EUROMAR 2025,
Oulu

The NMR-Purity and the new dimensions and standards of purity

Reino Laatikainen, Tuulia Tynkkynen[1], Pekka Laatikainen and Henri Martonen
Spin Discoveries Ltd., Kuopio, Finland
[1]Dept. of Pharmacy, University of Eastern Finland, Kuopio

Latest update Sept. 10th, 2025

CA

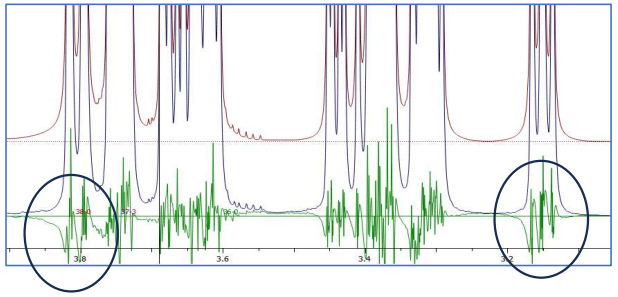
Integration vs. Total-Line-Shape Methods

- **Integration is sensitive to phase and baseline.** When the spectrum presents isolated signals, with **integration** is possible to achieve 95% confidence intervals as low as 1.5% of the compound concentrations, and that the **integration** of well-isolated peaks can lead to relative uncertainties of 11% when there are even slight phase and baseline errors[1]. ***This makes the integrals manipulable through phase, baseline and integration range, which are subjective parameters, but fortunately, easily from the original FID – in opposite to weighting and reference quality bias.***
- The most serious of problem of **integration** is the peak overlap. To be precise, the signals to be integrated need to be sufficiently isolated. Lorentzian signals decay slowly to infinity: for a maximum error of 1%, **integration** limits of 25 times the line width (± 25 Hz or 0.05 ppm with linewidth of 1 Hz at 500 MHz) are needed. For errors < 0.1%, the limits should be ±76 Hz (± 0.15 ppm)[2].
- The risk that an impurity signal is hiding under the integrated signal, increasing thus the integral – and purity – is serious with **integration**. The essential impurities can be expected to resemble to the target compound. The same problem but more serious, is with the MS and chromatographic methods.

1. Malz, F. and Jancke, H., *J.Pharm.Biomed.Analysis*, 2005, 38, 813–823. https://doi.org/10.1016/j.jpba.2005.01.043
2. Griffiths, L., *The Analyst*, 1998, 123, 1061–1068. https://doi.org/10.1039/a800625c

qQMSA sounds as an ideal tool for pNMR. However, to be perfect, the line-shape model should be perfect!

- The line width and line shape depend on compound, spectrometer and weather.
- In **ChemAdder**, the line shape can be described using **asymmetrical Lorenzian**, **Gaussian** and **Dispersion** functions, adding **out-of-coil corrections**, **virtual long-range couplings** and **isotope shifts** for ^{13}C and S, Cl.
- Despite the multi-term models, the fitting usually leads to typical **observed-calculated difference spectrum.**



The '*mysterious*' difference arises from line-shape artefacts, which are compound- and proton-specific (see inserts, similar for the two multiplets). The artefacts cannot be totally removed with line-shape tools or the approaches are impractical.

- If the **observed-calculated area** is not ZERO, as it should be, this leads to a small bias in concentrations. Unfortunately, the area is sensitive to the line-shape model!
- The bias can be minimized using the '**pNMR Option**' *(pNMRO),* which usually also smoothens the convergence (see 5b).

NMR Purity

- ***NMR Spectral Purity***, shortly ***NMR Purity*** or *Spectral Purity,* has dual meaning*: firstly, it means 'purity of the NMR spectrum',* secondly, *'purity of sample based on NMR spectral purity'!*
- *Spectral purity* can be based on any spectroscopy, but NMR-Purity emphasis the superiority of the NMR method in purity characterization.
- qQMSA+CTLS offers four quality parameters for the sample quality:
  - NMR-Purity
  - eR2 = essential R2: measures goodness of observed-calculated fit)
  - eRMS = essential RMS: measures spectral quality and analysis)
  - Parameter MATCH: measures fit of spectral parameters to the default values of chemical shifts and coupling constants. The conditions (solvent, concentration, pH, temperature, ..) become critical with use of this parameter.

The NMR-purity parameters offer superior measures for chemical quality control.

NaCl analogy

- In a NaCl-sample contains 0.001% of KCl, there is no method to determine the NaCl concentration so that one could say that its purity is 99.999%, but it is easy to show that the KCl concentrations is ca. 0.001%. For most of applications it is the same if NaCl purity is 99 or 100%, if the rest of weight comes from water, KCl or $MgCl_2$. For rare or dangerous impurities their concentration forms the issue. There are methods which allow (semi)quantitative profiling of the inorganic elements with one shot.
- Also, for organic compounds 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 to measure the **impurities**. There exists one method in which the analysis can be done by a single measurement and without calibration:

qNMR !

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, for example, solvent suppression, T2 editing or DISPEL pulse! The key parameter in this context is RESPONSE FACTOR (RF)!**

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.

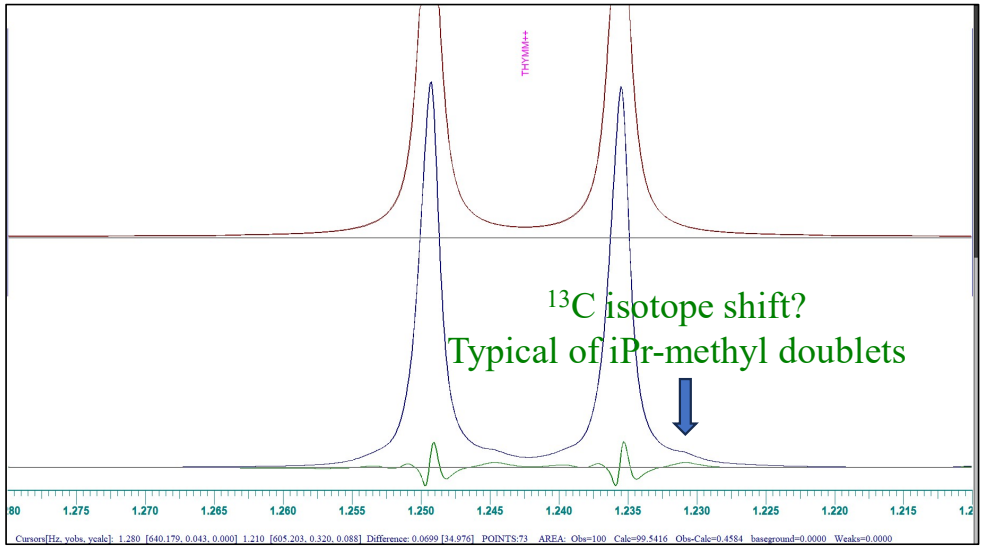
To show that the purity of a sample is (say) >99.8 %, demands ultrapure reference compounds, very accurate weighting of samples (enough sample and reference, and even *buoyancy* should be considered) - and a good integration protocol*!*

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

The novel **pNMR Options** of ChemAdder nearly remove the bias arising from line-shape (next page)

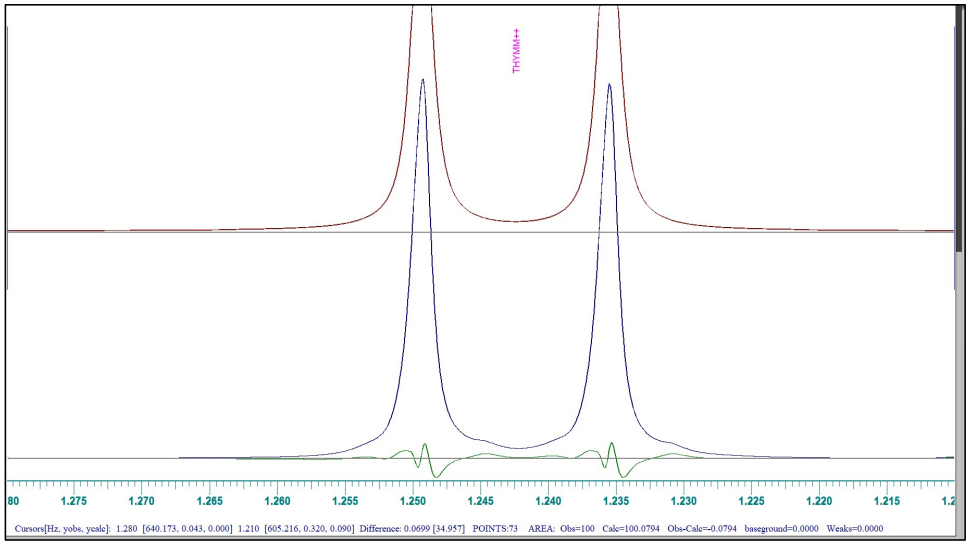
IT Supported QMSA and pNMR Options for pNMR: the bias is reduced by 83%

iPr-methyl doublet fit without the ‘pNMR Options’:



The observed-calculated area = 0.46% of the calculated area
Purity% = 97.6%

iPr-methyl doublet with the ‘pNMR Options’:



The observed-calculated area = 0.08% of the calculated area
Purity% = 99.8%

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 1-0.5%, and very accurate weighting (thus, in sufficient amounts) and very pure references for calibration are usually needed.
- 2. **Determine the impurities.** For impurities, 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%. Accurate weighting of small amounts and calibrations are not needed. The impurity concentrations can be determined with accuracy that is far better than demanded !

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.

QMSA is **The Tool** for impurity analysis – with minimal calibration and preparations! Also, **the relaxation delays can be shortened to < 10s** because the relaxation effects can be treated through RESPONSE FACTORS.

Essential is to know that there are not too much certain impurities !

ImPurity analysis

From the user's view-point it is usually the same if the material contains 99 or 100% of the key compound, if the impurities are known to be safe, like solvent or inert salt!

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

Essential is to know 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! Also, **the relaxation delays can be shortened to 5 - 10s** because the relaxation effects can be treated by optimizing the RESPONSE FACTORS.

While the result of integration depends on subjective decision (integration width, baseline correction, phasing, ..), an impurity signal cannot be easily removed by manipulating original FID.

The novel **pNMR Options** of ChemAdder nearly remove the bias arising from line-shape (next page)

The *'principle of 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 the 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!!
- NMR-impurity does give the impurity weight%, if the molar weights corresponding to the impurity signals are not known, which is often the case. However, it is sufficient to know the type of proton (CH, CH_2 or CH_3) to get the molarity% of the corresponding compound.

Why not RMSE and R2 ?

- *Residual root mean square Error* (RMSE or RMS) and R2-value, commonly used to describe the quality of QMSA, are poor descriptors because they depend on the spectral width, which can be 30 ppm while the spectrum occupies only 1 ppm!
- The RMSE and R2 should be independent of the range and RMSE equal to noise, but usually RMSE > noise, because the line-shape model is never perfect!

Why essential RMSE and essential R2 ?

- Essential range = the spectral range occupied by NMR information, defined by the range for which the intensity of the simulated spectrum > threshold. ChemAdder threshold:
  - thres = 0.1 * maximum intensity + 0.1* intensity of the nearest calculated signal.

Essential NMR Statistics

RMSE = RMS Error

R2 = The fraction of variance explained by model

Noise = RMSE for signal free parts of spectra!

} depend on spectral width

Essential Range = percentage of regions with NMR signals (calculated intensity > threshold)

Essential RMSE = RMSE for the Essential Range,

Essential R2 = R2 for the Essential Range,

NMR-Purity = percentage of spectral area explained by QMSA

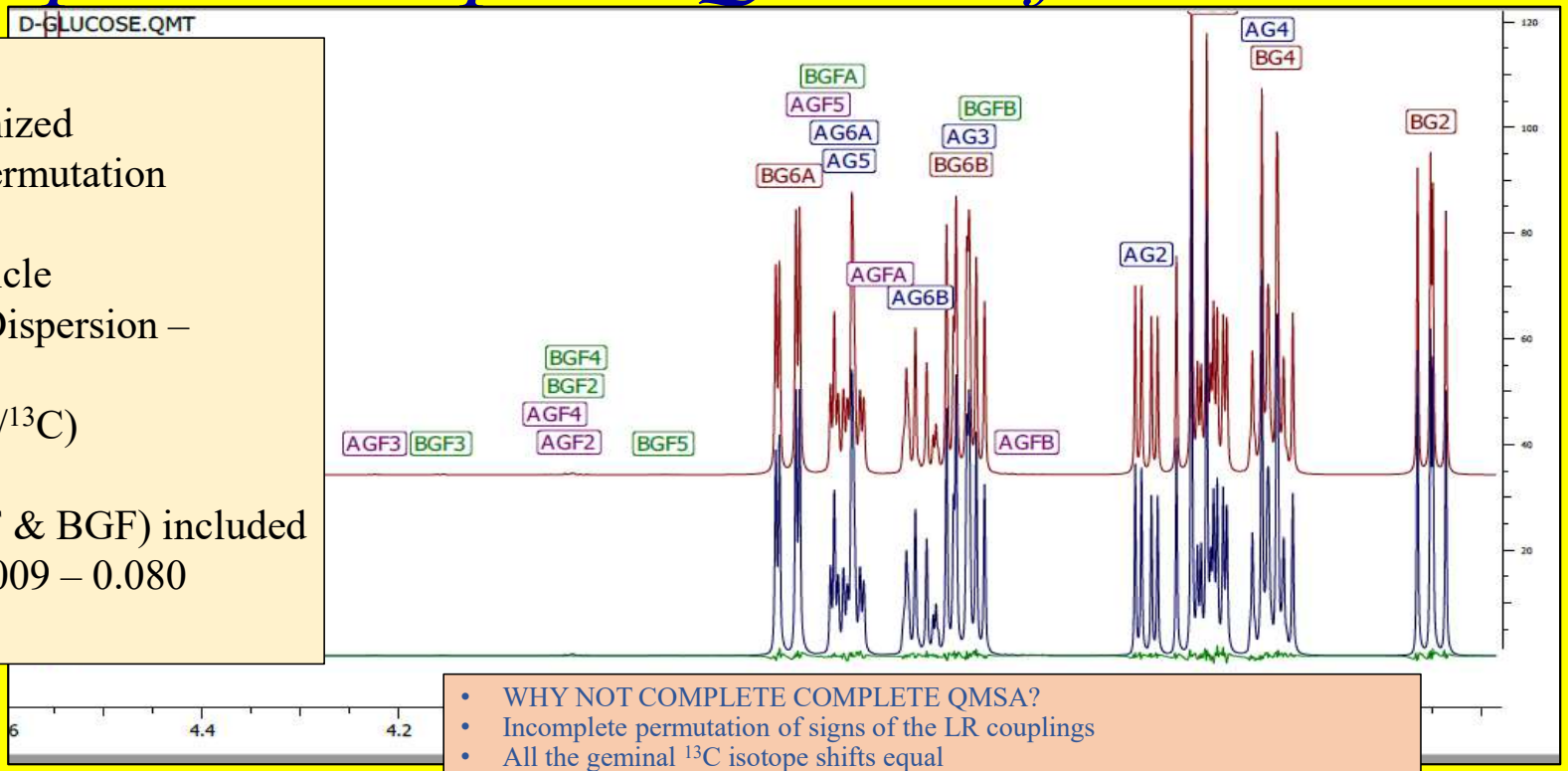
NMR-Impurity = percentage of spectral area NOT explained by QMSA

} independent of spectral width !!

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 (AGF & BGF) included
- Several other trace impurities (0.0009 – 0.080 mol%), see pages 15-17

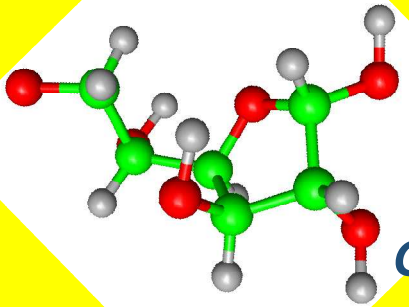


- Noise = 0.01%
- RMSE = 0.15%
- **Essential RMSE = 0.32%**
- Essential R2 = 99.96
- **NMR-purity = 99.86%**

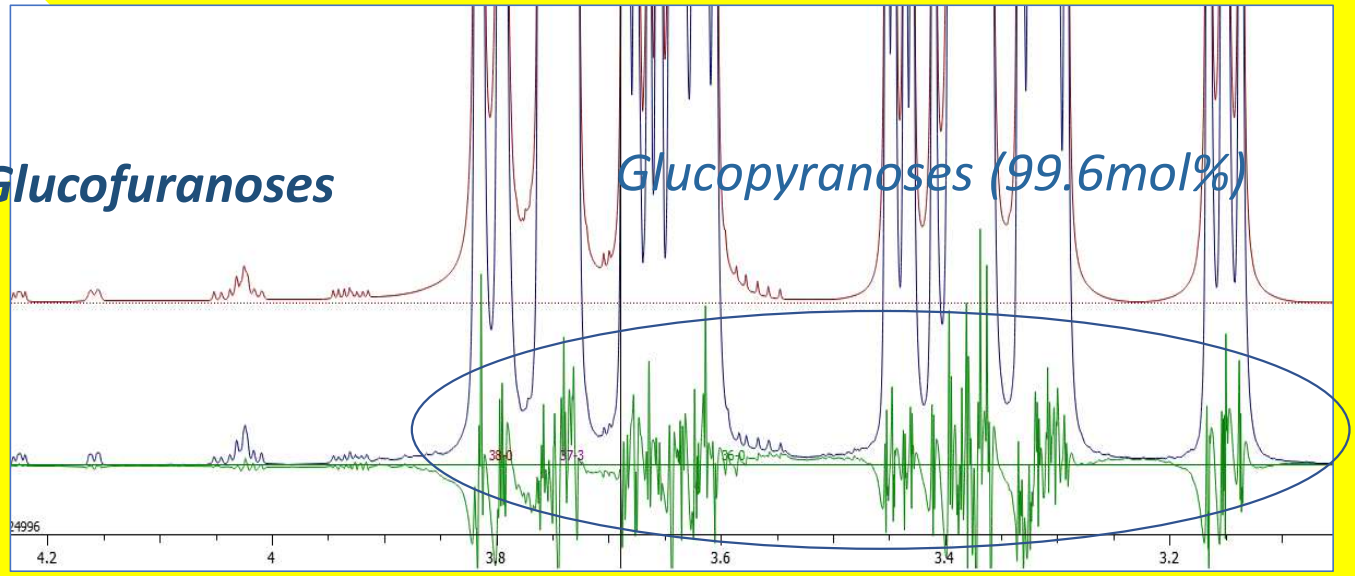
- WHY NOT COMPLETE COMPLETE QMSA?
- Incomplete permutation of signs of the LR couplings
- All the geminal ^{13}C isotope shifts equal
- No vicinal and long-range ^{13}C isotope shifts
- Other isotope effects ignored ($^{17,18}O$)
- Dipolar couplings ignored
- Minor (for example, open-chain tautomers) impuritities not treated with QMSA
- …

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

The *'Mysterious Line-Shape Problem'*



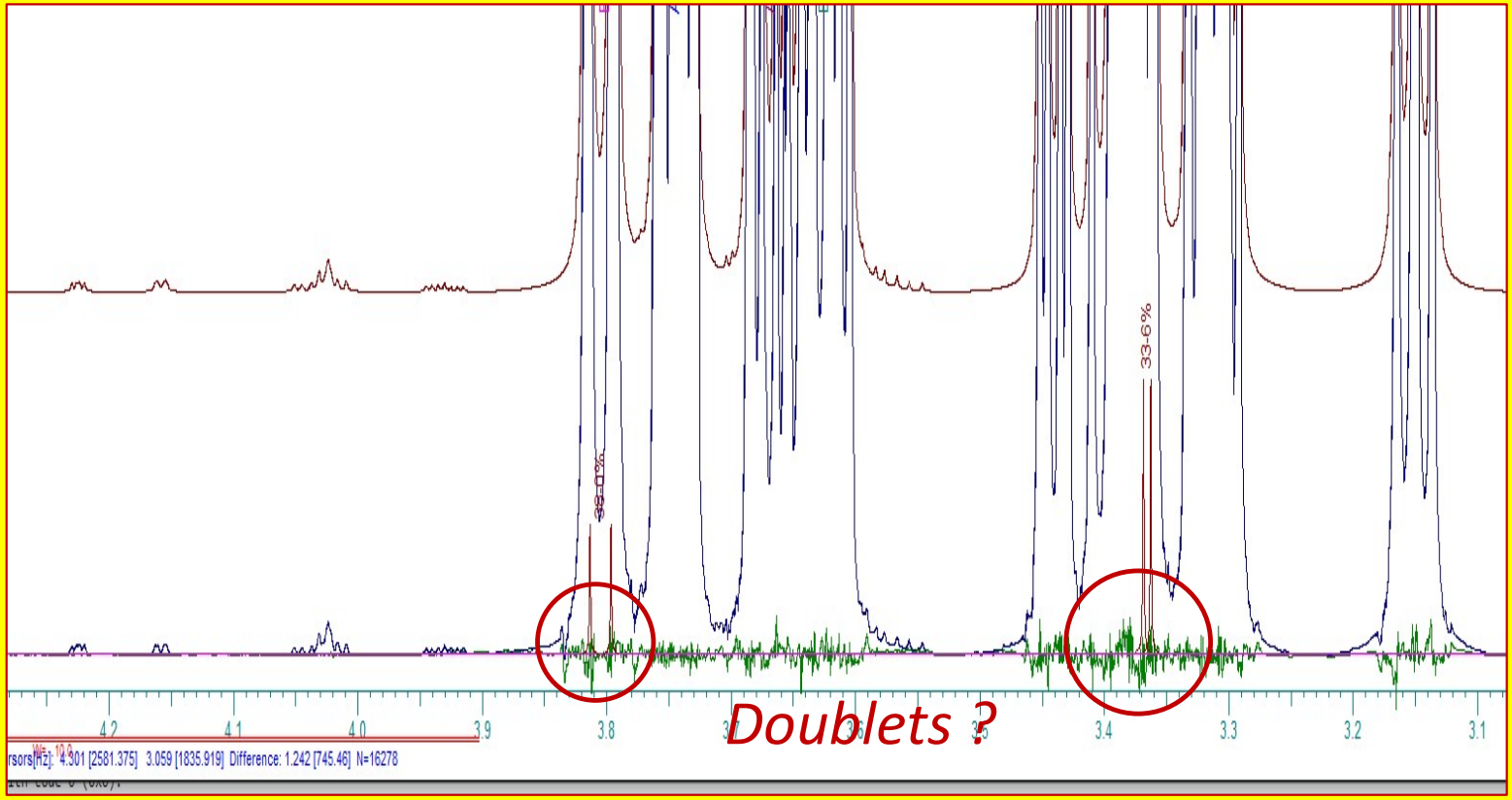
Glucose ^{1}H NMR 2-6-H region



Observed-calculated trace

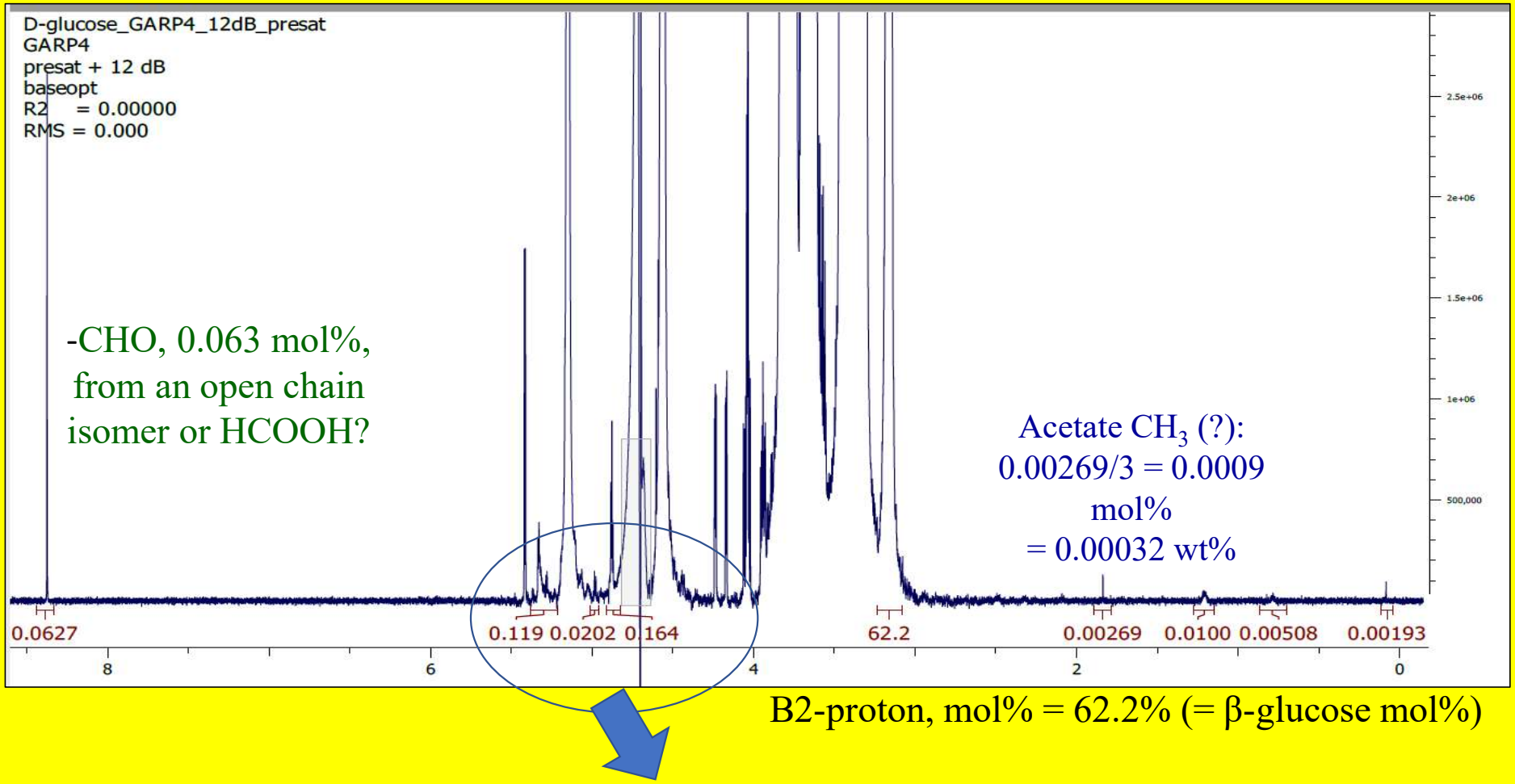
Essential rrms = 0.32%

The same after the '**Fourier Correction**', **out-of-coil correction** and when two **doublets** are added to the model

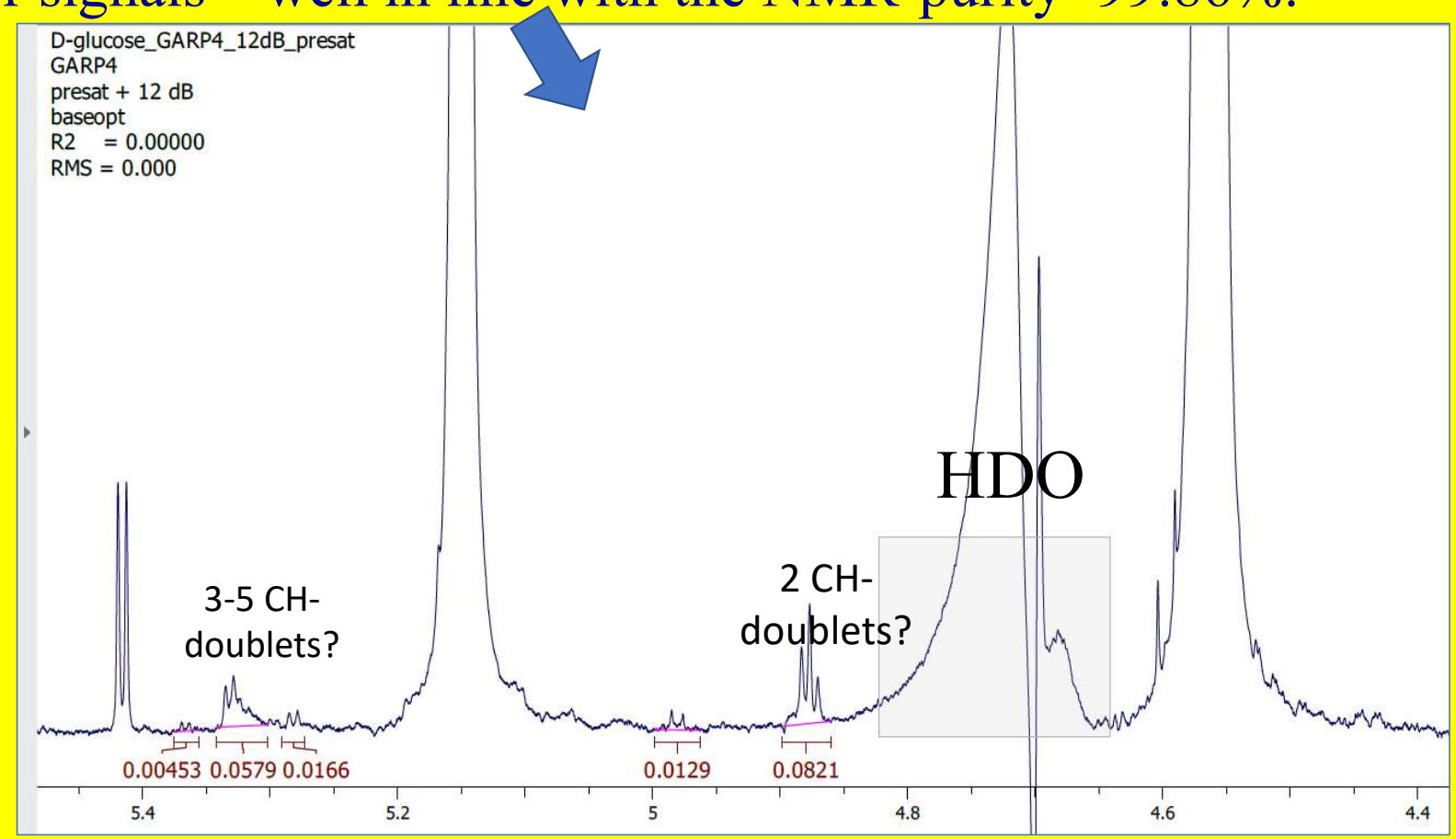


Essential rrms = 0.095%

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



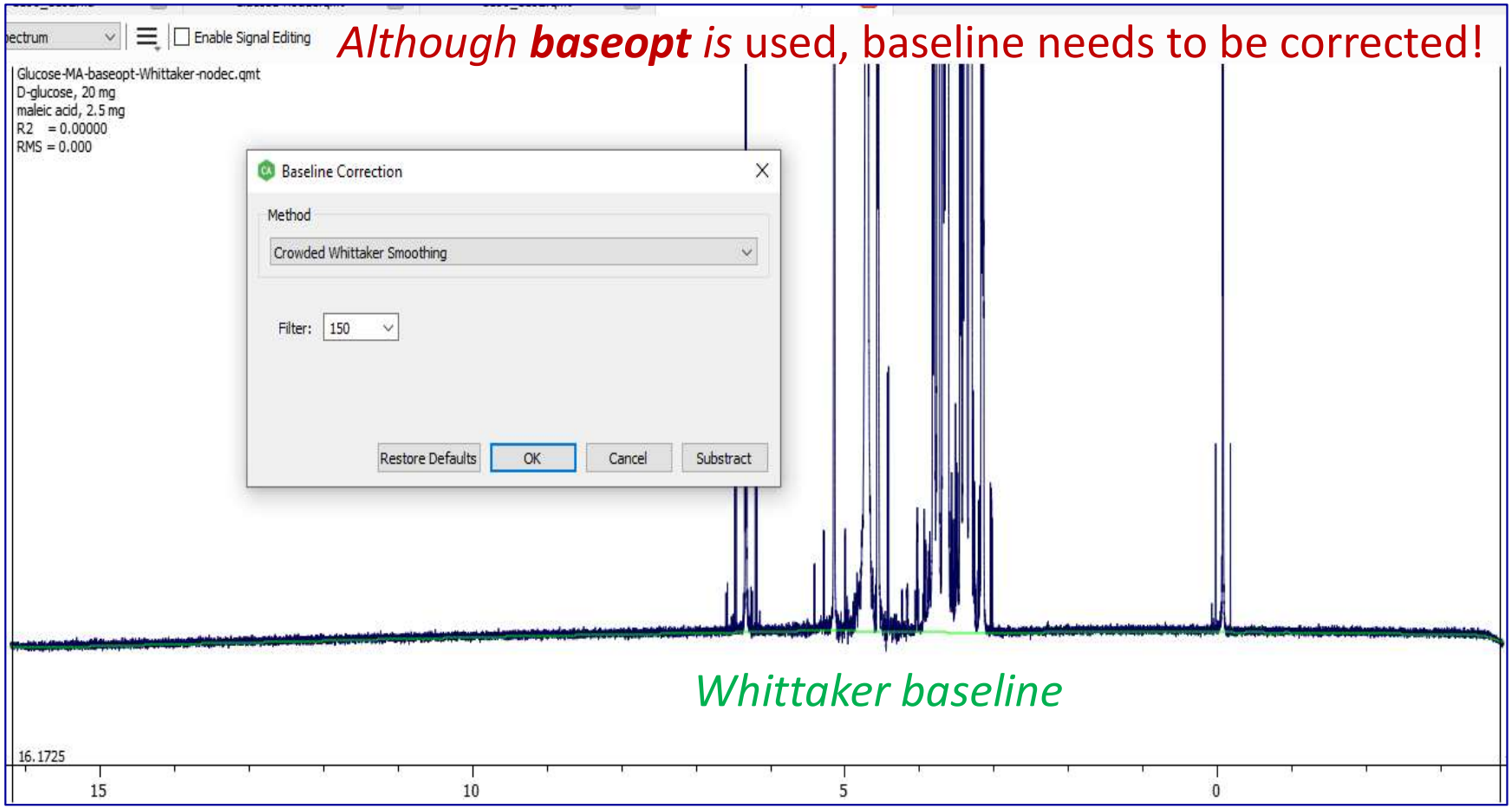
..also 5-8 glycoside CH-doublets (0.004 – 0.080 mol%) are observed, so that their total ca. 0.24 mol% - the rest of the spectra lies under the major signals – well in line with the NMR-purity=99.86%:



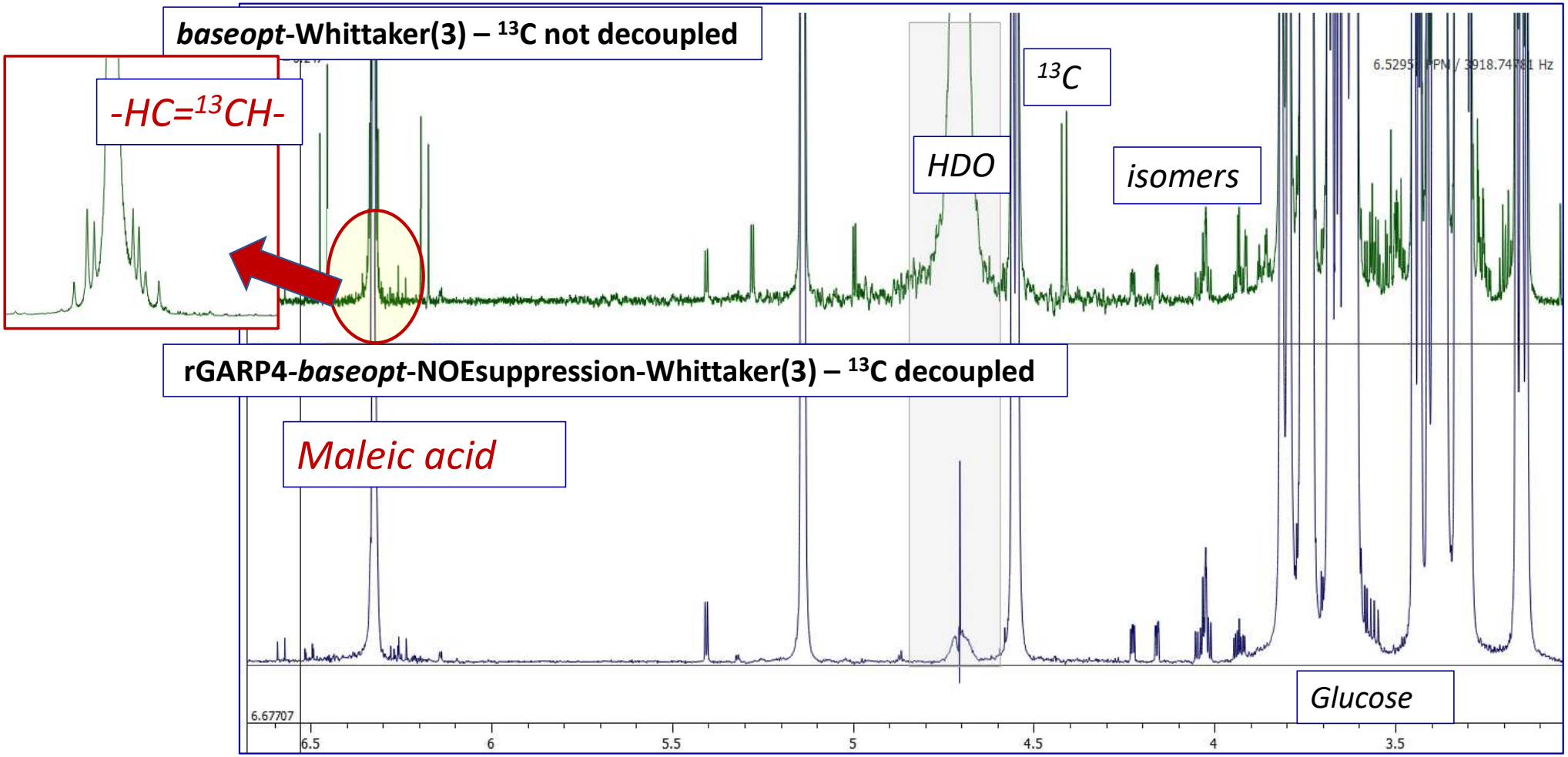
Measurement of ^{1}H NMR spectra for QMSA

- Long relaxation time (30 – 100 secs) or >5-10 s for NMR-Purity
- 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

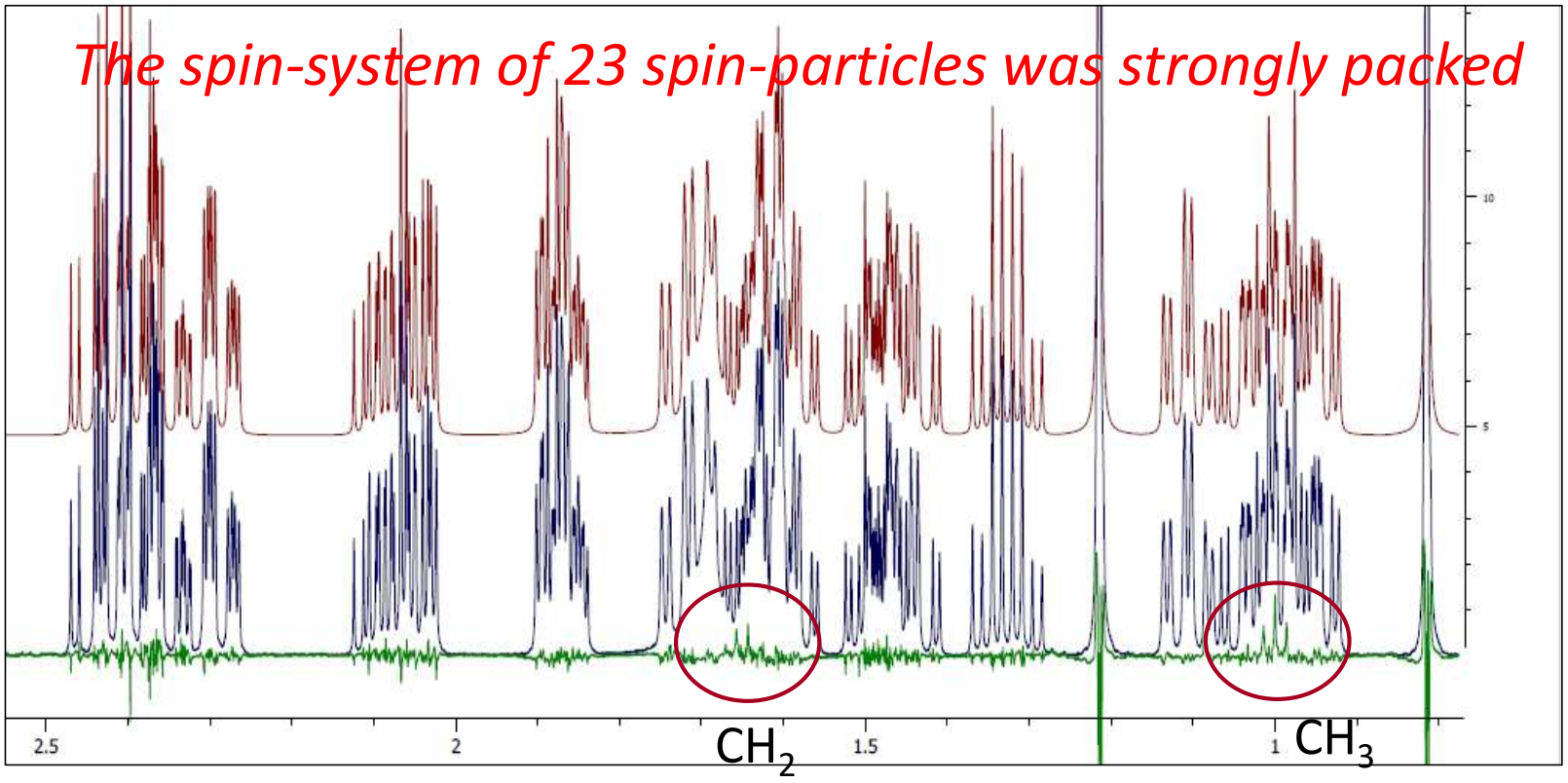


Removal of ^{13}C couplings using '*gentle GARP4 decoupling*'



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

The spectrum was fitted by ignoring the impurity from the model:



QMSA reveals the ethyl signals hiding in the peak jungle!

RESPONSE FACTORS: serious bias if ignored!

RF = 1.0 if the intensity of NMR signal area/proton = that of the reference signal
RF's of α-glucose in different experiments differ by up to 19% from the default (1.0)[$]:

	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), ds=4, ns = 8, 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.

[$] *J.Magn.Reson., 242(2014)67.*

Holistic quantitative QMSA(qQMSA) + CTLS
(CTLS = Constrained Total-Line-Shape)

A spectrum data may contain the 5 different type NMR signals :

- 1. *Quantum Mechanically modellable signals*
- 2. *Xtructures* (singlets, multiplets), like polymer and lipoprotein signals
- 3. *Xpectrals*, like albumin spectrum
- *4. Xpurities, see the following page*
- 5. *Integrals*, none of the above

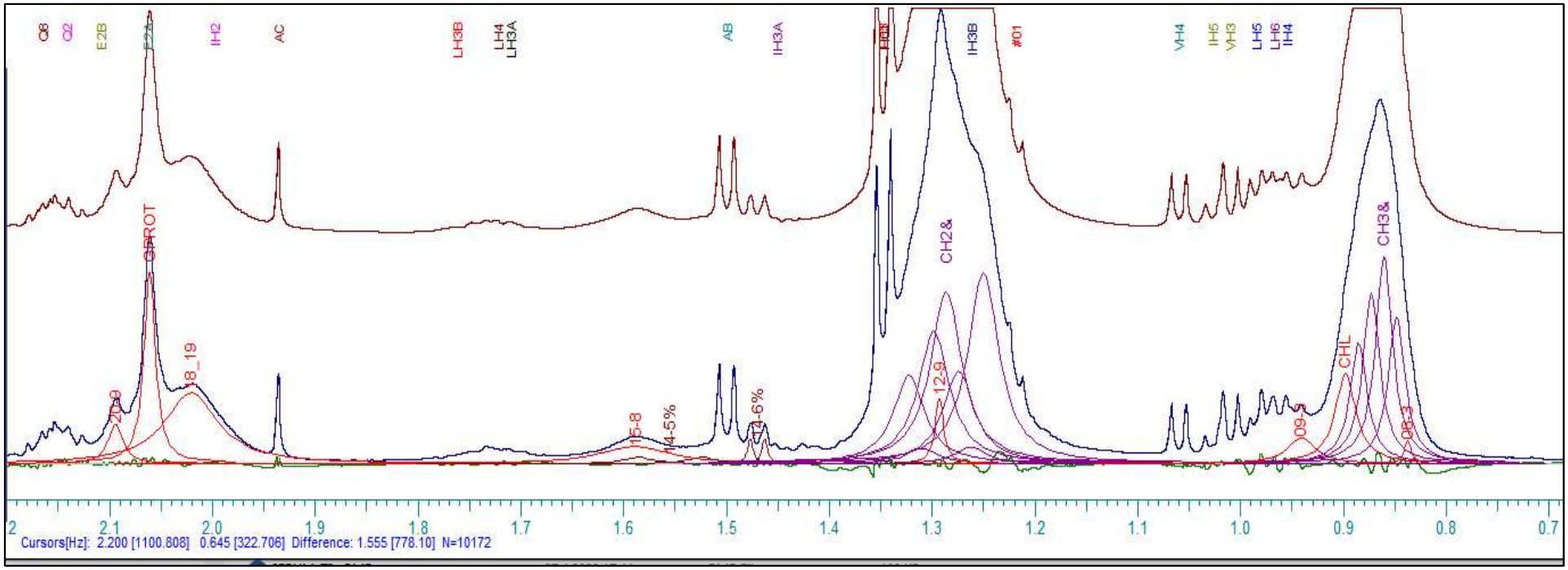
The common point is that the signal area/nucleus is the same:

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

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

QMSA + CTLS of *T2 edited* serum spectrum with three types of ***xtructures***

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



Xpurities

- A spectrum may contain weak well-defined signals like **^{13}C satellites** or **peaks arising from unknown impurities**.
- In not ignored QMSA, the weak signals are added to the baseline and, thus, lead to a positive bias - too high a purity !
- The **xpurities** can be found automatically, added to the model and subtracted from the major spectrum.
- The peak areas give an estimates of the impurity concentrations and more accurate picture about the sample purity.