Chemical structure drawing editors, coding of chemical structures. Use of coded chemical structures and various formula formats for searching in data bases / secondary resources and for calculation estimations of physico-chemical properties.

Why to use structure drawing editors?

- it´s difficult to draw structures by means of a general drawing editors (Photoshop, …)
- copying of structures as figures from various resources into your text or presentation brings differences in the look of individual structures and thus inconsistent and unsightly appearance
- only structures drawn with an structure editor can be transformed to universal codes or transport formates and / or used for structure searching in data bases or calculation estimation of physico-chemical or other properties

How to use structure drawing editors?

- if they are embedded into a web page, use it at this page and transfer into your text/presentation in a required formate
- or, you can download the page and use it locally on your PC
- example: JSME Molecular editor by Peter Ertel and Bruno Bienfait
  - rather suitable for usage with learning applications such as Moodle
  - you may know from Organic, Analytical, Medicinal Chemistry

How to use structure drawing editors?

- if they are installable software applications, (download and) install them on your PC and use together with your text or presentation editor
- you can use OLE function of your editor, or you can have opened windows with structure drawing and text/presentation editors simultaneously, draw the formula and copy it into your text or presentation

How to use structure drawing editors: an example of using the OLE function
ACD ChemSketch + MS PowerPoint

- Insert → Object → ACD ChemSketch → OK
  - ChemSketch opens a window; now you can draw a structure (or insert or import it from a file)
  - when you are done, select File → Update and then File → Close and return to MS Power Point
- now, switch back to your presentation and you can continue
- if you want to add/change/repair something in your structure:
  - double-click the formula in PowerPoint
  - you are back in ChemSketch and you can do what you want
  - again: when you are done, select File → Update and then File → Close and return to MS Power Point
- again, switch back to your presentation and you can continue

Examples of some available structure drawing and other chemical editors

- ACD ChemSketch
  – available for free 30days trial from webstore.acdlabs.com
  - registration required
- Marvin Sketch
  - available from https://chemaxon.com/products/marvin/download#download
  - the free license needs to be renewed every 2 months
  - for Windows, Mac, Linux (Debian, RedHat)
- Avogadro
  - available from https://sourceforge.net/projects/avogadro/
  - free
  - for Windows, Linux, Mac
  - enables direct constructing of 3D molecule models only
- Molsketch
  - available from https://sourceforge.net/projects/molsketch/
  - free
  - for Windows
  - only converts molecule codes to 3D models
  - cannot read SMILES from ChemId

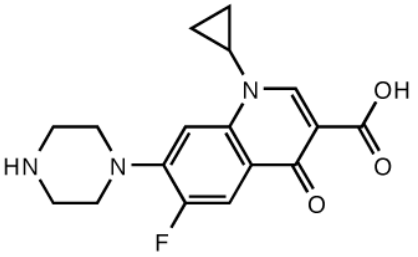
SMILES coding

SMILES = Simplified molecular-input line-entry system

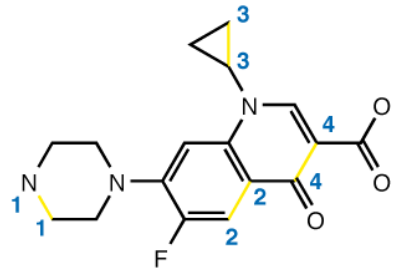
- a specification in the form of a line notation for describing the structure of chemical species (= compound) using short ASCII strings
- developed in late 1980th (David Weininger, Albert Leo, Corwin Hansch)

Brief principles of SMILES code creation:
SMILES generation algorithm for ciprofloxacin: break cycles, then write as branches off a main backbone

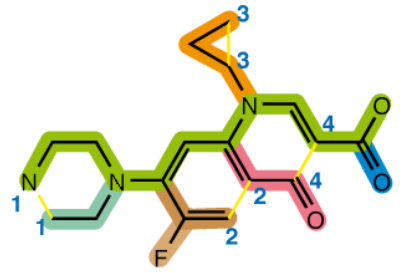
A



B



C



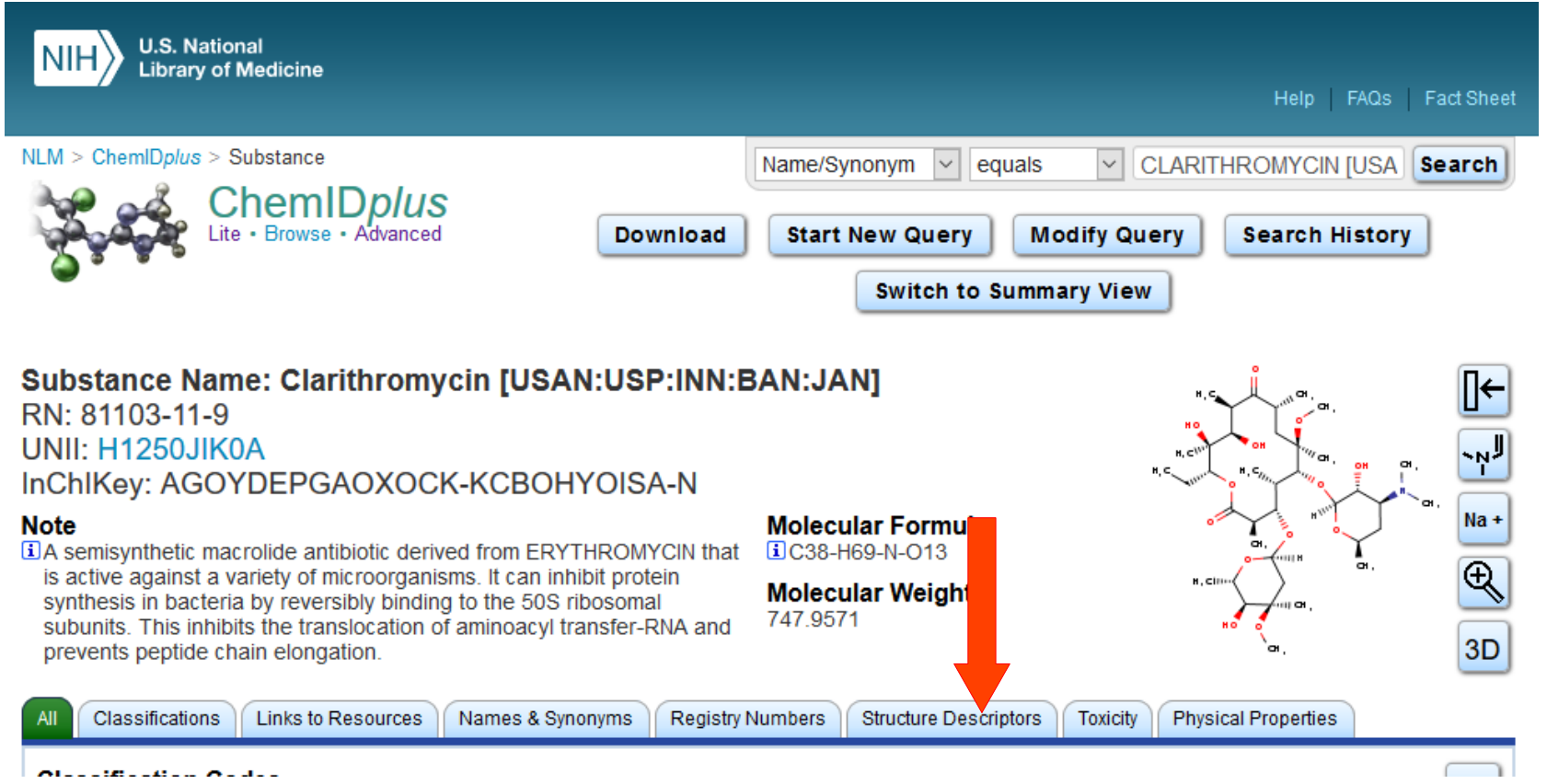
D

N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

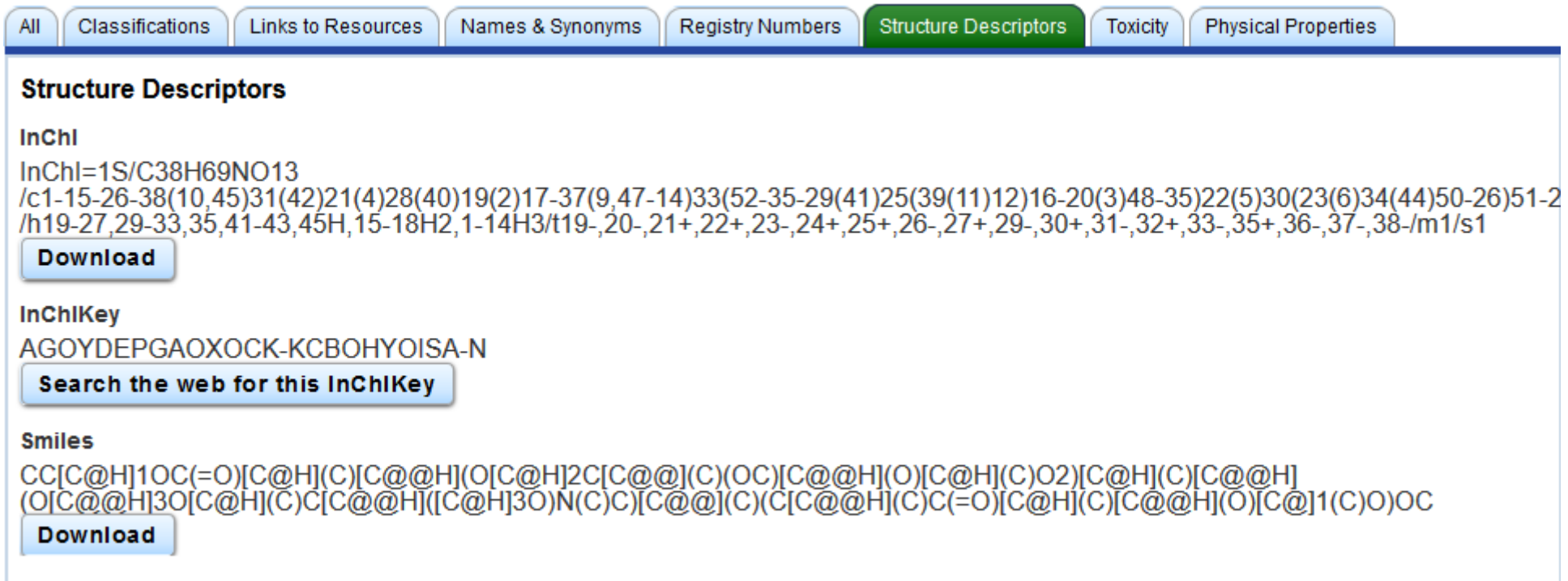
An example of use of SMILES: transfer of the structure of a bio-active compound from a data base to your text or presentation

- open a data base – e.g. ChemId Plus at https://chem.nlm.nih.gov/chemidplus/
- search for a drug by means of INN name: e.g. clarithromycin

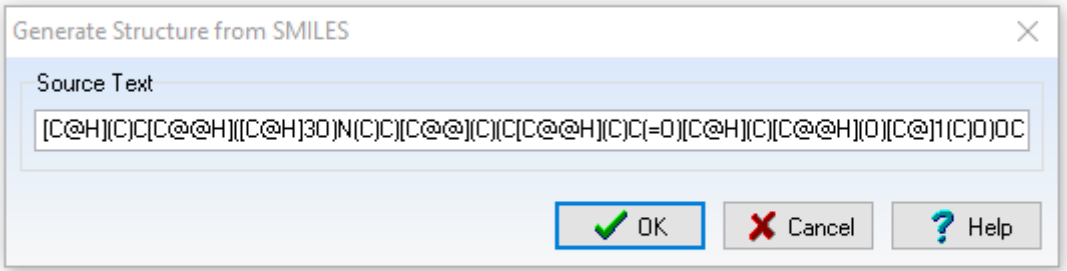
- now, you´ve got the record for clarithromycin
- if you want further work with the structure, you can either enlarge it and copy as the picture, or, better, you can choose Structure Descriptors



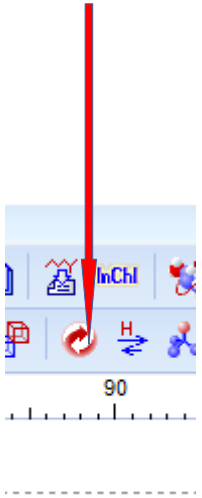
- now, you can see some structure codes
- use the Smiles
  - you can download and save it as ASCII test file, or simply copy it

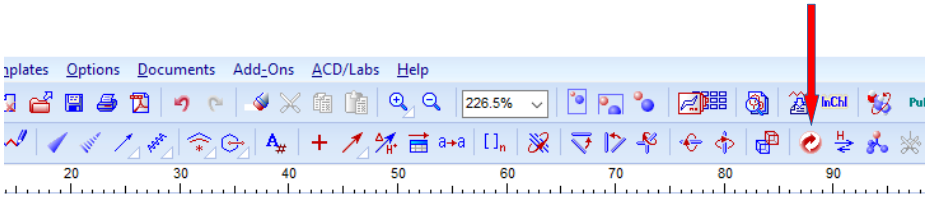


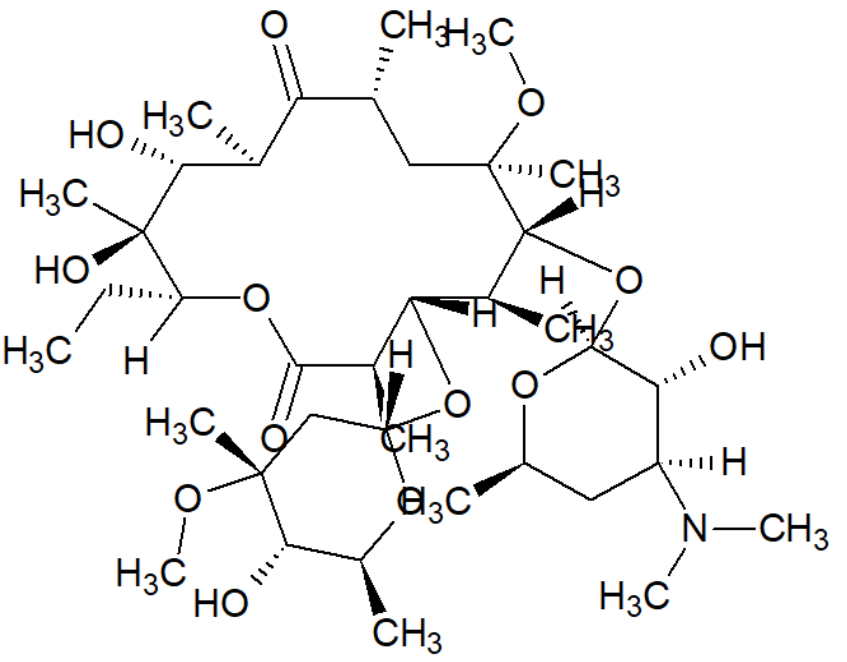
- now, you can paste it into the appropriate input raw in ChemSketch *(you must have it already opened either from an text/presentation editor by OLE function or independently)*:
  - Tools → Generate → Structure from Smiles



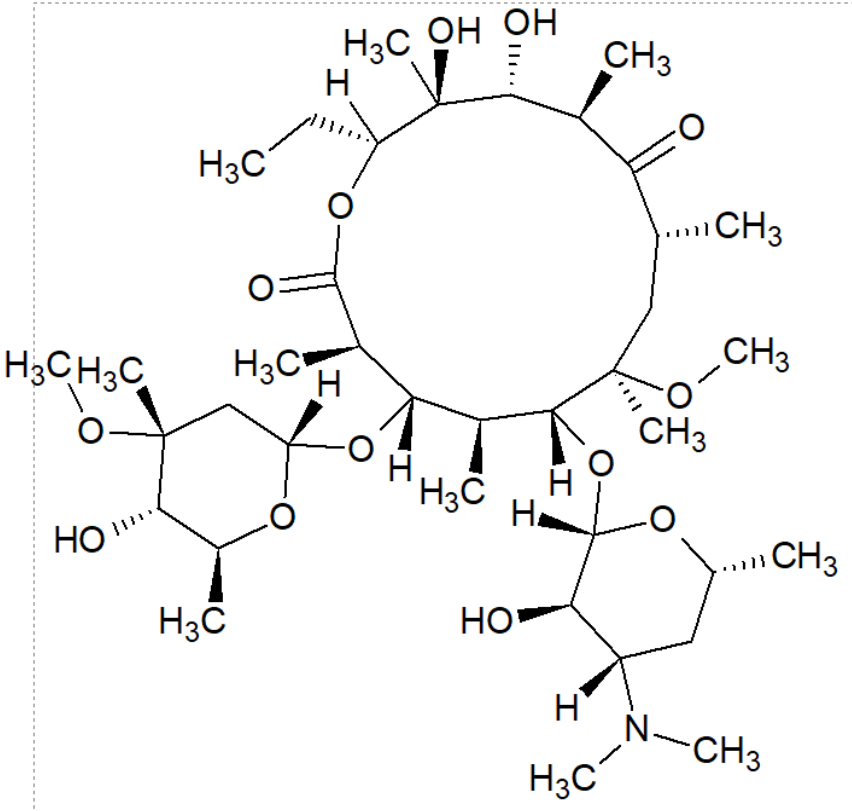
- now, the structrure of the antibiotic has appeared; but it requires to „Clean Structure“





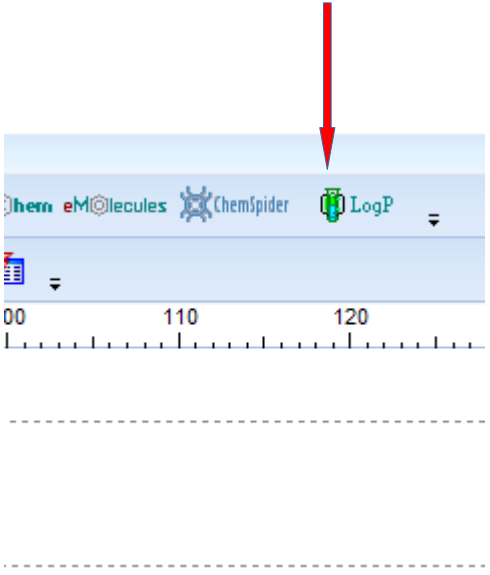


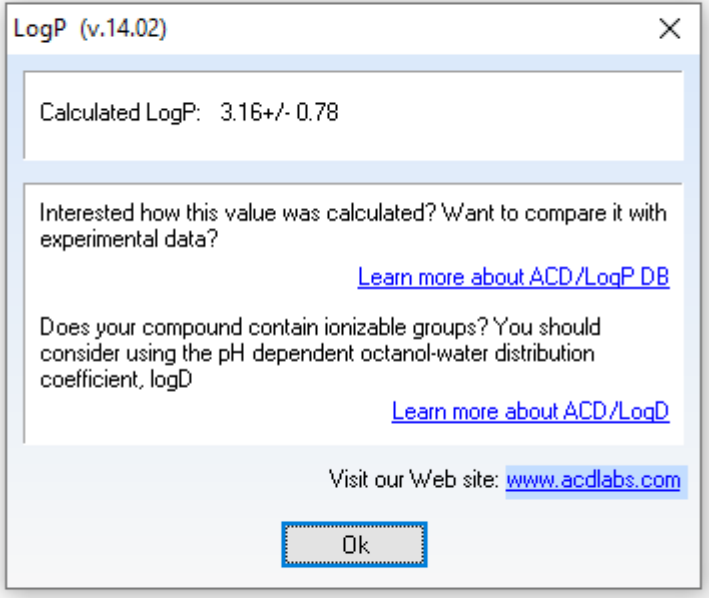
- … it´s better; you can try it more times
  - you can also try to Flip Top to Bottom or Flip Left to Right or select and rotate whole molecule either 2D or 3D
  - *but this all is only improvement of the appearance of the formula, not a structure optimization by means of a molecular mechanics*



Further use of the molecule structure imported via SMILES code

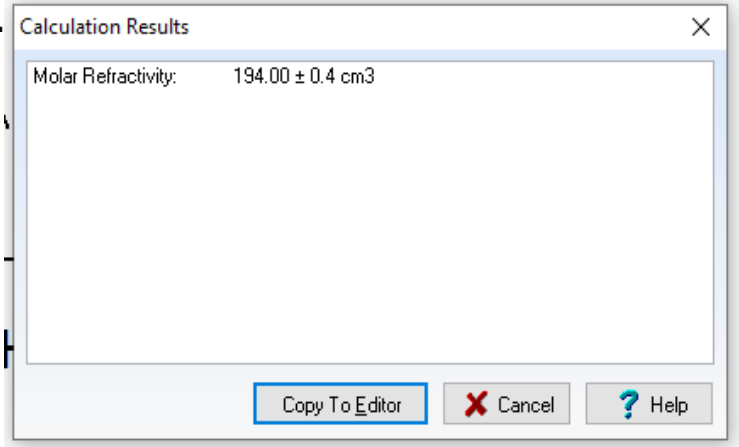
- calculation of estimated physico-chemical properties:
- **log P** ($\Rightarrow$ lipophilicity)
  - the green rightmost symbol





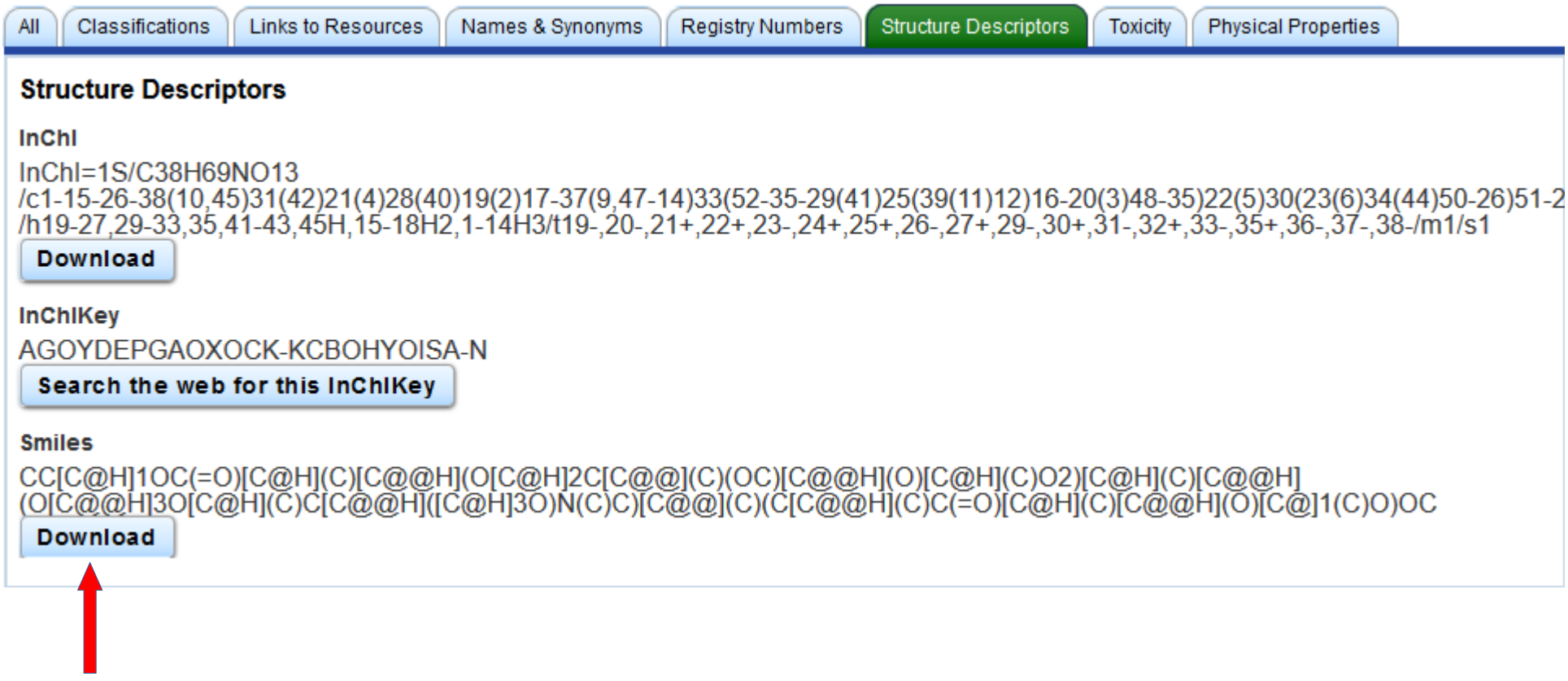
Further use of the molecule structure imported via SMILES code

- calculation of other estimated physico-chemical properties:
  - useful for QSAR calculations:
    - molar refractivity
    - molar volume
    - parachor
    - index of refractivity
    - density
    - dielectric constant
    - polarizability
  - Tools → Calculate → Molar refractivity …

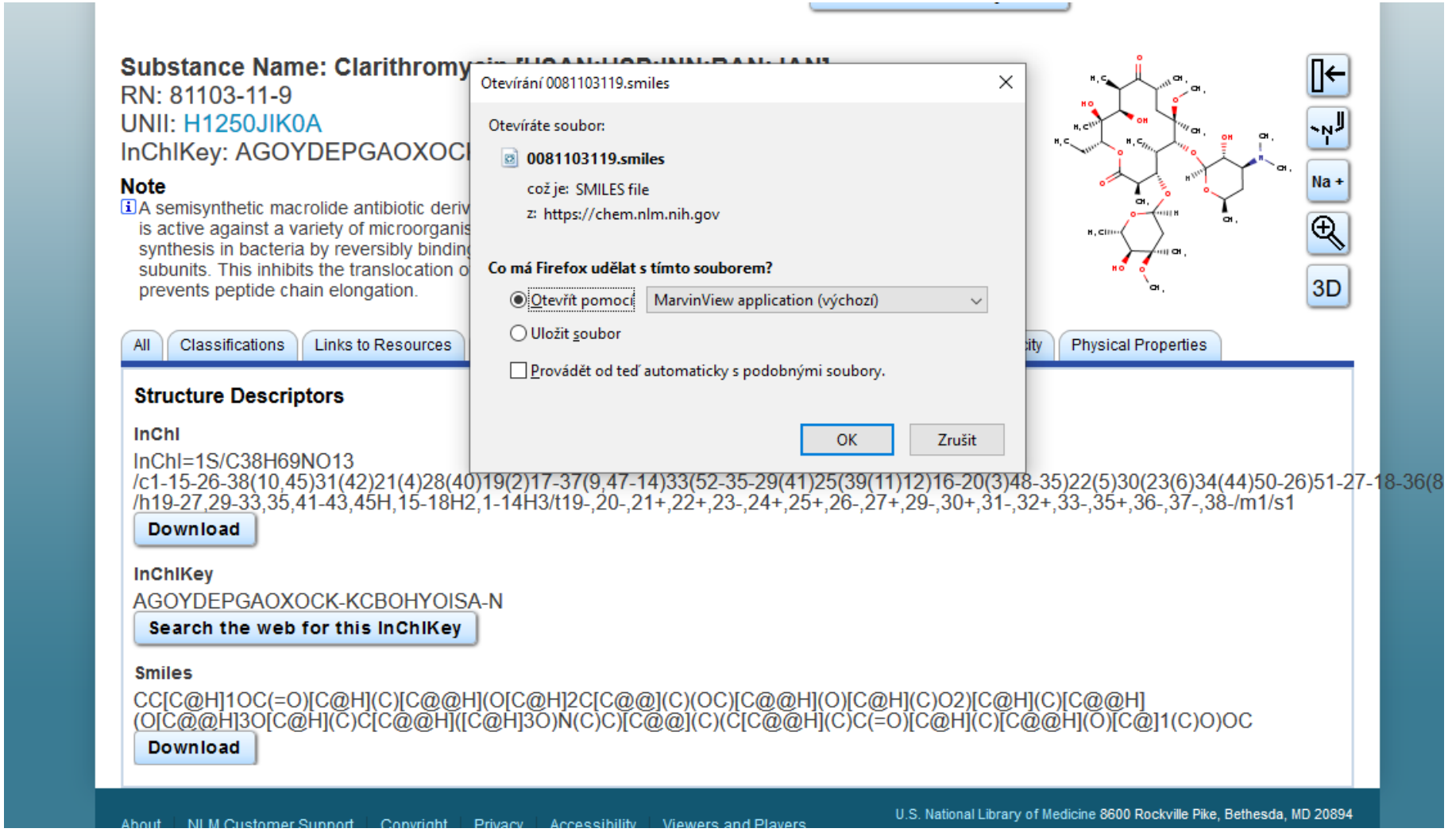


How to proceed with SMILES if you have Marvin installed

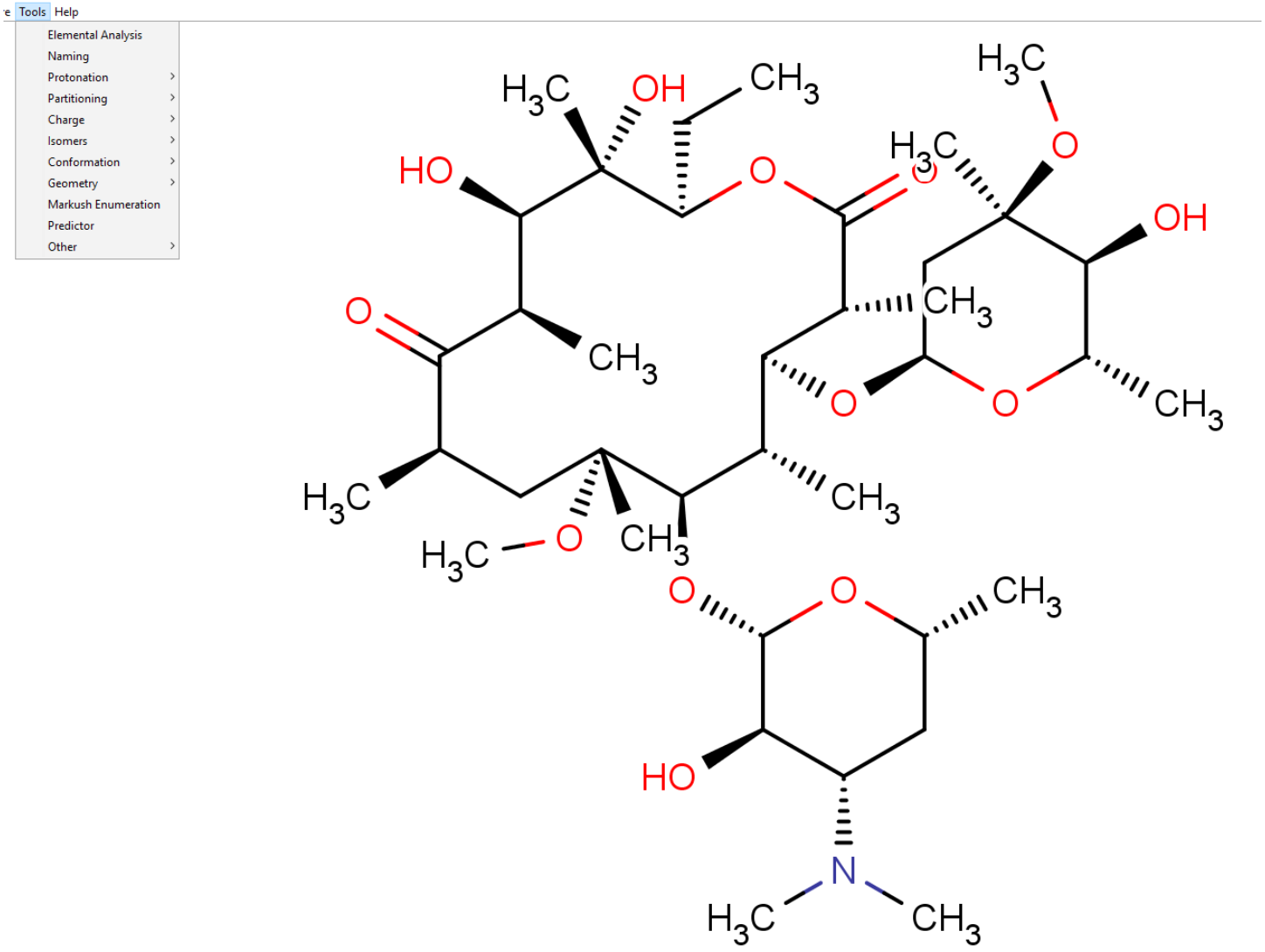
- simply click on Download



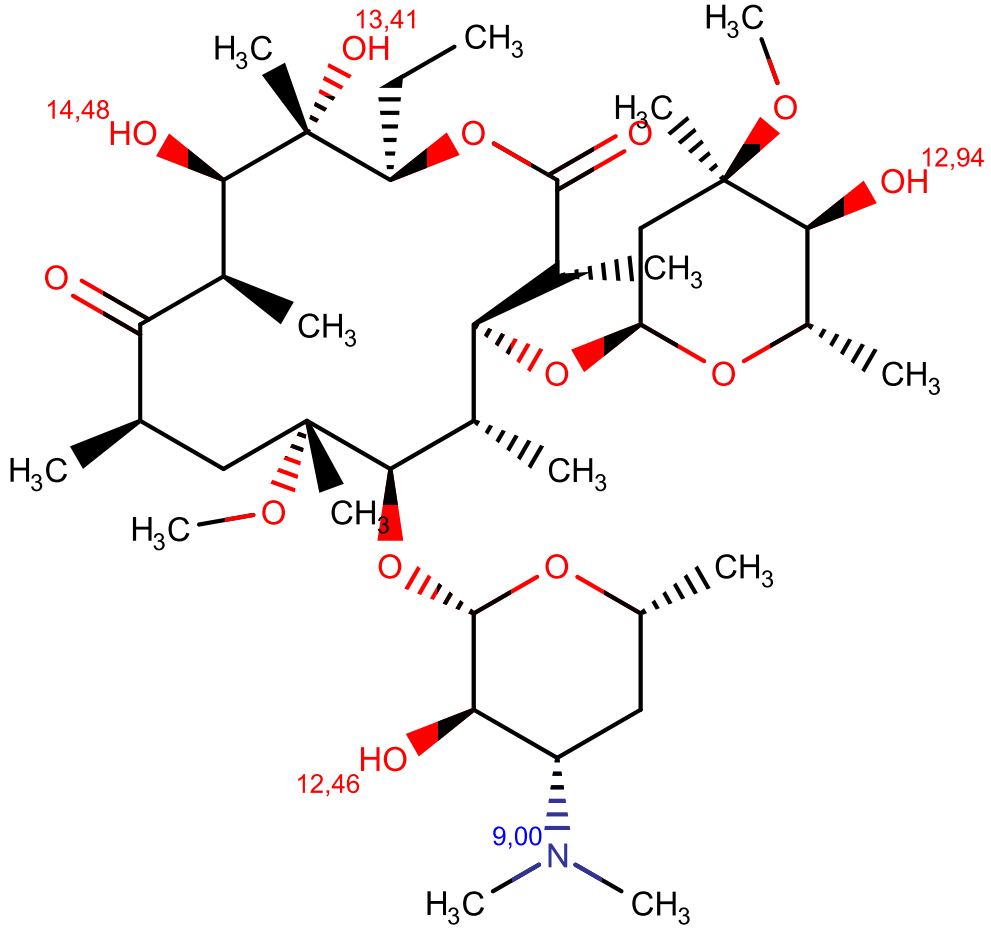
- Marvin directly opens SMILES as the structure

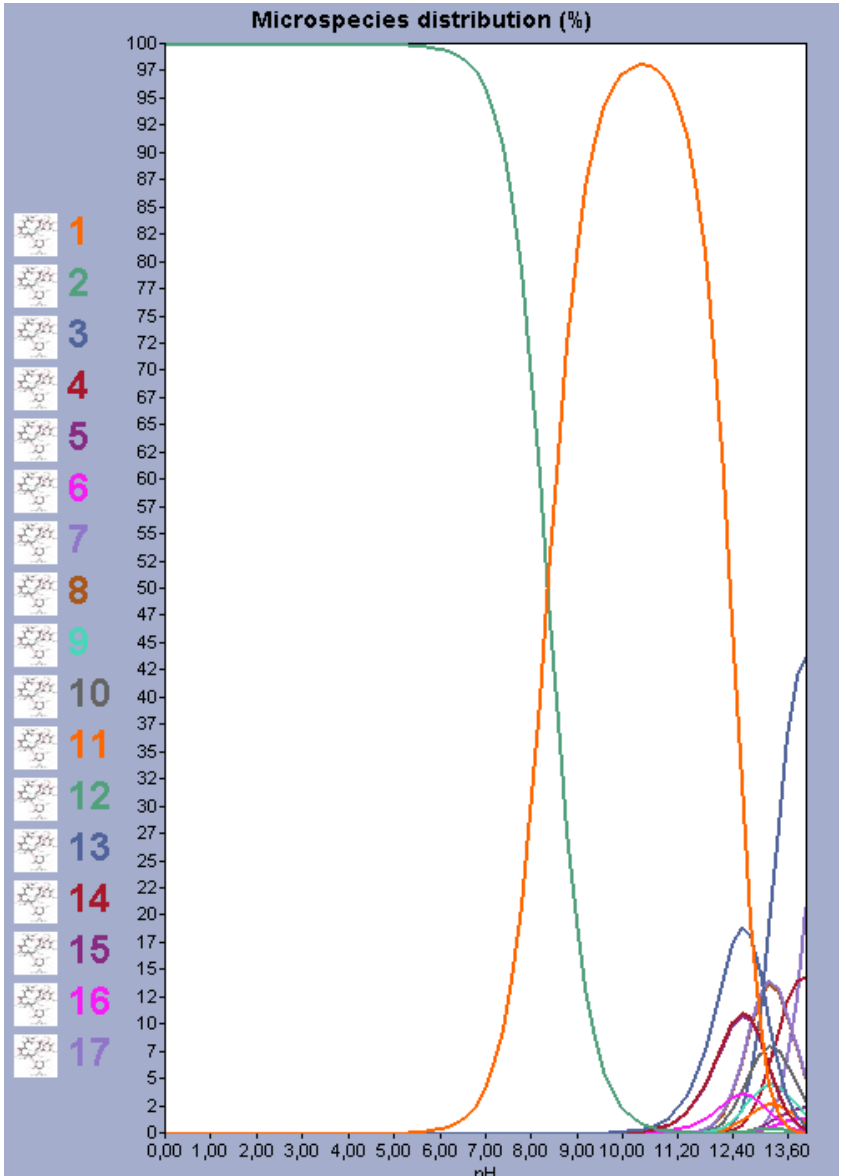


- further, you can get values, estimated by a calculation, for properties such as:
  - pK_a
  - isoelectric point (for amphoters)
  - log P (different algorithms to select)
  - log D
  - geometrical descriptors
    - polar surface area
    - molecular surface area
    - van der Waals volume

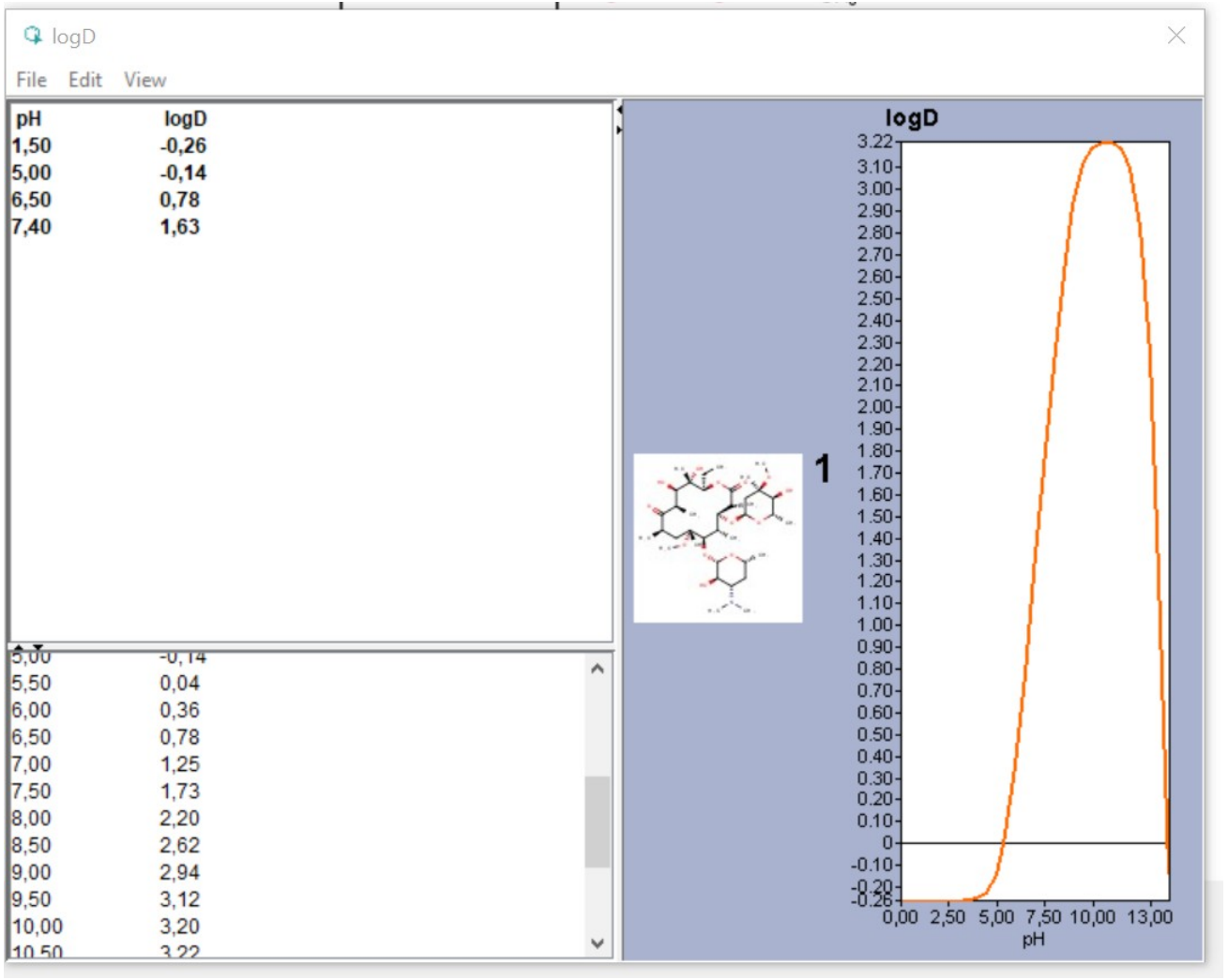


Main results of pKa estimation: structure formula with pKa values and plot of distribution of microspecies in dependence on pH

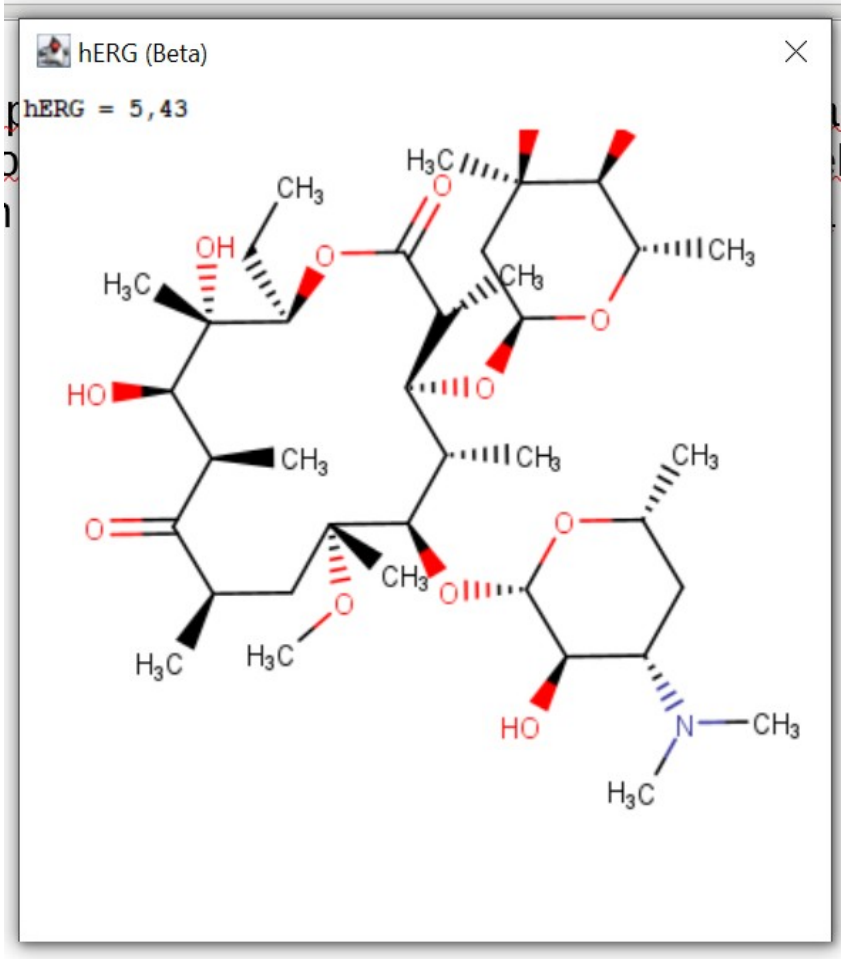




Results of log D calculation in Marvin



From ADMET properties, Marvin can estimate the hERG parameter, characterizing the affinity of a compound to a principal subunit of a K+ channel and thus the ability to prolong the QT interval on ECG and thus to cause a tachyarrhytmia of *torsade de point* type



InChI coding

InChI = International Chemical Identifier

- Initially developed by IUPAC (International Union of Pure and Applied Chemistry) and NIST (National Institute of Standards and Technology) from 2000 to 2005, the format and algorithms are non-proprietary.
- since supported 2010 by the not-for-profit InChI Trust, of which IUPAC is a member.

Format and layers

- Main layer
  - Chemical (empirical) formula (no prefix). This is the only sublayer that must occur in every InChI.
  - Atom connections (prefix: "c"). The atoms in the chemical formula (except for hydrogens) are numbered in sequence; this sublayer describes which atoms are connected by bonds to which other ones.
  - Hydrogen atoms (prefix: "h"). Describes how many hydrogen atoms are connected to each of the other atoms.
- Charge layer
  - charge sublayer (prefix: "q")
  - proton sublayer (prefix: "p" for "protons")

InChI coding (continued)

Stereochemical layer

- double bonds and cumulenes (prefix: "b")
- tetrahedral stereochemistry of atoms and allenes (prefixes: "t", "m")
- type of stereochemistry information (prefix: "s")

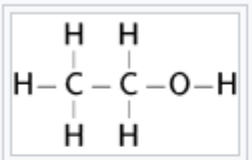
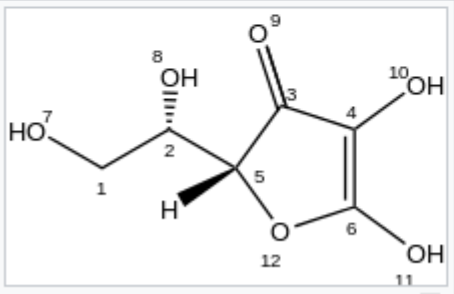
Isotopic layer (prefixes: "i", "h", as well as "b", "t", "m", "s" for isotopic stereochemistry)

Fixed-H layer (prefix: "f"); contains some or all of the above types of layers except atom connections; may end with "o" sublayer; never included in standard InChI

Reconnected layer (prefix: "r"); contains the whole InChI of a structure with reconnected metal atoms; never included in standard InChI

Simple examples of the standard InChI coding

Examples

Structural formula	standard InChI
 ethanol	`InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3`
 L-ascorbic acid	`InChI=1S/C6H8O6/c7-1-2(8)5-3(9)4(10)6(11)12-5/h2,5,7-8,10-11H,1H2/t2-,5+/m0/s1`

InChI Key

= 27 character condensed, „hashed“ version of InChI

- three parts separated by hyphens, of 14, 10 and one character(s), respectively, like XXXXXXXXXXXXX-YYYYYYYYFV-P
- the first 14 characters result from a SHA-256 hash of the connectivity information (the main layer and /q sublayer of the charge layer) of the InChI.
- the second part consists of 8 characters resulting from a hash of the remaining layers of the InChI, a single character indicating the kind of InChIKey (S for standard and N for nonstandard), and a character indicating the version of InChI used (currently A for version 1.)
- finally, the single character at the end indicates the protonation of the core parent structure, corresponding to the /p sublayer of the charge layer (N for no protonation, O, P, ... if protons should be added and M, L, ... if they should be removed.)

Use of InChI Key

- a fast way to find articles about the structure on the web, mainly patents

An example: **clarithromycin** at ChemId again

Structure Descriptors

InChI

InChI=1S/C38H69NO13
/c1-15-26-38(10,45)31(42)21(4)28(40)19(2)17-37(9,47-14)33(52-35-29(41)25(39(11)12)16-20(3)48-35)22(5)30(23(6)34(44)50-26)51-2
/h19-27,29-33,35,41-43,45H,15-18H2,1-14H3/t19-,20-,21+,22+,23-,24+,25+,26-,27+,29-,30+,31-,32+,33-,35+,36-,37-,38-/m1/s1

Download

InChIKey

AGOYDEPGAOXOCK-KCBOHYOISA-N

Search the web for this InChIKey

Smiles

CC[C@H]1OC(=O)[C@H](C)[C@@H](O[C@H]2C[C@@](C)(OC)[C@@H](O)[C@H](C)O2)[C@H](C)[C@@H]
(O[C@@H]3O[C@H](C)C[C@@H]([C@H]3O)N(C)C)[C@@](C)(C[C@@H](C)C(=O)[C@H](C)[C@@H](O)[C@]1(C)O)OC

Download