



New IUPAC Organic Nomenclature: From Bottle Label to Update of Databases

Andrey Yerin

*and all ACD/Labs team developing and
supporting nomenclature tools*

**ACS 248th National Meeting
Division of Chemical Information**

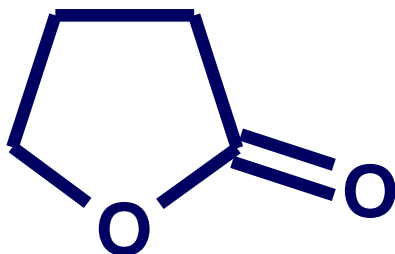
San Francisco, USA, August 11, 2014

Proliferation of Names

A chemical name is created according to nomenclature rules;
most commonly according to IUPAC rules.

But there are too many rules!

Several systematic names are possible for this structure



γ -Butyrolactone

Dihydro-2(3H)-furanone

Dihydrofuran-2(3H)-one

4-Hydroxybutanoic acid lactone

1-Oxacyclopentan-2-one

Oxolan-2-one

2-Oxotetrahydrofuran

Tetrahydro-2-furanon

All of these names correctly describe the structure.

But the multiplicity of systematic names is inconvenient and diminishes the role of nomenclature.

Is there a “most systematic” name?

The nice thing about standards is,
there are so many to choose from

C. Northcote Parkinson

... the rules of systematic nomenclature need not necessarily lead to a unique name for each compound...

To force the naming of all compounds into the Procrustean bed of one set of rules would not serve the needs of general communication.

A Guide to IUPAC Nomenclature of Organic Compounds, 1993

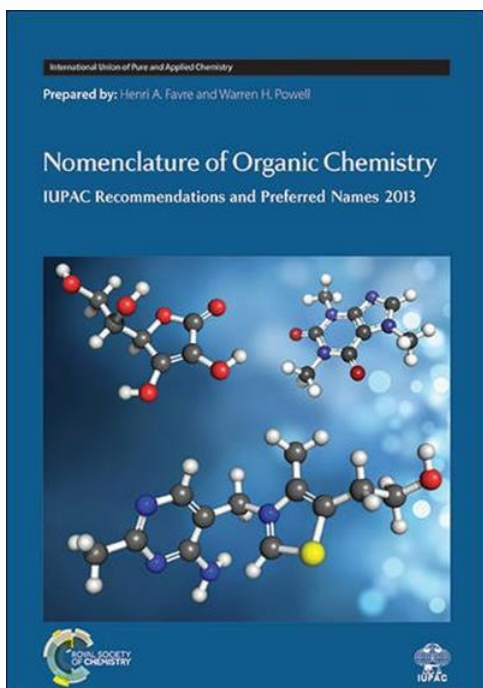
To enforce the role of systematic nomenclature the need of having one best name has been recognized by IUPAC:

...we have developed rules for assigning ‘preferred IUPAC names’, while continuing to allow alternative names in order to preserve the diversity and adaptability of the nomenclature...

Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Names 2013

New IUPAC Organic Nomenclature

In December 2013, the book “Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Names” was published.



It is truly a long awaited publication. The work started in 1992. The IUPAC project was initiated in 2001.

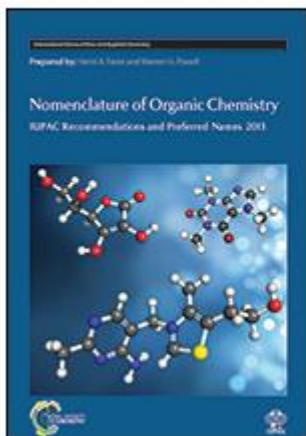
This major organic nomenclature publication is an answer to the rapid development of chemistry and appearance of new classes of chemicals we have seen over the past 20 years.

The book deals with naming principles that were unchanged for about 35 years since the IUPAC Blue book was published in 1979.

New IUPAC Organic Nomenclature

Practically all of the work was done by two authors – Henri Favre and Warren Powell, but many people participated in all stages.

Nomenclature of Organic Chemistry IUPAC Recommendations and Preferred Names 2013



Henri A Favre (Author), Warren H Powell (Author)
ISBN: 978-0-85404-182-4
Copyright: 2013
Format: Hardback
Extent: 1612
Price: £175.00

[BUY PRINT](#)

[View Contents](#)

It is important to note that the book was reviewed not only by classical nomenclature experts but experts in computerized nomenclature as well. The experts from ACD/Labs, CambridgeSoft and NextMove Software were involved.

The most principle innovation is the concept of preferred IUPAC name (PIN) selected by hierarchical sets of criteria.

There are many other changes compared to the Blue Book 1979. It takes ten pages just to list them. Lets see what ACD/Name provides for naming.

New IUPAC Organic Nomenclature

The new Blue Book introduced many changes; the most significant of which include:

- Less trivial names**
- Length is senior to unsaturation**
- More locants**
- More enclosing marks**
- Multiplication over substitution**
- Al, Ga, In and Tl are now 'organic elements'**
- many other specific changes**

Of high importance is the fact that the number of criteria to select the best name has increased significantly. The nomenclature became more systematic but at the same time more difficult for humans.

New IUPAC Organic Nomenclature

The systematic nomenclature assumes general naming principles and minimal number of trivial names.

Such nomenclature is easier to learn and implement in software.

The downside is that some very common names are changed.



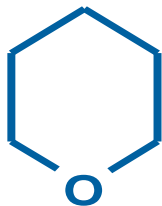
Was
PIN

tetrahydrofuran
oxolane

Tetrahydrofuran is still acceptable!
Some people would like to make it the preferred name, but we must agree to disagree.



Are there anybody who would argue with this change?



Was
PIN

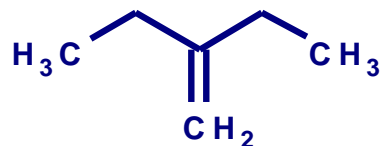
tetrahydro-2H-pyran
oxane

Why Wait to Program Our Software?

The first public draft of the new Blue Book was made available in 2004.

Some producers of naming software decided to implement new principles and claimed compatibility to '2004 rules'.

Being involved in the development of IUPAC recommendations, ACD/Labs decided to wait until later versions, keeping in mind that several principle procedures were still under consideration well before the actual publication. We did implement some earlier well-agreed upon changes, including the change of seniority of chain length to degree of unsaturation.



Was
PIN

2-ethylbut-1-ene
3-methylidenepentane

It means that changes concern not only advanced cases but some very simple cases as well.

Note that new name corresponds to CAS naming principles.

Time to Start Programming

ACD/Name is the most mature naming program on the market. The development was started in 1994 shortly after ACD/Labs was founded.

We support not only organic nomenclature but some inorganic and biochemical nomenclature as well.

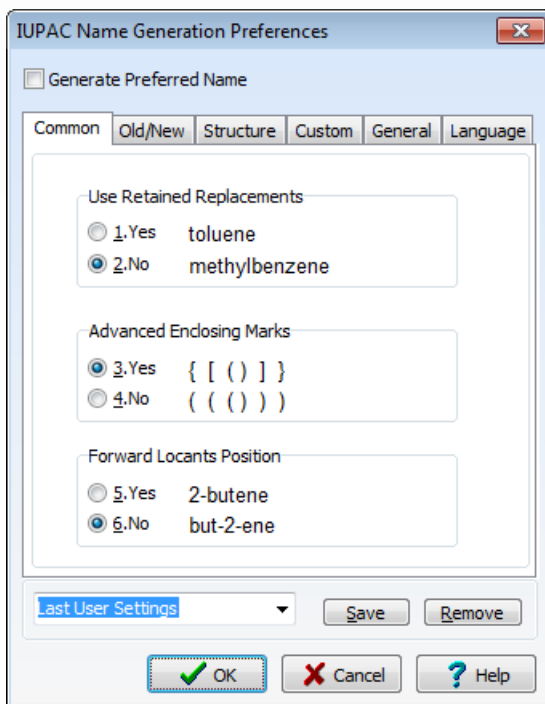
It is very important that the PIN concept does not make all other names unacceptable. Chemists still have a choice on how to name the compound providing that the chemical name correctly and unambiguously defines the chemical structure. Such names are considered belonging to general nomenclature instead of preferred IUPAC nomenclature.

Thus, while programming new procedures, we retained most old naming conventions and trivial or some even outdated names.

It makes implementation of new procedures even more difficult but allows us to increase the educational value of ACD/Name.

Preferred Name Leaves No Choice

While IUPAC rules allow the user to choose naming concepts and other several options, the generation of PIN assumes only one possible way to name a structure.



IUPAC Name Generation Preferences

Generate Preferred Name

Common | Old/New | Structure | Custom | General | Language

Use Retained Replacements

1.Yes toluene

2.No methylbenzene

Advanced Enclosing Marks

3.Yes { [()] }

4.No (())

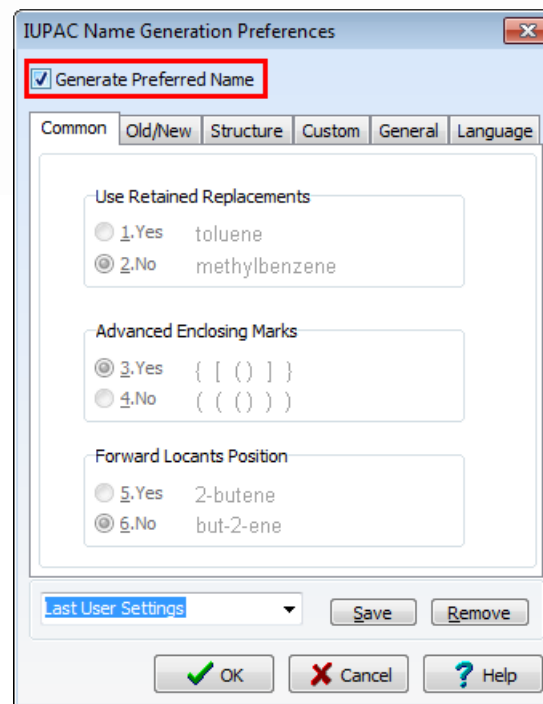
Forward Locants Position

5.Yes 2-butene

6.No but-2-ene

Last User Settings [v] Save Remove

OK Cancel Help



IUPAC Name Generation Preferences

Generate Preferred Name

Common | Old/New | Structure | Custom | General | Language

Use Retained Replacements

1.Yes toluene

2.No methylbenzene

Advanced Enclosing Marks

3.Yes { [()] }

4.No (())

Forward Locants Position

5.Yes 2-butene

6.No but-2-ene

Last User Settings [v] Save Remove

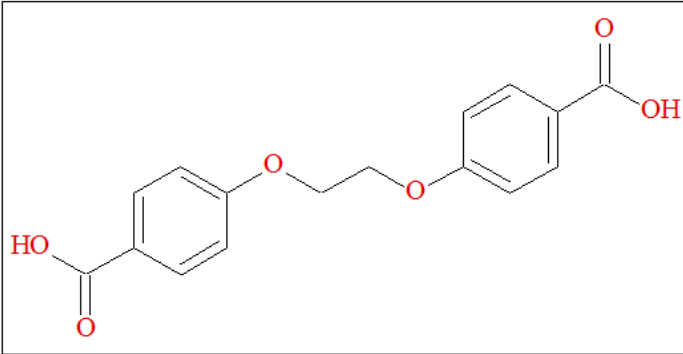
OK Cancel Help

The PIN generation option disables most other options, leaving the user with no choice.

How We Check New Procedures

The correspondence of names is compared by computer excluding even minor deviations from expected names.

We have a database of structures mentioned in the book and all names are copy-pasted. In this example no changes were found.



Formula:

FW:

STR_TYPE:

ACD/Name comment:

ACD_1401	<input type="text" value="4,4'-[ethane-1,2-diylbis(oxy)]dibenzoic acid"/>	COMP_1401	<input type="text" value="YES"/>
NAME_PNAME_PIN	<input type="text" value="4,4'-[ethane-1,2-diylbis(oxy)]dibenzoic acid"/>	REF_PNAME_PIN	<input type="text" value="P-51.2.2"/>
ACD_1500a	<input type="text" value="4,4'-[ethane-1,2-diylbis(oxy)]dibenzoic acid"/>	COMP_15a	<input type="text" value="YES"/>

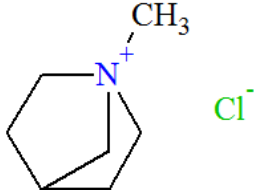
ACD 1401 – name corresponding to old rules, NAME_PNAME_PIN – name from book, ACD 1500a – our latest new name.

COMP scripted fields compare text strings for equivalence – YES or NO.

How We Check New Procedures

We can easily check what old names do not correspond to new principles – ‘NO’ value in COMP_1401.

‘YES’ in COMP_15a indicates that our new name corresponds to PIN.



Formula
C₇H₁₄ClN

FW
147.6458 (112.192)

STR_TYPE
FULL

ACD/Name comment
! loc ium

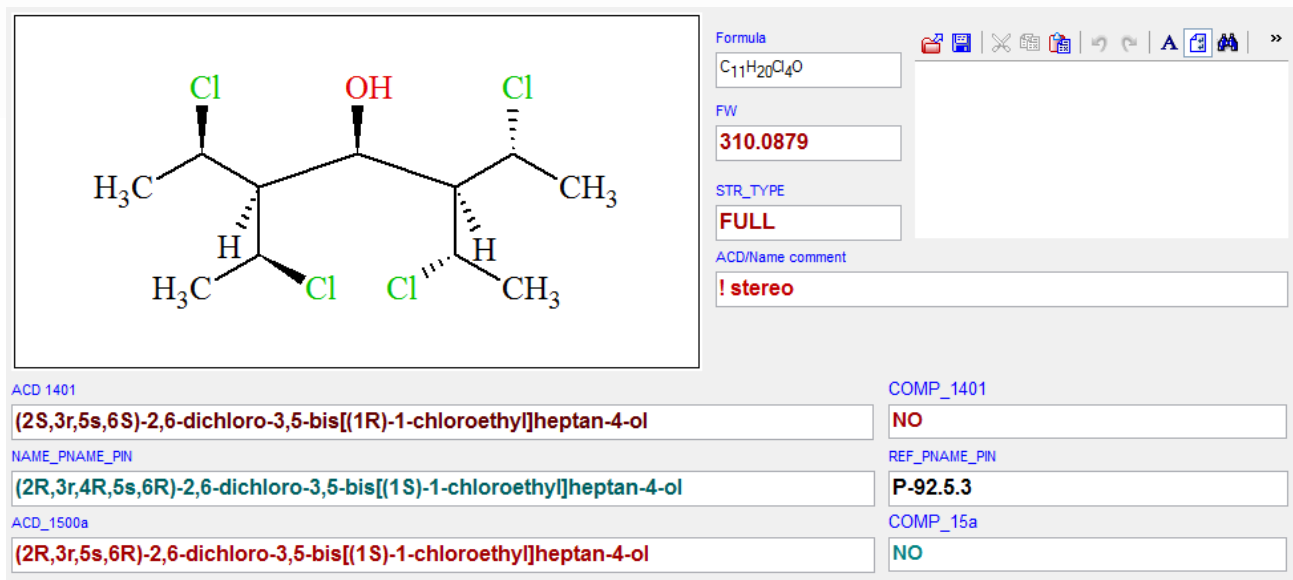
ACD_1401 1-methyl-1-azoniabicyclo[2.2.1]heptane chloride	COMP_1401 NO
NAME_PNAME_PIN 1-methyl-1-azabicyclo[2.2.1]heptan-1-ium chloride	REF_PNAME_PIN P-73.4
ACD_1500a 1-methyl-1-azabicyclo[2.2.1]heptan-1-ium chloride	COMP_15a YES

This example corresponds to ‘a’-terms for cationic heteroatoms discontinued in new rules.

How We Check New Procedures

The difference of in a new name with PIN most often indicates that certain procedures were not implemented or there were mistakes in the naming algorithms.

We check all such cases to minimize the differences.



This case indicates absent support of CIP rule 5.

Sometimes 'NO' indicates that error is not with ACD/Name but a mistake in the name provided in the book.

We Also Check the Book

Implementation of new procedures allows us to check the names provided in the book itself.

The new Blue Book is a very substantial publication. It has 1600 pages with more than 10,000 chemical names for 6000 structures.

All names have been created by the ultimate experts in nomenclature but it is too difficult for human beings to avoid mistakes and misprints. It is practically impossible to keep in mind all procedures and criteria and manually create fully correct chemical names.

Practically after each implemented procedure we find names in the book that include misprints or are simply inconsistent with the principles specified in the book.

Mistakes are hardly avoidable and it is especially so for the book with 10,000 chemical names.

Many mistakes are already found and ACD/Labs is probably a champion in spotting of errors. We do not search for mistakes, we just try to implement new rules.

Errata for New Blue Book

I arrived at the ACS Fall 2014 meeting just after the meeting of the IUPAC division of chemical nomenclature and structure representation (Division VIII) that took place in Bangor UK, August 1-4.

One of the most important discussions was devoted to publication of corrections for the new Blue Book.

Currently the list includes more than 500 corrections. While many of them are just textual corrections, a significant number deal with the correction of chemical names. In some cases these corrections refer to the principles of naming.

It means that even after the publication it is sometimes hard to decide what procedure to implement.

The errata can be expected about the end of this year. But IUPAC nomenclature continues its development along with chemistry itself and new recommendations should be expected.

Demand and Share for PINs

There are no doubts that preferred IUPAC name will come in widespread use. First of all for official documents and regulations.

We got the first request from our Japanese customers a month before the actual publication of new Blue book.

Obviously not every name is changed by the new rules but according to our estimations it affects about 25% of names; and chemists will want to know is a particular name is preferred or not.

Everyone can agree that chemical nomenclature is complex. It is now even more complex, including more procedures and criteria.

New principles of organic nomenclature should be learned and taught at least at a basic level to distinguish the correct name from incorrect versions.

New nomenclature rules significantly concern the automatic treatment of chemical information including algorithmic name generation. The chemistry and nomenclature has already become too complex to rely fully on human interpretation.

Lets control the quality of names

Humans cannot be trusted!

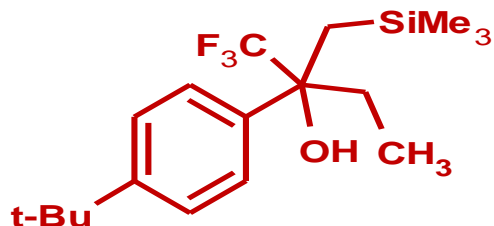
Manual naming and typing tends to produce mistakes and misprints.

The name below was found in recent publication. It looks quite systematic but once you try to derive the structure you understand that it is wrong.

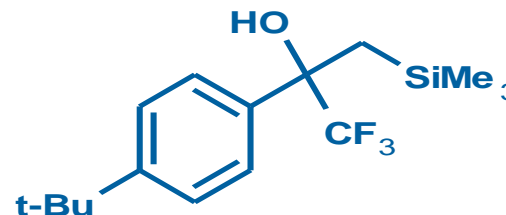
[dx.doi.org/10.1021/jo402577n](https://doi.org/10.1021/jo402577n)

J. Org. Chem. 2014, 79, 1145–1155

2-(4-(*tert*-Butyl)phenyl)-1,1,1-trifluoro-2-((trimethylsilyl)methyl)butan-2-ol



while intended



2-(4-*tert*-butylphenyl)-1,1,1-trifluoro-3-(trimethylsilyl)propan-2-ol

Currently the most important task is algorithmic control of chemical name quality. People should not be trusted to generate and type names. It should be done by software tools or controlled by such tools.

Thank you for attention

to this presentation

and to the quality

of chemical information