

INTERACTIVE COMPOUND EDITOR FOR CAREFUL QSAR BUILDERS...

...AND THE PSYCHOANALYSIS OF CHEMAXON PLUGINS

1

Dr. Dragos “Freud” Horvath

Laboratoire d’InfoChimie

UMR 7177 CNRS – Université de Strasbourg

THE GOAL

Take a set of molecules to be used for QSAR model building, and ensure that their chemical representation is compatible to the planned property modelling... that is, “standardize” them

- I. Remove counterions
- II. Remove too large species
- III. Render them neutral (except for quaternary cations), with split charges on N oxides
- IV. Visualize “problematic” species, in order to edit, interactively choose preferred possible structure[s], or just discard (and prevent headaches)
- V. Maintain the association of the structures to activities, in spite of potential duplications (multiple alternative renderings/structure) or structure deletions

THE MEANS

- Interactive *tcs*h script, taking a compound/activity file and a series of optional parameters as arguments..
 - Sequentially calls either the standardizer tool with specific rules, or...
 - various stand-alone java programs based on the ChemAxon plugins
 - If different options emerge, interactively displays the proposed renderings of that structure in a multi-cell *mview* window, and lets the user pick one or several of these
 - Eventually, the user may call up the structure in *msketch* and render it as desired.
 - Specified or edited structures are output, associated to the initially given activities (if several tautomeric forms are accepted, they will all have a common activity value)

I – PRELIMINARIES: GENERAL STANDARDIZATION ISSUES

PRELIMINARIES:

- Dehydrogenize
- Remove covalently bound alkaline & second group metals [they should have been drawn as ions!]
- Remove compounds above maximal tolerable size (100 heavy atoms by default)
- Interactively display compounds below the maximal tolerable, but above some practically acceptable size, and ask user whether they should be kept
- Detect counterion(s) as the accompanying species with less than 1/3 of atoms than the “main” species and remove them
- If multiple substructures of size ratio below 3:1 exist, show them and let the user pick the “active principle”

TYPE (2) TO KEEP ACETATE, (3) FOR IMIDAZOLE,
(I) TO DISCARD THIS SALT, (M) TO EDIT:

The image shows a terminal window on the left and a chemical structure viewer on the right. The terminal window, titled 'uuu_temp : interact_stand.', displays the following text in green: 'Processing compound ID=1, nr. 1 out of 1 dubious entries', 'Select any number to keep one of the mview options, i to ignore or m to manually edit structure', and 'Your choice: []'. The chemical structure viewer, titled 'STDIN - MarvinView 5.3.8', shows three panels labeled 1, 2, and 3. Panel 1 contains the chemical structure of imidazole. Panel 2 contains the chemical structure of acetic acid. Panel 3 contains the chemical structure of imidazole. The structures in panels 2 and 3 are drawn with red lines for the oxygen and hydroxyl groups, while the structures in panel 1 are drawn with blue lines.

uuu_temp : interact_stand.
File Edit View Scrollback Bookmarks Settings Help
Processing compound ID=1, nr. 1 out of 1 dubious entries
Select any number to keep one of the mview options, i to ignore or m to manually edit structure
Your choice: []

STDIN - MarvinView 5.3.8
File Edit View Table Tools Help

1 2 3

HN
N

O
OH

O
OH

N
NH

II.1 – TAUTOMERIZATION: SOME SUBLIMINAL ISSUES...

PROBLEMATIC SPECIES: 1. TAUTOMERS

- Compounds potentially appearing under several tautomeric forms
 - Enumerate [user toggle] (1) the “major” or (2) up to nine “dominant” tautomers, according to the *tautomerization* plugin (called by a standalone *Tautomerize.java*).
 - If the input form coincides with the one and only form returned by the tautomer enumeration tool, fine...
 - ... else, go interactive and ask the user to select one or more amongst the tautomers shown in a *mview* table.
 - Pray that the latter does not happen needlessly!!
 - If none of the shown tautomers satisfy your innate chemical intuition, interactively pass the structure to *msketch* and modify it at your will
 - If you're clueless, and no one watches you doing it, **ignore** the compound.

```

TautomerizationPlugin taut=new TautomerizationPlugin();
//taut.setTakeDominantTautomers(true);
taut.setDominantTautomerDistributionCalculation(true);
taut.setMaxStructureCount(9);

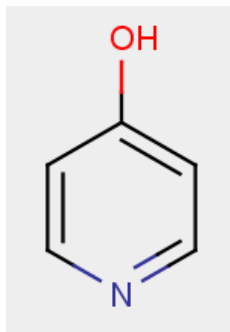
//retrieve filename and ph if given
for (int i=0;i<args.length;i++)
{
    if (args[i].equals("-f"))
    {
        if (args[i+1] != null)
        {
            fileName = args[i+1];
            file = new File(fileName);
        }
    }
    else if (args[i].equals("-pH"))
    {
        if (args[i+1] != null)
        {
            pH=Double.parseDouble(args[i+1]);
            taut.setpH(pH);
        }
    }
    else if (args[i].equals("-major"))
    {
        taut.setDominantTautomerDistributionCalculation(false);
        taut.setMaxStructureCount(1);
        taut.setTakeMajorTautomer(true);
    }
}
}

```

OOPS: TAUTOMERIZATION OUTPUT DEPENDS ON INPUT STRUCTURE RENDERING...

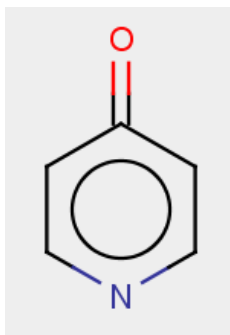
- echo "O=C1C=CNC=C1" | standardize -c "tautomerize"

- **OC1=CC=NC=C1**



- echo "O=c1ccncc1" | standardize -c "tautomerize"

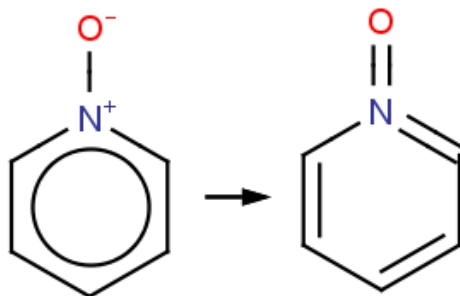
- **O=c1ccncc1**



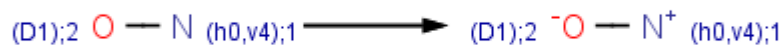
- Therefore, went back to the preliminaries and added a “dearomatize” directive to the standardizer at that step.

YET, THE TAUTOMERIZER DOESN'T LACK SELF CONFIDENCE AND DARES TO ALTER STRUCTURES!

- It *sometimes* (??) may dearomatize all by itself, even if not asked to, and will, by any means, restore five-legged N atoms...
- `echo "c1ccc[n+]([O-])c1" | standardize -c "tautomerize"`
 - `O=N1=CC=CC=C1`

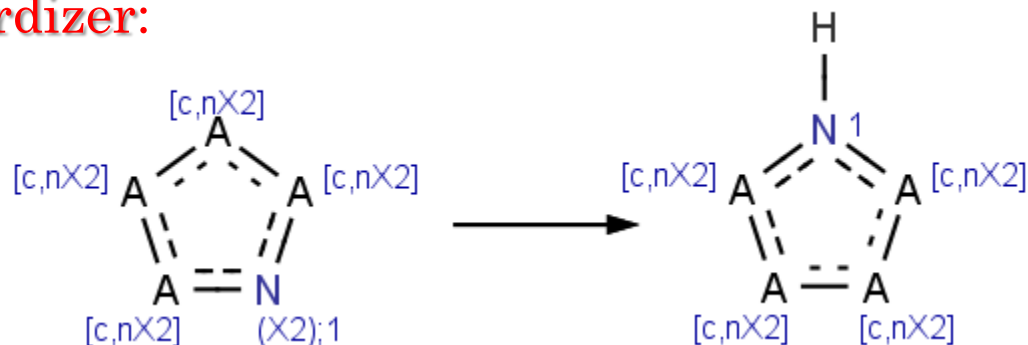


- Therefore, each of the enumerated dominant tautomers in *Tautomerize.java* is being restandardized immediately after its birth .



OOPS²: YOU CANNOT DEAROMATIZE IMIDAZOLE ?!

- echo "Cc1ncnc1" | standardize -f smiles:-a -c "dearomatize"
 - **Cc1ncnc1**
- echo "Cc1[nH]cnc1" | standardize -f smiles:-a -c "dearomatize"
 - **CC1=CN=CN1**
- Since one cannot assume that the Wise Chemist drew an explicit H on one of the imidazole nitrogens in order to "break the symmetry", one needs some pretty ugly rules to be added to the preliminary stage standardizer:



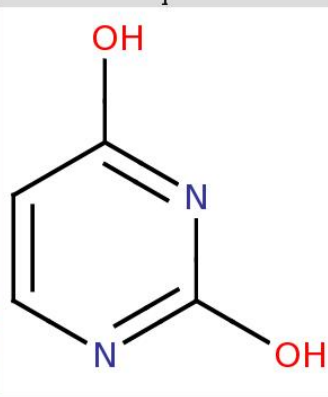
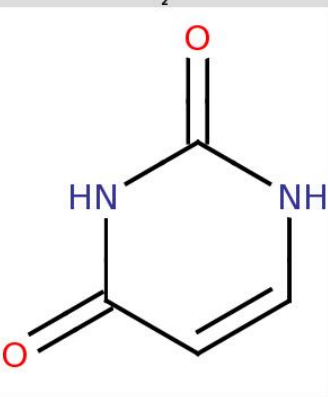
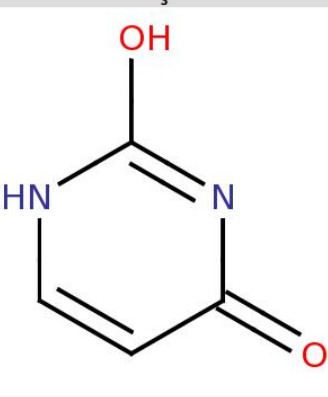
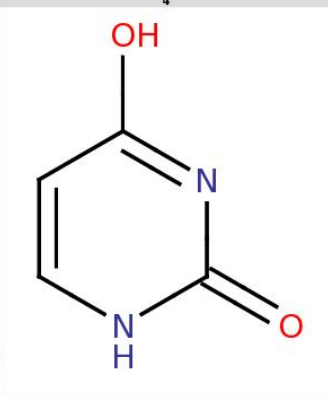
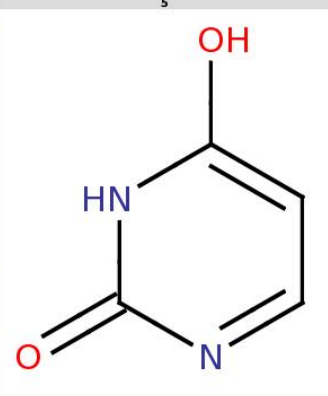
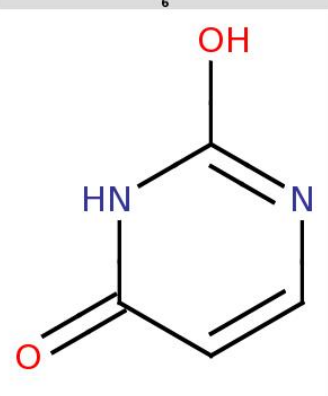
II.2 – TAUTOMERIZATION: THE REWARDS & FRUSTRATIONS OF TAUTOMER ANALYSIS

- In how far does the output of the tautomer plugin satisfy chemical intuition (the best tautomer relevance score known so far)?
- How often did it return absurd alternatives next to the correct ones, and thus forced us to go interactive for no reason?

```
test_temp : interact_stand.
File Edit View Scrollback Bookmarks Settings Help
Processing compound ID=1, nr. 1 out of 1 dubious entries
Select any number to keep one of the mview options, or multiple options separated by -,
i.e. type 1-2 to store both compounds displayed in cells 1 and 2 as independent molecules
Other options: i to ignore or m to manually edit structure (in this version, you may sketch multiple entries too)
```

Your choice:

STDIN - MarvinView 5.5.0.0

1 	2 	3 	4 
5 	6 	7	8

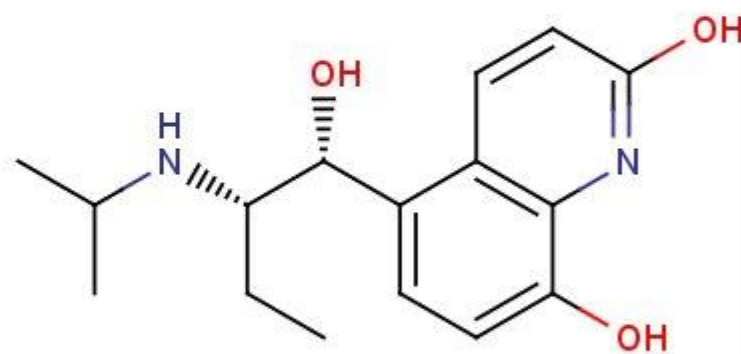
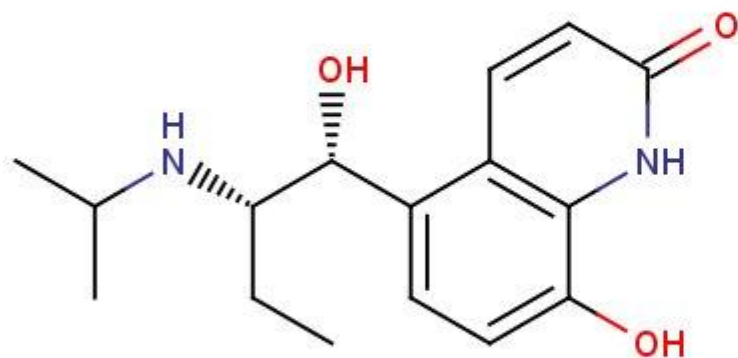


OFTEN, TAUTOMERIZATION COMES UP WITH VALID ALTERNATIVE(S) TO INPUT STRUCTURE:

File Edit View Table Tools Help

1

2



..WORKS FINE WITH LARGE COMPOUNDS:

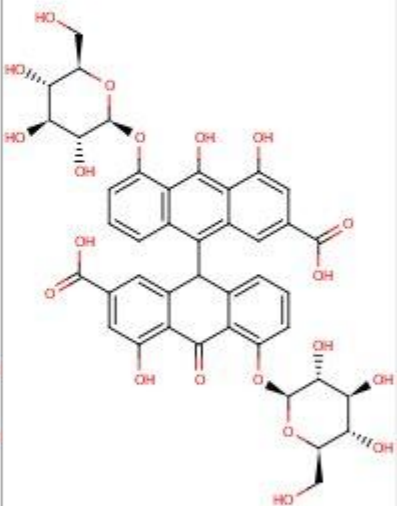
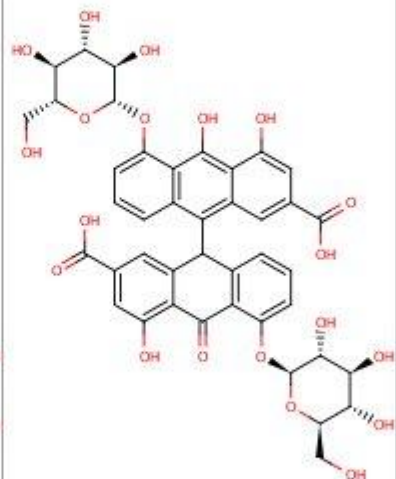
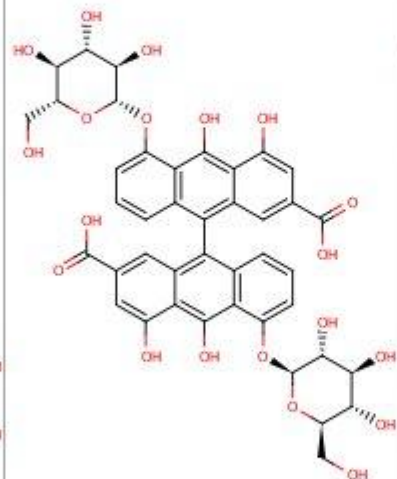
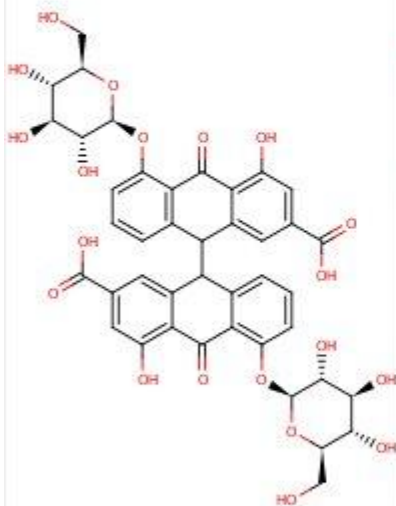
File Edit View Table Tools Help

1

2

3

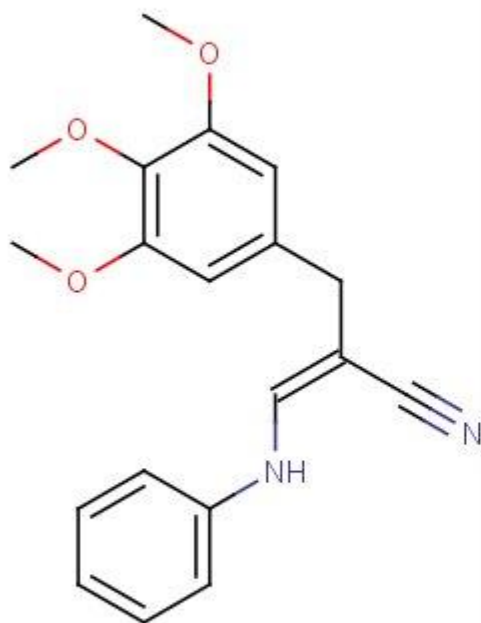
4



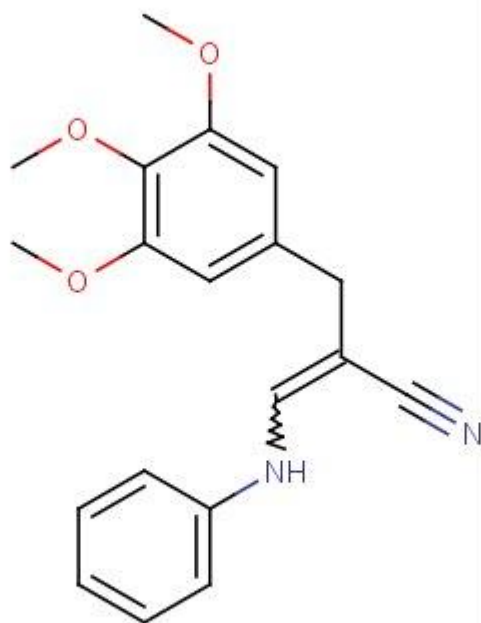
IT EVEN GOES AS FAR AS SIGNALLING THAT STEREOCHEMISTRY MAY CHANGE DUE TO TAUTOMERISM!

File Edit View Table Tools Help

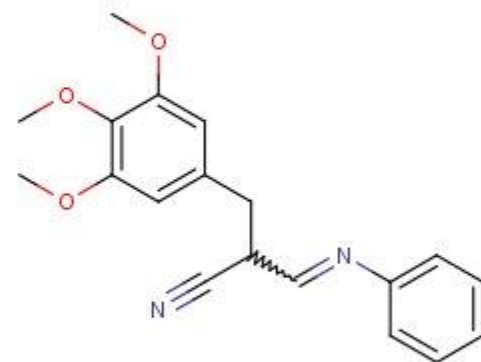
1



2



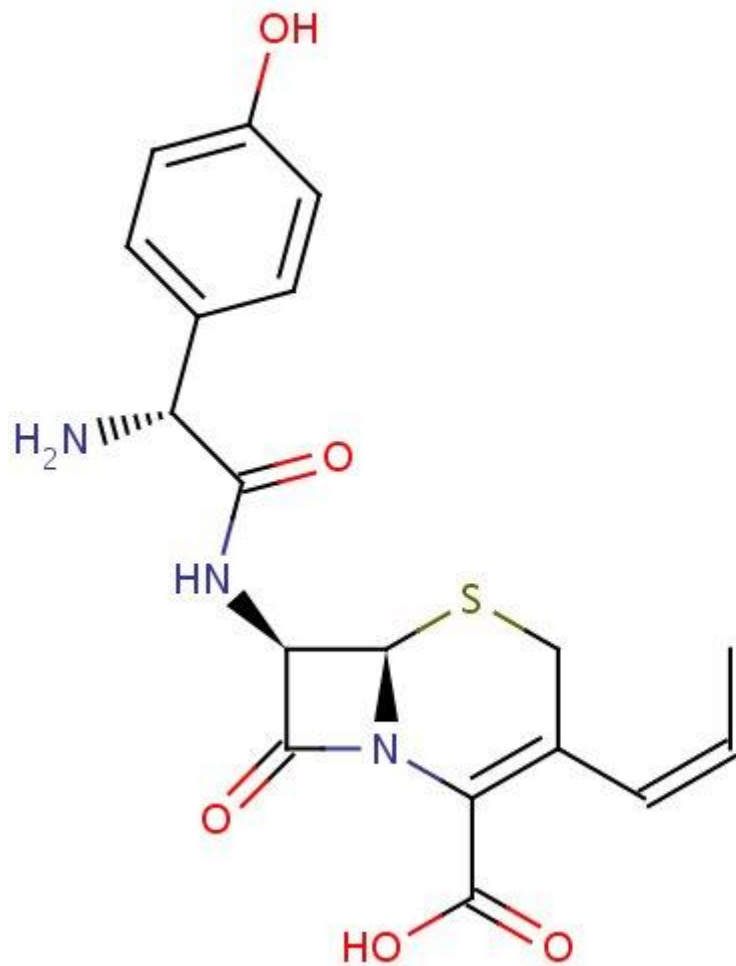
3



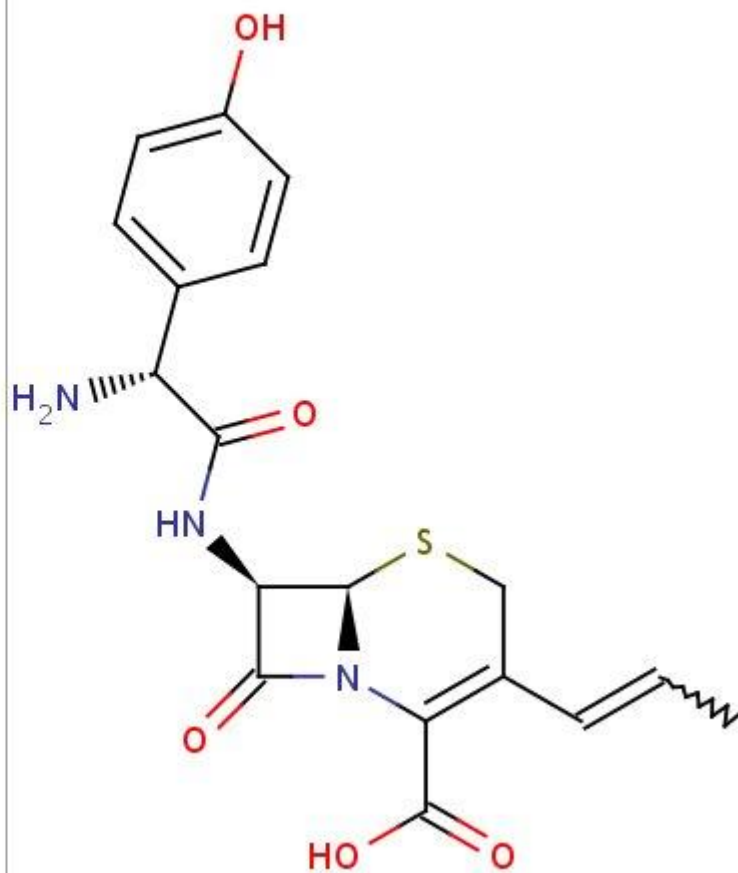
...BUT THEN IT GOES TOO FAR AND SEES
LABILE DOUBLE BONDS EVERYWHERE!

File Edit View Table Tools Help

1

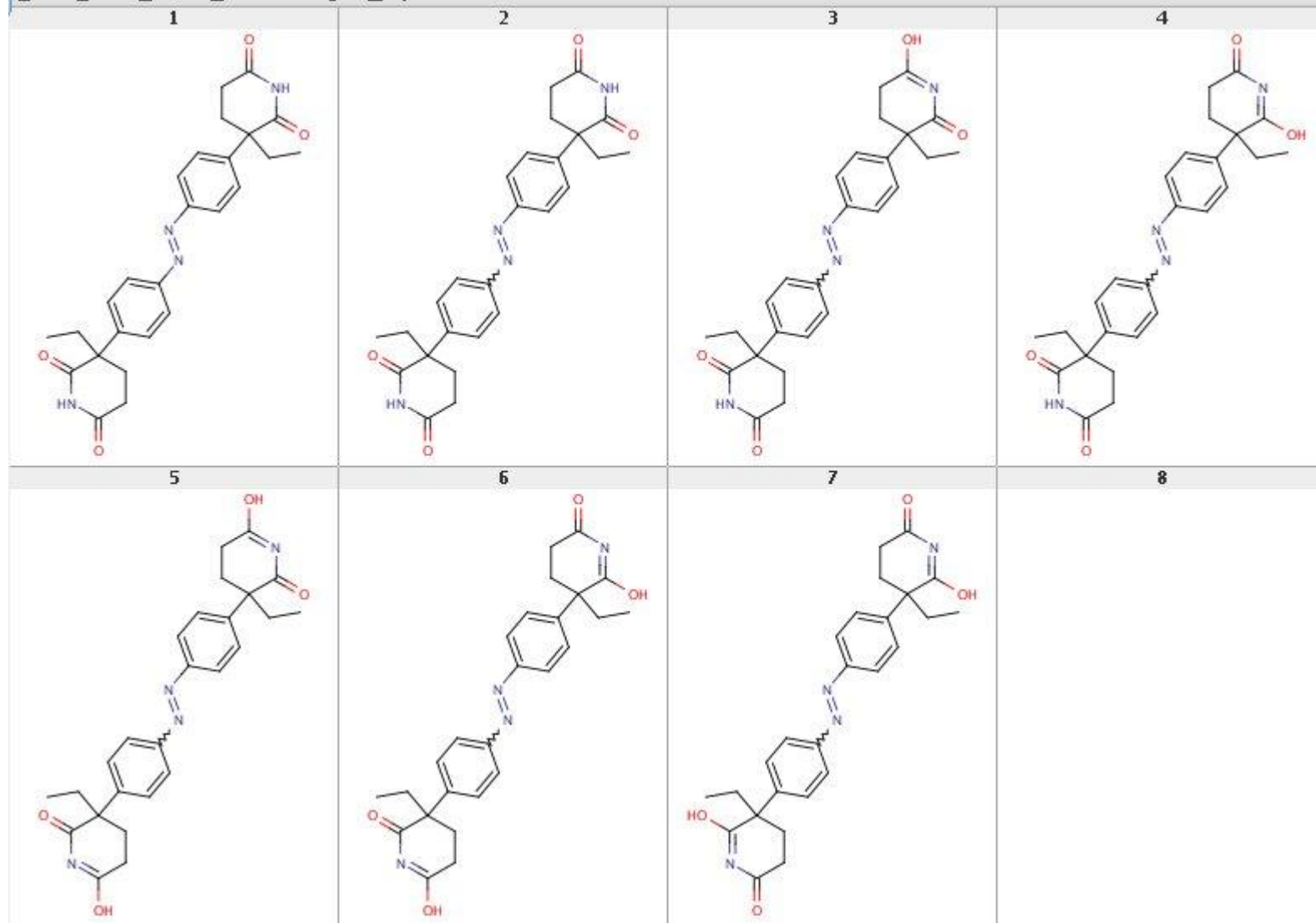


2



...REALLY EVERYWHERE, EVEN THOUGH THERE IS NO TAUTOMERISM!

File Edit View Table Tools Help

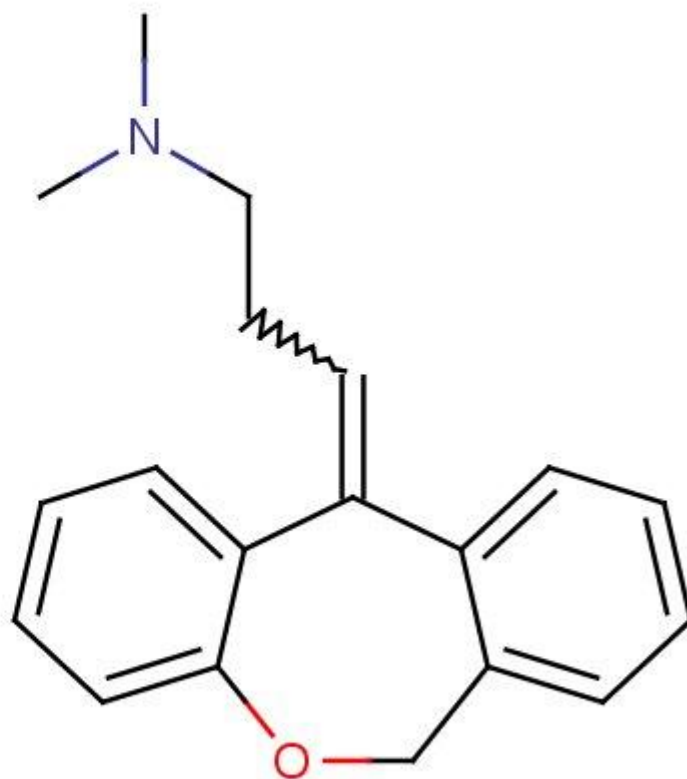
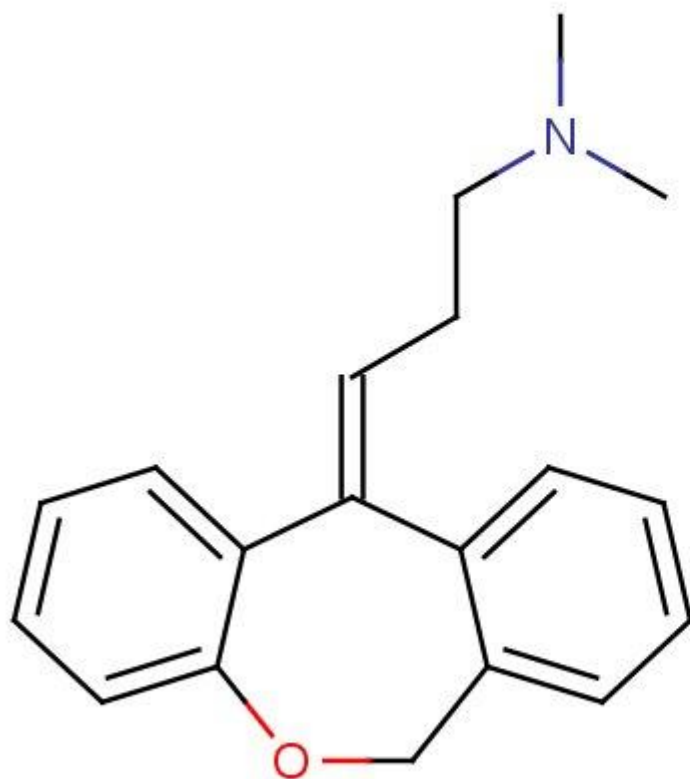


...REALLY NO TAUTOMERISM!

File Edit View Table Tools Help

1

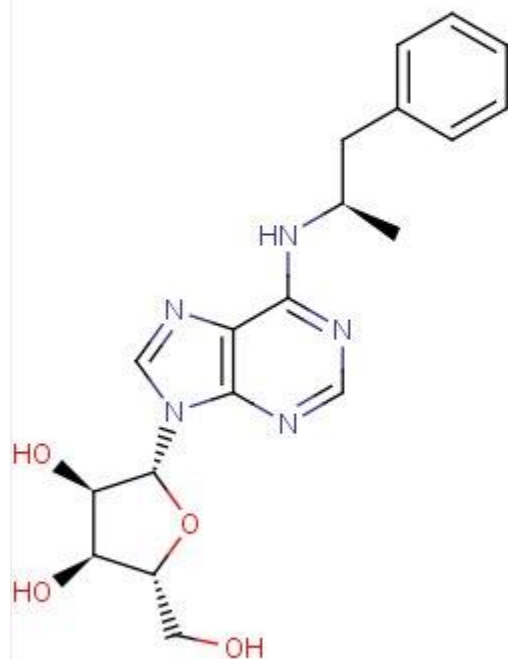
2



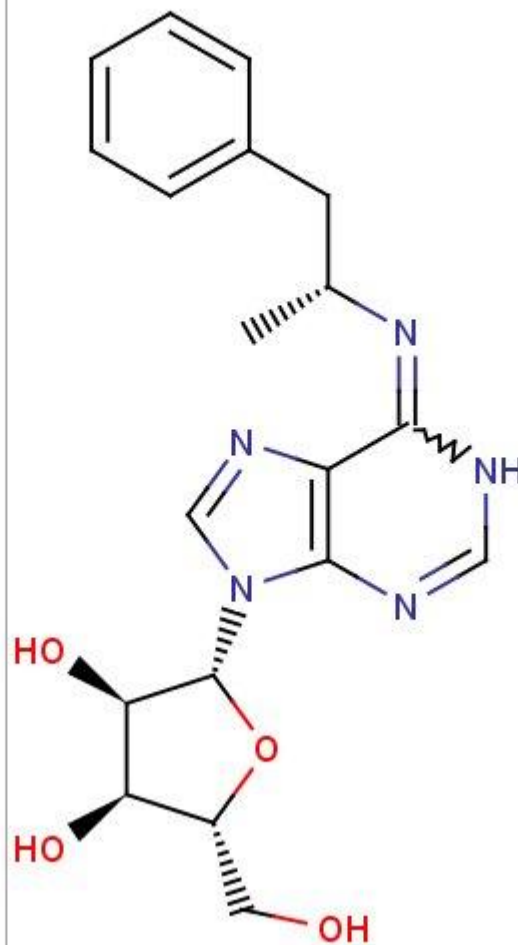
...OR HAS A HARD TIME WHEN STUMBLING UPON
DOUBLE BONDS OF UNKNOWN
STEREOCHEMISTRY.

File Edit View Table Tools Help

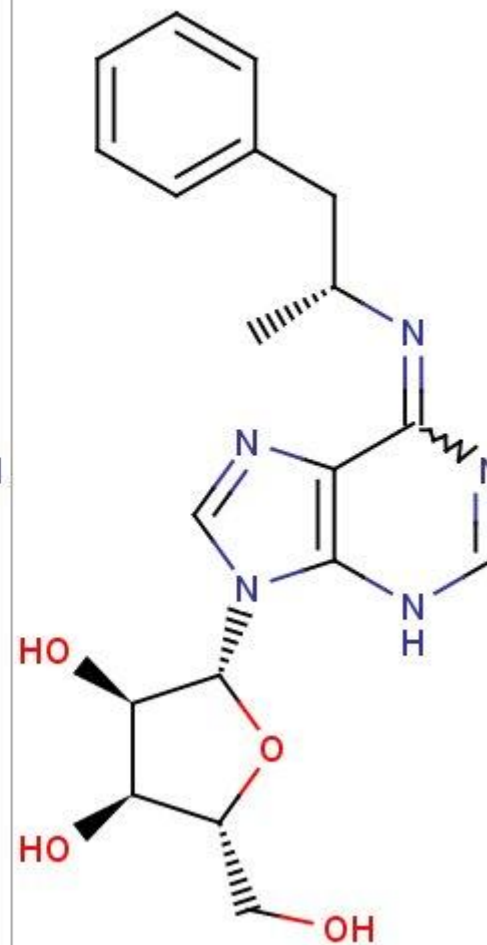
1



2



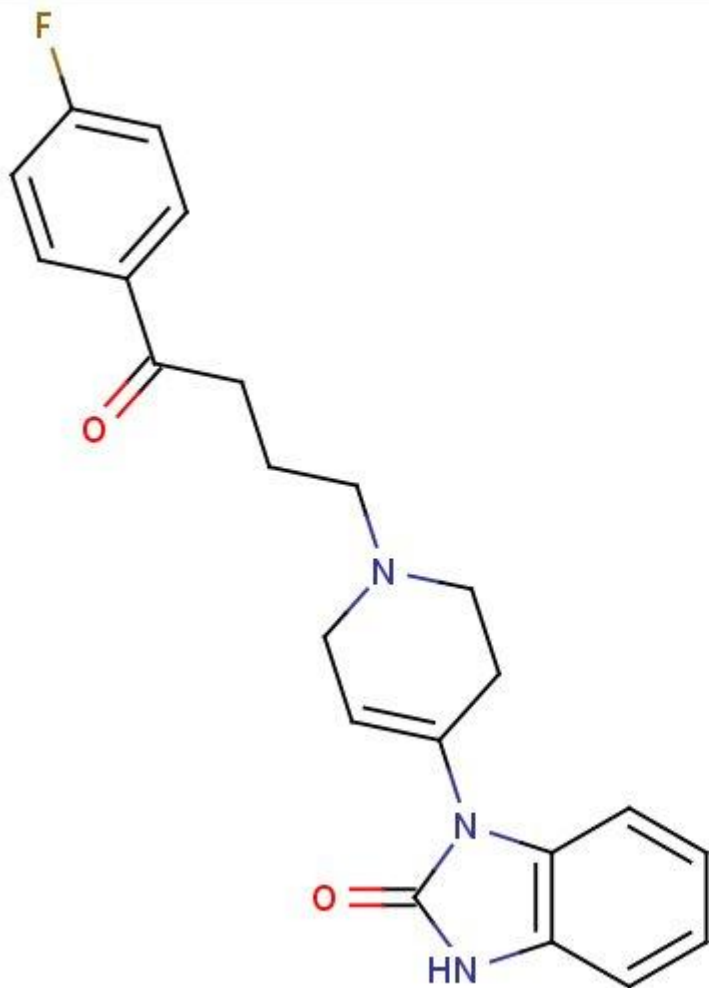
3



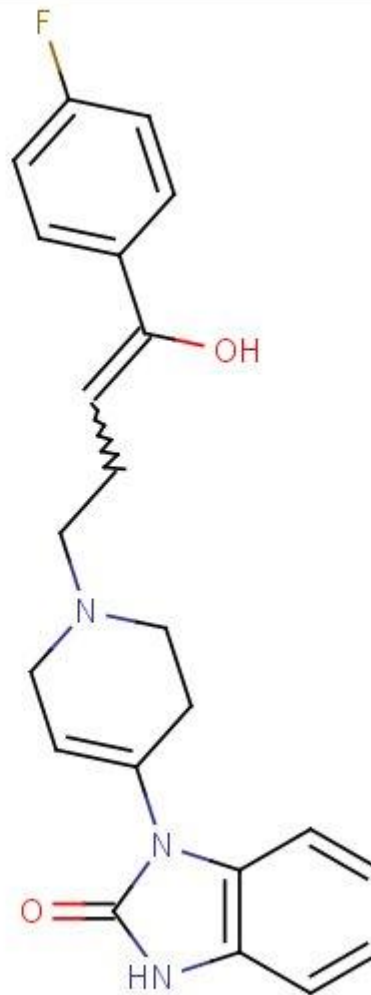
CHEMICALLY, IT IS OVERKEEN TO ENOLIZE
KETONES, ESPECIALLY THOSE WITH ELECTRON-
WITHDRAWING GROUPS.

File Edit View Table Tools Help

1



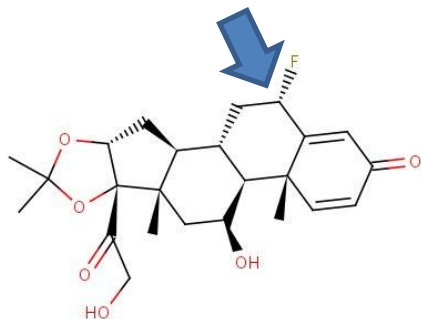
2



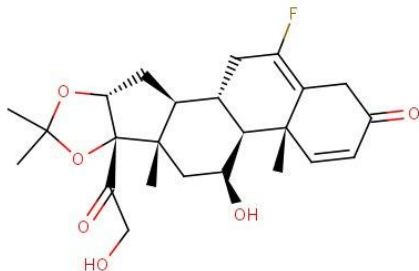
SOMETIMES IT WORKS ALL TOO WELL, AND HUNTS FOR IMPROBABLE GAMMA-TAUTOMERS

File Edit View Table Tools Help

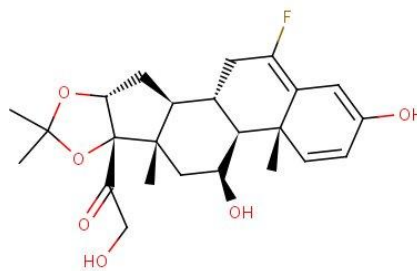
1



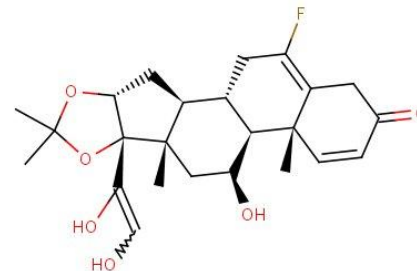
2



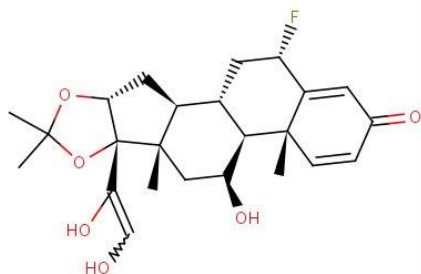
3



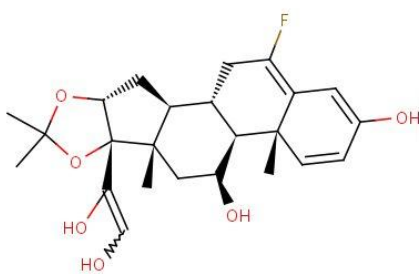
4



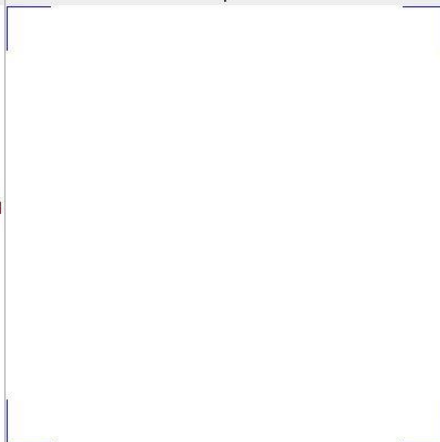
5



6



7

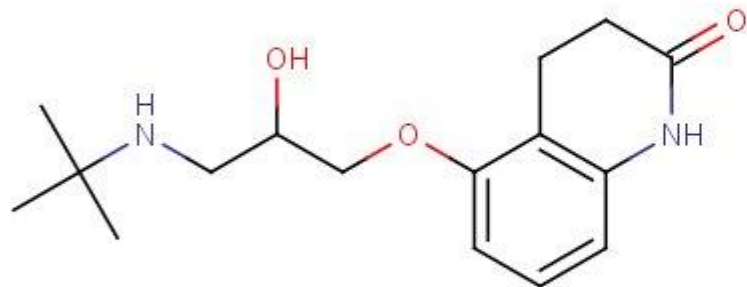


8

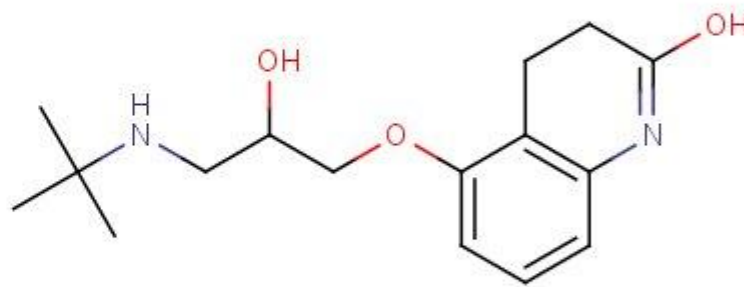
FURTHERMORE, IT DOES NOT HESITATE TO ATTACK HARMLESS CYCLIC AMIDES...

File Edit View Table Tools Help

1

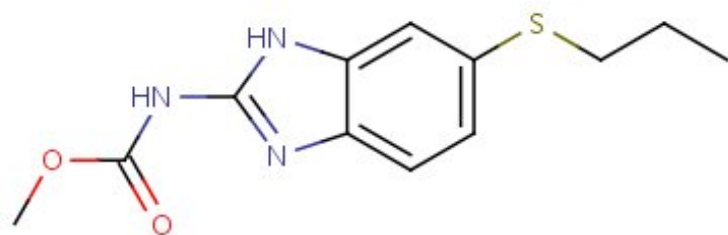


2

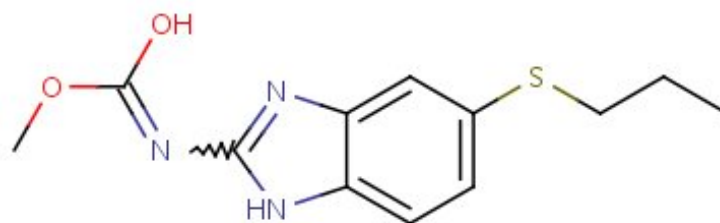


...EVEN ACYCLIC CARBAMATES...

3



4

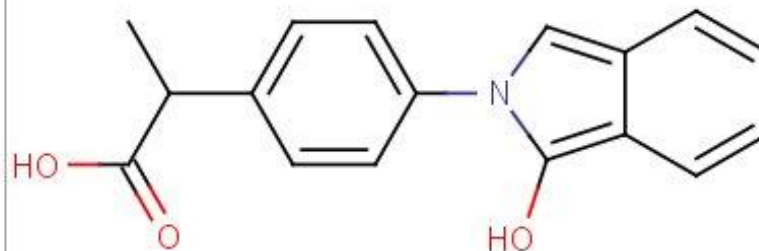
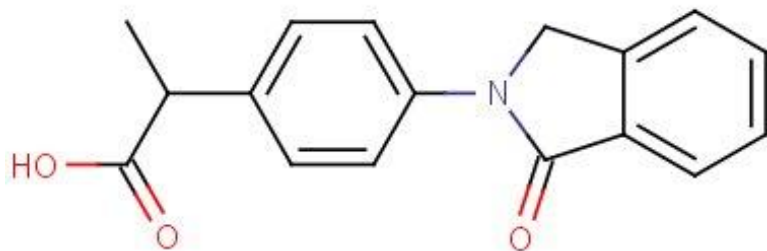


...SOMETIMES IT WOULD WALK OVER DEAD BODIES (OF BROKEN AROMATIC RINGS) TO GET THERE...

File Edit View Table Tools Help

1

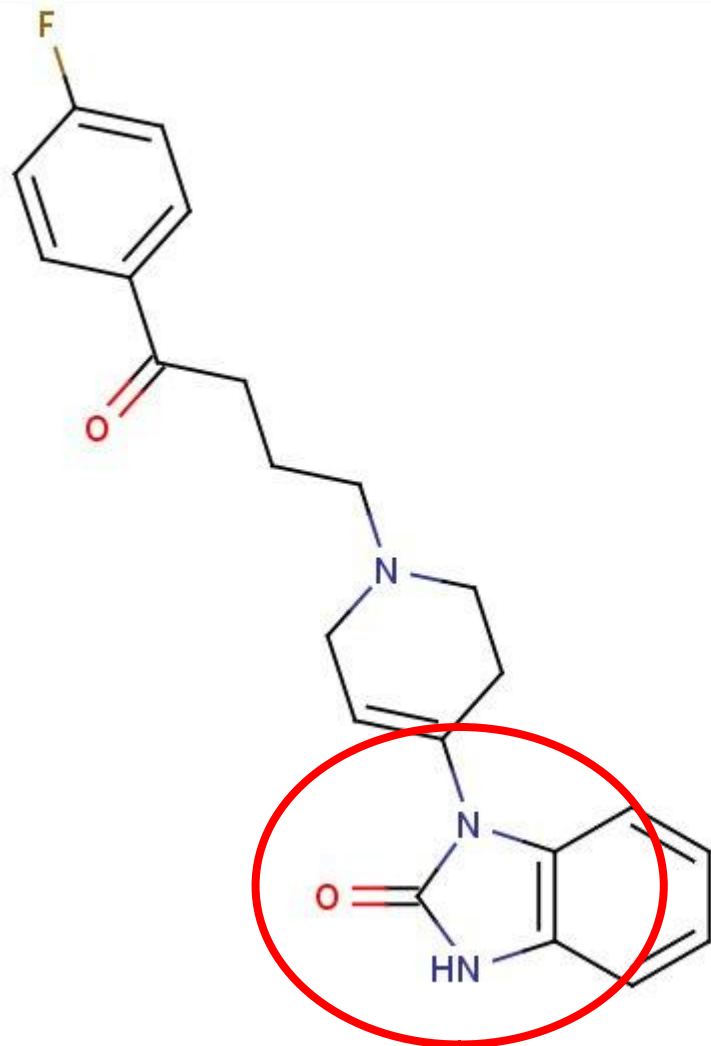
2



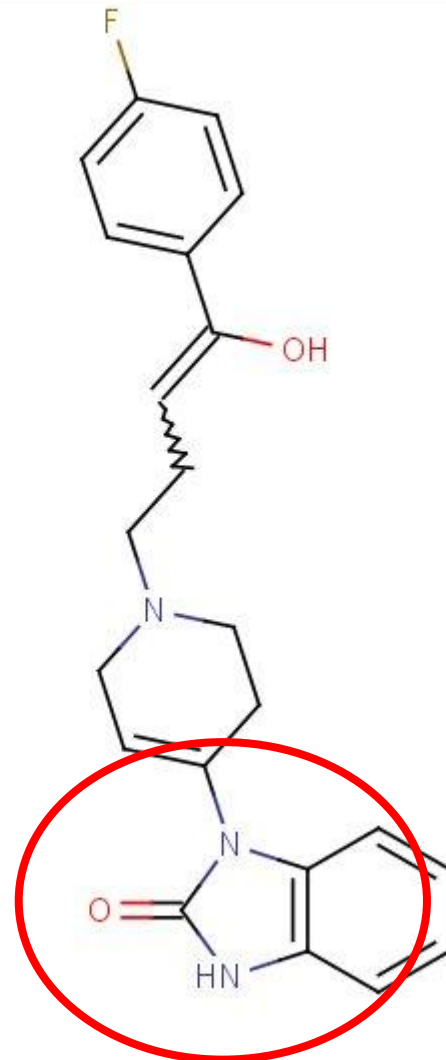
...YET, WHEN SOMETIMES ENOLIZING THE AMIDE
MIGHT SEEM TO MAKE SENSE (AROMATIC RING
CLOSURE), IT WON'T DO IT...

File Edit View Table Tools Help

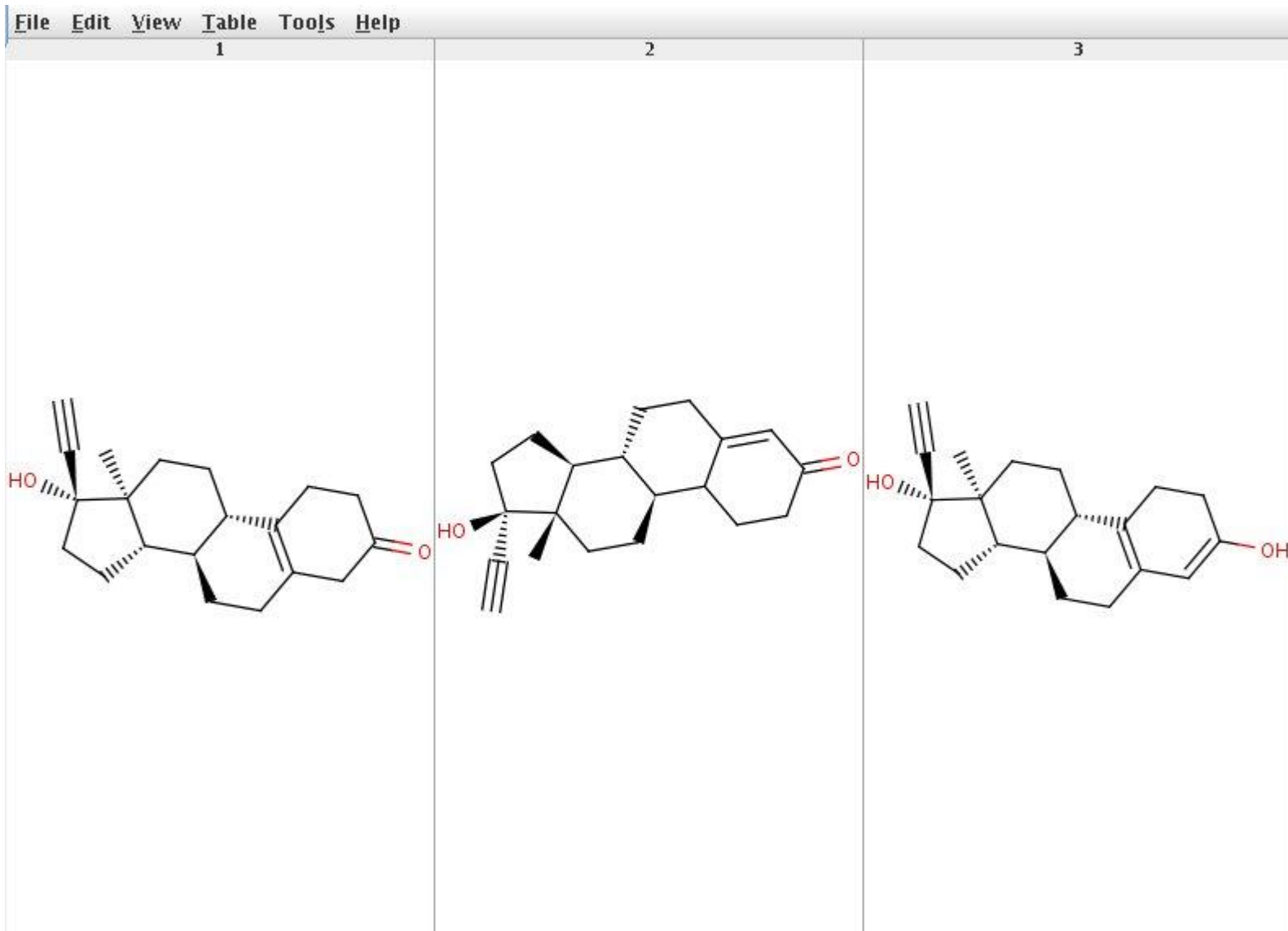
1



2



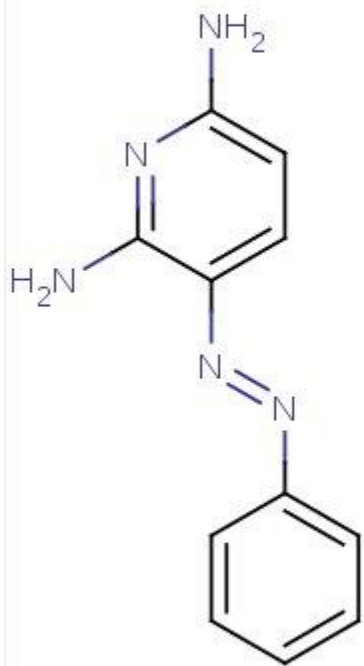
WORST OBSERVED SCENARIO: SOME CASES OF WILD DOUBLE BOND HOPPING...



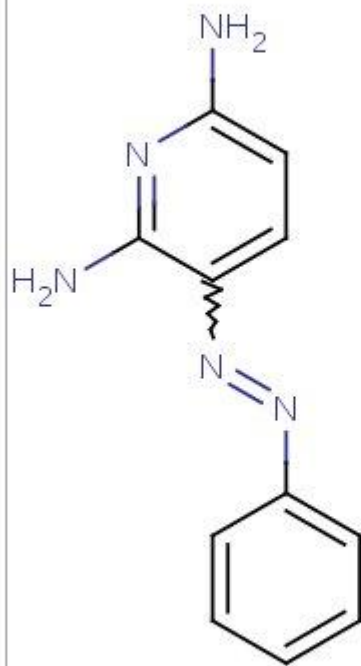
WOULD IT MAKE SENSE TO HAVE AMINOPYRIDINES ONLY AS DOMINANT TAUTOMERS?

File Edit View Table Tools Help

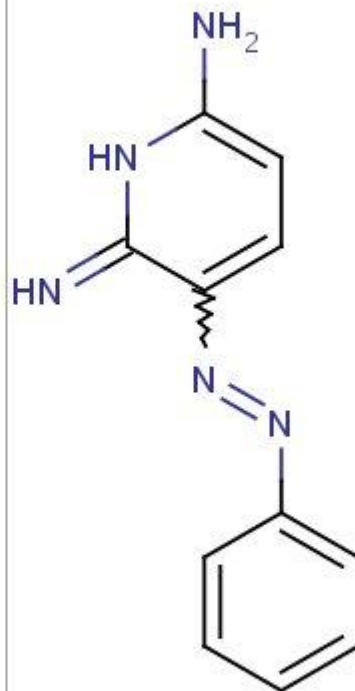
1



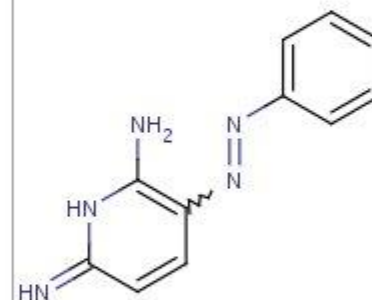
2



3

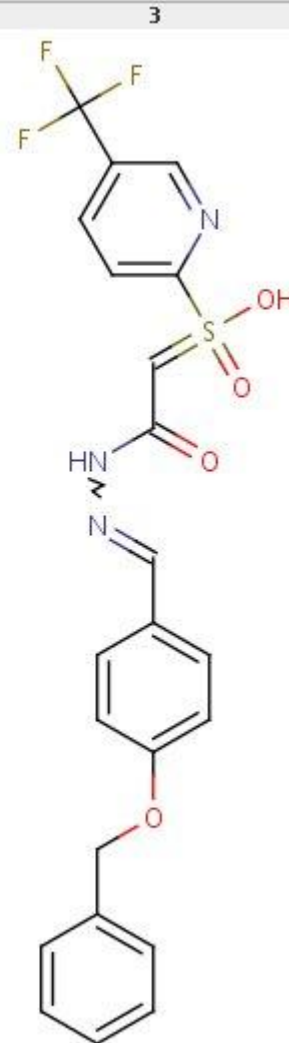
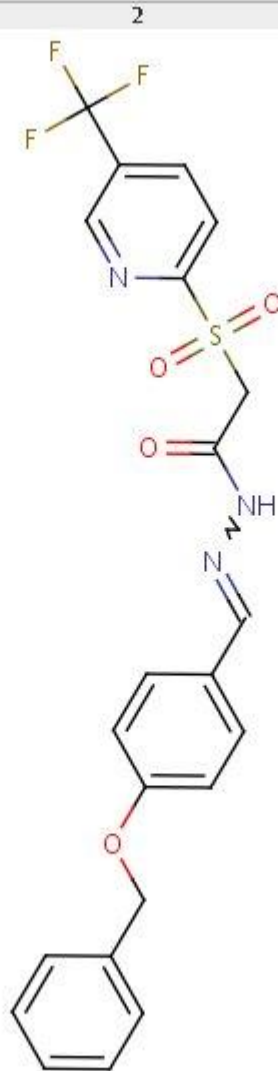
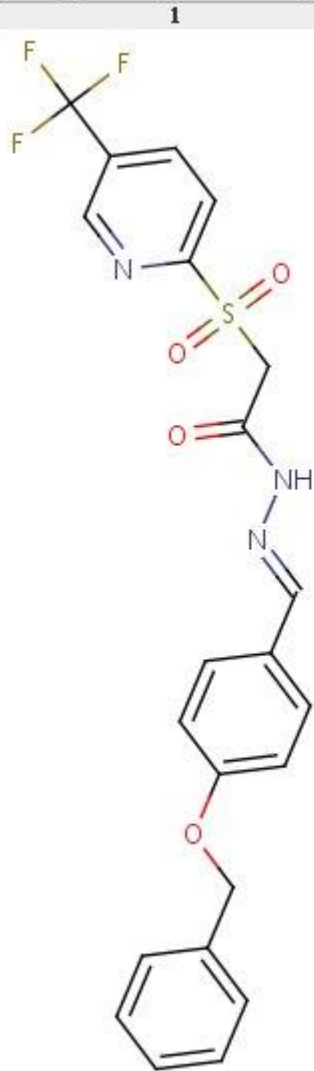


4



NO CLUE ABOUT S=C BONDS IN SULFONES, BUT WHY IGNORE THE REGULAR ENOL?

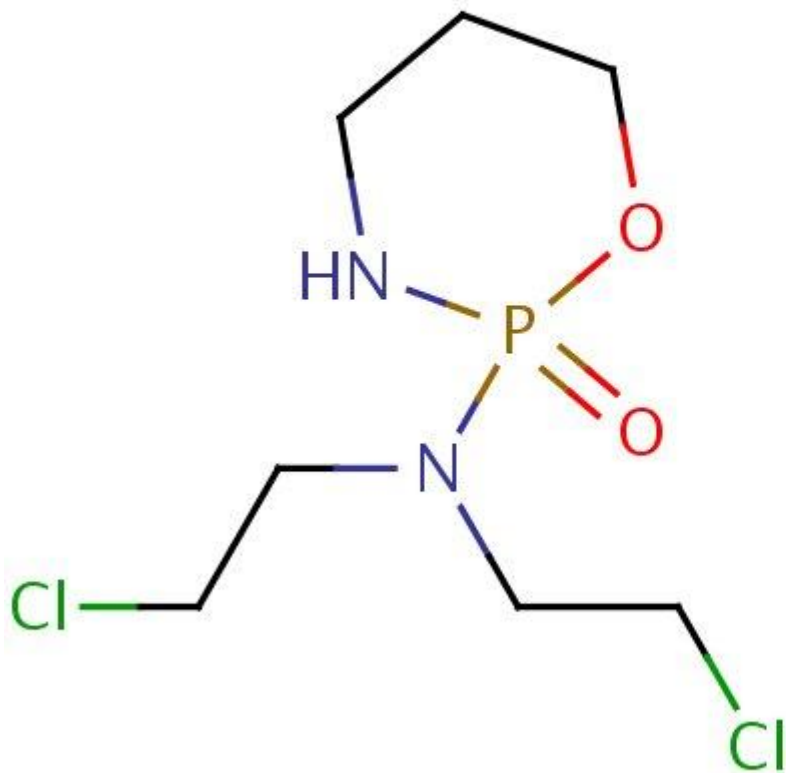
File Edit View Table Tools Help



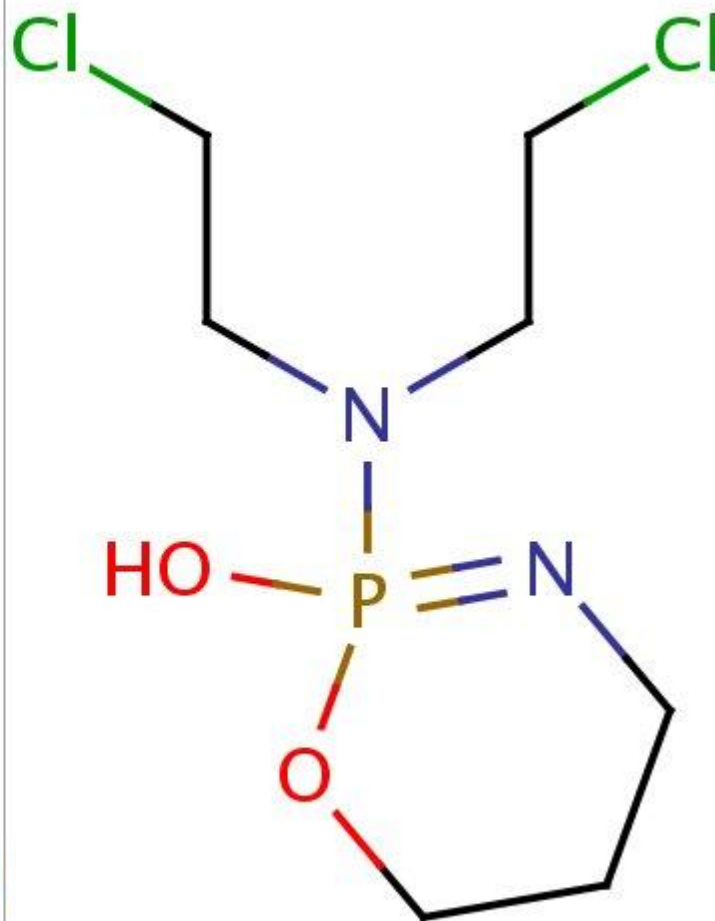
DO NOT ASK ME ABOUT THIS ONE, EITHER...

File Edit View Table Tools Help

1



2

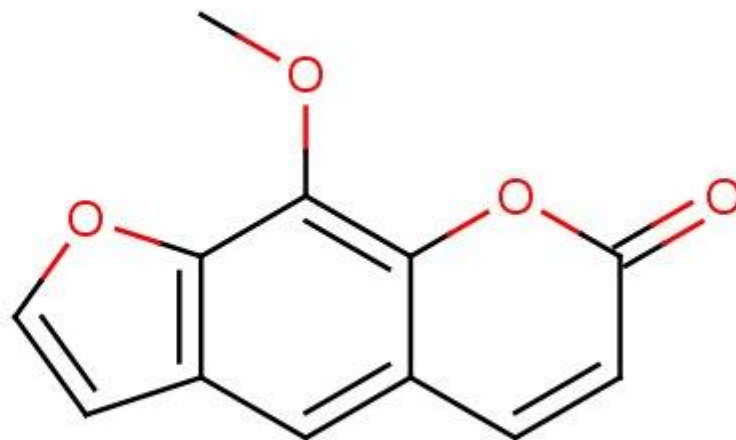
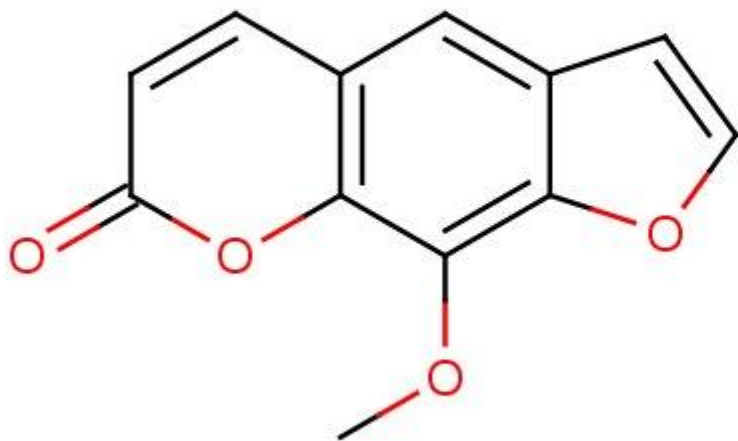


SOMETIMES THE DIFFERENT “TAUTOMERS” ARE DIFFERENT MESOMERIC FORMS, ALTHOUGH A “MESOMERIZE” DIRECTIVE WAS EMPLOYED...

File Edit View Table Tools Help

1

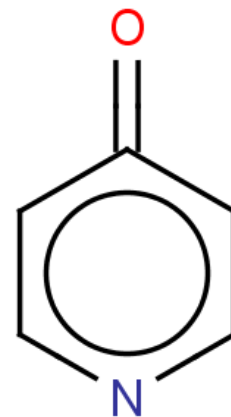
2



III – LAST BUT NOT LEAST: REAROMATIZATION

PROBLEMATIC SPECIES: 2. AROMATICS

- No “default” aggressive aromatization leading to structures with hypervalent atoms...
- Try both “basic” and “loose” aromatization approaches and, whenever their outputs diverge, show alternative renderings and let the user pick the preferred...
 - Fortunately, they rarely disagree
- On occasion, pass molecules through the pK_a plugin and flag those causing problems...



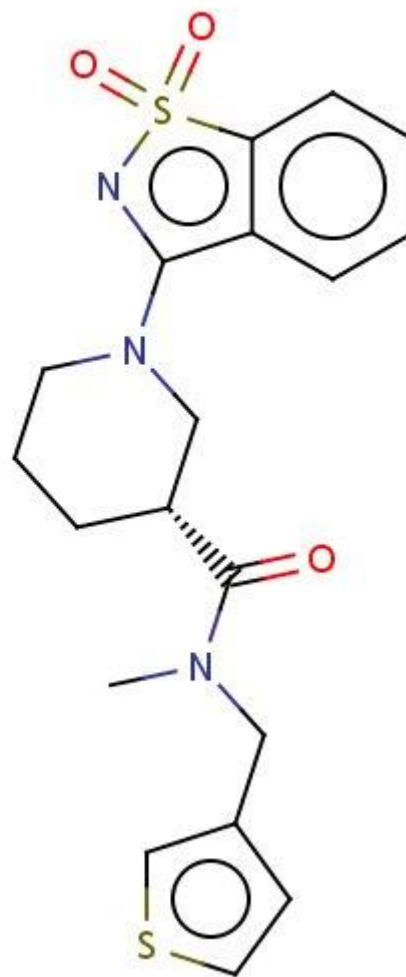
A RARE CASE OF AMBIGUOUS AROMATICITY STATUS...

File Edit View Table Tools Help

1



2

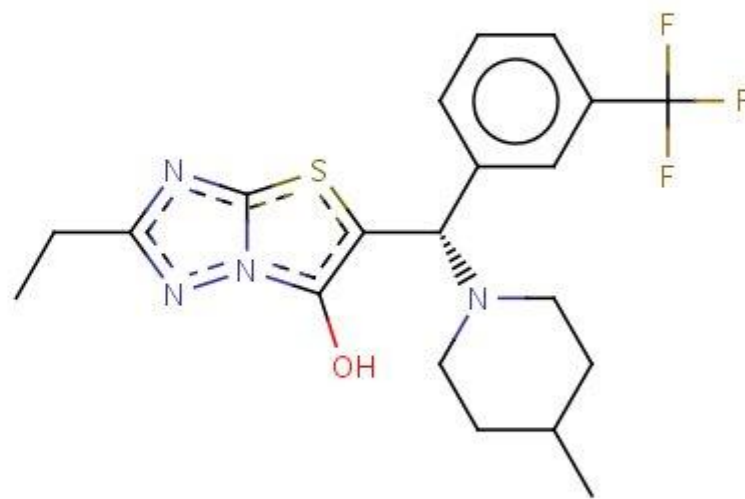
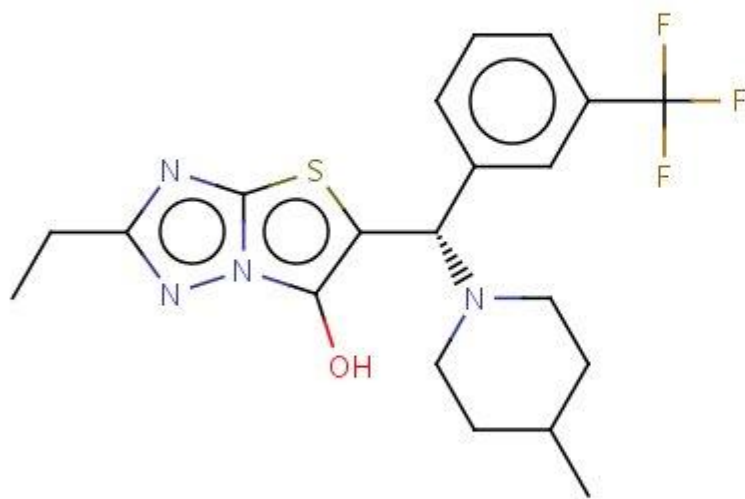


NOT QUITE SURE WHY “LOOSE” AND “BASIC” METHODS DISAGREE HERE...

File Edit View Table Tools Help

1

2



DIAGNOSIS...

- The tautomer plugin patient is generally in good mental health, although it has a leaning towards secrecy & paranoia, its behavior towards aromatic compounds being sometimes erratic.
- It also tends to uselessly complicate its existence by aiming at too subtle goals
 - impact of tautomerism on stereochemistry,
 - movement of γ -protons with respect to activating group
- The handling of aromaticity by the “basic” or “loose” procedures is fundamentally sane, avoiding Salvador Dali-style hypervalent structures.

THANK YOU FOR YOUR ATTENTION !