S-IN Soluzioni Informatiche
for Chemistry and Pharmaceutical Chemistry

Predizioni in silico: proprietà chimico-fisiche

Elena Fioravanzo: elena.fioravanzo@s-in.it

Outline

- Phys-chem properties
- Lipophilicity
  - What is lipophilicity?
  - LogP
  - *In silico* prediction methods
  - LogP vs LogD
  - Prediction methods
- Case Studies

Phys-chem properties

- pKa
- LogP/LogD
- Solubility
- Boiling point
- Vapour pressure
- Enthalpy of vaporization
- Adsorption coefficient
- Bioconcentration factor
- …….

What is "Lipophilicity"?

Log*P*, the partition coefficient, is a measure of the differential solubility of a compound in two immiscible solvents.

$$P = \frac{[\text{Compound}]_{\text{octanol}}}{[\text{Compound}]_{\text{water}}}$$

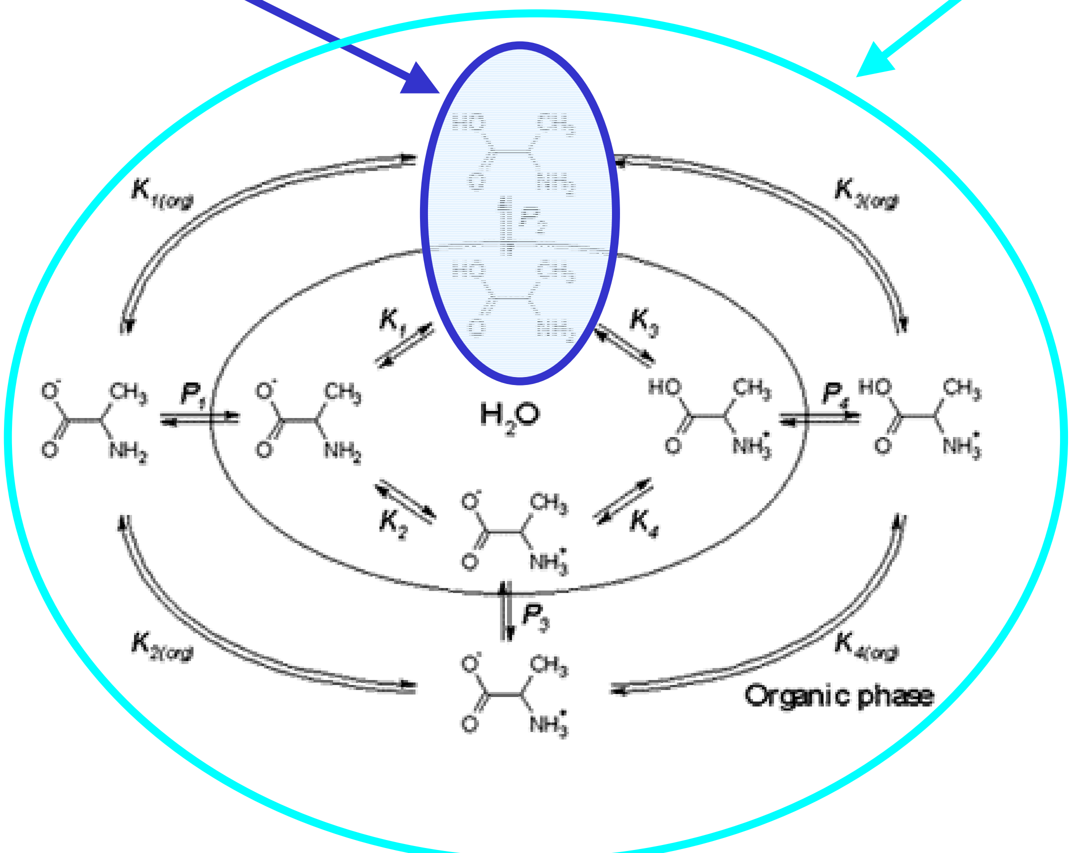
Log*P*

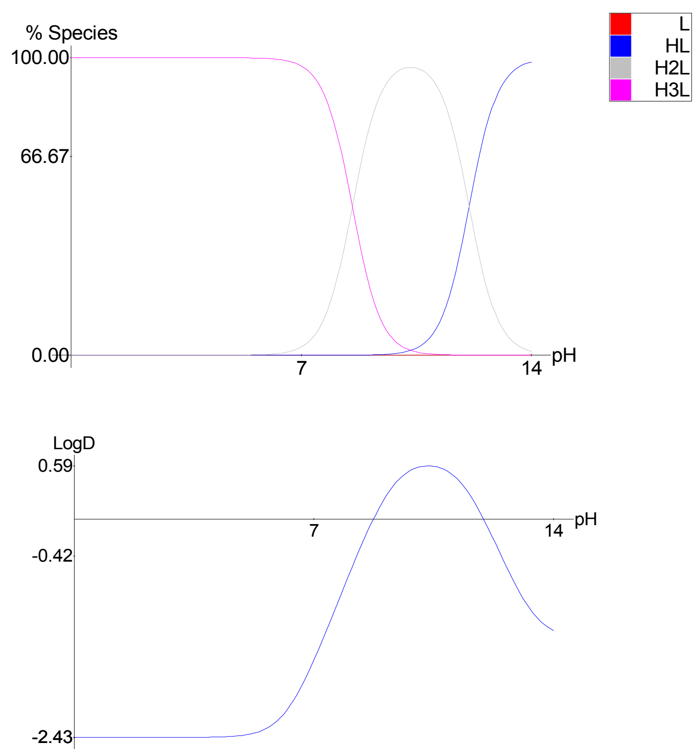
- Single Species
- Usually Refers to Neutral Form

Lipophilicity (log*P*/log*D*)

Log*P*-neutral compounds

LogD-ionizable compounds





PhysChem Properties

Measure or Predict ?

Measurement is almost always better

But…

- There are times when the cost savings (or time savings) of prediction over measurement justify the additional uncertainty
- There are times when measuring isn't feasible
  - Virtual compounds
  - No access to sample
  - Not enough sample
  - Sample is too impure

LogP Prediction Methods

- Fragment or Atom Based
  - Basic Principle: LogP is function of small contribution from fragment (or atom) and interaction
- Molecular descriptors and QSPR
  - Basic Principle: find correlation between selected molecular properties and experimental LogP

Fragment-based methods

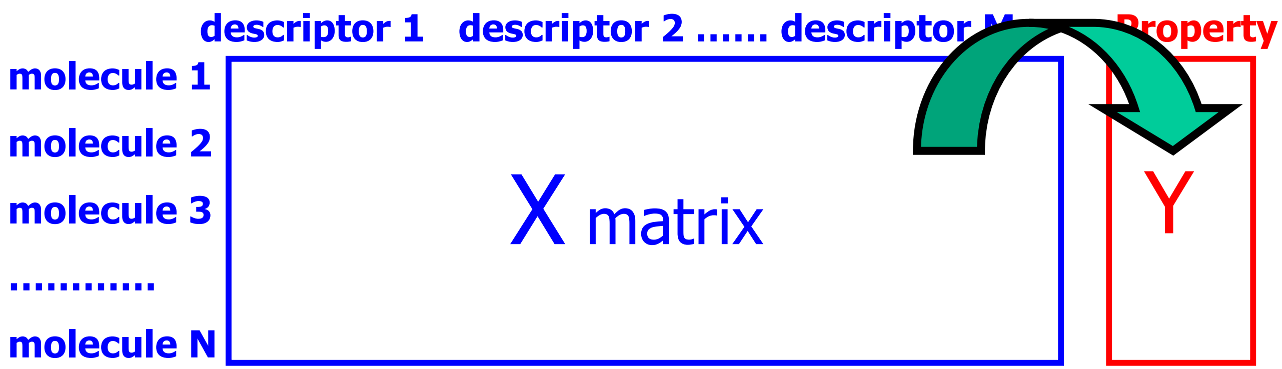
- **ClogP**
  - pioneered by Corwin Hansch and Al Leo
  - identify large fragments, whose contribution to logP value is known from their occurrence in other compounds with measured logP

Atom-based methods

- **AlogP, XlogP, SlogP**
  - pioneered by Gordon Crippen (Univ. Michigan)
  - based on identifying a series of "atom types" in the molecule

QSPR Methods

Quantitative **S**tructure **P**roperty **R**elationship



- Basic principle: Finding correlation between molecular descriptors (X) and molecular property (Y)

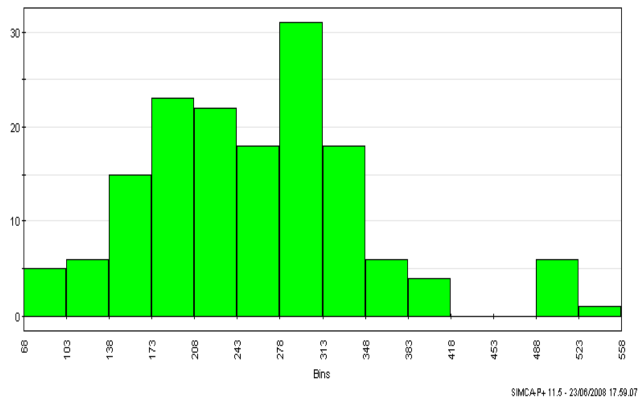
Advantages and Drawbacks

- **Fragment or Atom Based**
  - Advantages:
    - Highly interpretable results.
    - train calculation algorithm few experimental data
  - Drawbacks:
    - problems arise if test compound contains fragments that are "missing" from the training set
- **QSAR Based**
  - Advantages:
    - Less influenced by single fragment
  - Drawbacks:
    - Training is easy, but sometimes needs a lot of data…
    - Descriptors are often "chemically meaningless"; difficult interpretation of results and no suggestion about how to modify the property

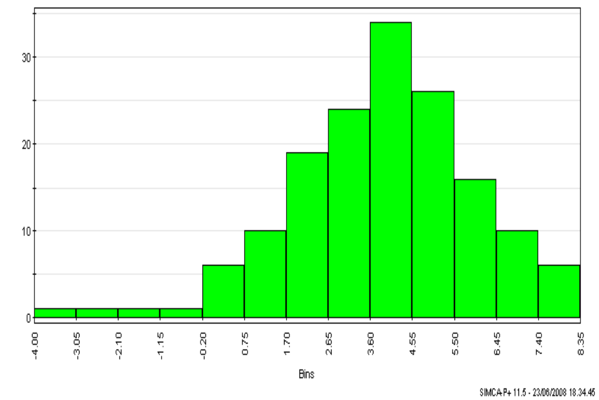
Case Studies

- Case Study 1: LogP
- Case Study 2: LogP
- Case Study 3: LogP
- Case Study 4: logD and pKa
- Case Study 5: Solubility

Case Study 1: LogP



MW distribution

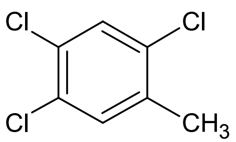


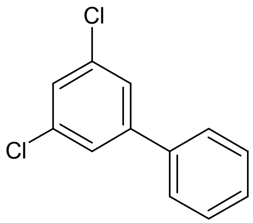
logP distribution

- 155 Molecules
- LogP experimental range from -3.99 to 8.24

Dataset Analysis - Substructures

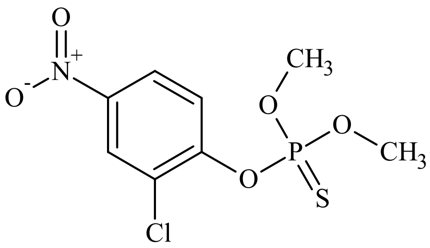
- Group 1 (~80 mol): phenyl or bi-phenyl substituted with alogens (approx 80 mol)





Representative structures of Group 1

- Group 2 (~ 70 mol): phenyl or etheroaromatics rings substituted with polar group (NO_2, PO_4, SO_3, etc)



Representative structures of Group 2

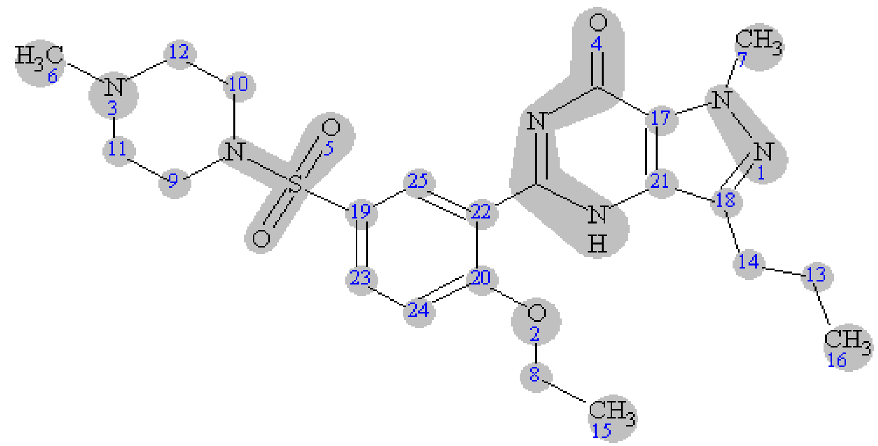
- Other Compounds: Silicates, rings with more than 8 atoms, 3-4 condensed rings...

LogP *in silico* prediction

- Fragment-based methods
  - ACDLabs/logP
  - PHA/logP
  - MLOGP
  - ALOGP
- QSPR methods
  - QikProp logPo/w
  - QikProp + PLS
  - DRAGON descriptors + PLS

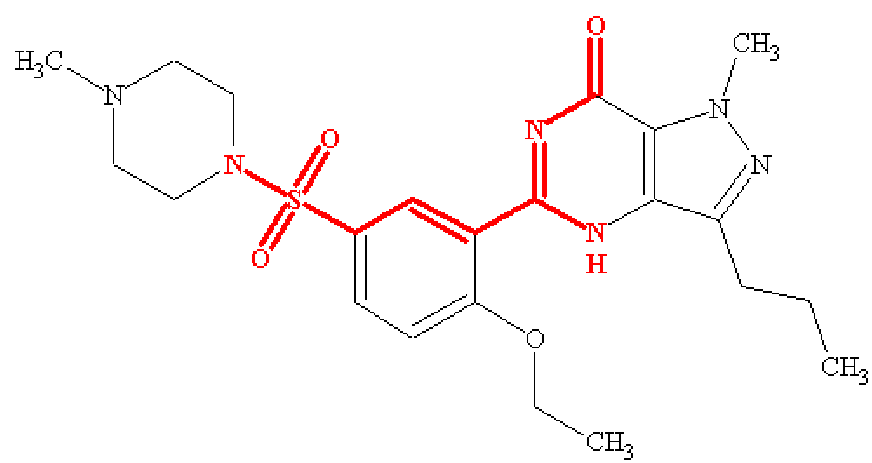
ACD/Labs LogP

- Decompose a molecule according to the principal of insulating carbons



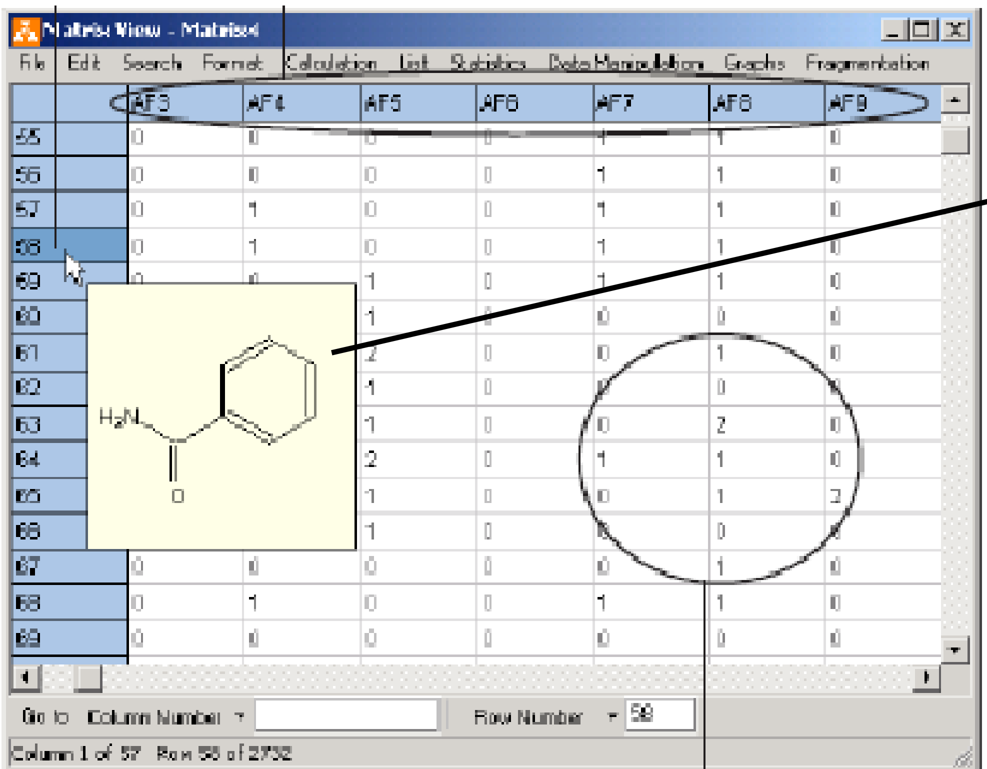
- **LogP** is a linear function of fragments and interactions

  Log P= Σ fn + ΣFm

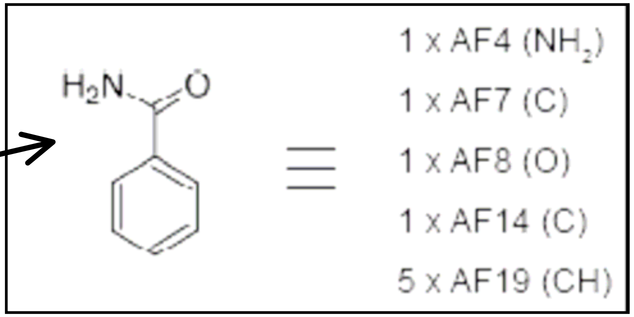


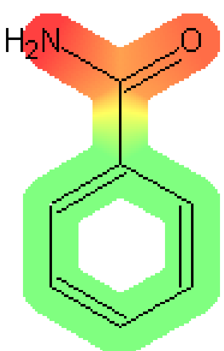
Pharma Algoritms LogP

$$\mathrm{Log}\ X = \Sigma\, a_i \times IC_i + \Sigma\, b_i \times F_j + \Sigma\, c_{ijk} \times Int_{ijk}$$



Number of instances
for each AF

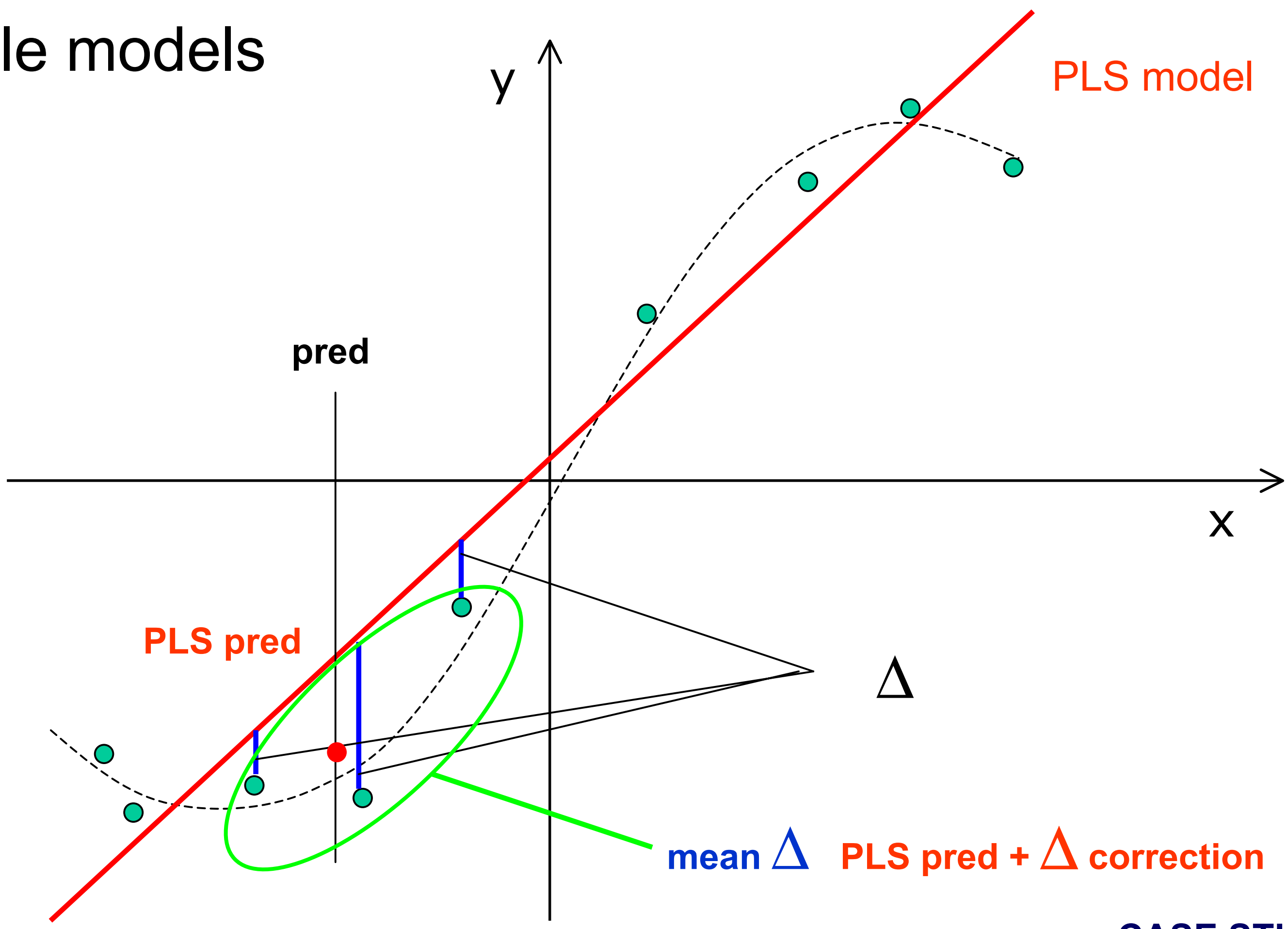




LogP (AB/LogP v2.0): 0.66
Reliability: High (RI = 0.88)

Pharma Algoritms LogP

Trainable models



ALOGP

- Ghose-Crippen-Viswanadhan logP
  - 120 atom types
  - **AlogP = $\sum n_i a_i$**
    - n_i = number of atom of type *i* and
    - a_i = corresponding hydrophobicity constant

MLOGP

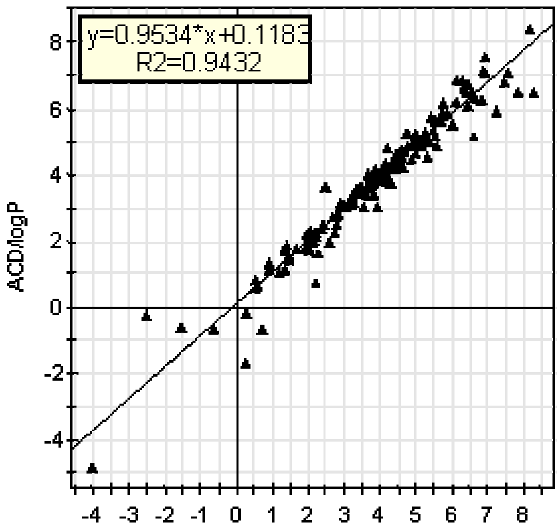
- Moriguchi MLogP
  - 13 structural parameters
    - CX carbon and halogen atoms
    - ON nitrogen and oxygen atoms
    - PRX proximity efffect
    - UB unsaturated bonds
    - HB intramolecular H bonds
    - POL Polar substituents
    - AMP amphotericity
    - ALK alkanes and alkenes
    - RNG ring structures
    - QN quarternary nitrogen
    - NO2 Nitro groups
    - NCS
    - BLM beta lactam

QikProp

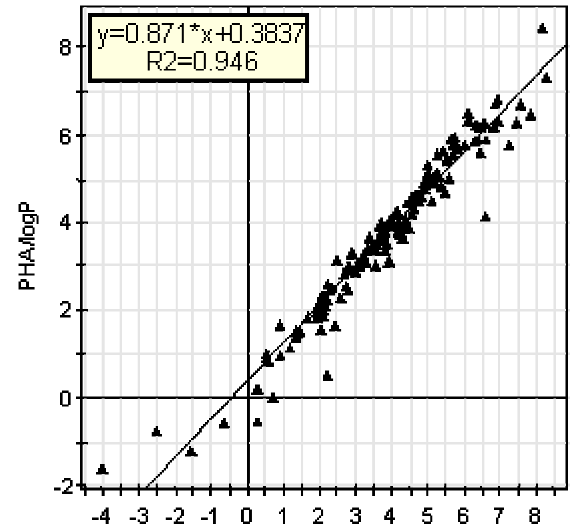
- Developed from W. Jorgensen, Yale
- The prediction is based on molecular descriptors calculated from Monte Carlo simulation in water
  - Surface area calculation and volume analysis.
  - Use algorithms to estimate hydrogen-bond counts and other simple descriptors

Results: Fragment-based methods

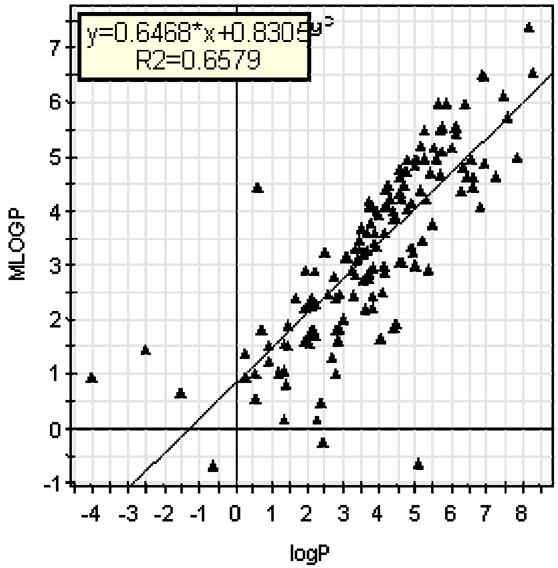
ACDLabs



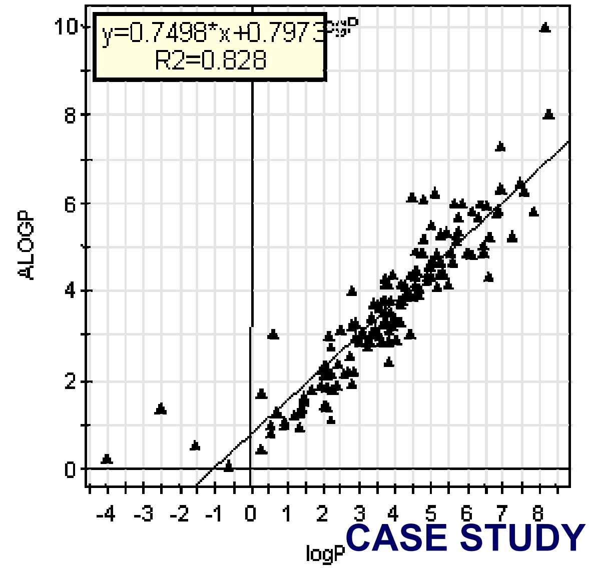
PHA



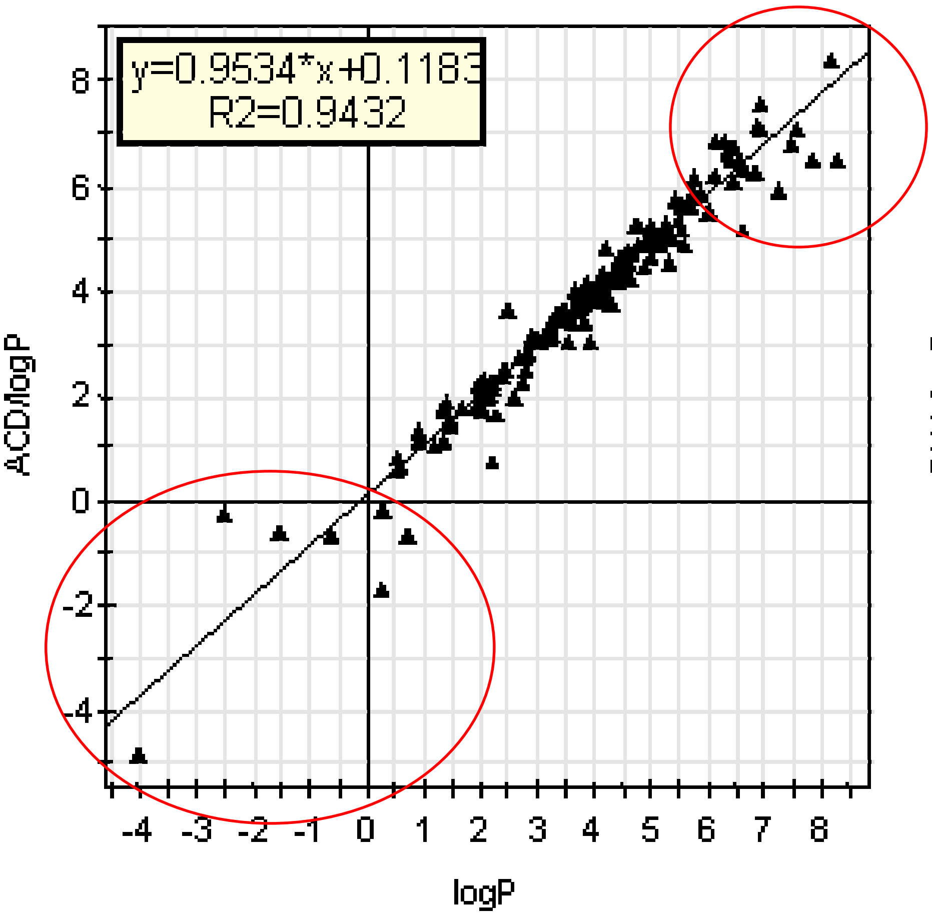
MLOGP



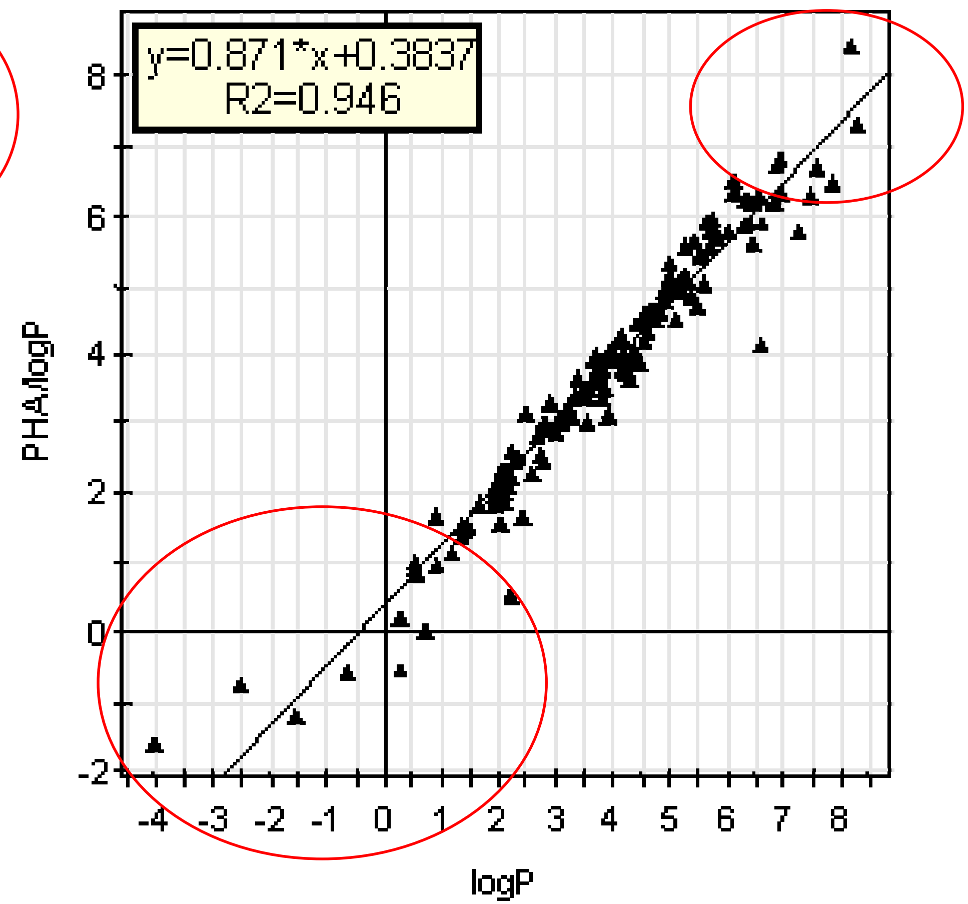
ALOGP



Results: Fragment-based methods

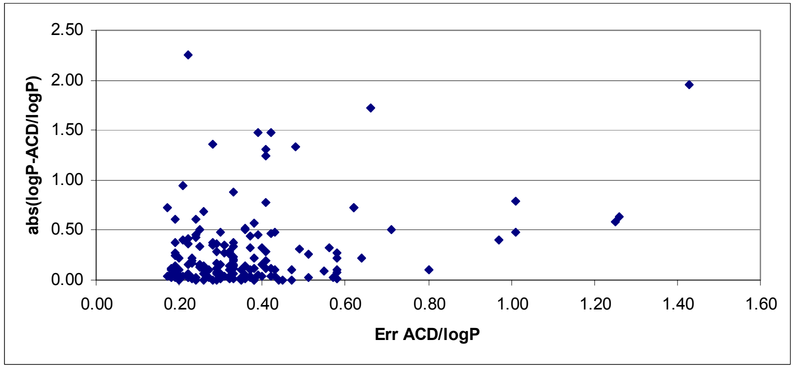


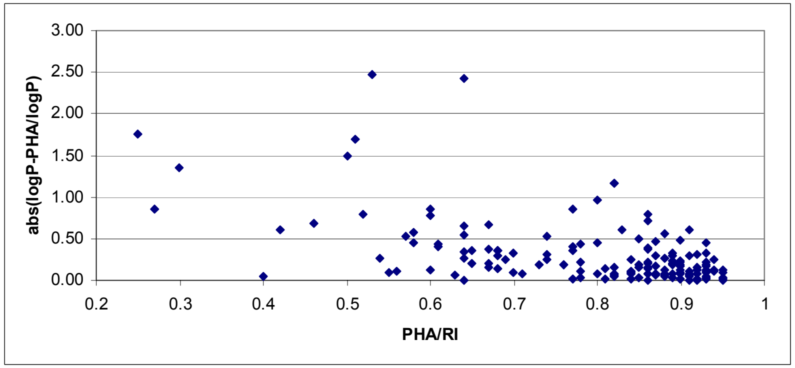
ACDLabs



PHA

Parameters to estimate Accuracy

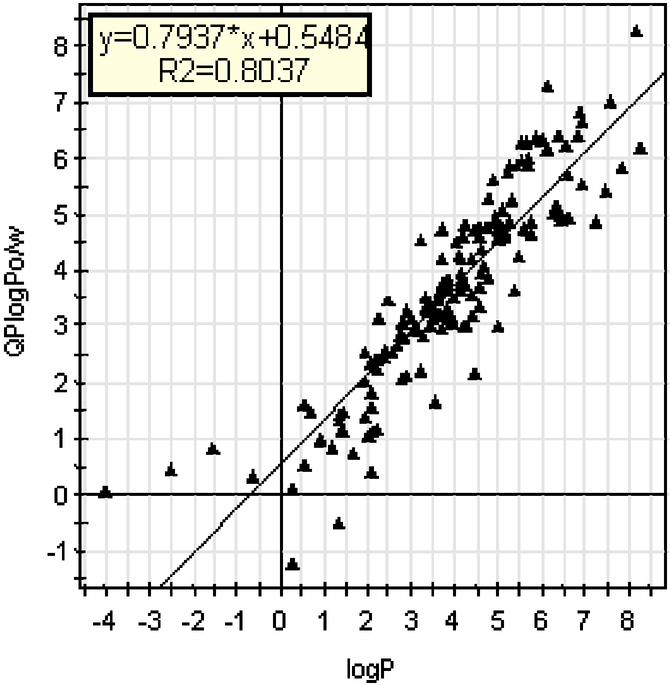




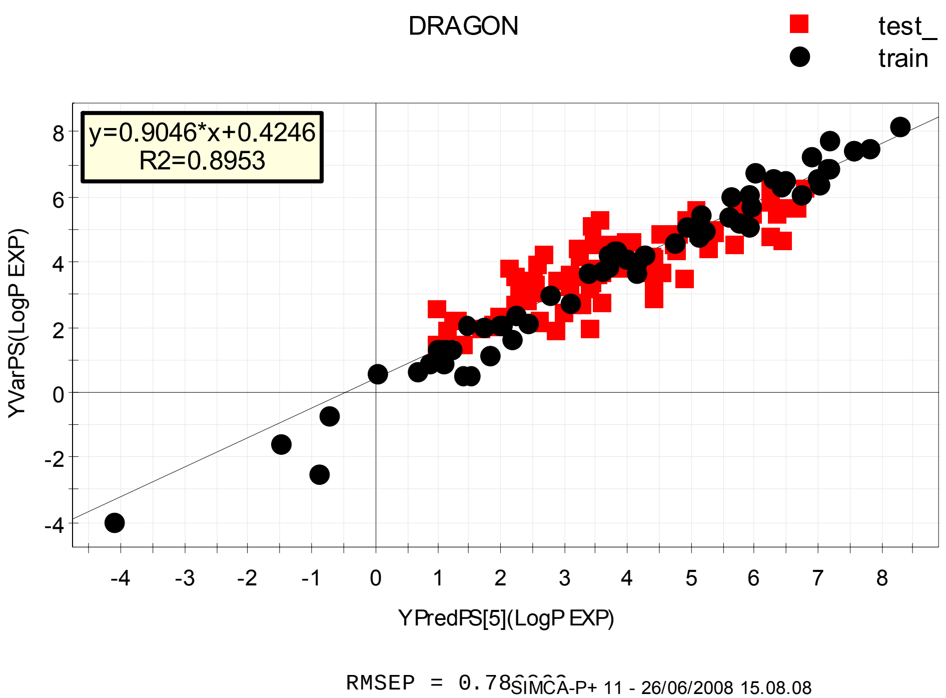
ACD: confidence interval

PHA: reliability Index

Results: QSPR models



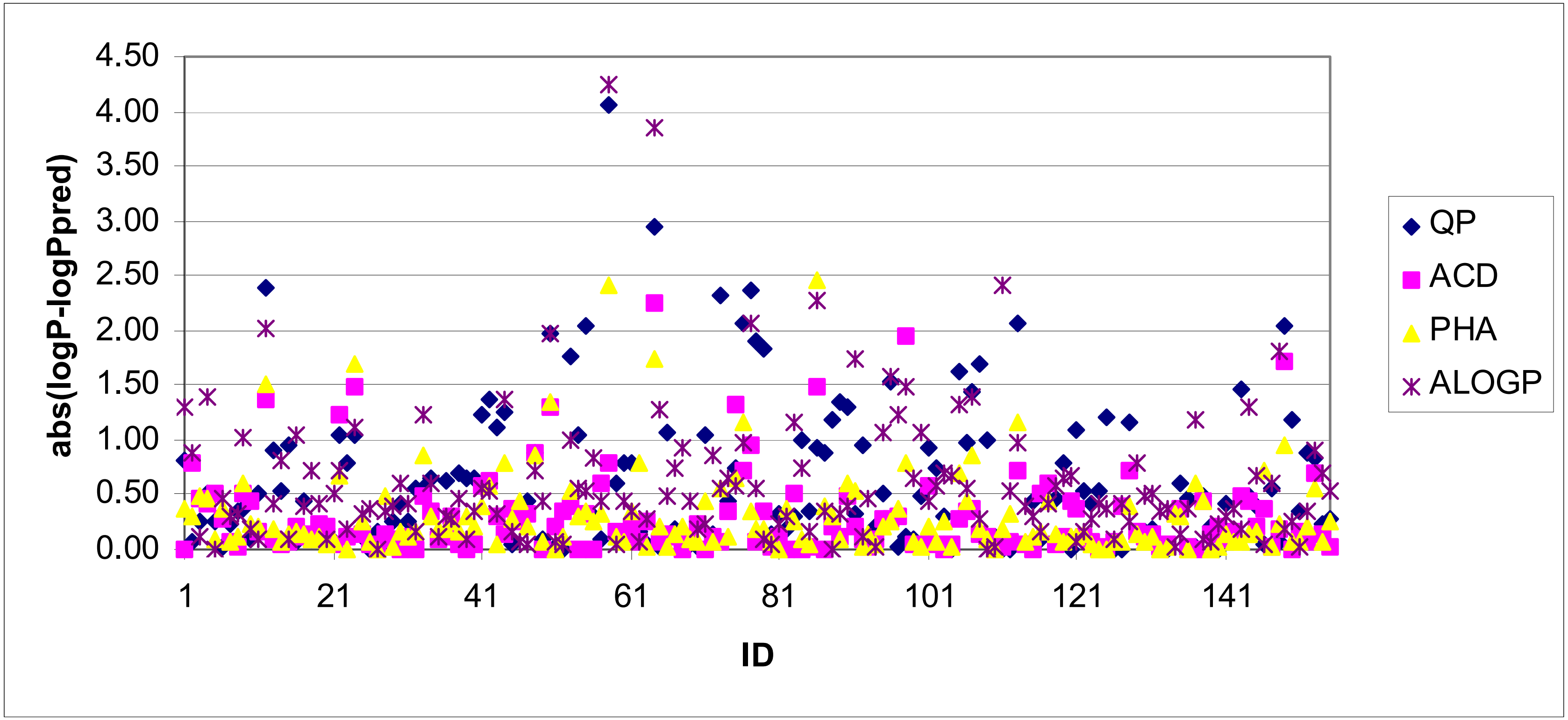
QikProp



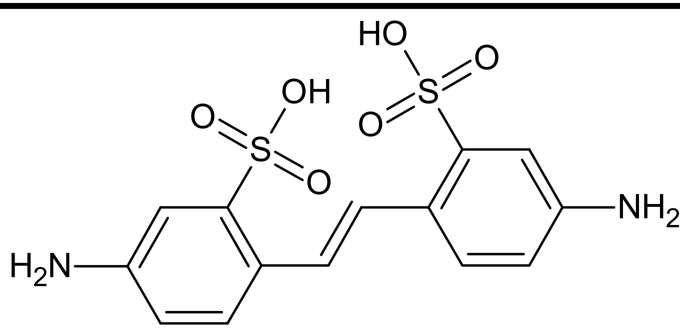
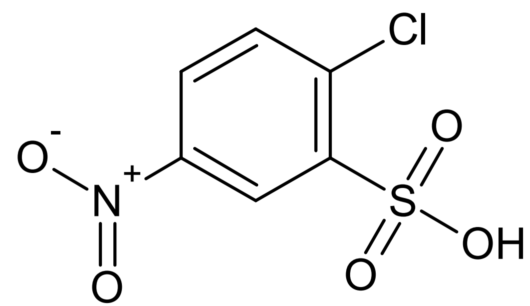
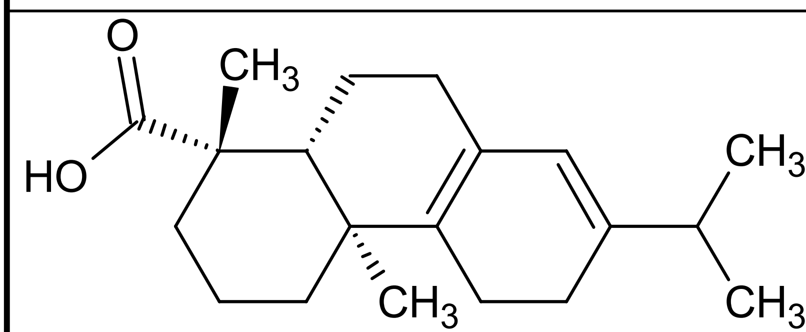
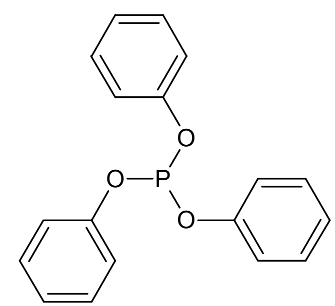
PLS DRAGON

CASE STUDY 1: LOGP

Analysis of errors



Analysis of errors

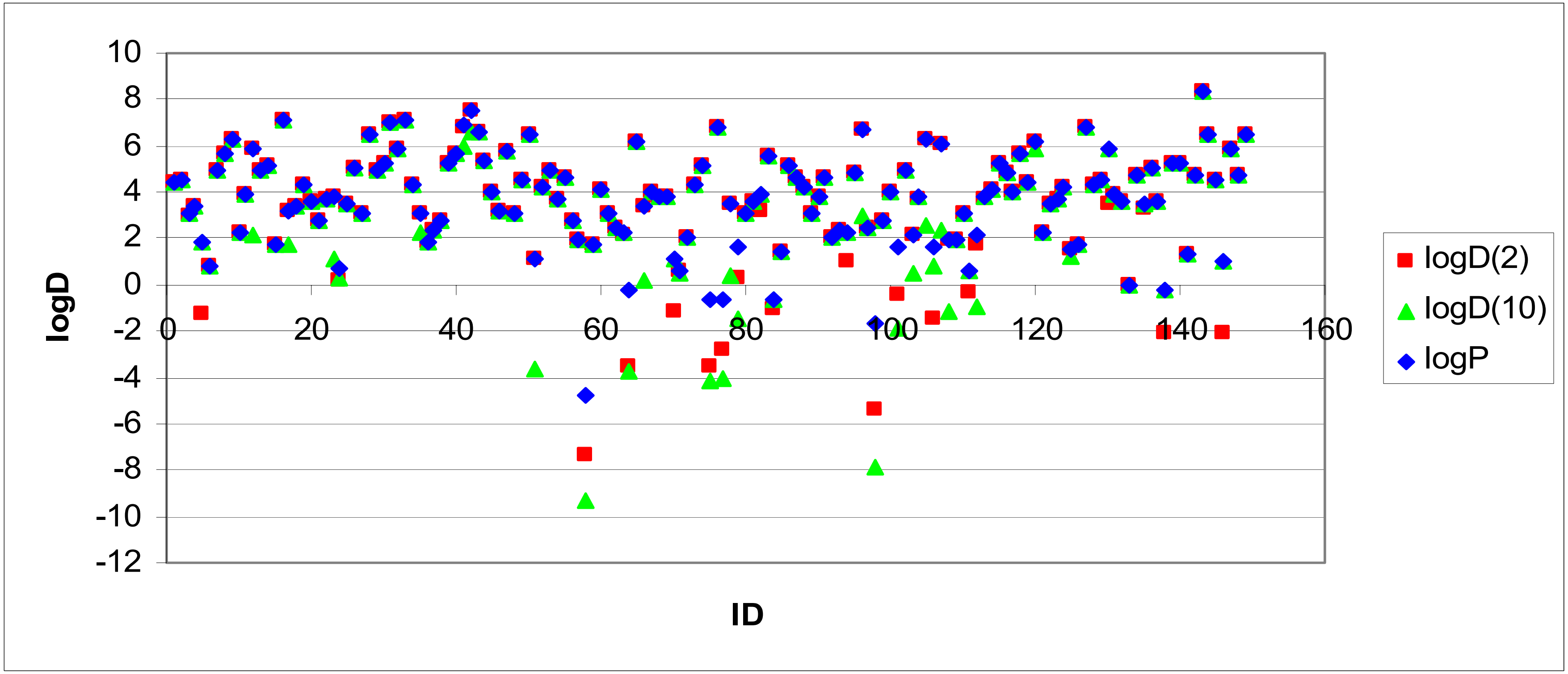
	
Unknown aromatic-S interactions	Unknown aromatic-S interactions
	
High logP value ~ 6	High logP value ~ 6

Results

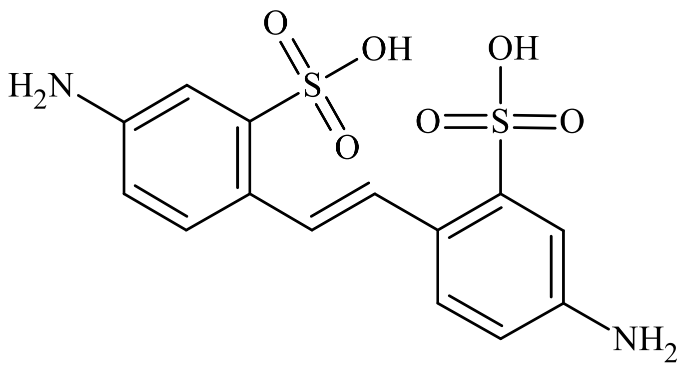
SDEP: Standard Error in Prediction

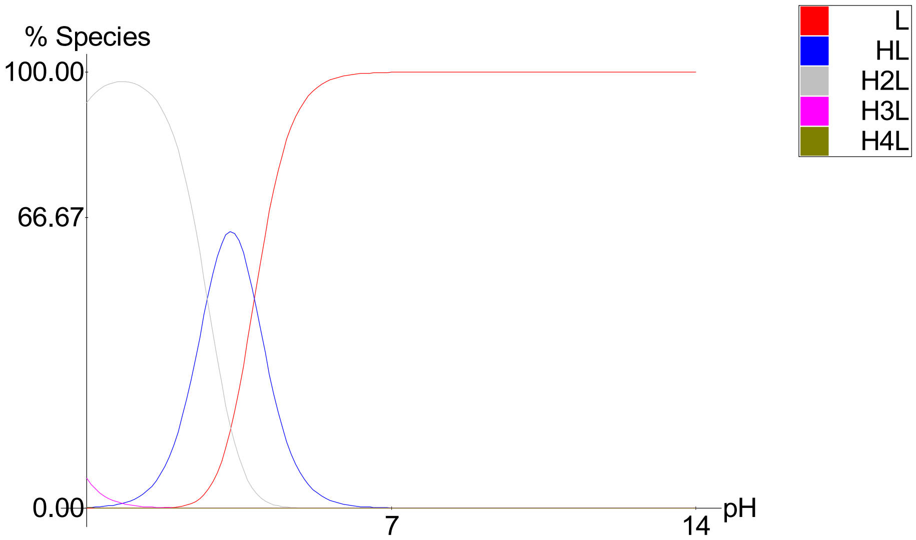
Approach	SDEP	Approach	SDEP
EXP	0.4	ALOGP	0.87
ACDLabs	0.49	QP	0.92
PHA	0.51	PLS QP	0.87
MLPGP	1.29	PLS DR	0.79

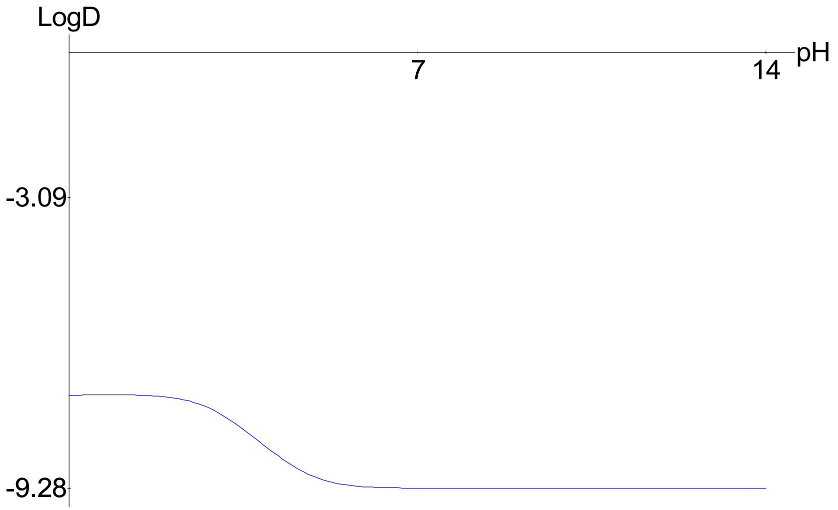
LogP vs logD


10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
logD
0
20
40
60
80
100
120
140
160
ID
logD(2)
logD(10)
logP

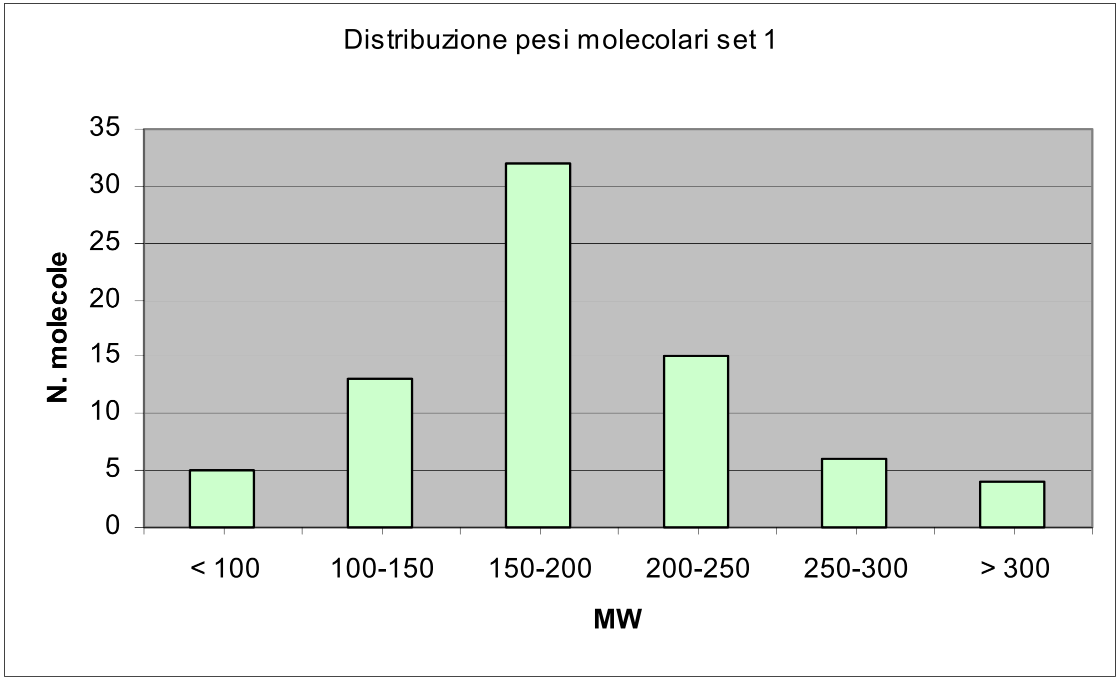
LogP vs LogD




% Species
100.00
66.67
0.00
7
14
pH
L
HL
H2L
H3L
H4L

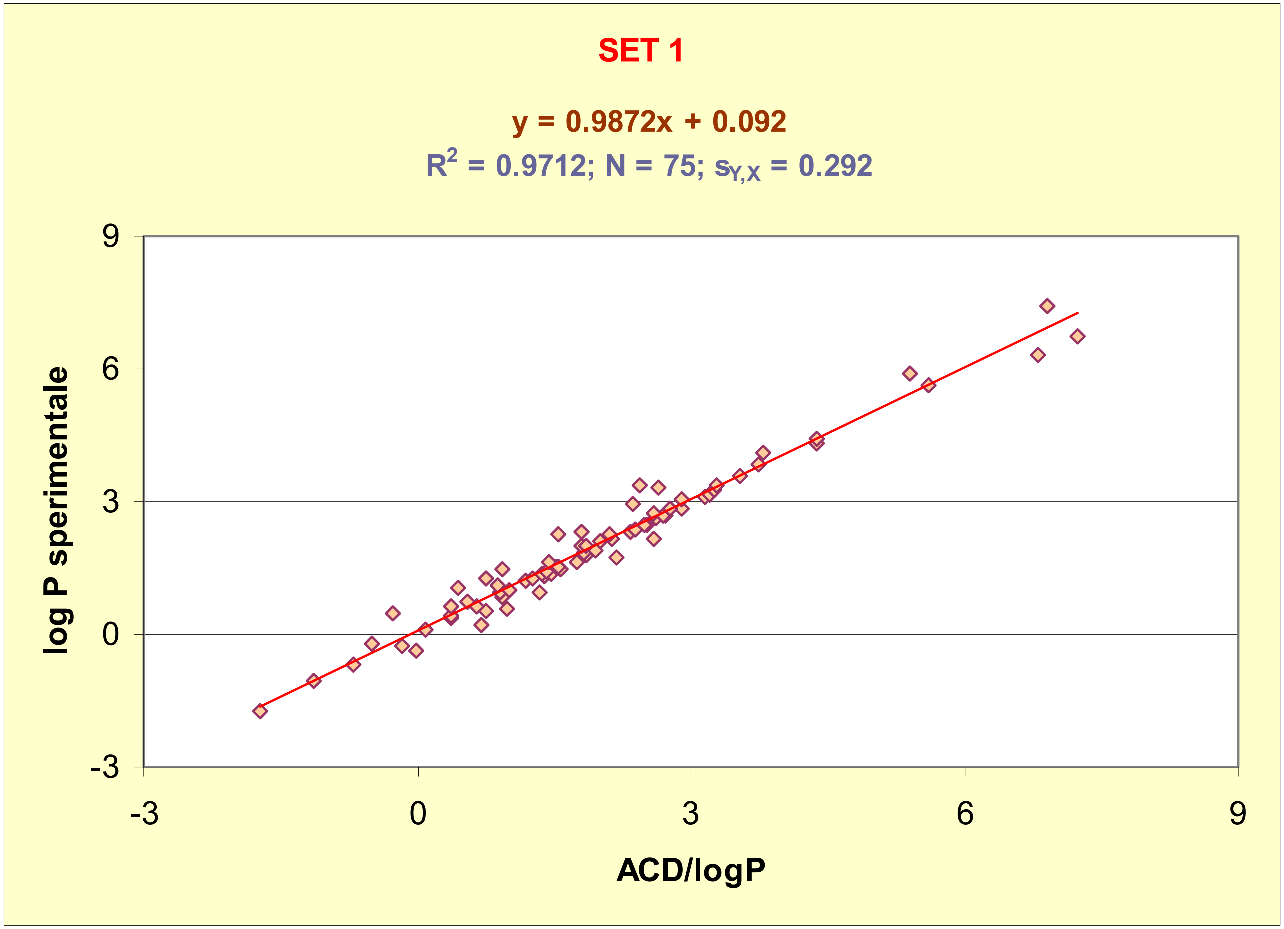

LogD
pH
7
14
-3.09
-9.28

Case Study 2: LogP



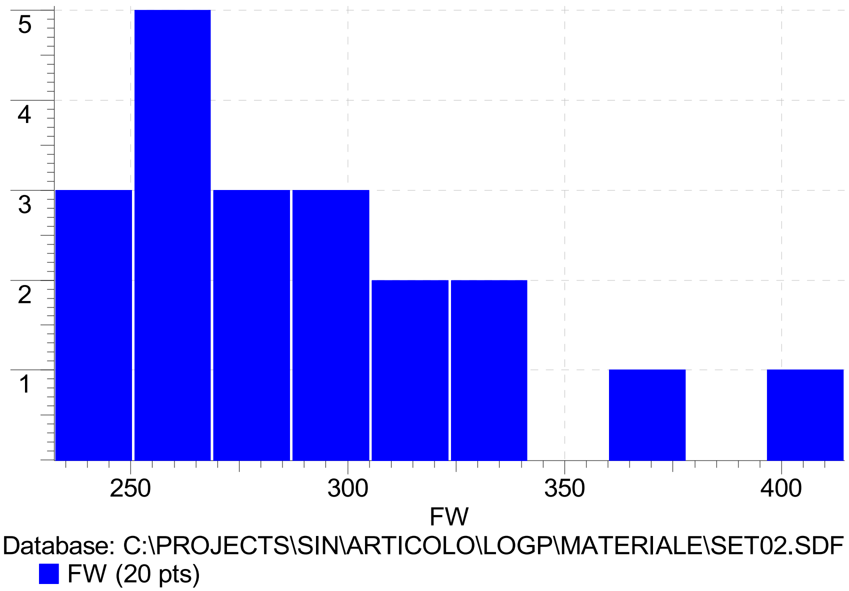
- 75 organic molecules

Case Study 2: Results

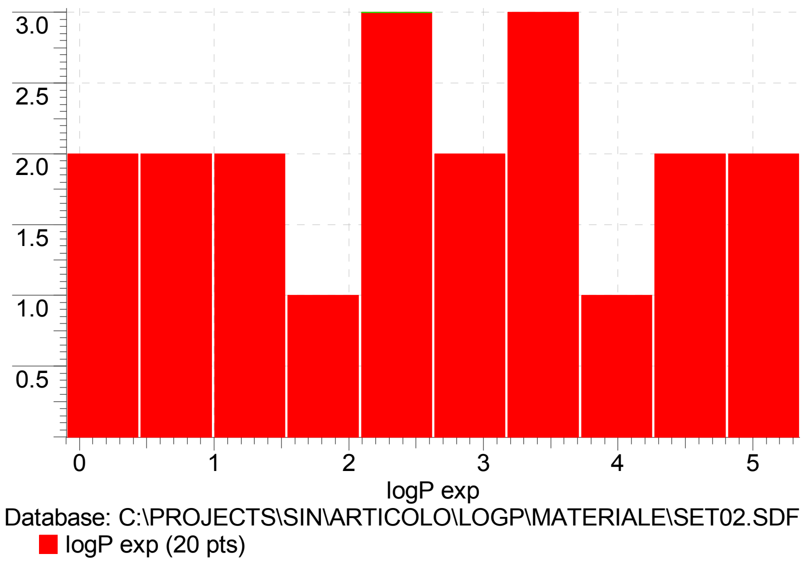


Case Study 3: LogP

- 20 drugs with experimental logP values



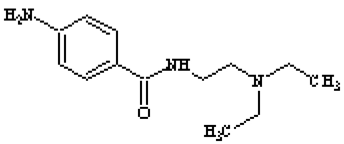
MW distribution



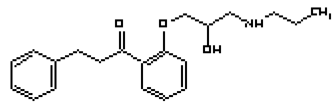
Exp. LogP values distribution

Database View Record Search Lists Options ACD/Labs Help

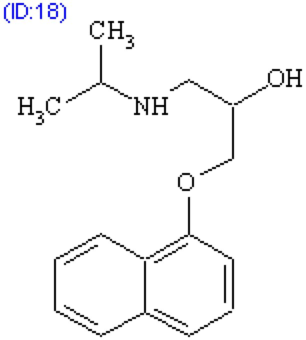
16 (ID:16)



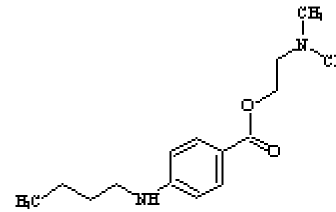
17 (ID:17)



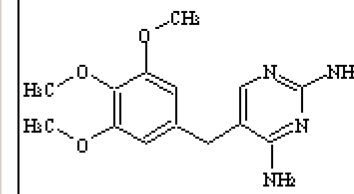
18 (ID:18)



19 (ID:19)



20 (ID:20)



ID: 20	A: 20/20 B: 20	Last Updated: 30/06/2008 14:43	Single DB

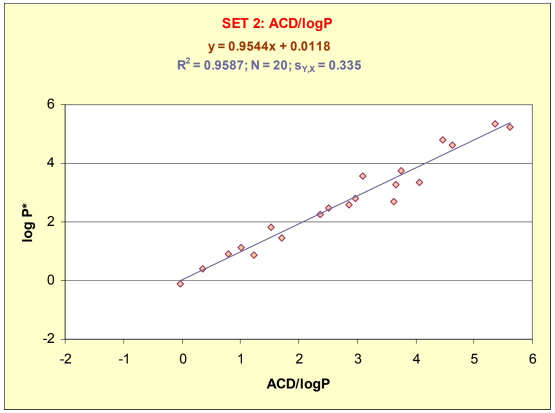
CASE STUDY 3: LOGP

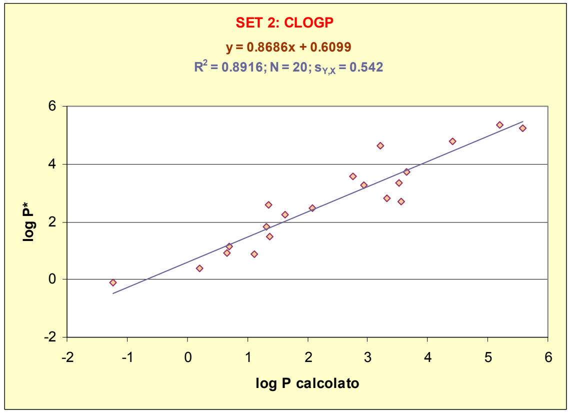
1-ChemSketch 2-Database

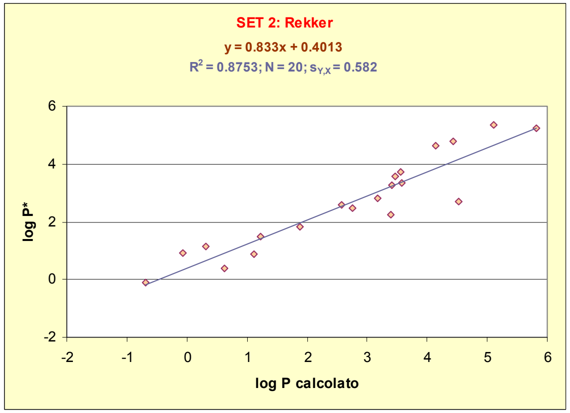
Case Study 3: Methods

- ACDLabs/LogP
- CLOGP (Leo Hansch)
- Rekker

Case Study 3: Results



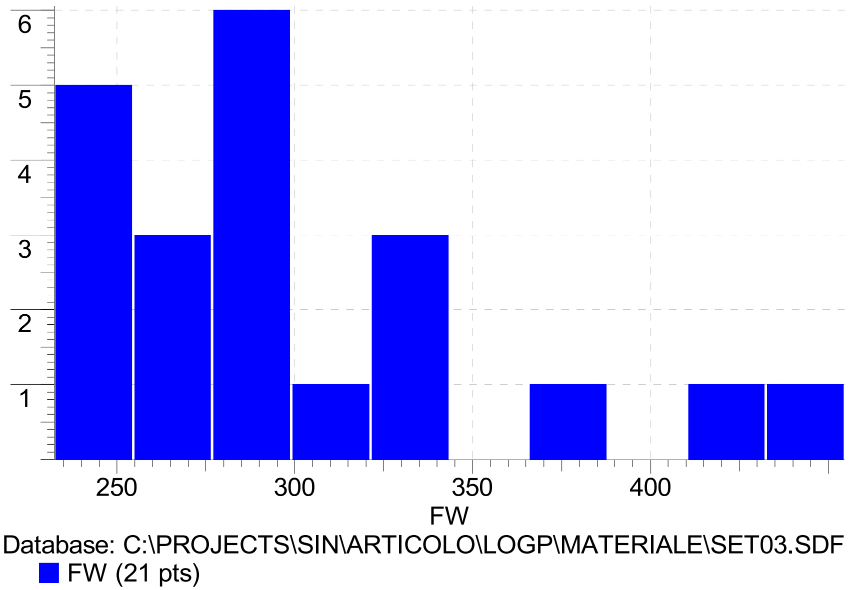




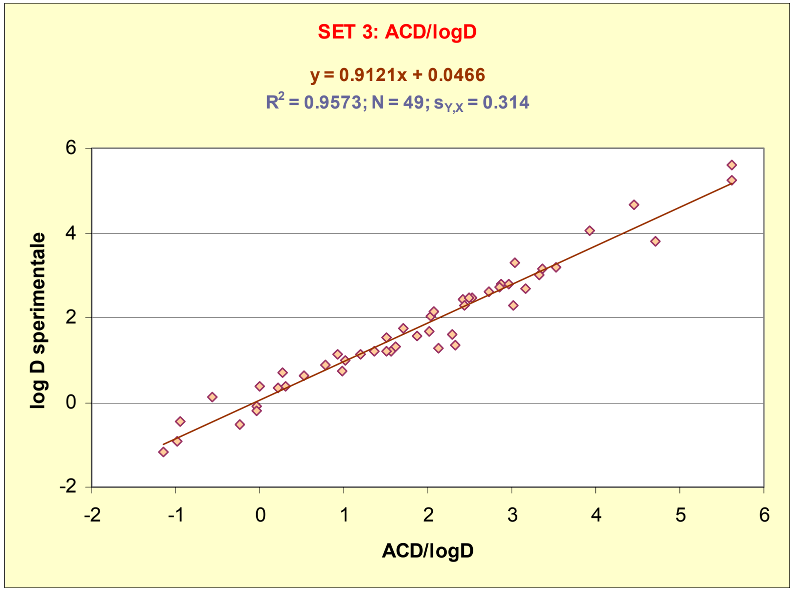
Methods	SDEP
ACDLabs	0.3
CLOGP	0.5
Rekker	0.6
EXP	0.4

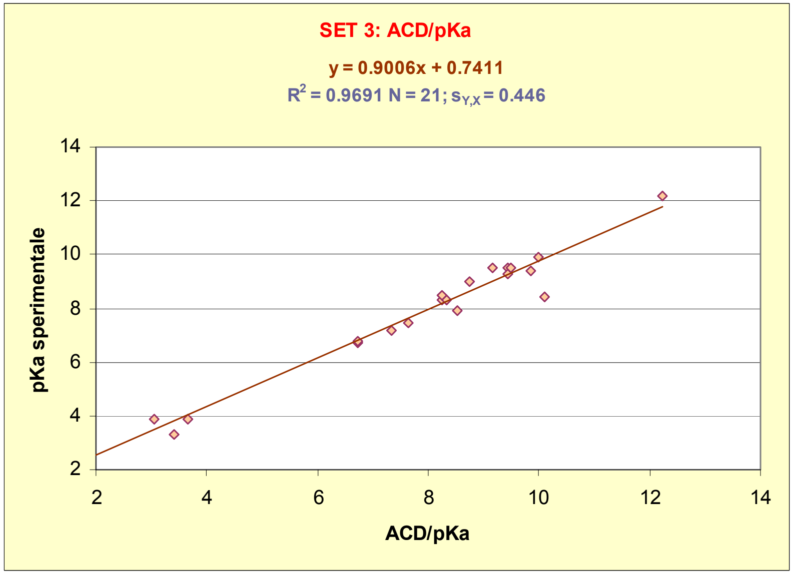
Case Study 4: LogD

- 21 drugs
  - Experimental logD values at different pH
  - Experimental pKa values



Case Study 4: Results





LogD – SDEP = 0.33

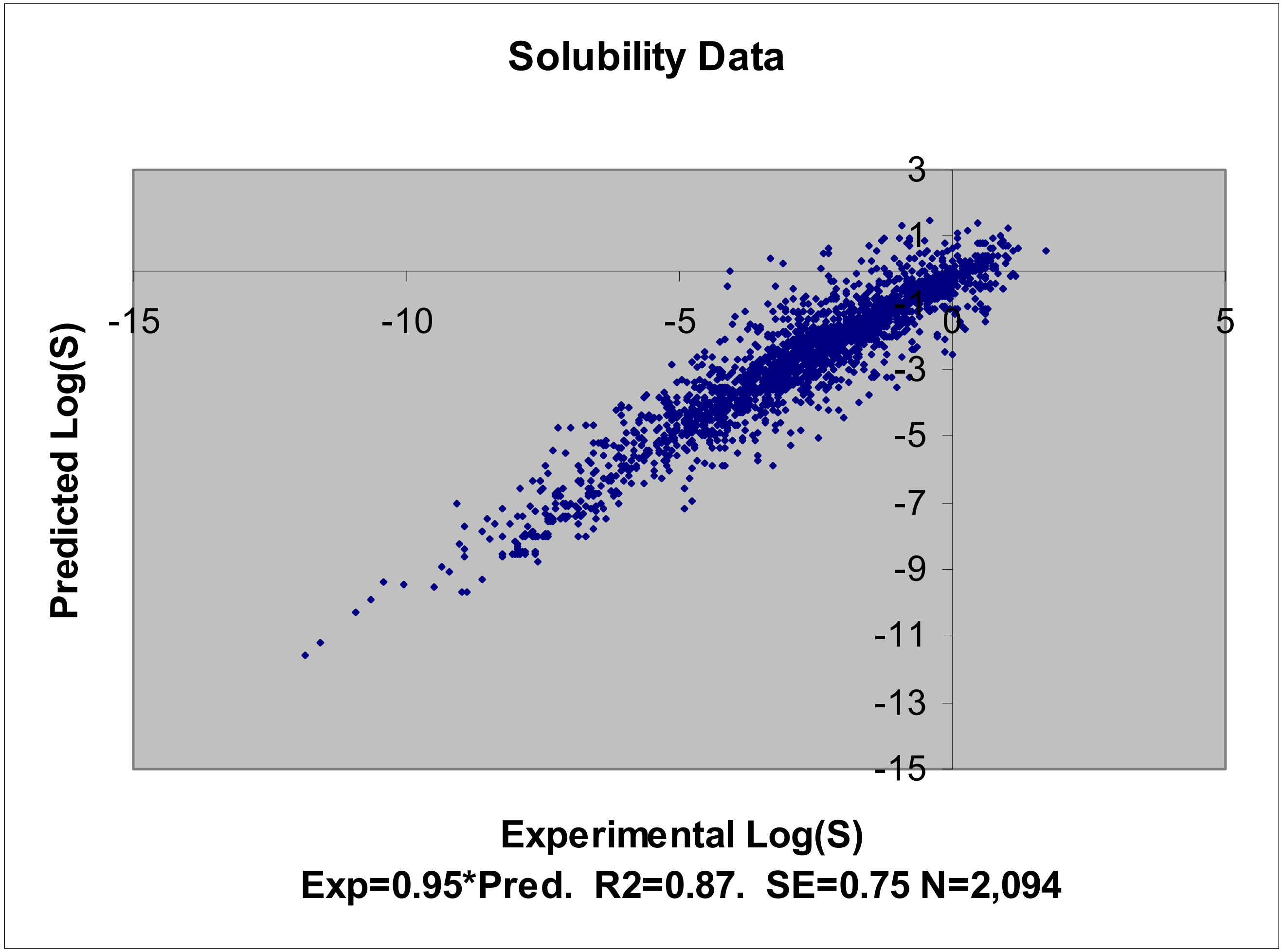
pKa – SDEP = 0.45

Case Study 5: Solubility

- **19 compounds were retrived from an article, with solubility value (pH 7) ranging from 2 to >250 uM**
- **Prediction results using ACD/Sol db:**
  - 7 were predicted correctly (error < 0.3 log unit)
  - 6 were predicted with an error < 1 log unit
  - 6 were predicted with an error > 1 log unit

- **Prediction results using ACD/Sol db + User Training db:**
  - 12 were predicted correctly (error < 0.3 log unit)
  - 4 were predicted with an error < 1 log unit
  - 3 were predicted with an error > 1 log unit

Case Study 5: Solubility



Summary

- Lipophilicity is not just LogP
- Different approaches to *in silico* prediction
- Mean errors in prediction ~ exp. Errors
- General models vs. localized models
- Experimental data useful to improve predictions

Acknowledgments

- S-IN Soluzioni Informatiche
  - Marco Parenti
  - Matteo Stocchero

- Istituto di Ricerche Farmacologiche Mario Negri
  - Emilio Benfenati
  - Elena Boriani
  - Alessandra Roncaglioni

Thanks for your attention

Elena Fioravanzo: elena.fioravanzo@s-in.it