



## INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

### InChI Software 1.02beta Summary

A beta-release of InChI version 1.02 software was issued in September 2007. The complete package contains the following:

Source code and Application Program Interface (API)  
Stand-alone executable (clnChI)  
Description of new features, with examples of using new functionality  
Copy of GNU LGPL licence  
Full InChI documentation, Windows executable (wInChI.exe), and validation set will be included in the final version 1.02 release.

The principal new features of this release are:

(1) A fixed-length (25-character) condensed digital representation of the Identifier to be known as InChIKey. In particular, this will

- facilitate web searching, previously complicated by unpredictable breaking of InChI character strings by search engines
- allow development of a web-based InChI lookup service
- permit an InChI representation to be stored in fixed length fields
- make chemical structure database indexing easier
- allow verification of InChI strings after network transmission.

An example of InChI with its InChKey equivalent is shown below. There is a finite, but very small probability of finding two structures with the same InChIKey. For duplication of only the first block of 14 characters this is 1.3% in  $10^9$ , equivalent to a single collision in one of 75 databases of  $10^9$  compounds each.

*Caffeine:*

*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): RYYVLZVUVIJVGH*

*Second block (8 letters), encodes proton positions (tautomers), stereochemistry, isotopes, reconnected layer: UHFFFAOY*

*Flag character, indicates InChI version, presence/absence of fixed H layer, isotopes, and stereochemistry: A*

*Check character: W*

(2) Restructured InChI generating software that separates key steps in its creation from an input chemical structure file. Among other uses, this allows checking of intermediate results to enable easier testing and development of InChI-based applications.

(3) Bug fixes designed to withstand malicious attempts to attack a Web server by providing a specially designed InChI string input to InChI binaries.  
Users are encouraged to report their experiences and any problems via the SourceForge website (<http://sourceforge.net/projects/inchi>).

Steve Heller  
Alan McNaught  
Igor Pletnev  
Steve Stein  
Dmitrii Tchekhovskoi

*September 2007*