InChI isotopologue and isotopomer specifications

Hunter N. B. Moseley[1*], Philippe Rocca-Serra[2], Reza M. Salek[3], Masanori Arita[4] and Emma L. Schymanski[5]

Abstract

This work presents a proposed extension to the International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI) standard that allows the representation of isotopically-resolved chemical entities at varying levels of ambiguity in isotope location. This extension includes an improved interpretation of the current isotopic layer within the InChI standard and a new isotopologue layer specification for representing chemical intensities with ambiguous isotope localization. Both improvements support the unique isotopically-resolved chemical identification of features detected and measured in analytical instrumentation, specifically nuclear magnetic resonance and mass spectrometry.

Scientific contribution

This new extension to the InChI standard would enable improved annotation of analytical datasets characterizing chemical entities, supporting the FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles of data stewardship for chemical datasets, ultimately promoting Open Science in chemistry.

Keywords International chemical identifier, InChI, Isotopologue, Isotopomer, Chemical annotation, Isotope-resolved

Introduction

The International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI) is an ASCII character string for uniquely identifying a chemical compound [1–3]. The InChI generation algorithm utilizes an augmented (highly modified) McKay's canonical graph labeling algorithm [4, 5] for the purpose of uniquely numbering atoms in a chemical compound graph regardless of the tautomer representation used in the InChI generation. The InChI generation algorithm was originally developed and implemented in the C programming language by the InChI Subcommittee of the IUPAC Chemical Nomenclature and Structure Representation Division (Division VIII). Subsequently in 2009, the InChI Trust, an international organization, was established to oversee the development of the algorithm and its implementations, leaving the IUPAC Division VIII InChI Subcommittee to oversee the conceptual design and specification of the InChI standard itself [6]. The current InChI libraries and software are freely available under the IUPAC-InChI Trust open source license [7] and includes indirect multi-language support through a compiled library that provides an application programming interface (API) [3]. Over the years, the InChI generation algorithm and software have been augmented to recognize more tautomeric forms of chemical compounds [3] as well as provide additional chemical description in the form of additional layers within the InChI string. InChI is also incorporated into larger identifier strings that

*Correspondence:
Hunter N. B. Moseley
hunter.moseley@uky.edu
[1] Department of Molecular & Cellular Biochemistry, University of Kentucky, Lexington, KY, USA
[2] Department of Engineering Science, University of Oxford E-Research Centre, University of Oxford, Oxford OX1 3QG, UK
[3] BiotechVision, Cambridge, UK
[4] Bioinformation and DDBJ Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan
[5] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg

represent other chemical phenomena including reactions (RInChI) [8], mixtures (MInChI) [9, 10], nanomaterials (NInChI) [11], and variable R-groups (VInChI and Mark-InChI) [3].

In its current form, the InChI standard cannot be used to accurately annotate spectral features for isotopically-resolved chemical entities detected in stable isotope-resolved metabolomics experiments, especially isotopologues. This was described to the IUPAC Division VIII InChI Subcommittee in 2018 as the "isotopically-resolved spectral feature annotation problem" (Fig. 1). An InChI Isotopologue Working Group was formed to address this challenge, with the following members: Drs. Hunter Moseley (chair), Philippe Rocca-Serra, Reza M. Salek, Masanori Arita, and Emma L. Schymanski. Within six months, the working group developed an initial proposal for extending the InChI standard for representing isotopes within chemical compounds, including isotopomers and isotopologues. The proposal was presented to the IUPAC Division VIII InChI Subcommittee and

to the scientific community at Metabolomics 2018, the annual conference of the International Metabolomics Society, in Seattle, Washington USA for feedback and public comment. The original proposal was thus developed into the InChI extension suggested here based on the feedback provided (see supplemental material).

This article presents the InChI Isotopologue and Isotopomer extension, which includes new layer specifications within the InChI standard for representing isotopes within chemical compounds. The extension builds upon the previous isotopic "/i" layer specification, but does not change the specification of this layer. Instead, the interpretation of this layer has been altered to better represent isotopically-resolved chemical information. In addition, a new isotopologue "/a" layer specification has been added to represent isotopologues, i.e., ambiguous isotope localization. The new isotopic "/i" layer interpretation and the new isotopologue "/a" layer specification support the unique identification of features detected and measured in analytical instrumentation corresponding
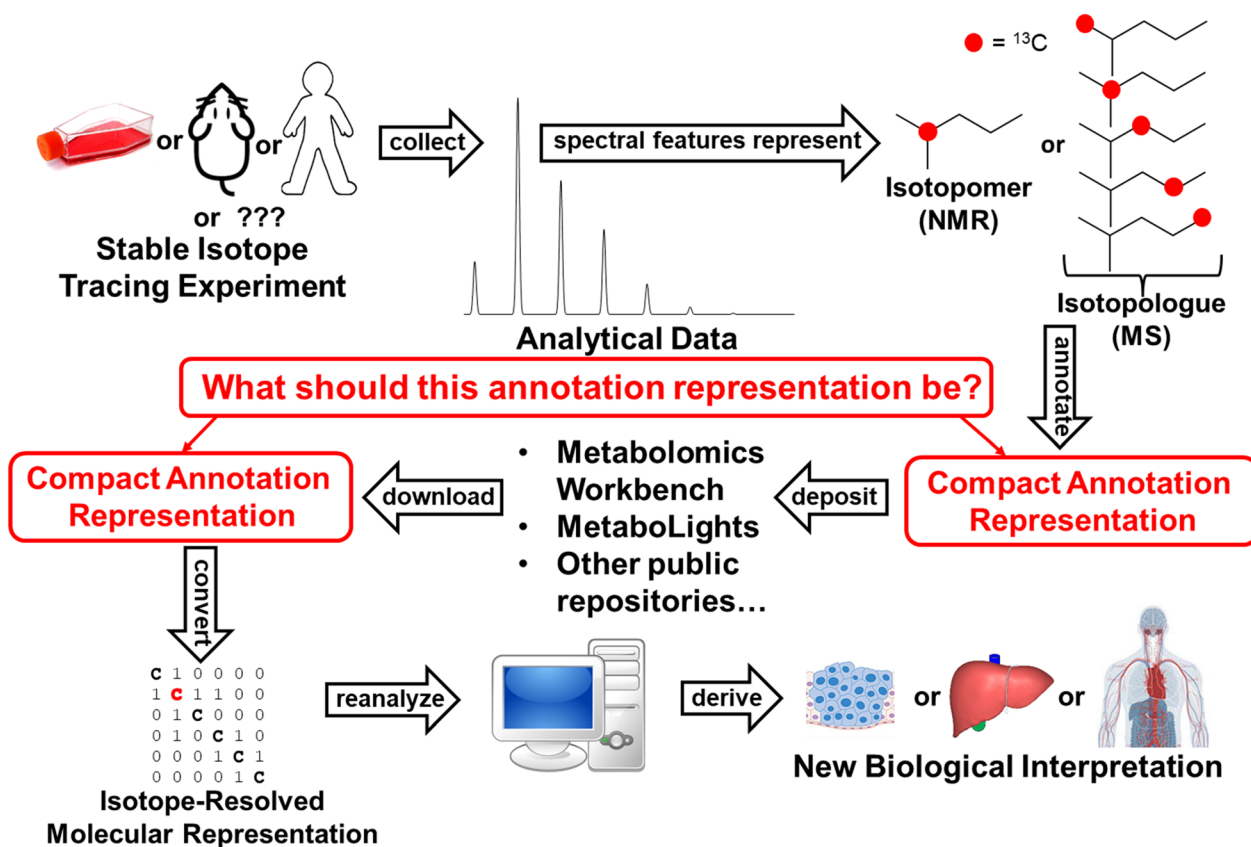

Fig. 1 Overview of the isotopically-resolved spectral feature annotation challenge. First, samples are taken from biological entities in stable isotope tracing experiments. Next, analytical data representing nuclear magnetic resonance (NMR) derived isotopomer spectral features and mass spectrometry (MS) isotopologue spectral features is collected. Spectral features require isotopically-resolved annotation before deposition into a scientific repository for later retrieval and reanalysis for deriving new biological interpretation within different experimental contexts. How can we best represent isotopically-resolved annotations? What should this annotation representation be?

Moseley *et al. Journal of Cheminformatics* (2024) 16:54

Page 3 of 9

to isotopically-resolved chemical entities. Thus, the new extension enables improved annotation of analytical datasets used for chemical characterization. Such expansion of chemical annotation capabilities will enhance accurate and specific documentation of the isotopic content of chemical features and entities, promoting the FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles of scientific data stewardship [12] and Open Science [13, 14] with respect to chemical characterization data.

Chemical terminology

Specific chemical terminology is needed to understand the new InChI standard specifications and their purpose, which we define below. These definitions draw heavily from definitions provided by the IUPAC Gold Book [15, 16] and Wikipedia. "*Structural isomer*" or "*constitutional isomerism*" is one of a set of isomeric molecules with the same molecular formula but with different bonding patterns and atomic organization (i.e. line formulae) [17, 18]. "*Stereoisomer*" is one of a set of isomeric molecules that have the same molecular formula and sequence of bonded atoms (identical constitution) but differ in the orientation of their atoms in three-dimensional space [19, 20]. "*(Extended) isotopomer*", also known as an isotopic isomer, is one of a set of isomeric molecules with the same number of each isotope of each element, but differing in their positions within the chemical structure. For example, CH_3CHDCH_3 (SMILES: CC([2H]) C) and $CH_3CH_2CH_2D$ (SMILES: CCC[2H]) are a pair of "*constitutional isotopomers*" of propane [21, 22]. Isotopomers can also be specific to stereochemistry, where they are known as "*isotopic stereoisomers*" (e.g. (*R*)- and (*S*)-CH_3CHDOH) [21, 22]. By definition, isotopomers have the same mass, since they have the same isotopic composition. However, the set of all isotopomers, which

span all possible isotopic compositions, are often collectively referred to as isotopomers (Fig. 2). While technically inaccurate, it is a pragmatic use of the term to refer to all possible isotopomers spanning all isotopic compositions for a given chemical entity. Therefore, we will refer to "*isotopomer*" as any isotope-position-distinct isomer of a given chemical entity. The term "*mass-equivalent isotopomer*" will be used to refer to the technical definition of isotopomer.

An "*isotopologue*" represents molecules of the same chemical entity and same isotopic composition. As shown in Fig. 2, related isotopologues represent the same chemical entity but differ in their isotopic composition (number of isotopic substitutions) [23, 24]. For example, the isotopologues of water that span the stable isotope compositions of hydrogen and oxygen are: $^1H_2^{16}O$, $^1H^2H^{16}O$, $^2H_2^{16}O$, $^1H_2^{17}O$, $^1H^2H^{17}O$, $^2H_2^{17}O$, $^1H_2^{18}O$, $^1H^2H^{18}O$, and $^2H_2^{18}O$. A single isotopologue represents a set of mass-equivalent isotopomers (e.g. $^{12}C_2^1H_6$ has one isotopomer; $^{12}C_2^1H_5^2H$ has two isotopomers). A "*nominal-mass isotopologue*" is a set of isotopologues with the same elemental and nucleon compositions. In many mass spectra, the detected spectral features are not isotopically-resolved but are elementally and nucleon-resolved, thus representing what we will term a nominal-mass isotopologue.

We now introduce two new chemical terms not previously defined to our knowledge before developing this extension to the InChI standard. "*Partial isotopomer*" is the part of an isotopomer where the isotopic content of specific atoms is known. In nuclear magnetic resonance spectroscopy (NMR) and in rare circumstances in tandem mass spectrometry, individual spectral features may directly indicate that certain atoms of a particular molecule have certain isotopes (see Fig. 3). Taking an example from NMR, certain analytical features may indicate that
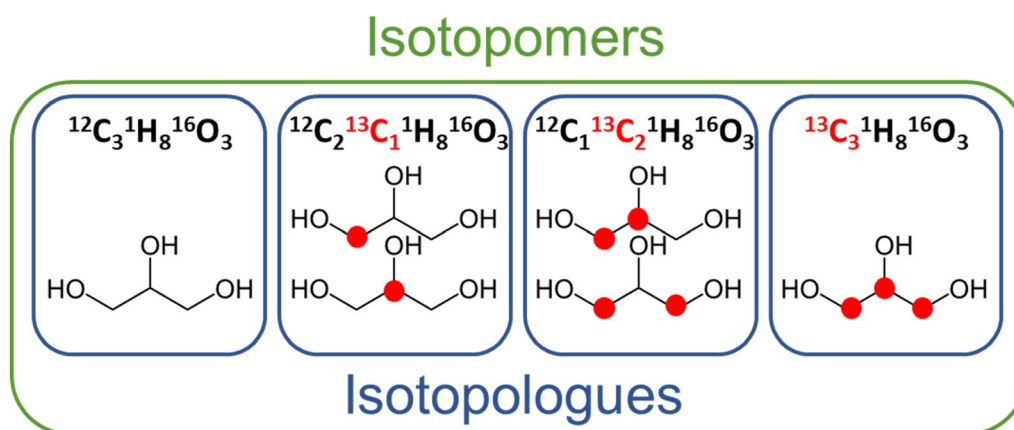

Fig. 2 Representation of ^{13}C isotopomers and isotopologues in glycerol. The red dot represents the location of ^{13}C within the glycerol structure

Moseley *et al. Journal of Cheminformatics*   (2024) 16:54
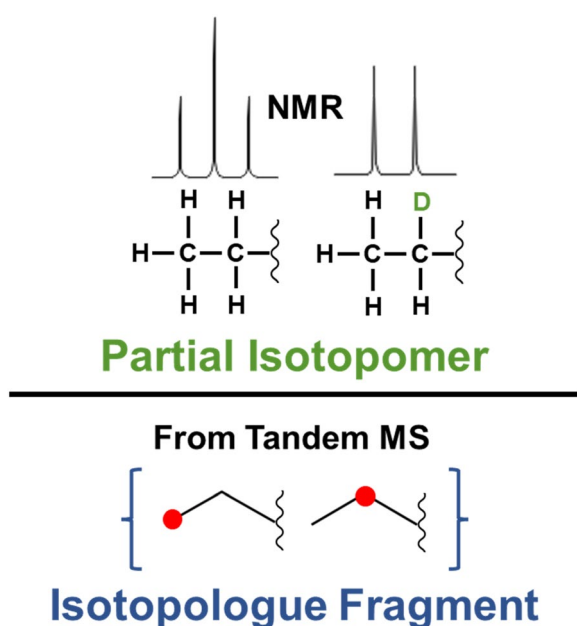
Page 4 of 9

Fig. 3 Representation of a partial isotopomer and isotopologue fragment. In NMR, spin–spin coupling (J-coupling or scalar coupling) can produce peaks that indicate the isotopic status of neighboring atoms representing a partial isotopomer. In the top example, the difference between a triplet vs doublet peak pattern indicates a partial isotopomer of H vs ^{2}H (D) bonded to the second carbon. In the bottom example, a peak in a tandem MS spectrum represents a single ^{13}C in the ethyl fragment of a larger compound

carbon 1 of alpha-D-glucopyranose is ^{13}C labeled without knowing the isotopic status of the other carbon atoms. "*Isotopologue fragment*" is a refined set of isotopomers where the ambiguity of isotope location is limited to a subset of the atoms. This is a subset of the mass-equivalent isotopomers representing the full isotopologue. Tandem mass spectrometry may provide spectral features that indicate where (possibly mixed) isotopes are localized within a chemical structure (see Fig. 3). For example, ^{13}C$_1$ may be limited to the glycerol part of a glycerophospholipid and not the fatty acyl chains.

InChI isotopic layer

The InChI isotopic layer was designed to represent the exact isotopomers of specific compounds. As described in Fig. 4, the isotopic layer starts with a "/i" followed by the atom number based on InChI canonicalization and then the isotope designation and hydrogen isotope designation, if present. This simplifies it to the atom number and hydrogen isotope designation if no isotope designation is needed. Previously, the InChI isotopic layer could only specify exact isotopomers, with limited ability to describe atomic location ambiguity for hydrogen

isotopes. Also, the previous interpretation of the isotopic layer for exact isotopomers was minimalistic, assuming that unspecified isotopes were the most abundant stable isotope.

The new interpretation of the InChI isotopic layer enables both exact isotopomers and partial isotopomers to be specified. Figure 5 illustrates examples of both exact and partial isotopomer interpretations of the InChI isotopic layer. Exact isotopomers require explicit definition of the isotope for each atom of an element (Fig. 5B(1)). In contrast, a partial isotopomer only requires the isotope specification for a subset of atoms for a given element (Fig. 5B(2, 3)). Furthermore, exact and partial isotopomers can be described in a stereospecific configuration (Fig. 5B(4)). However, as shown in Fig. 5B(5), exact and partial isotopomers with incomplete stereospecific configurations and involving deuterium isotopes can technically represent ambiguity in the location of the deuterium isotope, i.e., an isotopologue or isotopologue fragment. The previous interpretation of the isotopic layer would have interpreted partial isotopomers as exact isotopomers, thus preventing the representation of partial isotopomers. Since the isotopic layer specification remains unchanged, the new interpretation does not require any modifications to the underlying InChI algorithms or software and, thus, is currently implemented. Therefore, partial isotopomers can be specified using the InChI standard, especially for annotating most individual NMR spectroscopy spectral features. Moreover, the final example shown in Fig. 5B(6) demonstrates a partial isotopomer that would arise in H/D exchange experiments, when labile hydrogens are replaced with deuterium isotopes. This example illustrates the varied applications of partial isotopomer annotation in experiments utilizing stable isotopes, made possible by the refined specifications.

InChI isotopologue layer

To represent isotopologues, a new ambiguous isotope localization "/a" layer specification has been added. The isotopologue layer starts with a "/a" followed by an encoded description within parentheses. However, the specific encoding depends on the type of isotopologue or isotopologue fragment being specified: (a) isotopically-resolved structural isomer, (b) nominal-mass-resolved structural isomer, or (c) with only molecular formula resolution, i.e., unknown or ambiguous structural isomers.

For isotopically-resolved isotopologues of a constitutional isomer, the encoded description within the parentheses starts with the element code followed by the isotope count or number of atoms with the designated isotope, and then ending with the isotope designation, which is indicated by a sign (+ or −) and number

Moseley *et al. Journal of Cheminformatics*   (2024) 16:54

Page 5 of 9

Simple Definition: `/i#<+|->#[I#],..`

Complete Definition:

`/i<atom_number><isotope_designation>[<hydrogen_isotope><hydrogen_isotope_count>]`

`/i<atom_number>[<hydrogen_isotope><hydrogen_isotope_count>]`

Where:

`<atom_number>` - specific atom that has the specified isotope.

`<isotope_designation>` - isotope designation indicated by a sign (+ or -) and number indicating the unit mass difference from the rounded average atomic mass of the element. For example, the average atomic mass of Sn (118.710) is rounded to 119 and the most abundant isotope ^{120}Sn is specified by "+1".

`<hydrogen_isotope>` - D for deuterium. etc.

`<hydrogen_isotope_count>` - number of the specified hydrogen isotope bonded to the designated atom (<atom_number>); 1 if omitted.

Example:

Isotopomer of acetate containing one ^{13}C and two ^{2}H isotopes in the methyl group:
InChI=1S/C2H4O2/c1-2(3)4/h1H3,(H,3,4)/p-1/i1+1D2,2+0
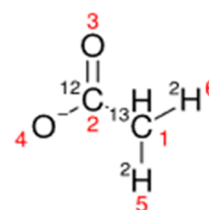
Fig. 4 Concise description of the InChI isotopic layer specification. Includes a simple definition, a complete definition with the description of terms, and an example. The example includes the line formula (2D representation) and InChI string of acetate containing one ^{13}C and two ^{2}H isotopes in the methyl group. The isotopic layer is highlighted in red

A
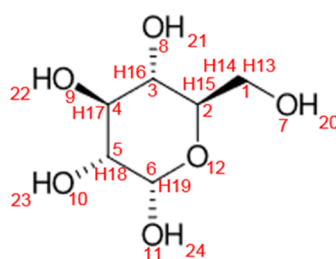
alpha-D-glucopyranose
InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1

1. Exact isotopomer of alpha-D-glucopyranose with ^{13}C at the 4th carbon:
   InChI=1/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i1+0,2+0,3+0,4+1,5+0,6+0   **B**

2. Partial isotopomer of alpha-D-glucopyranose with ^{13}C at the 4th carbon:
   InChI=1/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i4+1

3. Partial isotopomer of alpha-D-glucopyranose with ^{13}C at carbon 1 with bonded $^{2}H_2$:
   InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i1+1D2

4. Partial isotopomer of alpha-D-glucopyranose with $^{2}H_1$ bonded to carbon 1 in a stereospecific configuration:
   InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i1D/t1-,2-,3-,4+,5-,6+

5. Partial isotopomer of alpha-D-glucopyranose with $^{2}H_1$ bonded to carbon 1 with incomplete stereospecific configuration:
   InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i1D/t1?,2-,3-,4+,5-,6+

6. Partial isotopomer of alpha-D-glucopyranose with $^{2}H_1$ bonded to all heteroatoms:
   InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/i7D,8D,9D,10D,11D

Fig. 5 Examples of exact and partial isotopomer InChI strings. **A** Line formula (2D representation) of alpha-D-glucopyranose along with its InChI string. **B** Specific exact and partial isotopomer examples of alpha-D-glucopyranose, with the isotopic layer highlighted in red

Moseley *et al. Journal of Cheminformatics*   (2024) 16:54

Page 6 of 9

indicating the unit mass difference from the rounded average atomic mass of the element. Isotopically-resolved isotopologue fragments of a constitutional isomer include an additional comma-separated list of atom numbers designating specific atoms where the designated isotope could be. The encoding specification is fully described in Fig. 6.

Figure 7 illustrates examples of both isotopologues and isotopologue fragments. Specifically, Fig. 7B(1–3) illustrate ^{13}C, ^2H, and/or $^{17}O/^{18}O$ isotopologues of alpha-D-glucopyranose, whereas Fig. 7B(4–6) illustrate ^{13}C, ^2H, and/or ^{15}N isotopologue fragments. However, pay attention to Fig. 7B(6) where an isotopic layer must be used to specify the $^{15}N_1$ at atom 7 instead of being part

Fig. 6 Concise description of the InChI isotopically-resolved isotopologue layer specification of a constitutional isomer. Includes a simple definition and a complete definition with the description of the terms
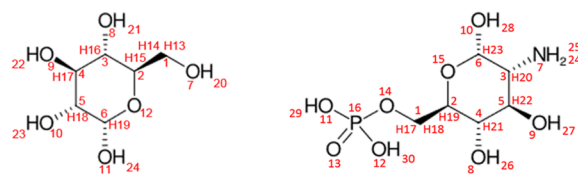
Fig. 7 Examples of isotopically-resolved isotopologues and isotopologue fragments. **A** Left is the line formula (2D representation) of alpha-D-glucopyranose along with its InChI string below. Right is the line formula of 2-amino-2-deoxy-6-*O*-phosphono-alpha-D-glucopyranose with its InChI string below. **B** Specific examples of isotopically-resolved isotopologues and isotopologue fragments of these two compounds, with the isotopologue layer highlighted in red

Moseley *et al. Journal of Cheminformatics*   (2024) 16:54

Page 7 of 9

Simple Definition: `/a(#n[,#]...)`

Complete Definition:

`/a(<neutron_count>n[,<atom_number>])`

Where:

`<neutron_count>` - number of neutrons of extra mass.

Notice the reversal of neutron_count relative to the "n" representing neutron. This is done so that no ambiguity will exist with the isotopically-resolved isotopologue specification, where the element is indicated first and the isotope_count second.

`<atom_number>` - specific atom(s) that may have the specified isotope. If empty, then all atoms of the given element may have the specified isotope (# in [,#]). If multiple atoms need to be specified, these are comma separated.

Examples:

1. M+3 isotopologue of alpha-D-glucopyranose:

   InChI=1/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/a(3n)

   The lack of a specified list of atoms indicates the whole molecule.

2. M+4 isotopologue fragment limited to carbon atoms of alpha-D-glucopyranose:

   InChI=1/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6+/m1/s1/a(4n,1,2,3,4,5,6)

Fig. 8 Concise description of the InChI nominal-mass-resolved isotopologue layer specification of a constitutional isomer. Includes a simple definition, a complete definition with the description of the terms, and examples. The example includes the line formula (2D representation) and InChI string of alpha-D-glucopyranose with 3 or 4 extra neutrons. The isotopic layer is highlighted in red

of the isotopologue layer, since the molecule has only one nitrogen atom. This is an important boundary condition between using the isotopic vs isotopologue layer for specifying the isotope content. The "/a" isotopologue layer is only used when the location of the isotope is "ambiguous". Besides the layer designation "/a" vs "/i", the use of parentheses in the isotopologue layer also makes it visually distinct from the isotopic layer.

For the nominal-mass-resolved isotopologues of a constitutional isomer, the description within the parentheses

---

Simple Definition:

`1/Ee#[Ee#].../a(Ee#<+|->#)`

        OR

`1/Ee#[Ee#].../a(#n)`

Complete Definition:

`/a(<element><isotope_count><isotope_designation>[,<atom_number>])`

        OR

`/a(<neutron_count>n[,<atom_number>])`

Examples:

1. $^{13}C_2$ isotopologue of molecules with molecular formula $C_6H_{12}O_6$:

   InChI=1/C6H12O6/a(C2+1)

2. $^{13}C_1{}^{18}O_1$ isotopologue of molecules with molecular formula $C_6H_{12}O_6$:

   InChI=1/C6H12O6/a(C2+1),(O1+2)

3. M+3 isotopologue of molecules with molecular formula $C_6H_{12}O_6$:

   InChI=1/C6H12O6/a(3n)

   This last example addresses a use case in mass spectrometry spectral feature assignment.

Fig. 9 Concise description of the InChI isotopologue layer specification of an unknown constitutional isomer. Includes a simple definition, a complete definition with the description of the terms, and examples. The first two examples involve the molecular formula $C_6H_{12}O_6$ with ^{13}C and ^{18}O isotopes. The last example involves the molecular formula $C_6H_{12}O_6$ with 3 extra neutrons. The isotopic layer is highlighted in red

Moseley *et al. Journal of Cheminformatics*   (2024) 16:54

Page 8 of 9

begins with the neutron count followed by "n," as shown in Fig. 8. Nominal-mass-resolved isotopologue fragments of a constitutional isomer include an additional comma-separated list of atom numbers designating specific atoms where a specific neutron could be located. Example 1 in Fig. 8 illustrates a nominal-mass isotopologue with 3 extra neutrons beyond what is expected for the most common stable isotopes present in alpha-D-glucopyranose. Example 2 in Fig. 8 illustrates an isotopologue fragment with 4 extra neutrons across the carbon atoms of alpha-D-glucopyranose.

For nominal-mass-resolved isotopologues where the constitutional isomer is unknown (or highly ambiguous), a truncated InChI string with only the molecular formula and the isotopologue layer is used with either an isotopically-resolved or nominal-mass encoded description. Examples 1 and 2 in Fig. 9 illustrate ^{13}C and [18]O containing isotope-resolved isotopologues of the molecular formula $C_6H_{12}O_6$, representing 6-carbon saccharides. Example 3 in Fig. 9 illustrates a nominal-mass-resolved isotopologue of the molecular formula $C_6H_{12}O_6$ containing 3 extra neutrons above what is expected for the most common stable isotope of the elements present. Due to the lack of chemical (sub)structure with atom identifiers, isotopologue fragments of an unknown constitutional isomer cannot be defined.

Conclusions

This article details a proposed extension to the InChI standard that allows the representation of isotopically-resolved chemical entities at varying levels of ambiguity in isotope location. This extension includes improved interpretation of the current isotopic "/i" layer for representing exact isotopomers and partial isotopomers. In addition, the extension includes an isotopologue "/a" layer specification for representing chemical entities with ambiguous isotope localization. This new isotopologue "/a" layer enables the representation of isotopologues and isotopologue fragments. Both improvements support the unique identification of features detected and measured in analytical instrumentation, enabling their interpretation as isotopically-resolved chemical entities. Thus, this new extension would enable improved annotation of analytical datasets characterizing chemical entities, and would better support FAIR data sharing and Open Science once extension-aware programs are available.

Abbreviations
IUPAC   International Union of Pure and Applied Chemistry
InChI   IUPAC International Chemical Identifier
API   Application programming interface
SMILES   Simplified Molecular Input Line Entry System
FAIR   Findable, Accessible, Interoperable, and Reusable

Declarations

Competing interests
The authors have no competing interests.

References
1. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I (2013) InChI-the worldwide chemical structure identifier standard. J Cheminform 5(1):7
2. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D (2015) InChI, the IUPAC international chemical identifier. J Cheminform 7(1):1–34
3. Goodman JM, Pletnev I, Thiessen P, Bolton E, Heller SR (2021) InChI version 1.06: now more than 99.99% reliable. J Cheminform 13(1):1–8
4. McKay BD (1978) Computing automorphisms and canonical labellings of graphs. Combinatorial mathematics. Springer, New York, pp 223–232
5. McKay BD, Piperno A (2014) Practical graph isomorphism, II. J Symb Comput 60:94–112
6. Heller S, McNaught A (2010) The status of the InChI project and the InChI trust. J Cheminform 2(1):1–1
7. IUPAC/InChI-trust licence for the international chemical identifier (InChI) software. https://www.inchi-trust.org/wp/download/106/LICENCE.pdf
8. Grethe G, Blanke G, Kraut H, Goodman JM (2018) International chemical identifier for reactions (RInChI). J Cheminform 10(1):1–9
9. Clark AM, McEwen LR, Gedeck P, Bunin BA (2019) Capturing mixture composition: an open machine-readable format for representing mixed substances. J Cheminform 11(1):1–17
10. Lai A, Clark AM, Escher BI, Fernandez M, McEwen LR, Tian Z, Wang Z, Schymanski EL (2022) The next frontier of environmental unknowns: substances of unknown or variable composition, complex reaction products, or biological materials (UVCBs). Environ Sci Technol 56(12):7448–7466
11. Lynch I, Afantitis A, Exner T, Himly M, Lobaskin V, Doganis P, Maier D, Sanabria N, Papadiamantis AG, Rybinska-Fryca A (2020) Can an InChI for nano address the need for a simplified representation of complex nanomaterials across experimental and nanoinformatics studies? Nanomaterials 10(12):2493
12. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE (2016) The FAIR

guiding principles for scientific data management and stewardship. Sci Data 3:160018

13. National Academies of Sciences E, Medicine (2018) Open science by design: realizing a vision for 21st century research

14. Vicente-Saez R, Martinez-Fuentes C (2018) Open science now: a systematic literature review for an integrated definition. J Bus Res 88:428–436

15. Chalk S, McEwen L (2017) The IUPAC gold book. Chem Int 39(3):25–30

16. Chalk SJ (2019) The IUPAC gold book website. https://goldbook.iupac.org/. Accessed 01 Feb 2018

17. Constitutional isomerism definition in the IUPAC gold book. https://doi.org/10.1351/goldbook.C01285. Accessed 01 Feb 2018

18. Structural isomer entry in Wikipedia. https://en.wikipedia.org/wiki/Structural_isomer. Accessed 01 Feb 2018

19. Stereoisomers definition in the IUPAC gold book. https://doi.org/10.1351/goldbook.S05984. Accessed 01 Feb 2018

20. Stereoisomerism entry in Wikipedia. https://en.wikipedia.org/wiki/Stereoisomerism. Accessed 01 Feb 2018

21. Isotopomer definition in the IUPAC gold book. https://doi.org/10.1351/goldbook.I03352. Accessed 01 Feb 2018

22. Isotopomers definition in Wikipedia. https://goldbook.iupac.org/. Accessed 01 Feb 2018

23. Isotopologue definition in the IUPAC gold book. https://doi.org/10.1351/goldbook.I03351. Accessed 01 Feb 2018

24. Isotopologue entry in Wikipedia. https://goldbook.iupac.org/. Accessed 01 Feb 2018

Publisher's Note

Hunter N. B. Moseley   Professor in the Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY USA; Bioinformatics and Data Analysis Section Editor-in-Chief of *Metabolites*; Associate Director, Institute for Biomedical Informatics, University of Kentucky, Lexington, KY USA.

Philippe Rocca-Serra   Associate Member of Faculty, senior researcher at the Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, UK.

Reza M. Salek   Consulting Director, BiotechVision, Cambridge, UK.

Masanori Arita   Professor and Head of DNA Data Bank of Japan, National Institute of Genetics, Japan; Team Leader of Metabolome Informatics, RIKEN Center for Sustainable Resource Science.

Emma L. Schymanski   Full Professor and Head of the Environmental Cheminformatics (ECI) group at the Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg; Special Advisor to the Rectorate for Open Science and Research Data Management.