Supporting Information

Accurate Identification of Unknown and Known Metabolic Mixture Components by Combining 3D NMR and Fourier Transform Ion Cyclotron Resonance Tandem Mass Spectrometry

Cheng Wang,[†] Lidong He,[⊥,] Da-Wei Li,[‡] Lei Bruschweiler-Li,[‡] Alan G. Marshall,[*,⊥,¶] and Rafael Brüschweiler[*,†,‡,§]

[†]Department of Chemistry and Biochemistry, [‡]Campus Chemical Instrument Center, and [§]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, United States

[⊥]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida, 32306, United States

[¶]Ion Cyclotron Resonance Program, The National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, United States

Content

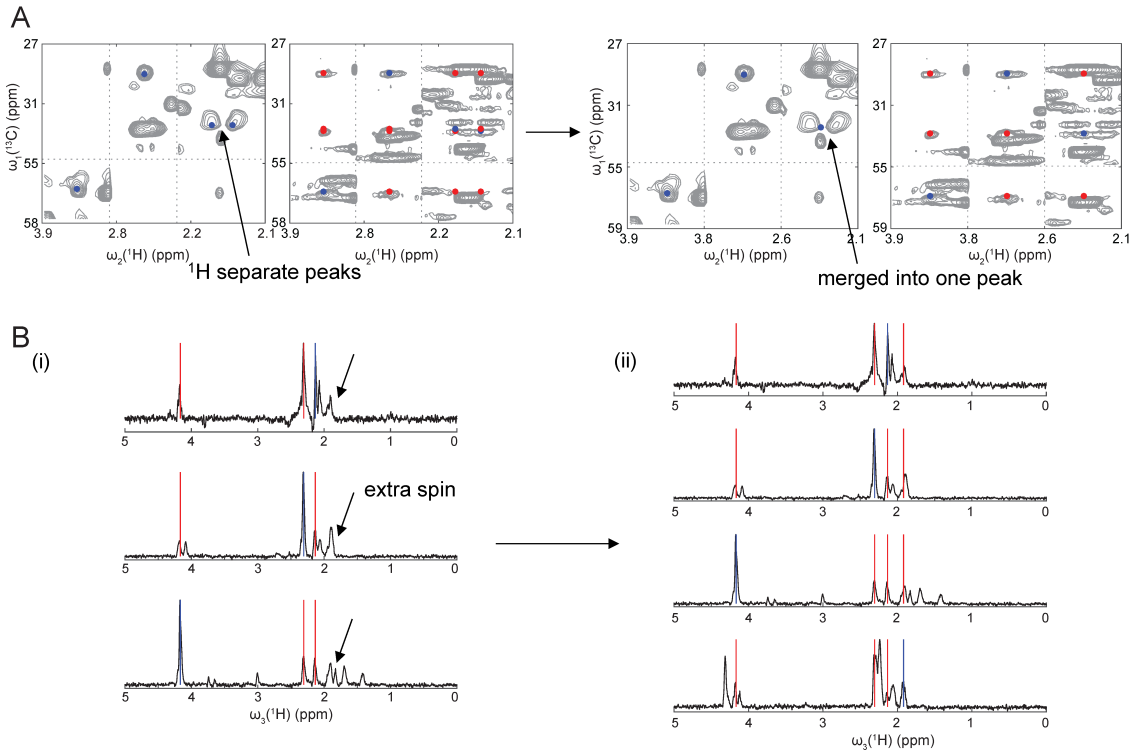


Figure S1. Illustration of the spin system refinement procedure based on 2D HSQC, 2D TOCSY, and 2D/3D HSQC-TOCSY NMR spectra. The following steps are depicted: a) merging of two separate ^{1}H peaks into one peak; b) identification of extra spins (indicated by arrows in (i)) by 1D ω_3 (^{1}H) trace comparison.

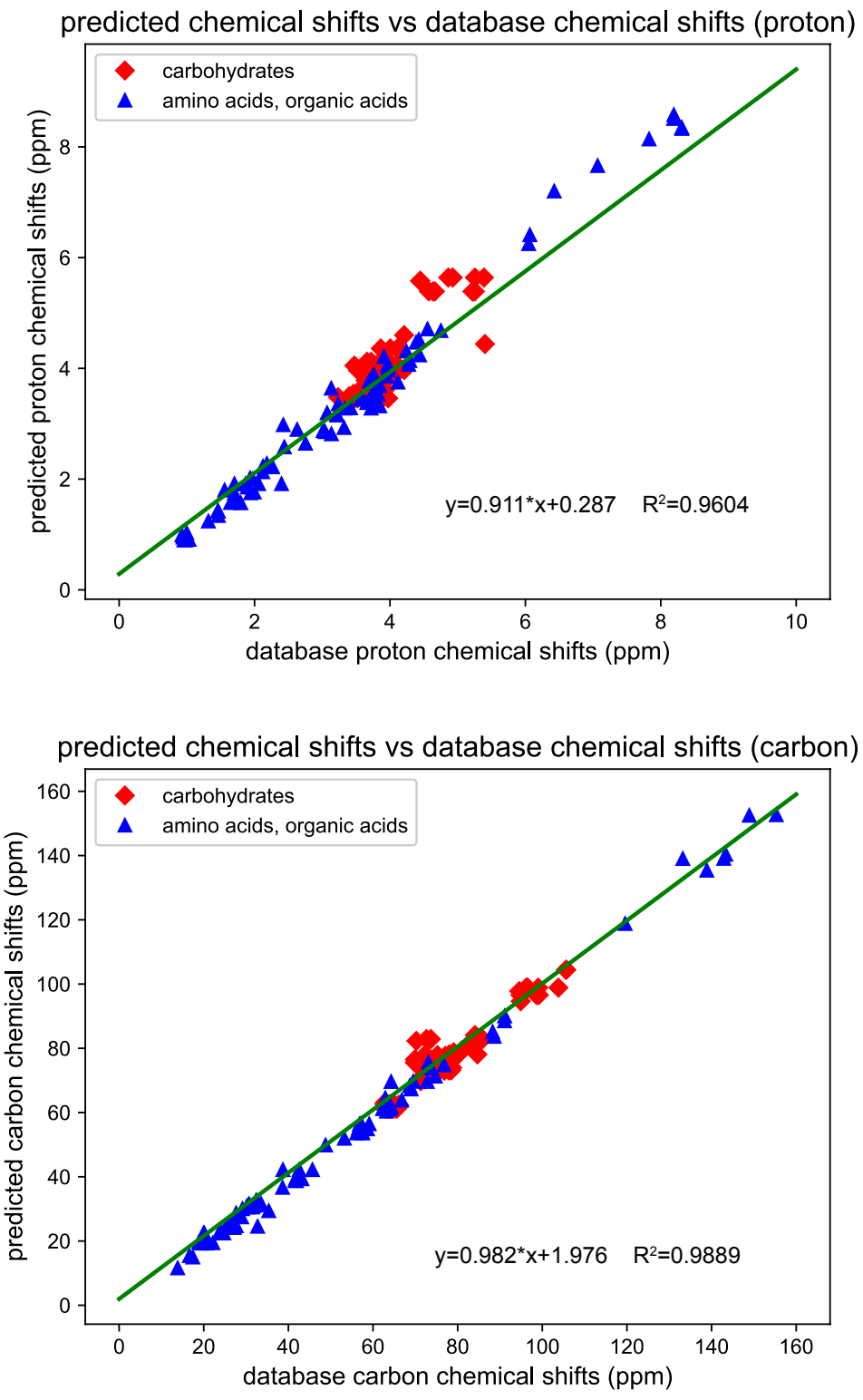


Figure S2. Predicted chemical shifts compared with their experimental chemical shifts of 25-compound model mixture. The RMSD between predicted and experimental chemical shifts of proton and carbon chemical shifts are 0.292 ppm and 2.903 ppm.

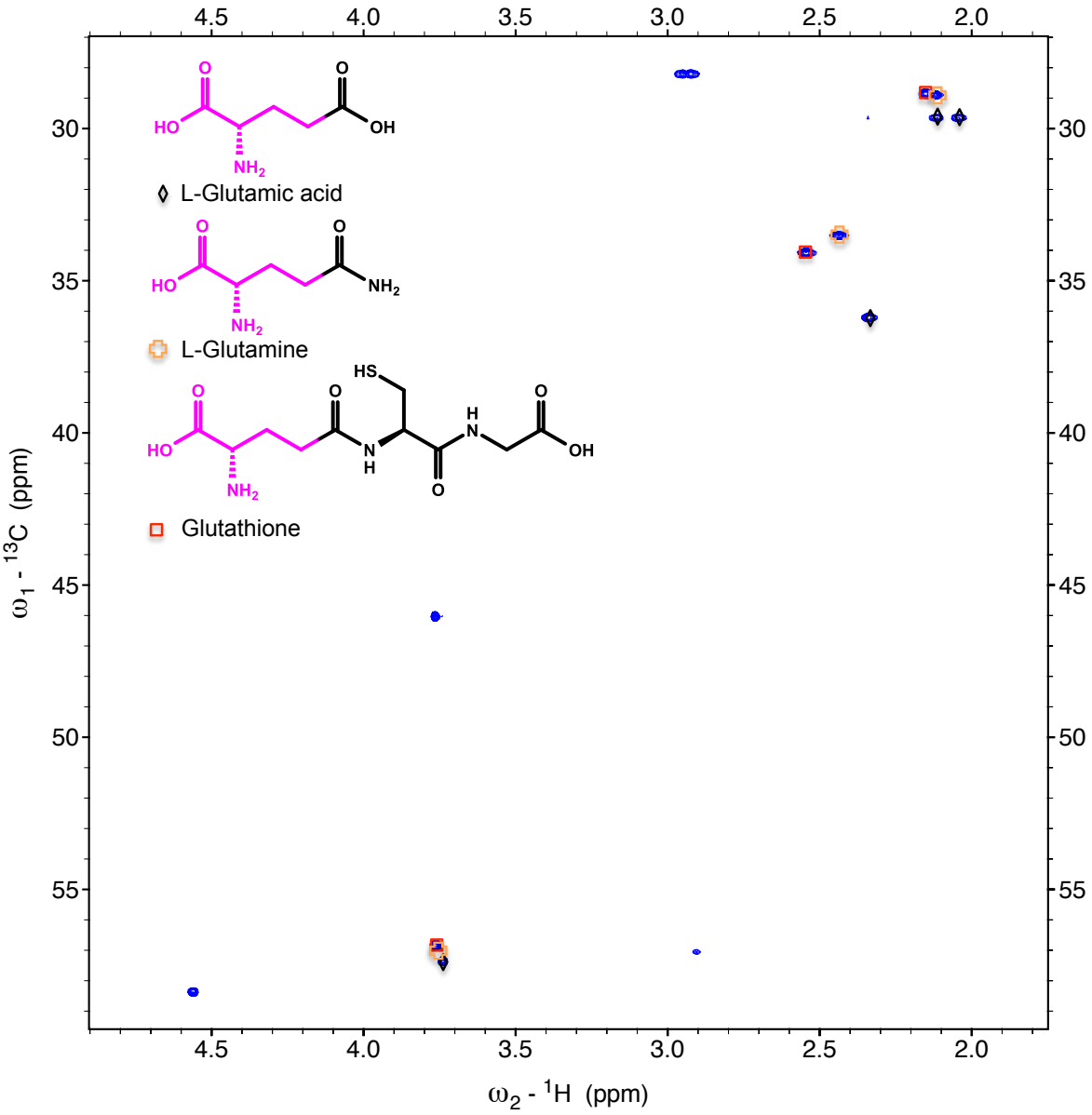


Figure S3. 2D ^{13}C-^{1}H HSQC of L-glutamine, glutathione, and L-glutamic acid mixture (peaks with the common motif ($HOOCCH(NH_2)CH_2CH_2CONH$-) are highlighted by symbols). The spectrum illustrates the similarity of chemical shifts of identical fragments (colored in magenta) that are part of different molecules. Only cross-peaks that belong to the motif are labeled.

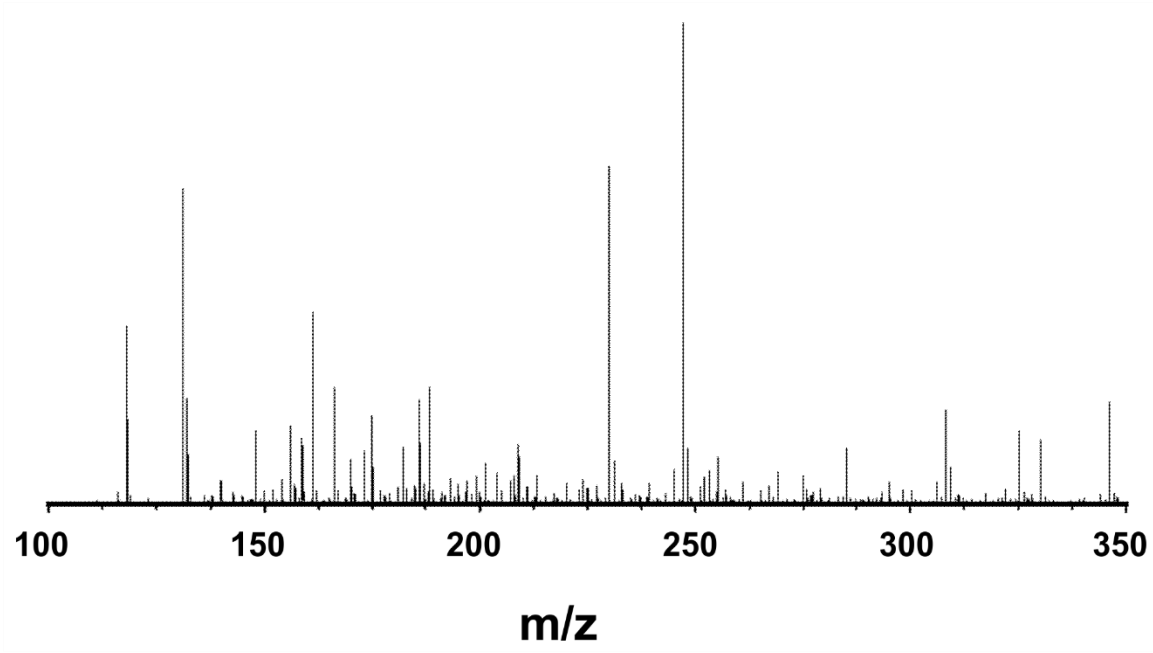


Figure S4. Positive electrospray ionization 9.4 T FT-ICR broadband mass spectrum of *E. coli* cell lysate.

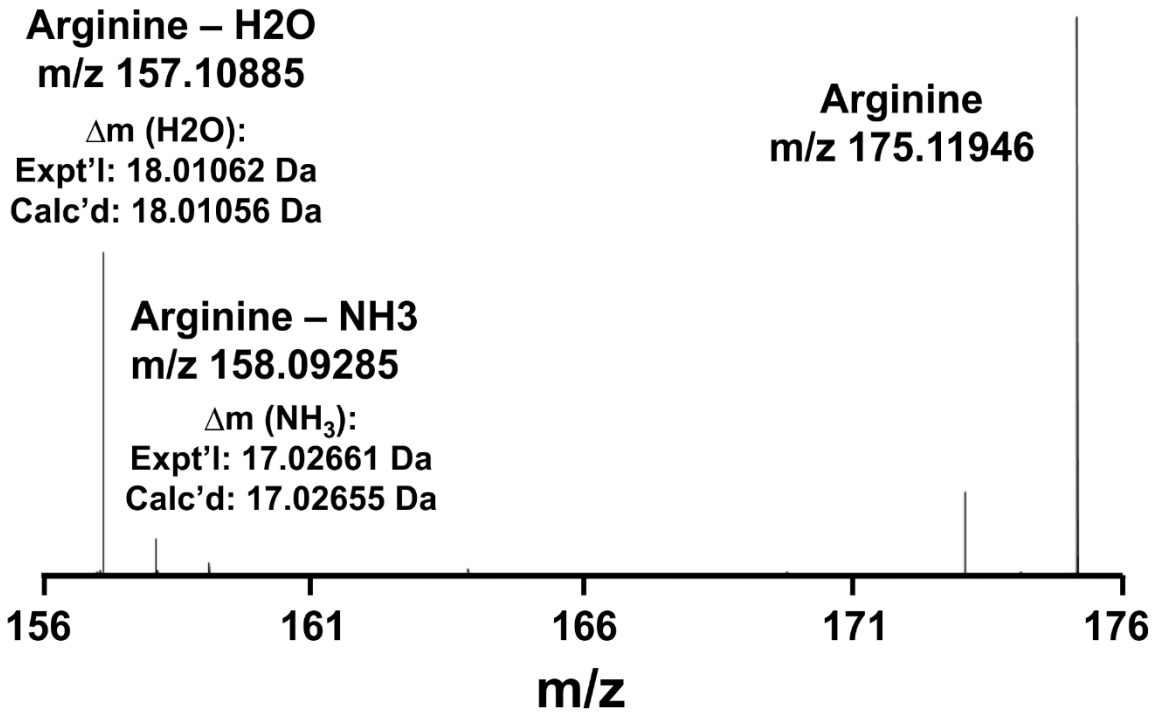


Figure S5. Infrared multiphoton dissociation positive electrospray ionization 9.4 T FT-ICR product ion mass spectrum of arginine.

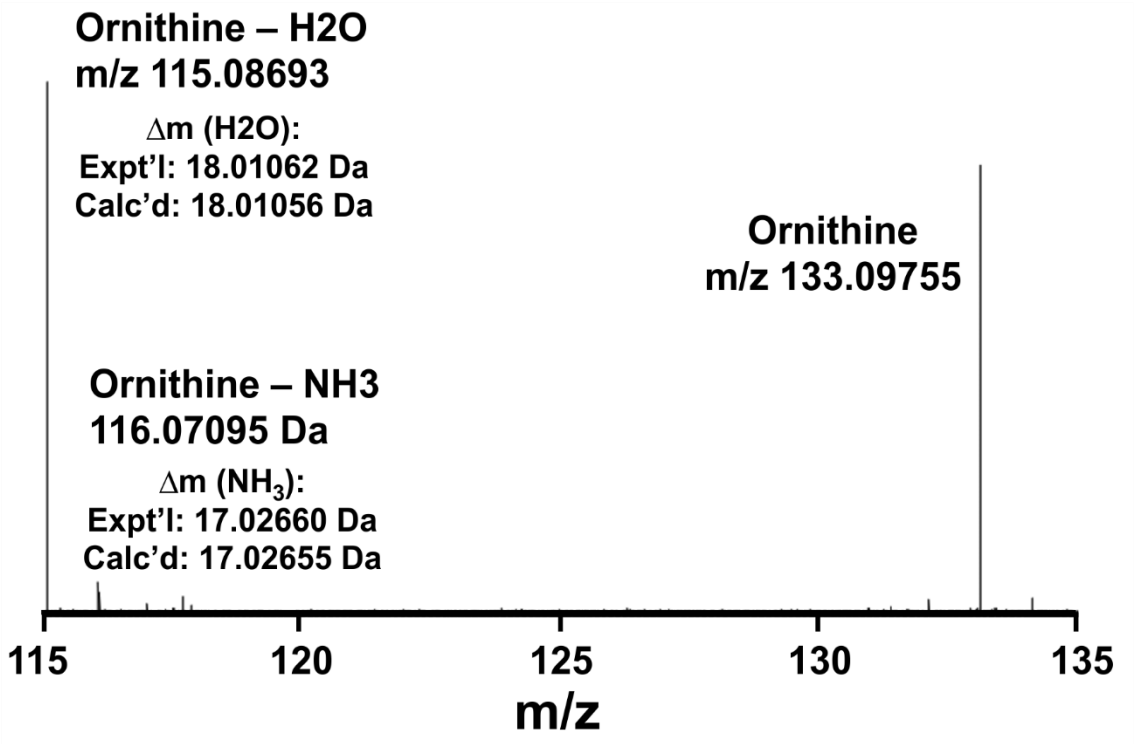


Figure S6. Infrared multiphoton dissociation positive electrospray ionization 9.4 T FT-ICR product ion mass spectrum of ornithine.

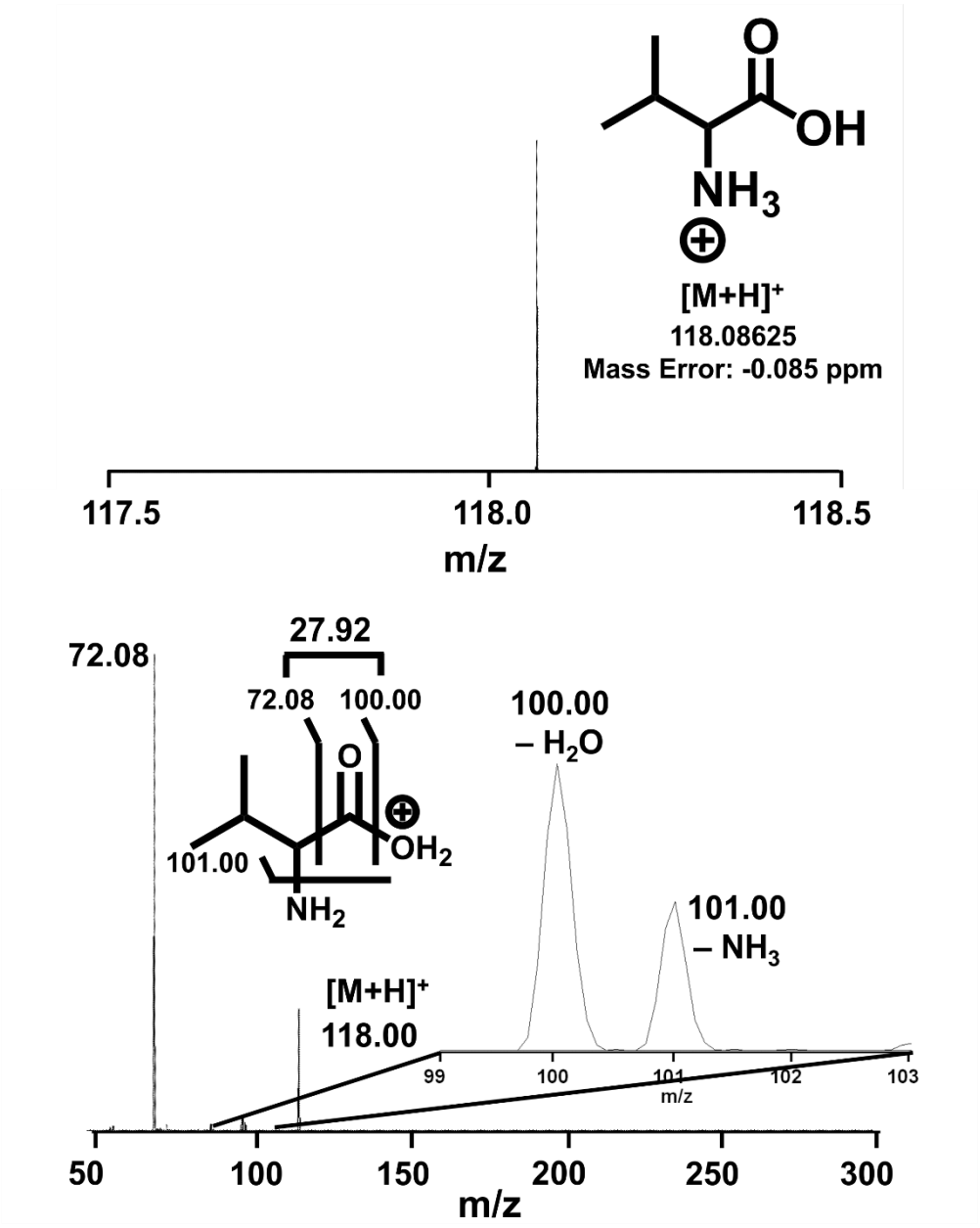


Figure S7. Collision-induced dissociation (normalized collision energy 22) Velos Pro product ion mass spectrum of valine.

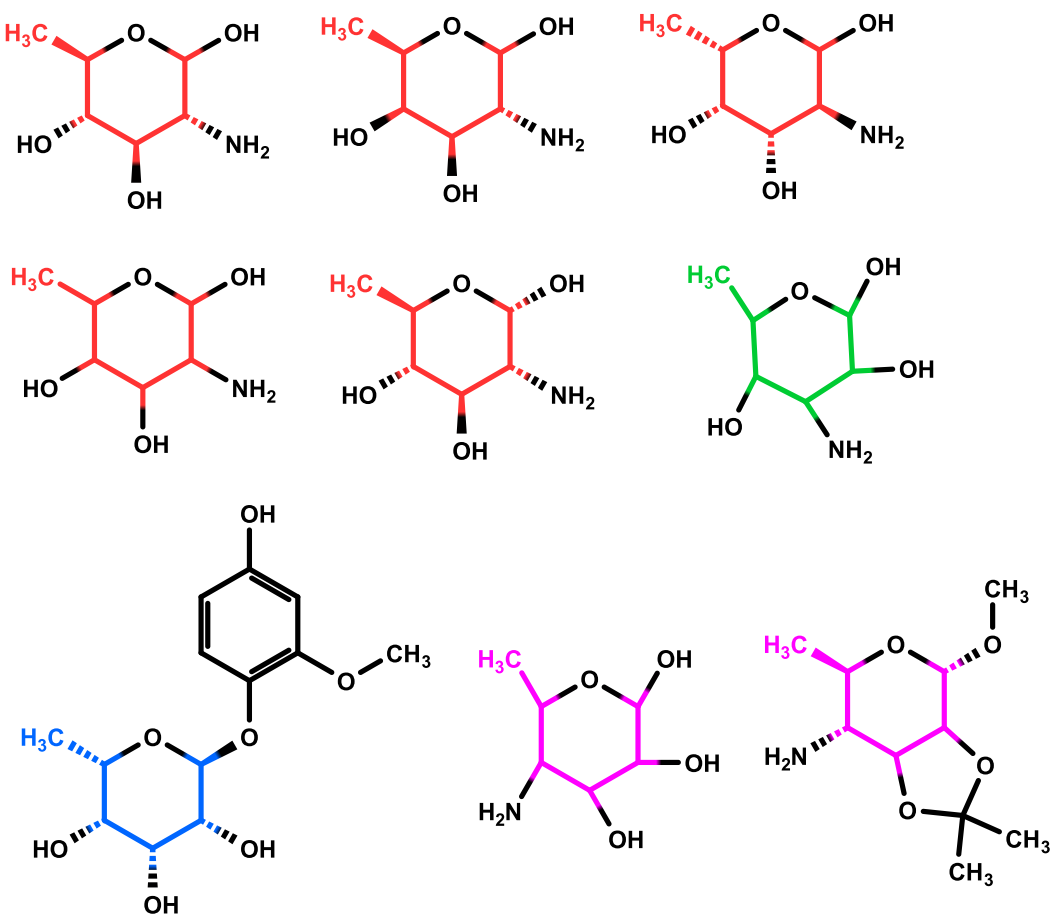


Figure S8. Motif identification of compound candidates for the spin system of an unknown compound from *E. coli* cell lysate. Four different motifs were identified, highlighted in red, green, blue and magenta, which are consistent with the NMR data.

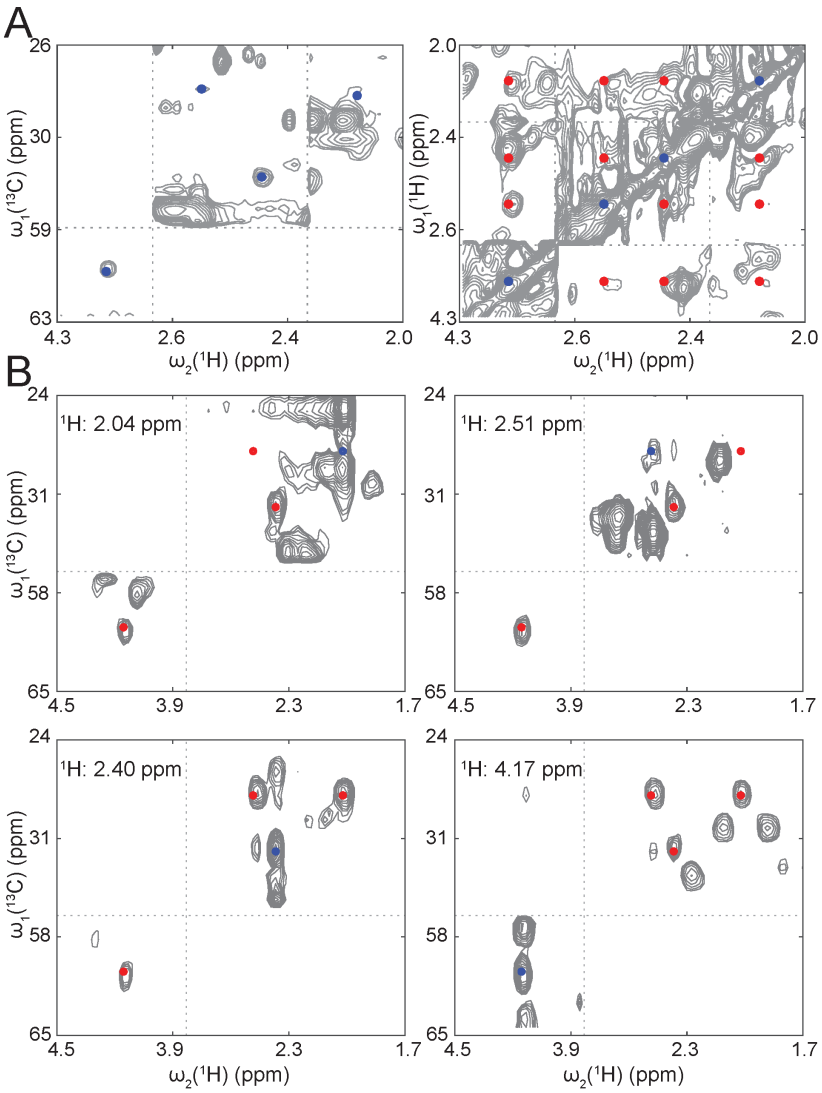


Figure S9. Spin system of pyroglutamic acid spin system extracted from 3D HSQC-TOCSY and confirmed by 2D TOCSY and 2D HSQC-TOCSY. a) Cross-peaks of pyroglutamic acid shown in the 2D HSQC (blue peaks) and 2D TOCSY spectra (blue and red). b) Four cross-peaks of pyroglutamic acid shown in multiple 2D slices of 3D HSQC-TOCSY. Cross-peaks (δ_H, δ_C) at (2.04, 28.0) ppm and (2.51, 28.0) ppm are two separate peaks of the same CH_2 group.

3D HSQC-TOCSY mixing time optimization

To obtain a high quality 3D HSQC-TOCSY spectrum with maximal proton-proton magnetization transfer efficiency, optimization of the isotropic mixing time is critical for spin-system assignment and the sequential spin-system extraction process.[1,2] Because 3D HSQC-TOCSY and 2D ^{1}H-^{1}H TOCSY share the same pulse sequence element for isotropic mixing, the 25-compound model mixture was used for the 2D ^{1}H-^{1}H TOCSY experiment to obtain the optimal mixing time with good overall magnetization transfer efficiency between resonances at extreme ends of a spin system. A series of TOCSY experiments was performed with 40 ms, 80 ms, 100 ms, 110 ms, 120 ms and 130 ms mixing times. The effect of the TOCSY mixing time on proton-proton magnetization transfer efficiency is shown in Figure S10. Based on these results a mixing time of 120 ms was chosen for both 2D ^{1}H-^{1}H TOCSY and 3D HSQC-TOCSY experiment.

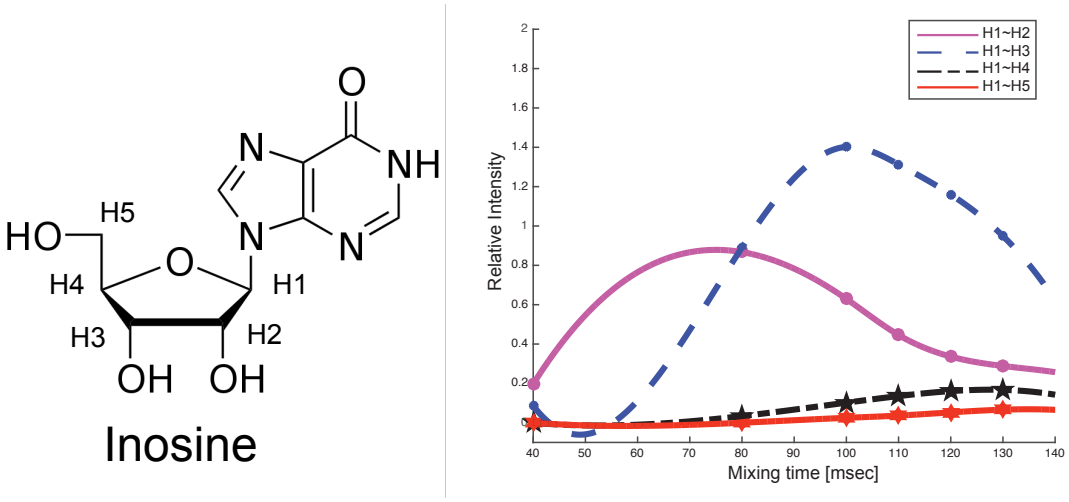


Figure S10. Effect of TOCSY-mixing time on magnetization transfer efficiency of inosine. The figure depicts magnetization transfer across the whole spin system as a function of the mixing time.

Table S1. Metabolites of *E. coli* cell lysate identified by COLMARm web server and database

Compound identified by COLMARm	Formula	^{13}C RMSD (ppm)	^{1}H RMSD (ppm)
L-glutamine	C5H10N2O3	0.050	0.019
L-valine	C5H11NO2	0.070	0.013
maltose	C12H22O11	0.130	0.010
cellobiose	C12H22O11	0.120	0.010
N-acetyl-putrescine	C6H14N2O	0.180	0.016
L-glutamic acid	C5H9NO4	0.070	0.016
D-glucose	C6H12O6	0.130	0.011
spermidine	C7H19N3	0.080	0.016
L-phenylalanine	C9H11NO2	0.130	0.015
L-tyrosine	C9H11NO3	0.180	0.020
N-α-acetyl-L-lysine	C8H16N2O3	0.160	0.016
L-glutathione-reduced	C10H17N3O6S	0.090	0.008
L-methionine	C5H11NO2S	0.050	0.010
adenosine	C10H13N5O4	0.070	0.011
inosine	C10H12N4O5	0.170	0.019
L-proline	C5H9NO2	0.090	0.015
leucine	C6H13NO2	0.060	0.012
pyridoxamine-5-phosphate-1	C8H13N2O5P	0.170	0.014
guanosine	C10H13N5O5	0.130	0.005
DL-α-glycerol-phosphoric acid	C3H9O6P	0.110	0.010
N-acetyl-L-glutamine	C7H12N2O4	0.230	0.017
D-glucuronic acid	C6H10O7	0.200	0.011
methyl-uridine	C10H14N2O6	0.160	0.018
dAMP	C10H14N5O6P	0.150	0.014
dTMP	C10H15N2O8P	0.140	0.019
UMP	C9H13N2O9P	0.180	0.013
ethanolamine	C2H7NO	0.130	0.006
uracil	C4H4N2O2	0.220	0.021
N-acetyl-L-alanine	C5H9NO3	0.090	0.012
malic acid	C4H6O5	0.080	0.007
rhamnose	C6H12O5	0.110	0.013
N1-acetyl-spermine	C12H28N4O	0.090	0.010
cysteine-glutathione-disulfide	C13H22N4O8S2	0.180	0.011
dTTP	C10H17N2O14P3	0.170	0.022
UDP-GlcNAc	C17H27N3O17P2	0.110	0.005

ADP	C10H15N5O10P2	0.140	0.016
ATP	C10H16N5O13P3	0.110	0.019
ITP	C10H15N4O14P3	0.160	0.022
UDP-glucose	C15H24N2O17P2	0.080	0.018
NAD	C21H27N7O14P2	0.060	0.018
NADP+	C21H29N7O17P3	0.170	0.017

Table S2. Metabolites identified in *E. coli* cell lysate and verified by COLMAR web server and SUMMIT MS/NMR. (Highlighted metabolites in **green** were consistent with identification by COLMARm (based on HSQC for query and TOCSY + HSQC-TOCSY for validation); highlighted metabolites in **blue** were identified based on the COLMAR HSQC database alone.

Compound	Rank	Total hits (RMSD<5.0 ppm)	Mass error (ppm)	RMSD (ppm)	RMSD (ppm) (COLMAR)
L-glutamine	1	6268	0.14	1.27	0.14
L-valine	5	5089	0.13	0.99	0.10
maltose	14	641	0.08	1.52	0.12
cellobiose	29	641	0.08	1.73	0.11
N-acetyl-putrescine	45	1697	0.08	1.87	0.17
L-glutamic acid	79	5549	0.14	1.58	0.12
D-glucose	170	659	-0.07	2.17	0.12
spermidine	213	1697	0.14	2.82	0.13
L-phenylalanine	230	7685	0.06	2.06	0.14
L-tyrosine	251	7685	0.11	2.07	0.19
N-alpha-acetyl-L-lysine	278	576	0.11	2.94	0.16
L-glutathione-reduced	313	6268	0.10	2.55	0.09
L-methionine	417	4982	0.13	2.49	0.08
D-xylose	1	96	0.09	1.85	0.47
D-sorbose	3	285	-0.07	1.59	0.16
lactose	6	744	0.08	1.71	0.18
N-alpha-acetyl-ornithine	7	1552	0.18	2.18	0.18
AMP	19	36	0.12	2.69	0.76
D-ribose	26	142	0.09	2.51	0.19
D-galactose	33	659	-0.07	1.74	0.64
2-pyrrolidinone-5-carboxylate	275	5549	0.08	2.02	0.30
lysine	281	576	0.20	2.95	0.12
isovalerylglutamic acid	289	3773	0.13	1.8	0.14
S-adenosyl-L-homocysteine	332	4982	0.07	2.3	0.62
gamma-glutamylcysteine	367	6268	-0.05	2.55	0.19
D-glucosamine	9	144	0.12	2.46	0.17
xylulose	10	1629	0.09	1.69	0.74

L-arabinose	25	96	0.09	2.58	0.71
D-tagatose	26	285	-0.07	1.84	0.65
muramic acid	32	81	0.01	3.09	0.41
melibiose	36	99	0.08	2.22	0.28
isomaltose	38	99	0.08	2.41	0.19
D-salicin	80	744	-0.06	2.29	0.25
sucrose	136	168	0.08	3.26	0.29
beta-gentiobiose	345	956	0.08	2.48	0.40
D-trehalose	378	1139	0.08	2.42	0.23
N-acetyl-lactosamine	427	1079	0.09	2.45	0.48

Table S3. Quantum-chemical calculations of NMR chemical shifts for two selected compounds, together with matching results from experimental spectra. The quantum-chemical calculations were performed according to our MOSS-DFT protocol published recently.[3]

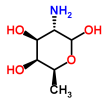
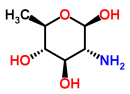
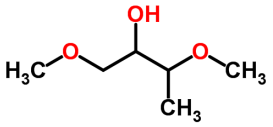
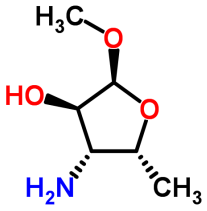
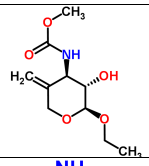
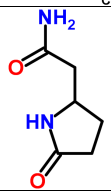
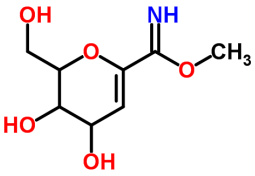
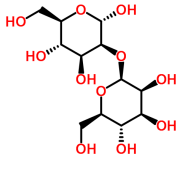
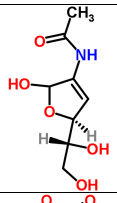
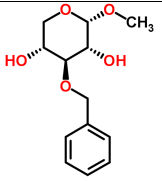
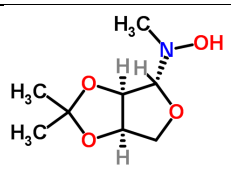
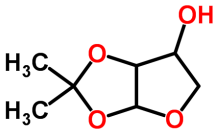
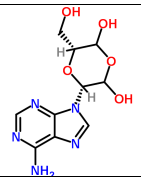
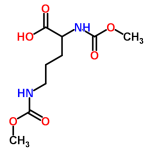
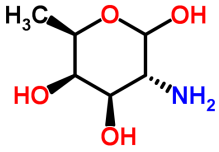
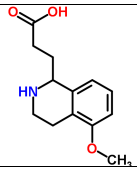
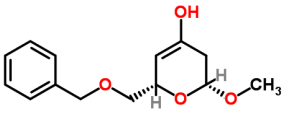
Compound	Peak index	Predicted ^{1}H c.s. (ppm)[a]	Predicted ^{13}C c.s. (ppm)[a]	Expt. ^{1}H c.s. (ppm)	Expt. ^{13}C c.s. (ppm)
L-fucosamine 	1	1.342	17.506	1.167	19.542
	2	3.252	60.607	3.618	59.207
	3	3.835	71.816	4.032	70.725
	4	3.893	73.883	3.632	75.944
	5	3.901	72.945	3.767	73.296
	6	4.887	94.099	5.573	98.048
	RMSD (ppm)	2.9335			
6-desoxy-D-glucosamine 	1	1.337	18.311	1.167	19.542
	2	3.207	61.619	3.618	59.207
	3	3.737	73.449	4.032	70.725
	4	3.326	77.943	3.632	75.944
	5	3.565	74.916	3.767	73.296
	6	4.949	93.834	5.573	98.048
	RMSD (ppm)	3.1622			

Table S4. Unknown spin systems identified in *E. coli* cell lysate by 3D HSQC-TOCSY and compound candidates (top hits) returned by SUMMIT.

Unknown spin system peak list (δ_H, δ_C) ppm	Top hit	ChemSpider ID	RMSD (ppm)
(4.093, 73.531) (3.434, 75.502) (3.387, 74.313) (1.315, 19.470)		36208103	1.98
(1.167, 19.542) (3.767, 73.296) (4.032, 70.725) (3.618, 59.207) (5.573, 98.048)		32971157	3.23
(4.523, 55.624) (4.063, 74.956) (3.809, 80.235) (4.864, 100.947)		26596688	2.69
(2.058, 30.448) (2.283, 36.248) (4.305, 56.298) (1.918, 30.569)		32970114	2.58
(3.948, 63.316) (3.481, 71.620) (3.801, 78.104) (3.781, 63.146) (4.858, 105.234)		4324706	2.38
(4.058, 67.096) (4.189, 67.116) (3.592, 77.226) (3.391, 75.683) (3.481, 71.620) (4.643, 105.803)		27471393	2.36
(3.481, 71.620) (3.801, 78.104) (3.864, 63.231) (4.858, 105.234)		29369739	2.81
(4.376, 65.593) (3.714, 75.454) (3.703, 76.472) (4.464, 65.664) (4.841, 105.758)		10264285	3.96

(3.864, 63.231) (3.607, 84.515) (3.772, 78.845) (4.858, 105.234)		9334888	3.03
(3.781, 63.146) (3.607, 84.515) (4.858, 105.234) (3.772, 78.845)		263432	4.23
(3.607, 84.515) (3.948, 63.316) (5.418, 94.247)		4981062	3.68
(1.413, 24.783) (3.002, 41.766) (4.166, 57.181) (1.766, 33.713)		9314583	2.03
(1.167, 19.542) (3.767, 73.296) (4.032, 70.725) (3.618, 59.207) (5.573, 98.048) (3.598, 75.523)		9919658	3.22
(2.705, 29.164) (3.864, 56.460) (2.162, 32.873)		37527314	0.86
(3.817, 70.012) (3.593, 72.692) (5.241, 94.598)		17253887	2.39

References

(1) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J.; Rance, M., *Protein NMR Spectroscopy (Second Edition): Principles and Practice.* Academic Press, **2007**; pp 280-289.

(2) Braunschweiler, L.; Ernst, R.R., Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Mag. Reson.* **1983**, 53, 521-528.

(3) Hoffmann, F.; Li, D. W.; Sebastiani, D.; Brüschweiler, R., Improved quantum chemical NMR chemical shift prediction of metabolites in aqueous solution toward the validation of unknowns. *J. Phys. Chem. A* **2017**, 121, 3071-3078.