InChI's in LaTex and CDK News

Egon Willighagen ⓘD

Citation

Keywords

Abstract

An InChI (or see the FAQ) is a line notation for a molecular structure that was recently developed by the NIST and the IUPAC. Principally they can be applied to protein too (see below), but because proteins would give lenghty InChI's and are quite well defined in terms of connectivity anyway, those can better be described by their amino acid sequence.

Copyright

chem-bla-ics

An InChI (or see the FAQ) is a line notation for a molecular structure that was recently developed by the NIST and the IUPAC. Principally they can be applied to protein too (see below), but because proteins would give lenghty InChI's and are quite well defined in terms of connectivity anyway, those can better be described by their amino acid sequence.

The March 2006 issue of CDK News, the Chemistry Development Kit project newsletter, will be released later today, and had, for the second time, the requirment that authors provide InChI's for molecular structures mentioned in the articles. Different from the previous issue is how InChI's are marked up in LaTeX. I've setup a `\inchi{}` for this that automatically creates a Google search query as link behind the InChI:

```
\newcommand{
   \inchi}[1]{\href{http://www.google.com/search?q=#1}
                  {\normalfont\texttt{InChI=#1}
            }
}
```

Now, googling for InChI's only works if one removes the `InChI=` part of the InChI. As an example I will show how it works for methane. The InChI for this compound is `InChI=1/CH4/h1H4`, so in LaTex one enters `\inchi{1/CH4/h1H4}`. This will create a link like: InChI=1/CH4/h1H4.

BTW, if you are interested in InChI's for proteins, here is the InChI for 1CRN, created with OpenBabel:

```
InChI=1/C202H439N55O64S6/c1-28-92(12)149-188(308)237-127-84-323-324-
85-128(176(296)225-114(46-37-63-212-202(209)210)165(285)232-122(69-89(6)7)195(315)253-64-
47-132(253)179(299)215-80-143(274)241-158(107(27)265)199(319)257-68-42-51-136(257)182(302
115(60-61-144(275)276)164(284)218-100(20)162(282)244-149)236-187(307)148(91(10)11)242-172
120(74-138(204)269)229-168(288)117(70-108-43-34-33-35-44-108)228-169(289)119(73-137(203)2
230-173(293)124(81-258)234-166(286)113(45-36-62-211-201(207)208)224-159(279)99(19)221-186
147(90(8)9)243-189(309)150(93(13)29-2)245-174(294)125(82-259)235-183(303)135-50-41-66-255
196(316)130-87-326-322-83-126(223-142(273)79-216-185(305)154(103(23)261)251-171(291)118(7
110-54-58-112(267)59-55-110)231-192(312)155(104(24)262)250-163(283)101(21)220-175(127)295
(298)246-151(94(14)30-3)190(310)247-152(95(15)31-4)191(311)248-153(96(16)32-5)198(318)256
40-49-134(256)181(301)213-77-140(271)217-97(17)161(281)249-156(105(25)263)194(314)240-131
(88-327-325-86-129(177(297)239-130)238-193(313)157(106(26)264)252-184(304)146(206)102(22)
(317)254-65-39-48-133(254)180(300)214-78-141(272)222-121(76-145(277)278)170(290)227-116(7
109-52-56-111(266)57-53-109)167(287)219-98(18)160(280)233-123(200(320)321)75-139(205)270/
202,211-252,258-321H,28-88,203-210H2,1-27H3/t92-,93-,94-,95-,96-,97-,98-,99-,100-,101-,10
103+,104+,105+,106+,107+,109-,110-,111+,112+,113-,114-,115-,116-,117-,118-,119-,120-,121-
123-,124-,125-,126-,127-,128-,129-,130-,131-,132-,133-,134-,135-,136-,137?,138-,139-,140-
142-,143+,146-,147-,148-,149-,150-,151-,152-,153-,154-,155-,156-,157-,158-,159+,160?,161-
163-,164-,165?,166+,167?,168+,169+,170+,171-,172+,173+,174+,175?,176-,177?,178+,179+,180-
181?,182-,183+,184?,185+,186+,187-,188-,189-,190+,191?,192-,193?,194-,195-,196-,197-,198-
```