Subject: Stereoisomers conformer generation and Inchikeys
Posted by mattiafelice.palermo on Tue, 31 Aug 2021 15:08:34 GMT
View Forum Message <> Reply to Message

Dear Thomas,

I'm not 100% sure this is a bug as I am not particularly educated about InChIKeys generation and I am a bit rusty regarding stereochemistry.

I have calculated the lowest energy conformers for a molecule with two cis/trans stereocenters (limiting to 1 conformer per stereoisomer):

Datawarrior automatically generates the 3D geometry for four stereoisomers. Then I have generated the InChIKeys for the four molecules and Datawarrior returns the same key, RUSZGEIGALJIOH-UHFFFAOYSA-N, for every molecule (see figure).

I have tried generating the InChIKeys with RDKit in Python and instead, I have obtained the following:

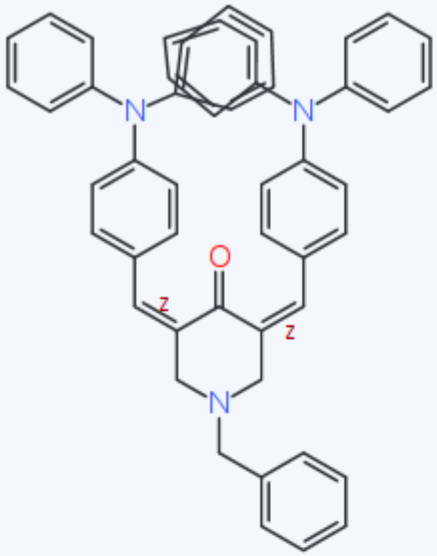
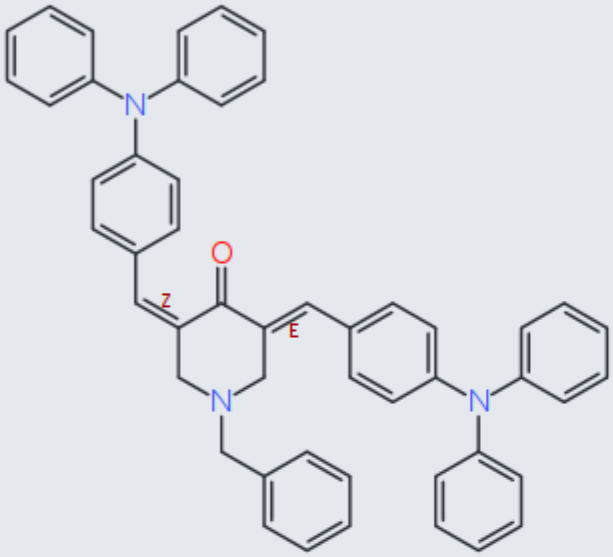
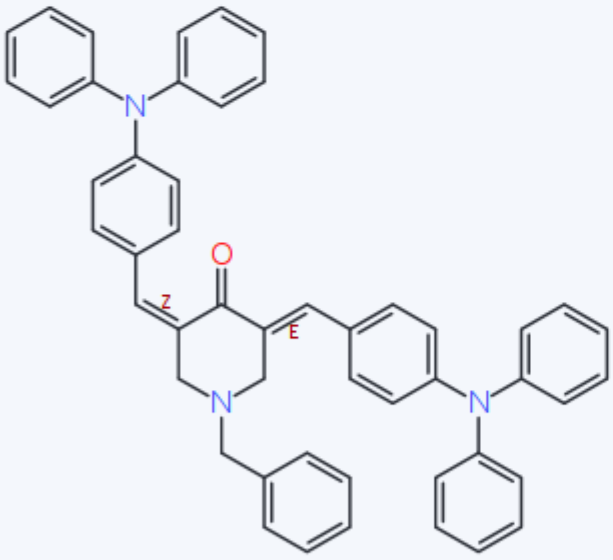
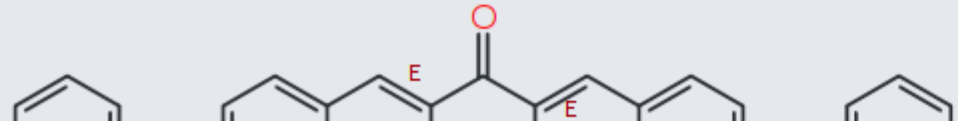
ID | InChI-Key | Energy
1) RUSZGEIGALJIOH-MRSMTKAOSA-N 156.35
2) RUSZGEIGALJIOH-JYYJLMHASA-N 155.66
3) RUSZGEIGALJIOH-JYYJLMHASA-N 155.64
4) RUSZGEIGALJIOH-LMKHWIEJSA-N 148.88

I have two questions:
Which InChIKeys are the correct ones, DataWarrior or RDKit?
Aren't molecules 2 and 3 (E, Z and Z, E) the same stereoisomer in this particular case? RDKit InChIKeys seem to confirm that. If that is the case, perhaps DW conformer routine should only keep one of the two, or is this an intended behaviour?

Thank you for your help!

Mattia

File Attachments
1) stereoisomers.png, downloaded 2281 times

	Structure	ID	Stereo Isomer
1		665	4
2		665	2
3		665	3
		665	1

Subject: Re: Stereoisomers conformer generation and Inchikeys
Posted by nbehrnd on Thu, 02 Sep 2021 20:18:30 GMT
View Forum Message <> Reply to Message

Dear Mattia,

based on visual inspection of the structure formulae, entries #2 and #3 depict the same isomer. Thus, there are three different compounds in the table for which one anticipates three different SMILES strings, InChis, or eventually InChiKeys.

Based on the visual representation, I redrew the structures in PerkinElmer's ChemDraw test page.[1] This page equally permits the export in SMILES, InChi, and InChiKey format. There is no match of the later with those you report as generated by DW. The ChemDraw InChiKeys match the ones you report by RDKit. It is possible to process ChemDraw's SMILES strings into InChiKeys by OpenBabel, too, yet with the same result; match with RDKit/ChemDraw, no match with DW.

Importing the three different ChemDraw SMILES strings into DW yields different structure representations. However already the InChi for these match each other. Since the InChikeys only are a hash of the InChi, it is not surprising that these derivates equally match each other (again). If interested, compare with attached .dwar and Emacs .org file.

Based on these observations, pending correction of this local problem for DW, I would suggest to use the InChiKeys by RDKit. On occasion, implementation of the underlying rules indeed may cause problems which remain unidentified for quite some time until comparison with other programs.[2]

Norwid

[1] https://chemdrawdirect.perkinelmer.cloud/js/sample/index.htm l#
[2] https://mattermodeling.stackexchange.com/questions/6460/rdki
t-and-pysmiles-results-differ-on-some-smiles-strings

Addition:

Since DW is able to generate a random library of molecules and to assign for these SMILES and InChiKeys, I wrote a little Python script to compare InChiKeys by DW with those generated by OpenBabel. (This only is low level of concept/doodle only.) For two runs (10 and 250 molecules), the ratio of SMILES for which both programs assign different InChiKeys over the grand total of structures submitted is a little bit higher, than anticipated. Perhaps a complementary check with RDKit is useful before Thomas steps in.

File Attachments

1) 2021-09-02_InChi_test.dwar, downloaded 760 times
2) 2021-09-02_DW.org, downloaded 799 times
3) dissenter_check.zip, downloaded 628 times

Subject: Re: Stereoisomers conformer generation and Inchikeys
Posted by nbehrnd on Tue, 14 Sep 2021 07:01:37 GMT
View Forum Message <> Reply to Message

An improvement is on the way. Based on a separate exchange, Thomas mentioned the assignments of InChI/InChiKey provided by DataWarrior are revised and, where necessary, are going to be corrected.

For the moment, alternatives may be the online conversion of individual of structures,[1] or one of the reference implementations distributed by the InChI Trust[2] which, by a command like

./inchi-1 test.sdf -AuxNone -Tabbed -Key

write InChI and InChIKeys of structures in test.sdf into a tabulator separated .txt file.

Norwid

[1], examples
http://www.cheminfo.org/Chemistry/Cheminformatics/Generate_I nChI/index.html
https://chemdrawdirect.perkinelmer.cloud/js/sample/index.htm l# (Structure -> Get InChI)

[2] https://www.inchi-trust.org/downloads/

Subject: Re: Stereoisomers conformer generation and Inchikeys
Posted by thomas on Mon, 20 Sep 2021 22:39:41 GMT
View Forum Message <> Reply to Message

Many thanks to Norwid and Mattia for pointing out the discrepancies of created Inchis and Inchi-Keys with the ones produced by OpenBabel and also for providing a test data set. The issues were mainly caused by a bug in the conversion of the tetrahedral stereo configuration for the jni-inchi (1.0.3) library from 2010 that DataWarrior used to create inchis. During the recent days DataWarrior was updated to using a new open-source project jna-inchi from Daniel Lowe, which is based on the inchi-trust code 1.0.6 from 2020. The stereo issues have also been fixed. The current development update contains all fixes and a new jniinchi.jar (which actually is the jnainchi.jar). The dev update link appears on the download page after checking the 'Read and Understood' checkbox.

Subject: Re: Stereoisomers conformer generation and Inchikeys
Posted by nbehrnd on Sun, 31 Oct 2021 19:03:32 GMT
View Forum Message <> Reply to Message

It should be noted that tautomers share standard an InChI and standard InChIKey in common.

One FAQ by the InChI trust (though by 2012-05-12) explains this by

«different tautomers have the same connectivity/hydrogen layer»[1]

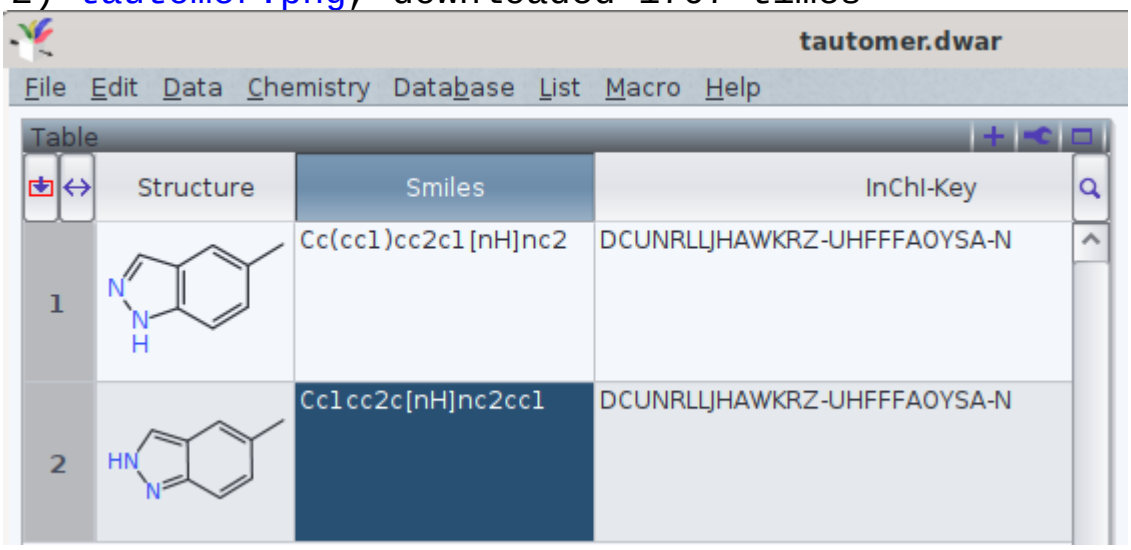
With the current version of InChI executables (1.06, December 2020), it is possible e.g., to add what InChI defines as «Fixed H layer», which offers to discern the tautomers. However, currently this is may yield a non-standard InChI (i.e., the string does not start with InChI=1S/, but InChI=1/). Thus, their use in a search in databases[2] may refer to more than one chemical entity.

Norwid

[1] https://www.inchi-trust.org/technical-faq-2/
[2] «Many InChIs and quite some feat» by Warr, W.A., J. Comput. Aided Mol. Des., 2015, 29: 681. https://doi.org/10.1007/s10822-015-9854-3

File Attachments

1) tautomer.dwar, downloaded 698 times
2) tautomer.png, downloaded 1707 times



Subject: Re: Stereoisomers conformer generation and Inchikeys
Posted by nbehrnd on Tue, 12 Apr 2022 07:32:34 GMT
View Forum Message <> Reply to Message

On a tangent for those curious about InChI:

On April 5th/6th 2022, InChI trust organized a public zoom meeting to showcase existing applications and ongoing developments with/of InChI including topics like Markush formulae, tautomerism, polymers, nanoparticles, trading with mixtures and reactions as well as the «precision FDA contest» anticipated for fall 2022. Since today, the .pdf (total of 16, about 31 MB in volume) are mirrored on

https://www.inchi-trust.org/decks-open-inchi-days-2022/

Norwid