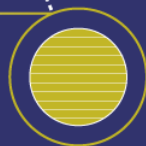


Chemonaut

on-line compound sourcing platform

Anna Rzepiela





Pyxis Discovery

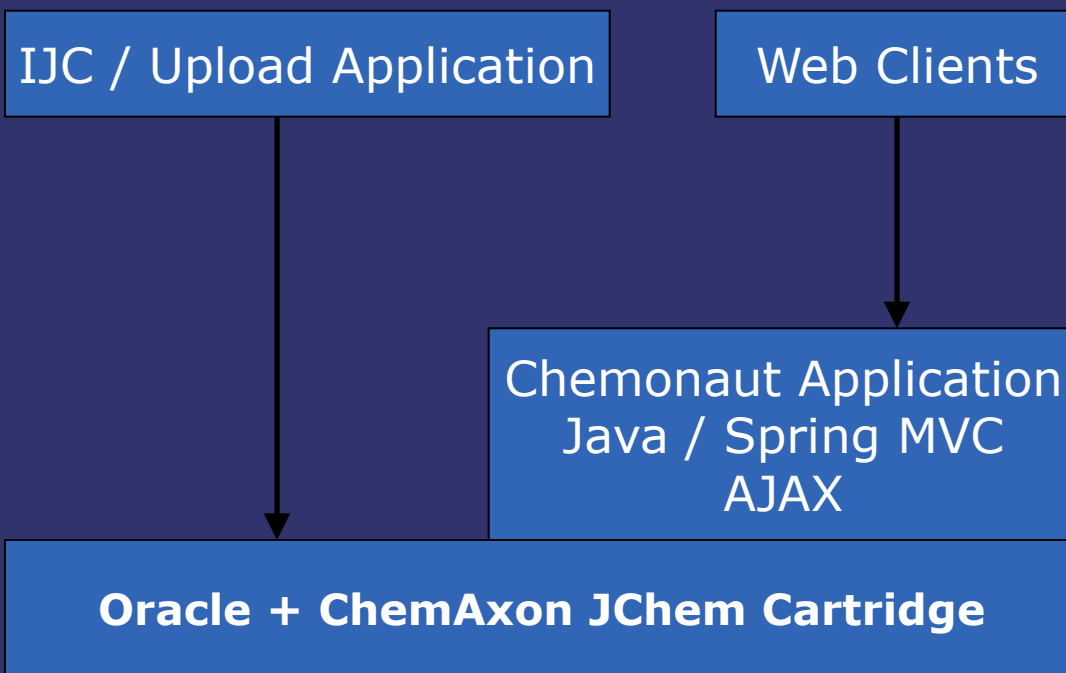
- Chemistry based lead discovery services
- Effective application of **computational chemistry** tools for faster lead identification using fewer compounds
- **3 focus areas**
 - Chemistry-based lead discovery services
 - Smart Libraries
 - Chemonaut



Chemonaut

- Database of screening compounds and building blocks from reliable suppliers
- Standardized structures and data
- Leadlike filtering
- Physico-chemical properties calculated

Chemonaut: application



Chemonaut: web view

Welcome Dr. Chemonaut

[sign out](#)

[home](#)

[in focus](#)

>> explore and select

[procurement service](#)


[expert support](#)


[chemonaut community](#)

[become a member](#)

[downloads](#)

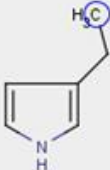
[contact](#)





Search Results Queries Selections Inquiry

File Edit View Insert Atom Bond Structure Tools Help



Search Option

Type

- Substructure
- Superstructure
- Full
- Duplicate
- Similarity

Database

Screening and Building Blocks

Screening Building Blocks

Hits Limit: Time Limit:

Hits Coloring Hits Alignment

Filters

Advanced Options


SMARTS:


Chemonaut: web view

Welcome Dr. Chemonaut

[sign out](#)

[home](#)
[in focus](#)
>> explore and select
[procurement service](#)
[expert support](#)
[chemonaut community](#)
[become a member](#)
[downloads](#)
[contact](#)





Search Results Queries Selections Inquiry

File Edit View Insert Atom Bond Structure Tools Help

Search Option

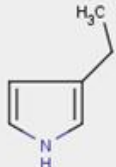
Filters

Physico-Chemical

<input type="text"/>	<	MW	<	<input type="text"/>
<input type="text"/>	<	cLogP	<=	<input type="text"/>
<input type="text"/>	<	cLogWS	>=	<input type="text"/>
<input type="text"/>	<	TPSA	<	<input type="text"/>
<input type="text"/>	<	R-Bonds	<	<input type="text"/>
<input type="text"/>	<	H-Donors	<	<input type="text"/>
<input type="text"/>	<	H-Acceptors	<	<input type="text"/>

Advanced Options

SMARTS:





Chemonaut: web view

Welcome Dr. Chemonaut

[sign out](#)

[home](#)
[in focus](#)
[>> explore and select](#)
[procurement service](#)
[expert support](#)
[chemonaut community](#)
[become a member](#)
[downloads](#)
[contact](#)





Search Results Queries Selections Inquiry

File Edit View Insert Atom Bond Structure Tools Help

100%


Search Option
Filters
Advanced Options

Stereochemistry
 On Exact Ignore
Double bond stereo check:
 All Marked Off

Atom Matching
Charges: On Exact Ignore
Isotopes: On Exact Ignore
Radicals: On Exact Ignore
Valence: On Ignore

The options above are ignored for Similarity search

H₃C




SMARTS:


Chemonaut: web view

Welcome Dr. Chemonaut

[sign out](#)

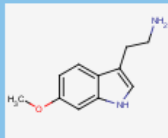
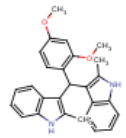
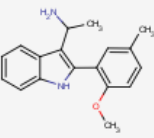
[home](#)
[in focus](#)
>> explore and select
[procurement service](#)
[expert support](#)
[chemonaut community](#)
[become a member](#)
[downloads](#)
[contact](#)



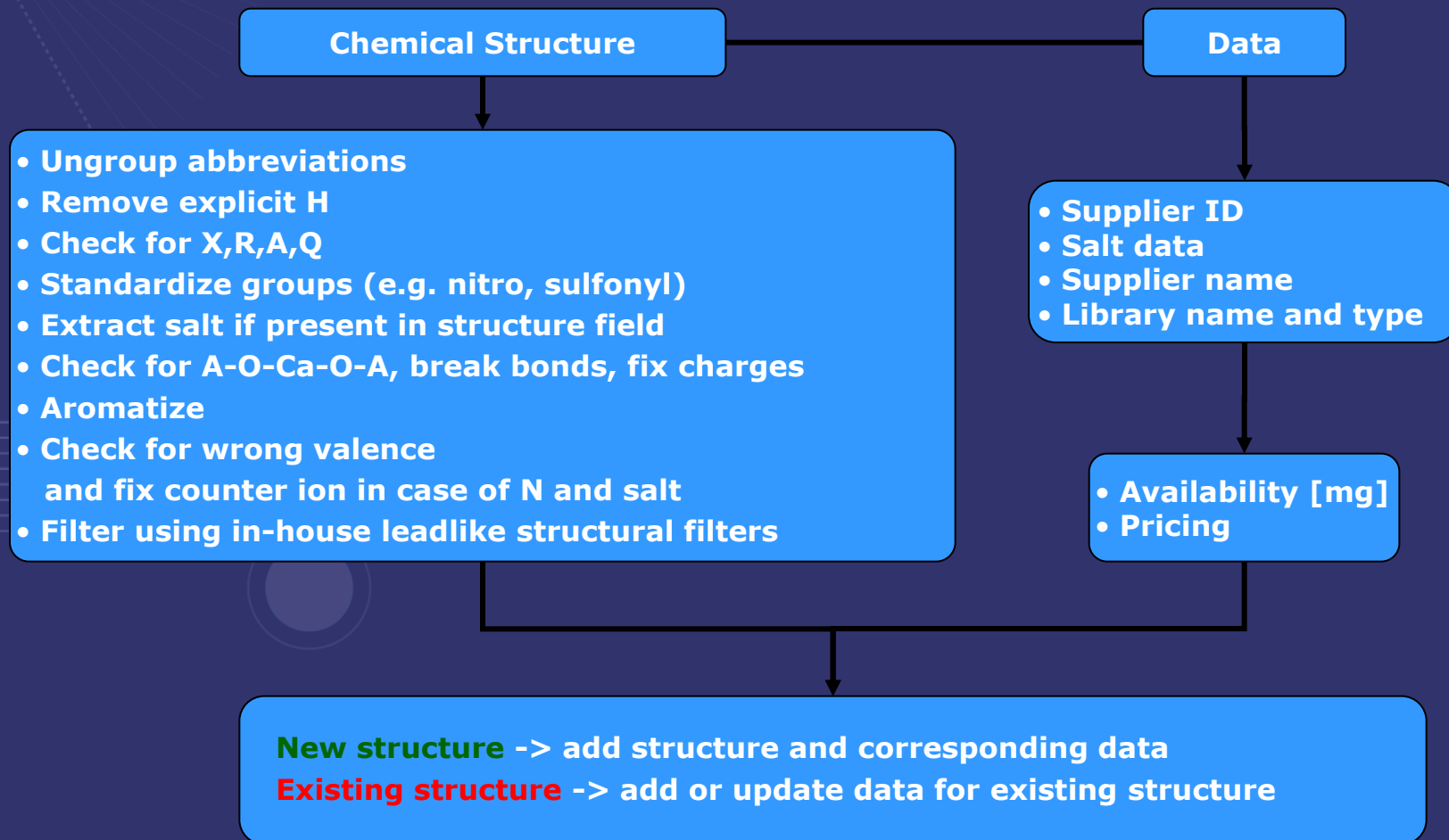


Search Results Queries Selections Inquiry

Hits: 10000 Selection Sort Currency Add to Inquiry Page 3 of 1000

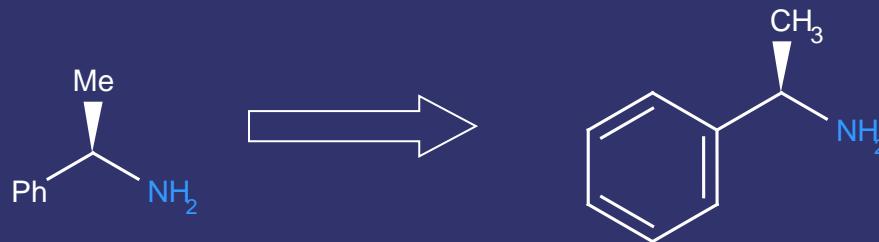
<input checked="" type="checkbox"/>	 <chem>COC1=CC=C(C=C1)CCN2C=CC=C2</chem>	Chemonaut ID: C000611903 Supplier ID: 5107324 MW: 190.24 Salt: HCl Price EUR: 1mg: 30 5mg: 44	Supplier: ChemBridge cLogP: ChemBridge TPSA: Bosche Scientific H-Acc: ChemBridge 10mg: 59 Availability: Fluorochem InterBioScreen Otava Princeton Biomolecular TimTec Alchem Pharmtech Interchim
<input checked="" type="checkbox"/>	 <chem>COC1=CC=C(C=C1)C2=CC=C(C=C2)C3=CC=C(C=C3)N4C=CC=C4</chem>	Chemonaut ID: C000621E01 Supplier ID: T5472809 MW: 410.51 Salt: Price EUR: 1mg: 46 5mg: 50	Supplier: cLogP: 5.43 cLogWS: -5.57 TPSA: 50.04 R-Bonds: 5 H-Acc: 2 H-Don: 2 10mg: 66 20mg: 78 50mg: 120 Availability: 120 mg
<input checked="" type="checkbox"/>	 <chem>COC1=CC=C(C=C1)C2=CC=C(C=C2)N3C=CC=C3</chem>	Chemonaut ID: C0006C1B01 Supplier ID: 1201085 MW: 280.36 Salt: Price EUR: 100mg: POR 250mg: POR 500mg: POR 1g: POR 5g: POR	Supplier: Otava cLogP: 2.93 cLogWS: -4.15 TPSA: 51.04 R-Bonds: 3 H-Acc: 2 H-Don: 2 Availability: >1 mg

Chemonaut: procedure

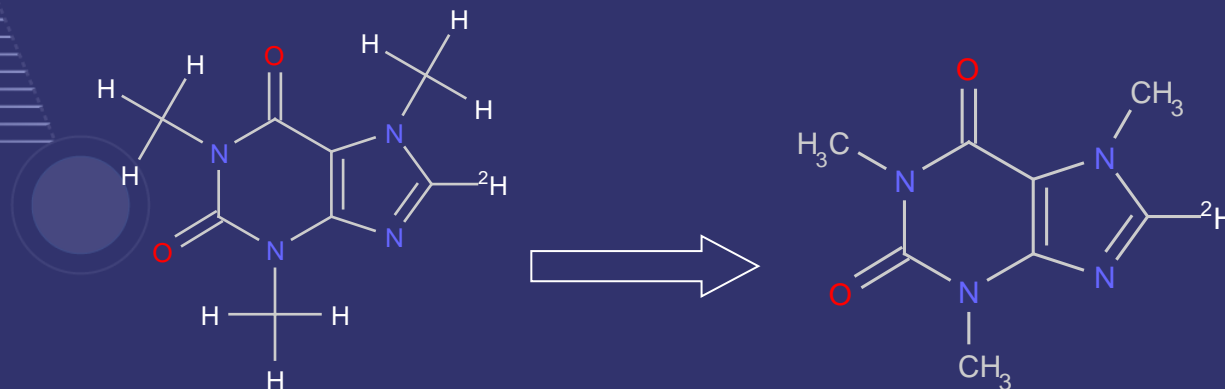


Structure import: standardize

- Ungroup abbreviations



- Remove explicit Hydrogen atoms



Converts explicit hydrogens to implicit (removes hydrogen atoms from the molecule graph). Charged, radical, mapped, isotope or wedged hydrogens are not removed.

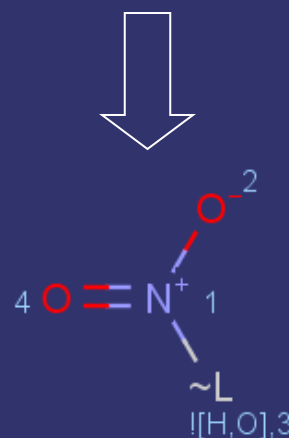
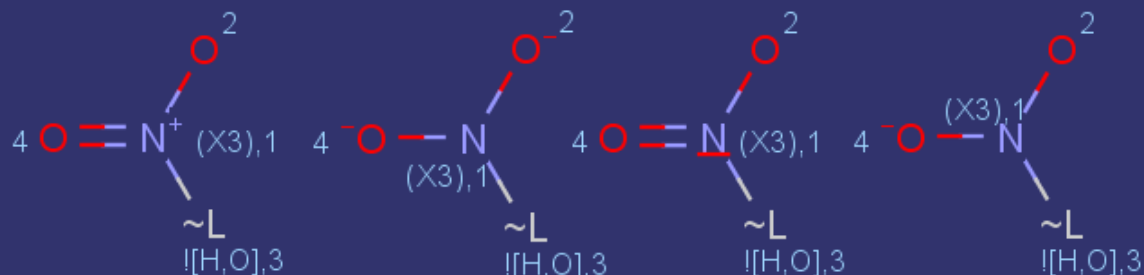
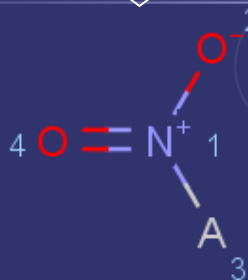
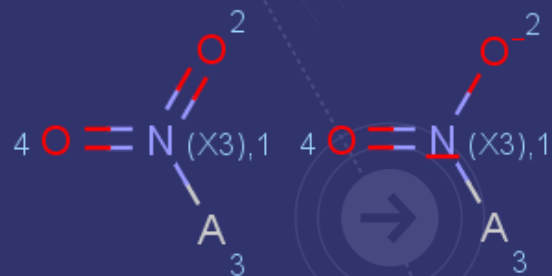
Structure import: wash

- Check for X,R,A,Q and discard if found



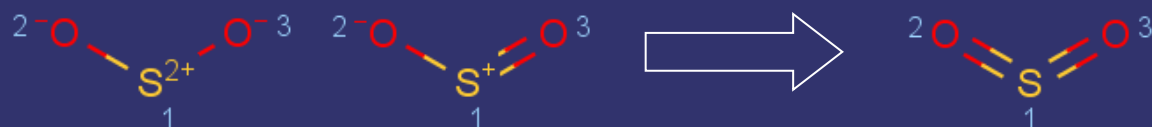
Structure import: standardize

- Standardize groups, e.g. nitro group

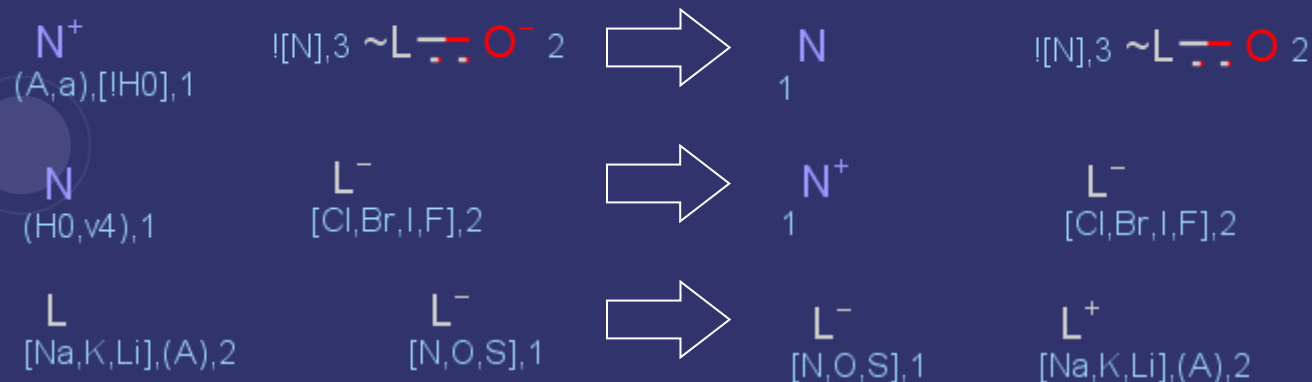


Structure import: standardize

- Standardize groups, e.g. sulfonyl



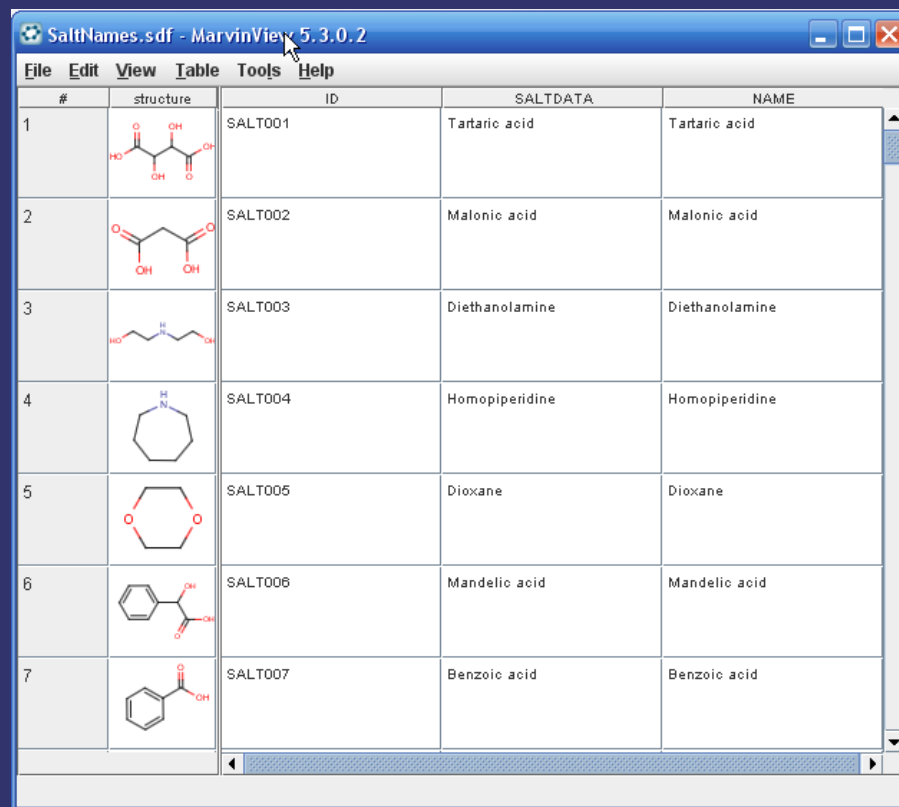
- Fix charges, e.g.: match charge on structure and salt, protonate acids, deprotonate bases:



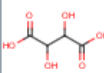
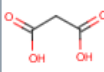
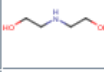
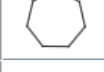



Structure import: wash

- Move salt data from structure field to Salt data field

for records with more than one structure in the structure field check if exists in the salt table store salt and ratio in salt field

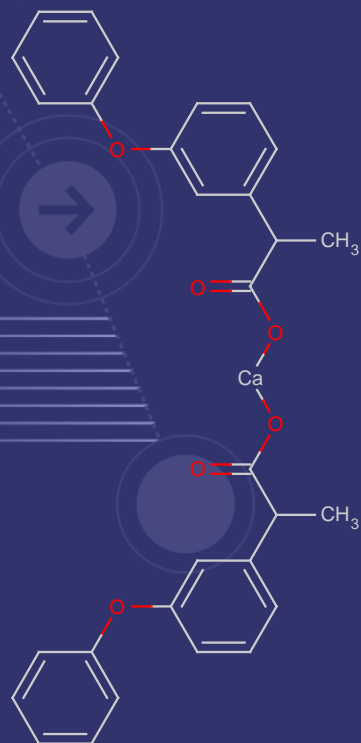


The screenshot shows the MarvinView 5.3.0.2 software window with a table of chemical structures and their corresponding salt data. The table has five columns: #, structure, ID, SALTDATA, and NAME. The structures are: 1. Tartaric acid, 2. Malonic acid, 3. Diethanolamine, 4. Homopiperidine, 5. Dioxane, 6. Mandelic acid, and 7. Benzoic acid.

#	structure	ID	SALTDATA	NAME
1		SALT001	Tartaric acid	Tartaric acid
2		SALT002	Malonic acid	Malonic acid
3		SALT003	Diethanolamine	Diethanolamine
4		SALT004	Homopiperidine	Homopiperidine
5		SALT005	Dioxane	Dioxane
6		SALT006	Mandelic acid	Mandelic acid
7		SALT007	Benzoic acid	Benzoic acid

Structure import: standardize

- Check for A-O-Ca-O-A, break bonds, fix charges



The screenshot shows the ChemAxon software interface. The 'Structure' window displays the standardized chemical structure of the biphenyl ether derivative. The 'STRUCTURES' table shows the following data:

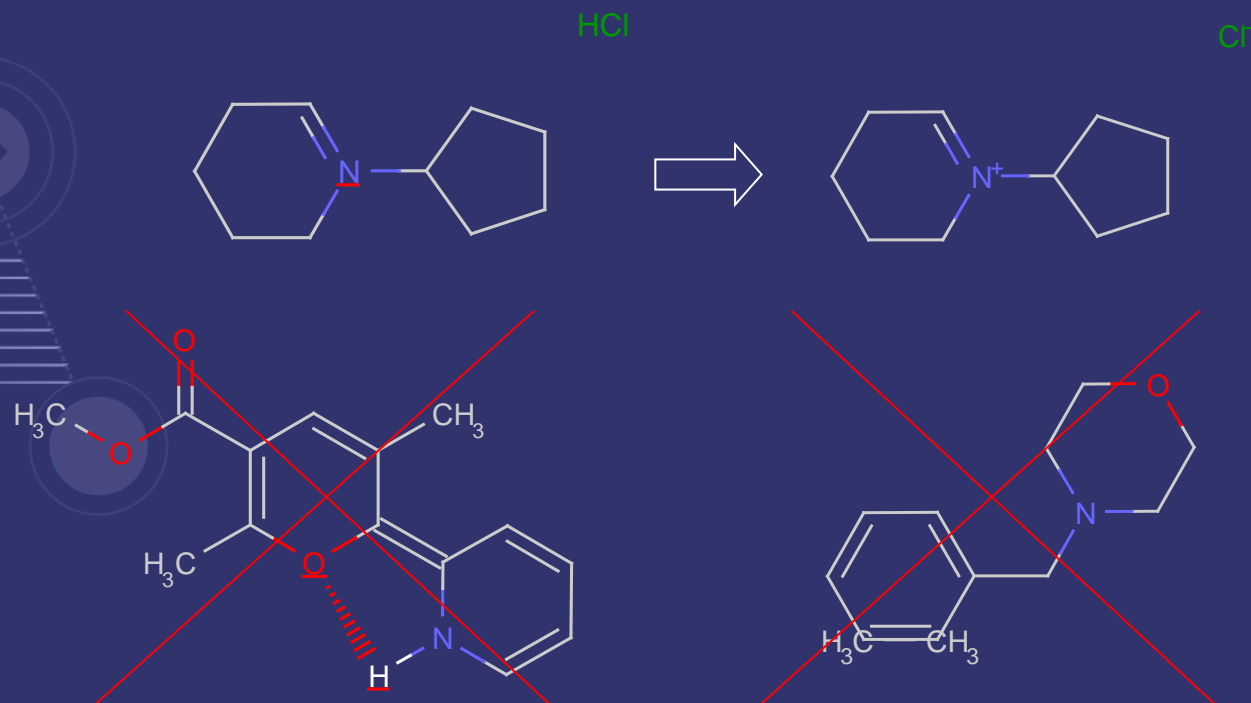
	CdId	Px CdId	PX_LIB	Mol Weight	Formula
1	1786931	1796527	2	241.26	C15H13O3

The 'Data' table shows the following data:

	ID	Px CdId	Px Cdsub	Px Sid	Px Supplier ID	Px Av	Px Lib	Px Salt
1	2554432	1796527	1	5	OR4881	0	2	0.5 Ca2+/H2O
2	2555087	1796527	2	14	F1676	0	2	0.5 Ca2+

Structure import: standardize

- Check for wrong valence and fix counter ion in case of N and salt, discard the rest



Chemonaut: leadlike filtering

remove the false positives, the covalent-acting electrophiles, the 'suicide inhibitors', the 'magic bullets', the 'warheads', etc., as well as the dyes, fertilizers, photographic chemicals, etc.

In total, a set of over 80 structural filters

1. Rishton, G.M., Reactive compounds and in vitro false positives in HTS, *Drug Discovery Today*, 2 (1997) 382-384.
2. Rishton, G.M., Nonleadlikeness and leadlikeness in biochemical screening, *Drug Discovery Today*, 8 (2003) 86-96.
3. Rishton, G.M., Failure and success in modern drug discovery: guiding principles in the establishment of high probability of success drug discovery organizations, *Medicinal Chemistry*, 1 (2005) 519-527.
4. Olah, M.M. *et al.*, Strategies for Compound Selection, *Current Drug Discovery Technologies*, 1 (2004), 211-220.
5. Huth, J.R. *et al.*, ALARM NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens, *J. Am. Chem. Soc.*, 127 (2005) 217-224.
6. McGovern, S.L. *et al.*, A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening, *J. Med. Chem.*, 45 (2002) 1712-1722.
7. Seidler, J. *et al.*, Identification and prediction of promiscuous aggregating inhibitors among known drugs, *J. Med. Chem.*, 46 (2003) 4477-4486.
8. Roche, O. *et al.*, Development of a virtual screening method for identification of 'frequent hitters' in compound libraries, *J. Med. Chem.*, 45 (2002) 137-142.

Chemonaut: leadlike filtering

1. Chemically reactive (sub)structures

- ...aldehydes, isocyanates, hydrazones and hydrazides, maleimides...
...[C;H1](*)=O, N=[C,S]=[N,O,S]...

2. Biologically unstable, cytotoxic (sub)structures and frequent hitters

- ...thiols, quinones, thioketones, aminothiophenes, diphenyl diazenes...
...[S;H1]*, O=C1C=CC=CC1=O...

3. Chemical structures difficult in follow up

- ...structures containing: x CF₂ groups, nitrate and sulfate groups, >2 nitro groups, no hetero-atoms...
...FC(F)(*)C(F)(F)C(F)(F)*, [#8-][N+](=O)O!@-*...

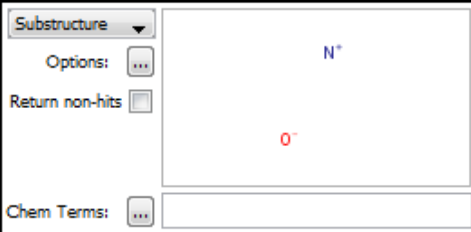
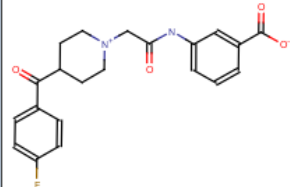
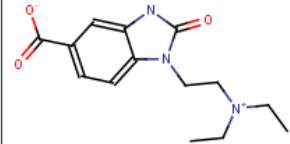
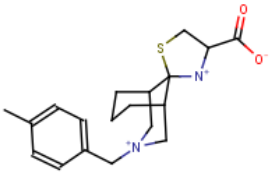


Chemonaut: future developments

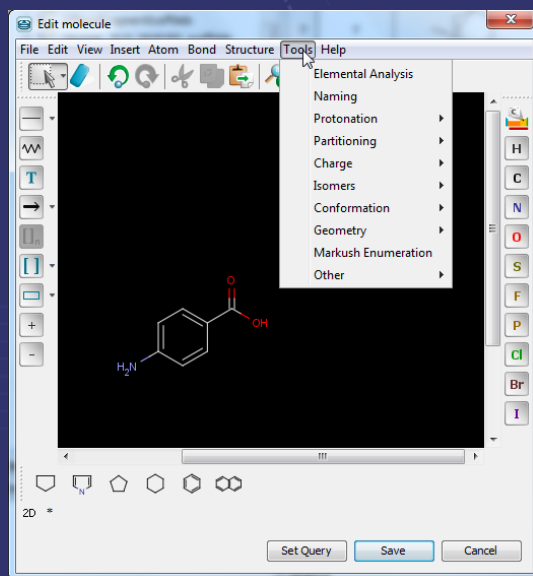
- **Tautomer enumeration**
- **pKa prediction and enumeration of ionized macrospecies**
- **Conformer enumeration and 3D substructure search**

pKa prediction and enumeration of ionized macrospecies

- advanced search options
- structure preparation for virtual screening

Structure		CdId	Structure	Mol Weight	Formula	IDNUMBER
Substructure		1		384.40	C21H21FN2O4	PYX 00004098
Options: ...		2		277.32	C14H19N3O3	PYX 00005416
Return non-hits <input type="checkbox"/>		3		347.49	C19H27N2O2S	PYX 00011410
Chem Terms: ...						

Protomers: ChemAxon pKa predictor



pKa Options

Decimal places: 2

Mode: macro

Acid/base prefix: dynamic

Min basic pKa: -10

Max acidic pKa: 20

Temperature (K): 298

Show microspecies distribution

Microspecies

pH lower limit: 0

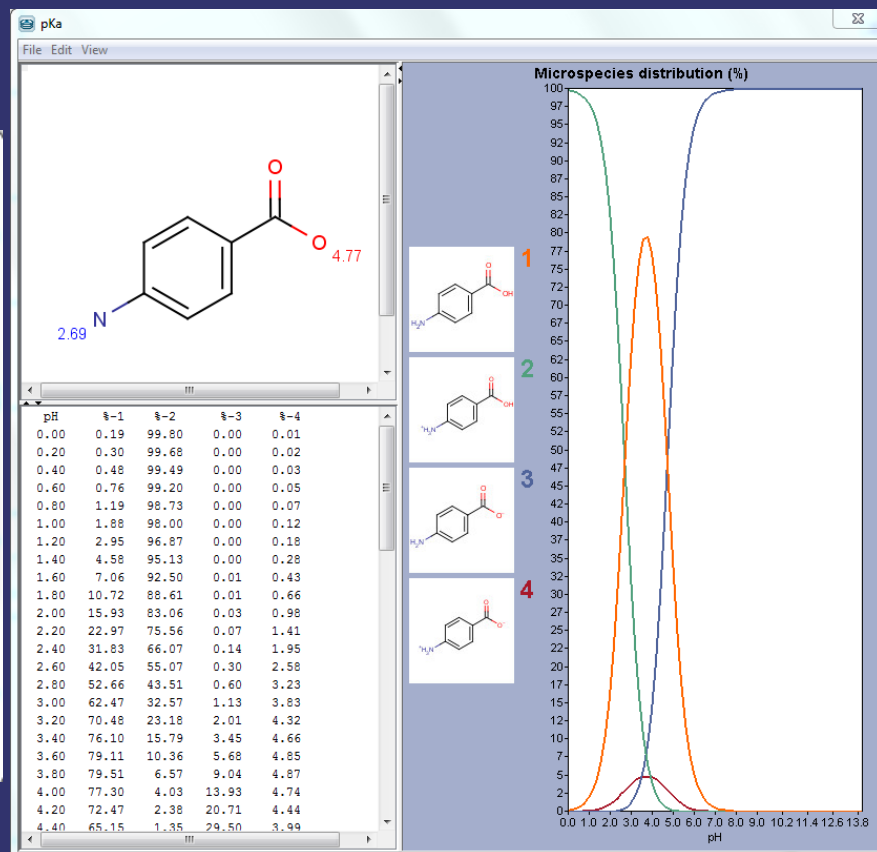
pH upper limit: 14

pH step size: 0.2

Use correction library

Take major tautomeric form

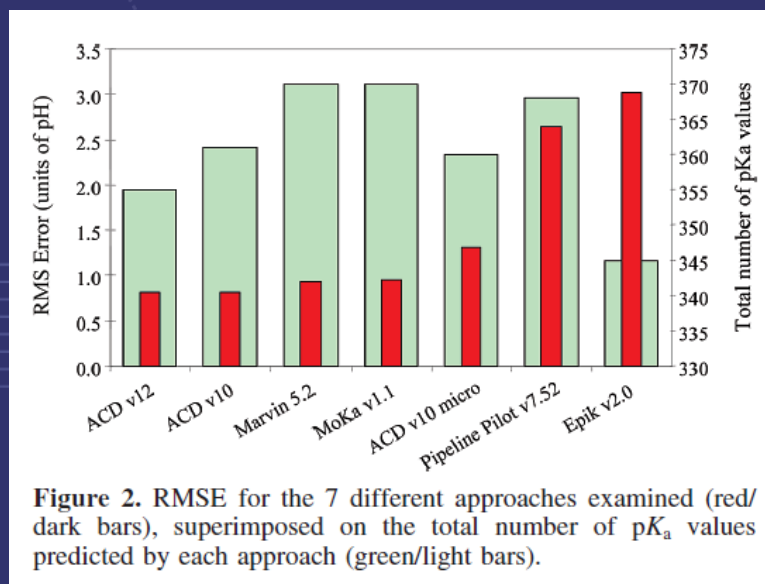
OK Cancel Restore Defaults



`cxcalc pka -i -10 -x 20 -a 3 -b 3 Input.sdf >Output.tbl`

Protomers: ChemAxon pKa predictor

- Good agreement with experimental values in terms of RMSE
- Good detection success rate
- Predicts many protomers



	measured	batch pKa (standard settings)	batch pKa (-i -2 -x 12)
basic groups	135	249	230
acidic groups	7	43	47

Manchester *et al.* J. Chem. Inf. Model. 2010 xxx xxx

Lee *et al.* J. Chem. Inf. Model. 2009 49 2013

Liao *et al.* J. Chem. Inf. Model. 2009 49 2801

Protomers: ionized forms enumeration

- `"cxcalc msdistr -H 7.4 Input.sdf > Output.sdf"`
- command line to write the microspecies distribution of an "Input.sdf" at given pH into an "Output.sdf"
- "Output.sdf" contains protonated microspecies at pH 7.4 and [%] distribution value of microspecies at this pH
- **no -i, -x, -a, -b nor pH range or [%] distribution min**
- **~30 protomers per molecule (leadlike molecules)**



Chemonaut

Screening - 3,916,796 structures from which 2,539,616 are UNIQUE
19 suppliers

Building Blocks - 622,279 structures from which 452,770 are UNIQUE
46 suppliers

Special - 887,156 structures from which 591,336 are UNIQUE
20 suppliers

Overall - 5,426,231 structures from which 3,151,539 are UNIQUE
51 suppliers

Virtual - 11 million structures
9 suppliers

An abstract graphic on a dark blue background. It features a central circle with a right-pointing arrow. From this central circle, several lines radiate outwards to other circles: one to a top-left circle with a downward arrow, one to a bottom circle, and a series of horizontal lines to the left. The text "Thank you" is overlaid on the central circle.

Thank you