Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2018

Supplemental Information

Collision Cross Section Compendium to Annotate and Predict Multi-omic Compound Identities

Jaqueline A. Picache, Bailey S. Rose, Andrzej Balinski, Katrina L. Leaptrot, Stacy D. Sherrod, Jody C. May, John A. McLean*

Department of Chemistry, Center for Innovative Technology, Vanderbilt Institute of Chemical Biology, Vanderbilt Institute for Integrative Biosystems Research and Education, Vanderbilt-Ingram Cancer Center; Vanderbilt University, Nashville, Tennessee

*Corresponding Author Email: john.a.mclean@vanderbilt.edu

Abstract

Ion mobility mass spectrometry (IM-MS) expands the analyte coverage of existing multi-omic workflows by providing an additional separation dimension as well as a parameter for characterization and identification of molecules – the collision cross section (CCS). This work presents a large, Unified CCS Compendium of > 3800 experimentally acquired CCS values obtained from traceable molecular standards and measured with drift tube-mass spectrometers. An interactive visualization of this Compendium along with data analytic tools have been made openly accessible. Represented in the Compendium are 14 structurally-based chemical super classes, consisting of a total of 80 classes and 157 subclasses. Using this large data set, regression fitting and predictive statistics have been performed to describe mass-CCS correlations specific to each chemical ontology. These structural trends provide a rapid and effective filtering method in the traditional untargeted workflow for identification of unknown biochemical species. The utility of the approach is illustrated by an application to metabolites in human serum, quantified trends of which were used to assess the probability of an unknown compound belonging to a given class. CCS-based filtering narrowed the chemical search space by 60% while increasing the confidence in the remaining isomeric identifications from a single class, thus demonstrating the value of integrating predictive analyses into untargeted experiments to assist in identification workflows. The predictive abilities of this Compendium will improve in specificity and expand to more chemical classes as additional data from the IM-MS community is contributed. Instructions for data submission to the Compendium and criteria for inclusion are provided.

Table of Contents

Section S1. Ion Mobility Peak Annotation

In data sets in which multiple ion mobility peaks were observed for a single analyte, mobility peaks were annotated by assigning a peak number to each mobility peak. Peak number assignments begin with “1”, which refers to the smallest observed CCS or shortest drift time. Each subsequent mobility peak is assigned in numerical order (2, 3, etc.). If only one peak is observed, the CCS value is assigned a ”1”.

An example is shown in Figure S1. If only one peak is observed, the peak is assigned the number “1”.

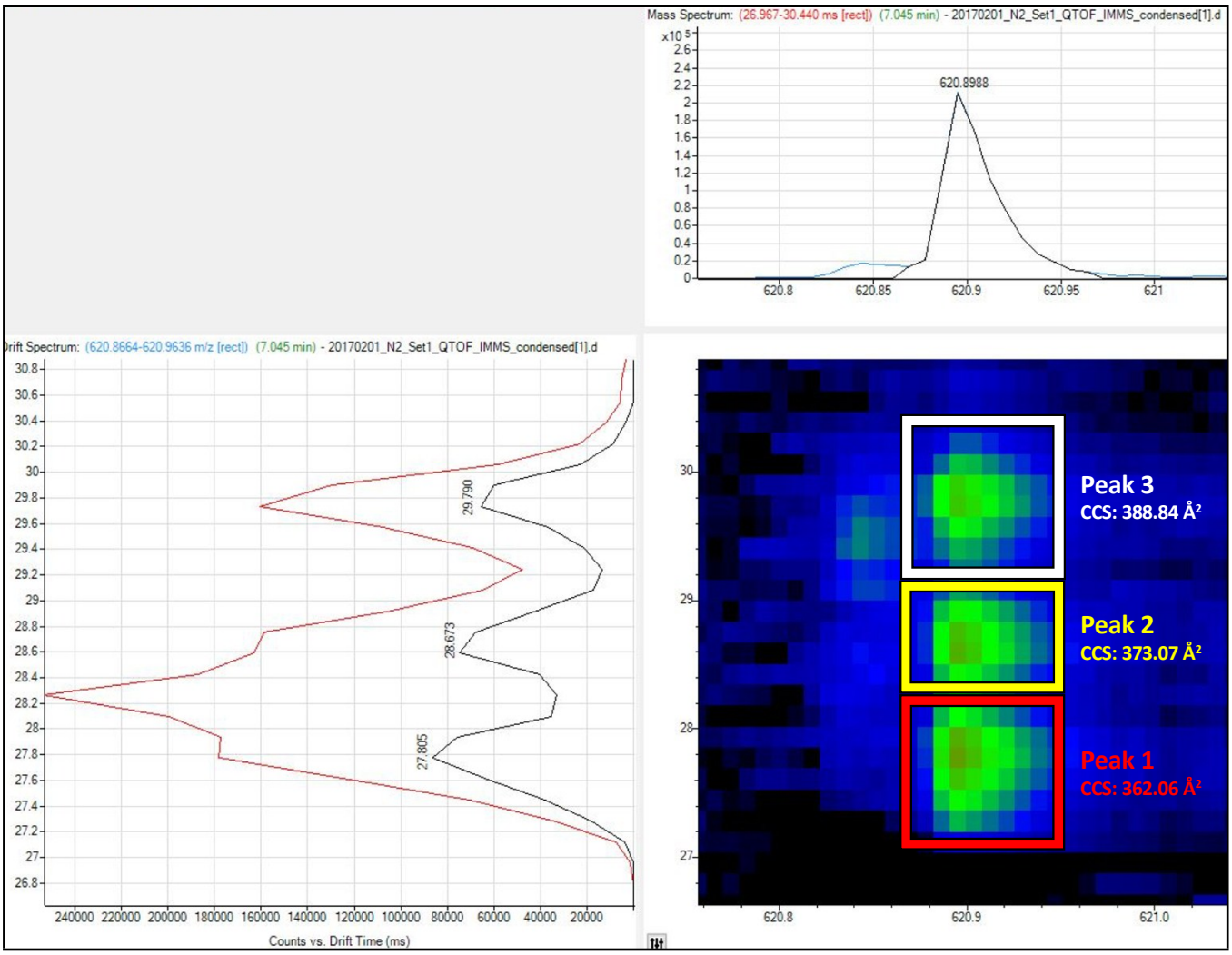

Figure S1. Illustration of CCS value annotations for analytes with multiple mobility peaks

Section S2. Data Inclusion Criteria

The Unified CCS Compendium is anticipated to be a collaborative effort of the IM-MS community; and the authors would like to invite contributions to this open-access repository for quality-controlled CCS measurements. Contributions towards the Unified CCS Compendium will improve informatics tools within the Compendium to aid in IM-MS based multi-omic analyte identification workflows. For consistency, please follow the guidelines below. These guidelines are aimed at standardizing the data submission process and will expedite data quality assessment. Please note that these guidelines are subject to change; and the most up-to-date procedures can be downloaded from the online Compendium.[1] Currently, only the submission of CCS data obtained from drift tube measurements is accepted. In the future, it is anticipated that the Compendium will be expanded to support CCS measurements obtained from other IM techniques.

Guidelines for data submission into the Unified CCS Compendium

Single Field Data

The supplemental information packet includes a file entitled "SI_SingleField_DataFormat.xlsx". The two spreadsheets ("Single Field Reference Standards" and "Single Field Data Format") within the Excel file will need to be populated prior to submission of single field data to the Unified CCS Compendium. *Caution: If the Excel file is opened in read-only mode, the spreadsheet will not be editable. Please click 'enable editing' to proceed.*

Step 1: Collect, at minimum, triplicate measurements of the reference standards for each day samples are acquired.

- Recommended strategy: Infuse reference standards simultaneously with the analyte(s) of interest (i.e., as an internal reference). This allows the reference standards to be measured under the same conditions that the analytes are exposed to.
- If measuring reference standards independently, it is advised to acquire CCS measurements of the reference standards before, during, and after each set of acquisitions to assess any systematic deviations in mass and mobility measurements. This will also allow profiles of pressure, temperature, and electric field to be constructed for each acquisition set, to assist in assessing measurement quality.

Step 2: Collect, at minimum, triplicate measurements of the experimental analytes of interest (if not acquired in step 1). Include $\geq$ 5 compounds from the quality assessment (QA) compounds list (Table S1g) to assess data.

- Experimental values for analytes chosen from the QA compounds list must meet the following criteria:
  - Average CCS percent error of $\leq$ 0.5%

$$percent\ error = \frac{(CCS_{experimental} - CCS_{QA})}{CCS_{QA}} \cdot 100$$

  - Maximum individual CCS percent error $\leq$ 1%

Step 3: Populate columns A-F in the spreadsheet entitled "Single Field Reference Standards" (see Fig. S2) with experimental data generated from step 1 for each replicate.

- Use rows 5-14 for positive ion data and/or rows 15-24 for negative ion data.
- Single-field CCS and corresponding *m/z* values can be obtained using the "CCS Calibration (Single-Field)" method implemented in IM-MS Browser (Agilent Technologies). Alternately, the single-field drift time/CCS relationship can be calculated directly from drift time measurements of reference standards using equations described in previous work.[2]
- Columns G-P in the spreadsheet will be auto-populated. Important: The data in this spreadsheet is intended to be used ONLY for reference standards that were measured, NOT the analytes being submitted to the CCS Compendium.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
4	Experimental Replicate 1 *m/z*	Experimental Replicate 2 *m/z*	Experimental Replicate 3 *m/z*	Experimental Replicate 1 CCS	Experimental Replicate 2 CCS	Experimental Replicate 3 CCS	Reference Standards *m/z*	Reference Standards CCS	Experimental Average *m/z*	*M/z* error (ppm)	Experimental Average CCS	CCS Std Dev	CCS % RSD	% CCS Difference	Polarity
5	118.086	118.088	118.085	121.40	121.30	121.32	118.086	121.30	118.09	2.82	121.34	0.05	0.04%	0.03%	+
6	322.048	322.050	322.047	153.80	153.73	153.75	322.048	153.73	322.05	1.04	153.76	0.04	0.02%	0.02%	+
7	622.029	622.031	622.028	203.00	202.96	202.98	622.029	202.96	622.03	0.54	202.98	0.02	0.01%	0.01%	+
8	922.010	922.012	922.009	243.70	243.64	243.66	922.010	243.64	922.01	0.36	243.67	0.03	0.01%	0.01%	+
9	1221.991	1221.993	1221.990	282.30	282.20	282.22	1221.991	282.20	1221.99	0.27	282.24	0.05	0.02%	0.01%	+
10	1521.971	1521.973	1521.970	317.10	316.96	316.98	1521.971	316.96	1521.97	0.22	317.01	0.08	0.02%	0.02%	+
11							1821.952	351.25							+
12							2121.933	383.03							+
13							2421.914	412.96							+
14							2721.895	441.21							+
15							112.986	108.23							-
16							301.998	140.04							-
17							601.979	180.77							-
18							1033.969	255.34							-
19							1333.969	284.76							-
20							1633.950	319.03							-
21							1933.931	352.55							-
22							2233.911	380.74							-
23							2533.892	412.99							-
24							2833.873	432.62							-
25											Averages:	0.04	0.02%	0.02%	

Figure S2. "Single Field Reference Standards" spreadsheet

Step 4: Populate Columns A-K and Column Q of the spreadsheet entitled "Single Field Data Format" (see Fig. S3) with experimental data acquired in step 2.

- Columns L-P will be auto-populated.
- Assign peak numbers in Column Q as necessary: smallest CCS = 1, next smallest = 2, etc. If only one peak is observed, assign a "1".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Compound	Formula	CAS	Adduct	Charge	Experimental Replicate 1 m/z	Experimental Replicate 2 m/z	Experimental Replicate 3 m/z	Experimental Replicate 1 CCS	Experimental Replicate 2 CCS	Experimental Replicate 3 CCS	Average Experimental m/z	Average Experimental CCS	Std. Dev	%RSD	CCS/z	Peak Number
2	Example Lipid	C45H73NO8P	5634-86-6	[M-H]	-1	786.5070	786.5074	786.5109	277.30	277.33	277.27	μ m/z	μ CCS	σ	$=\sigma/\mu$CCS*100	277.30	1
3	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	361.68	361.76	362.75	620.91	362.06	0.60	0.17	181.03	1
4	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	373.66	373.09	372.45	620.91	373.07	0.60	0.16	186.53	2
5	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	388.97	389.27	388.27	620.91	388.84	0.51	0.13	194.42	3
6	Your	Data	Here!														

Figure S3. "Single Field Data Format" spreadsheet (Columns A-K)

Step 5: Classify each compound using the ClassyFire web application (found at http://classyfire.wishartlab.com/).[3]

- Populate Columns R-W of the spreadsheet entitled "Stepped Field Data Format" (shown in Fig. S4) with the classification information, source (e.g. research group), and DOI (if published).

	A	B	C	D	E	R	S	T	U	V	W
1	Compound	Formula	CAS	Adduct	Charge	Kingdom	Super.Class	Class	Subclass	Source	DOI
2	Example Lipid	C45H73NO8P	5634-86-6	[M-H]	-1	Organic compounds	Lipids and lipid-like molecules	Glycerophospholipids	Glycerophosphoethanolamines	Research Group	If unpublished, please fill in as "Unpublished".
3	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
4	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
5	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
6	Your	Data	Here!			Your	Classifications	Here!			

Figure S4. "Single Field Data Format" spreadsheet (Columns R-W)

Step 6: Calculate the average RSD for *all* experimental values (QA compounds as well as all analytes/compounds that are being submitted for inclusion into the Unified CCS Compendium).

- The CCS values submitted must meet the following criteria:
  - Average RSD $\leq$ 0.5% for all experimental data set

- Individual compound RSD ≤ 0.7%

Step 7: Double check to ensure that steps 1-6 were performed.

- Step 1: Triplicate measurements acquired for reference standards each day sample measurements were collected.
- Step 2: Triplicate measurements acquired for ***all*** experimental values, including at least five compounds from the QA compound list.
- Step 3 & 4: Enter all data into the formatted spreadsheets.
- Step 5: Classify all compounds in the provided columns of the spreadsheets.
- Step 6: Calculate average RSD and individual RSDs.

Step 8: Submit spreadsheet for quality assessment.

Data must be submitted by emailing the completed spreadsheet ("SI_SingleField_DataFormat.xlsx") to ccscompendium@vanderbilt.edu.

- Please include the following information with each submission.
  - Institution
  - Research group
  - Instrument source type
  - Solvent/buffer system
  - List of reference compounds included in experimental data set

Upon data submission, the data will temporarily be quarantined and a quality control assessment will be performed. The quality control assessment includes: (1) verifying that all inclusion criteria is met, (2) confirming that all pertinent information is provided, and (3) checking that data is formatted properly. After the authors have processed a dataset (typically less than 10 days), collaborators will be notified which values will be accepted or if any revisions are needed. Data will be made available as soon as the quality control assessment is complete.

Stepped Field Data

The supplemental information packet includes a file entitled "SI_SteppedField_ScaleAndDataFormat.xlsx". The two spreadsheets ("Stepped Field Reference Standards and Scale" and "Stepped Field Data Format") within the Excel file will need to be populated prior to submission of stepped field data to the Unified CCS Compendium. *Caution: If the Excel file is opened in read-only mode, the spreadsheet will not be editable. Please click 'enable editing' to proceed.*

Step 1: Collect, at minimum, triplicate measurements of the reference standards for each day samples are acquired.

- Recommended strategy: Infuse reference standards simultaneously with the analyte(s) of interest (i.e., as an internal reference). This allows the reference standards to be measured under the same conditions that the analytes are exposed to.
- If measuring reference standards independently, it is advised to acquire CCS measurements of the reference standards before, during, and after each set of acquisitions to assess any systematic deviations in mass and mobility measurements. This will also allow profiles of pressure, temperature, and electric field to be constructed for each acquisition set, to assist in assessing measurement quality.

Step 2: Collect, at minimum, triplicate measurements of the experimental analytes of interest (if not acquired in step 1). Include $\geq$ 5 compounds from the quality assessment (QA) compounds list (Table S1) to assess data quality.

- Experimental values for analytes chosen from the QA compounds list must meet the following criteria:
  - Average CCS percent error of $\leq$ 0.5%

$$percent\ error = \frac{(CCS_{experimental} - CCS_{QA})}{CCS_{QA}} \cdot 100$$

  - Maximum individual CCS percent error $\leq$ 1%

Step 3: Populate columns A-F in the spreadsheet entitled "Stepped Field Reference Standards and Scale" (see Fig. S5) with data generated from step 1 for each replicate.

- True effective lengths for data collected in Step 1 must be calculated using the "Stepped Field Reference Standards and Scale" spreadsheet. Further detail addressing the purpose of scaling as well as the scaling procedure are discussed in supplemental Section S3.
- Use rows 7-16 for positive ion mode and/or rows 17-26 for negative ion mode.
- CCS and *m/z* values can be obtained using the "CCS Calculator (Stepped-Field)" method in IM-MS Browser (Agilent Technologies). Alternately, stepped-field CCS values can be calculated from corrected drift times using the fundamental low-field ion mobility equation.[4,5] Drift time correction requires a linear regression analysis incorporating the raw drift time measured at each of the drift fields surveyed, as described previously.[6]
- The experimental effective length (in cm) needs to be entered in the yellow box (Cell D4) located at the top of the spreadsheet. This length can be found in the "BaseDataAccess.dll.config" file located in the Mass Hunter Workstation (Agilent Technologies) install directory (typically: C Drive > Program Files > Agilent > MassHunter > Workstation > IMS > B.07.02 > Bin). Alternately, this is the length value used in the initial CCS calculation that is to be scaled.
- Columns G-P in the spreadsheet will be auto-populated.
- Important: The data in this spreadsheet is ONLY for reference standards that were measured, NOT the analytes being submitted to the CCS Compendium.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
4	**Experimental Effective Length (cm):**			78.24												
5																
6	**Experimental Replicate 1 *m/z***	**Experimental Replicate 2 *m/z***	**Experimental Replicate 3 *m/z***	**Experimental Replicate 1 CCS**	**Experimental Replicate 2 CCS**	**Experimental Replicate 3 CCS**	**Reference Standard *m/z***	**Reference Standard CCS**	**Experimental Average *m/z***	***M/z* error (ppm)**	**Experimental Average CCS**	**CCS Std Dev**	**CCS % RSD**	**Scale Factor**	**New Effective Length**	**Polarity**
7	118.086	118.088	118.085	121.40	121.30	121.32	118.086	121.30	118.09	2.82	121.34	0.05	0.04%	1.000164867	78.25289918	+
8	322.048	322.050	322.047	152.95	152.95	152.78	322.048	153.73	322.05	1.04	152.89	0.10	0.07%	0.997278864	78.02709834	+
9	622.029	622.031	622.028	202.44	202.48	202.29	622.029	202.96	622.03	0.54	202.40	0.10	0.05%	0.998627769	78.13263665	+
10	922.010	922.012	922.009	242.09	242.42	242.34	922.010	243.64	922.01	0.36	242.28	0.17	0.07%	0.997213979	78.02202172	+
11	1221.991	1221.993	1221.990	280.82	280.84	280.77	1221.991	282.20	1221.99	0.27	280.81	0.03	0.01%	0.997533326	78.04700741	+
12	1521.971	1521.973	1521.970	316.34	316.40	316.42	1521.971	316.96	1521.97	0.22	316.39	0.04	0.01%	0.999095148	78.1692044	+
13							1821.952	351.25								+
14							2121.933	383.03								+
15							2421.914	412.96								+
16							2721.895	441.21								+
17							112.986	108.23								-
18							301.998	140.04								-
19							601.979	180.77								-
20							1033.969	255.34								-
21							1333.969	284.76								-
22							1633.950	319.03								-
23							1933.931	352.55								-
24							2233.911	380.74								-
25							2533.892	412.99								-
26							2833.873	432.62								-
27											**Averages:**	**0.08**	**0.04%**	**Averages:**	**78.10847795**	

Figure S5. "Stepped Field Reference Standards and Scale" spreadsheet

Step 4: Populate Columns A-K and Column R of the spreadsheet entitled "Stepped Field Data Format" with experimental data acquired in step 2 (see Fig. S6).

- Columns L-Q will be auto-populated.
- Assign peak numbers in Column R as necessary: smallest CCS = 1, next smallest = 2, etc. If only one peak is observed, assign a "1".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	**Compound**	**Formula**	**CAS**	**Adduct**	**Charge**	**Experimental Replicate 1 *m/z***	**Experimental Replicate 2 *m/z***	**Experimental Replicate 3 *m/z***	**Experimental Replicate 1 CCS**	**Experimental Replicate 2 CCS**	**Experimental Replicate 3 CCS**	**Average Experimental *m/z***	**Average Experimental CCS**	**Std. Dev**	**% RSD**	**Scaled CCS**	**CCS/z**	**Peak Number**
2	Example Lipid	C45H73NO8P	5634-86-6	[M-H]	-1	786.5070	786.5074	786.5109	277.30	277.33	277.27	μ *m/z*	μ CCS	σ	=σ/μCCS*100	=μ*(old effective length/ new effective length)^2	277.30	1
3	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	361.68	361.76	362.75	620.91	362.06	0.60	0.17	363.28	181.03	1
4	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	373.66	373.09	372.45	620.91	373.07	0.60	0.16	374.32	186.53	2
5	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	388.97	389.27	388.27	620.91	388.84	0.51	0.13	390.15	194.42	3
6	**Your**	**Data**	**Here!**															

Figure S6. "Stepped Field Data Format" spreadsheet (Columns A-R)

Step 5: Classify each compound using the ClassyFire web application (found at http://classyfire.wishartlab.com/).[3]

- Populate Columns S-X of the spreadsheet entitled "Stepped Field Data Format" (Fig. S7) with the classification information, source (e.g. research group), and DOI (if published).

	A	B	C	D	E	S	T	U	V	W	X
1	**Compound**	**Formula**	**CAS**	**Adduct**	**Charge**	**Kingdom**	**Super.Class**	**Class**	**Subclass**	**Source**	**DOI**
2	Example Lipid	C45H73NO8P	5634-86-6	[M-H]	-1	Organic compounds	Lipids and lipid-like molecules	Glycerophospholipids	Glycerophosphoethanolamines	Research Group	If unpublished, please fill in as "unpublished".
3	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
4	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
5	Cyclosporin	C62H111N11O12		[M+H+K]	+2	Organic compounds	Organic polymers	Polypeptides		McLean	Unpublished
6	**Your**	**Data**	**Here!**			**Your**	**Classifications**	**Here!**			

Figure S7. "Stepped Field Data Format" spreadsheet (Columns S-X)

Step 6: Calculate the average RSD for *all* experimental values (QA compounds as well as all analytes/compounds that are being submitted for inclusion into the unified CCS compendium).

- The CCS values submitted must meet the following criteria:
  - Average RSD $\leq$ 0.5% for all experimental data set
  - Individual compound RSD $\leq$ 0.7%

Step 7: Check to ensure that steps 1-6 were performed.

- Step 1: At minimum, triplicate measurements were acquired for reference standards for each day that sample measurements were collected.
- Step 2: At minimum, triplicate measurements were acquired for ***all*** experimental values, including at least five compounds from the QA compound list.
- Step 3 & 4: Enter all data into the formatted spreadsheets.
- Step 5: Classify all compounds in the provided columns of the spreadsheets.
- Step 6: Calculate average RSD and individual RSD.

Step 8: Submit spreadsheet for quality assessment.

Data must be submitted by emailing the completed spreadsheet ("SI_SteppedField_ScaleAndDataFormat.xlsx") to ccscompendium@vanderbilt.edu.

- Please include the following information with each submission.
  - Institution
  - Research group
  - Instrument source type
  - Solvent/buffer system
  - List of reference compounds included in experimental data set

Upon data submission, the data will temporarily be quarantined and a quality control assessment will be performed. The quality control assessment includes: (1) verifying that all inclusion criteria is met, (2) confirming that all pertinent information is provided, and (3) checking that data is formatted properly. After the authors have processed a dataset (typically less than 10 days), collaborators will be notified which values will be accepted or if any revisions are needed. Data will be made available as soon as the quality control assessment is complete.

Table S1. Quality Assessment (QA) Compound List

Standard reference CCS values obtained on a specially-modified drift tube instrument as previously reported.[2]

Compound	*m/z*	Ion Species	Stepped Field CCS (Å²)	Single Field CCS (Å²)
Small Molecules				
Cortisol	363.22	M+H	189.27 ± 0.10	188.34 ± 0.00
Cortisol	385.20	M+Na	213.72 ± 0.00	212.79 ± 0.07
Creatinine	112.05	M-H	120.69 ± 0.15	118.84 ± 0.07
Creatinine	114.07	M+H	123.86 ± 0.00	122.98 ± 0.02
Creatinine	136.05	M+Na	132.99 ± 0.35	132.61 ± 0.36
Glucose	203.05	M+Na	147.34 ± 0.29	146.94 ± 0.07
Homocysteine	136.04	M+H	130.77 ± 0.05	129.58 ± 0.63
L-arginine	173.10	M-H	138.03 ± 0.05	137.08 ± 0.01
L-arginine	175.12	M+H	136.84 ± 0.05	136.45 ± 0.00
L-aspartic acid	132.03	M-H	120.39 ± 0.40	119.15 ± 0.04
L-cystine	239.02	M-H	144.38 ± 0.09	143.58 ± 0.01
L-cystine	241.03	M+H	150.07 ± 0.05	149.48 ± 0.03
L-cystine	263.01	M+Na	151.81 ± 0.10	151.26 ± 0.13
L-glutamic acid	146.05	M-H	125.65 ± 0.15	124.47 ± 0.00
L-histidine	154.06	M-H	130.01 ± 0.09	128.83 ± 0.00
L-histidine	156.08	M+H	132.74 ± 0.11	131.93 ± 0.02
L-histidine	178.06	M+Na	135.47 ± 0.50	134.39 ± 0.44
L-isoleucine	130.09	M-H	131.28 ± 0.05	129.83 ± 0.01
L-isoleucine	132.10	M+H	133.81 ± 0.04	132.88 ± 0.03
L-leucine	130.09	M-H	132.51 ± 0.01	131.14 ± 0.00
L-leucine	132.10	M+H	135.55 ± 0.06	134.57 ± 0.03
L-lysine	147.11	M+H	131.62 ± 0.52	131.22 ± 0.14
L-methionine	150.06	M+H	134.07 ± 0.40	133.02 ± 0.47
L-phenylalanine	164.07	M-H	141.29 ± 0.19	139.94 ± 0.03
L-phenylalanine	166.09	M+H	141.27 ± 0.05	140.30 ± 0.12
L-proline	116.07	M+H	126.21 ± 0.20	125.38 ± 0.08
L-tyrosine	180.07	M-H	145.58 ± 0.34	144.42 ± 0.07
L-tyrosine	182.08	M+H	146.44 ± 0.20	145.58 ± 0.12
Levomefolic Acid	458.18	M-H	200.56 ± 0.11	198.99 ± 0.01
Levomefolic Acid	460.19	M+H	197.52 ± 0.26	197.17 ± 0.04
Pyridoxal Phosphate	246.02	M-H	150.80 ± 0.10	149.35 ± 0.04
Pyridoxal Phosphate	248.03	M+H	151.94 ± 0.10	151.37 ± 0.02
Pyridoxal Phosphate	270.01	M+Na	161.40 ± 0.20	161.46 ± 0.20
Uric Acid	167.02	M-H	126.92 ± 0.05	125.55 ± 0.07
Peptides				
Angiotensin1	1296.69	M+H	357.31 ± 0.26	355.62 ± 0.41
Angiotensin1	648.85	M+2H	387.29 ± 0.20	388.41 ± 0.10
Angiotensin1	432.90	M+3H	474.70 ± 0.15	477.05 ± 0.04
Angiotensin1	324.93	M+4H	549.23 ± 0.05	550.98 ± 0.07
Angiotensin2	1046.54	M+H	314.38 ± 0.15	313.65 ± 0.03
Angiotensin2	523.78	M+2H	353.79 ± 0.17	355.09 ± 0.03
Angiotensin2	349.52	M+3H	436.23 ± 0.20	437.30 ± 0.12
Bradykinin	1060.57	M+H	315.25 ± 0.30	314.00 ± 0.12
Bradykinin	530.79	M+2H	343.32 ± 0.10	344.99 ± 0.03

Compound	***m/z***	**Ion Species**	**Stepped Field CCS (Å²)**	**Single Field CCS (Å²)**
Bradykinin	354.19	M+3H	447.60 ± 0.11	449.07 ± 0.38
Melittin	1423.38	M+2H	613.36 ± 0.11	614.26 ± 0.02
Melittin	949.26	M+3H	721.06 ± 0.53	722.45 ± 0.02
Melittin	712.20	M+4H	756.78 ± 0.53	760.82 ± 0.12
Melittin	569.96	M+5H	808.60 ± 0.60	815.39 ± 0.10
Melittin	569.96	M+5H	844.39 ± 0.25	854.37 ± 0.15
Neurotensin	836.96	M+2H	434.32 ± 0.20	435.42 ± 0.06
Renin Substrate	879.97	M+2H	460.38 ± 0.40	461.11 ± 0.03
Renin Substrate	586.98	M+3H	518.81 ± 0.36	524.12 ± 0.07
Renin Substrate	440.49	M+4H	634.59 ± 0.35	637.65 ± 0.23
Substance P	1347.74	M+H	362.51 ± 0.20	361.44 ± 0.04
Substance P	674.37	M+2H	399.87 ± 0.20	400.09 ± 0.05
Substance P	449.92	M+3H	495.73 ± 1.29	496.51 ± 0.37
Proteins				
Cytochrome C	773.39	M+16H	3403.2 ± 2.10	3420.2 ± 2.38
Cytochrome C	727.96	M+17H	3538.1 ± 0.28	3554.7 ± 0.70
Cytochrome C	687.57	M+18H	3655.3 ± 1.57	3670.4 ± 0.74
Cytochrome C	651.44	M+19H	3741.8 ± 0.82	3757.9 ± 0.00
Cytochrome C	618.92	M+20H	3816.1 ± 0.79	3832.3 ± 0.00
Ubiquitin	856.98	M+10H	2192.3 ± 0.60	2204.8 ± 0.41
Ubiquitin	779.16	M+11H	2349.1 ± 0.77	2362.3 ± 0.00
Ubiquitin	714.32	M+12H	2424.6 ± 0.88	2444.2 ± 00
Ubiquitin	659.45	M+13H	2577.7 ± 0.63	2594.3 ± 0.53
Ubiquitin	612.41	M+14H	2727.4 ± 4.94	2728.8 ± 1.74
Ubiquitin	1223.80	M+7	1773.2 ± 1.26	1785.4 ± 0.29
Ubiquitin	1223.80	M+7	1875.7 ± 1.03	1884.3 ± 0.29
Ubiquitin	1070.96	M+8	1950.9 ± 0.24	1960.5 ± 0.33
Ubiquitin	952.08	M+9	2052.4 ± 0.64	2063.4 ± 0.00

Section S3. True Effective Length Calculation and CCS Scaling to Reference Values for Stepped Field DTIMS

The reference CCS values summarized in Table S1g were measured on a drift tube modified with grids at the entrance and exit of the drift region to minimize electric field penetration effects (i.e., fringing fields);[7,8] and allow for CCS calculations to be performed using a precise, geometric length of the drift tube. Thus, these values are considered Standard Reference Values for purposes of accuracy comparisons. Commercially-available drift tubes are prone to drift field inhomogeneity and imprecise determination of the drift length, and therefore, a true effective length must be calculated in order for CCS measurements to correspond to the Standard Reference Values. More detailed information on the sources of measurement variability and a propagation of uncertainty for drift tubes is discussed elsewhere.[2]

The supplemental information packet includes a file entitled "SI_SteppedField_ScaleAndDataFormat.xlsx". The two spreadsheets ("Stepped Field Reference Standards and Scale" and "Stepped Field Data Format") within the Excel file will need to be populated prior to submission of stepped field data to the unified CCS Compendium. *Caution: If the Excel file is opened in read-only mode, the spreadsheet will not be editable. Please click 'enable editing' to proceed.*

The following true effective length calculation and scaling procedure must be followed for each acquisition period (at least one reference standards calibration and true effective length calculation per day of data acquisition).

Step 1: Collect at minimum, triplicate measurements of the reference standards for each day samples are acquired.

- Ideal strategy: Infuse reference standards simultaneously with the analyte(s) of interest. This allows the reference standards to be observed in the same conditions that analytes are exposed to.
- If measuring reference standards independently, acquire reference standards CCS measurements before, during, and after each set of acquisition to assess drift in mass and mobility measurements.

Step 2: Populate columns A and B in the spreadsheet entitled "Stepped Field Reference Standards and Scale" (see Fig. S8) with data generated from step 1 for each replicate.

- Use rows 7-16 for positive ion mode and/or rows 17-26 for negative ion mode.
- CCS and *m/z* values can be obtained using the "CCS Calculator (Stepped-Field)" method in IM-MS Browser (Agilent Technologies). Alternately, these can be calculated from the fundamental equation as discussed in the previous section.
- The experimental effective length (in cm) needs to be entered in the yellow box (Cell D4) located at the top of the spreadsheet. This length can be found in the "BaseDataAccess.dll.config" file located in the Mass Hunter Workstation (Agilent Technologies) install directory (typically: C Drive > Program Files > Agilent > MassHunter > Workstation > IMS > B.07.02 > Bin). Alternately, this is the length value used in the initial CCS calculation that is to be scaled.

Step 3: Columns C-G will be auto-populated.

- The scale factor in column F is calculated for each reference compound CCS using the following equation:

$$scale\ factor = \sqrt{\frac{experimental\ CCS}{reference\ standard\ CCS}}$$

- New effective length values for each reference standards ion in Column G are calculated using the equation:

new effective length = scale factor * experimental effective length

Step 4: The true effective length (yellow cell G27, at the bottom of the "New Effective Length" column) is automatically calculated.

- This value is the average of all calculated effective lengths originating from each reference standards ion.
- The true effective length value is automatically carried over to the "Stepped Field Data Format spreadsheet" to scale experimental data.

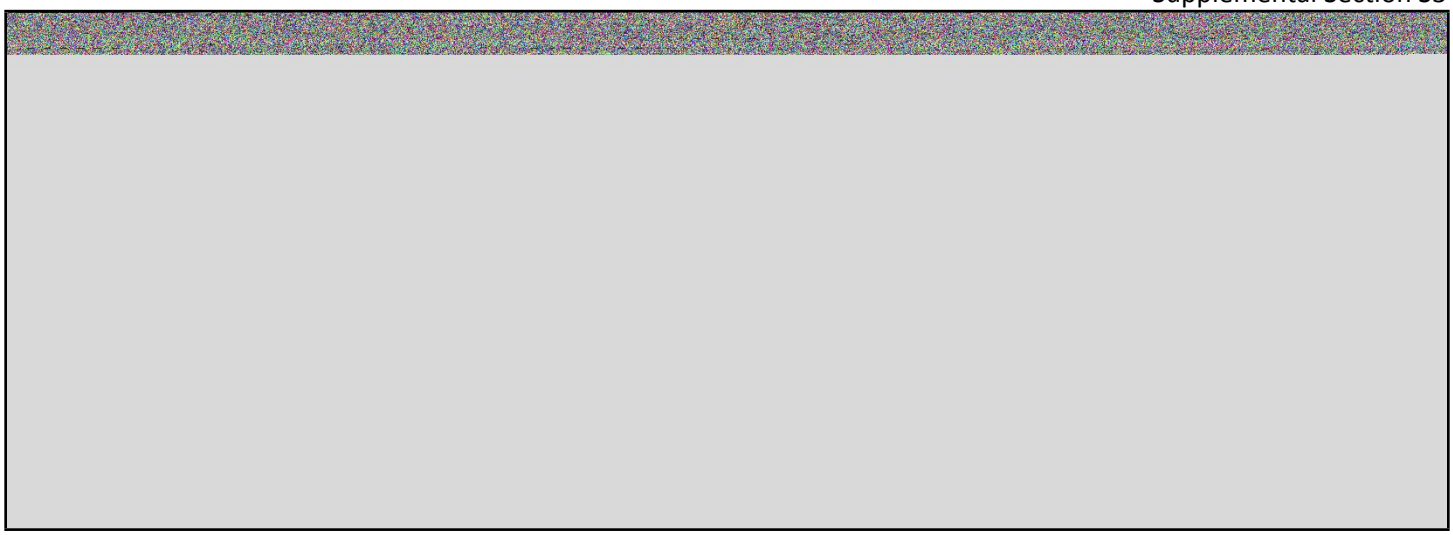

Figure S8. "Stepped Field Reference Standards and Scale" spreadsheet

Step 5: To scale experimental CCS values, follow steps 2-6 of the Stepped Field Data Guidelines for Data Submission (Section S1) before continuing.

- Scaled CCS values are auto-populated in column P using the equation:

$$scaled\ CCS = \ experimental\ CCS \cdot \left(\frac{experimental\ effective\ length}{true\ effective\ length}\right)^{\wedge}2$$

- An example is shown in Fig. S9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Compound	Formula	CAS	Adduct	Charge	Experimental Replicate 1 *m/z*	Experimental Replicate 2 *m/z*	Experimental Replicate 3 *m/z*	Experimental Replicate 1 CCS	Experimental Replicate 2 CCS	Experimental Replicate 3 CCS	Average Experimental *m/z*	Average Experimental CCS	Std. Dev	% RSD	Scaled CCS	CCS/z	Peak Number
2	Example Lipid	C45H73NO8P	5634-86-6	[M-H]	-1	786.5070	786.5074	786.5109	277.30	277.33	277.27	μ *m/z*	μ CCS	σ	=σ/μCCS*100	=μ*(old effective length/ new effective length)^2	277.30	1
3	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	361.68	361.76	362.75	620.91	362.06	0.60	0.17	363.28	181.03	1
4	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	373.66	373.09	372.45	620.91	373.07	0.60	0.16	374.32	186.53	2
5	Cyclosporin	C62H111N11O12		[M+H+K]	+2	620.9060	620.9068	620.9072	388.97	389.27	388.27	620.91	388.84	0.51	0.13	390.15	194.42	3
6	Your	Data	Here!															

Figure S9. "Stepped Field Data Format" spreadsheet (Columns A-Q)

Section S4

Table S2. All super classes and classes represented in the Unified CCS Compendium at date of submission

Super Class	Class	*m/z* Range	N
Alkaloids and derivatives	Yohimbine alkaloids	609	1
		138 – 164	3
Benzenoids	Anthracenes	178 – 271	9
	Benzene and substituted derivatives	108 – 886	159
	Fluorenes	166	1
	Indanes	300	1
	Naphthalenes	128 – 254	25
	Pentacenes	278, 280	2
	Phenanthrenes and derivatives	178 – 303	19
	Phenols	109 – 208	31
	Pyrenes	202 – 304	22
Homogeneous metal compounds	Homogeneous transition metal compounds	132 – 2991	62
Homogeneous non-metal compounds	Non-metal oxoanionic compounds	200	1
Lipids and lipid-like molecules	Fatty acyls	125 – 935	223
	Glycerolipids	253 – 746	8
	Glycerophospholipids	171 – 1017	334
	Prenol lipids	137 – 886	21
	Sphingolipids	548 – 8989	146
	Steroids and steroid derivatives	287 – 648	78
Nucleosides, nucleotides, and analogues	(5'->5')-dinucleotides	662 – 783	29
	5'-deoxyribonucleosides	250 – 408	19
	Flavin nucleotides	455 – 809	9
	Imidazole ribonucleosides and ribonucleotides	337 – 362	4
	Nucleoside and nucleotide analogues	24, 268	2
	Purine nucleosides	250 – 613	49
	Purine nucleotides	280 – 790	125
	Pyrimidine nucleosides	226 – 281	23
	Pyrimidine nucleotides	304 – 646	124
Organic acids and derivatives	Carboximidic acids and derivatives	131, 154	2
	Carboxylic acids and derivatives	89 – 2110	623
	Keto acids and derivatives	115 – 184	11
	Hydroxy acids and derivatives	103 – 239	12
	Organic carbonic acids and derivatives	155	1
	Organic phosphonic acids and derivatives	124 – 205	13
	Organic sulfonic acids and derivatives	124 – 213	4
	Peptidomimetics	225 – 1317	23

Super Class	Class	*m/z* Range	N
Organic acids and derivatives	Proteins	493 – 3302	139
	Sulfinic acids and derivatives	108, 111	2
	Tryptic peptides	288 – 1580	254
Organic nitrogen compounds	Organonitrogen compounds	74 – 1233	101
Organic oxygen compounds	Organic oxoanionic compounds	227 – 411	6
	Organooxygen compounds	105 – 1505	340
Organic polymers	Cyclic Peptides	1111 – 1704	20
	Polypeptides	294 – 1724	230
Organohalogen compounds	Organofluorides	301 – 2834	66
Organoheterocyclic compounds	Azoles	127 – 458	12
	Benzimidazoles	145 – 225	4
	Benzodioxoles	191 – 272	4
	Benzopyrans	421, 424	2
	Dihydrofurans	173 – 350	6
	Dithiolanes	228	1
	Furofurans	199, 200	2
	Imidazopyrimidines	119 – 218	60
	Indoles and derivatives	148 – 381	69
	Lactams	348 – 738	10
	Lactones	153 – 350	6
	Naphthofurans	821 – 1684	14
	Pteridines and derivatives	162 – 483	28
	Pyridinecarboxylic acids and derivatives	140	1
	Pyridines and derivatives	96 – 285	52
	Pyrroles	110	1
	Quinolines and derivatives	172 – 431	17
	Tetrahydroisoquinolines	178 – 181	2
	Tetrapyrroles and derivatives	563 – 1378	9
	Triazines	215 – 325	8
Phenylpropanoids and polyketides	Anthracyclines	540 – 1320	6
	Cinnamaldehydes	133, 134	2
	Cinnamic acids and derivatives	149 – 360	5
	Coumarins and derivatives	161, 186	2
	Flavonoids	269 – 1424	44
	Isoflavonoids	140 – 418	16
	Linear 1,3-diarylpropanoids	255 – 280	4
	Macrolactams	786 - 825	3
	Macrolides and analogues	661– 955	15
	Phenylpropanoic acids	165 – 284	11
	Tetracyclines	410	8
Polyhedralcarbon molecules		720, 840	2

Section S5. Nonlinear Regression Equations

Power Fit

$$y = a \cdot x^{-k} + y_0 \tag{1}$$

a is the curve max – curve min; k is the curve rate

Four-Parameter Sigmoidal Fit (4P)

$$y = y_o + \frac{y_{max} - y_0}{1 + 10^{(\log y_{50} - x) \cdot H}} \tag{2}$$

y_{50} is x at curve half-maximum; H is the Hill Slope

Five-Parameter Sigmoidal Fit (5P)

$$y = y_o + \frac{y_{max} - y_0}{\left(1 + 10^{((\log y_{50} - x) \cdot H}\right)^S} \tag{3}$$

S is the curve symmetry parameter

Confidence Interval

$$z \cdot s_{y,x} \cdot \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}\right)^{1/2} \tag{4}$$

z is standard deviations z score based on interval percentage (z-score for 99% is 2.576);
$S_{y,x}$ is the standard error of the x and y data inputs;
SS_x is the sum of the squared deviations from the x input mean

Predictive Intervals

$$z \cdot s_{y,x} \cdot \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}\right)^{1/2} \tag{5}$$

Section S6. Supplemental Experimental Methods: LC-MS and LC-IM-MS Acquisition Parameters

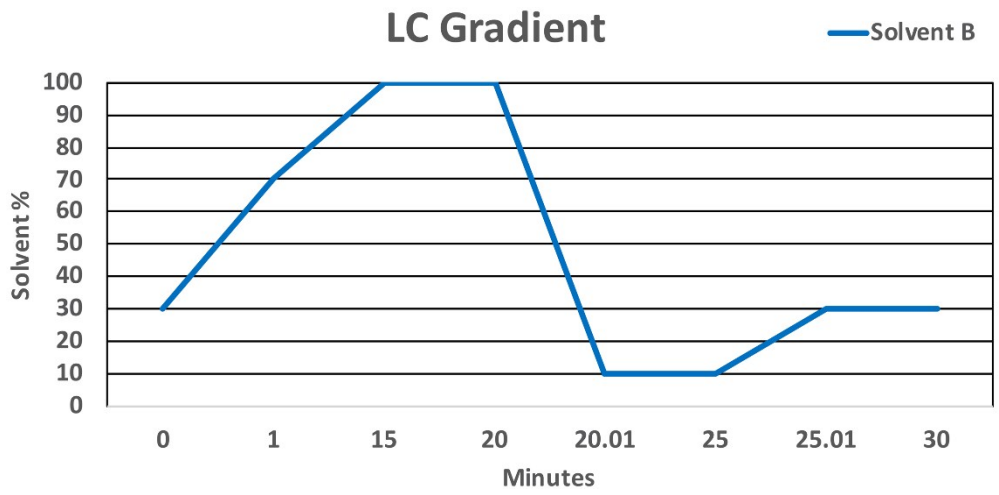

Figure S10. LC Gradient

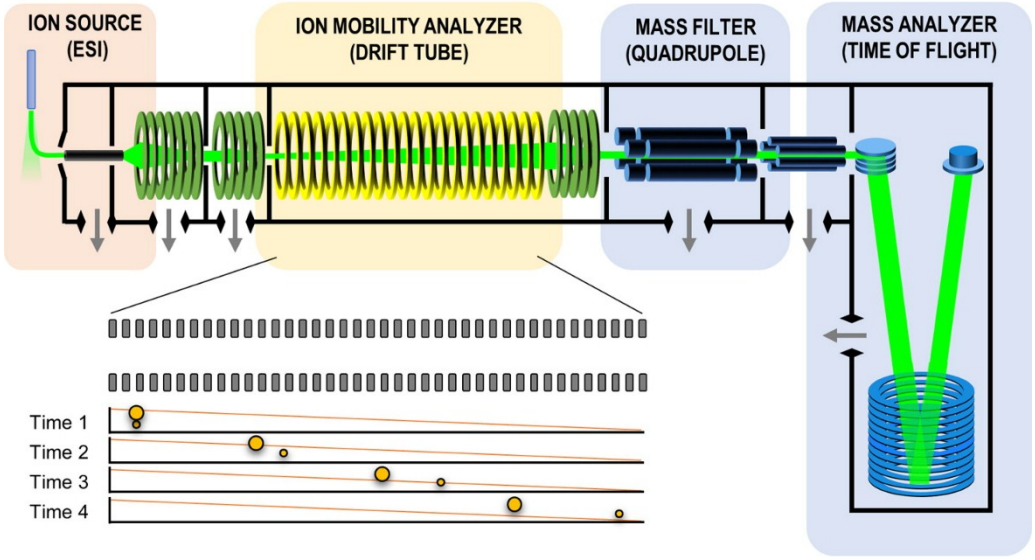

Figure S11. Instrument schematic from reference [6].

Serum samples were analyzed via liquid chromatographic separation using an C18 Zorbax RRHD (1.8µm) column on a 1290 Infinity LC system (Agilent Technologies). Solvent A was water with 0.1% formic acid; and Solvent B was 3:2 isopropanol:acetonitrile with 0.1% formic acid. 2 µl of sample were injected via autosampler and separations occurred using a 30 min gradient described in Fig. S5a at 200 µl/min. Post-LC separation, analytes were ionized using an electrospray ionization source (Jet Stream, Agilent Technologies) at 300°C and a VCap voltage of 3500 V. The drying gas flow rate was 8 L/min, while the sheath gas flowed at 11 L/min. When data was acquired using LC-IM-MS mode, ion mobility separations were performed using a uniform field drift tube with high-purity nitrogen drift gas at 3.95 Torr at room temperature (~298 K). A single field analysis at 17.26 V/cm was performed on a standardized calibrant mixture (Agilent Tune Mix) to normalize sample drift times. Time-of-flight scan range was 100 *m/z* to 1700 *m/z*. Further information can be found in previous work.[2]

References

1 Mclean Research Group, CCS Compendium, https://www.vanderbilt.edu/AnS/Chemistry/groups/mcleanlab/ccs.html.

2 S. M. Stow, T. J. Causon, X. Zheng, R. T. Kurulugama, T. Mairinger, J. C. May, E. E. Rennie, E. S. Baker, R. D. Smith, J. A. McLean, S. Hann and J. C. Fjeldsted, *Anal. Chem.*, 2017, **89**, 9048–9055.

3 Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner and D. S. Wishart, *J. Cheminform.*, 2016, **8**, 1–20.

4 E. A. Mason and E. W. McDaniel, *Transport Properties of Ions in Gases*, John Wiley & Sons, Ltd., New York City, NY, 1988.

5 W. F. Siems, L. A. Viehland and H. H. Hill, *Anal. Chem.*, 2012, **84**, 9782–9791.

6 J. C. May, C. R. Goodwin, N. M. Lareau, K. L. Leaptrot, C. B. Morris, R. T. Kurulugama, A. Mordehai, C. Klein, W. Barry, E. Darland, G. Overney, K. Imatani, G. C. Stafford, J. C. Fjeldsted and J. A. McLean, *Anal. Chem.*, 2014, **86**, 2107–2116.

7 T. Wyttenbach, P. R. Kemper and M. T. Bowers, *Int J Mass Spectrom*, 2001, **212**, 13–23.

8 K. Kaplan, S. Graf, C. Tanner, M. Gonin, K. Fuhrer, R. Knochenmuss, P. Dwivedi and H. H. Hill, *Anal. Chem.*, 2010, **82**, 9336–9343.