# IUPAC International Chemical Identifier (InChI)
# InChI version 1, software version 1.04 (September 2011)

# Release Notes

Last revision date: September 12, 2011

---

This document is a part of the release of the IUPAC International Chemical Identifier with InChIKey, version 1, software version 1.04.

---

## CONTENTS



This document summarizes and briefly explains the new features of InChI software v. 1.04.

# Summary

The current version of InChI Identifier is 1; the current version of InChI software is 1.04 (September 2011). Previously released software versions 1.01 (2006), 1.02-beta (2007), 1.02-standard (2009), and 1.03 (June 2010) as well as all earlier versions are now considered obsolete.

The current software is able of producing both standard and non-standard InChI/InChIKey.

The summary of the features of the distrubuted package is as follows:

- the software licence is changed from GNU Lesser General Public Licence, LGPL version 2.1 to the more permissive IUPAC/InChI-Trust Licence for the International Chemical Identifier (InChI) Software version 1.0;
- the software now supports the chemical elements up to 112, copernicium (the last one which is currently recognised by IUPAC; support for elements 105-112 is newly added);
- the possibility of processing multiple input files at a single run is added to the inchi-1 executable (both Windows and Linux versions); the common file name wildcards are recognized;
- the support for CML input files is no longer provided; related source codes were removed ;
- several minor bugfixes/changes were made to the source codes;
- InChI executable and InChI API library binaries are supplied in 32 and 64 bit versions for both Windows and Linux.

# Licensing

Previously, InChI software has been published under the GNU Lesser General Public Licence, LGPL version 2.1. Now it is replaced with the more permissive IUPAC/InChI-Trust Licence for the International Chemical Identifier (InChI) Software version 1.0.

The text of the IUPAC/InChI-Trust Licence is given below (it is also included in the distribution package as the file LICENCE).

## IUPAC/InChI-Trust Licence for the International Chemical Identifier (InChI) Software version 1.0

Preamble

This licence is released by the International Union of Pure and Applied Chemistry (IUPAC) and InChI Trust Limited.

This licence is a modified version of the Terms and Conditions of the GNU Lesser General Public Licence version 2.1. Section 6(a) of the Terms and Conditions has been amended to allow a more permissive use of the InChI software and sections 2(c),13-16 (inclusive) and the instructions for use have been amended accordingly.

Terms and Conditions for Copying, Distribution and Modification of the InChI Software

0. This Licence Agreement applies to any software library or other program which contains a notice placed by the copyright holder or other authorized party saying it may be distributed under the terms of this Licence. The Licensee is addressed as "you".
'IUPAC' means the International Union of Pure and Applied Chemistry.

A "library" means a collection of software functions and/or data prepared so as to be conveniently linked with application programs (which use some of those functions and data) to form executables.

The "Library", below, refers to any such software library or work which has been distributed under these terms. A "work based on the Library" means either the Library or any derivative work under copyright law: that is to say, a work containing the Library or a portion of it, either verbatim or with modifications and/or translated straightforwardly into another language. (Hereinafter, translation is included without limitation in the term "modification".)

"Source code" for a work means the preferred form of the work for making modifications to it. For a library, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the library.

Activities other than copying, distribution and modification are not covered by this Licence; they are outside its scope. The act of running a program using the Library is not restricted, and output from such a program is covered only if its contents constitute a work based on the Library (independent of the use of the Library in a tool for writing it). Whether that is true depends on what the Library does and what the program that uses the Library does.

1. You may copy and distribute verbatim copies of the Library's complete source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate

copyright notice and disclaimer of warranty; keep intact all the notices that refer to this Licence and to the absence of any warranty; and distribute a copy of this Licence along with the Library.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Library or any portion of it, thus forming a work based on the Library, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) The modified work must itself be a software library.

b) You must cause the files modified to carry prominent notices stating that you changed the files and the date of any change.

c) You must cause the whole of the work to be licensed at no charge to all third parties under the terms of this Licence. This requirement does not extend to any "work that uses the Library" that might also be compiled or linked against the "work based on the Library."

d) If a facility in the modified Library refers to a function or a table of data to be supplied by an application program that uses the facility, other than as an argument passed when the facility is invoked, then you must make a good faith effort to ensure that, in the event an application does not supply such function or table, the facility still operates, and performs whatever part of its purpose remains meaningful.

(For example, a function in a library to compute square roots has a purpose that is entirely well-defined independent of the application. Therefore, Subsection 2d requires that any application-supplied function or table used by this function must be optional: if the application does not supply it, the square root function must still compute square roots.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Library, and can be reasonably considered independent and separate works in themselves, then this Licence, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Library, the distribution of the whole must be on the terms of this Licence, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Library.

In addition, mere aggregation of another work not based on the Library with the Library (or with a work based on the Library) on a volume of a storage or distribution medium does not bring the other work under the scope of this Licence.

3. You may opt to apply the terms of the ordinary GNU General Public Licence instead of this Licence to a given copy of the Library. To do this, you must alter all the notices that refer to this Licence, so that they refer to the ordinary GNU General Public Licence, version 2, instead of to this Licence.
(If a newer version than version 2 of the ordinary GNU General Public Licence has appeared, then you can specify that version instead if you wish.)
Do not make any other change in these notices.

Once this change is made in a given copy, it is irreversible for that copy, so the ordinary GNU General Public Licence applies to all subsequent copies and derivative works made from that copy.

This option is useful when you wish to copy part of the code of the Library into a program that is not a library.

4. You may copy and distribute the Library (or a portion or derivative of it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you accompany it with the complete

corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange.

If distribution of object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place satisfies the requirement to distribute the source code, even though third parties are not compelled to copy the source along with the object code.

5. A program that contains no derivative of any portion of the Library, but is designed to work with the Library by being compiled or linked with it, is called a "work that uses the Library". Such a work, in isolation, is not a derivative work of the Library, and therefore falls outside the scope of this Licence.

6. You may combine or link a "work that uses the Library" with the Library to produce a work containing portions of the Library, and distribute that work under terms of your choice.

You must give prominent notice with each copy of the work that the Library is used in it and that the Library and its use are covered by this Licence.

You must supply a copy of this Licence. If the work during execution displays copyright notices, you must include the copyright notice for the Library among them, as well as a reference directing the user to the copy of this Licence.

Also, you must do one of these things:

a) Accompany the work with the complete corresponding machine-readable source code for the Library including whatever changes to the Library were used in the work (which must be distributed under Sections 1 and 2 above).

b) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (1) uses at run time a copy of the library already present on the user's computer system, rather than copying library functions into the executable, and (2) will operate properly with a modified version of the library, if the user installs one, as long as the modified version is interface-compatible with the version that the work was made with.

c) Accompany the work with a written offer, valid for at least three years, to give the same user the materials specified in Subsection 6a, above, for a charge no more than the cost of performing this distribution.

d) If distribution of the work is made by offering access to copy from a designated place, offer equivalent access to copy the above specified materials from the same place.

e) Verify that the user has already received a copy of these materials or that you have already sent this user a copy.

7. You may place library facilities that are a work based on the Library side-by-side in a single library together with other library facilities not covered by this Licence, and distribute such a combined library, provided that the separate distribution of the work based on the Library and of the other library facilities is otherwise permitted, and provided that you do these two things:

a) Accompany the combined library with a copy of the same work based on the Library, uncombined with any other library facilities. This must be distributed under the terms of the Sections above.

b) Give prominent notice with the combined library of the fact that part of it is a work based on the Library, and explaining where to find the accompanying uncombined form of the same work.

8. You may not copy, modify, sublicense, link with, or distribute the Library except as expressly provided under this Licence. Any attempt otherwise to copy, modify, sublicense, link with, or distribute the Library is void, and will automatically terminate your rights under this Licence. However, parties who have received copies, or rights, from you under this Licence will not have their licences terminated so long as such parties remain in full compliance.

9. You are not required to accept this Licence, since you have not signed it. However, nothing else grants you permission to modify or distribute the Library or its derivative works. These actions are prohibited by law if you do not accept this Licence. Therefore, by modifying or distributing the Library (or any work based on the Library), you indicate your acceptance of this Licence to do so, and all its terms and conditions for copying, distributing or modifying the Library or works based on it.

10. Each time you redistribute the Library (or any work based on the Library), the recipient automatically receives a licence from the original licensor to copy, distribute, link with or modify the Library subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties with this Licence.

11. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this Licence, they do not excuse you from the conditions of this Licence. If you cannot distribute so as to satisfy simultaneously your obligations under this Licence and any other pertinent obligations, then as a consequence you may not distribute the Library at all. For example, if a patent licence would not permit royalty-free redistribution of the Library by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this Licence would be to refrain entirely from distribution of the Library.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply, and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system which is implemented by public licence practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this Licence.

12. If the distribution and/or use of the Library is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Library under this Licence may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this Licence incorporates the limitation as if written in the body of this Licence.

13. IUPAC and the InChI Trust Limited may publish revised and/or new versions of the IUPAC/InChI Trust Licence for the International Chemical Identifier (InChI) Software from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Library specifies a version number of this Licence which applied to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by IUPAC and the InChI Trust Limited.

14. If you wish to incorporate parts of the Library into other free programs whose distribution conditions are incompatible with these, write to the author to ask for permission.

15. If you modify the Library in any way whatsoever, the output from any such modified Library may not be referred to as 'InChI' or any similar name. Any attempt to refer to such output as 'InChI' will automatically terminate your rights under this Licence.

NO WARRANTY

15. Because the Library is licensed free of charge, there is no warranty for the Library, to the extent permitted by applicable law. Except when otherwise stated in writing the copyright holders and other parties provide the Library "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the

implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the Library is with you. Should the Library prove defective, you assume the cost of all necessary servicing, repair or correction.

16. In no event unless required by applicable law or agreed to in writing will any copyright holder, or any party who may modify and/or redistribute the Library as permitted above, be liable to you for damages, including any general, special, incidental or consequential damages arising out of the use or inability to use the Library (including but not limited to loss of data or data being rendered inaccurate or losses sustained by you or third parties or a failure of the Library to operate with any other software), even if such holder or other party has been advised of the possibility of such damages.

END OF TERMS AND CONDITIONS

Instructions for Use

To apply this Licence, attach the following notices to the library. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

IUPAC/InChI-Trust Licence for the International Chemical Identifier (InChI)
Software version 1.0.
Copyright ©IUPAC and InChI Trust Limited

This library is free software; you can redistribute it and/or modify it under the terms of the IUPAC/InChI Trust Licence for the International Chemical Identifier (InChI) Software version 1.0; either version 1.0 of the Licence, or (at your option) any later version.

This library is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the IUPAC/InChI Trust Licence for the International Chemical Identifier (InChI) Software version 1.0 for more details.

You should have received a copy of the IUPAC/InChI Trust Licence for the International Chemical Identifier (InChI) Software version 1.0 along with this library; if not, write to:

The InChI Trust
c/o FIZ CHEMIE Berlin
Franklinstrasse 11
10587 Berlin
GERMANY

Also add information on how to contact you by electronic and paper mail.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the library if necessary. Here is a sample:

XXXX Limited, hereby disclaims all copyright interest in the library 'YYYY'
(a library for 'ZZZZ') written by A N Other.

Signed by: _____
Name:
Position:

# New and updated features

## Support for chemical elements 105-112

The previous releases of InChI software supported chemical elements up to rutherfordium (Rf, atomic number 104).

In the current release, support was added for the elements 105-112, dubnium to copernicium (element 112, copernicium, is the last which is currently recognised by IUPAC).

To incorporate the elements into InChI software, one needs to define some parameters. In particular, these are:

(a) average atomic mass M1 and atomic mass of the most abundant isotope M2 (as integer numbers), and the exact mass M3 of the most abundant isotope (this is necessary for InChI to make an "isotopic" part of the identifier);

(b) valence schema, that is, typical valence(s) exhibited by the element as ion or "neutral" atom, V.

Elements 105-112 are artificial and have no stable isotopes. The atomic mass of the longest-lived isotope was selected for M2, instead of the mass of the most abundant one, and M1 was set equal to M2. Masses of isotopes were extracted from: *M. E. Wieser and T. B. Coplen. Atomic weights of the elements 2009 (IUPAC Technical Report). Pure Appl. Chem., 2011, v. 83, No. 2, pp. 359–396*.

All the elements 105-112 were described, in internal tables, as metals. As for their valences, one should note that the chemistry of the elements of the end of the Periodic Table is far from established; often, existing knowledge is based mainly on analogies rather than on exact knowledge.

Based primarily on the discussion in *"The Chemistry of the Actinide and Transactinide Elements". Ed. by L.R. Morss, N.M. Edelstein, J.Fuger. Springer; 3rd edition, 2007*, the standard valence (for the formally neutral atom) of Db was set to 5, Sg to 6 and Bh to 7. For the other newly added elements, 108-112, the formal value of 1 was used.[1]

The final parameters are listed in Table 1.

[1] In the previous versions of InChI software the value of 1 was used for the valence of supported post-Md elements, i.e. No to Rf (while actinides up to Md had the usual value of 3). In v. 1.04 these values were changed and set to: 2 for No (well-known exception in the actinide series), 3 for Lr, and 4 for Rf. Changing valences for indicated elements may change related InChI strings; however, the impact of this should be evidently quite small.

Table 1. Selected parameters of chemical elements 105-112.

| Element | Name | Element number | Atomic mass M1, M2 | Atomic mass M3 | Standard valence V (for charge 0) |
|---|---|---|---|---|---|
| Db | Dubnium | 105 | 268 | 268.1250 | 5 |
| Sg | Seaborgium | 106 | 271 | 271.1330 | 6 |
| Bh | Bohrium | 107 | 267 | 267.1277 | 7 |
| Hs | Hassium | 108 | 277 | 277.1500 | 1* |
| Mt | Meitnerium | 109 | 276 | 276.1510 | 1* |
| Ds | Darmstadtium | 110 | 281 | 281.1620 | 1* |
| Rg | Roentgenium | 111 | 280 | 280.1640 | 1* |
| Cn | Copernicium | 112 | 285 | 285.1740 | 1* |

* Formal value, indicates “unknown” or “uncertain”.

## AMI (Allow Multiple Inputs) mode

There were concerns regarding the performance of the inchi-1 executable when working with a large number (thousands and tens of thousands) of single-record MOL files, in comparison with a multi-record SDF file containing exactly the same molecular data.

The slow-down when working with multiple (MOL) files with respect to a large container (SDF) file is in principle natural. Running the executable multiple times assumes a multiply increased startup and file handling overhead.

It was observed that this overhead is especially significant in the case of working in MS Windows, as compared to Linux (while the difference in performance for treating an SDF file is small or absent).

To reduce the negative impact of this slowdown for those who work with a multitude of small files instead of a large container file, the possibility of processing multiple input files at a single run was added to the inchi-1 executable (both Windows and Linux versions).

This mode is activated by the inchi-1 command line option "/AMI" (Windows) or "–AMI" (Linux; AMI stands for "Allow Multiple Inputs"). In this mode, all the file names supplied in the command line are considered as the names of separate input files.

For further convenience, the common file name wildcards ("*" and "?") are supported.

For example, issuing a command

```
inchi-1 *.mol /AMI (Windows)
```

```
inchi-1 *.mol -AMI (Linux)
```

will instruct the executable to process all the mol-files in the current directory.

Note, that omitting the switch "AMI" assumes working in a conventional single-input mode which may result in undesired treatment of wildcards[2].

In AMI mode, the names of output, log and problem files could not be individually specified. Instead, they are formed, for each of multiple inputs, by appending the file name with suffixes ".txt", ".log" and ".prb". However, to partially mimic the behavior of inchi-1 in conventional single-input mode, three additional command line options are introduced (see Table 2). They allow one to redirect the output to `stdout`, log to `stderr`, as well as to suppress creation of problem files.

Examples *(Windows, Linux)*:

```
inchi-1 nci*.mol /AMI /AMIOutStd /AMIPrbNone /AuxNone /Key
```

```
./inchi-1 /home/me/mol/nci/*.mol -AMI -AMILogStd -AMIPrbNone
-RecMet –FixedH
```

[2] There is an important difference in wildcard expansion under Windows and Linux.
Under Windows, inchi-1 executable makes an expansion itself (if "AMI" switch is specified). That is, if "AMI" is omitted, no expansion occurs and "*.mol" is just considered as an invalid file name.
Under Linux, wildcards are always expanded by shell. That is, if "AMI" is omitted, "*.mol" will be expanded to the list of file names; the first four of them will be treated by inchi-1, according to single-input rules, as input, output, log and problem file names (which means that the last three files will be over-written).

Table 2. AMI mode options of inchi-1 executable.

| Option | Meaning |
|---|---|
| **AMI** | Allow multiple input files |
| **AMIOutStd** | Write output to `stdout` |
| **AMILogStd** | Write log messages to `stderr` |
| **AMIPrbNone** | Suppress creation of problem files |

As indicated by tests, processing of multiple MOL files in AMI mode may be several times faster (the exact speed-up depends on many details; anyway the processing time is still significantly longer than that for a single SDF file containing the same data).

### (No) CML support

Since 2004, InChI software had support for CML (Chemical Markup Language).

Since v. 1.04 of September 2011, this support for CML (ability to read CML input files) is no longer provided. However the possibility of developing an InChI interface to a separate CML module will be explored.

Accordingly, source codes for CML support have been removed from the distribution.

## Distribution package

### Binaries

This package includes 'command line' InChI executable and InChI API library binaries (32 and 64 bit versions are supplied for both Windows and Linux).

Also included is winchi-1.exe, a graphical Windows application (a 32 bit version which will also run under 64 bit Windows).

File/directory INCHI-1-BIN

windows/

| | |
|---|---|
| winchi-1.exe | InChI graphical Windows application |
| windows/32bit/ | |
| inchi-1.exe | InChI stand-alone command line executable, 32 bit |
| windows/32bit/dll/ | |
| libinchi.dll | InChI dynamic-link library, 32 bit |
| windows/64bit/ | |
| inchi-1.exe | InChI stand-alone command line executable, 64 bit |
| windows/64bit/dll | |
| libinchi.dll | InChI dynamic-link library, 64 bit |
| linux/ | |
| linux/32bit/ | |
| inchi-1.gz | InChI stand-alone command line executable, 32 bit; gzipped |
| linux/32bit/so/ | |
| libinchi.so.1.04.00.gz | shared library for InChI API, 32 bit; gzipped |
| linux/64bit/ | |
| inchi-1.gz | InChI stand-alone command line executable, 64 bit; gzipped |
| linux/64bit/so | |
| libinchi.so.1.04.00.gz | shared library for InChI API, 64 bit; gzipped |

Note that InChI stand-alone executable inchi-1[.exe] does not require dll/so libraries.

File/directory INCHI-1-API/INCHI_API

| | | |
|---|---|---|
| gcc_so_makefile/result/ | | |
| inchi_main | (Linux) | the InChI software library demo application |
| vc9/inchi_dll/Release/ | | |
| inchi_main.exe | (Windows) | the InChI software library demo application |

## Examples

Example programs which use InChI API library for both Windows (dll) and Linux (so) are supplied in INCHI-1-API/INCHI_API section of this distribution package. For the details, please refer to 'readme.txt' files in respective sub-directories.

There are examples for C ('inchi_main' application, see projects for MS Visual Studio 2008 'vc9' and gcc 'gcc_so_makefile' ) and Python ('python_sample'). Also supplied are InChI API library sorce codes and related MS Visual Studio 2008/gcc projects.

Note that to use the shared library, you may wish to create 'libinchi.so.1' as a symbolic link to 'libinchi.so.1.04.00'

### Other

The source codes and MS Visual Studio 2008/gcc projects for inchi-1 executable are supplied in INCHI-1-API/INCHI_API section of this distribution package. For further details, please refer to 'readme.txt' files in respective sub-directories.

The documentation (Release Notes; InChI Technical Manual; InChI User Guide; InChI API Reference) is supplied in INCHI-1-DOC section of this distribution package in PDFformat.

## Bugfixes/changes

There were several minor fixes/changes made after software release v. 1.03 (June 2010).

In particular:
(1) the bug in the normalization procedure for some structures (containing a radical at an atom in an aromatic ring) which may result in different InChI strings for the same molecule, depending on the original order of the atomic numbers (reported by Dmitry Pavlov, SciTouch) was fixed;
(2) the standard valences for No, Lr, and Rf were updated, see above;
(3) the unnecessary dependence of the inchi-1 executable for Windows on Microsoft run-time dynamic libraries was eliminated.

The combined impact of these changes on InChIs for real-world datasets is nearly negligible.

The new software has been extensively regression-tested against standard and non-standard InChIs generated with v. 1.03 (2010) software, in both Windows and Linux environments, with the various option combinations for standard and non-standard InChI. The test sets included:

1. "InChI-101" (public). This is a test set of 2,186 structures which has been created previously and included in software v. 1.01 distribution as "InChI validation suite". The structures include some very tricky and "chemically strange" ones, to verify InChI behaviour in exotic cases.

2. "NCI" (public). 249,081 structures from "NCI Open Database Compounds", retrieved from:

http://cactus.nci.nih.gov/ncidb2/download.html

3. "MSL-NIST" (proprietary). 191,436 structures.

4. "MDB" (proprietary). 100,000 structures.

5. "PubChem Compound" (public). 31,467,170 structures. Retrieved from PubChem on 2011-02-19.

6. "PubChem Substance" (public). 74,387,591 structures. This dataset, in contrast with "PubChem Compound", contains non-normalized structures just as they were deposited in PubChem. Retrieved from PubChem on 2011-02-19.

No changes in InChI strings in comparison with those produced with v. 1.03 software were observed[3].

[3] The dataset "PubChem Substance" contains elements 105-112 which raise an error "Unknown element(s)" with software v. 1.03; with current software v. 1.04 these elements are treated successfully.