QMSA Letters, 2025

The Art of QMSA

Almost Complete Complete QMSA – the Concepts of NMR-Purity and Impurity

Reino Laatikainen[1], Pekka Laatikainen[1], Henri Martonen[1], Tuulia Tynkkynen[2] and Jani Rahkila[3]
[1]Spin Discoveries Ltd., Kuopio, Finland, University of Eastern Finland, Finland, [3]Åbo Academi, Finland (presently Bruker Biospin)

Last update June 14th, 2025

ChemAdder, Yess!

LAOCOON group, which gave the name for the early QMSA program

The Adders, yeah!

From LAOCOON to ChemAdder

1B

What's new ?

This presentation is updated regularly ..

The last update Feb. 12th, 2024

If you have read this before, see pages (number in right upper corner):

- Purity or Impurity – that's the question, page 2
- Demysterification, pages 11A-11C*
- Xpectrals, pages 38-39*
- Xpurities, page 40*

* Only in SpinAdder/ChemAdder, so far.

NaCl analogy

- If one has a NaCl-sample which may contain 0.001% of KCl, there is no method to determine NaCl so that one could say that its purity is 99.999%, but it is easy to measure the KCl concentrations at < 0.001% level.
- With the best analytical methods, one might be able to say that the NaCl concentration is 100±1%. In the case of Na, there is no such accurate method – so far, we know.
- The NaCl purity analysis is analogical for organic compounds, for which it is difficult to find a method, which could give concentration with accuracy better than 1%. If a compound has an isolated NMR signal, it may be possible – although the Response Factors (RF), baseline and signal overlap form a challenge.
- The solution for organic compounds is, like for NaCl, to measure the impurities. There exists one method in which the analysis can be done by a single measurement and without calibration - qNMR !

Purity or Impurity – that's the question

There are two principal ways to characterize the chemical **purity of a compound**:

- 1. **Determine the purity** = the main component concentration. Even if the measurement is done very carefully, the accuracy is seldom better than 0.1%, and very accurate weighting (thus, in sufficient amounts) and pure references for calibration are needed.
- 2. **Determine the impurities.** For them, the demanded accuracy of the method is seldom an issue- it may be sufficient to say that the impurity% is <0.1% or 0.1-0.2% . No very accurate weightings of small amounts and calibrations are not needed!

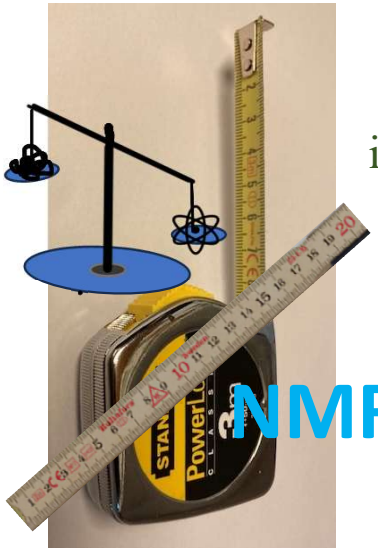
There is one method in which one single measurement gives simultaneously the concentrations of the main component and the impurities with relative accuracies of 1% and (say) 1-10%, respectively:

qQMSA

One can say, for example, that the purity of a sample is >99±1mol% and that the sample contains two impurities with concentrations of 0.10±0.04 mol% and 0.40±0.04 mol% - corresponding to ca. 0.50±0.07 mol% total impurity, so that **NMR-purity** is ca**.** 99.65 mol%! In addition, the impurities can be identified or at least characterized.

The NMR method is *Primary Ratio Method* (PRM) or *METRIC*

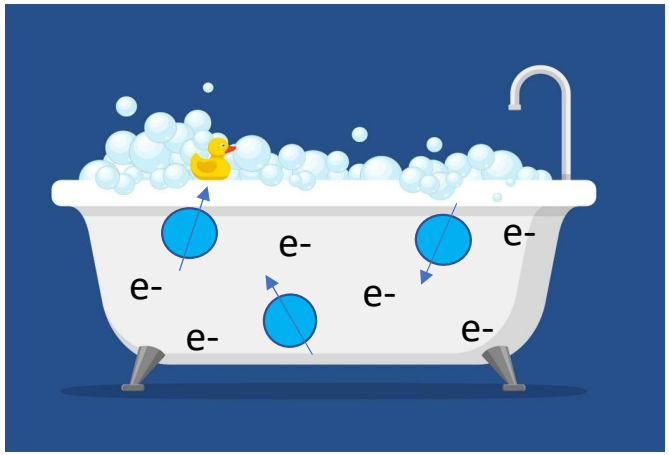
The *METRIC* means, that like a ***metric measure,*** *a measure* can be used without calibration. Also, NMR signal area as compared to a reference signal – **the *measure*** – is the same if the measurements are done in the same way anywhere, ..even in different magnetic field strengths.



The reason is that the intensity of an NMR signal is proportional to number of protons, independently of the compound where the nucleus locates …**but only on the condition that the relaxation delays are long enough, and that the experimental procedure (pulse program) does not contain certain phases! The key parameter in this context is RESPONSE FACTOR (RF) – see below!**

very SPECIAL OFFER

- We are looking for examples of **qQMSA Purity Analysis**, to gather experiences and to make our tools more automatic.
- If you think that you have a candidate for such an analysis, we are ready to make the analysis for **free**!
- To ensure the quality of the data, the spectra are measured by our collaborator.
- The better, if there would be a group of a few related compounds and if, not necessarily, the analyses can be used later as our examples.
- The **qQMSA Purity Analysis** menu: ^{13}C decoupled spectra in 2 - 3 solvents (to reveal impurity signals hiding under major multiplets), chiral purity test, 1D TOCSY for characterization of impurities, …etc.
- Contact us through our homepages!



qQMSA – quantitative Quantum Mechanical Spectral Analysis

- Nuclear spins hang or float in molecular electron cloud. They interact only weakly with the environment, so that the life-times of the spin-states in magnetic field are very long, seconds...even minutes. If the molecules are in isotropic environment, rotating very quickly, a high-resolution NMR spectrum can be observed. The magnetic interactions (couplings) between the floating spins lead to the diagnostic fine-structure of the signals.
- The **energetics** of the spin-states of coupled protons, and the **relative transition intensities within multiplets** obey the laws of Quantum Mechanics, perfectly .
- The ^{1}H NMR spectrum of a compound can be calculated into *very details*, using the only NMR parameters (chemical shifts and coupling constants, which do not depend on the instrument) and parameters describing line-shapes.
- The rest of the spectral area, not explained by the model, represents to other compounds (solvent, reference..) and impurities which are not in the model.

Purity analysis

The chemical purity of a sample is The Issue in the drug compound or starting material purity analysis. The conventional strategy is to measure area of a selected NMR signal and to compare it to a reference compound.

To show that the purity of a sample is (say) >99.8 %, demands ultrapure references compounds, very accurate weighting of samples (enough sample and reference, and even *buoyancy* should be considered), and a good integration protocol*!*

Even 0.2% uncertainty in weighting, purity of reference or integration is then too much ! And if an impurity signal is hiding under the selected signal, the integration may give too high purity, using all the signals decreases odds of this bias.

ImPurity analysis

From the user's view-point it is usually the same if the material contains 99 or 100% of the key compound – in the 1% includes water, salts, etc.!

The determination of the percentage (comparing the target signal to quantitative reference) with accuracy better than 1% is demanding !

Essential is that there are not too much certain impurities !

The impurity concentrations can be determined by accuracy that is far better than demanded !

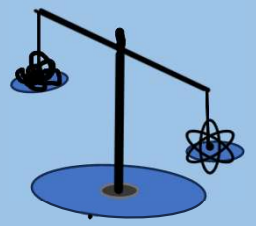
QMSA of whole the spectrum is **The Tool** for impurity analysis – with minimal calibration and preparations!

The *'principle of the very details'*

- The ^{1}H NMR spectral area of a compound can be modelled into *very details* - the rest of the area represents to impurities or compounds which are not in the model:

  NMR-impurity = 100 * [Observed area-Simulated area]/[Observed area]

- Then [**100 - NMR-Impurity**] represents the **NMR-purity** of the sample.
- **NMR-purity** of 100% means that there are no other compounds that give ^{1}H NMR signals, but not that there would be no salts.
- Also, impurity signals hiding under complex spectra can be revealed by QMSA!



IMPURITY ANALYSIS

ALMOST COMPLETE COMPLETE ANALYSIS OF GLUCOSE ^{1}H NMR SPECTRUM

Impurity concentrations of 0.1-0.3 wt% should be measurable !

NMR Statistics

RMSE = RMS Error - depends on spectral width !!

R2 = The fraction of variance explained by model - depends on spectral width !!

Noise = RMSE for signal free parts of spectra

Essential Range = percentage of regions with NMR signals (<4 x Line-Width from nearest QM line, larger than 1%)

Essential RMSE = RMSE for the Essential Range, independent of spectral width !!

Essential R2 = R2 for the Essential Range, independent of spectral width !!

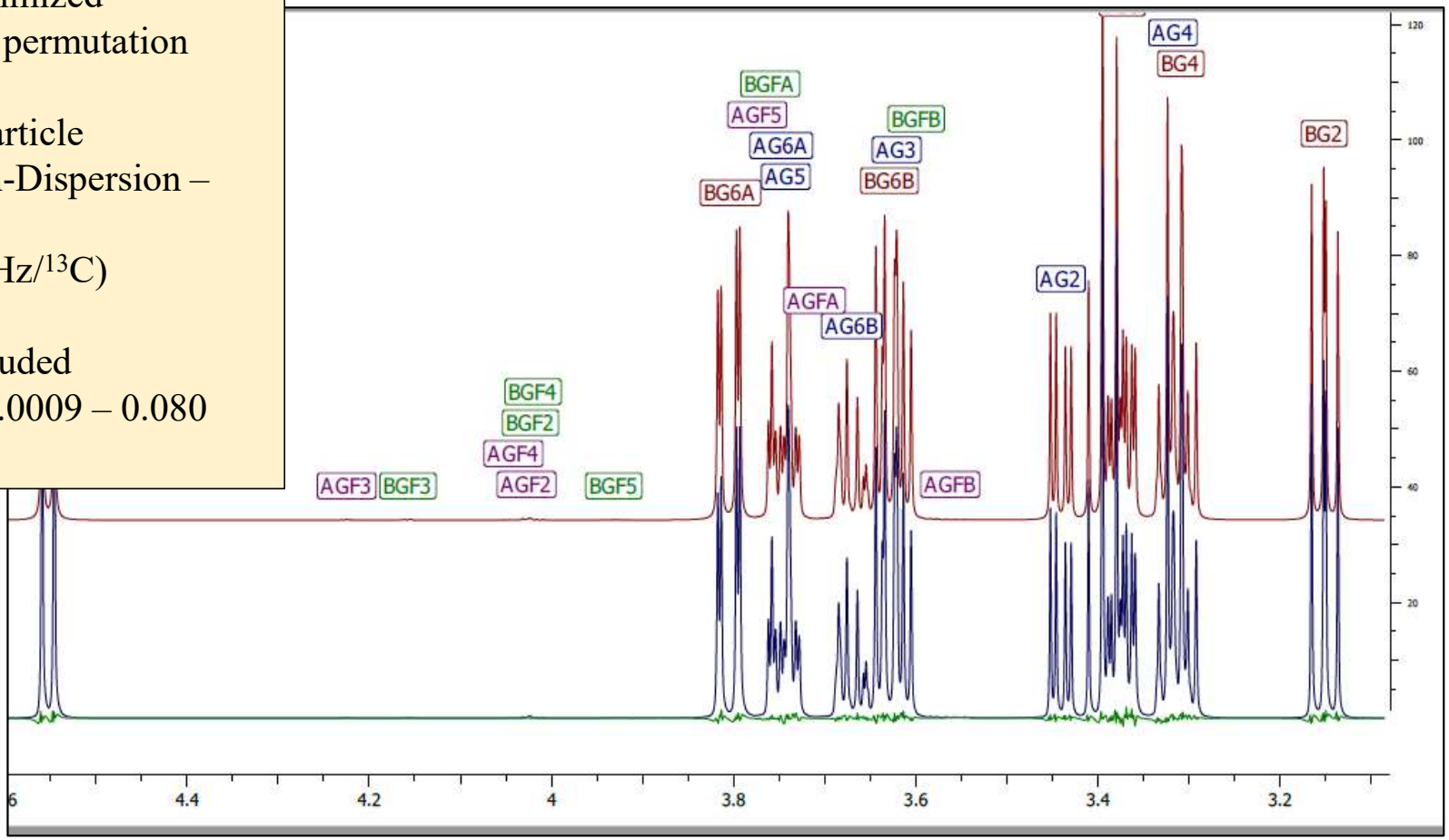
NMR-Purity = percentage of spectral explained by QMSA

NMR-Impurity = percentage of spectral NOT explained by QMSA

eRMSE is used in error analyses and a good measure of spectral quality !

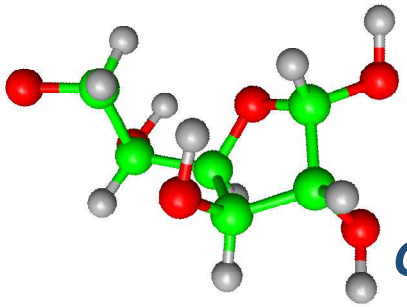
Almost Complete Complete QMSA of Glucose

- ^{13}C Decoupled ^{1}H spectrum
- Chemical shifts & couplings optimized
- Long-range couplings after sign permutation analysis
- Different line-widths for each particle
- Line-shape: Lorenzian-Gaussian-Dispersion – Asymmetry
- Geminal ^{13}C isotope shifts (0.7 Hz/^{13}C)
- Response Factors optimized
- 2 Minor furanose tautomers included
- Several other trace impurities (0.0009 – 0.080 mol%), see pages 15-17



- Noise = 0.01%
- RMSE = 0.15%
- **Essential RMSE = 0.32%**
- Essential R2 = 99.96
- **NMR-purity = 99.86%**

The *'Mysterious Line-Shape Problem'*



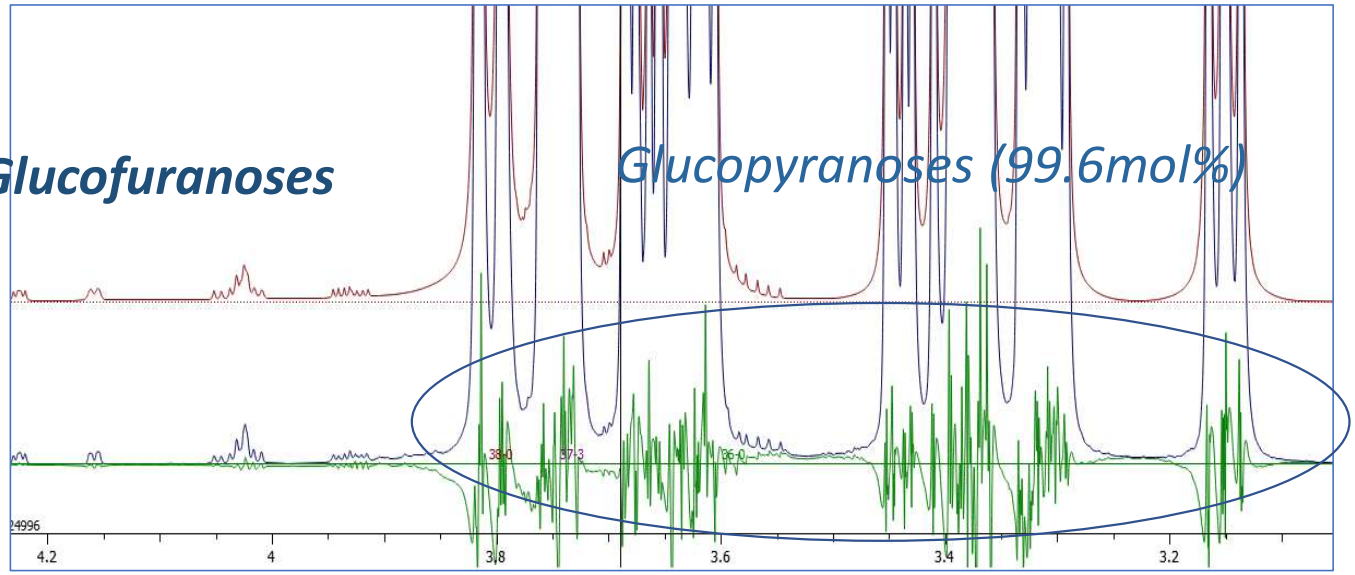
Glucose ^{1}H NMR 2-6-H region

Glucofuranoses

Glucopyranoses (99.6mol%)

Calculated

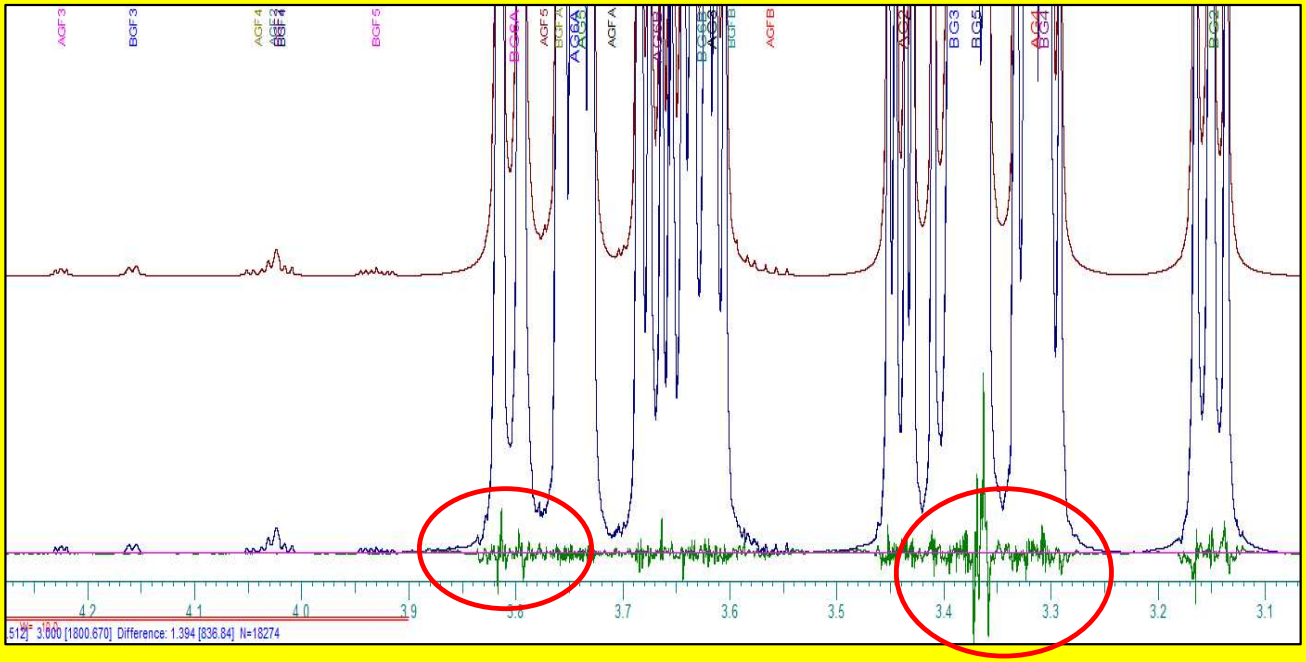
Observed



Observed-calculated trace

Essential rrms = 0.32%

…The same after the ‘demysterification’ – still under testing !

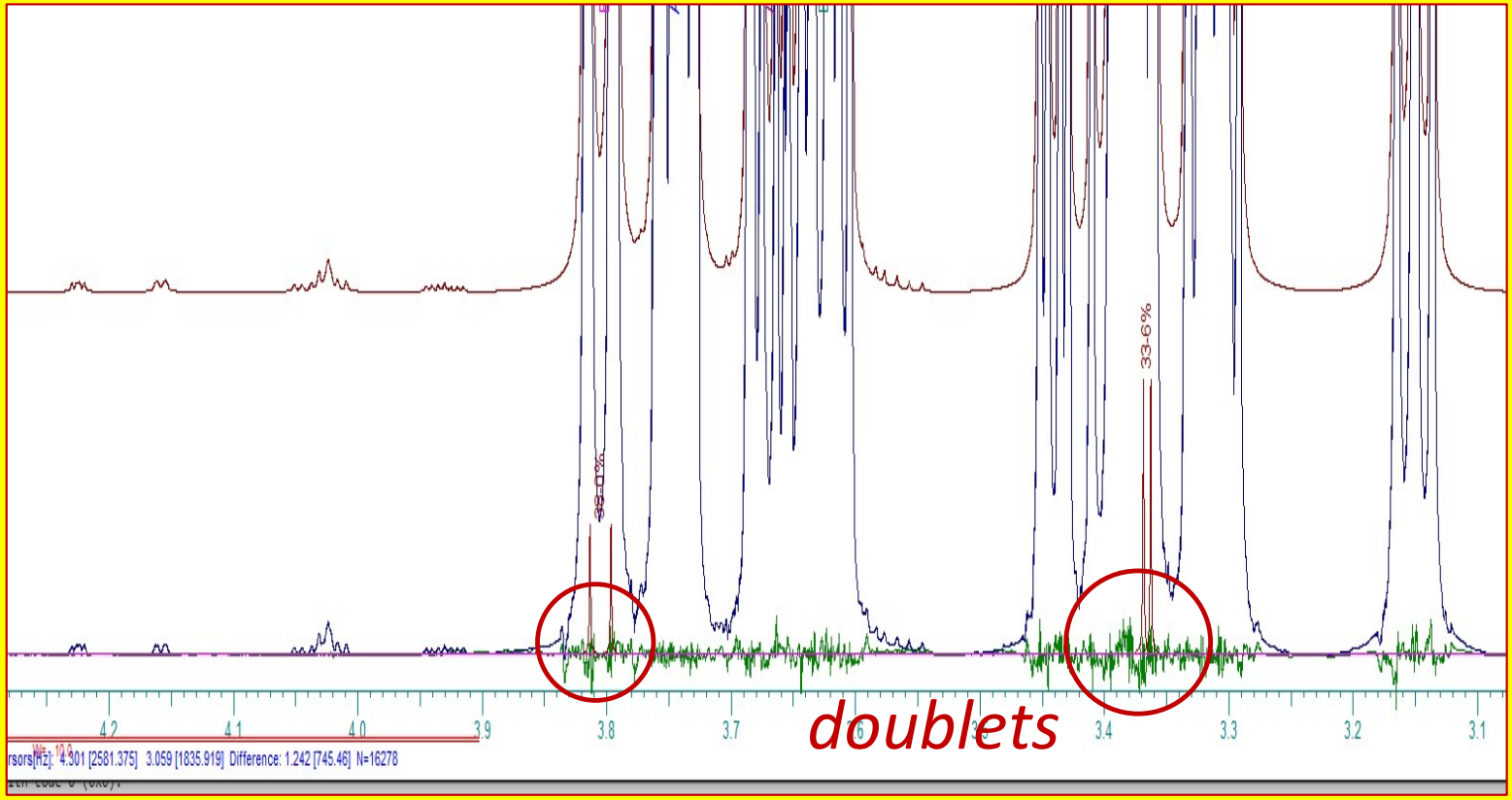


Impurities?

Essential rrms = 0.11%

Updated Oct. 10th, 2023

…and when two doublets are added to the model

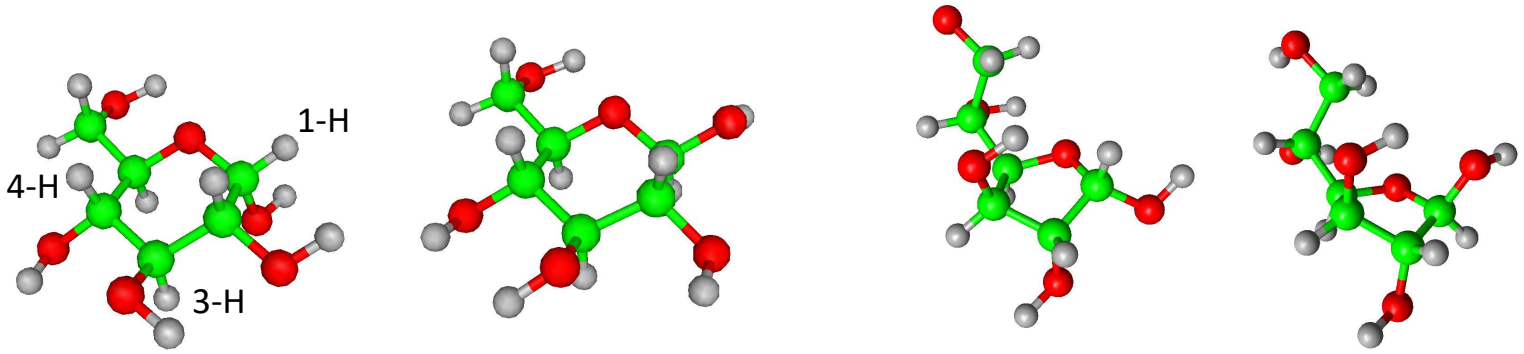


Essential rrms = 0.095%

NEW: Upgraded Oct. 10th, 2023

QMSA of 1D TOCSY of glucofuranoses

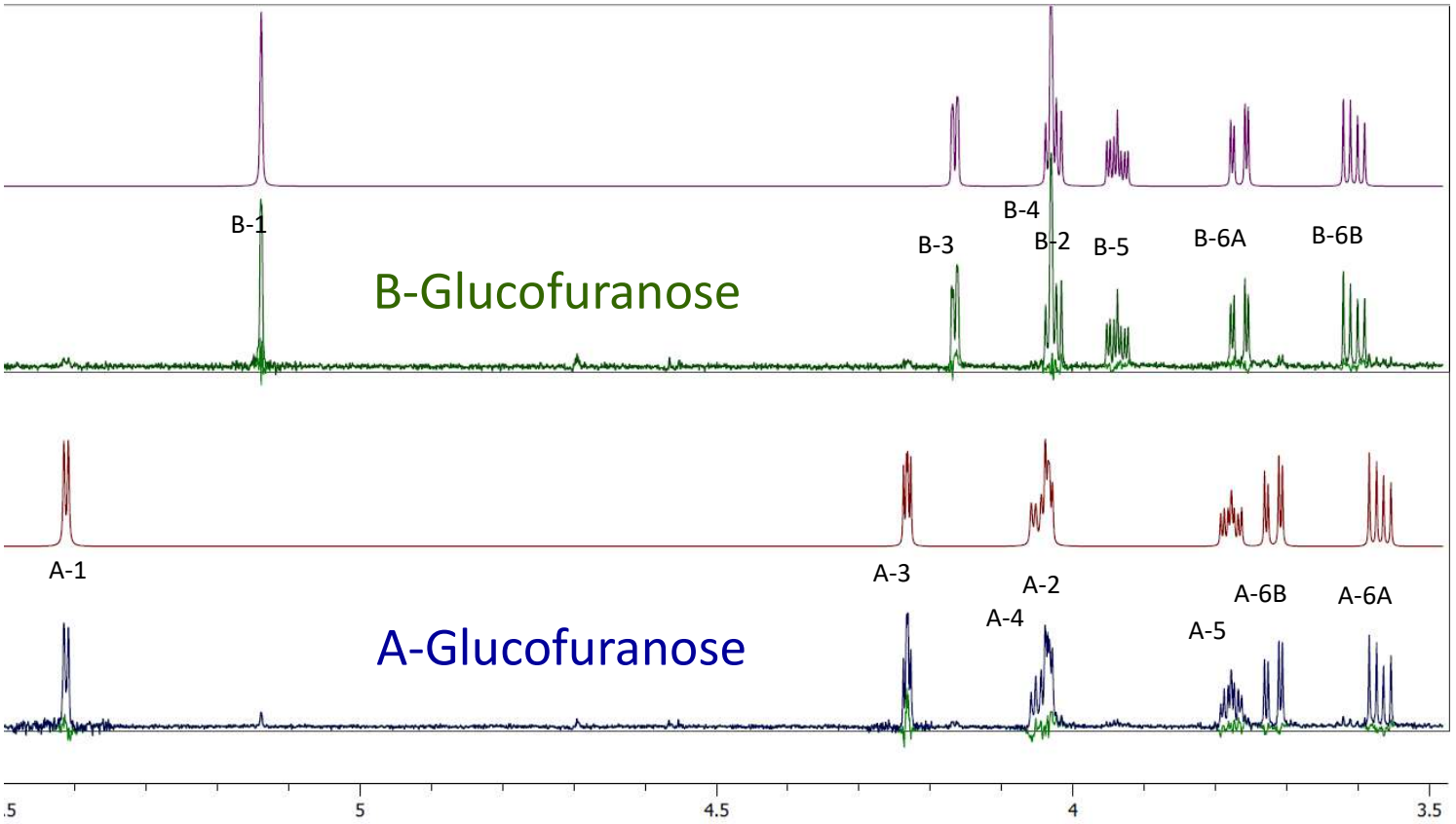
Glucose has 2 major glucopyranose + 2 notable glucofuranose + 2 minor open-chain tautomers



A-Glucopyranose **B-glucopyranose** A-glucofuranose B-glucofuranose

- 1D TOCSY spectra can be analyzed with QMSA – like normal 1D spectra
- The populations of 0.20 and 0.15 mol% are obtained for the glucofuranoses
- The spectra parameters can be exploited in analysis of glucose spectra
- **The isomers can be thought as typical low concentration impurities !**
- **1D TOCSY is an ideal tool for identification of impurity structures !!**

QMSA of ^{1}H TOCSY of glucofuranoses



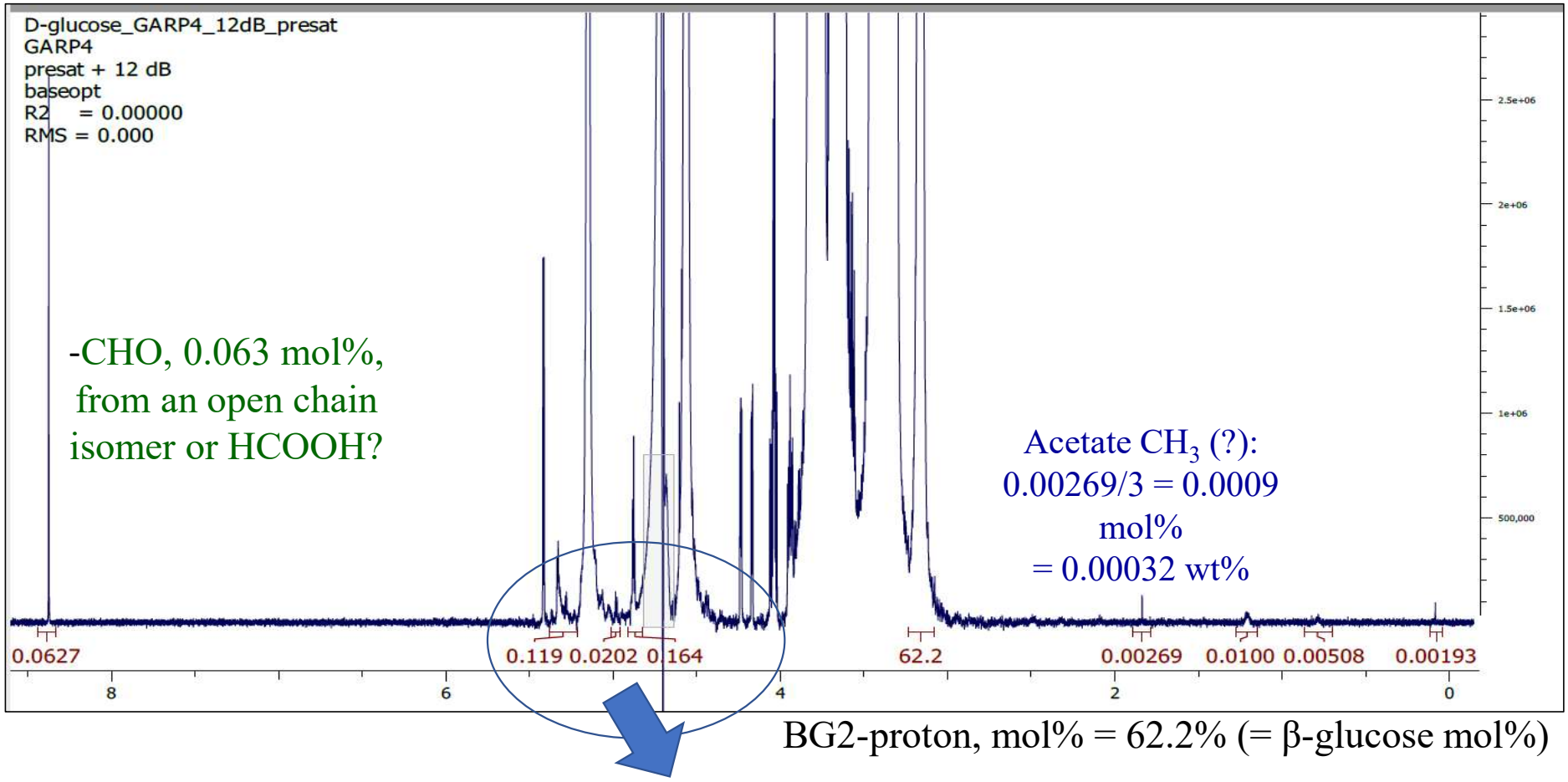
Almost Complete Complete QMSA of Glucose

QMSA is not necessary if all the impurities are visible – which cannot be decided before QMSA !

The impurity concentrations can be quantitated also by using the **traditional integration** – or with **Total-Line-Shape fitting** using **spectral xtructures**

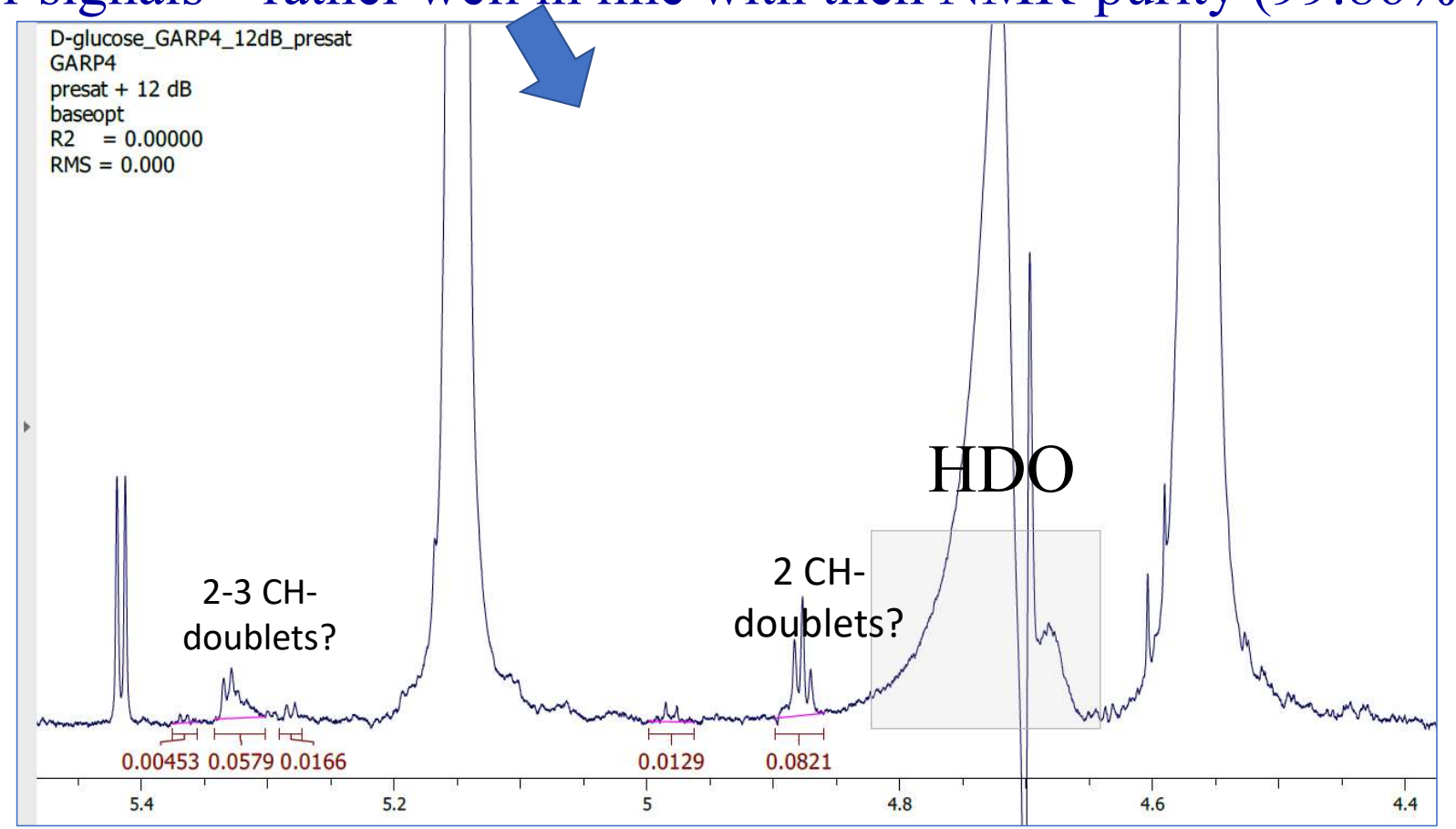
BUT TO BE SURE THAT THERE ARE NO IMPURITIES HIDING UNDER MAJOR SIGNALS, QMSA IS THE WAY !!

The mol%'s of rest of impurities can be determined by **integration**, …here using the B2-proton signal as the quantitative reference:

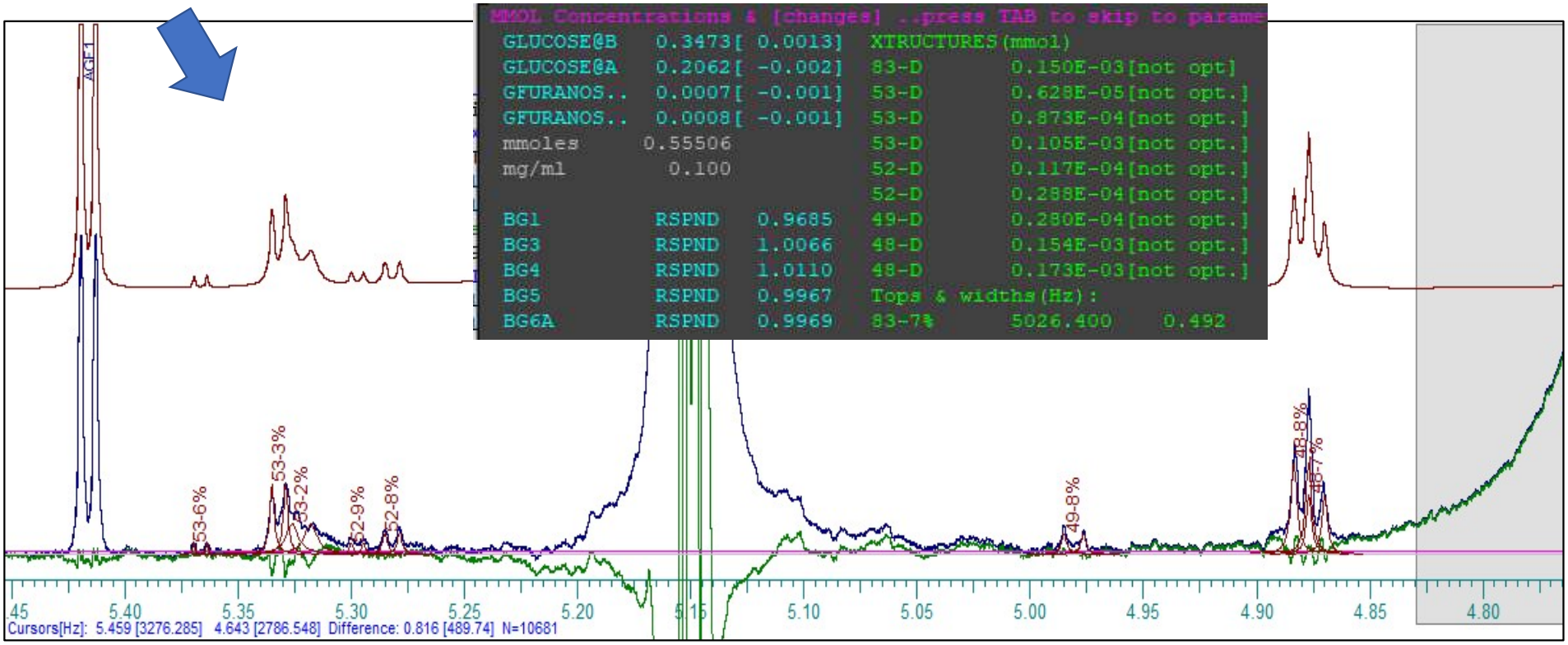


BG2-proton, mol% = 62.2% (= β-glucose mol%)

..also 5-8 glycoside CH-doublets (0.004 – 0.080 mol%) are observed, so that their total ca. 0.24 mol% - the rest of the spectra lies under the major signals – rather well in line with then NMR-purity (99.86%):



…or the quantitation can be done using '*spectral xtructures*' (= TLS-fitting) …the analysis gives also the splitting's (couplings, here 4.6 Hz)%



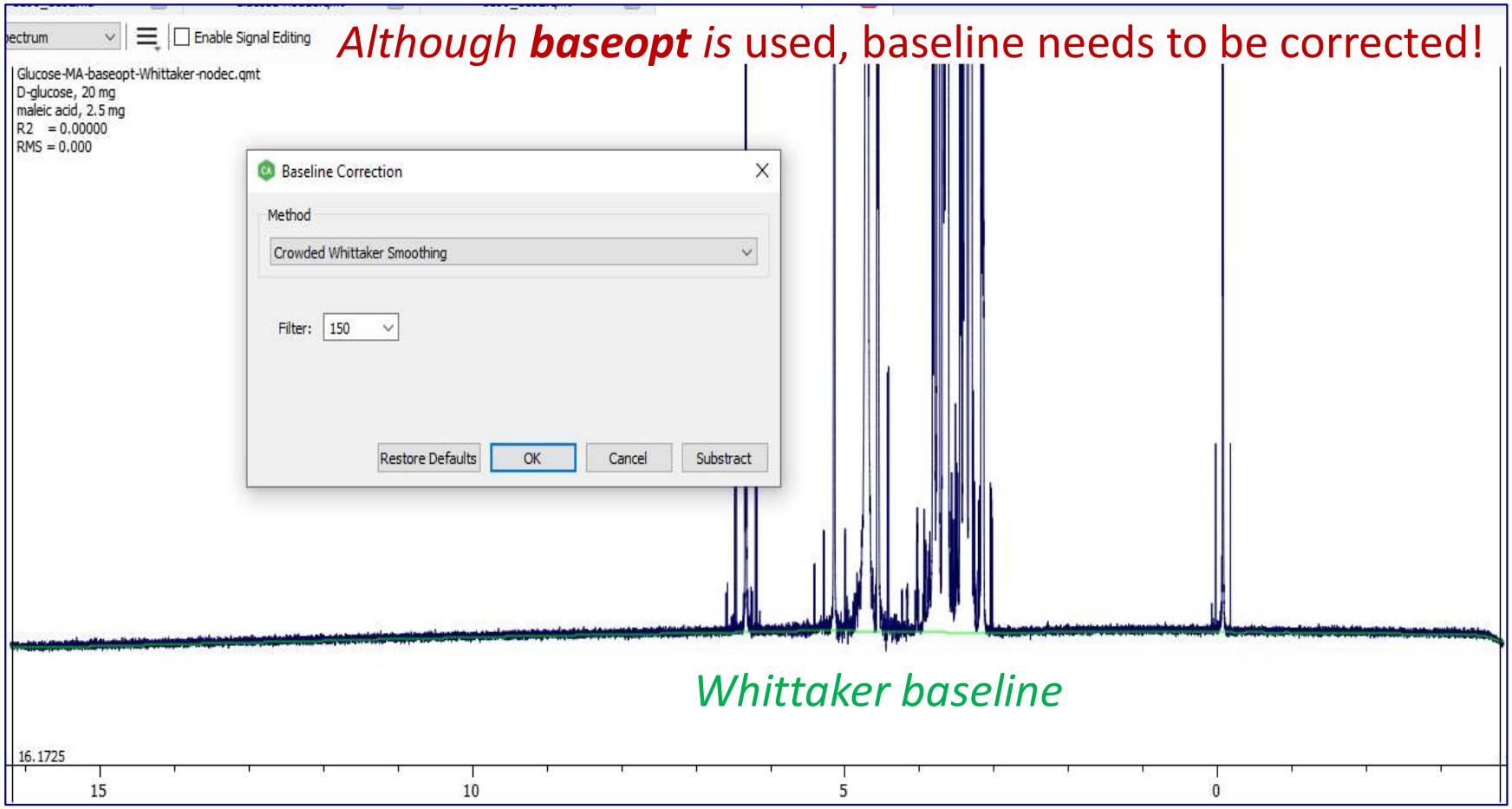
% The % in the names of the xtructures (for example, 53-6%) tells that the **xtructure** at 5.36 ppm is regular type – there are four different (more or less regular) **xtructure** types in SpinAdder– see slide 39.

EXPERIMENTAL

Measurement of ^{1}H NMR spectra for qQMSA

- Long relaxation time (10 – 100 secs)
- Remove of **^{13}C satellites** (DISPEL, or random GARP4 ^{13}C decoupling)
- Remove **long-range ^{13}C,H couplings** by decoupling – ^{13}C isotope shifts remain and must be taken into account!
- Use ***baseopt*** option in preparation of the spectrum - yields a better baseline!
- Use Whittaker base-line smoothing

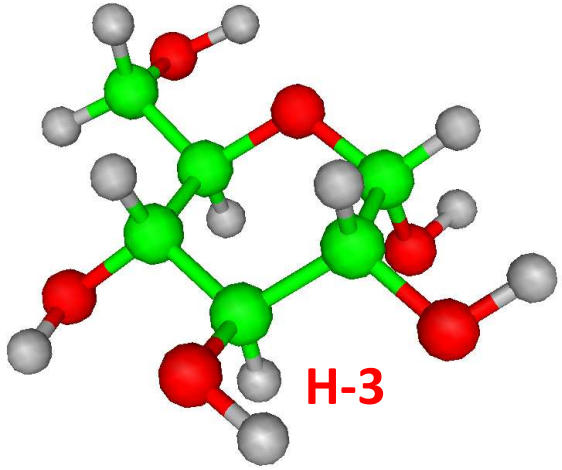
Whittaker baseline correction & smoothing



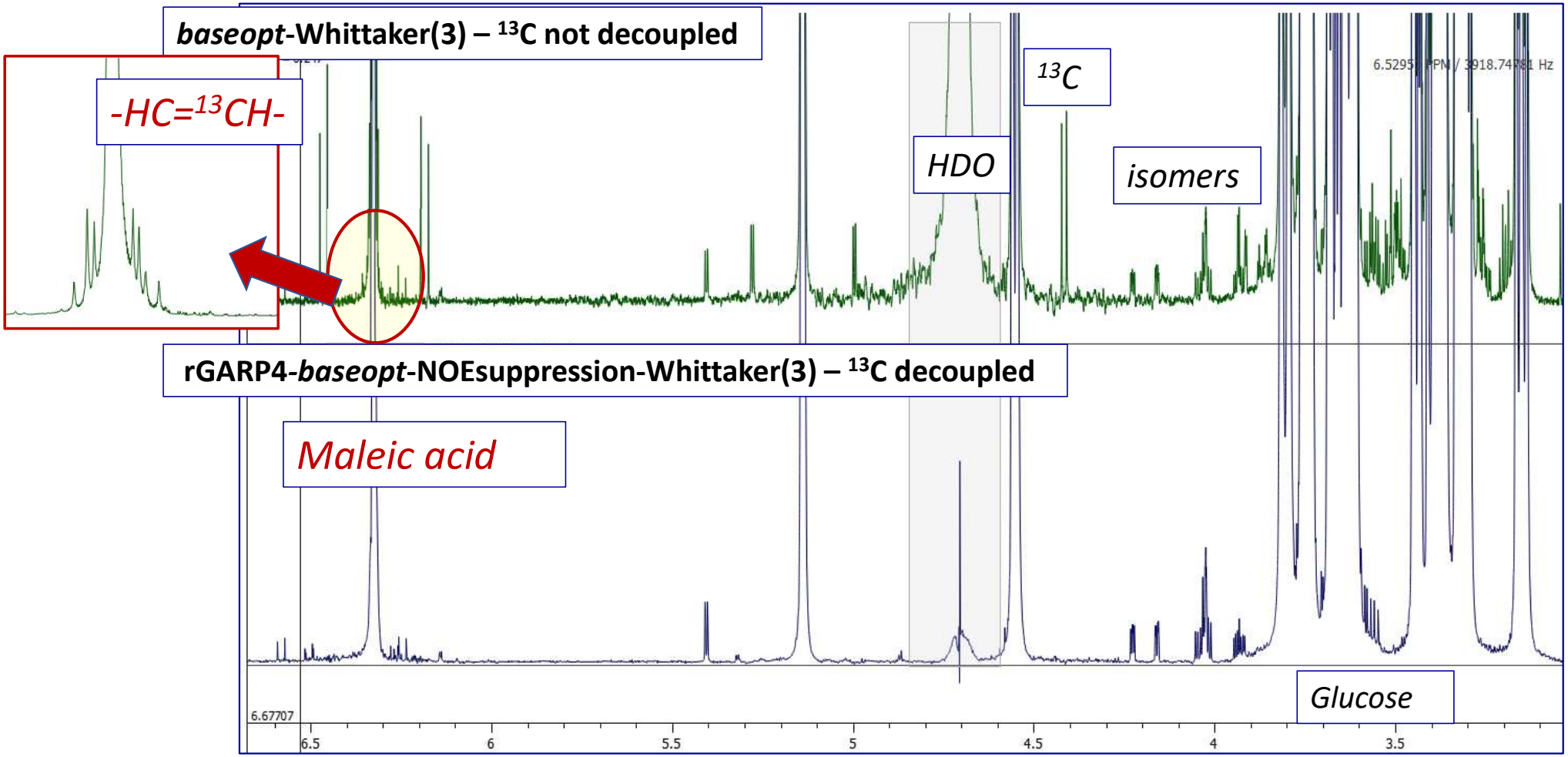
Long-range $^{13}C,^{1}H$ couplings produce a complex background to ^{1}H multiplets

For example, the glucose H-3-proton signal is composed of

- ^{1}H multiplet without ^{13}C couplings (**94.4%**)
- *Satellite spectrum* from 1-bond coupling to ^{13}C-3 (1.1%)
- 2 Spectra arising from geminal (4-5 Hz) couplings to ^{13}C-2 & ^{13}C-4 (2.2%)
- 2 Spectra arising from vicinal (1-2 Hz) couplings to ^{13}C-1 & ^{13}C-5 (2.2%)
- Spectrum arising from 4-bond (< 1Hz) coupling to ^{13}C-6 (1.1%)



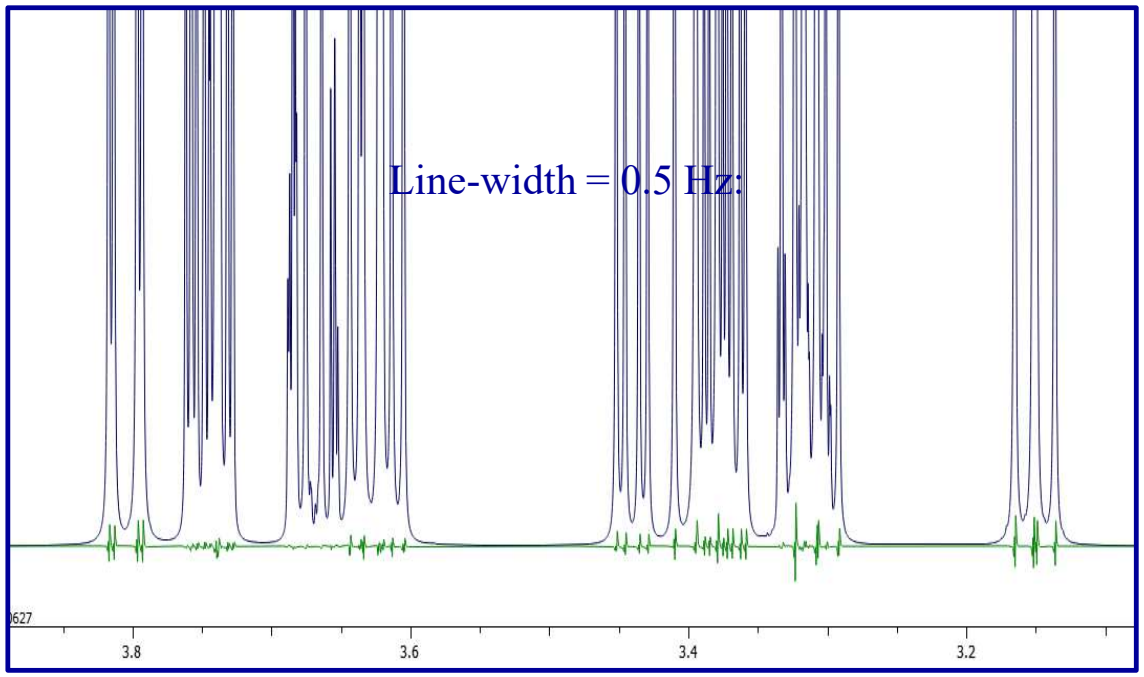
Removal of ^{13}C couplings using '*random GARP4 decoupling*'



Isotope shifts, see also QMSA Letters 1(2022)

The **effective*** ^{13}**C contributions** to the calculated 600 MHz spectrum of glucose:

*The 'effective' means that after adding isotope effects to a synthetic spectrum, the line-shape parameters were optimized so that the ***essential RMSE*** dropped from 0.25 to 0.14%. This means that a part of the isotope effect is compensated by asymmetry and out-of-coil correction.



Conclusion: the ^{13}C isotope shoulders cannot be completely described with **Lorenzian-Gaussian-Asymmetry-Dispersion** line-shape.

The effects may be significant in impurity analyses - but less at benchtop.

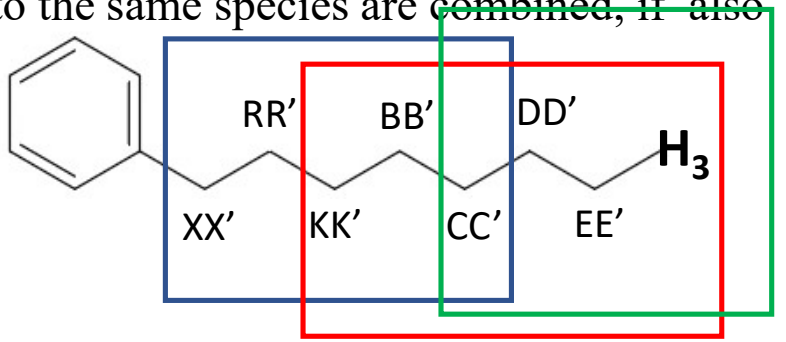
Why only *Almost Complete Complete QMSA*!

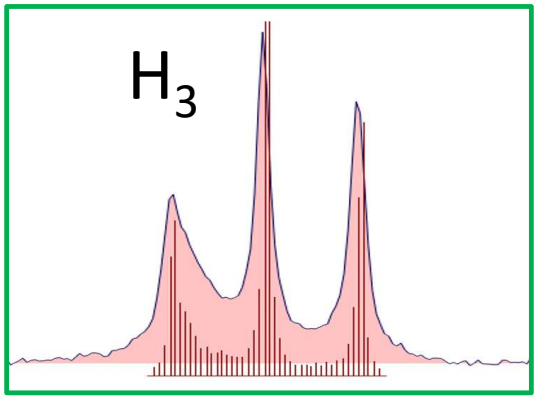
- Incomplete permutation of signs of the LR couplings
- All the geminal ^{13}C isotope shifts equal
- No vicinal and long-range ^{13}C isotope shifts
- Other isotope effects ignored ($^{17,18}O$) – see *QMSA Letters 2(2022)*
- Dipolar couplings ignored
- Minor (for example, open-chain tautomers) impurtities not treated with QMSA
- **Spin-system-packing** approximations - none in the glucose case
- …

The above effects are **spin-particle specific**, while the instrumental artefacts are similar for all the species !!

Spin-system packing (ignoring long-range couplings and second-order effects) may lead to proton specific bias?

- Compression of spin-networks
  - XX'RR'KK'BB'CC'DD'EE'H_3 → **XX'RR'KK'**BB'CC' + KK'**BB'CC'DD'EE**H_3 + DD'EE'**H_3**
  - Transitions separated by < 0.01 Hz and belonging to the same species are combined, if also their derivatives are similar.
  - 58 000 000 → 24 000 Transition, from 64 to **4** sec
- Multithreading
  - Analysis of 16 serum samples: 16 min → 4 min.





"Spin dust"

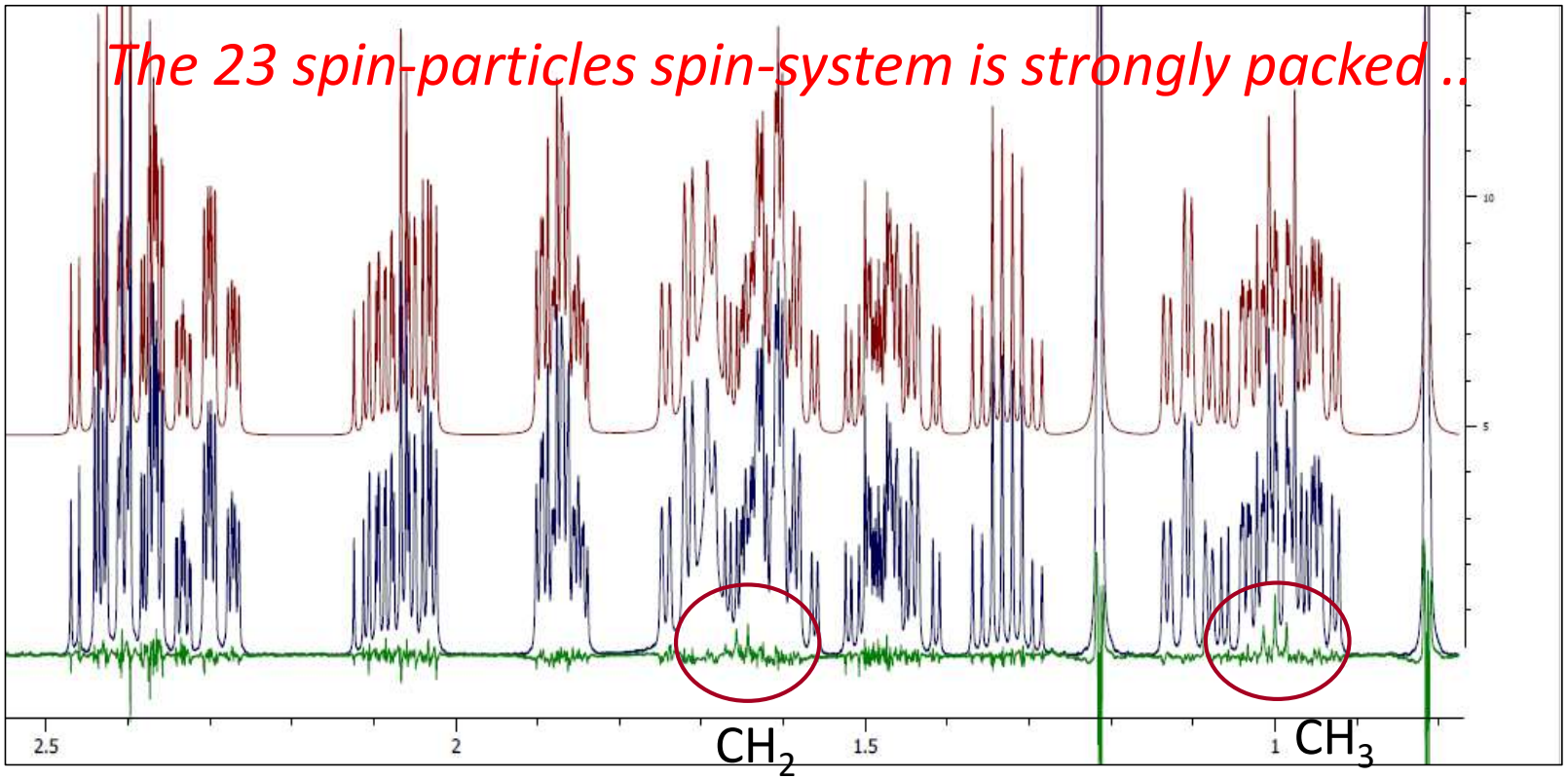
H_3-signal is composed of thousands of non-degenerate transitions – which yield its diagnostic outlook.

The effect is not rare, for example leucine methyl signal has a similar shape.

How to describe the shape in another way than QMSA, we ask?!

QMSA of Testosterone + ca. 0.40 wt% CH_3CH_2R-impurity

- fitted by ignoring the impurity:



..but still QMSA reveals impurity signals which are hiding in the peak jungle!

RESPONSE FACTORS

RF = 1.000 if the intensity of NMR signal area/proton = that of the reference signal

RF's of α-glucose in different experiments (*JMR 242(2014)67)*

	qH[a] D_2O	H[b] D_2O	qpresat[c] D_2O	presat[d] D_2O	qpresat[c] H_2O+D_2O	presat[d] H_2O+D_2O
H1	**0.962**	**0.875**	0.960	0.880	0.950	0.924
H2	**0.974**	0.993	0.965	0.993	0.904	0.909
H3	**1.000**	0.910	**1.000**	0.920	0.969	**1.000**
H4	**0.978**	0.953	0.990	0.990	**1.000**	0.978
H5	**0.965**	0.997	0.975	**1.000**	0.850	0.885
H6A	**0.977**	0.997	0.953	0.994	0.884	0.868
H6B	**0.975**	**1.000**	0.955	0.981	**0.811**	**0.840**

[a] **Basic proton spectrum** (zg): 128k data points (td), 4 dummy scans (ds), 8 transients (ns), AQ=7.7s, **RD(d1)=52.3s** and 90° pulse.

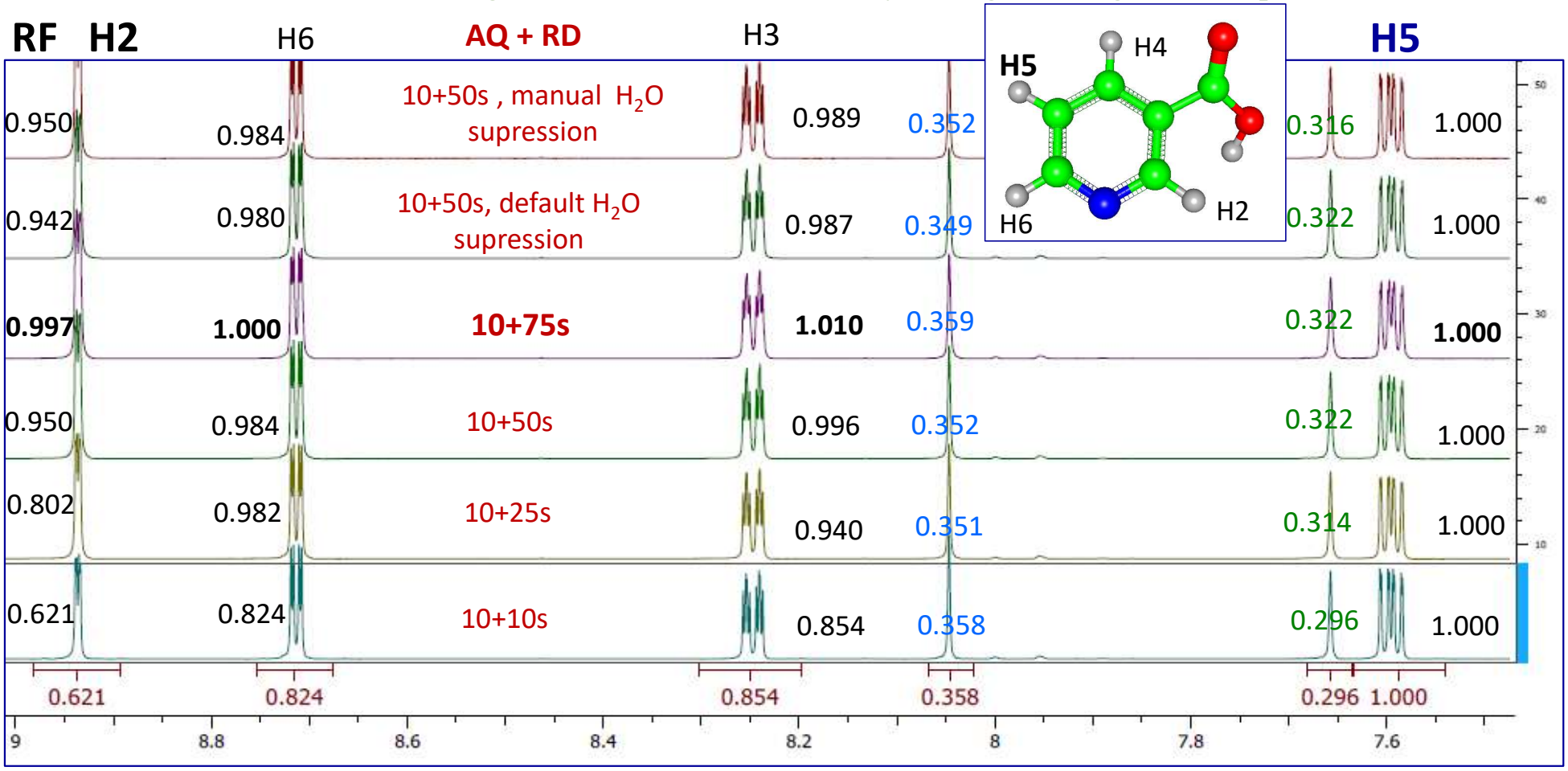
[b] Basic proton spectrum (zg): td=128k, ds=4, ns=32, AQ=7.7s, d1=2.3s and 90° pulse.

[c] Noesypresat pulse sequence (noesygppr1d): 10 ms mixing time, td=128k, ds=4, ns=8, AQ=7.7s, d1=3.0s, additional delay before suppression(d2)=49.3s and 90° pulse.

[d] As in c, but d2=0.

Nicotinic acid in ***B-FORCE tablets*** – the **macromolecular** (CMC) effect to RF!

RF's were obtained with integration tool of ChemAdder, by setting the integral of H5-proton to 1.000



DRUG PREPARATION ANALYSIS

Drug preparation analysis

ADVANTAGES:

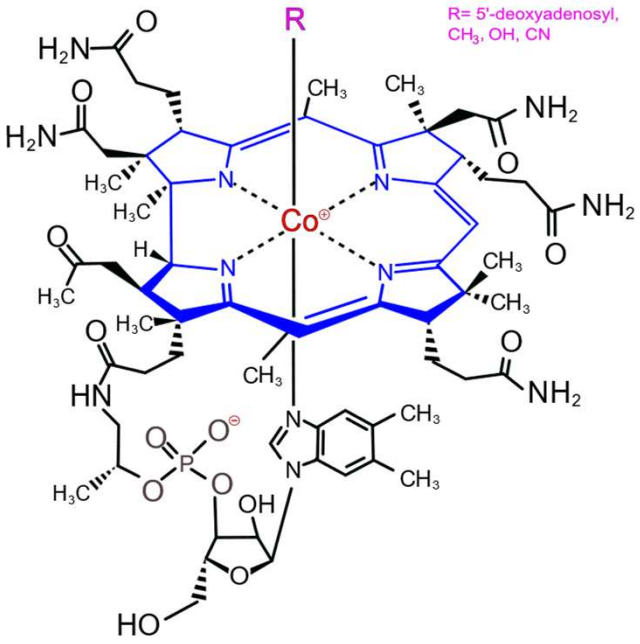
- One sample, all the components, including excipients, impurities and compounds not mentioned in label
- Amount of sample not a problem - spectroscopic cost reasonable
- Modest number of components – in contrast to biofluids

CHALLENGES

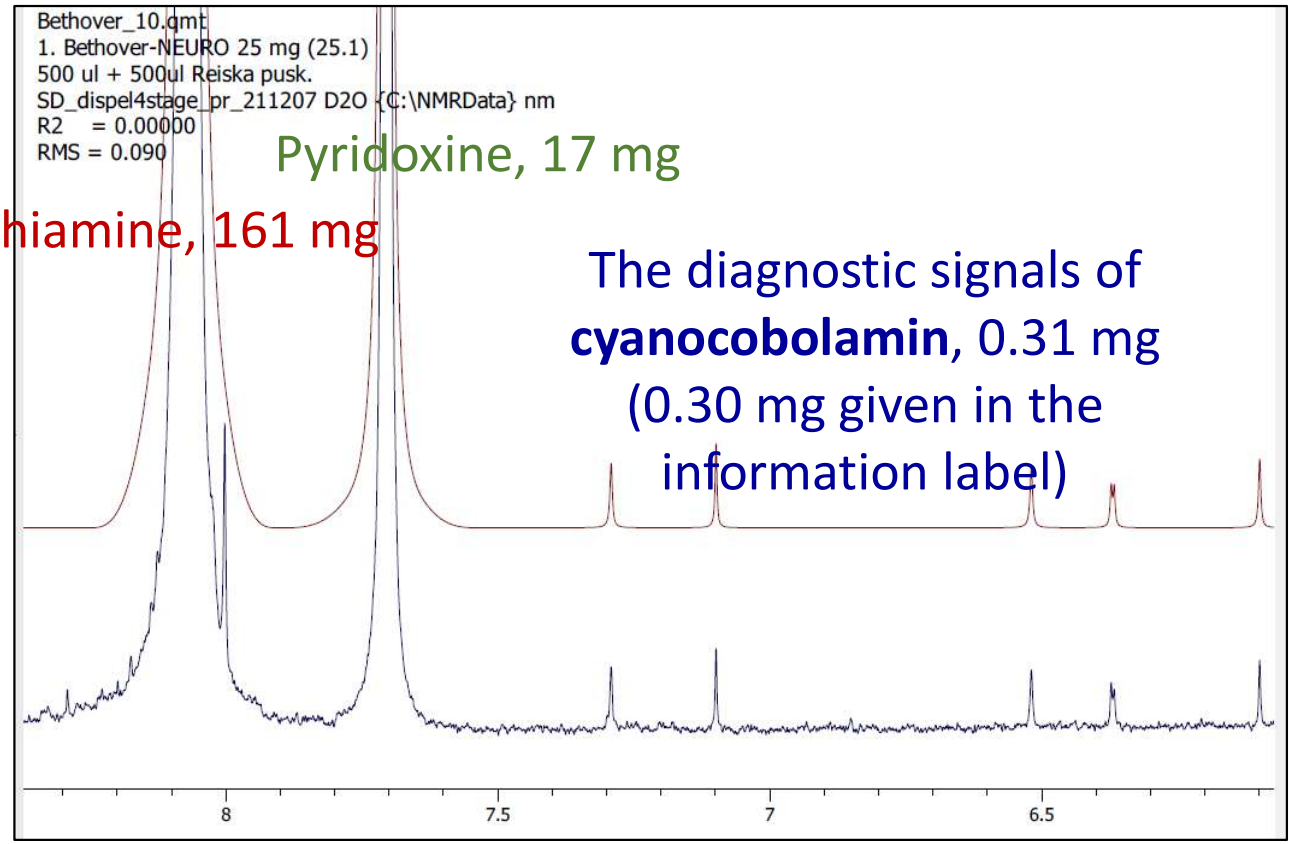
- Sample extraction: gels, microfilters may contain glycerol, etc.
- Excipients, like CMC, bind components – for example, no riboflavin was detected in extracts of a preparate containing 35 mg of it/ tablet (of 500 mg) !
- The dynamic range: in B_{12}-vitamin preparate the content of B_{12}-vitamin can be 10 - 1000 ug/tablet, while other B-vitamins vary from tens to hundreds mg.
- Response factors.

The solution to the RF-problem: (simultaneous) QMSA of the basic and spiked spectra …helps also to identify singlets in biofluid spectra

Analysis of B_{12}-vitamin in *Bethover NEURO* capsules for B-vitamin supplementary



Cyanocobolamin (B_{12}vitamin):
MW =1355,
52 chemical shifts



QMSA at BENCHTOP(40-80 MHz)

QMSA at BENCHTOP(40-80 MHz)

Solve parameters at high field – apply at low field

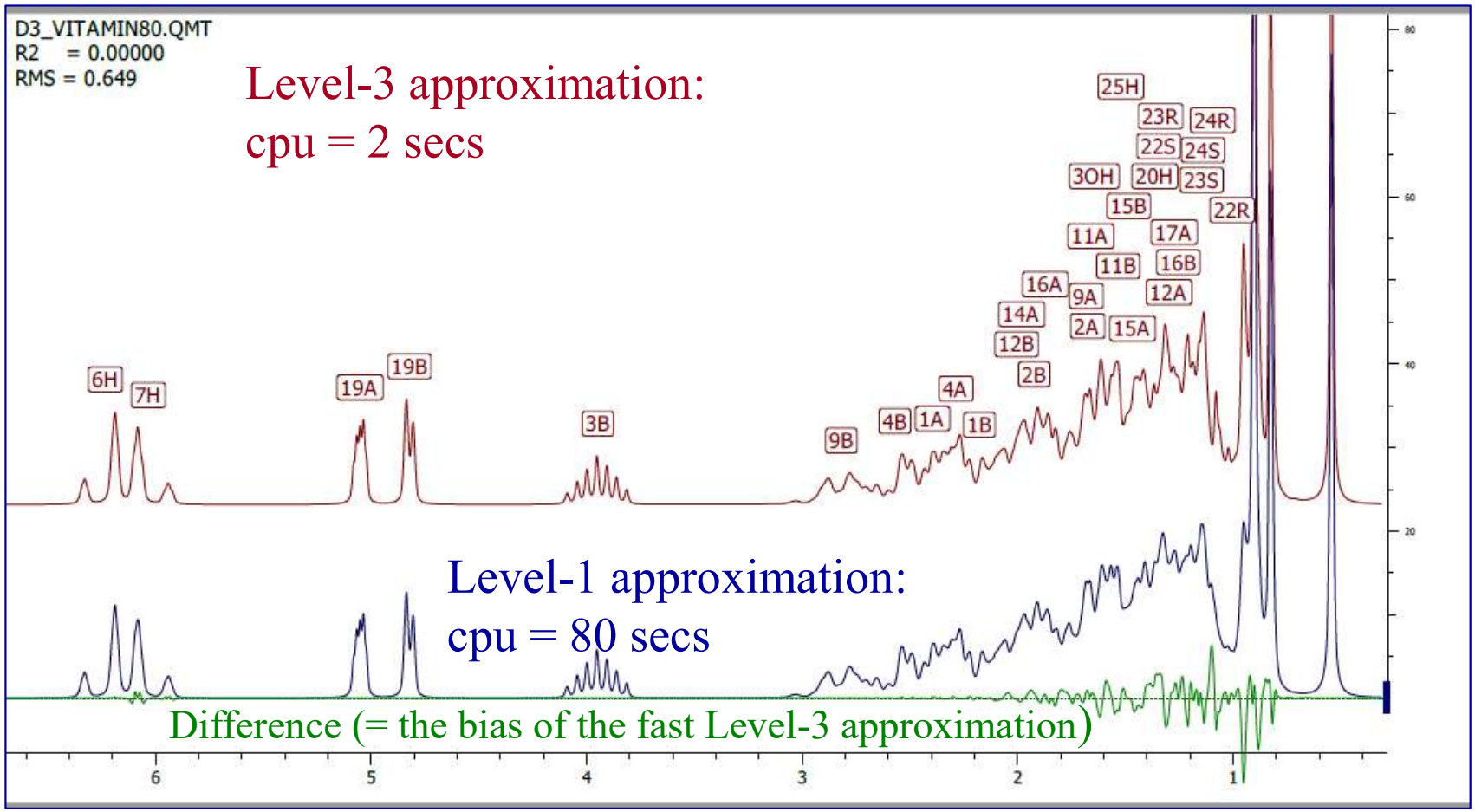
PROS:

- Chemical shifts variations less significant – recalculation of spectrum is not needed so often – if at all – during QMSA
- Costs ??

CONS:

- Sensitivity – higher concentrations – longer measurement times – concentration effects to shifts
- Simulation times for large spin-systems
- Overlap of signals
- Limited number of analyzable components – simple mixtures of simple compounds

D_3-vitamin 80 MHz simulation quality and time depend on *level of approximation* (1-3)

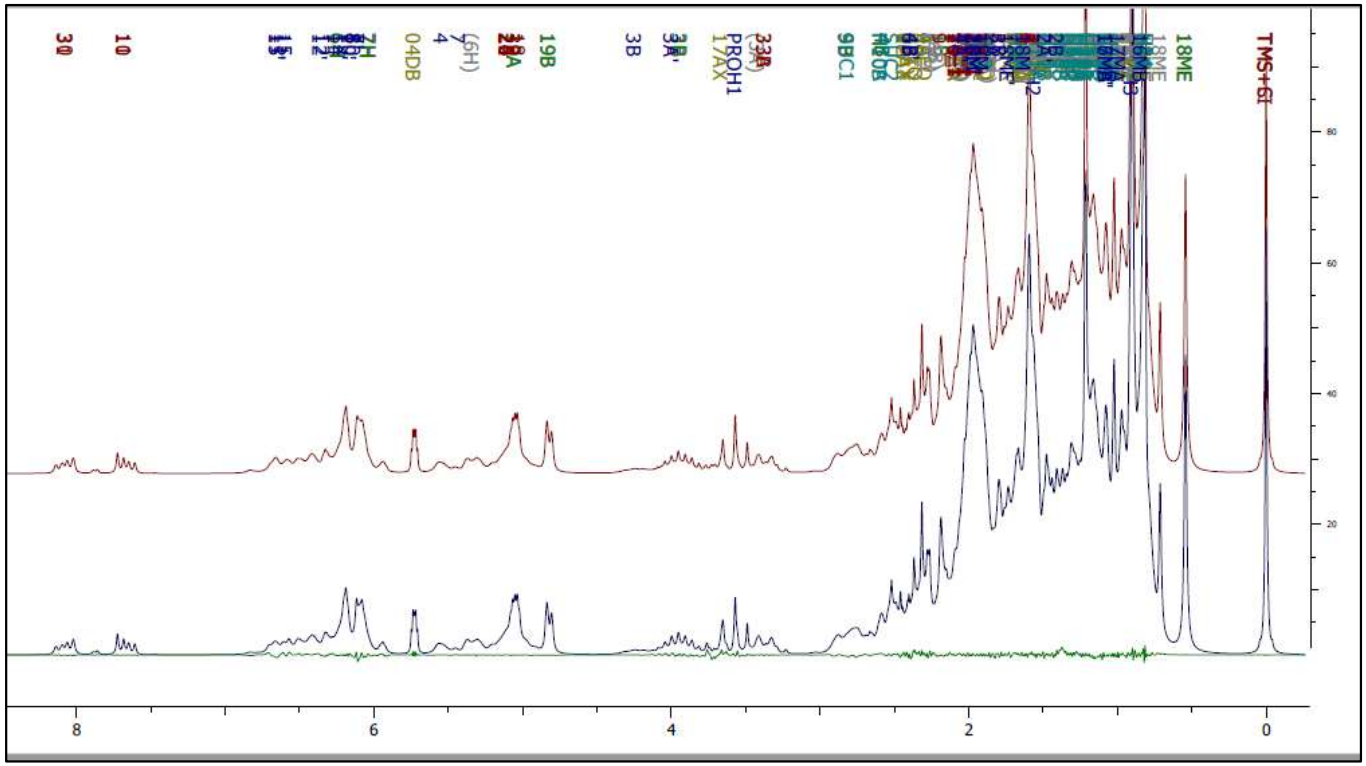


ANALYSIS of MIXTURE of D3-VITAMIN, CALCIFEROL, TESTOSTERONE, CHOLESTEROL, A-TOCOPHEROL, MENAQUINONE-7, LUTEIN, DOCOSAHEXAENOICACIDETHYL ESTER (10 mg each) + 1 mg of PROPANOL, SYNTHETIC SPECTRUM at 80 MHZ

The spectrum was simulated at approximation level 2 (5 min), then analyzed with highest approximation level 3 (< 5 sec/cycle).

When the 250 chemical shifts & line-widths were optimized (to compensate the approximation bias), RMSE dropped from 0.42 to 0.14% and gave the concentrations within bias < 3%.

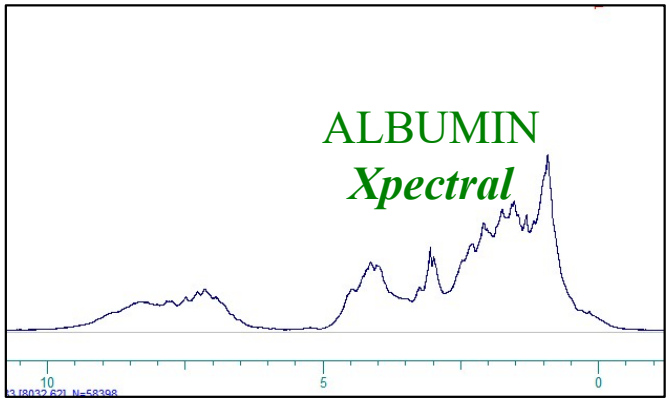
Conclusion: benchtop analyses of large molecule systems can be done in acceptable time.



Biofluids

Biofluids.. Xpectrals

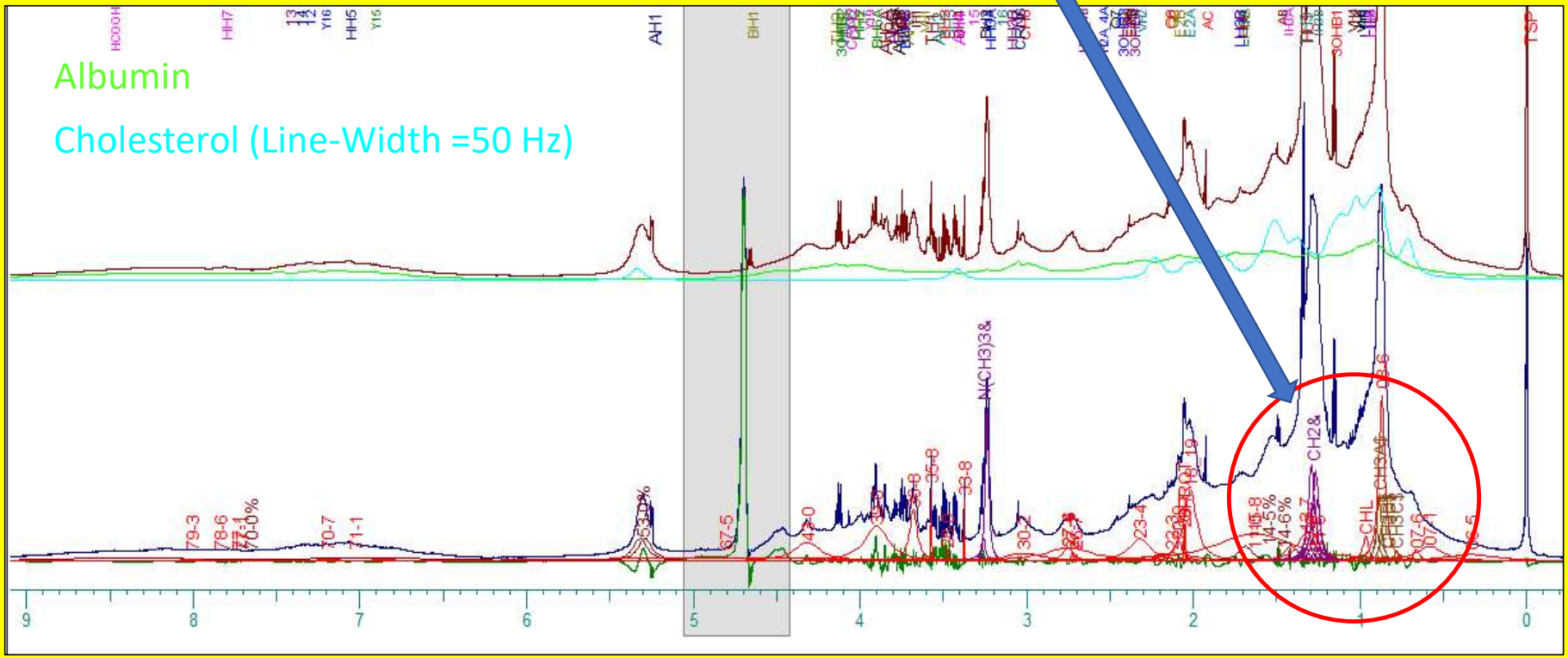
- ^{1}H NMR of serum = *QM-spectra* + *Xtructures* + *Xpectrals*
  - QM-spectra: glucose and other small molecule metabolites
  - Xtructures: lipoproteins
  - Xpectrals: albumin, cholesterol
  - ***Spiked QMSA (sQMSA)*** …to confirm assignments of singlets and to compensate RF bias!



The **spectral xtructures** can be singlets, regular or less regular (several options) multiplets. In the less regular multiplets, either line-spacings, intensities and/or line-widths are allowed to vary.

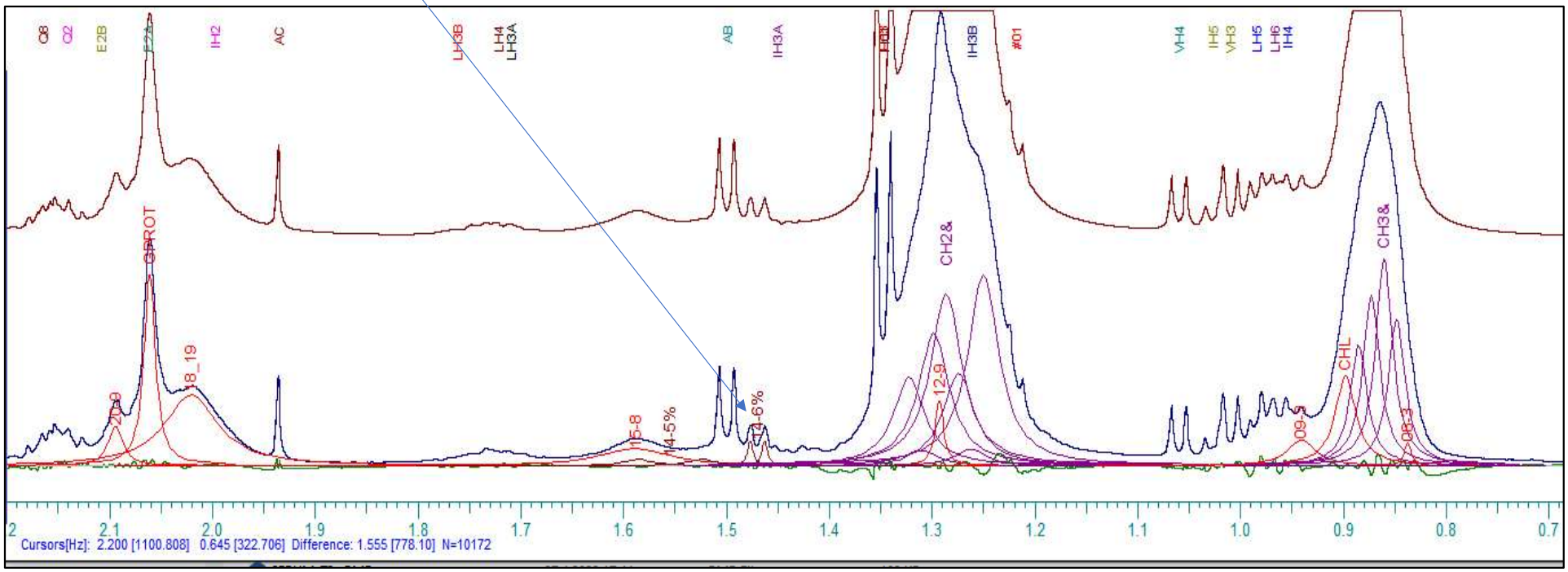
QM-spectra + Xtructures + Xpectrals

Yields more pure lipoprotein signals ?



T2 edited serum spectrum with three types of ***xtructures***

- Singlets
- Regular doublet (X%)
- Multiplets with varying line-intensities, constant line-spacings and line-widths (lipoproteins, CH2& and CH3&)



Analysis of 1001 spin particles: urine

1001 spin particles, 1546 nuclei (^{1}H,^{31}P,^{14}N)

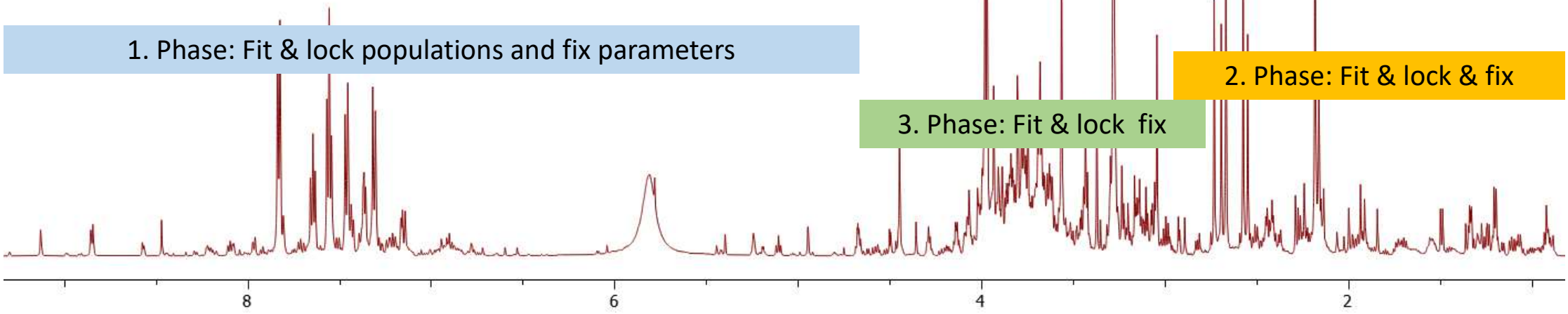
214 compounds with 270 spin systems

Simulation time < 2 s, << 2 s with multithreading

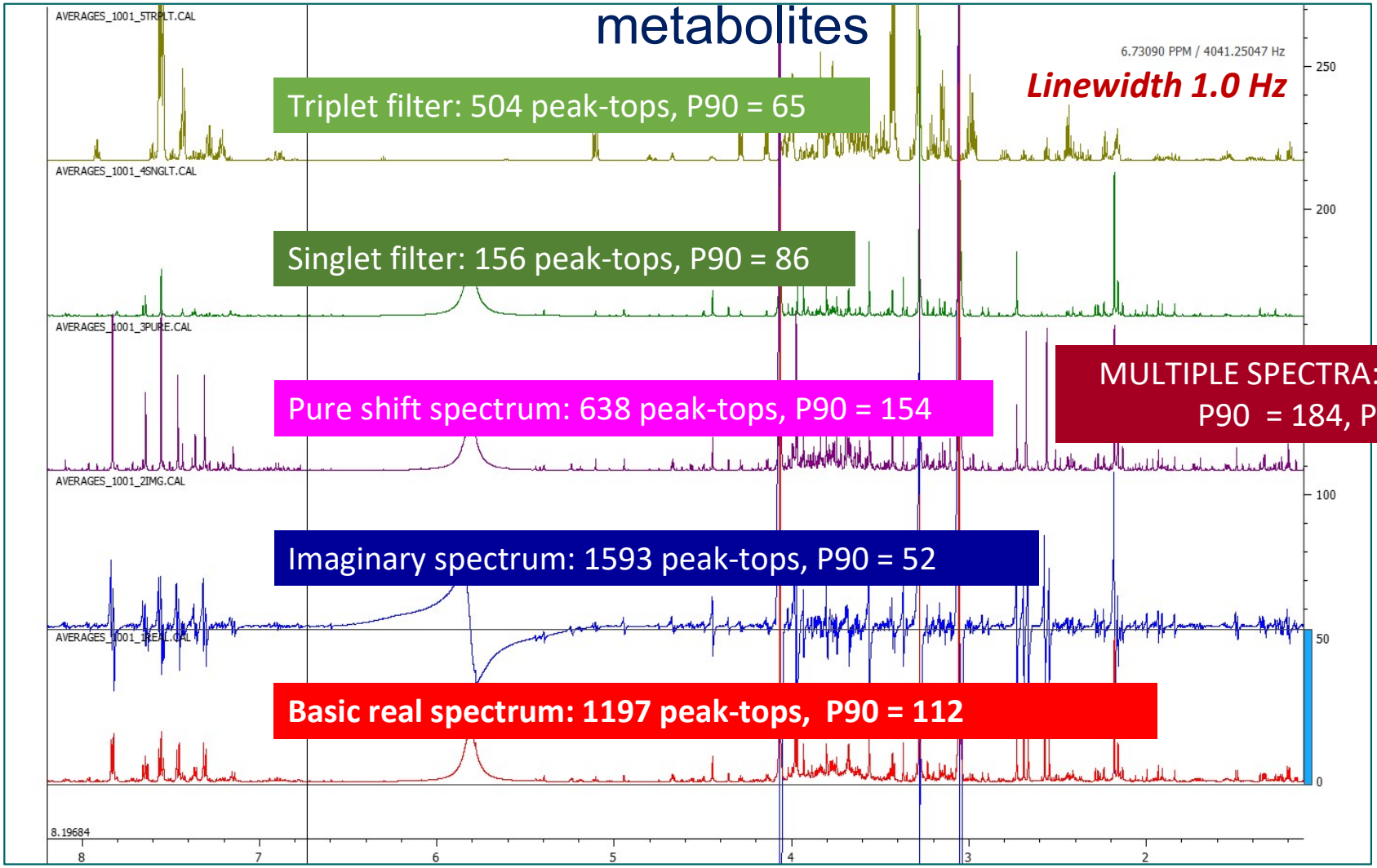
1. Phase: Fit & lock populations and fix parameters

2. Phase: Fit & lock & fix

3. Phase: Fit & lock fix



URINE 1001 particles, multiple spectra QMSA (mQMSA) of 214 metabolites



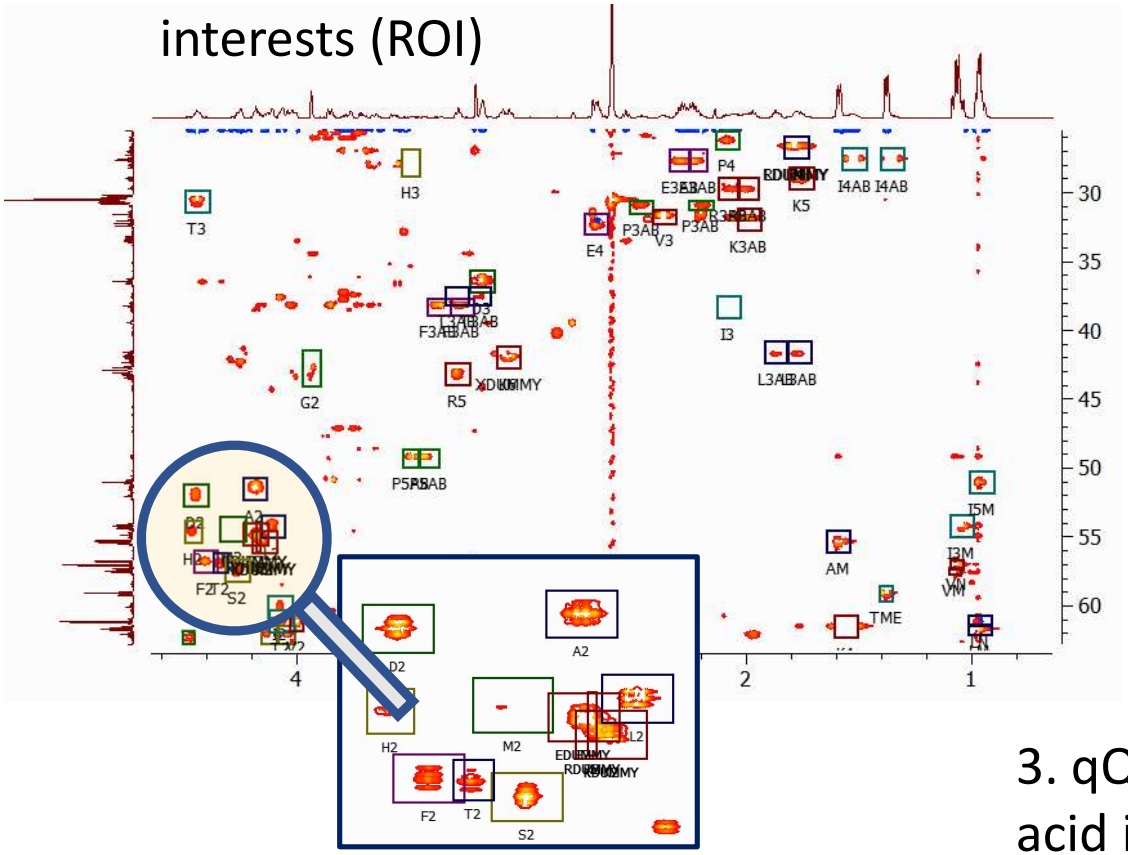
P90 = No. of compounds having at least one **DIAGNOSTIC** 90% purity signal

QMSA of 2D spectra

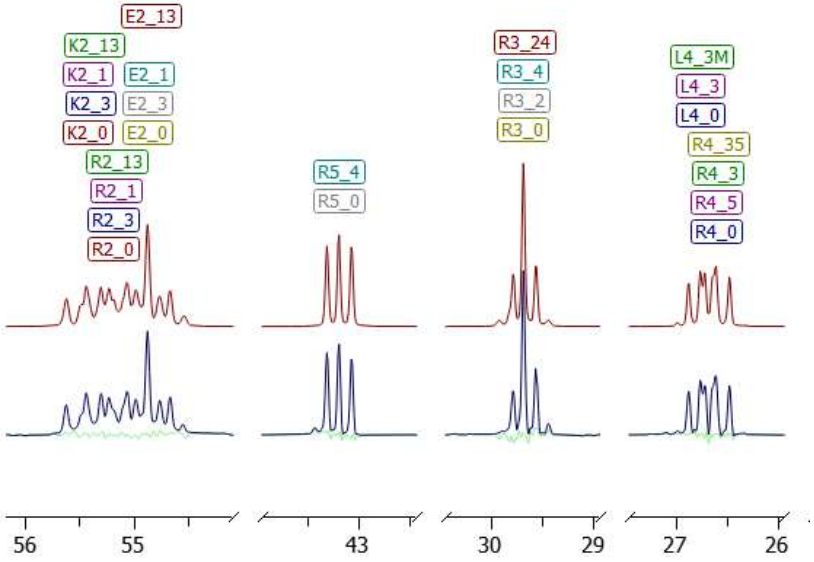
HSQC of amino acid ^{13}C isotopomers 2D spectrum to VIRTUAL 1D spectra: metabolic flux analysis 1/2

Collaboration with Technical Research Centre of Finland (VTT)

1. Extraction of regions of interests (ROI)



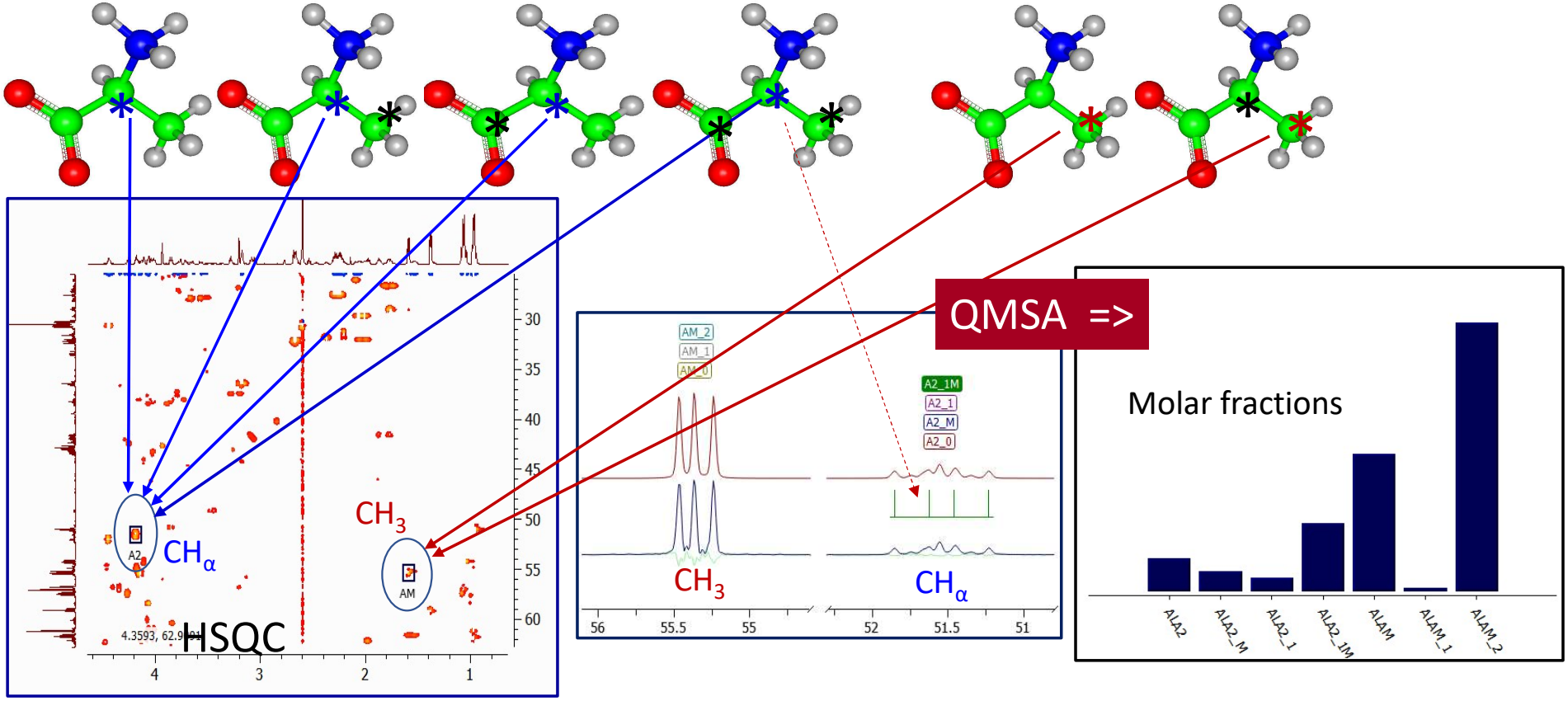
2. 2D → 1D F2 Projections



3. qQMSA → molar fractions of amino acid isotopomers

HSQC of amino acid ^{13}C isotopomers 2D spectrum to VIRTUAL 1D spectra: metabolic flux analysis 2/2

Alanine ^{13}C isotopomers:



QMSA - pros

- Complete QMSA in a few minutes!
- Overlapping signals and variation of shifts – a challenge for integration protocols.
- Complex or second-order spectral structures – a challenge for deconvolution protocols.
- From spectral storage to qNMR and special applications.
- ASL's: one spectrum – one file – any field & line-shape & shifts – even from poor spectra and mixtures - no experimental artefacts – compression factor of > 90% - prior knowledge
- Chemical confidence – not only concentrations – also unknown compounds can be characterized.
- ^{13}C Satellites can be defined in ASL files – like creatinine in urine.
- Accurate peak-lists – pattern search, etc..
- Integral transforms – iteration of poor trial parameters – fast screening for maximum amount of a compound.
- Achieving and export of NMR data to journals and their supplementary... instead of raw spectra – an opportunity!

Maximum amount of information with minimum number of parameters!

qQMSA + CTLS

A spectrum data may contain different type NMR signals, needing different models:

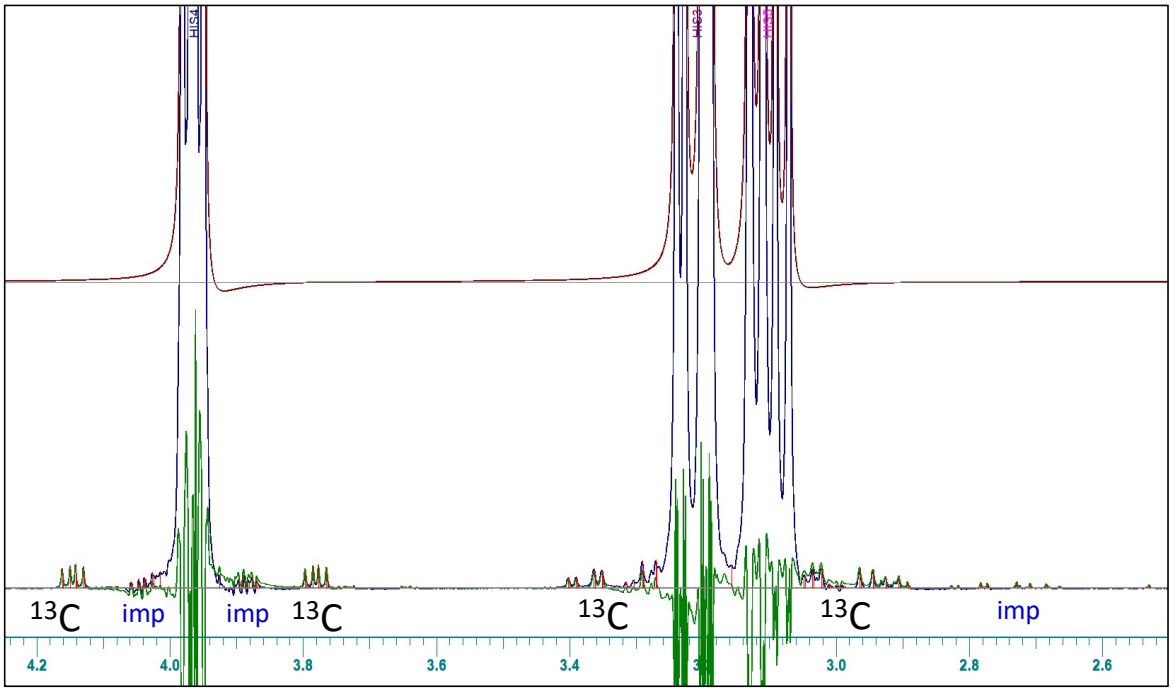
- Quantum Mechanically modellable signals
- Xtructures (singlets, multiplets), like polymer and lipoprotein signals
- Xpectrals, like albumin spectrum
- *Xpurities, see the following page*
- Signals defined by Integral ranges
- The common point is that the signal area/nucleus is the same:

  Total area = QM + Xtructures + Xpectrals + Xpurities + Integrals

All the species can be handled in one model by ChemAdder !

Xpurities – new 2024

- A spectrum may contain weak well-defined signals like **^{13}C satellites** or **signals (peaks) arising from unknown impurities**.
- In QMSA, the weak signals are added to the baseline and, thus, lead to a positive bias - too high a purity !
- The **xpurity** peaks can now be found automatically to the model and subtracted from the major spectrum.
- The peak areas give (rough) estimates of the concentrations, to be removed from the total area and thus concentration – which are thus improved.



Acknowledgements

Pekka Laatikainen, Spin Discoveries Ltd.
Henri Martonen, Spin Discoveries Ltd.
Tuulia Tynkkynen, UEF
Jani Rahkila, Åbo Academi (presently Bruker BioSpin)
Hannu Maaheimo, VTT
Ben Shapiro, US Pharmacopeia, Rockville, MD, USA
Sunil Paudel, US Pharmacopeia, Rockville, MD, USA
Yang Liu, US Pharmacopeia, Rockville, MD, USA