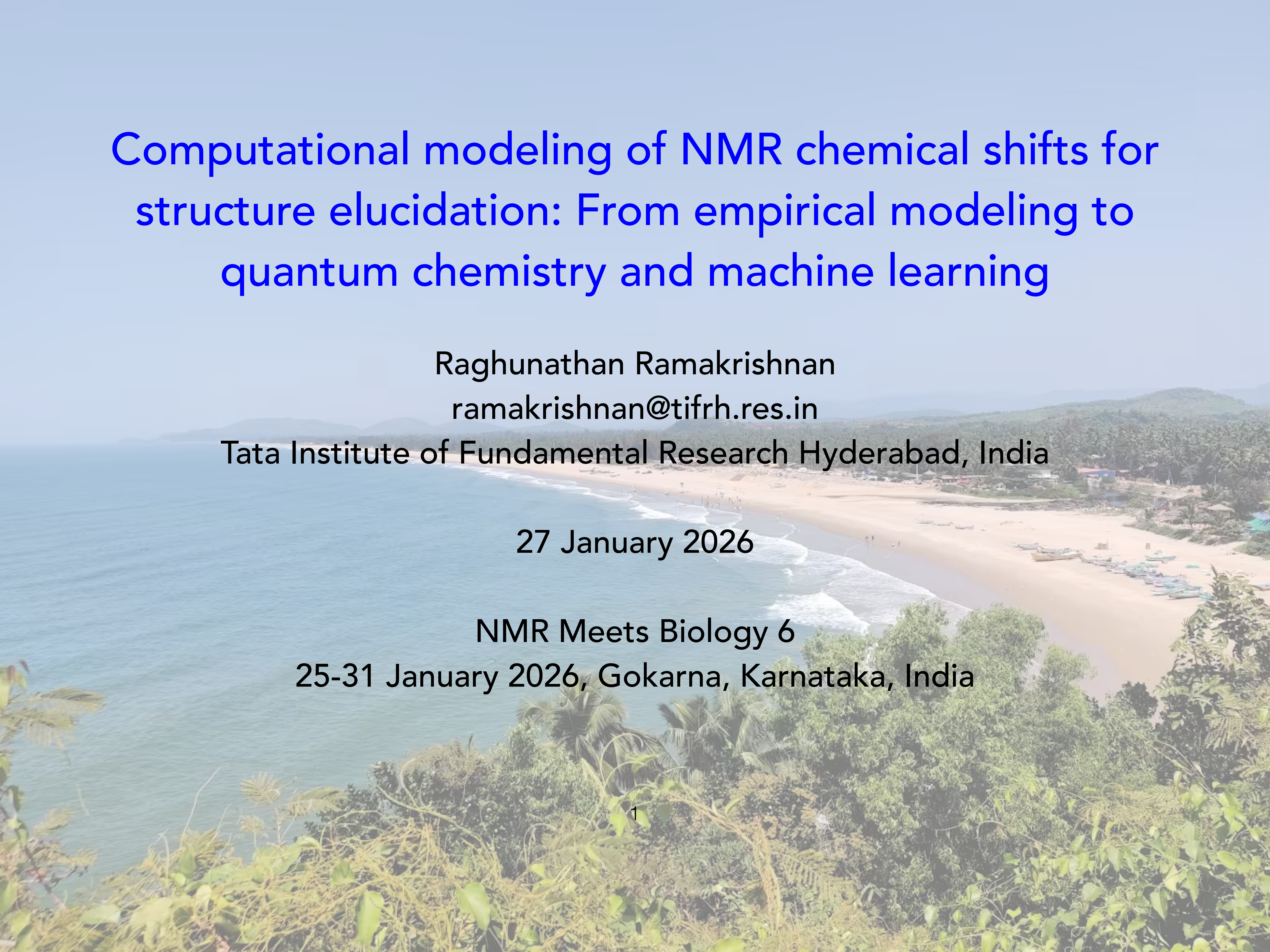


Computational modeling of NMR chemical shifts for structure elucidation: From empirical modeling to quantum chemistry and machine learning

Raghunathan Ramakrishnan
ramakrishnan@tifrh.res.in
Tata Institute of Fundamental Research Hyderabad, India

27 January 2026

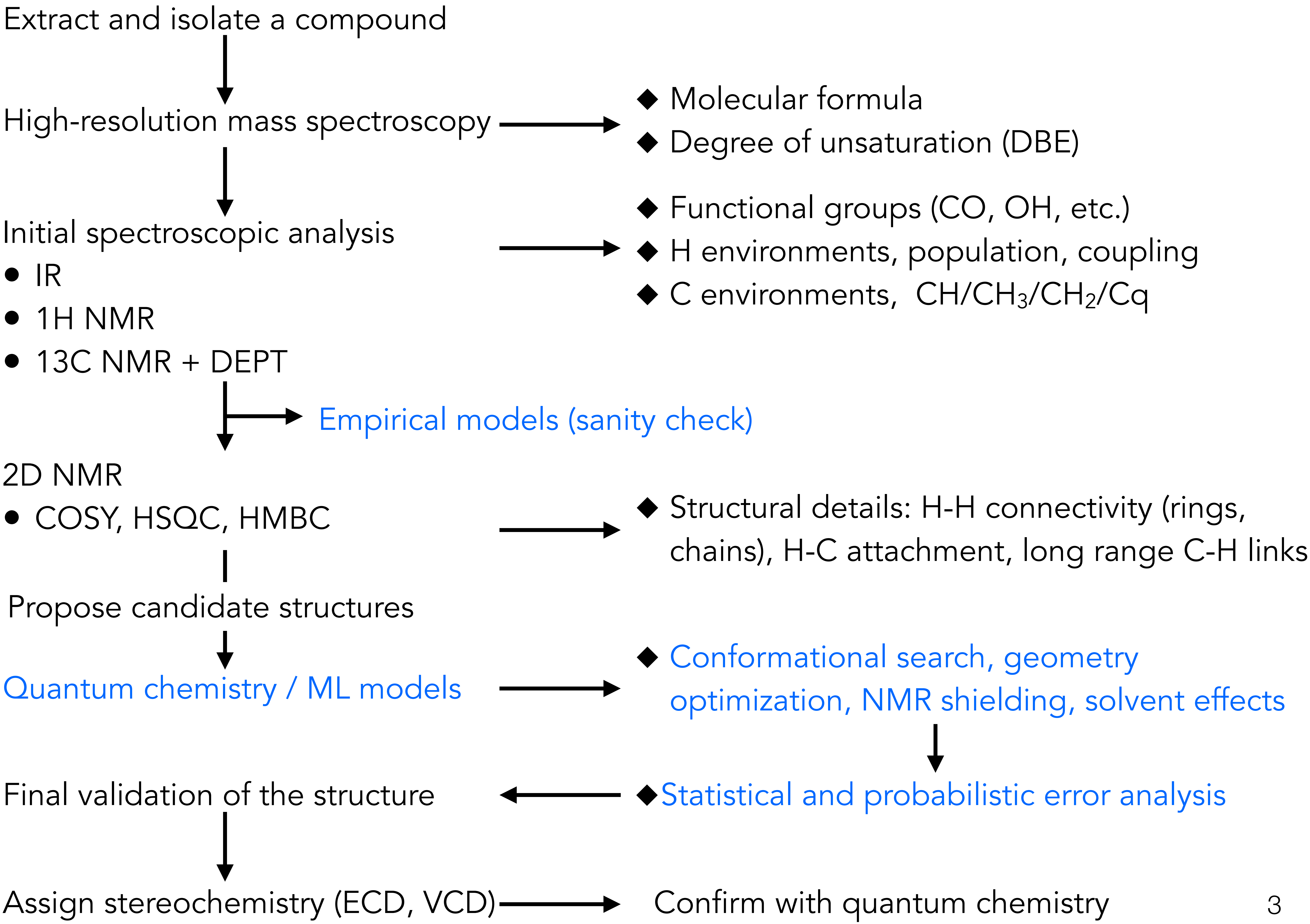
NMR Meets Biology 6
25-31 January 2026, Gokarna, Karnataka, India

Some concepts behind the methods we will use tomorrow during the hands-on session

- Empirical additivity-based models
- Quantum chemistry
- Machine learning

For students, ORCA quantum chemistry software can be installed on laptop after the talk.
Don't Leave!

(From synthesis to) Structure elucidation workflow



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- ● IR
- ● 1H NMR
- ● 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, CH/CH$_3$/CH$_2$/Cq

Empirical models (sanity check)

2D NMR

- ● COSY, HSQC, HMBC

- ◆ Structural details: H-H connectivity (rings, chains), H-C attachment, long range C-H links

Propose candidate structures

Quantum chemistry / ML models

- ◆ Conformational search, geometry optimization, NMR shielding, solvent effects

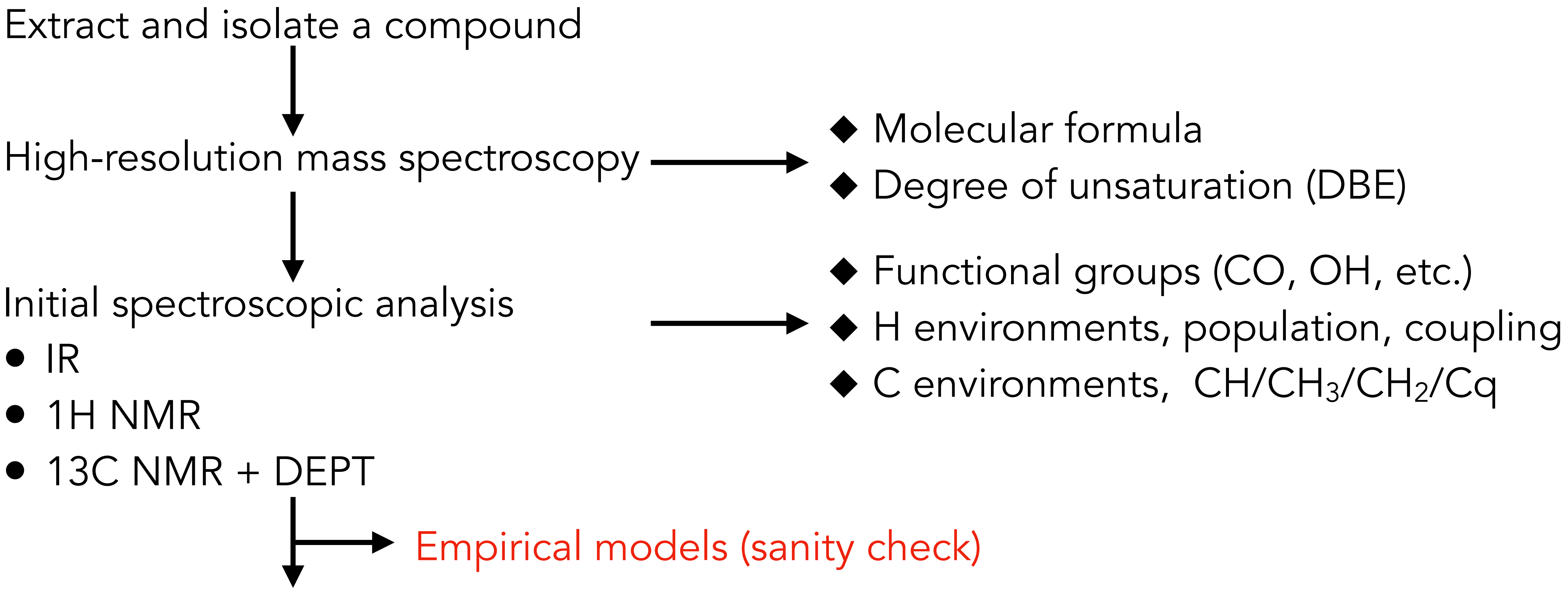
- ◆ Statistical and probabilistic error analysis

Final validation of the structure

Assign stereochemistry (ECD, VCD)

Confirm with quantum chemistry

Empirical models: power and ambiguity



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

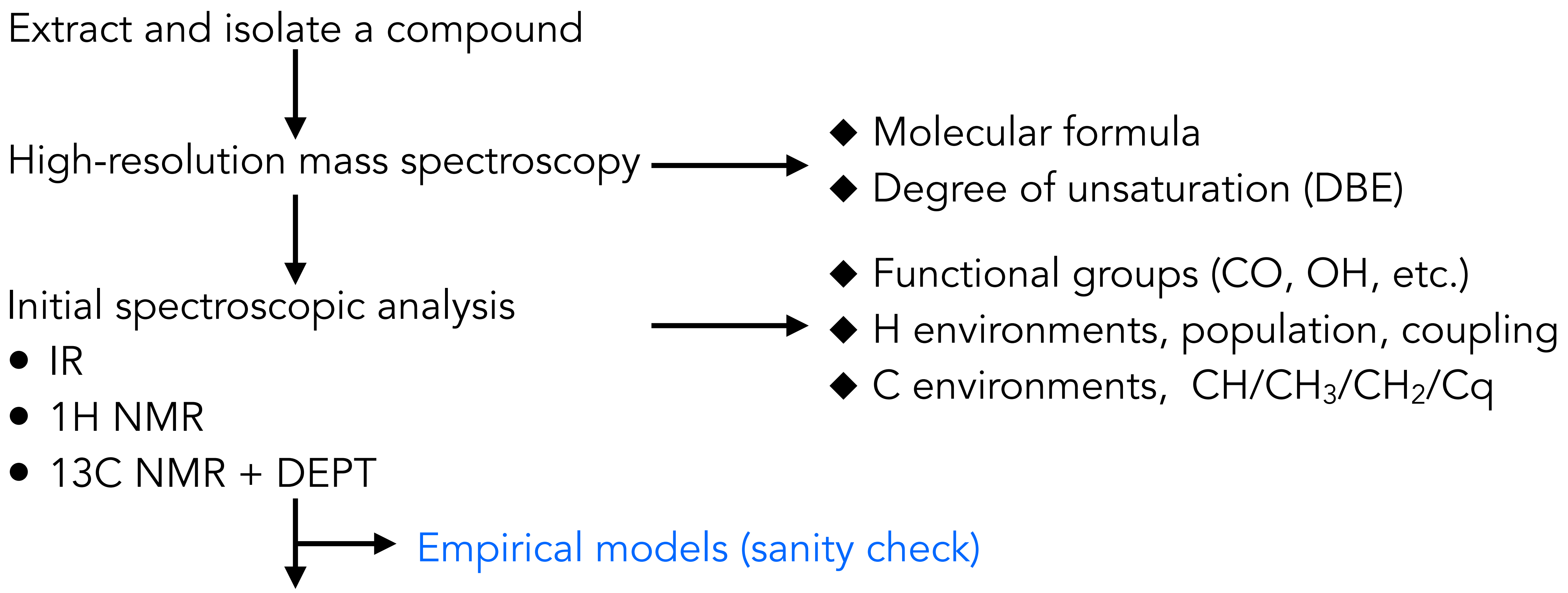
Initial spectroscopic analysis

- IR
- 1H NMR
- 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, CH/CH$_3$/CH$_2$/Cq

Empirical models (sanity check)

Empirical models: power and ambiguity



Extract and isolate a compound

High-resolution mass spectroscopy

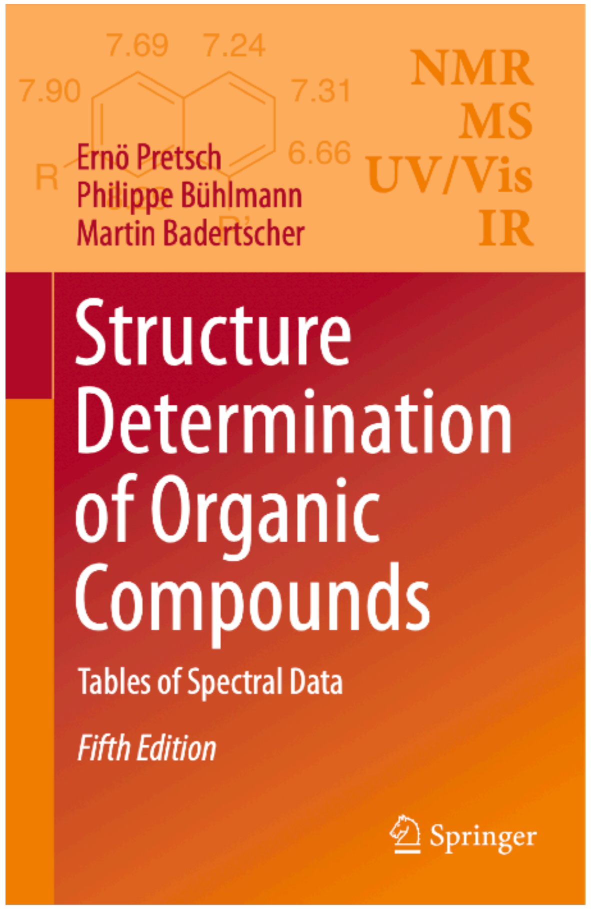
- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- IR
- 1H NMR
- 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, $CH/CH_3/CH_2/Cq$

Empirical models (sanity check)



Ernö Pretsch
Philippe Bühlmann
Martin Badertscher

NMR
MS
UV/Vis
IR

Structure Determination of Organic Compounds

Tables of Spectral Data

Fifth Edition

Springer

Parameters for H and C environments

No dependence on dihedral angles, stereochemistry, or 3D structure

Empirical additivity models for ^{1}H chemical shifts

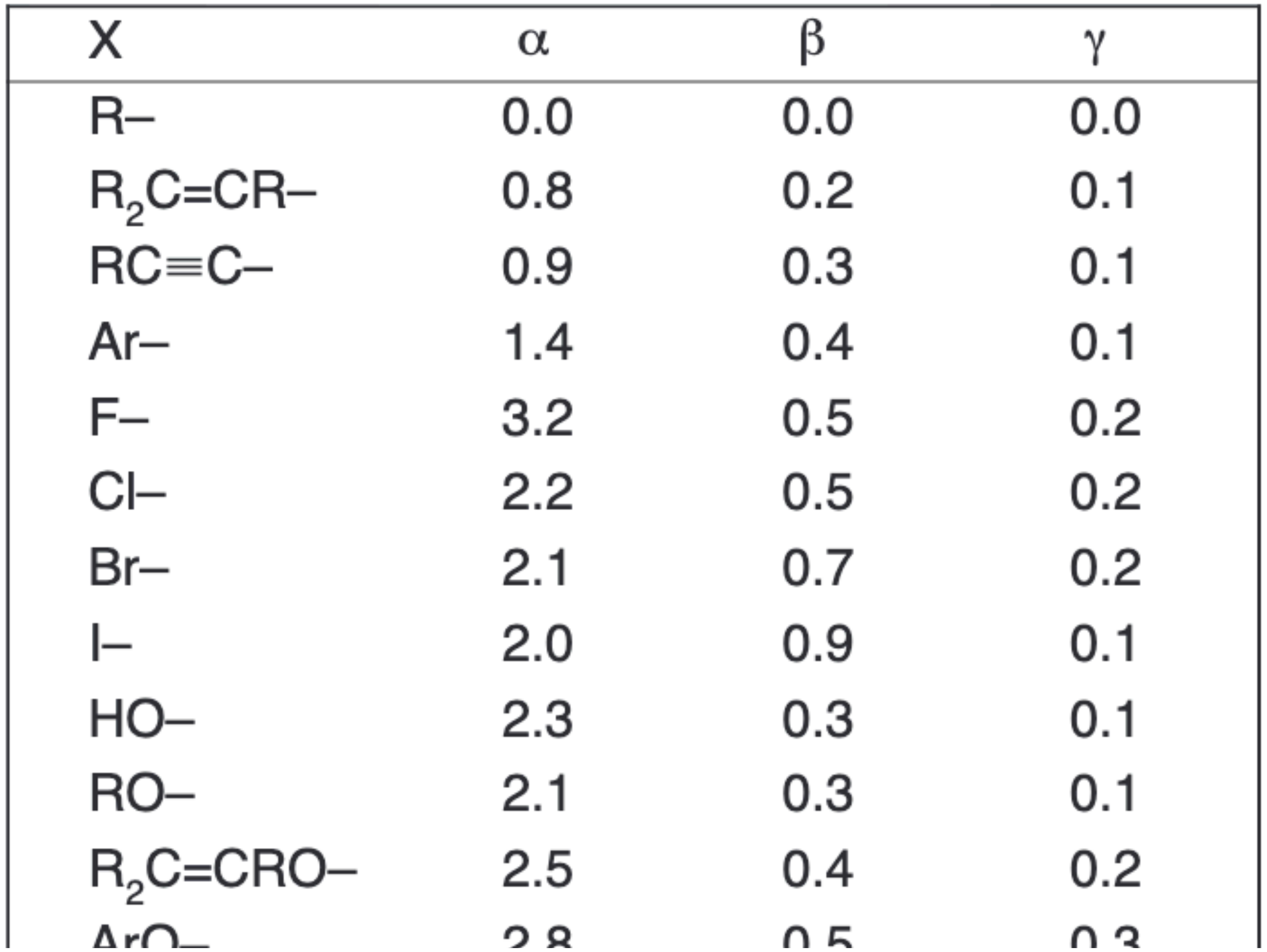
methyl $\delta_{CH_3} = 0.9 + \sum(\beta + \gamma)$ $\quad CH_3-\underset{\beta}{C}-\underset{\gamma}{C}-$

methylene $\delta CH_2 = 1.2 + \sum(\alpha + \beta + \gamma)$

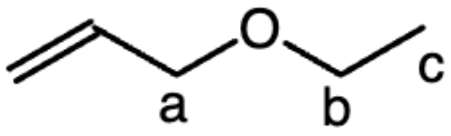
$$-\underset{\alpha}{CH_2}-\underset{\beta}{C}-\underset{\gamma}{C}-$$

methine $\delta CH = 1.5 + \sum(\alpha + \beta + \gamma)$

$$H-\underset{\alpha}{C}-\underset{\beta}{C}-\underset{\gamma}{C}-$$



X	α	β	γ
R–	0.0	0.0	0.0
R_2C=CR–	0.8	0.2	0.1
RC≡C–	0.9	0.3	0.1
Ar–	1.4	0.4	0.1
F–	3.2	0.5	0.2
Cl–	2.2	0.5	0.2
Br–	2.1	0.7	0.2
I–	2.0	0.9	0.1
HO–	2.3	0.3	0.1
RO–	2.1	0.3	0.1
R_2C=CRO–	2.5	0.4	0.2
ArO–	2.8	0.5	0.3



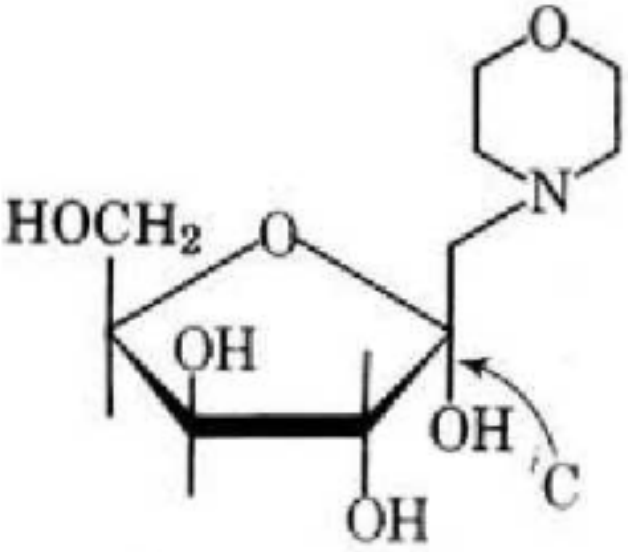
	a	b	c
Actual δ (ppm)	4.0	3.5	1.2
Calculated δ (ppm)	1.2	1.2	0.9
C=C–	0.8	0.2	0.1
R–O–	2.1	2.1	0.3
Total	4.1	3.5	1.3

P. S. Beauchamp, R. Marquez, *J. Chem. Educ.*, 74, 1483 (1997).

Empirical additivity models for ^{13}C chemical shifts

A Short Set of ^{13}C-NMR Correlation Tables

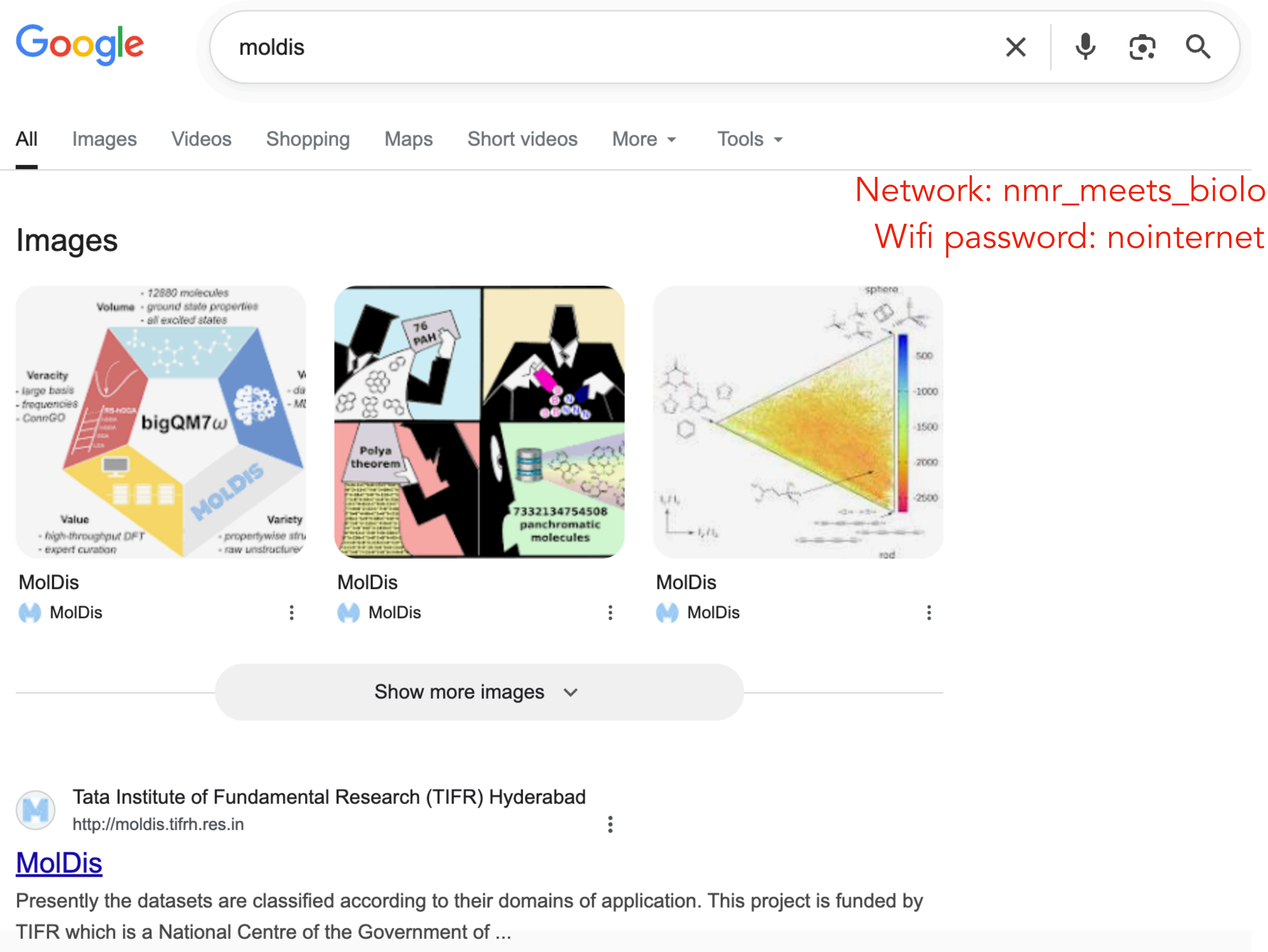
D. W. Brown
University of Bath, Bath, BA2 7AY, Avon, England



$$
\begin{aligned}
\delta &= -2.3 + (2\alpha^1 + 2\alpha^2 + 2\beta^1 + \beta^2 + \beta^3 + 5\gamma^1 + \gamma^2) \\
&\quad + (4° \rightarrow 3° + 4° \rightarrow 2° + 4° \rightarrow 2° + 4° \rightarrow 1°) \\
&= -2.3(18.2 + 98.0 + 18.4 + 10.1 + 11.3 - 12.5 - 6.2) \\
&\quad + (-15.0 - 8.4 - 8.4 - 1.5) = 101.7 \text{ (observed 97.6)}
\end{aligned}
$$

D. W. Brown, *J. Chem. Educ.*, 62, 209 (1985).

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab



Tata Institute of Fundamental Research (TIFR) Hyderabad
http://moldis.tifrh.res.in

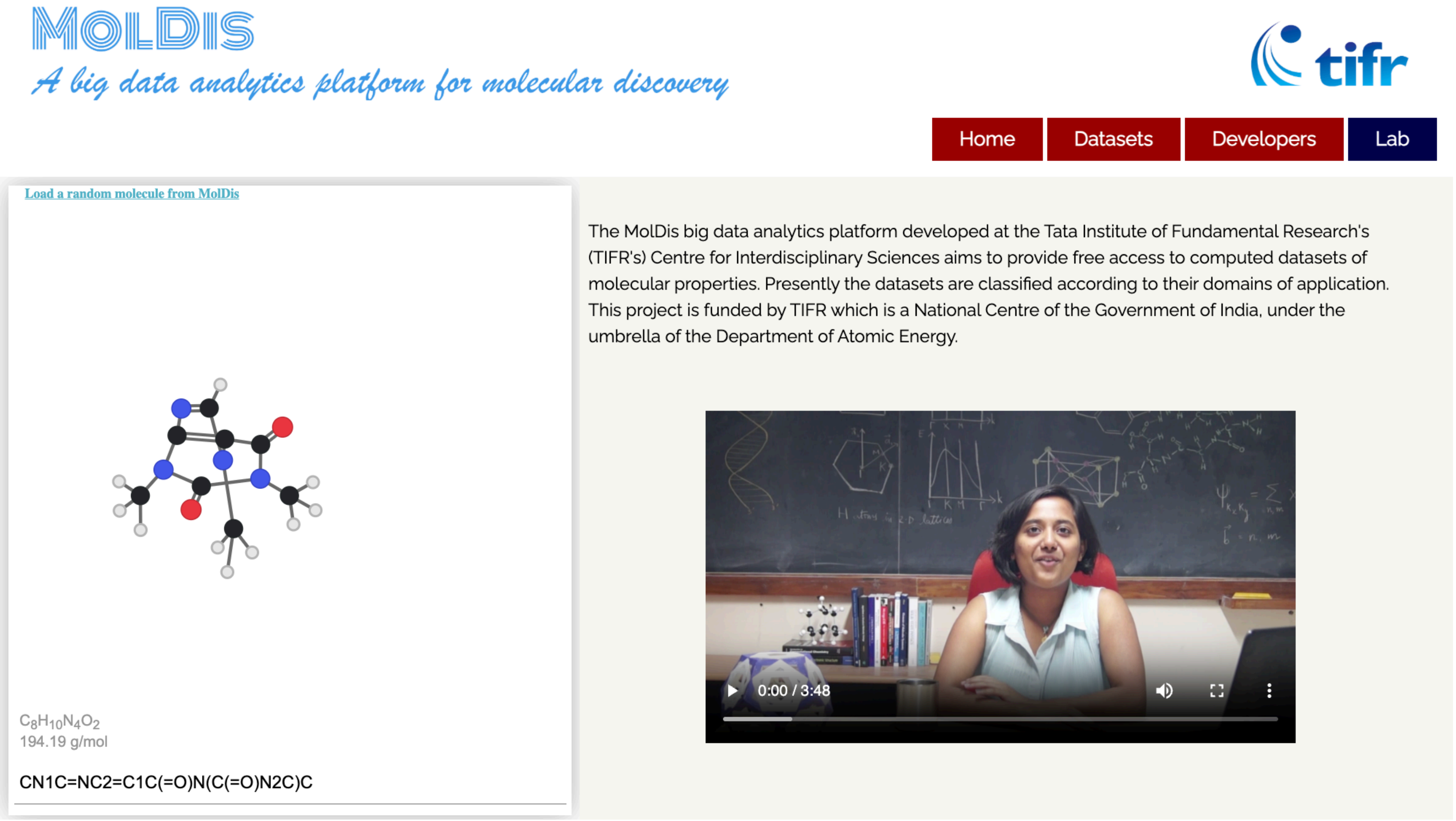
MolDis

Presently the datasets are classified according to their domains of application. This project is funded by TIFR which is a National Centre of the Government of ...

Network: nmr_meets_biology

Wifi password: nointernet1

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab



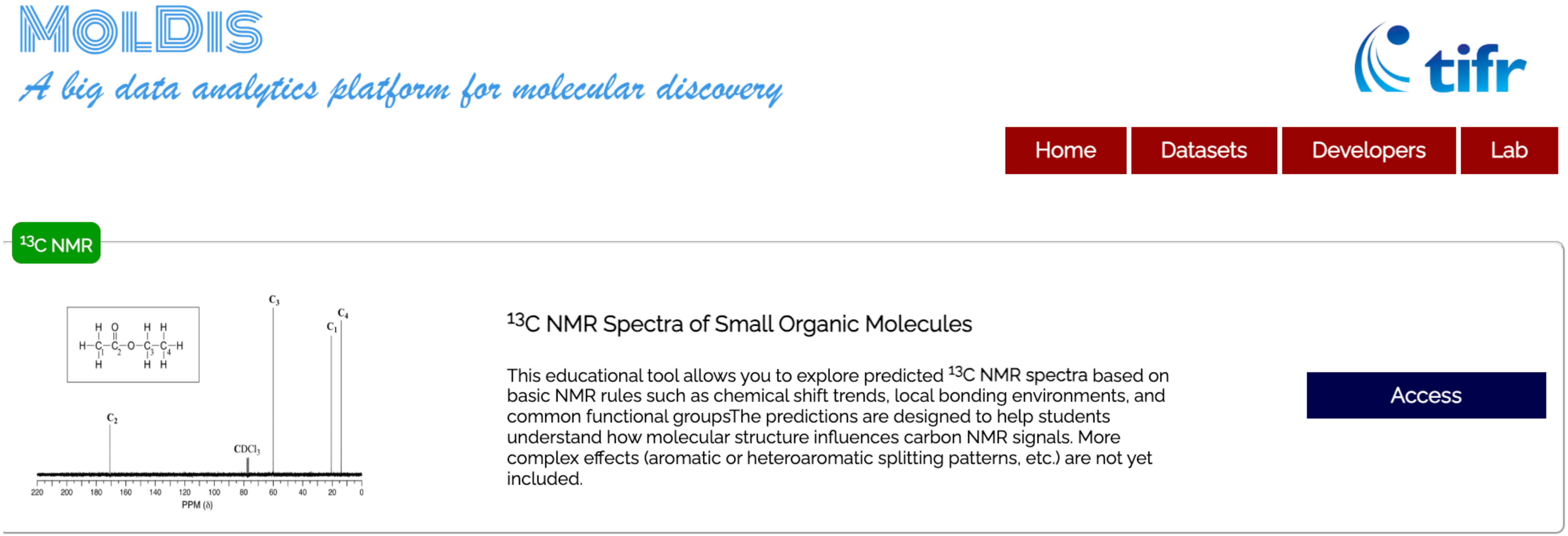
Load a random molecule from MolDis

$C_8H_{10}N_4O_2$
194.19 g/mol

CN1C=NC2=C1C(=O)N(C(=O)N2C)C

The MolDis big data analytics platform developed at the Tata Institute of Fundamental Research's (TIFR's) Centre for Interdisciplinary Sciences aims to provide free access to computed datasets of molecular properties. Presently the datasets are classified according to their domains of application. This project is funded by TIFR which is a National Centre of the Government of India, under the umbrella of the Department of Atomic Energy.

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab



^{13}C NMR Spectra of Small Organic Molecules

This educational tool allows you to explore predicted ^{13}C NMR spectra based on basic NMR rules such as chemical shift trends, local bonding environments, and common functional groupsThe predictions are designed to help students understand how molecular structure influences carbon NMR signals. More complex effects (aromatic or heteroaromatic splitting patterns, etc.) are not yet included.

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab

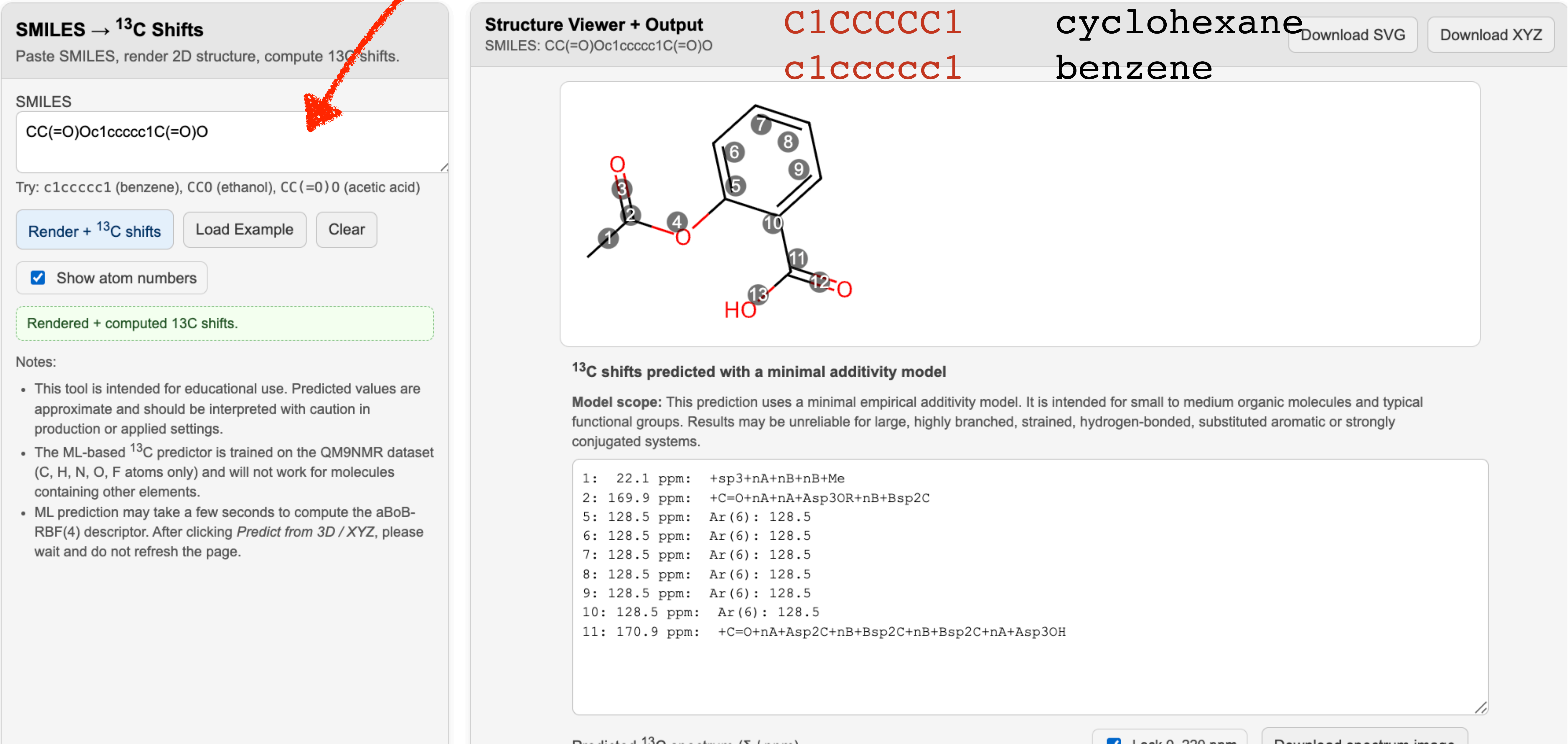
MOLDIS

A big data analytics platform for molecular discovery

SMILES representation

C	methane
CC	ethane
CCCCC	pentane
CC(C)CC	isopentane
CC(C)(C)C	neopentane
C1CC1	cyclopropane
C1CCCCC1	cyclohexane
c1ccccc1	benzene

tifr



SMILES → ^{13}C Shifts

Paste SMILES, render 2D structure, compute 13C shifts.

SMILES

CC(=O)Oc1ccccc1C(=O)O

Try: c1ccccc1 (benzene), CCO (ethanol), CC(=O)O (acetic acid)

☑ Show atom numbers

Rendered + computed 13C shifts.

Notes:

- This tool is intended for educational use. Predicted values are approximate and should be interpreted with caution in production or applied settings.
- The ML-based ^{13}C predictor is trained on the QM9NMR dataset (C, H, N, O, F atoms only) and will not work for molecules containing other elements.
- ML prediction may take a few seconds to compute the aBoB-RBF(4) descriptor. After clicking *Predict from 3D / XYZ*, please wait and do not refresh the page.

Structure Viewer + Output

SMILES: CC(=O)Oc1ccccc1C(=O)O

^{13}C shifts predicted with a minimal additivity model

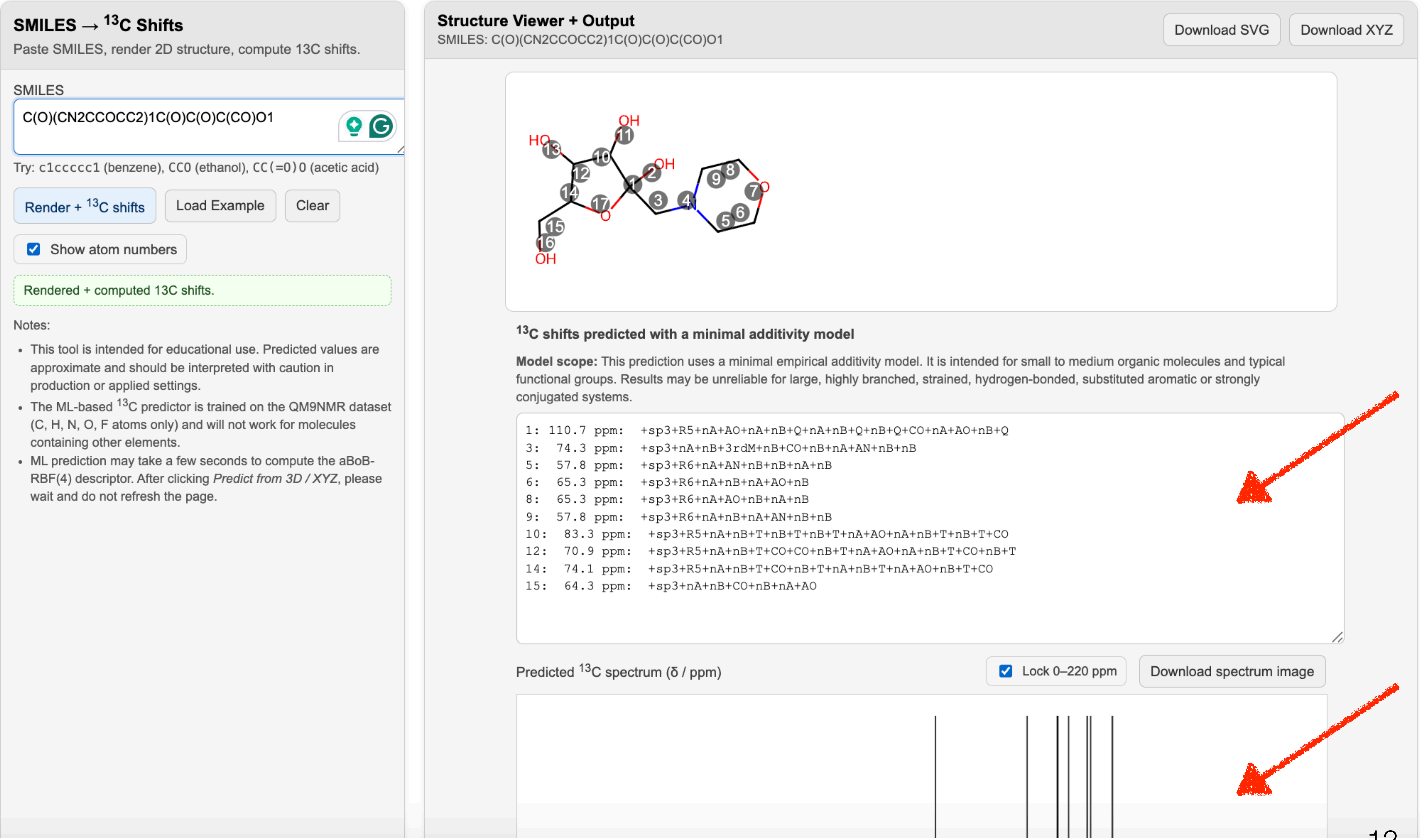
Model scope: This prediction uses a minimal empirical additivity model. It is intended for small to medium organic molecules and typical functional groups. Results may be unreliable for large, highly branched, strained, hydrogen-bonded, substituted aromatic or strongly conjugated systems.

```
1:  22.1 ppm:  +sp3+nA+nB+nB+Me
2: 169.9 ppm:  +C=O+nA+nA+Asp3OR+nB+Bsp2C
5: 128.5 ppm:  Ar(6): 128.5
6: 128.5 ppm:  Ar(6): 128.5
7: 128.5 ppm:  Ar(6): 128.5
8: 128.5 ppm:  Ar(6): 128.5
9: 128.5 ppm:  Ar(6): 128.5
10: 128.5 ppm:  Ar(6): 128.5
11: 170.9 ppm:  +C=O+nA+Asp2C+nB+Bsp2C+nB+Bsp2C+nA+Asp3OH
```

Predicted ^{13}C spectrum (δ / ppm)

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab

A big data analytics platform for molecular discovery



SMILES → ^{13}C Shifts

Paste SMILES, render 2D structure, compute 13C shifts.

SMILES

C(O)(CN2CCOCC2)1C(O)C(O)C(CO)O1

Try: `c1ccccc1` (benzene), `CCO` (ethanol), `CC(=O)O` (acetic acid)

☑ Show atom numbers

Rendered + computed 13C shifts.

Notes:

- This tool is intended for educational use. Predicted values are approximate and should be interpreted with caution in production or applied settings.
- The ML-based ^{13}C predictor is trained on the QM9NMR dataset (C, H, N, O, F atoms only) and will not work for molecules containing other elements.
- ML prediction may take a few seconds to compute the aBoB-RBF(4) descriptor. After clicking *Predict from 3D / XYZ*, please wait and do not refresh the page.

Structure Viewer + Output

SMILES: C(O)(CN2CCOCC2)1C(O)C(O)C(CO)O1

^{13}C shifts predicted with a minimal additivity model

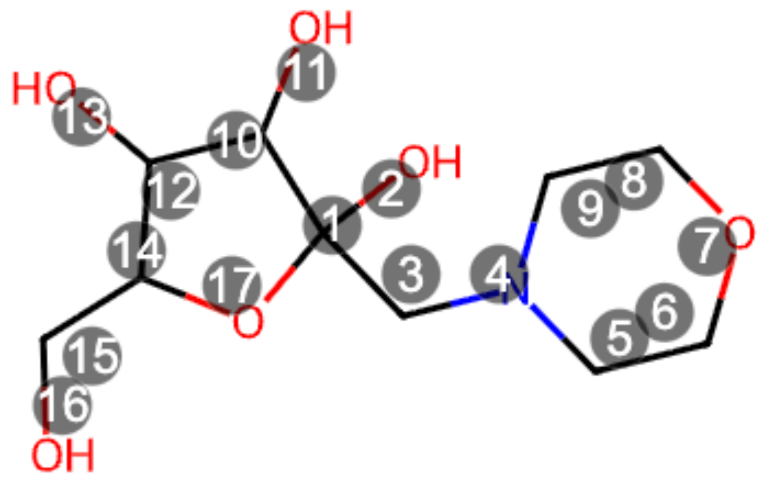
Model scope: This prediction uses a minimal empirical additivity model. It is intended for small to medium organic molecules and typical functional groups. Results may be unreliable for large, highly branched, strained, hydrogen-bonded, substituted aromatic or strongly conjugated systems.

```
1: 110.7 ppm:  +sp3+R5+nA+AO+nA+nB+Q+nA+nB+Q+nB+Q+CO+nA+AO+nB+Q
3:  74.3 ppm:  +sp3+nA+nB+3rdM+nB+CO+nB+nA+AN+nB+nB
5:  57.8 ppm:  +sp3+R6+nA+AN+nB+nB+nA+nB
6:  65.3 ppm:  +sp3+R6+nA+nB+nA+AO+nB
8:  65.3 ppm:  +sp3+R6+nA+AO+nB+nA+nB
9:  57.8 ppm:  +sp3+R6+nA+nB+nA+AN+nB+nB
10:  83.3 ppm:  +sp3+R5+nA+nB+T+nB+T+nB+T+nA+AO+nA+nB+T+nB+T+CO
12:  70.9 ppm:  +sp3+R5+nA+nB+T+CO+CO+nB+T+nA+AO+nA+nB+T+CO+nB+T
14:  74.1 ppm:  +sp3+R5+nA+nB+T+CO+nB+T+nA+nB+T+nA+AO+nB+T+CO
15:  64.3 ppm:  +sp3+nA+nB+CO+nB+nA+AO
```

Predicted ^{13}C spectrum (δ / ppm)

☑ Lock 0–220 ppm

Empirical additivity model for ^{13}C chemical shifts on MolDis-Lab



D. W. Brown, *J. Chem. Educ.*, 62, 209 (1985).

$$\begin{aligned}\delta &= -2.3 + (2\alpha^1 + 2\alpha^2 + 2\beta^1 + \beta^2 + \beta^3 + 5\gamma^1 + \gamma^2)\\ &\quad + (4° \rightarrow 3° + 4° \rightarrow 2° + 4° \rightarrow 2° + 4° \rightarrow 1°)\\ &= -2.3(18.2 + 98.0 + 18.4 + 10.1 + 11.3 - 12.5 - 6.2)\\ &\quad + (-15.0 - 8.4 - 8.4 - 1.5) = 101.7 \text{ (observed 97.6)}\end{aligned}$$

^{13}C shifts predicted with a minimal additivity model

Model scope: This prediction uses a minimal empirical additivity model. It is intended for small to m functional groups. Results may be unreliable for large, highly branched, strained, hydrogen-bonded, conjugated systems.

```
1: 110.7 ppm:  +sp3+R5+nA+AO+nA+nB+Q+nA+nB+Q+nB+Q+CO+nA+AO+nB+Q
3:  74.3 ppm:  +sp3+nA+nB+3rdM+nB+CO+nB+nA
5:  57.8 ppm:  +sp3+R6+nA+AN+nB+nB+nA+nB
6:  65.3 ppm:  +sp3+R6+nA+nB+nA+AO+nB
8:  65.3 ppm:  +sp3+R6+nA+AO+nB+nA+nB
9:  57.8 ppm:  +sp3+R6+nA+nB+nA+AN+nB+nB
10:  83.3 ppm:  +sp3+R5+nA+nB+T+nB+T+nB+T+
12:  70.9 ppm:  +sp3+R5+nA+nB+T+CO+CO+nB+T
14:  74.1 ppm:  +sp3+R5+nA+nB+T+CO+nB+T+nA
15:  64.3 ppm:  +sp3+nA+nB+CO+nB+nA+AO
```

A Very Brief, Rapid, Simple, and Unified Method for Estimating Carbon-13 NMR Chemical Shifts

The BS Method[1]

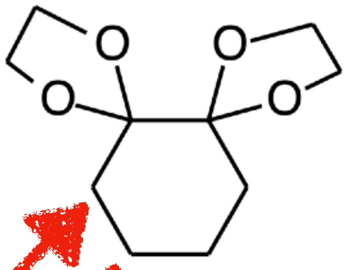
J. Chem. Educ., 64, 915 (1987).

Ben Shoulders
The University of Texas, Austin, TX 78712
Steven C. Welch[2]
University of Houston, Houston, TX 77004

Empirical models offer instantaneous predictions for sanity checking

Empirical degeneracy

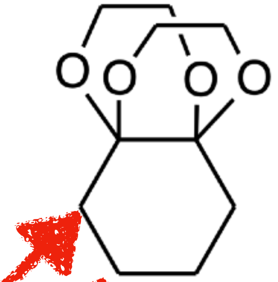
C12(OCCO2)C2(OCCO2)CCCC1



1

```
11:  36.4 ppm:  +sp3+R6+nA+nB+CO+CO+nB+nB+nA+nB
12:  21.4 ppm:  +sp3+R6+nA+nB+CO+CO+nA+nB
13:  21.4 ppm:  +sp3+R6+nA+nB+nA+nB+CO+CO
14:  36.4 ppm:  +sp3+R6+nA+nB+nA+nB+nB+nB+CO+CO
```

C123C(OCCO2)(OCCO3)CCCC1



2

```
11:  36.4 ppm:  +sp3+R6+nA+nB+CO+CO+nB+nB+nA+nB
12:  21.4 ppm:  +sp3+R6+nA+nB+CO+CO+nA+nB
13:  21.4 ppm:  +sp3+R6+nA+nB+nA+nB+CO+CO
14:  36.4 ppm:  +sp3+R6+nA+nB+nA+nB+CO+CO+nB+nB
```

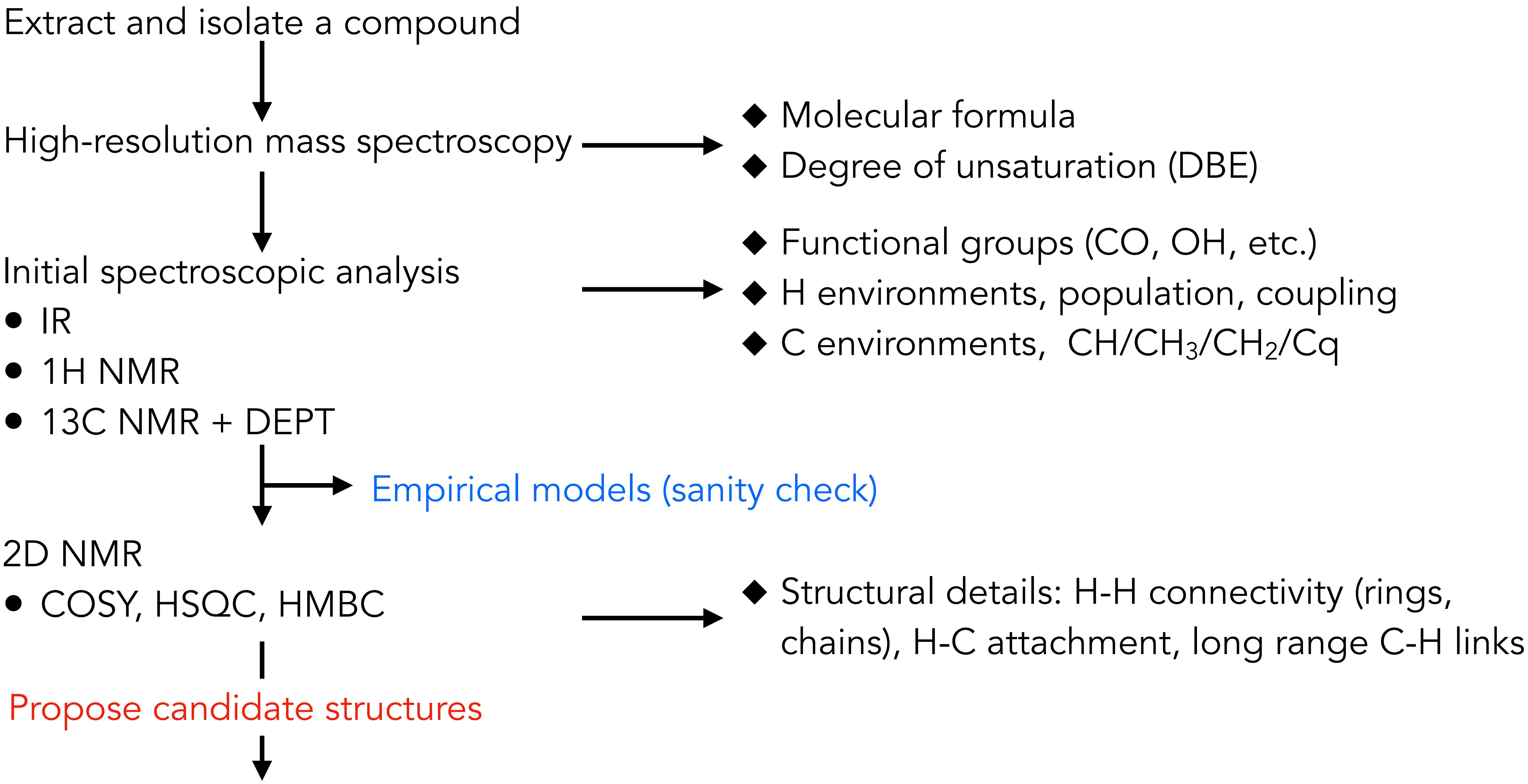
Identical empirical chemical shifts for different structures

(*lack of long-range effects and 3D interactions*)

Perhaps both structures have same ^{1}H (sub) spectrum?

L. J. Tilley et al., *J. Chem. Educ. 79, 593* (2002)

Can the proposed structure go wrong?



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- IR
- 1H NMR
- 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, CH/CH$_3$/CH$_2$/Cq

Empirical models (sanity check)

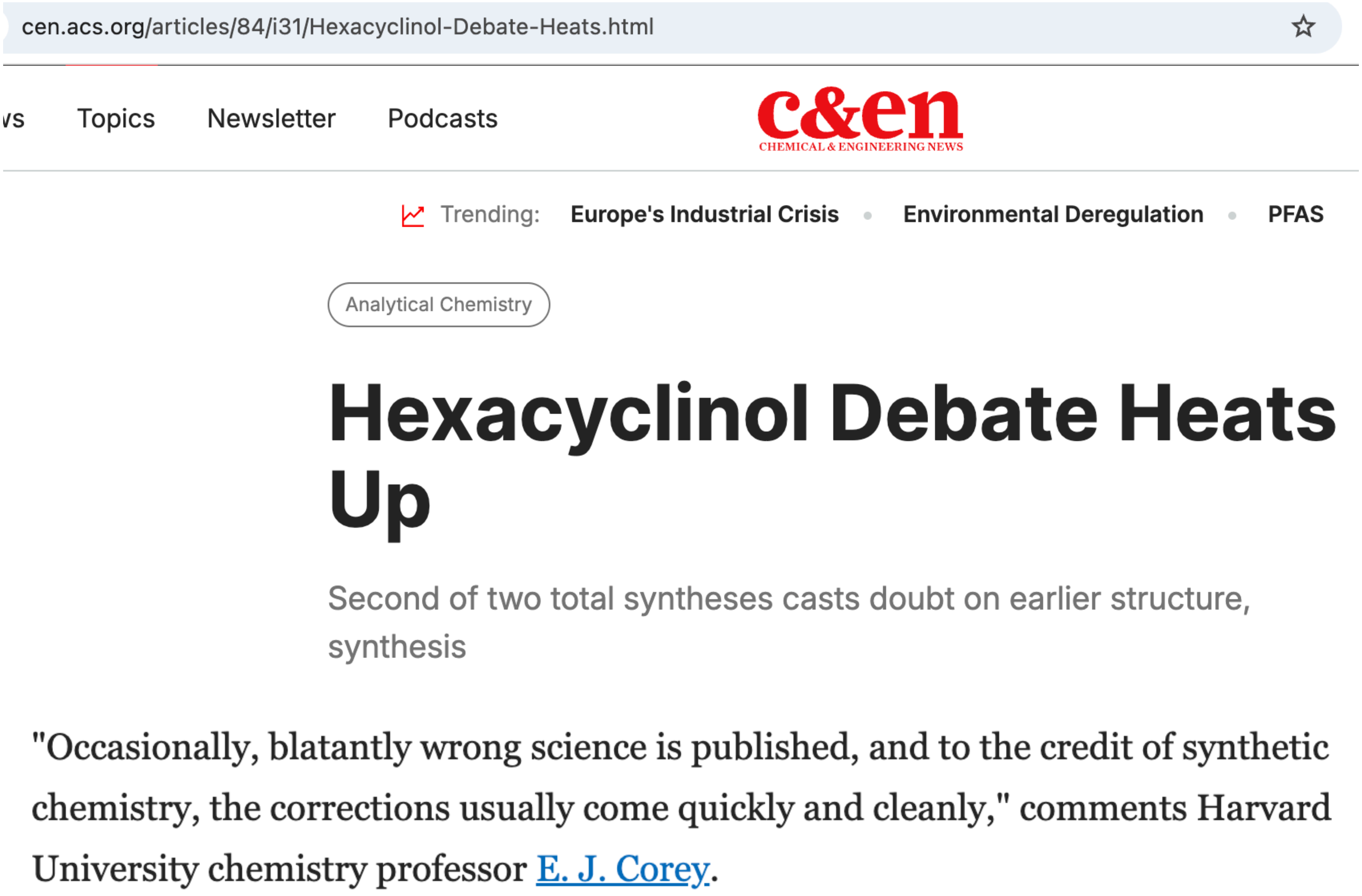
2D NMR

- COSY, HSQC, HMBC

- ◆ Structural details: H-H connectivity (rings, chains), H-C attachment, long range C-H links

Propose candidate structures

Structure assignment going wrong for hexacyclinol



ws
Topics
Newsletter
Podcasts
c&en
CHEMICAL & ENGINEERING NEWS
Trending:
Europe's Industrial Crisis
Environmental Deregulation
PFAS
Analytical Chemistry

cen.acs.org/articles/84/i31/Hexacyclinol-Debate-Heats.html

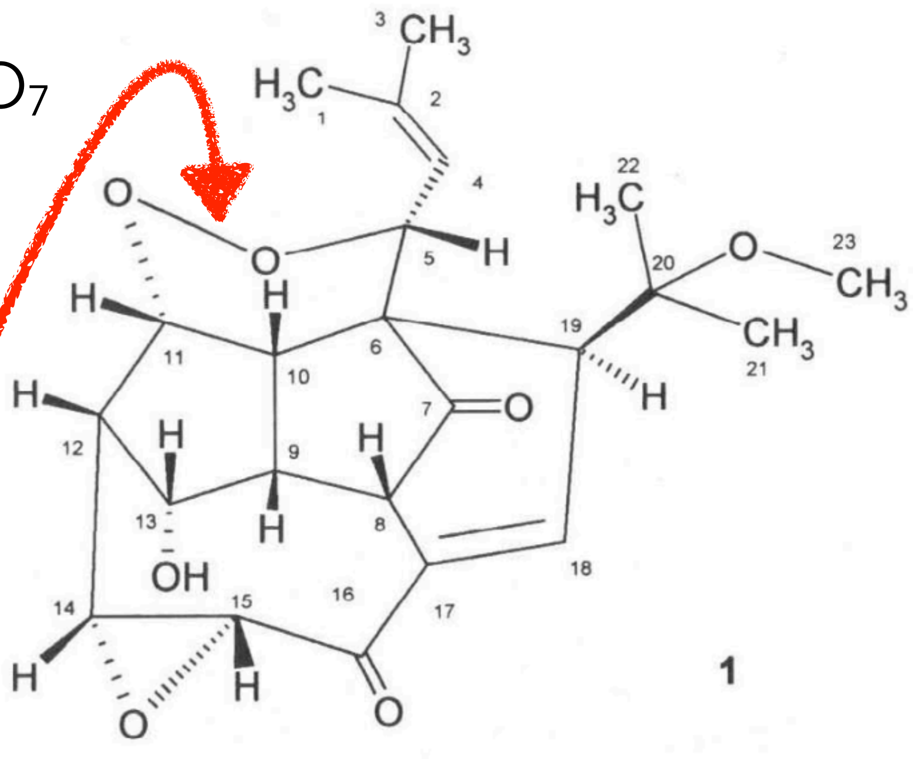
Hexacyclinol Debate Heats Up

Second of two total syntheses casts doubt on earlier structure, synthesis

"Occasionally, blatantly wrong science is published, and to the credit of synthetic chemistry, the corrections usually come quickly and cleanly," comments Harvard University chemistry professor E. J. Corey.

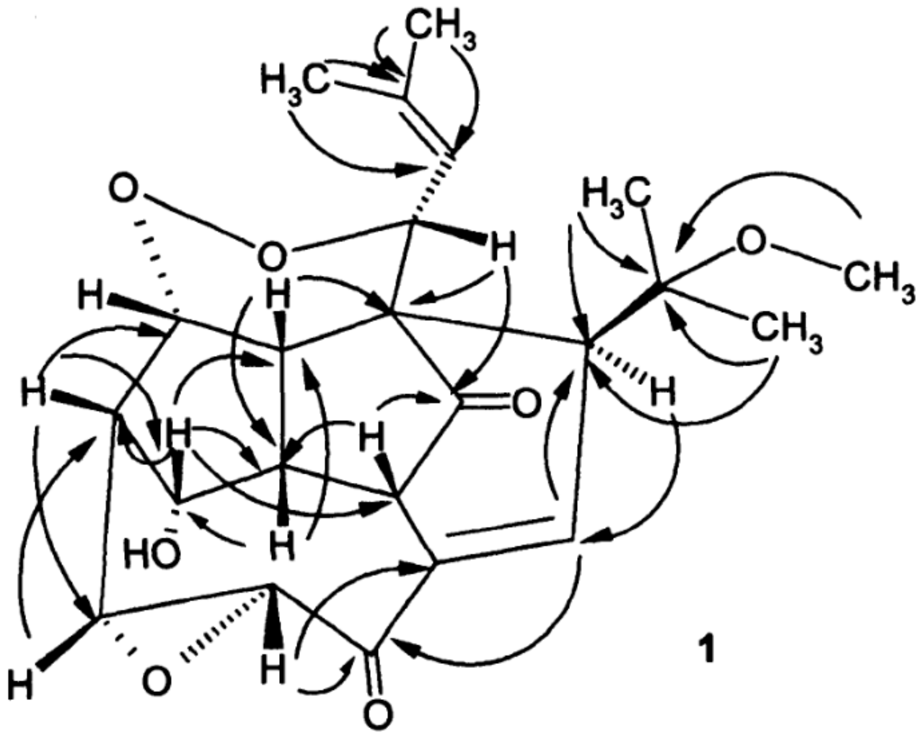
Initial structure assignment of hexacyclinol *extracted* (in 2002)

$C_{23}H_{28}O_7$

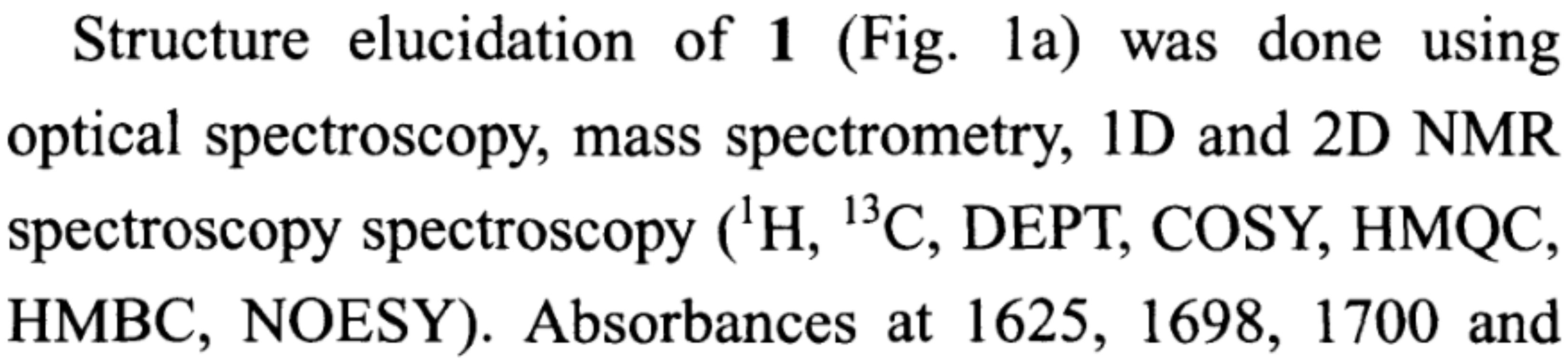


1

strained endoperoxide



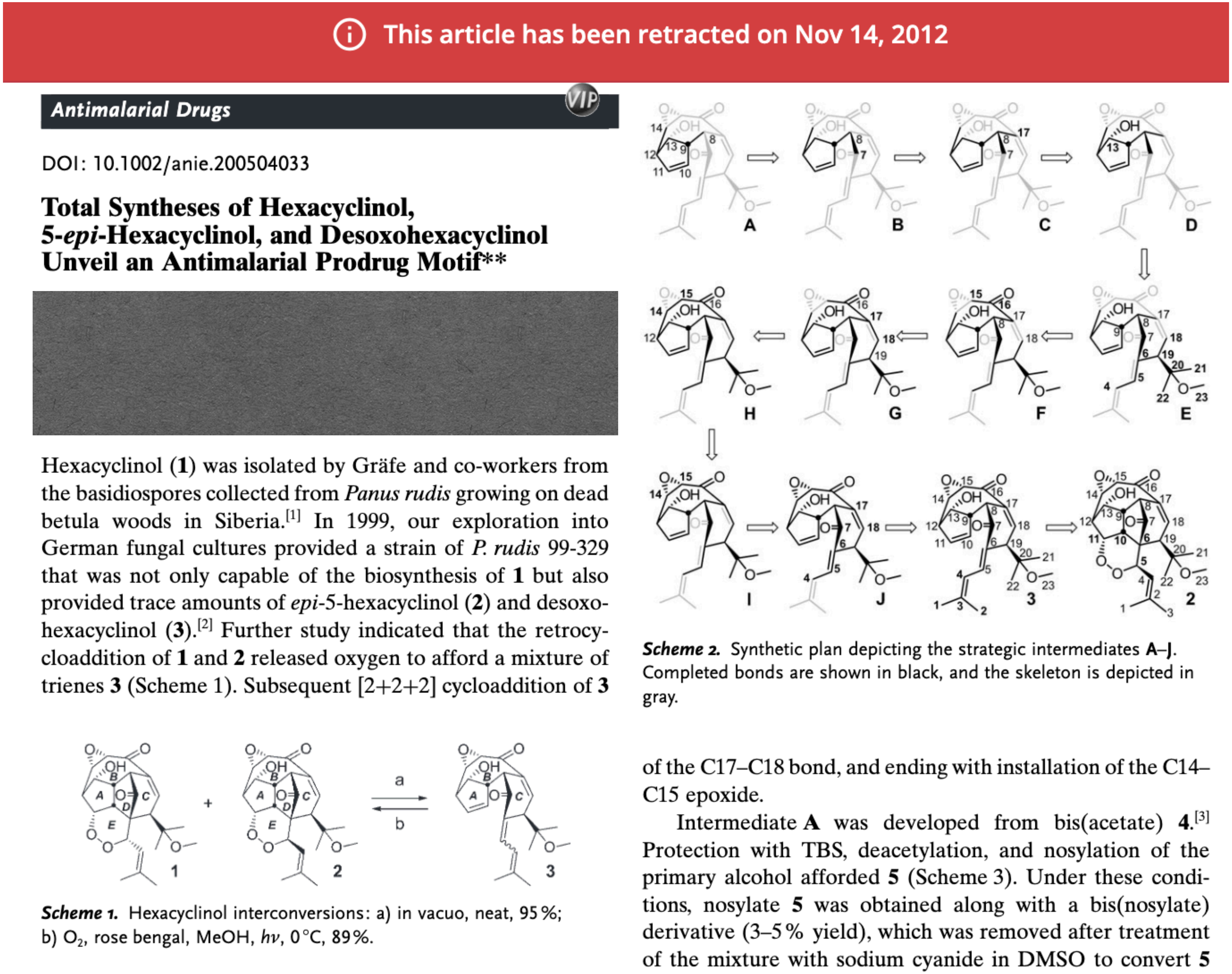
1



Structure elucidation of **1** (Fig. 1a) was done using optical spectroscopy, mass spectrometry, 1D and 2D NMR spectroscopy spectroscopy (^{1}H, ^{13}C, DEPT, COSY, HMQC, HMBC, NOESY). Absorbances at 1625, 1698, 1700 and

B. Schlegel et al., J. Antibiot, 55, 814 (2002).

Initial structure assignment of hexacyclinol *synthesized* (in 2006)



ⓘ This article has been retracted on Nov 14, 2012

Antimalarial Drugs

DOI: 10.1002/anie.200504033

Total Syntheses of Hexacyclinol, 5-*epi*-Hexacyclinol, and Desoxohexacyclinol Unveil an Antimalarial Prodrug Motif**

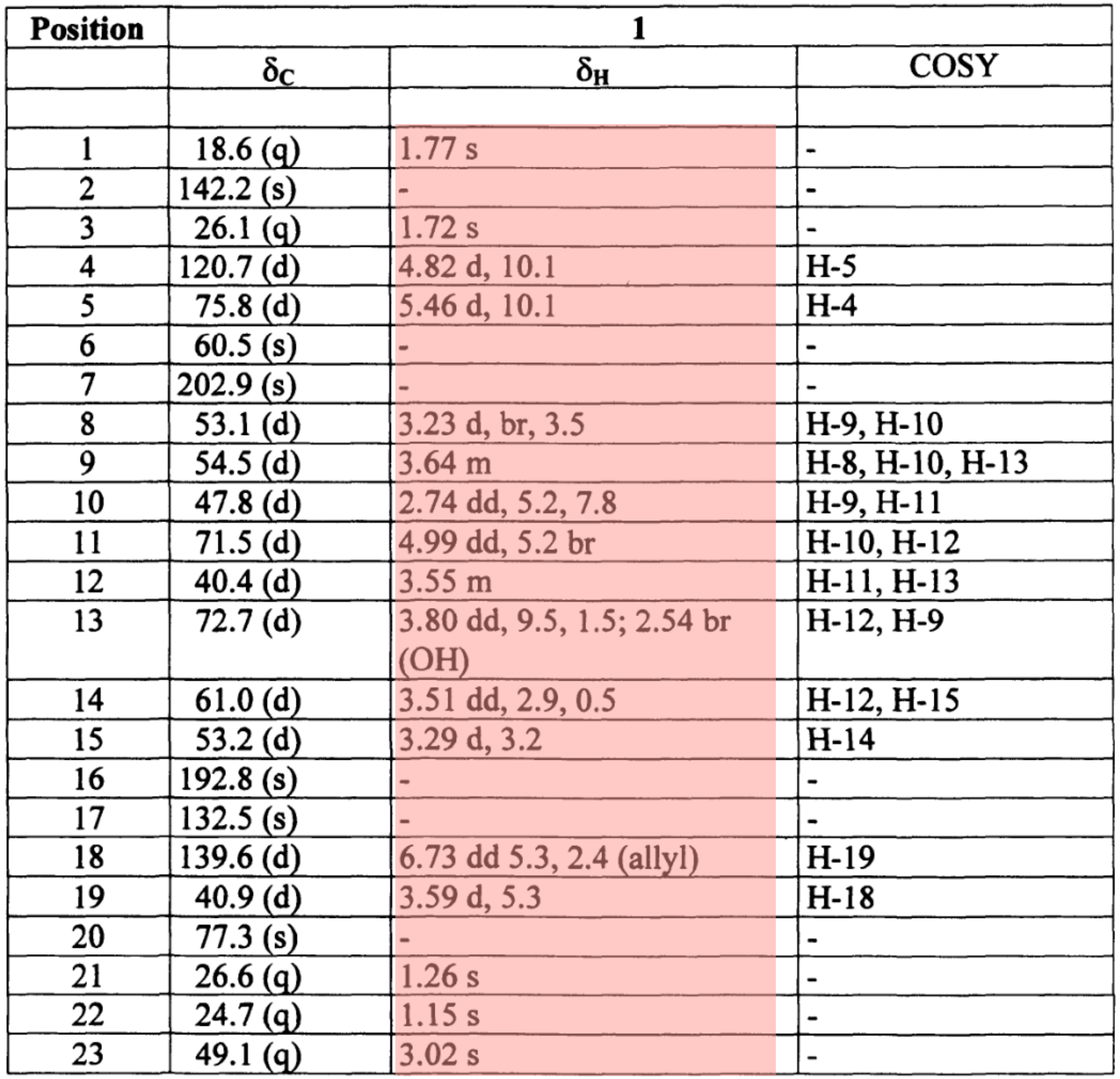
Hexacyclinol (**1**) was isolated by Gräfe and co-workers from the basidiospores collected from *Panus rudis* growing on dead betula woods in Siberia.[1] In 1999, our exploration into German fungal cultures provided a strain of *P. rudis* 99-329 that was not only capable of the biosynthesis of **1** but also provided trace amounts of *epi*-5-hexacyclinol (**2**) and desoxohexacyclinol (**3**).[2] Further study indicated that the retrocycloaddition of **1** and **2** released oxygen to afford a mixture of trienes **3** (Scheme 1). Subsequent [2+2+2] cycloaddition of **3** of the C17–C18 bond, and ending with installation of the C14–C15 epoxide.

Scheme 1. Hexacyclinol interconversions: a) in vacuo, neat, 95%; b) O_2, rose bengal, MeOH, *hν*, 0°C, 89%.

Scheme 2. Synthetic plan depicting the strategic intermediates **A**–**J**. Completed bonds are shown in black, and the skeleton is depicted in gray.

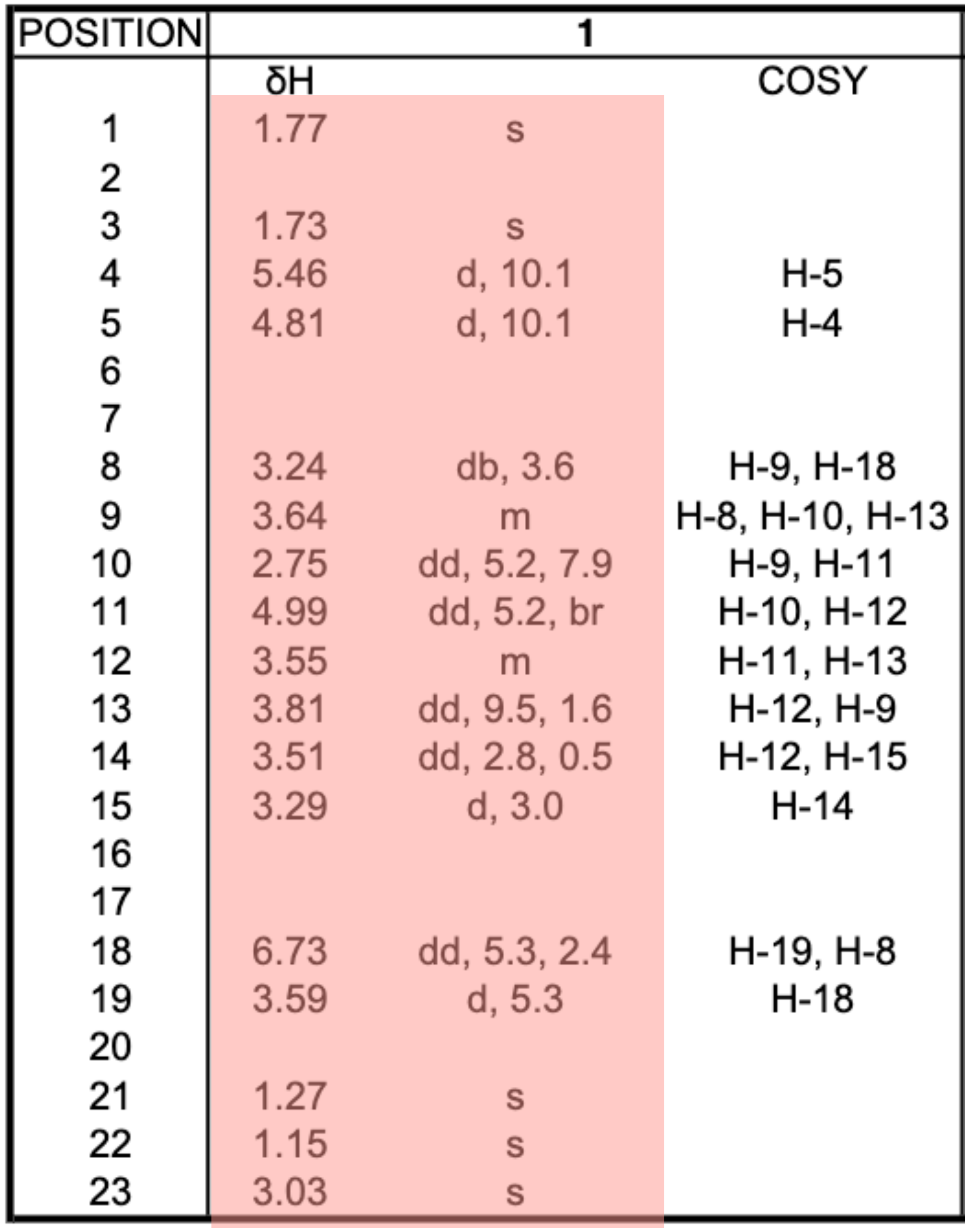
Intermediate **A** was developed from bis(acetate) **4**.[3] Protection with TBS, deacetylation, and nosylation of the primary alcohol afforded **5** (Scheme 3). Under these conditions, nosylate **5** was obtained along with a bis(nosylate) derivative (3–5% yield), which was removed after treatment of the mixture with sodium cyanide in DMSO to convert **5**

NMR data



Position	1		
	δ_C	δ_H	COSY
1	18.6 (q)	1.77 s	-
2	142.2 (s)	-	-
3	26.1 (q)	1.72 s	-
4	120.7 (d)	4.82 d, 10.1	H-5
5	75.8 (d)	5.46 d, 10.1	H-4
6	60.5 (s)	-	-
7	202.9 (s)	-	-
8	53.1 (d)	3.23 d, br, 3.5	H-9, H-10
9	54.5 (d)	3.64 m	H-8, H-10, H-13
10	47.8 (d)	2.74 dd, 5.2, 7.8	H-9, H-11
11	71.5 (d)	4.99 dd, 5.2 br	H-10, H-12
12	40.4 (d)	3.55 m	H-11, H-13
13	72.7 (d)	3.80 dd, 9.5, 1.5; 2.54 br (OH)	H-12, H-9
14	61.0 (d)	3.51 dd, 2.9, 0.5	H-12, H-15
15	53.2 (d)	3.29 d, 3.2	H-14
16	192.8 (s)	-	-
17	132.5 (s)	-	-
18	139.6 (d)	6.73 dd 5.3, 2.4 (allyl)	H-19
19	40.9 (d)	3.59 d, 5.3	H-18
20	77.3 (s)	-	-
21	26.6 (q)	1.26 s	-
22	24.7 (q)	1.15 s	-
23	49.1 (q)	3.02 s	-

2002

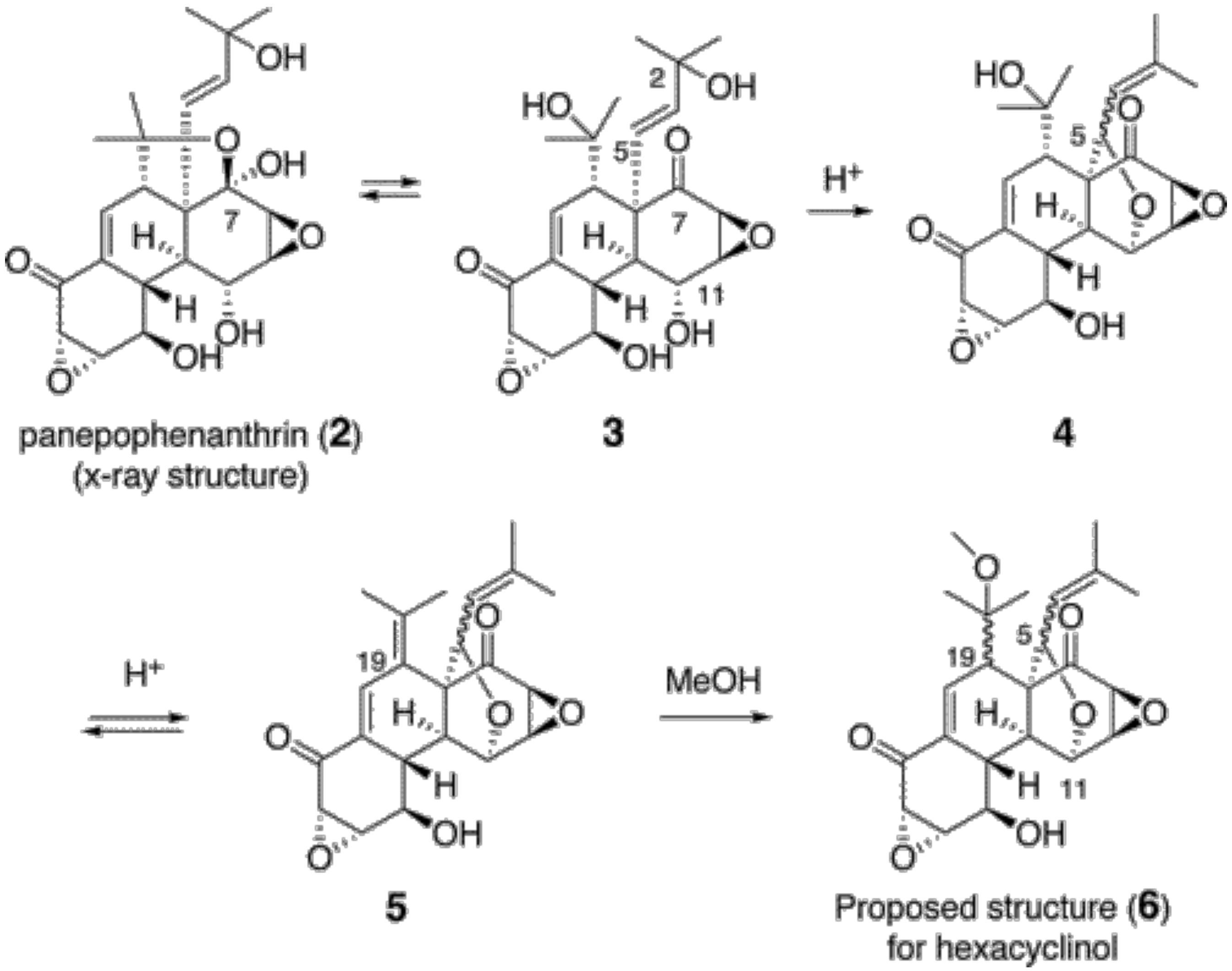


POSITION	1		
	δH		COSY
1	1.77	s	
2			
3	1.73	s	
4	5.46	d, 10.1	H-5
5	4.81	d, 10.1	H-4
6			
7			
8	3.24	db, 3.6	H-9, H-18
9	3.64	m	H-8, H-10, H-13
10	2.75	dd, 5.2, 7.9	H-9, H-11
11	4.99	dd, 5.2, br	H-10, H-12
12	3.55	m	H-11, H-13
13	3.81	dd, 9.5, 1.6	H-12, H-9
14	3.51	dd, 2.8, 0.5	H-12, H-15
15	3.29	d, 3.0	H-14
16			
17			
18	6.73	dd, 5.3, 2.4	H-19, H-8
19	3.59	d, 5.3	H-18
20			
21	1.27	s	
22	1.15	s	
23	3.03	s	

2006

Reassignment of the structure of hexacyclinol with DFT-based NMR

synthetic route



panepophenanthrin (2)
(x-ray structure)

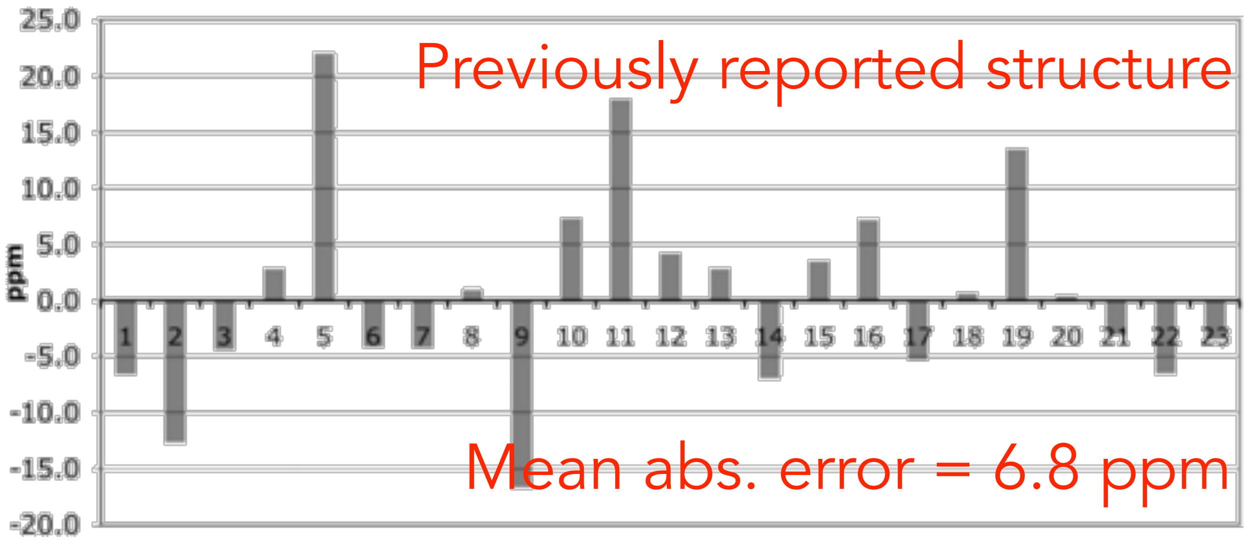
3

4

5

Proposed structure (6)
for hexacyclinol

DFT-exp. ^{13}C chemical shifts



Previously reported structure

Mean abs. error = 6.8 ppm

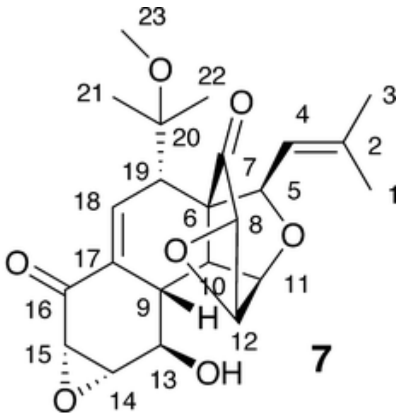
DFT modeling:

Geometry: HF/3-21G

^{13}C NMR: mPW1PW91/6-31G(d,p)

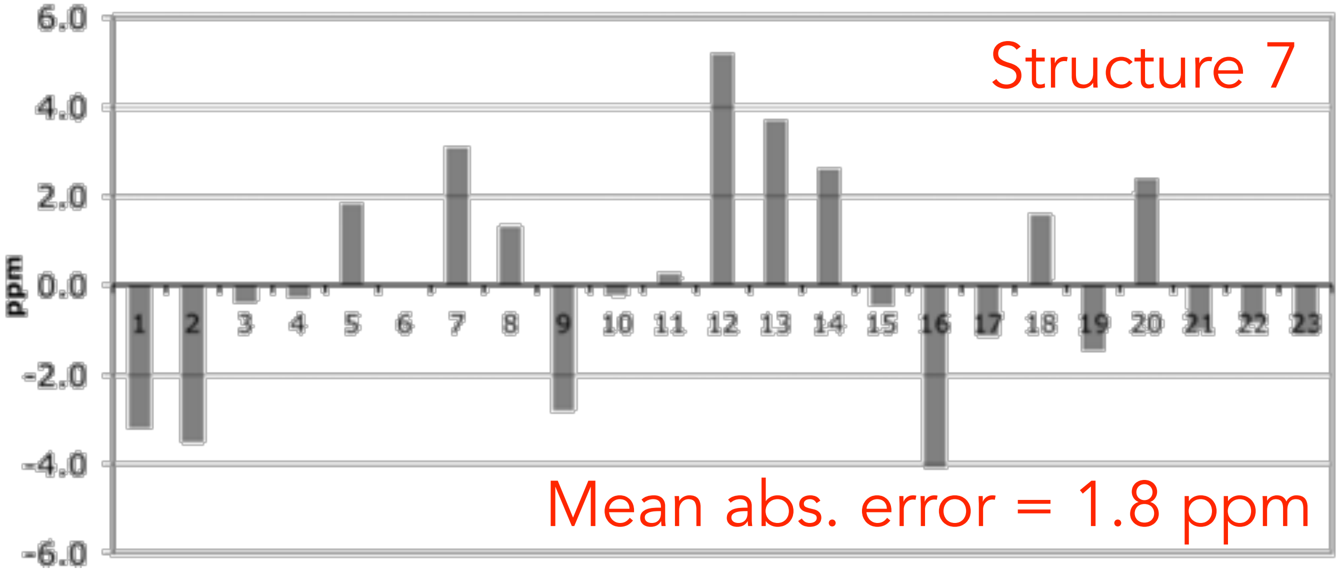
Methanol solvent

Has an error of 1-2 ppm for similar compounds



7

Conformer of 6



Structure 7

Mean abs. error = 1.8 ppm

S. D. Rychnovsky, *Org. Lett.* 8, 2895 (2006).

ORGANIC LETTERS

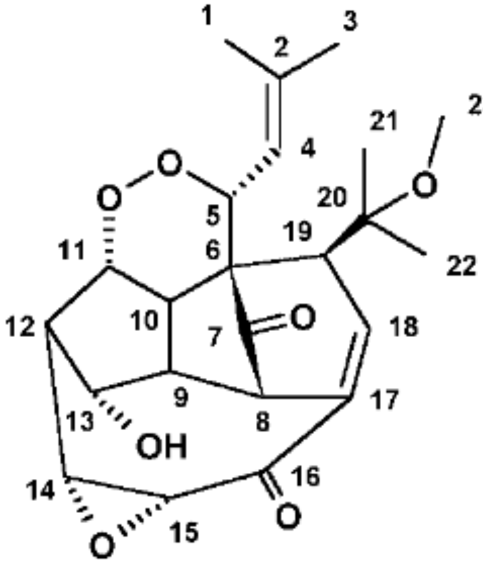
2009
Vol. 11, No. 6
1409–1412

Can Two Molecules Have the Same NMR Spectrum? Hexacyclinol Revisited

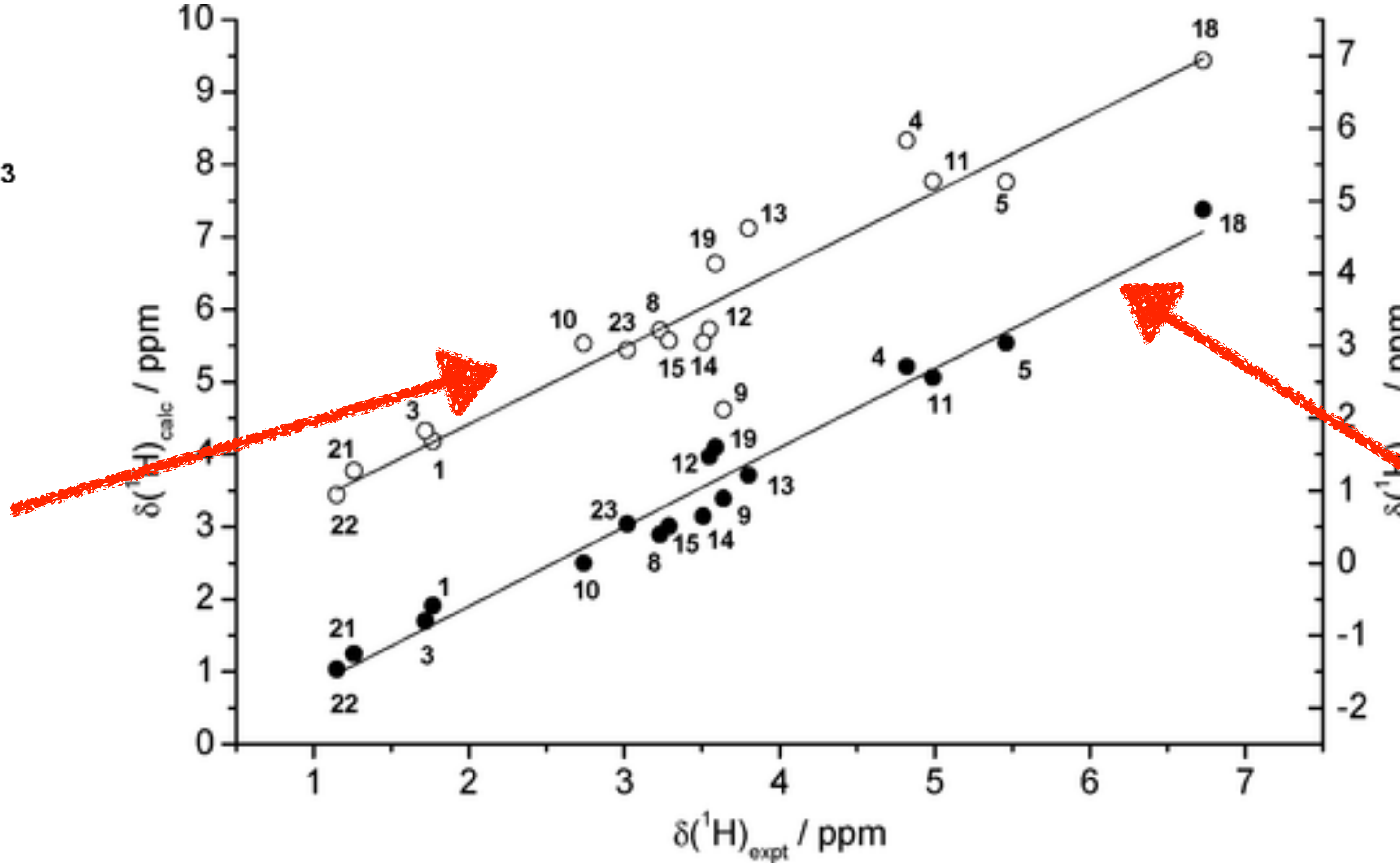
Giacomo Saielli[†] and Alessandro Bagno[*,‡]

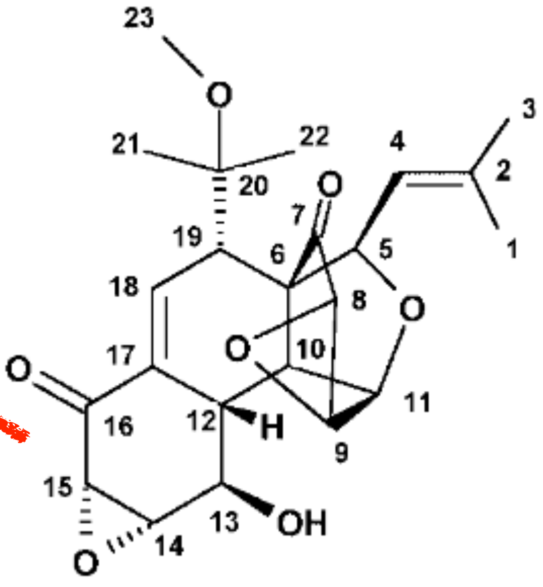
DFT modeling:
Geometry: B3LYP/6-31G(d,p)
^{1}H NMR: B97-2/cc-pVTZ

averaged over two conformers



1





2

Mean abs. error = 0.4 ppm

Mean abs. error = 0.2 ppm

Probabilistic error metrics for high confidence structure assignment

- DP4 score: Penalizes a structure with outliers in computed chemical shifts

Bayesian probability that a candidate structure i (with N centers) is correctly assigned to the experimental data

$$P\left(i \mid \delta_1, \delta_2, \cdots, \delta_N\right) = \frac{\Pi_{k=1}^{N}\left[1 - T_\nu(t)\right]}{\sum_{j=1}^{m} \Pi_{k=1}^{N}\left[1 - T_\nu(t)\right]}$$

$T_\nu(t)$: Cumulative probabilities for Student's t-distribution with ν degrees of freedom

t: standard score of scaled prediction error

$$t = \frac{\left|\left(\delta^{i}_{\mathrm{scaled},k} - \delta_{\exp,k}\right) - \mu\right|}{\sigma}$$

$\delta^{j}_{\mathrm{scaled},k}$: Linearly corrected calculated values against experimental values

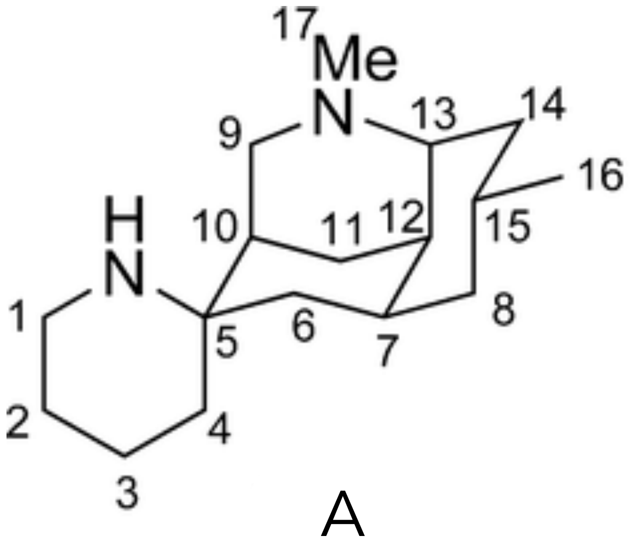
μ, σ, ν: Free parameters (unlike in hypothesis testing)

For partially or unassigned peaks, permute computed values to maximize accuracy

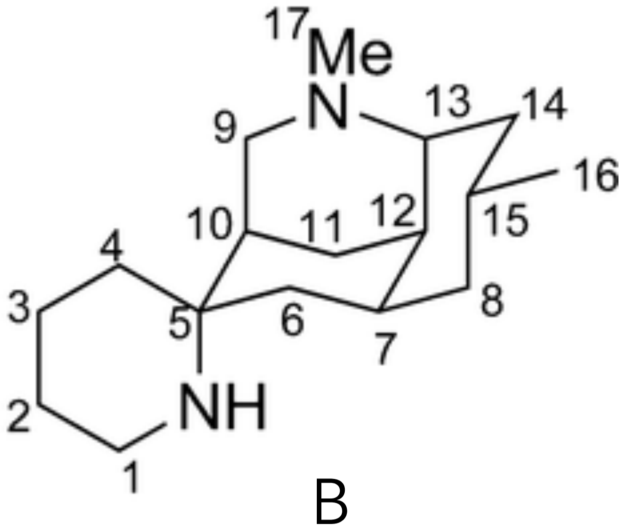
S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.* 132, 12946 (2010).

DP4 probability vs. statistical metrics

Error metrics for comparing computed chemical shifts of two structures A and B against experimental values



A

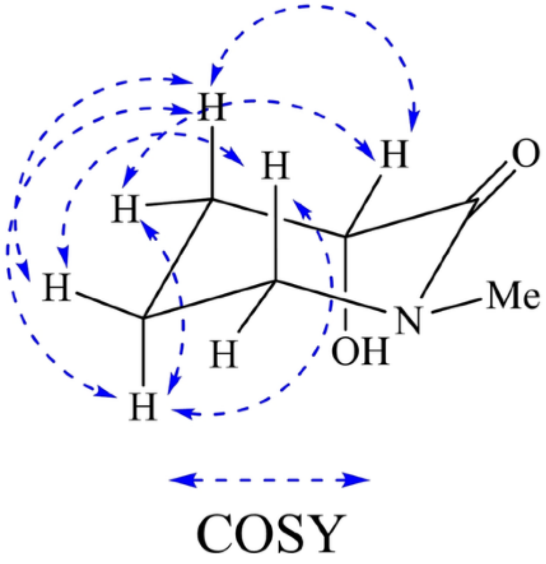


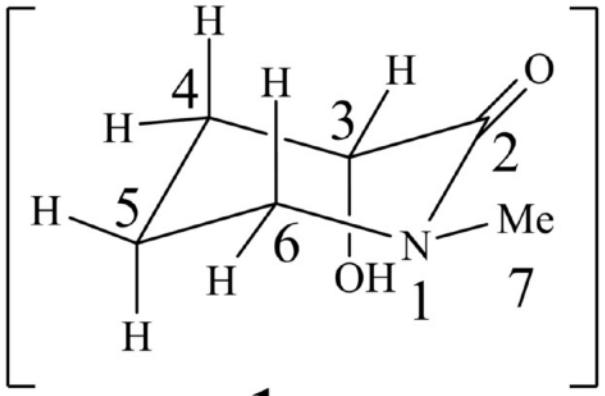
B

^{13}C NMR		
Mean absolute error	1.50	1.62
Standard deviation of the error	1.59	1.83
DP4 probability	79.5%	20.5%
^{1}H NMR		
Mean absolute error	0.11	0.18
Standard deviation of the error	0.13	0.22
DP4 probability	100%	0%

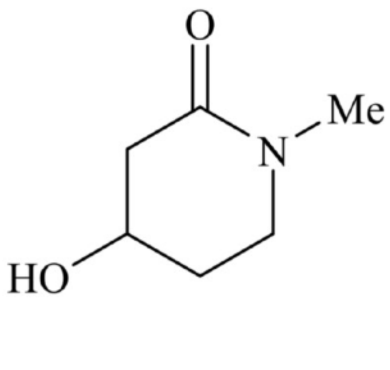
S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.* 132, 12946 (2010).

Example structure assignment of a natural product

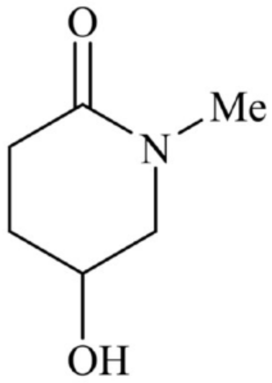




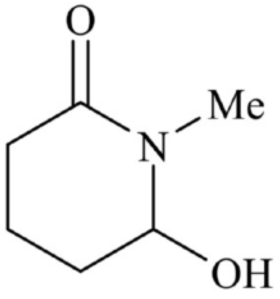
1a



1b

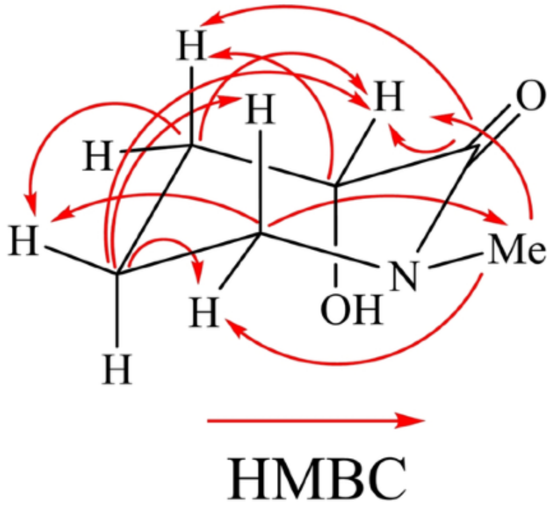


1c

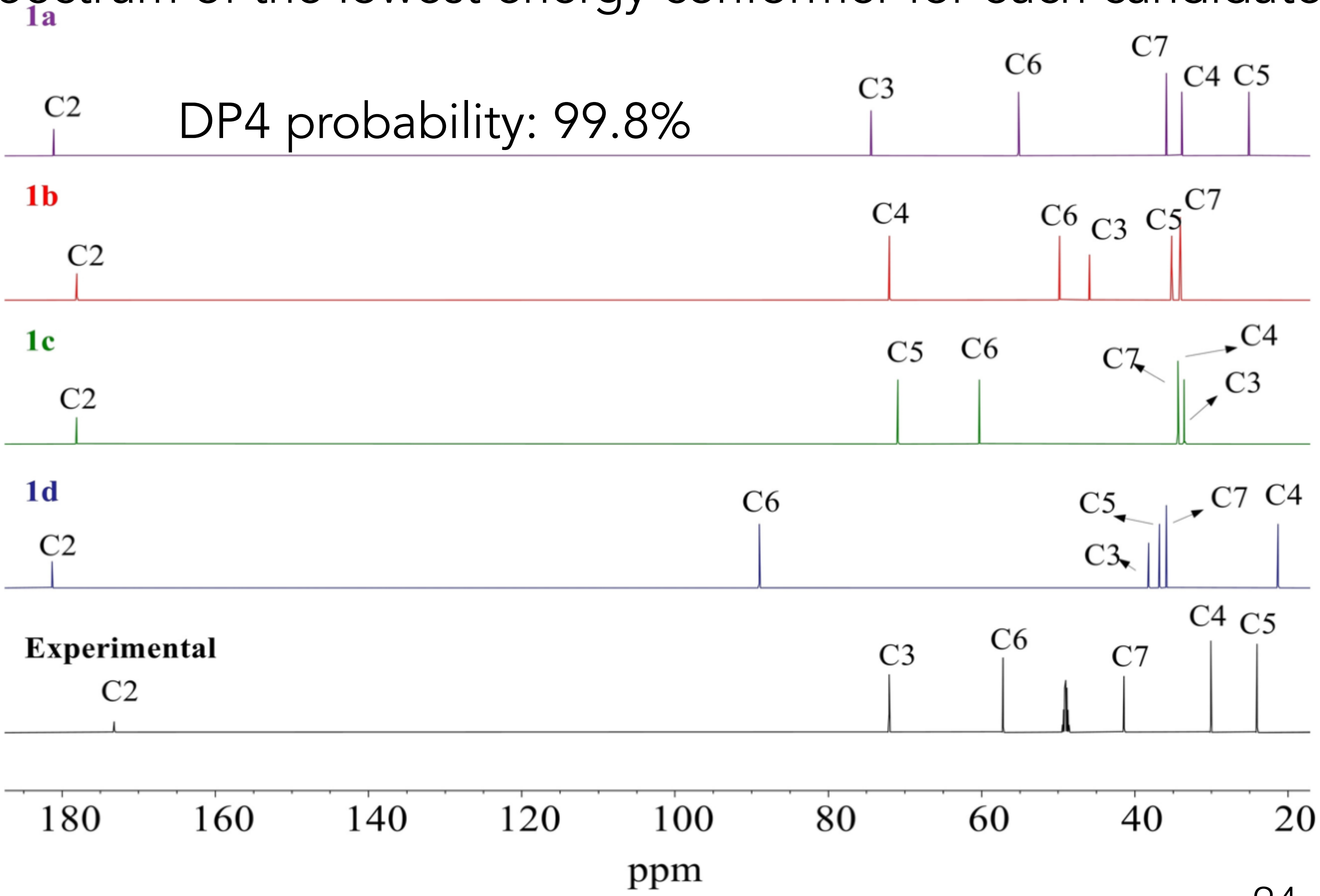


1d

Proposed candidate structures for a natural product



DFT spectrum of the lowest energy conformer for each candidate


DP4 probability: 99.8%

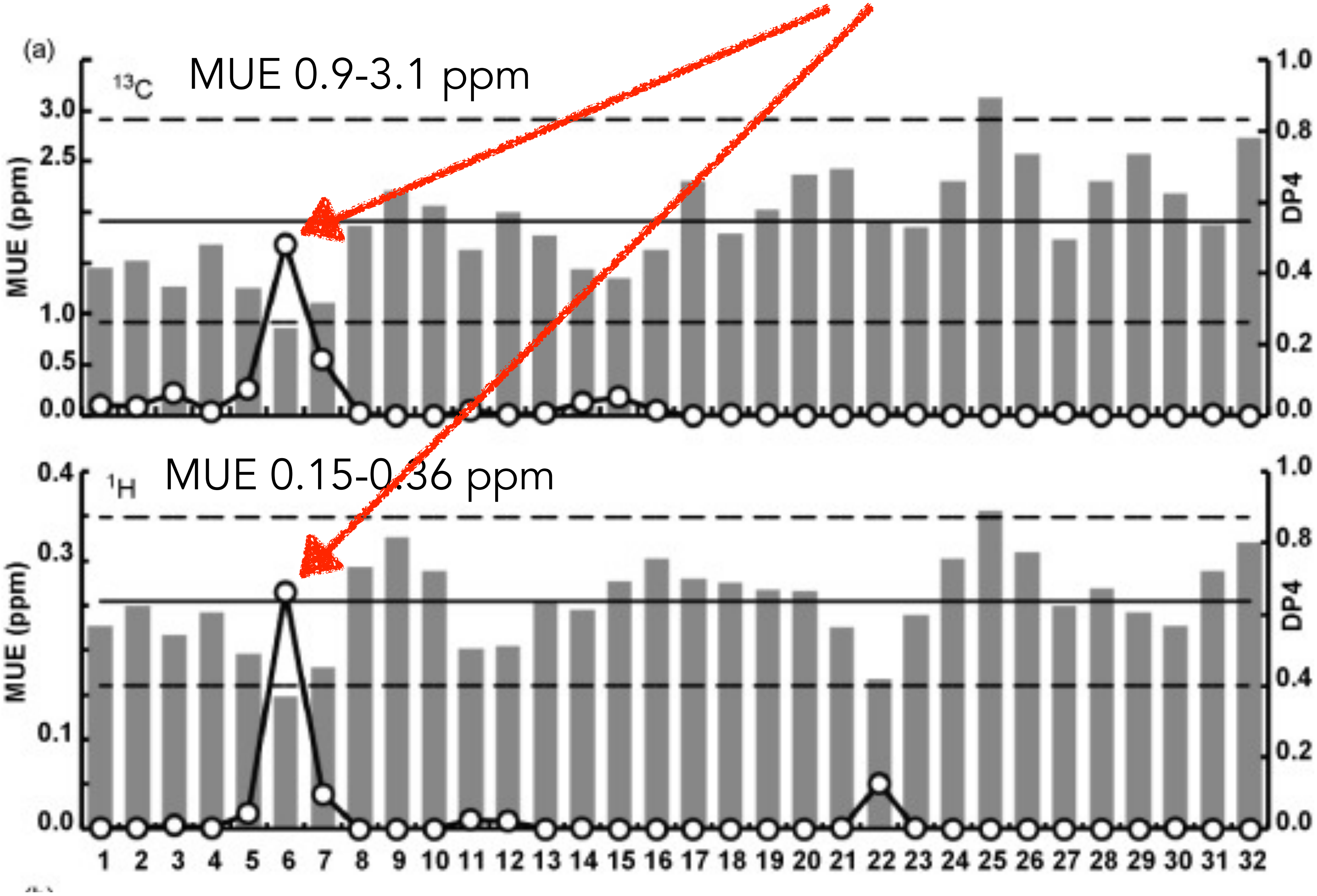
DFT modeling:
Geometry: B3LYP/6-311+G(2d,p)
NMR: M06-2X/6-31+G(d,p)
Methanol solvent

B. N. S. Pinto et al., *Asian J. Org. Chem.*, 11, e202200182 (2022).

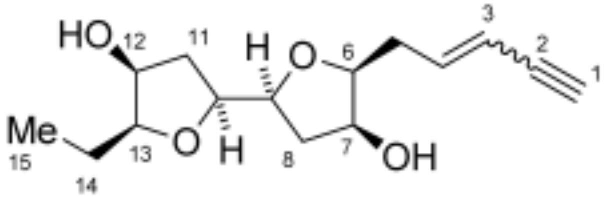
High-confidence structure assignment of laurefurenyne (32 diastereomers)

DFT modeling: Geometry: wB97XD/6-31G(d), Shielding: mPW1PW91/6-311G(d,p)
Boltzmann weighing of conformers sampled with Monte Carlo search (and MMFF)

Diastereomer 5 has the highest DP4 probability

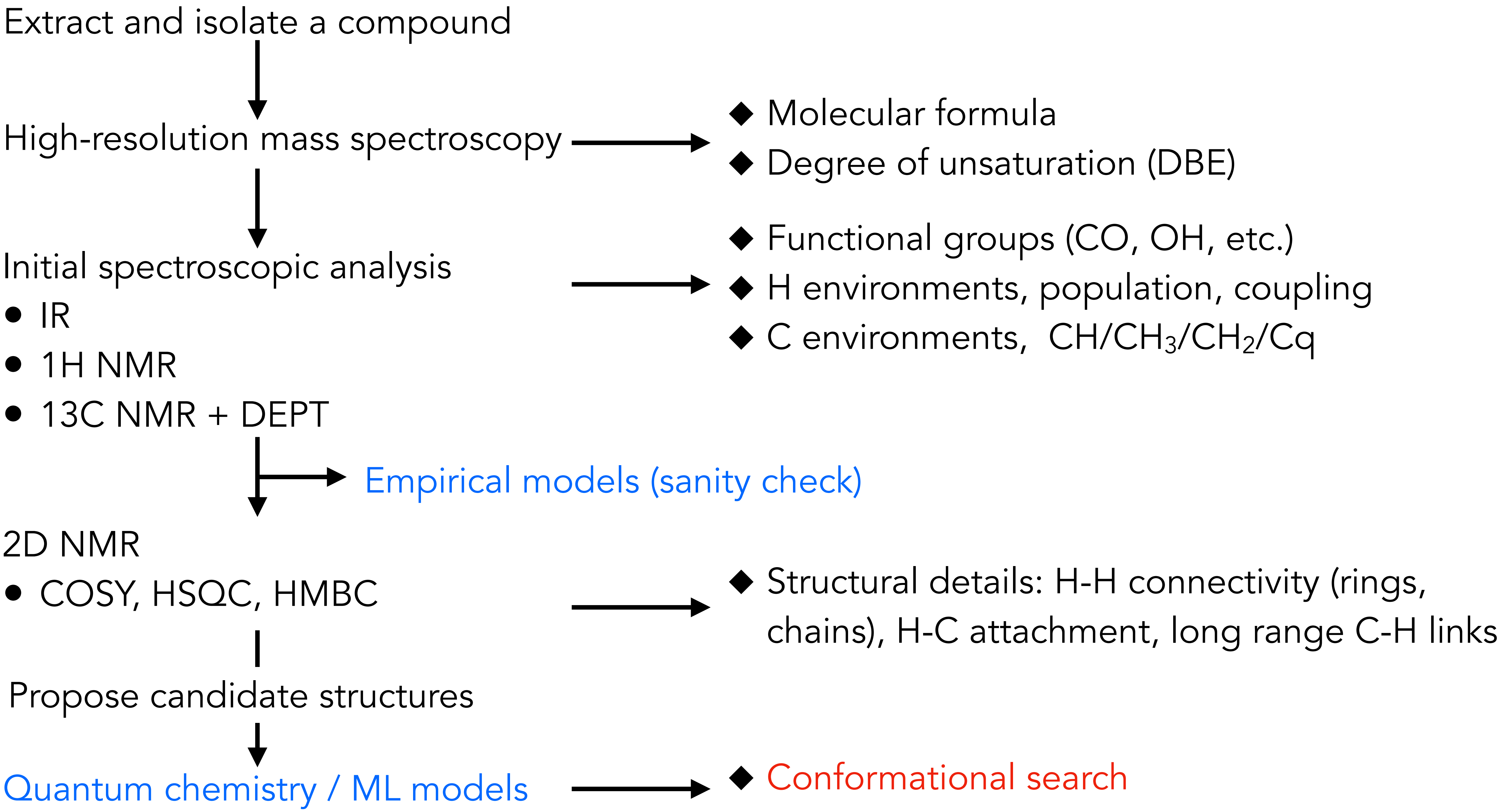

MUE 0.9-3.1 ppm

MUE 0.15-0.36 ppm



D. J. Shepherd, *Chem. Eur. J.*19, 12644 (2013).

Quantum chemistry of NMR parameters



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- ● IR
- ● 1H NMR
- ● 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, CH/CH_3/CH_2/Cq

Empirical models (sanity check)

2D NMR

- ● COSY, HSQC, HMBC

- ◆ Structural details: H-H connectivity (rings, chains), H-C attachment, long range C-H links

Propose candidate structures

Quantum chemistry / ML models

- ◆ Conformational search

Ensemble averaging of chemical shifts: conformational / vibrational effects

- Quantum mechanical (nuclear QM) ensemble

$$\langle\sigma\rangle_T = \int d\mathbf{R} \sum_n \frac{e^{-\beta E_n}}{Z} |\Psi_n(\mathbf{R})|^2 \sigma(\mathbf{R})$$

- Classical-/ab initio-MD ensemble (classical nuclei)

$$\langle\sigma\rangle_T = \frac{1}{Z}\int d\mathbf{R}d\mathbf{P}\, e^{-\beta H_{\mathrm{nuc}}^{cl}(\mathbf{R},\mathbf{P})} \sigma(\mathbf{R})$$

- Discrete conformer approximation (common practice)

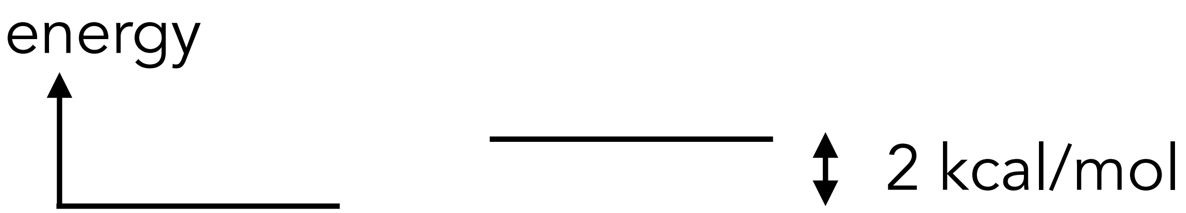
$$\langle\sigma\rangle_T = \sum_i w_i \sigma_i; \qquad w_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}$$

- Local vibrational correction for anharmonic effects (within each conformer)

$$\sigma(\mathbf{R}) \approx \sigma_0 + \Sigma_k \left(\frac{\partial\sigma}{\partial Q_k}\right) Q_k + \frac{1}{2}\Sigma_k \left(\frac{\partial^2\sigma}{\partial Q_k^2}\right) Q_k^2 + \ldots$$

Needed for high-level benchmarking of computer NMR against precise experimental data

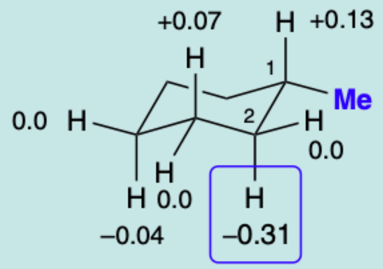
Boltzmann weighing of conformers



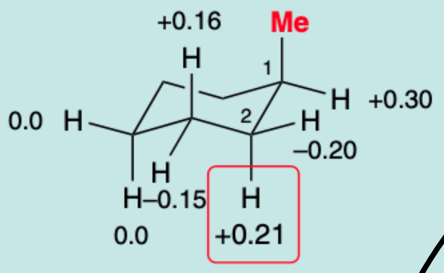
$(T = 298.15\text{K})$

weights $= \exp(-E/RT)$ $w_1 = 1.0$ $w_2 = 0.034$ $Z = w_1 + w_2 = 1.034$

$p_1 = w_1/Z = 96.7\,\%$ $p_2 = w_2/Z = 3.3\,\%$



7eq-Me



7ax-Me

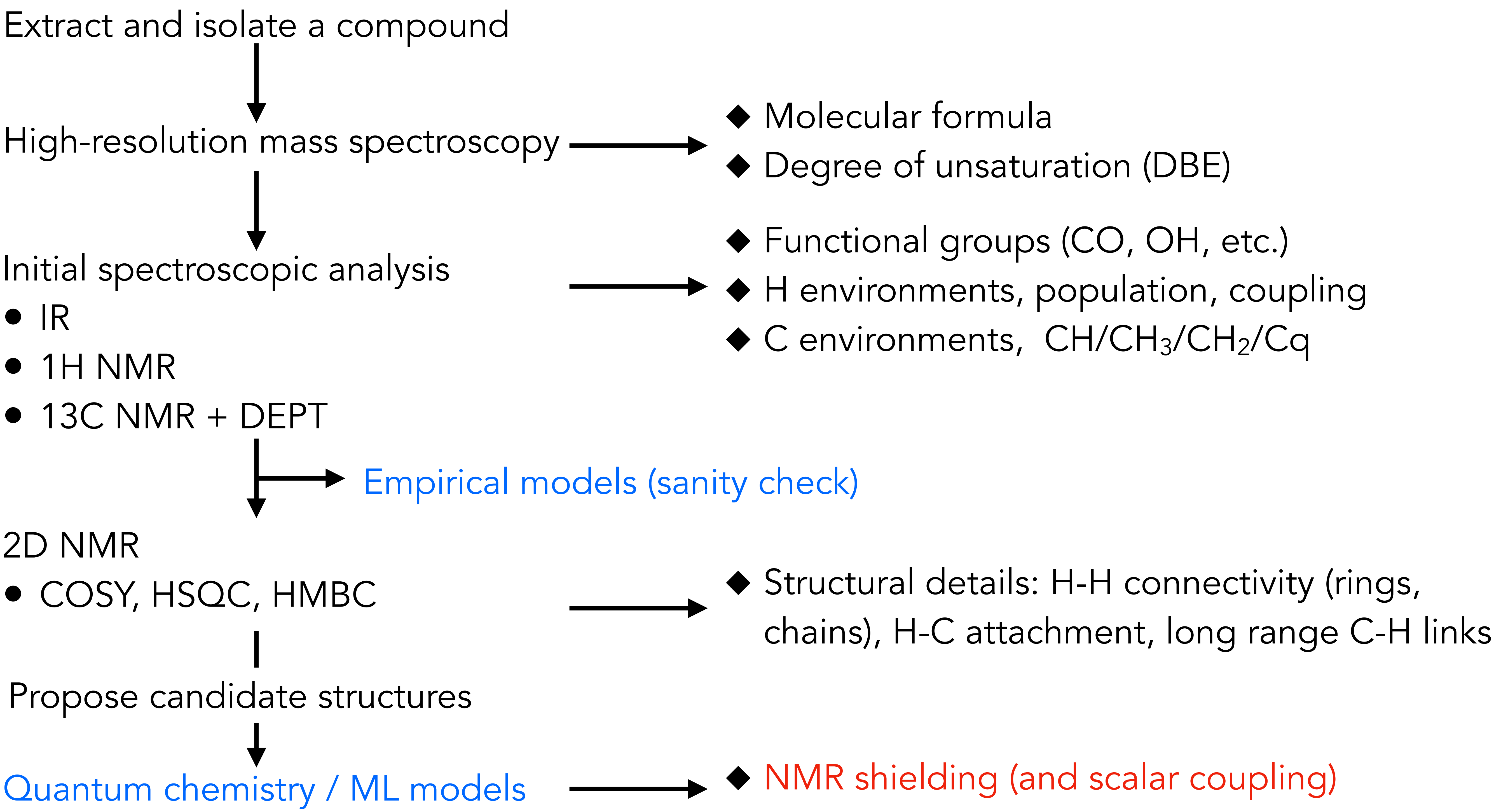
$p_1\delta_1 + p_2\delta_2$

Boltzmann-weighted chemical shifts

	δ(comp)	Proton	δ(comp)	δ(comp)		δ(exp)[25]
	1.29	H1	1.84	1.30	–	1.34
	1.63	$H2_b$	1.54	1.63	–	1.65
	0.88	$H2_a$	1.49	0.90	–	0.88
	1.26	$H3_b$	1.36	1.27	–	1.23
	1.67	$H3_a$	1.54	1.67	–	1.68
	1.63	$H4_b$	1.22	1.62	–	1.62
	1.15	$H4_a$	1.64	1.16	–	1.14
	0.86	Me	1.03	0.87	–	0.86
MAE	0.02		0.32	0.02		

P. H. Willoughby et al., *Nature Protocols*, 9, 644 (2022).

NMR parameters



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- IR
- 1H NMR
- 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, $CH/CH_3/CH_2/Cq$

Empirical models (sanity check)

2D NMR

- COSY, HSQC, HMBC

- ◆ Structural details: H-H connectivity (rings, chains), H-C attachment, long range C-H links

Propose candidate structures

Quantum chemistry / ML models

- ◆ NMR shielding (and scalar coupling)

Molecular properties as derivatives of electronic total energy

- In quantum chemistry, the total energy is the central observable.
- Molecular properties emerge as a response (of the electron density and nuclei) to an external perturbation (ε) can be represented as a Taylor expansion of the energy around the unperturbed value

$$E(\varepsilon) = E(\varepsilon = 0) + \left.\frac{dE}{d\varepsilon}\right|_{\varepsilon=0} \varepsilon + \frac{1}{2!}\left.\frac{d^2E}{d\varepsilon^2}\right|_{\varepsilon=0} \varepsilon^2 + \dots$$

$$E(\varepsilon_1, \varepsilon_2) = E(\varepsilon_1 = 0, \varepsilon_2 = 0) + \left.\frac{dE}{d\varepsilon_1}\right|_{\varepsilon_1=0} \varepsilon_1 + \left.\frac{dE}{d\varepsilon_2}\right|_{\varepsilon_2=0} \varepsilon_2 + \frac{1}{2!}\left.\frac{d^2E}{d\varepsilon_1^2}\right|_{\varepsilon_1=0} \varepsilon_1^2 + \frac{1}{2!}\left.\frac{d^2E}{d\varepsilon_1 d\varepsilon_2}\right|_{\varepsilon_1=0,\varepsilon_2=0} \varepsilon_1\varepsilon_2 + \frac{1}{2!}\left.\frac{d^2E}{d\varepsilon_2^2}\right|_{\varepsilon_2=0} \varepsilon_2^2 + \dots$$

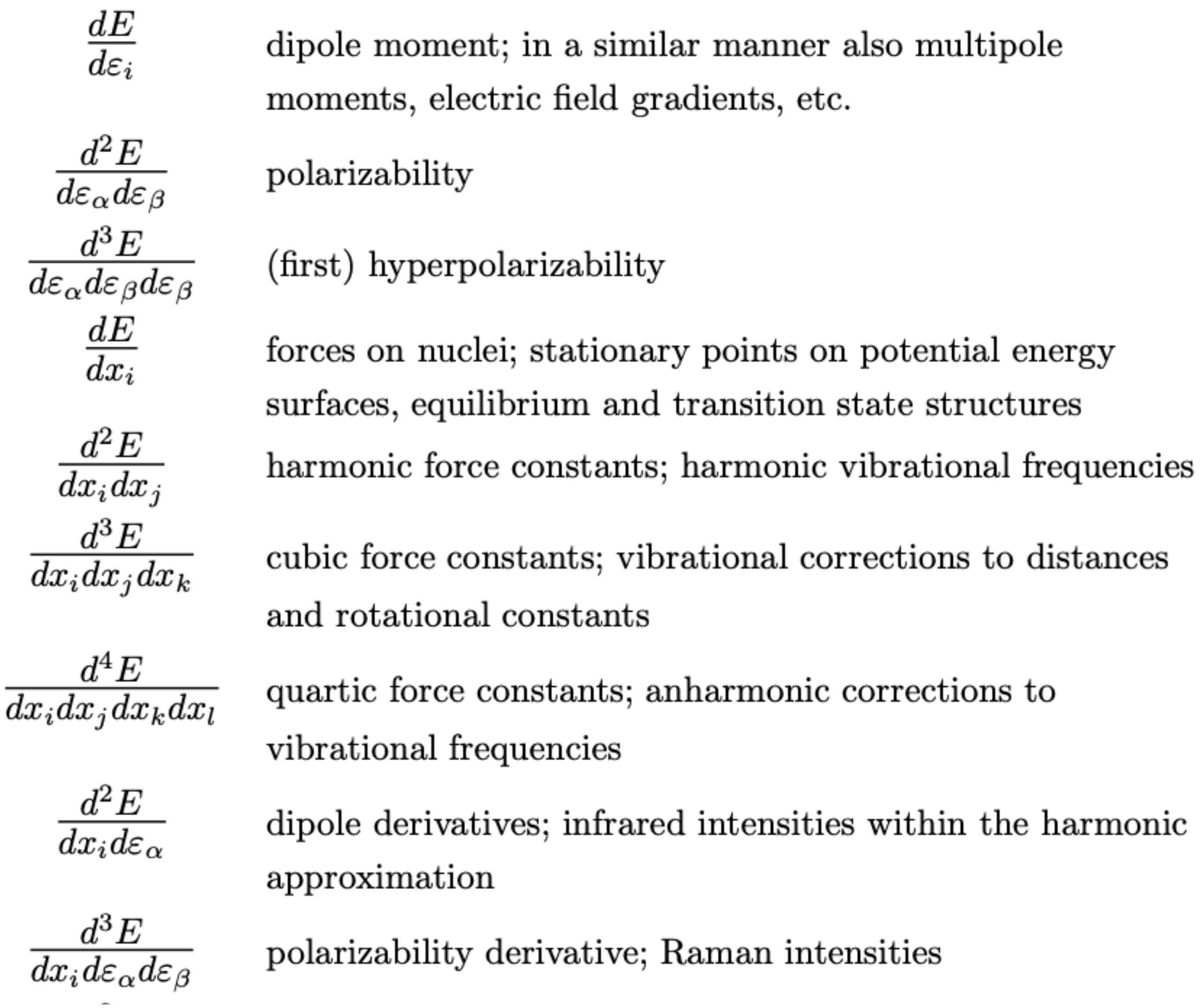
- When the external perturbation (ε) is the electric field

$$\begin{array}{llll}
\text{dipole moment } (\mu) & \hat{=} & -\left.\frac{dE}{d\varepsilon}\right|_{\varepsilon=0} & \text{(first derivative)} \\
\text{polarizability } (\alpha) & \hat{=} & -\left.\frac{d^2E}{d\varepsilon^2}\right|_{\varepsilon=0} & \text{(second derivative)} \\
\text{first hyperpolarizability } (\beta) & \hat{=} & -\left.\frac{d^3E}{d\varepsilon^3}\right|_{\varepsilon=0} & \text{(third derivative)}
\end{array}$$

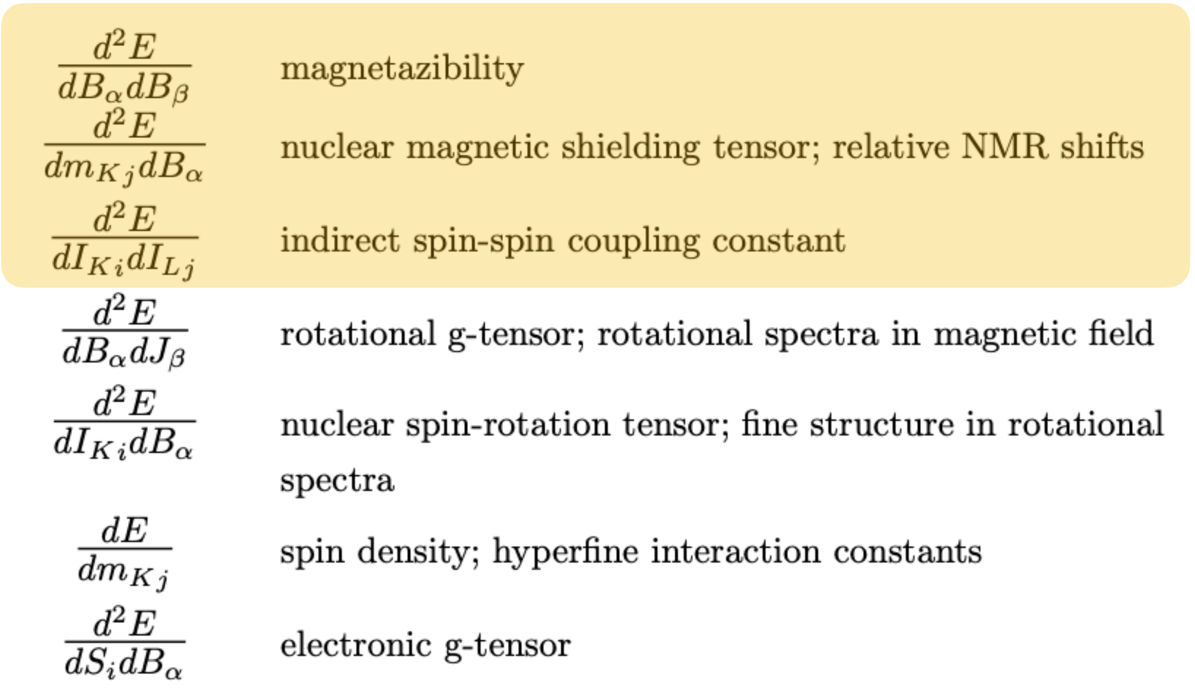
J. Gauss, in http://www.fz-juelich.de/nic-series/ (2000)

A. Hinchcliffe, *Ab initio determination of Molecular Properties* (1987)

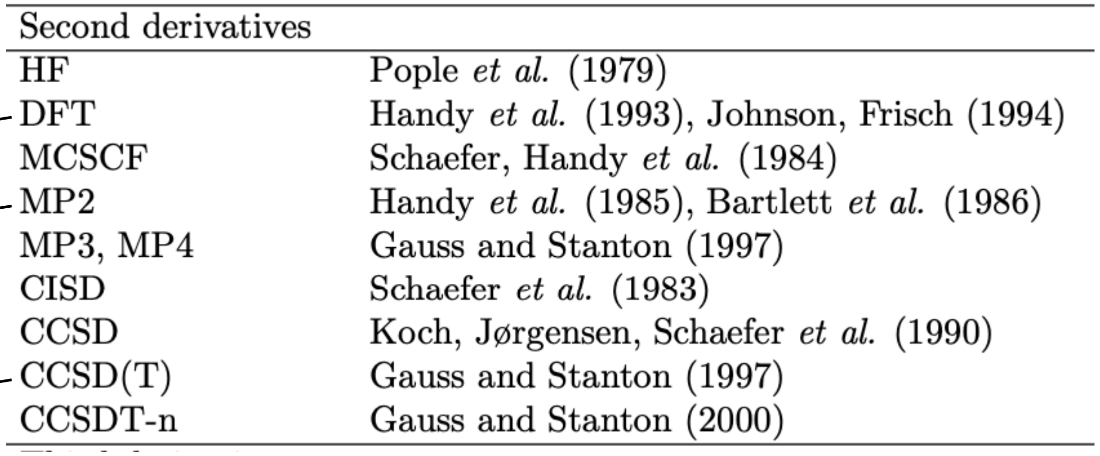
Availability of response equations (analytic second derivatives) for NMR



$\frac{dE}{d\varepsilon_i}$	dipole moment; in a similar manner also multipole moments, electric field gradients, etc.
$\frac{d^2E}{d\varepsilon_\alpha d\varepsilon_\beta}$	polarizability
$\frac{d^3E}{d\varepsilon_\alpha d\varepsilon_\beta d\varepsilon_\beta}$	(first) hyperpolarizability
$\frac{dE}{dx_i}$	forces on nuclei; stationary points on potential energy surfaces, equilibrium and transition state structures
$\frac{d^2E}{dx_i dx_j}$	harmonic force constants; harmonic vibrational frequencies
$\frac{d^3E}{dx_i dx_j dx_k}$	cubic force constants; vibrational corrections to distances and rotational constants
$\frac{d^4E}{dx_i dx_j dx_k dx_l}$	quartic force constants; anharmonic corrections to vibrational frequencies
$\frac{d^2E}{dx_i d\varepsilon_\alpha}$	dipole derivatives; infrared intensities within the harmonic approximation
$\frac{d^3E}{dx_i d\varepsilon_\alpha d\varepsilon_\beta}$	polarizability derivative; Raman intensities



$\frac{d^2E}{dB_\alpha dB_\beta}$	magnetazibility
$\frac{d^2E}{dm_{K_j} dB_\alpha}$	nuclear magnetic shielding tensor; relative NMR shifts
$\frac{d^2E}{dI_{K_i} dI_{L_j}}$	indirect spin-spin coupling constant
$\frac{d^2E}{dB_\alpha dJ_\beta}$	rotational g-tensor; rotational spectra in magnetic field
$\frac{d^2E}{dI_{K_i} dB_\alpha}$	nuclear spin-rotation tensor; fine structure in rotational spectra
$\frac{dE}{dm_{K_j}}$	spin density; hyperfine interaction constants
$\frac{d^2E}{dS_i dB_\alpha}$	electronic g-tensor



Second derivatives	
HF	Pople *et al.* (1979)
DFT	Handy *et al.* (1993), Johnson, Frisch (1994)
MCSCF	Schaefer, Handy *et al.* (1984)
MP2	Handy *et al.* (1985), Bartlett *et al.* (1986)
MP3, MP4	Gauss and Stanton (1997)
CISD	Schaefer *et al.* (1983)
CCSD	Koch, Jørgensen, Schaefer *et al.* (1990)
CCSD(T)	Gauss and Stanton (1997)
CCSDT-n	Gauss and Stanton (2000)

Gaussian, NWCHEM, GAMESS, Orca, Molpro, Qchem, and other programs for molecules CASTEP, VASP, Quantum Espresso for solids

CFOUR for molecules

J. Gauss, in http://www.fz-juelich.de/nic-series/ (2000)

Calculation of NMR shielding

Nuclear magnetic shielding tensor of nucleus *A*

$$\sigma_{\alpha\beta}^{(A)} = \frac{\partial^2 E}{\partial B_\alpha \, \partial m_{A,\beta}}$$

E : total electronic energy

α, β : components x, y, z

$\mathbf{B}$: external magnetic field vector

$\mathbf{m}_A$: magnetic moment of nucleus A

$$\sigma_{\alpha\beta} = \sigma_{\alpha\beta}^{\text{dia}} + \sigma_{\alpha\beta}^{\text{para}}$$

$$\sigma_{\alpha\beta}^{\text{dia}} = \langle \Psi_0 | \hat{O}_{\text{dia}}^{\alpha\beta} | \Psi_0 \rangle$$

$$\sigma_{\alpha\beta}^{\text{para}} = -2 \sum_{n \neq 0} \frac{\langle \Psi_0 | \hat{P}_{\text{para}}^{\alpha} | \Psi_n \rangle \langle \Psi_n | \hat{Q}_{\text{para}}^{\beta} | \Psi_0 \rangle + \alpha \leftrightarrow \beta}{E_n - E_0}$$

$$\hat{O}_{\text{dia}}^{\alpha\beta} = \sum_{i \in \text{electrons}} \left(\mathbf{r}_i \cdot \mathbf{r}_{iA} \, \delta_{\alpha\beta} - r_{i\alpha} r_{iA\beta} \right) r_{iA}^{-3}; \quad \mathbf{r}_{iA} = \mathbf{r}_i - \mathbf{R}_A$$

$$\hat{P}_{\text{para}}^{\alpha} = \sum_{i \in \text{electrons}} L_{i\alpha}; \qquad \mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i$$

$$\hat{Q}_{\text{para}}^{\alpha} = \sum_{i \in \text{electrons}} L_{iA\alpha} r_{iA}^{-3}; \qquad \mathbf{L}_{iA} = \left(\mathbf{r}_i - \mathbf{R}_A \right) \times \mathbf{p}_i$$

$$\sigma_{\alpha\beta}^{(A)} = \sum_{\mu,\nu \in \text{AOs}} \frac{\partial^2 h_{\mu\nu}}{\partial B_\alpha \partial m_{A,\beta}} + \sum_{\mu\nu \in \text{AOs}} \frac{\partial D_{\mu\nu}}{\partial B_\alpha} \frac{\partial h_{\mu\nu}}{\partial m_{A,\beta}}$$

Linear response form used in calculations

R. Ditchfield, *Molecular Physics* (1973)

Electronic Hamiltonian of a molecule in magnetic field

- Sum of one- and two-electron terms

$$\hat{H} = \sum_i \hat{h}(i) + \sum_{i<j} \hat{g}(i,j)$$

  - ○ $\hat{g}(i,j) = 1/r_{ij}$: two-electron operator, Coulomb
  - ○ $\hat{h}(i)$: one-electron operator contains magnetic field

$$\hat{h}(i) = \frac{1}{2}\left(\hat{\mathbf{p}}_i + \mathbf{A}(\mathbf{r}_i)\right)^2 - \sum_A \left[\frac{Z_A}{|\,\mathbf{r}_i - \mathbf{R}_A\,|} - \mathbf{m}_A \cdot \mathbf{B}(\mathbf{r}_i)\right]$$

magnetic vector potential, $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})$

$$\frac{1}{2}\left(\hat{\mathbf{p}} + \mathbf{A}\right)^2 = \frac{\hat{\mathbf{p}}^2}{2} + \underbrace{\frac{1}{2}\left(\hat{\mathbf{p}} \cdot \mathbf{A} + \mathbf{A} \cdot \hat{\mathbf{p}}\right)}_{\textbf{paramagnetic term}} + \underbrace{\frac{1}{2}\mathbf{A}^2}_{\textbf{diamagnetic term}}$$

- Pure gauge transformation (uniform field) $\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r}) + \nabla\chi(\mathbf{r})$, $\mathbf{B} = \mathbf{B}_0$ unchanged

  - ○ $\mathbf{B}' = \nabla \mathbf{A}' = \nabla \times \left(\mathbf{A} + \nabla\chi\right) = \nabla \times \mathbf{A} + \nabla \times \nabla\chi = \nabla \times \mathbf{A} + 0 = \mathbf{B}$

Gauge-origin dependence in finite basis set

- Gauge independence of exact wavefunction
  - Let $\hat{H}\Psi = \frac{1}{2}\left(\hat{\mathbf{p}} + \mathbf{A}\right)^2 \Psi + V(\mathbf{r})\Psi$
  - $\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r}) + \nabla\chi(\mathbf{r}) \longrightarrow \hat{H}(\mathbf{A}')\Psi' = \frac{1}{2}\left(\hat{\mathbf{p}} + \mathbf{A} + \nabla\chi\right)^2 \Psi' + V(\mathbf{r})\Psi'$
  - $\nabla \times \mathbf{A}' = \nabla \times \mathbf{A} \longrightarrow \Psi'(\mathbf{r}) = e^{i\chi(\mathbf{r})}\Psi(\mathbf{r}) \implies \langle\Psi|\hat{H}|\Psi\rangle = \langle\Psi'|\hat{H}'|\Psi'\rangle$

- Shifting the global origin of the vector potential

  Let $\mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{r} - \mathbf{R}_0)$, and $\mathbf{A}'(\mathbf{r}) = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{r} - \mathbf{R}_0 + d\mathbf{R}_0) = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{r} - \mathbf{R}_0')$

  $\mathbf{A}'(\mathbf{r}) - \mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{B}_0 \times (\mathbf{R}_0 - \mathbf{R}_0') = \nabla\chi(\mathbf{r})$; where $\chi(\mathbf{r}) = \frac{1}{2}\left[\mathbf{B}_0 \times \left(\mathbf{R}_0 - \mathbf{R}_0'\right)\right] \cdot \mathbf{r}$

- MOs expanded on a finite basis of AOs: $\psi(\mathbf{r}) = \sum_{\mu} c_\mu \phi_\mu(\mathbf{r})$
  - Finite AO basis cannot represent $e^{i\chi(\mathbf{r})}\phi_\mu(\mathbf{r})$, i.e., cannot be expanded as $\sum_{\mu} d_\mu \phi_\mu(\mathbf{r})$
  - For magnetic properties, gauge-including AOs (GIAOs) are used

    $$\phi_\mu^{\mathrm{GIAO}}(\mathbf{r}) = \exp\left[\frac{i}{2}\left(\mathbf{B}_0 \times \mathbf{R}_\mu\right) \cdot \mathbf{r}\right] \phi_\mu(\mathbf{r})$$

    GIAO removes dependence on $\mathbf{R}_0$ by attaching the correct phase to each AO.

Linear response and coupled-perturbed equations

- For Hartree-Fock formalism

$$\sigma^{(A)}_{\alpha\beta} = \sum_{\mu,\nu\in\mathrm{AOs}} \frac{\partial^2 h_{\mu\nu}}{\partial B_\alpha \partial m_{\mathrm{A},\beta}} + \sum_{\mu\nu\in\mathrm{AOs}} \boxed{\frac{\partial D^{\mathrm{HF}}_{\mu\nu}}{\partial B_\alpha}} \frac{\partial h_{\mu\nu}}{\partial m_{\mathrm{A},\beta}}$$

implicit dependence on $\mathbf{B}$

$$D^{\mathrm{HF}}_{\mu\nu} = \sum_i^{\mathrm{occ}} C_{\mu i} C_{\nu i}$$

charge-density bond-order matrix

$$\psi_i^{\mathrm{MO}}(\mathbf{r}) = \sum_\mu c_{\mu i} \phi_\mu^{\mathrm{basis}}(\mathbf{r})$$

e.g. LCGTO (Gaussian-type orbitals)

$$\frac{D^{\mathrm{HF}}_{\mu\nu}}{\partial B_\alpha} = \sum_i^{\mathrm{occ}} \left(C^{(0)}_{\mu i} \boxed{\frac{\partial C_{\nu i}}{\partial B_\alpha}} + \frac{\partial C_{\mu i}}{\partial B_\alpha} C^{(0)}_{\nu i} \right)$$

solution of coupled-perturbed Hartree-Fock (CPHF) equations

$$\left(\mathbf{F} - \varepsilon_i \mathbf{S}\right) \frac{\partial \mathbf{C}_i}{\partial B_\alpha} = - \left(\frac{\partial \mathbf{F}}{\partial B_\alpha} - \frac{\partial \varepsilon_i}{\partial B_\alpha} \mathbf{S} \right) \mathbf{C}_i^{(0)}$$

In Kohn-Sham DFT, coupled-perturbed Kohn-Sham equations are solved

Post-Hartree-Fock corrections (follow energy corrections)

- MP2 correlation correction to HF energy

$$E^{(2)} = \frac{1}{4}\sum_{ijab} \frac{|\langle ij||ab\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}$$

  - i, j: occupied MOs from HF
  - a, b: virtual MOs from HF
  - ε_p: energies of MOs from HF
  - $\langle ij||ab\rangle$: antisymmetrized two-electron integrals

- Using double-excitation amplitudes

$$t_{ij}^{ab} = \frac{\langle ij||ab\rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} = \frac{V_{ijab}}{\Delta_{ij}^{ab}} \qquad E^{(2)} = \frac{1}{4}\sum_{ijab} t_{ij}^{ab}$$

- MP2 correlation correction to HF shielding tensor

$$\sigma_{\alpha\beta}^{(2)} = \frac{\partial^2 E^{(2)}}{\partial B_\alpha \partial m_{A,\beta}} = \frac{1}{4}\sum_{ijab} \frac{\partial^2}{\partial B_\alpha \partial m_{A,\beta}} \left[t_{ij}^{ab} V_{ijab} \right]$$

$$\sigma_{\alpha\beta}^{(2)} = \frac{1}{4}\sum_{ijab} \left[\frac{\partial^2 t_{ij}^{ab}}{\partial B_\alpha \partial m_{A,\beta}} V_{ijab} + \frac{\partial t_{ij}^{ab}}{\partial B_\alpha} \frac{\partial V_{ijab}}{\partial m_{A,\beta}} + \frac{\partial V_{ijab}}{\partial B_\alpha} \frac{\partial t_{ij}^{ab}}{\partial m_{A,\beta}} + \frac{\partial^2 V_{ijab}}{\partial B_\alpha \partial m_{A,\beta}} t_{ij}^{ab} \right]$$

Requires CP-MP2 equations.

E. C. Vauthier, *Can. J. Chem.* 66, 1781 (1988)

Coupled-cluster theory for high-precision quantum chemistry

$|\Phi\rangle = e^{\hat{C}}|\Psi_0\rangle$ $\hat{C}$ is the cluster operator, and $e^{\hat{C}} = 1 + \hat{C} + \frac{1}{2!}\hat{C}^2 + \frac{1}{3!}\hat{C}^3 + \dots$

The effect of the cluster operator is defined as the sum

$$\hat{C}|\Psi_0\rangle = \hat{C}_1|\Psi_0\rangle + \hat{C}_2|\Psi_0\rangle + \ldots + \hat{C}_N|\Psi_0\rangle$$

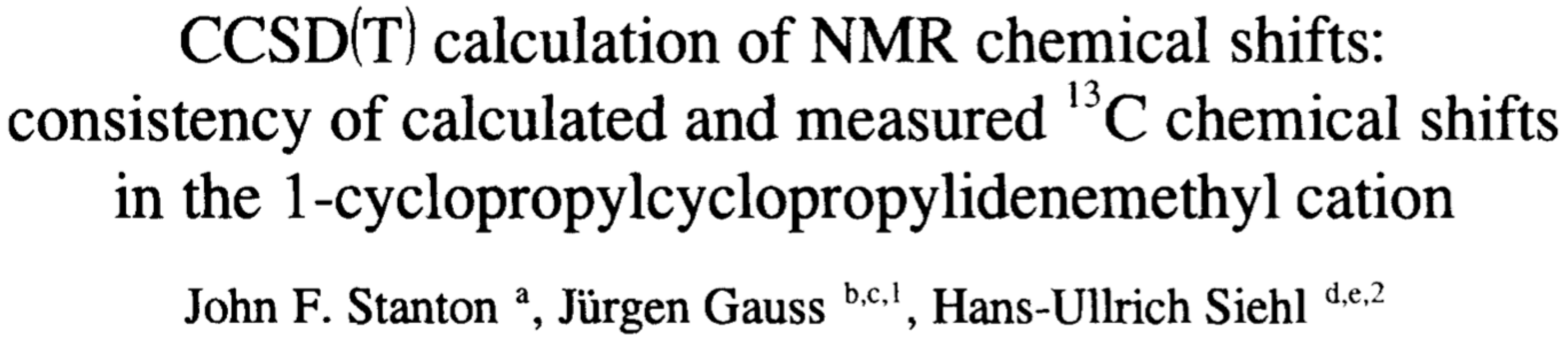
with

$$\hat{C}_1|\Psi_0\rangle = \sum_{a,r} t_a^r|\Psi_a^r\rangle, \quad \hat{C}_2|\Psi_0\rangle = \sum_{a,b,r,s} t_{a,b}^{r,s}|\Psi_{a,b}^{r,s}\rangle$$

where the excitation amplitudes, t, are determined by a non-linear iterative optimization

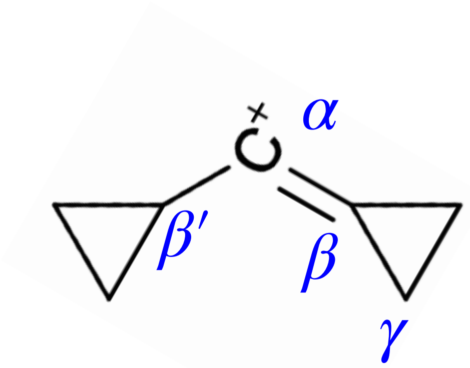
- Truncating the sum gives rise to CCSD, CCSDT, etc., approximations. One of the most popular approximations is CCSD(T), where the triples energy correction is estimated using the perturbation theory. CCSD(T) with a large LCAO expansion is commonly used for accurate modelling of spectroscopic properties (very) small molecules.

CCSD(T) settles the disputes based on approximate modeling



CCSD(T) calculation of NMR chemical shifts: consistency of calculated and measured ^{13}C chemical shifts in the 1-cyclopropylcyclopropylidenemethyl cation

John F. Stanton [a], Jürgen Gauss [b,c,1], Hans-Ullrich Siehl [d,e,2]



- An earlier MP2 calculation suggested geometric reorganization in the solvent to influence the ^{13}C shift of C_α

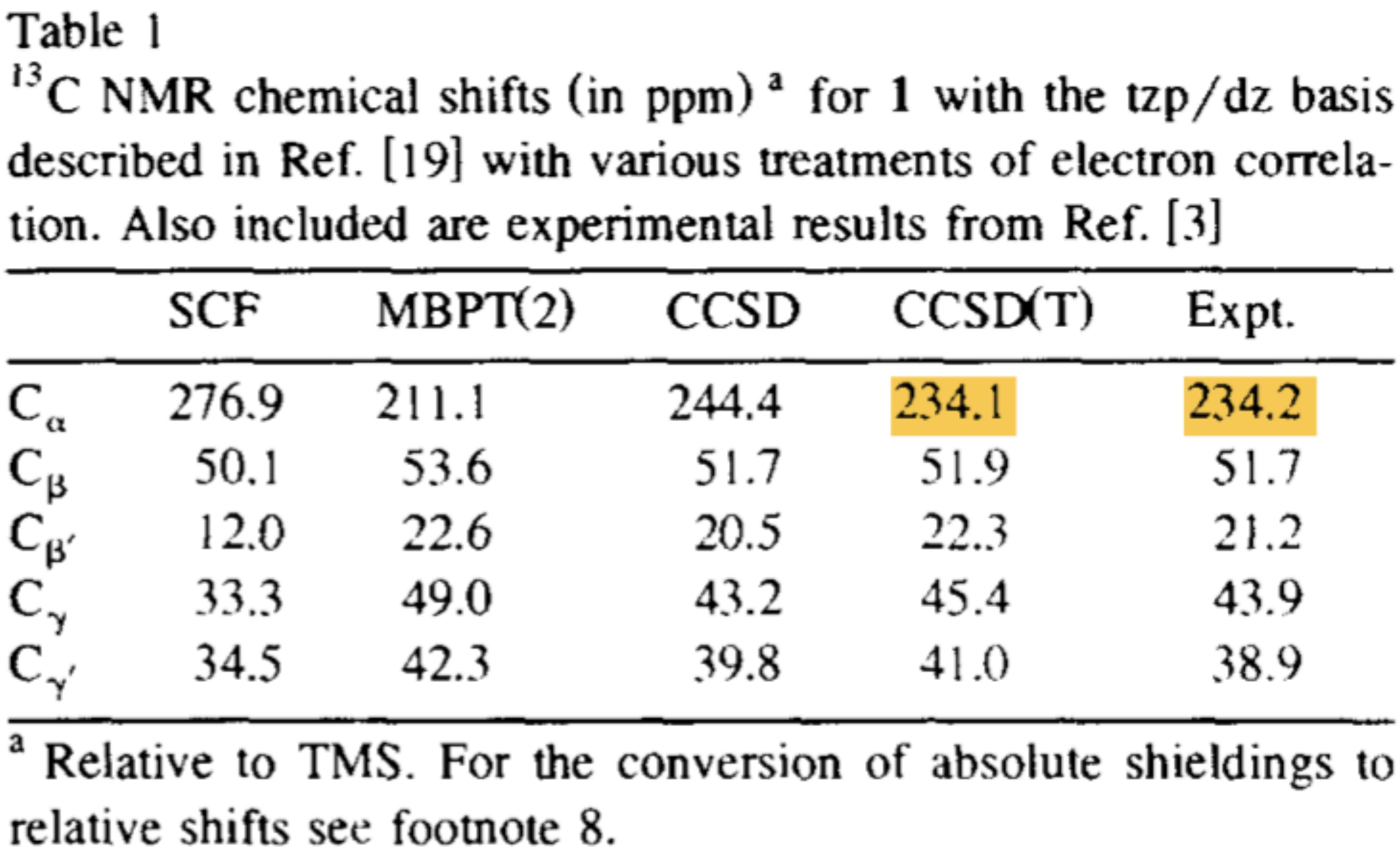


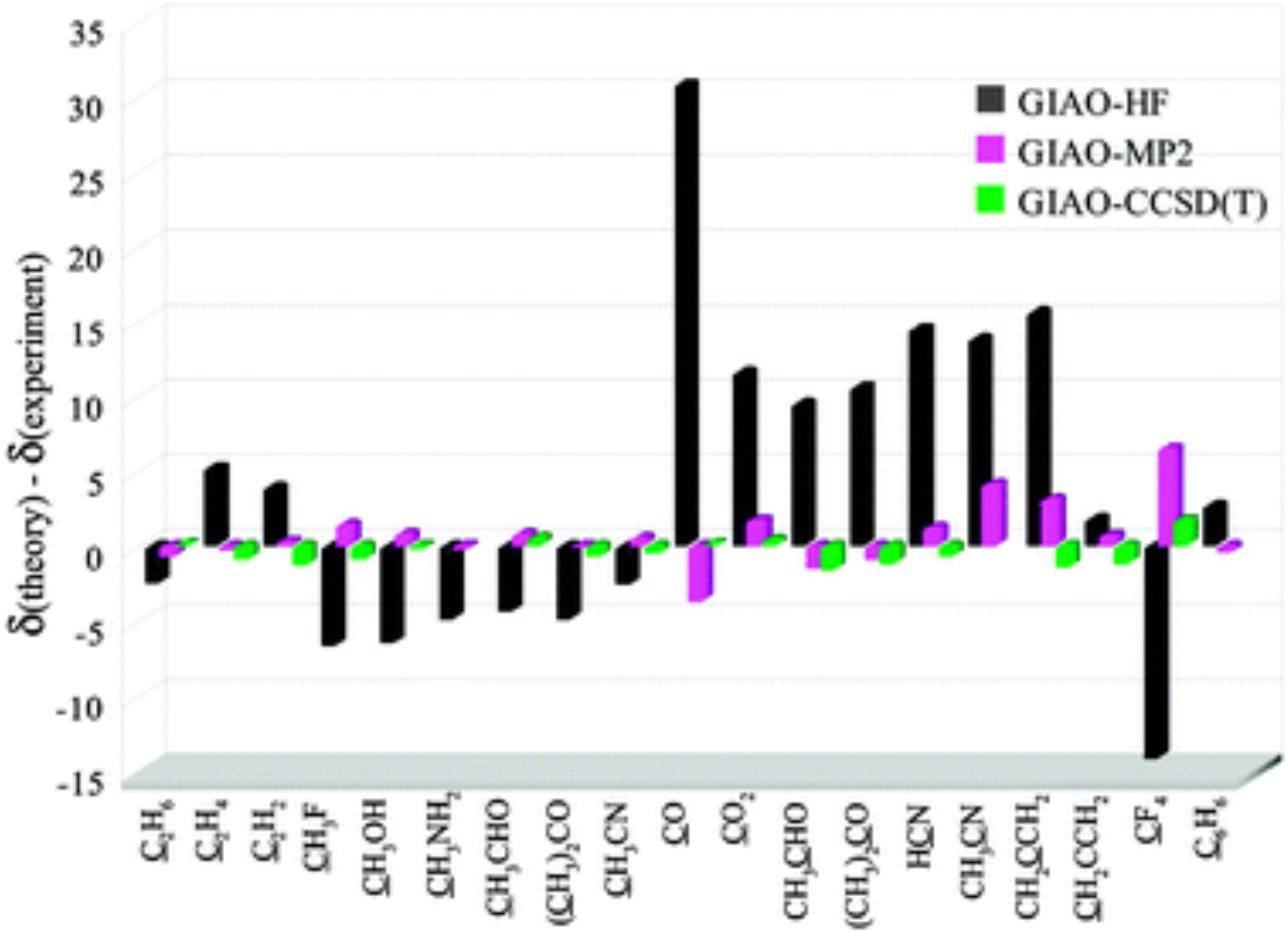
Table 1
^{13}C NMR chemical shifts (in ppm) [a] for **1** with the tzp/dz basis described in Ref. [19] with various treatments of electron correlation. Also included are experimental results from Ref. [3]

	SCF	MBPT(2)	CCSD	CCSD(T)	Expt.
C_α	276.9	211.1	244.4	234.1	234.2
C_β	50.1	53.6	51.7	51.9	51.7
$C_{\beta'}$	12.0	22.6	20.5	22.3	21.2
C_γ	33.3	49.0	43.2	45.4	43.9
$C_{\gamma'}$	34.5	42.3	39.8	41.0	38.9

[a] Relative to TMS. For the conversion of absolute shieldings to relative shifts see footnote 8.

J. F. Stanton et al., *Chem. Phys. Lett.* 262, 183 (1996)

Some quantum chemistry benchmarks



MAE

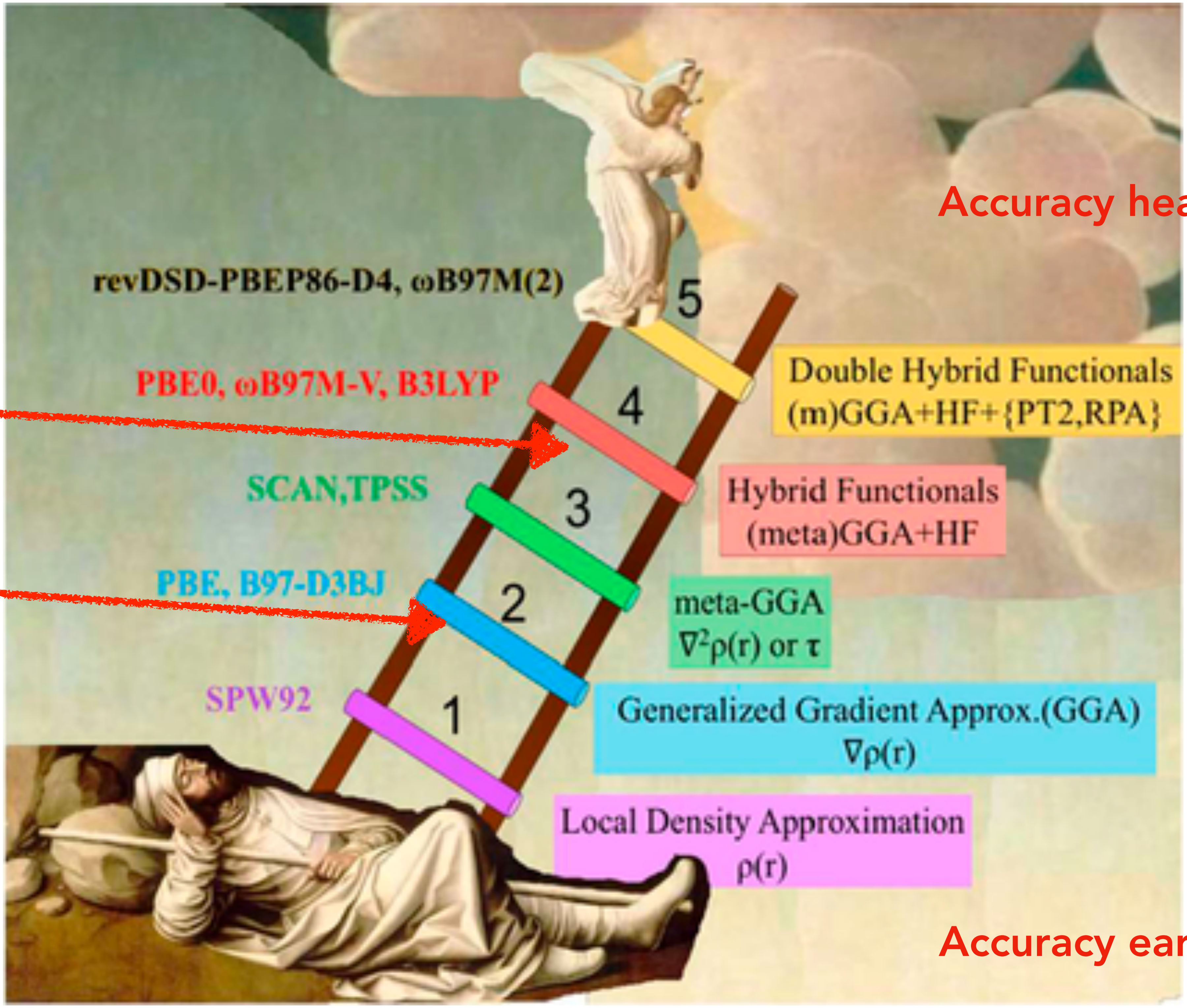
8.8 ppm
1.5 ppm
0.7 ppm

F. A. A. Mulder, *Chem. Soc. Rev.*, 39, 528 (2010)

Jacob’s ladder in DFT approximations

workhorse for molecules

workhorse for solids



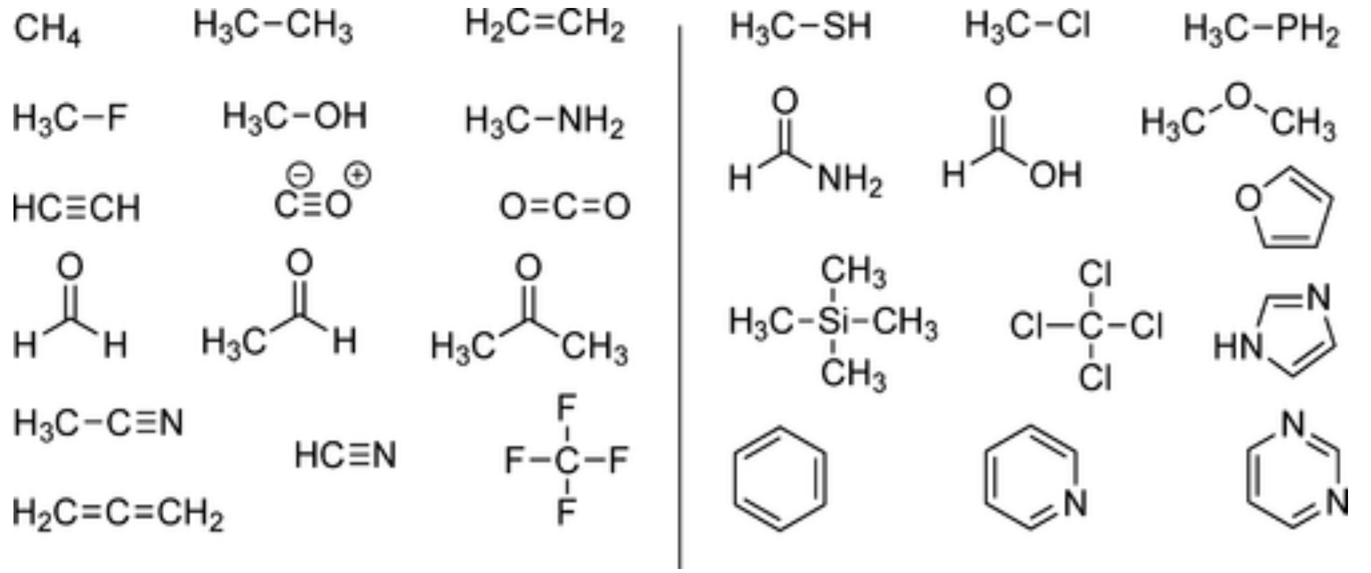
Accuracy heaven

Accuracy earth

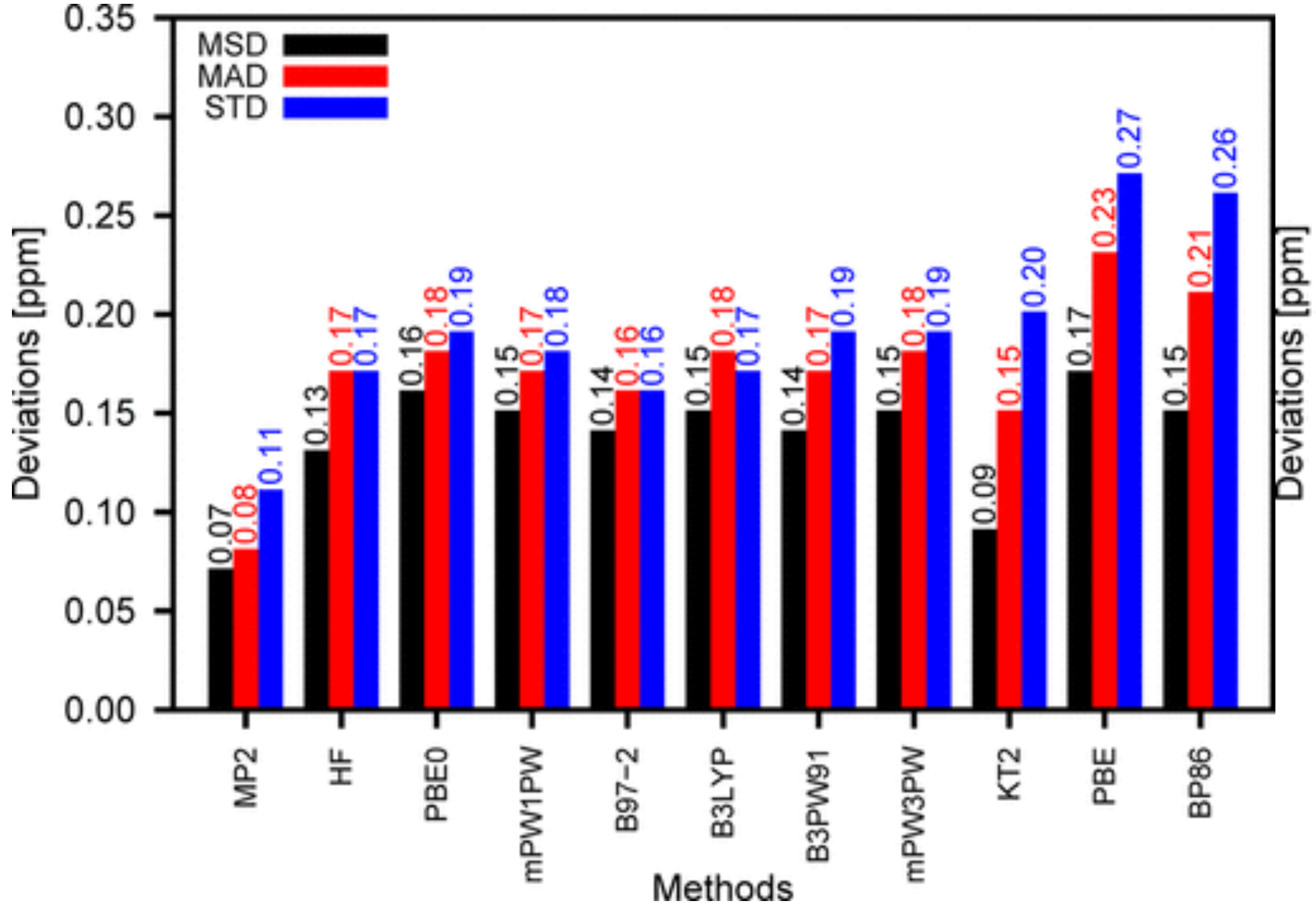
J. M. L. Martin, G. Santra, *Israel J. Chem.* 60, 787 (2020)

J. P. Perdew, K. Schmidt, *AIP Conf. Proc.* 577, 1 (2001)

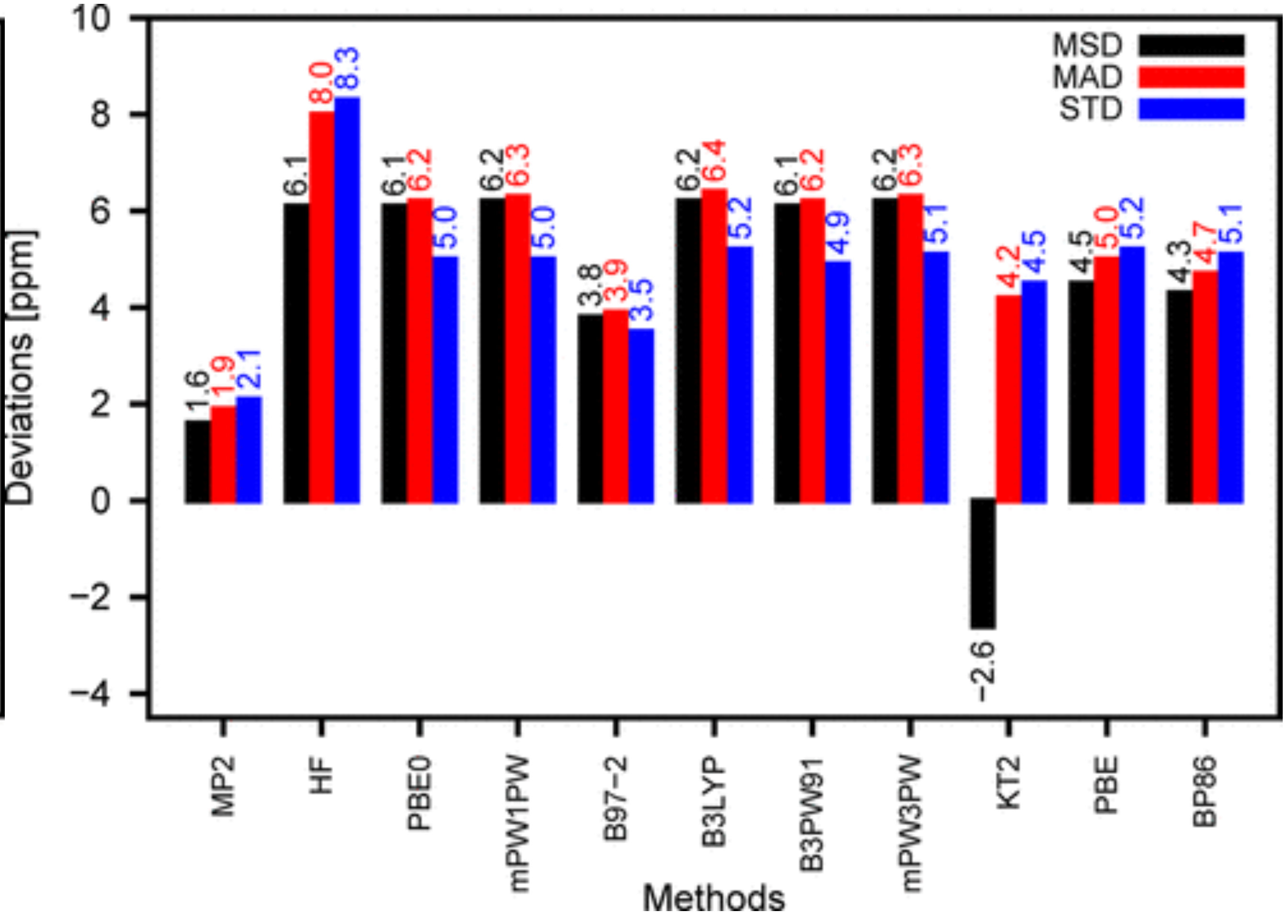
HF, MP2 and DFT benchmarks with CCSD(T) reference



^{1}H

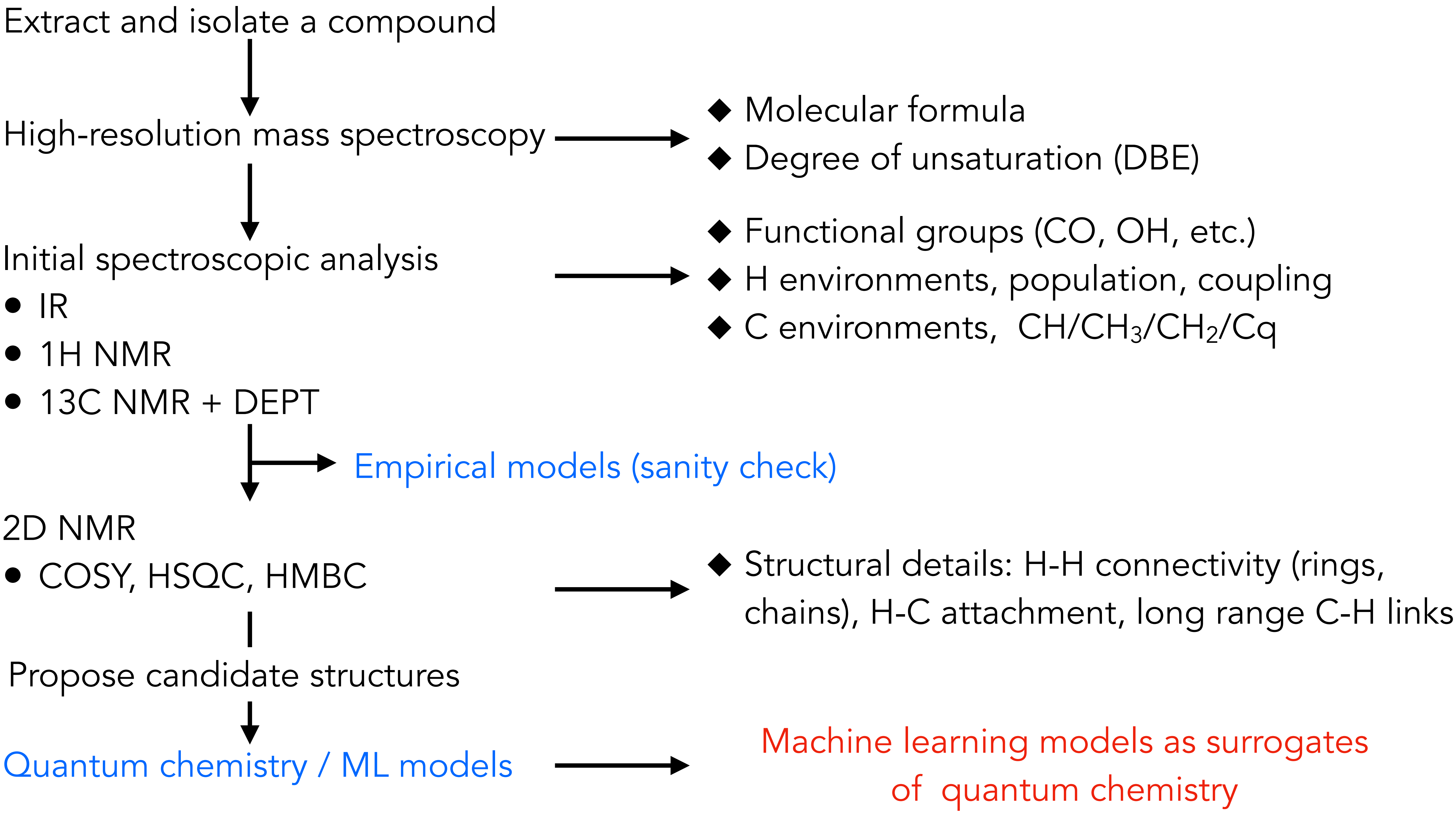


^{13}C



D. Flaig, *J. Chem. Theory Comput.* 10, 572 (2014)

(From synthesis to) Structure elucidation workflow



Extract and isolate a compound

High-resolution mass spectroscopy

- ◆ Molecular formula
- ◆ Degree of unsaturation (DBE)

Initial spectroscopic analysis

- ● IR
- ● 1H NMR
- ● 13C NMR + DEPT

- ◆ Functional groups (CO, OH, etc.)
- ◆ H environments, population, coupling
- ◆ C environments, CH/CH$_3$/CH$_2$/Cq

Empirical models (sanity check)

2D NMR

- ● COSY, HSQC, HMBC

- ◆ Structural details: H-H connectivity (rings, chains), H-C attachment, long range C-H links

Propose candidate structures

Quantum chemistry / ML models

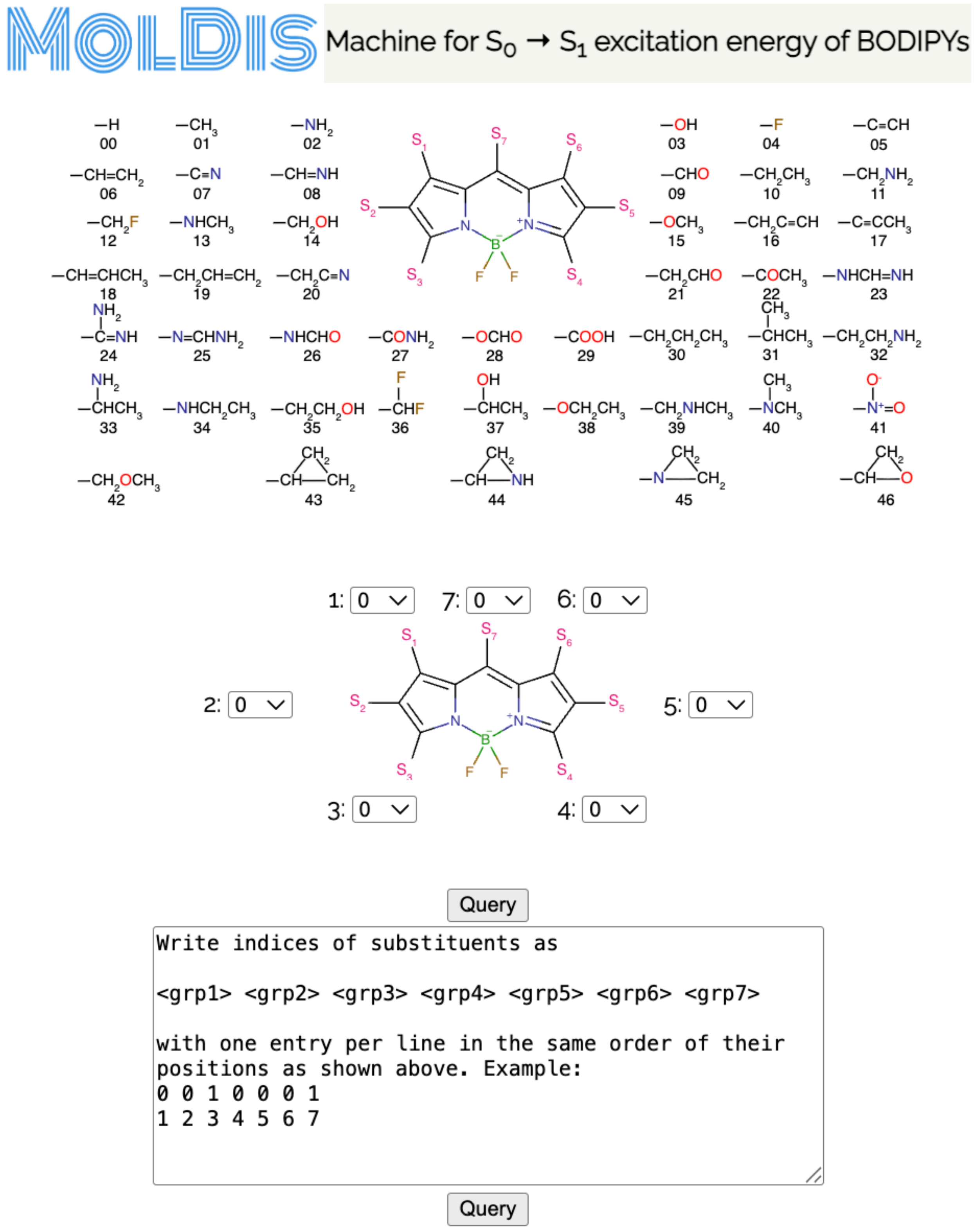
Machine learning models as surrogates of quantum chemistry

Quantum chemistry calculations are time-consuming

The goal is to develop ML models trained on large datasets that offer quantum chemistry accuracy at empirical speed

Example web-app for electronic excitation energy of BODIPY dyes

https://moldis.tifrh.res.in/db/bodipy



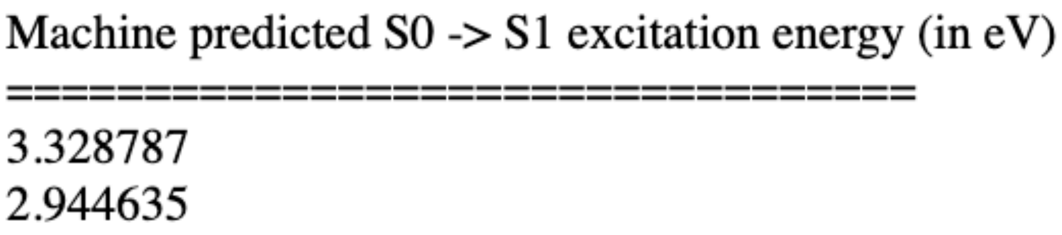
MOLDIS Machine for S_0 → S_1 excitation energy of BODIPYs

1: 0 7: 0 6: 0

2: 0 5: 0

3: 0 4: 0

```
Write indices of substituents as

<grp1> <grp2> <grp3> <grp4> <grp5> <grp6> <grp7>

with one entry per line in the same order of their
positions as shown above. Example:
0 0 1 0 0 0 1
1 2 3 4 5 6 7
```

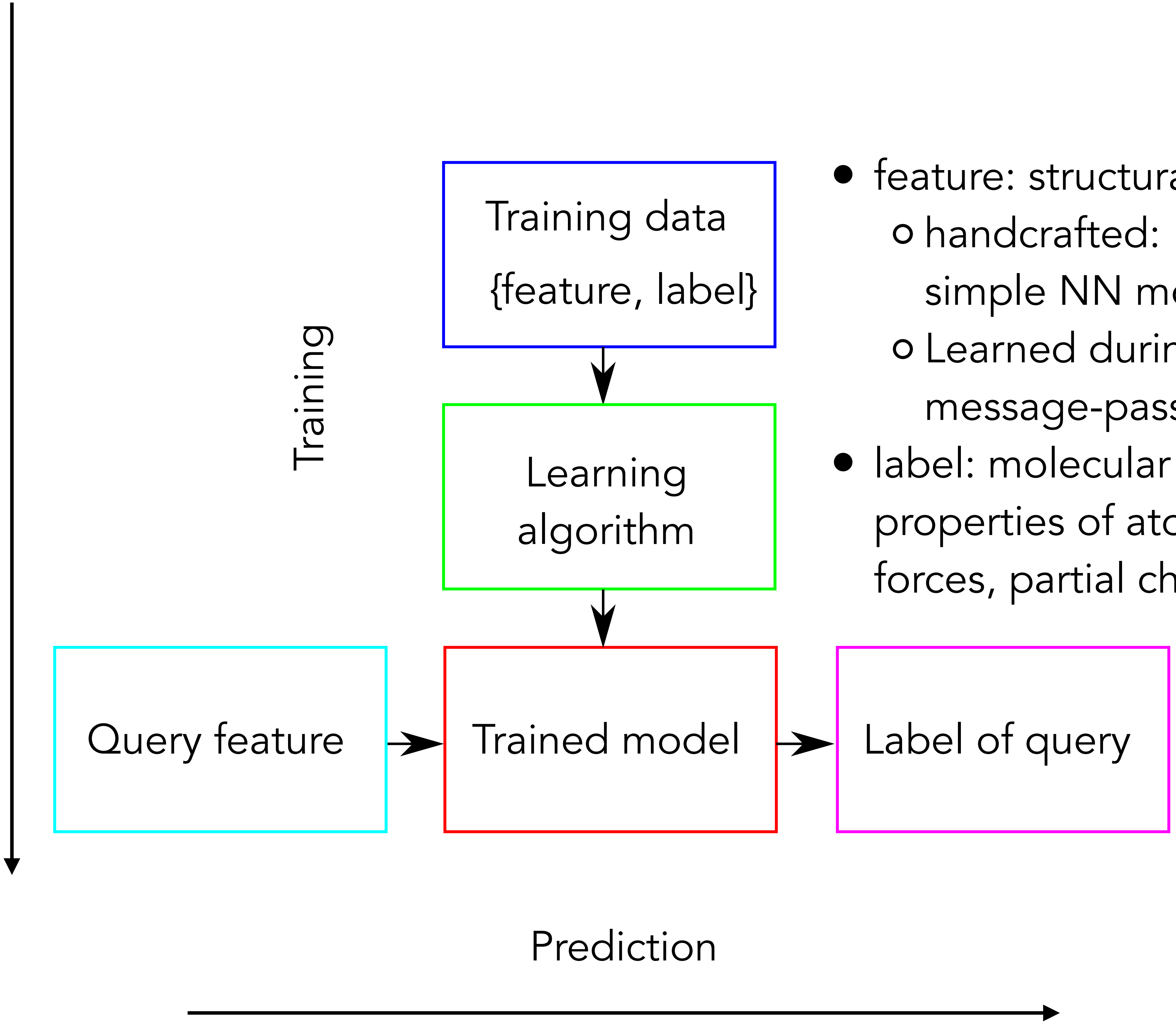


```
Machine predicted S0 -> S1 excitation energy (in eV)
==================================
3.328787
2.944635
```

- ML model trained on TDDFT data with 5-10% error
- Enables rapid screening of 253 Billion possible combinations (empirical speed)

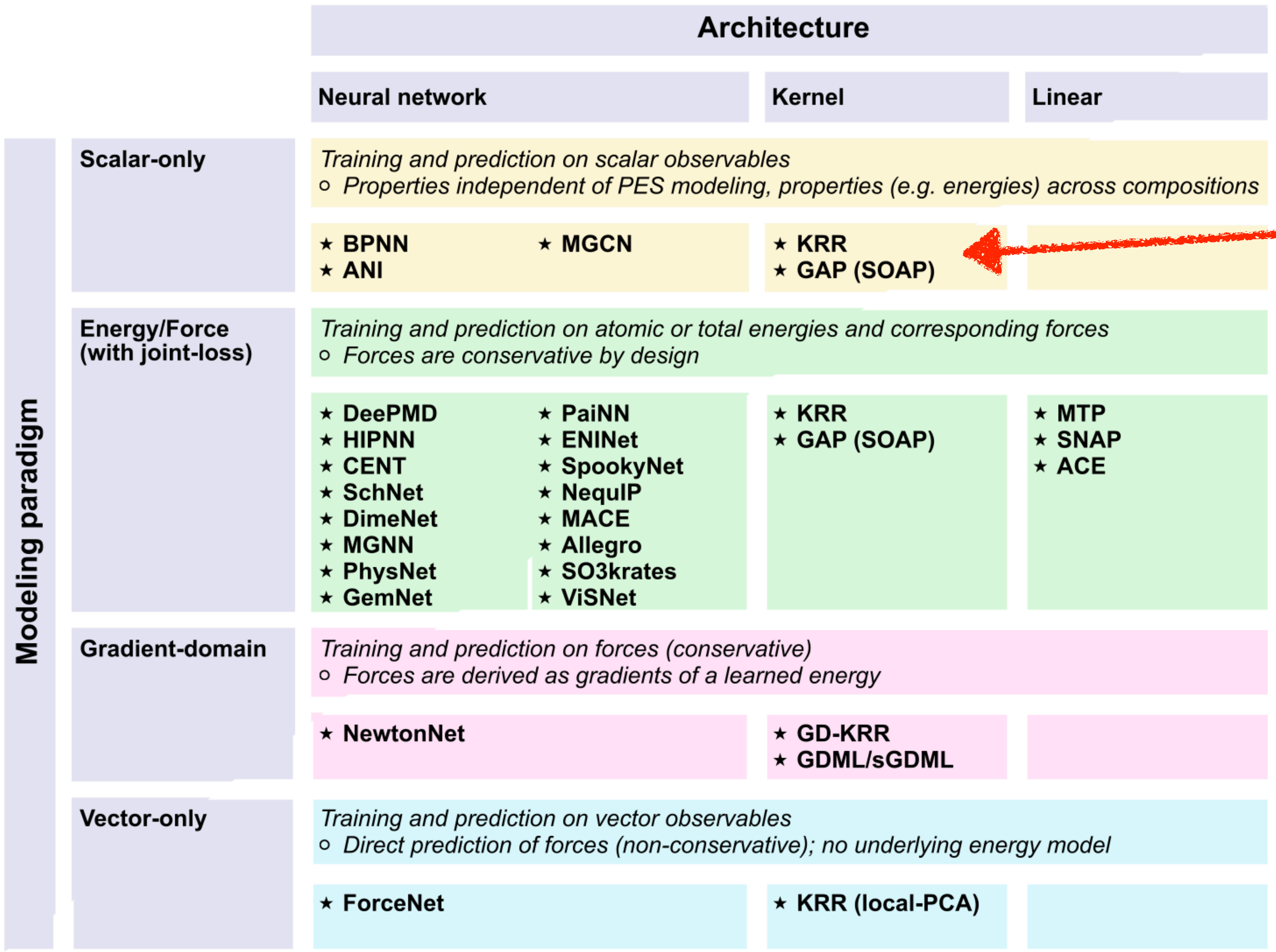
A. Gupta, *J. Chem. Phys.* 155, 244102 (2021)

ML workflow



- feature: structural descriptor of an atom
    - handcrafted:
      simple NN models, Kernel-based models
    - Learned during training:
      message-passing NN
- label: molecular properties (energy), properties of atoms-in-molecules (atomic forces, partial charges, NMR shielding)

Taxonomy of ML models used in chemistry problems

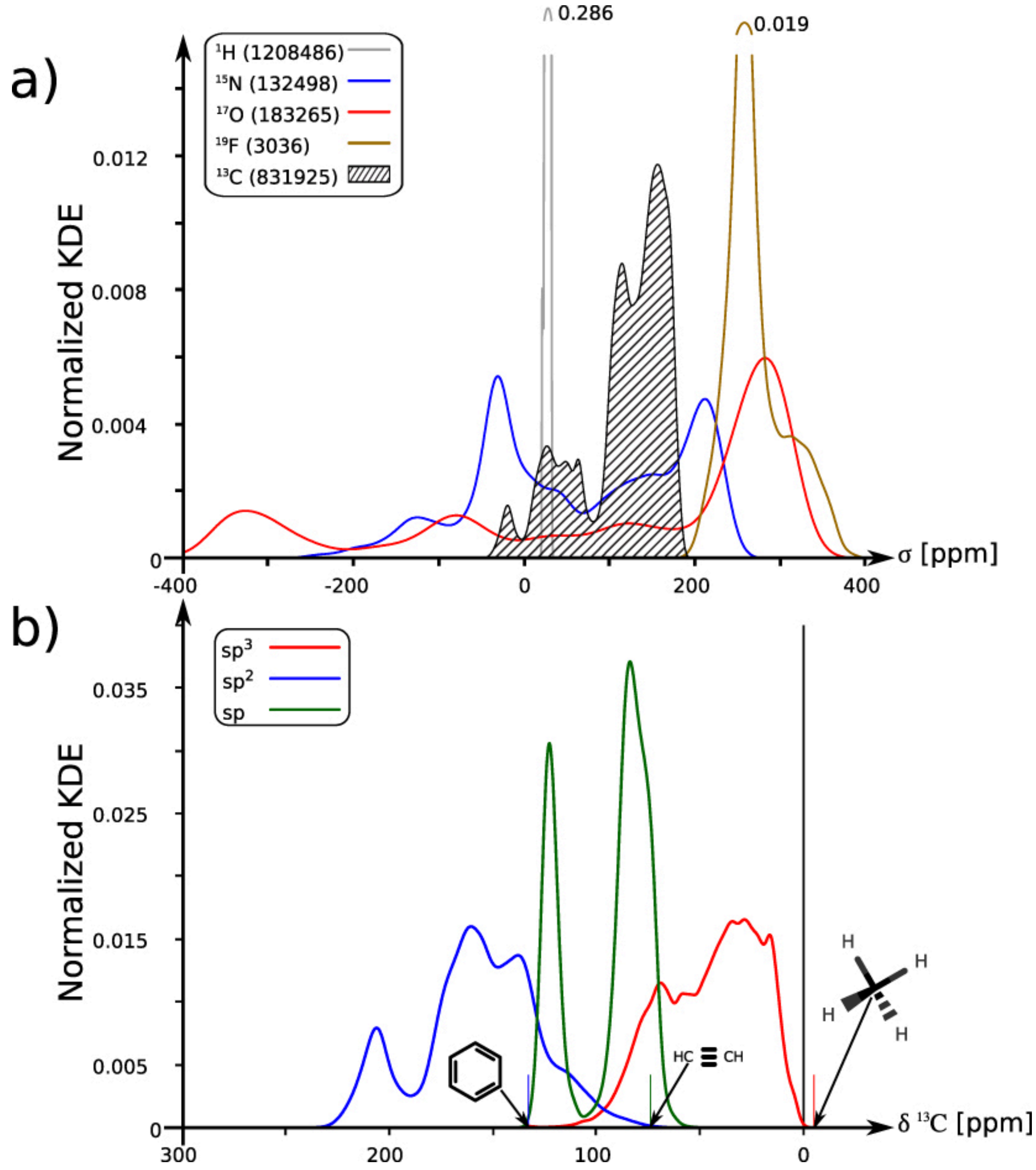


		Architecture		
		Neural network	Kernel	Linear
Modeling paradigm	Scalar-only	*Training and prediction on scalar observables* ○ *Properties independent of PES modeling, properties (e.g. energies) across compositions*		
		★ BPNN ★ ANI ★ MGCN	★ KRR ★ GAP (SOAP)	
	Energy/Force (with joint-loss)	*Training and prediction on atomic or total energies and corresponding forces* ○ *Forces are conservative by design*		
		★ DeePMD ★ HIPNN ★ CENT ★ SchNet ★ DimeNet ★ MGNN ★ PhysNet ★ GemNet ★ PaiNN ★ ENINet ★ SpookyNet ★ NequIP ★ MACE ★ Allegro ★ SO3krates ★ ViSNet	★ KRR ★ GAP (SOAP)	★ MTP ★ SNAP ★ ACE
	Gradient-domain	*Training and prediction on forces (conservative)* ○ *Forces are derived as gradients of a learned energy*		
		★ NewtonNet	★ GD-KRR ★ GDML/sGDML	
	Vector-only	*Training and prediction on vector observables* ○ *Direct prediction of forces (non-conservative); no underlying energy model*		
		★ ForceNet	★ KRR (local-PCA)	

Figure 6. A logical classification of machine-learned potentials (MLPs) organized by modeling paradigm (rows) and architectural class (columns) that distinguishes conservative, force-only, and scalar models while grouping them under neural, kernel, or linear frameworks. All acronyms used here are defined in the main text.

(Topical review) *J..Phys.: Condens. Matter* 38, 013002 (2026)

QM9NMR dataset



- Contains isotropic shielding of CHONF atoms in 134,000 QM9 molecules (with upto 9 CONF atoms).

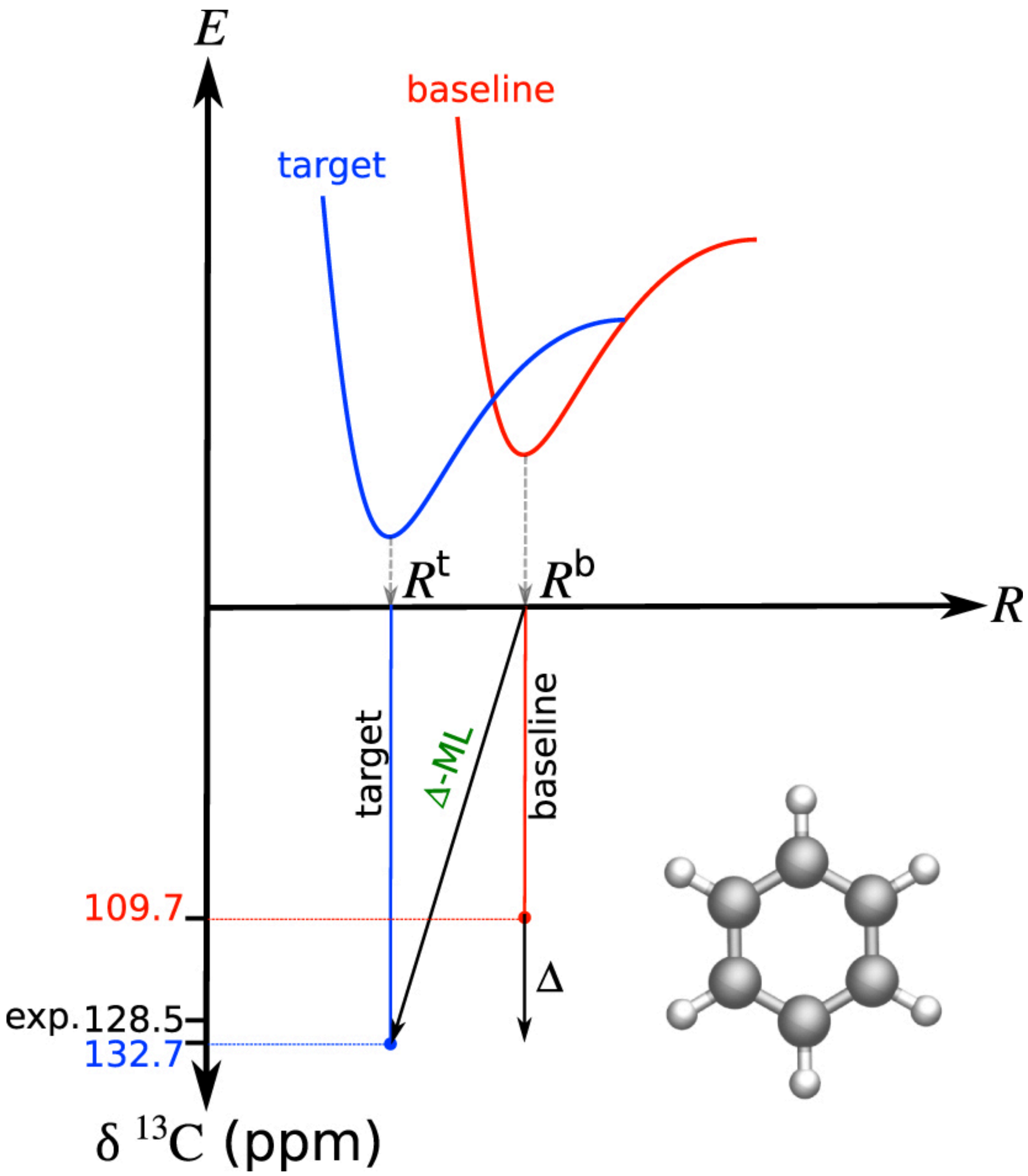
- mPW1PW91/6-311+G(2d,p)@B3LYP/6-31G(2df,p) level DFT modelling in vacuum and with continuum models of five commonly used polar and non-polar organic solvents: acetone, CCl4, DMSO, methanol, and THF

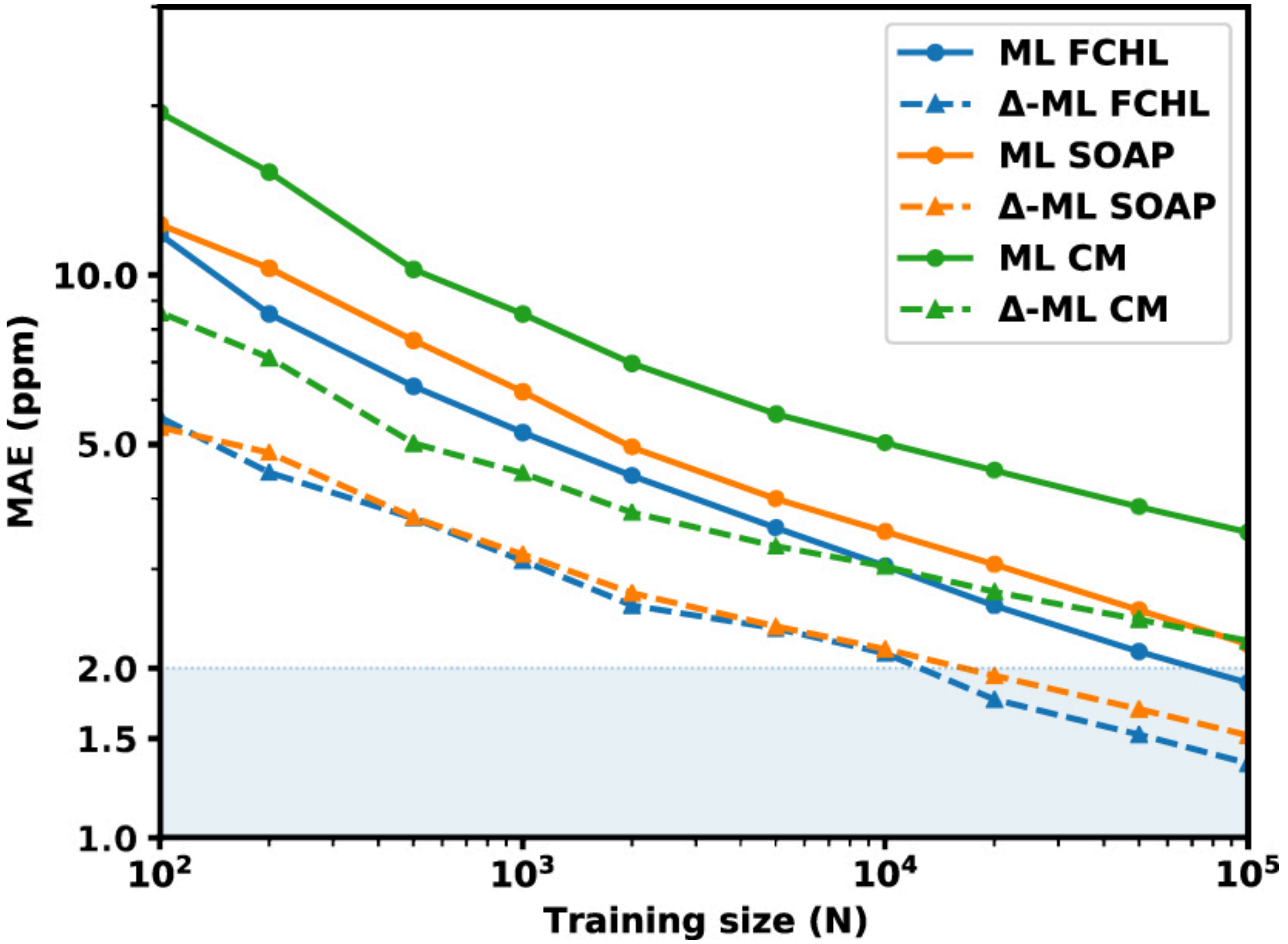
- The dataset contains 0.8 Million ^{13}C shielding and 1.2 Million ^{1}H shielding in each phase

R. Ramakrishnan, et al. Scientific Data (2014).

A. Gupta, et al. Mach. Learn.: Sci. Technol. (2021).

ΔML modeling of ^{13}C isotropic shielding

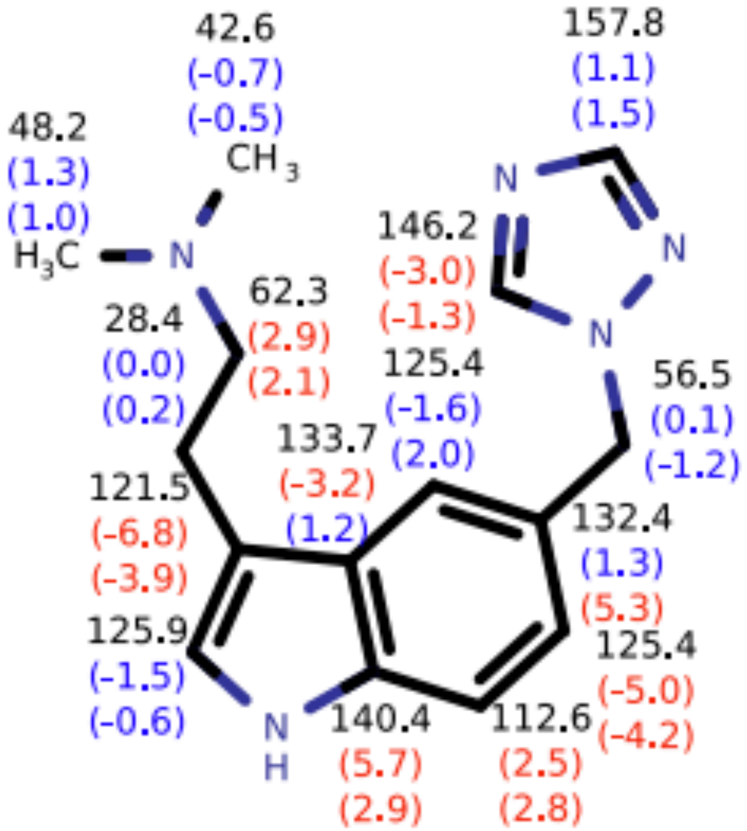




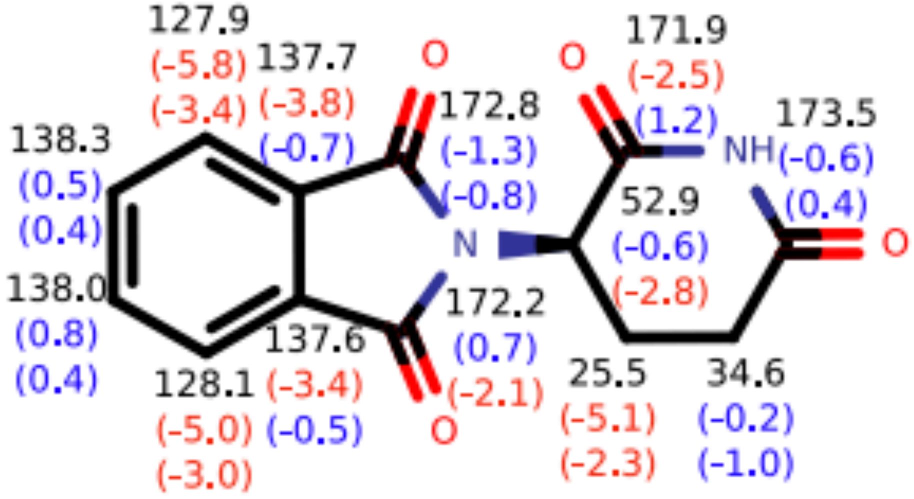
- ΔML modeling is done with PM7 geometries for generating structural descriptors (FCHL, SOAP, CM)
- B3LYP/STO3G level shielding is the baseline
- The targetline is mPW1PW91/6-311+G(2d,p)@B3LYP/6-31G(2df,p)

A. Gupta, et al. Mach. Learn.: Sci. Technol. (2021).

Transferability to large drug molecules unknown to the model



Rizatriptan
ML (2.4, 0.97)
Δ-ML (2.0, 0.97)



Thalidomide
ML (2.3, 0.91)
Δ-ML (1.5, 0.93)

standard deviation of the error

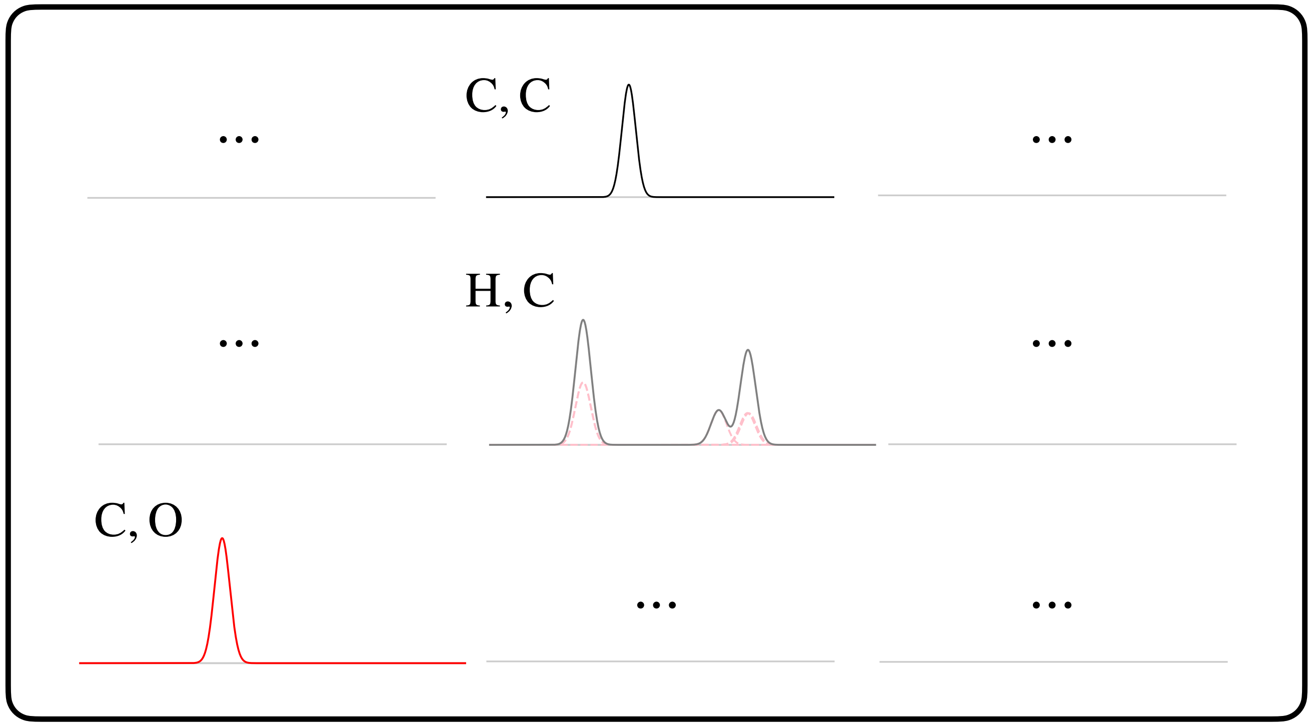
mean absolute error

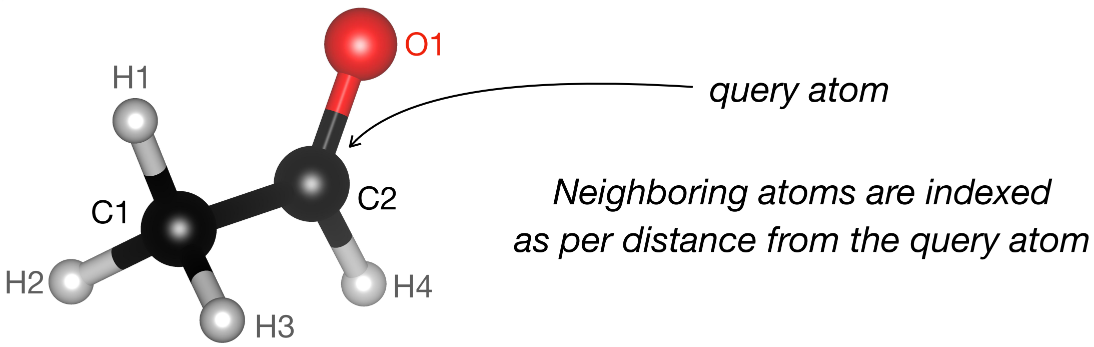
Neighborhood-informed features for ^{13}C chemical shifts

Scaling term: $\frac{1}{R^4}$ (like dipole-induced dipole)

Pairwise functions:

$$\mathbf{d}^{(A,B)}(r) = \sum_{J \neq I, I \in A, J \in B} g_{IJ}(r) \cdot \frac{Z_I Z_J}{R_{IJ}} \cdot s(R_{IJ})$$





query atom

Neighboring atoms are indexed as per distance from the query atom

$\mathbf{d}(0) = \mathbf{d}_{\mathrm{C2}}$ *Query atom's descriptor vector*

$\mathbf{d}(1) = \mathbf{d}_{\mathrm{C2}} \ \mathbf{d}_{\mathrm{H4}}$ *First neighbour's descriptor is padded*

$\mathbf{d}(2) = \mathbf{d}_{\mathrm{C2}} \ \mathbf{d}_{\mathrm{H4}} \ \mathbf{d}_{\mathrm{O1}}$ *Second neighbour*

$\mathbf{d}(3) = \mathbf{d}_{\mathrm{C2}} \ \mathbf{d}_{\mathrm{H4}} \ \mathbf{d}_{\mathrm{O1}} \ \mathbf{d}_{\mathrm{C1}}$ *Third neighbour*

$\mathbf{d}(4) = \mathbf{d}_{\mathrm{C2}} \ \mathbf{d}_{\mathrm{H4}} \ \mathbf{d}_{\mathrm{O1}} \ \mathbf{d}_{\mathrm{C1}} \ \mathbf{d}_{\mathrm{H1}}$ *Fourth neighbour*

$\mathbf{d}(5) = \mathbf{d}_{\mathrm{C2}} \ \mathbf{d}_{\mathrm{H4}} \ \mathbf{d}_{\mathrm{O1}} \ \mathbf{d}_{\mathrm{C1}} \ \mathbf{d}_{\mathrm{H1}} \ \mathbf{d}_{\mathrm{H2}}$ *Fifth neighbour*

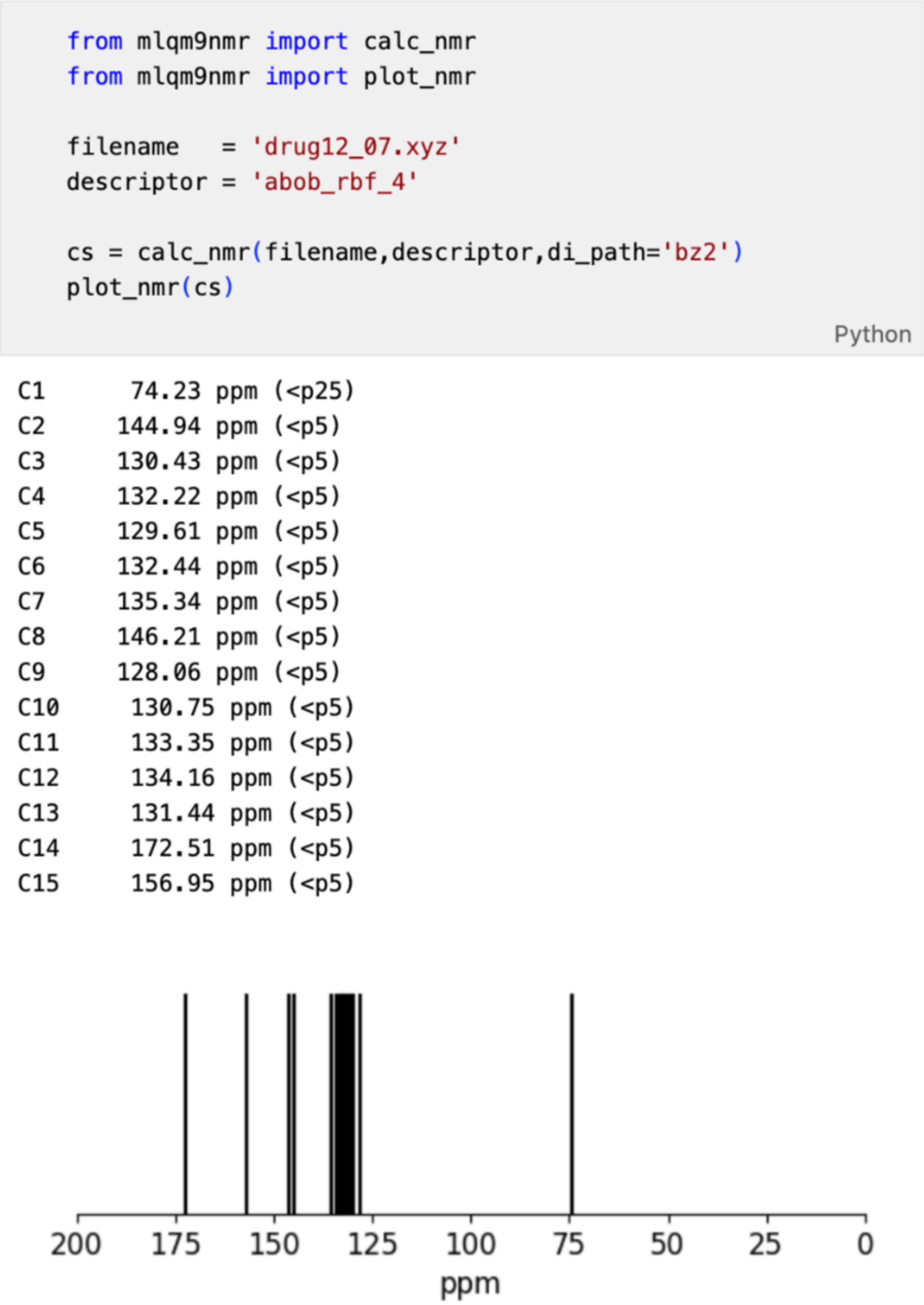
Concatenated pairwise functions:

$$\mathbf{d}(r) = \left[\mathbf{d}^{(\mathrm{H,H})}(r), \mathbf{d}^{(\mathrm{C,C})}(r), \cdots, \mathbf{d}^{(\mathrm{H,C})}(r), \cdots, \mathbf{d}^{(\mathrm{C,N})}(r), \mathbf{d}^{(\mathrm{C,O})}(r), \cdots\right]$$

aBoB-RBF(4): Out-of-sample mean prediction error 1.69 ppm

S. Das, et al. *J. Chem. Phys.* (2026).

MLQM9NMR Python module (XYZ to ^{13}C shifts)



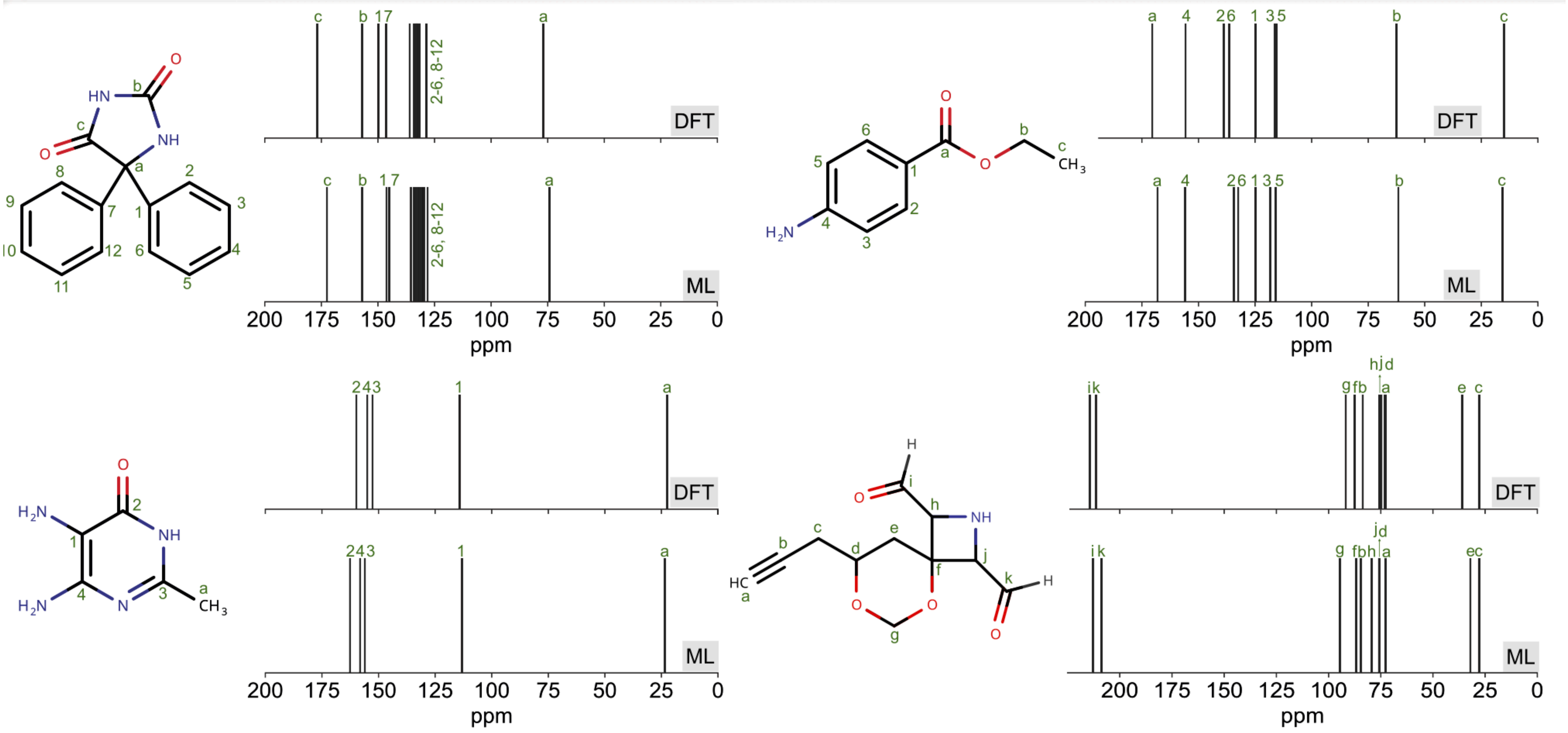
```python
from mlqm9nmr import calc_nmr
from mlqm9nmr import plot_nmr

filename   = 'drug12_07.xyz'
descriptor = 'abob_rbf_4'

cs = calc_nmr(filename,descriptor,di_path='bz2')
plot_nmr(cs)
```

Python

```
C1      74.23 ppm (<p25)
C2     144.94 ppm (<p5)
C3     130.43 ppm (<p5)
C4     132.22 ppm (<p5)
C5     129.61 ppm (<p5)
C6     132.44 ppm (<p5)
C7     135.34 ppm (<p5)
C8     146.21 ppm (<p5)
C9     128.06 ppm (<p5)
C10     130.75 ppm (<p5)
C11     133.35 ppm (<p5)
C12     134.16 ppm (<p5)
C13     131.44 ppm (<p5)
C14     172.51 ppm (<p5)
C15     156.95 ppm (<p5)
```

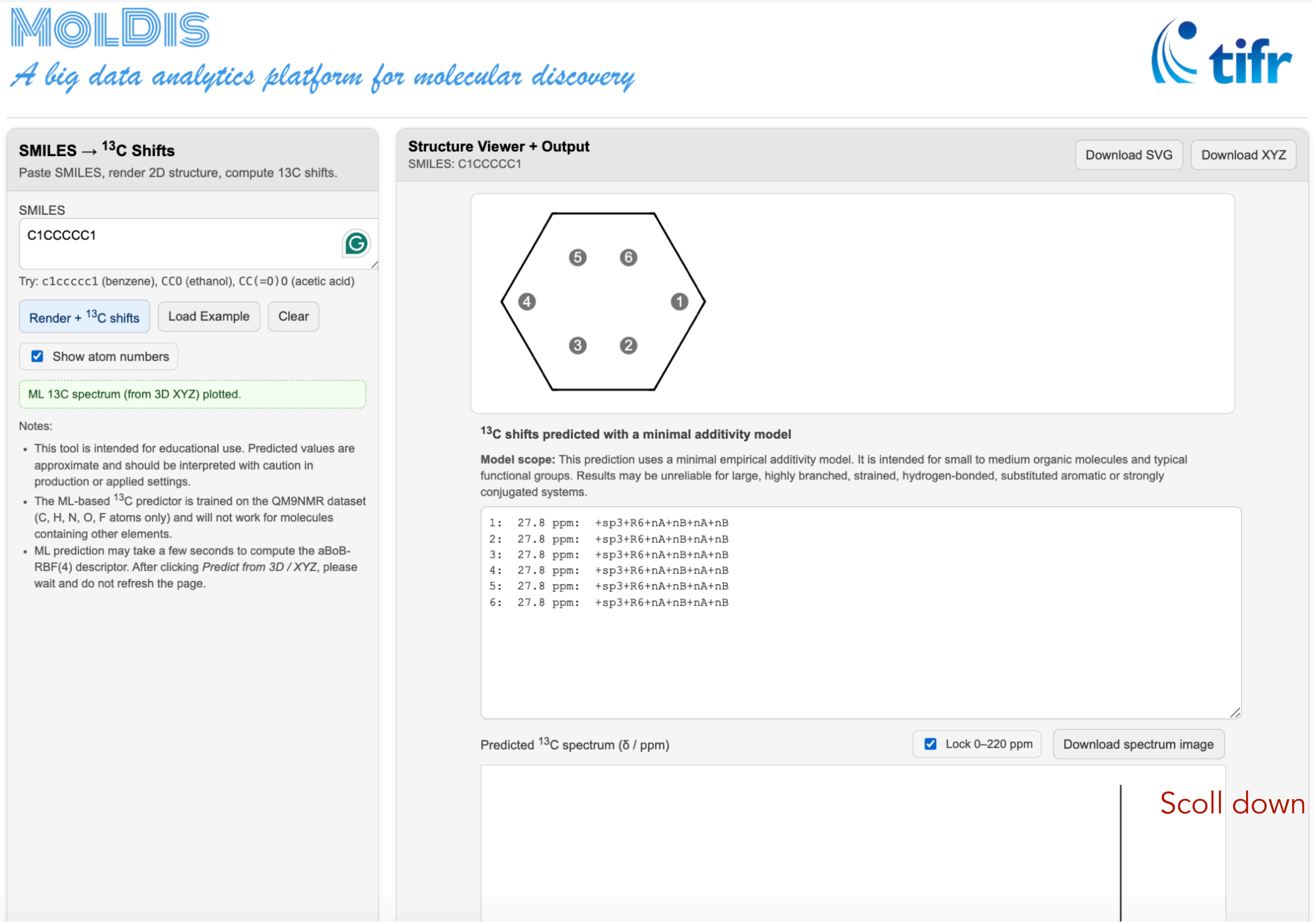
S. Das, et al. *J. Chem. Phys.* (2026).

Comparison with target-level data for larger drug molecules



S. Das, et al. *J. Chem. Phys.* (2026).

DFT-level ^{13}C chemical shifts predicted with ML on MolDis-Lab



MOLDIS

A big data analytics platform for molecular discovery

tifr

SMILES → ^{13}C Shifts

Paste SMILES, render 2D structure, compute 13C shifts.

SMILES

C1CCCCC1

Try: `c1ccccc1` (benzene), `CCO` (ethanol), `CC(=O)O` (acetic acid)

Render + ^{13}C shifts | Load Example | Clear

☑ Show atom numbers

ML 13C spectrum (from 3D XYZ) plotted.

Notes:

- This tool is intended for educational use. Predicted values are approximate and should be interpreted with caution in production or applied settings.
- The ML-based ^{13}C predictor is trained on the QM9NMR dataset (C, H, N, O, F atoms only) and will not work for molecules containing other elements.
- ML prediction may take a few seconds to compute the aBoB-RBF(4) descriptor. After clicking *Predict from 3D / XYZ*, please wait and do not refresh the page.

Structure Viewer + Output

SMILES: C1CCCCC1

Download SVG | Download XYZ

^{13}C shifts predicted with a minimal additivity model

Model scope: This prediction uses a minimal empirical additivity model. It is intended for small to medium organic molecules and typical functional groups. Results may be unreliable for large, highly branched, strained, hydrogen-bonded, substituted aromatic or strongly conjugated systems.

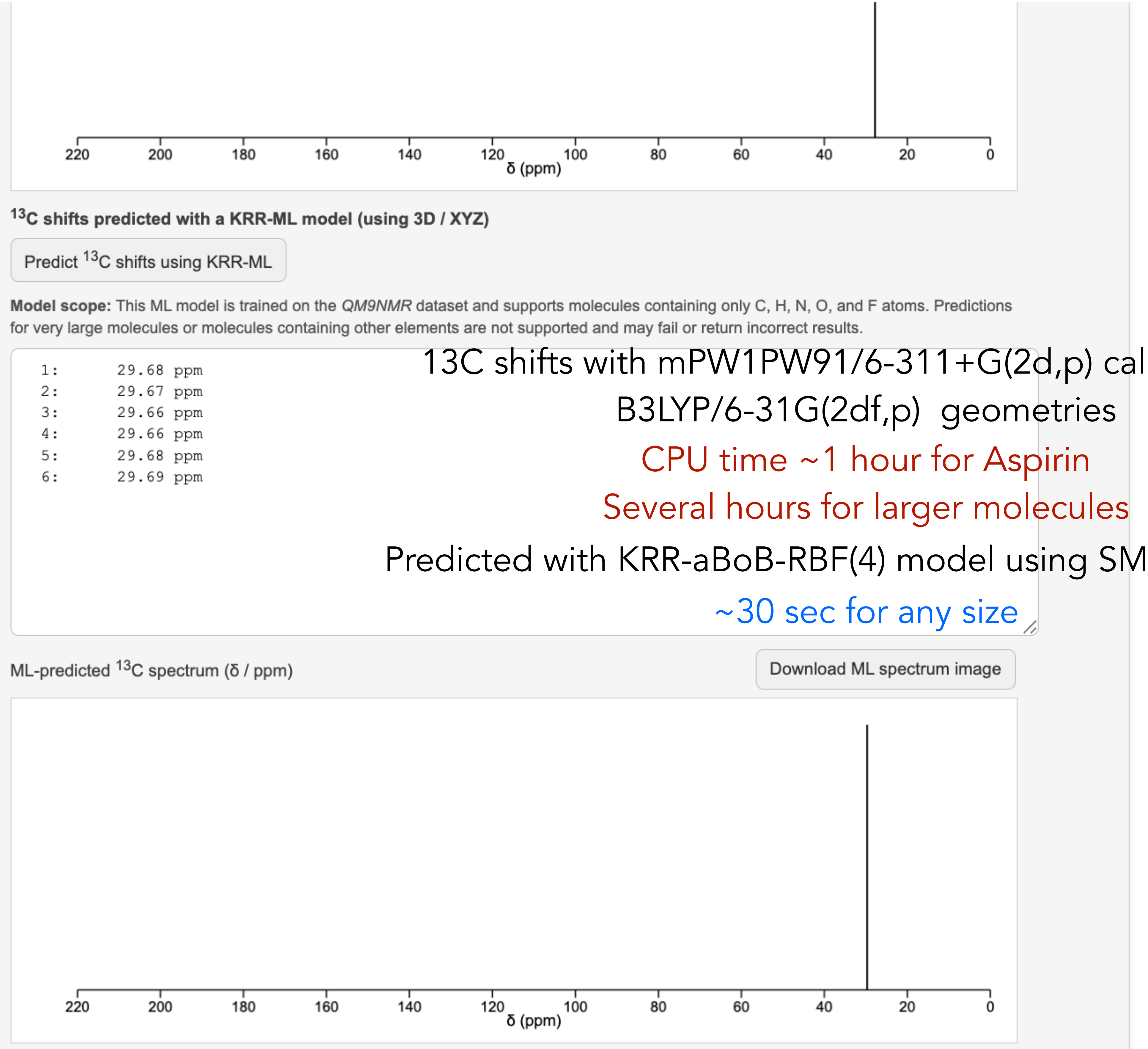
```
1:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
2:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
3:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
4:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
5:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
6:  27.8 ppm:  +sp3+R6+nA+nB+nA+nB
```

Predicted ^{13}C spectrum (δ / ppm)

☑ Lock 0–220 ppm | Download spectrum image

Scoll down

DFT-level ^{13}C chemical shifts predicted with ML on MolDis-Lab



^{13}C shifts predicted with a KRR-ML model (using 3D / XYZ)

Model scope: This ML model is trained on the *QM9NMR* dataset and supports molecules containing only C, H, N, O, and F atoms. Predictions for very large molecules or molecules containing other elements are not supported and may fail or return incorrect results.

```
  1:      29.68 ppm
  2:      29.67 ppm
  3:      29.66 ppm
  4:      29.66 ppm
  5:      29.68 ppm
  6:      29.69 ppm
```

ML-predicted ^{13}C spectrum (δ / ppm)

13C shifts with mPW1PW91/6-311+G(2d,p) calculated on B3LYP/6-31G(2df,p) geometries

CPU time ~1 hour for Aspirin

Several hours for larger molecules

Predicted with KRR-aBoB-RBF(4) model using SMILES as input

~30 sec for any size

Pics from Kang-Yatse-2 expedition, September 2025

See you tomorrow for the hands-on!