Spinning up new predictors: NMR and aqueous solubility

Csaba Fábri

ChemAxon

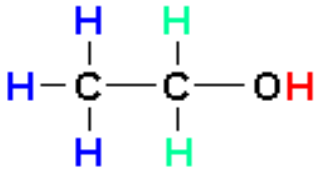
Solutions for Cheminformatics

Agenda

- Nuclear magnetic resonance (NMR) spectroscopy
  - Introduction of NMR spectroscopy
  - Prediction of spectrum parameters
  - Validation of the chemical shift prediction model
  - NMR Predictor demonstration
  - NMR future plans
- Solubility (development has been started)
  - Brief summary of the solubility model
  - Solubility Predictor demonstration

NMR introduction: ethanol ^{1}H NMR

Ethanol



Chemaxon's novel NMR Predictor overview

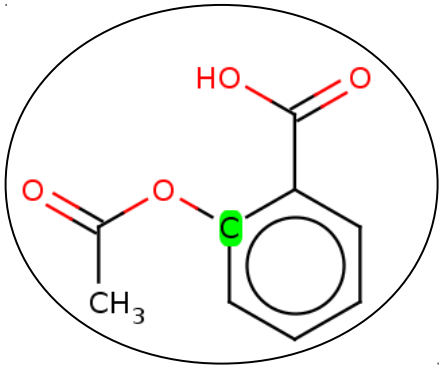
- Fast and accurate ^{1}H and ^{13}C NMR prediction for standard organic molecules
- Easy-to-use graphical user interface
- Comparison of experimental and predicted spectra
- Ready for web (MarvinSketch applet)
- Another useful predictor in Chemaxon's Discovery Toolkit

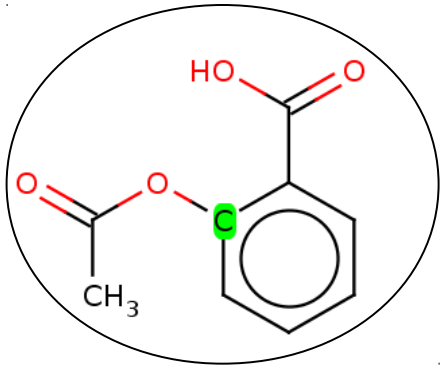
Chemical shift prediction overview

- Aim: fast and accurate prediction of chemical shifts
- First approach: HOSE code technology
- Second approach: QSPR
- Mixed model: combination of HOSE and QSPR

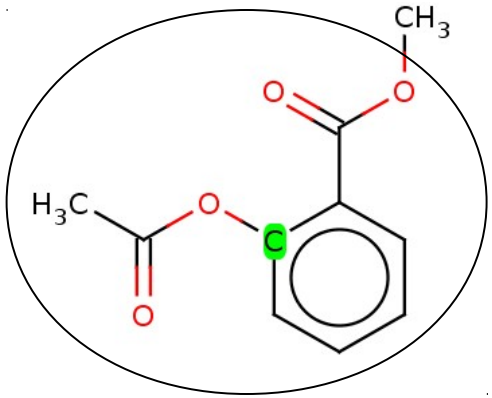
HOSE code-based chemical shift prediction

query: ? ppm





1. hit: 150.4 ppm

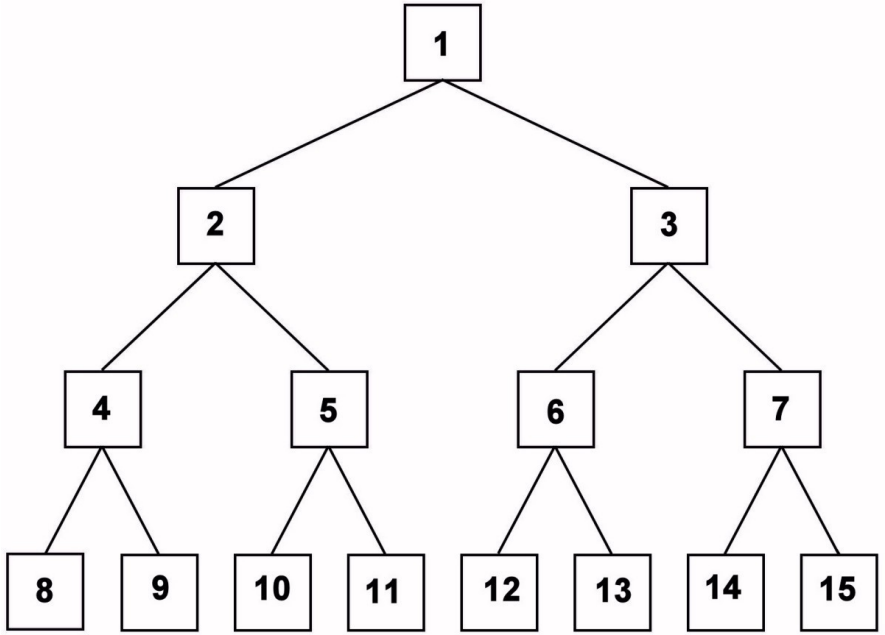


2. hit: 150.8 ppm

C-3;*C*CO(*CC,*C,C/*C,=OO,*&,=OC/)

Decision tree-based chemical shift prediction

- Physicochemical and topological descriptors
- Model fitting: Decision tree with multilinear least-squares regression (MLR)
- NMRShiftDB (http://nmrshiftdb.nmr.uni-koeln.de)



Chemical shift descriptor set

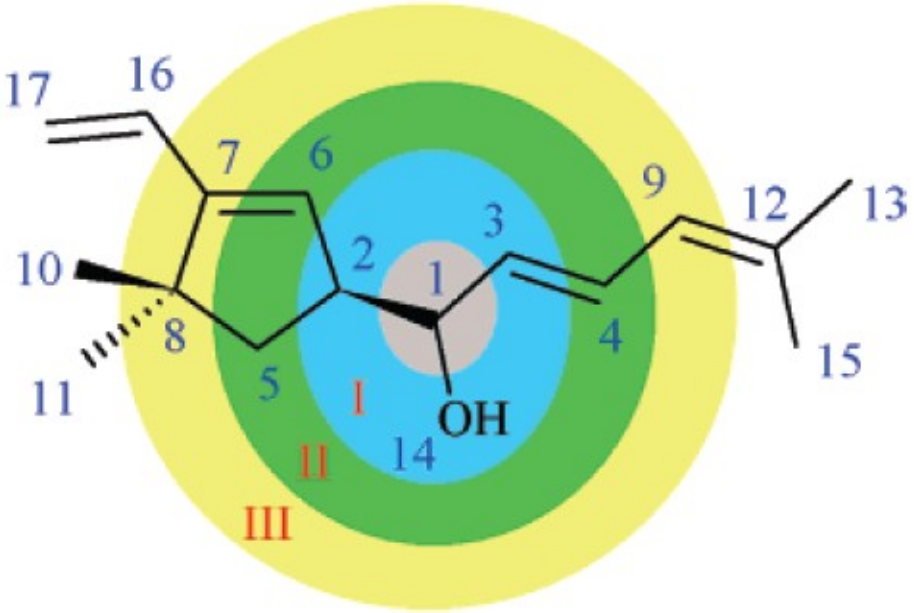


Figure 2. Encoding of the atom environment. The three nearest spheres are marked as cyan, green, and yellow circles, and the Roman numerals denoting the number of each sphere are shown. Atom no. 1 is the central atom. The blue numbers are assigned arbitrarily and serve as references (see the text for more details).

J. Chem. Inf. Comput. Sci. **40**, 1169 (2000)

Mixed chemical shift model

HOSE search is started with an initial radius of 6,
radius is consecutively decreased until a threshold is reached

Are there any hits with a radius greater equal than a threshold?
(threshold: 3 for ^{13}C NMR, 4 for ^{1}H NMR)

yes

Hits with the largest radius are averaged

no

Decision tree-based model is invoked

Chemical shift model validation: 5.10 vs 5.11.1

^{13}C NMR	5.10	5.11.1
< 5 ppm	77 %	88 %
< 10 ppm	93 %	97 %

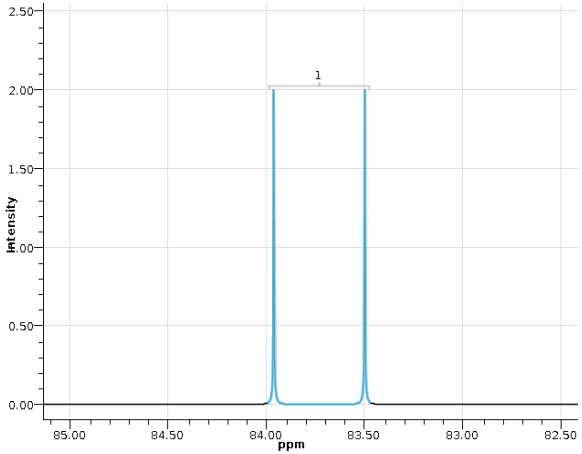
^{1}H NMR	5.10	5.11.1
< 0.5 ppm	81 %	91 %
< 1.0 ppm	93 %	98 %

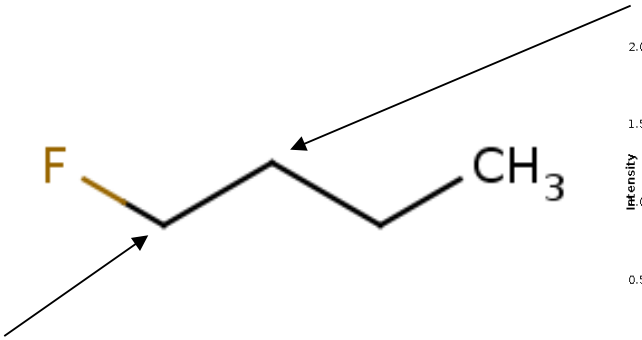
Coupling constant prediction

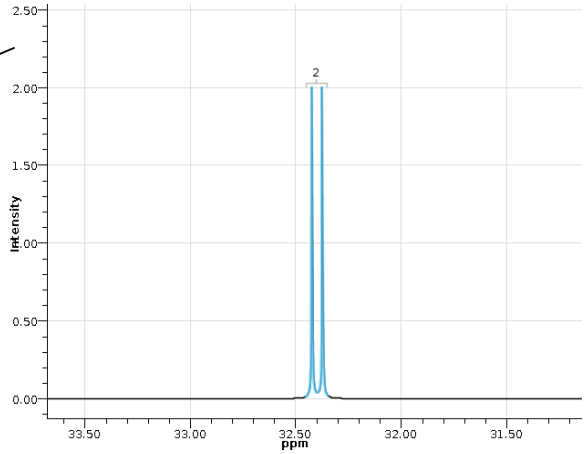
HH coupling (^{3}J)

CF coupling (^{1}J and ^{2}J, new in 5.11)

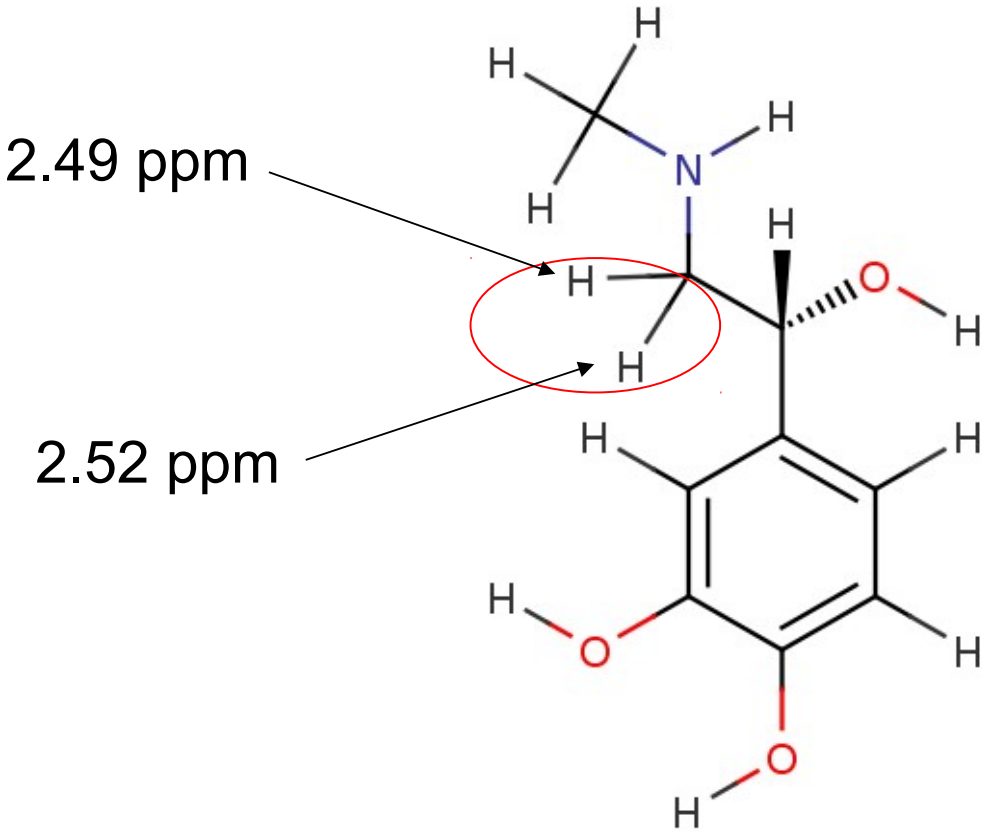
HF coupling (^{2}J, ^{3}J and ^{4}J, new in 5.11)



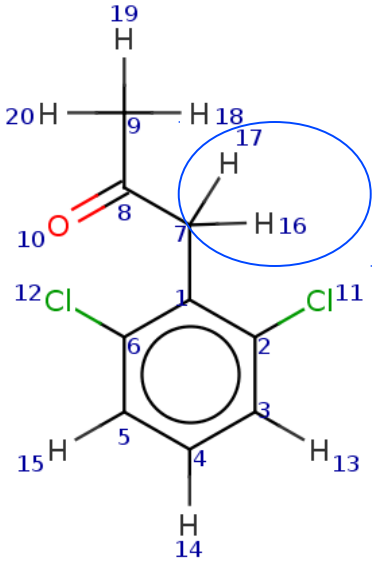




Diastereotopic protons are distinguished



Example: mixed chemical shift model

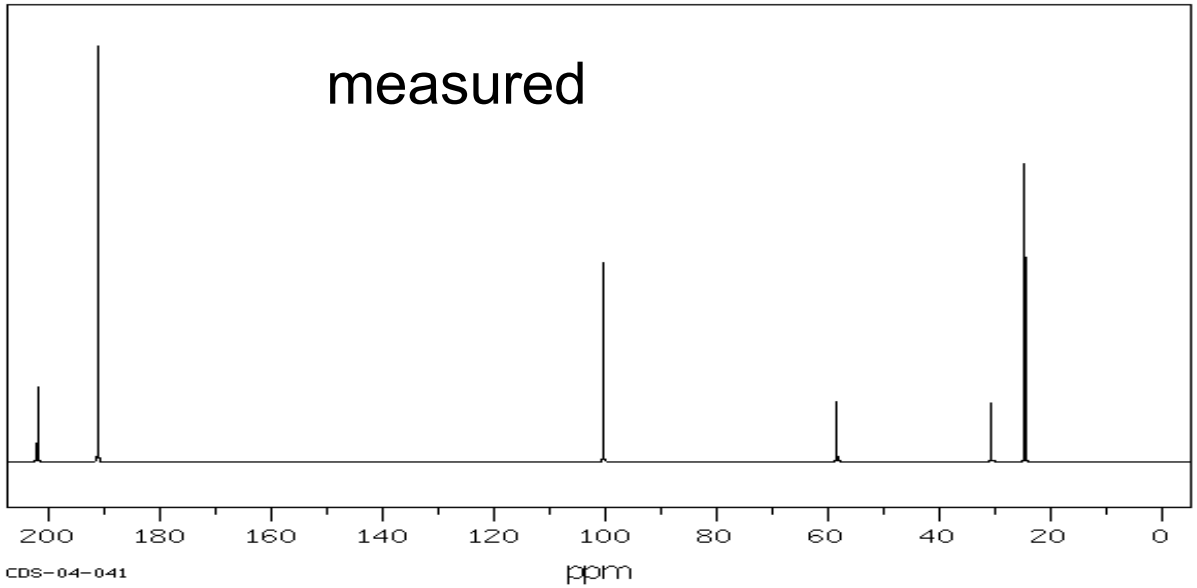


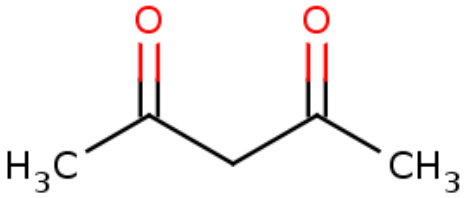
^{1}H NMR chemical shifts (ppm)

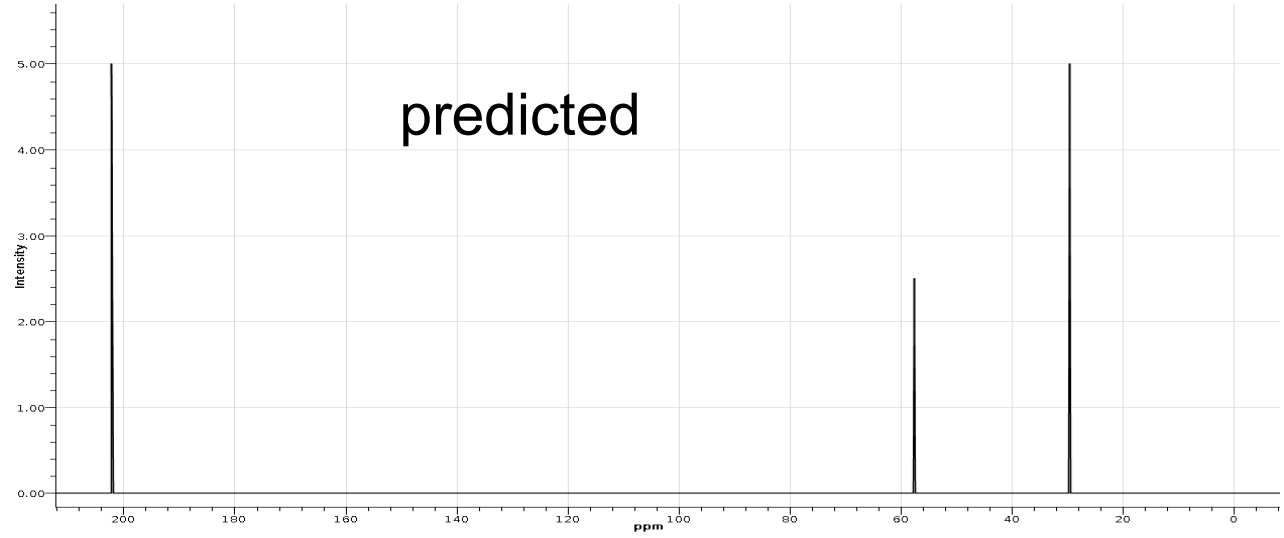
Atom	Pred.	Exp.	HOSE radius	HOSE hits
18,19, 20	2.20	2.24	5	4
16,17	3.89	4.09	-	-
14	7.17	7.16	6	2
13,15	7.33	7.32	5	2

Decision tree-based shift model is used if there are no HOSE hits!

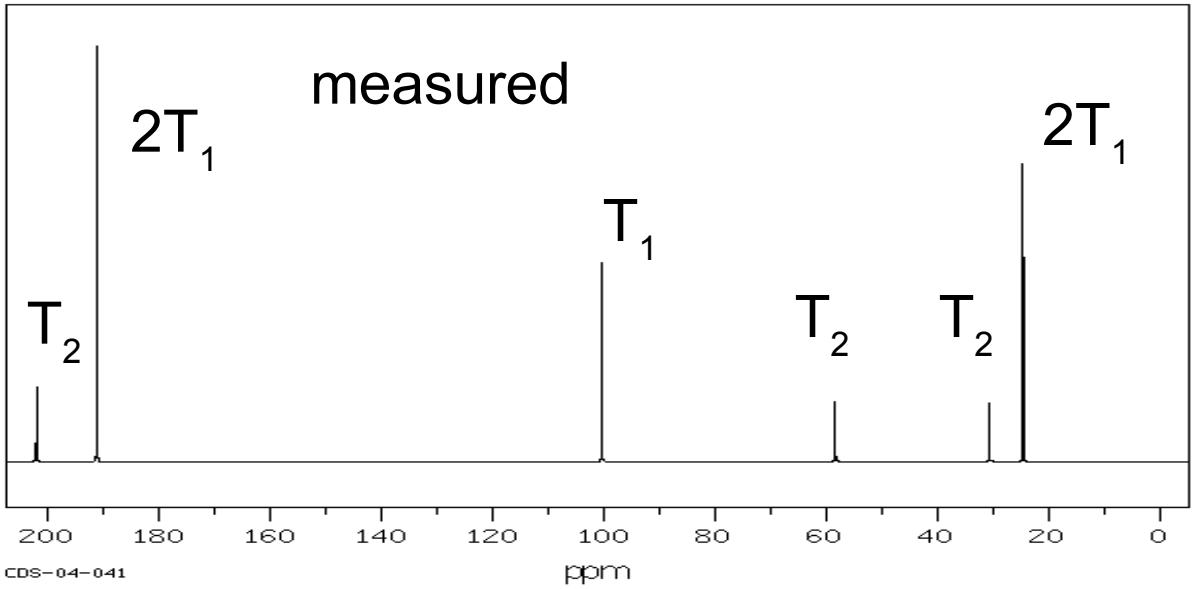
Case study: 2,4-pentanedione ^{13}C NMR



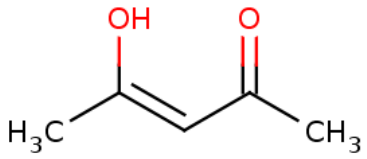


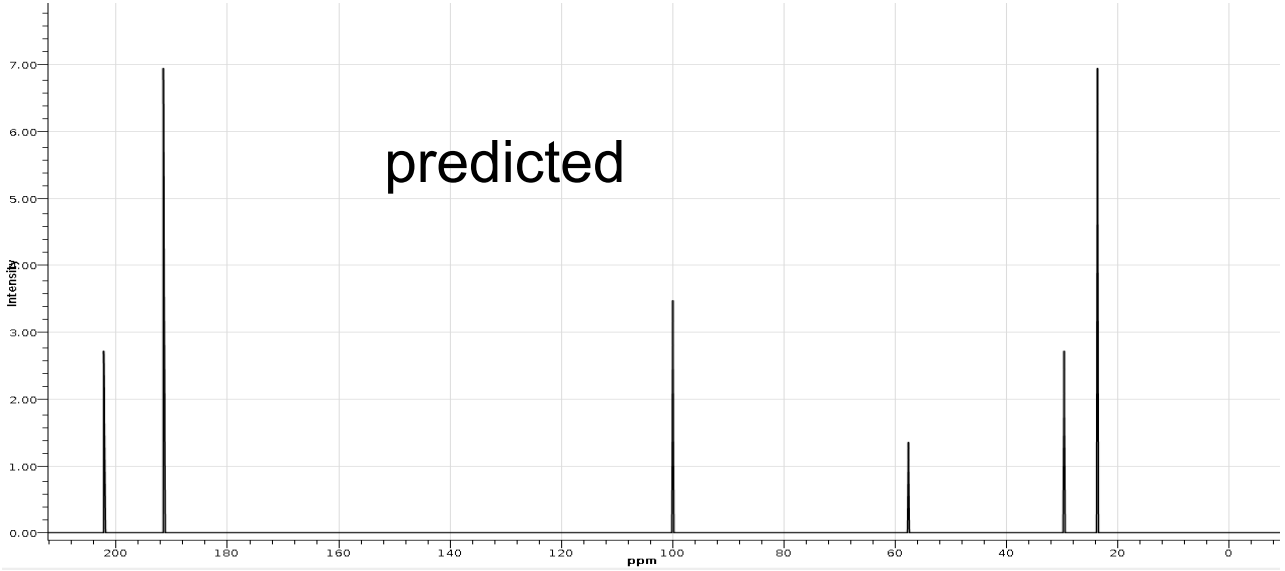


Tautomerization effect

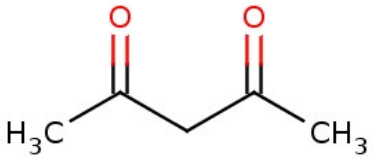


T_1: 72 %





T_2: 28 %



NMR future plans

- Extend the NMR chemical shift training set
- Support more NMR data formats for import/export
- NMR prediction on www.chemicalize.org
- Give the user the possibility to train the NMR chemical shift model

Solubility Predictor

- Brief summary:
  - QSPR model
  - Physicochemical and topological descriptors
  - Training for experimental intrinsic solubilities
- List of relevant features:
  - Prediction of intrinsic solubilities
  - Prediction of pH-dependent solubility curves
- Demonstration

Summary

- Fast and accurate ^{1}H and ^{13}C NMR spectrum prediction for standard organic molecules
- Development of the Solubility Predictor has been started
- Any feedback is appreciated:

calculators-support@chemaxon.com

Acknowledgements

Calculations team members:
Zsolt Mohácsi, Imre Barna, Rita Veréb, István Rábel,
József Szegezdi, Viktória Pálfi

NMR consultant:
Csaba Peltz

NMR icons:
János Papdeák

Application scientists:
Iván Solt, Attila Szabó, Krisztián Niesz

NMRShiftDB:
Stefan Kuhn