Supporting information

Extractive Ratio Analysis NMR Spectroscopy for Metabolite Identification in Complex Biological Mixtures

Liladhar Paudel,[1,2] G. A. Nagana Gowda,[1] and Daniel Raftery[1,3,4*]

[1]*Northwest Metabolomics Research Center, Anesthesiology and Pain Medicine, and*
[3]*Department of Chemistry, University of Washington, Seattle, WA 98109,*
[4]*Fred Hutchinson Cancer Research Center, Seattle, WA 98109*
[2]Currently at *Lubrizol Advanced Materials, 9911 Brecksville Road, Brecksville, OH 44141*

* Corresponding Author:

Professor Daniel Raftery
Northwest Metabolomics Research Center
Department of Anesthesiology and Pain Medicine
University of Washington
850 Republican St.
Seattle, WA 98109
Tel: 206-543-9709
Fax: 206-616-4819
Email: draftery@uw.edu

This file contains 13 pages, S1 to S13. The first page is the Title page; pages S2 to S4 contain Table S1, Table S2 and Table S3, respectively; pages S5 and S6 contain MatLab script for data smoothing and data normalization, respectively; pages S7 to S13 contain Figure S1 to Figure S7, respectively.

Table S1. List of standard compounds used for initial testing of E-RANSY analysis.

Number	Compound Name	Concentration (μM)
1	2-Hydroxybutyric acid	311.3
2	2-Hydroxyisovalerate	315.0
3	2-Oxoisocaproate	317.2
4	2-Oxoisovalerate	316.2
5	3-Hydroxyisovalerate	315.2
6	3-Methyl-2-oxovalerate	311.7
7	Acetate	313.5
8	Benzoic acid	314.4
9	Citrate	312.3
10	Formate	313.5
11	Fumarate	316.8
12	Hippuric acid	312.5
13	Isobutyric acid	314.1
14	Isovalerate	316.2
15	Lactate	312.0
16	Succinate	313.7
17	Aconitate, trans-	314.0
18	Alpha-ketoglutarate, sodium salt	312.6
19	Isocitrate, trisodium salt hydrate	311.9
20	Malate	314.6
21	Oxaloacetate	317.9
22	Pyruvate, sodium salt	317.1
23	Nicotinic acid	317.3
24	3-Methylhistidine	313.2
25	Arginine	312.5
26	Asparagine	318.4
27	Aspartate	312.2
28	Histidine	315.1
29	Isoleucine	316.4
30	Methionine	313.5
31	Phenylalanine	315.0
32	Serine	317.0
33	Tryptophan	312.0
34	Tyrosine	313.9
35	Valine	315.4
36	Acetylcarnitine	313.6
37	Betaine	312.9
38	Carnitine	312.5
39	Choline	312.5
40	Creatine	315.9
41	Creatinine	314.6
42	1,2-Propanediol	317.5
43	2-Propanol	314.5
44	Acetone	316.0
45	Dimethylamine	313.9
46	Dimethylglycine	314.4
47	Ethanol	318.4
48	Glycerol	312.0
49	Uridine	314.5

Table S2. Parameters used in solvent-solvent extraction of compounds from standard mixtures for E-RANSY analysis.

Sample No.	Sample volume (μL)	Sample volume after pH adjustment (μL)	pH	Ethyl Acetate volume (μL)	Ethyl Acetate Extract taken for drying (μL)	DI water Added (μL)	Time used to completely dry (hrs)
1	500	550	0.57	1833	1375	200	10
2	500	550	1.07	1833	1375	200	10
3	500	550	1.76	1833	1375	200	10
4	500	550	2.29	1833	1375	200	10
5	500	550	2.52	1833	1375	200	10
6	500	550	3.23	1833	1375	200	10
7	500	550	3.69	1833	1375	200	10
8	500	550	4.25	1833	1375	200	10
9	500	550	4.60	1833	1375	200	10
10	500	550	5.20	1833	1375	200	10
11	500	550	5.69	1833	1375	200	10
12	500	550	6.17	1833	1375	200	10
13	500	550	7.07	1833	1375	200	10
14	500	550	7.06	1833	1375	200	10
15	500	550	7.69	1833	1375	200	10
16	500	550	8.83	1833	1375	200	10
17	500	550	10.96	1833	1375	200	10
18	500	550	11.40	1833	1375	200	10
19	500	550	12.01	1833	1375	200	10

Table S3. Parameters used in solvent-solvent extraction of urine metabolites for E-RANSY analysis.

Sample number	Sample volume (μL)	Sample volume after pH adjustment (μL)	pH	Ethyl Acetate volume (μL)	Ethyl Acetate volume used for drying (μL)	Time to dry (hrs)
1	600	725	0.54	2400	2000	8
2	600	725	1.00	2400	2000	8
3	600	725	1.41	2400	2000	8
4	600	725	2.01	2400	2000	8
5	600	725	2.46	2400	2000	8
6	600	725	3.03	2400	2000	8
7	600	725	3.36	2400	2000	8
8	600	725	3.97	2400	2000	8
9	600	725	4.56	2400	2000	8
10	600	725	4.99	2400	2000	8
11	600	725	5.55	2400	2000	8
12	600	725	5.95	2400	2000	8
13	600	725	6.33	2400	2000	8
14	600	725	7.36	2400	2000	8

MatLab script for the binomial smoothing function algorithm

```
function binomial_smothing(k)
clearvars -global;
file2read='testdata.xlsx'; % file name of Excel Workbook
sheet2read='testdata';   % sheet name of dataset
range2read='C:P'; %data range; here data are found in column C to
column P

[data,text,all]=xlsread(file2read,sheet2read,range2read);
ppmdata=data(1:end,1);
data=data(1:end,2:end);
[rownum,colnum]=size(data);

%%smoothing function matrix
if k==2
binomat=[1 1];
elseif k==3
binomat=[1 2 1];
elseif k==4
binomat=[1 3 3 1];
elseif k==6
binomat=[1 5 10 10 5 1];
elseif k==12
binomat=[1 11 55 165 330 462 462 330 165 55 11 1];
elseif k==18
binomat=[1 17 136 680 2380 6188 12376 19448 24310 24310 19448 12376
6188 2380 680 136 17 1];
else
return
end
N1=length(binomat);

N2=(rownum*colnum/N1);
data2=reshape(data,[N1,N2]);

data3=transpose(repmat(binomat,N2,1));
data4=data2.*data3;
data5=sum(data4,1);
N3=rownum/N1;
sm_data=reshape(data5,[N3,colnum]);

sm_ppm=ppmdata(ceil(N1/2):N1:end,:);

sheetidx=round(N3/1000);
output_sheet=sprintf('%s_binosm_%dk',sheet2read,sheetidx);
%saving output_sheet variable to .mat file
%save(output_sheet);

xlswrite(file2read,text,output_sheet,'A1');
xlswrite(file2read,[sm_ppm,sm_data],output_sheet,'A2');
end
```

MatLab script for the spectral normalization algorithm

```
%%norm2CS (normalization to column sum i.e. normalization to total
sum)
function norm2CS()
if nargin~=0
    error('Must not enter any argument')
end
clearvars -global;
%%
tic
file2read='test.xlsx'; % file name of Excel Workbook
sheet2read='sheet1';    % Sheet name of the dataset
range2read='A:N'; %%ppm data and all ten samples initially taken

[data,text,~]=xlsread(file2read,sheet2read,range2read);
ppmdata=data(1:end,1:3);
data=data(1:end,4:end);
outdata =(bsxfun (@rdivide, data, sum(data)));
outdata=[ppmdata,outdata];
toc
%%
output_sheet=sprintf('%s_norm2CS',sheet2read);
xlswrite(file2read,text,output_sheet,'A1');
xlswrite(file2read,outdata,output_sheet,'A2');

 %%
 toc
end
```

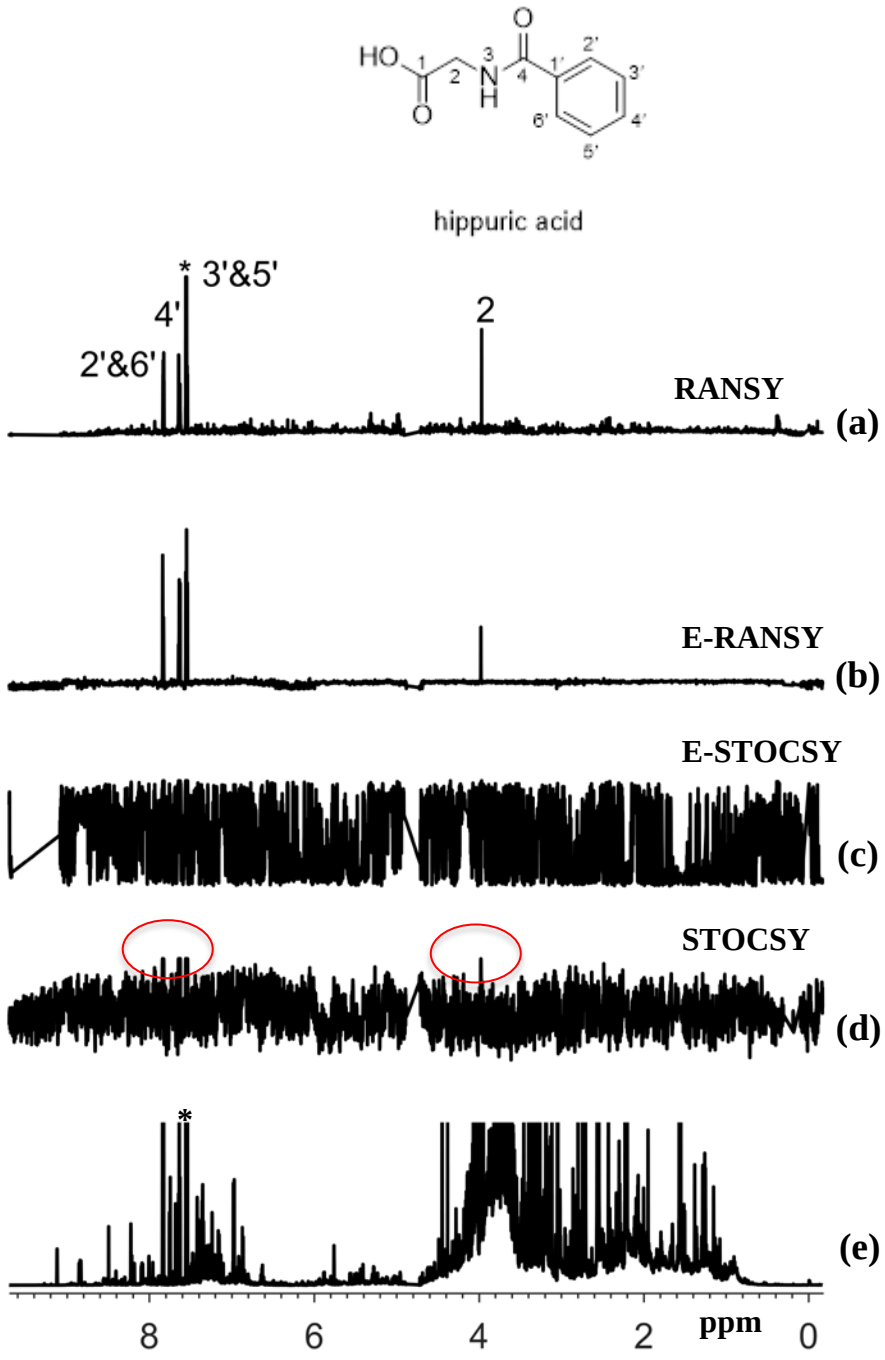

Figure S1. Comparison of the results of ratio analysis and correlation analysis of either extracted urine or intact urine spectra using the driving peak as indicated by asterisk (*). The spectra shown are (a) RANSY; (b) E-RANSY; (c) E-STOCSY, (d) STOCSY and (e) the intact urine 1D ^{1}H NMR spectrum. The two red ovals indicate the weak hippuric acid STOCY peaks. The inset shows the structure of hippuric acid identified based on RANSY/E-RANSY. Peaks in the RANSY spectrum are labeled with corresponding protons as labeled in the structure of the metabolite. For RANSY and STOCSY, intact urine NMR spectra were used, and for E-RANSY and E-STOCSY, ethyl acetate extracted urine NMR spectra were used.

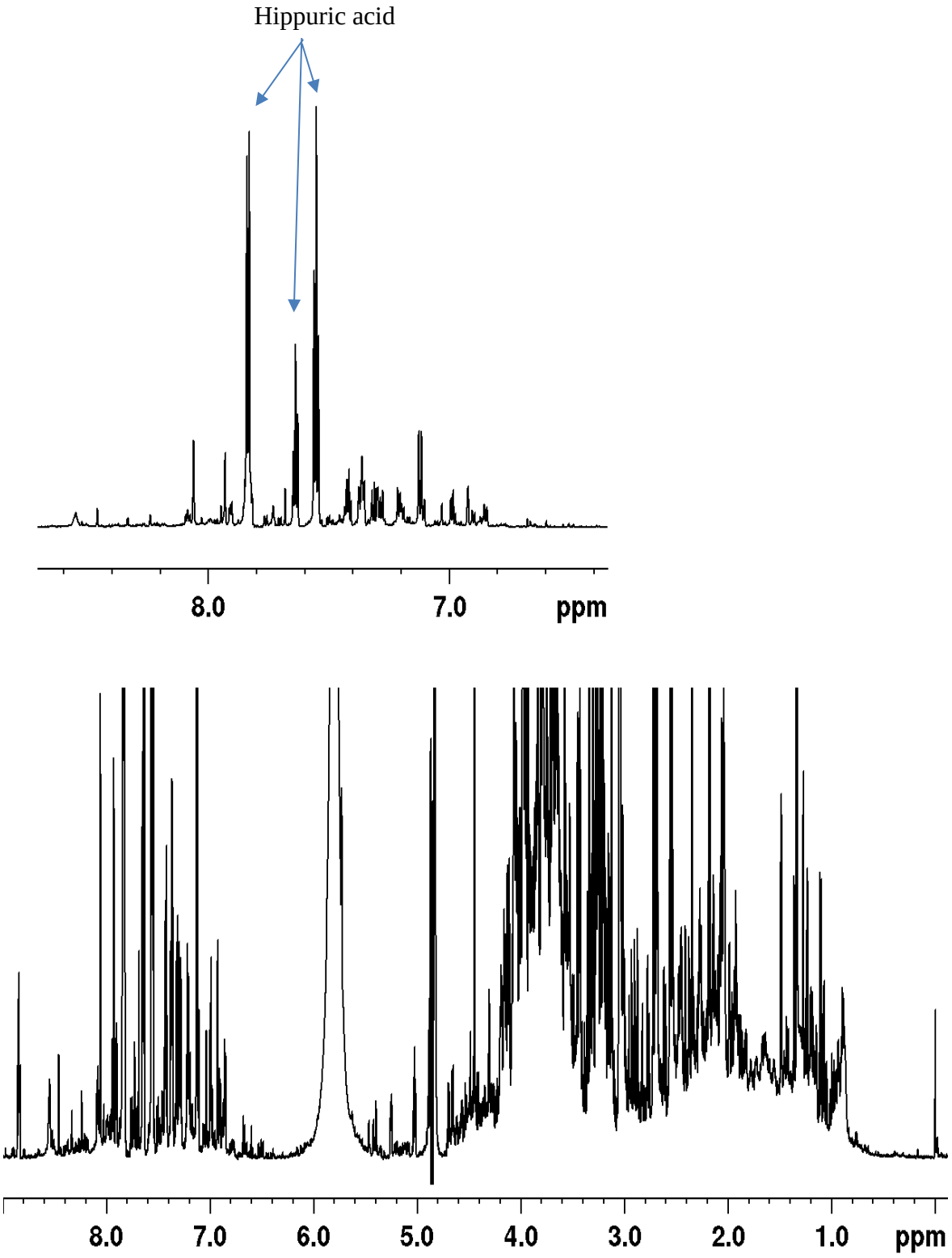

Figure S2. A typical human urine ^{1}H NMR spectrum highlighting the generally dominant peaks from hippuric acid as shown in the expanded region of the spectrum (inset).

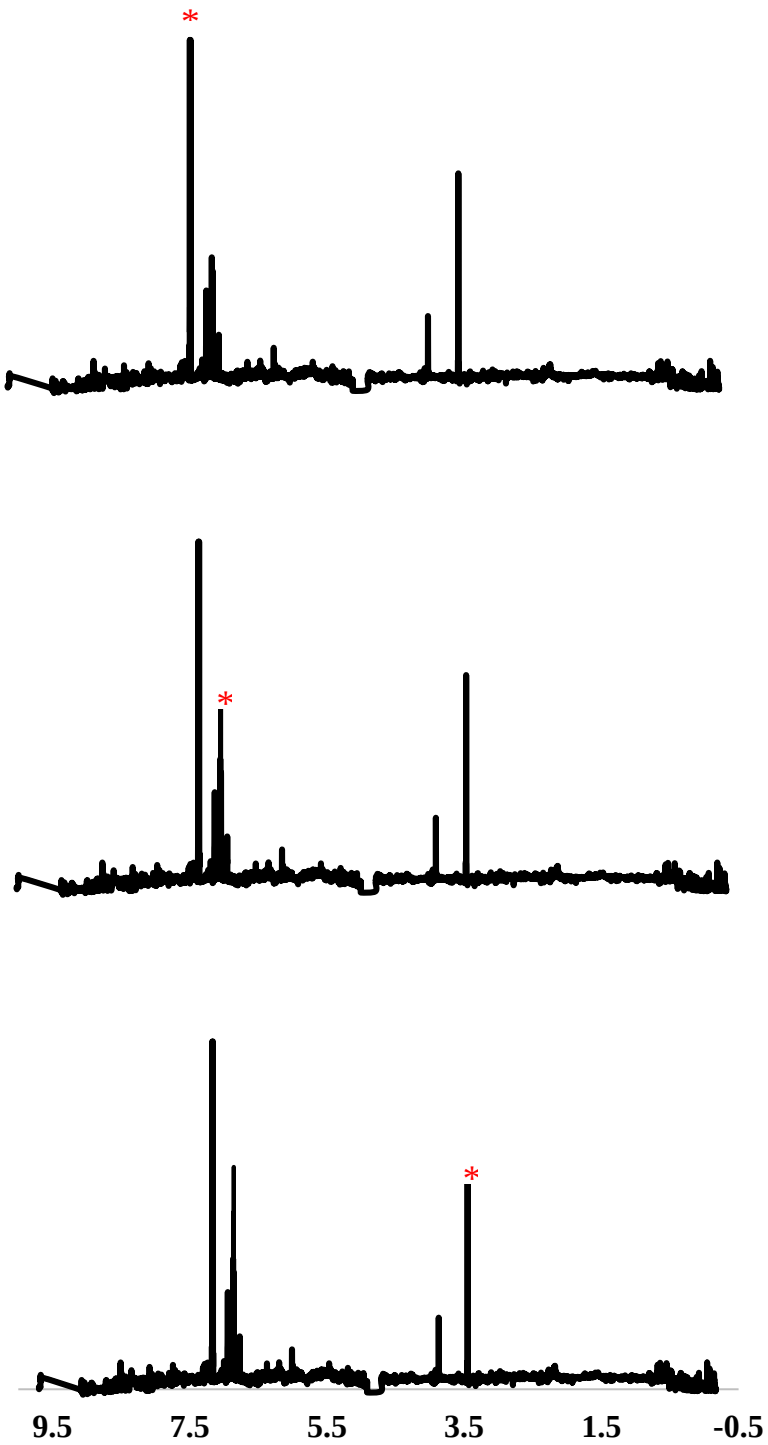

Figure S3. E-RANSY spectra of 4-hydroxyphenylacetic acid obtained using different peaks as the driver peak, separately. The driver peak for each spectrum is indicated by *****. Note, virtually identical spectra are obtained irrespective of the peak chosen as the driver peak.

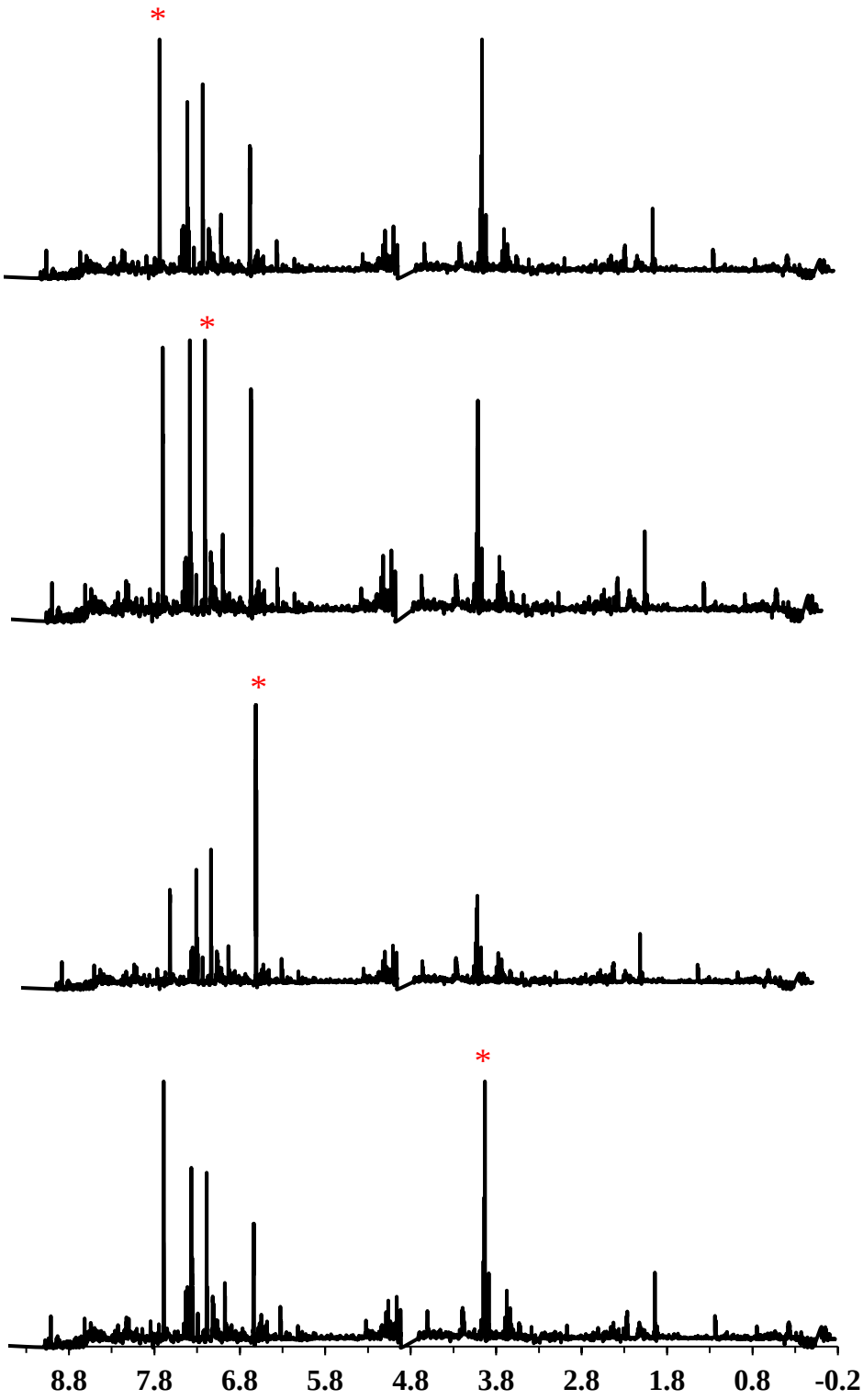

Figure S4. E-RANSY spectra of 2-furoylglycine obtained using different peaks as the driver peak, separately. The driver peak for each spectrum is indicated by *****. Note, virtually identical spectra (with some variation in amplitudes) are obtained irrespective of the peak chosen as the driver peak.

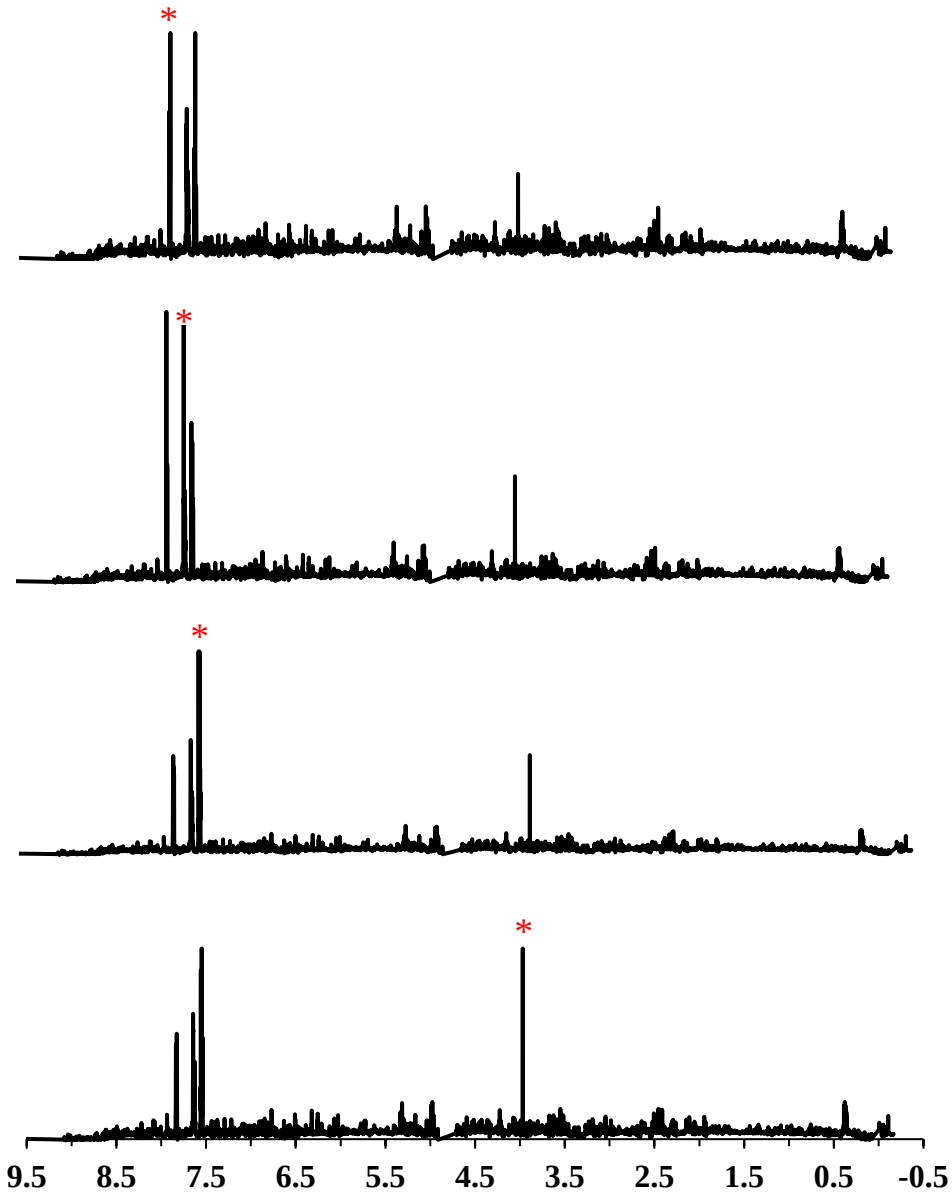

Figure S5. E-RANSY spectra of hippuric acid obtained using different peaks as the driver peak, separately. The driver peak for each spectrum is indicated by *****. Note, virtually identical spectra (with some variation in amplitudes) are obtained irrespective of the peak chosen as the driver peak.

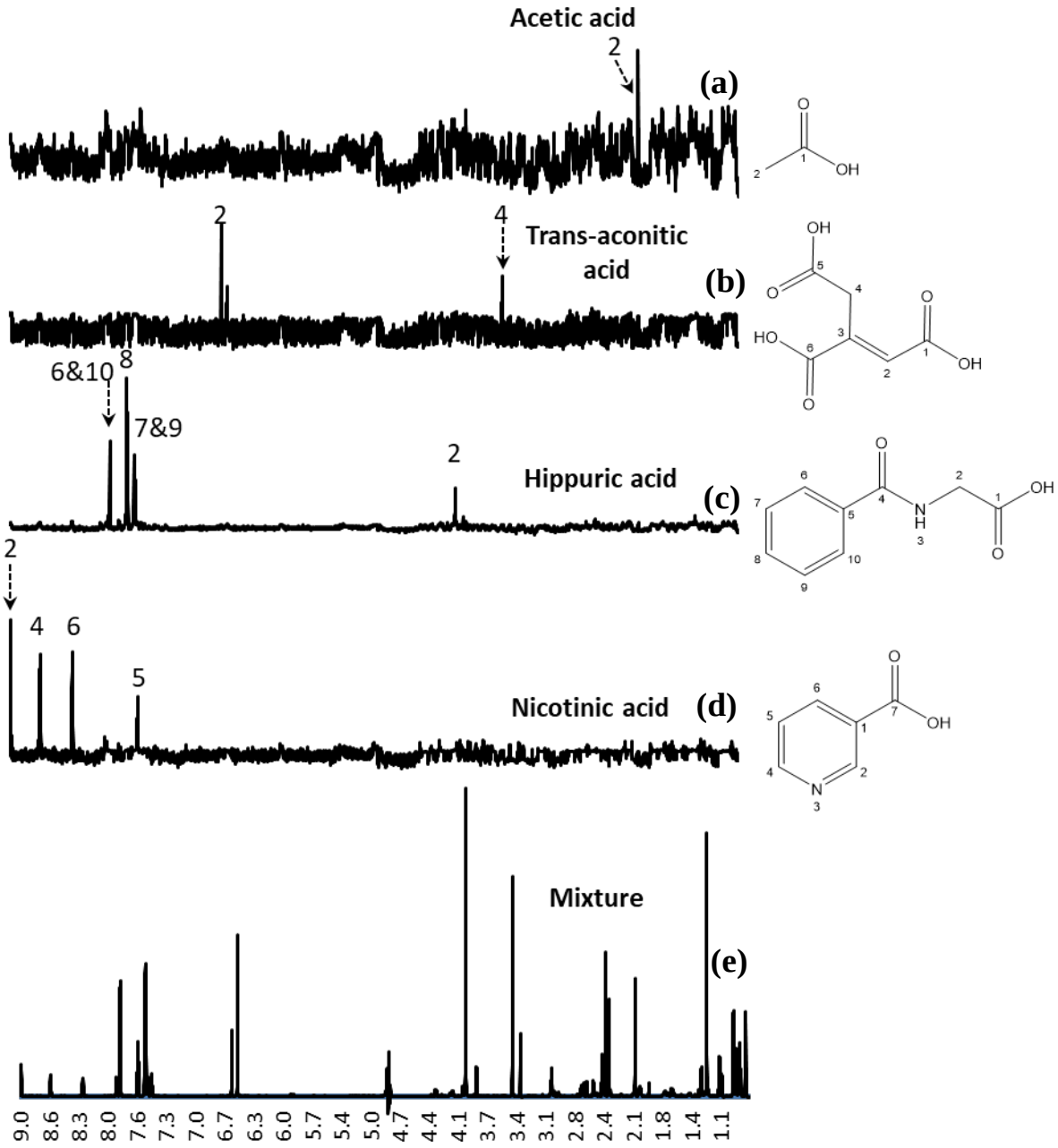

Figure S6. E-RANSY spectra (a-d) for a few standard compounds obtained after the analysis of an ethyl acetate extracted standard mixture of 49 compounds (see Table S1); (e) ^{1}H 800 MHz spectrum of a mixture of the standards.

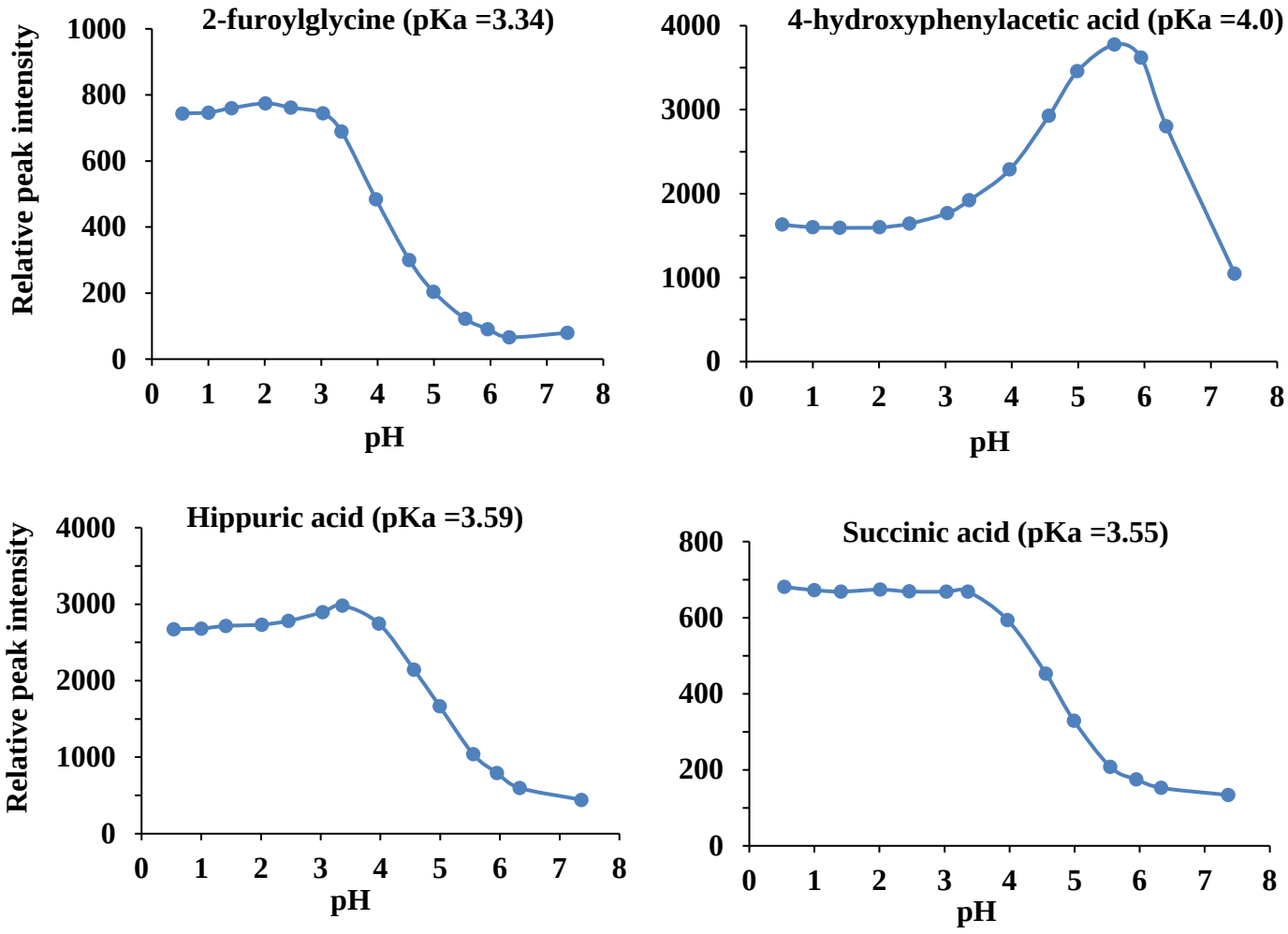

Figure S7. Variation of the relative NMR peak intensities for typical metabolites from the carboxylic acid class, 2-furoylglcyine, 4-phenylacetic acid, hippuric acid and succinic acid, from urine after extraction into ethyl acetate at different pH values.