Supporting Information: Analysis of Small Molecule Mixtures by Super-Resolved ^{1}H NMR Spectroscopy

Aritro Sinha Roy[†] and Madhur Srivastava[*,†,‡]

†*Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA*

‡*National Biomedical Resources for Advanced ESR Technologies (ACERT), Ithaca, NY, USA*

* E-mail: ms2736@cornell.edu

Methods

Wavelet transform

A continuous wavelet transform can be defined as,[S1,S2]

$$F(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(\delta)\psi^* \left(\frac{\delta - \tau}{s}\right) dt \tag{1}$$

where s is the inverse frequency (or frequency range) parameter, τ is the signal localization parameter, δ represents the chemical shift, $f(\delta)$ is the spectrum, $F(\tau, s)$ is the wavelet-transformed signal at a given signal localization and frequency and $\psi^* \left(\frac{\delta-\tau}{s}\right)$ is the signal probing function called "wavelet." Different wavelets are used to vary selectivity or sensitivity of adjacent frequencies with respect to signal localization. They are not dependent on *a priori* information of the signal or its characteristics.

Discrete wavelet transform (DWT) is expressed by two sets of wavelet components (Detail and Approximation) in the following way:[S1]

$$D_j[n] = \sum_{m=0}^{p-1} f[\delta_m] 2^{\frac{j}{2}} \psi[2^j \delta_m - n] \tag{2}$$

$$A_j[n] = \sum_{m=0}^{p-1} f[\delta_m] 2^{\frac{j}{2}} \phi[2^j \delta_m - n] \tag{3}$$

where $f[\delta_m]$ is the discrete input spectrum, p is the length of input signal $f[\delta_m]$, $D_j[n]$ and $A_j[n]$ are the Detail and Approximation components, respectively, at the j^{th} decomposition level, and $\psi[2^j \delta_m - n]$ and $\phi[2^j \delta_m - n]$ are wavelet and scaling functions, respectively. The maximum number of decomposition levels that can be obtained is N, where $N = \log_2 p$, and $1 \leq j \leq N$. The scaling and wavelet functions, at a decomposition level, are orthogonal to each other, as they represent non-overlapping frequency information. Similarly, wavelet functions at different decomposition levels are orthogonal to each other.

The Detail component $D_j[n]$ is the discrete form of Equation 1, where j and n are associated with s and τ, respectively. The Approximation component $A_j[n]$ represent the remaining frequency bands not covered by the Detail components till j^{th} level. The signal $f[\delta_m]$ can be reconstructed using the inverse discrete wavelet transform as follows,

$$f[\delta_m] = \sum_{k=0}^{p-1} A_{j_0}[k] \phi_{j_0,k}[\delta_m] + \sum_{j=1}^{j_0} \sum_{k=0}^{p-1} D_j[k] \psi_{j,k}[\delta_m] \tag{4}$$

where j_0 is the maximum decomposition level from which input signal needs to be reconstructed. Compared to that, both the Approximation and Detail components at each level are further decomposed into a set of Approximation and Detail components. A schematic diagram of DWT and WPT decomposition against increasing levels are shown for comparison in Figure S1.

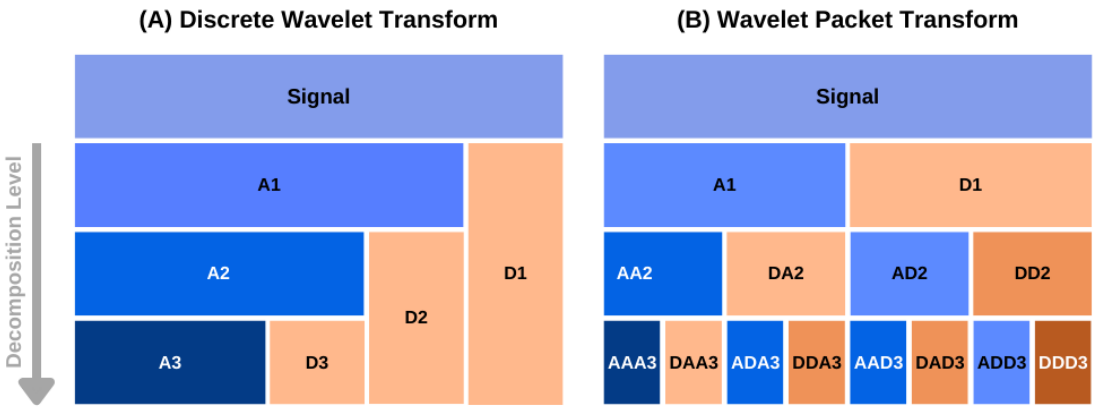

Figure S1: A schematic diagram of data decomposition in discrete (A) and packet wavelet transform (B) methods. The Approximation and Detain components at level k is denoted as A_k and D_k in (A). In case of wavelet packet transform, the Approximation and Detail component at a decomposition level are denoted by the component name of the previous level followed by A_k or D_k, respectively.

Tables

This section tabulates the ^{1}H NMR cluster midpoints and the WPT shift peak positions for all the saccharides and amino acids used in this work.

Table S1: ^{1}H NMR peaks vs. WPT shift peaks for the carbohydrates

Molecule	Multiplet Type	^{1}H Cluster Midpoint (ppm)	WPT Shift Peaks (ppm)
Cellobiose	d	5.24	5.24
	d	4.52	4.52
	m	3.95	3.95
	m	3.88	3.89
	m	3.83	3.83
	dd	3.75	3.75
	m	3.65	3.65
	m	3.60	3.61
	m	3.51	3.50
	m	3.43	3.44
	m	3.32	3.32
Fructose	m	4.12	4.12
	dd	4.03	4.04
	m	4.01	4.01
	dd	3.90	3.90
	m	3.82	3.81
	m	3.70	3.71
	m	3.58	3.57
Sucrose	d	5.40	5.40
	d	4.21	4.21
	t	4.04	4.04
	dd	3.89	–
	dd	3.87	3.87
	m	3.82	3.82
	t	3.75	3.74
	s	3.67	3.67
	dd	3.55	3.55
	t	3.46	3.46

Table S2: ^{1}H NMR peaks vs. WPT shift peaks for the amino acids

Molecule	Multiplet Type	^{1}H Cluster Midpoint (ppm)	WPT Shift Peaks (ppm)
Glutamine	t	3.77	3.76
	m	2.45	2.44
	m	2.13	2.13
Glycine	s	3.54	3.54
Isoleucine	d	3.66	3.66
	m	1.97	1.97
	m	1.46	1.45
	m	1.25	1.25
	d	1.00	1.00
	t	0.93	0.92
Leucine	m	3.72	3.73
	m	1.70	1.70
	t	0.95	0.95
Threonine	m	4.24	4.24
	d	3.58	3.58
	d	1.32	1.32
Valine	d	3.60	3.60
	d	2.26	2.26
	m	1.03	1.03
	d	0.98	0.97

References

(S1) Addison, P. *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, 2nd ed.; CRC Press: London, UK, 2016.

(S2) Srivastava, M. Improving Signal Resolution and Reducing Experiment Time in Electron Spin Resonance Spectroscopy via Data Processing Methods. Ph.D. thesis, Cornell University, 2018.