Co-funded by the Prevention of and Fight
against Crime Programme of the European Union

Substance descriptors: structure, common name, systematic name, systematic chemical identifiers and their consistency

EUROPEAN PROJECT RESPONSE
to challenges in forensic drugs analyses

Dr. Sonja Klemenc
E-Mail: sonja.klemenc@policija.si
Presented by: Denis Saboti
MNZ GPU, National Forensic Laboratory, Vodovodna 95, 1000 Ljubljana, Slovenia

NFL NACIONALNI FORENZIČNI LABORATORIJ
NATIONAL FORENSIC LABORATORY

INTRODUCTION

In the framework of the EU co-funded project RESPONSE over 500 NPSs (test purchased, collected and seized samples as well as reference materials) were chemically characterized in the last two years. The p**ublic open database** (NPS and related compounds) is accessible here:

http://www.policija.si/apps/nfl_response_web/seznam.php

Database description and guidelines for use you can find here:

draft
http://www.policija.si/eng/images/stories/GPUNFL/PDF/2016/DrugsMonographsDatabase-DescriptionAndGuidelines-draft.pdf
new versions will be published here:
http://www.policija.si/eng/index.php/generalpolicedirectorate/1669-nfl-page-response

This talk focuses on one single element of the RESPONSE project database:

SUBSTANCE DESCRIPTORS and some chem-informatic tools applied ..
..and what we have learned about this topic (so far ;-)…

SUBSTANCE DESRIPTORS/ IDENTIFIERS:

REQUREMENTS

RESPONSE developed the idea that each structure should be:

- described by several generally accepted descriptors in the NPS field investigations
  - structure (picture -primary descriptor)
  - structure related descriptor – IUPAC name (English only)
  - structure encoding that machines and humans can understand and/or compare – i.e. systematic chemical identifier should be given and it should be "unique", if possible
  - other descriptors (common names and/or synonyms)…

The "unique" descriptor should:

- enhance interoperability options with other public open databases
- be "googlable" (it can serve as self-updating info tool – i. e. new info about particular compound can be reached in real time)

Validation of key identifiers (structure-name-structure) should be performed by chem-informatic tools.

SUBSTANCE DESRIPTORS/ IDENTIFIERS:

Applied

- Structure (primary descriptor)
- Common name
- Sytematic name (IUPAC)
- Other names (IUPAC based and/or synonymes, common names at the time of compound entry)
- Std InChI Key (Standardized International Chemical Identifier Key)
- Molecular formula and weight (not discused further)

RESPONSE — NPS AND RELATED COMPOUNDS - ANALYTICAL REPORTS

Co-funded by the Prevention of and Fight against Crime Programme of the European Union

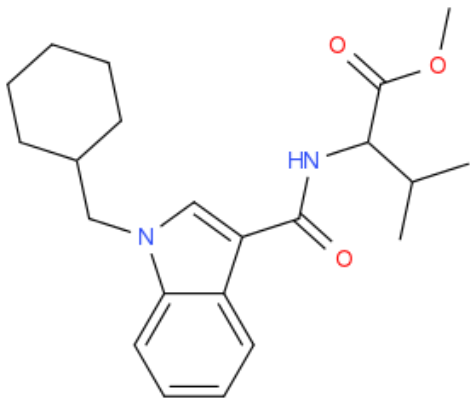
substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	Mw (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChIKey	Type of detection	comments
AMB-CHMICA		methyl 2-[[1-(cyclohexylmethyl)-1H-indol-3-yl]formamido]-3-methylbutanoate	methyl (1-(cyclohexylmethyl)-1H-indole-3-carbonyl)valinate ; methyl 2-(1-(cyclohexylmethyl)-1H-indole-3-carbonylamino)-3-methylbutanoate ; MMB-CHMICA Cayman (L-stereoisomere defined)	C22H30N2O3	370.48	240	256	144	13.53	show	ROWZIXRLVUOMCJ-UHFFFAOYSA-N	test purchase	purchased as MA-CHMINACA; and another one as MAB-CHMINACA
MMB-CHMICA		methyl 2-[[1-(cyclohexylmethyl)-1H-indol-3-yl]formamido]-3-methylbutanoate	AMB-CHMICA; methyl (1-(cyclohexylmethyl)-1H-indole-3-carbonyl)valinate ; methyl 2-(1-(cyclohexylmethyl)-1H-indole-3-carbonylamino)-3-methylbutanoate ;	C22H30N2O3	370.49	240	256	144	13.48	show	ROWZIXRLVUOMCJ-UHFFFAOYSA-N	RM-reference material	

Fig.: A part of database window (several columns and two rows only)

SUBSTANCE DESRIPTORS/ IDENTIFIERS -EXAMPLE:

Structure – primary descriptor

Substance is identified only when its structure is defined!



Structure: picture in .png (Portable Network Graphics) format

<u>In the RESPONSE database structure is given</u>:

- ❑ for base form of compound
- ❑ for non charged state of compound
- ❑ stereochemistry (cis , trans; E, Z, S,R , tautomerism)

is not taken into account (with some exceptions for reference materials - this decision is currently under re-consideration)

Structures of small molecules are easily recognizable and compared by humans and are understandable regardless of the language used.

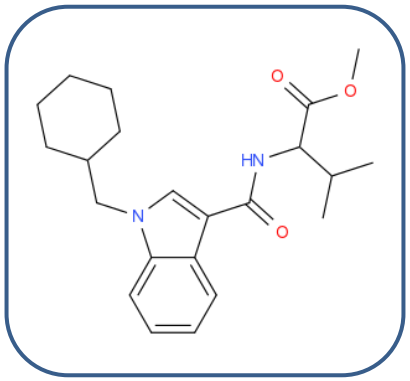
Not very useful for oral/text communications.

SUBSTANCE DESRIPTORS/ IDENTIFIERS -EXAMPLES:

Common names and synonyms - widely used , practical, but..

NOT UNIQUE and not necessarily related to the structure, low power for intra and interoperability of databases.

AMB-CHMICA



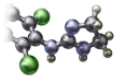
MMB-CHMICA

European Monitoring Centre for Drugs and Drug Addiction

EDND – European information system and database on new drugs

Designer Drugs 2017

New Synthetic Drugs Database



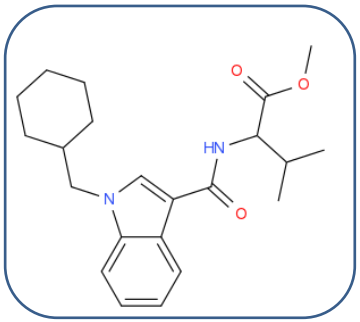
ChemIDplus
A TOXNET DATABASE
Lite • Browse • Advanced

PubChem OPEN CHEMISTRY DATABASE

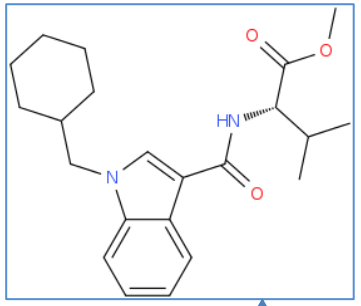
Cayman CHEMICAL

SUBSTANCE DESRIPTORS/ IDENTIFIERS - EXAMPLES:

Systematic names - IUPAC, structure related, but many variations..)



NOT UNIQUE
complex, language dependant, low power for intra and interoperability of databases. ..



methyl 2-[[1-(cyclohexylmethyl)indole-3-carbonyl]amino]-3-methyl-butanoate

methyl (1-(cyclohexylmethyl)-1H-indole-3-carbonyl)-L-valinate* (Cayman RM)
methyl (1-(cyclohexylmethyl)-1H-indole-3-carbonyl)-valinate*

methyl 2-{[1-(cyclohexylmethyl)-1H-indol-3-yl]formamido}-3-methylbutanoate

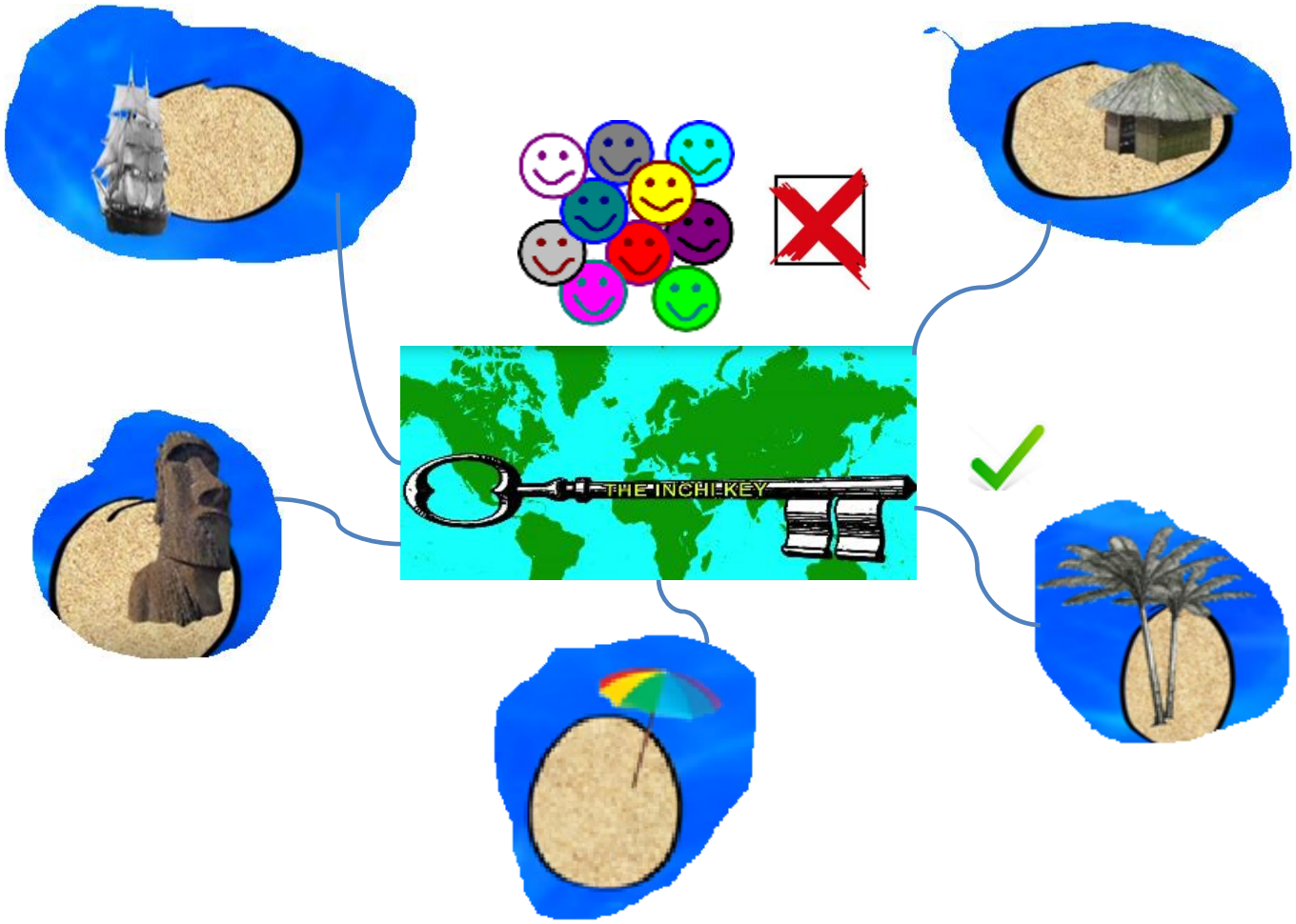
(Methyl 2-[[1-Cyclohexylmethyl)-1H-indol-3-yl]formamido]-3-methylbutanoate

ERROR- structure cannot be generated from this name – VALIDATION NEEDED!

https://www.chemograph.de/dd2017/dd2017-24483-366805.html

SUBSTANCE DESRIPTORS/ IDENTIFIERS - How to conect the islands?

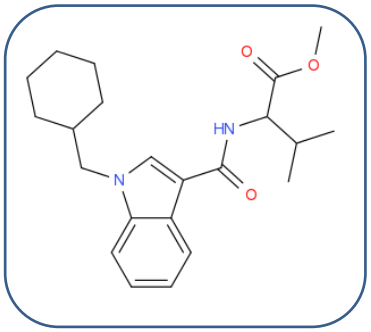
SMILES and Std InChI key have been explored



SUBSTANCE DESRIPTORS/ IDENTIFIERS - THE "UNIQUE DESCRIPTOR" :
SMILES explored

SMILES (Simplified molecular-input line-entry system, 1980) and canonical SMILES

At the begining RESPONSE used canonical SMILES, which are not unique. Because of this SMILES were replaced by Standard InChI Key in all records.



a) C1(CCCCC1)CN1C=C(C2=CC=CC=C12)C(=O)NC(C(=O)OC)C(C)C

b) COC(=O)C(NC(=O)C1=CN(CC2CCCCC2)C2=CC=CC=C12)C(C)C

c) O=C(OC)C(NC(=O)c3cn(CC1CCCCC1)c2ccccc23)C(C)C

ONE STRUCTURE

SMILES generated by different chem-informatic tools

Cheminformatic tools: a) OPSINE; b Marvin Sketch; c) ACD/Chem Sketch

InChI code and InChI key (in short)

InChI (string) is the International Chemical Identifier developed (2000) and its standardized version in 2008.

GOOD:

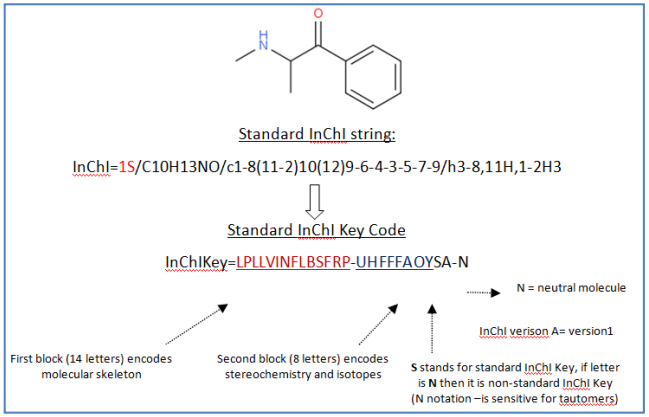
- InChI is produced from just the structural formula of a chemical substance.
- strict uniqueness of identifier - the same label always means the same substance, and the same substance always receives the same label (under the same labeling conditions – for the same level of structure details!).
- non-proprietary, open source code, free access.
- most cheminformatic tools can produce structure from the code

BAD:

- codes are of flexible lenght, sometimes very long
- google does not like such codes
- substructure search is not posible

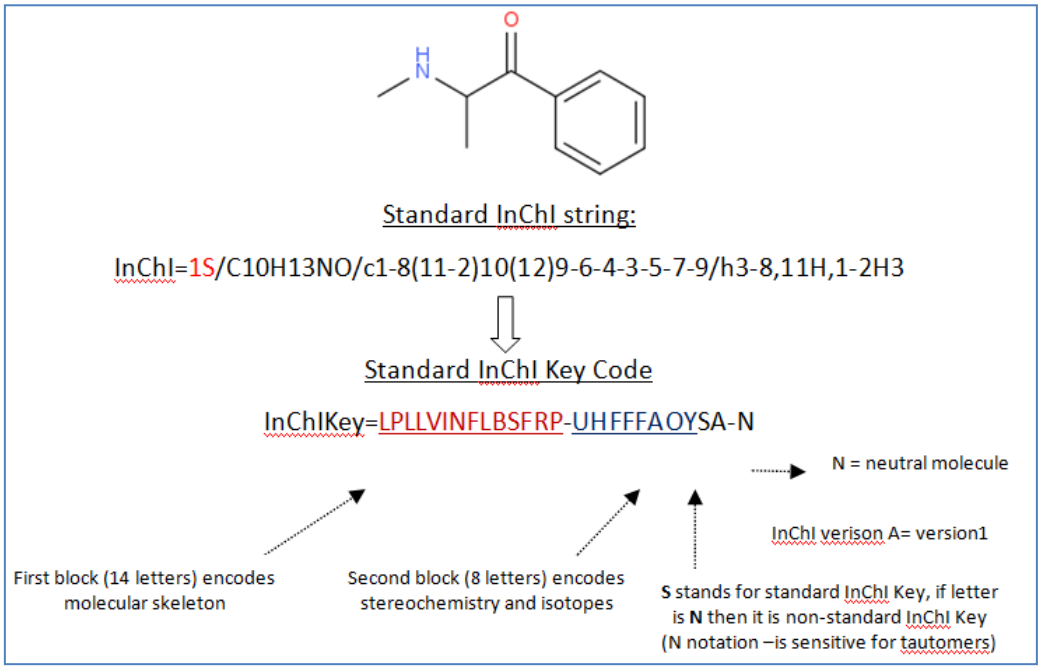
InChI key is a hashed version of InChI string

GOOD: googlable
BAD: structure can not be recreated, a little loose of the uniqunes



SUBSTANCE DESRIPTORS/ IDENTIFIERS - THE “UNIQUE DESCRIPTOR” in the RESPONSE :

STANDARD InChI key



More about InChI :

http://www.inchi-trust.org/technical-faq/#2.6

Standard InChI string and corresponding standard InChI Key – example for methcathinone (2-(methylamino)-1-phenyl-propan-1-one

Note: InChI has a layered structure which allows one to represent molecular structure with a desired level of detail (for example accounting for tautomerism or not). This flexibility, however, may sometimes appear to be a drawback, with respect to standardization and interoperability. So the Standard InChI was launched in 2008, in response to these concerns.

SUBSTANCE DESRIPTORS/ IDENTIFIERS - THE “UNIQUE” DESCRIPTOR :

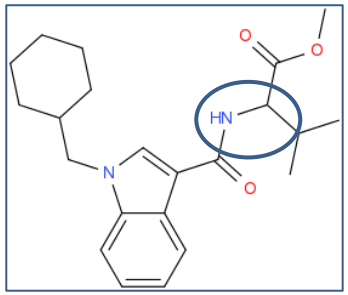
Standard InChI key explored - Example

The original InChI trust software (free software and free public open source code) was applied for Standard InChI key generation from the structure (sdf or mol format).

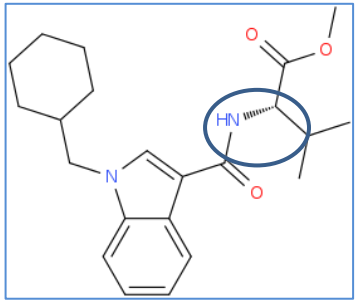
InChI TRUST http://www.inchi-trust.org/downloads/

*The same label always means the same substance, and the same substance always receives the same label (**under the same labeling conditions – for the same level of structure details!**).*

(a)



(b)



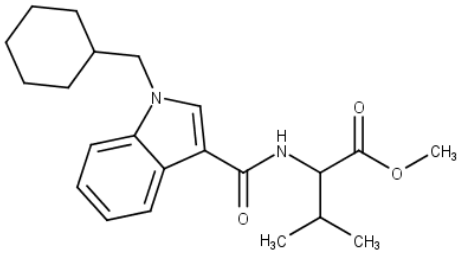
The same compound with different levels of structure details (a) and (b) gives different codes.

(a) ROWZIXRLVUOMCJ-UHFFFAOYSA-N

(b) ROWZIXRLVUOMCJ-FQEVSTJZSA-N

Structure – IUPAC name - structure and oposite: EXAMPLE

MarvinSketch : free (Chem-Axon) tool: converts structure to IUPAC name and oposite (RESPONSE adopted this cheminformatic tool)



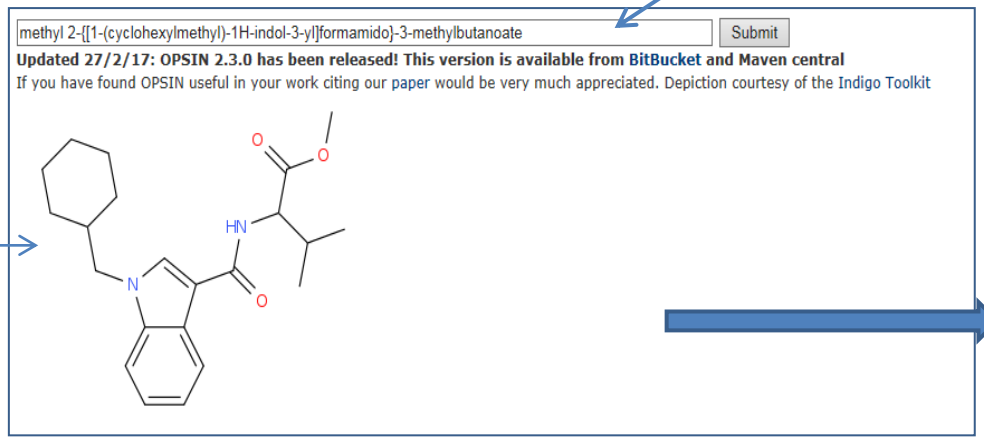
methyl 2-{[1-(cyclohexylmethyl)-1H-indol-3-yl]formamido}-3-methylbutanoate

Validation by OPSINE (free online tool):

http://opsin.ch.cam.ac.uk/

UNIVERSITY OF CAMBRIDGE — Centre for Molecular Informatics

OPSIN: Open Parser for Systematic IUPAC nomenclature



It also generates *Std InChI string*, *Std InChI key* and smiles.

SUBSTANCE DESRIPTORS/ IDENTIFIERS - Validation in RESPONSE:

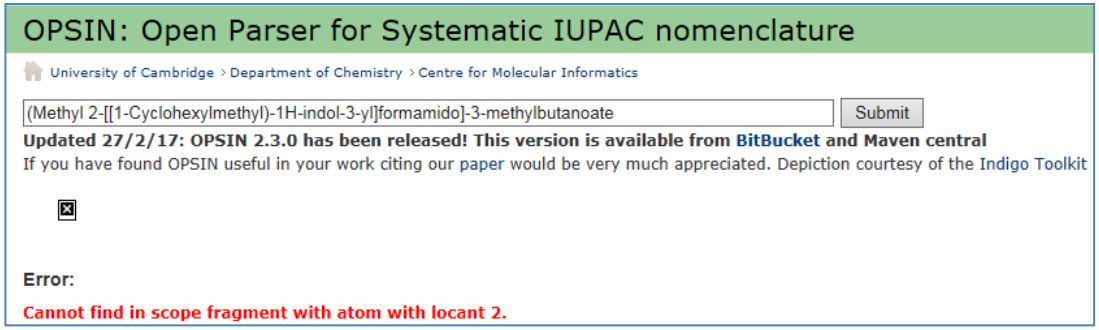
IUPAC name - structure by OPSINE: EXAMPLES

(Methyl 2-[[1-Cyclohexylmethyl)-1H-indol-3-yl]formamido]-3-methylbutanoate

ERROR- structure cannot be generated from this name – VALIDATION NEEDED!

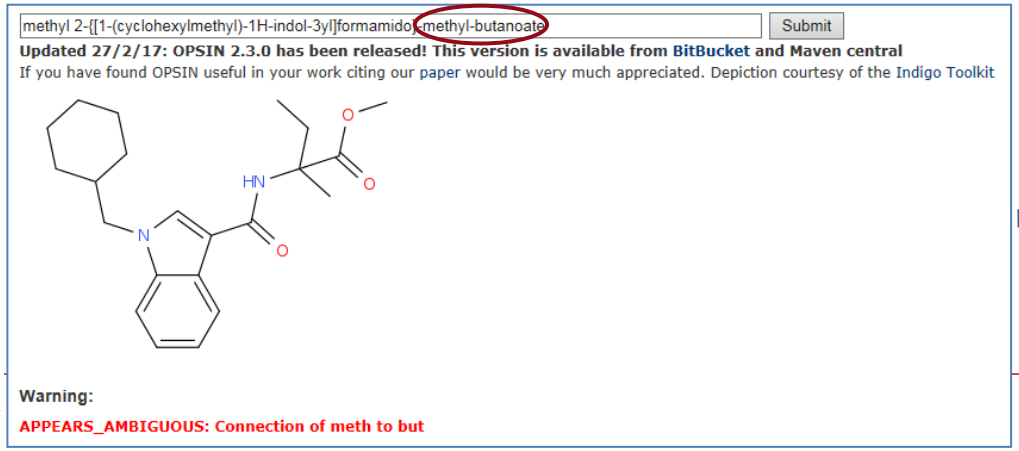
https://www.chemograph.de/dd2017/dd2017-24483-366805.html

OPSINE locate and comments a error



STRUCTURE IS NOT GIVEN

and/or AMBIGOUS names



STRUCTURE IS GENERATED, but with warning!

Std InChI string or key/ InChI string/Key – validation of some cheminformatic tools

The best way to generate InChI keys is the use of the original InChI trust software!

Anyhow, many other cheminformatic tools generate InChIs as well and we have validated this part a bit as well.

Some of our findings:

OPSINE: so far results were always consistent by those obtained by original software from mol file.

CACTUS (NCI/CADD Cheminformatic group), we observed some inconsistencies (example is substance W-19 (N-[(2Z)-1-[2-(4-aminophenyl)ethyl]piperidin-2-ylidene]-4-chlorobenzene-1-sulfonamide)

Marvin Sketch and Chemicalize (both from Chemaxon):
For some compounds they return InChi Key and and for the other Std InChI key. User has no influence on this.

SUBSTANCE DESRIPTORS/ IDENTIFIERS:
Validation in RESPONSE

Names and identifiers

Common names	methcathinone
IUPAC name	2-(methylamino)-1-phenylpropan-1-one
SMILES	CNC(C)C(=O)C1=CC=CC=C1
InChI	InChI=1/C10H13NO/c1-8(11-2)10(12)9-6-4-3-5-7-9/h3-8,11H,1-2H3
InChIKey	InChIKey=LPLLVINFLBSFRP-UHFFFAOYNA-N
CAS	28521-94-0,5650-44-2

A

Names and identifiers

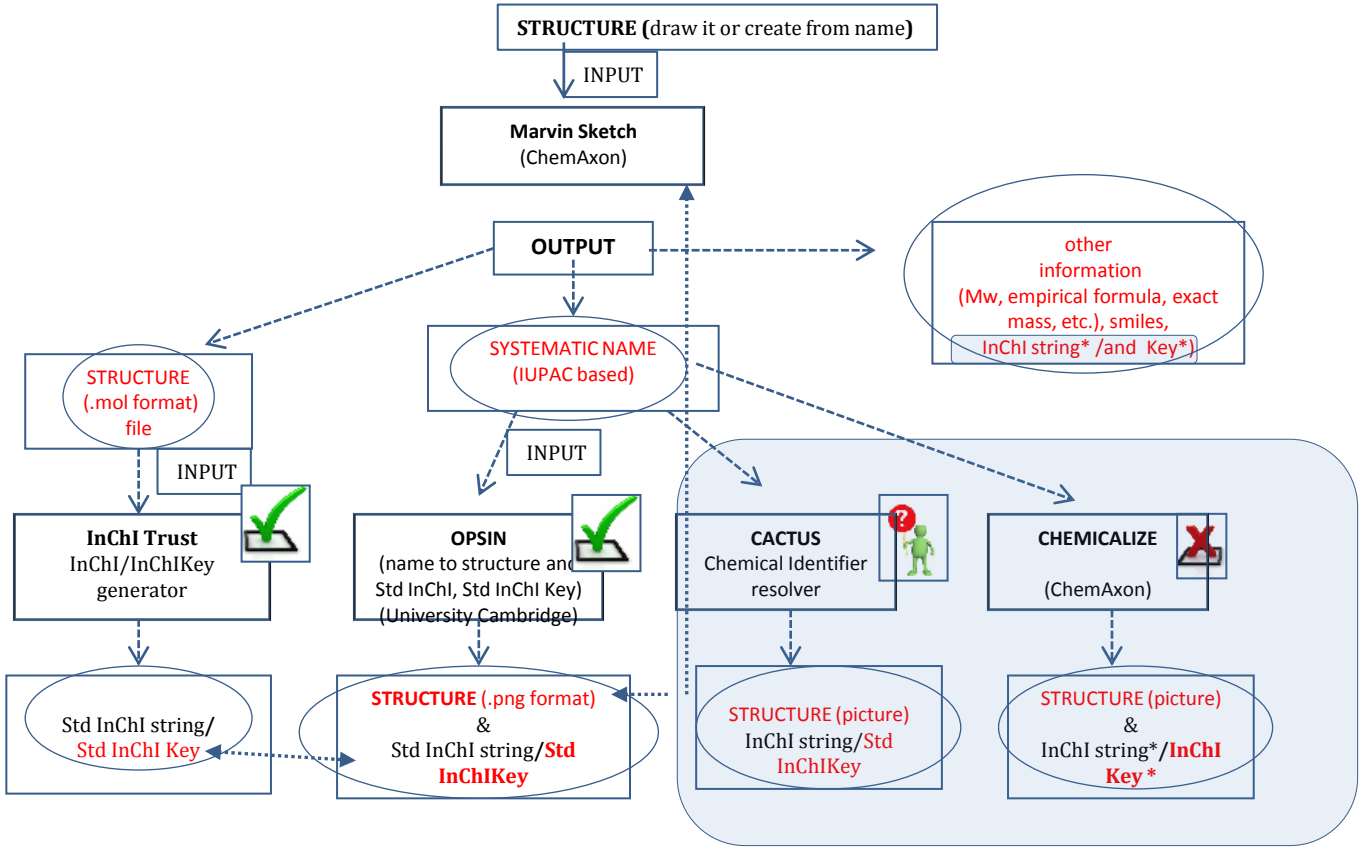
Common names	N/A
IUPAC name	1-pentyl-N-phenyl-1H-indole-3-carboxamide
SMILES	CCCCCN1C=C(C(=O)NC2=CC=CC=C2)C2=CC=CC=C12
InChI	InChI=1S/C20H22N2O/c1-2-3-9-14-22-15-18(17-12-7-8-13-19(17)22)20(23)21-16-10-5-4-6-11-16/h4-8,10-13,15H,2-3,9,14H2,1H3,(H,21,23)
InChIKey	InChIKey=SPUQBSMPSCYQLS-UHFFFAOYSA-N
CAS	N/A

B

Example: Chemicalize (ChemAxon) outputs for two compounds: A) for methcathinone (2-(methylamino)-1-phenyl-propan-1-one) **non standard** InChI string and Key were generated; B9) for SDB-006-N-phenyl-analog (1-pentyl-N-phenyl-1H-indole-3-carboxamide) **standard** forms of string and Key were the output.

SUBSTANCE DESRIPTORS/ IDENTIFIERS:

general validation scheme in RESPONSE



Conclusions

❑ Several substance descriptors and their advantages/ disadvantages for interoperability between different databases have been explored

❑ With some limitations **googlable Std InChi Key** seems to have a good potential to improve interoperability within and between databases and provide up to date information in real time

- main limitations which one should have in mind:
  - uniqueness of the key is limited to the same level of structure information included into the key (e.g. stereochemistry, base or salt form)
  - the extent of structural information given in particular database will most probably be related to the intended purpose of database use

❑ Validation of some structure related chemical identifiers is highly recommended!

References

- http://www.policija.si/eng/index.php/generalpolicedirectorate/1669-nfl-page-response
- http://www.policija.si/apps/nfl_response_web/seznam.php
- http://www.emcdda.europa.eu/
- https://en.wikipedia.org/wiki/International_Chemical_Identifier
- Stephen R Heller, Alan McNaught,Igor Pletnev,Stephen Stein and Dmitrii Tchekhovskoi, InChI, the IUPAC International Chemical Identifier, Journal of Cheminformatics 2015, 7:23, DOI: 10.1186/s13321-015-0068-4; https://jcheminf.springeropen.com/articles/10.1186/s13321-015-0068-4
- https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system
- Wendy Warr, Representation of chemical structures . Wiley Interdiscip Rev Comput Mol Sci. 2011;1:557–79, DOI: 10.1002/wcms.36
- https://www.chemaxon.com/products/marvin/marvinsketch/?gclid=CjwKEAiAgavBBRCA7ZbggrLSkUcSJACWDexAN5B9vnh01oHXQKgvgUvcMGZY065FgDNVurKrfl9VWxoC-sjw_wcB
- http://www.inchi-trust.org/
- IUPAC International Chemical Identifier (InChI) Programs InChI version 1, software version 1.04 (September 2011) User's Guide, download able here: http://www.inchi-trust.org/downloads/ (*INCHI-1-DOC.zip*)
- http://opsin.ch.cam.ac.uk/
- https://cactus.nci.nih.gov/
- https://chemicalize.com/welcome

Acknowledgement

❑ *The NFL activities implemented in the RESPONSE project have been financially supported by the Prevention of and fight against crime Programme of the European Union (grant agreement number JUST/2013/ISEC/DRUGS/AG/6413). We kindly acknowledge this! The content of this presentation is the sole responsibility of the author and can in no way be taken to reflect the views of the European Commission.*

❑ *Personally, I warmly acknowledge my colleague Denis Saboti who willingly took over the oral presentation of this work . It was really a brave decission because he works in the lab for less than five months.*

Sonja