

# MassBank Record Format 2.10 (draft)

MassBank Consortium (March 1, 2016)

## Updated

The default Creative Commons license of MassBank record is defined as CC BY.

Two new tags are added, CH\$LINK: INCHIKEY and PK\$SPLASH.

InChI key in CH\$LINK: INCHIKEY is a hashed version of InChI code and used as an optional, common link based on chemical structures. SPLASH in PK\$SPLASH (Section 2.6.1) is a mandatory, hashed identifier of mass spectra.

## 1. Overview

- Each MassBank Record has one-to-one relation to a specific mass spectrum.
  - It is assumed that the sample of measurement of each mass spectrum is a single chemical substance.
- MassBank Record Information is classified into
  - single line information and multiple line information.
  - mandatory and optional
  - unique and iterative.

### 1.1 Syntax Rules

- Single line information is either one of the followings:
  - Tag [ ] [space] Value ( [ ] [space] Value)
  - Tag [ ] [space] subtag [space] Value ( [ ] [space] Value)
- Multiple line information
  - First line is Tag [ ] [space]
  - Following lines are [space] [space] Value
- Last line of a MassBank Record is [//].

### 1.2 Order of Information

- MassBank Record Information in a MassBank Record is arranged in a fixed order (see Section 2).

### 1.3 Others

- [MS] [ ] [space] Value [ ] is the mzOntology ID in OLS (<http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MS>)

Table 1. MassBank Record Format (Summary)

Tag	Mandatory / Optional	Unique/ Iterative	Single line/ Multiple line	Description	Subsection in manual	Note
<b>Record Specific Information</b>						
ACCESSION	M	U	S	Record identifier	2.1.1	
RECORD_TITLE	M	U	S	Short title of the record	2.1.2	
DATE	M	U	S	Date of creation or last modification of record	2.1.3	
AUTHORS	M	U	S	Name and affiliation of authors	2.1.4	
LICENSE	M	U	S	Creative Commons License or its compatible terms	2.1.5	
COPYRIGHT	O	U	S	Copyright	2.1.6	
PUBLICATION	O	U	S	Bibliographic information of reference	2.1.7	
COMMENT	O	I	S	Comments	2.1.8	
<b>Information of Chemical Compound Analyzed</b>						
CH\$NAME	M	I	S	Chemical name	2.2.1	
CH\$COMPOUND_CLASS	M	U	S	Chemical category	2.2.2	
CH\$FORMULA	M	U	S	Chemical formula	2.2.3	
CH\$EXACT_MASS	M	U	S	Exact mass	2.2.4	
CH\$SMILES	M	U	S	SMILES code	2.2.5	
CH\$IUPAC	M	U	S	InChI code	2.2.6	
CH\$LINK: subtag identifier	O	U	S	External database name with identifier	2.2.7	Note 1
<b>Information of Biological Sample</b>						
SP\$SCIENTIFIC_NAME	O	U	S	Scientific name of biological species	2.3.1	
SP\$LINEAGE	O	U	S	Lineage of species		
SP\$LINK: subtag identifier	O	U	S	External database name with identifier		Note 1
SP\$SAMPLE	O	U	S	Information of sample preparations		
<b>Analytical Methods and Conditions</b>						
AC\$INSTRUMENT	M	U	S	Commercial name and manufacturer of instrument	2.4.1	
AC\$INSTRUMENT_TYPE	M	U	S	Type of instrument	2.4.2	Note 2
AC\$MASS_SPECTROMETRY: MS_TYPE	M	U	S	MSn type of data	2.4.3	Note 3

Table 1 (Continued).

AC\$MASS_SPECTROMETRY: ION_MODE	M	U	S	Positive or negative mode of ion detection	2.4.4	
AC\$MASS_SPECTROMETRY: subtag	O	U	S	Analytical conditions of mass spectrometry	2.4.5	Note 1
AC\$CHROMATOGRAPHY: subtag	O	U/I	S	Analytical conditions of chromatographic separation	2.4.6	Note 1,4
Description of mass spectral data						
MS\$FOCUSED_ION: subtag	O	U	S	Precursor ion and $m/z$	2.5.1	Note 1
MS\$DATA_PROCESSING: subtag	O	U	S	Data processing method		Note 1
Peak Information						
PK\$SPLASH	M	U	S	Hashed identifier of mass spectra	2.6.1	
PK\$ANNOTATION	O	U	M	Chemical annotation of peaks by molecular formula	2.6.2	
PK\$NUM_PEAK	M	U	S	Total number of peaks	2.6.3	
PK\$PEAK	M	U	M	Peak ( $m/z$ , intensity, and relative intensity)	2.6.4	
Supplementary Definitions						
Description of isotope-labeled chemical compound					2.7.1	

(Last modified: March 1, 2016)

General note. Decimal point should be a period, “.”, but not a comma, “,”. For example, “ $m/z$  425.7”. No thousand separator is inserted.

Note 1. Within each tag, data fields should be arranged by the alphabetical order of subtag names.

Note 2. Data field of AC\$INSTRUMENT\_TYPE consists of "(Separation tool type-)Ionization method-Ion analyzer type(Ion analyzer type)". Examples of the data are "ESI-QTOF", "ESI-QQ", "GC-EI-EB", "LC-ESI-ITTOF".

Note 3. Data field of AC\$MASS\_SPECTROMETRY: MS\_TYPE is either "MS", "MS2", "MS3", "MS4", , , , .

Note 4. Tag AC\$CHROMATOGRAPHY: SOLVENT is iterative.

## 2. MassBank Record Information

- Table 1 summarizes the current MassBank Record Information.
- MassBank Record Information consists of 6 groups (Table 2).

Table 2. Groups of MassBank Record Information.

<i>Information groups</i>	<i>Sections</i>
Record Specific Information	2.1
Chemical Information (Tag starts with <code>CH\$</code> )	2.2
Sample Information (Tag starts with <code>SP\$</code> )	2.3
Analytical Chemistry Information (Tag starts with <code>AC\$</code> )	2.4
Mass Spectral Data Information (Tag starts with <code>MS\$</code> )	2.5
Mass Spectral Peak Data (Tag starts with <code>PK\$</code> )	2.6

- Information should be arranged by the order shown in Table 2.

### 2.1 Record Specific Information

#### 2.1.1 ACCESSION \*

- Identifier of the MassBank Record. Mandatory
- Example  
ACCESSION: ZMS00006
  - 8-character fix-length string.
  - Prefix two or three alphabetical capital characters specify the site, *i.e.* database, where the record is submitted and stored.
  - Prefixes currently used are listed in the “Prefix of ID” column of the MassBank “Statistics” table (<http://www.massbank.jp/en/statistics.html>).
  - Rest of the field is decimal letters which are the identifier of the record at each site.

#### 2.1.2 RECORD\_TITLE \*

- Brief Description of MassBank Record. Mandatory
- Example  
RECORD\_TITLE: (-)-Nicotine; ESI-QQ; MS2; CE 40 V; [M+H]<sup>+</sup>
  - It consists of the values of `CH$NAME` [ ; ] `AC$INSTRUMENT_TYPE` [ ; ] `AC$MASS_SPECTROMETRY: MS_TYPE`.

### 2.1.3 DATE \*

- Date of the Creation or the Last Modification of MassBank Record. Mandatory
- Example  
DATE: 2011.02.21 (Created 2007.07.07)

### 2.1.4 AUTHORS \*

- Authors and Affiliations of MassBank Record. Mandatory
- Example  
AUTHORS: Akimoto N, Grad Sch Pharm Sci, Kyoto Univ and Maoka T, Res Inst Prod Dev.
  - Only single-byte characters are allowed. For example, ö is not allowed.

### 2.1.5 LICENSE \*

- Creative Commons License of Re-use of MassBank Record. Mandatory
- Example  
LICENSE: CC BY
  - Contributors to MassBank are encouraged to show the license “CC BY”. This license mean that others are free to “share” (copy and redistribute the MassBank record in any medium or format) and to “adapt” (remix, transform, and build upon the MassBank record) for any purpose, even commercially. The contributors cannot revoke these freedoms as long as the others follow the license terms.

### 2.1.6 COPYRIGHT

- Copyright of MassBank Record. Optional
- Example  
COPYRIGHT: Keio University

### 2.1.7 PUBLICATION

- Reference of the Mass Spectral Data. Optional
- Example  
PUBLICATION: Iida T, Tamura T, et al, J Lipid Res. 29, 165-71 (1988). [PMID: 3367086]
  - Citation with PubMed ID is recommended.

### 2.1.8 COMMENT

- Comments. Optional and Iterative
- In MassBank, COMMENT fields are often used to show the relations of the present

record with other MassBank records and with data files. In these cases, the terms in brackets [ ] and [ ] are reserved for the comments specific to the following five examples.

#### Example 1

COMMENT: This record is a MS3 spectrum. Link to the MS2 spectrum is added in the following comment field.

COMMENT: [MS2] KO008089

#### Example 2

COMMENT: This record was generated by merging the following three MassBank records.

COMMENT: [Merging] KO006229 Tiglate; ESI-QTOF; MS2; CE:10 V [M-H]-.

COMMENT: [Merging] KO006230 Tiglate; ESI-QTOF; MS2; CE:20 V [M-H]-.

COMMENT: [Merging] KO006231 Tiglate; ESI-QTOF; MS2; CE:30 V [M-H]-.

#### Example 3

COMMENT: This record was merged into a MassBank record, KOX00012, with other records.

COMMENT: [Merged] KOX00012

#### Example 4

COMMENT: Analytical conditions of LC-MS were described in separate files.

COMMENT: [Mass spectrometry] ms1.txt

COMMENT: [Chromatography] lc1.txt.

#### Example 5

COMMENT: Profile spectrum of this record is given as a JPEG file.

COMMENT: [Profile] CA000185.jpg

## 2.2 Information of Chemical Compound Analyzed

### 2.2.1 CH\$NAME \*

- Name of the Chemical Compound Analyzed. Mandatory and Iterative
- Example  
CH\$NAME: D-Tartaric acid  
CH\$NAME: (2S,3S)-Tartaric acid
- No prosthetic molecule of adducts (HCl, H<sub>2</sub>SO<sub>3</sub>, H<sub>2</sub>O, *etc*), conjugate ions (Chloride, *etc*), and protecting groups (TMS, *etc.*) is included.
- Chemical names which are listed in the compound list are recommended.  
Synonyms could be added.
- If chemical compound is a stereoisomer, stereochemistry should be indicated.

### 2.2.2 CH\$COMPOUND\_CLASS \*

- Category of Chemical Compound. Mandatory
- Example  
CH\$COMPOUND\_CLASS: Natural Product; Carotenoid; Terpenoid; Lipid
- Either `Natural Product` or `Non-Natural Product` should be precedes the other class names .

### 2.2.3 CH\$FORMULA \*

- Molecular Formula of Chemical Compound. Mandatory
- Example  
CH\$FORMULA: C9H10CINO3
- It follows the Hill's System.
- No prosthetic molecule is included (see 2.2.1 CH\$NAME).
- Molecular formulae of derivatives by chemical modification with TMS, *etc.* should be given in the MS\$FOCUSED\_ION: DERIVATIVE\_FORM (2.5.1) field.

### 2.2.4 CH\$EXACT\_MASS \*

- Monoisotopic Mass of Chemical Compound. Mandatory
- Example  
CH\$EXACT\_MASS: 430.38108
- A value with 5 digits after the decimal point is recommended.

### 2.2.5 CH\$SMILES \*

- SMILES String. Mandatory
- Example  
CH\$SMILES: NCC(O)=O
- Isomeric SMILES but not a canonical one.

### 2.2.6 CH\$IUPAC \*

- IUPAC International Chemical Identifier (InChI Code). Mandatory
- Example  
CH\$IUPAC: InChI=1S/C2H5NO2/c3-1-2(4)5/h1,3H2,(H,4,5)
- Not IUPAC name.

### 2.2.7 CH\$LINK: subtag identifier

- Identifier and Link of Chemical Compound to External Databases.

- Optional and Iterative
- Example
  - CH\$LINK: CAS 56-40-6
  - CH\$LINK: INCHIKEY UFFBMTHBGFGIHF-UHFFFAOYSA-N
  - CH\$LINK: KEGG C00037
  - CH\$LINK: PUBCHEM SID: 11916 [CID:182232](#)
- Currently MassBank records have links to the following external databases :
  - CAS
  - CHEBI
  - CHEMPDB
  - CHEMSPIDER
  - INCHIKEY
  - KEGG
  - KNAPSACK
  - LIPIDBANK
  - LIPIDMAPS
  - PUBCHEM
- CH\$LINK fields should be arranged by the alphabetical order of database names.
- InChI Key, a hashed version of InChI code, is a common link by chemical structures.

## 2.3 Information of Biological Sample

### 2.3.1 SP\$SCIENTIFIC\_NAME

- Scientific Name of Biological Species, from Which Sample was Prepared. Optional
- Example
  - SP\$SCIENTIFIC\_NAME: Mus musculus

### 2.3.2 SP\$LINK subtag identifier

- Identifier of Biological Species in External Databases. Optional
- Example
  - SP\$LINK: NCBI-TAXONOMY 10090

### 2.3.3 SP\$SAMPLE

- Tissue or Cell, from which Sample was Prepared. Optional and iterative
- Example
  - SP\$SAMPLE: Liver extracts



## 2.4 Analytical Method and Conditions

### 2.4.1 AC\$INSTRUMENT \*

- Commercial Name and Model of (Chromatographic Separation Instrument, if any were coupled, and) Mass Spectrometer and Manufacturer. Mandatory
- Example  
AC\$INSTRUMENT: LC-10ADVPmicro HPLC, Shimadzu; LTQ Orbitrap, Thermo Electron.
- Cross-reference to mzOntology: Instrument model [MS:1000031]
- All the instruments are given together in a single line. This record is not iterative.

### 2.4.2 AC\$INSTRUMENT\_TYPE \*

- General Type of Instrument. Mandatory
- Example  
AC\$INSTRUMENT\_TYPE: LC-ESI-QTOF
- Format is  
(Separation tool type)-Ionization method-Ion analyzer type(Ion analyzer type).
- Separation tool types are CE, GC, LC.
- Ionization methods are APCI, APPI, EI, ESI, FAB, MALDI.
- Ion analyzer types are B, E, FT, IT, Q, TOF.
- In tandem mass analyzers, no “-“ is inserted between ion analyzers.
- FT includes FTICR and other type analyzers using FT, such as Orbitrap<sup>(R)</sup>.
- IT comprises quadrupole ion trap analyzers such as 3D ion trap and linear ion trap.
- Other examples of AC\$INSTRUMENT\_TYPE data are as follows.
  - ESI-QQ
  - ESI-QTOF
  - GC-EI-EB
  - LC-ESI-ITFT
- Cross-reference to mzOntology: Ionization methods [MS:1000008]; APCI [MS:1000070]; APPI [MS:1000382]; EI [MS:1000389]; ESI [MS:1000073]; B [MS:1000080]; IT [MS:1000264], Q [MS:1000081], TOF [MS:1000084].

### 2.4.3 AC\$MASS SPECTROMETRY: MS\_TYPE \*

- Data Type. Mandatory
- Example  
AC\$MASS\_SPECTROMETRY: MS\_TYPE MS2
- Either of MS, MS2, MS3, MS4, , , .

- Brief definition of terms used in MS\_TYPE
  - `MS2` is 1<sup>st</sup> generation product ion spectrum(of `MS`)
  - `MS3` is 2<sup>nd</sup> generation product ion spectrum(of `MS`)
  - `MS2` is the precursor ion spectrum of `MS3`
    - IUPAC Recommendations 2006  
([http://old.iupac.org/reports/provisional/abstract06/murray\\_prs.pdf](http://old.iupac.org/reports/provisional/abstract06/murray_prs.pdf))

#### 2.4.4 AC\$MASS SPECTROMETRY: ION\_MODE \*

- Polarity of Ion Detection. Mandatory
- Example  
AC\$MASS\_SPECTROMETRY: ION\_MODE POSITIVE
- Either of `POSITIVE` or `NEGATIVE` is allowed.
- Cross-reference to mzOntology: POSITIVE [MS:1000030] or NEGATIVE [MS:1000129]; Ion mode [MS:1000465]

#### 2.4.5 AC\$MASS SPECTROMETRY: subtag Description

- Other Optional Experimental Methods and Conditions of Mass Spectrometry.
- Description is a list of numerical values with/without unit or a sentence.
- AC\$MASS SPECTROMETRY fields should be arranged by the alphabetical order of subtag names.

#### 2.4.5 Subtag: COLLISION ENERGY

- Collision Energy for Dissociation.
- Example 1  
AC\$MASS\_SPECTROMETRY: COLLISION\_ENERGY 20 kV
- Example 2  
AC\$MASS\_SPECTROMETRY: COLLISION\_ENERGY Ramp 10-50 kV

#### 2.4.5 Subtag: COLLISION GAS

- Name of Collision Gas.
- Example  
AC\$MASS\_SPECTROMETRY: COLLISION\_GAS N2
- Cross-reference to mzOntology: Collision gas [MS:1000419]

#### 2.4.5 Subtag: DATE

- Date of Analysis.

2.4.5 Subtag: DESOLVATION GAS FLOW

- Flow Rate of Desolvation Gas.

- Example

AC\$MASS\_SPECTROMETRY: DESOLVATION\_GAS\_FLOW 600.0 l/h

2.4.5 Subtag: DESOLVATION TEMPERATURE

- Temperature of Desolvation Gas.

- Example

AC\$MASS\_SPECTROMETRY: DESOLVATION\_TEMPERATURE 400 C

2.4.5 Subtag: IONIZATION ENERGY

- Energy of Ionization.

- Example

AC\$MASS\_SPECTROMETRY: IONIZATION\_ENERGY 70 eV

2.4.5 Subtag: LASER

- Desorption /Ionization Conditions in MALDI.

- Example

AC\$MASS\_SPECTROMETRY: LASER 337 nm nitrogen laser, 20 Hz, 10 nsec

2.4.5 Subtag: MATRIX

- Matrix Used in MALDI and FAB.

- Example

AC\$MASS\_SPECTROMETRY: MATRIX 1-2 uL m-NBA

2.4.5. Subtag : MASS ACCURACY

- Relative Mass Accuracy.

- Example

AC\$MASS\_SPECTROMETRY: MASS\_ACCURACY 50 ppm over a range of about m/z 100-1000

2.4.5 Subtag: REAGENT GAS

- Name of Reagent Gas.

- Example

AC\$MASS\_SPECTROMETRY: REAGENT\_GAS ammonia

#### 2.4.5 Subtag: SCANNING

- Scan Cycle and Range.
- Example

AC\$MASS\_SPECTROMETRY: SCANNING 0.2 sec/scan (m/z 50-500)

#### 2.4.6 AC\$CHROMATOGRAPHY: subtag Description

- Experimental Method and Conditions of Chromatographic Separation. Optional
- AC\$CHROMATOGRAPHY fields should be arranged by the alphabetical order of subtag names.

#### 2.4.6 Subtag: CAPILLARY VOLTAGE

- Voltage Applied to Capillary Electrophoresis or Voltage Applied to the Interface of LC-MS.
- Example

AC\$CHROMATOGRAPHY: CAPILLARY\_VOLTAGE 4 kV

#### 2.4.6 Subtag: COLUMN\_NAME

- Commercial Name of Chromatography Column and Manufacture.
- Example of LC

AC\$CHROMATOGRAPHY: COLUMN\_NAME Acquity UPLC BEH C18 2.1 by 50 mm (Waters, Milford, MA, USA)

- Example of CE

AC\$CHROMATOGRAPHY: COLUMN\_NAME Fused silica capillary id=50 um L=100 cm (HMT, Tsuruoka, Japan)

#### 2.4.6 Subtag: COLUMN TEMPERATURE

- Column Temperature.
- Example

AC\$CHROMATOGRAPHY: COLUMN\_TEMPERATURE 40 C

#### 2.4.6 Subtag: FLOW GRADIENT

- Gradient of Elution Solutions.
- Example

AC\$CHROMATOGRAPHY: FLOW\_GRADIENT 0/100 at 0 min, 15/85 at 5 min, 21/79 at 20 min, 90/10 at 24 min, 95/5 at 26 min, 0/100, 30 min

#### 2.4.6 Subtag: FLOW\_RATE

- Flow Rate of Migration Phase.
- Example  
AC\$CHROMATOGRAPHY: FLOW\_RATE 0.25 ml/min

#### 2.4.6 Subtag: RETENTION\_TIME

- Retention Time on Chromatography.
- Example  
AC\$CHROMATOGRAPHY: RETENTION\_TIME 40.3 min
- Cross-reference to mzOntology: Retention time [MS:1000016]

#### 2.4.6 Subtag: SOLVENT

- Chemical Composition of Buffer Solution. Iterative
- Example  
AC\$CHROMATOGRAPHY: SOLVENT A acetonitrile-methanol-water (19:19:2) with 0.1% acetic acid  
AC\$CHROMATOGRAPHY: SOLVENT B 2-propanol with 0.1% acetic acid and 0.1% ammonium hydroxide (28%)

### 2.5 Description of Mass Spectral Data

#### 2.5.1 MS\$FOCUSED\_ION: subtag Description

- Information of Precursor or Molecular Ion.
- MS\$FOCUSED\_ION fields should be arranged by the alphabetical order of subtag names.

#### 2.5.1 Subtag: BASE\_PEAK

- *m/z* of Base Peak.
- Example  
MS\$FOCUSED\_ION: BASE\_PEAK 73

#### 2.5.1 Subtag: DERIVATIVE\_FORM

- Molecular Formula of Derivative for GC-MS.
- Example 1  
MS\$FOCUSED\_ION: DERIVATIVE\_FORM C19H42O5Si4
- Example 2  
MS\$FOCUSED\_ION: DERIVATIVE\_FORM C{9+3\*n}H{16+8\*n}NO5Si{n}

### 2.5.1 Subtag: DERIVATIVE MASS

- Exact Mass of Derivative for GC-MS.
- Example  
MS\$FOCUSED\_ION: DERIVATIVE\_MASS 462.21093

### 2.5.1 Subtag: DERIVATIVE TYPE

- Type of Derivative for GC-MS.
- Example  
MS\$FOCUSED\_ION: DERIVATIVE\_TYPE 4 TMS

### 2.5.1 Subtag: ION TYPE

- Type of Focused Ion.
- Example  
MS\$FOCUSED\_ION: ION\_TYPE [M+H]<sup>+</sup>
- Types currently used in MassBank are [M]<sup>+</sup>, [M]<sup>+\*</sup>, [M+H]<sup>+</sup>, [2M+H]<sup>+</sup>, [M+Na]<sup>+</sup>, [M-H+Na]<sup>+</sup>, [2M+Na]<sup>+</sup>, [M+2Na-H]<sup>+</sup>, [(M+NH<sub>3</sub>)+H]<sup>+</sup>, [M+H-H<sub>2</sub>O]<sup>+</sup>, [M+H-C<sub>6</sub>H<sub>10</sub>O<sub>4</sub>]<sup>+</sup>, [M+H-C<sub>6</sub>H<sub>10</sub>O<sub>5</sub>]<sup>+</sup>, [M]<sup>-</sup>, [M-H]<sup>-</sup>, [M-2H]<sup>-</sup>, [M-2H+H<sub>2</sub>O]<sup>-</sup>, [M-H+OH]<sup>-</sup>, [2M-H]<sup>-</sup>, [M+HCOO]<sup>-</sup>, [(M+CH<sub>3</sub>COOH)-H]<sup>-</sup>, [2M-H-CO<sub>2</sub>]<sup>-</sup> and [2M-H-C<sub>6</sub>H<sub>10</sub>O<sub>5</sub>]<sup>-</sup>.

### 2.5.1 Subtag: PRECURSOR M/Z

- *m/z* of Precursor Ion in MS<sub>n</sub> spectrum.
- Example  
MS\$FOCUSED\_ION: PRECURSOR\_M/Z 289.07123
- Calculated exact mass is preferred to the measured accurate mass of the precursor ion.
- Cross-reference to mzOntology: precursor *m/z* [MS:1000504]

### 2.5.1 Subtag: PRECURSOR TYPE

- Type of Precursor Ion in MS<sup>n</sup>.
- Example  
MS\$FOCUSED\_ION: PRECURSOR\_TYPE [M-H]<sup>-</sup>
- Types currently used in MassBank are [M]<sup>+</sup>, [M]<sup>+\*</sup>, [M+H]<sup>+</sup>, [2M+H]<sup>+</sup>, [M+Na]<sup>+</sup>, [M-H+Na]<sup>+</sup>, [2M+Na]<sup>+</sup>, [M+2Na-H]<sup>+</sup>, [(M+NH<sub>3</sub>)+H]<sup>+</sup>, [M+H-H<sub>2</sub>O]<sup>+</sup>, [M+H-C<sub>6</sub>H<sub>10</sub>O<sub>4</sub>]<sup>+</sup>, [M+H-C<sub>6</sub>H<sub>10</sub>O<sub>5</sub>]<sup>+</sup>, [M]<sup>-</sup>, [M-H]<sup>-</sup>, [M-2H]<sup>-</sup>, [M-2H+H<sub>2</sub>O]<sup>-</sup>, [M-H+OH]<sup>-</sup>, [2M-H]<sup>-</sup>, [M+HCOO]<sup>-</sup>, [(M+CH<sub>3</sub>COOH)-H]<sup>-</sup>, [2M-H-CO<sub>2</sub>]<sup>-</sup> and

[2M·H·C6H10O5].

- Cross-reference to mzOntology: Precursor type [MS: 1000792]

### 2.5.3 MS\$DATA PROCESSING: subtag

- Data Processing Method of Peak Detection. Optional
- MS\$DATA PROCESSING fields should be arranged by the alphabetical order of subtag names.
- Cross-reference to mzOntology: Data processing [MS:1000543]

#### 2.5.3 Subtag: FIND\_PEAK

- Peak Detection.
- Example  
MS\$DATA\_PROCESSING: FIND\_PEAK convexity search; threshold = 9.1

#### 2.5.3 Subtag: WHOLE

- Whole Process in Single Method / Software.
- Example  
MS\$DATA\_PROCESSING: WHOLE Analyst 1.4.2

## 2.6 Information of Mass Spectral Peaks

### 2.6.1 PK\$\$SPLASH \*

- Hashed Identifier of Mass Spectra. Mandatory and Single Line Information
- Example  
PK\$\$SPLASH: splash10-z200000000-87bb3c76b8e5f33dd07f

### 2.6.2 PK\$ANNOTATION

- Chemical Annotation of Peaks with Molecular Formula. Optional and Multiple Line Information
- Example 1  
PK\$ANNOTATION: m/z annotation exact\_mass error(ppm) formula  
794.76 [PC(18:0,20:4)-CH3]- 794.56998 239 C45H81NO8P
- Example 2  
PK\$ANNOTATION: m/z {annotation formula exact\_mass error(ppm)}  
494.34  
[lyso PC(alkyl-18:0,-)]- C25H53NO6P 494.36105 -42
- Example 3

PK\$ANNOTATION: m/z formula annotation exact\_mass error(ppm)

167.08947 C9H12O2N [M+1]+(13C) 167.08961 0.81

168.08681 C9H12O2N [M+1]+(13C, 15N) 168.08664 1.04

- Line 1 defines the record format of the annotation blocks. Contributors freely define the record format by using appropriate terms.
- Line 2 or later : sequence of multiple line annotation blocks.
- The first line of each annotation block should be indented by `space space`. The second or later line in each annotation block should be indented by `space space space space`.
- See Section 2.7.2 about more details of Example 3.

### 2.6.3 PK\$NUM\_PEAK \*

- Total Number of Peaks in PK\$PEAK (2.6.3) . Mandatory
- Example  
PK\$NUM\_PEAK 86

### 2.6.4 PK\$PEAK \*

- Peak Data. Mandatory and Multiple line Information
- Example  
PK\$PEAK: m/z int. rel.int.  
326.65 5.3 5  
328.28 7.6 7
- Line 1: fixed string which denotes the format of Line 2 or later.  
PK\$PEAK: m/z int. rel.int.
- Format of Line 2 or later: `space space MZ space INT space REL`
  - MZ: m/z of the peak.
  - INT: intensity of the peak.
  - REL: an integer from 1 to 999 which denotes relative intensity of the peak.
- Peaks are arranged in the ascending order of *m/z*.

## 2.7 Supplementary Definitions

### 2.7.1 Description of Isotope-Labeled Compounds

- This section defines the chemical information of isotope-labeled chemical compounds.
- CH\$NAME is Chemical Name followed by ”–[(Labeled Positions-)Isotopic Atom Name with the Number of Isotopic Atoms]”.
- Example



CH\$NAME: Glycine-[2-13C, 15N]

CH\$NAME: L-Aspartic acid-[2-15N][3,3-d2]

CH\$NAME: Benzene-[d6]

- MOLFILE depends on whether the labeled position is specified. If the labeled position is specified, molfile defines the isotopic atom name and the labeled position. Otherwise molfile should be the same to that of the non-labeled chemical compound.
- CH\$FORMULA should be the same to that of the non-labeled chemical compound.
- CH\$EXACT\_MASS is the monoisotopic mass, but not the sum of the mass of the isotopes. Thus CH\$EXACT\_MASS should be equal to that of the non-labeled chemical compound.
- CH\$SMILES is the same to that of the non-labeled chemical compound.
- CH\$IUPAC, which is InChI code, should define the isotope name and the labeled positions if these two are specified. If not, InChI code is the same to that of the non-labeled chemical compound.
- MS\$FOCUSED\_ION: PRECURSOR\_M/Z should be the value that was actually used in the mass spectrometry.
- MS\$FOCUSED\_ION: PRECURSOR\_TYPE should be the same to that of non-labeled chemical compound.
- Example  
MS\$FOCUSED\_ION: PRECURSOR\_TYPE [M+H]<sup>+</sup>
- Record Editor correctly generates CH\$FORMULA, CH\$EXACT\_MASS, CH\$SMILES, and CH\$IUPAC from the molfile of the isotope-labeled chemical compound.
- 

### 2.7.2 PK\$ANNOTATION of Natural Abundant Isotopic Peaks

- This section describes the annotation of natural abundant isotopic peaks. Optional and Multiple Line Information
- Example  
PK\$ANNOTATION: m/z formula annotation exact\_mass error(ppm)  
167.08947 C9H12O2N [M+1]<sup>+</sup>(13C) 167.08961 0.81  
168.08681 C9H12O2N [M+1]<sup>+</sup>(13C, 15N) 168.08664 1.04
- Line 1 defines the record format of Line 2 or later lines.
- The first line of each annotation block should be indented by space space.