Int. J. Ion Mobil. Spec. (2008) 11:71–76
DOI 10.1007/s12127-008-0010-9

SOFTWARE APPLICATIONS

Recommendation of a standard format for data sets from GC/IMS with sensor-controlled sampling

Wolfgang Vautz · Bertram Bödeker · Sabine Bader · Jörg Ingo Baumbach

Received: 20 June 2008 / Accepted: 29 July 2008 / Published online: 19 August 2008

Abstract The harmonization of data formats is always under discussion, especially with respect to the increasing application of ion mobility spectrometry in metabolomics and different other life sciences. To organise the exchange between different types of ion mobility spectrometers (IMS) using various pre-separation techniques [gas-chromatography (GC), e.g. multi-capillary columns (MCC)] applied and several sensors for a controlled sampling and to start a uniform visualisation procedure, a data format is recommended with respect to further use in data acquisition, visualisation, peak finding, signal comparison and data mining. Although the format is optimised for MCC/IMS and GC/IMS with sampling control by CO_2 or flow sensors for breath analysis, its flexibility is ensured by the possibility of version-controlled modifications. The data format proposed will be described in detail.

Keywords Ion mobility spectrometry · Gas-chromatography · Multi-capillary column · Sampling · Data format · Data acquisition · Visualisation · Peak finding · Data mining

W. Vautz (✉) · B. Bödeker · J. I. Baumbach
Department of Metabolomics,
ISAS-Institute for Analytical Sciences,
Bunsen-Kirchhoff-Straße 11,
44139 Dortmund, Germany
e-mail: vautz@isas.de

S. Bader
Roche Diagnostics GmbH,
Nonnenwald 2,
82377 Penzberg, Germany

Introduction

The challenge of harmonization of data generated by different instruments of the same method starts mostly in a single lab. Variations in the experimental setup or in the design need to be recorded to enable a later comparison of the results. In recent years typically rather broad application of instruments under different conditions and with a high repetition rate are required, particularly outside the controlled conditions of the laboratory and no longer only the exemplary detection of particular analytes in traces in air. The resulting problems with respect to the parameters to be controlled and recorded is discussed continuously from the first conferences on ion mobility spectrometry since the 1990s [1]. The need becomes much more essential with the change of the specific question on IMS “Is a particular analyte present?” like known explosives, drugs or chemical warfare agents—to the global question to MCC/IMS e.g. in breath analysis “Which analytes are present in which concentration?”[2]. The progress of the instrumentation towards on-site applications becomes more visible since 1999 [3], but little progress was made considering an IUPAC standard proposed between 1998 and 2002 [4–7].

Therefore, a data format is proposed as used in the Department of Metabolomics of ISAS—Institute for Analytical Sciences. A comprehensive software package was developed and will be described soon in this journal, including visualization, peak comparison, peak finding and data mining. It should be possible to transform other formats used into a data format compatible to this software, including different types developed outside of ISAS. Furthermore, data from other, related methods which potentially could be of interest for a comparison such as data from DMS, FAIMS, GC, GC/MS, could easily be imported by a suitable data evaluation software together with the standard IMS data. However, a standard data format for the mentioned methods does not exist as well and therefore the import procedure has to be adapted for each particular format.

Data format

To support the interaction of scientists in different laboratories using different types of IMS, a standard protocol as basic data format was proposed and will be described in detail. First of all the described format was developed for raw data and therefore should not be changed for any reason after recording. If, for any reason, the data is found to be invalid during the following evaluation process or if any corrections in the header are required, e.g. due to faulty insertion by the operator, the raw data file has to be converted into a corrected data file which can be indicated by extension of the original file name. This evolution of the original data file has to be recorded in a suitable data bases to guarantee a consistent data set always.

The entire data file is structured into the following sections:

- header (all header lines are marked by a starting "#")
  - general information line (1–11)
  - sample information line (12–22)
  - IMS—information line (23–78)
  - external sampling control line (79–97)[1]
  - statistics (line 98–130)
- data matrix (starting from line 131).

[1] A controlled sample is needed, when the sample should not be introduced continuously into the IMS to avoid contamination. This can be done by help of a loop which is filled with the sample and then will be introduced into the IMS. If human breath should be analysed, inhalation and expiration have to be differentiated. This can be done e.g. by help of flow or CO2 sensors. Their signal can be used to control e.g. a magnetic valve which enables filling of the sample loop only when e.g. the subject exhales.

The data are stored in a monthly directory YYYYMM (YYYY—year, MM—month) and the filenames are restricted to the following format:

NNNN_YYMMDDhhmm_ims.csv

with: NNNNIMS short name

YY	year e.g. "08" for 2008
MM	month
DD	day
Hh	hour (start of data acquisition)
Mm	minute (start of data acquisition)
_ims	fixed to describe that the data are obtained from IMS (if a different method is used (e.g. DMS) this has to be changed to e.g. "_dms")
.csv	describes an ASCII format using "," as separator which can be used by all common software applications.

An additional file containing data about the optional sensor controlled sampling has the name "NNNN_YYMMDDhhmm_exsc.csv", where "_exsc" means **ex**ternal **s**ampling **c**ontrol. It is linked from the main data file and is structured in the sections:

- header (all header lines are marked by a starting "#")
  - general information
- data.

The information available in the different sections of both files is described in detail in the following table. It should be mentioned, that the variable names may be changed in the templates manually or some of them, in particular related to the sensors used for temperature, pressure or controlled sampling, ideally in the data acquisition software.

IMS data file format—NNNN_YYMMDDhhmm_ims.csv

Header: General information

Line		Name	Type or value	Comment
1	#	data type	Text	Data type depending on application, e.g. raw data, exsc, …
2	#	version	Text	DA-software version
3	#	template version	Text	template version
4	#	AD-board type	Text	AD-board typ
5	#	ser.-no.	Text	AD-board serial no.

Line		Name	Type or value	Comment
6	#			Free
7	#	date	Date	MM/DD/YYYY
8	#	time	Time	hh:mm:ss, start of data acquisition
9	#	file	NNNN_ YYMMDDhhmm_ ims.csv	file name
10	#			Free
11	#			Free

Header: Sample information

Line		Name	Type or value	Comment
12	#	SAMPLE INFORMATION		Title
13	#			Free
14	#	sample type	Sample, reference, ublank, test signal, series	Values depending on the application
15	#	sample ID	20 char	Sample ID
16	#	comment	Text	Comment
17	#	location	4 char	Location short
18	#	location name	Text	Location
19	#	height ASL/m	Integer	Location: height ASL
20	#	total data acquisition time/s	Real	Duration of the measurement in s
21	#			Free
22	#			Free

Header: IMS information

Line		Name	Type or value	Comment
23	#	IMS - INFORMATION		Title
24	#			Free
25	#	operator	2 char	Operator: short
26	#	operator name	Text	Operator: name
27	#	IMS	4 char	IMS: short
28	#			Free
29	#	K0 RIP positive/ cm^2/Vs	Real	K_o of the RIP in the positive mode
30	#	K0 RIP negative/ cm^2/Vs	Real	K_o of the RIP in the negative mode
31	#	polarity	Positive, negative	Detection mode
32	#	grid opening time/us	Integer	–
33	#			Free
34	#	pause/s	Integer	Delay between 2 spectra
35	#	tD interval (corr.)/ ms from	Real	Recorded interval of drift time from …
36	#	tD interval (corr.)/ ms to	Real	…. to
37	#	1/K0 interval/Vs/ cm^2 from	Real	Recorded $1/K_o$ interval from
38	#	1/K0 interval/Vs/ cm^2 to	Real	…. to
39	#	no. of data points per spectra	Integer	–
40	#	no. of spectra	Integer	–
41	#	no. averaged spectra	Integer	–
42	#	baseline/signal units	Integer	Base line in signal units
43	#	baseline/V	Real	Base line in volt
44	#	V/signal unit	Real	Volt/signal unit
45	#			Free
46	#	drift length/mm	Integer	–
47	#	HV/kV	Real	High voltage applied to drift tube in kilovolt
48	#	amplification/ V/nA	Real	–
49	#			Free
50	#	drift gas	Text	Type
51	#	drift gas flow/ mL/min	Integer	Flow
52	#	sample gas	Text	Type
53	#	sample flow/ mL/min	Integer	Flow
54	#	carrier gas	Text	Type
55	#	carrier gas flow/ mL/min	Integer	Flow
56	#	pre-separation type	Text	E.g. MCC, GC and characteristics of the column
57	#	pre-separation T/deg C	Real	–
58	#	sample loop T/deg C	Real	Optional if sample loop is used instead of direct introduction
59	#	sample loop volume/mL	Real	–
60	#			Free
61	#	ambient T source	Sensor, manual	Temperature—source may be a sensor or manual input
62	#	ambient T/deg C	Real	–
63	#	ambient T x^2	Real	For sensor: conversion from signal to degree Celsius
64	#	ambient T x^1	Real	For sensor: conversion from signal to degree Celsius
65	#	ambient T x^0	Real	For sensor: conversion from signal to degree Celsius
66	#	ambient T x^−1	Real	For sensor: conversion from signal to degree Celsius
67	#	ambient T x^−2	Real	For sensor: conversion from signal to degree Celsius
68	#	ambient p source	Sensor, manual	Pressure—source may be a sensor or manual input

Line		Name	Type or value	Comment
69	#	ambient p/hPa	Real	–
70	#	ambient p x^2	Real	For sensor: conversion from signal to hectopascal
71	#	ambient p x^1	Real	For sensor: conversion from signal to hectopascal
72	#	ambient p x^0	Real	For sensor: conversion from signal to hectopascal
73	#	ambient p x^−1	Real	For sensor: conversion from signal to hectopascal
74	#	ambient p x^−2	Real	For sensor: conversion from signal to hectopascal
75	#			Free
76	#	6-way valve	Manual, auto	When using a sample loop, introduction automatic/manual
77	#			Free
78	#			Free

Header: External sampling control

Line		Name	Type or value	Comment
79	#	EXTERNAL SAMPLING CONTROL		Title
80	#			Free
81	#	control status	Off, on	Controlled sampling
82	#	control zero/ signal units	Integer	Baseline in signal units
83	#	control zero/V	Real	Baseline in volts
84	#	control threshold/signal units	Integer	Threshold sampling on in signal units
85	#	control threshold/V	Real	Threshold sampling on in volts
86	#	control threshold2/signal units	Integer	Threshold sampling off in signal units
87	#	control threshold2/V	Real	Threshold sampling off in volts
88	#	control sampling time/s	Integer	sampling duration in s
89	#	control variable	Text	Control variable
90	#	control dimension	Text	Dimension of control variable
91	#	control x^2	Real	Conversion signal/ dimension of control variable
92	#	control x^1	Real	Conversion signal/ dimension of control variable
93	#	control x^0	Real	Conversion signal/ dimension of control variable
94	#	control x^−1	Real	Conversion signal/ dimension of control variable
95	#	control x^−2	Real	Conversion signal/dimension of control variable
96	#			Free
97	#			Free

Header: Statistics

Line		Name	Type or value	Comment
98	#	STATISTICS		Title
99	#			Free
100	#	RIP detection	Enabled, disabled	Automatic RIP detection
101	#	tD (RIP corr.)/ ms	Real	Drift time RIP in ms
102	#	1/K0 (RIP)/ Vs/cm^2	Real	$1/K_o$ RIP
103	#	K0 (RIP)/ cm^2/Vs	Real	K_o RIP
104	#	SNR (RIP)	Real	Signal-noise ratio RIP
105	#	WHM (RIP)/ Vs/cm^2	Real	Width of half maximum RIP
106	#	res. power (RIP)	Real	Resolving power RIP (drift time/WHM)
107	#			Free
108	#	tD (preRIP corr.)/ms	Real	Drift time preRIP in ms
109	#	1/K0 (preRIP)/ Vs/cm^2	Real	$1/K_o$ preRIP
110	#	K0 (preRIP)/ cm^2/Vs	Real	K_o preRIP
111	#	SNR (preRIP)	Real	Signal-noise ratio preRIP
112	#	WHM (preRIP)/Vs/cm^2	Real	Width of half maximum preRIP
113	#	res. power (preRIP)	Real	Resolving power preRIP (drift time/WHM)
114	#			
115	#	signal RIP/V	Real	Signal height RIP/V
116	#	signal preRIP/ V	Real	Signal height preRIP/V
117	#	RIP/preRIP	Real	Relation RIP/preRIP
118	#			Free
119	#			Free
120	#	Fims/cm^2/kV	Real	Instrument constant ($1/K_o=t_D$/Fims)
121	#			Free
122	#			Free
123	#			Free
124	#			Free
125	#			Free
126	#			Free
127	#			Free
128	#			Free
129	#			Free
130	#			Free

Data matrix

The data matrix starts from line 131 with:

line 131	retention time t_R/s (0 –…)
line 132	spectra no. (0 –…)
1. column	inverse mobility $1/K_o$/V s/cm^2
2. column	corr. drift time $t_{D,corr}$/ms (corr. means corrected with respect to the grid opening time)

Line	1st column: inv. mobility	2nd column: corr. drift time	3rd column: 1st spectrum	4th column: 2nd Spektrum	5th column: 3rd Spektrum	….	…	…	…	…	Comment
131	\	tR	0	0.99	1.99	2.98	3.98	4.98	5.97	6.97	Retention time
132	1/K0	tDcorr.\SNr	0	1	2	3	4	5	6	7	Spectra no.
133	−0.004246	−0.16	−17	−6	−7	−6	−6	−7	−6	−6	1st chromatogram
134	−0.003715	−0.14	−44	−23	−24	−22	−23	−24	−23	−24	2nd chromatogram
135	−0.003185	−0.12	−39	−28	−25	−25	−26	−27	−27	−26	:
136	−0.002654	−0.1	−31	−27	−24	−24	−23	−26	−26	−24	:
137	−0.002123	−0.08	−21	−23	−19	−21	−19	−20	−21	−21	:
138	−0.001592	−0.06	−12	−16	−14	−16	−14	−15	−15	−15	:
139	−0.001062	−0.04	−2	−10	−8	−10	−8	−9	−9	−10	:
140	−0.000531	−0.02	4	−4	−3	−5	−4	−4	−3	−6	:
141	0	0	10	0	0	−1	0	1	2	−2	:
142	0.000531	0.02	19	1	2	2	2	4	4	1	:
143	0.001062	0.04	25	3	5	4	3	6	6	4	:
144	0.001592	0.06	30	3	6	5	5	7	7	4	:
145	0.002123	0.08	33	4	6	5	8	7	8	4	:
146	0.002654	0.1	33	4	5	5	8	7	8	4	:
147	0.003185	0.12	30	4	5	6	7	6	7	3	:
148	0.003715	0.14	36	16	16	18	16	17	17	15	:
:	0.004246	0.16	38	26	26	28	25	28	28	26	:
:	0.004777	0.18	31	30	30	29	30	30	31	29	:
:	:	:	:	:	:	:	:	:	:	:	:

Sensor controlled sampling file format—NNNN_YYMMDDhhmm_exsc.csv

Header: General information

Line		Variable	Format or value	Comment
1	#	data type	IMS exsc data	Data type
2	#	version	Text	DA software version
3	#	exsc template version	Text	Template version
4	#			Free
5	#	date	MM/DD/YYYY	–
6	#	time	hh:mm:ss	hh:mm:ss, start of data acquisition
7	#	file		Linked IMS data file
8	#			Free

Data

Line	Sampling time/s	Control variable	Control status: 0—sampling off 1—sampling on
9	time/s	Flow/L/min	exsc_status
10	0.04	−0.120356	0
11	0.08	−0.060178	0
12	0.12	−0.120356	0
13	0.16	0.000244	0
14	0.2	−0.120356	0
15	0.24	−0.120356	0
16	0.28	0.000244	0
17	0.32	−0.120356	0
18	0.36	−0.060178	0
:	:	:	:
:	:	:	:

Conclusions

With the data format proposed, all major needs considering the experimental conditions and the spectra themselves are stored together, including the information if and how the sampling was controlled by help of an external sensor. Thus, considering the time gap between the measurement and the evaluation, all information needed and normally found in laboratory or instrumentation books is still available as stored in the data file(s). In addition, an assessment of the measurement, the instrument and the data somewhere else will be supported. Therefore, especially in emergency cases and to reduce false alarms, all data could be considered and compared with former data of the same instrumentation. However, the improvement of comparability using suitable normalisation procedures for ion mobility, retention time and signal intensity needs intensive efforts in the near future. Furthermore, time series could be considered with respect to the instrument and to the subject/object of investigation. As an example for application in medical health care: patients staying at home could be controlled by the medical doctors by help of automatically transferred data files with respect to characterisation of exhaled breath for medical purpose (remote diagnosis).

The data format is open for further improvement and hopefully will support the development of a platform to harmonize data input into larger data bases to be built for applications in life sciences, especially to close the gap between different methods in metabolomics. It could be a step forward to bring IMS-data and mass spectrometric data together if applied on the same sample. Thus, GC/MS measurements of samples of human breath could coach MCC/IMS data with respect to validate MCC/IMS findings. Data from MCC/IMS without pre-enrichment could be compared with GC/MS data rather fast and direct. Finally, the application of different methods developed in the fields of bio-informatics and statistics could be available for different GC/IMS applications.

Acknowledgements The financial support of the European Union, the Bundesministerium für Bildung und Forschung and the Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes NRW is thankfully acknowledged.

References

1. Baumbach JI, Davies AN, Irmer Av (1995) Lampen PH Exchange, interpretation, and database-search of ion mobility spectra supported by data format JCAMP-DX. NASA Conf. Publ. FIELD Full Journal Title:NASA Conference Publication 3301:94–111
2. Davies AN, Baumbach JI (1999) Multidimensional data analysis-quantifying the hidden dimension. Spectrosc Eur 11:23–24
3. Baumbach JI, Eiceman GA (1999) Ion mobility spectrometry: arriving on site and moving beyond a low profile. Appl Spectrosc 53:338A–355A
4. Baumbach JI, Lampen P, Davies A (1998) IUPAC/JCAMP-DX: an international standard for the exchange of ion mobility spectrometry data. Int J Ion Mobility Spectrom 1:64–67
5. Baumbach JI, Davies A, Lampen P, Schmidt H (2001) JCAMP-DX. A standard format for the exchange of ion mobility spectrometry data. Pure Appl Chem 73:1765–1782
6. Davies AN, Baumbach JI, Lampen P, Schmidt H (2001) Finalisation of a IUPAC/JCAMP-DX data transfer standard for ion mobility spectrometry data. Int J Ion Mobility Spectrom 4:84–108
7. Davies AN, Lampen P, Schmidt H, Baumbach JI (2002) Reporting ion mobility spectrometry data and the IUPAC/JCAMP-DX international data standard. Int J Ion Mobility Spectrom 5:47–50