How InChI Can Support the EPO

Stephen Heller
InChI-Trust Project Director
steve@inchi-trust.org

The main web sites for the IUPAC InChI project are:
http://www.iupac.org/inchi
and
http://www.inchi-trust.org

5/6/2011

Outline

1. Background/History/Objective/Why InChI?
2. InChI and the EPO
3. InChI Technical Details and Examples
4. InChI Trust
5. Current and Future InChI activities
6. Acknowledgements

Objective

The objective of the IUPAC Chemical Identifier Project is to create a unique label, the IUPAC Chemical Identifier (InChI), which will be an Open Source, freely available, non-proprietary identifier for well defined chemical substances that can be used in printed and electronic data sources thus enabling easier LINKING of and working with diverse data and information compilations.

Why InChI? - Too Many Identifiers

Structure diagrams
- various conventions
- contain 'too much' information

Connection Tables
- MolFiles, Smiles, ROSDAL, …

Pronounceable names
- IUPAC, CAS 8th CI name, CAS 9th CI name, trivial, WHO INN

Index Numbers
- EINECS, FEMA, DOT, RTECS, CAS, Beilstein, USP, RTECS, EEC, RCRA, NCI, UN, USAF , EC

CML and InChI

CML (Chemical Markup Language) is an approach to managing molecular information using tools such as XML and Java.

CML is capable of supporting a wide range of chemical concepts including:

MOLECULES
reactions
spectra and analytical data
computational chemistry
chemical crystallography and materials

InChI fits within CML framework and structure and is the way molecules are described.

Why Use InChI ?

For publishers, database providers, and organizations (such as the EPO) with one or more databases and with customers and stakeholders needing to access this information, using InChI gives one an advantage being able to LINK content from multiple sources. It offers the EPO and users the ability to more easily find existing information and data by easily being able to integrate, remix, and retell. InChI is a small, but vital, part of new organization models and technologies involving chemicals that will lead to improved efficiencies new discoveries. Combinability increases the value of information and data.

Critical factors for the success of InChI project

1. Technically competent staff
2. Fulfill a real community need
3. Political and Financial Support

Technical: **InChI is a unique representation/identifier for defined chemical structures. Probably marginally better than previous ones. The InChI algorithm was built on the shoulders of giants, starting with Euler in 1736.**
http://en.wikipedia.org/wiki/Graph_theory

Practical: **InChI and the related hash-code compressed InChIKey are the only available universal LINKs for in-house and public databases of defined chemical structures. The adoption and use of InChI by the vast majority of publishers and database providers assure it will be widely used.**

Why InChI is becoming a success

1. Organizations need a structure representation for their content (databases, journals, patents, chemicals for sale, products, and so on) so that their content can be LINKED to and combined with other content on the Internet.

2. InChI is a public domain algorithm that anyone, anywhere can freely use. The other major representations are proprietary and hence not affordable for the world-wide community.

3. InChI is not a replacement for any internal structure representations. InChI is IN ADDITION to what one uses internally. Its value to most organizations is in LINKING information.

How do we know the InChI project is beneficial?

Success is uncoerced adoption

InChI – What is missing?

While we believe InChI covers some 99% of the chemicals found in computer readable databases, there are areas of chemistry not yet covered by the InChI algorithm. Some are currently being addressed, while others of lesser importance will be addressed in the next few years. But these gaps have not impeded the widespread adoption and support of InChI.

One main gap is the area of generic or Markush structures as has been pointed out by a number of people.

InChI have some advantages over other chemical identifiers developed before:

(1) They are freely useable and non-proprietary.

(2) They allow a more advanced representation of chemical information than other codes (such as the SMILES code).

(3) They are unambiguous, i.e. conversion of chemical structures using standardized algorithms only leads to one InChI. (see next slides)

(4) They are precisely indexed by major search engines such as Google.

However, InChI are not applicable to generic formats often disclosed in patent literature, such as Markush structures, since they were rather designed to represent specific chemical structures and compounds. InChI therefore are not yet useful for comprehensive retrieval of patent literature.

Excerpt taken from:
Full-text prior art and chemical structure searching in e-journals and on the internet – A patent information professional's perspective
World Patent Information, Volume 31, Issue 4, December 2009, Pages 278-284
Maik Annies (Syngenta)

Re: [CHMINF-L] Inchi and chemical databases

You forwarded this message on 9/15/2010 5:37 PM.

CHEMICAL INFORMATION SOURCES DISCUSSION LIST [CHMINF-L@LISTSERV.INDIANA.EDU] on behalf of Ian A Watson

Sent: Wednesday, September 15, 2010 3:24 PM

To: CHMINF-L@LISTSERV.INDIANA.EDU

Interesting example of Caffeine smiles on the web site. I was able to generate 172 different smiles for the Caffeine molecule (email me if you'd like them). Presumably each one of these could be a unique smiles in somebody's implementation.

But when I converted each of those 172 different smiles to InChI, I got the exact same InChI string for each one. That's exactly how things are supposed to work. Nice.

Ian Watson

```
c1(=O)c2c(n(C)c(=O)n1C)ncn2C
c12c(n(C)c(=O)n(C)c1=O)ncn2C
O=c1n(C)c(=O)c2c(ncn2C)n1C
Cn1c2c(nc1)n(C)c(=O)n(C)c2=O
c12c(ncn1C)n(C)c(=O)n(c2=O)C
O=c1c2c(ncn2C)n(c(=O)n1C)C
c12c(n(cn1)C)c(=O)n(C)c(=O)n2C
Cn1c2c(nc1)n(c(=O)n(c2=O)C)C
c12c(ncn1C)n(c(=O)n(C)c2=O)C
c12c(ncn1C)n(C)c(=O)n(C)c2=O
Cn1c(=O)n(C)c(=O)c2c1ncn2C
n1(c2c(nc1)n(C)c(=O)n(C)c2=O)C
c12c(n(C)cn1)c(=O)n(c(=O)n2C)C
Cn1c(=O)c2c(ncn2C)n(c1=O)C
n1cn(C)c2c1n(c(=O)n(c2=O)C)C
n1cn(c2c1n(C)c(=O)n(c2=O)C)C
c12c(c(=O)n(c(=O)n1C)C)n(C)cn2
c1nc2c(n1C)c(=O)n(C)c(=O)n2C
c1(=O)n(C)c(=O)c2c(ncn2C)n1C
O=c1n(c(=O)c2c(ncn2C)n1C)C
Cn1cnc2c1c(=O)n(C)c(=O)n2C
n1(c(=O)n(c(=O)c2c1ncn2C)C)C
c1(=O)n(C)c(=O)c2c(n1C)ncn2C
O=c1n(c2c(n(cn2)C)c(=O)n1C)C
Cn1c2c(n(cn2)C)c(=O)n(c1=O)C
Cn1c(=O)c2c(n(c1=O)C)ncn2C
Cn1cnc2c1c(=O)n(c(=O)n2C)C
c1nc2c(c(=O)n(C)c(=O)n2C)n1C
c12c(ncn1C)n(c(=O)n(c2=O)C)C
c1nc2c(n1C)c(=O)n(c(=O)n2C)C
Cn1c2c(n(cn2)C)c(=O)n(C)c1=O
n1(C)c2c(n(C)c(=O)n(c2=O)C)nc1
n1(C)c2c(nc1)n(C)c(=O)n(c2=O)C
n1(c(=O)c2c(n(c1=O)C)ncn2C)C
n1(c(=O)c2c(n(C)c1=O)ncn2C)C
Cn1c(=O)n(c2c(c1=O)n(C)cn2)C
n1(C)c(=O)n(C)c(=O)c2c1ncn2C
c1(=O)n(C)c(=O)c2c(ncn2C)n1C
n1(cnc2c1c(=O)n(c(=O)n2C)C)C
n1(C)c(=O)n(C)c2c(n(cn2)C)c1=O
n1(c2c(n(cn2)C)c(=O)n(C)cn2)C
n1(C)cnc2c1c(=O)n(C)c(=O)n2C
O=c1c2c(n(C)c(=O)n1C)ncn2C
n1(c2c(nc1)n(c(=O)n(c2=O)C)C)C
n1(C)c(=O)c2c(n(c1=O)C)ncn2C
n1(c2c(c(=O)n(c1=O)C)ncn2C
n1cn(C)c2c1n(C)c(=O)n(c2=O)C
c12c(c(=O)n(C)c(=O)n1C)n(cn2)C
Cn1c2c(n(C)cn2)c(=O)n(c1=O)C
n1(c(=O)n(C)c2c(n(cn2)C)c1=O)C
n1cn(c2c1n(C)c(=O)n(C)c2=O)C
c1(=O)n(c2c(c(=O)n1C)C)n(cn2)C
Cn1c(=O)n(c(=O)c2c1ncn2C)C
O=c1n(c(=O)c2c(n(C)cn2)C)C)C
n1(c2c(c(=O)n(c1=O)C)n(C)cn2)C
c12c(n(cn1)C)c(=O)n(c(=O)n2C)C
c12c(c(=O)n(C)c(=O)n1C)n(C)cn2
Cn1c(=O)c2c(n(C)c1=O)ncn2C
```

```
c1(=O)n(C)c2c(n(cn2)C)c(=O)n1C
O=c1n(C)c2c(c(=O)n1C)n(C)cn2
n1(C)c2c(c(=O)n(C)c1=O)n(C)cn2
n1cn(c2c1n(c(=O)n(C)c2=O)C)C
O=c1n(c(=O)n(C)c2c1n(cn2)C)C
c1(=O)c2c(n(c(=O)n1C)C)ncn2C
c1(=O)n(c2c(n(cn2)C)c(=O)n1C)C
Cn1c2c(c(=O)n(c1=O)C)n(cn2)C
c1(=O)n(c(=O)c2c(n1C)ncn2C)C
O=c1n(c(=O)c2c(n1C)ncn2C)C
n1cn(C)c2c1n(c(=O)n(C)c2=O)C
n1(c(=O)n(C)c2c(c1=O)n(C)cn2)C
O=c1c2c(ncn2C)n(C)c(=O)n1C
n1cnc2c1c(=O)n(C)c(=O)n2C)C
n1(C)cnc2c1c(=O)n(C)c(=O)n2C)C
n1cn(C)c2c1n(C)c(=O)n(C)c2=O
O=c1n(C)c(=O)n(C)c2c1n(C)cn2
n1(C)c(=O)n(c2c(c1=O)n(C)cn2)C
Cn1c(=O)c2c(ncn2C)n(C)c1=O
n1(c2c(n(cn2)C)c(=O)n(c1=O)C)C
n1(C)c2c(n(C)c(=O)n(c1=O)nc1
Cn1c2c(n(c(=O)n(c2=O)C)C)nc1
n1(c(=O)n(C)c(=O)c2c1ncn2C)C
O=c1n(C)c2c(n(C)cn2)c(=O)n1C
n1(C)c2c(n(cn2)C)c(=O)n(C)c1=O
c1(=O)c2c(ncn2C)n(c(=O)n1C)C
O=c1n(c2c(c(=O)n1C)n(cn2)C)C
Cn1c2c(n(C)c(=O)n(C)c2=O)nc1
Cn1c2c(nc1)n(c(=O)n(C)c2=O)C
Cn1c2c(n(C)cn2)c(=O)n(C)c1=O
c12c(n(C)c(=O)n(c1=O)C)ncn2C
n1(c2c(c(=O)n(c1=O)C)n(cn2)C)C
c1(=O)n(C)c(=O)n(c2c1n(cn2)C)C
n1(c2c(n(C)cn2)c(=O)n(c1=O)C)C
c1(=O)n(c2c(n(C)cn2)c(=O)n1C)C
n1(c2c(nc1)n(C)c(=O)n(c2=O)C)C
Cn1c2c(nc1)n(C)c(=O)n(c2=O)C
c12c(c(=O)n(c(=O)n1C)C)n(cn2)C
Cn1c2c(n(c(=O)n(C)c2=O)C)nc1
c1(=O)n(c(=O)n(C)c2c1n(C)cn2)C
c1(=O)n(C)c2c(n(C)cn2)c(=O)n1C
n1(c(=O)c2c(n(C)cn2)C)c(=O)n1C
n1(C)c2c(n(C)cn2)c(=O)n(c1=O)C
n1(C)c2c(nc1)n(C)c(=O)n(c2=O)C
n1(C)c2c(n(cn2)C)c(=O)n(c1=O)C
n1(C)c2c(c(=O)n(C)c1=O)n(cn2)C
n1(c(=O)n(c2c(n(C)cn2)c1=O)C)C
n1(c(=O)n(c2c(c1=O)n(cn2)C)C)C
n1(C)c2c(n(C)cn2)c(=O)n(C)c1=O
n1(C)c2c(c(=O)n(c1=O)C)n(cn2)C
n1(C)c2c(n(c(=O)n(C)c2=O)C)nc1
n1(C)c2c(nc1)n(c(=O)n(c2=O)C)C
```

```
O=c1c2c(n(c(=O)n1C)C)ncn2C
O=c1n(C)c2c(n(cn2)C)c(=O)n1C
n1(C)c(=O)n(c2c(n(C)cn2)c1=O)C
n1(C)c2c(c(=O)n(c1=O)C)n(cn2)C
Cn1c2c(c(=O)n(C)c1=O)n(C)cn2
c1(=O)n(c2c(c(=O)n1C)n(cn2)C)C
n1(c2c(n(C)c(=O)n(c2=O)C)nc1)C
n1(c2c(c(=O)n(C)c1=O)n(C)cn2)C
n1(C)c(=O)c2c(ncn2C)n(c1=O)C
Cn1c(=O)n(C)c2c(n(cn2)C)c1=O
O=c1n(C)c(=O)c2c(n1C)ncn2C
n1(c(=O)n(c2c(c1=O)n(cn2)C)C)C
O=c1n(c(=O)n(C)c2c1n(C)cn2)C
n1(C)c(=O)n(c2c(n(cn2)C)c1=O)C
n1(c(=O)n(C)c2c(n(C)cn2)c1=O)C
c1(=O)n(C)c(=O)n(C)c2c1n(cn2)C
n1(c(=O)n(c2c(c1=O)n(cn2)C)C)C
O=c1n(c(=O)n(C)c2c1n(cn2)C)C
n1(c(=O)c2c(ncn2C)n(c1=O)C)C
c1(=O)c2c(ncn2C)n(C)c(=O)n1C
Cn1c2c(n(C)c(=O)n(c2=O)C)nc1
n1(C)c(=O)c2c(n(C)c1=O)ncn2C
n1(C)c(=O)n(C)c2c(c1=O)n(C)cn2
Cn1c2c(c(=O)n(C)c1=O)n(cn2)C
n1(C)c(=O)n(C)c2c(n(C)cn2)c1=O
n1(c2c(n(C)cn2)c(=O)n(C)c1=O)C
n1(C)c(=O)n(c(=O)n2c1n(cn2)C)C
c1(=O)n(c(=O)n(c2c1n(C)cn2)C)C
n1(C)c2c(nc1)n(c(=O)n(C)c2=O)C
Cn1c(=O)n(C)c2c(c1=O)n(C)cn2
O=c1n(c2c(c(=O)n1C)n(C)cn2)C
n1(C)c2c(n(c(=O)n(C)c2=O)C)nc1
n1(C)c(=O)n(C)c2c(c1=O)n(cn2)C
n1(C)c2c(nc1)n(C)c(=O)n(c1=O)C
n1(C)c2c(n(cn2)C)c(=O)n(c1=O)C
n1(C)c2c(c(=O)n(C)c1=O)n(cn2)C
n1(c(=O)n(c2c(n(C)cn2)c1=O)C)C
n1(c(=O)n(c2c(c1=O)n(C)cn2)C)C
n1(C)c2c(n(C)cn2)c(=O)n(C)c1=O
n1(C)c2c(c(=O)n(c1=O)C)n(C)cn2
n1(C)c2c(n(c(=O)n(C)c2=O)C)nc1
n1(C)c2c(nc1)n(c(=O)n(c2=O)C)C
```

Current Progress for Markush/Generic Structures

The InChI project has initiated an RFP to expand the system capabilities to include Markush and generic structures, as well as polymers and mixtures. This work is expected to be completed by the end of 2011 or early 2012. It is hoped that organizations that have a vested interest in these areas will participate in these efforts to assure their rapid adoption. The EPO is one organization that the InChI developers hope will be part of this solution.

What is in it for the EPO?

While extending InChI into Markush and polymers is of value to the community, there is particular value to the EPO. The justification (or perhaps a better way to put it - the return on investment - ROI) is that the patent examiners can more easily, cost effectively, and correctly evaluate patent applications. This will improve quality and the quantity they can process. In addition, applicants and searchers can be more efficient and cost effective in their prior art checking and searching since InChIs are being added to databases and scientific journals at a increasing pace. There are no other notations now being used, e.g., SMILES or CAS numbers, that can make this statement, since both are proprietary, not widely readily available, and not likely to ever be non-proprietary.

Put very explicitly, today there already are more InChIs in databases and information resources than any other chemical identifier because of two factors. One is that InChIs are free. The second is that the Internet allows one to find information associated with an InChI.

Besides these practical and political benefits, should the EPO decide to ADD InChIs to what it uses to process patent applications, it will show leadership amongst the patent offices of the world.

Ways to add and use InChIs at the EPO

InChIs can be added (with the emphasis on ADDING and not being a replacement for anything) to the patents the EPO processes that have chemical structures. Adding chemical structures, such as InChIs can be done in a number of ways, depending on the technical, political, and regulatory issues involved.

1. The EPO can ask the applicants to send in the computer files they have ALREADY created to prepare their pdf applications they send in.

2. The EPO can tell/order applicants to do so (this is probably not so easy).

3.The EPO can hire some people at the EPO (or contract out) to redraw the pdf structures and create the InChIs for an internal EPO database that the EPO creates.

How difficult is it to create an InChI?

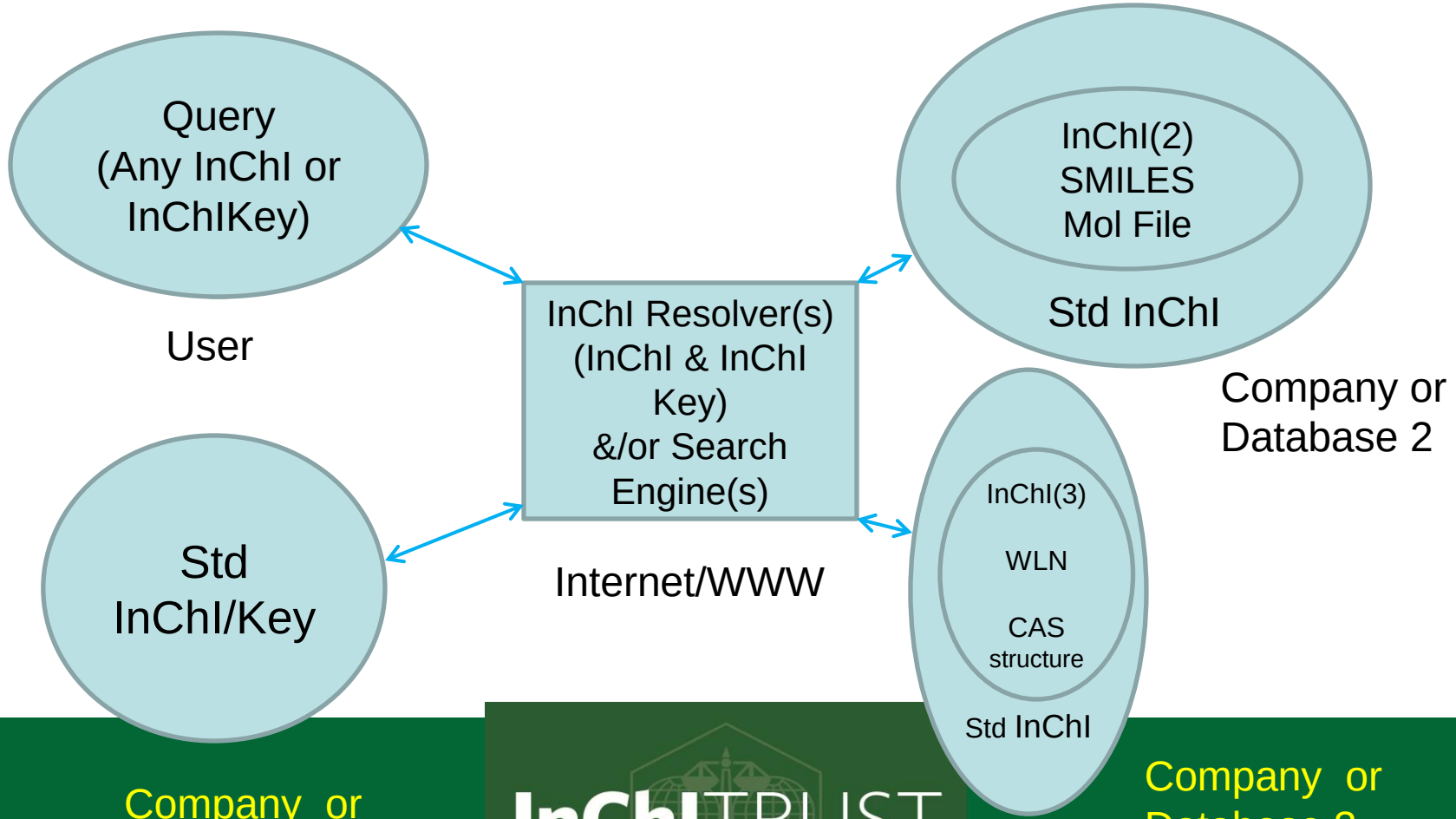
Today, all the major structure drawing programs (ChemDraw, MDL/Symyx/Accelrys Draw, ISIS Draw, ChemAxon Marvin Sketch, ACD Labs ChemSketch, Jmol and so on) have incorporated the InChI algorithm in their products, with usually an "InChI" button for generating the InChI. Since no patent application structure is now hand drawn, all patent applicants have effectively created InChIs and can just easily make an InChI string as create a pdf for their filing application.

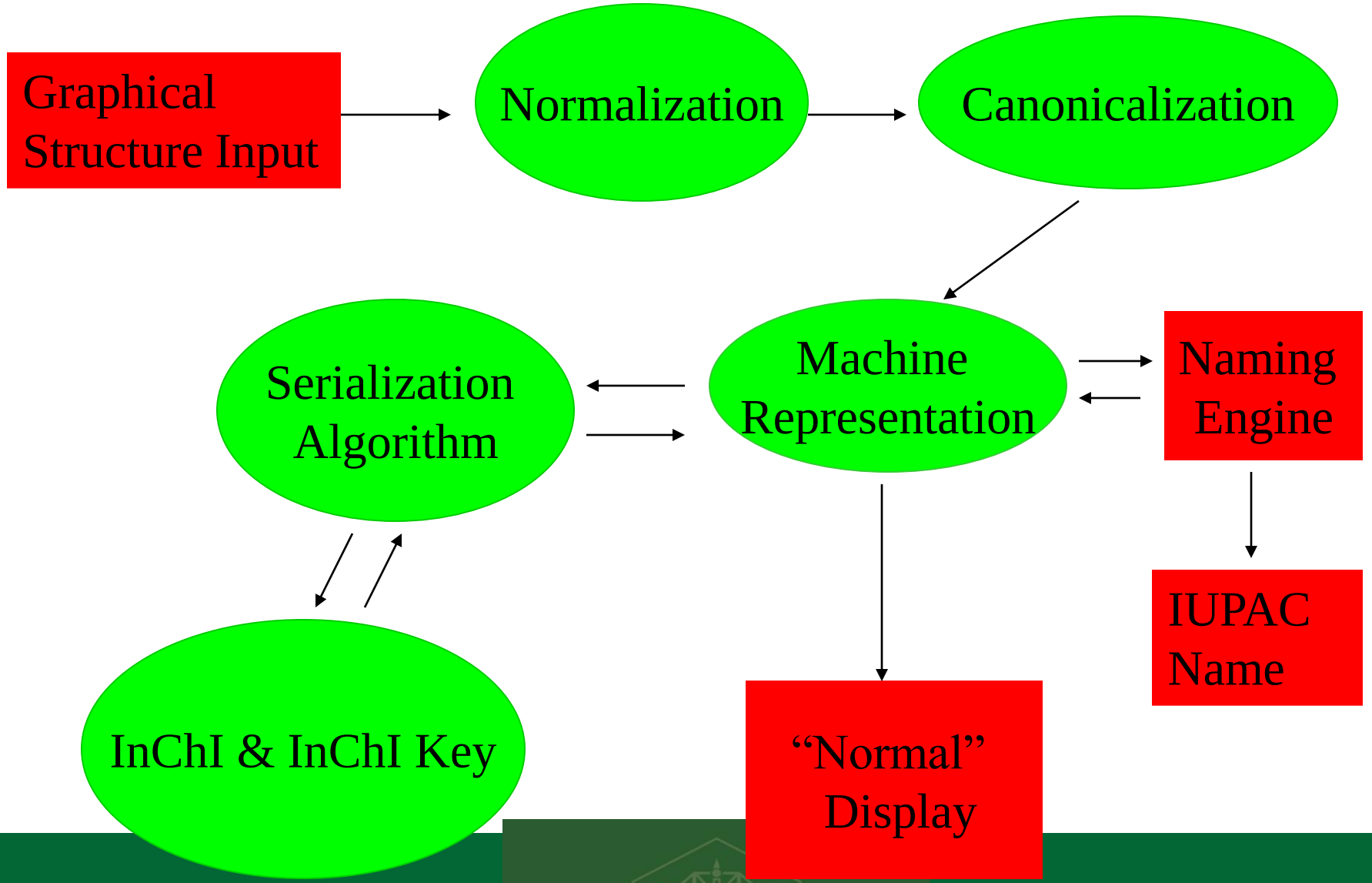
Who uses/searches InChIs?

InChIs are now found in virtually all major chemical databases, particularly in the very large ones. Databases such as Reaxys (30 million structures), NIH/PubChem (25 million structures), NIH/NCI (60 million structures), and SciFinder (55 million structures) all have InChIs and allow for InChIs as input for a search.

The next slide shows different databases from different organizations can link together and find ALL available information ONLY by using InChIs.

The LINKED and Interoperable and Combinable World of InChI




Graphical Structure Input
Normalization
Canonicalization
Machine Representation
Serialization Algorithm
Naming Engine
IUPAC Name
InChI & InChI Key
"Normal" Display

InChI layered structure design

The current InChI layers are:

1. **Formula**
2. **Connectivity (no formal bond orders)**
    a. **disconnected metals**
    b. **connected metals**
3. **Isotopes**
4. **Stereochemistry**
    a. **double bond (*Z/E)***
    b. **tetrahedral (sp3)**
5. **Tautomers (on or off)**

Charges are added to end of the string

InChI Characteristics

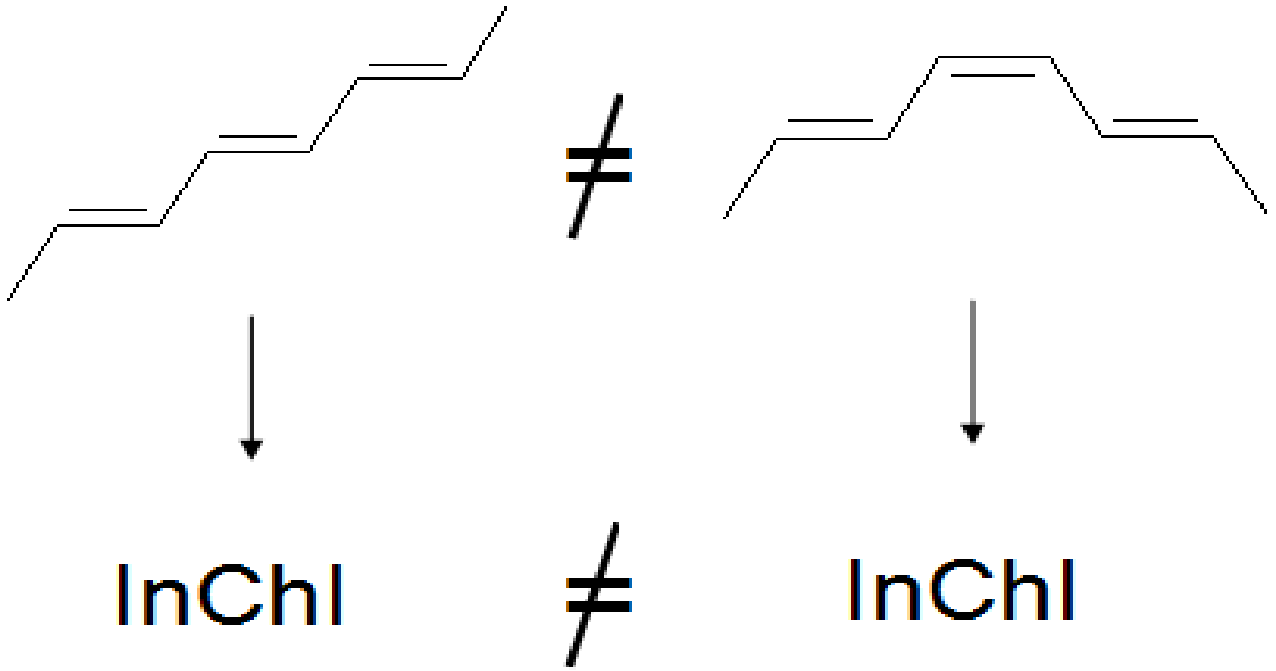
1. Easy to generate (It will use existing software.)

2. Expressive (It will contain structural information.)

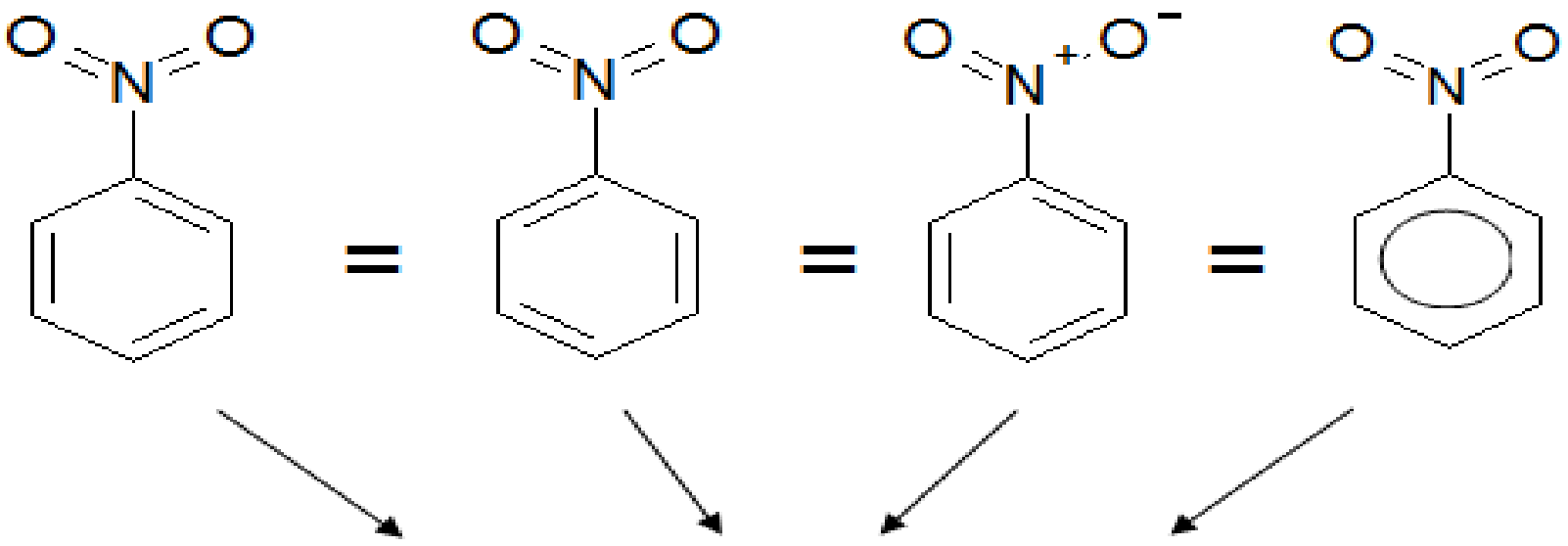
3. Unique/Unambiguous

4. Easy to search for structure via Internet search engines (Google, Yahoo, Microsoft Bing, etc.) using the InChI (hash) Key.

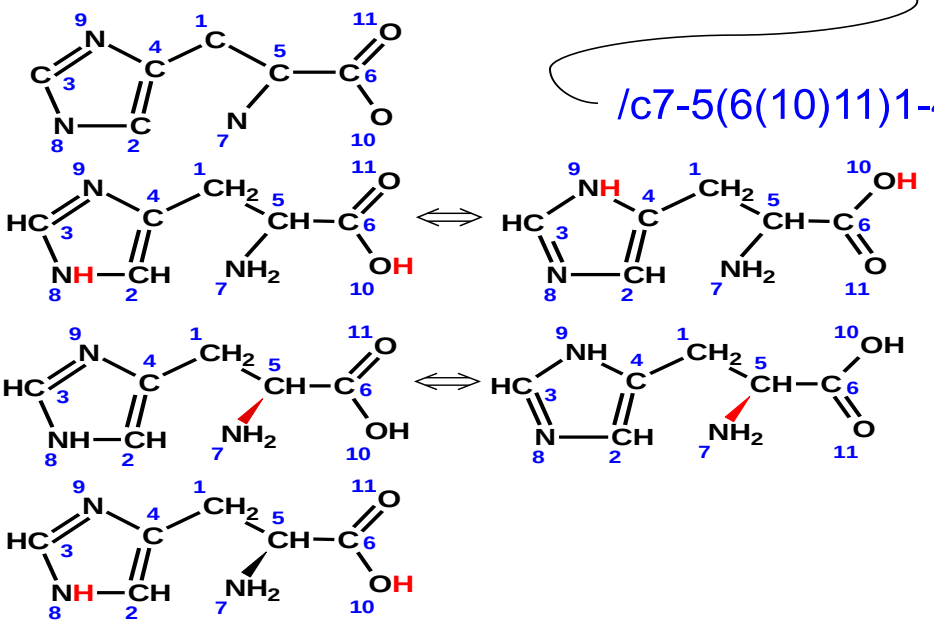
Different geometric or stereo isomers have different identifiers
–All distinguishing structural information is included



One compound has only ONE InChI



InChI Layers: L-Histidine



InChI=1/C6H9N3O2

Formula

Connections

/c7-5(6(10)11)1-4-2-8-3-9-4

Hydrogens (mobile)

/h2-3,5H,1,7H2,(H,8,9)(H,10,11)

Stereo

/t5-/m0/s1

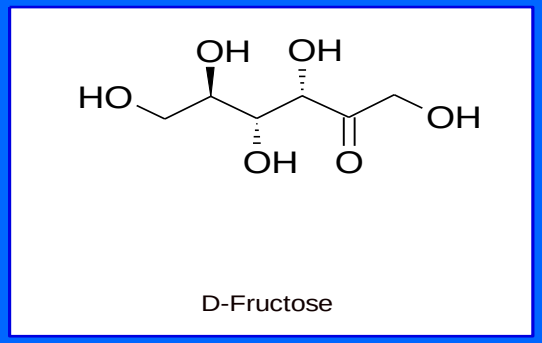
Hydrogens (fixed)

/f/h8,10H

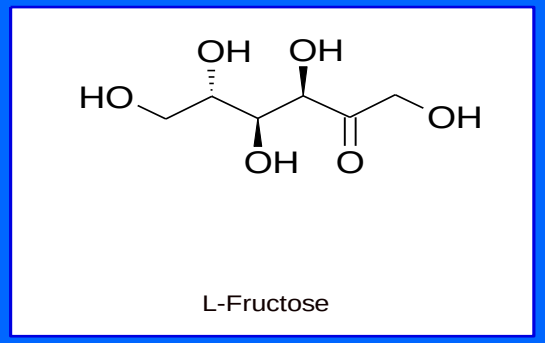
InChI=1/C6H9N3O2/c7-5(6(10)11)1-4-2-8-3-9-4/h2-3,5H,1,7H2,(H,8,9)(H,10,11)/t5-/m0/s1/f/h8,10H

InChIKey=HNDVDQJCIGZPNO-QLMCEAFFNA-N **InChIKey=HNDVDQJCIGZPNO-YFKPBYRVSA-N**

InChI TRUST



D-Fructose



L-Fructose

D-Fructose (natural)

InChI=1S/C6H12O6/c7-1-3(9)5(11)6(12)4(10)2-8/h3,5-9,11-12H,1-2H2/t3-,5-,6-/m1/s1

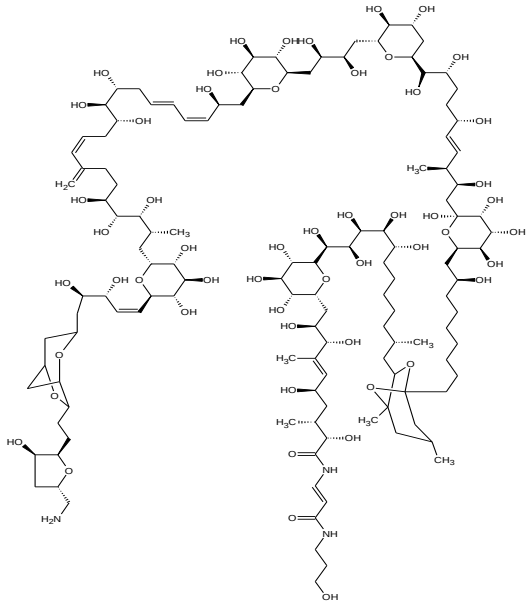
InChIKey=BJHIKXHVCXFQLS-UYFOZJQFSA-N

L-Fructose

InChI=1S/C6H12O6/c7-1-3(9)5(11)6(12)4(10)2-8/h3,5-9,11-12H,1-2H2/t3-,5-,6-/m0/s1

InChIKey=BJHIKXHVCXFQLS-FUTKDDECSA-N

Really long InChI (Palytoxin)



Palytoxin

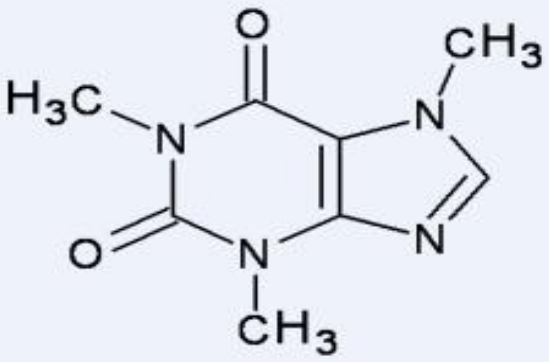
Isolated from Hawaiian soft coral

One of the most toxic non-peptide substances

Contains >70 stereochemical elements

InChI=1S/C129H223N3O54/c1-62(29-33-81(143)108(158)103(153)68(7)47-93-111(161)117(167)110(160)91(180-93)36-35-76(138)82(144)51-73-50-74-53-92(178-73)90(177-74)38-37-89-85(147)52-75(61-130)179-89)23-20-28-78(140)105(155)77(139)26-18-13-16-25-70(135)48-94-112(162)118(168)113(163)97(181-94)55-84(146)83(145)54-95-107(157)87(149)57-96(182-95)106(156)80(142)34-32-69(134)31-30-65(4)88(150)60-129(176)125(174)123(173)115(165)99(184-129)49-71(136)24-15-10-9-11-19-40-128-59-64(3)58-127(8,186-128)100(185-128)44-63(2)22-14-12-17-27-79(141)109(159)116(166)120(170)122(172)124-121(171)119(169)114(164)98(183-124)56-86(148)102(152)66(5)45-72(137)46-67(6)104(154)126(175)132-42-39-101(151)131-41-21-43-133/h13,16,18,20,23,25,30-31,35-36,39,42,45,63-65,67-100,102-125,133-150,152-174,176H,1,9-12,14-15,17,19,21-22,24,26-29,32-34,37-38,40-41,43-44,46-61,130H2,2-8H3,(H,131,151)(H,132,175)/b18-13+,23-20-,25-16-,31-30+,36-35-,42-39+,66-45+/t63-,64?,65-,67+,68+,69+,70+,71-,72-,73?,74?,75-,76+,77+,78+,79+,80+,81-,82+,83+,84+,85+,86-,87+,88-,89+,90?,91+,92?,93+,94-,95+,96-,97+,98+,99+,100?,102+,103+,104-,105-,106?,107-,108+,109-,110+,111-,112-,113+,114-,115-,116-,117-,118+,119+,120+,121-,122-,123+,124?,125+,127?,128?,129-/m0/s1

InChIKey=CWODDUGJZSCNGB-DCBUCRFRSA-N



InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3 (caffeine)

character indicating the number of protons ('N' means neutral)

InChIKey=RYYVLZVUVIJVGH-UHFFFAOYSA-N

flag character for InChI version: 'A' for version 1

First block (14 letters)

Encodes molecular skeleton (connectivity)

Second block (8 letters)

Encodes stereochemistry and isotopes

flag character ('S') indicates standard InChIKey (produced out of standard InChI)

1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3 Search

About 5,780 results (0.36 seconds) Advanced search

Everything
Images
Videos
More

Show search tools

InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3**
InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H**,1-3H3 ... reveals an inhibitor of **Mre11**-Rad50-Nbs1 complex , Nature Chemical Biology, 2008 ...
www.chemspider.com/InChIKey=RYYVLZVUVIJVGH-UHFFFAOYAW - Cached - Similar

Caffeine - Wikipedia, the free encyclopedia
1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3. InChI key, RYYVLZVUVIJVGH-UHFFFAOYAW. Properties. Molecular formula, C8H10N4O2 ...
en.wikipedia.org/wiki/Caffeine - Cached - Similar

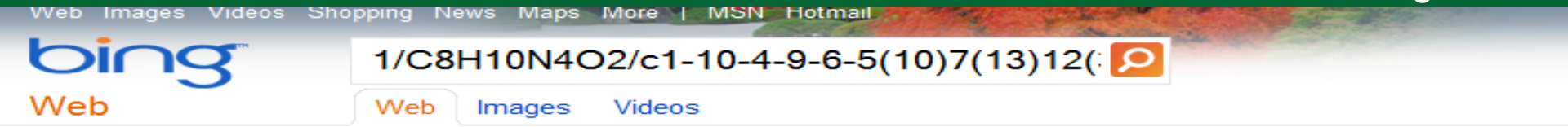
Compound **7** : Moonlighting proteins Hal3 and Vhs3 form a ...
Nov 1, 2009 ... InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3**. InChIKey: RYYVLZVUVIJVGH-UHFFFAOYAW ...
www.nature.com › Journal home › Archive › Article › Full text

caffeine (CHEBI:27732)
Oct 17, 2009 ... InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3**. InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** ...
www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:**27732** - Cached

InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3**
InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3**. ... reveals an inhibitor of Mre11-Rad50-Nbs1 complex , Nature Chemical Biology, 2008 ...
mesh.chemspider.com/Chemical-Structure.2424.html - Cached

Caffeine Mass Spectrum
CH$NAME: Caffeine CH$FORMULA: C8H10N4O2 CH$EXACT_MASS: 194.08038 CH$SMILES: ... CH$IUPAC: **1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** ...
www.massbank.jp/jsp/Dispatcher.jsp?type=disp&id...**1** - Cached - Similar

caffeine 58-08-**2**
Aug 3, 2010 ... IUPAC Name -, 1,3,7-trimethylpurine-2,6-dione. InChI -, InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** ...
www.thegoodscentscompany.com/data/rw**1014161**.html - Cached - Similar



SEARCH HISTORY
Search more to see your history

See all
Clear all · Turn off

ALL RESULTS 1-10 of 60 results · Advar

NMRanalyst Sample Application: Caffeine
... **1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** ... 1D Proton Resonances From Web Site: $> cat ...
www.sciencesoft.net/caffeine/index.html · Cached page

ChemSpider News » ChemSpider Integrations
The InChiÂ andÂ InChIKey for caffeine are shown below: InChI=**1/C8H10N4O2 /c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** InChIKey=RYYVLZVUVIJVGH-UHFFFAOYAW
www.chemspider.com/news/category/integration · Cached page

InChI=1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
Log Octanol-Water Partition Coef (SRC): Log Kow (KOWWIN v1.67 estimate) = 0.16 Log Kow (Exper. database match) = -0.07 Exper.
www.chemspider.com/Chemical-Structure.2424.html · Cached page

caffeine 58-08-2
1,3,7-trimethylpurine-2,6-dione: InChI - InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2 /h4H,1-3H3**: InChIKey - RYYVLZVUVIJVGH-UHFFFAOYAW
www.thegoodscentscompany.com/data/rw1014161.html · Cached page

Chemistry and Biology support, KDE/Strigi GSoC project: August 2007
InChI=**1/C8H10N4O2/ c1-10-4-9-6-5(10)7(13) 12(3)8(14)11(6)2/ h4H,1-3H3** The solution was to add a special flag to chemistry.inchi ontology field property that would indicate that a ...
neksa.blogspot.com/2007_08_01_archive.html · Cached page

International Union of Pure and Applied Chemistry
InChI=**1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3** InChIKey=RYYVLZVUVIJVGH-UHFFFAOYAW First block (**14** letters), encodes molecular skeleton (connectivity ...
www.iupac.org/inchi/release102.html · Cached page

Caffeine Mass Spectrum
... name: caffeine ch$formula: c8h10n4o2 ch$exact_mass: 194.08038 ch$smiles: cn(c2)c(c(=o)1)c(n2)n(c)c(=o)n(c)1 ch$iupac: **1/c8h10n4o2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2 /h4h,1-3h3** ...
www.massbank.jp/jsp/Dispatcher.jsp?type=disp&id=PR010011&site=1 · Cached page

InChI TRUST

Scientific Articles Mentioning InChI

"The Chemical Translation Service (CTS) - a web-based tool to improve standardization of metabolomic reports"
Gert Wohlgemuth, Pradeep Kumar Haldiya, Egon Willighagen, Tobias Kind, and Oliver Fiehn
Bioinformatics, published 9 September 2010 (Open Access)

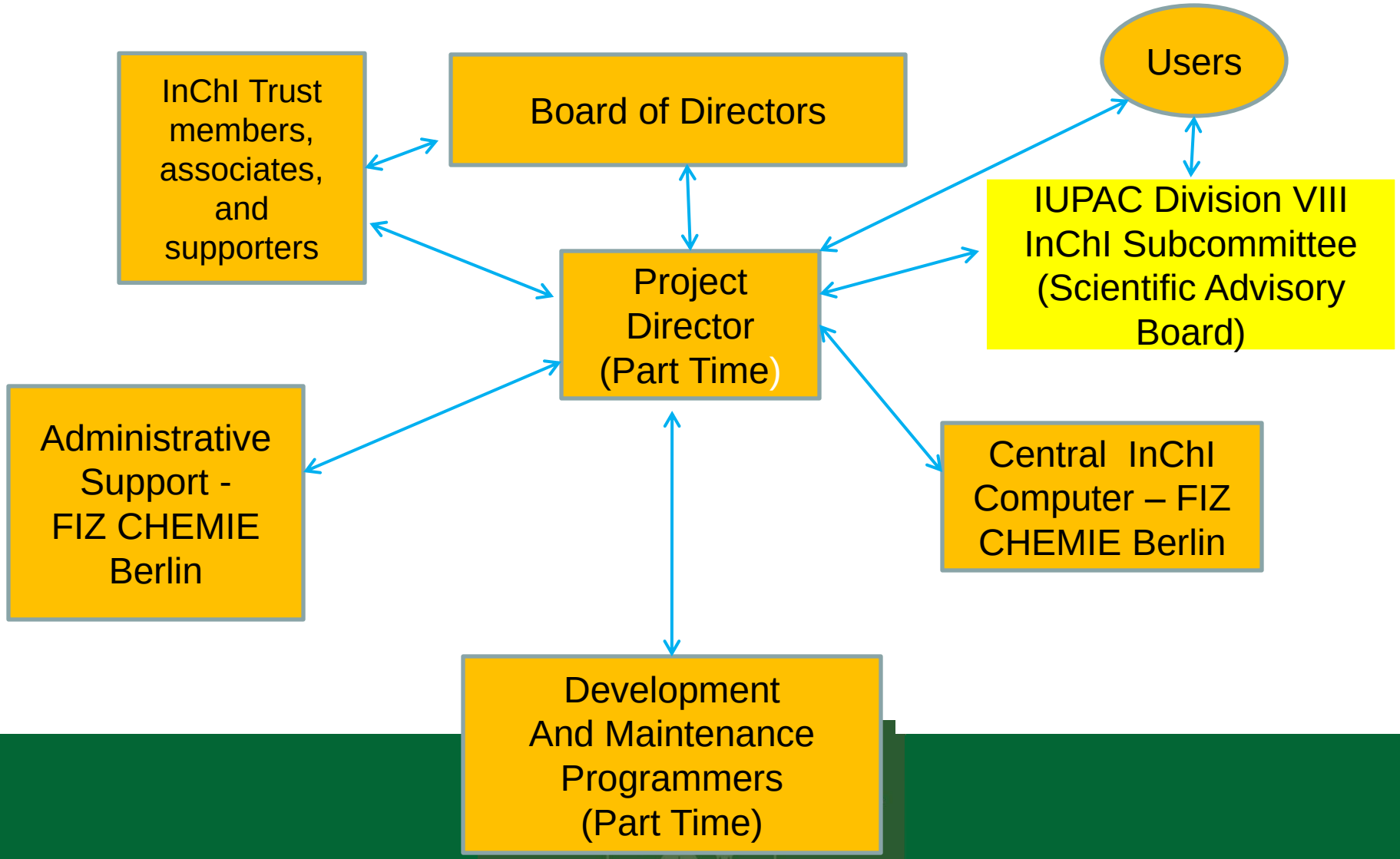
"PathwayAccess: CellDesigner plugins for pathway databases"
John L. Van Hemert and Julie A. Dickerson
Bioinformatics **2010**, *26(18)*, 2345-2346 (Open Access)

"Utopia documents: linking scholarly literature with research data"
T. K. Attwood, D. B. Kell, P. McDermott, J. Marsh, S. R. Pettifer, and D. Thorne
Bioinformatics **2010**, *26(18)*, 568-574 (Open Access)

The InChI Trust

With the requirements met of what areas of chemistry InChIs were needed for NIST databases, and since IUPAC is fundamentally and culturally a volunteer organization, a way had to be found to continue development of InChI, and maintain the InChI algorithm. InChI had to be "institutionalized" and turned over to an entity that would ensure its ongoing activities and be acceptable to the community. It was concluded that a not-for-profit organization would best fit the ongoing and future project needs. Thus the decision to create and incorporate the "InChI Trust". As there is no "free lunch", the Trust needs resources to continue to operate. Membership in the InChI Trust requires annual dues. The income from these revenues will be used exclusively for InChI development, maintenance, and educational activities associated with the project. Membership will entitle a member to influence the direction, priority, and speed of further Trust activities. Those organizations which do not join the InChI Trust will still have free access to the InChI algorithms but will not participate in any decision-making or direction - setting activities.

InChI Trust Organization



Current InChI Trust Members

Accelrys
ACD/Labs
ChemAxon
CSIRO
Dialog
Elsevier Properties SA
FIZ CHEMIE
IBM Research
IUPAC
Informa / Taylor & Francis
Nature Publishing Group
OpenEye
Royal Society of Chemistry
Springer
Wiley

15 as of 5/2011

Current InChI Trust Supporters

American Chemical Society Division of Chemical Information (CINF) (Carmen Nitsche)
Caltech Library Services, Pasadena, CA, USA (Dana Roth)
Chemistry Department, University of California, Riverside, CA, USA (Chris Reed)
Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, NC, USA (Alex Tropsha)
ETH Zürich, Chemistry Biology Pharmacy Information Center, Switzerland (Martin Brändle)
Faculty of Science, University of Paderborn, Germany (Gregor Fels)
Imperial College London, UK (Henry Rzepa)
Institute for Chemoinformatics and Bioinformatics, University of Applied Sciences Gelsenkirchen, Recklinghausen Section, Germany (Achim Zielesny)
Leadscope, Columbus, OH, USA (Michael Conley)
Ludwig-Maximilians-Universität München, Munich, Germany (Thomas Engel)
National Center for Biomedical Ontology, Stanford University, CA, USA (Mark Musen)
National Chemical Laboratory, Pune, India (Muthukumarasamy Karthikeyan)
National Institute of Chemistry, Ljubljana, Slovenia (Dusanka Janezic)
NextMove Software, Santa Fe, NM, USA (Roger Sayle)
Open Babel (Noel O'Boyle)
SciencePoint, Redmond, WA, USA (Rudy Potenzone)
Technical University of Vienna, Austria (Ulrich Jordis)
The Chem21 Group, Inc., Lake Forest, IL, USA (Tony Hopfinger)
Trinity University, San Antonio, TX, USA (Steven Bachrach)
Unilever Centre for Molecular Science Informatics, Cambridge University, UK (Robert Glen)
University of California, Davis, Genome Center, CA, USA (Oliver Fiehn)
University of California, San Francisco, CA, USA (John Irwin)
University of Indiana, Bloomington, IN, USA (David Wild)
University of the West Indies, Mona Campus, Jamaica (Robert Lancashire)
Xemistry GmbH, Königstein, Germany (Wolf-Dietrich Ihlenfeldt)

25 as of 5/2011

Current IUPAC Working Groups

In Progress:

Organometallics
InChI Resolver
Electronic States
RInChI –InChI for Reactions

Completed:

Markush
Polymers/Mixtures

To be started in 2012:

Inorganics

Possible Future Enhancements

1. Transrutherfordium elements
2. Transition states and Excited states.
3. Work with IUCr for 3D information
4. Proteins, Peptides & Biopolymers
5. Mac supported version
6. Java version
7. VS2010 .NET compilation support
8. Integrate with Microsoft Chem4Word

The Future

InChI has become mainstream for publishers, databases providers, and software developers. Over the next 5-10 years, publishers will use data mining to create both better abstracts, useful indexing, and concept terms. Search engines will be able to search for appropriate text and structures and direct users to the original (fee or free/Open Access/Open Data) sources.

Acknowledgements

(Primarily members for the IUPAC InChI subcommittee and associated InChI working groups)

Steve Bachrach, Colin Batchelor, John Barnard, Evan Bolton, Steve Boyer, Steve Bryant, Szabolcs Csepregi , Rene Deplanque, Nicko Goncharoff, Jonathan Goodman, Guenter Grethe, Richard Hartshorn, Jaroslav Kahovec , Richard Kidd, Hans Kraut, Alexander Lawson , Peter Linstrom, Bill Milne, Gerry Moss, Peter Murray-Rust, Heike Nau , Marc Nicklaus, Carmen Nitsche, Matthias Nolte , Igor Pletnev, Josep Prous, Hinnerk Rey, Ulrich Roessler, Roger Schenck , Martin Schmidt, Steve Stein, Peter Shepherd, Markus Sitzmann, Chris Steinbeck, Keith Taylor, Dmitrii Tchekhovskoi, Bill Town, Wendy Warr, Jason Wilde, Tony Williams, Andrey Yerin.

Special Acknowledgement: Ted Becker& Alan McNaught for their vision and leadership of the future of IUPAC nomenclature.