Supporting Information

Customized Metabolomics Database for the Analysis of NMR ^{1}H-^{1}H TOCSY and ^{13}C-^{1}H HSQC-TOCSY Spectra of Complex Mixtures

Kerem Bingol , Lei Bruschweiler-Li, Da-Wei Li, and Rafael Brüschweiler*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210

Campus Chemical Instrument Center, The Ohio State University, Columbus, OH 43210, United States

National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306

*To whom correspondence should be addressed:
Rafael Brüschweiler, Ph.D.

Newman & Wolfrom Lab of Chemistry, Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210

E-mail: bruschweiler.1@osu.edu

Materials and Methods

Sample preparation. *E. coli DH5α* cells were cultured at 37 °C, at 250 rpm in M9 minimum medium with glucose (natural abundance, 5g/L) added as sole carbon source. One liter of culture at OD ~3 was centrifuged at 5000x*g* for 20 min at 4 °C, and the cell pellet was resuspended in 50 mL of 50 mM phosphate buffer at pH 7.0. Cell suspension was then subjected to centrifugation for cell pellet collection. The cell pellet was resuspended in 10 mL of ice cold water and exposed to freeze-thaw procedure 3 times. The sample was centrifuged at 20000x*g* at 4°C for 15 min to remove the cell debris. Prechilled methanol and chloroform were sequentially added to the supernatant under vigorous vortex at H_2O:methanol:chloroform ratios of 1:1:1.[S1] The mixture was then left at -20 °C overnight for phase separation. Next, it was centrifuged at 4000x*g* for 20 min at 4 °C, and the clear top hydrophilic phase was collected and subjected to rotary evaporator processing to have the methanol content reduced. Finally, the liquid was lyophilized. The NMR sample was prepared by dissolving the lyophilized material in D_2O and transferred to a 5-mm NMR tube.

NMR experiments and Processing. The 2D ^{1}H-^{1}H TOCSY spectrum was collected with N_1=512 and N_2=1024 complex data points. The spectral widths for the indirect and the direct dimensions were 7702.5 Hz and 7692.3 Hz, respectively. The number of scans per t_1 increment was set to 8. The transmitter frequency offset was 4.7 ppm in both ^{1}H dimensions. 2D ^{13}C-^{1}H HSQC-TOCSY spectrum was collected with N_1=512 and N_2=1024 complex points. The spectral widths for the indirect and the direct dimensions were 29934.5 Hz and 7692.3 Hz, respectively. The number of scans per t_1 increment was set to 64. The transmitter frequency offset were 85 ppm in the ^{13}C dimension and 4.7 ppm in the ^{1}H dimension. The TOCSY mixing times were set to 90 ms for ^{1}H-^{1}H TOCSY and ^{13}C-^{1}H HSQC-TOCSY. The experimental time for ^{1}H-^{1}H TOCSY was 12 hours and for ^{13}C-^{1}H HSQC-TOCSY 36 hours. A 1D ^{1}H NMR spectrum was collected with 16384 complex data points and 16 number of scans. The spectral width and the transmitter frequency offset were set to 11261.3 Hz and 4.7 ppm, respectively. All NMR spectra

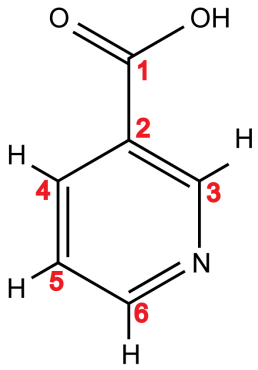
were collected using a cryoprobe at 700 MHz proton frequency at 298 K. The NMR data were zero-filled, Fourier transformed, phase and baseline corrected using NMRPipe[S2] and converted to a Matlab-compatible format for further processing and analysis.

References

[S1] Bingol, K.; Brüschweiler, R. *Anal. Chem.* **2011**, *83*, 7412-7417.

[S2] Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

A

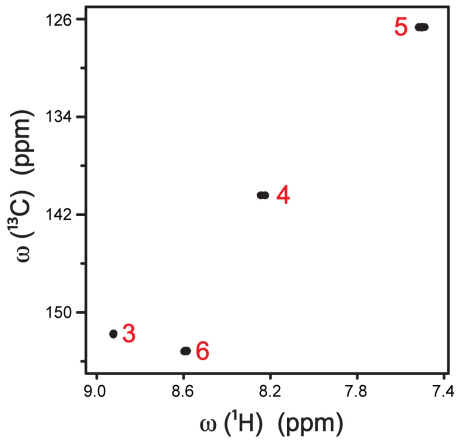


B

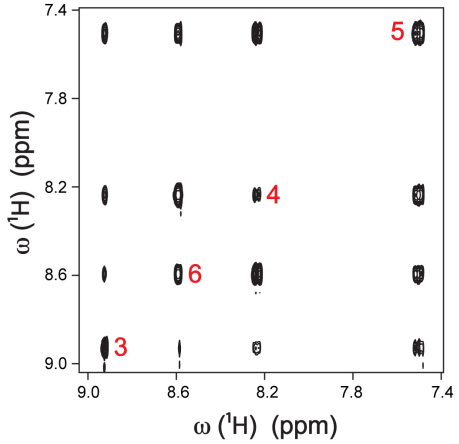
Assignment

	^{13}C ppm	^{1}H ppm
1	176.052	—
2	135.032	—
3	151.726	8.9256
4	140.383	8.2359
5	126.639	7.5047
6	153.140	8.5931

C



D



E

^{1}H (^{13}C)-TOCCATA

^{1}H ppm	^{13}C ppm
8.9256	151.726
8.2359	140.383
7.5047	126.639
8.5931	153.140

F

^{13}C-TOCCATA

^{13}C ppm
176.052
135.032
151.726
140.383
126.639
153.140

Figure S-1. Construction of a new database entry exemplified for nicotinic acid. (A) Carbon atoms in the structure of nicotinic acid are labeled by red numbers. (B) The chemical shift assignments are performed by co-analysis of multiple 1D and 2D NMR spectra, in particular 2D ^{13}C-^{1}H HSQC (C) and 2D ^{1}H-^{1}H TOCSY (D) from the BMRB database. The red cross-peak numbers in these spectra correspond to the positions of the ^{13}C and ^{1}H nuclei in the structure of nicotinic acid (A). The nicotinic acid ^{1}H(^{13}C) TOCCATA data is stored as shown in Panel E, which is different from the corresponding data in the ^{13}C-TOCCATA database shown in Panel F.

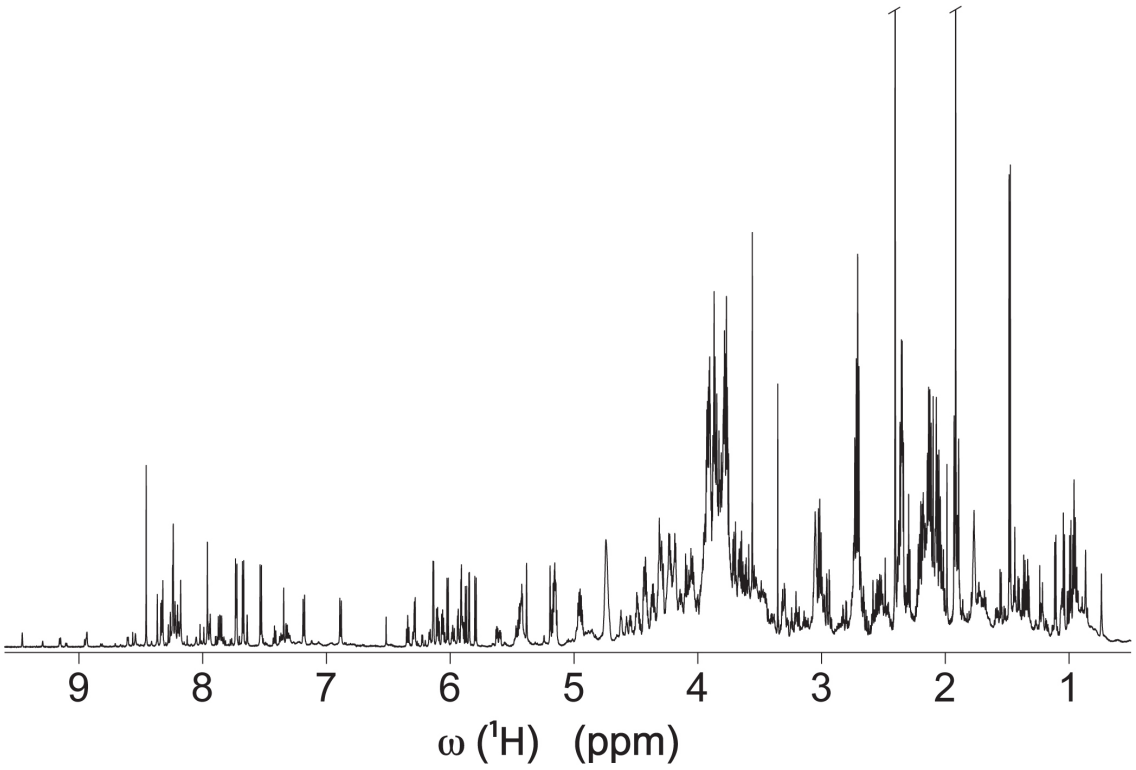


Figure S-2. 1D ^{1}H NMR spectrum of *E. coli* cell lysate used in this study. The spectrum is acquired at 700 MHz proton frequency at 298 K with 16,384 complex data points.

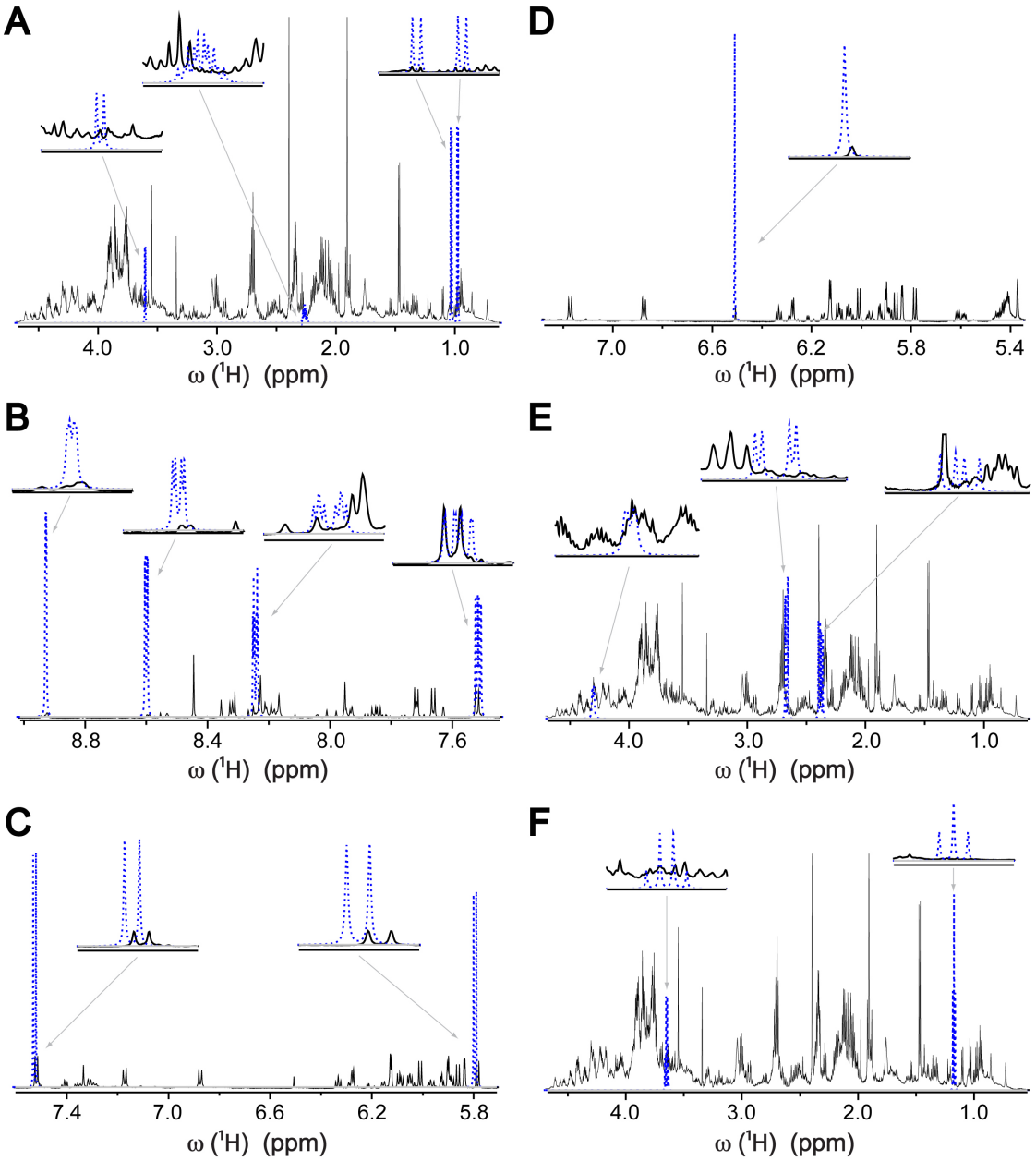


Figure S-3. Identification of metabolites by 1D ^{1}H NMR spectral matching at the example of *E. coli* cell lysate using the Chenomx NMR software. Overlay of 1D ^{1}H NMR spectra of metabolites from Chenomx database (blue) on 1D ^{1}H NMR spectrum of *E. coli* cell lysate (black) shows at least one peak that matches for valine (A), nicotinic acid (B), uracil (C), fumarate (D), but not for malate (E) and ethanol (F).

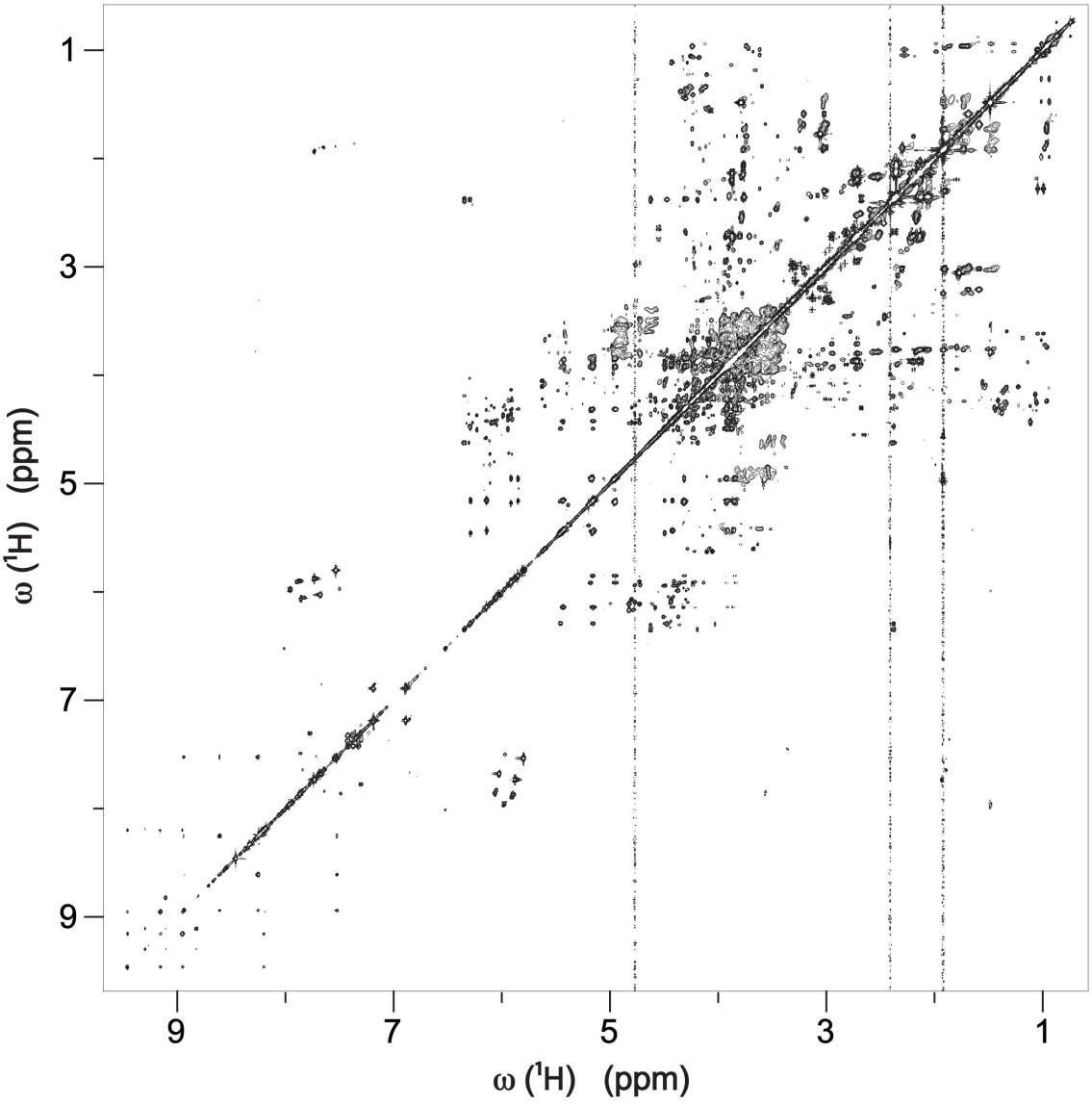


Figure S-4. 2D ^{1}H-^{1}H TOCSY spectrum of *E. coli* cell lysate used in this study. The spectrum is acquired at 700 MHz proton frequency at 298 K with 512 and 1024 complex data points along the indirect and direct dimensions, respectively.

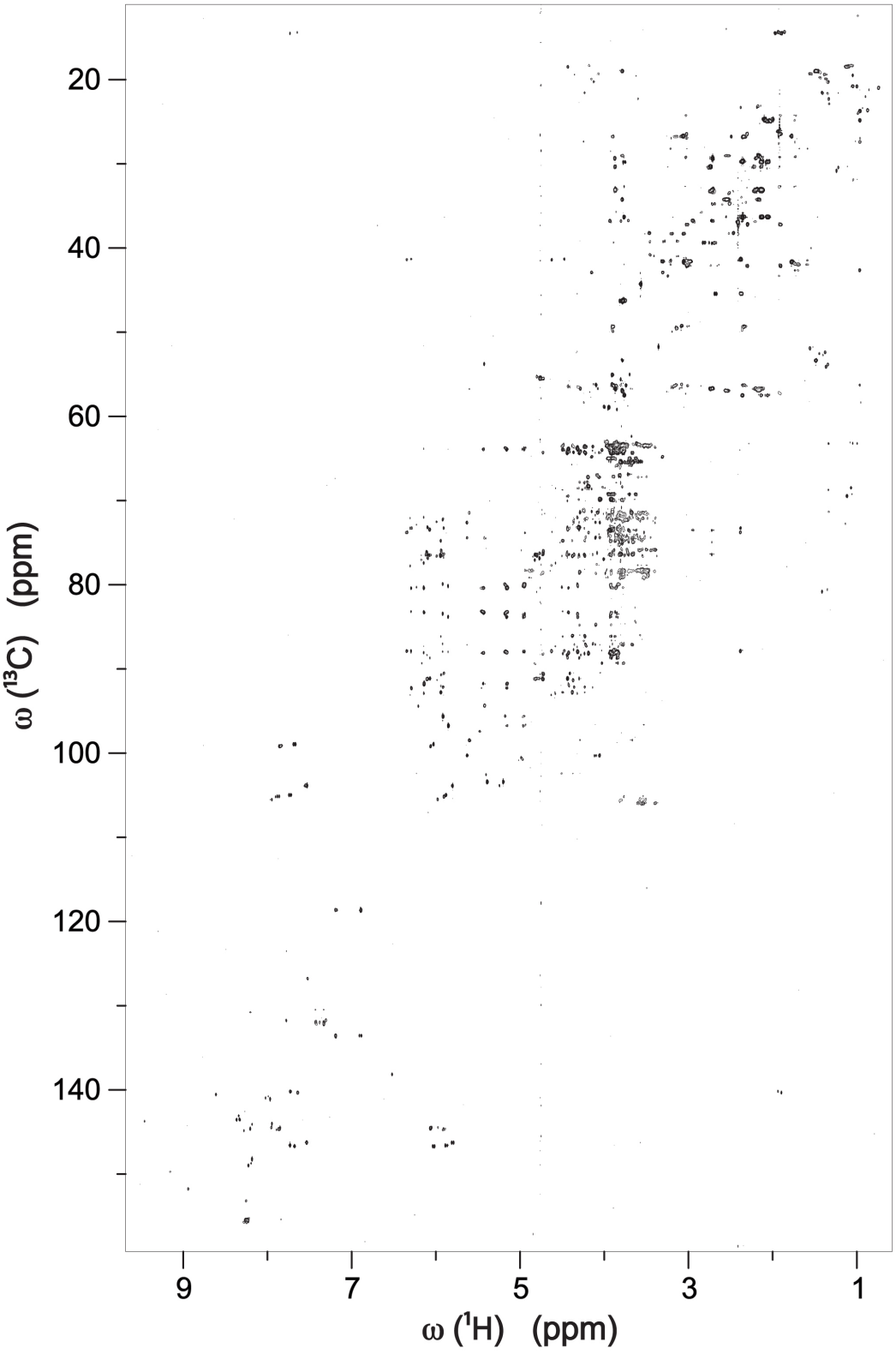


Figure S-5. 2D ^{13}C-^{1}H HSQC-TOCSY spectrum of *E. coli* cell lysate used in this study. The spectrum is acquired at 700 MHz proton frequency at 298 K with 512 and 1024 complex data points along the indirect and direct dimensions, respectively.

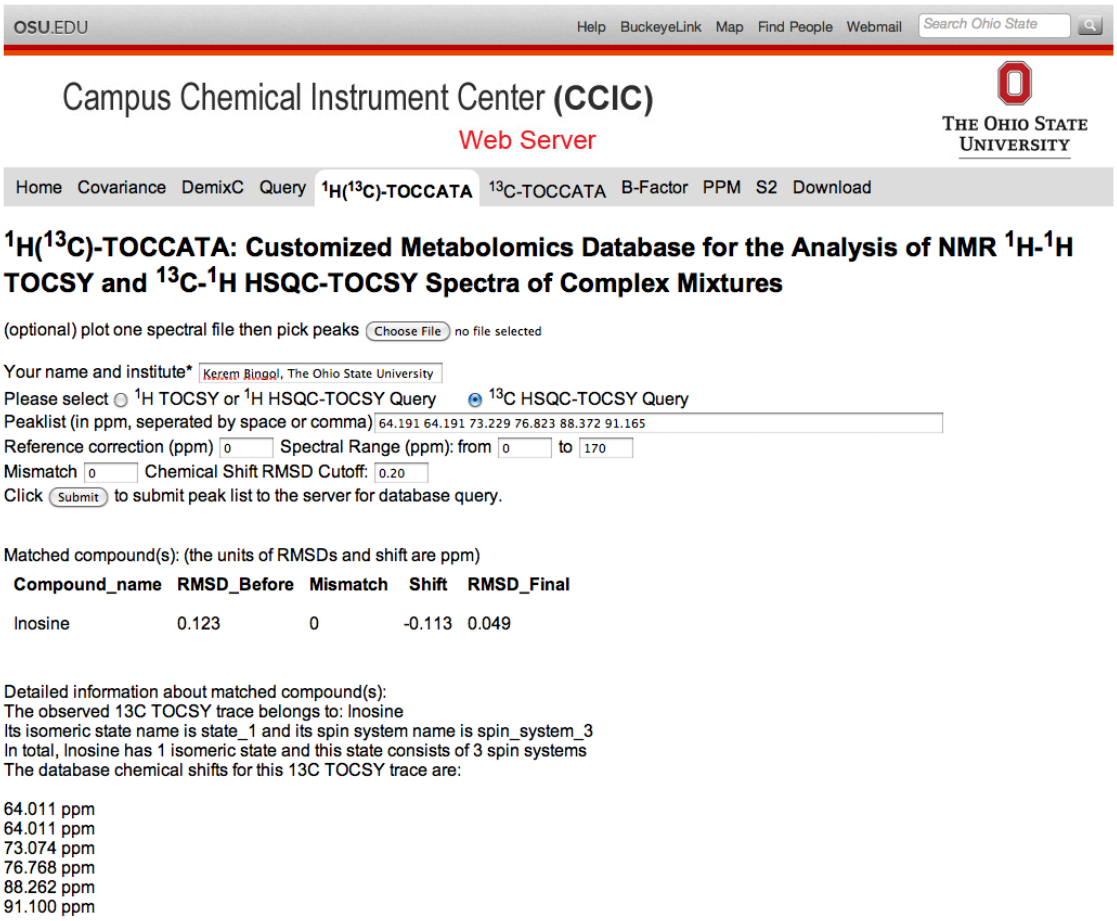


Figure S-6. Screenshot of the ^{1}H(^{13}C)-TOCCATA web server. A peak list of the ^{13}C HSQC-TOCSY trace from 2D ^{13}C-^{1}H HSQC-TOCSY spectrum is queried against the database. Query returns the best matching compound (in this case ribose ring of inosine) with the chemical shift root mean square difference (rmsd) before and after a uniform shift of -0.113 ppm was applied. A mismatch number $M = 0$ indicates that the number of query peaks and database peaks for inosine were the same.

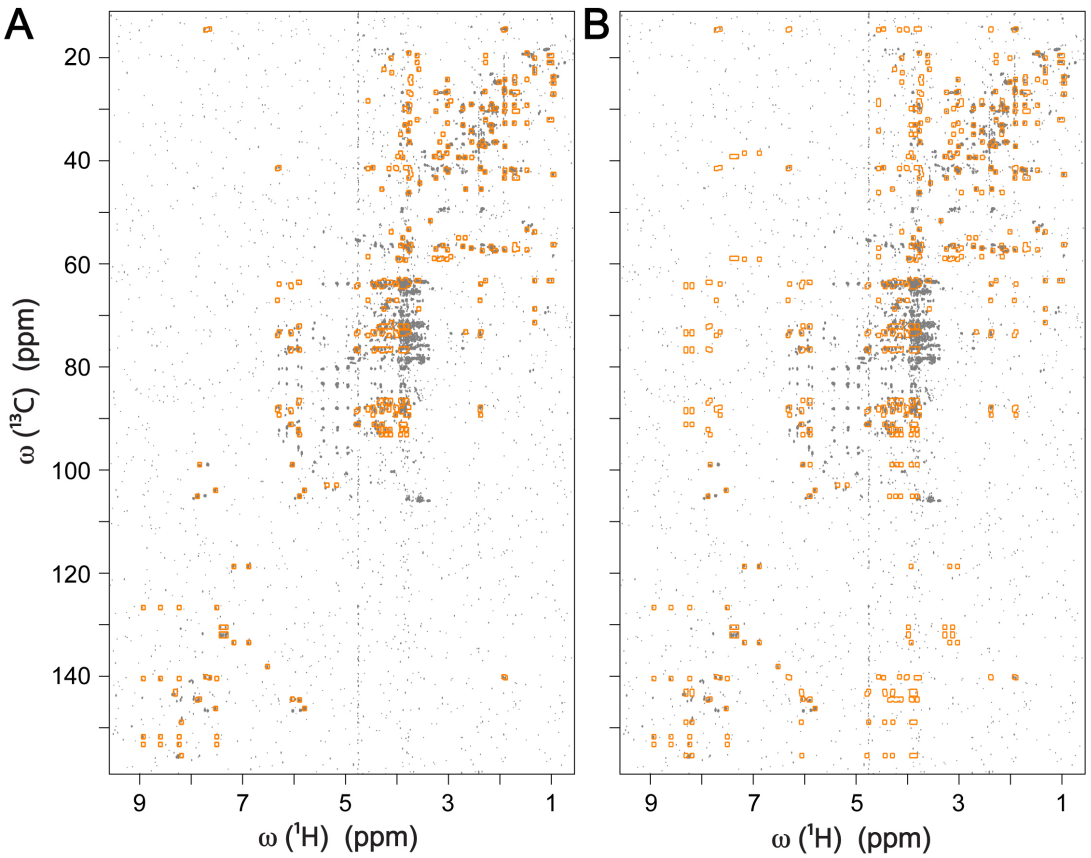


Figure S-7. Overlay of reconstructions of ^{13}C-^{1}H HSQC-TOCSY spectra from databases (orange) with the experimental ^{13}C-^{1}H HSQC-TOCSY spectrum of *E. coli* cell lysate (black). (A) The reconstruction of the ^{13}C-^{1}H HSQC-TOCSY spectrum (orange) is based on spin-system information from ^{1}H(^{13}C)-TOCCATA database. (B) The reconstruction of the ^{13}C-^{1}H HSQC-TOCSY spectrum (orange) is based on entire 1D ^{1}H and 1D ^{13}C NMR spectra (from BMRB database). A list with all 33 metabolites used for reconstruction in both panels is given in Table 2.

Table S-1. List of compounds currently contained in the $^{1}H(^{13}C)$-TOCCATA database. For each compound, the number of their isomeric states recorded in the database is given in the "Number of isomers" column. The number of spin systems of each isomer is given in the "Number of spin systems of each isomer" column (note that all isomers of each of the compounds of the database has the same number of spin systems).

Number of isomers	Number of spin systems of each isomer	Compound name
2	1	Acetaldehyde_oxime_mixture_of_syn_and_anti
1	1	Acetic_acid
1	2	Acetophenone
1	3	Kanamycin
1	2	N_Acetyl_L_Glutamine
1	2	5-(Acetylamino)-2-nitrobenzoic_acid
1	3	Acetylcarnitine
1	2	N_Acetyl_L_glutamic_acid
1	2	Nepsilon_Acetyl_L_Lysine
1	2	N_acetyl_L_aspartic_acid
1	2	N_(2_Acetamido)iminodiacetic_acid
1	3	Adenosine_3_monophosphate
1	1	Adipic_acid
1	1	Alanine
1	2	Alanine-Alanine
2	2	(1_Amino_1_phenylmethyl)phosphonic_acid
1	3	4_Aminoantipyrine
1	1	gamma_Aminobutyric_acid
1	1	2_Aminoethyl_dihydrogen_phosphate
1	1	6_Aminohexanoic_acid
1	1	2_Aminophenol
1	3	AMP
1	1	1,6-anhydro_beta_d_glucose
1	1	L_Arabitol
1	1	D_Aspartate
1	1	Acetamide
1	1	Dihydroxyacetone
1	1	Acetyl_phosphate
1	3	Acetylcholine
1	2	N_Acetylglycine
1	2	N_Acetylneuraminic_acid
1	3	dAMP
1	4	S_Adenosyl_L_homocysteine
1	1	Adonitol
1	1	Agmatine
1	1	Allantoin
1	1	alpha_Ketoglutaric_acid
1	3	alpha_((Methylamino)methyl)benzylalcohol
1	2	2_Amino_1_phenylethanol
2	1	2_Amino_3_phosphonopropionic_acid
1	1	4_Aminobenzoic_acid
1	1	2_Aminobutyric_acid
1	2	4_(2_Aminoethyl)morpholine
1	1	3_Aminoisobutyric_acid
1	1	4_Aminophenol
1	4	L_Anserine
1	2	Arbutin
1	1	L_Ascorbate
1	2	4_Acetamidobutyric_acid
1	1	Acetonitrile
1	2	N_Acetyl_L_alanine
1	2	N_alpha_Acetyl_L_Lysine
1	3	N_acetyl_D_Phenyl_alanine
1	1	Acrylamide
1	3	Adenosine
1	3	ADP
1	1	beta_Alanine
1	1	D_Allose
1	1	2_Amino_5_ethyl_1,3,4_thiadiazole

1	1	1_amino_1_cyclohexanecarboxylic_acid
1	1	2_Aminoadipic_acid
1	2	Methyl_4_aminobutyrate
1	1	2_Aminocaprylic_acid
2	1	2_Aminoethylphosphonic_acid
1	4	6_Aminopenicillanic_acid
1	1	1_Amino_2_propanol
1	2	p_Anisic_acid
1	1	Anthranilic_acid
2	1	L_Arabinose
1	1	L_Arginine
1	1	L_Asparagine
1	3	ATP
1	1	Benzoate
1	1	Betaine
1	1	Biotin
1	1	4_Hydroxy_benzoic_acid
1	2	Benzyl_alcohol
1	2	Betaine_aldehyde
1	1	3_Hydroxybutyrate
1	1	1,2,4_Benzenetriol
1	1	Benzonitrile
1	1	Butanol
1	1	Butyric_acid
1	3	cAMP
1	1	N_carbamyl_L_Glutamic_acid
1	2	L_Carnitine
1	3	Chlorogenic_acid
1	2	trans_Cinnamic_acid
1	2	R_Citramalic_acid
1	2	Creatinine
1	2	p_Cresol
1	1	Cyclohexanol
1	2	N_cyclohexylformamide
1	2	L_Cystathionine
1	1	L_Cysteine
1	2	CMP
1	1	Cadaverine
1	1	L_Canavanine
1	2	CDP
1	2	Choline
1	2	4_Hydroxycinnamic_acid
1	1	Citrate
2	9	CoA
1	3	Cotinine
1	1	Creatine
1	2	m_Cresol
1	2	CTP
1	1	cis_1,2_cyclohexanediol
1	1	Cyclohexanone
1	1	N_cyclohexylsulfamic_acid
2	1	Cysteamine
1	1	L_Cystine
1	1	Cytosine
1	1	epsilon_Caprolactam
1	2	3_Carboxypropyl_trimethyl_ammonium
1	3	L_Carnosine
1	2	Chorismic_acid
1	2	trans_3_Hydroxycinnamic_acid
1	1	cis_Aconitic_acid
1	1	Citraconic_acid
1	1	L_Citrulline
1	2	Creatine_phosphate
1	2	o_Cresol
1	1	Cyclohexylamine
1	1	L_Cysteic_acid
1	2	Cytidine
1	3	dADP
1	2	dGTP
1	2	dGMP

1	1	1,3_Diaminopropane
1	1	Diethanolamine
1	1	2,5_Dihydroxybenzoic_acid
1	2	3,4_Dihydroxyphenylacetic_acid
1	1	1,3_Dimethylurea
1	2	DSS
1	3	dAMP
1	3	3_Deazauridine
1	2	2_Deoxycytidine
1	2	2_Deoxyuridine
1	2	Dihydrocoumarin
1	2	cis_1,2_dihydronaphthalene_1,2_diol
1	1	Dihydrouracil
1	1	3,4_Dihydroxybenzoic_acid
1	2	3,4_Dihydroxy_L_phenylalanine
1	2	N,N_Dimethylglycine
1	2	dTMP
1	2	Decamethonium
1	2	2_Deoxyguanosine
1	1	d_Desthiobiotin
2	1	alpha,epsilon_Diaminopimelic_acid
1	2	2,3_Dideoxycytidine
1	1	O,O_diethyl_thiophosphate
1	1	L_Dihydroorotic_acid
1	2	trans_2,3_Dimethylacrylic_acid
1	2	2,2_Dimethylsuccinic_acid
1	1	meso_Erythritol
1	1	Ethanol
1	1	Ethanolamine
1	1	Ethylmalonic_acid
1	1	2_Ethylpiperidine
1	3	Epinephrine
1	1	Ethanesulfonic_acid
1	8	FAD
1	3	Ferulic_acid
1	4	Folate
1	1	Formate
2	2	Fructose_1,6_bisphosphate
2	1	L_Fucose
1	1	Formaldehyde
3	2	D_Fructose
1	2	D_Fructose_1_phosphate
1	1	Fumaric_acid
1	2	2_Furoylglycine
1	5	FMN
1	1	Formamide
1	2	D_Fructose_2,6_bisphosphate
2	2	D_Fructose_6_phosphate
1	1	2_Furoic_acid
1	3	N_(2_furoyl)glycine_methylester
1	1	alpha_D_Galactose_1_phosphate
1	1	D_Glucono_1,5_lactone
2	1	D_Glucosamine_6_phosphate
1	1	D_Glucosaminic_acid
1	1	alpha_D_Glucose_1,6_bisphosphate
2	1	D_Glucuronate
1	1	L_Glutamine
1	3	L_Glutathione_reduced
1	1	2,3_Diphospho_D_glyceric_acid
1	1	Glycolate
1	2	Guaiacol
1	1	Guanidineacetic_acid
2	1	D_Galactono_1,4_lactone
2	1	D_Galacturonic_acid
2	1	D_Glucosamine
1	2	N_Acetyl_D_glucosamine_1_phosphate
1	3	Methyl_N_acetyl_alpha_D_glucosaminide
2	1	D_Glucose_6_phosphate
1	1	Glutaconic_acid
1	1	Glutaric_acid

1	1	Glyceraldehyde
1	1	Glycerol
1	1	Glycine
1	1	Glyoxylic_acid
1	1	4_Guanidinobutyric_acid
2	1	L_Gulonolactone
2	1	D_Galactose
1	1	Gallic_acid
1	2	GDP
1	1	Gluconic_acid
2	2	N_Acetyl_D_glucosamine
2	2	N_Acetyl_D_glucosamine_6_phosphate
2	1	D_Glucose
1	1	alpha_D_Glucose_1_phosphate
1	1	L_Glutamic_acid
1	3	L_Glutathione_oxidized
1	1	D_Glyceraldehyde
1	1	DL_alpha_Glycerol_phosphate
1	1	Glycolaldehyde
2	2	Gly_Pro
1	1	Hexanoic_acid
1	1	cis_3_hexen_1_ol
1	2	L_Histidine
1	1	L_Homocitrulline
1	1	L_Homoserine
1	1	3_Hydroxyanthranilic_acid
1	2	4_Hydroxybenzyl_alcohol
1	2	trans_2_Hydroxycinnamic_acid
1	1	2_Hydroxyisocaproic_acid
1	1	2_Hydroxy_3_methylbutyric_acid
1	2	5_(Hydroxymethyl)uracil
1	2	4_Hydroxyphenethyl_alcohol
1	2	2_Hydroxyphenylacetic_acid
1	2	3_(2_Hydroxyphenyl)propionic_acid
1	3	4_Hydroxy_3_methoxybenzyl_alcohol
1	3	HEPES
1	2	Hippuric_acid
1	2	L_Histidinol
1	1	Homocysteine
1	1	O_Succinyl_L_homoserine
1	3	4_Hydroxy_3_methoxymandelic_acid
1	2	4_Hydroxybenzaldehyde
1	1	2_Hydroxyhexanoic_acid
1	2	3_Hydroxymandelic_acid
1	2	3_Hydroxy_3_methylglutaric_acid
1	1	6_Hydroxynicotinic_acid
1	2	4_Hydroxyphenylacetic_acid
1	3	4_Hydroxy_3_methoxycinnamaldehyde
1	1	Hypotaurine
1	1	trans_3_Hexenedioic_acid
1	2	Histamine
1	1	Homoarginine
1	2	Homogentisic_acid
1	3	Homovanillic_acid
1	1	Hydroquinone
1	1	3_Hydroxybenzoic_acid
1	1	2_Hydroxybutyric_acid
1	1	3_Butyn_1_ol
1	3	5_Hydroxyindole_3_acetic_acid
1	3	3_Hydroxy_4_methoxycinnamic_acid
1	1	2_Hydroxyoctanoic_acid
1	2	3_Hydroxyphenylacetic_acid
1	2	4_Hydroxyphenylacetonitrile
1	2	4_Hydroxyphenylglycine
1	3	5_Hydroxy_L_tryptophan
1	1	Imidazole
1	3	Indole_3_lactic_acid
1	3	Inosine_5_monophosphate
1	1	scyllo_Inositol
1	1	Isethionic_acid

1	1	Threo_isocitric_acid
1	1	L_Isoleucine
1	1	Itaconic_acid
1	3	Indole_3_acetic_acid
1	1	muco_Inositol
1	1	Isobutyric_acid
1	1	Isonicotinic_acid
1	3	ITP
1	3	Inosine
1	1	myo_Inositol
1	2	IPTG
1	1	Isovaleric_acid
1	2	Kynurenic_acid
1	1	2_Ketobutyric_acid
1	2	L_Kynurenine
1	1	beta_Leucine
1	1	Lactic_acid
1	1	Cysteinesulfinic_acid
1	1	Leucine
1	1	Lysine
1	1	Maleamic_acid
1	1	Malic_acid
2	2	Methionine_sulfoxide
1	2	4_Methylcatechol
1	1	5_Methylcytosine
1	1	3_Methylglutaric_acid
1	1	Methylmalonic_acid
1	1	4_Methyl_2_oxovaleric_acid
1	3	3_Methylphenylacetate
1	2	4_Methylsalicylic_acid
1	1	N_Methylurea
1	1	trans,trans_Muconic_acid
1	1	Maleic_acid
2	1	D_Mannose
1	2	MES
1	1	Methanol
1	1	3_Methyl_2_butenoic_acid
1	1	3_Methyladipic_acid
1	3	1_Methyl_L_Histidine
1	1	3_Methyl_2_oxobutinoic_acid
1	3	4_Methyl_5_thiazoleethanol
1	2	5_Methyluridine
1	2	Monoethyl_malonate
1	1	D_Mannitol
1	1	Mesaconic_acid
1	2	L_Methionine
1	2	3_Methylcatechol
2	1	3_Methylcrotonaldehyde
1	1	2_Methylglutaric_acid
1	2	N_methyl_L_Aspartic_acid
1	1	3_Methyl_2_oxopentanoic_acid
1	2	3_Methylsalicylic_acid
1	2	MOPS
1	1	Nicotinamide
1	2	Nicotinuric_acid
1	3	N_acetyl_Histidine
1	3	Neostigmine
1	1	4_Nitrophenol
1	2	Norepinephrine
1	1	Nicotinic_acid
1	3	Nicotine
1	1	4_Nitrocatechol
1	1	4_Nitrophenyl_phosphate
1	1	L_Norleucine
1	1	L_Ornithine
1	1	Orthanilic_acid
1	1	2_Octenoic_acid
1	2	N(alpha)_Acetyl_ornithine
1	4	Pantolactone
1	5	Pantothenate

1	1	Phenol
2	3	Phenylacetylglycine
1	1	Phenylglyoxylic_acid
1	1	6_Phosphogluconic_acid
1	1	O_phospho_threonine
1	2	D_Pinitol
1	1	Pipecolic_acid
1	1	3,4_Dehydro_proline
1	1	1,3_Propanediol
1	1	Propionic_acid
1	3	Purine
1	2	3_Pyridine_aldoxime
2	4	Pyridoxal_5_phosphate
1	4	Pyridoxine
1	1	2_Pyrrolidinone_5_carboxylate
1	5	Panthenol
1	2	L_Phenylalanine
1	2	2_Phenylbutyric_acid
1	1	3_Phosphoglyceric_acid
1	1	O_Phospho_L_serine
1	1	Phthalic_acid
1	1	Polygalacturonic_acid
1	1	trans_4_Hydroxy_L_proline
1	1	Propanol
1	1	Putrescine
1	1	3_Pyridinecarbonitrile
1	4	Pyridoxamine
1	1	Pyrocatechol
2	1	Pyruvic_acid
1	2	Phenethylamine
1	2	Phenoxyacetic_acid
1	2	Phenylacetic_acid
1	2	2_Phenylethanol
1	1	Phosphoenolpyruvic_acid
2	1	Phosphonoacetic_acid
1	2	Phthalaldehydic_acid
1	1	Pimelic_acid
1	1	L_Proline
1	1	1,2_Propanediol
1	1	Pyridine
1	2	Pyridoxal
1	2	Pyridoxamine_5_phosphate
1	2	Quinaldic_acid
1	1	Quinolinic_acid
1	1	Quinic_acid
2	1	D_Ribose_5_phosphate
1	1	Resorcinol
1	1	D_Saccharate
1	1	Salicylate
1	2	Sarcosine
1	2	Selenomethionine
1	3	Serotonin
1	2	D_Sorbose
1	2	Spermine
1	3	D_Salicin
1	1	L_Serine
1	1	Shikimic_acid
1	3	Sinapic_acid
1	1	Suberic_acid
1	1	Sulfanilic_acid
1	2	Saligenin
1	1	D_Sorbitol
1	2	Spermidine
1	1	Succinic_acid
1	1	Sulfoacetic_acid
1	6	Thiamine_monophosphate
1	1	L_Threitol
1	2	Thymidine
1	1	Trigonelline
1	1	Trimethylamine_N_oxide

1	3	Tryptamine
1	2	L_Tyrosine
1	1	L_Tartaric_acid
1	1	Tetramethylammonium
1	1	Thioacetamide
1	1	Threo_beta_methylaspartate
1	2	p_Toluic_acid
1	1	Triethanolamine
2	1	Trimethyl_phosphate
1	3	L_Tryptophan
1	1	Taurine
1	6	Thiamin
1	1	L_Threonine
1	1	D_Trehalose
1	2	Triethylenetetramine
1	1	Trimethylamine
1	2	Tropic_acid
1	2	Tyramine
1	2	UDP
1	1	3_Ureidopropionic_acid
1	2	Uridine
1	2	UTP
1	2	UMP
1	1	Uracil
1	2	Urocanate
1	1	Ureidosuccinic_acid
1	4	Uridine_5_diphospho_N_acetylglucosamine
1	3	UDP_glucuronate
1	1	Valeric_acid
1	1	L_Valine
1	3	Vanillin
1	2	Vanillic_acid
1	1	Xylitol
2	1	D_Xylose
1	1	D_Xylonate
1	3	Caffeine
1	5	NADH
1	5	NAD
1	5	NADPH
1	5	NADP+
1	2	Nicotinamide_D_ribonucleotide_oxidized
1	2	Nicotinamide_D_ribonucleotide_reduced
1	3	Sucrose
2	2	Lactose
2	2	Maltose
4	1	D_Ribose

Table S-2. List of *E. coli* metabolites that have at least one matching peak in 1D ^{1}H NMR spectrum and Chenomx NMR software Suite[a].

	# of matching peaks/ total # of peaks	Chemical shifts of the matching peaks (ppm)
Glutamate	3/4	2.044, 2.119, 2.342
Succinate	1/1	2.395
Formate	1/1	8.445
Alanine	1/2	1.470
Glycine	1/1	3.549
Acetate	1/1	1.906
4-Aminobutyrate	2/3	2.285, 3.004
Methanol	1/1	3.343
Putrescine	2/2	1.758, 3.040
Leucine	1/3	0.949
Valine	2/4	0.980, 1.031
Isoleucine	1/6	0.999
Nicotinate	2/4	8.594, 8.928
Fumarate	1/1	6.507
Phenylalanine	4/6	3.115, 7.311, 7.357, 7.409
$NADP^{+}$	5/17	8.139, 8.402, 8.807, 9.092, 9.282
Uracil	2/2	5.785, 7.520
Tyrosine	2/5	6.875, 7.172
Aspartate	1/3	2.802

[a]Chenomx NMR Suite 7.7 evaluation (version 8). The magnetic field strength of Chenomx profiler is set to the experimental field strength at 700 MHz. The pH is set to 7.00 +/- 0.50.

Table S-3. Performance of 1D ^{1}H NMR databases BMRB, MMCD, COLMAR, HMDB and the new database $^{1}H(^{13}C)$-TOCCATA for the query of 45 ^{1}H TOCSY traces extracted from cell lysate 2D ^{1}H-^{1}H TOCSY spectrum. Identification is counted as correct (+) if the query returned the name of the metabolite as the first hit.

	BMRB[a]	MMCD[b]	COLMAR	HMDB[c]	TOCCATA
Valine (4)	+	+	+	+	+
Lysine (5)	+	-	+	-	+
Isoleucine (6)	+	+	+	+	+
Leucine (3)	-	+	+	+	+
Proline (6)	+	+	+	-	+
Alanine (2)	-	+	+	-	+
Ethanol (2)	-	+	-	+	+
Arginine (5)	-	-	+	-	+
β-Alanine (2)	+	+	+	-	+
γ−Aminobutyrate (3)	-	+	+	-	+
Nicotinic acid (4)	+	+	+	+	+
Tyrosine (2)	-	-	-	-	+
Phenylalanine (3)	+	-	-	-	+
Uracil (2)	-	+	+	-	+
Lactate (2)	-	+	-	+	+
Phosphoenolpyruvate (2)	-	+	+	-	+
Putrescine (2)	-	+	-	+	+
Thymidine 1 (6)	+	-	+	-	+
Thymidine 2 (2)	-	-	+	-	+
2-Deoxycytidine 1 (2)	-	-	-	-	+
2-Deoxycytidine 2 (7)	-	-	+	-	+
$NADP^{+}$ (4)	-	-	-	-	+
Tryptophan (4)	+	-	+	-	+
p-Toluic acid (2)	-	-	+	-	+
Cytosine (2)	-	+	+	+	+
Propionic acid (2)	+	+	+	-	+
Ethanolamine (2)	-	+	+	-	+
N-acetyl-glutamate (4)	-	-	+	+	+
Citrulline (4)	+	+	+	-	+
Cytidine (2)	-	-	+	-	+
Spermidine (2)	+	-	-	-	+
2-Aminobutyrate (3)	-	-	+	+	+
Threonine (3)	-	+	+	+	+
Uridine (6)	+	-	+	-	+
N-α-acetyl-ornithine (4)	+	+	+	-	+
N-acetyl-glutamine (4)	+	+	+	-	+
Methionine-sulfoxide 1 (3)	-	-	+	-	+
Methionine-sulfoxide 2 (4)	-	-	-	-	+
Coenzyme A 1 (2)	-	-	-	-	+
Coenzyme A 2 (2)	-	-	-	-	+
Pantothenate (2)	+	-	-	+	+
Glutamate (3)	-	-	-	+	+
Adenosine (6)	-	-	-	-	+
Adenosine-3-monophosphate (5)	-	-	-	-	+
Inosine (6)	+	-	-	-	+
Total # of correct hits[d]	17	20	29	13	45

[a] "H range" parameter of 0.02 was used in all queries.
[b] "H_tol" parameter of 0.05 ppm and "Threshold" parameter of 80% were used in all queries (default values).
[c] "Peak Tolerance" parameter 0.02 was used in all queries.
[d] Total number of correct identifications.

Table S-4. Performance of 2D ^{1}H-^{1}H TOCSY NMR databases BMRB, MMCD and Metabominer and the new database ^{1}H(^{13}C)-TOCCATA for the query of 45 ^{1}H TOCSY traces extracted from cell lysate 2D ^{1}H-^{1}H TOCSY spectrum. Identification is counted as correct (+) if the query returned the name of the metabolite as the first hit.

	MMCD[a]	HMDB[b]	Metabominer[c]	TOCCATA
Valine (4)	+	+	+	+
Lysine (5)	-	+	-	+
Isoleucine (6)	+	+	+	+
Leucine (3)	+	-	+	+
Proline (6)	+	-	+	+
Alanine (2)	+	+	+	+
Ethanol (2)	+	+	+	+
Arginine (5)	-	+	+	+
β-Alanine (2)	+	+	+	+
γ−Aminobutyrate (3)	+	-	+	+
Nicotinic acid (4)	+	-	+	+
Tyrosine (2)	-	-	-	+
Phenylalanine (3)	-	+	-	+
Uracil (2)	+	+	+	+
Lactate (2)	+	+	+	+
Phosphoenolpyruvate (2)	+	-	-	+
Putrescine (2)	+	-	-	+
Thymidine 1 (6)	-	-	+	+
Thymidine 2 (2)	-	+	+	+
2-Deoxycytidine 1 (2)	-	-	-	+
2-Deoxycytidine 2 (7)	-	-	-	+
$NADP^{+}$ (4)	-	-	-	+
Tryptophan (4)	-	+	+	+
p-Toluic acid (2)	-	-	-	+
Cytosine (2)	+	+	+	+
Propionic acid (2)	+	+	+	+
Ethanolamine (2)	+	+	+	+
N-acetyl-glutamate (4)	-	-	+	+
Citrulline (4)	+	+	+	+
Cytidine (2)	-	+	+	+
Spermidine (2)	-	-	-	+
2-Aminobutyrate (3)	-	-	-	+
Threonine (3)	+	+	+	+
Uridine (6)	-	+	-	+
N-α-acetyl-ornithine (4)	+	-	-	+
N-acetyl-glutamine (4)	+	-	-	+
Methionine-sulfoxide 1 (3)	-	-	-	+
Methionine-sulfoxide 2 (4)	-	-	-	+
Coenzyme A 1 (2)	-	-	+	+
Coenzyme A 2 (2)	-	-	+	+
Pantothenate (2)	-	-	-	+
Glutamate (3)	-	+	-	+
Adenosine (6)	-	-	-	+
Adenosine-3-monophosphate (5)	-	-	-	+
Inosine (6)	-	-	-	+
Total # of correct hits[d]	20	20	24	45

[a] "H_tol" parameter of 0.05 ppm and "Threshold" parameter of 80% were used in all queries (default values).
[b] "X-axis Peak Tolerance" and "Y-axis Peak Tolerance" parameters were set to 0.2 in all queries.
[c] "Tolerance" parameter of 0.03 ppm and Biofluid (all) [223] "Library" were used in all queries.
[d] Total number of correct identifications.

Table S-5. Performance of 1D ^{1}H and 1D ^{13}C NMR databases; BMRB, MMCD, COLMAR, HMDB and the new database $^{1}H(^{13}C)$-TOCCATA for the query of 38 ^{1}H and ^{13}C HSQC-TOCSY traces extracted from cell lysate 2D ^{13}C-^{1}H HSQC-TOCSY spectrum. Identification is counted as correct (+) if the query returned the name of the metabolite as first hit.

	BMRB		MMCD		COLMAR		HMDB		TOCCATA	
	^{1}H[a]	^{13}C[b]	^{1}H[c]	^{13}C[d]	^{1}H	^{13}C	^{1}H[e]	^{13}C[f]	^{1}H	^{13}C
Valine	-	+	+	+	+	+	+	+	+	+
Lysine	+	+	-	+	+	+	-	+	+	+
Malate	+	-	+	-	+	+	+	+	+	+
Alanine	-	-	+	-	+	+	-	+	+	+
Leucine	+	+	+	-	+	+	+	+	+	+
Threonine	-	-	+	-	+	+	+	+	+	+
β-Alanine	-	+	+	-	+	+	-	+	+	+
Uracil	+	-	+	-	+	+	-	+	+	+
Tyrosine 1	-	-	-	-	-	-	-	+	+	+
Tyrosine 2	-	-	-	-	-	-	+	+	+	-
Phenylalanine 1	-	+	-	-	+	+	-	+	+	+
Phenylalanine 2	+	+	-	-	-	+	-	+	+	+
Arginine	+	-	-	-	+	+	+	+	+	+
γ−Aminobutyrate	-	-	+	-	+	-	+	+	+	+
Aspartate	+	+	+	-	+	+	+	+	+	+
Glutamate	-	-	-	-	-	+	+	+	+	+
Lactate	-	-	+	-	+	+	-	+	+	+
Nicotinic acid	+	+	+	-	+	+	+	-	+	+
Fumarate	-	-	+	-	+	-	-	+	+	+
Phosphoenolpyruvate	-	-	+	-	+	-	-	-	+	+
Serine	+	+	+	-	+	+	-	-	+	+
Methanol	-	+	+	+	-	-	-	-	-	+
Glycine	-	+	+	-	-	-	-	+	-	+
Succinate	-	-	+	-	-	-	-	+	-	+
N-acetyl-alanine	-	-	-	-	-	+	-	+	-	+
Acetic acid	-	-	+	-	-	-	-	+	+	-
Putrescine	-	-	+	+	-	-	-	+	+	+
Thymidine 1	-	-	-	-	+	-	-	-	+	-
Thymidine 2	+	+	-	-	+	+	-	-	+	+
Cytidine	-	-	-	-	+	+	-	-	-	+
dTMP 1	-	-	-	-	-	+	-	-	+	+
dTMP 2	-	+	-	-	+	+	+	-	+	+
Uridine 1	+	+	+	-	+	+	+	+	+	+
Uridine 2	+	-	-	-	+	+	-	+	-	+
Adenosine	-	-	-	-	-	+	+	+	+	+
Inosine	+	+	-	-	-	+	-	-	+	+
Glutathione reduced	+	+	-	-	+	-	-	+	+	+
Cystathionine	-	-	-	-	-	-	-	-	+	-
Total # of correct hits[g]	14	16	20	4	24	25	13	27	32	34

[a] "H range" parameter of 0.02 was used in all queries.

[b] "C range" parameter of 0.2 was used in all queries.

[c] "H_tol" parameter of 0.05 ppm and "Threshold" parameter of 80% were used in all queries (default values).

[d] "C_tol" parameter of 0.5 ppm and "Threshold" parameter of 80% were used in all queries.
[e] "Peak Tolerance" parameter 0.02 was used in all queries.
[f] "Peak Tolerance" parameter 0.2 was used in all queries.
[g] Total number of correct identifications.