CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Prediction of pH-Dependent Solubility of Drugs and Drug Candidates

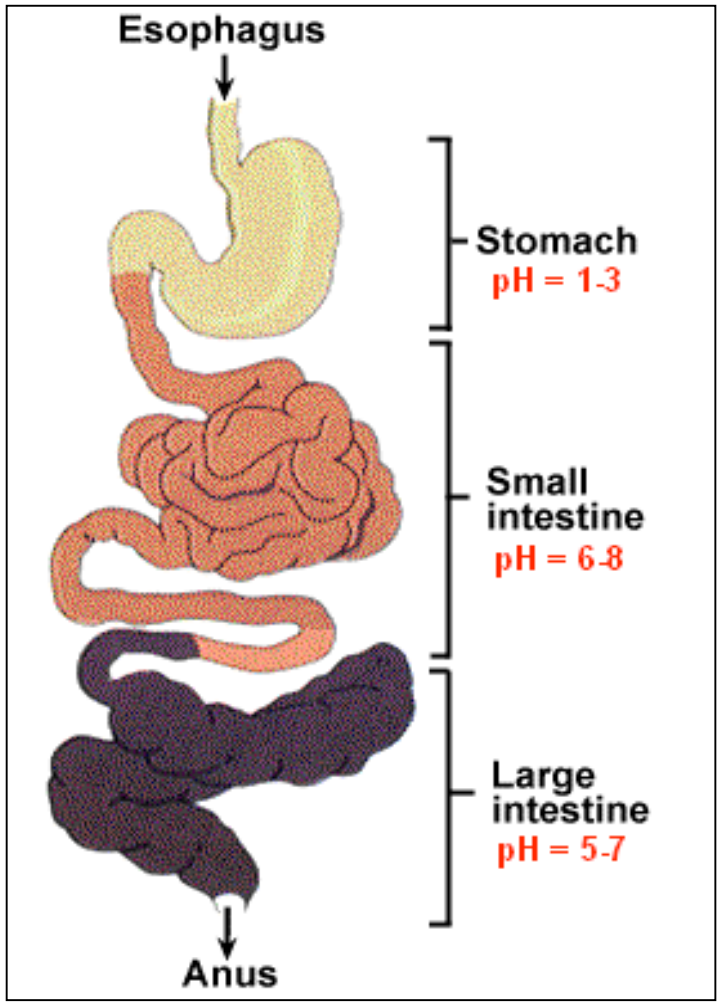
Niclas Tue Hansen

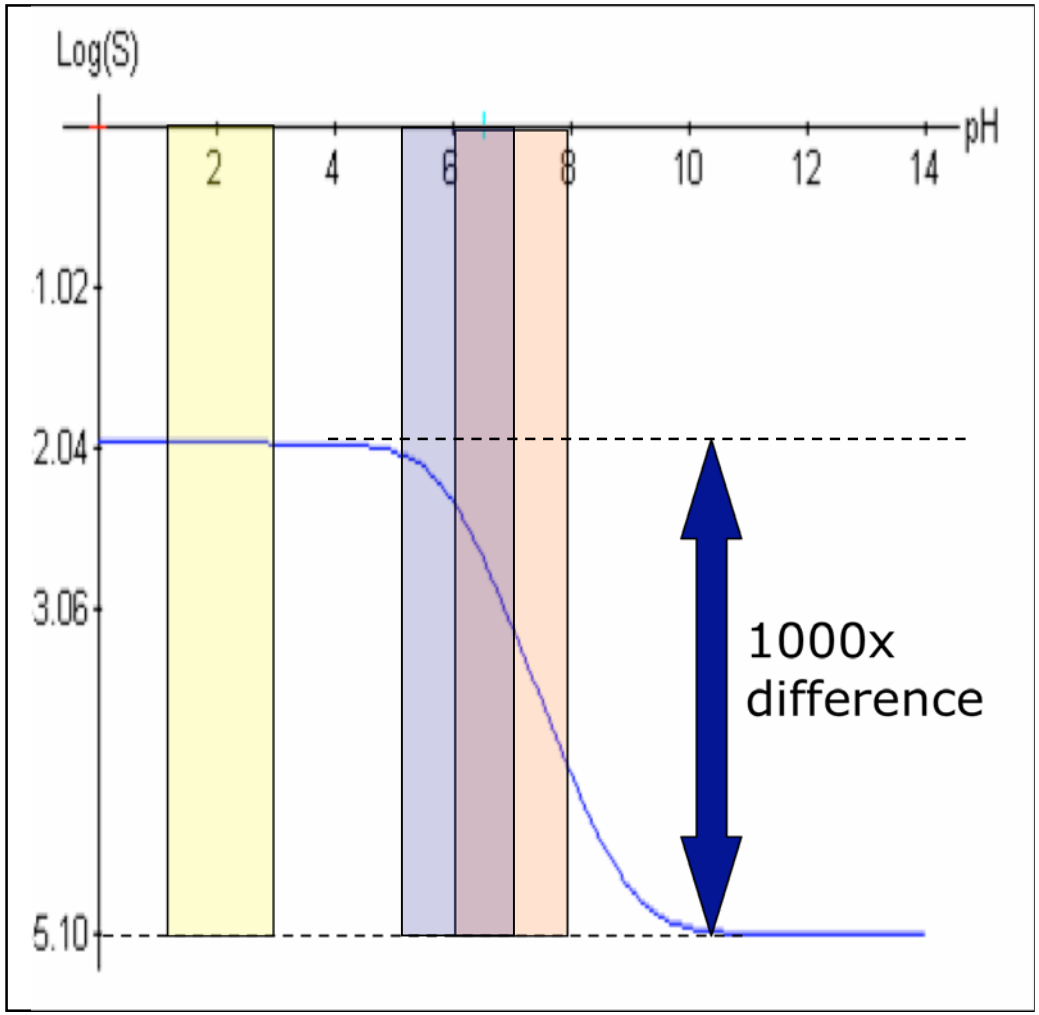
Centre for Biological Sequence Analysis
Technical University of Denmark

Workshop Chemoinformatics in Europe, Obernai, France, 29 May – 1 June 2006

DTU

Why consider pH-dependency?

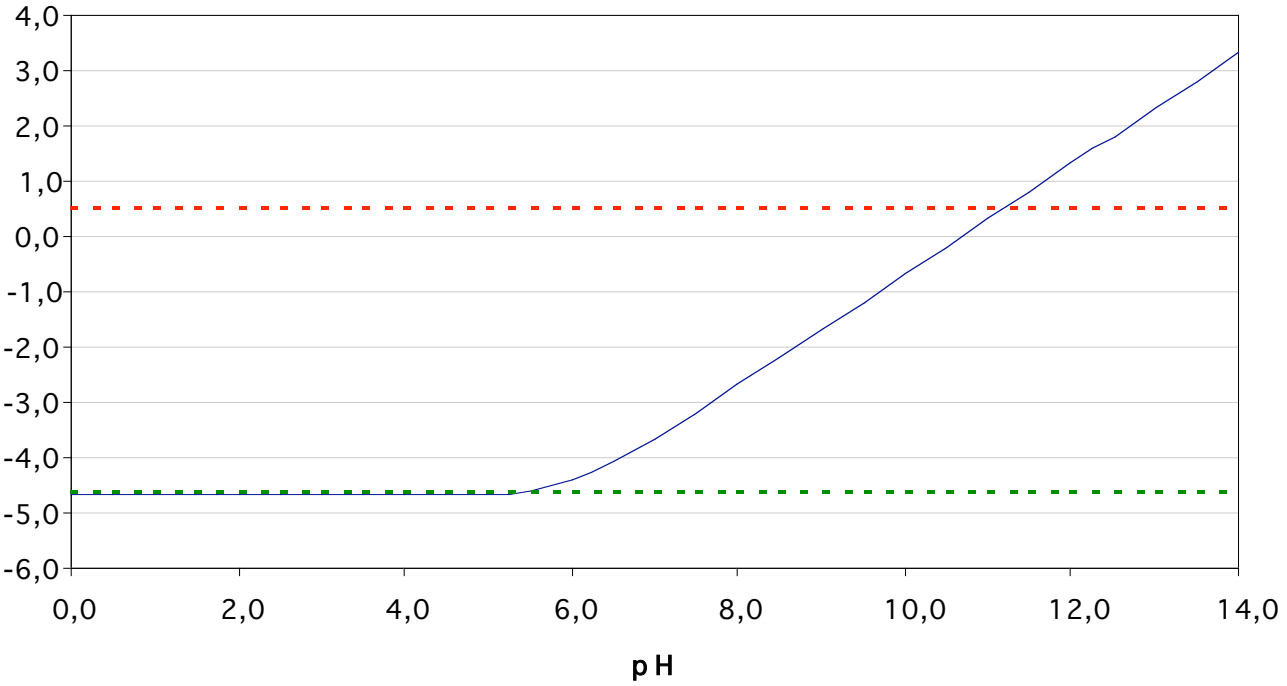




In Theory: Henderson-Hasselbalch

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

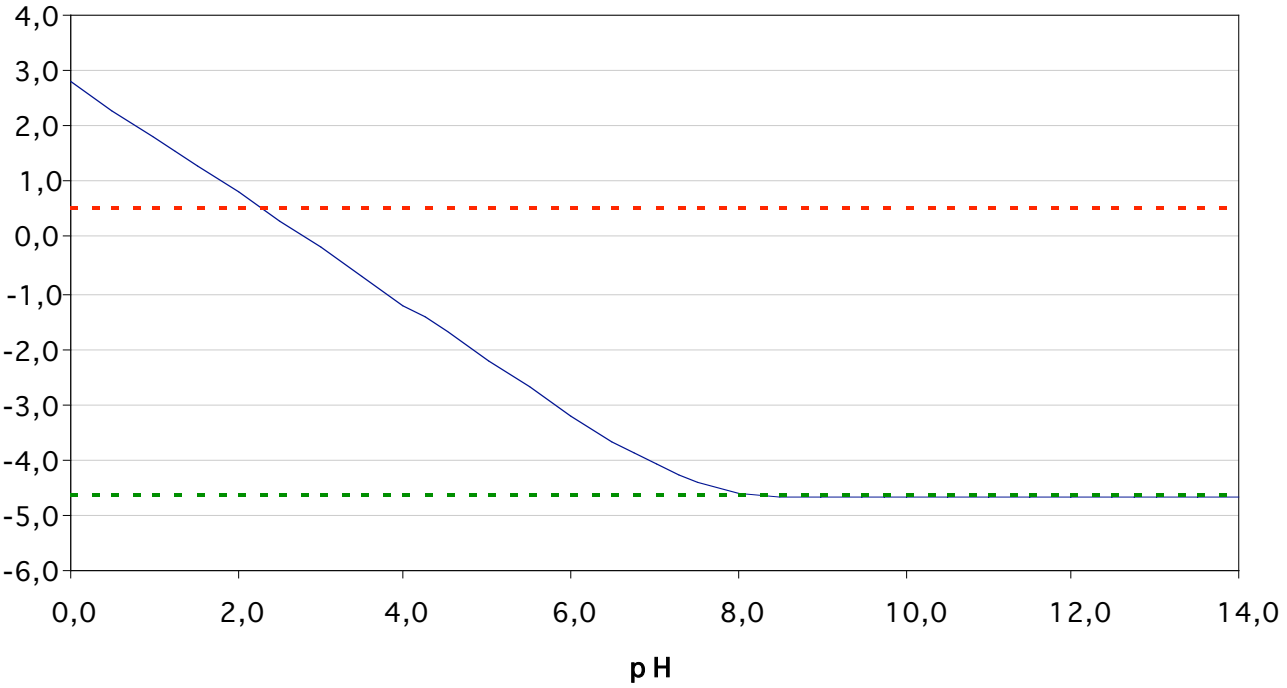
Acids: $$LogS = LogS_0 + Log\left(1 + 10^{pH - pK_a}\right)$$



In Theory: Henderson-Hasselbalch

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Bases: $$LogS = LogS_0 + Log\left(1 + 10^{pK_b - pH}\right)$$

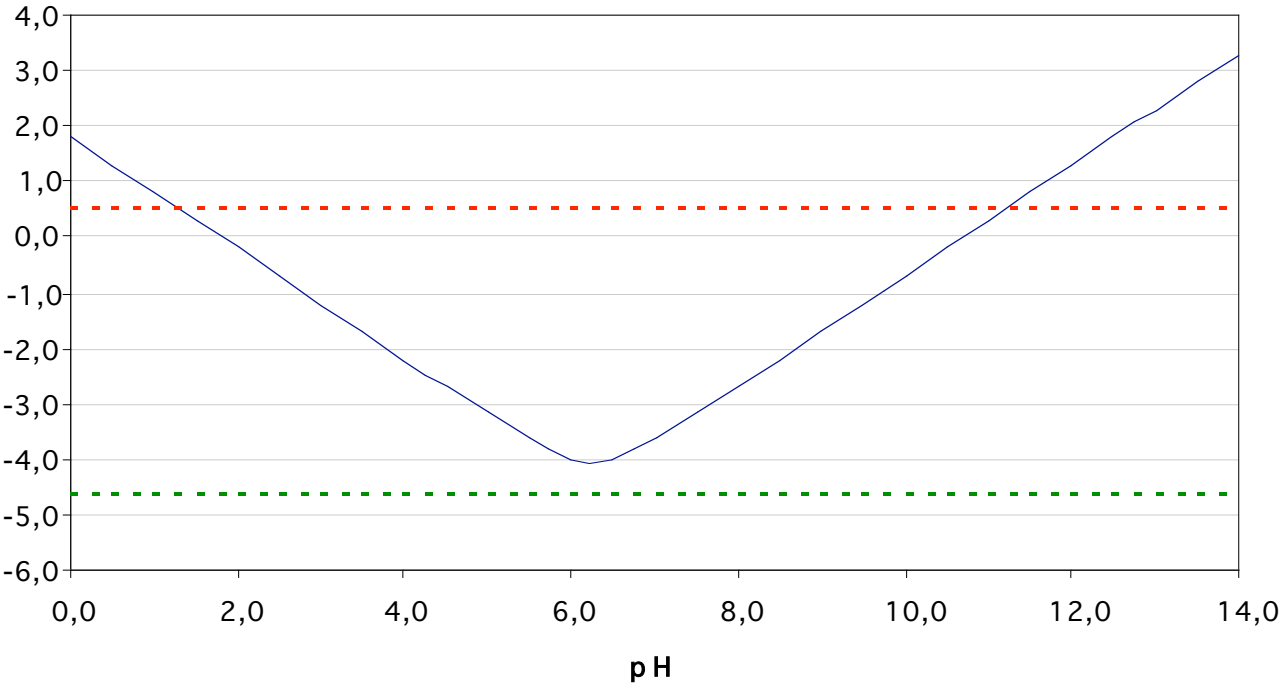


In Theory: Henderson-Hasselbalch

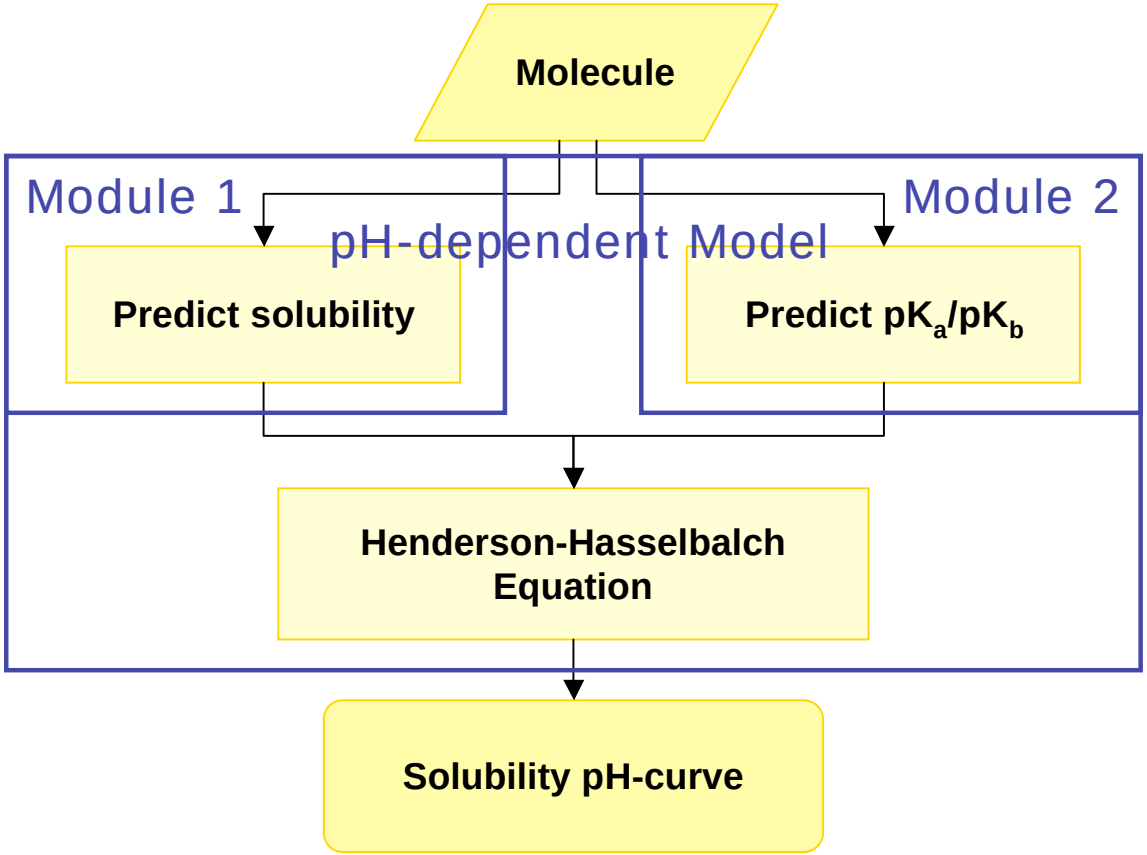
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Ampholytes:

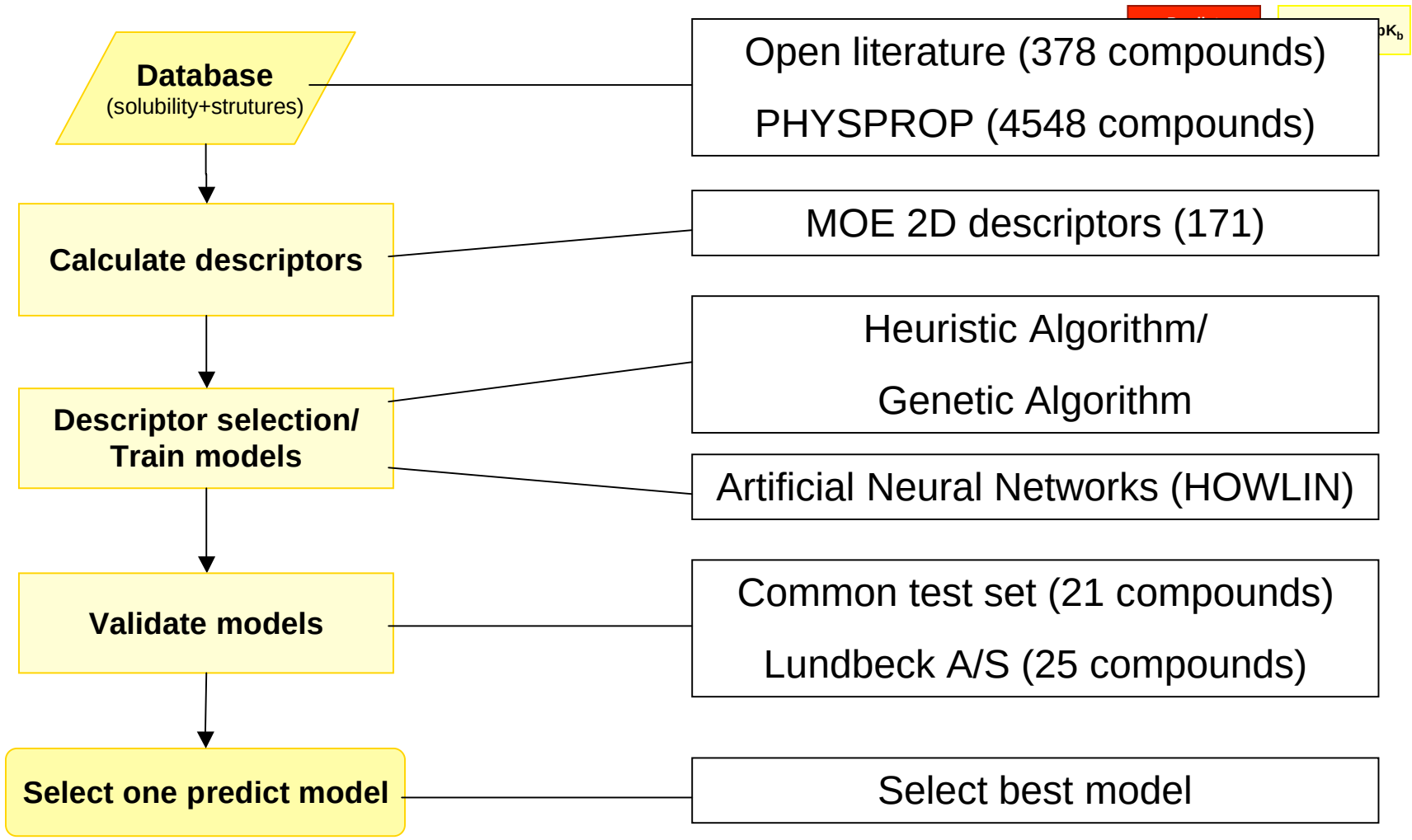
$$LogS = LogS_0 + Log\left(1 + 10^{pH - pK_a} + 10^{pK_b - pH}\right)$$



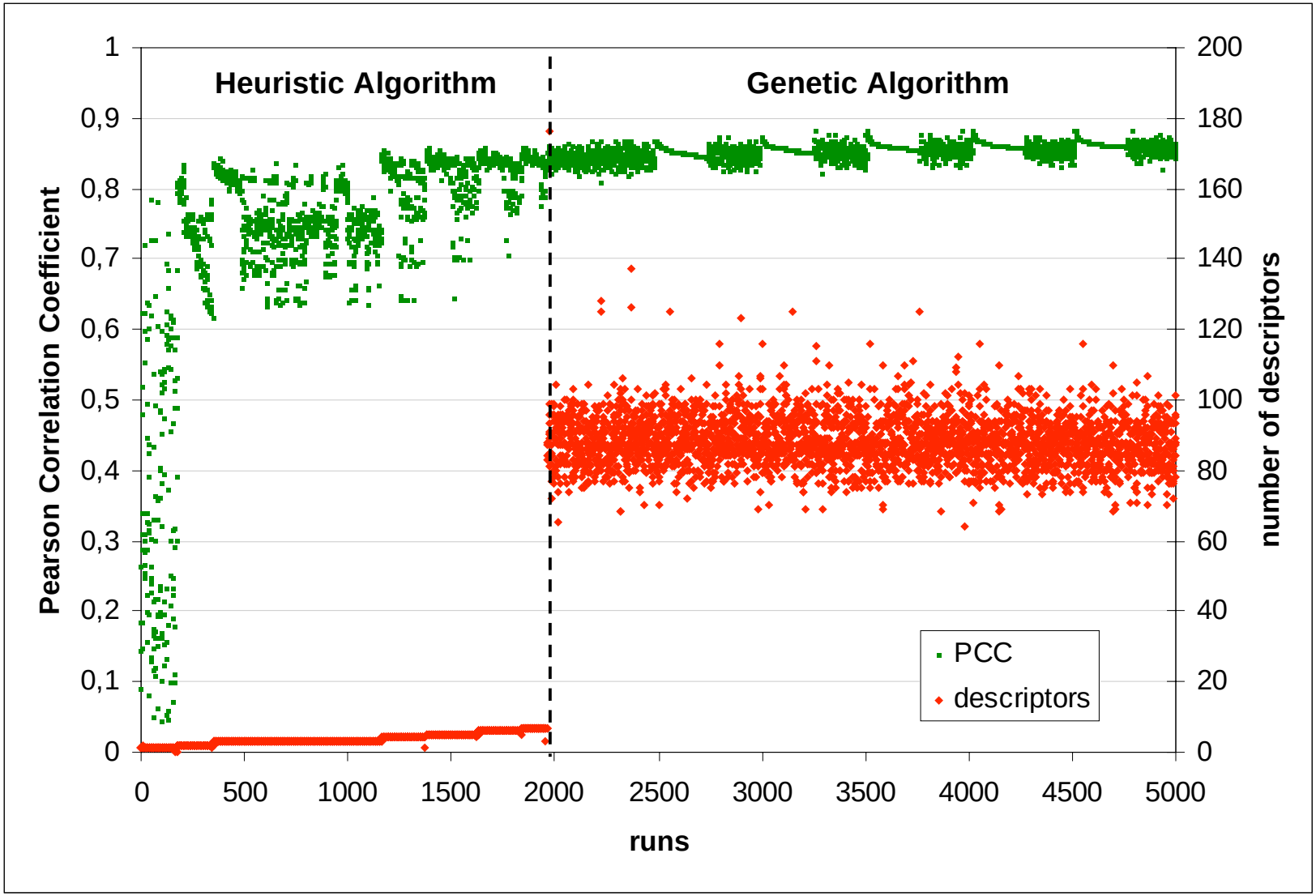
Modular pH-dependent model



Intrinsic Solubility Module



Descriptor selection



Results

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

	No. Desc.	Training (3-part validation)	
		R^2	RMSE
378/Heu	5	0.86	0.97
378/GA	>50	0.90	0.87
4548/Heu	9	0.94	0.64
4548/GA	>50	0.94	0.64

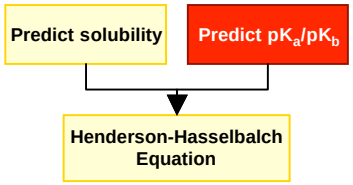
External validation: Others

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

	Huusk.	Klopman	Kuhne*	This work
Train/ Test set	1297	482	694	4548
# desc.	30	46	50	9
R^2	0.91	0.70	0.75	0.85
RMSE	0.63	1.24	1.06	0.97

* Based on 19 of the 21 molecules

pK_a Prediction Module



Two models were validated:

- ACD/Labs pK_a module
- Marvin pK_a plug-in

Validation set (PhysProp) with filter:

- Experimental pK_a values
- Temperature range 25±5°C
- pKa range 0-13

Result: 467 experimental pK_a values

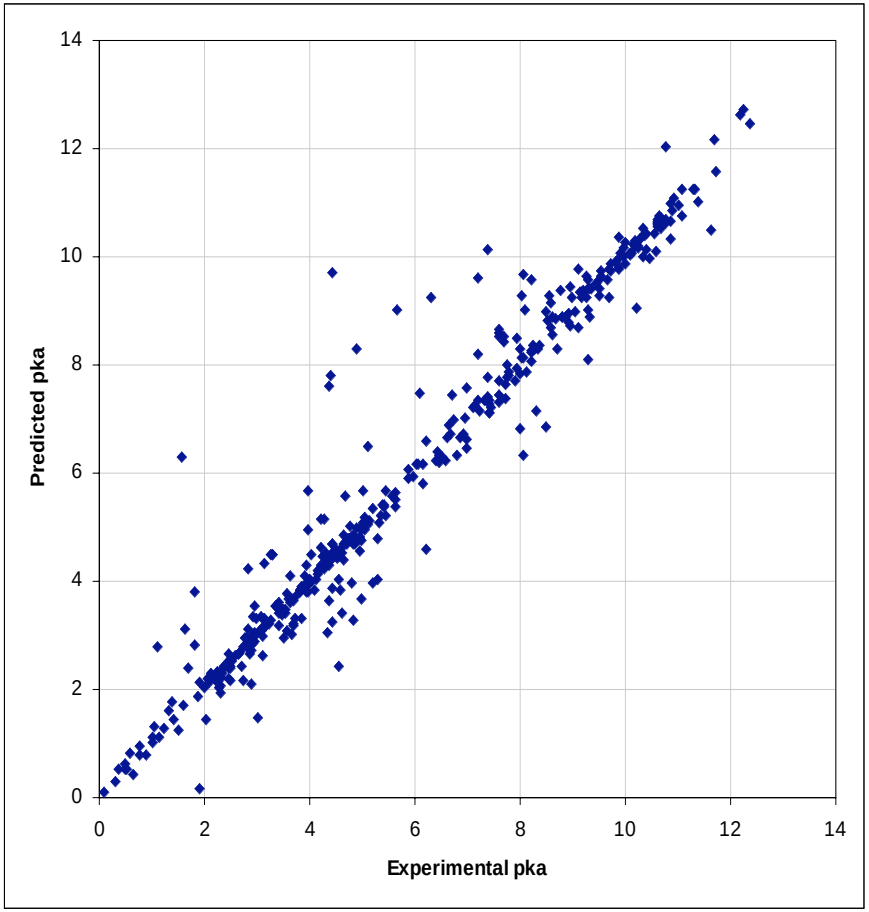
Validation: ACD/Labs

ACD/Labs model:

- Expensive
- Limited rights

Statistics:

- N = 454 (13 err.)
- MAE = 0.34
- RMSE = 0.70



Validation: Marvin

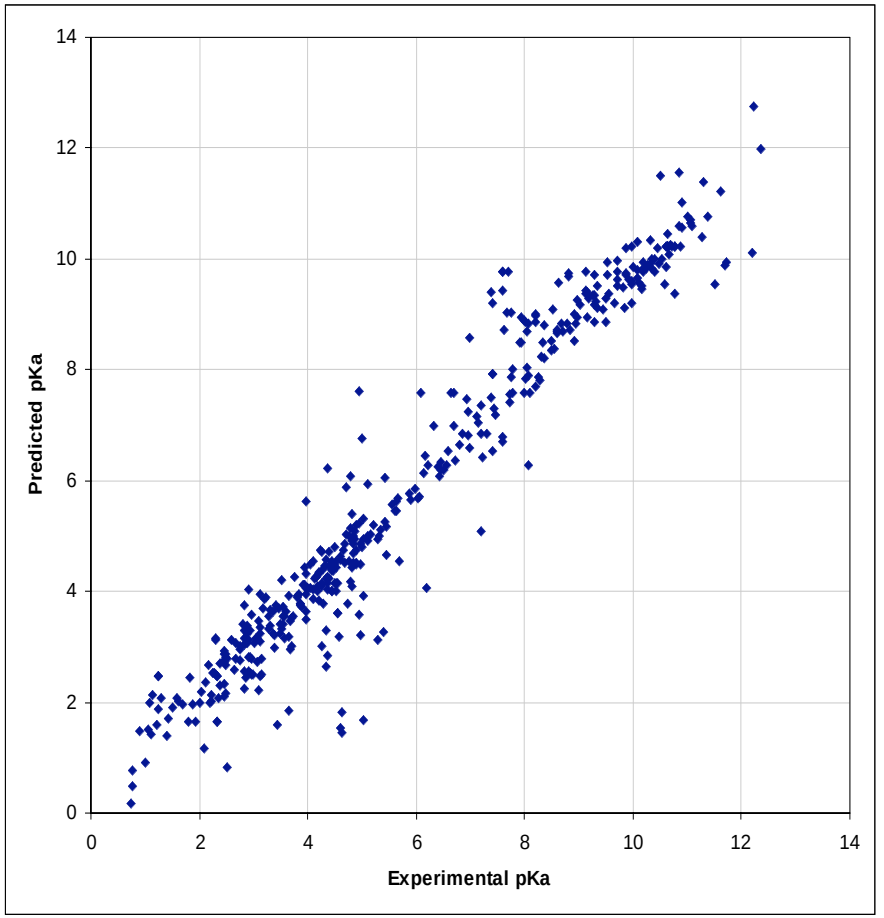
CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Marvin model:

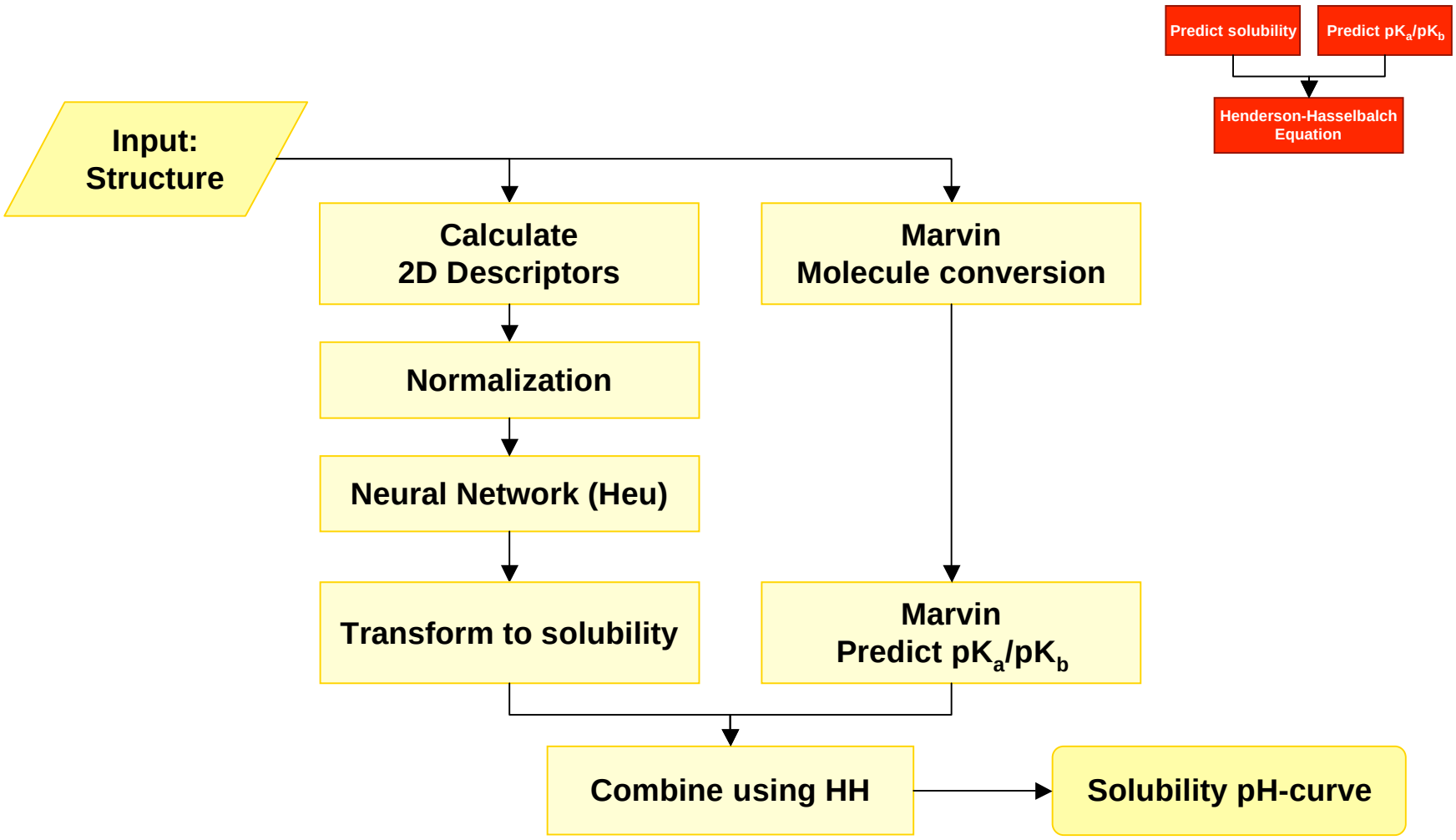
- Free for academia
- Accessible Java-classes

Statistics:

- N = 458 (9 err.)
- MAE = 0.48
- RMSE = 0.71



Combined model



Experimental Solubility Curves

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Source: literature (~70 drugs)

Filter:

- Exact solubility values
- Experimental pK_a and S_0
- No co-solvents

Result:

- 27 experimental solubility curves

Validation of Theory and Prediction

Henderson-Hasselbalch relationship:

$$LogS = LogS_0 + Log\left(1 + 10^{pH - pK_a}\right)$$

- Validation of Theoretical model:
  with **experimental parameters**

- Validation of Prediction model:
  with **predicted parameters**

Validation with Ibuprofen

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Experimental:

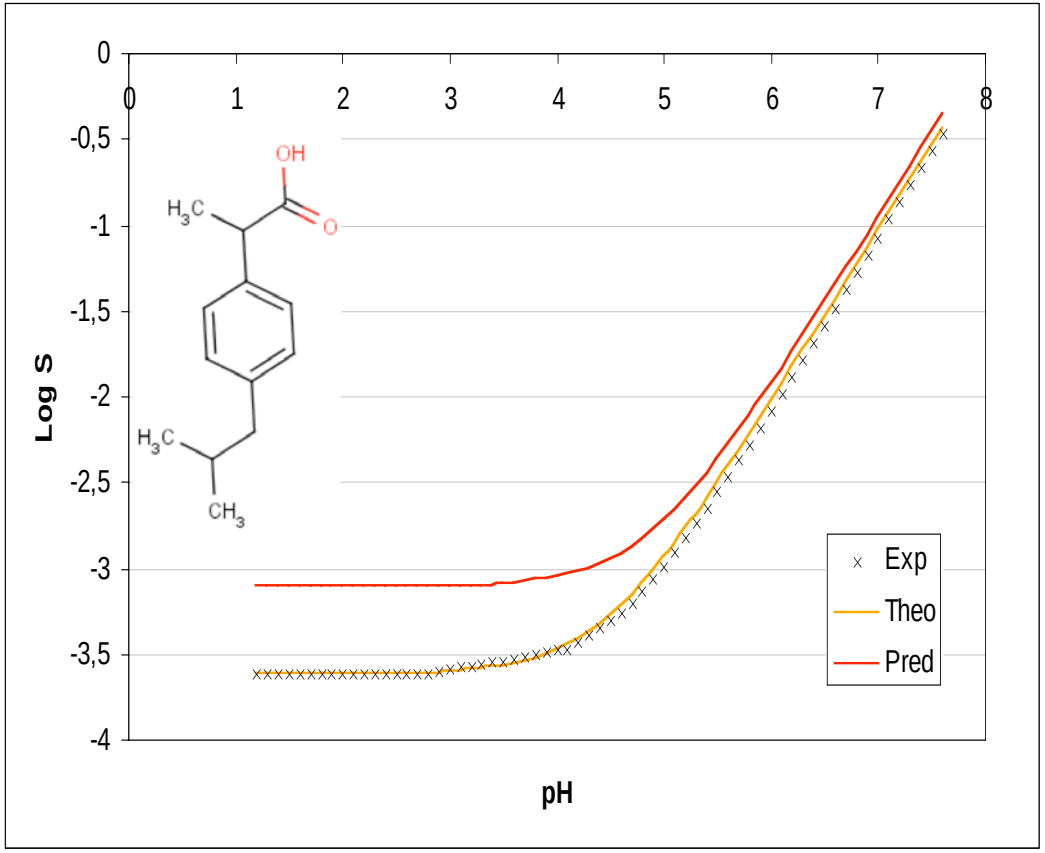
- pK_a = 4.42
- Log S_0 = -3.62

Predicted:

- pK_a = 4.85
- Log S_0 = -3.11

Statistics

- $RMSE_{theo}$ = 0.03
- $RMSE_{pred}$ = 0.37

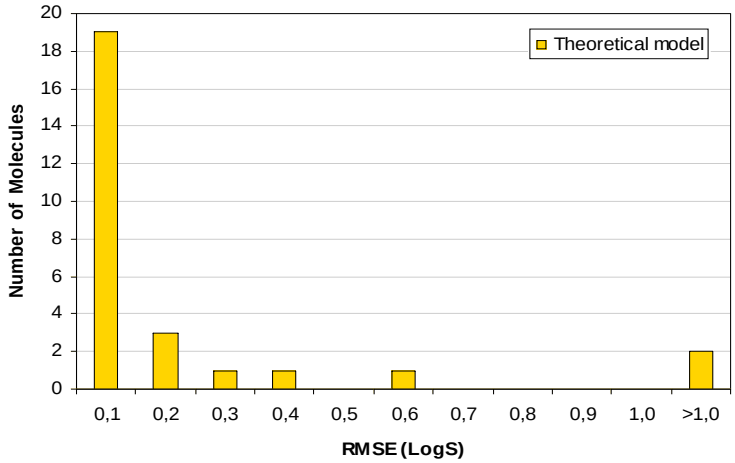


Validation of combined model

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

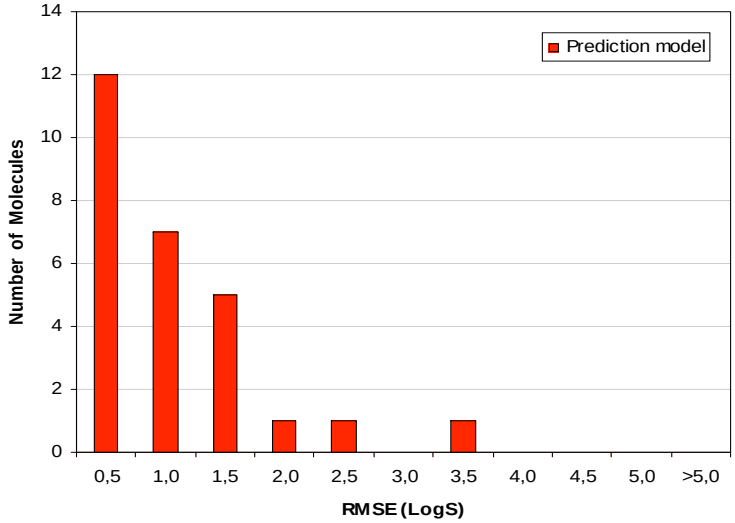
Theory vs. Exp.:

- RMSE = 0.25 (27 drugs)
- RMSE = 0.09 (25 drugs)

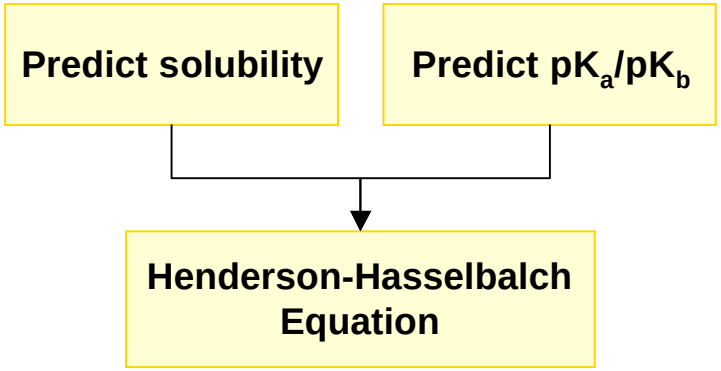


Prediction vs. Exp.:

- RMSE = 0.88 (27 drugs)
- RMSE = 0.80 (26 drugs)



Summary



This model seems adequate for prediction of pH-dependent solubility

Good pK_a/pKb predictions enable us to to predict pH-dependent solubilities without increasing the error significantly

Acknowledgements

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Supervisors:

- Irene Kouskoumvekaki
- Svava Ósk Jónsdóttir
- Flemming Steen Jørgensen (DFU)
- Søren Brunak

Others:

- Thomas Pontén
- Olivier Taboureau
- Jens Pontoppidan

Thanks… Questions?!