The Generally Useful Estimate of Solvent Systems (GUESS) method enables the rapid purification of methylpyridoxine regioisomers by countercurrent separation

Supplementary Data

Yang Liu[a], J. Brent Friesen[a,b,*], Larry L. Klein[c], James B. McAlpine[a,c], David C. Lankin[a], Guido F. Pauli[a,c], and Shao-Nong Chen[a,**]

[a] Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

[b] Physical Sciences Department, Rosary College of Arts and Sciences, Dominican University, River Forest, IL 60305, USA

[c] Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, USA

* Corresponding Author. Tel.: (312) 413-5894; Fax: (312) 996-8803.

** Corresponding Author. Tel.: (312) 355-1949; Fax: (312) 355-2693.

Email address: jbfriesen@dom.edu (J.B. Friesen), sc4sa@uic.edu (S.N. Chen).

Table of Contents

S1. NMR Identification of Compounds **1**, **2** and **3**.

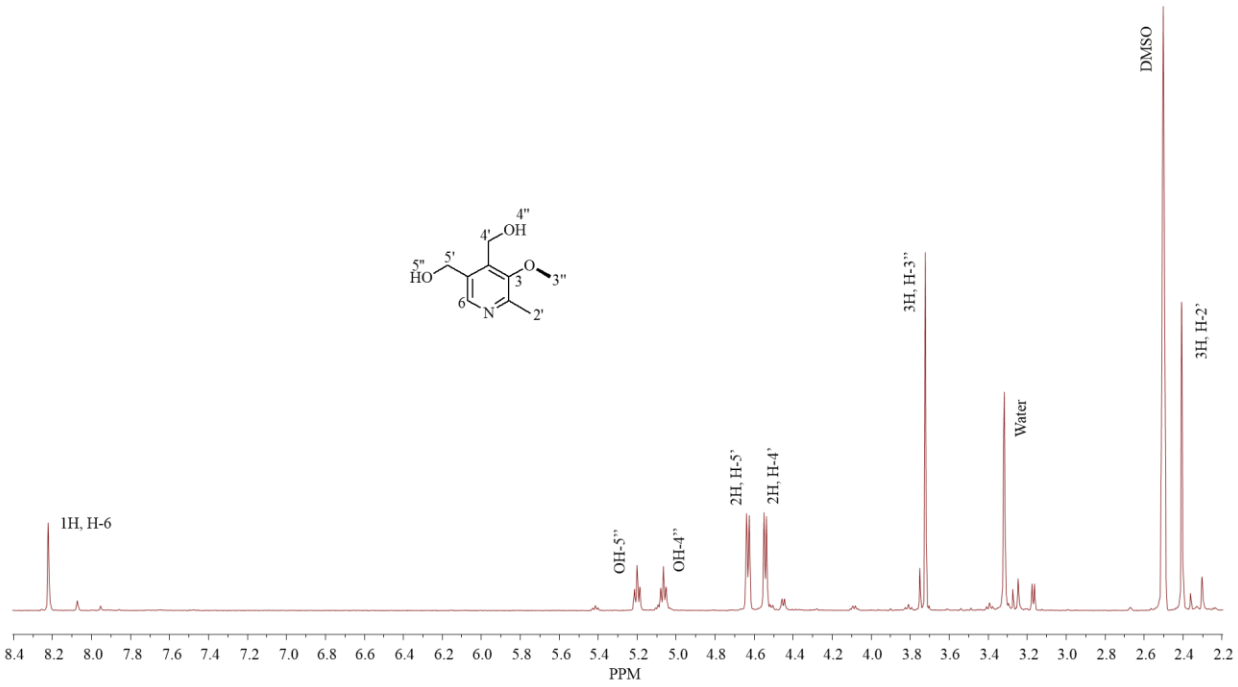
General 1D and 2D NMR – Data Acquisition. A Bruker DPX-400 NMR spectrometer (Karlsruhe, Germany), equipped with a 5 mm, 4-nucleus direct observe probe was used. The probe was frequency tuned and impedance matched prior to all data collection. The ^{1}H-NMR measurements were recorded at 400.17 MHz in 5 mm NMR tubes (Norell, Landisville, NJ, USA) at 298 K (25 °C). Chemical shifts are expressed in ppm (δ) with reference to the residual solvent signal (δ 7.240 ppm for $CHCl_3$) and relative to TMS (δ 0.000 ppm). Identical acquisition and processing parameters for quantitative ^{1}H NMR (qHNMR) spectra were used for all samples: pulse program 30 degree excitation pulse (zg30); 65536 (64k) time domain data; spectral width (sw) 20.69 ppm; acquisition time 4.0 s; a relaxation delay (d1) 30 s; number of scans (ns) 16. Depending on the concentration of each samples, receiver gains were obtained using the receiver gain adjust data (rga) feature. **Processing**. All processing was performed using MestReNova v9.0.1 (Mestrelab Research). Line resolution was improved by Lorentzian-Gaussian (LG) window functions (LB -0.3, GB 0.05) and zero-filling performed (256 K) and baseline correction using a 5th order polynomial function. Phase correction was performed manually.

The assessment used the 100% quantitative ^{1}H NMR (qHNMR) method as described in:

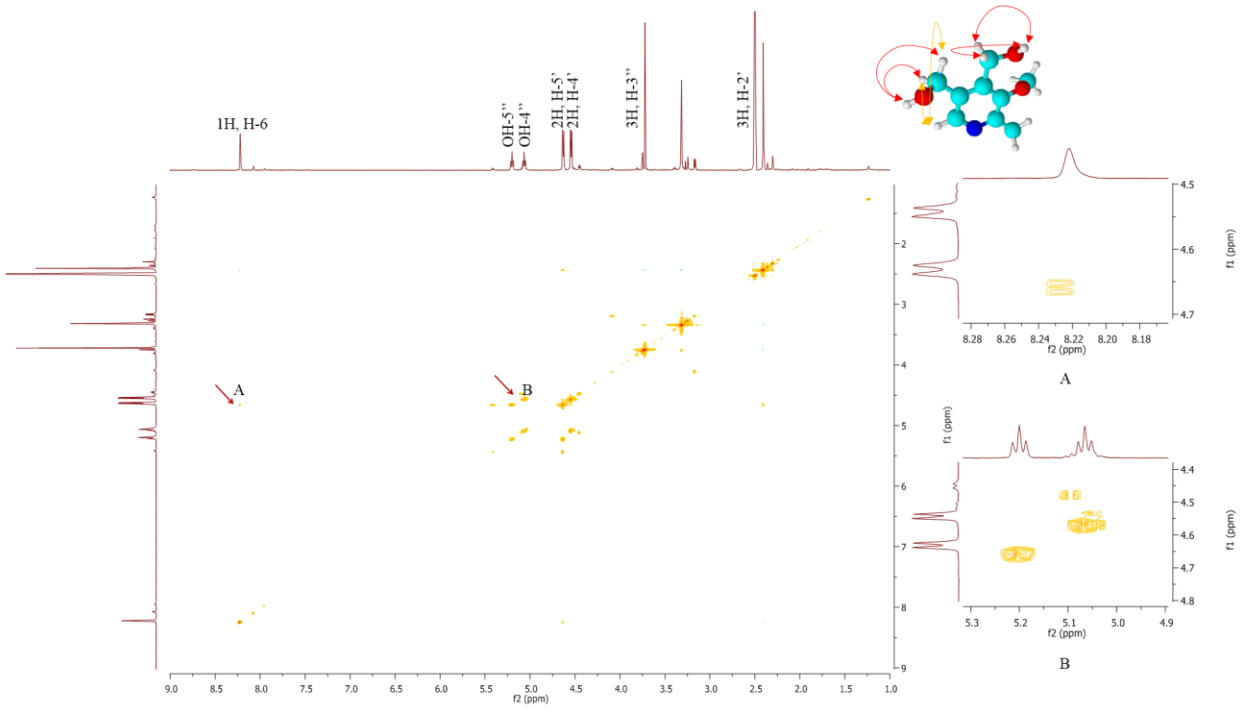
- Guido F. Pauli, Shao-Nong Chen, Charlotte Simmler, David C. Lankin, Tanja Gödecke, Birgit U. Jaki, J. Brent Friesen, James B. McAlpine, and José G. Napolitano. Importance of Purity Evaluation and the Potential of Quantitative ^{1}H NMR as a Purity Assay. *J. Med. Chem.*, **2014**, 57, pp 9220–9231 [DOI: 10.1021/jm500734a]

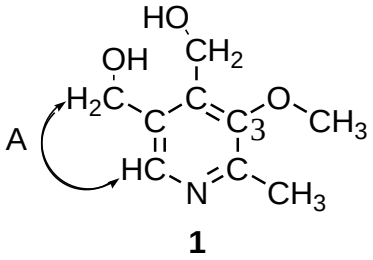
^{1}H and 2D NMR Data of Compound 1. As reported in the Experimental section, the ^{1}H NMR spectra of the compound **1** matched previously published data. The 1D ^{1}H and 2D NMR data of Compound **1** can further distinguish it from other isomers. Due to the presence of intermolecular hydrogen bonding of the OH group to DMSO-*d*$_6$, there should be 2 methylene groups appearing as doublets in the ^{1}H NMR spectrum. Moreover, the proton from pyridine ring (H-6) should exhibit long-rang coupling with the neighboring protons of the methylene (H-5') group on the pyridine ring in a ^{1}H, ^{1}H -COSY spectrum.

1D ^{1}H NMR Spectrum of Compound 1. Chemical shifts are expressed in ppm (δ) with reference to the residual DMSO signal (δ 2.500 ppm) relative to TMS (δ 0.000 ppm). Acquisition and processing parameters for quantitative ^{1}H NMR spectra that were used: pulse program 90 degree excitation pulse (zgig); 96 k time domain data; spectral width (sw) 30 ppm; acquisition time (aq) 4.0 s; a relaxation delay (d1) 60 s; number of scans (ns) 8; receiver gain (rg) 256. All processing was performed with MestReNova v9.0.1 (Mestrelab Research). Line resolution was improved by application of Lorentzian-Gaussian (LG) window functions (LB -0.5, GB 1.0) with zero-filling 256 K, prior to Fourier transformation of the FID data, baseline correct using a 5th order polynomial function, and phase correction was performed manually.

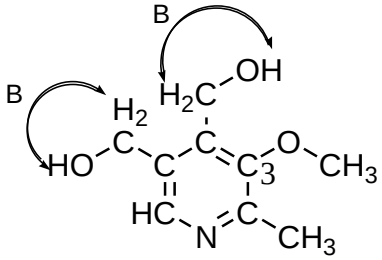


2D ^{1}H, ^{1}H -COSY Spectrum of Compound 1. The 2D ^{1}H, ^{1}H -COSY spectrum was acquired in the absolute value (magnitude) mode. Acquisition and processing parameters for COSY spectrum that were used: spectral width (sw) 7 ppm in each dimension; acquisition time 2 s; a relaxation delay (d1) 1 s; number of scans (ns) 4. All processing was performed with MestReNova v9.0.1 (Mestrelab Research).





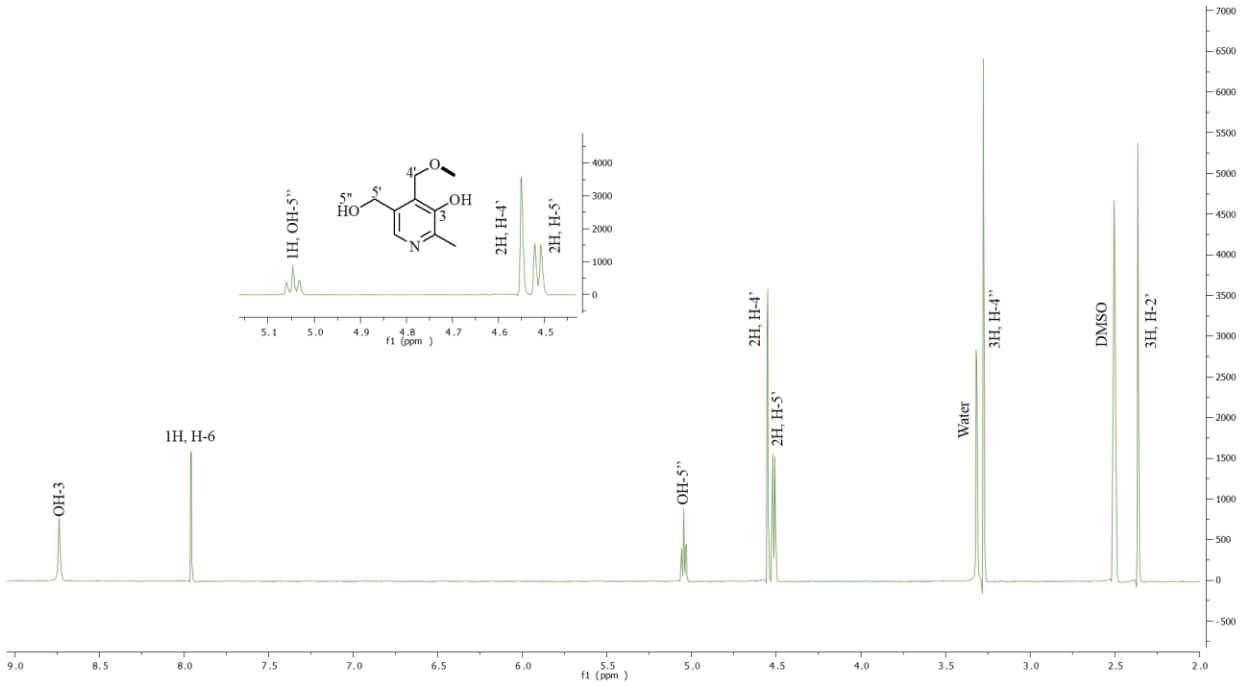
1



1

^{1}H, ^{13}C and 2D NMR Data of Compound 2. All data is provided in the Supporting Information (SI) of the previous study {Liu, 2014 #4}. Compound **2** was identified as ginkgotoxin in DMSO-*d*$_6$.

1D ^{1}H NMR Spectrum of Compound 2 {Liu, 2014 #4} - Due to the presence of intermolecular hydrogen bonding of the OH group to DMSO-*d*$_6$, there should be 1 methylene group appearing as doublets in the ^{1}H NMR spectrum.



^{1}H NMR Spectrum using $CDCl_3$ and HR-ESI-MS of Compound 1, 2 and 3

Characterization of the three isomers was performed by High Resolution-Electro Spray Ionization Mass Spectrometry (HR-ESI-MS; calibrant: reserpine) on a Waters Synapt Mass Spectrometer (Waters, Milford, MA, USA) and NMR performed by BRUKER DPX-400 NMR spectrometer (Bruker, Billerica, MA, USA) in $CDCl_3$.

3-*O*-methylpyridoxine (**1**, yield = 77%): ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.143 (s, 1H, H-6), 4.814 (s, 2H, H-5), 4.744 (s, 2H, H-4), 3.825 (s, 3H, H-4), 2.521 (s, 3H, H-2); HR-ESI-MS m/z 184.08954 $[MH]^+$ (calculated for $C_9H_{13}NO_3$, 183.08956).

4'-*O*-methylpyridoxine (**2**, yield = 57%): ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.773 (s, 1H, H-6), 4.895 (s, 2H, H-5), 4.569 (s, 2H, H-4), 3.518 (s, 3H, H-4), 2.442 (s, 3H, H-2); HR-ESI-MS m/z 184.08954 $[MH]^+$ (calculated for $C_9H_{13}NO_3$, 183.08956). The ^{1}H NMR spectrum was consistent with previously published data.

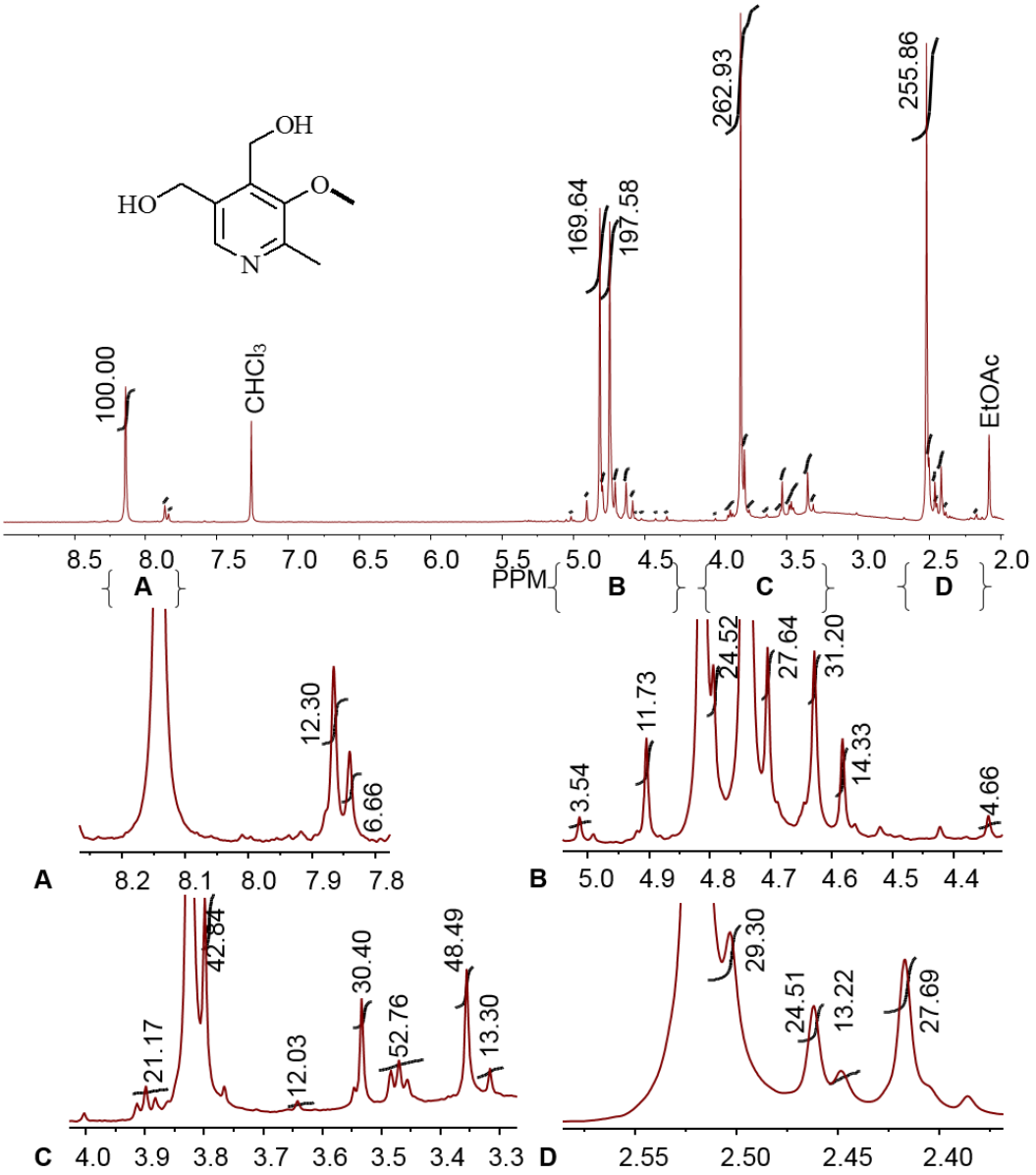
5'-*O*-methylpyridoxine (**3**, yield = 61%): ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.850 (s, 1H, H-6), 5.036 (s, 2H, H-5), 4.344 (s, 2H, H-4), 3.307 (s, 3H, H-4), 2.460 (s, 3H, H-2); HR-ESI-MS m/z 184.08954 $[MH]^+$ (calculated for $C_9H_{13}NO_3$, 183.08956).

S2-S14 Purification analysis by quantitative ^{1}H NMR

As can be seen in Fig. 2, the ^{1}H NMR signals of the three regioisomers do not overlap. This enabled the ready recognition of impurities. For example, the qHNMR spectrum shows that the silica gel purified product **1** contains a small amount of **2**. Visual inspection of the qHNMR spectra of the isomers **1**-**3** purified by silica gel column chromatography or by precipitation (Fig. 2A) readily revealed the presence of relatively larger amounts of impurities when compared to the CCC purified products (Fig. 2B). A common synthetic impurity is the presence of starting material in the product as can be seen in the spectrum of **2** purified by silica gel chromatography. Full assessment of the qHNMR spectral data provides quantitative information based on the mole ratios of the proton signals associated with the components. Definitive identification of impurities such as isomers, starting material, and residual solvents is readily achieved. In this study, qHNMR analysis indicated that CCC achieved the purification of **1**, **2** and **3** at 92%, 98% and 97% purity, respectively.

S2. ^{1}H NMR of compound **1** purified by NP silica gel column chromatography.

The regions denoted by A, B, C, and D correspond to the regions of the ^{1}H NMR spectrum which pertain to (A) aryl protons, (B) methylene protons, (C) methoxy groups, and (D) aryl methyl groups of the pyridoxine derivatives, respectively.



S3. Quantitative ^{1}H NMR purity calculation for compound **1** purified by NP silica gel column chromatography.

The following spreadsheet summarizes the assignments, integral values, and the calculations performed in order to determine the molar-based and mass-based purity values.

QUANTITATIVE ANALYSIS

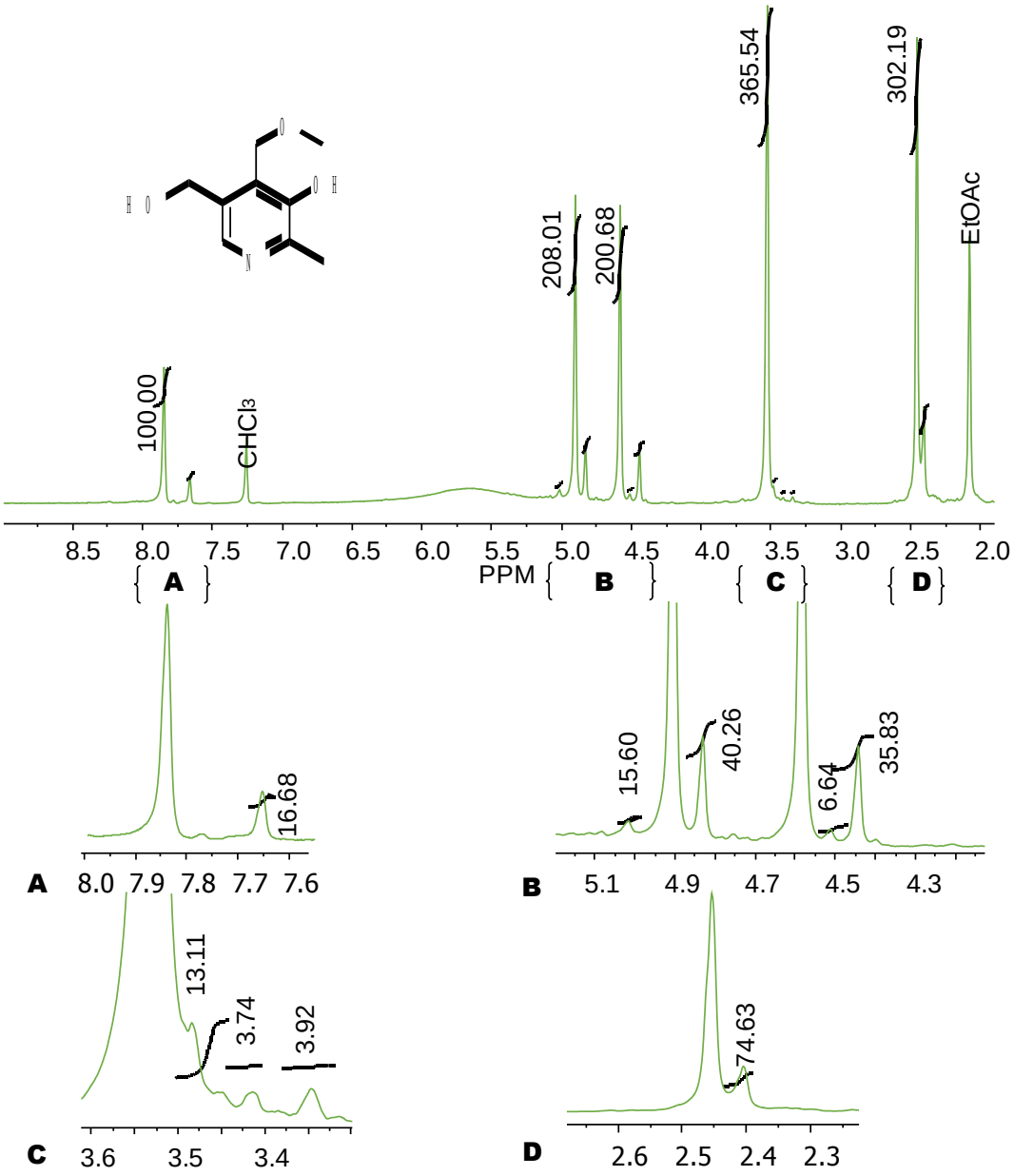
100% CDCl3

COMPOUND	MW	δH [ppm]	Integral	# H	Integral for 1H	molar Average	Mass ratio	mass Average	
1	183	8.15	100	1	100.00	91.63	1.000	91.63	A
		4.82	169.64	2	84.82				B
		4.75	197.58	2	98.79				B
		3.83	262.93	3	87.64				C
		2.53	260.64	3	86.88				D
2	183	7.86	12.30	1	12.30	9.75	1.000	9.75	A
		4.90	11.73	2	5.87				B
		4.79	24.52	2	12.26				B
		3.52	30.40	3	10.13				C
		2.51	24.51	3	8.17				D
3	183	7.86	6.66	1	6.66	3.92	1.000	3.92	A
		5.06	3.54	2	1.77				B
		4.35	4.66	2	2.33				B
		3.30	13.30	3	4.43				C
		2.46	13.22	3	4.41				D
congener 1	197.00	4.71	27.64	2	13.82	11.50	1.077	12.38	B
		4.63	31.20	2	15.60				B
		3.79	42.84	3	14.28				C
		3.60	12.03	3	4.01				C
		2.50	29.31	3	9.77				D
congener 2	197.00	4.56	14.33	2	7.17	8.97	1.077	9.66	B
		3.55	10.00	3	3.33				C
		3.35	48.49	3	16.16				C
		2.41	27.69	3	9.23				D
unknown 1	88.00	3.90	21.70	2	10.85	18.62	0.481	8.95	?
		3.47	52.76	2	26.38				?

MOLAR (%mol/mol)		**MASS (%w/w)**	
	91.63	**Pure compound:**	**91.63**
Sum of 1H Average for Impurities:	**52.75**	**Sum of 1H Average for Impurities:**	**44.65**
	144.38	**Sum:**	**136.28**
% PURITY	**63.5%**	**% PURITY**	**67.2%**
% IMPURITY	**36.54%**	**% IMPURITY**	**32.77%**

S4. ^{1}H NMR of compound **2** purified by NP silica gel column chromatography.

The regions denoted by A, B, C, and D correspond to the regions of the ^{1}H NMR spectrum which pertain to (A) aryl protons, (B) methylene protons, (C) methoxy groups, and (D) aryl methyl groups of the pyridoxine derivatives, respectively.



S5. Quantitative ^{1}H NMR purity calculation for compound **2** purified by NP silica gel column chromatography.

The following spreadsheet summarizes the assignments, integral values, and calculations performed in order to determine the molar-based and mass-based purity values.

QUANTITATIVE ANALYSIS

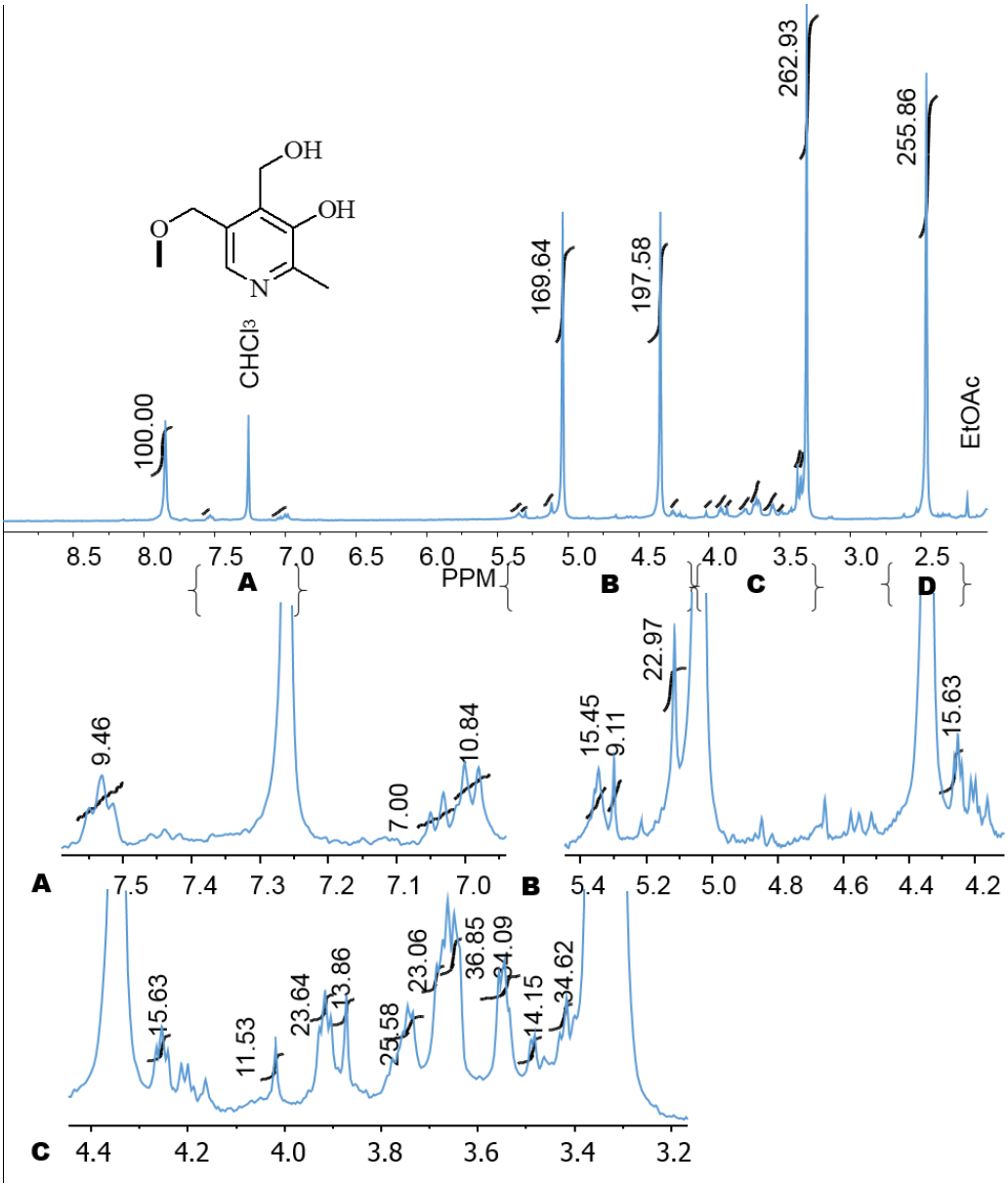
100% CDCl3 **INTEGRATION ON ACD**

						molar		*mass*	
COMPOUND	**MW**	**δH [ppm]**	**Integral**	**# H**	**Integral for 1H**	**Average**	**Mass ratio**	**Average**	
2	183	7.86	100	1	100.00	104.75	1.000	104.75	A
		4.92	201.66	2	100.83				B
		4.60	200.68	2	100.34				B
		3.54	365.54	3	121.85				C
		2.47	302.19	3	100.73				D
pyridoxine	169	7.68	16.68	1	16.68	19.90	0.923	18.38	A
		4.85	40.26	2	20.13				B
		4.46	35.83	2	17.92				B
		2.42	74.63	3	24.88				D
congener	197	5.08	15.60	2	7.80	3.61	1.077	3.88	B
		4.53	6.64	2	3.32				B
		3.50	13.11	3	4.37				C
		3.43	3.74	3	1.25				C
		3.36	3.92	3	1.31				C

MOLAR (%mol/mol)		**MASS (%w/w)**	
	104.75		**104.75**
Sum of 1H Average for Impurities:	**23.51**		**22.26**
Sum:	**128.26**	**Sum:**	**127.01**
% PURITY	**81.7%**	**% PURITY**	**82.5%**
% IMPURITY	**18.33%**	**% IMPURITY**	**17.53%**

S6. ^{1}H NMR of compound **3** purified by precipitation.

The regions denoted by A, B, C, and D correspond to the regions of the ^{1}H NMR spectrum which pertain to (A) aryl protons, (B) methylene protons, and (C) methoxy groups of the pyridoxine derivatives, respectively.



S7. Quantitative ^{1}H NMR purity calculation for compound **3** purified by precipitation.

The following spreadsheet summarizes the assignments, integral values, and calculations performed in order to determine the molar-based and mass-based purity values.

QUANTITATIVE ANALYSIS

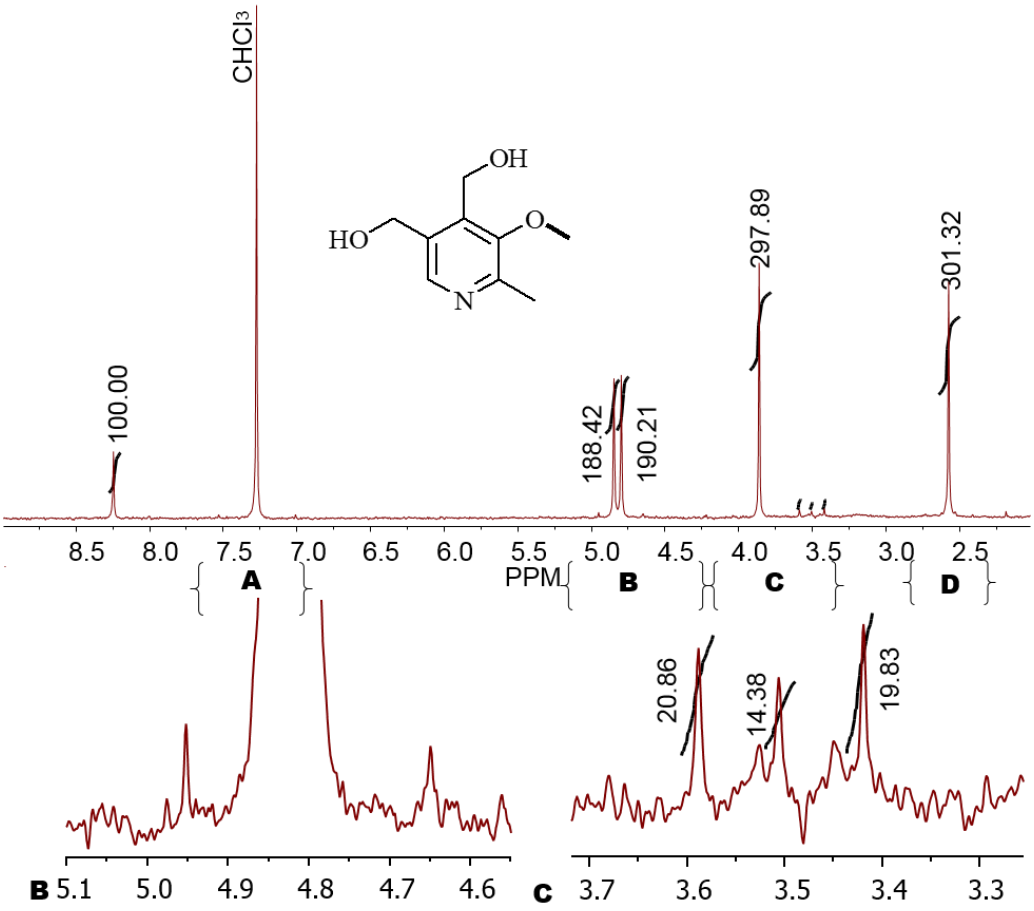
100% CDCl3

						molar		*mass*	
COMPOUND	**MW**	**δH [ppm]**	**Integral**	**# H**	**ntegral for 1H**	**Average**	**Mass ratio**	**Average**	
3	183	7.86	100	1	100.00	91.31	1.000	91.31	A
		5.05	169.64	2	84.82				B
		4.36	197.58	2	98.79				B
		3.32	262.93	3	87.64				C
		2.47	255.86	3	85.29				D
toluene sulfonic acid	172	7.05	7.00	1	7.00	8.92	0.941	8.39	
		7.00	10.84	1	10.84				
		2.34		3					
methanol	32	3.48	14.15	3	4.72	4.72	0.175	0.83	
congener	197	7.54	9.46	1	9.46	8.33	1.077	8.97	A
		5.31	9.11	2	4.56				B
		5.13	22.97	2	11.49				B
		3.89	13.86	3	4.62				C
		3.38	34.62	3	11.54				C
		2.55		3					D
glycerol	92.00	4.27	15.63	1	15.63	15.29	0.503	7.69	
		3.93	23.64	2	11.82				
		3.65	36.85	2	18.43				
acetone	58.00	2.17	36.55	6	6.09	6.09	0.317	1.93	
congener	197.00	5.36	15.45	1	15.45	10.61	1.077	11.42	A
		4.03	11.53	2	5.77				B
		3.76	25.58	2	12.79				B
		3.68	23.06	3	7.69				C
		3.55	34.09	3	11.36				C
		2.55		3					D

MOLAR (%mol/mol)		**MASS (%w/w)**	
	91.31	**Pure compound:**	**91.31**
Sum of 1H Average for Impurities:	**53.96**	**um of 1H Average for Impurities:**	**39.23**
	145.27	**Sum:**	**130.54**
% PURITY	**62.9%**	**% PURITY**	**69.9%**
% IMPURITY	**37.15%**	**% IMPURITY**	**30.05%**

S8. ^{1}H NMR of compound **1** purified by countercurrent separation.

The regions denoted by A, B, C, and D correspond to the regions of the ^{1}H NMR spectrum which pertain to (B) methylene protons and (C) methoxy groups, respectively.



S9. Quantitative ^{1}H NMR purity calculation for compound **1** purified by countercurrent separation.

The following spreadsheet summarizes the assignments, integral values, and calculations performed in order to determine the molar-based and mass-based purity values.

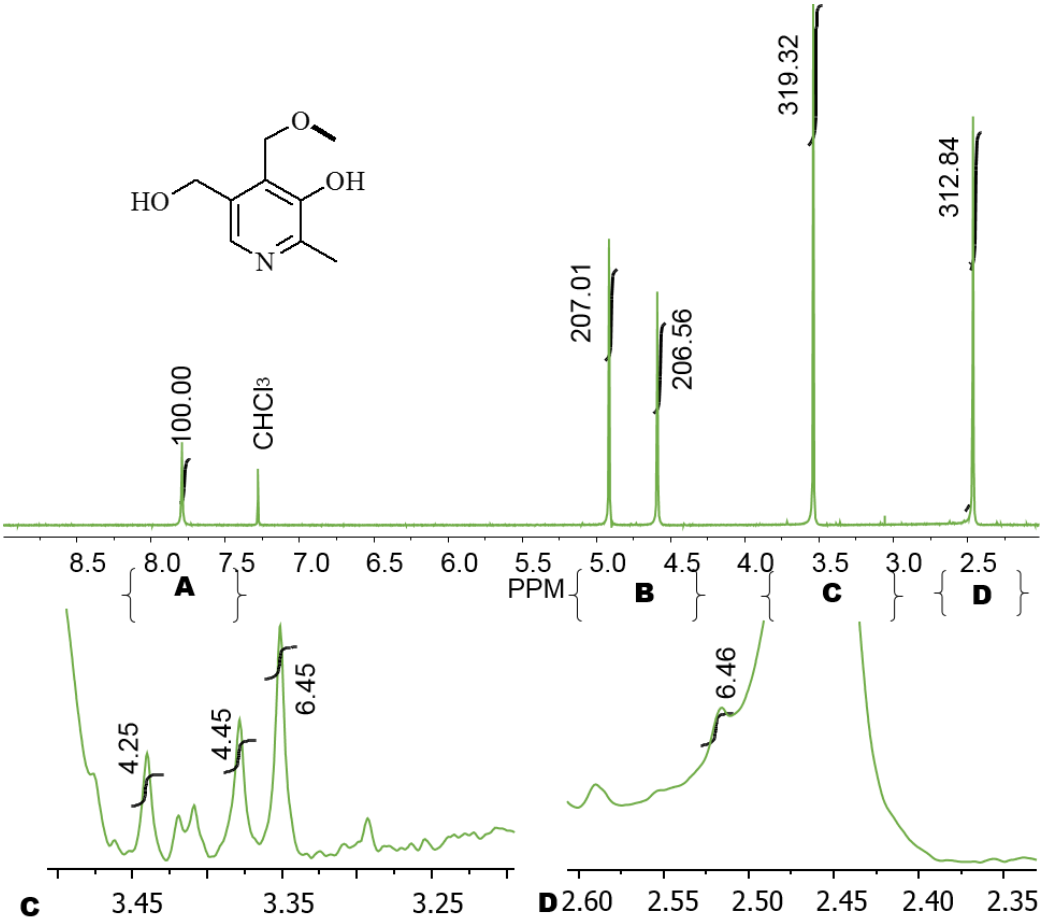
QUANTITATIVE ANALYSIS

COMPOUND	MW	δH [ppm]	Integral	# H	Integral for 1H	molar Average	Mass ratio	mass Average	
1	183	8.14	100	1	100.00	97.81	1.000	97.81	A
		4.81	188.42	2	94.21				B
		4.74	190.21	2	95.11				B
		3.82	297.89	3	99.30				C
		2.52	301.32	3	100.44				D
congener	225.00	3.59	20.86	3	6.95	6.12	1.230	7.52	C
		3.51	14.38	3	4.79				C
		3.42	19.83	3	6.61				C

MOLAR (%mol/mol)			MASS (%w/w)	
Pure compound:	97.81		Pure compound:	97.81
Sum of 1H Average for Impurities:	6.12		Sum of 1H Average for Impurities:	7.52
Sum:	103.93		Sum:	105.33
% PURITY	94.1%	92.4%	% PURITY	92.9%
% IMPURITY	5.81%		% IMPURITY	7.12%

S10. ^{1}H NMR of compound **2** purified by countercurrent separation.

The regions denoted by A, B, C and D correspond to the regions of the ^{1}H NMR spectrum which pertain to (C) methoxy groups and (D) aryl methyl groups of the pyridoxine derivatives, respectively.



S11. Quantitative ^{1}H NMR purity calculation for compound **2** purified by countercurrent separation.

The following spreadsheet summarizes the assignments, integral values, and calculations performed in order to determine the molar-based and mass-based purity values.

QUANTITATIVE ANALYSIS

100% CDCl3 **INTEGRATION ON ACD**

						molar		*mass*	
COMPOUND	**MW**	**δH [ppm]**	**Integral**	**# H**	**tegral for 1H**	**Average**	**Mass ratio**	**Average**	
2	183	7.86	100	1	100.00	103.50	1.000	103.50	A
		4.92	207.01	2	103.51				B
		4.60	206.56	2	103.28				B
		3.54	319.32	3	106.44				C
		2.47	312.84	3	104.28				D
congener	225.00	3.44	4.25	3	1.42	1.80	1.230	2.21	C
		3.38	4.45	3	1.48				C
		3.35	6.45	3	2.15				C
		2.52	6.46	3	2.15				D

MOLAR (%mol/mol)

Pure compound:	**103.50**
	1.80
Sum:	**105.30**

% PURITY	**98.3%**
% IMPURITY	**1.71%**

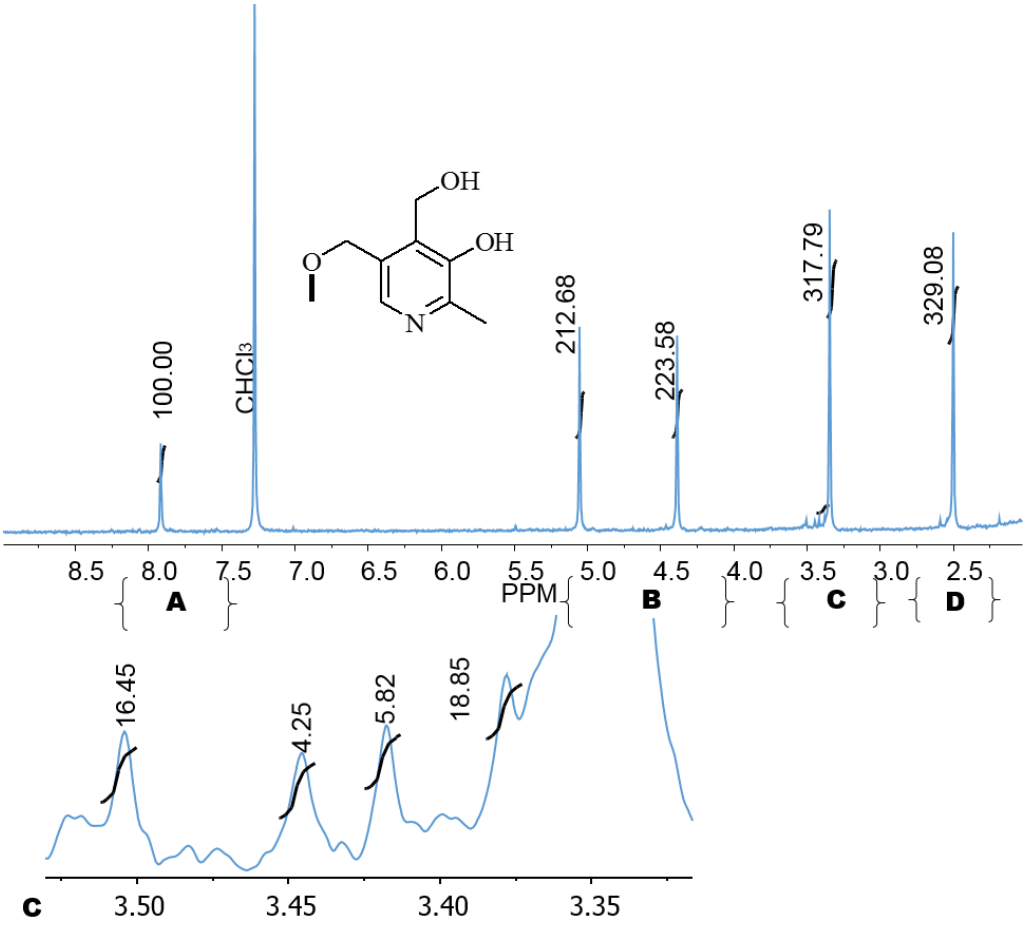
MASS (%w/w)

Pure compound:	**103.50**
Sum of 1H Average for Impurities:	**2.21**
Sum:	**105.72**

% PURITY	**97.9%**
% IMPURITY	**2.09%**

S12. ^{1}H NMR of compound **3** purified by countercurrent separation.

The region denoted by A, B, C and D corresponds to the region of the ^{1}H NMR spectrum which pertains to (C) methoxy groups.



S13. Quantitative ^{1}H NMR purity calculation for compound **3** purified by countercurrent separation.

The following spreadsheet summarizes the assignments, integral values, and calculations performed in order to determine the molar-based and mass-based purity values.

QUANTITATIVE ANALYSIS

						molar		*mass*
COMPOUND	**MW**	**δH [ppm]**	**Integral**	**# H**	**Integral for 1H**	**Average**	**Mass ratio**	**Average**
3	183	7.85	100	1	100.00	106.75	1.000	106.75
		5.03	212.68	2	106.34			
		4.34	223.58	2	111.79			
		3.31	317.79	3	105.93			
		2.46	329.08	3	109.69			
congener	225.00	3.50	16.45	3	5.48	5.65	1.230	6.95
		3.42	15.85	3	5.28			
		3.38	18.58	3	6.19			
methanol	32.00	3.45	4.25	3	1.42	1.42	0.175	0.25

MOLAR (%mol/mol)		**MASS (%w/w)**	
Pure compound:	**106.75**	**Pure compound:**	**106.75**
Sum of 1H Average for Impurities:	**7.07**	**Sum of 1H Average for Impurities:**	**7.20**
Sum:	**113.82**	**Sum:**	**113.95**
% PURITY	**93.8%**	**% PURITY**	**93.7%**
% IMPURITY	**6.71%**	**% IMPURITY**	**6.81%**