



INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

InChI Software 1.02 Summary

The final InChI version 1.02 software was issued in January 2009 as an implementation for generating Standard InChI (see below) and the corresponding Standard InChIKey. The complete package, downloadable from www.iupac.org/inchi/download/ contains the following:

- source code and Application Program Interface (API)
- stand-alone executables for Windows and Linux (stdinchi-1.exe, stdwinchi-1.exe, and stdinchi-1.gz)
- description of new features, with examples of using new functionality
- copy of GNU LGPL licence

Standard InChI

In response to user requests, a Standard InChI (i.e. without options for properties such as tautomerism and stereoconfiguration) has been defined as follows:

Standard InChI is for the purposes of interoperability/compatibility between large databases/web searching and information exchange.

Standard InChI and non-standard InChI are always distinguishable.

Standard InChI is a stable identifier; however, periodic updates may be necessary; they are reflected in the identifier version designation, which is included in the InChI string.

Any shortcomings in standard InChI may be addressed using non-standard InChI (currently obtainable using InChI version 1.02beta).

Standard InChIKey

In response to user feedback the format of InChIKey has been changed; it is different from that in InChI software v. 1.02-beta, having 27 characters rather than 25.

Standard InChIKey has five distinct components.

1. 14-character hash of the basic (Mobile-H) InChI layer;
2. 8-character hash of the remaining layers (except for the “/p” segment, which accounts for added or removed protons: it is not hashed at all; the number of protons is encoded at the end of the standard InChIKey.)
3. 1 flag character,
4. 1 version character
5. the last character is a [de]protonation indicator.

The overall length of InChIKey is fixed at 27 characters, including separators (dashes):

AAAAAAAAAAAAA-BBBBBBBBFV-P

This is significantly shorter than a typical InChI string.

Here

(1) **AAAAAAAAAAAAAAAA** is a 14-character hash.

(2) **BBBBBBBB** is an 8-character hash

(3) **F** is a flag indicating standard InChIKey (produced out of standard InChI): it always has the value 'S'.

(4) **V** is a flag for InChI version character: 'A' for version 1, 'B' for version 2, etc.

P is an indicator for the number of protons; this number is not encoded in the hash but is indicated as a separate 2-character block at the end, where one character is a hyphen, as -N for neutral, -M for -1 hydrogen, -O for +1 hydrogen, etc.

Full details and examples are provided in the documentation accompanying the software download.

Software implementing the final InChI version 1.02 for non-standard InChI (i.e. with all previous options retained and with the 27-character InChIKey) will be issued in due course.

Users are encouraged to report their experiences and any problems via the SourceForge website (<http://sourceforge.net/projects/inchi>).

Steve Heller (Chair, IUPAC InChI Subcommittee)
Alan McNaught (Coordinator, IUPAC InChI project)
Igor Pletnev
Steve Stein
Dmitrii Tchekhovskoi

January 2009