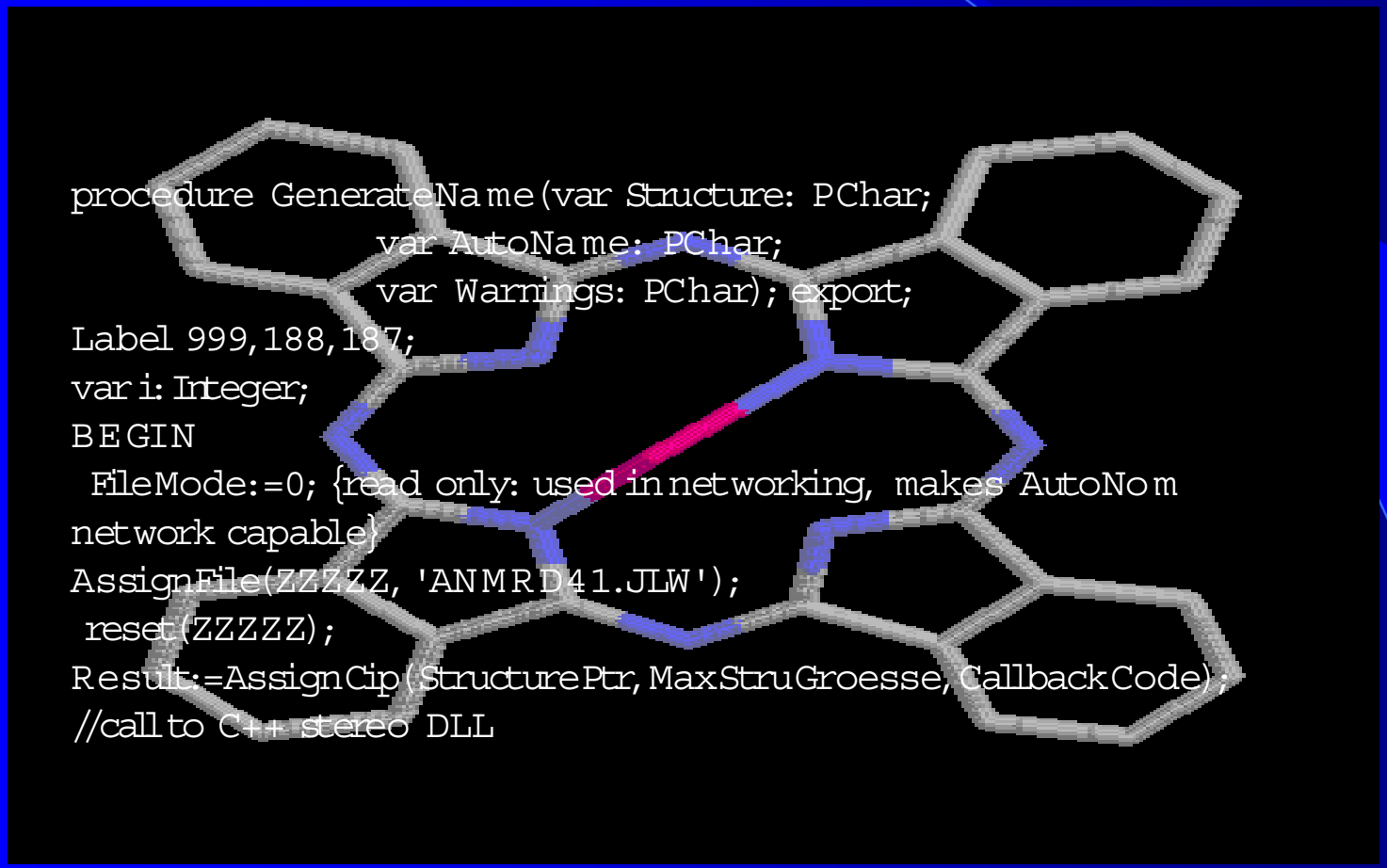
COMPUTER AIDED ORGANIC NOMENCLATURE : AutoNom (TM) AS EFFECTIVE TOOL FOR AUTOMATIC NAMING AT REGISTRATION AND PUBLICATION OF CHEMICAL STRUCTURES.



```
procedure GenerateName(var Structure: PChar;
                       var AutoName: PChar;
                       var Warnings: PChar); export;
Label 999,188,187;
var i: Integer;
BEGIN
 FileMode:=0; {read only: used in networking, makes AutoNom
network capable}
AssignFile(ZZZZZ, 'ANMRD41.JLW');
 reset(ZZZZZ);
Result:=AssignCip(StructurePtr, MaxStruGroesse, CallbackCode);
//call to C++ stereo DLL
```

Dr Janusz L. Wiśniewski

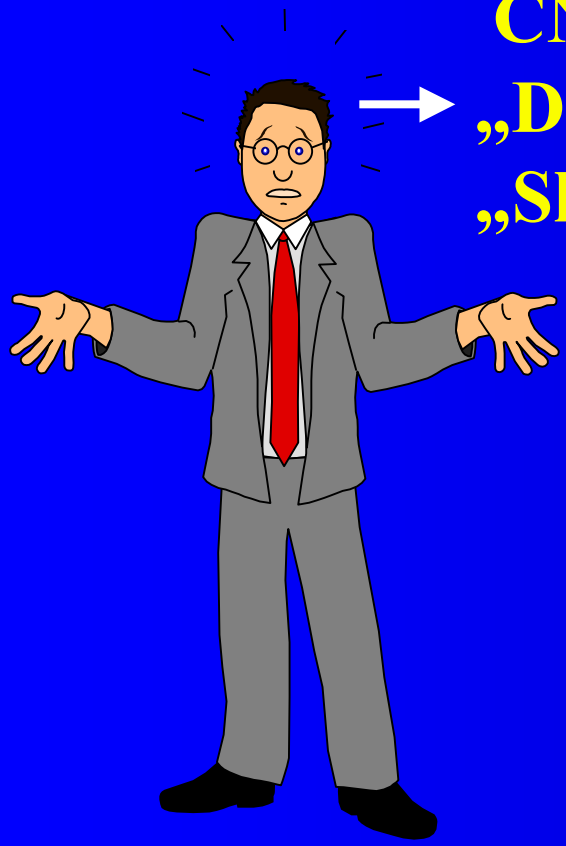
MDL Information Systems GmbH

Frankfurt/Main, Germany

Nomenclature Ambiguity

IUPAC and its CNOC

P-name Project (95-01)



CNOC: retained, preferred, alternative....
„Dialects”: Beilstein, CAS,
„Slang”: Trade names, Catalog Names

"A man with one watch knows what time it is, a man with two watches is never sure....."

"Everywhere is a walking distance...
If you have the time"
- Steven Wright -

AutoNom's Story

Wednesday, 21 January 1987 approx. 10.30 a.m.

Sandy Lawson's office on the tenth floor of the Carl-Bosch-Haus, Beilstein Institute, Frankfurt/Main, Germany:

Sandy Lawson (R & D at MDL Information Systems GmbH, Frankfurt/Main) meets Janusz L. Wisniewski (currently Senior Software Developer) and while handing him the IUPAC Blue Book containing the rules of Organic Nomenclature says:
"Implement this book, don't talk to chemists too much, simply program it."

more... >>>

AUTONOM – a Chemist's Dream

Some of us love it, some of us hate it. Some of us do it well, others badly, but none of us do it quite the same as the next man (or woman). CAS does it, BEILSTEIN does it, IUPAC does it, and they all do it slightly (or even distinctly) differently, depending on when they do it.

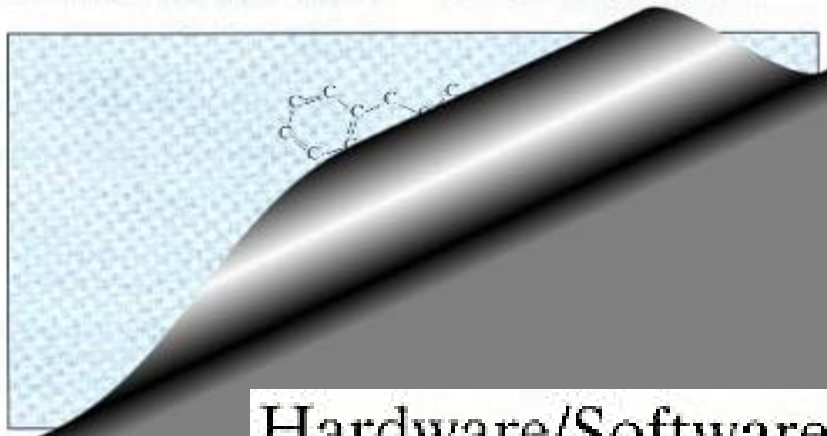
If you haven't guessed already, we are referring to systematic organic nomenclature. Let's be frank, most of us are fairly weak in this essential discipline. It's often much easier to synthesize a new compound than name it, and the chances that someone else will find your compound in indexes using your nomenclature can be slimmer than you think.

A helping hand

To help you a little in this respect, the R&D Division of BEILSTEIN is working on a computer program that sounds like the chemist's dream. It's called AUTONOM, and works like this: you draw your structure on your PC-compatible using a mouse, and AUTONOM gives you an IUPAC-compatible name within a few seconds!

It's as simple as that.

Here's an example of a compound which has never been made (yet):



Hardware/Software Requirements: IBM-AT (or compatible, also 80386-based, naturally), CGA, EGA, or Hercules Graphics, MS-mouse (or compatible), 640K RAM, Winchester with at least 2 MB free.

Reference:
[Beilstein Brief, **2/1989**]

AutoNom's Story (cntd.)

... structural diagram...

- ✓ organic species
- ⇘ neglects
  - ✓ polymers
  - ✓ coordination compounds
  - ✓ inorganic materials
  - ✓ alloys
- ✓ structures up to **125** (heavy) atoms
- ✓ rings and chains up to **44** atoms

... appropriate systematic name...

1. AutoNom uses IUPAC conventions and generates single, unique (always the same no matter who is currently in charge of your Nomenclature Department !) name.

2. The program is built around the principles of substitutive nomenclature, with a limited application of replacement nomenclature (only for certain heterocycles, not for chains), additive nomenclature (hydro prefixes) and subtractive nomenclature (unsaturation in aliphatic compounds and cycloalkanes). The use of the alternative radicofunctional nomenclature is limited to primary amines and the prefix form of some acid derivatives. The conjunctive nomenclature is not supported, and (at present) this also applies to multiplicative nomenclature.

3. Broad framework of IUPAC recommendations allows sometimes distinct variations in usage (frequently according to taste (!!) of the user):
"aniline = "amino-benzene“ = "phenyl-amine“ = “benzeneamine”
"acetic acid" = "ethanoic acid"
"pyridine“ = "azabenzene“

4. There is no "best" name for any given structure

… if possible…

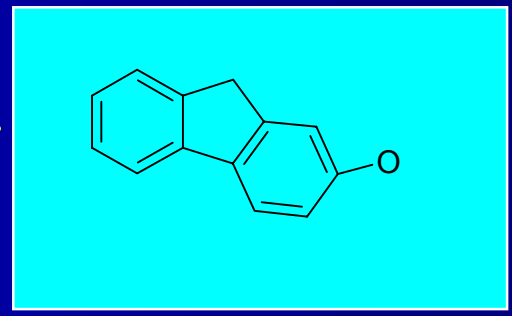
Avoid giving an incorect name !!!

If the program is "in doubt" as to whether the name is correct, it will politely refuse naming, giving a diagnostic instead

"Philosophy":

Better no name than a wrong name !

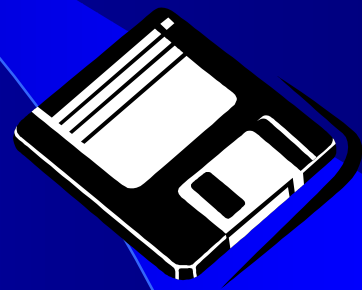
The Algorithm:



structure initialization

functional group identification

ring perception and recognition



parent structure selection

binary name tree processing

chemical name assembly

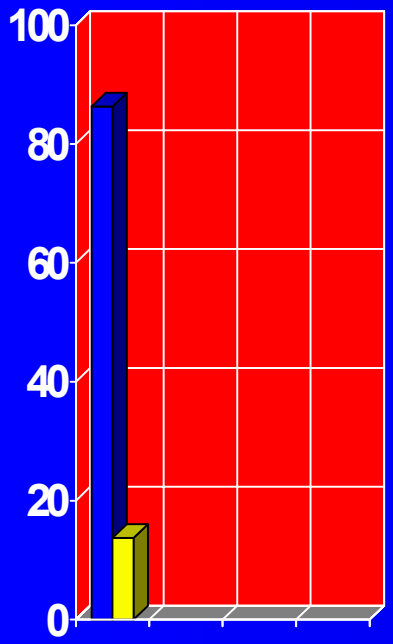
9H-Fluorene-2-ol

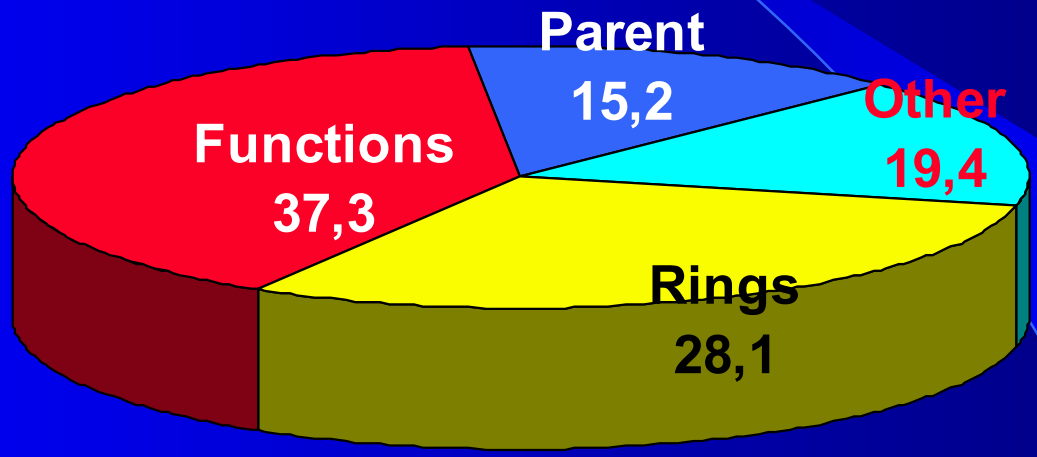
ELSEVIER SCIENCE

MDL Information Systems GmbH

AutoNom's Efficiency:

- success rate (measured over 63,040.0 sample): **86.3%**





Distribution of diagnostic

- naming rate: **40-60 structures/minute**

Most up-to-date:

Chemistry

1. CAS ring nomenclature
2. stereo-chemistry: (R,S) & (E/Z)
3. optional "on-screen" numbering

Internet

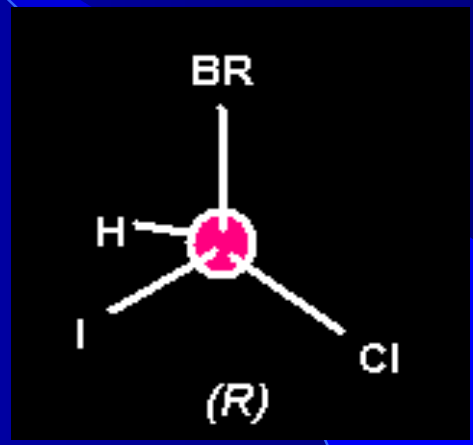
4. AutoNom as (free) CGI script: **www.chemweb.com**
5. AutoNom Standard as (free) download: **www.mdli.com**
6. AutoNom 2000-TT DLL as customizable DLL

Stereochemistry

* Implementation of **CIP** (Cahn-Ingold-Prelog) convention to order ligands (for both chiral centers as well as topological E,Z)

1. **"1982 Set of Rules" implemented** (Prelog, Helmchen, "*Angew. Chem*", 1982, 21, 567) : it is new; ("1966 set" is mostly used)

* atomic number rule
* atomic mass rule
* Z over E (cis over trans)
* asymmetric (R,S) over pseudo-asymmetric (r,s)
* like pair (R,R)/(S,S) over unlike pair (R,S)/(S,R)
* r over s
•R over S



(! This set fully complies with newly adopted *P-Name* recommendations of the IUPAC Organic Nomenclature Committee !)

2. efficient handling of localized and MANCUDE type double bonds in ring systems (atom duplicating; phantom atoms, non-integer atomic numbers)

AutoNom - Publication of Structures

- Publishing for the online databases (Dialog, STN)

* Beilstein Institute case study ($$$)

* Beilstein Information ---→ MDL Information
(Compound Warehouse Project, ISIS/Draw Add-On, ISIS/Base Add-On)

- Publishing for the XFire

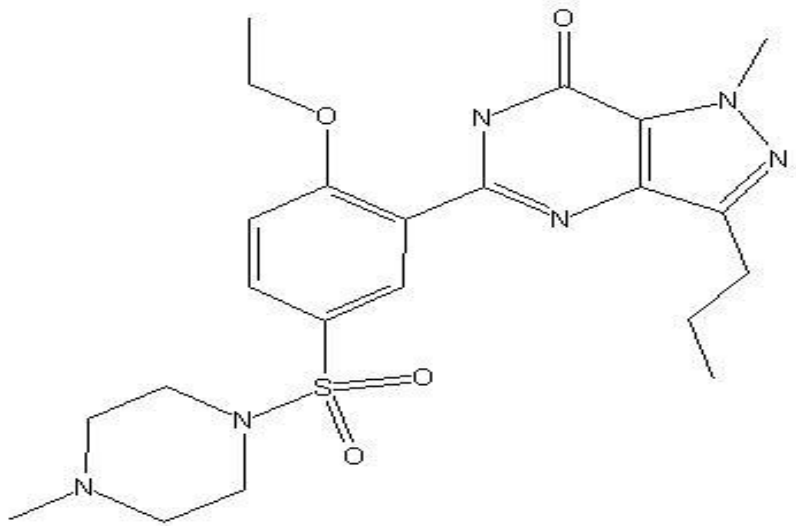
Q06:BS0101PR hit 1 of 1

Substance

Copy

Beilstein Registry Number	7673458
Chemical Name	sildenafil
Autoname	5-[2-ethoxy-5-(4-methyl-piperazine-1-sulfonyl)-phenyl]-1-methyl-3-propyl-1,6-dihydro-pyrazolo[4,3-*d*]pyrimidin-7-one
Molecular Formula	$C_{22}H_{30}N_6O_4S$
Molecular Weight	474.58
Lawson Number	30672, 28000, 3633, 2817, 298
Compound Type	heterocyclic
Constitution ID	6545075
Tautomer ID	7250545
Beilstein Reference	6-26
Entry Date	1997/07/31
Update Date	2001/01/30

Q06: Hit 1



Field Availability List 1-5 of 5 Home

Code	Field Name	Occ.
RX	Reaction	2
DE	Dissociation Exponent	2
POT	Electrochemical Characteristics	1
POW	Partition octan-1-ol/water (MCS)	1
PHAR	Pharmacological Data	18
M		

Reaction 1 of 2 Home

Reaction ID	4604481
Reactant BRN	7658133 5-(2-ethoxy-phenyl)-1-methyl-3-propyl-1,6-dihydro-pyrazolo[4,3-*d*]pyrimidin-7-one
	102724 1-methyl-piperazine
Product BRN	7673458 5-[2-ethoxy-5-(4-methyl-piperazine-1-sulfonyl)-phenyl]-1-methyl-3-propyl-1,6-dihydro-pyrazolo[4,3-*d*]pyrimidin-

AutoNom @ Registration of Structures

- **28+% of the compounds registered at the Beilstein file are sterical and in the incoming new structures the ratio is even higher**
- **Users search for THE specific (stereo)isomer and this must be differentiated by the registry numbers (BRNs)**
- **CIP/Parity - for tetrahedral centers ans topological sterochemistry of ligands at double bonds – relies exclusively on the spatial arrangement of atoms and bonds and is independent of nomenclature**

So what's the point ?

1. Mainly for the (double) quality control: Morgan numbering gets translated into nomenclature numbering (AutoNom/human) and stereo descriptors in the chemical name are compared with those calculated (CIP/Parity) by the algorithm. Only if total identity can be (algorithmically) confirmed, the next steps of the registrations are launched.

2. Secondly, the nomenclature locants are necessary for the stereo display; After the check @1. They are either taken from the AutoNom or the human (nomenclature specialist) name.

Future

Nomenclature:

- **Major** upgrade in the area of stereochemistry
  - Stereochemistry of amino acids, steroids, carbohydrates, terpenes
  - Handling of D- and L- prefixes for amino acids
  - ξ (Xi) stereo descriptor support (for wavy bond(s) or "uncertain" stereochemistry

Software:

- AutoNom COM* (AutoNom code as object runable everywhere)

* COM = Component Object Model (Microsoft Nomenclature !)

Thank you ☺

- **Thank you Libuse !!!**

(Dr . Libuse Goebels, Quality Control @ MDL Information Systems, Frankfurt/Main;
She is like a talking Blue Book and…. knows computers)

- **Thank you Sandy !!!**

(Prof. Dr. Alexander Lawson, R&D @ MDL Information Systems Frankfurt/Main)
For having the vision of computerized nomenclature in the time of 640kB of RAM and making me believe that it can come true.