LogD and pKa enhancements: Applications in ADME profiling

ACD UK users meeting, Windsor, 2007

Jordi Munoz-Muriedas, GSK, Stevenage
logD and pKa in ADME

Related with almost every ADME process

**Hydrophobicity generally:**
- Increases absorption, volume of distribution, plasma protein binding, Pgp efflux, CNS penetration, Cytochrome inhibition, intrinsic clearance
- Decreases solubility

**Acidity generally:**
- Increases plasma protein binding, solubility
- Decreases in vivo clearance, CNS penetration, PGP efflux, volume of distribution

**Basicity generally:**
- Increases in vivo clearance, solubility
- Decreases plasma protein binding
pKa predictors performance

ACD pKA vs. other software (n=1000)

Measured - Calculated pKa within specified range

Cumulative %

+/-0.25  +/-0.5  +/-1  +/-2  >+/-2  not_calc

- External 1
- ACD_v4.1
- External 2
- ACD_v7
- External 3
logD predictors performance

ACD logD vs. other software (n=4693)

Measured - Calculated logD within specified range

- +/-0.5
- +/-1
- +/-2
- >+/2
- not_calc

Cumulative %

- in-house method
- ACD_v4.1_logd
- external 1
- ACD_v7_logd
- external 2
ACD logD prediction: logD range [-1.5,3.0] n=17006
logD predictors performance

ACD logD prediction n=22903

Scatterplot (Results_def_LogD_comp_250406_11v*22093c)
A.LOGD = 0.8652 + 0.5052*x; 0.95 Conf.Int.

LogD_ACD9:A.LOGD: r² = 0.5179; r = 0.7197, p = 0.00000; y = 0.885163858 + 0.505190306*x
Quick implementation in ADMET models

Absorption predicted by logD/CMR model

![Graph showing absorption predicted by logD/CMR model]
Quick implementation in ADMET models

Necessity of good logD/pKa predictions

Colours based on predicted Absorption:
Green: Good
Yellow: Good (borderline)
Orange: Poor borderline
Red: Poor

Expt Absorb: High
Pred. Absorb: Fail (high confidence)
Expt. logD= 1.89
ACD logD= -1.5

Expt Absorb: High
Pred. Absorb: Fail (Borderline)
Expt. logD= 0.75
ACD logD= -3.8
Training pKa, logP and logD

ACD physchem and pKa accuracy extenders

- pKa
- logP
- Solubility

pKA and logP trained values can be used together to increase accuracy in logD prediction
pKa training
pKa training

Global dataset training

ApKa1 Prediction

N=369

Cumulative %

Error (log units)

BpKa1 Prediction

N=684

Cumulative %

Error (log units)
pKa training: Basic compounds

Global dataset training: Basic pKas

Before Training

After training
pKa training

Acidic compounds: New reaction center

Most acidic pKa: Measured vs. ACD v9

Before Training

After Training

Most acidic pKa: Measured vs. ACD v9_user

Before Training

After Training
pKa training

Acidic compounds: Absorption prediction

Colours based on predicted Absorption:
Green: Good
Yellow: Good (borderline)
Orange: Poor borderline
Red: Poor

Expt Absorb: High
Pred. Absorb: Fail (Borderline)
Expt. logD= 2.25
ACD logD= 1.20

ACD logD (pKa trained) = 2.84
New absorption prediction: High
pKa training

LogD correction using trained pKas (n=17006)

LogD prediction

Cumulative %

V9_User
V9

Error (log units)

<0.5 37 36
<1 64 63
<2 90 90
>=2 100 100
Solubility correction using trained pKas (n=909)

Error Sol(pH7)-Error Sol(pH1.2) / Error Sol(pH7)* - Error Sol(pH1.2)*

<table>
<thead>
<tr>
<th></th>
<th>Error&lt;0 .5</th>
<th>Error&lt;1 .0</th>
<th>Error&lt;1 .5</th>
<th>Error&lt;2 .0</th>
<th>Error&lt;3 .0</th>
<th>Error&gt;3 .0</th>
</tr>
</thead>
<tbody>
<tr>
<td>v9</td>
<td>385</td>
<td>575</td>
<td>697</td>
<td>784</td>
<td>867</td>
<td>909</td>
</tr>
<tr>
<td></td>
<td>42.4%</td>
<td>63.3%</td>
<td>76.7%</td>
<td>86.2%</td>
<td>95.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>v9_pKa_user</td>
<td>386</td>
<td>581</td>
<td>711</td>
<td>796</td>
<td>876</td>
<td>909</td>
</tr>
<tr>
<td></td>
<td>42.5%</td>
<td>63.9%</td>
<td>78.2%</td>
<td>87.6%</td>
<td>96.4%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
**pKa training**

**Application to projects**

\[
\text{Before training: pKa} = 1.62 \text{pKa}_\text{acd} - 7.00 \\
R^2 = 0.72
\]

\[
\text{After training: pKa} = 1.07 \text{pKa}_\text{acd} - 0.57 \\
R^2 = 0.87
\]

N=16
pKa training

Application to ADME

Bioavailability and CYP inhibition issues related with the pKa of the imidazopyridine group

Poor pKa prediction by ACD

Results improved significantly after pKa AE training

Validation set
N=28

Before training

\[
pKa = 0.09pKa_{acd} + 6.03
\]
\[R2=0.02\]

After training

\[
pKa = 1.14pKa_{acd} - 0.93
\]
\[R2=0.63\]
pKa training

2C19 pIC50 vs. Basic pKa (n=77)

Expt pKa

ACD pKa

ACD trained pKa

Experimental pKa
R2=0.40

Before Training
R2=0.09

After Training
R2=0.33
“logD training”
Using logP training to improve logD prediction

Training process leads to a better correlation but to a systematic underestimation of logD

Expt LogD = 0.68 ACD logD + 0.23
R2 = 0.39
Average error 0.17
Overprediction

Expt LogD = 0.69 ACD logD + 0.57
R2 = 0.64
Average error -0.35
Underprediction
logD training

Relating error with specific chemical groups

Exp LogD vs Trained ACD logD

Red: Amides

N=26
logD training

Relating error with specific chemical groups

- logP underpredicted for series with amides
- Training process corrected this partially
logD training

Relating error with pKa prediction

The training process leads to an underprediction of the logP. Probably due to a underprediction of the basicity of the compounds
The figure shows a scatter plot with a linear regression line. The equation of the line is $Y = 0.37x + 5.46$, with a coefficient of determination $R^2 = 0.26$. The x-axis represents the logD values, while the y-axis represents the pKa underprediction.
pKa influence in logP training

\[ \log D = \log P - \sum_{i=1}^{N_a} \log \left( \frac{1 + 10^{\Delta_i^A - ApKa_i}}{1 + 10^{pH - ApKa_i}} \right) - \sum_{i=1}^{N_b} \log \left( \frac{1 + 10^{BpKa_i - \Delta_i^B}}{1 + 10^{BpKa_i - pH}} \right) \]

**Current algorithm:**
- Experimental logD and predicted ACD pKa values are used to calculate estimation of “experimental” logP
- These estimations are used to train logP fragments
- Trained logP can be used to calculate logD
- LogD predictions can be enhanced by using (independently) trained pKas
- Problem: Errors in pKa prediction in the first step will affect logP training and logD prediction

**Ideal:**
- It should be possible to introduce Experimental logD and experimental pKa (or at least trained pKas) to get a better prediction of logP in the first step
- This logP trained using pKa information and trained pKa should lead to a better logD prediction
Conclusions

- pKa accuracy extender is a very useful tool to enhance pKa predictions when applied to improve predictions for local sets.

- pKa AE power would be increased and be of great help if it could be used simultaneously with the logP training algorithm.

- logP training algorithm is very dependent of pKa prediction and should not be used with datasets where errors in pKa are expected unless experimental logP values could be provided.
Acknowledgements

- **ADMET In Silico Team, GSK, Stevenage:**
  Anne Hersey, Paul Gleeson, Sandeep Modi

- **ACD Labs:**
  Ed Kovolanov, Greg Pearl