{7.4}$

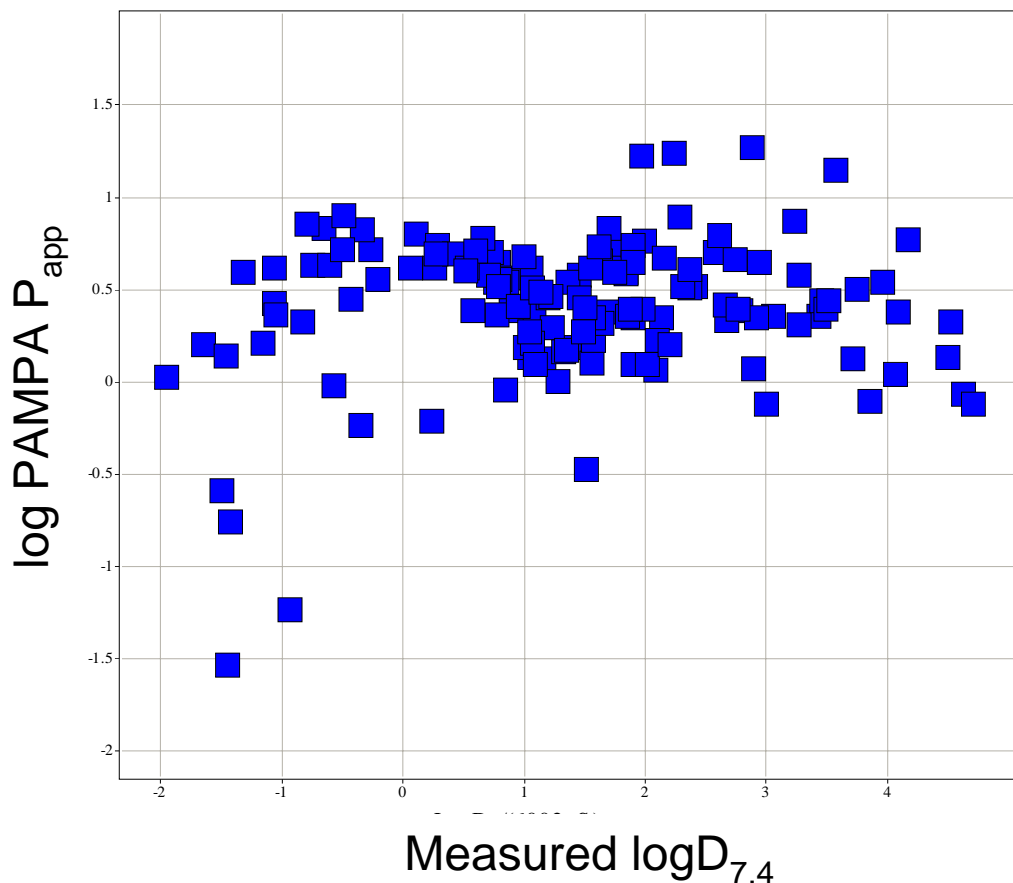
- PAMPA permeability exhibits a non-linear dependence upon measured $\log D_{7.4}$.



- For acids ■ and neutrals ■.
- Significant parabolic relationship.
- $\log P_{app} = 0.73 + 0.29^* \log D - 1.01^*(\log D - 1.01)^2$
- $R^2 = 0.55$, $RMSE = 0.50$, $F = 95$, $p < 0.0001$

Correlation with $\log D_{7.4}$

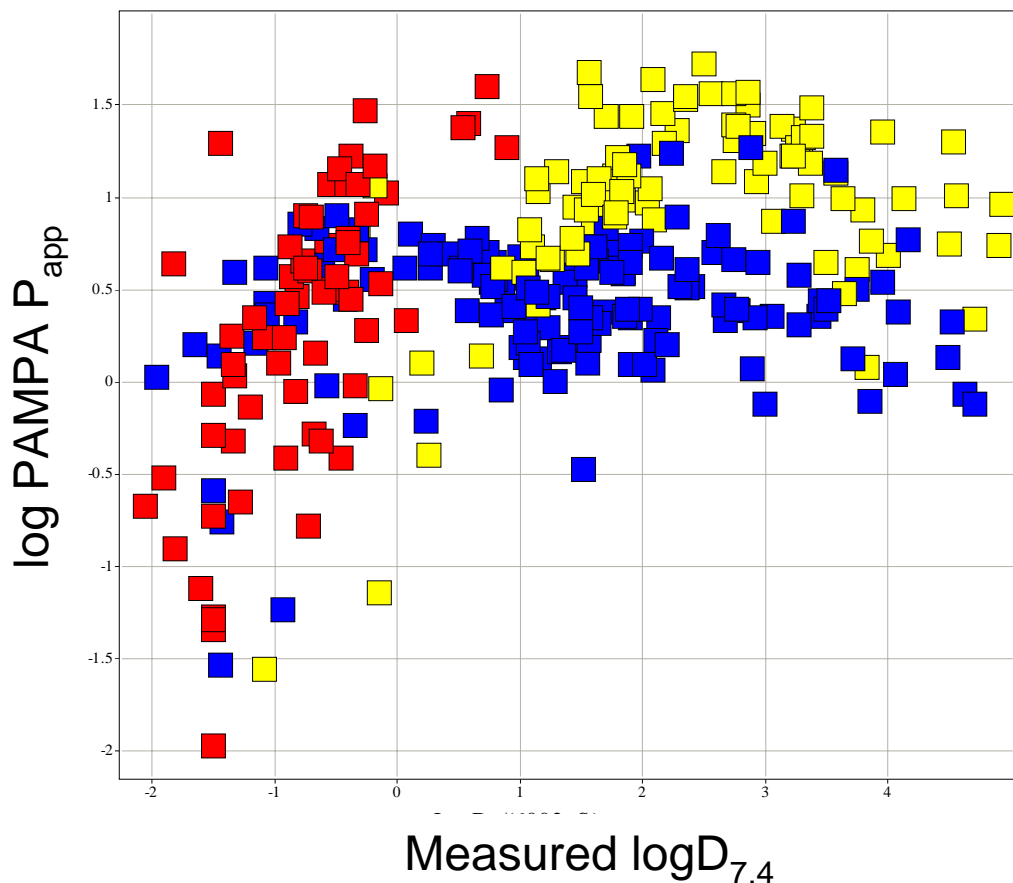
- PAMPA permeability exhibits a non-linear dependence upon measured $\log D_{7.4}$.



- For bases ■.
- No significant linear, poor polynomial fit.
- $R^2 = 0.12$, $RMSE = 0.37$, $F = 9$, $p = 0.0001$
- Sub series dependent relationship.

Correlation with $\log D_{7.4}$

- PAMPA permeability exhibits a non-linear dependence upon measured $\log D_{7.4}$.

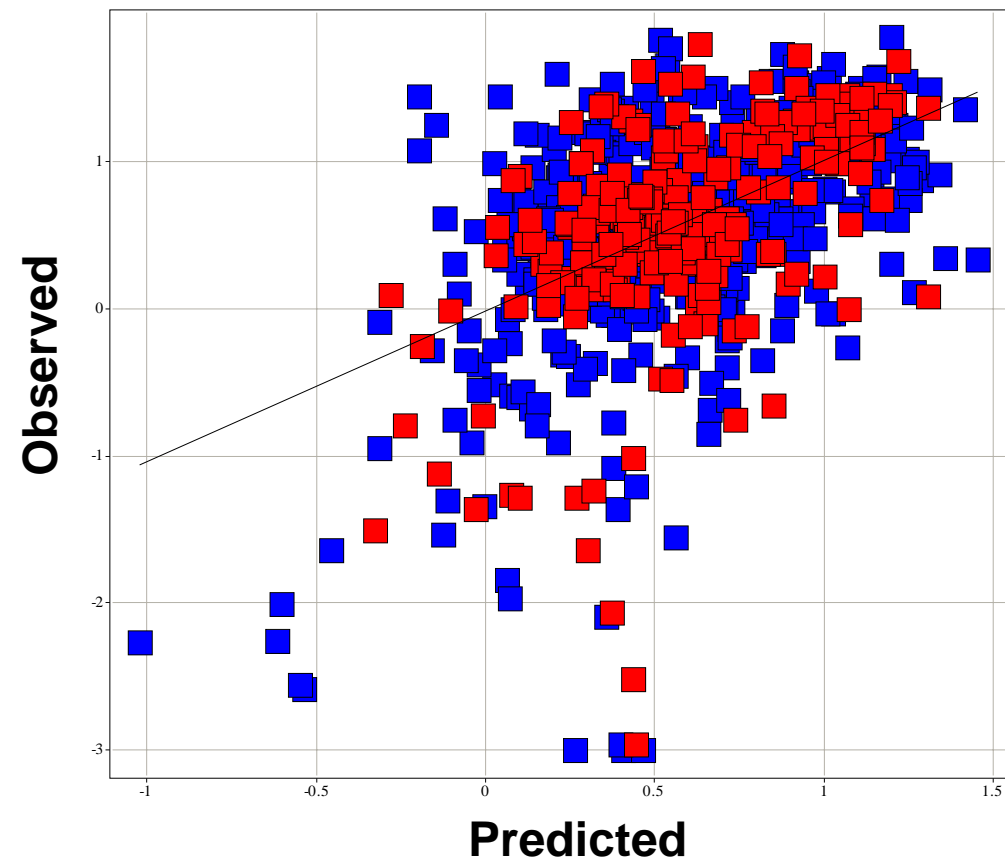


- For acids ■, bases ■, and neutrals ■.
- Zwitterions excluded.
- Significant parabolic relationship.
- $\log P_{app} = 0.54 + 0.17 * \log D - 0.07 * (\log D - 1.20)^2$
- $R^2 = 0.30$, RMSE = 0.52, F = 62, p < 0.0001

Methodology

- QSAR models built from single dataset containing chemistry from 3 sites (n = 854).
- Calculated Drone MFF, Selma and Volsurf descriptors, combined with ACD lipophilicity descriptors.
- Removed compounds where ACD failed (-17).
- Dataset of 837 results split randomly into Training set (70 %) and Test set (30 %).
- Sub-set with measured $\log D_{7.4}$ instead of ACD.
- Dataset of 331 results using original global split gives Training (76 %) and Test (24 %).
- Compared models by means of RMSE.

PLS PAMPA Models: Drone

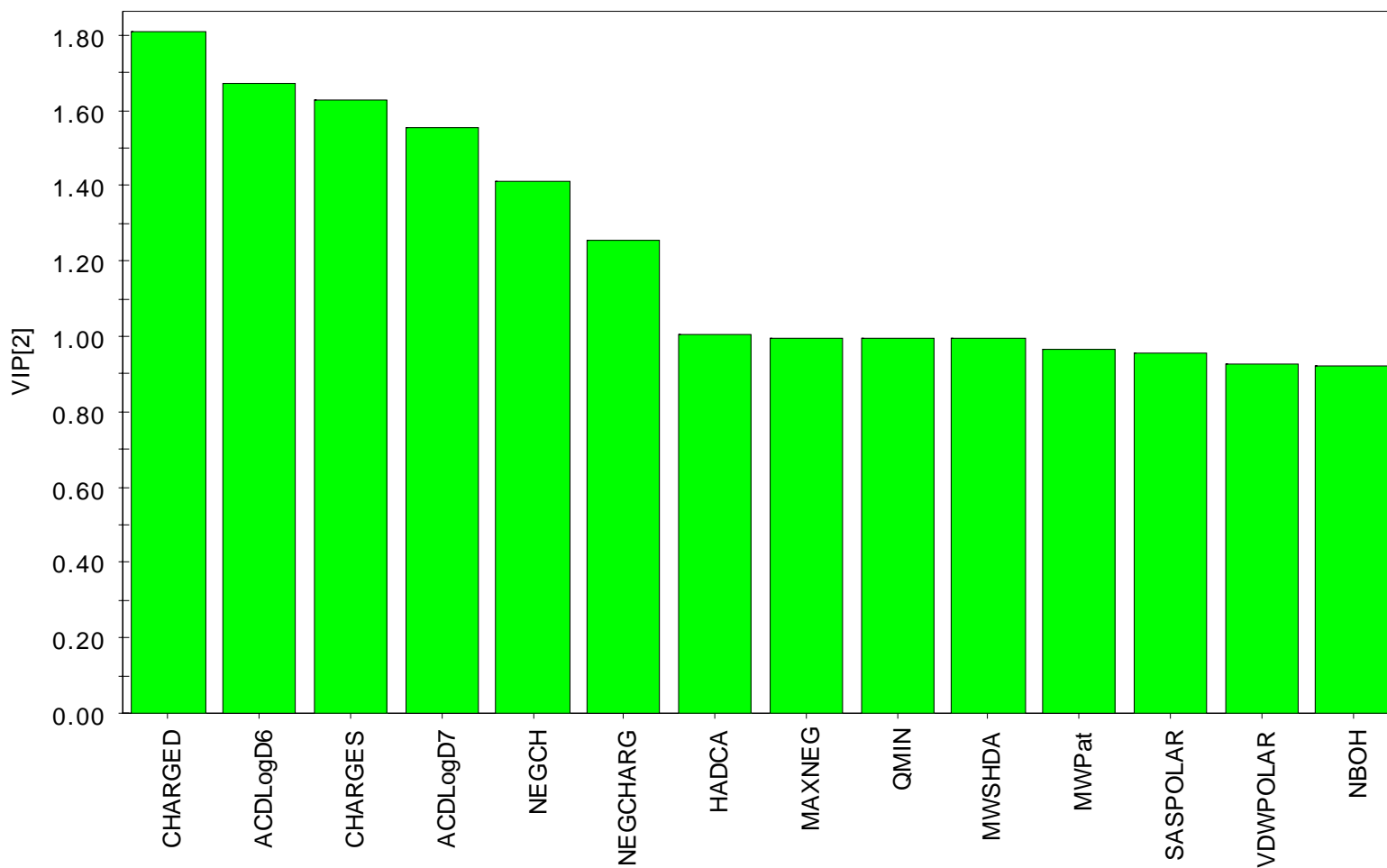


- Two component model.
- Further components not statistically significant.
- Validates as significant but weakly predictive.
- Especially poor in low P_{app} region.
- Would not fit statistically significant components with Selma or Volsurf descriptors.

Train ■ (n=590) RMSE = 0.62 $R^2 = 0.25$, $Q^2 = 0.23$

Test ■ (n=247) RMSE = 0.61

Drone PLS: VIP plot

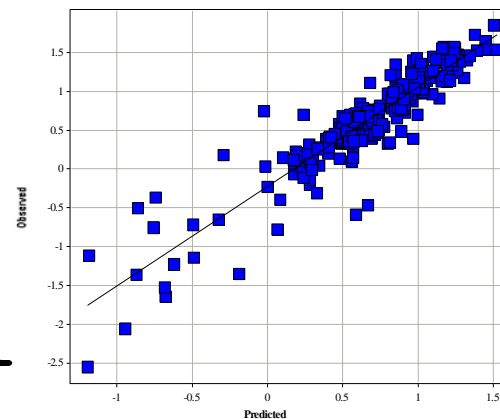
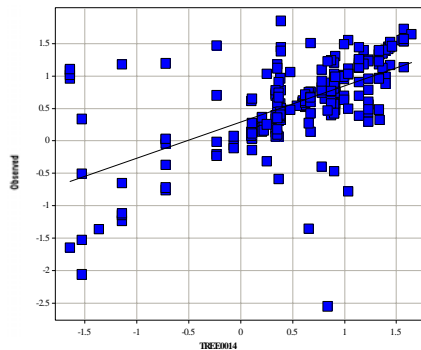
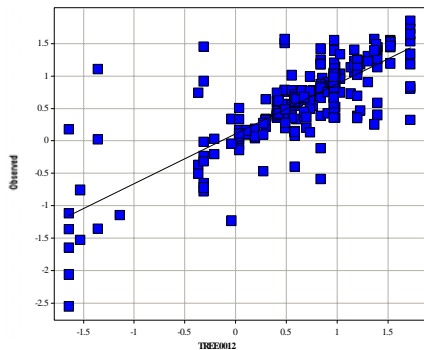
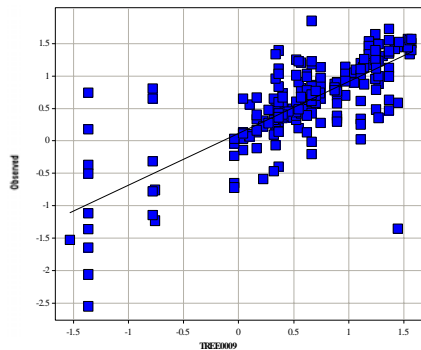
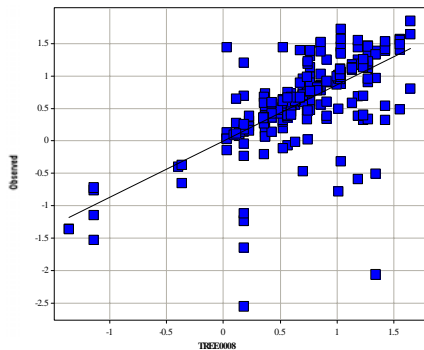
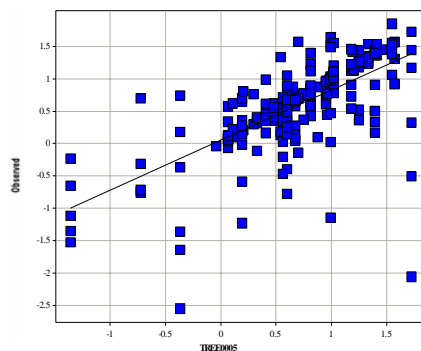
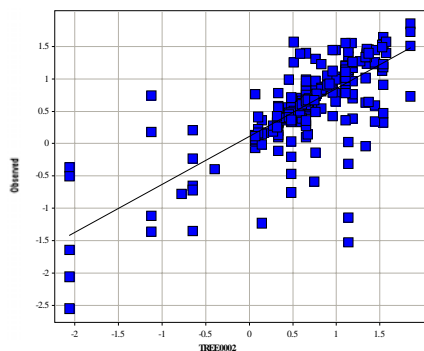


Simca-P 8.0 by Umetrics AB 2003-08-08 10:00

CART

- Non-linear relationships difficult to model in a linear way with MLR / PLS.
- Could be modelled by “linearising” logD *via* use of cross terms.
- Another option is to use CART: Classification and Regression Trees (Binary recursive partitioning).
- “Parent” groups of compounds are split into “child” groups with descriptor questions that have Yes or No answers.
- Multiple questions eventually lead to discrete bins of compounds, can be a regression.

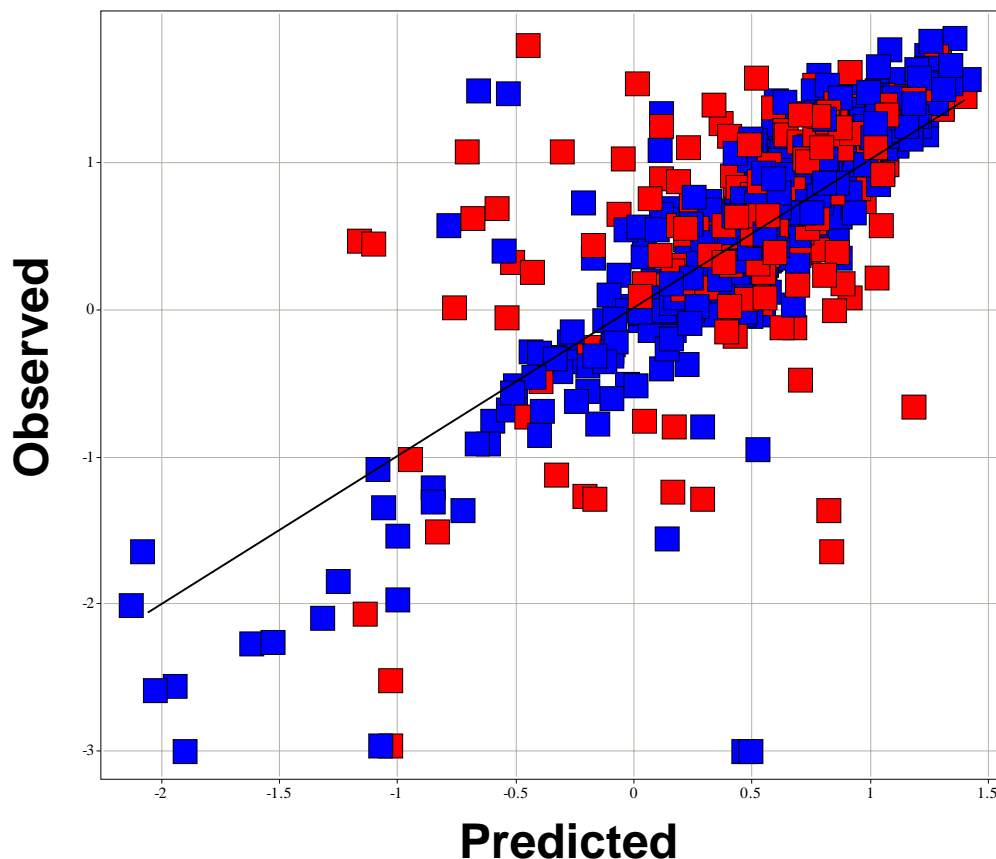
Consensus approach



- Consensus of 15 individual regression trees is used for final model.
- This results in separate bins forming continuous data.

CART PAMPA Models: Drone

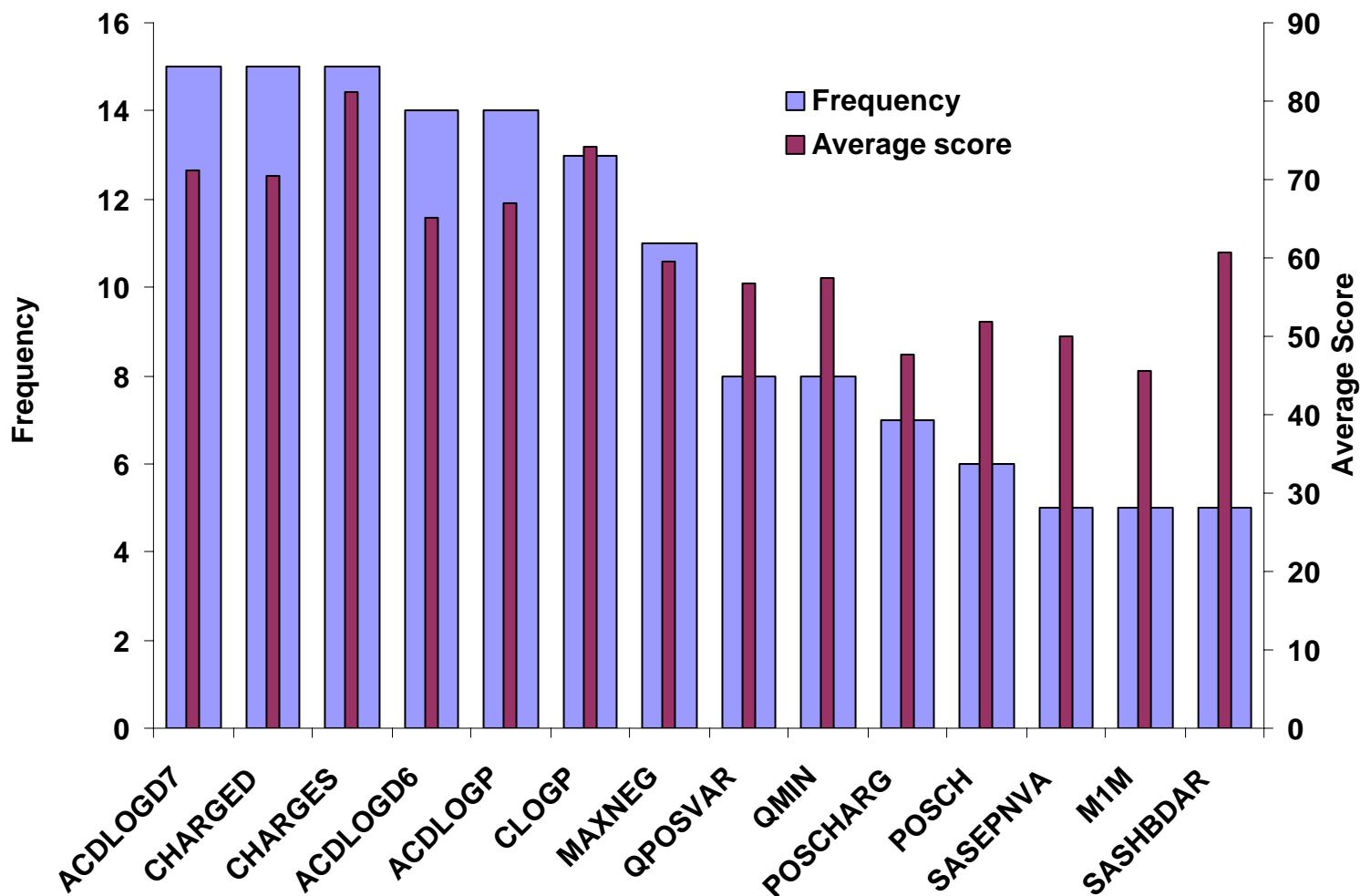
- 97 % within 1 log units.
- 86 % within 0.5 log units.



Train ■ (n=590) RMSE = 0.39 $R^2 = 0.73$

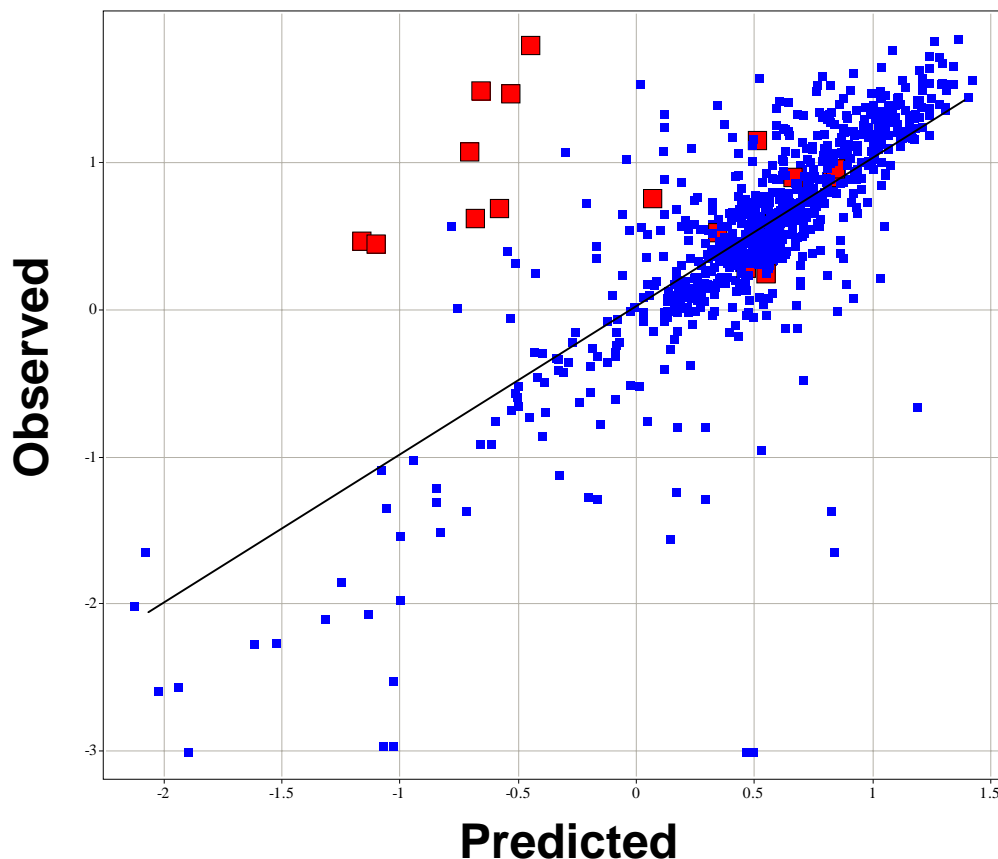
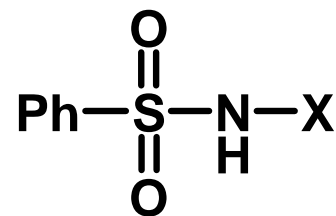
Test ■ (n=247) RMSE = 0.59

Important Drone variables



CART PAMPA Models: Drone

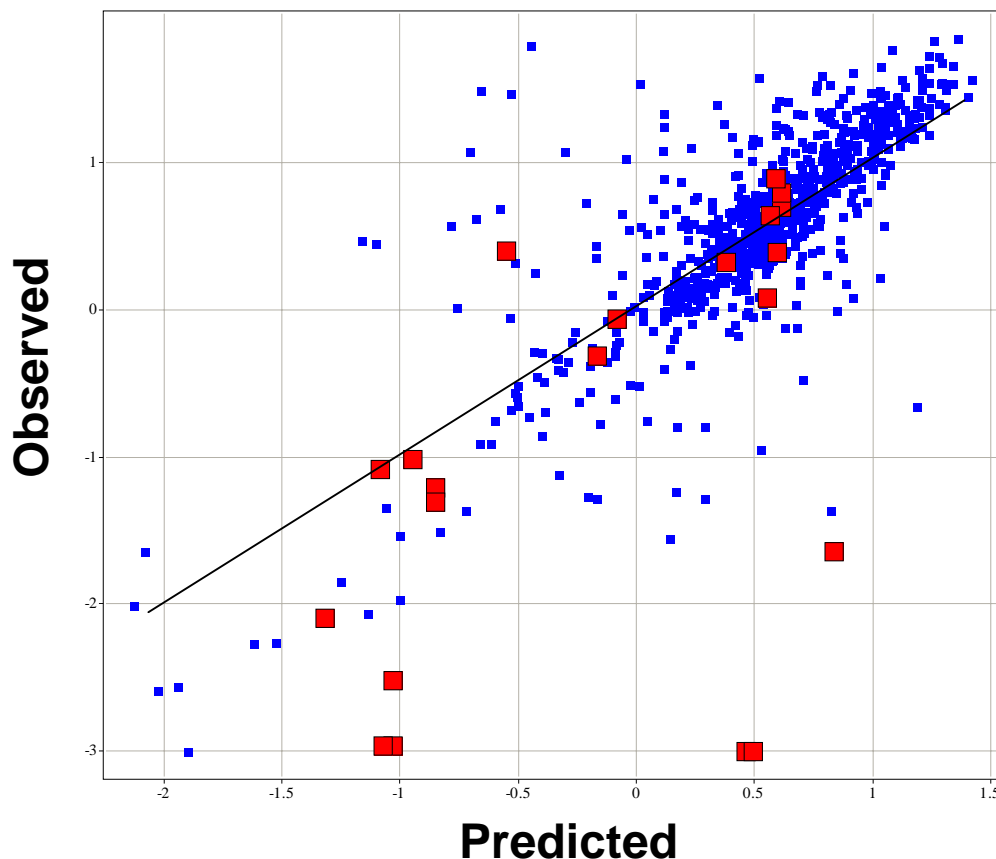
- 97 % within 1 log units.
- 86 % within 0.5 log units.
- Sulfonamides are underpredicted.



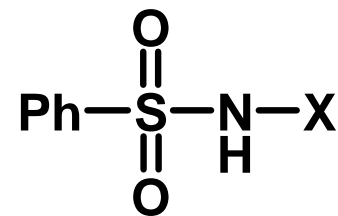
Train (n=590) RMSE = 0.39 $R^2 = 0.73$

Test (n=247) RMSE = 0.59

CART PAMPA Models: Drone



- 97 % within 1 log units.
- 86 % within 0.5 log units.
- Sulfonamides are underpredicted.



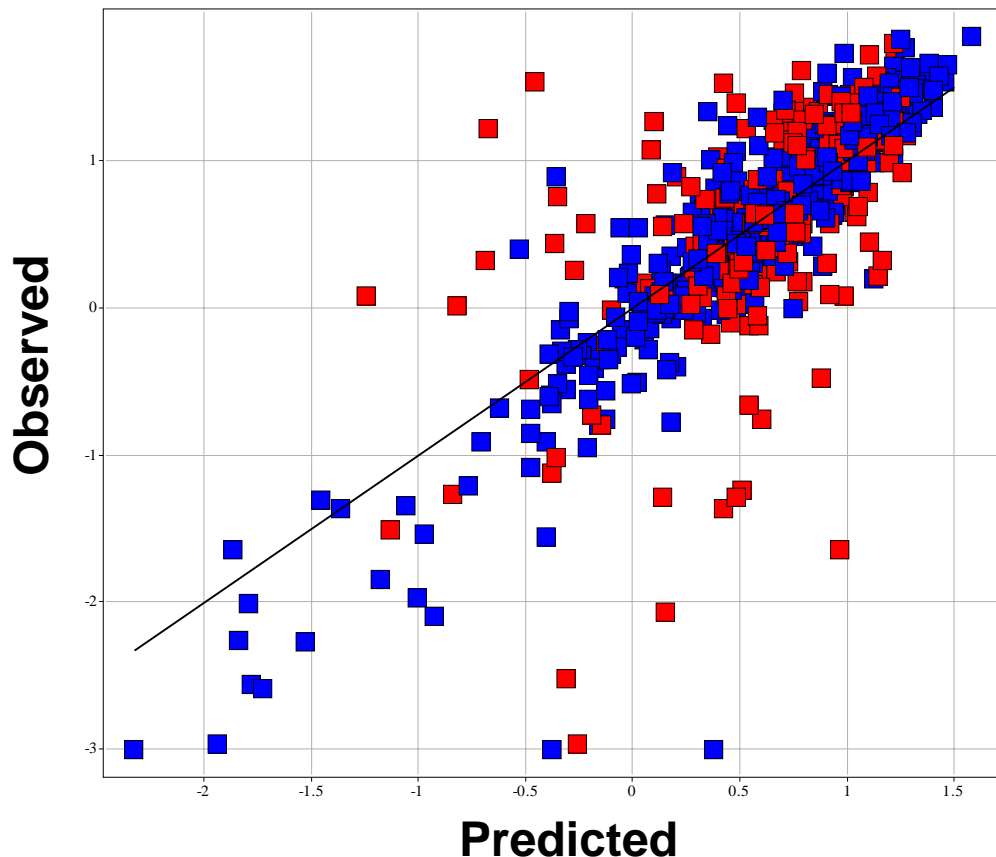
- Compounds where ACD misses the ionisation and over predicts logD are problematic.

Train (n=590) RMSE = 0.39 $R^2 = 0.73$

Test (n=247) RMSE = 0.59

CART PAMPA Models: Selma

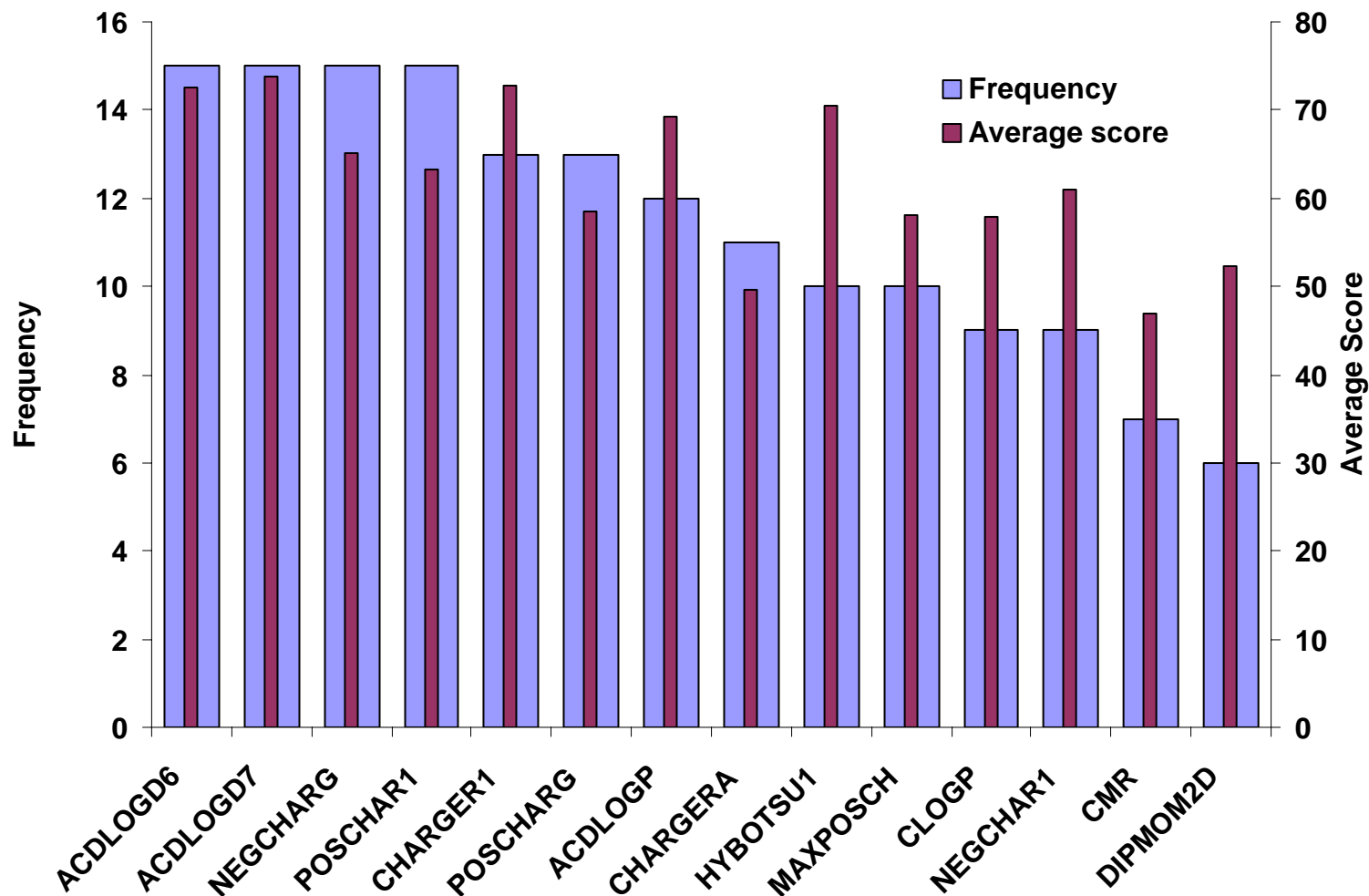
- 97 % within 1 log units.
- 86 % within 0.5 log units.



Train ■ (n=590) RMSE = 0.32 $R^2 = 0.83$

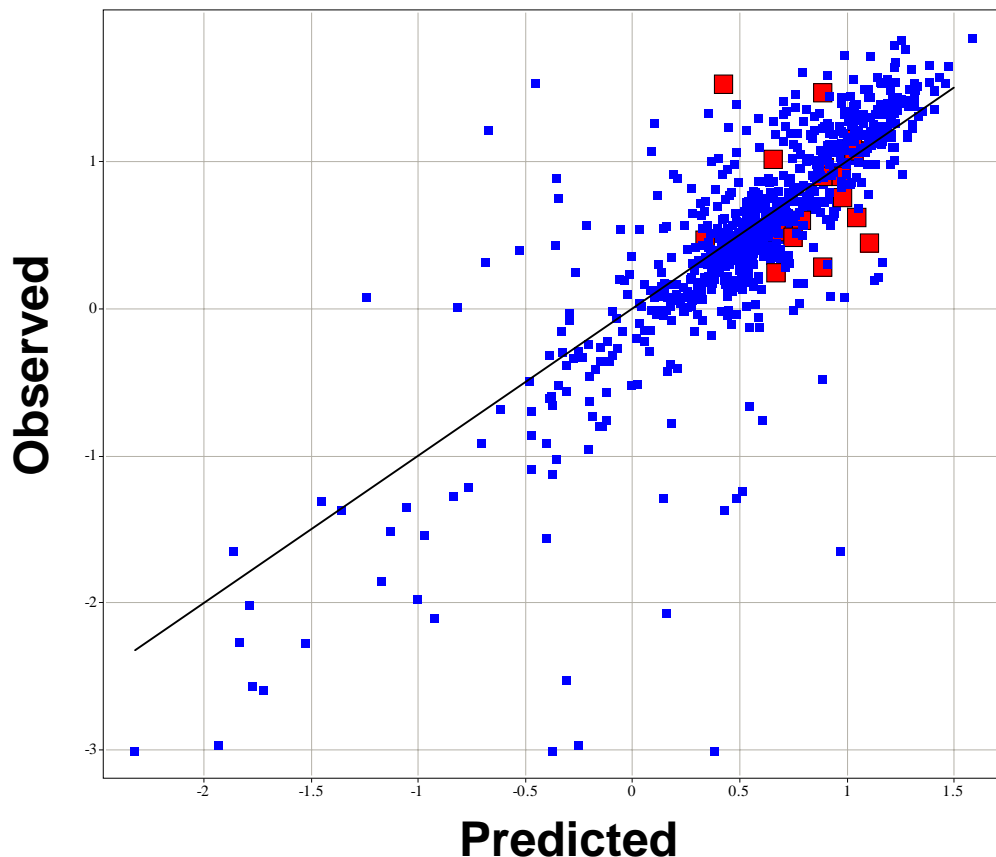
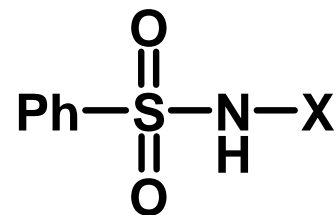
Test ■ (n=247) RMSE = 0.59

Important Selma variables



CART PAMPA Models: Selma

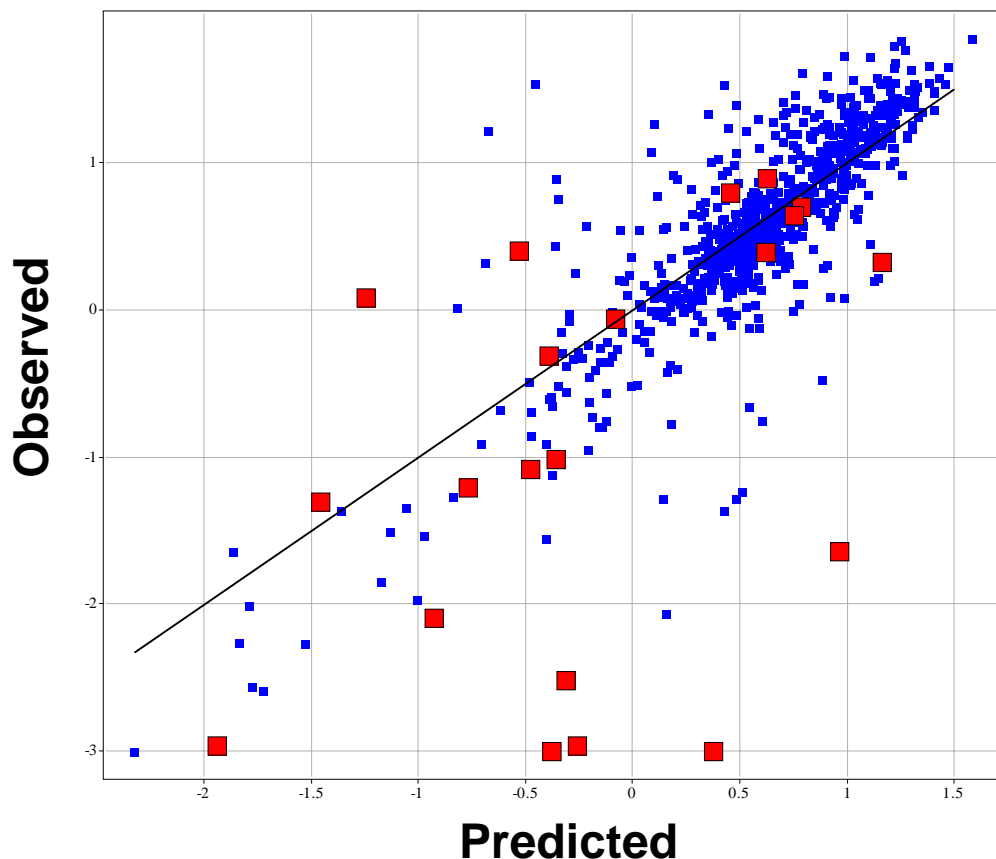
- 97 % within 1 log units.
- 88 % within 0.5 log units.
- Sulfonamides are handled better.



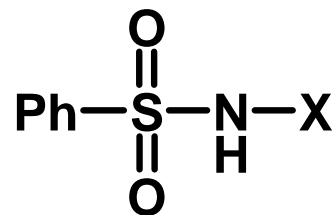
Train (n=590) RMSE = 0.32 $R^2 = 0.83$

Test (n=247) RMSE = 0.59

CART PAMPA Models: Selma



- 97 % within 1 log units.
- 88 % within 0.5 log units.
- Sulfonamides are handled better.



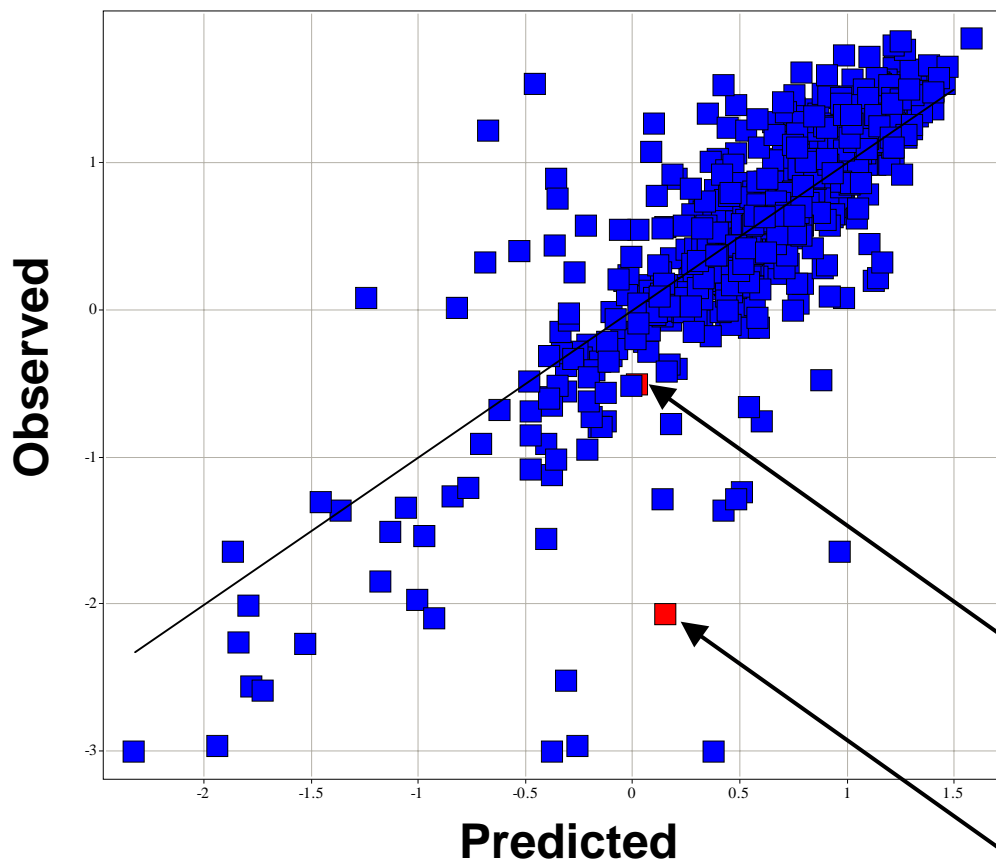
- Compounds where ACD misses the ionisation and over predicts logD are still problematic.

Train (n=590) RMSE = 0.32 $R^2 = 0.83$

Test (n=247) RMSE = 0.59

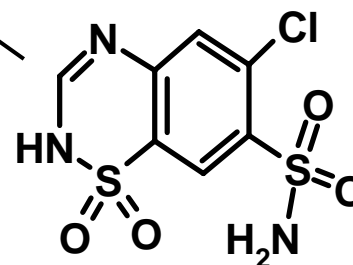
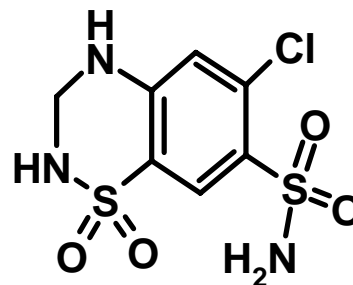
CART PAMPA Models: Selma

- Not all descriptor sets pick up structural subtleties.
- Hydrochlorothiazide and chlorothiazide are challenging pair for PAMPA QSAR.



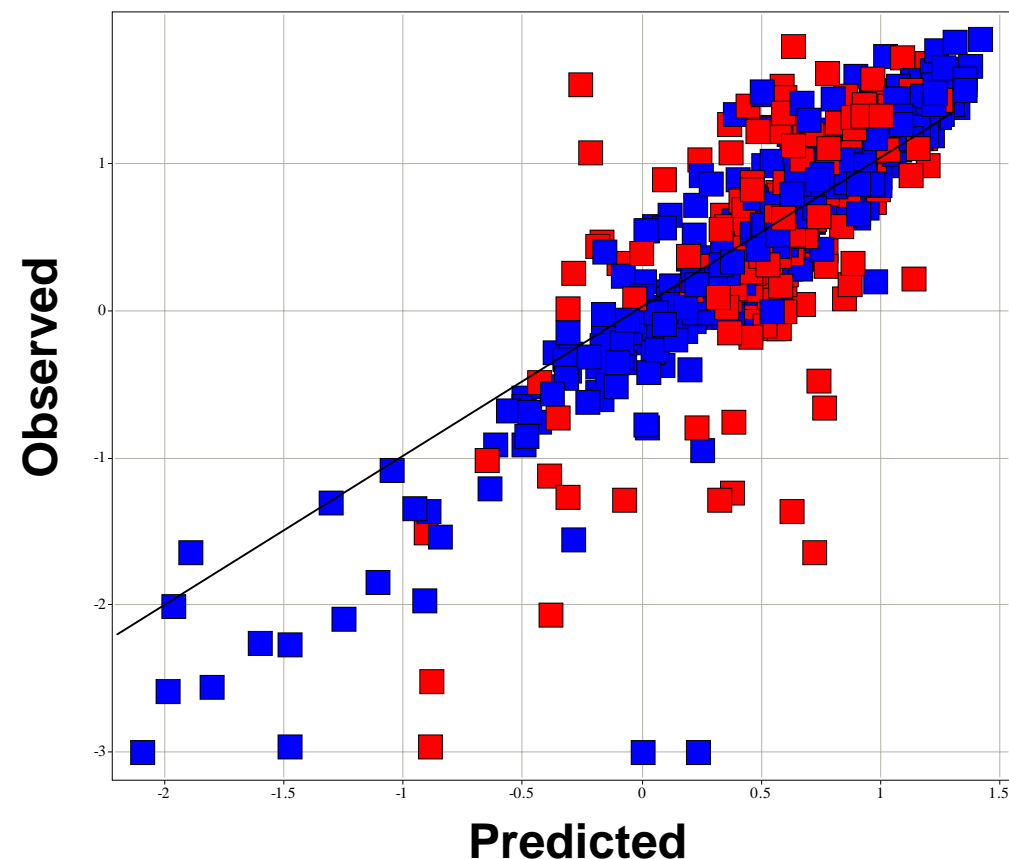
Train (n=590) RMSE = 0.32 $R^2 = 0.83$

Test (n=247) RMSE = 0.59



Two H atoms different, OK with Drone

CART PAMPA Models: Consensus

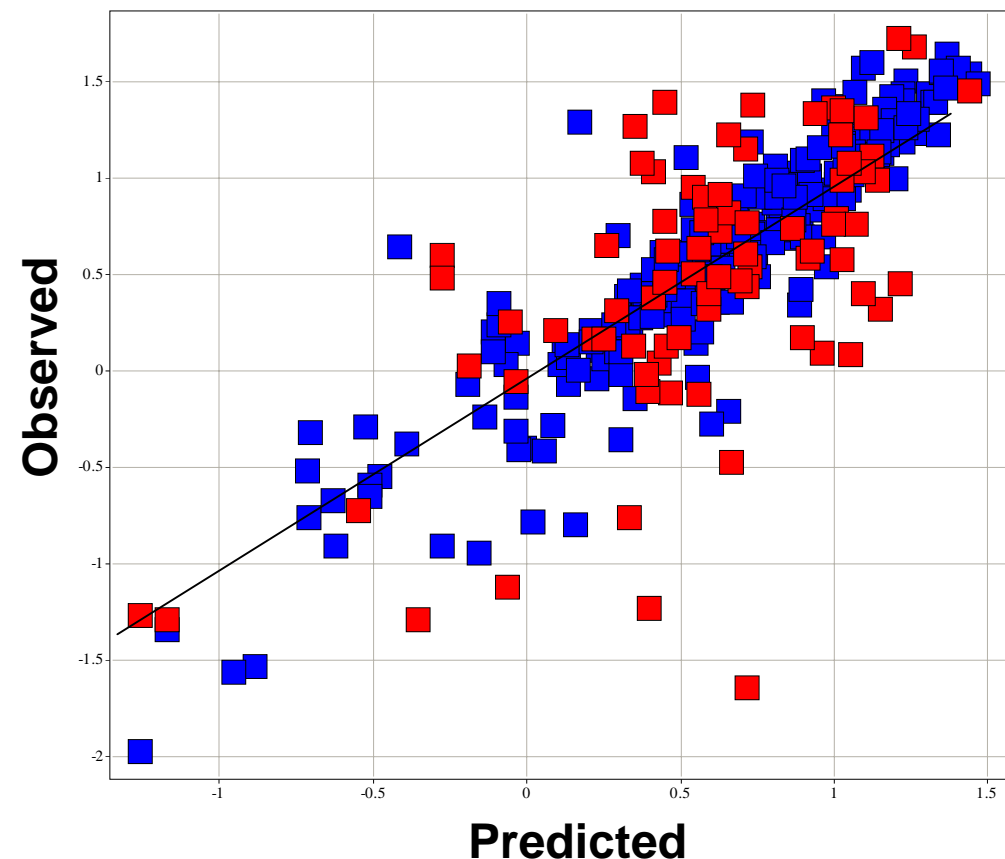


- Utilises average observed P_{app} from all descriptor sets.
- 98 % within 1 log units.
- 89 % within 0.5 log units.
- Gives better combination of Train / Test RMSE values than Volsurf alone.
- Major outliers in all models are reproduced.

Train ■ (n=590) RMSE = 0.32 $R^2 = 0.84$

Test ■ (n=247) RMSE = 0.52

CART PAMPA Models: mlogD

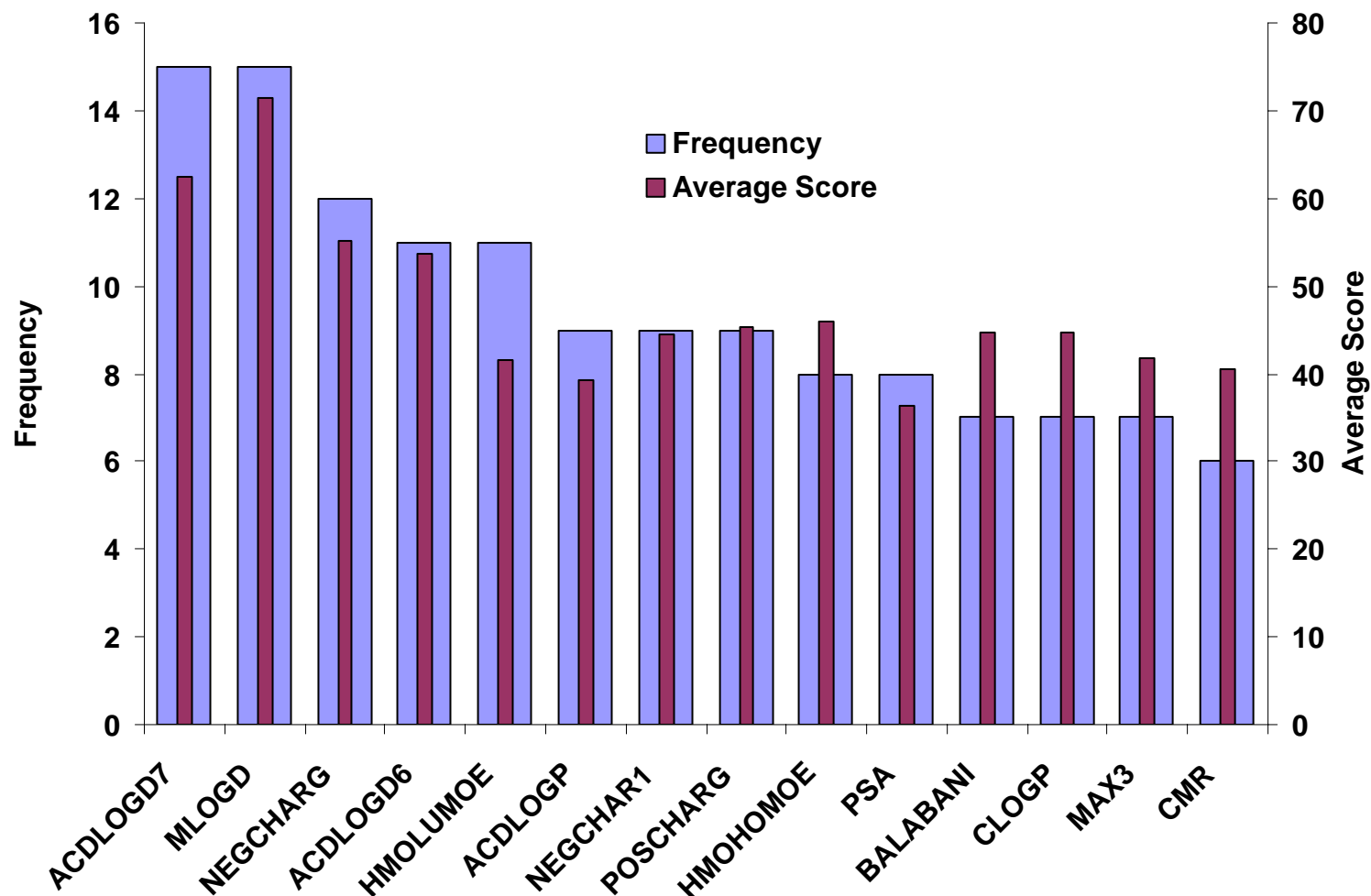


- Uses measured $\log D_{7.4}$ in conjunction with ACD.
- Built with Selma descriptor set.
- 98 % within 1 log units.
- 89 % within 0.5 log units.
- Range of structural classes within low region is limited by available $\log D$ values.

Train ■ (n=251) RMSE = 0.25 $R^2 = 0.84$

Test ■ (n=80) RMSE = 0.57

Important variables



Conclusion

- Vast majority of compounds are predicted well by CART models.
- Importance of lipophilicity in defining PAMPA QSAR model requires careful examination of compound series: pK_a assignment.
- Mixed descriptor sets and techniques: consensus models, may improve prediction quality.
- Low permeability area is still poorly trained and care should be taken with new chemical series.
- But, does the results distribution indicate that poor *passive* permeability isn't a common problem or is the dataset not representative?

Acknowledgements

POC Team, AstraZeneca R&D Charnwood

Nicola Colclough

Andy Davis

Pete Kenny

Vanessa Moore

Jeff Morris

Alan Wait