# What can we do with RAW NMR data and spin parameters

Jonathan Bisson, University of Illinois at Chicago
bjo@uic.edu

NMReDATA Symposium 2019, Porto, Portugal

CENAPT

https://cenapt.pharm.uic.edu


UIC COLLEGE OF PHARMACY
UNIVERSITY OF ILLINOIS AT CHICAGO

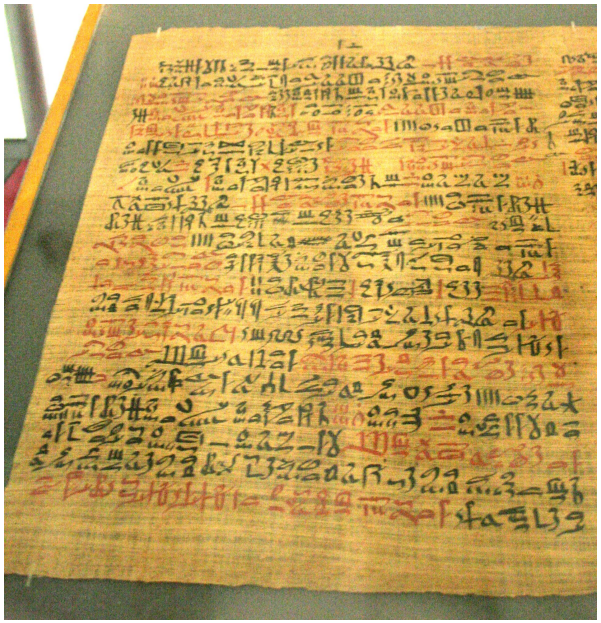
# RAW NMR data and spin parameters

- What do we do? Pharmacognosy
- History… we keep reinventing the wheel
- Applications of spin simulation
- RAW data initiative and the benefits of RAW data
- How could NMReData benefit us



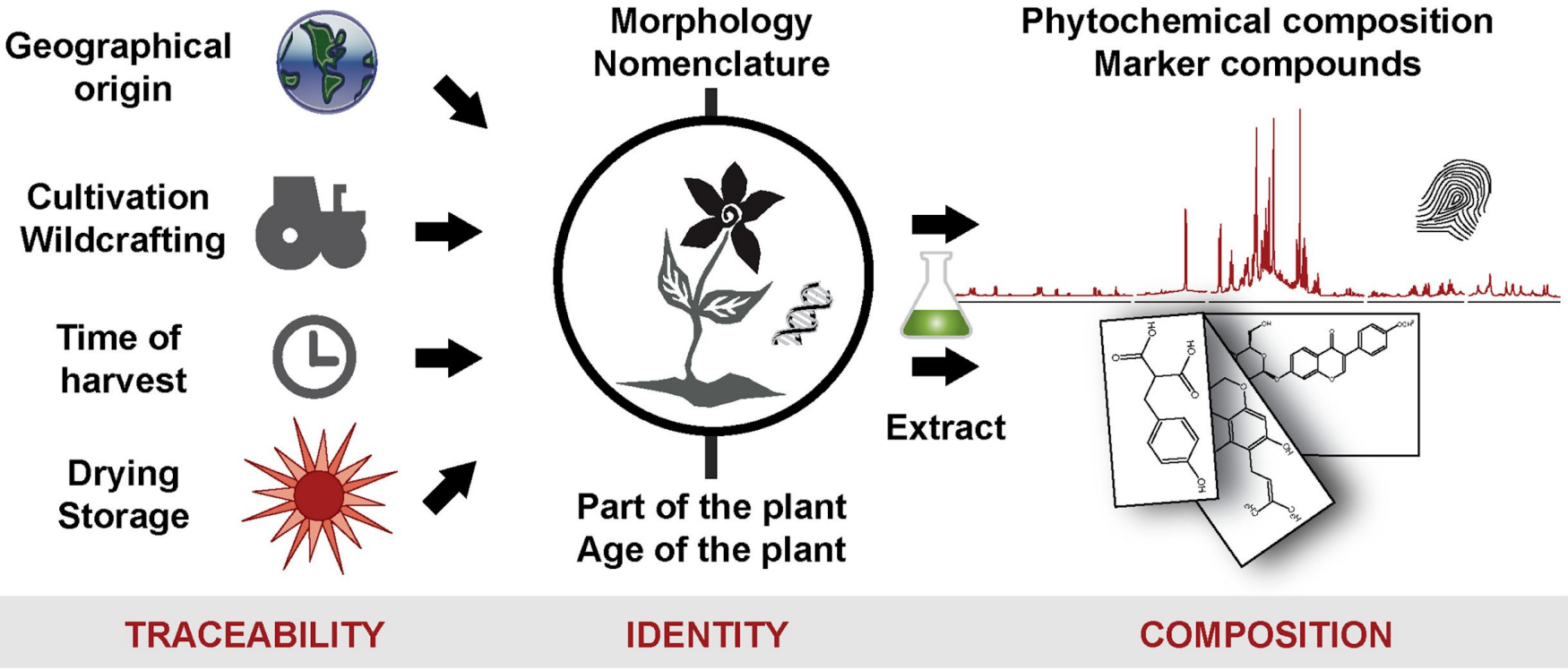
# Pharmacognosy

"[…]the study of the physical, chemical, biochemical and biological properties of drugs, drug substances or potential drugs or drug substances of natural origin as well as the search for new drugs from natural sources"



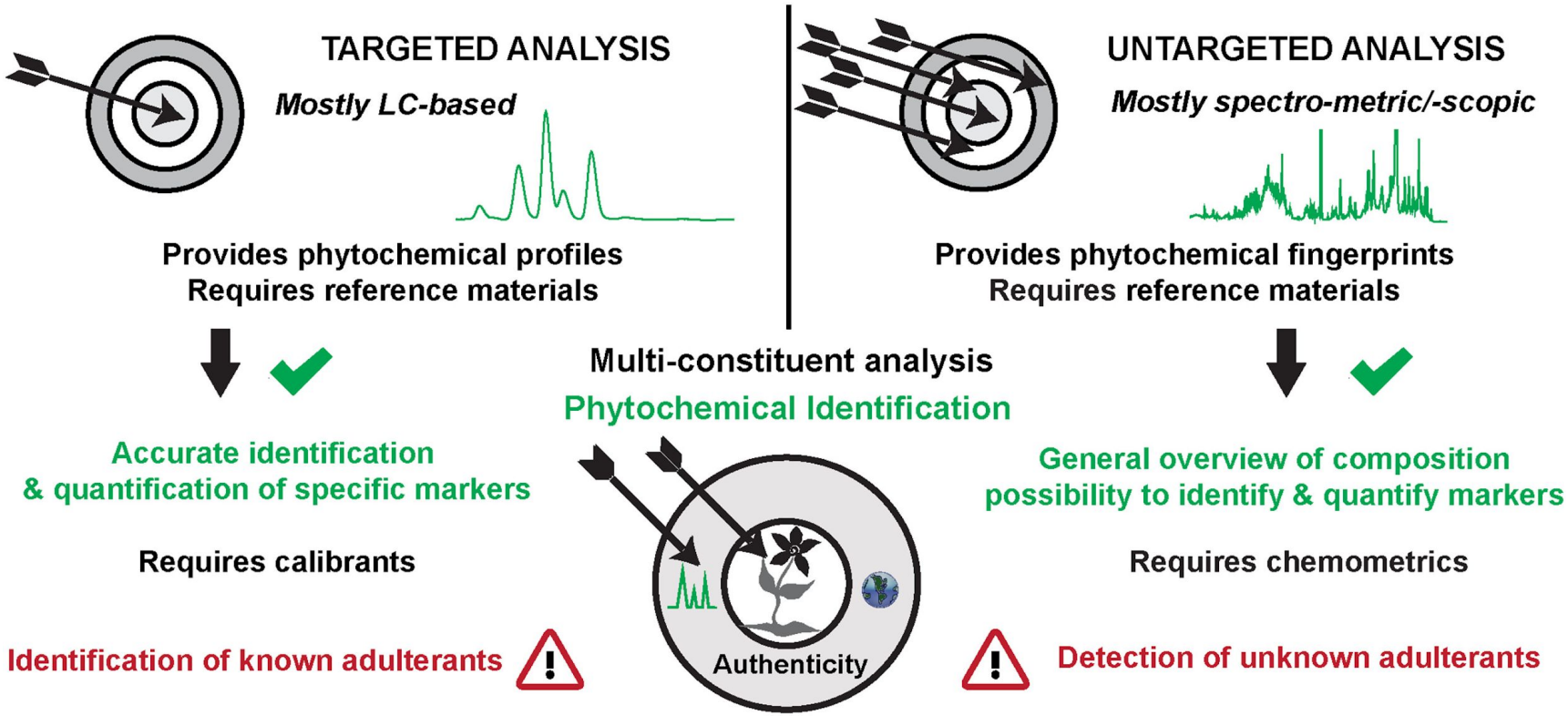
(Former definition by the American Society of Pharmacognosy).

# Pharmacognosy and adulteration



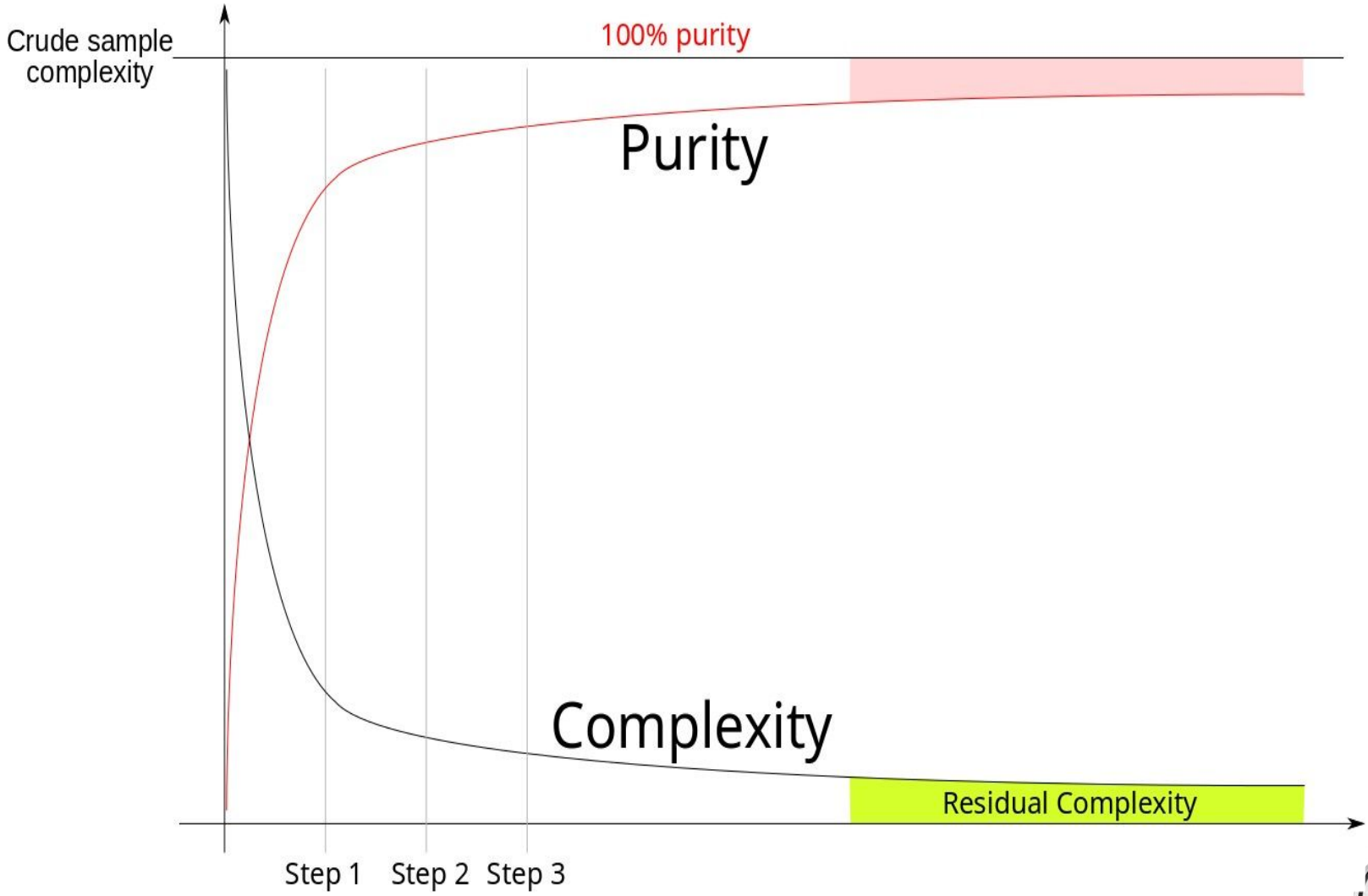

Simmler et al, https://doi.org/10.1016/j.fitote.2017.11.017

# Pharmacognosy and adulteration



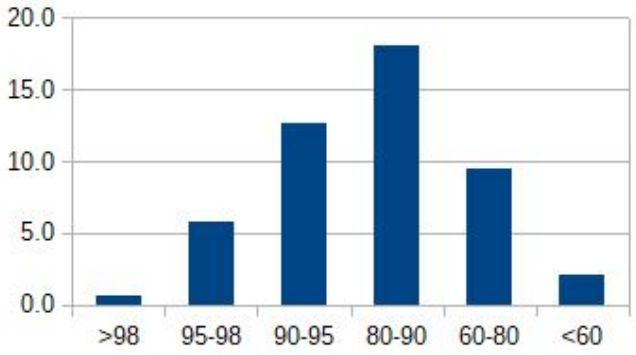

Simmler et al, https://doi.org/10.1016/j.fitote.2017.11.017

# The problem of residual complexity



# AnaPurNa[1]

- 2012: AnaPurNa: meta study of NP analysis and purification methodology
  - 2,000 publications 1998-2010, 13 journals, 80,000 pages
  - Gold Standard according to the world literature on NPs?
    - Average number of isolation steps post partitioning: **2.4**
    - Purity determined for **0.5%** of reported NPs
- 2017-: qNMR meta analysis of published SI data
  - Purity estimate using visual 100% qNMR, six tiers
  - Majority of isolates falls into 80-90% purity window

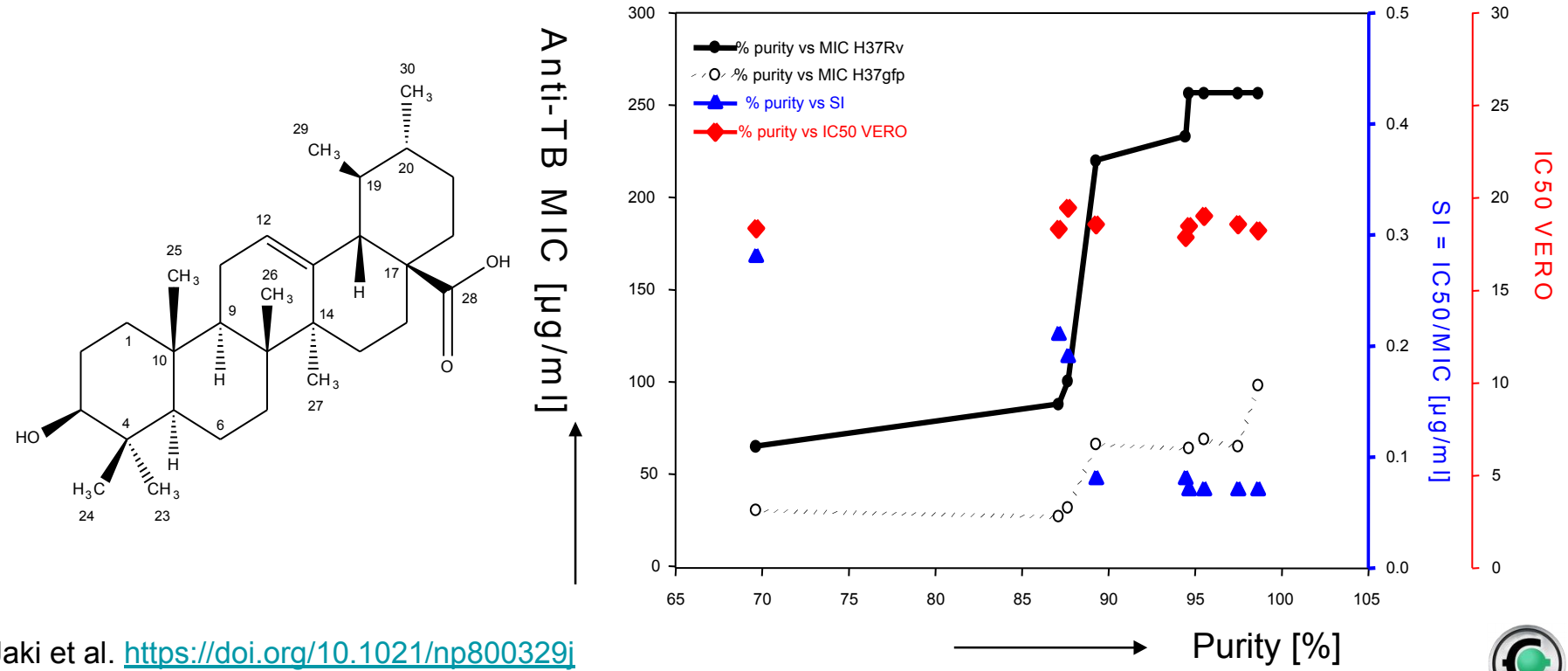



[1]Pauli et al. https://doi.org/10.1021/np300066q

# Purity-Activity relationships

Is **ursolic acid** (MIC 32-128 μg/mL) a viable anti-TB lead?

- qNMR Answer: Inverse correlation between purity and activity
- **qNMR Net Outcome Pure** UA is essentially **in**active




Jaki et al. https://doi.org/10.1021/np800329j

# Spin simulation

« While computer simulation of complex NMR spectra is not new, we have developed a modern version of an NMR simulation program that incorporates modern software standards in a user-friendly interface and graphic representation of calculated spectrum. »



# Spin simulation

« While computer simulation of complex NMR spectra is not new, we have developed a modern version of an NMR simulation program that incorporates modern software standards in a user-friendly interface and graphic representation of calculated spectrum. »

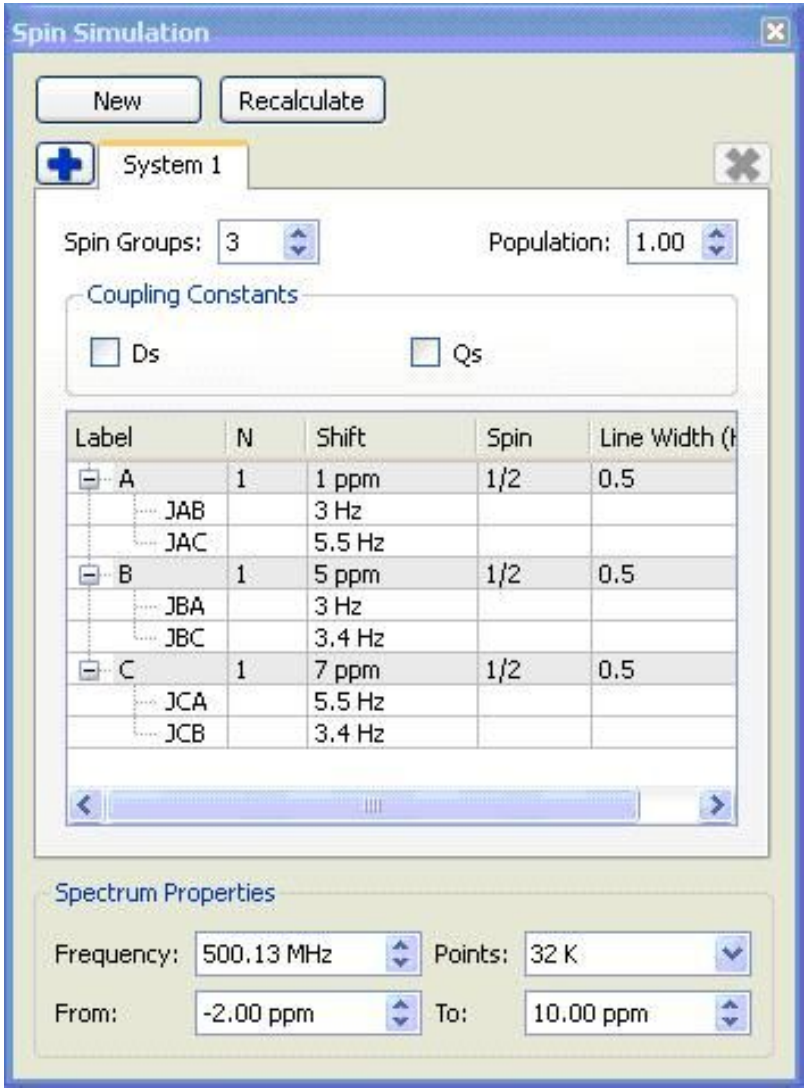
Clark & Thrasher about LAOCOON. Journal of Chemical Education 1990

http://dx.doi.org/10.1021/ed067p235.2

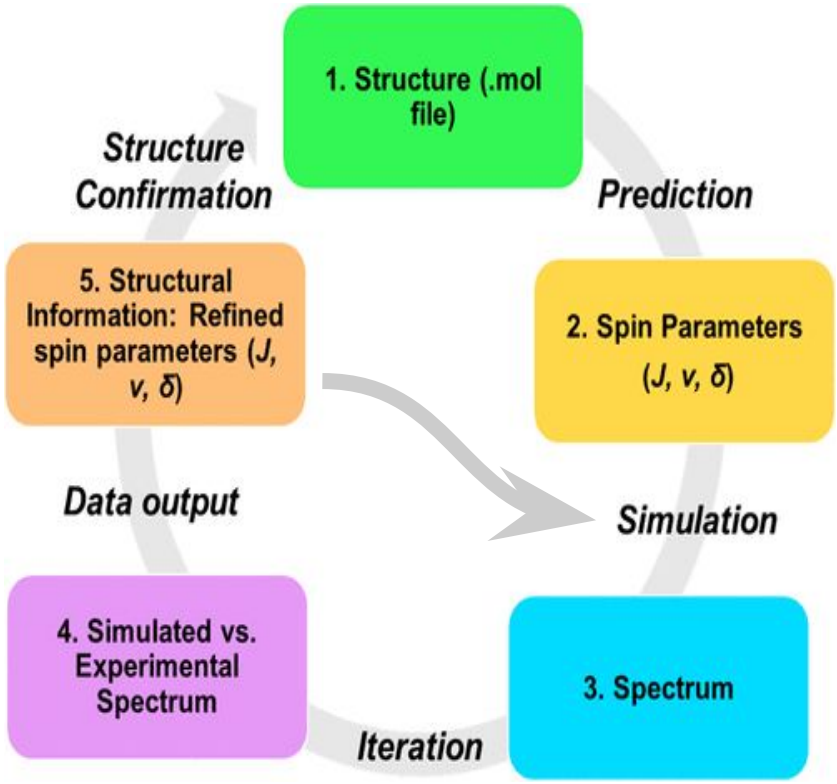
**Table 2. Timings for NMR Simulations Using LAOCOON PC**

| Number of Nuclei | "Average", s | "Worst Case", s |
|---|---|---|
| 2 | 5 | 5 |
| 3 | 6 | 6 |
| 4 | 8 | 9 |
| 5 | 19 | 24 |
| 6 | 106 | 147 |
| 7 | 1350 | 1985 |

# First experience with spin simulation in 2019 for a chemist?



# Spin simulation and iteration




- Good success with PERCH.
  - What do we do now?
- The way we worked with PERCH:
  - Prediction from structure with MD
  - Simulation <-> Iteration → RMS
  - All in one software
- What now?
  - Put the focus on better predictions to reduce the need for iterations? (DFT…)
  - Put the focus on better quality simulations? (spin dynamics, exchanges…)
  - Write our own solution?
- We need an open and integrated solution!
  - Most of our users have no idea how to write a text file with spin parameters even less write code.
  - Many developers keep source code hidden, project dies and others have to reinvent the wheel. And this did not happen only once…

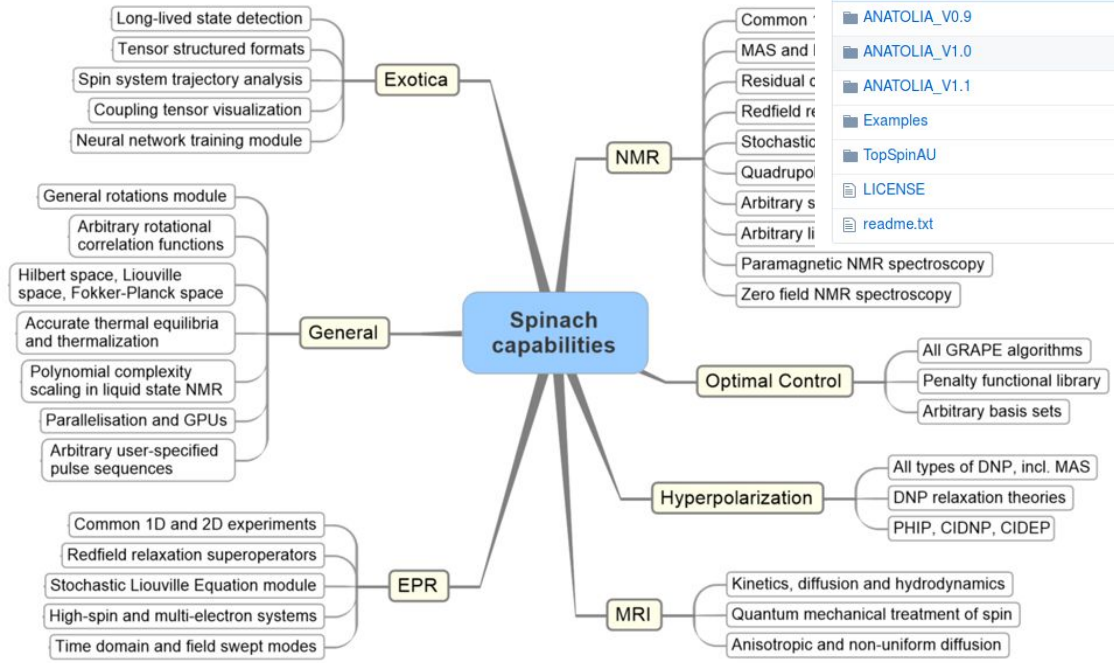
# But there are other solutions!

- Is it an integrated solution?
  - from structure to optimized parameters
- Is it a fast solution?
  - PERCH iterate small molecules in a couple of minutes
  - Some solutions we tried took >1h for 9 spins, crashed…
- Is it an accurate solution?
  - We get ~1% errors even with weak couplings
- Free?
  - Most academic projects are not supported once the student that coded it graduates
  - Industry? You kill a project and do not have a competing one, make it Free
  - Give your users freedom

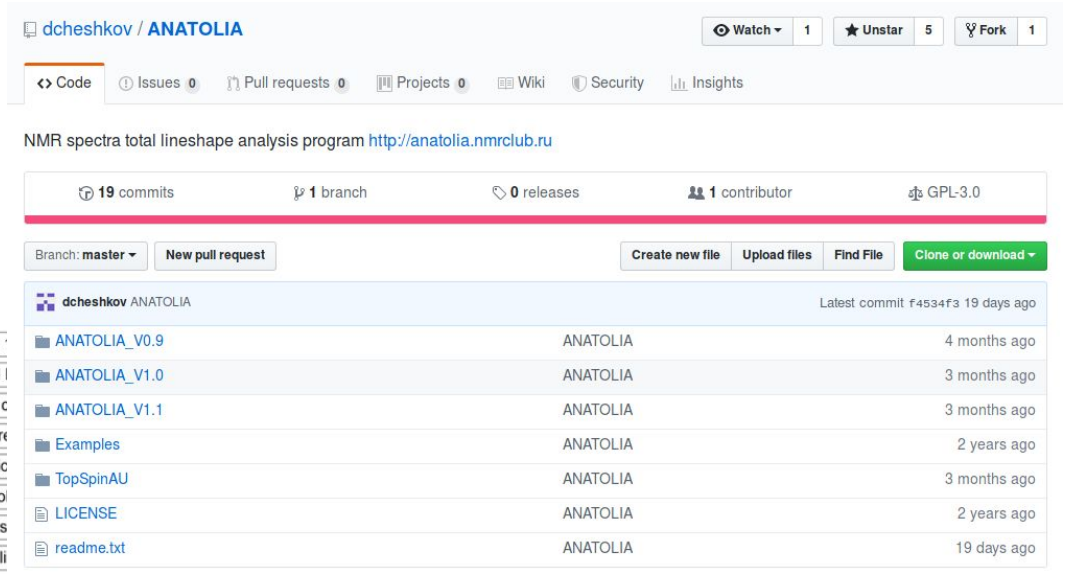


# Spin simulation: The fine way or

## The brute but fast way




https://spindynamics.org




https://github.com/dcheshkov/ANATOLIA

# Spin Analysis (a really incomplete list)

**Table S7.** List of software tools for NMR spin simulation, QM-based and iterative analysis.

| Name | Simulation | QM-based | Iteration/Fit | Format | Format documented | Structure in format | Availability | Commercial | URL/contact | Comment |
|---|---|---|---|---|---|---|---|---|---|---|
| ACDlabs | Y | Y | | | | | Y | Y | | |
| ChemAdder | Y | Y | Y | | | | | | http://chemadder.com | |
| ChemInfo.org | Y | Y | | | | | Y | N | https://www.cheminfo.org/Spectra/NMR/Tools/Simulate_spin_system/index.html | |
| ChenomX | Y | Y | | | | | Y | Y | https://www.chenomx.com/ | "quantum-mechanical rules-based simulations of compound lineshapes" |
| CT | Y | Y | Y | JSON | Y | Y | Beta | Y | ct@nmrsolutions.fi | |
| DsymPC | Y | Y | Y | | | | N | | ftp://ftp.rz.uni-duesseldorf.de/pub/msdos/chemie | |
| Gamma | Y | Y | | code | Y | | Y | N | http://scion.duhs.duke.edu/vespa/gamma | |
| iNMR | Y | | Y | | | | Y | Y | http://inmr.net/ | |
| Mnova | Y | Y | | MnovaXML | | N | Y | Y | https://mestrelab.com | Mnova-SpinSim XML Format |
| NMRSIM | Y | Y | | | | | Y | | http://science.widener.edu/svb/nmr/nmr_soft.html | |
| NSS | Y | | | code | Y | | Y | | http://eos.univ-reims.fr/LSD/JmnSoft/ | Simulates FIDs |
| NUTS | Y | Y | | NS | N | | Y | Y | https://www.acornnmr.com/ | |
| PERCH | Y | Y | Y | MMS | Y | Y | N | Y | | |
| pNMRSIM | Y | | | | | | | | | Designed for solid state |
| Simpson | Y | | | | | | | | http://www.bionmr.chem.au.dk/bionmr/software/index.php | Designed for solid state |
| Spinach | Y | Y | Y | SpinXML | Y | | Y | N | http://spindynamics.org/Spinach.php | |
| SpinEvolution | Y | Y | Y | | Y | | Y | Y | https://spinevolution.com/ | |
| SpinWorks | Y | Y | Y | | | | Y | N | https://home.cc.umanitoba.ca/~wolowiec/spinworks/ | |
| Topspin/Daisy | Y | Y | Y | | | N | Y | Y | https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/MagneticResonance/NMR/brochures/TopSpin.pdf | |
| VNMR/LAME | Y | Y | Y | | | N | | Y | http://openvnmrj.org/ | |
| WinDNMR | Y | Y | | | | | Y | N | http://www.chem.wisc.edu/areas/reich/plt/windnmr.htm | LAOCOON |
| Gissmo | Y | Y | Y | | | Y | | Y | N | http://gissmo.nmrfam.wisc.edu/ | |
| NMRpipe | Y | Y | N? | | | | | Y | | http://www.nmrscience.com/nmrpipe.html | |

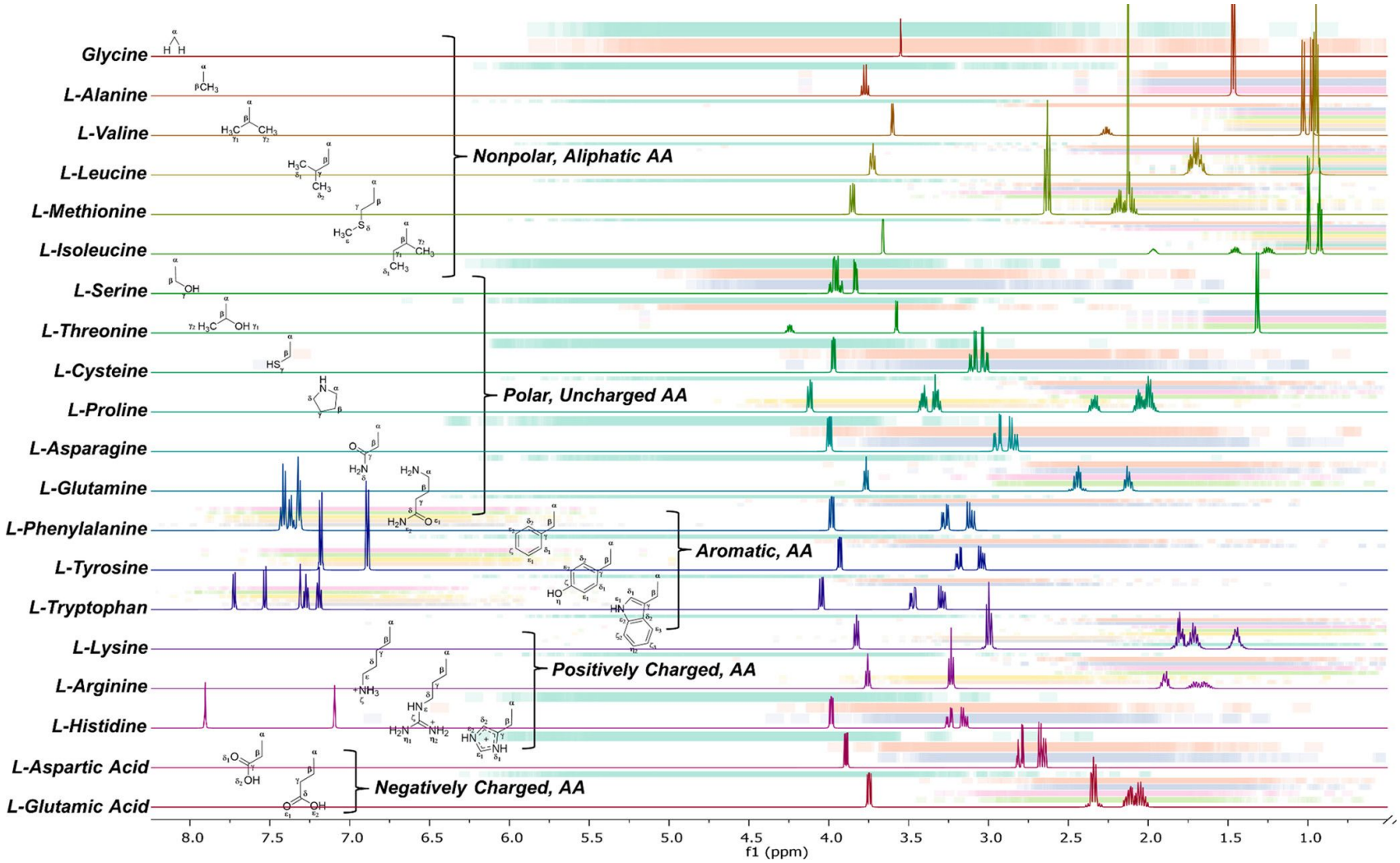
Free-software

Open Source

# Applications



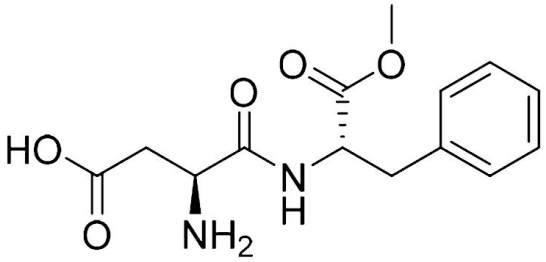
# Building peptides block by block



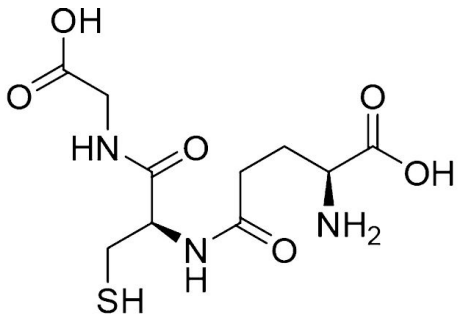

Choules et al. https://doi.org/10.1021/acs.joc.8b02704

# Quality control of peptides: the studied peptides

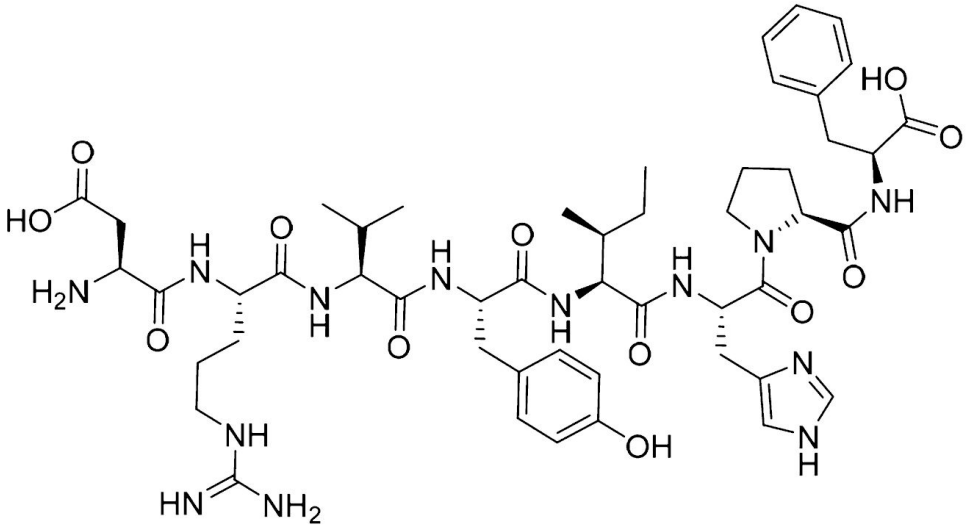
**1**



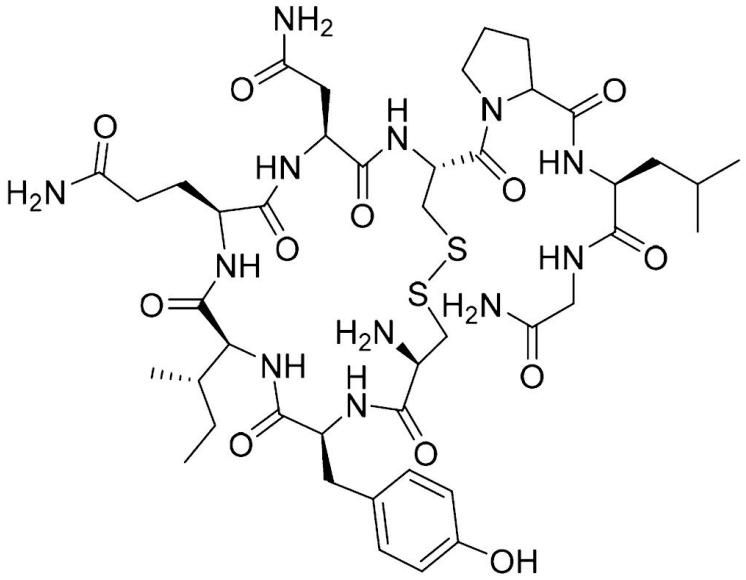
**2**



**3**

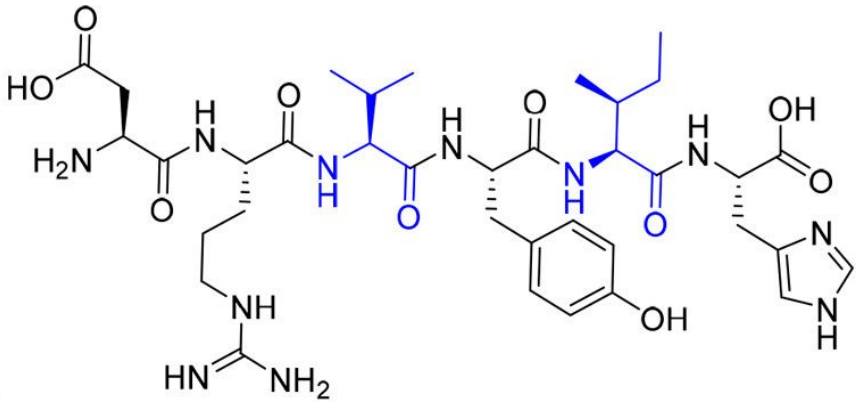


**4**

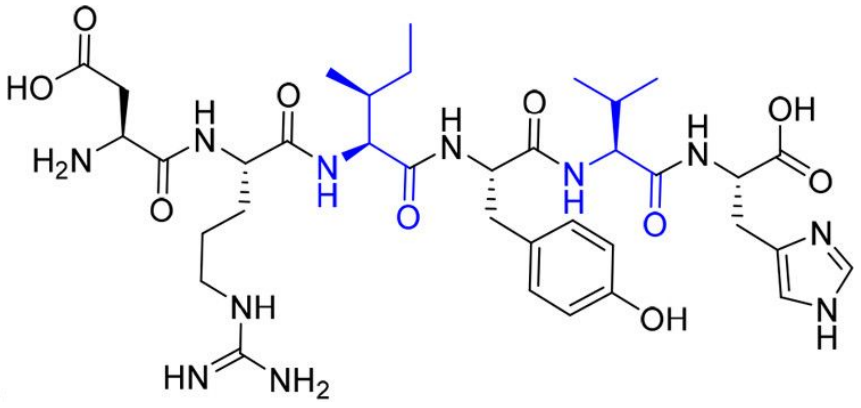


Choules et al. https://doi.org/10.1021/acs.joc.8b02704

# Quality control of peptides: check for syn. errors



**5.**



**6.**

Choules et al. https://doi.org/10.1021/acs.joc.8b02704

# Quality control of peptide: Field scaling and cheap magnets

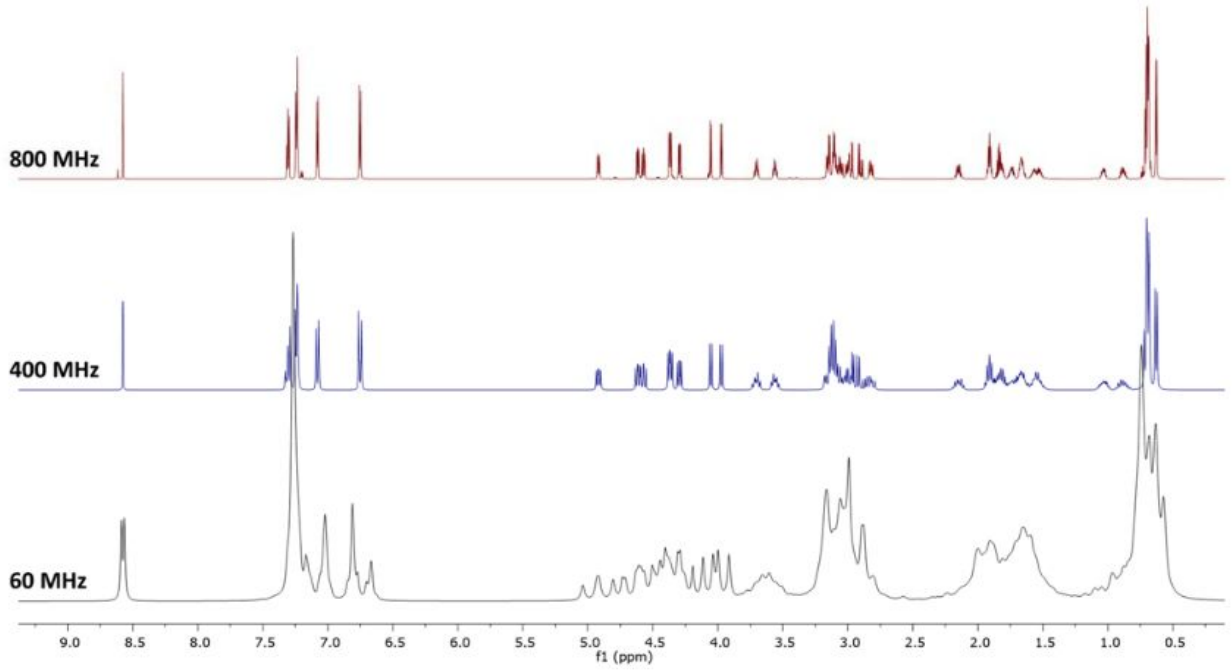




**Figure S52**. $^{1}$H NMR HiFSA generated spectra of D-Tyr angiotensin II at 800 MHz (red), 400 MHz (blue), and 60 MHz (black). Spectra generated from 800 MHz experimental parameters.

Generation of the spectrum at any field can be scripted, takes a couple of seconds.
We also have a PMS (Perch format) to Mnova Spin simulation converter (works with any field)
https://github.com/bjonnh/spinverter

Choules et al. https://doi.org/10.1021/acs.joc.8b02704

# Quality control of peptides: Field scaling

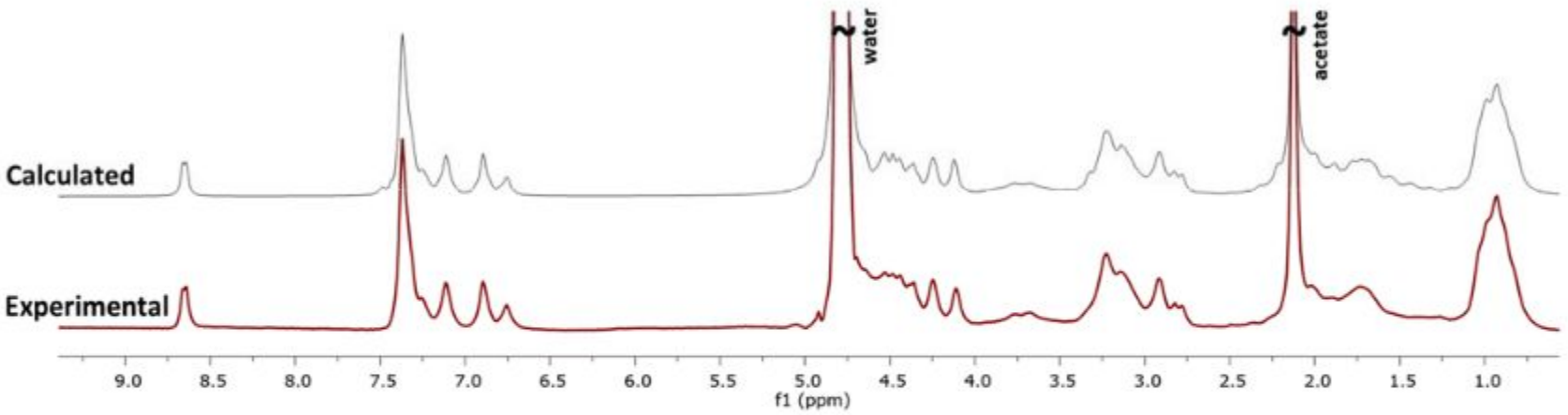
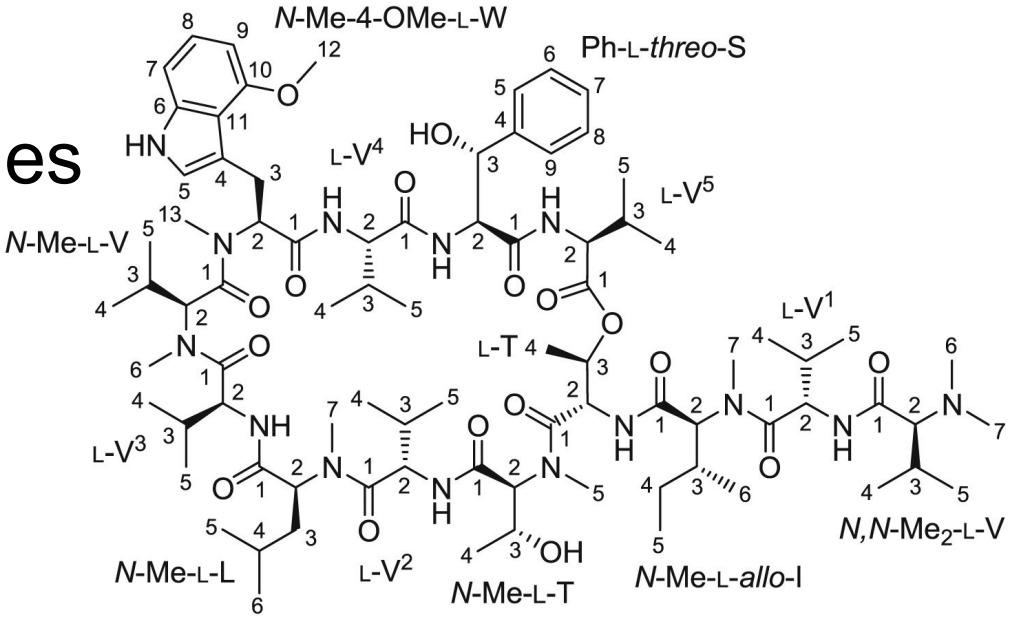
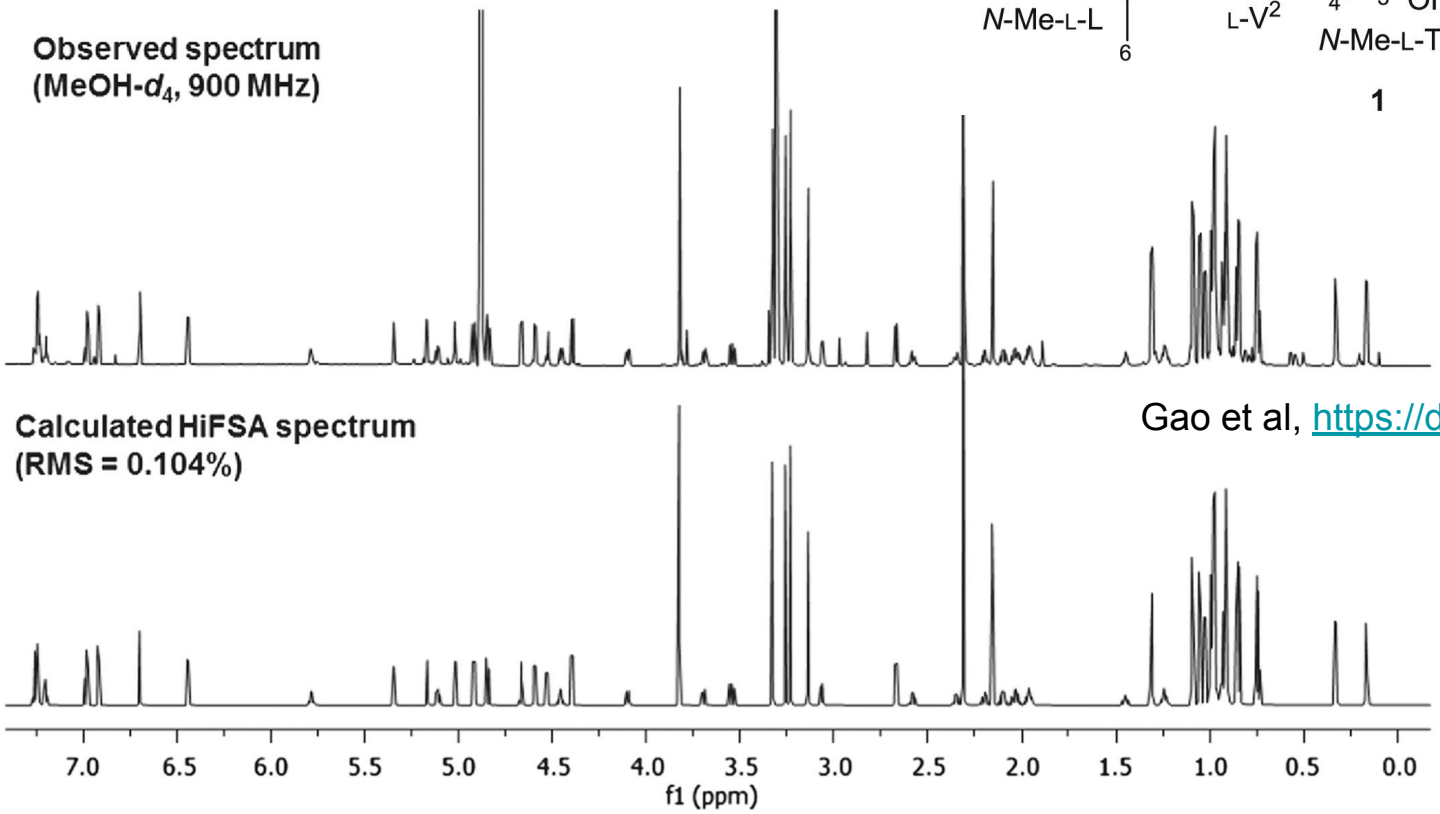



**Figure S49.** (**A**) Experimental $^{1}$H NMR spectra of angiotensin II at 60 MHz in $D_2O$ at 305 K. (**B**) HiFSA generated spectrum from 800 MHz PMS file (top, black) vs. experimental spectrum (bottom, red). Solvent signals included due to overlap with compound signals.

Choules et al. https://doi.org/10.1021/acs.joc.8b02704

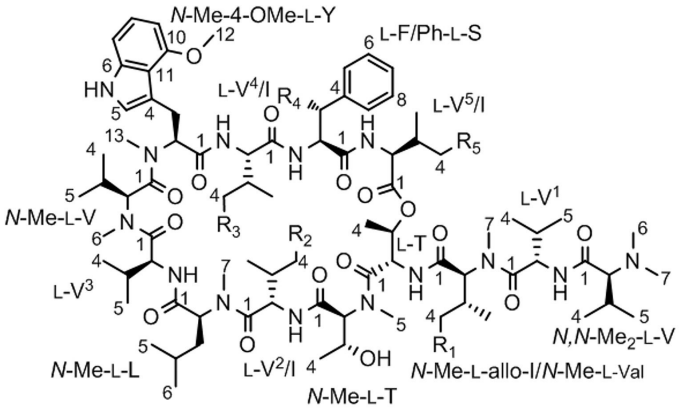
# Anti-Tuberculosis Peptides



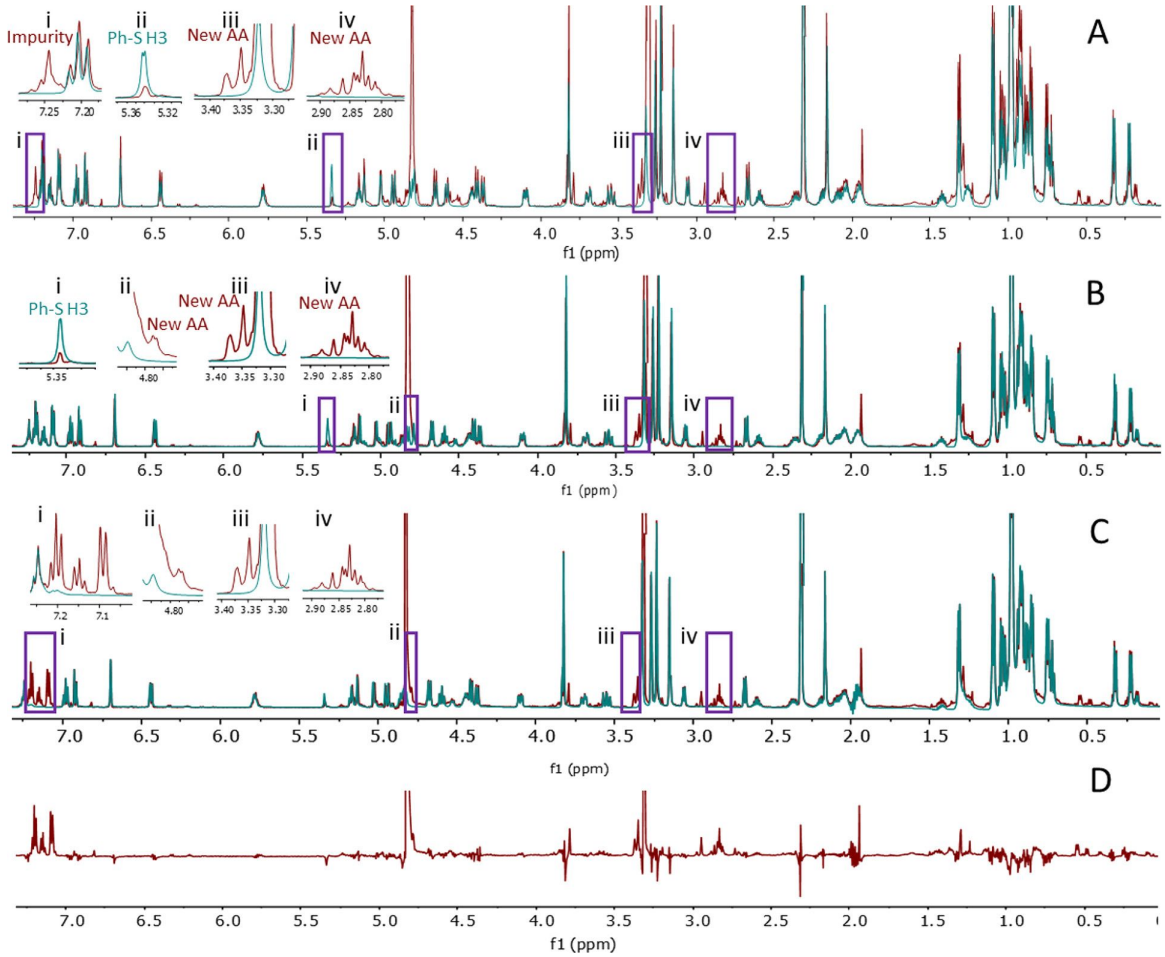




Gao et al, https://doi.org/10.1002/mrc.4425

# Sequencing peptides




| | |
|---|---|
| **Ecumicin (1)** | $R_1$=$CH_3$, $R_4$=OH, $R_2$,$R_3$,$R_5$=H |
| **Deoxyecumicin (2)** | $R_1$=$CH_3$, $R_2$,$R_3$,$R_4$,$R_5$=H |
| **Norecumicin (3)** | $R_4$=OH, $R_1$,$R_2$,$R_3$,$R_5$=H |
| **Nordeoxyecumicin (4)** | $R_1$,$R_2$,$R_3$,$R_4$,$R_5$=H |
| **Homoecumicin-ile11 (5)** | $R_1$,$R_3$=$CH_3$, $R_4$=OH, $R_2$,$R_5$=H |
| **Homoecumicin-ile13 (6)** | $R_1$,$R_5$=$CH_3$, $R_4$=OH, $R_2$,$R_3$=H |
| **Homoecumicin-ile6 (7)** | $R_1$,$R_2$=$CH_3$, $R_4$=OH, $R_2$,$R_5$=H |

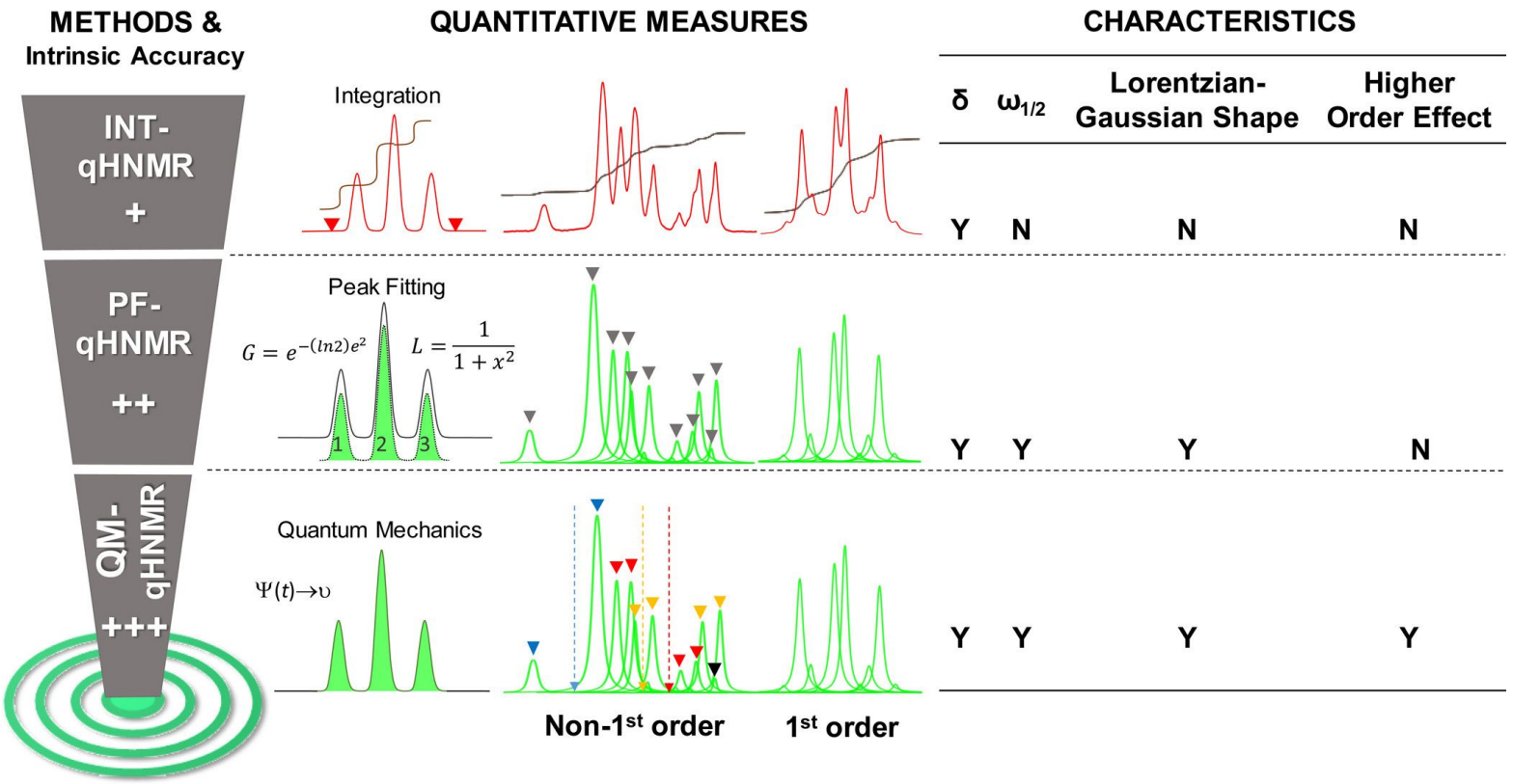



*Gao et al*, [10.1021/acs.jnatprod.7b00207](https://doi.org/10.1021/acs.jnatprod.7b00207)

# Why simulation based methods:

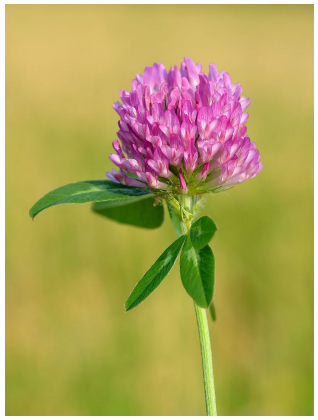
we want to use the full spectrum to quantify

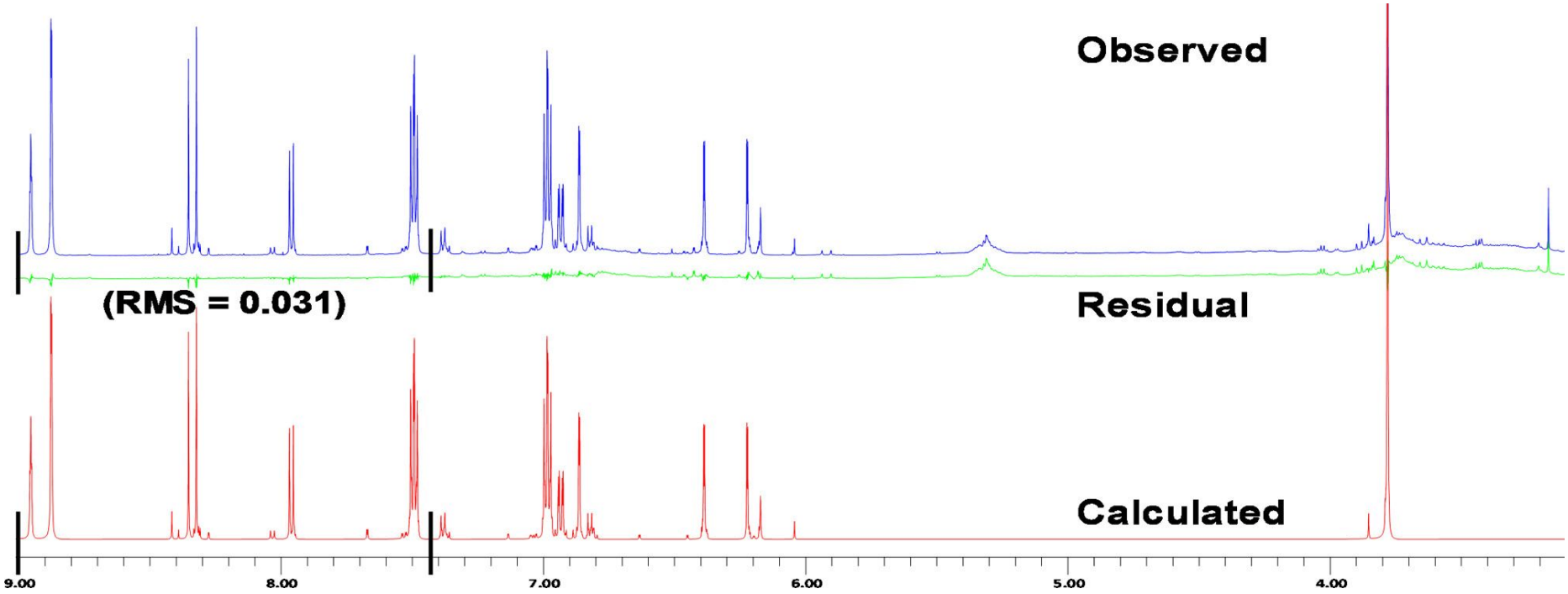
we want to see what is left.




Phansalkar et al. https://doi.org/10.1021/acs.jnatprod.6b00923

# Why simulation based methods:
## Fitting and quantifying 10 compounds





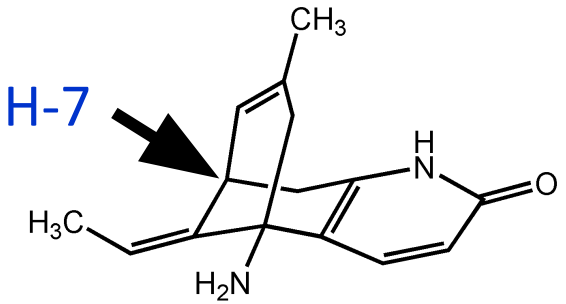

Phansalkar et al. https://doi.org/10.1021/acs.jnatprod.6b00923

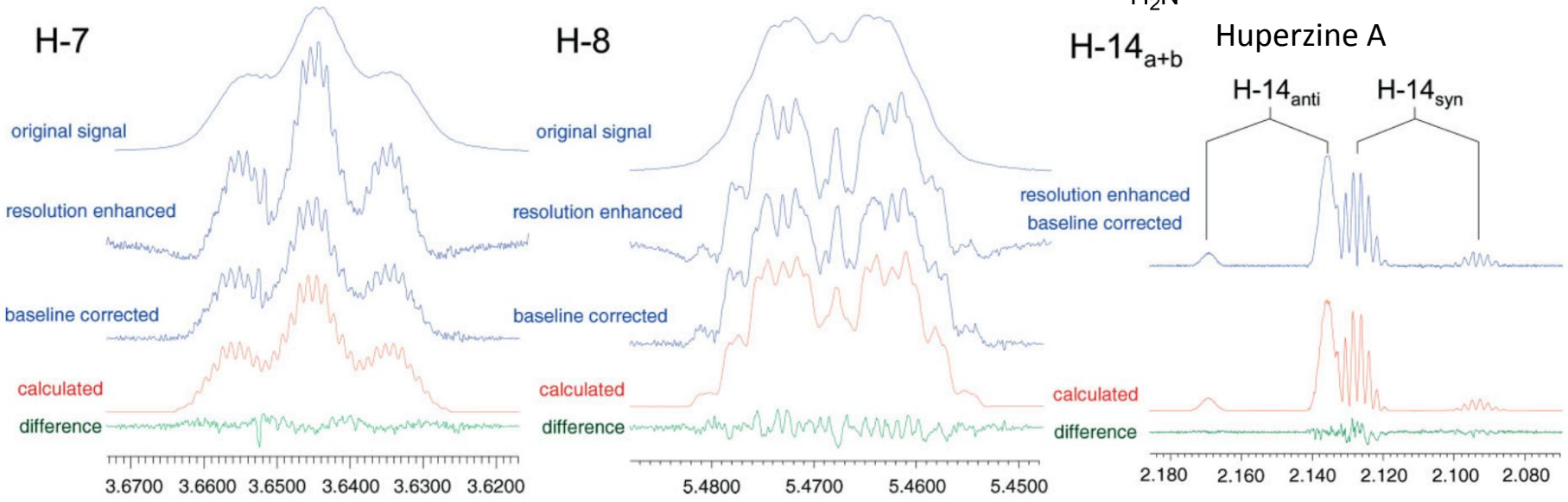
Picture: Ivar Leidus - Own work CC-BY-SA 4.0 Mediawiki

# Why does it matter?

ddddddqq !

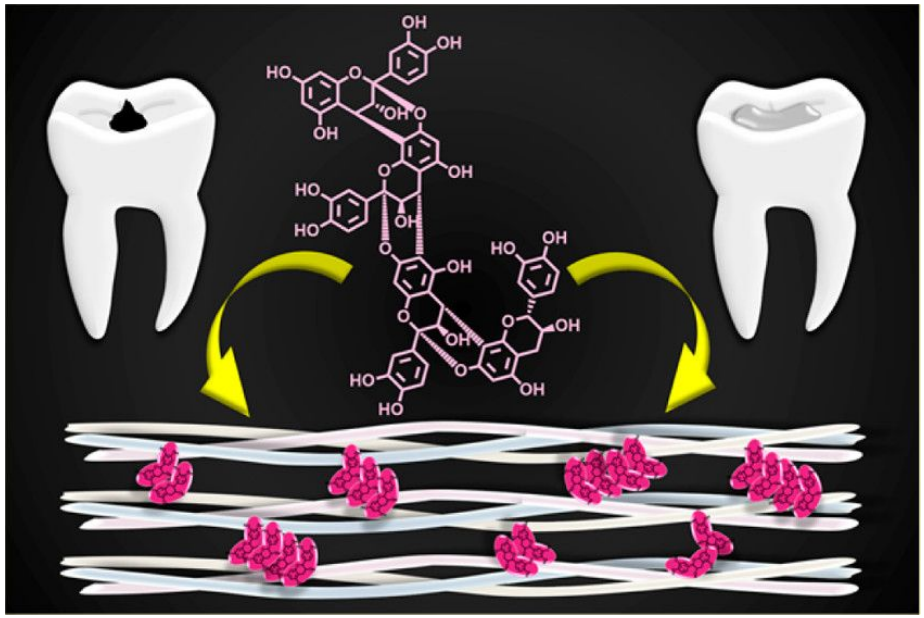
- HupA: "Nootropic", supposed to help with AD, PD
- ABCD(E)(MN)(OP)$X_3Y_3$ spin system
  - 15H/11 NCE spins, 38 *J*-couplings, including **31(!)** long-range ($^{4\text{-}6}J$)

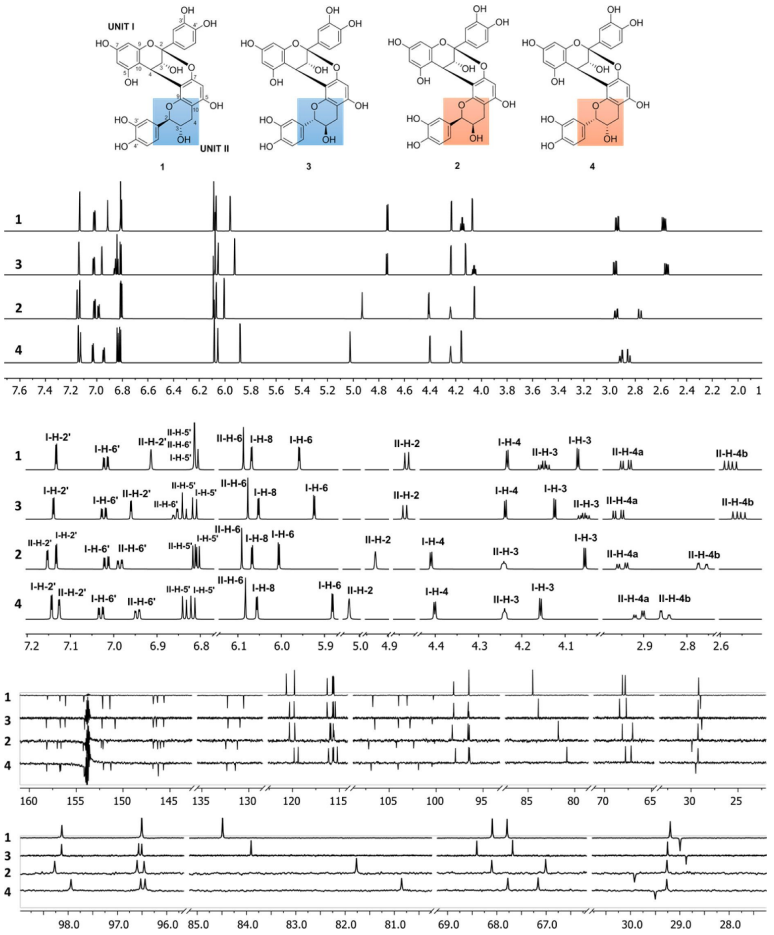



Niemitz et al, https://doi.org/10.1002/mrc.2061

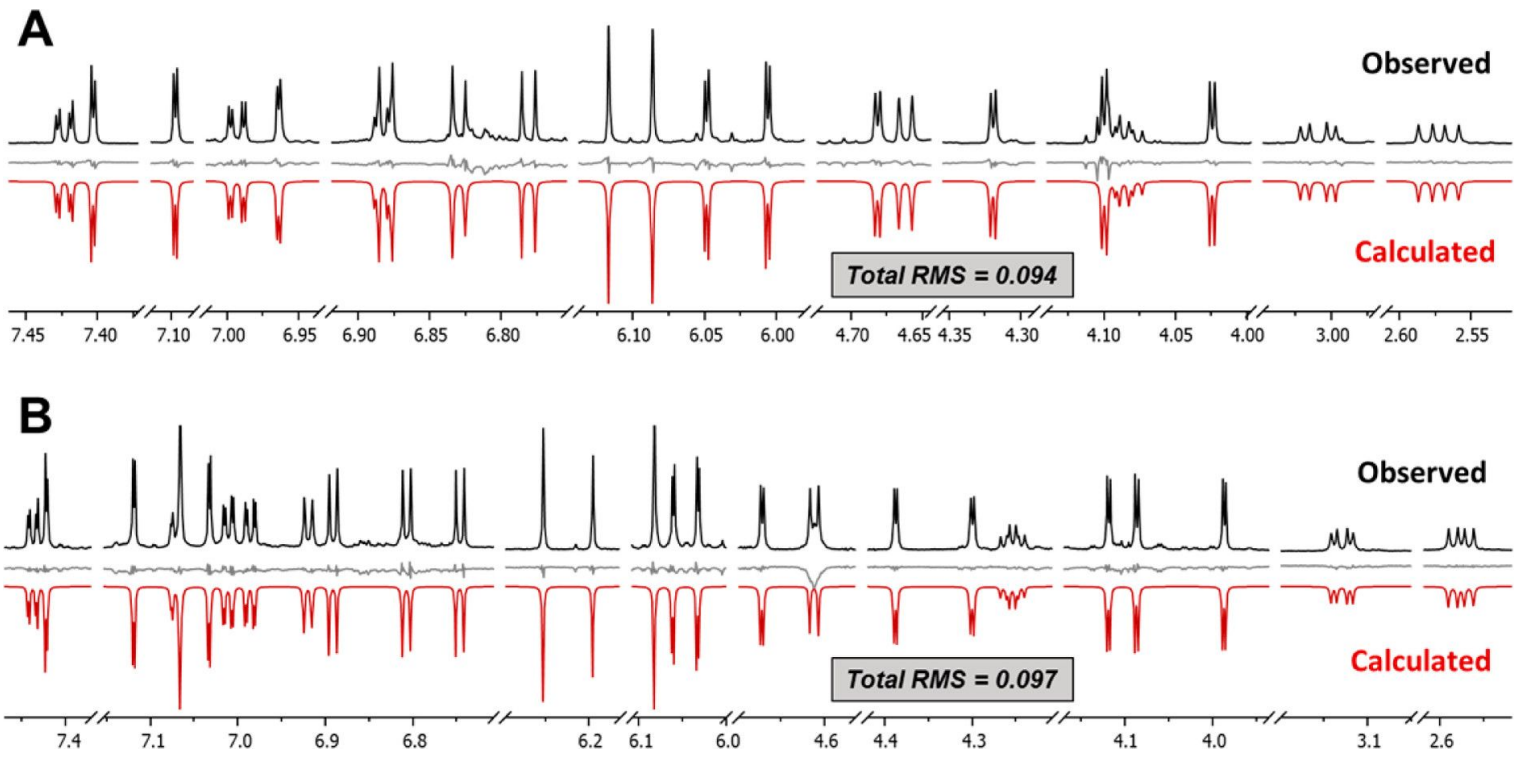
# Dentistry applications






Nam et al, https://doi.org/10.1021/acs.joc.6b02161

# Dentistry applications




Nam et al, https://doi.org/10.1021/acs.joc.6b02161

# qHNMR is Free

- Price of quantitative conditions in 1D $^{1}$H NMR (qHNMR): $0.00
- p90, D1, TD, etc. are a matter of awareness, not cost
- HNMR is essentially **already** quantitative
  - Adjust parameters to run qHNMR routinely!
- Dynamic range
  - Instrument time: $^{1}$H 5 min vs 2D/ $^{13}$C 5 hrs
  - For ~1% level, need to see $^{13}$C satellites

# 100% Method: because we now have flat baselines

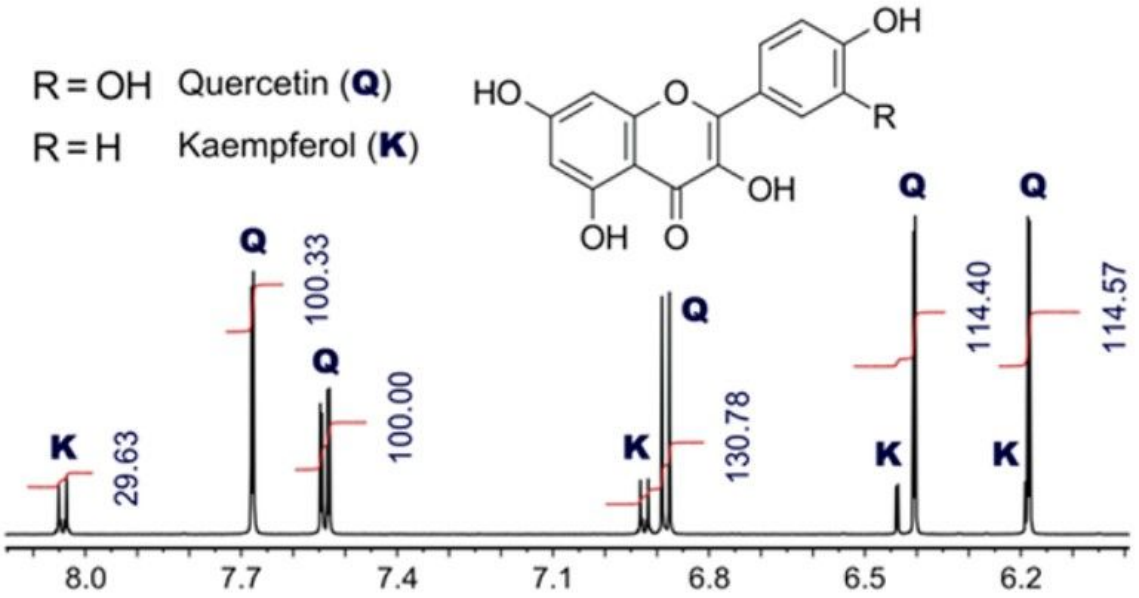
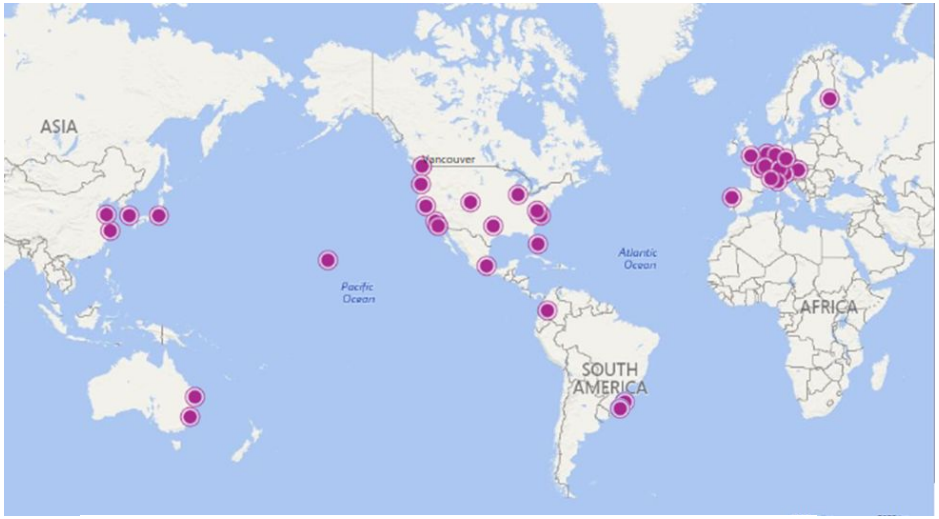



**Figure 1.** Application of the relative (100%) qHNMR method (see also S2, Supporting Information). A commercial sample of quercetin (Q; declared purity >99%; 24.67 mg/mL [not required for purity calculation] in DMSO-$d_6$, 600 MHz) was analyzed. A structurally related compound, kaempferol (K), was identified as an impurity. On the basis of the relative integral ratios, the content of quercetin and kaempferol in the sample was determined as 87.8% and 12.2% w/w, respectively.

Pauli et al. https://dx.doi.org/10.1021/jm500734a

# Community Article
# NMR Raw Data Initiative




- Authors: 73
  (some in this room)
- Laboratories: 38

- **Impossible without Jim McAlpine**
- Manuscript Stats: 60+ major, 100+ minor versions, 113 NP structures, 38 figures, 400 references, 130 pages
- Started 04/2017 published 07/2018

http://dx.doi.org/10.1039/C7NP00064B

The Value of Universally Available Raw NMR Data for Transparency, Reproducibility, and Integrity in Natural Product Research

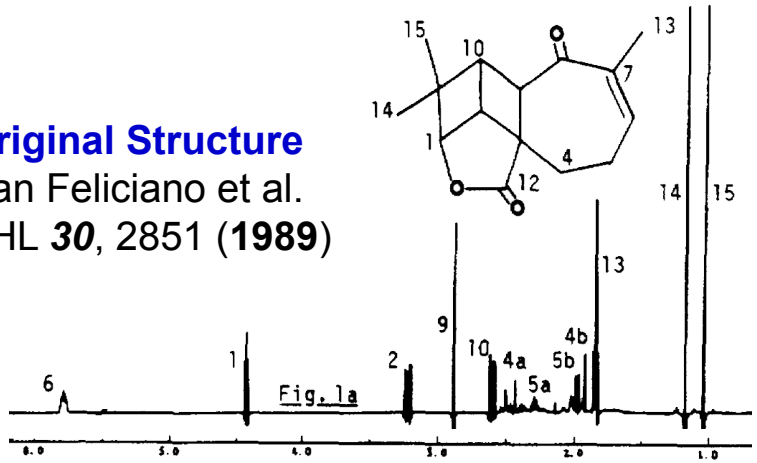
James B. McAlpine,[a*] Shao-Nong Chen,[a] Andrei Kutateladze,[b] John B. MacMillan,[c] Giovanni Appendino,[d] Andersson Barison,[e] Mehdi A. Beniddir,[f] Maique W. Biavatti,[g] Stefan Bluml,[h] Asmaa Boufridi,[i] Mark S. Butler,[j] Robert J. Capon,[j] Young H. Choi,[k] David Coppage,[c] Phillip Crews,[c] Michael T. Crimmins,[l] Marie Csete,[m] Pradeep Dewapriya,[j] Joseph M. Egan,[n] Mary J. Garson,[o] Grégory Genta-Jouve,[p] William H. Gerwick,[q,r] Harald Gross,[s] Mary Kay Harper,[t] Precilia Hermanto,[u] James M. Hook,[u] Luke Hunter,[u] Damien Jeannerat,[v] Nai-Yun Ji,[w] Tyler A. Johnson,[c] David G. I. Kingston,[x] Hiroyuki Koshino,[y] Hsiau-Wei Lee,[c] Guy Lewin,[f] Jie Li,[r] Roger G. Linington,[n] Miaomiao Liu,[i] Kerry L. McPhail,[z] Tadeusz F. Molinski,[aa] Bradley S. Moore,[q,r] Joo-Won Nam,[ab] Ram P. Neupane,[ac] Matthias Niemitz,[ad] Jean-Marc Nuzillard,[ae] Nicholas H. Oberlies,[af] Fernanda M. M. Ocampos,[e] Guohui Pan,[ag] Ronald J. Quinn,[i] D. Sai Reddy,[b] Jean-Hugues Renault,[ae] José Rivera-Chávez,[ah] Wolfgang Robien,[ai] Carla M. Saunders,[aj] Thomas J. Schmidt,[ak] Christoph Seger,[al] Ben Shen,[ag] Christoph Steinbeck,[am] Hermann Stuppner,[al] Sonja Sturm,[al] Orazio Taglialatela-Scafati,[an] Dean J. Tantillo,[aj] Robert Verpoorte,[k] Bin-Gui Wang,[ar] Craig M. Williams,[o] Philip G. Williams,[ac] Julien Wist,[ao] Jian-Min Yue,[ap] Chen Zhang,[aq] Zhengren Xu,[ag] Charlotte Simmler,[a] David C. Lankin,[a] Jonathan Bisson,[a] Guido F. Pauli[a*]

Abstract

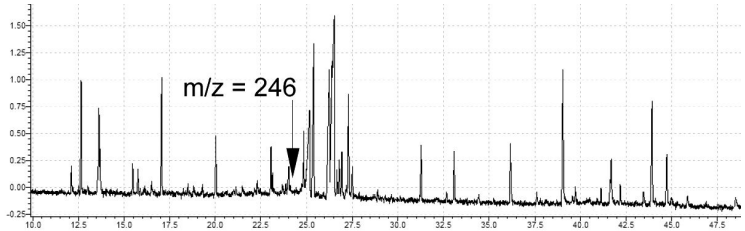
With contributions from the global natural product (NP) research community, and continuing the Raw Data Initiative, this review collects a comprehensive demonstration of the immense scientific value of disseminating raw nuclear magnetic resonance (NMR) data, independently of, and in parallel with, classical publishing outlets. A comprehensive compilation of historic to present-day cases as well as contemporary and future applications show that addressing the urgent need for a repository of publicly accessible raw NMR data has the potential to transform natural products (NPs) and associated fields of chemical and biomedical research. The call for advancing open sharing mechanisms for raw data is intended to enhance the transparency of experimental protocols, augment the reproducibility of reported outcomes, including biological studies, become a regular component of responsible research, and thereby enrich the integrity of NP research and related fields.

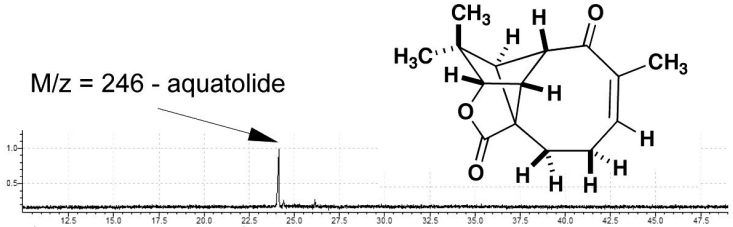
# The Value of Raw Data: The Case of the Aquatolide

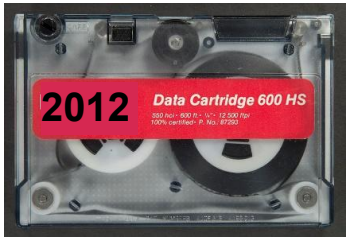
**Original Structure**
San Feliciano et al.
THL ***30***, 2851 (**1989**)



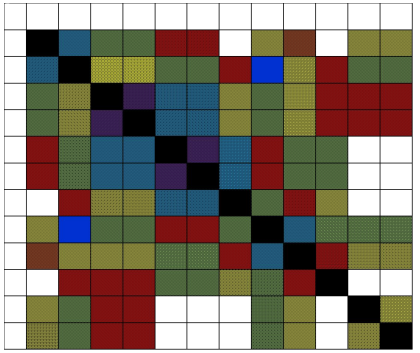

**Revised Structure**
Lodewyk et al. JACS ***134***, 18550 (**2012**)



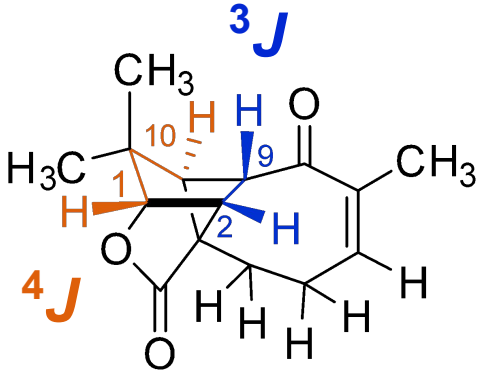




$^{1}$H NMR
QuILT

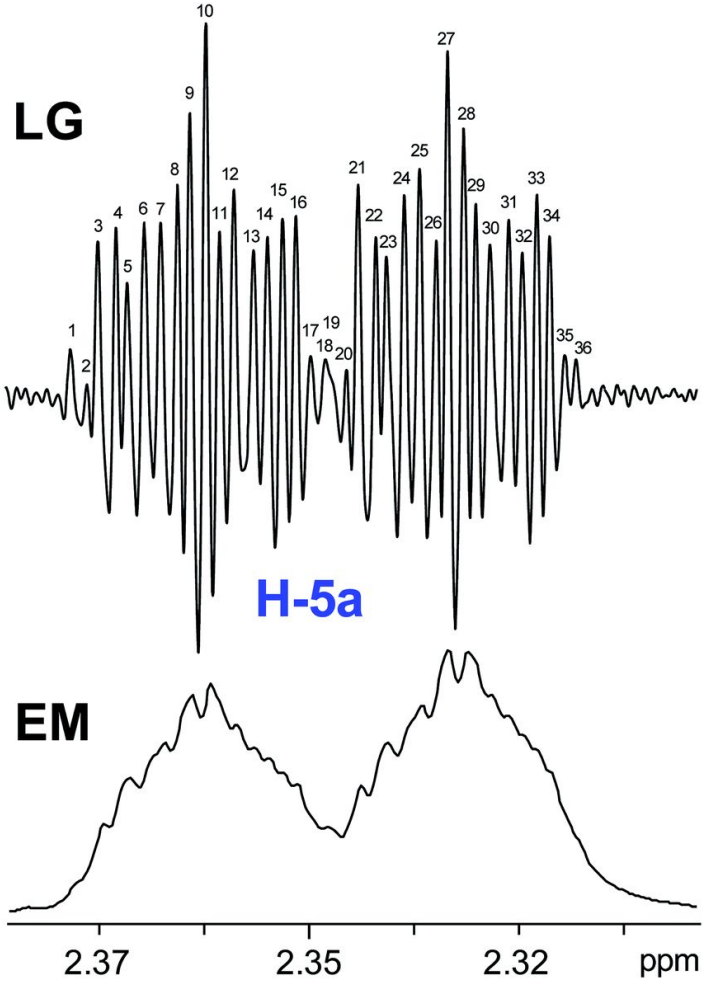


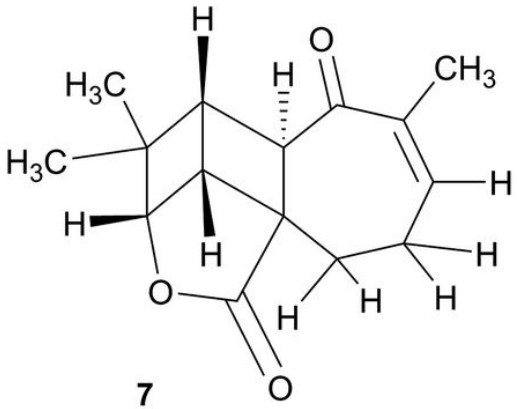
<u>all</u> $^{2\text{-}6}J$
<u>all</u> $\delta$




Pauli et al., https://doi.org/10.1021/acs.joc.5b02456

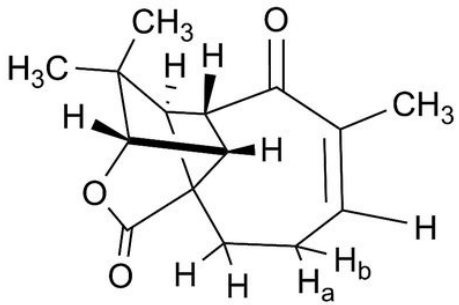
# What Can be Done With RAW Data



7

Aquatolide Original

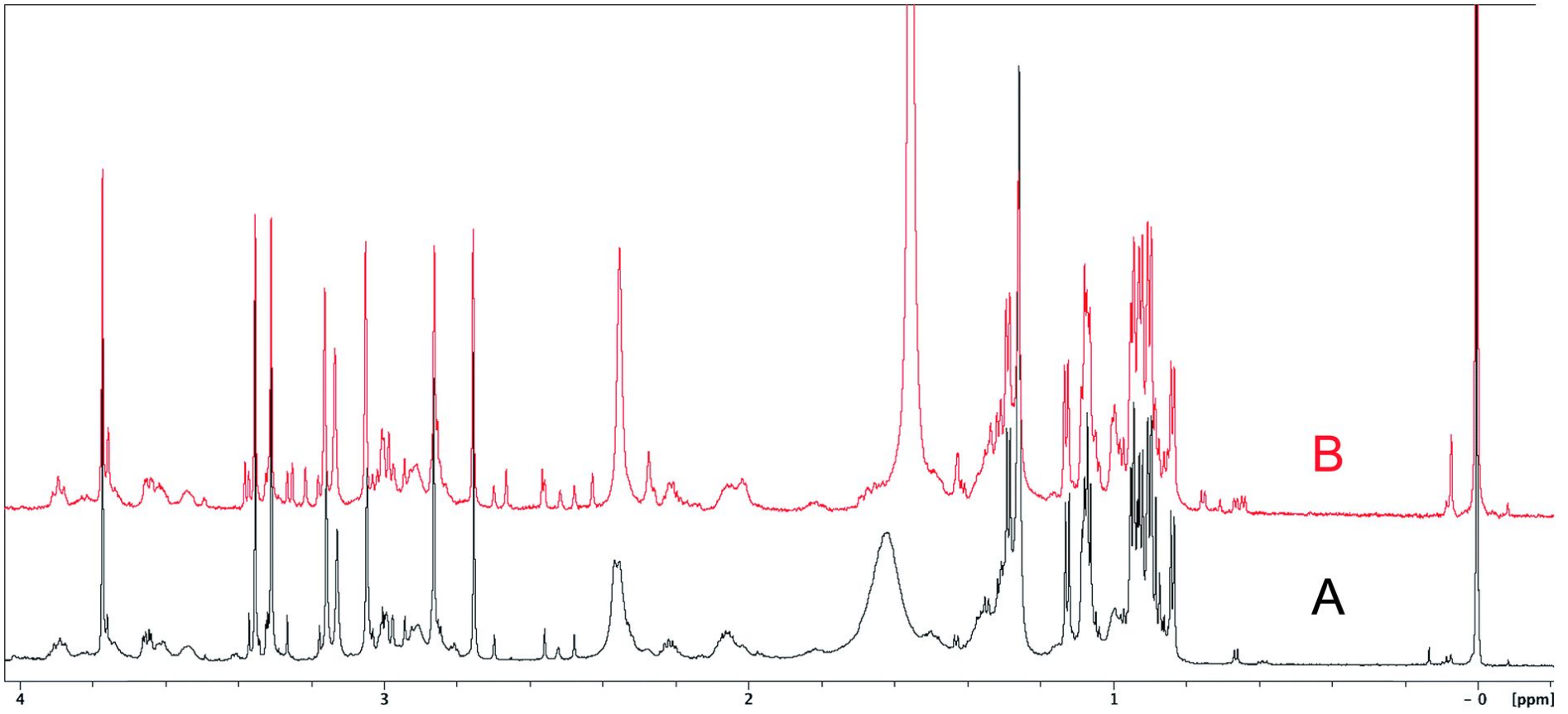


8

Aquatolide Revised

McAlpine et al.https://dx.doi.org/10.1039/C7NP00064B

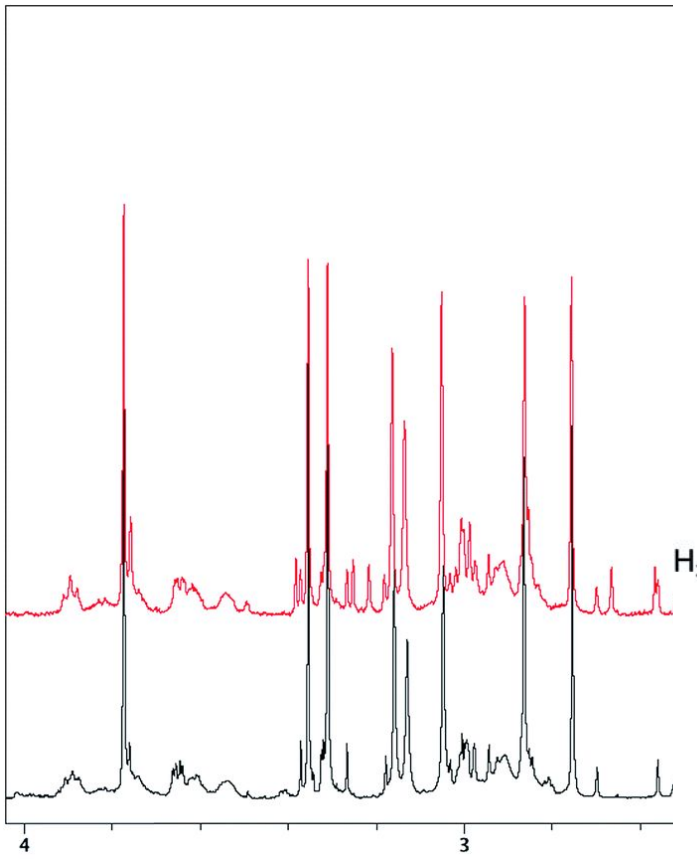
# Are They the Same Compounds?

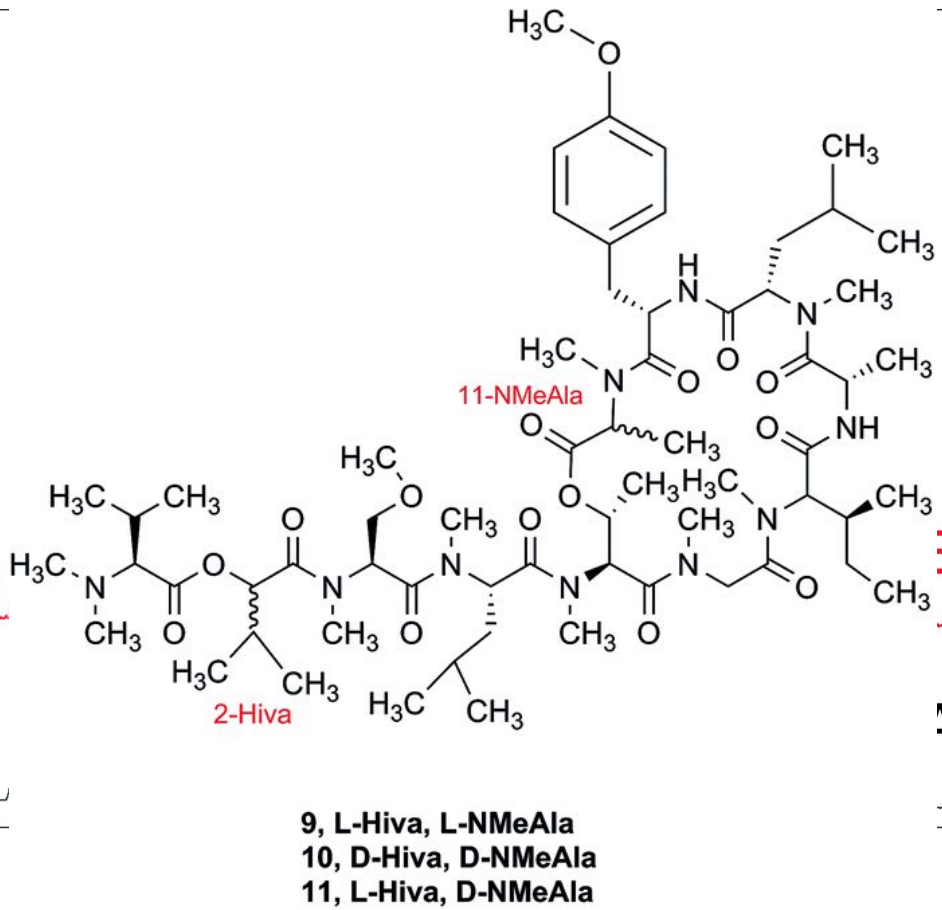



McAlpine et al.https://dx.doi.org/10.1039/C7NP00064B

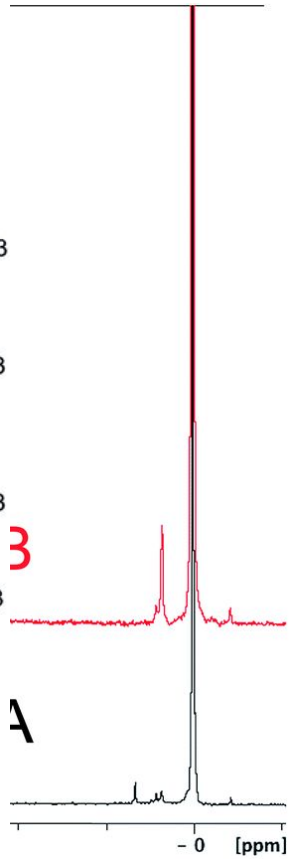
# Are They the Same Compounds?

## If you just look at the table, probably…



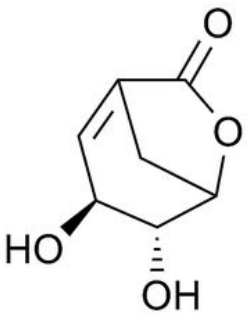

**9, L-Hiva, L-NMeAla**
**10, D-Hiva, D-NMeAla**
**11, L-Hiva, D-NMeAla**




**Coibamide A Diastereomers**

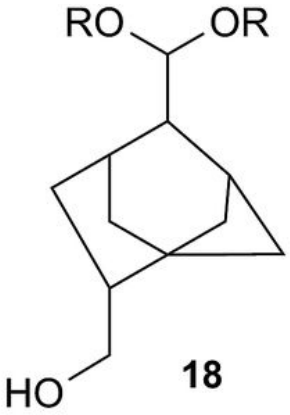
McAlpine et al.https://dx.doi.org/10.1039/C7NP00064B

# Impossible Structures

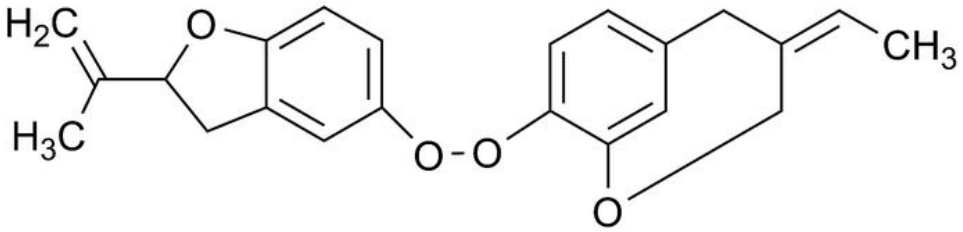



17

**Folenolide**




18




19

**Geometrically Impossible Structures**

McAlpine et al.https://dx.doi.org/10.1039/C7NP00064B

# Why RAW


- Wrong structures: Frequency of Structural Revisions
  DFT calculations[1,2] indicate a potential ~15% error rate in some classes
  Coherent with Wolfgang Robien's results[3]
- Dereplication, avoid working on already known compounds
  Need to balance that with structural revisions…
- Research integrity

<u>NMR is critical for progress in NP research, food research, clinical trials quality…</u>

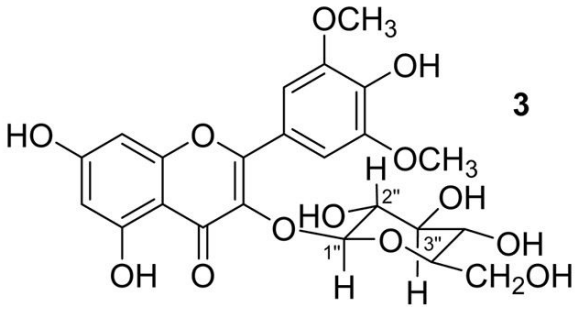
Outcome of NCCIH "Natural Products Data Repository Roundtable Discussion" 06/2017

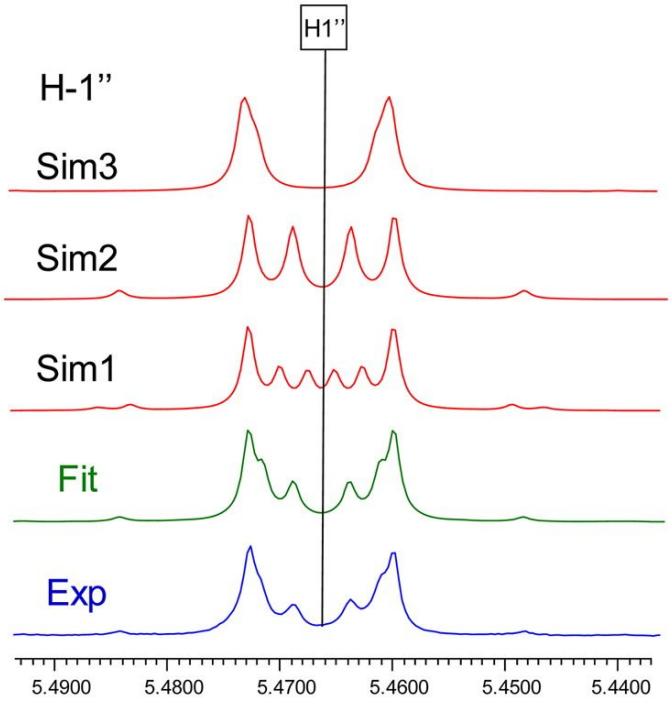
- Enhance or even enables peer-review
- Samples can (will!) be lost, but data can last (if we care)

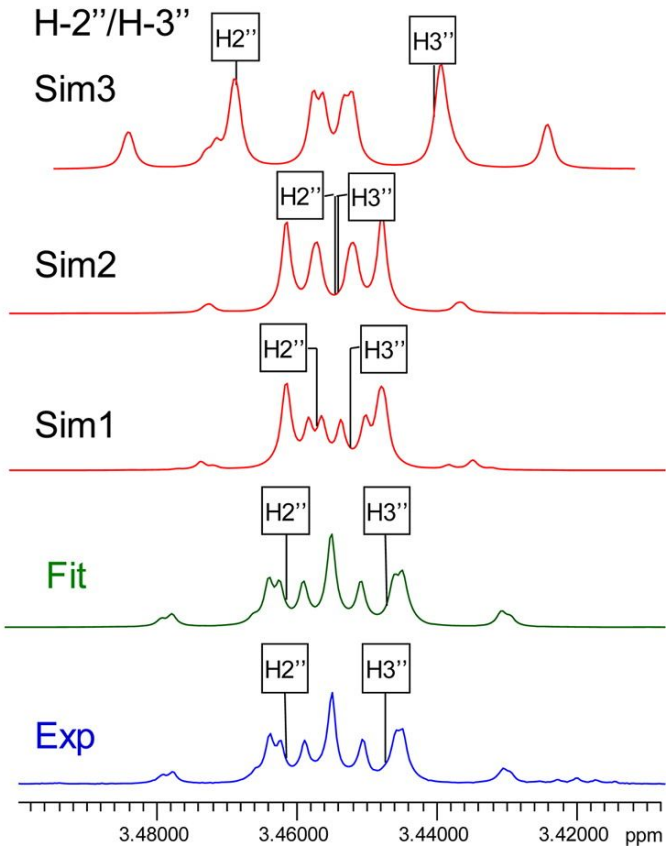
[1,2] Kutateladze et al. JOC 82, 10795 (2017) and JOC 82, 3368 (2017)
[3] Robien. Progress in the Chemistry of Organic Natural Products 105, 137 (2017)

# What (most) published NMR tables do not allow.

| | H-1'' | H-2'' | H-3'' |
|---|---|---|---|
| Sim3 | 5.4660 | 3.4687<br>+0.0070 | 3.4403<br>-0.0070 |
| Sim2 | 5.4660 | 3.4547<br>-0.0070 | 3.4543<br>+0.0070 |
| Sim1 | 5.4660 | 3.4567<br>-0.0050 | 3.4523<br>+0.0050 |
| Fit to Exp | 5.4660 | 3.4617 | 3.4473 |
| 2-decim | 5.47 | 3.46 | 3.45 |



# Quantitative NMR (qNMR) | Validity & Potential

- Implementing qNMR methodology since 1998
  - qNMR for NPs & biomedical research
- NP and pharmaceutical applications
  - Validation of isolates, drug leads, APIs
  - QC of reference materials
  - QC of botanical products
- Highly reproducible and value added
  - 1D $^{1}$H NMR includes qHNMR for free
  - **Raw NMR Data Initiative**[1]
  - Dereplication/ID tool
  - **qNMR Summits** since 2016


**5th qNMR Summit**
USP Headquarters
Oct 2+3, 2019

USP-CENAPT

## qNMR@UIC
30+ publications

PCA 12, 28-42 (2001)
JNP 68, 133-149 (2005)
JNP 75, 834-851 (2012)
GARP {$^{13}$C}$^{1}$H qNMR
JNP 70, 589-585 (2007)
COBiot 25, 51-59 (2014)
AdvNutr 7, 179-189 (2016)

Validation of qHNMR for NPs
HiFSA-qHNMR
JNP 75, 238-248 (2012)
PCA 24,581-597 (2013)
JOC 78, 2827-2839 (2013)
JPBA 93, 59-67 (2014)
JMC 57, 9220-9231(2014)
JNP 80, 634-647 (2017)
JOC 83, 6664-6672 (2019)
JOC 84, 3055-3073 (2019)

gfp@uic.edu

[1]McAlpine et al.https://dx.doi.org/10.1039/C7NP00064B

# How to achieve that?

- Active dissemination and publication
  - When you review papers, ask for the RAW data to be published with the article
  - When you use papers and find doubtful results, ask for the RAW data to compare, do not stay with tables data
  - When you write papers, publish the data with it
  - When you teach, talk to your students about the importance of RAW data

# Some things we hope to do with NMReDATA

- Spin simulations/predictions
  - How to describe spin systems? How far should we go? How to annotate simulations?
- Formal structure validation
  - What are the missing couplings? Is the spin description coherent with the structure? With the spectra?
- Industry support
  - Integration w/ industry formats such as Allotrope's ADF (at least import?)
- Coding/Integration
  - Libraries, we have Java with Stefan Kuhn's work, but what about Python,R
  - Reference implementation? coverage score for implementations
- Data integration support
  - Integration with OBO and ADF (Allotrope) ontologies?
    - Why reinvent terminology (again) ? NMRSTAR, NMRml…
- A Free and Open repository for data, or at least a common protocol…
- We are making a natural products ontology (that extends beyond) that can describe organisms, methods of obtention etc… How can we integrate?



# Thank you

The Guido F. Pauli group

Charlotte Simmler, Shao-Nong Chen, James B. (Jim) McAlpine, David Lankin, Joseph G. (Joe) Ray…

Institute for Tuberculosis Research

Scott G. Franzblau, Sang H. Cho, Mary Choules, Wei Gao, Birgit Jaki…

NMR Solutions

Matthias Niemitz, Samuli-Petrus Korhonen…

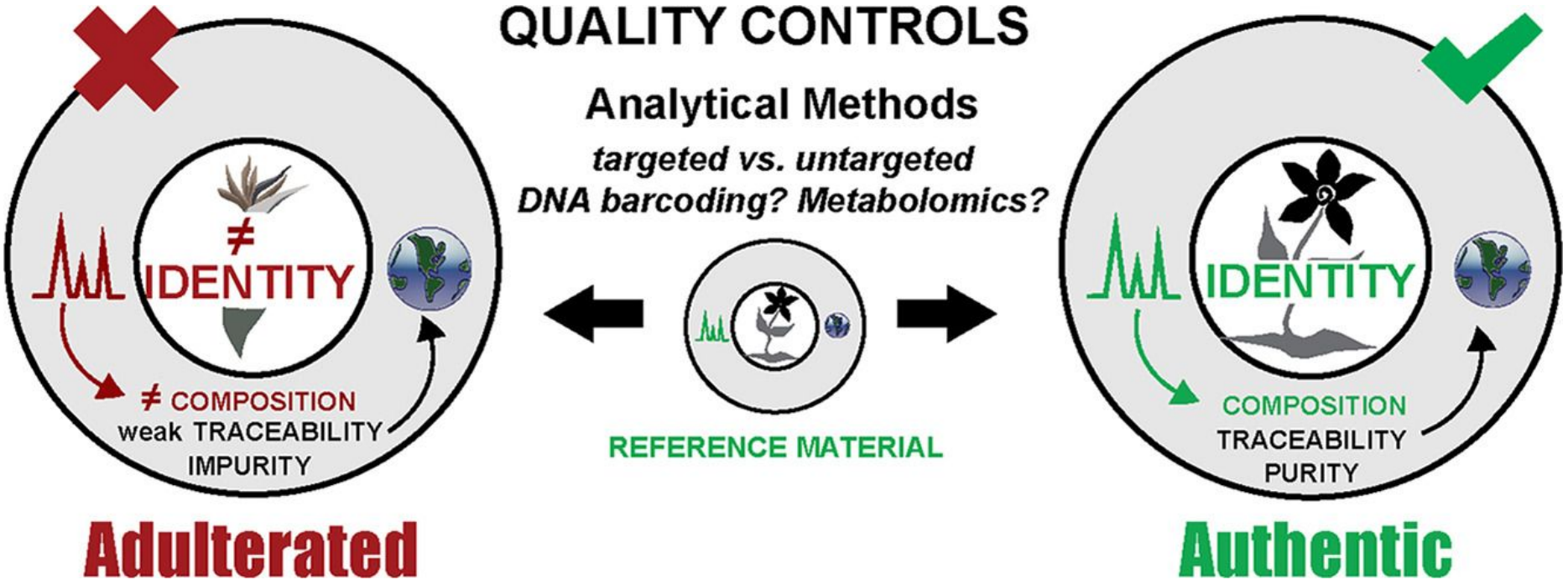
The NMReDATA committee and the sponsors

NIH - NCCIH - ODS

grant U41 AT008706



# Pharmacognosy and adulteration




Simmler et al, https://doi.org/10.1016/j.mote.2014.11.007