HHS Public Access

Author manuscript

J Phys Chem A. Author manuscript; available in PMC 2023 December 08.

Published in final edited form as:

J Phys Chem A. 2022 December 08; 126(48): 9108–9113. doi:10.1021/acs.jpca.2c06858.

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

Aritro Sinha Roy,

Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-0001, United States

Madhur Srivastava

Chemistry and Chemical Biology,Cornell University, Ithaca, New York 14853-0001, United States; National Biomedical Resources for Advanced ESR Technologies (ACERT), Ithaca, New York 14853, United States

Abstract

Analysis of small molecules is essential to metabolomics, natural products, drug discovery, food technology, and many other areas of interest. Current barriers preclude from identifying the constituent molecules in a mixture as overlapping clusters of NMR lines pose a major challenge in resolving signature frequencies for individual molecules. While homonuclear decoupling techniques produce much simplified *pure shift* spectra, they often affect sensitivity. Conversion of typical NMR spectra to pure shift spectra by signal processing without *a priori* knowledge about the coupling patterns is essential for accurate analysis. We developed a super-resolved wavelet packet transform based ^{1}H NMR spectroscopy that can be used in high-throughput studies to reliably decouple individual constituents of small molecule mixtures. We demonstrate the efficacy of the method on the model mixtures of saccharides and amino acids in the presence of significant noise.

Graphical Abstract

Corresponding Author Madhur Srivastava – ms2736@cornell.edu.

AUTHOR INFORMATION

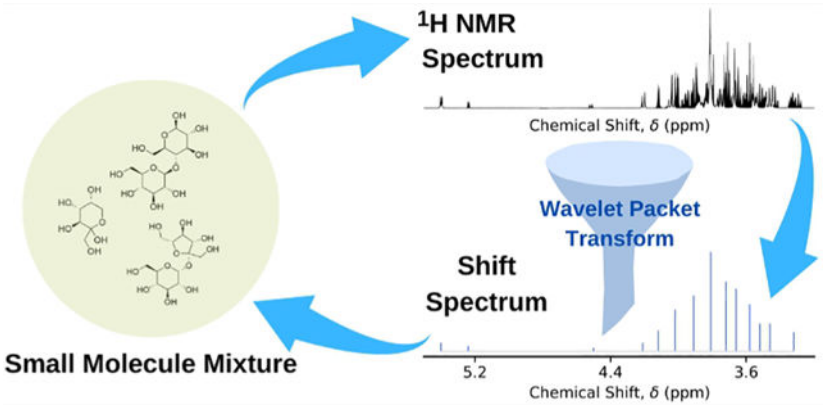
Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c06858

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c06858.

Overview of the wavelet transform theory and tables of peaks from clusters in ^{1}H NMR and WPT shift spectra of individual molecules (PDF)

The authors declare no competing financial interest.



INTRODUCTION

Analysis of small molecules across a wide range of studies of metabolites, organic and natural products, food, potential drug molecules rely heavily on nuclear magnetic resonance (NMR) spectroscopy, along with mass spectrometry (MS) and liquid chromatography (LC).[1–9] Out of the three methods, NMR is ideal for identification of novel molecules and large-scale studies, given the ease of sample preparation, its non-destructive nature, access to instruments and high reproducibility.[10] However, ^{1}H NMR spectroscopy, which is the most widely used version, suffers from limited resolution in analyzing small-molecule mixtures.[11,12] In recent years, sensitivity of NMR has been improved greatly,[13–15] but resolution enhancement efforts have faced an immovable barrier due to the presence of overlapping multiplets, originating from the relatively small chemical shift windows and scalar couplings between protons. While multidimensional NMR can enhance resolution, these are often not feasible because of the long signal acquisition time associated with those. Collapsing the multiplets in a typical ^{1}H NMR spectrum into singlets produces a *pure shift* spectrum,[16,17] which can greatly improve resolution for analysis of small molecule mixtures. Homonuclear decoupled ^{1}H NMR yields pure shift spectra,[16,18–21] but the resolution increment in those experiments often comes at the price of reduced sensitivity and spectral artifacts.[16,22] However, deconvolution of NMR spectra to shift spectra, which includes the maximum entropy method, can only be applied successfully when the scalar coupling patterns and the coupling constants are known.[23,24] Many other sampling, spectral editing, and fingerprint matching techniques have been developed for both 1D and 2D NMR spectroscopic analysis,[25–29] but those have either limited applicability or limited reliability, especially for highly complex or novel molecular mixtures. In that regard, NMR shift spectroscopy provides more control to the users in spectral interpretation and error analysis. Therefore, our objective was to propose a signal processing technique that can be applied to any ^{1}H NMR spectrum of a mixture of small molecules without *a priori* knowledge and convert it to a shift spectrum for accurate identification of the constituting molecules.

Magnetic resonance spectra are made of multiple components corresponding to different types of interactions among nuclear and/or electronic spins. A multilevel filtration technique is required to separate the components by their frequency and timestamps. Wavelet transform is such a technique that decomposes a spectrum into a pair of approximation and

detail components, which corresponds to the low- and high-frequency spectral components, respectively. In converting NMR spectra to shift spectra, we used successive wavelet decomposition of the original spectra to produce an approximation component which provides the overall envelope encompassing a multiplet, essentially converting it to a singlet. Previously, we applied the wavelet transform based algorithm[30] to continuous wave electron spin resonance (cw-ESR) spectra, in order to enhance the resolution and extract hidden features.[31] Its applications in removing spectral noise[30,32–34] and isolating weak signal components in ESR[35] have been extremely successful.

In the discrete form of wavelet transform (DWT), the approximation component from a decomposition level is passed through to the next level of decomposition, providing a coarse filtering of the original signal into sub-bands of unequal frequency ranges. In wavelet packet transform (WPT), both the approximation and detail components are decomposed at each level, resulting in a much finer filtering over 2^K equal-width sub-bands at a decomposition level of K. That has been our motivation to use WPT for conversion of NMR to shift spectra without any loss of information. In this regard, we should note that detail components extract both high-frequency spectral features as well as noise. That is why wavelet transform based conversion of NMR to shift spectra is practically insensitive to noise.

THEORETICAL METHODS

A high level summary of the process of generating shift spectra from NMR using WPT is shown in Figure 1 using glutathione ^{1}H NMR as an example. In the first step, the input NMR spectrum was decomposed by WPT, producing an approximation component that is virtually the original spectrum, while most of the spectral noise showed up in the detail component. The decomposition was continued until all multiplets in the spectrum were converted to singlets in the resultant approximation component (A^K). For this particular example, K was equal to 7. The second step in the analysis poses two sets of challenges: (i) identifying artifacts and (ii) locating peak positions. We figured that the artifacts can be separated from actual peaks by comparing the final approximation component to the original spectrum as well as the approximation components from the previous decomposition levels. In the second step of Figure 1, the relevant regions are highlighted (light purple), while peaks showing up outside those regions were identified as artifacts and those are highlighted in yellow circles. The peaks in the relevant regions are pointed out by downward triangles (blue). Consequently the stick spectrum is produced combining the peak positions and heights. In this work, we selected the peaks by visual inspection, however one can develop an automated method based on well-justified criteria.[36] The details of the WPT theory is provided in the Supporting Information.

RESULTS AND DISCUSSION

Model Data Preparation.

Quantitative analysis of carbohydrates is an essential step in food and pharmaceutical industries,[37–41] while identification of amino acids in biological samples is crucial in probing various disease pathophysiology.[42–46] We analyzed two model mixtures of (I) saccharides and (II) amino acids in this work. An equal mixture of cellobiose, fructose

and sucrose was used in (I). In our second model (II), we used a mixture of six amino acids, namely, glutamine, glycine, isoleucine, leucine, threonine, and valine. Two different varieties of (II) were prepared by mixing the individual amino acid spectra in different proportions.

All the NMR spectra in this work were obtained from the Human Metabolome Database (HMDB).[47] Impurity peaks at chemical shifts values smaller than 2.4 ppm in the cellobiose spectrum were removed. Normalization of individual spectra was carried out by dividing each data set by the sum of the NMR intensities. We concatenated individual components' spectra to create the NMR spectra for the model mixtures. The process of concatenation increased the number of spectral data points compared to the individual spectra or artificially enhanced spectral resolution. In order to resolve the issue, the resultant spectral domain size was reduced to the mean length of the individual spectra by binning of the NMR frequencies into discrete intervals. For mixture (I), the noisy spectra was obtained by adding 2.5% random noise to the noise-free version, and 10% random noise was added in case of mixture (II).

Spectral Analysis.

In analyzing a model mixture of saccharides (I), WPT shift spectra for each of the components, cellobiose, fructose, and sucrose, were obtained first from the corresponding experimental ^{1}H NMR spectra, recorded at 500 MHz. Cellobiose is a disaccharide and is an excellent test case given that its NMR spectrum is highly overlapped in the region of 3.2–4.0 ppm. The WPT spectra were obtained from the corresponding ^{1}H NMR spectra by using Daubechies (Db9) wavelet at the decomposition levels of 7 (cellobiose), 7 (fructose), and 8 (sucrose), respectively. For all the three molecules, the ^{1}H NMR multiplet midpoints and the WPT shift peaks are given in Table S1. It can be seen that each of the peaks were replicated in the WPT shift spectra except for the peak at 3.89 ppm for sucrose. The NMR spectrum of sucrose has two double doublets centered at 3.87 and 3.89 ppm. The WPT analysis in this case treated the two closely spaced clusters as a single multiplet and yielded a peak at 3.87 ppm. The original NMR and the WPT shift spectra for the three saccharides are shown in Figure 2. Next, the individual NMR spectra were mixed in an appropriate fashion to create the NMR spectra for a mixture of the saccharides. Figure 3 shows the conversion and analysis of the mixed NMR spectra. The key takeaways were the replication of the signature NMR peaks for all three components in the highly overlapped region of 3.5 ppm < chemical shift ≤ 4.2 ppm and the exact match between the WPT shift spectra for the noise-free (A) and noisy (B) cases. Comparing with the original spectra, we inferred that the peak at 3.81 ppm was resulted from unresolved peaks centered around 3.81 ppm (fructose), 3.82 ppm (sucrose), and 3.83 ppm (cellobiose). A separate analysis of the region between 3.78 and 3.94 ppm at a wavelet decomposition level of 7 resolved the cellobiose peak at 3.83 ppm and previously unresolved peaks at 3.87 ppm (sucrose) and 3.90 ppm (fructose).

Two sets of NMR spectra for model mixture (II) were produced by mixing the spectra of glutamine, glycine, isoleucine, leucine, threonine, and valine at equal (II.A) and varying proportions (II.B), given in Figure 4. The noise level was set to 10% compared to the highest NMR peak intensity in both the cases. The area under an NMR peak is proportional to the number of associated protons and in a mixed spectra, separated constituent peaks can

be used in quantification of the composition of the mixture. Given that the area under a peak is proportional to both its height and width, we compared the WPT shift spectra of (II.A) and (II.B) for correlation with the compositions of the mixtures. Glancing through the two WPT shift spectra shown in Figure 4A,B, we can notice significant changes in the peak intensities for glycine, threonine and valine, which correlates to the change in the compositions of the two mixtures. By comparing the peak heights in the second mixture with the first, we calculated the amount of those three amino acids to be 13% (absolute error = +3), 24% (−6), and 14% (+4), respectively. For glutamine, isoleucine and leucine, the calculated concentrations from WPT spectra were 18% (−2), 17% (+2), and 15% (0). While the calculated relative concentrations were significantly off from the actual values given that peak heights were used instead of peak areas, the changes in peak intensities in the two model mixtures correlated with the changes in their compositions. This property of WPT shift spectra can be useful in semiquantitative analysis of small molecule mixtures and probing relative changes in their compositions. WPT shift peaks for the individual amino acids, obtained at a wavelet decomposition level of 7, are tabulated in Table S2.

CONCLUSIONS

In summary, we showed that wavelet transform or more specifically wavelet packet transform produces highly resolved shift spectra from ^{1}H NMR spectra of small molecule mixtures in a reliable fashion. Our method extracted the signature NMR lines for the constituent molecules in two model mixtures of saccharides and amino acids, even in the presence of 10% noise. Additionally, correlation between the composition of a mixture and its WPT shift peak intensities revealed the quantitative nature of the method. It should be noted that WPT analysis resolves NMR spectra fully where the multiplets are not indistinguishably overlapped.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This research was funded by NSF Grant No. 2044599 and in part by the Cornell internal funding.

Data Availability Statement

Data used in this work have been stored in the public github repository (https://github.com/Signal-Science-Lab), and the software is available at https://denoising.cornell.edu.

REFERENCES

(1). Worley B; Powers R. Multivariate analysis in metabolomics. Curr. Metabolomics 2012, 1, 92–107.
(2). Idle JR; Gonzalez FJ Metabolomics. Cell Metabolism 2007, 6, 348–351. [PubMed: 17983580]
(3). Pellecchia M; Sem DS; Wüthrich K. NMR in drug discovery. Nat. Rev. Drug Discovery 2002, 1, 211–219. [PubMed: 12120505]
(4). Softley CA; Bostock MJ; Popowicz GM; Sattler M. Paramagnetic NMR in drug discovery. Journal of Biomolecular NMR 2020, 74, 287–309. [PubMed: 32524233]

(5). Pauli GF; Jaki BU; Lankin DC Quantitative 1H NMR: Development and potential of a method for natural products analysis. J. Nat. Prod 2005, 68, 133–149. [PubMed: 15679337]

(6). Breton RC; Reynolds WF Using NMR to identify and characterize natural products. Natural Product Reports 2013, 30, 501–524. [PubMed: 23291908]

(7). Wen J-L; Sun S-L; Xue B-L; Sun R-C Recent advances in characterization of lignin polymer by solution-state nuclear magnetic resonance (NMR) methodology. Materials 2013, 6, 359–391. [PubMed: 28809313]

(8). Spyros A; Dais P. 31P NMR spectroscopy in food analysis. Prog. Nucl. Magn. Reson. Spectrosc 2009, 54, 195–207.

(9). Capitani D; Sobolev AP; Di Tullio V; Mannina L; Proietti N. Portable NMR in food analysis. Chem. Biol. Technol. Agric 2017, 4, 17.

(10). Emwas A-H; Roy R; McKay RT; Tenori L; Saccenti E; Gowda GN; Raftery D; Alahmari F; Jaremko L; Jaremko M; et al. NMR spectroscopy for metabolomics research. Metabolites 2019, 9, 123. [PubMed: 31252628]

(11). Markley JL; Brüschweiler R; Edison AS; Eghbalnia HR; Powers R; Raftery D; Wishart DS The future of NMR-based metabolomics. Curr. Opin. Biotechnol 2017, 43, 34–40. [PubMed: 27580257]

(12). Robinette SL; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. NMR in metabolomics and natural products research: Two sides of the same coin. Acc. Chem. Res 2012, 45, 288–297. [PubMed: 21888316]

(13). Lee JH; Okuno Y; Cavagnero S. Sensitivity enhancement in solution NMR: Emerging ideas and new frontiers. J. Magn. Reson 2014, 241, 18–31. [PubMed: 24656077]

(14). Mompeán M; Sánchez-Donoso RM; De La Hoz A; Saggiomo V; Velders AH; Gomez M. Pushing nuclear magnetic resonance sensitivity limits with microfluidics and photo-chemically induced dynamic nuclear polarization. Nat. Commun 2018, 9, 108. [PubMed: 29317665]

(15). Kovacs H; Moskau D; Spraul M. Cryogenically cooled probesA leap in NMR technology. Prog. Nucl. Magn. Reson. Spectrosc 2005, 46, 131–155.

(16). Zangger K. Pure shift NMR. Prog. Nucl. Magn. Reson. Spectrosc 2015, 86, 1–20. [PubMed: 25919196]

(17). Aguilar JA; Faulkner S; Nilsson M; Morris GA Pure shift 1H NMR: A resolution of the resolution problem? Angew. Chem., Int. Ed 2010, 49, 3901–3903.

(18). Foroozandeh M; Morris GA; Nilsson M. PSYCHE pure shift NMR spectroscopy. Chem. - Eur. J 2018, 24, 13988–14000. [PubMed: 29532969]

(19). Aguilar JA; Nilsson M; Morris GA Simple proton spectra from complex spin systems: Pure shift NMR spectroscopy using BIRD. Angew. Chem 2011, 123, 9890–9891.

(20). Lupulescu A; Olsen GL; Frydman L. Toward single-shot pure-shift solution 1H NMR by trains of BIRD-based homonuclear decoupling. J. Magn. Reson 2012, 218, 141–146. [PubMed: 22446507]

(21). Castañar L; Parella T. Broadband 1H homodecoupled NMR experiments: Recent developments, methods and applications. Magn. Reson. Chem 2015, 53, 399–426. [PubMed: 25899911]

(22). Giraudeau P. Challenges and perspectives in quantitative NMR. Magn. Reson. Chem 2017, 55, 61–69. [PubMed: 27370178]

(23). Delsuc MA; Levy GC The application of maximum entropy processing to the deconvolution of coupling patterns in NMR. Journal of Magnetic Resonance (1969) 1988, 76, 306–315.

(24). Shimba N; Stern AS; Craik CS; Hoch JC; Dötsch V. Elimination of 13Cα splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. J. Am. Chem. Soc 2003, 125, 2382–2383. [PubMed: 12603112]

(25). Hansen AL; Li D; Wang C; Brüschweiler, R. Absolute minimal sampling of homonuclear 2D NMR TOCSY spectra for high-throughput applications of complex mixtures. Angew. Chem 2017, 129, 8261–8264.

(26). Zhang F; Brüschweiler, R. Robust deconvolution of complex mixtures by covariance TOCSY spectroscopy. Angew. Chem 2007, 119, 2693–2696.

(27). Novoa-Carballal R; Fernandez-Megia E; Jimenez C; Riguera R. NMR methods for unravelling the spectra of complex mixtures. Natural Product Reports 2011, 28, 78–98. [PubMed: 20936238]

(28). Robinette SL; Zhang F; Brüschweiler-Li, L.; Brüschweiler, R. Web server based complex mixture analysis by NMR. Anal. Chem 2008, 80, 3606–3611. [PubMed: 18422338]

(29). Wei S; Zhang J; Liu L; Ye T; Gowda GN; Tayyari F; Raftery D. Ratio analysis nuclear magnetic resonance spectroscopy for selective metabolite identification in complex samples. Anal. Chem 2011, 83, 7616–7623. [PubMed: 21894988]

(30). Srivastava M; Anderson CL; Freed JH A new wavelet denoising method for selecting decomposition levels and noise thresholds. IEEE Access 2016, 4, 3862–3877. [PubMed: 27795877]

(31). Roy AS; Srivastava M. Hyperfine decoupling of ESR spectra using wavelet transform. Magnetochemistry 2022, 8, 32.

(32). Srivastava M. Improving signal resolution and reducing experiment time in electron spin resonance spectroscopy via data processing methods. Ph.D. thesis, Cornell University, 2018.

(33). Srivastava M, Freed JH, Anderson CL Systems, methods and programs for denoising signals using wavelets. US Patent US10,891,719, 2021.

(34). Bekerman W; Srivastava M. Determining decomposition levels for wavelet denoising using sparsity plot. IEEE Access 2021, 9, 110582–110591.

(35). Srivastava M; Dzikovski B; Freed JH Extraction of weak spectroscopic signals with high fidelity: Examples from ESR. J. Phys. Chem. A 2021, 125, 4480–4487. [PubMed: 34009996]

(36). Li D-W; Hansen AL; Yuan C; Bruschweiler-Li L; Brüschweiler, R. DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra. Nat. Commun 2021, 12, 5229. [PubMed: 34471142]

(37). Grembecka M. Sugar alcohols–Their role in the modern world of sweeteners: A review. European Food Research and Technology 2015, 241, 1–14.

(38). BeMiller JN Food Analysis; Springer, 2010; pp 147–177.

(39). Brummer Y; Cui SW Understanding carbohydrate analysis. In Food Carbohydrates: Chemistry, Physical Properties and Applications; Cui SW, Ed.; CRC Press, 2005; pp 1–38.

(40). Petersen BO; Hindsgaul O; Meier S. Profiling of carbohydrate mixtures at unprecedented resolution using highprecision 1 H-13 C chemical shift measurements and a reference library. Analyst 2014, 139, 401–406. [PubMed: 24301019]

(41). Harazono A; Kobayashi T; Kawasaki N; Itoh S; Tada M; Hashii N; Ishii A; Arato T; Yanagihara S; Yagi Y; et al. A comparative study of monosaccharide composition analysis as a carbohydrate test for biopharmaceuticals. Biologicals 2011, 39, 171–180. [PubMed: 21549615]

(42). Sakamoto T; Qiu Z; Inagaki M; Fujimoto K. Simultaneous amino acid analysis based on 19F NMR using a modified OPA-derivatization method. Anal. Chem 2020, 92, 1669–1673. [PubMed: 31886647]

(43). Armenta JM; Cortes DF; Pisciotta JM; Shuman JL; Blakeslee K; Rasoloson D; Ogunbiyi O; Sullivan DJ, Jr; Shulaev, V. Sensitive and rapid method for amino acid quantitation in malaria biological samples using AccQ• Tag ultra performance liquid chromatography-electrospray ionization-MS/MS with multiple reaction monitoring. Anal. Chem 2010, 82, 548–558. [PubMed: 20038084]

(44). Burton AS; Stern JC; Elsila JE; Glavin DP; Dworkin JP Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. Chem. Soc. Rev 2012, 41, 5459–5472. [PubMed: 22706603]

(45). Lu K; Ye W; Zhou L; Collins LB; Chen X; Gold A; Ball LM; Swenberg JA Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. J. Am. Chem. Soc 2010, 132, 3388–3399. [PubMed: 20178313]

(46). Hanff E; Ruben S; Kreuzer M; Bollenbach A; Kayacelebi AA; Das AM; von Versen-Höynck F; von Kaisenberg C; Haffner D; Ückert S; et al. Development and validation of GC–MS methods for the comprehensive analysis of amino acids in plasma and urine and applications to the HELLP syndrome and pediatric kidney transplantation: Evidence of altered methylation, transamidination, and arginase activity. Amino Acids 2019, 51, 529–547. [PubMed: 30604095]

(47). Wishart DS; Knox C; Guo AC; Eisner R; Young N; Gautam B; Hau DD; Psychogios N; Dong E; Bouatra S; et al. HMDB: A knowledgebase for the human metabolome. Nucleic Acids Res. 2009, 37, D603–D610. [PubMed: 18953024]

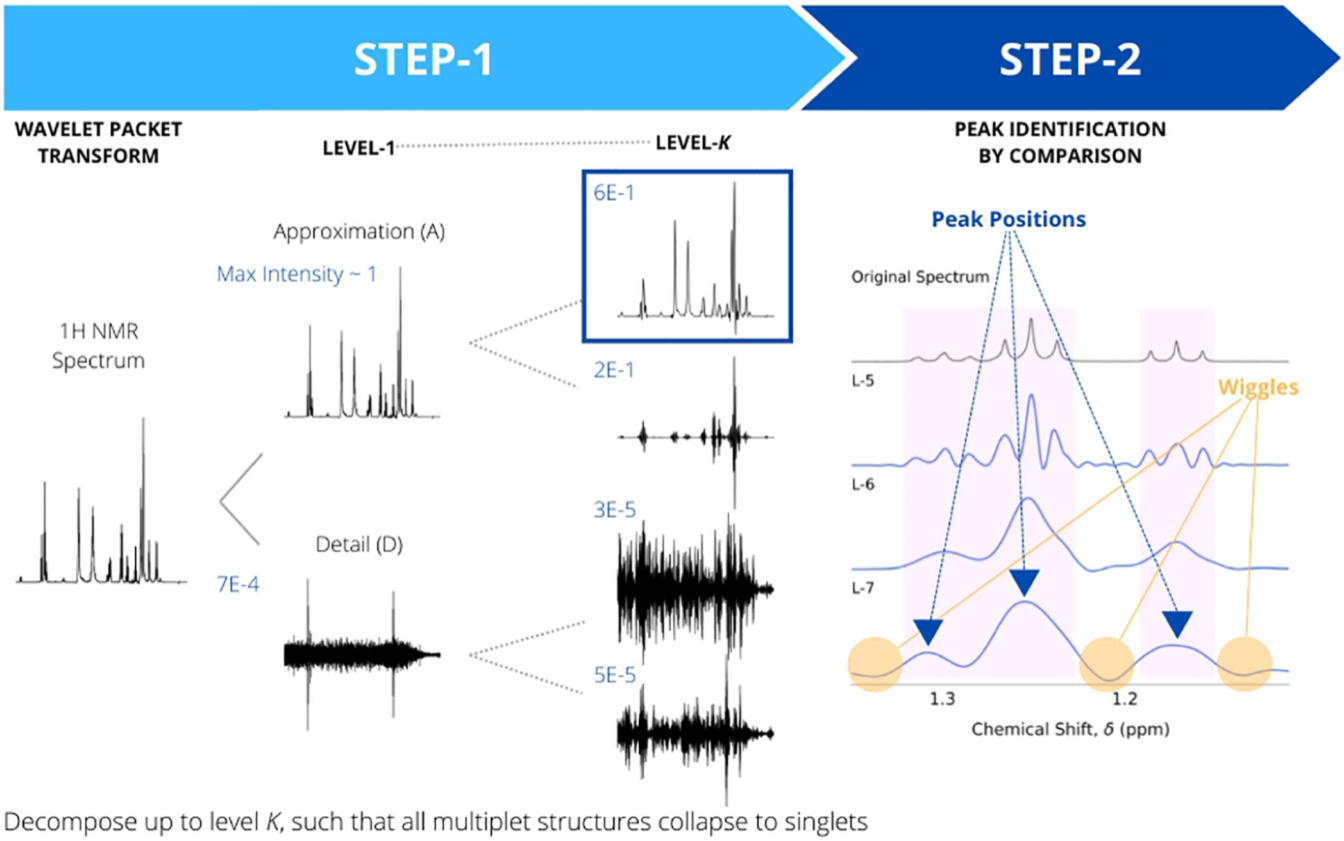


Figure 1.

Conversion of ^{1}H NMR spectra to WPT shift spectra. Glutathione ^{1}H NMR recorded at 500 MHz in H_2O is shown in the left. In the first step, all multiplets in the spectrum are collapsed to singlets in the approximation component (blue boxed) at WPT decomposition level *K*. The peaks are separated from artifacts (yellow circles) of WPT in the second step. The peak positions (blue colored triangles) and heights are used to produce the stick spectrum. The maximum intensity of the original NMR spectrum was ~1, and the maximum intensities of each of the components are printed next to the spectra in blue for comparison.

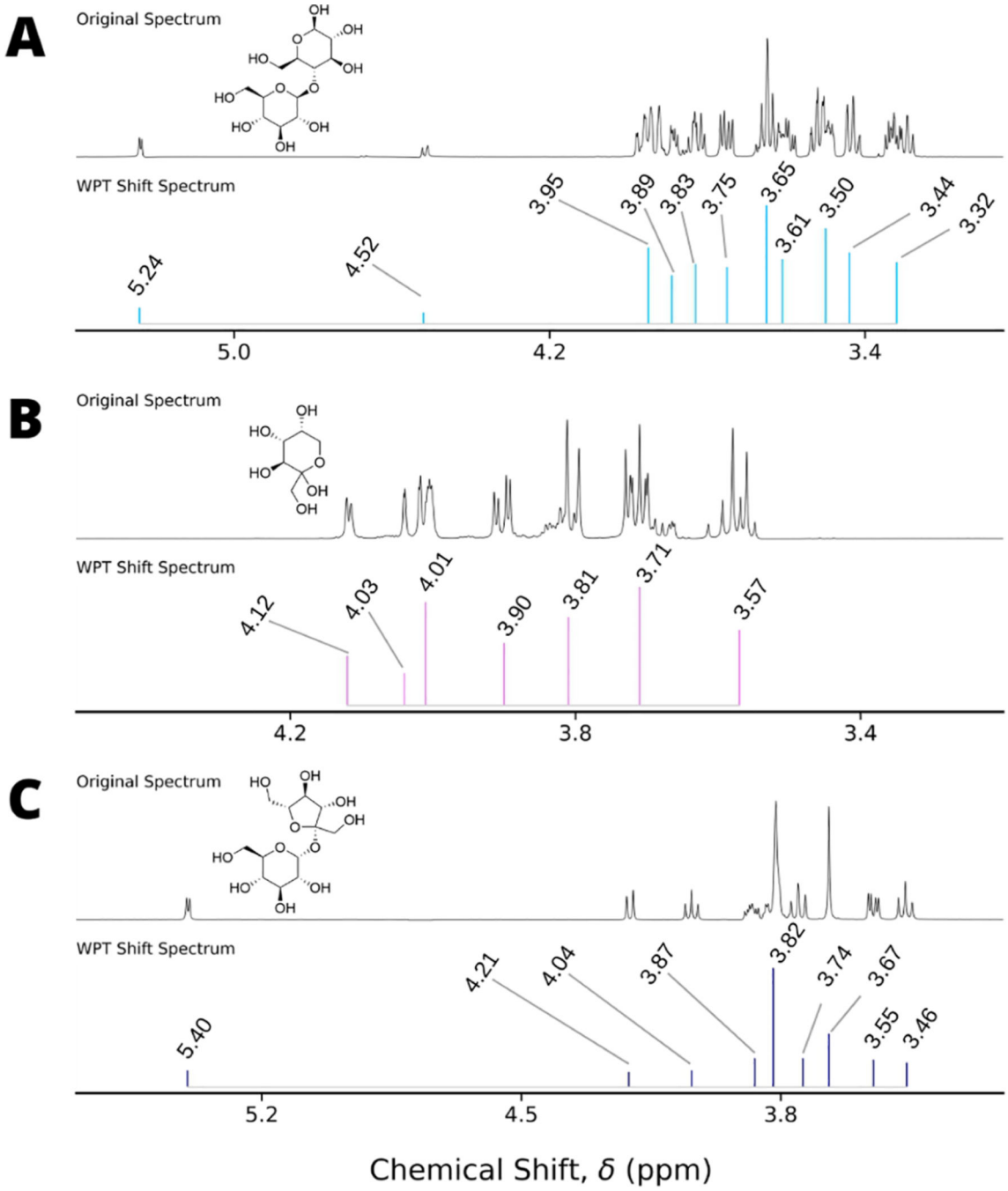


Figure 2.

Conversions of ^{1}H NMR spectra of cellobiose, fructose, and sucrose to WPT shift spectra. The WPT shift spectra were produced from the approximation components at levels 7 (A), 7 (B), and 8 (C) of wavelet decomposition of the respective NMR spectra by using Db9 wavelet.

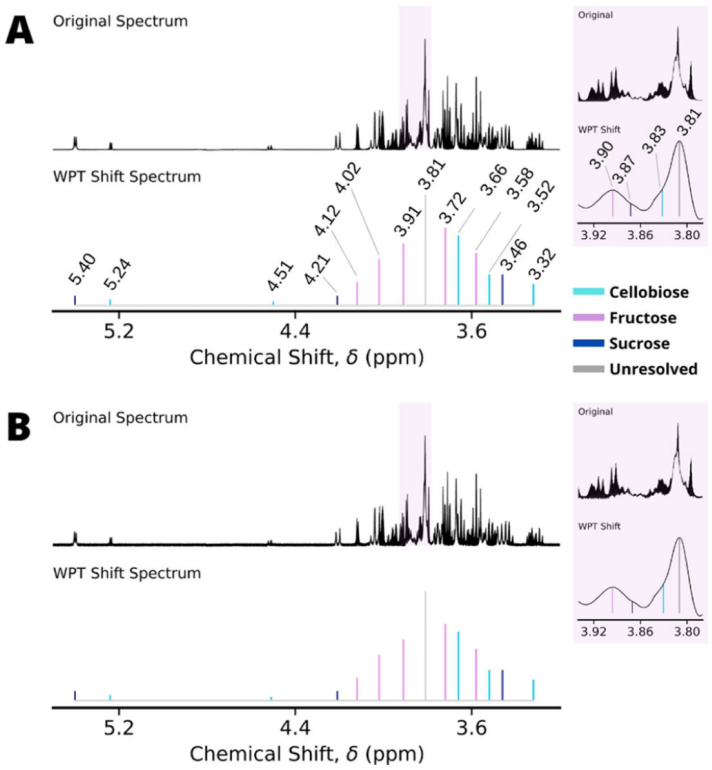


Figure 3.

Analysis of model mixtures of cellobiose, fructose, and sucrose using Db9 wavelet at a wavelet decomposition level of 8 are illustrated for the noise-free (A) and noisy (B) data sets. The resolved peaks corresponding to cellobiose (light blue), fructose (violet), and sucrose (navy) are color-coded, while the line at 3.81 ppm is marked unresolved (gray). Analysis of the spectral slice between 3.78 and 3.94 ppm at a decomposition level of 7 resolved the peaks at 3.83 ppm (cellobiose), 3.87 ppm (sucrose) and 3.90 ppm (fructose).

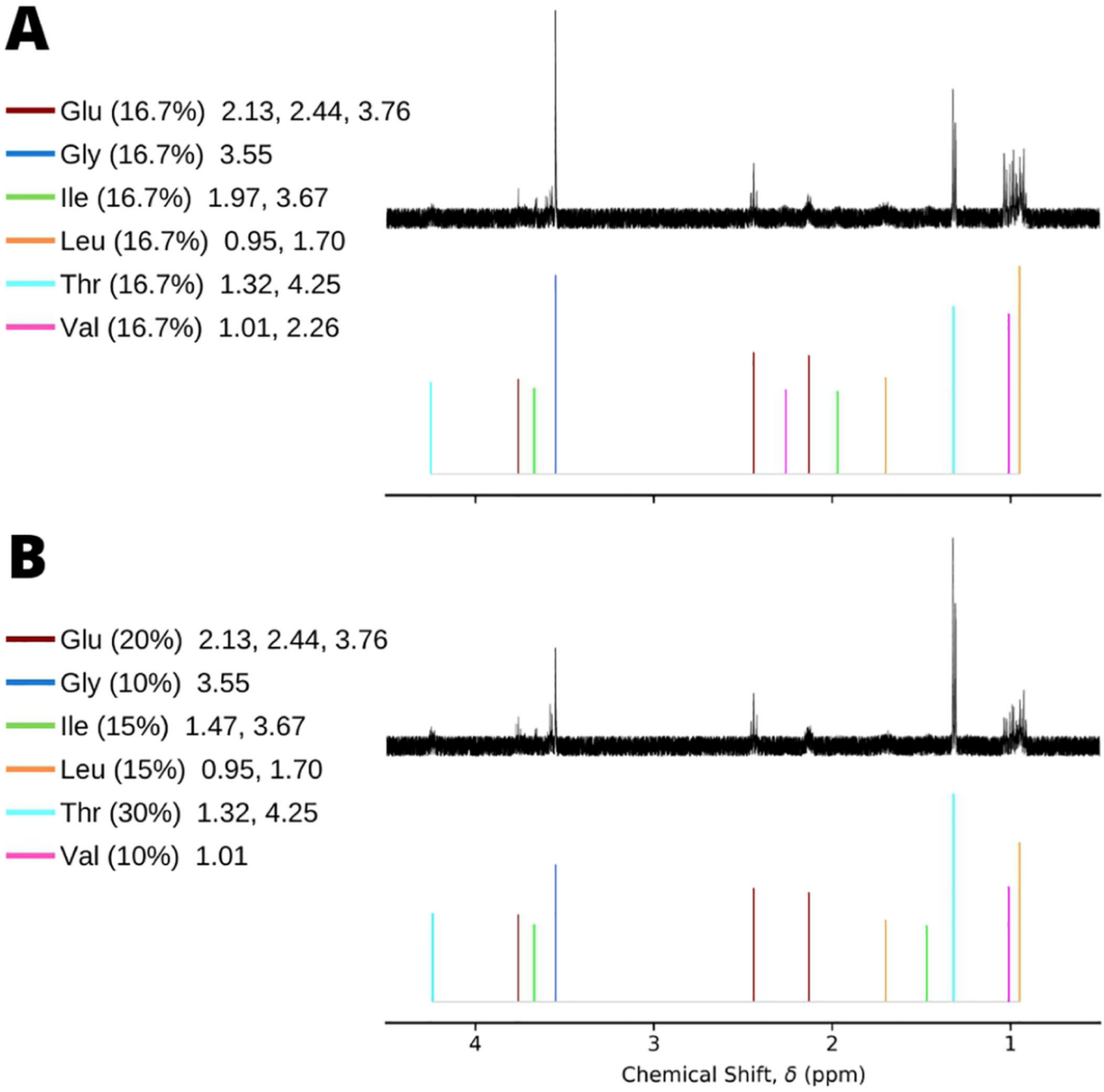


Figure 4.

Analysis of the NMR spectra of two different compositions of model mixture (II) of six amino acids at 10% noise are illustrated. Conversion of the NMR spectra using Db9 wavelet at a decomposition level of 8 are shown (A, B). The peaks corresponding to glutamine (red), glycine (blue), isoleucine (green), leucine (orange), threonine (cyan), and valine (magenta) were resolved in both the WPT shift spectra. The compositions of the mixtures and the resolved lines in the WPT shift spectra are tabulated on the left.