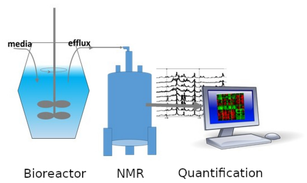
Towards qNMR for metabolomics: NMR signal modelling

Roy Chih Chung Wang[1], David A. Campbell[1,], Miroslava Čuperlović-Culf[2,3]
[1] *School of Mathematics and Statistics, Carleton University, Ottawa, Canada*
[2] *Digital Technologies Research Center, National Research Council, Ottawa, Canada*
[3] *Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Canada*
Correspondance: Miroslava.Cuperlovic-Culf@nrc-cnrc.gc.ca

Artificial Intelligence for Design Challenge Program



Bioreactors are a widely used method for producing cells for gene therapies, vaccines, and other bioprocessing applications. Quantitative NMR (qNMR) is a promising spectrometry method that could eventually help organizations to assess and optimize bioreactor performance in an on-line setting. The 1D ^{1}H NMR is the fastest and least expensive NMR technique for most metabolomic settings. However, many metabolites have overlapping 1D ^{1}H signatures in the spectral window of this type of NMR. Existing publicly-available automated qNMR algorithms have issues when the number of metabolites increase. Our approach is to first work characterizing and reducing the degrees of freedoms of a generative model of the NMR signal, before devising the quantification algorithm.

See [1] for a survey and review of qNMR methods. We are actively developing Julia packages for this work at https://github.com/AI4DBiological-Systems.

We use the conventional spin Hamiltonian theory to simulate the resonance frequencies and amplitudes, given only the spectrometer-invariant physical chemistry parameters of the compound.
Here, N is the number of nuclei in the spin system.

$$H = \sum_{j=1}^{N} (\omega_{cs})_j \mathfrak{J}_{jz} + \sum_{j<k} 2\pi (J)_{j,k} \mathfrak{J}_{j\cdot k}$$

All operators expressed in matrix-form in the Zeeman basis.
- $\mathfrak{J}_{jz}$ is angular momentum operator of j-th nucleus in the z-axis.
- $\mathfrak{J}_{j\cdot z}$ is the total product angular momentum operator of *j*-th and *k*-th nuclei.
- $\mathfrak{J}_z^{(N)}$ is the total angular momentum in the z-axis, $\mathfrak{J}_x^{(N)}$ for x-axis, $\mathfrak{J}_y^{(N)}$ for y-axis.
- ω_{cs} is the chemical shift, depends on NMR machine settings.
- $(J)_{j,k}$ J-coupling constant between the *j*-th and *k*-th nuclei.

$$\Omega_l = \boldsymbol{\lambda}_s - \boldsymbol{\lambda}_r$$
$$\alpha_l = -\tfrac{1}{2i} \left\langle v_r^*, \mathfrak{J}_y^{(N)} v_s \right\rangle \left\langle v_s^*, \mathfrak{J}_x^{(N)} v_r \right\rangle$$

$$\mathfrak{J}_z^{(N)} = \sum_{j=1}^{N} \mathfrak{J}_{jz}$$

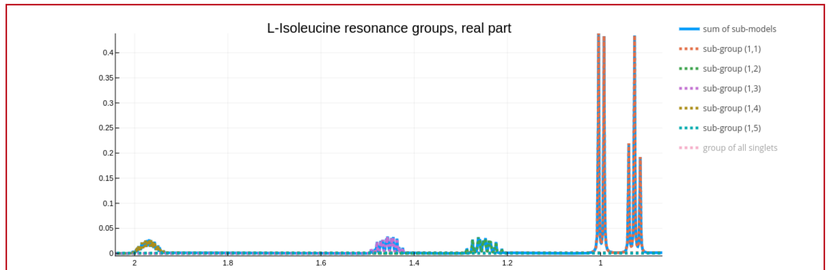
- The *s*-th eigenvalue of the Hamiltonian *H* is $\boldsymbol{\lambda}_s$ and its eigenvector is v_s.
- An eigenpair generates an observable resonance component if $\rho_{rs} = -1$.
- The amplitude α and frequency Ω respectively for the *l*-th component are generated by the *(r,s)* eigenpair.

$$M_r = \frac{\langle \mathfrak{J}_z^{(N)} v_r, v_r \rangle}{\langle v_r, v_r \rangle} \qquad \rho_{rs} = M_r - M_s, \quad \forall r, s \in \{1, 2, \cdots, 2^N\}$$

We devised a resonance group model from the individual nucleus angular momentum $\mathfrak{J}_{jz}$ to create a "feature" $\Delta c_{(r,s)}$ for each component, then partition or cluster the number of contributing resonance components into a small number of groups that share the same degrees of freedom.

$$c_{j,r} := \frac{\langle \mathfrak{J}_{jz} v_r, v_r \rangle}{\langle v_r, v_r \rangle} \qquad \sum_j \Delta c_{(r,s),j} = -1 \qquad \Delta c_{(r,s)} := \begin{bmatrix} \Delta c_{(r,s),1} \\ \Delta c_{(r,s),2} \\ \vdots \\ \Delta c_{(r,s),N} \end{bmatrix}$$

$$\Delta c_{(r,s),j} := c_{j,r} - c_{j,s}$$



interactive illustrations at https://ai4dbiological-systems.github.io/ResonanceGroupIllustration/

Our models are an extension of the free-induction decay (FID) NMR signal model. The FID model has the parameters *w* (relative concentration), α (amplitude), Ω (frequency), λ (T-2 decay), *β* (phase) for each resonance component. Existing qNMR methods use an "autophasing" step to remove the phase parameter, but it only allows Fourier domain processing and distorts the original signal. Since our models can be interpreted as an FID model, we open the possibility of fitting the data in the time-domain for the concentration quantification step, after the alignment step.

Let N be the number of metabolites, ι be the imaginary number, L_n be the number of resonance components of the *n*-th compound, *u* be frequency, and *d* be a shift variable. To align the model to the NMR data in the Fourier domain, we fix λ from a fast preliminary estimate of T-2 from data, simulate α and Ω via Hamiltonian theory using the machine settings for the data and tabulated J-coupling values from [2]. The conventional Fourier-domain FID model seeks to fit *w*, *u*, *d*, *β*:

$$g_{\text{FFID}}(u, w, d, \lambda, \beta) = \sum_{n=1}^{N} w_n \sum_{l=1}^{L_n} \frac{\alpha_{n,l} e^{\iota \beta_{n,l}}}{\lambda_{n,l} + \iota(2\pi u + d_{n,l} - \Omega_{n,l})}.$$

To reduce L_n, we use the "feature" Δc to partition the components for each metabolite into the groups: G_n, the number of spin systems of metabolite *n*, $K_{n,i}$ the number of resonance groups for system *i*. Our spectral model is

$$g_{mFFID}(u, d, \kappa, w, \lambda, \beta) = \sum_{n=1}^{N} w_n \sum_{i=1}^{G_n} \sum_{k=1}^{K_{n,i}} \sum_{l=1}^{L_{n,i,k}} \frac{\kappa_{\alpha,n,i,k} \alpha_{n,i,k,l} e^{\iota \langle \kappa_{\beta,n,i}, \Delta c_{n,i,k,l} \rangle}}{\lambda_{n,i} + \iota(2\pi u + d_{n,i} - \Omega_{n,i,k,l})}$$

In our model, we assume there are $L_{n,i,k}$ components indexed by (n,i,k,l) that share the same degrees of freedom for a number of parameters. We model phase as an inner product of κ_β and Δc, and add a multiplier variables κ_α and ξ because the simulated α does not correspond to the experimental data for some regions for some metabolites in natural abundance.

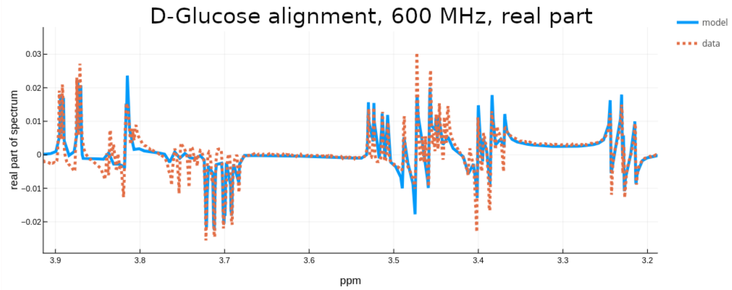
$$q_{n,i,k}(r, h) \approx \sum_{l=1}^{L_{n,i,k}} \frac{\alpha_{n,i,k,l}}{h + \iota(r - \Omega_{n,i,k,l})}$$

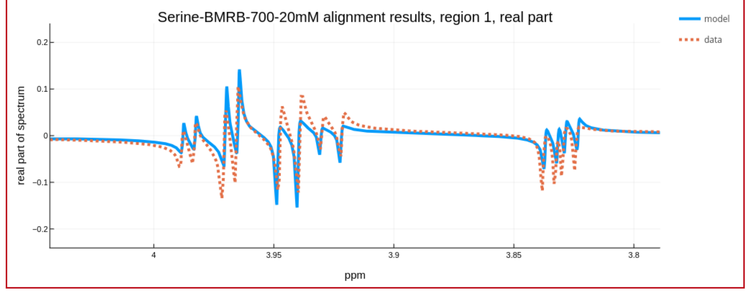
$$q(u, w, \kappa, \lambda, \beta) = \sum_{n=1}^{N} w_n \sum_{i=1}^{G_n} \sum_{k=1}^{K_{n,i}} z_{n,i,k} q_{n,i,k}(2\pi u - d_{n,i}, \xi_{n,i} \lambda_0)$$

$$z_{n,i,k} := \kappa_{\alpha,n,i,k} e^{\iota \langle \kappa_{\beta,n,i}, \Delta \bar{c}_{n,i,k} \rangle} \in \mathbb{C}$$

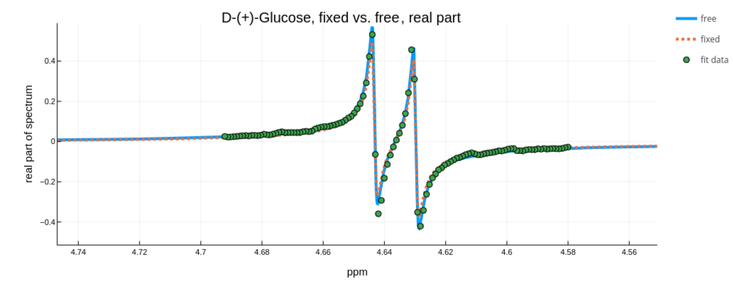
We approximated $\Delta c_{n,i,k,l}$'s with a single $\Delta c_{n,i,k}$, then used 2-D spline surrogate functions $q_{n,i,k}$'s so that we do not need to compute the sum involving $L_{n,i,k}$ when we evaluate the model. The surrogate model for the entire metabolite mixture is q, and is the actual model used to fit the data.

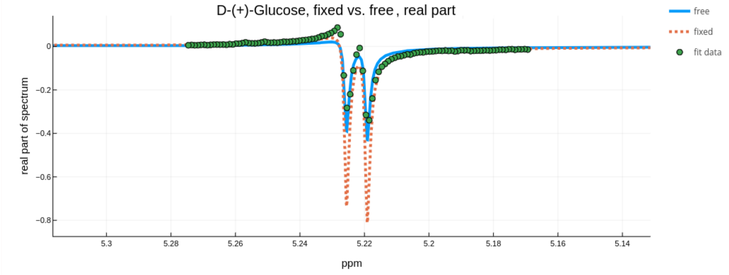
Since the accuracy of estimating concentration by fitting sharp resonance spectral lines is very sensitive to modelling error, so we only do frequency alignment in the Fourier domain.





The following plots of D-Glucose at different regions with a fixed and free (to fit) compensation κ_α show that the simulation need to be adjusted in amplitude (via κ_α) or T-2 (spectral peak width, via ξ). We plan to explore and learn the κ_α and ξ variables from data over different spectrometer frequencies and metabolite concentrations.





References: [1]: Wang et. al. Automatic 1D 1H NMR Metabolite Quantification for Bioreactor Monitoring. Metabolites, 2021, DOI: 10.1002/biot.202000320. [2]: Dashti et. al., Applications of Parametrized NMR Spin Systems of Small Molecules, Analytic Chemistry, 2018, DOI: 10.3390/metabo11030157.