

Supporting Information for “QC metrics from CPTAC raw LC-MS/MS data interpreted through multivariate statistics”

Xia Wang¹, Matthew C. Chambers², Lorenzo J. Vega-Montoto², David M. Bunk³, Stephen E. Stein³, David L. Tabb²

1 Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221

2 Vanderbilt University Medical School, Biomedical Informatics, Nashville, TN 37232-8575

3 National Institute of Standards and Technology, Gaithersburg, MD 20899

Contents

Supporting Information for “QC metrics from CPTAC raw LC-MS/MS data interpreted through multivariate statistics”	1
Table S1: Study 1 instrument inclusion	2
Table S2: Mass spectrometer introduction dates	3
Table S3: QuaMeter IDFree metrics and execution.....	4
Table S4: Loading matrices for factor analysis in Studies 1 and 5.....	6
Table S5: Medians and interquartile ranges (IQR) of dissimilarity measures.	8
Table S6: ANOVA association with individual metrics.....	9
Figure S1: Additional dissimilarity values from Study 1	11
Figure S2: Time points for each experiment in Study 5.....	12

Table S1: Study 1 instrument inclusion

In Study 1 data, instruments colored blue were included in the data analysis, while instruments that employed MALDI / Peptide Mass Fingerprinting or that lacked raw data were excluded. The Waters QTOF was excluded because of problems in data export.

Instrument and Laboratory	Peptide Mass Fingerprinting	Raw available	Ion Source
AB4700@90	No	Yes	MALDI
AB4800@54	No	No	MALDI
AB4800@96	No	No	MALDI
AB-Voyager@52	Yes	No	nanoESI
AB-Voyager@54	Yes	No	nanoESI
LTQ@73	No	Yes	nanoESI
LTQ@90	No	Yes	nanoESI
LTQ1@95	No	Yes	microESI
LTQ2@95	No	Yes	nanoESI
LTQb@65	No	Yes	nanoESI
Orbi@86	No	Yes	nanoESI
OrbiA@56	No	Yes	nanoESI
OrbiO@65	No	Yes	nanoESI
OrbiW@56	No	Yes	nanoESI
QSTARp@52	No	Yes	nanoESI
QSTARp@90	No	Yes	nanoESI
QSTARx@54	No	Yes	nanoESI
QTOFp@86	No	Yes	nanoESI
QTRAP@52	No	Yes	nanoESI
QTRAP@73	No	Yes	nanoESI
QTRAP@95	No	Yes	nanoESI
vMALDI-LTQ@52	No	Yes	MALDI
XCTp@90	No	Yes	nanoESI

Table S2: Mass spectrometer introduction dates

The following page provides a list of mass spectrometers that were used widely in proteomics in the 1990s and the first decade of the 2000s. The instruments included in Studies 1 and 5 are included, along with older and newer instruments for context.

Vendor	Model	Design	Introduced
Thermo	TSQ 700	QqQ	Jun-90
Thermo	TSQ 7000	QqQ	May-94
Thermo	LCQ	QIT	Mar-96
Waters	QTof 1	QqTOF	May-96
Waters	QTof 2	QqTOF	Jun-96
Thermo	LCQ Deca	QIT	Mar-99
AB SCIEX	QSTAR Pulsar	QqTOF	Mar-00
Waters	QTof Ultima	QqTOF	Jun-00
Thermo	LCQ Deca XP	QIT	Mar-01
Waters	QTof Micro	QqTOF	Mar-01
Thermo	TSQ Quantum	QqQ	May-01
Bruker Daltonics	ultraflex	TOF-TOF	Jul-01
AB SCIEX	4700	TOF-TOF	Jan-02
AB SCIEX	QSTAR XL	QqTOF	Sep-02
Bruker Daltonics	esquire HCT	QIT	Mar-03
AB SCIEX	4000 Q TRAP	QqQIT	Jun-03
Agilent	XCT	QIT	Jun-03
Thermo	LTQ	QIT	Jun-03
Thermo	LTQ FT	QIT-FT	Jun-03
Waters	QTof Premier	QqTOF	May-04
Bruker Daltonics	ultraflex II	TOF-TOF	Oct-04
AB SCIEX	3200 Q TRAP	QqQIT	Apr-05
AB SCIEX	4800	TOF-TOF	May-05
Thermo	LTQ Orbitrap	QIT-FT	Jun-05
AB SCIEX	QSTAR Elite	QqTOF	Jan-06
Agilent	6410	QqQ	Jan-06
Agilent	6510	QqTOF	Jan-06
Waters	Synapt	IMS-QqTOF	May-06
Bruker Daltonics	Ultraflex III	TOF-TOF	Aug-06
Agilent	6520	QqTOF	Jun-07
Agilent	6530	QqTOF	Jun-08
Bruker Daltonics	maXis	QqTOF	Jun-08
Thermo	TSQ Vantage	QqQ	Jun-08
Waters	Xevo QTof	QqTOF	Jan-09
AB SCIEX	5500 Q TRAP	QqQIT	Mar-09
AB SCIEX	5800	TOF-TOF	May-09
Agilent	6430	QqQ	May-09
Agilent	6540	QqTOF	May-09
Bruker Daltonics	amaZon	QIT	May-09
Bruker Daltonics	ultrafleXtreme	TOF-TOF	May-09
Thermo	LTQ Velos	QIT	May-09
Thermo	LTQ Orbitrap Velos	QIT-FT	May-09
Waters	Synapt G2	IMS-QqTOF	May-09

Table S3: QuaMeter IDFree metrics and execution

The following page gives a comprehensive list of the metrics produced in the “IDFree” mode of QuaMeter. The software was executed with this command line:

```
quameter.exe *.raw -MetricsType idfree -OutputFilepath metrics.tsv
```

The quameter.cfg files used for each type of instrument included the following:

Orbitrap configuration:

```
ChromatogramMzLowerOffset = "0.01mz"  
ChromatogramMzUpperOffset = "0.01mz"  
Instrument = "orbi"
```

Ion Trap configuration:

```
ChromatogramMzLowerOffset = "1.5mz"  
ChromatogramMzUpperOffset = "1.5mz"  
Instrument = "LTQ"
```

QqTOF configuration:

```
ChromatogramMzLowerOffset = "0.1mz"  
ChromatogramMzUpperOffset = "0.1mz"  
Instrument = "orbi" #Here, “Orbi” means “can resolve isotopes.”
```

Filename	What is the name of the file from which the metrics were computed?
StartTimeStamp	At what time did acquisition begin for this experiment?
XIC-WideFrac	What fraction of precursor ions account for the top half of all peak width?
XIC-FWHM-Q1	What is the 25%ile of peak widths for the wide XICs?
XIC-FWHM-Q2	What is the 50%ile of peak widths for the wide XICs?
XIC-FWHM-Q3	What is the 75%ile of peak widths for the wide XICs?
XIC-Height-Q2	The log ratio for 50%ile of wide XIC heights over 25%ile of heights.
XIC-Height-Q3	The log ratio for 75%ile of wide XIC heights over 50%ile of heights.
XIC-Height-Q4	The log ratio for maximum of wide XIC heights over 75%ile of heights.
RT-Duration	What is the highest scan time observed minus the lowest scan time observed?
RT-TIC-Q1	The interval when the first 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q2	The interval when the second 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q3	The interval when the third 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q4	The interval when the fourth 25% of TIC accumulates divided by RT-Duration
RT-MS-Q1	The interval for the first 25% of all MS events divided by RT-Duration
RT-MS-Q2	The interval for the second 25% of all MS events divided by RT-Duration
RT-MS-Q3	The interval for the third 25% of all MS events divided by RT-Duration
RT-MS-Q4	The interval for the fourth 25% of all MS events divided by RT-Duration
RT-MSMS-Q1	The interval for the first 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q2	The interval for the second 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q3	The interval for the third 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q4	The interval for the fourth 25% of all MS/MS events divided by RT-Duration
MS1-TIC-Change-Q2	The log ratio for 50%ile of TIC changes over 25%ile of TIC changes
MS1-TIC-Change-Q3	The log ratio for 75%ile of TIC changes over 50%ile of TIC changes
MS1-TIC-Change-Q4	The log ratio for largest TIC change over 75%ile of TIC changes
MS1-TIC-Q2	The log ratio for 50%ile of TIC over 25%ile of TIC
MS1-TIC-Q3	The log ratio for 75%ile of TIC over 50%ile of TIC
MS1-TIC-Q4	The log ratio for largest TIC over 75%ile TIC
MS1-Count	How many MS scans were collected?
MS1-Freq-Max	What was the fastest frequency for MS collection in any minute? (Hz)
MS1-Density-Q1	What was the 25%ile of MS scan peak counts?
MS1-Density-Q2	What was the 50%ile of MS scan peak counts?
MS1-Density-Q3	What was the 75%ile of MS scan peak counts?
MS2-Count	How many MS/MS scans were collected?
MS2-Freq-Max	What was the fastest frequency for MS/MS collection in any minute? (Hz)
MS2-Density-Q1	What was the 25%ile of MS/MS scan peak counts?
MS2-Density-Q2	What was the 50%ile of MS/MS scan peak counts?
MS2-Density-Q3	What was the 75%ile of MS/MS scan peak counts?
MS2-PrecZ-1	What fraction of MS/MS precursors is singly charged?
MS2-PrecZ-2	What fraction of MS/MS precursors is doubly charged?
MS2-PrecZ-3	What fraction of MS/MS precursors is triply charged?
MS2-PrecZ-4	What fraction of MS/MS precursors is quadruply charged?
MS2-PrecZ-5	What fraction of MS/MS precursors is quintuply charged?
MS2-PrecZ-more	What fraction of MS/MS precursors is charged higher than +5?
MS2-PrecZ-likely-1	What fraction of MS/MS precursors lack known charge but look like +1s?
MS2-PrecZ-likely-multi	What fraction of MS/MS precursors lack known charge but look like >+1s?

Table S4: Loading matrices for factor analysis in Studies 1 and 5

A: This table reports the loadings resulting from the Study 1 factor analysis. Sample 1A (NC120 lab digestion protocols) and Sample 1B (NCI 20 centrally digested samples) were combined for this factor analysis. The final row reports the fraction of total variability described by each factor.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
XIC.WideFrac						
XIC.FWHM.Q2		0.599042				
XIC.Height.Q2	0.78069					
XIC.Height.Q3	0.756443					
XIC.Height.Q4	0.779741					
RT.TIC.Q1						
RT.TIC.Q2	-0.51697					
RT.TIC.Q3	-0.62854					
RT.TIC.Q4						
RT.MS.Q1						-0.70514
RT.MS.Q2					-0.77146	
RT.MS.Q3						0.653192
RT.MS.Q4					0.868972	
RT.MSMS.Q1						0.720215
RT.MSMS.Q2				-0.59798		
RT.MSMS.Q3				-0.56632		
RT.MSMS.Q4				0.70771		
MS1.TIC.Change.Q2						
MS1.TIC.Change.Q3						0.542975
MS1.TIC.Change.Q4				0.615974		
MS1.TIC.Q2						
MS1.TIC.Q3					0.640952	
MS1.TIC.Q4				0.713715		
MS1.Count						
MS1.Freq.Max			0.53585			
MS1.Density.Q1		0.856041				
MS1.Density.Q2		0.882536				
MS1.Density.Q3						
MS2.Count	0.882953					
MS2.Freq.Max	0.922741					
MS2.Density.Q1			0.893332			
MS2.Density.Q2			0.95602			
MS2.Density.Q3			0.888335			
Proportion	0.18	0.106	0.104	0.088	0.087	0.08

B: This table reports the loadings resulting from the Study 5 factor analysis. Samples 3A (yeast) and 3B (yeast + BSA) were combined for this factor analysis.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
XIC.WideFrac		0.931				
XIC.FWHM.Q2	-0.939					
XIC.Height.Q2		-0.684		0.611		
XIC.Height.Q3	0.91					
XIC.Height.Q4	-0.621				0.522	
RT.TIC.Q1	-0.678			-0.538		
RT.TIC.Q2		0.848				
RT.TIC.Q3		0.828				
RT.TIC.Q4		-0.878				
RT.MS.Q1			0.93			
RT.MS.Q2				0.887		
RT.MS.Q3			-0.953			
RT.MS.Q4				-0.908		
RT.MSMS.Q1	-0.961					
RT.MSMS.Q2	-0.935					
RT.MSMS.Q3	-0.925					
MS1.TIC.Change.Q2			0.903			
MS1.TIC.Change.Q3	0.638					
MS1.TIC.Change.Q4			0.636			
MS1.TIC.Q2			0.896			
MS1.TIC.Q3	0.913					
MS1.TIC.Q4	0.765					
MS1.Count	0.931					
MS1.Freq.Max			0.652	0.684		
MS1.Density.Q1	-0.612		-0.629			
MS1.Density.Q2	-0.787	0.528				
MS1.Density.Q3	-0.631	0.659				
MS2.Freq.Max		0.566			0.773	
MS2.Density.Q1	0.957					
MS2.Density.Q2	0.967					
MS2.Density.Q3	0.958					
Proportion	0.403	0.176	0.172	0.138	0.052	0.008

Table S5: Medians and interquartile ranges (IQR) of dissimilarity measures.

A: Medians and interquartile ranges of dissimilarity measures for Study 1 experiments, corresponding to Figure 2A and 2B in the main text.

	Sample Pairs	Median	IQR
Figure 2A: Study 1 experiments from the same mass spectrometer (in the same lab)	1A	3.71	3.20
	1B	3.63	4.51
	1A:1B	6.31	3.95
Figure 2B: Study 1 experiments from the same type of instruments but different laboratories	1A	7.91	3.20
	1B	8.16	3.51
	1A:1B	8.47	3.83

B: Medians and interquartile ranges of dissimilarity measures for Study 5 experiments, corresponding to Figure 2C and 2D in the main text. Results on QIT and Orbitrap instruments in Study 1 are also listed for comparison with Study 5, which used LTQ and Orbitrap instruments.

	same mass spectrometer (Figure 2A and 2C)				same type of instruments but different laboratories (Figure 2B and 2D)		
	Sample Pairs	Median	IQR		Sample Pairs	Median	IQR
Figure 2A: Study 1 experiments from QIT and Orbitrap instruments	1A	2.27	3.46	Figure 2B: Study 1 experiments from QIT and Orbitrap instruments	1A	8.22	3.46
	1B	2.23	3.50		1B	7.75	3.38
	1A:1B	4.99	2.74		1A:1B	8.35	3.77
Figure 2C: Study 5 experiments from the same mass spectrometer	3A	1.86	1.61	Figure 2D: Study 5 experiments from the same type of instruments but different laboratories	3A	7.36	1.62
	3B	1.78	1.15		3B	7.47	2.20

Table S6: ANOVA association with individual metrics

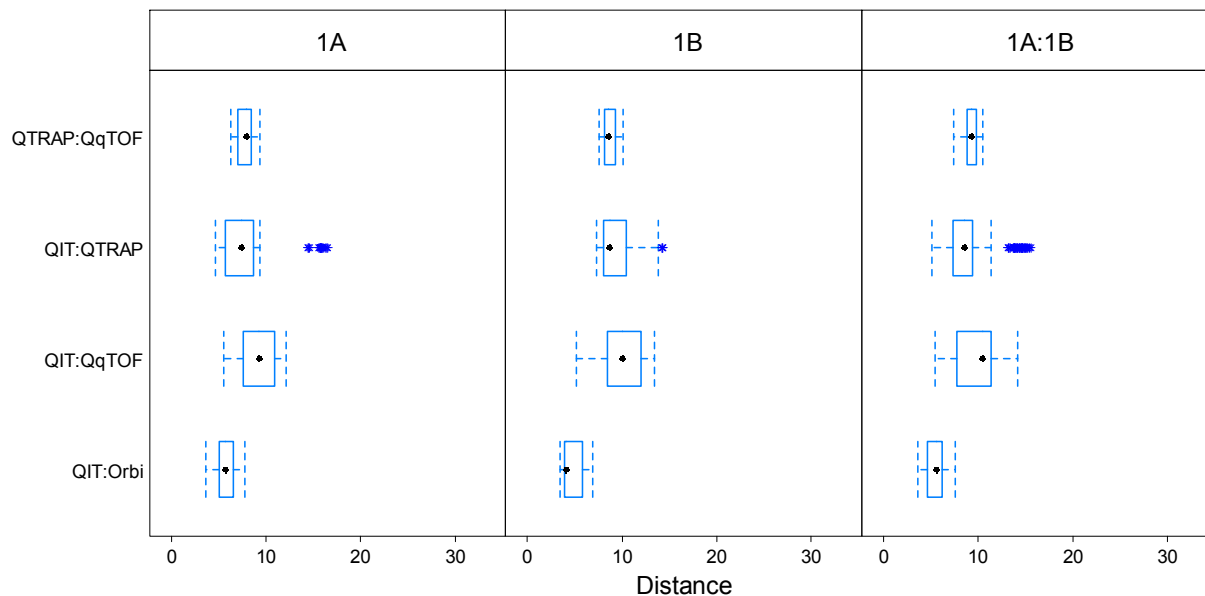
The Bonferroni method adjusted the p-values for multiple hypotheses testing. An “X” in a given cell indicates the corresponding factors (Mass Spectrometer or Batch) had a significant impact on the variability of the given metric.

QuaMeter Metrics	Study 1 Sample 1B		Study 5 Sample 3A		Study 5 Sample 3B	
	Mass Spectrometer	Nested Batch	Mass Spectrometer	Nested Batch	Mass Spectrometer	Nested Batch
XIC.WideFrac		X	X	X	X	
XIC.FWHM.Q2		X	X	X	X	
XIC.Height.Q2		X	X		X	
XIC.Height.Q3		X	X		X	
XIC.Height.Q4	X		X		X	
RT.TIC.Q1	X	X	X			
RT.TIC.Q2		X				
RT.TIC.Q3		X		X		X
RT.TIC.Q4		X				
RT.MS.Q1	X		X			
RT.MS.Q2	X				X	
RT.MS.Q3					X	
RT.MS.Q4	X				X	
RT.MSMS.Q1			X			
RT.MSMS.Q2			X	X		
RT.MSMS.Q3		X	X	X		
RT.MSMS.Q4						
MS1.TIC.Change.Q2	X		X	X	X	
MS1.TIC.Change.Q3		X			X	
MS1.TIC.Change.Q4		X	-----	-----	-----	-----
MS1.TIC.Q2			X		X	
MS1.TIC.Q3		X	X		X	
MS1.TIC.Q4	X		X			
MS1.Count	X	X	X		X	
MS1.Freq.Max			X		X	X
MS1.Density.Q1	X	X		X	X	
MS1.Density.Q2		X	X	X	X	
MS1.Density.Q3		X	X	X	X	
MS2.Count	X	X	-----	-----	-----	-----
MS2.Freq.Max	X	X	X		X	
MS2.Density.Q1	X	X	X	X	X	
MS2.Density.Q2	X	X	X	X	X	X
MS2.Density.Q3	X	X	X	X	X	

We examined how the individual metrics were affected by the mass spectrometer and batch factors. The nested ANOVA model was evaluated on each of the metrics, separately. For Study 1, 14 of 33 metrics were significantly affected by the mass spectrometer factor, and 22 were significantly affected by the nested batch factor after controlling for multiple comparisons by the Bonferroni method. For Study 5, there are 22 metrics significantly affected by the mass spectrometer effects and 12 metrics significantly affected by the batch effects for sample 3A and 24 and 3 for sample 3B. This table summarizes the ANOVA results for each metric in Study 1 and Study 5. It shows that there are more metrics that are significantly affected by nested batch factor in Study 1 than in Study 5. The possible reason is that the SOP reduced the variability of metrics related to experimental run orders. In Study 5, the major source of variability in the experimental performance was from the mass spectrometer effects, thus a larger proportion of metrics were significantly affected by the mass spectrometer factor. Also, some metrics were consistently affected by one or two factors in both studies. For example, the changes in mass spectrometer led to significant variation in XIC.Height.Q4, MS1.TIC.Change.Q2, MS1.Count, MS2.Freq.Max and MS2.Density.Qx. The batch has a significant impact on RT.TIC.Q3 and MS2.Density.Q2 in both studies. On the other hand, other metrics were only significant in one study (e.g. RT.TIC.Q2).

Figure S1: Additional dissimilarity values from Study 1

A: The following chart gives the distribution of dissimilarity values for instruments of different types in the same laboratory. Sample 1A: NC1-20 mixture digested by lab digestion protocols; Sample 1B: NCI-20 mixture centrally digested.



B: The following chart gives the distribution of dissimilarity values for instruments of different types and laboratories, representing a worst-case scenario for similarity. Sample 1A: NC1-20 mixture digested by lab digestion protocols; Sample 1B: NCI-20 mixture centrally digested.

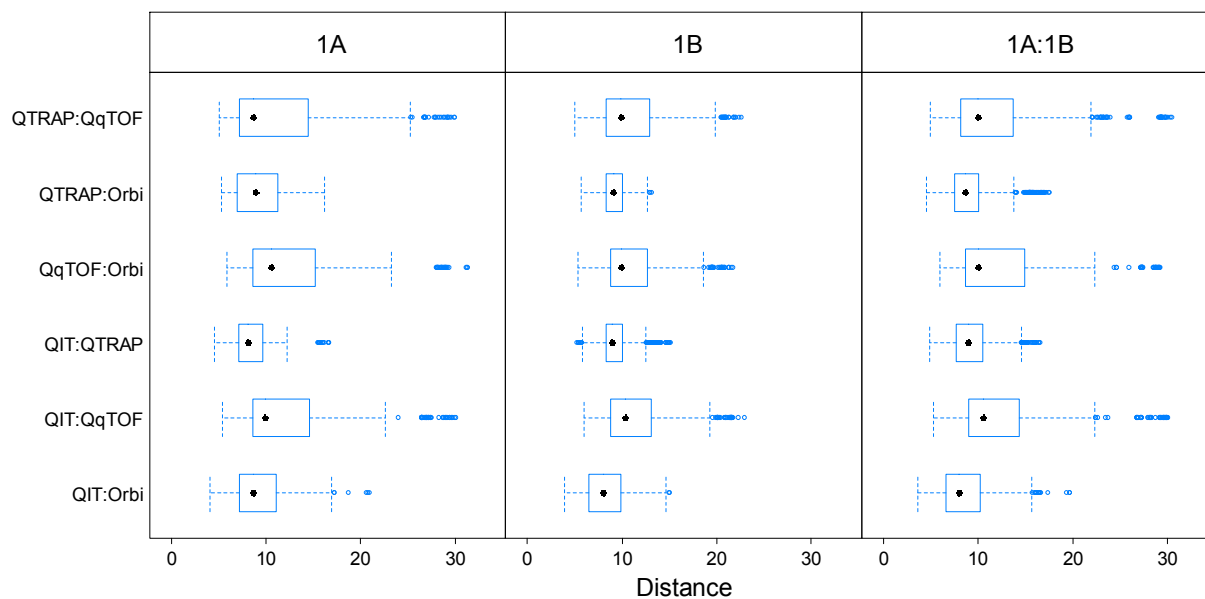


Figure S2: Time points for each experiment in Study 5

The following chart illustrates the run times for each LC-MS/MS experiment in the course of Study 5 across six instruments. Run order was prescribed by the SOP under which Study 5 was conducted. Sample 1B: digested NCI-20 mixture; Samples 3A: yeast; 3B: yeast + BSA.

