[InChI Design Details 1](#_Toc33460816)

[is there a way to get the canonical heavy atom numbers? 1](#_Toc33460817)

[Re: [InChI-discuss] original atoms question 5](#_Toc33460818)

# InChI Design Details

## is there a way to get the canonical heavy atom numbers?

<https://sourceforge.net/p/inchi/mailman/message/1620016/>

From: inchi-discuss-bounces@...

[mailto:inchi-discuss-bounces@...] On Behalf Of D.

Tchekhovskoi

Sent: Monday, August 07, 2006 11:14 AM

To: Discussions related to the IUPAC International Chemical Identifier

(InChI)

Subject: Re: [InChI-discuss] If I'm using InChI as a library/dll through

the API, is there a way to get the canonical heavy atom numbers?

=20

George,

InChI AuxInfo string provides one or several mappings of canonical

numbers on the input structure atom numbers except terminal hydrogen

atoms. The input structure atom numbering, as well as canonical

numbering for each component, start from 1.

Specifically, each mapping is a list of input structure atom numbers in

order of their canonical numbers.

For example,

/N:3,1,5;2,3,7

means atom 3 (numbering starts from 1) in the input structure has

canonical number 1, atom 1 in the input structure has canonical number

2, atom 5 in the input structure has canonical number 3. Semicolon ";"

separates components (parts of the structure that are not interconnected

by bonds). Atom 2 in the input structure belongs to the 2nd component

where it has canonical number 1; atom 3 has canonical number 2 in the

2nd component, and atom 7 has canonical number 3 in the 2nd component.

Atoms 4 and 6 are not present because they are terminal hydrogen atoms,

each being connected to (one of) these components. No conclusion may be

made about atom 8 in the input structure unless you look at /rA: segment

of AuxInfo; the first number there is the number of input atoms. If and

only if this number is 8 or greater then atom 8 is present in the input

structure and it is a terminal H, including isotopes D or T.

Note: In case of symmetrical structures this mapping may be ambiguous.

However, AuxInfo also provides lists of constitutionally identical atoms

(that is, stereo is not taken into account): AuxInfo segment /E:

contains sets of canonical numbers of equivalent atoms; each set is in

parentheses; components are separated by semicolons. Stereochemistry may

break this equivalence; this fact is not reflected in AuxInfo.

You need to understand how to parse AuxInfo. Please take a look at the

section "Auxiliary Information Output" in Chapter VI, "OUTPUT TEXT

FORMAT", of the InChI User's Guide. It may be instructive to compare it

to the "InChI Output" section in the same chapter. To properly implement

AuxInfo parsing you also need to read section "b. Abbreviations" in the

"Appendix 2. Abbreviations and Layer Precedence" of the InChI Technical

Manual.

AuxInfo, in general, is made out of one or two parts. The two parts may

be understood as two separate AuxInfo strings, each of them may include

up to all possible segments (except /rA:, /rB:, and /rC: that are only

in the first part) The two parts may appear only in case of /Recmet or

"include bonds to metal atoms" option. They correspond to the two parts

of InChI. To see this correspondence please compare

InChI=3D1/<InChI part 1>/r<InChI part 2>

to

AuxInfo=3D1/<AuxInfo part 1>/R:<AuxInfo part 2>

Some of InChI layer delimiting prefixes unambiguously correspond to some

of AuxInfo prefixes:

InChI AuxInfo

/r /R:

/f /F:

/i /I:

For the purpose of this discussion you may ignore the following AuxInfo

prefixes and what follows them until the next slash or the end of the

AuxInfo string, whichever comes first:

/CRV:

/rA:<atoms>

/rB:<bonds>

/rC:<coordinates>

/it:<sp3 stereo of the alternatively inverted structure>

/gE:<mobile H groups equivalence>

The prefixes to recognize the lists of input atom numbers in canonical

order may be found in the previously cited InChI User's Guide section

"Auxiliary Information Output"; they are:=20

/N: - main layer

/iN: - main layer, inverted stereo

/I: - main layer, isotopic

/I:.../iN: - main layer, isotopic, inverted stereo

/F: - fixed H layer

/F:.../iN: - fixed H layer, inverted stereo

/F:.../I: - fixed H layer, isotopic

/F:.../I:.../iN: - fixed H layer, isotopic, inverted stereo

Here each "..." mean omitted segments or empty strings (nothing). There

is only one or no /F: in a single AuxInfo part. There is only one or no

/I: both on the left and on the right side of /F: in a single AuxInfo

part. There is only one or no /iN: between each pair:

/N: - /I:; /I: - /F:; /F: - /I:; /I: - end of a single AuxInfo part.

Please keep in mind that /F: and/or any of /I: may be missing, as well

as the whole AuxInfo part that starts with /R:

Components in AuxInfo are separated by semicolons; their order is same

as in InChI except when /o segment is present in InChI. In the latter

case /o contains permutation that maps components in the last fixed-H

layer of InChI on the components in the last Fixed-H layer of AuxInfo.

See Figure A3-3 in the InChI Technical Manual and related to it text for

the explanation of /o segment.

Please notice that the stereochemical sp3 (/t) layer of InChI actually

describes inverted component(s) that is/are the enantiomer(s) of the

input component(s) if there is a "1" in the /m segment of InChI. In this

case /it: and /iN: in the AuxInfo describe the stereochemistry and

numbering of the input (uninverted) component =20

If there is a "0" in /m segment then InChI contains the stereochemistry

of the uninverted (input) component and /it: contains the

stereochemistry of the inverted component.

In general, if the spatial inversion changes the stereochemistry then

the alternative sp3 stereo segment is preceded by /it: in the AuxInfo;

the corresponding numbering follows /iN:.

Regards

Dmitrii

At 09:18 AM 8/4/2006, you wrote:

Content-class: urn:content-classes:message

Content-Type: multipart/alternative;

boundary=3D"=3D\_aebef7fe9c17146da9082db014bfe10a"

Hi,

=20

I just started using the InChI library in our tools as a way to compare

structures to see if they are the same compound. It works great. I had

a question I wanted to be able to, given a bunch of structures that have

identical INCHI=3D& strings, know that heavy (non-H) atoms are the same

atom across all the structures.

=20

I don t see this functionality exposed in the API, if I were to go

looking for a place to pick up this information, where would I find it?

I would like to expose this through the public API somehow, and would

gladly contribute any changes that I make back to the project.

=20

Thanks for your time.

=20

--

George T. Talbot

Senior Scientific Software Engineer

Locus Pharmaceuticals, Inc.

gtalbot@...

=20

------------------------------------------------------------------------

-

Take Surveys. Earn Cash. Influence the Future of IT

Join SourceForge.net's Techsay panel and you'll get the chance to share

your

opinions on IT & business topics through brief surveys -- and earn cash

<http://www.techsay.com/default.php?page=3Djoin.php&p=3Dsourceforge&CID=3D=>

DEVDE

V\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

InChI-discuss mailing list

InChI-discuss@...

<https://lists.sourceforge.net/lists/listinfo/inchi-discuss>

## Re: [InChI-discuss] original atoms question

<https://sourceforge.net/p/inchi/mailman/message/28292914/>

[**Re: [InChI-discuss] original atoms question**](https://sourceforge.net/p/inchi/mailman/message/28292914/)

From: Igor Pletnev <igor.pletnev@gm...> - 2011-10-28 08:36:11

**Attachments:** [Message as HTML](https://sourceforge.net/p/inchi/mailman/attachment/CA%2Bpv_NR1PYgdgjN7EuE%2BpiEEKwEt%2BWgAjgFdqKh4KRdV3dfpHA%40mail.gmail.com/1/)

Mark,

you found the bug in AuxInfo output for isotopically substituted

molecule(s).

Indeed, in your example mapping of canonical numbers to original ones does

change upon introducing isotopic information.

The generated InChI string itself is correct; however, in the associated

AuxInfo "isotopic mapping" is, unfortunately, represented incorrectly.

In particular, "isotopic mapping" part of AuxInfo is encoded as "/I:m" -

which means that it repeats "non-isotopic" one (i.e., "7,8,2,6,3,5,1,4" ).

This is incorrect. The correct AuxInfo should contain changed mapping info,

namely, "/I:8,7,3,5,2,6,4,1".

We already have a preliminary fix for this bug. Currently, it is under

testing; we are also examining the related issues.

Thanks for reporting this issue.

Regards,

Igor

On Tue, Oct 25, 2011 at 1:28 PM, Sandor Mark <sanmark@...> wrote:

> I would like to assign properties to original atoms (atoms in the input MOL

> structure) based on the inchi and the auxinfo string to track structure

> changes (hydrogen addition, parity loss

> etc.) made by inchi.

>

> What I need would be the original atom numbers in the auxinfo string. But

> not, as it is shown in the attached example where position of the isotopic

> carbon is lost.

>

> Here is the inchi and the auxinfo generated:

>

> InChI=1S/C8H16/c1-7-3-5-8(2)6-4-7/h7-8H,3-6H2,1-2H3/t7-,8?/i7+2/t7?,8-

>

> AuxInfo=1/0/N:7,8,2,6,3,5,1,4/E:(1,2)(3,4,5,6)(7,8)/I:m/E:(3,4)(5,6)/rA:8nCCCC.i14CCCC/rB:s1;s2;s3;s4;s1s5;N1;s4;/rC:-.3536,-1.2964,0;-1.068,-1.7089,0;-1.068,-2.5339,0;-.3536,-2.9464,0;.3609,-2.5339,0;.3609,-1.7089,0;-.3536,-.4714,0;-.3536,-3.7714,0;

>

> Introducing isotopic atoms, canonical numbers change between

> formerly equivalent carbon atoms 7 and 8. In the auxinfo string, however,

> there is no change in the 'I:' layer. This way C-14 atom is mapped to the

> non-isotopic original atom 1.

>

> Is this the intended behaviour, or it is a bug?

>

> As I see additional information from the input structure is printed to the

> /rA, /rB, /rC part of the auxinfo string. Perhaps /rA and /rB contain the

> information I need. If so, how can I interpret them?

>

> Regards,

> Mark

>

>

> ------------------------------------------------------------------------------

> The demand for IT networking professionals continues to grow, and the

> demand for specialized networking skills is growing even more rapidly.

> Take a complimentary Learning@... Self-Assessment and learn

> about Cisco certifications, training, and career opportunities.

> <http://p.sf.net/sfu/cisco-dev2dev>

> \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

> InChI-discuss mailing list

> InChI-discuss@...

> <https://lists.sourceforge.net/lists/listinfo/inchi-discuss>

>

>