Electronic Supplementary Information (ESI)

SMolESY: An Efficient and Quantitative Alternative to On-Instrument Macromolecular ^{1}H-NMR Signal Suppression†

Panteleimon G. Takis,*[a,b] Beatriz Jiménez, [a,b] Caroline J. Sands, [a,b] Elena Chekmeneva [a,b] and Matthew R. Lewis [a,b]

[a]Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom

[b]National Phenome Centre, Department of Metabolism, Digestion and Reproduction, IRDB Building, Imperial College London, Hammersmith Campus, London, W12 0NN, United Kingdom

*Correspondence and any requests for materials should be addressed to Dr. Panteleimon G. Takis,

e-mail: p.takis@imperial.ac.uk

Table of Contents

SMolESY impact on time and cost for the NMR metabolomics pipeline

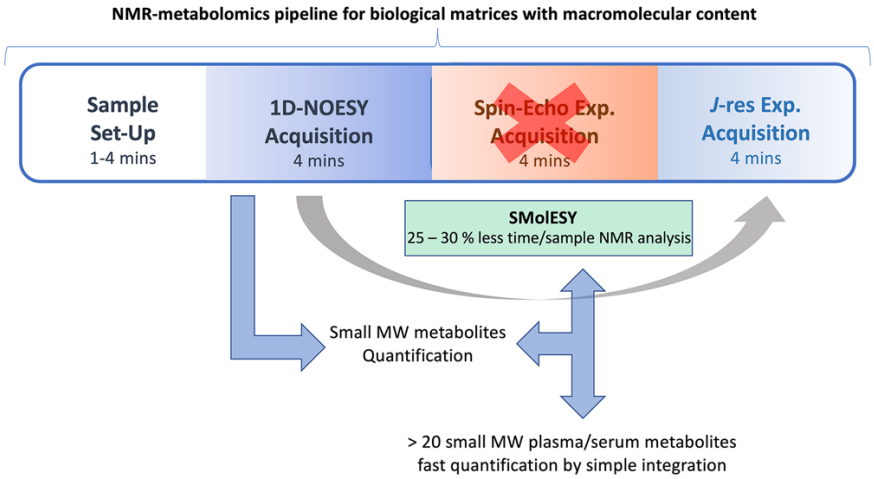


Figure S1. The experimental scheme of NMR based metabolomics pipeline for biofluids with macromolecular content (e.g. proteins, lipoproteins, lipids etc.) – SMolESY contribution. The total amount of experimental time is up to 16 minutes. The 1D-NOESY experiment [containing both macromolecules and low molecular weight (MW) metabolites signals] is followed by a one-dimensional spin-echo experiment (e.g. CPMG), which suppresses the broad signals of macromolecules by T_2 relaxation times filtering. Finally, a pseudo-2D experiment is acquired, *J*-res, employed for the assignment facilitation of low MW metabolites overlapped signals. As shown, spin-echo experiments (given the limited acquisition time) do not efficiently attenuate broad signals of macromolecules, and they cannot be employed for absolute quantification of metabolites as they are highly modulated by T_2 values (the same stands for the projections of J-res experiments that are highly modulated by the J-coupling constant values). On the other hand, SMolESY production requires < 1 sec per spectrum, supersedes spin-echo experiments in broad signals attenuation as well as maintains its quantitative ability for low MW metabolites absolute quantification. Consequently, a substitution of a spin-echo experiment by SMolESY could lead to up to 30% decrease of acquisition time per spectrum (or up to 30% increase of acquired spectra/samples), whereas it could speed up assignment/quantification of small MW metabolites. Consequently, SMolESY could significantly reduce the cost of NMR analyses for large scale epidemiological studies and provide more details for the low MW metabolites content.

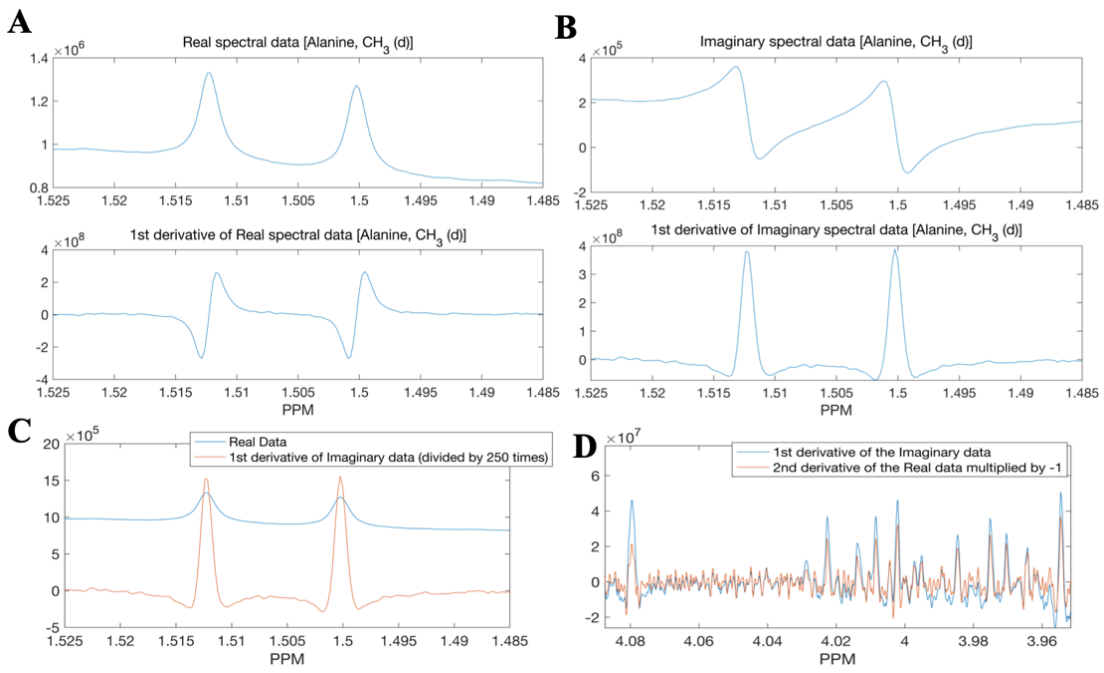
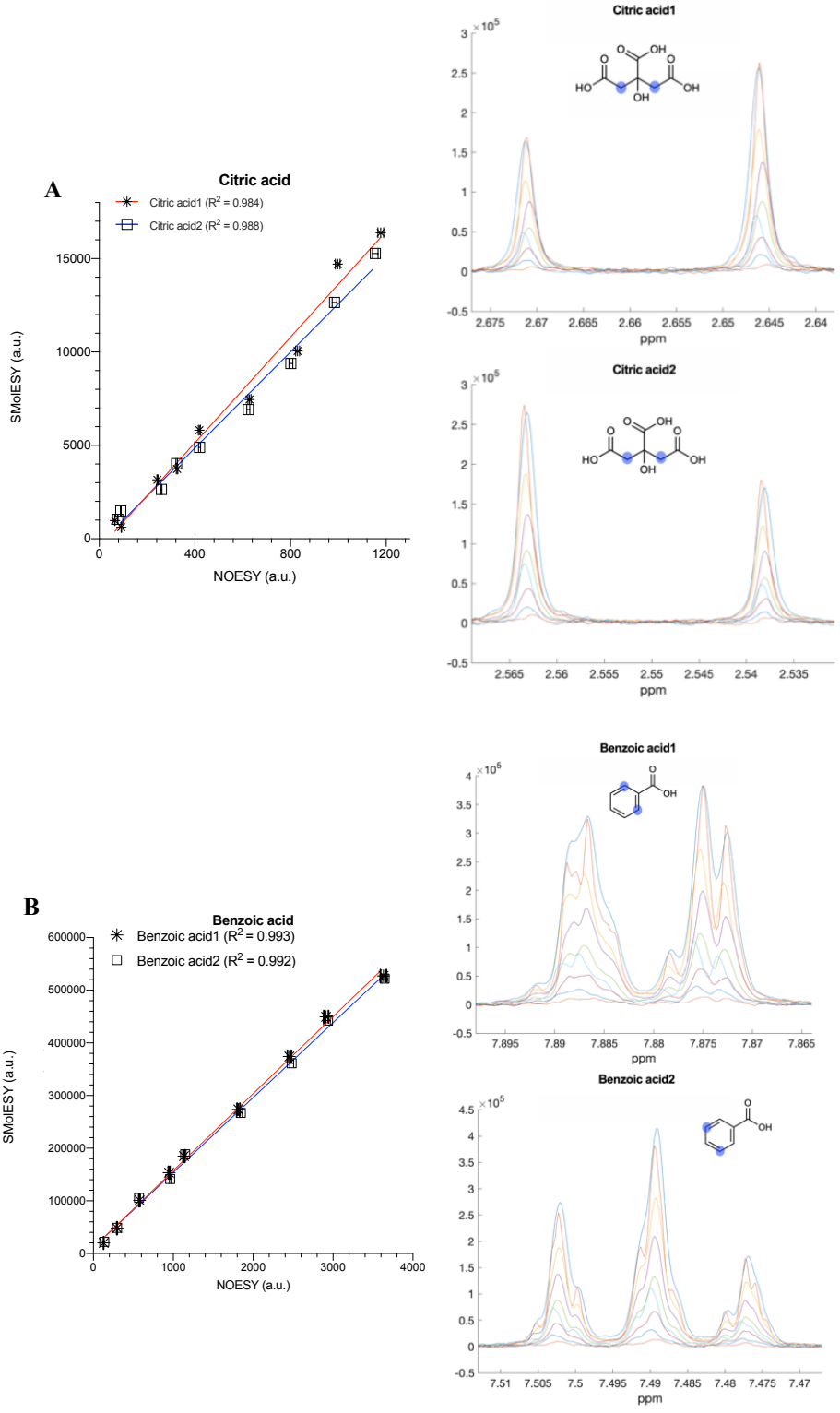
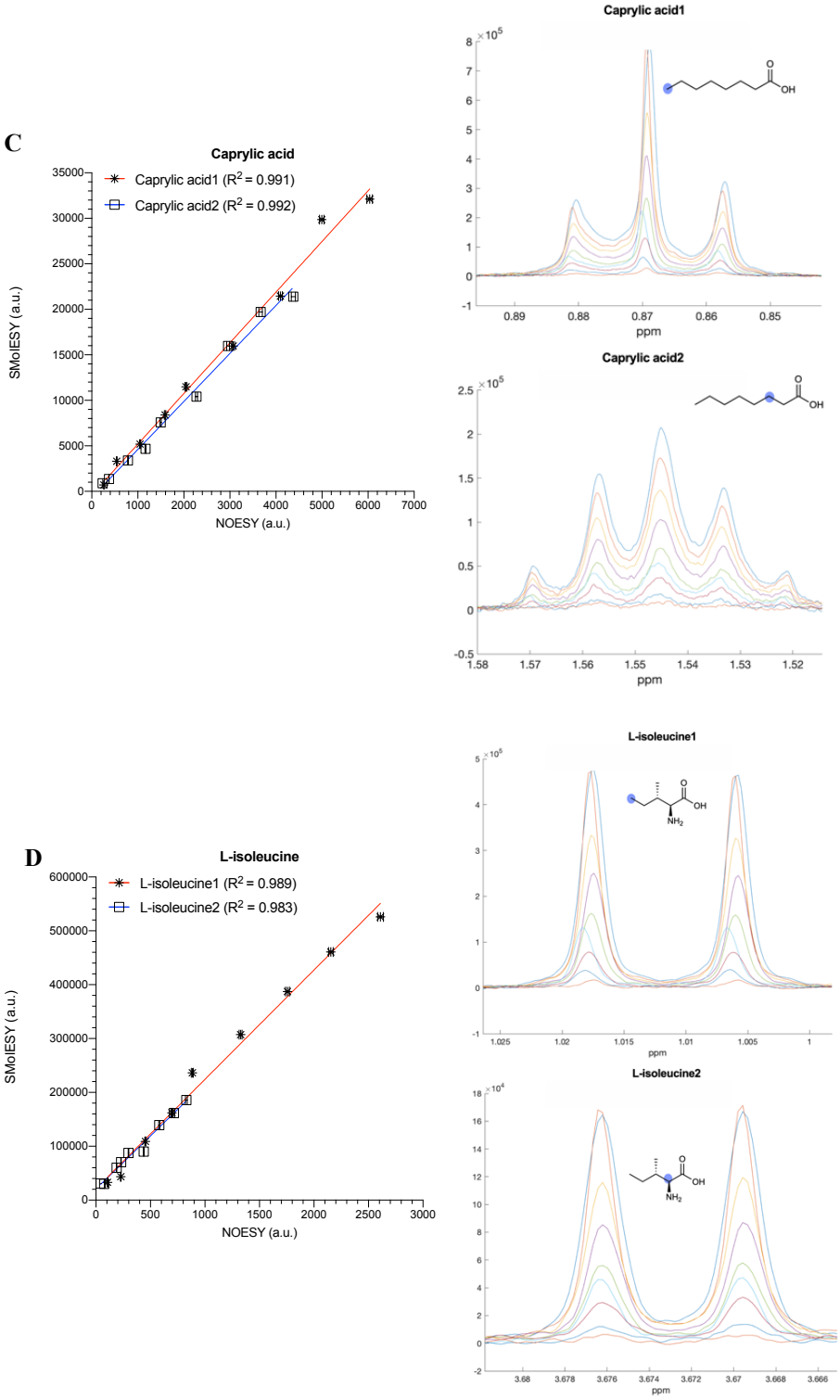


Figure S2. Examples of enhanced spectral resolution by the imaginary NMR spectral part differentiation. (A) The real spectral data (i.e. doublet, d) of the spin system from the -CH_3 group of L-alanine in a typical plasma/serum matrix (upper panel). The 1st numerical derivative of the real data from the L-alanine -CH_3 ^{1}H-NMR signal (after Fourier transform and phase correction) (bottom panel), produces an antisymmetric signal (positive on one side and negative on the other). (B) The imaginary spectral data of the spin system from the -CH_3 group of L-alanine in a typical plasma/serum matrix (upper panel). In contrary to the real data, the 1st derivative of the imaginary data, due to its gradient (namely positive-negative maxima per signal) (bottom panel), produces a positive transformed signal. (C) Overlaid real and 1st derivative of the imaginary part of the L-alanine -CH_3 doublet spectral regions, show no chemical shifting, without the need of applying any symmetrisation algorithms. The transformed signal from the imaginary spectral data could be immediately employed for any NMR-based metabolomics or analytical study. (D) Comparison between the 2nd derivative of the real data of the NMR spectrum multiplied by -1 (this could be the same for the 2nd power derivative) and the 1st derivative of the imaginary part of the same spectral region, taken from a ^{1}H-NMR plasma spectrum. It is immediately appreciated that the signal-to-noise ratio of the 2nd derivative of the real spectral data is decreased compared to the 1st derivative of the imaginary part.

Testing metabolites intra-reproducibility by artificial mixtures of metabolites




C
Caprylic acid
Caprylic acid1 (R^2 = 0.991)
Caprylic acid2 (R^2 = 0.992)
SMoIESY (a.u.)
NOESY (a.u.)
Caprylic acid1
Caprylic acid2
ppm
D
L-isoleucine
L-isoleucine1 (R^2 = 0.989)
L-isoleucine2 (R^2 = 0.983)
SMoIESY (a.u.)
NOESY (a.u.)
L-isoleucine1
L-isoleucine2
ppm

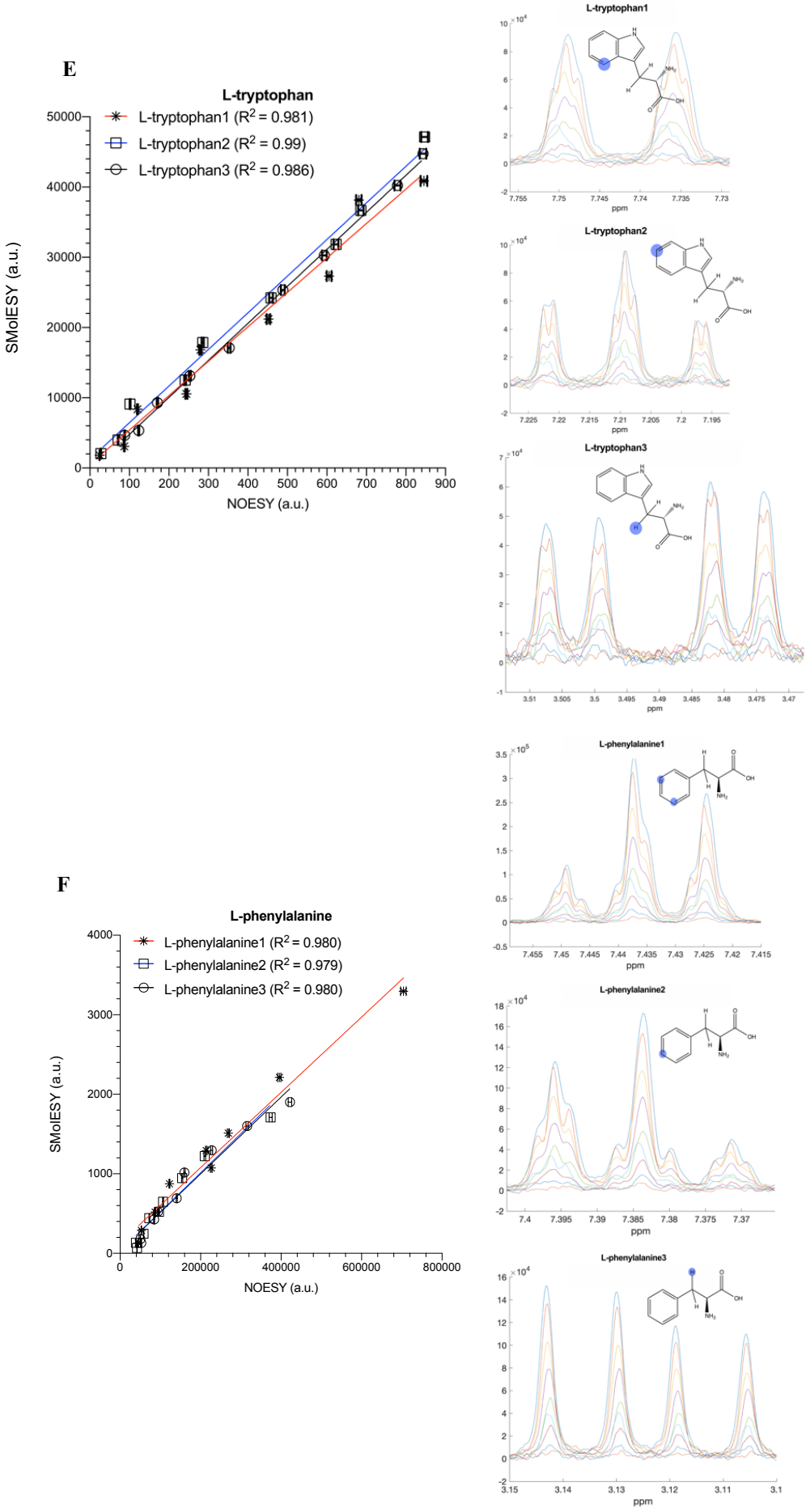

E
L-tryptophan
L-tryptophan1 (R2 = 0.981)
L-tryptophan2 (R2 = 0.99)
L-tryptophan3 (R2 = 0.986)
SMolESY (a.u.)
NOESY (a.u.)
L-tryptophan1
L-tryptophan2
L-tryptophan3
ppm
F
L-phenylalanine
L-phenylalanine1 (R2 = 0.980)
L-phenylalanine2 (R2 = 0.979)
L-phenylalanine3 (R2 = 0.980)
SMolESY (a.u.)
NOESY (a.u.)
L-phenylalanine1
L-phenylalanine2
L-phenylalanine3
ppm

Figure S3. Validation of SMolESY intra-metabolites signals reproducibility. Plots of the fitted linear regression curves ($R^2 \geq 0.98$) based upon the SMolESY versus 1D-NOESY integrals from various ^1H spin systems (highlighted by light blue circles) with different multiplicities (horizontal/vertical error bars point at ±1 % error in integration) of (**A**) Citric acid, (**B**) Benzoic acid, (**C**) Caprylic acid, (**D**) L-isoleucine, (**E**) L-tryptophan and (**F**) L-phenylalanine in 9 concentrations in the artificial metabolites mixtures. All slope, intercepts as well as one-way ANOVA tests for the confirmed all intercepts/slopes coincidence for each metabolite (see below Table S1). These results highlight the intra-metabolites SMolESY signals reproducibility.

Table S1. Statistical analyses data for SMolESY intra-metabolites signals reproducibility tests. Statistical data of the fitted linear regression curves for the SMolESY versus 1D-NOESY integrals of NMR signals from various ^{1}H spin systems from several metabolites in 9 concetrations in the artificial metabolites mixtures. All linear curves statistically pass through the zero point and according to the one-way ANOVA tests both slopes and intercepts of all curves for each metabolite coincide.

Best-fit values	**Cytidine 1**	**Cytidine 2**	**Cytidine 3**	**Cytidine 4**	**Cytidine 5**	**one-way ANOVA results**
Slope	65.41	66.63	68.62	67.53	65.03	
Y-intercept	3672	30.63	-3676	-1133	-2303	For the slopes
X-intercept	-56.14	-0.4597	53.57	16.77	35.41	F = 0.265
1/slope	0.01529	0.01501	0.01457	0.01481	0.01538	DFn† = 4, DFd‡ = 35
Std. Error						P = 0.8984
Slope	1.854	3.673	2.827	2.406	3.347	The pooled slope equals 66.59
Y-intercept	2413	4415	3544	3079	4310	
95% Confidence Intervals						For the intercepts
Slope	61.02 to 69.79	57.94 to 75.32	61.93 to 75.31	61.84 to 73.22	57.12 to 72.94	F = 1.291
Y-intercept	-2034 to 9377	-10409 to 10470	-12056 to 4704	-8413 to 6148	-12493 to 7887	DFn† = 4, DFd‡ = 39
X-intercept	-151.6 to 29.54	-175.8 to 142	-74.53 to 163.1	-97.89 to 116.7	-134.9 to 175.3	P = 0.2905 The pooled intercept equals 630

Best-fit values	**Citric acid 1**	**Citric acid 2**				**one-way ANOVA results**
Slope	14.15	12.87				
Y-intercept	-523.2	-287.5				For the slopes
X-intercept	36.98	22.35				F = 2.165
1/slope	0.07068	0.07772				DFn† = 1, DFd‡ = 14
Std. Error						P = 0.1633
Slope	0.6842	0.5335				The pooled slope equals 13.53
Y-intercept	445.7	341.9				
95% Confidence Intervals						For the intercepts
Slope	12.53 to 15.77	11.61 to 14.13				F = 1.728
Y-intercept	-1577 to 530.6	-1096 to 520.9				DFn† = 1, DFd‡ = 15
X-intercept	-41.41 to 102.3	-44.08 to 78.99				P = 0.2084 The pooled intercept equals 414.2

Best-fit values	**Benzoic acid 1**	**Benzoic acid 2**				**one-way ANOVA results**
Slope	146.2	142.3				
Y-intercept	11191	11549				For the slopes
X-intercept	-76.55	-81.15				F = 0.91
1/slope	0.00684	0.007027				DFn† = 1, DFd‡ = 14
Std. Error						P = 0.3563
Slope	2.747	3.021				The pooled slope equals 144.2
Y-intercept	5276	5887				
95% Confidence Intervals						For the intercepts
Slope	139.7 to 152.6	135.2 to 149.4				F = 1.434
Y-intercept	-1285 to 23667	-2371 to 25468				DFn† = 1, DFd‡ = 15
X-intercept	-167.9 to 8.492	-186.6 to 16.02				P = 0.2496 The pooled intercept equals 11376

Best-fit values	**Caprylic acid 1**	**Caprylic acid 2**				**one-way ANOVA results**
Slope	5.566	5.273				For the slopes
Y-intercept	-378.0	-671.1				F = 0.9582
X-intercept	67.92	127.3				DFn† = 1, DFd‡ = 14
1/slope	0.1797	0.1896				P = 0.3443
Std. Error						The pooled slope equals 5.465
Slope	0.2014	0.1998				
Y-intercept	657.2	475.5				For the intercepts
95% Confidence Intervals						F = 3.619.
Slope	5.090 to 6.042	4.801 to 5.746				DFn† = 1, DFd‡ = 15
Y-intercept	-1932 to 1176	-1795 to 453.3				P = 0.0765
X-intercept	-227.2 to 325.2	-92.74 to 318.1				The pooled intercept equals -576.7

Best-fit values	**L-isoleucine 1**	**L-isoleucine 2**				**one-way ANOVA results**
Slope	203.0	197.0				For the slopes
Y-intercept	21257	20708				F = 0.08475
X-intercept	-104.7	-105.1				DFn† = 1, DFd‡ = 14
1/slope	0.004927	0.005076				P = 0.7752
Std. Error						The pooled slope equals 202.4
Slope	8.231	9.841				
Y-intercept	11589	4539				For the intercepts
95% Confidence Intervals						F = 0.15
Slope	183.5 to 222.4	173.7 to 220.3				DFn† = 1, DFd‡ = 15
Y-intercept	-6147 to 48661	-9975 to 31441				P = 0.7039
X-intercept	-260.6 to 28.12	-177.8 to -46.09				The pooled intercept equals 20266

Best-fit values	**L-tryptophan 1**	**L-tryptophan 2**	**L-tryptophan 3**			**one-way ANOVA results**
Slope	48.99	52.14	52.68			
Y-intercept	508.7	1243	-483.6			For the slopes

X-intercept	-10.38	-23.85	9.18			F = 0.787 DFn† = 2, DFd‡ = 21 P = 0.4682 The pooled slope equals 51.26 <u>For the intercepts</u> F = 2.722 DFn† = 2, DFd‡ = 23 P = 0.0869 The pooled intercept equals 445.7
1/slope	0.02041	0.01918	0.01898			
Std. Error						
Slope	3.162	2.022	0.8736			
Y-intercept	1456	939.6	426.8			
95% Confidence Intervals						
Slope	41.51 to 56.47	47.36 to 56.92	50.62 to 54.75			
Y-intercept	-2935 to 3953	-978.3 to 3465	-1493 to 525.7			
X-intercept	-92.48 to 53.52	-71.89 to 17.49	-10.32 to 27.44			
Best-fit values	**L-phenylalanine 1**	**L-phenylalanine 2**	**L-phenylalanine 3**			**one-way ANOVA results**
Slope	0.004732	0.004916	0.004833			<u>For the slopes</u> F = 0.0642 DFn† = 2, DFd‡ = 21 P = 0.9380 The pooled slope equals 0.004785 <u>For the intercepts</u> F = 0.7911 DFn† = 2, DFd‡ = 23 P = 0.4653 The pooled intercept equals 68.56
Y-intercept	131.5	29.58	31.80			
X-intercept	-27794	-6016	-6580			
1/slope	211.3	203.4	206.9			
Std. Error						
Slope	0.0002872	0.0004470	0.0003886			
Y-intercept	88.23	73.07	81.53			
95% Confidence Intervals						
Slope	0.004053 to 0.005411	0.003859 to 0.005973	0.003914 to 0.005751			
Y-intercept	-77.10 to 340.1	-143.2 to 202.4	-161.0 to 224.6			
X-intercept	-81350 to 14700	-50122 to 25080	-55497 to 28945			

†DFn: degrees of freedom numerator.

‡DFd: degrees of freedom denominator.

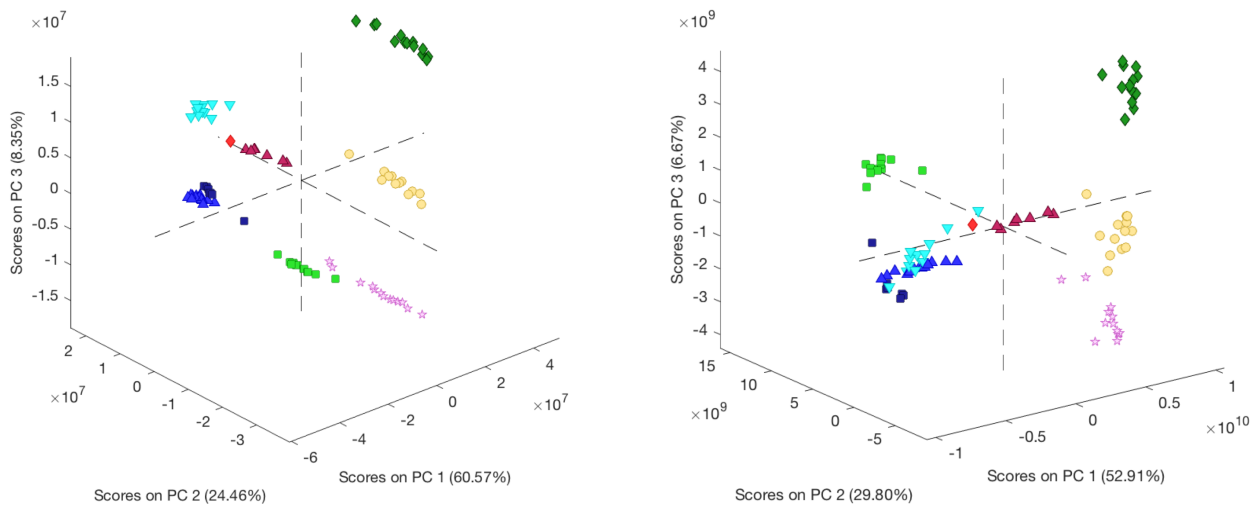
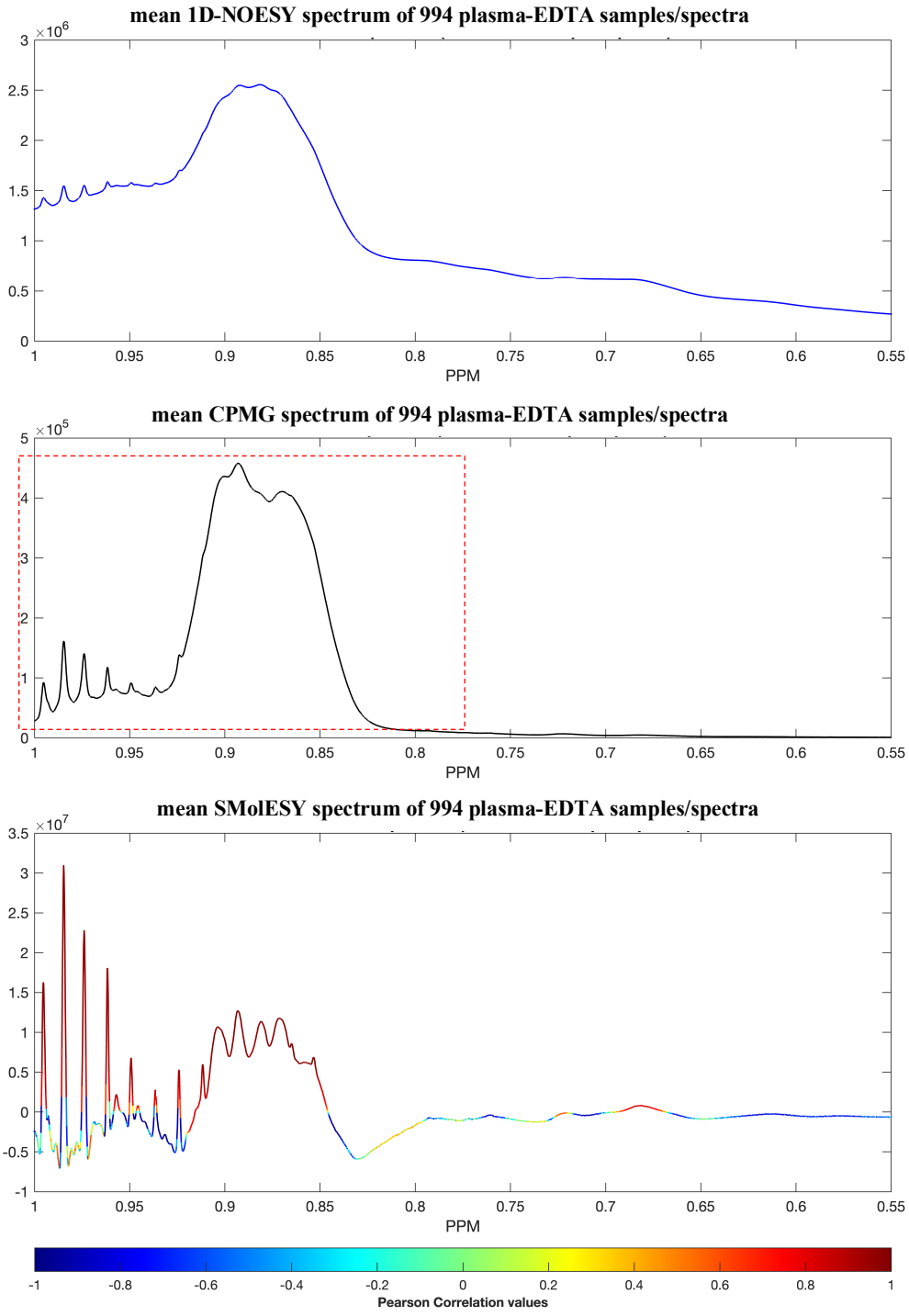


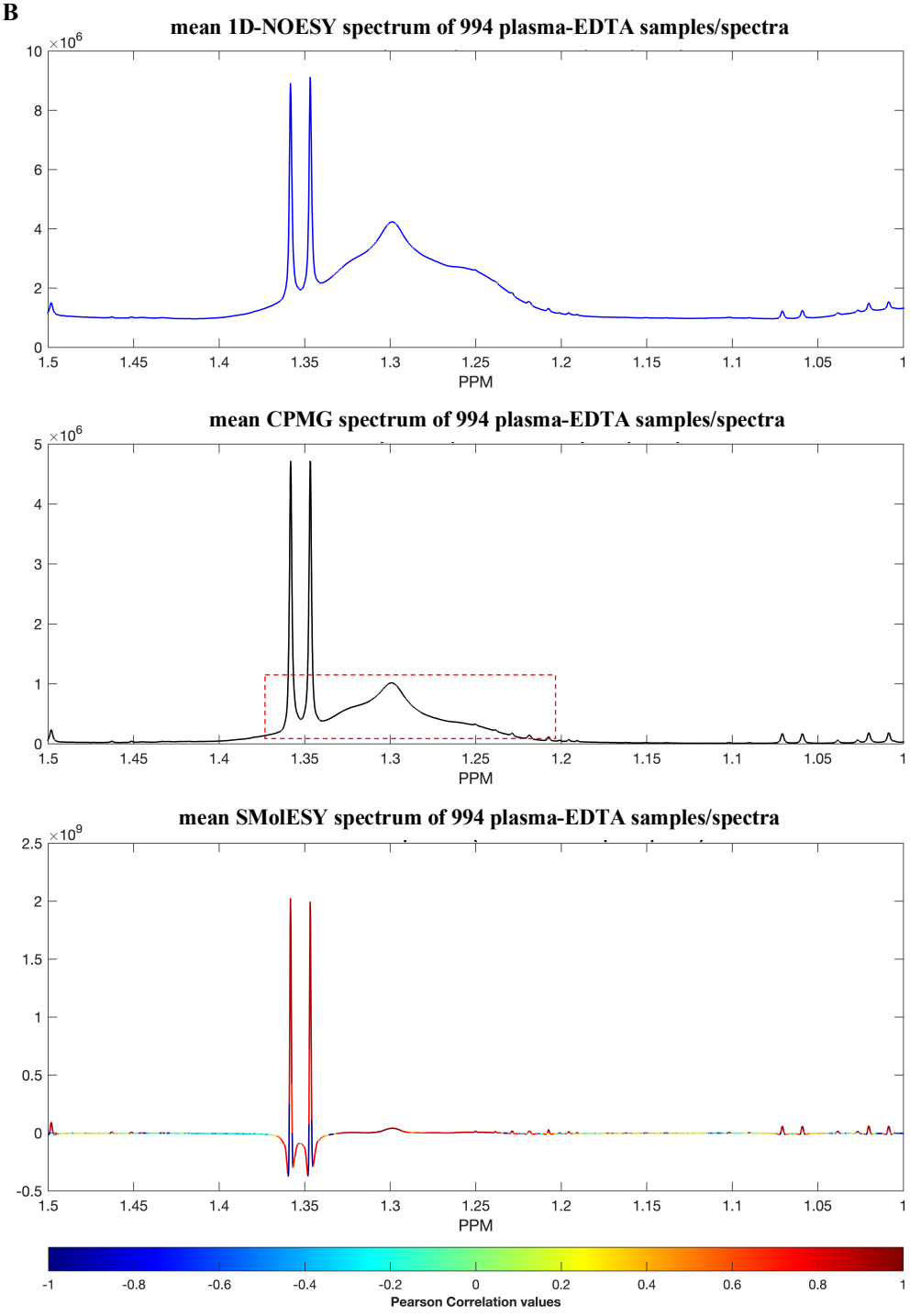
Figure S4. 3D scores plots of the first three components from the PCA of a urine dataset for both 1D-NOESY and SMolESY spectra. From the analysis, it is observed that both NOESY and SMolESY spectra capture similar cumulative variability, 93.4% and 89.4% respectively, and provide similar discrimination of sample groups.

Correlation of SMolESY vs. CPMG spectra for 994 plasma-EDTA samples

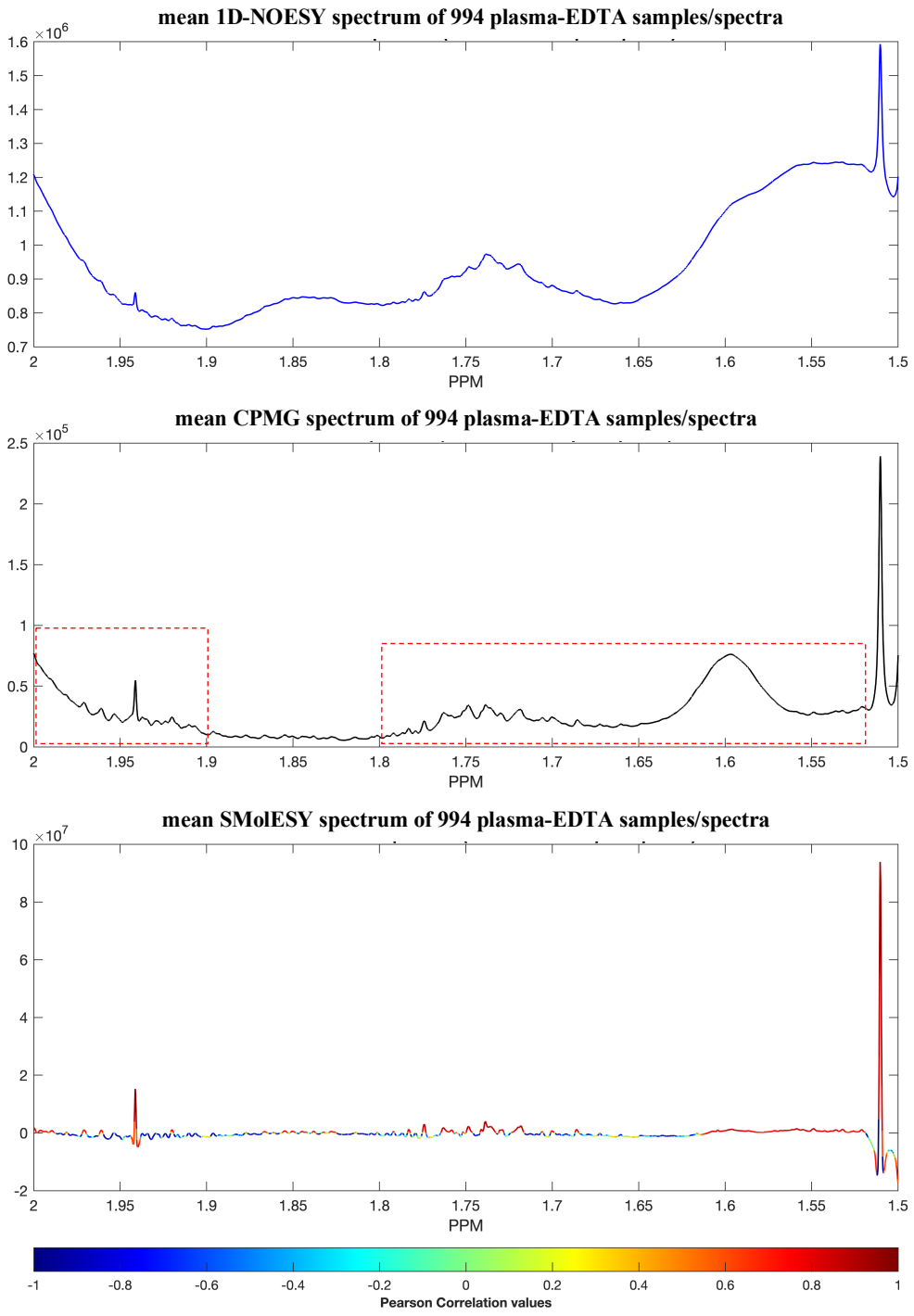
A



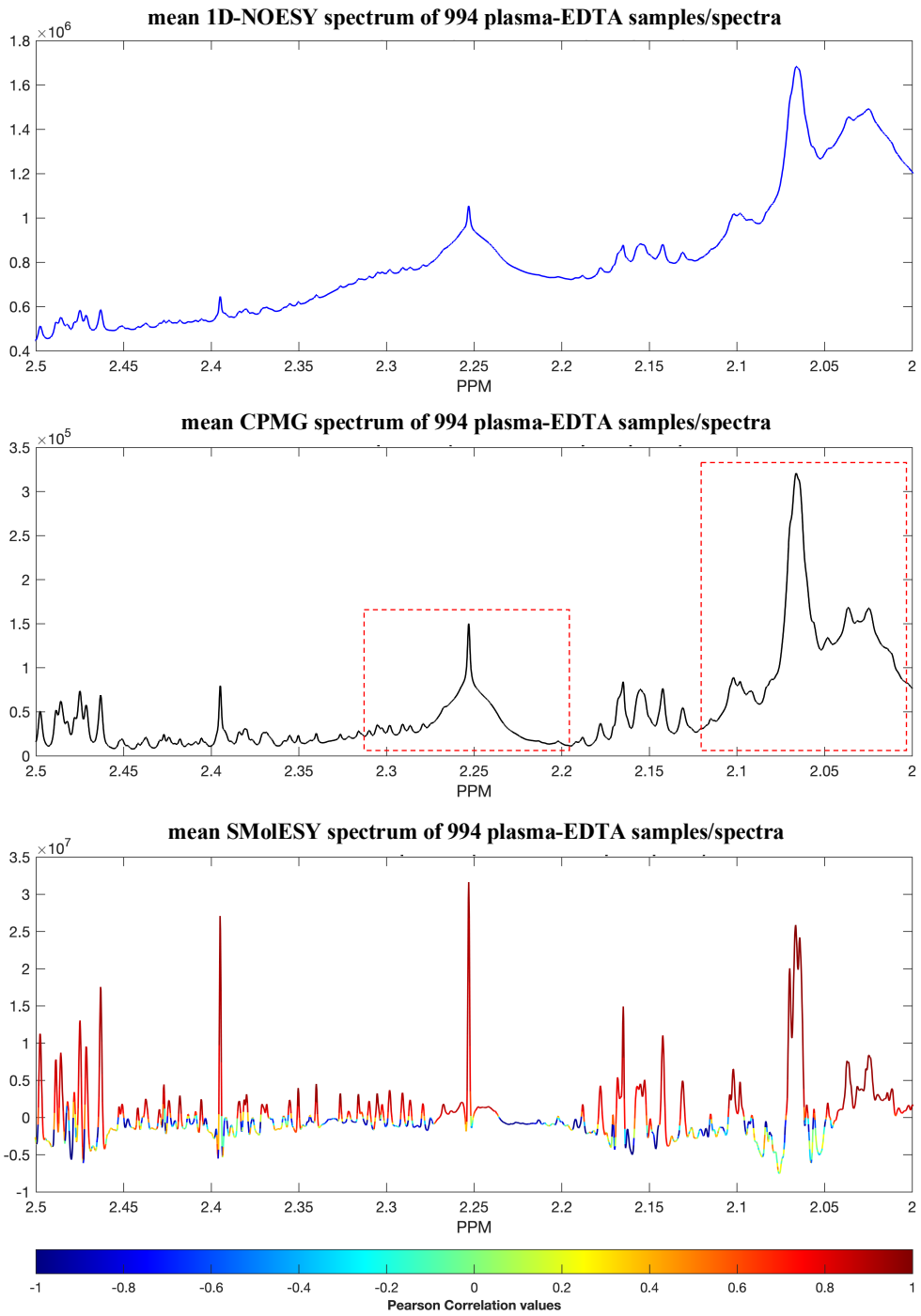
B


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^6
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^9
PPM
Pearson Correlation values

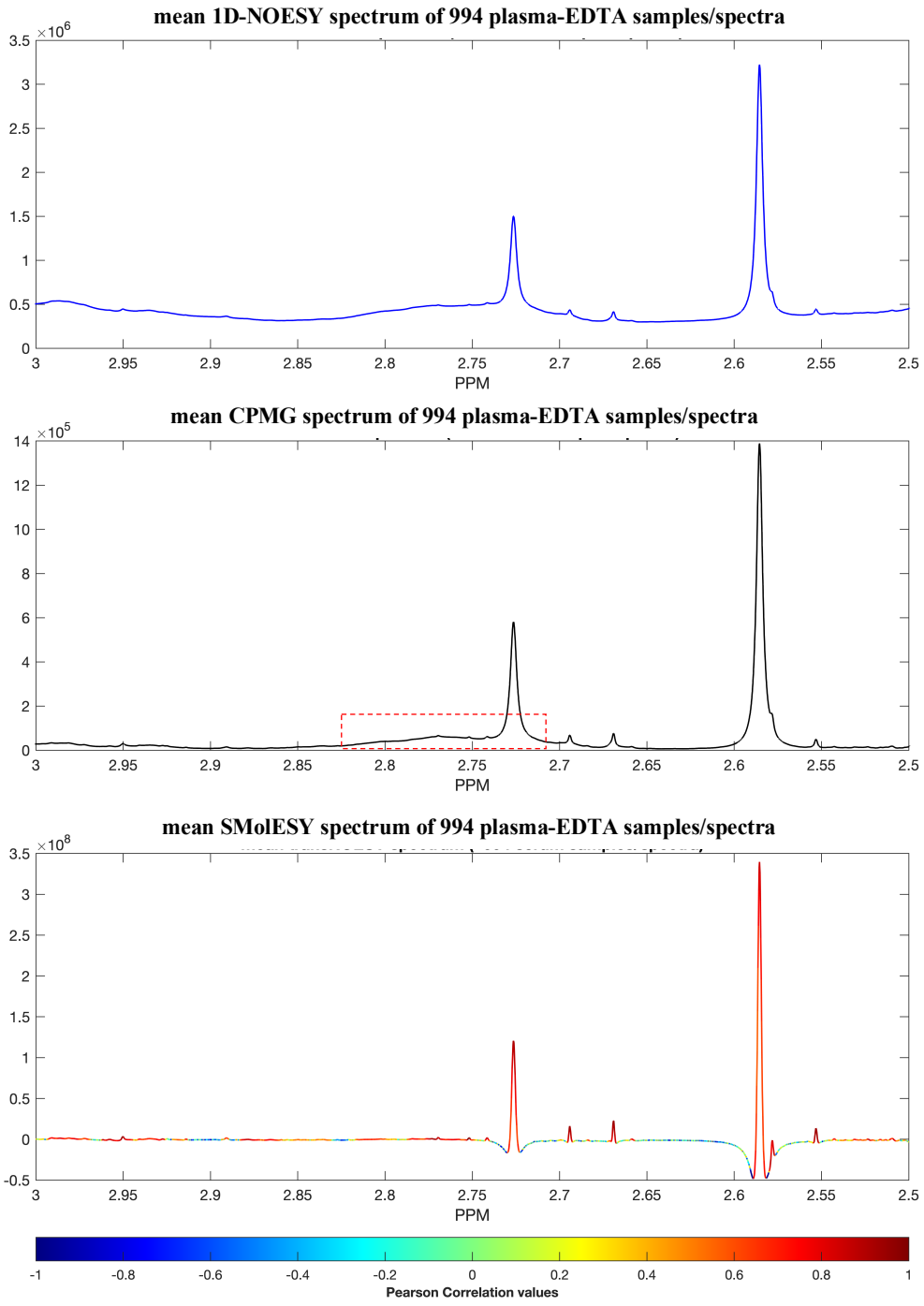
C


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
2
1.95
1.9
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
2.5
2
1.5
1
0.5
0
2
1.95
1.9
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^7
10
8
6
4
2
0
-2
2
1.95
1.9
1.85
1.8
1.75
1.7
1.65
1.6
1.55
1.5
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

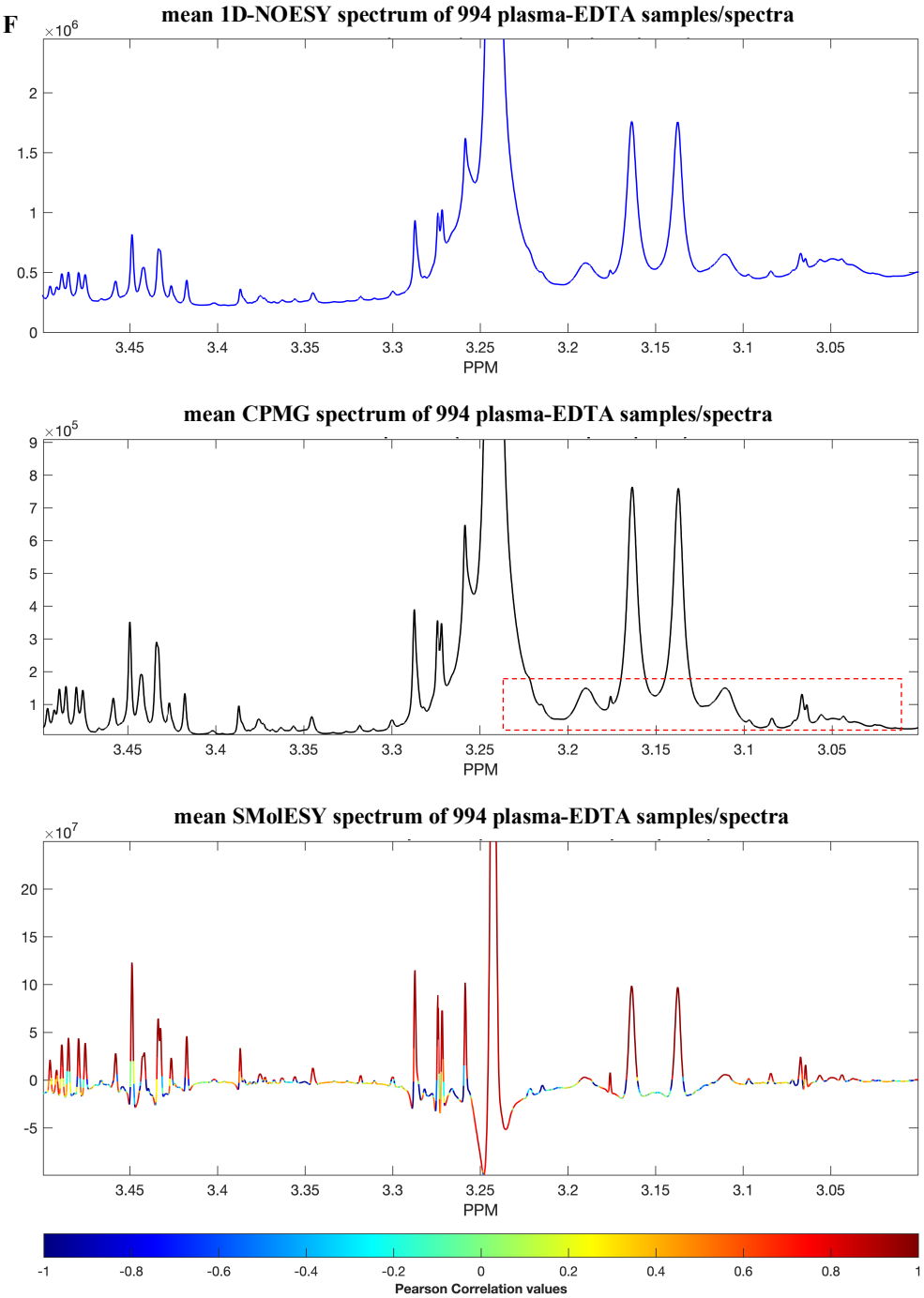
D


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
2.5 2.45 2.4 2.35 2.3 2.25 2.2 2.15 2.1 2.05 2
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
3.5
3
2.5
2
1.5
1
0.5
0
2.5 2.45 2.4 2.35 2.3 2.25 2.2 2.15 2.1 2.05 2
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^7
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
2.5 2.45 2.4 2.35 2.3 2.25 2.2 2.15 2.1 2.05 2
PPM
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Pearson Correlation values

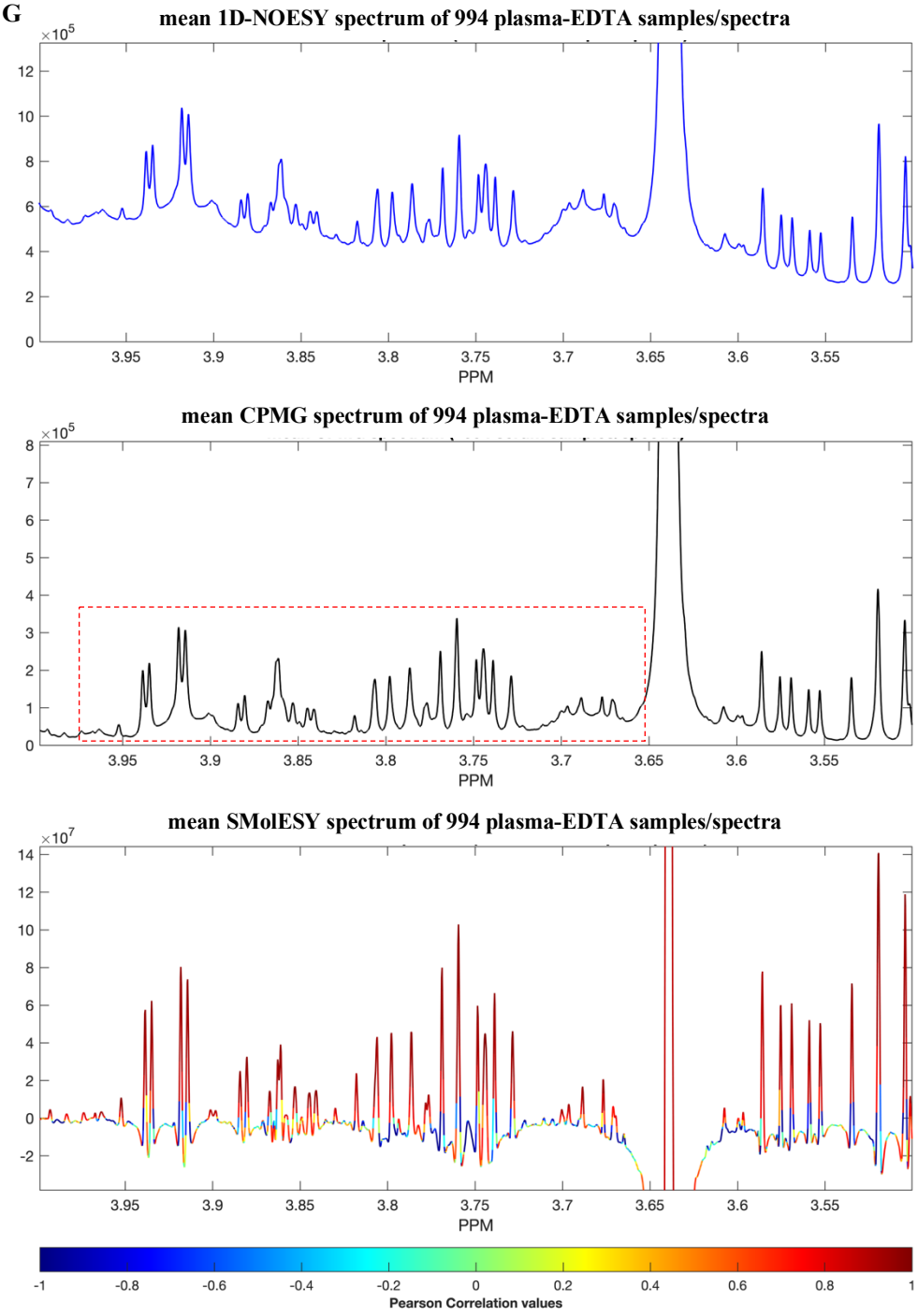
E


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
3.5
3
2.5
2
1.5
1
0.5
0
3 2.95 2.9 2.85 2.8 2.75 2.7 2.65 2.6 2.55 2.5
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
14
12
10
8
6
4
2
0
3 2.95 2.9 2.85 2.8 2.75 2.7 2.65 2.6 2.55 2.5
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^8
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
3 2.95 2.9 2.85 2.8 2.75 2.7 2.65 2.6 2.55 2.5
PPM
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Pearson Correlation values

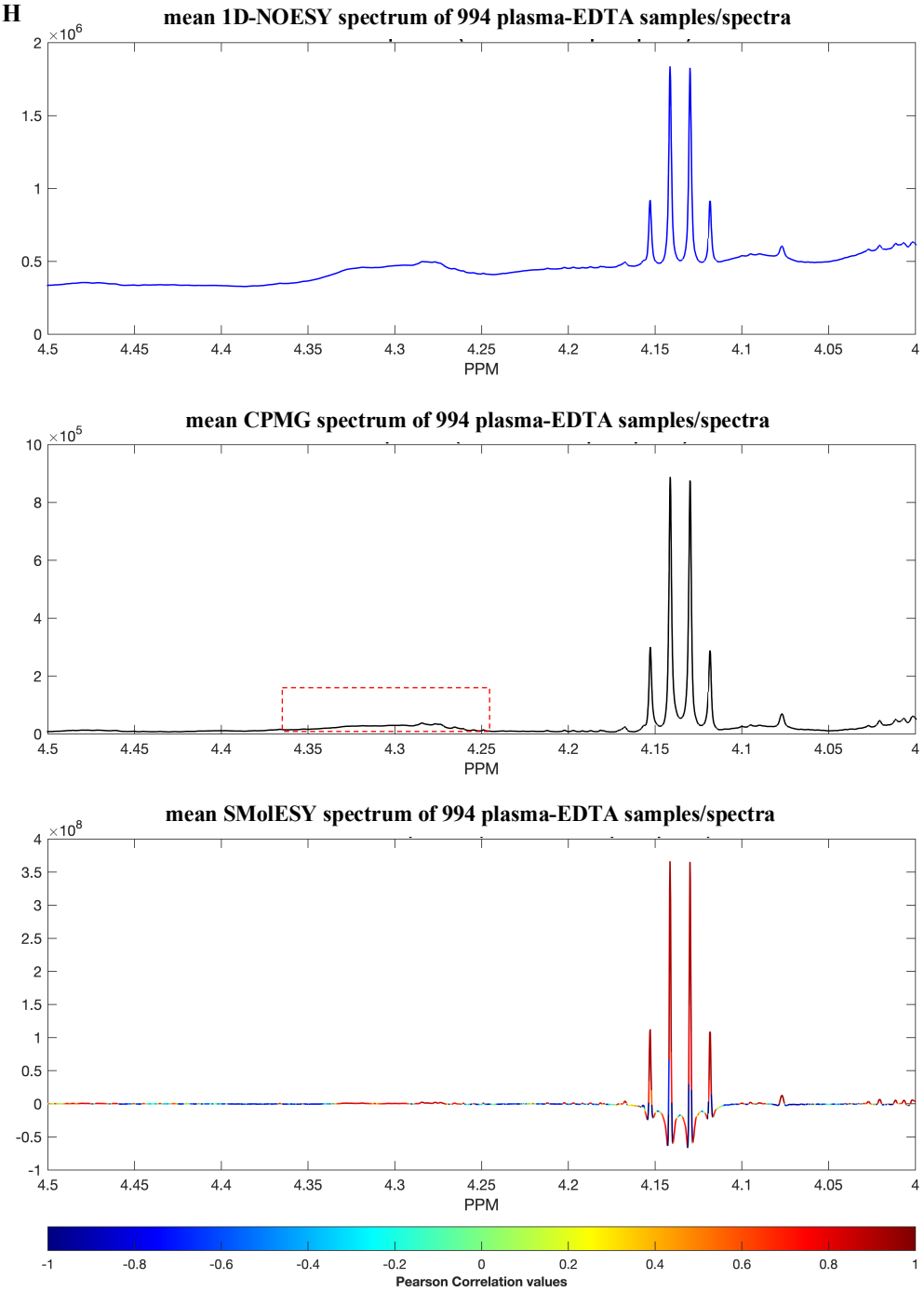
F


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
2
1.5
1
0.5
0
3.45
3.4
3.35
3.3
3.25
3.2
3.15
3.1
3.05
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
9
8
7
6
5
4
3
2
1
3.45
3.4
3.35
3.3
3.25
3.2
3.15
3.1
3.05
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^7
20
15
10
5
0
-5
3.45
3.4
3.35
3.3
3.25
3.2
3.15
3.1
3.05
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

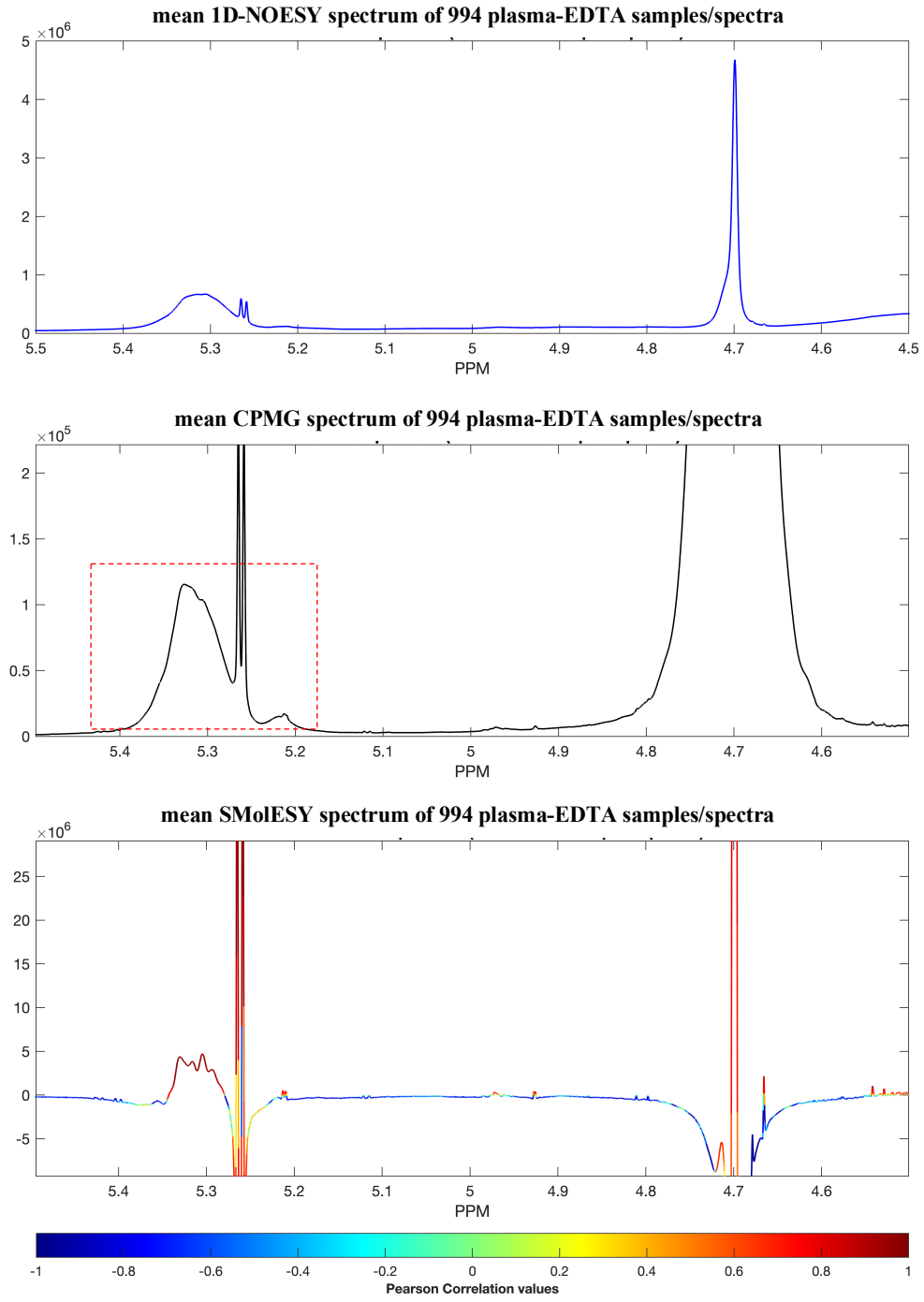
G


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^5
12
10
8
6
4
2
0
3.95
3.9
3.85
3.8
3.75
3.7
3.65
3.6
3.55
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
8
7
6
5
4
3
2
1
0
3.95
3.9
3.85
3.8
3.75
3.7
3.65
3.6
3.55
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^7
14
12
10
8
6
4
2
0
-2
3.95
3.9
3.85
3.8
3.75
3.7
3.65
3.6
3.55
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

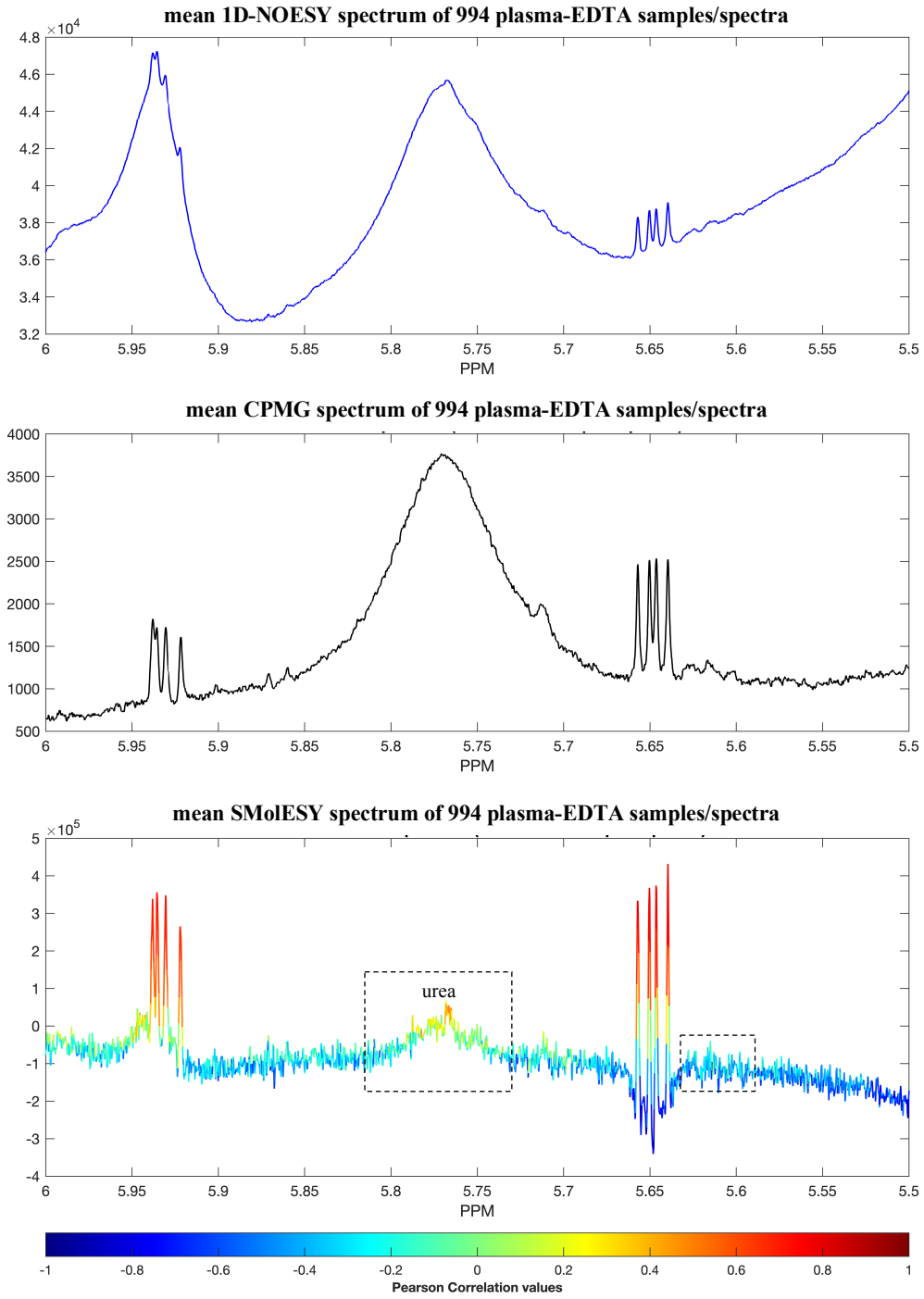
H


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
2
1.5
1
0.5
0
4.5 4.45 4.4 4.35 4.3 4.25 4.2 4.15 4.1 4.05 4
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
10
8
6
4
2
0
4.5 4.45 4.4 4.35 4.3 4.25 4.2 4.15 4.1 4.05 4
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^8
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
4.5 4.45 4.4 4.35 4.3 4.25 4.2 4.15 4.1 4.05 4
PPM
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
Pearson Correlation values

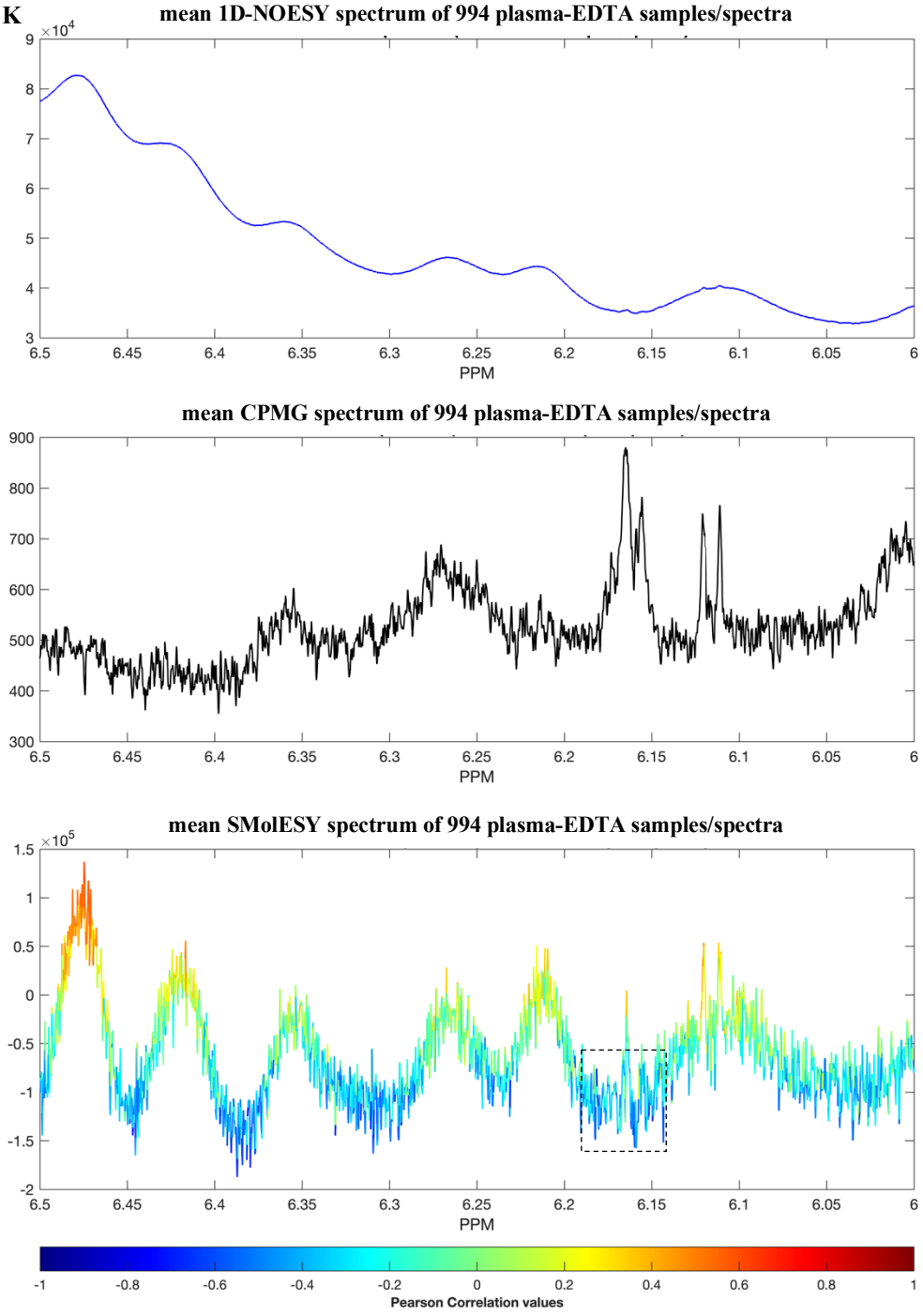
I


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^5
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
PPM
Pearson Correlation values

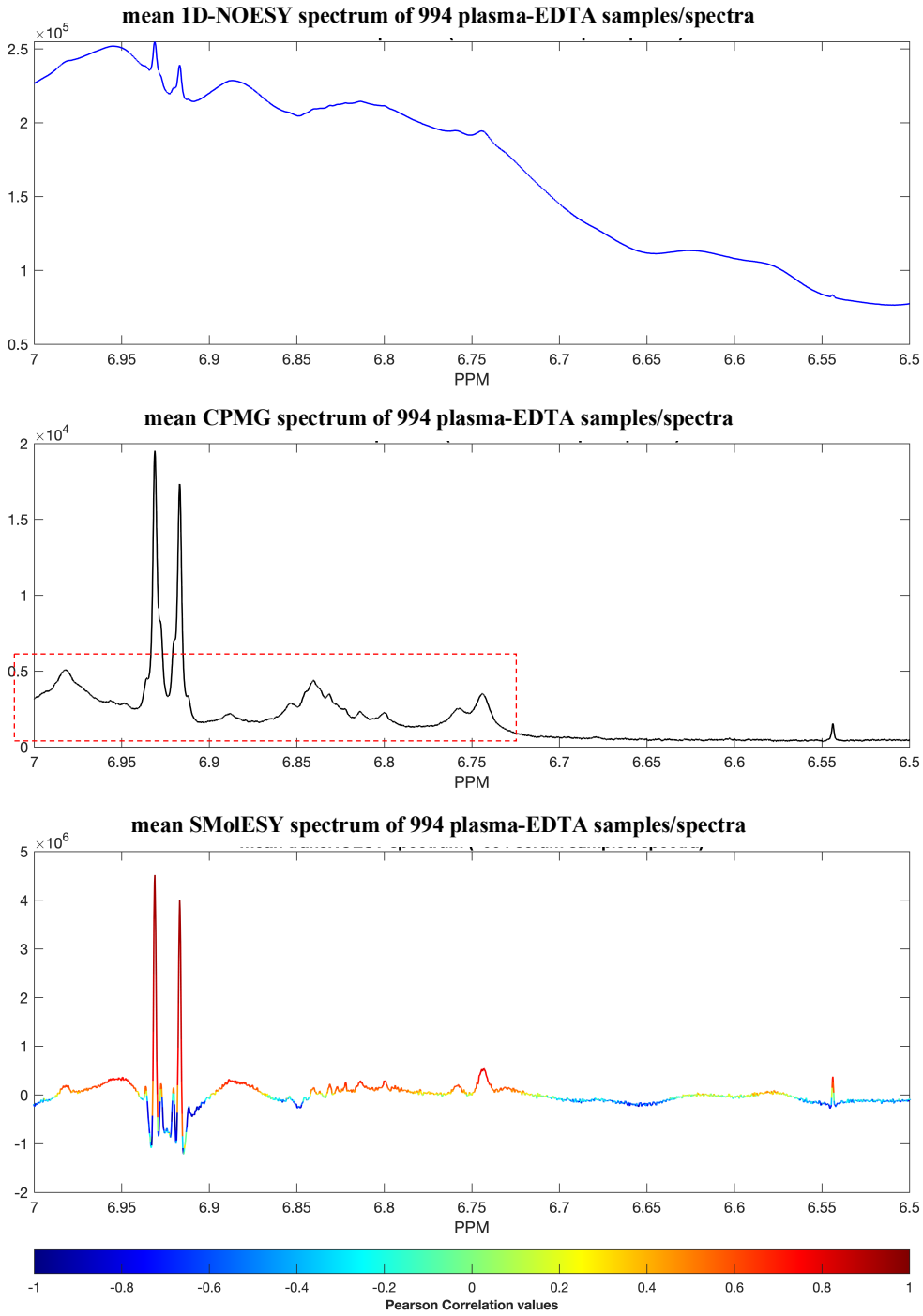
J


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^4
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
3.2
6
5.95
5.9
5.85
5.8
5.75
5.7
5.65
5.6
5.55
5.5
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
4000
3500
3000
2500
2000
1500
1000
500
6
5.95
5.9
5.85
5.8
5.75
5.7
5.65
5.6
5.55
5.5
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^5
5
4
3
2
1
0
-1
-2
-3
-4
urea
6
5.95
5.9
5.85
5.8
5.75
5.7
5.65
5.6
5.55
5.5
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

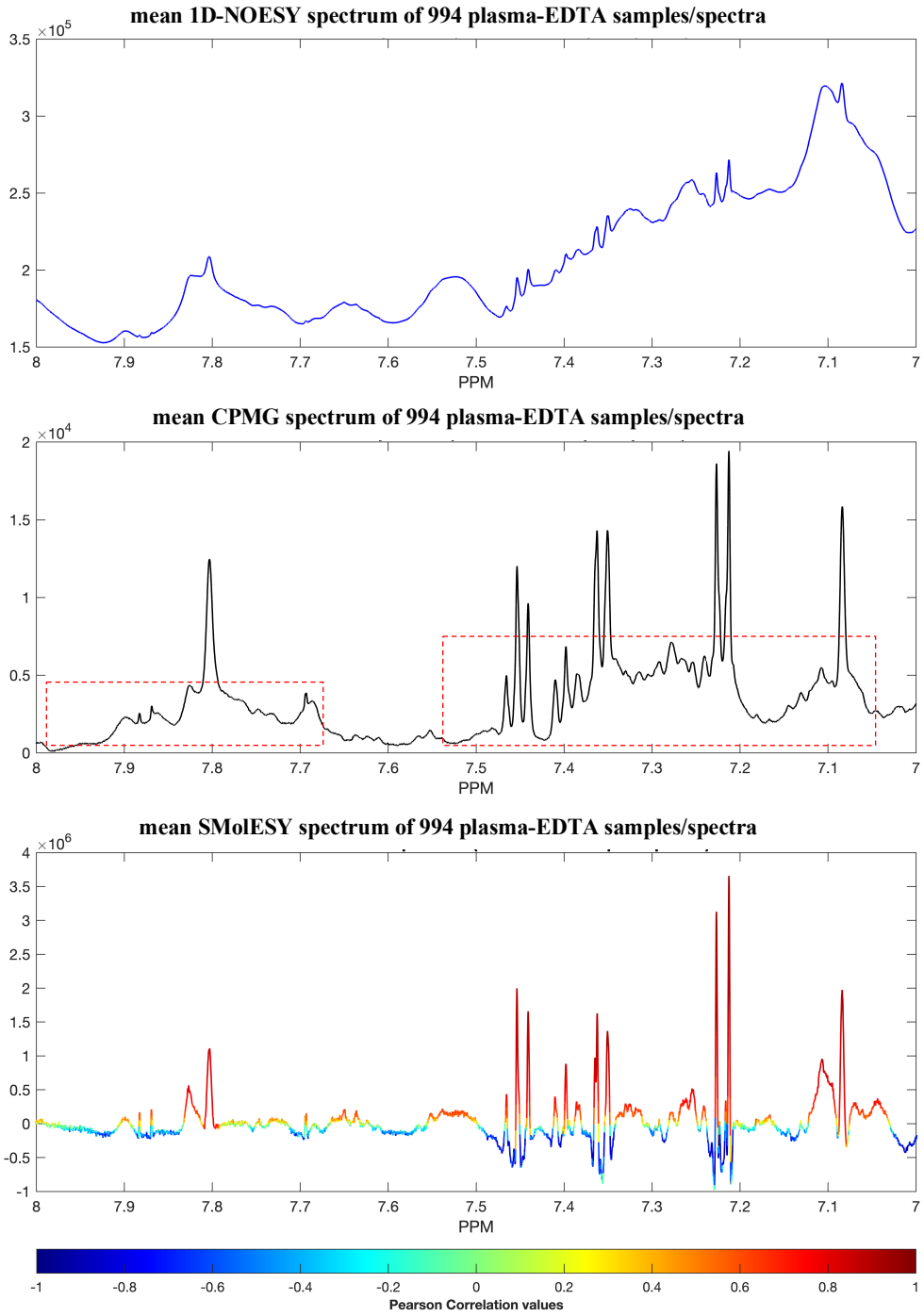
K


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^4
9
8
7
6
5
4
3
6.5
6.45
6.4
6.35
6.3
6.25
6.2
6.15
6.1
6.05
6
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
900
800
700
600
500
400
300
6.5
6.45
6.4
6.35
6.3
6.25
6.2
6.15
6.1
6.05
6
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^5
1.5
1
0.5
0
-0.5
-1
-1.5
-2
6.5
6.45
6.4
6.35
6.3
6.25
6.2
6.15
6.1
6.05
6
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

L


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^5
2.5
2
1.5
1
0.5
7
6.95
6.9
6.85
6.8
6.75
6.7
6.65
6.6
6.55
6.5
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^4
2
1.5
1
0.5
0
7
6.95
6.9
6.85
6.8
6.75
6.7
6.65
6.6
6.55
6.5
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
5
4
3
2
1
0
-1
-2
7
6.95
6.9
6.85
6.8
6.75
6.7
6.65
6.6
6.55
6.5
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

M


mean 1D-NOESY spectrum of 994 plasma-EDTA samples/spectra
×10^5
3.5
3
2.5
2
1.5
8
7.9
7.8
7.7
7.6
7.5
7.4
7.3
7.2
7.1
7
PPM
mean CPMG spectrum of 994 plasma-EDTA samples/spectra
×10^4
2
1.5
1
0.5
0
PPM
mean SMolESY spectrum of 994 plasma-EDTA samples/spectra
×10^6
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
PPM
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Pearson Correlation values

N

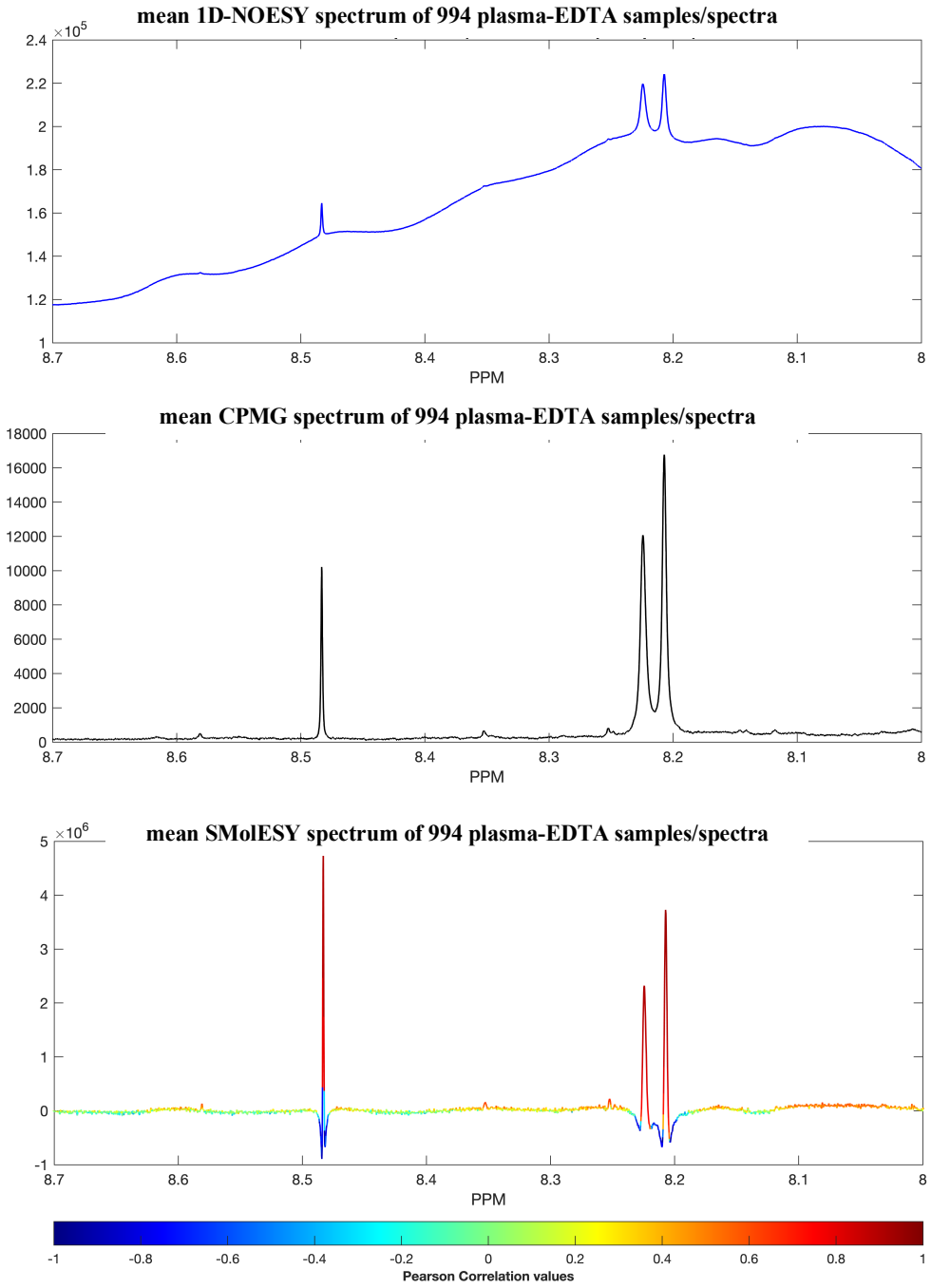


Figure S5. SMolESY performance in 994 plasma-EDTA samples. (A-N) Mean spectrum of 994 plasma-EDTA 1D-NOESY, CPMG and SMolESY ^{1}H-NMR spectra (0.55–8.7 ppm) zoomed at ~0.5-0.7 ppm windows. SMolESY is colored according to the Pearson coefficients from SMolESY versus CPMG signals correlation in 994 spectra. The majority of highly resolved SMolESY signals are linearly correlated to the CPMG and > 99.5 % of CPMG features of low molecular weight metabolites are maintained, while successfully suppressing the broad signals of macromolecules in contrast to CPMG (examples of unsuppressed CPMG broad signals are highlighted by red dashed boxes in panels A-I, L, M). It is noted that broad signal of urea along with 3 broad signals of very low abundance metabolites (< 1.5 times the signal-noise-ratio) are highly suppressed and low correlated to the CPMG (black dashed boxes in panels K, J), even though being recovered by the SMolESY.

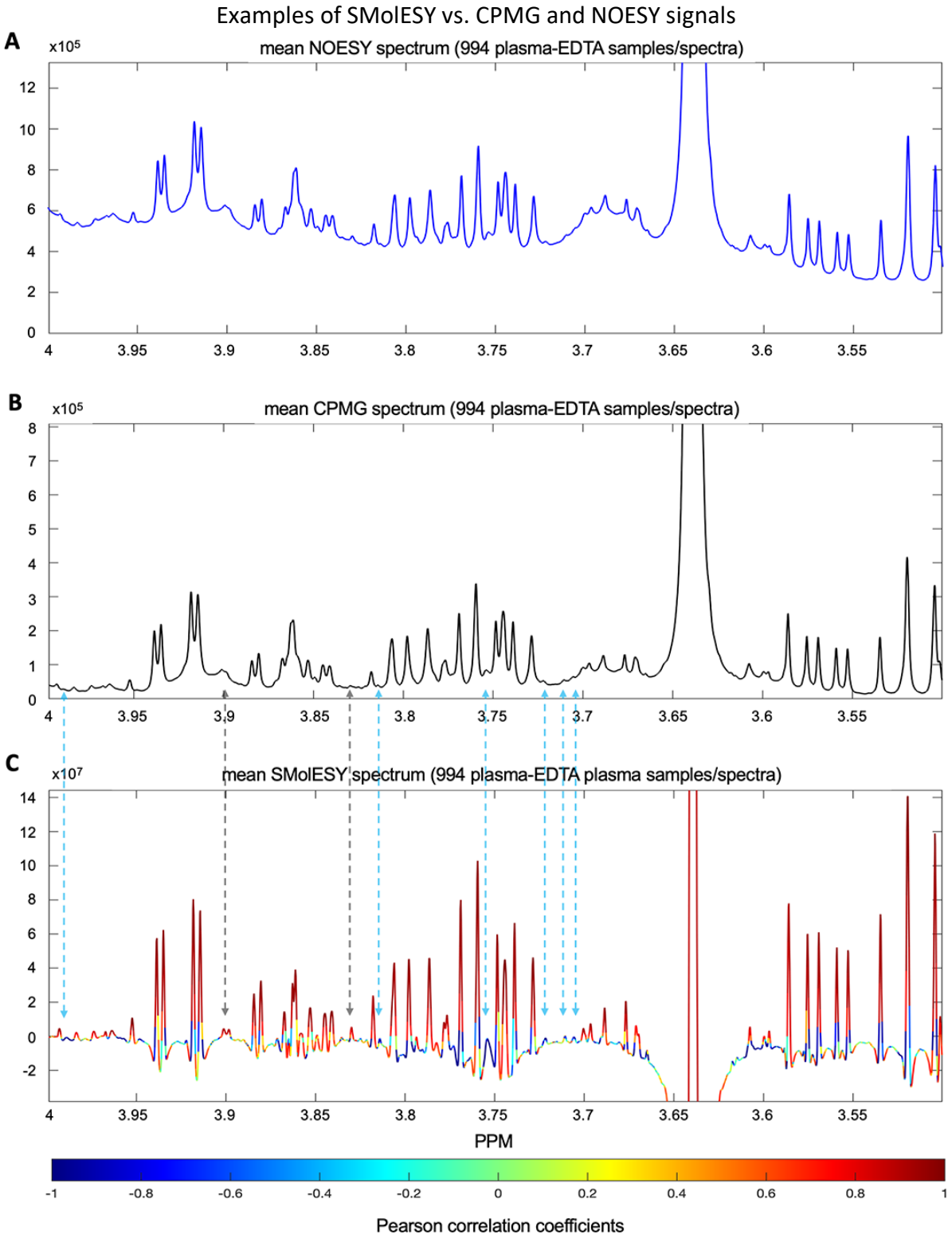
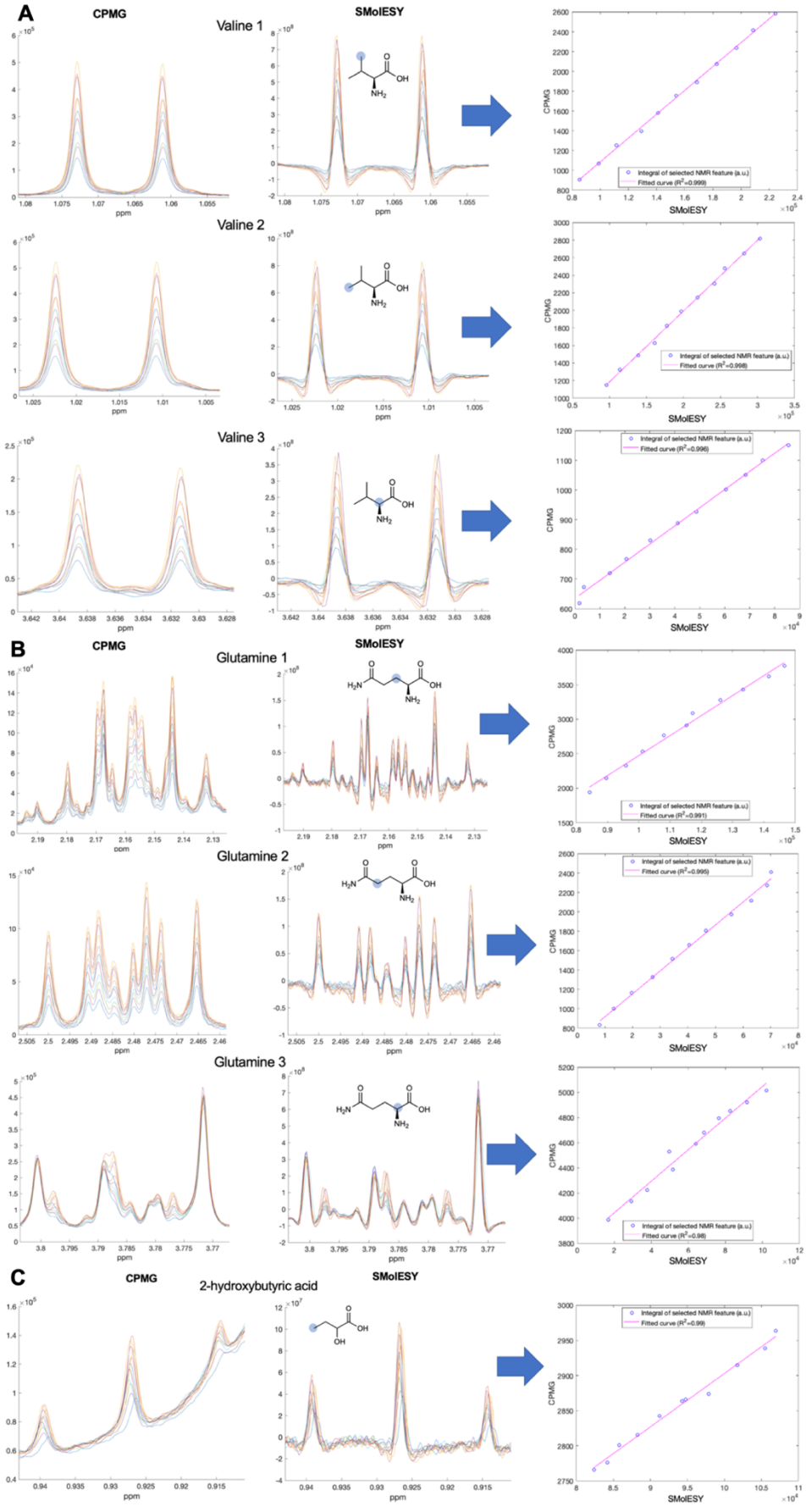
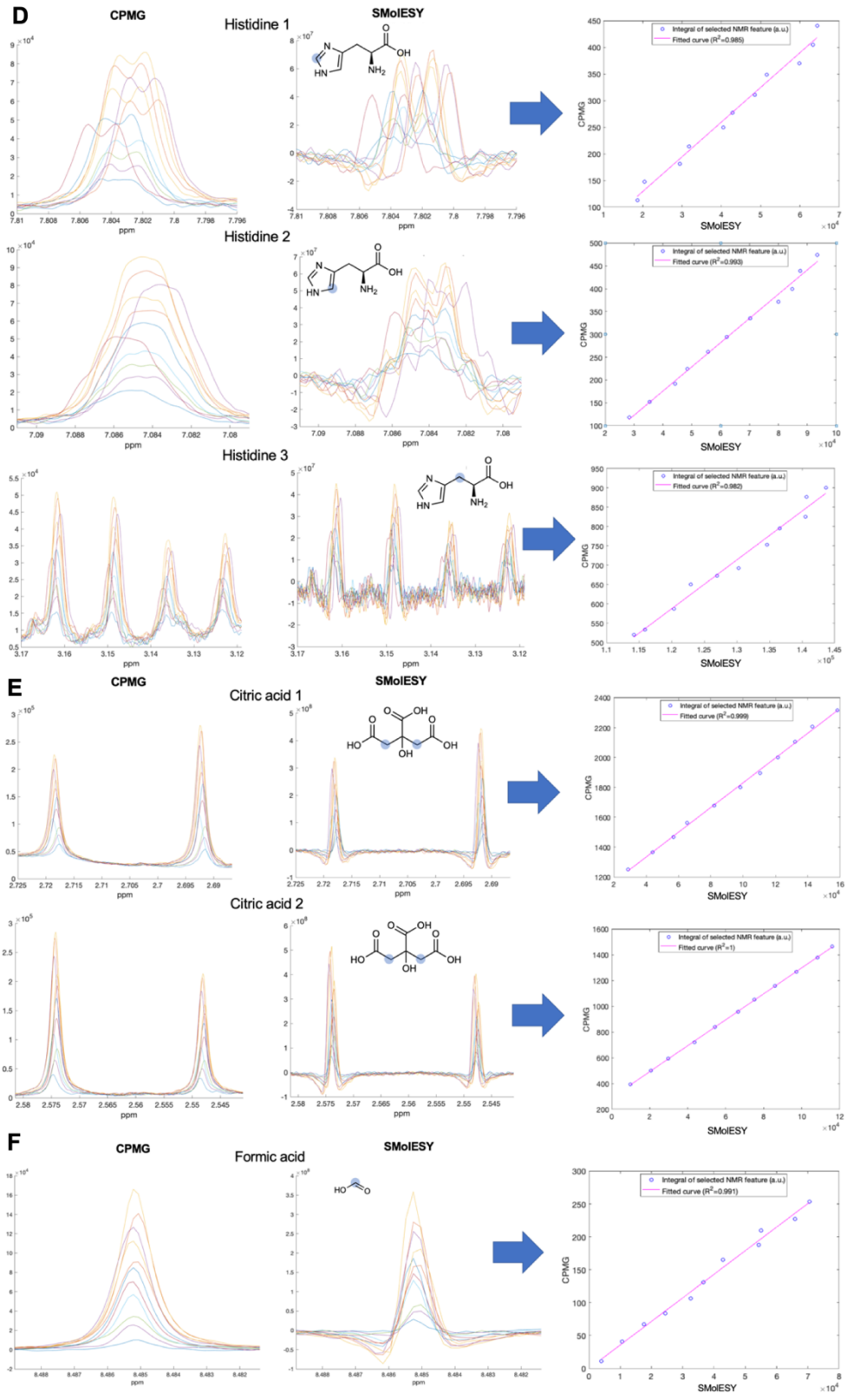
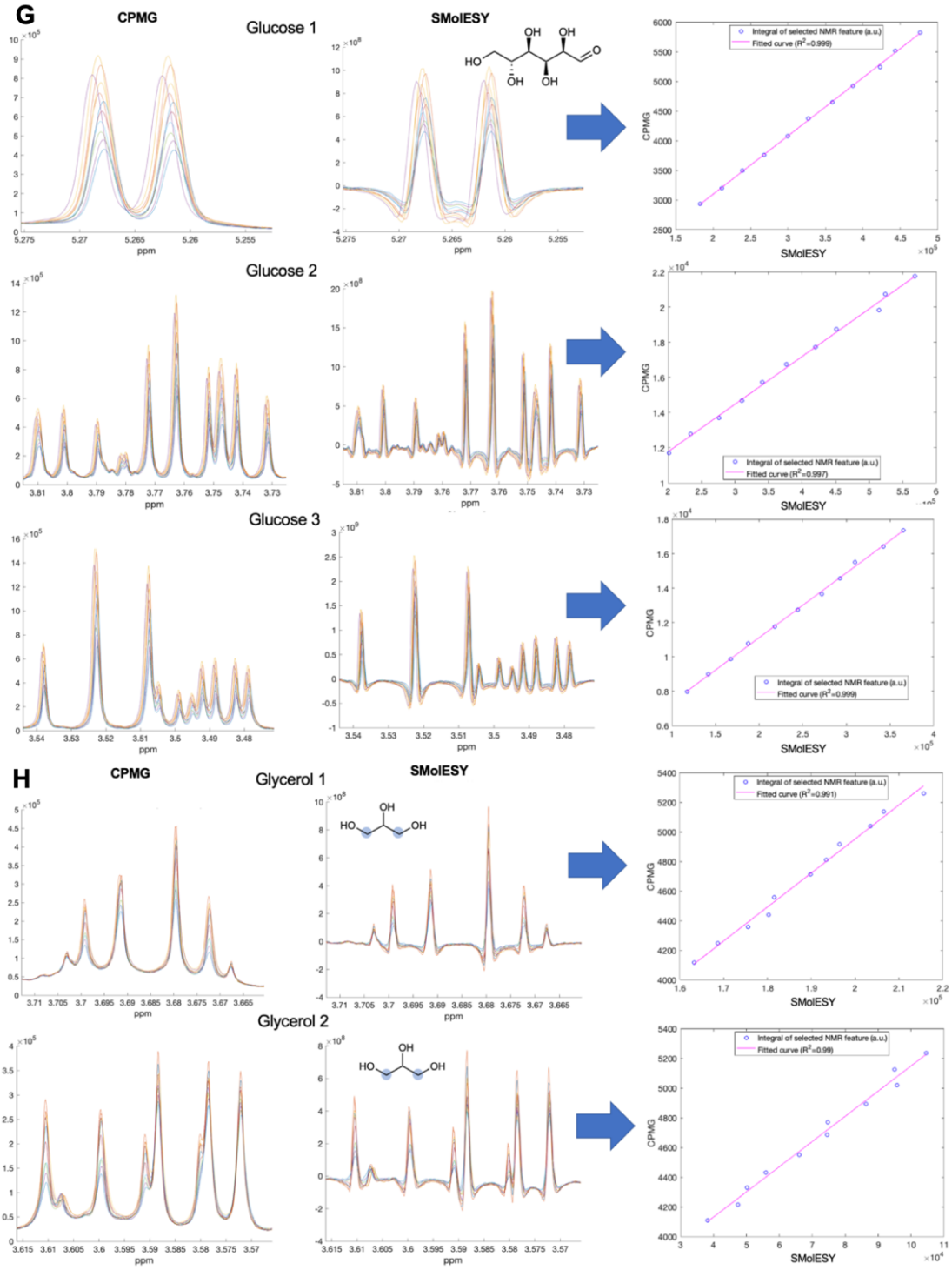


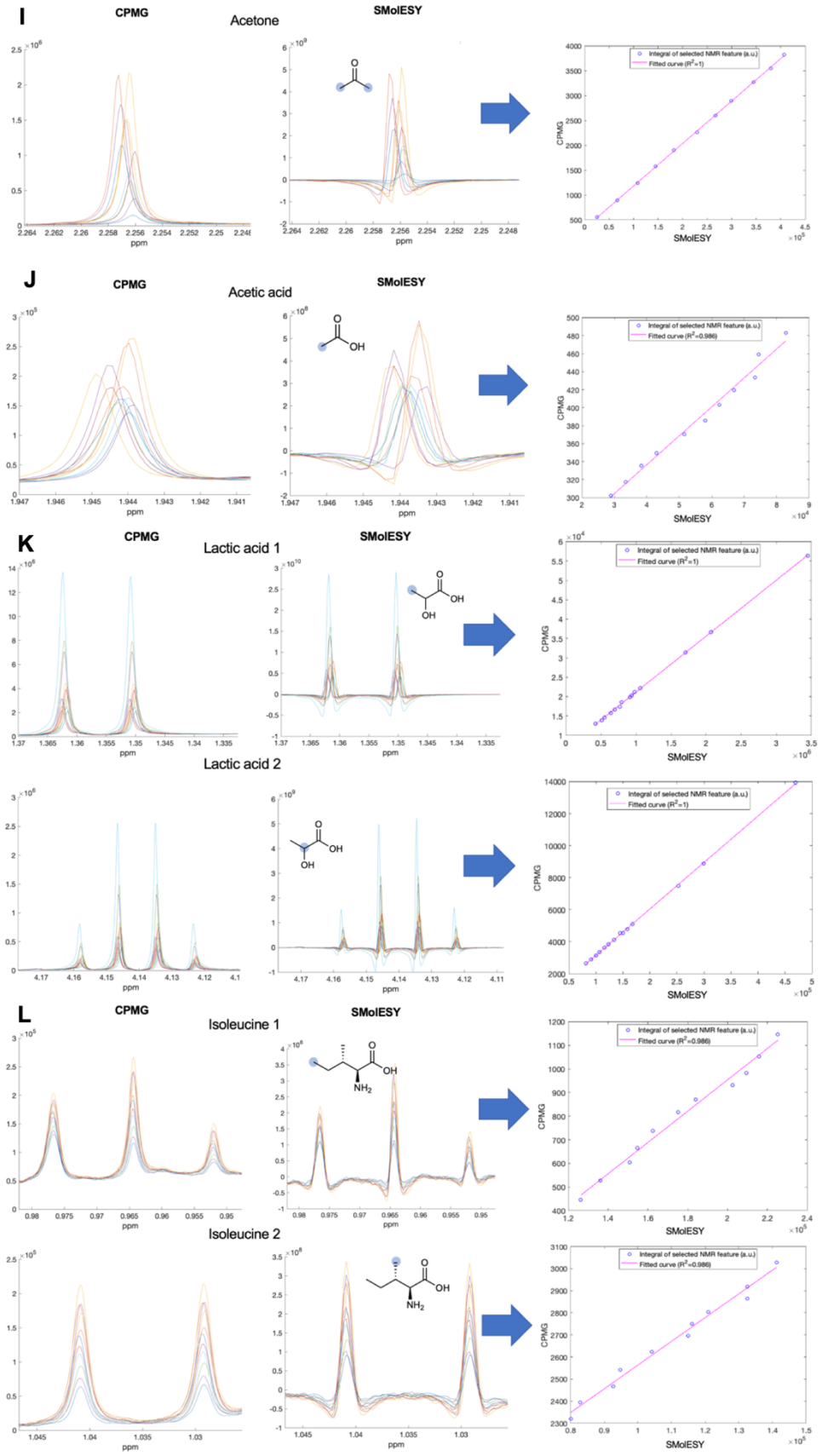
Figure S6. Mean spectrum of 994 plasma-EDTA samples spectra focusing on the 3.5-4.0 ^{1}H-NMR ppm region. Mean spectrum is calculated by the (**A**) 1D-NOESY, (**B**) CPMG and (**C**) SMolESY 994 spectra. SMolESY is colored according to the Pearson coefficients from SMolESY versus CPMG signals correlation for the 994 spectra. The majority of highly resolved (e.g. pointed by grey arrows) SMolESY signals are linearly correlated to the CPMG. Few signals are zero or anti-correlated to the CPMG spectra (light blue arrows), due to partial overlapping with other signals and/or baseline distortions in the CPMG (i.e. ineffective suppression of broad signals).

SMolESY vs CPMG spectral bins correlation – spiking experiments in real plasma matrix




D
CPMG
Histidine 1
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.985)
CPMG
SMolESY
Histidine 2
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.993)
CPMG
SMolESY
Histidine 3
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.982)
CPMG
SMolESY
E
CPMG
Citric acid 1
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.999)
CPMG
SMolESY
Citric acid 2
Integral of selected NMR feature (a.u.)
Fitted curve (R²=1)
CPMG
SMolESY
F
CPMG
Formic acid
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.991)
CPMG
SMolESY


G
CPMG
SMolESY
Glucose 1
Glucose 2
Glucose 3
H
Glycerol 1
Glycerol 2
ppm
CPMG
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.999)
Fitted curve (R²=0.997)
Fitted curve (R²=0.999)
Fitted curve (R²=0.991)
Fitted curve (R²=0.99)


I
CPMG
Acetone
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=1)
CPMG
SMolESY
J
CPMG
Acetic acid
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.986)
CPMG
SMolESY
K
CPMG
Lactic acid 1
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=1)
CPMG
SMolESY
Lactic acid 2
Integral of selected NMR feature (a.u.)
Fitted curve (R²=1)
CPMG
SMolESY
L
CPMG
Isoleucine 1
SMolESY
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.986)
CPMG
SMolESY
Isoleucine 2
Integral of selected NMR feature (a.u.)
Fitted curve (R²=0.986)
CPMG
SMolESY
ppm

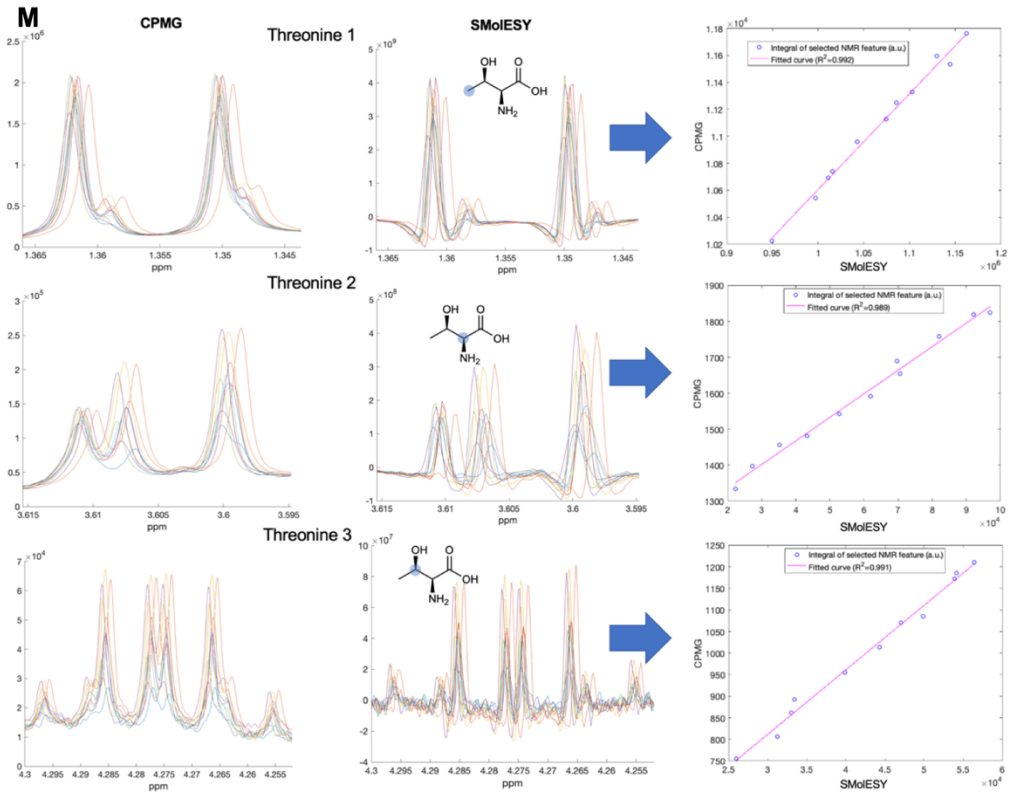


Figure S7. SMolESY application and reproducibility validation to spectra binning. Comparison between CPMG (left panel) and SMolESY (middle panel) spectral bins including various ^{1}H-NMR signals of 13 spiked metabolites (11 concentrations) in a plasma sample: (A) L-valine, (B) L-glutamine, (C) 2-hydroxybutyric acid, (D) L-histidine, (E) Citric acid, (F) Formic acid, (G) Glucose, (H) Glycerol, (I) Acetone, (J) Acetic acid (K) L-lactic acid, (L) L-isoleucine and (M) L-threonine. Linear regression curves (right panel) exhibit $0.98 < R^2 < 1$ for all metabolites, indicate that SMolESY is highly reproducible while superseding CPMG in broad signals suppression resulting in zero baseline distortions. Due to ~0 error in bin integration, horizontal and vertical error bars are not plotted in all linear regression plots. Light blue circles indicate the proton spin system of each metabolite included in each spectral bin which might include signals of other (not spiked) metabolites.

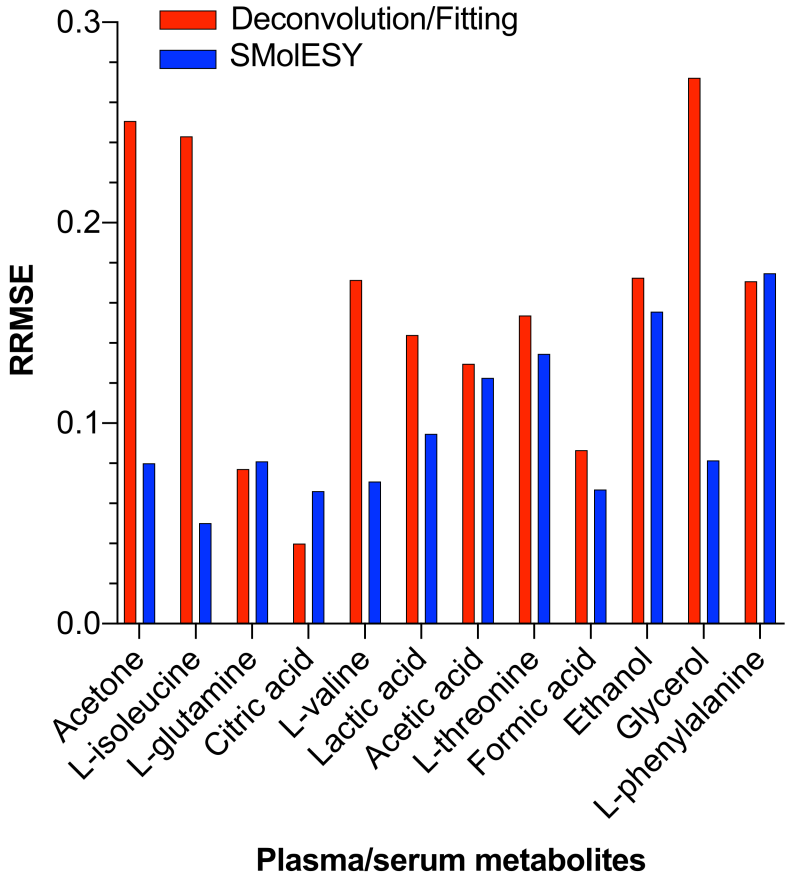


Figure S8. The relative root mean square errors (RRMSE) of SMolESY and deconvolution algorithms used for 12 spiked plasma metabolites absolute quantification were calculated based upon the regression analyses (see Fig. 5). RRMSE values show that absolute quantification via SMolESY signals integration provides less errors compared to the fitting procedures followed by deconvolution algorithms.

An overview of the SMolESY_platform graphical user interface (GUI) toolbox.

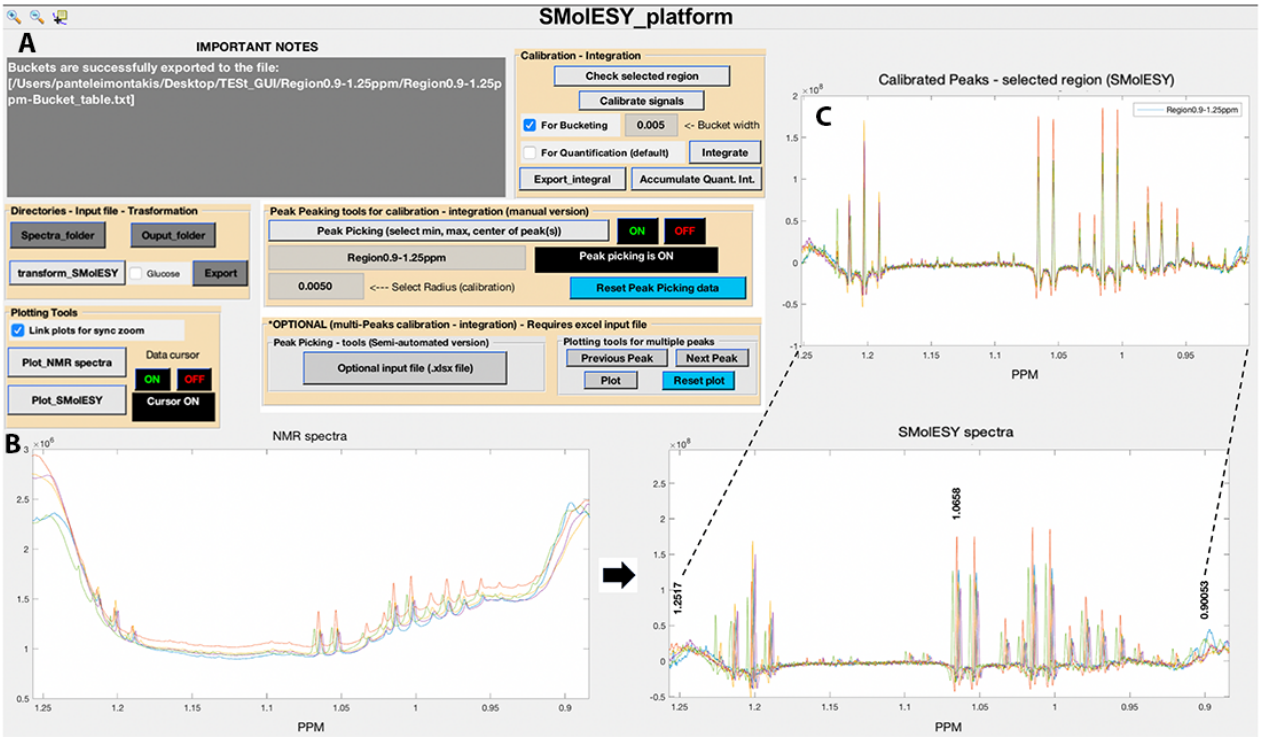


Figure S9. (A) The development of SMolESY_platfrom provides the opportunity for any user to load 1D ^{1}H-NMR raw spectra, so as to transform them into SMolESY and export them into a .txt file. In addition, the user has the opportunity to calibrate the SMolESY spectra to the doublet of the anomeric proton of glucose (~5.25 ppm) in case of plasma/serum/CSF etc. acquired spectra. (B) It offers the possibility to plot both 1D ^{1}H-NMR and SMolESY spectra for a synchronized zoom in both panels, (C) as well as to align into specific reference peaks a set of spectral bins or individual signals so as to integrate SMolESY features either for qualitative (i.e. option of variable-size SMolESY spectra binning) or quantitative purposes (i.e. option of SMolESY signals integration for quantification). The alignment of the signals could be performed both manually or in a semi-automated way upon users experience and request. More details and user guidelines of the software could be found at: https://github.com/pantakis/SMolESY_platform.

Computer code for the calculation of the Pearson correlation coefficients.

```
function [SMolESY ints,CPMG ints,Corv,Pv] = pearson cor SMolESYvsCPMG(ppm SMolESY,ppm CPMG,SMolESY data,CPMG data)

%
% Calculating the integrals of each feature/point of SMolESY for a
% comparison to CPMG equal width of ppm.
%
% NOTE: SMolESY - CPMG spectral data should be aligned to the same
%       reference peak (e.g. glucose, TSP etc.)
%
% Inputs
% ppm SMolESY: PPM data of all SMolESY data (one vector)
% ppm_CPMG: PPM data of all CPMG data (one vector)
% SMolESY data: SMolESY data (intensities) (matrix, rows: spectra, columns:
%               intensities)
% CPMG data: CPMG data (intensities) (matrix, rows: spectra, columns:
%               intensities)
%
% Ouyputs
% SMolESY ints: Integral of each SMolESY feature
% CPMG_ints: Integral of each CPMG feature (discribing the same ppm region
%           as the corresponding SMolESY feature)
% Corv: calculated Pearson's Linear Correlation Coefficient
%
%
%
    num_of_spectra = size(SMolESY_data,1);
    XAXIS CPMG rounded = round(ppm CPMG,4);
    XAXIS_SMolESY_rounded = round(ppm_SMolESY,4);

    % equal Xaxis buckets for both CPMG and SMolESY/NOESY
    for l = 2:length(ppm SMolESY)
    [~,ii] = find(ppm_CPMG <= ppm_SMolESY(:,l-1) & ppm_CPMG >= ppm_SMolESY(:,l));
    KL(l-1).aa = ii;
    clearvars ii
        for i = 1:num of spectra
            SMolESY_ints(i,l-1) = trapz(fliplr(XAXIS_SMolESY_rounded(1,l-1:l)),fliplr(SMolESY_data(i,l-1:l)));
            if length(KL(l-1).aa) > 1
                CPMG_ints(i,l-1) = trapz(fliplr(XAXIS_CPMG_rounded(1,min(KL(l-1).aa):max(KL(l-1).aa))),fliplr(CPMG_data(i,min(KL(l-1).aa):max(KL(l-1).aa))));
            else
                CPMG ints(i,l-1) = CPMG data(i,KL(l-1).aa);
            end
        end
    [rho,pval] = corr(CPMG_ints(:,l-1),SMolESY_ints(:,l-1),'Type','Pearson');
    Corv(1,l-1) = rho;
    Pv(1,l-1) = pval;
    end
end
```

Data S1 The described MATLAB code was followed to correlate each SMolESY feature intensity (i.e. integral) versus its corresponding CPMG feature for more than 3000 plasma spectra form different individuals as depicted in Fig. 2 and Fig. S5.

17 spiked metabolites 11 different concentrations spiked in real plasma sample

Table S2

Metabolites	Concentration (mM) [#]										
Lactic acid	0	0.333	0.666	0.999	1.332	1.665	1.998	2.331	2.664	2.997	3.330
2-hydroxybutric acid sodium salt	0	0.016	0.032	0.048	0.063	0.079	0.095	0.111	0.127	0.143	0.159
Acetone	0	0.034	0.069	0.103	0.138	0.172	0.207	0.241	0.275	0.310	0.344
Citric acid	0	0.078	0.156	0.234	0.312	0.390	0.469	0.547	0.625	0.703	0.781
D-glucose	0	0.611	1.221	1.832	2.442	3.053	3.663	4.274	4.885	5.495	6.106
Ethanol	0	0.043	0.087	0.130	0.174	0.217	0.261	0.304	0.347	0.391	0.434
Glycerol	0	0.054	0.109	0.163	0.217	0.271	0.326	0.380	0.434	0.489	0.543
L-aspartic acid	0	0.038	0.075	0.113	0.150	0.188	0.225	0.263	0.300	0.338	0.376
L-glutamine	0	0.082	0.164	0.246	0.328	0.411	0.493	0.575	0.657	0.739	0.821
L-histidine	0	0.045	0.090	0.135	0.181	0.226	0.271	0.316	0.361	0.406	0.451
L-isoleucine	0	0.023	0.046	0.069	0.091	0.114	0.137	0.160	0.183	0.206	0.229
L-phenylalanine	0	0.049	0.097	0.145	0.194	0.242	0.291	0.339	0.387	0.436	0.484
L-threonine	0	0.025	0.050	0.076	0.101	0.126	0.151	0.176	0.201	0.227	0.252
L-tryptophan	0	0.073	0.147	0.220	0.294	0.367	0.441	0.514	0.588	0.661	0.735
L-valine	0	0.085	0.170	0.256	0.341	0.427	0.512	0.598	0.683	0.768	0.854
Sodium acetate	0	0.024	0.049	0.073	0.098	0.122	0.146	0.171	0.195	0.219	0.244
Sodium formate	0	0.044	0.088	0.132	0.176	0.221	0.265	0.309	0.353	0.397	0.441

[#]Zero values correspond to the non-spiked plasma sample.

Example of a smoothing filter for denoising SMolESY

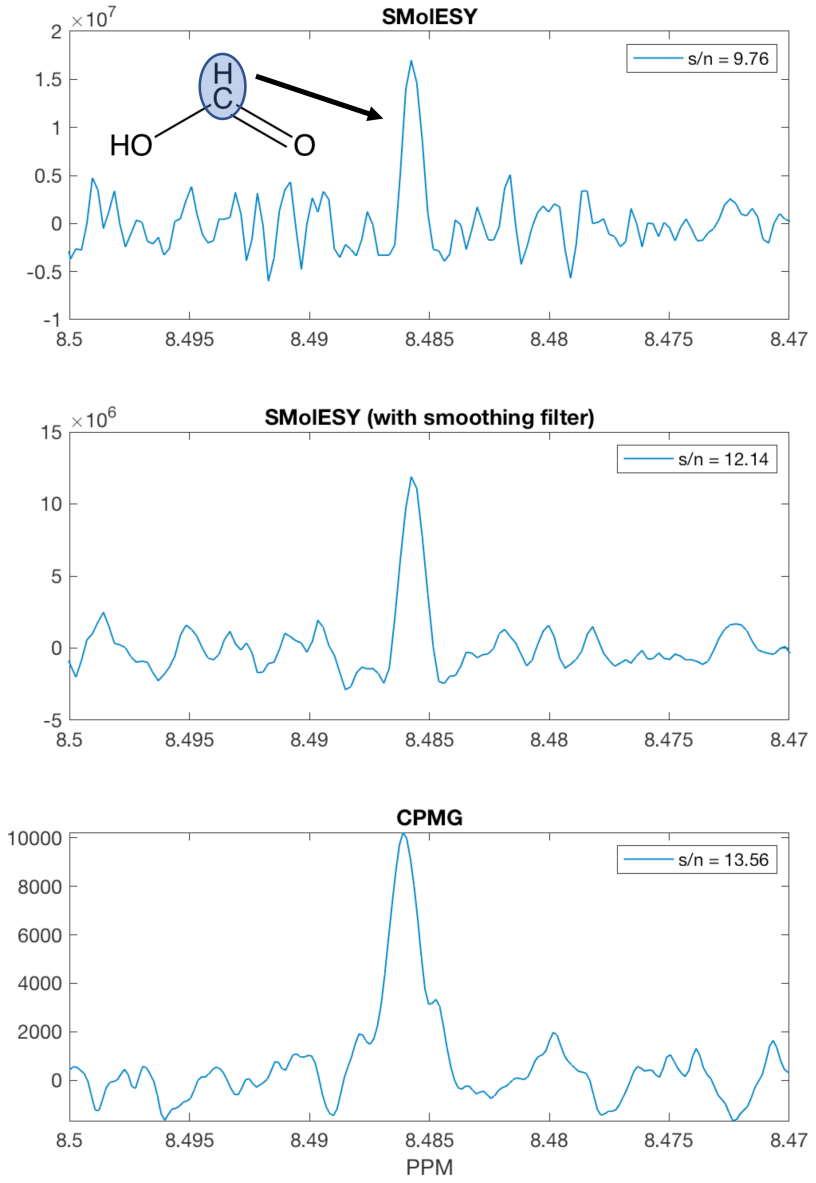


Figure S10. SMolESY (upper panel), noise filtered SMolESY (middle panel) and the CPMG (lower panel) spectral regions, focusing on the ^{1}H-NMR signal of the proton from Formic acid (at ~0.02 mM) in a plasma sample. The selected singlet resonates in usually noisy spectral region of a plasma ^{1}H-NMR profile, consequently a quite large s/n decrease is expected after its transformation from the 1D-NOESY spectrum. Indeed, SMolESY spectrum shows the lowest s/n (~27% decrease compared to CPMG), whereas the application of a simple lowpass filter (i.e. filter coefficients equal to the reciprocal of the window span) results in ~18% increase of s/n compared to pure SMolESY signal, and exhibits quite similar s/n with CPMG (~10% lower than CPMG).