Tutorial:
Molecular Docking intro

Karel Berka
Václav Bazgier

Tutorial preparation

Programs installation:

- Pymol

http://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol

- Autodock Tools

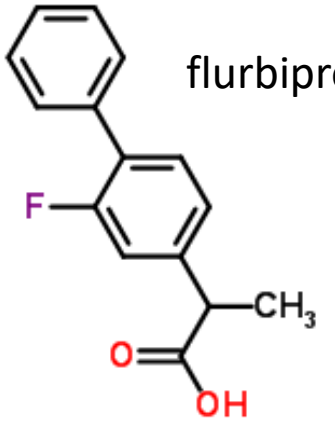
http://mgltools.scripps.edu/

- Autodock Vina

http://vina.scripps.edu/

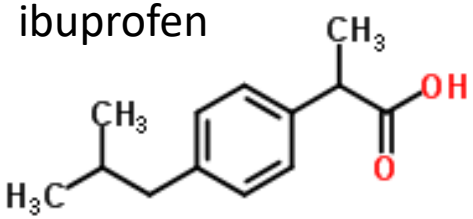
Task:

- Predict **site-of-metabolism** of some NSAID drugs (flurbiprofen, ibuprofen) which are metabolized by Cytochrome P450 2C9 (CYP2C9) (i) by **docking** into the active site of CYP2C9 or (ii) by **ligand-based** SMARTCYP web service



flurbiprofen

ibuprofen



Background: Cytochrome P450 (CYP)

Diverse superfamily of hemoproteins

=> **Drug metabolism**

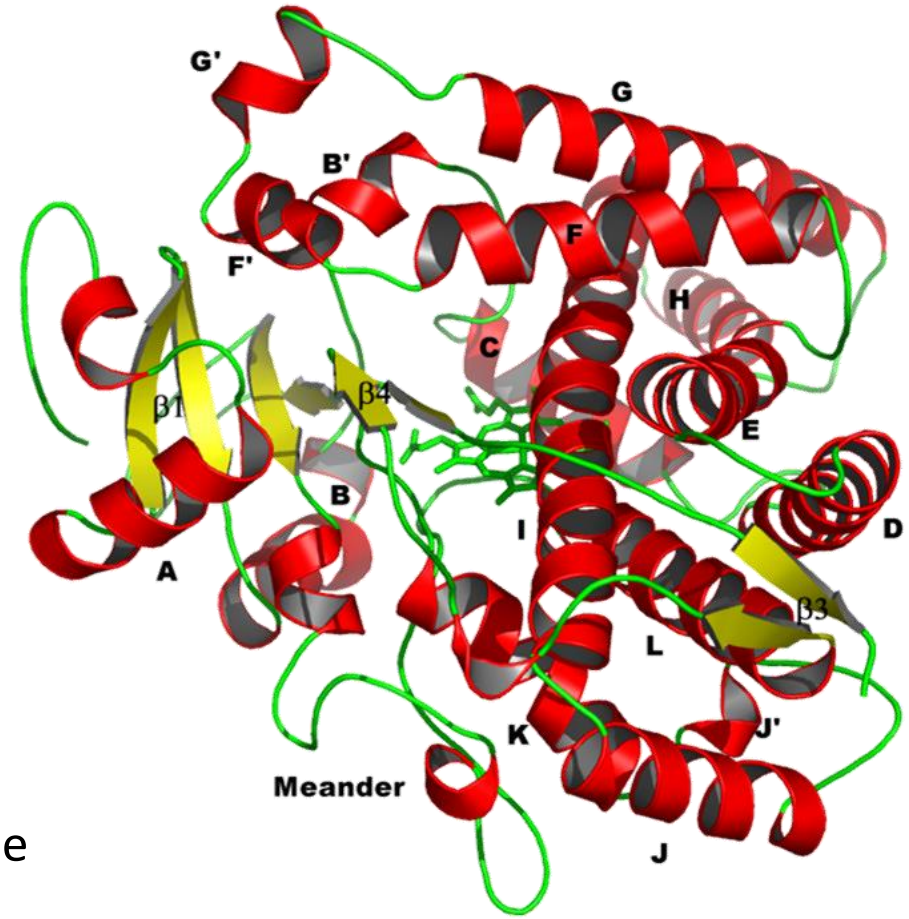
Most common reaction is a monooxygenase reaction **on heme iron**

$RH + O_2 + 2H^+ + 2e^- \rightarrow ROH + H_2O$

substrates become more polar

Features

- Highly promiscous (metabolize multiple substrates)
- Highly regio and stereospecific



CYP2C9 (1OG2)

Williams et al., Nature 424, 464, 2003

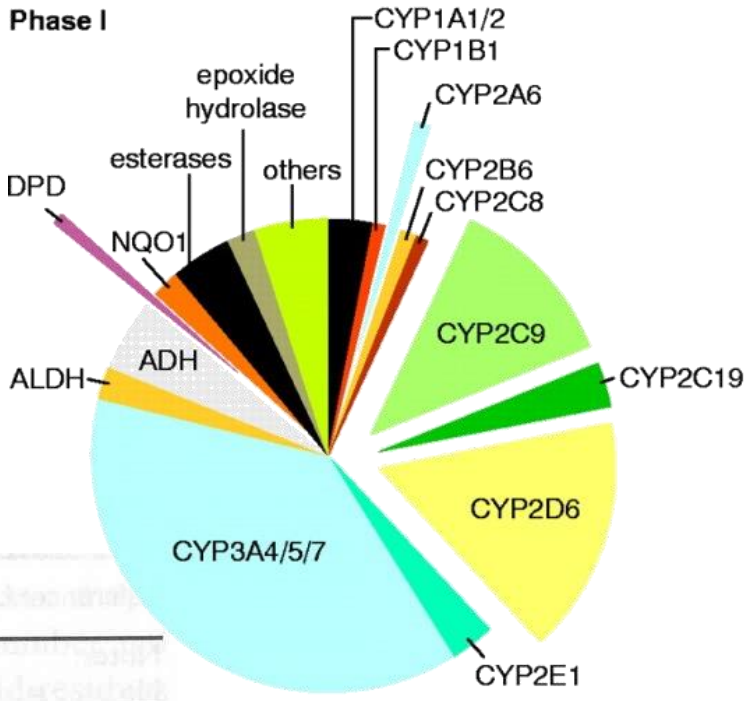
CYP Substrate Specificity

CYP family

– 75% of phase I metabolism

– One of the main reasons for failures of clinical testing => Predictions of possible metabolism needed ASAP

- Drug-Drug Interactions (grapefruit)



Evans, Relling
Science 286, 487,1999

Table 4.18 Characteristics of human P450 substrates

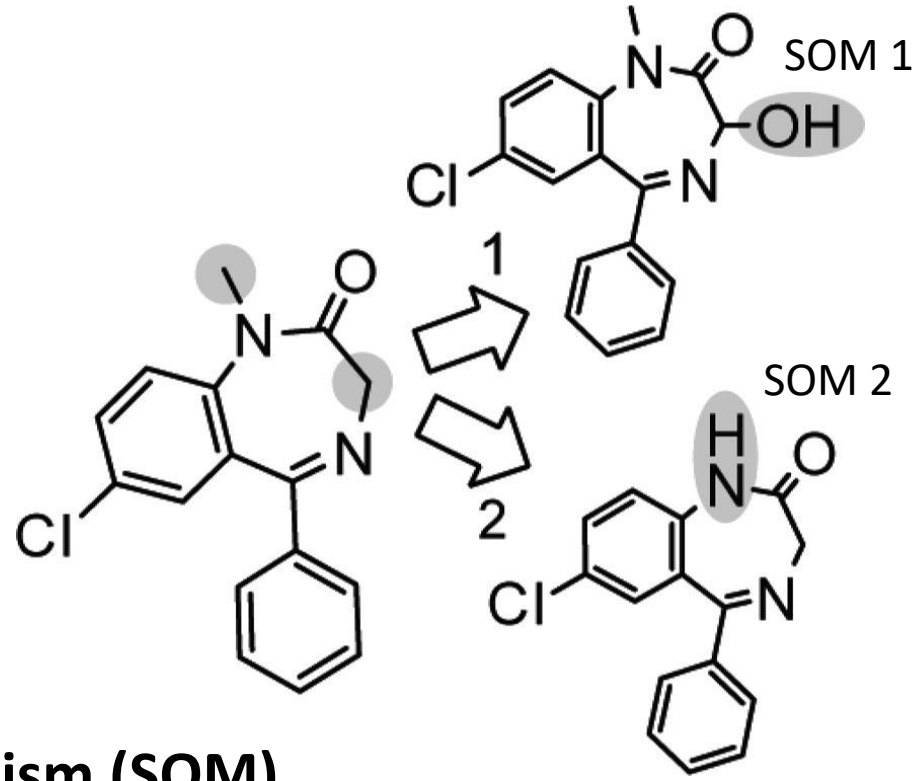
CYP	*General structural and physicochemical characteristics*
1A2	Planar (poly)aromatic/heterocyclic amines and amides with 2 or 3 hydrogen bond acceptors
2A6	Compounds usually contain ketonic or nitroso groups, generally polar
2B6	Non-planar (often V-shaped) molecules, usually lipophilic with hydrogen bond donor/acceptors
2C9	Generally weakly acidic compounds with hydrogen bond donor/acceptors
2C19	Generally neutral or basic compounds with hydrogen bond donor/acceptors
2D6	Nitrogenous bases with sites of metabolism 4–7 Å from basic nitrogen
2E1	Structurally diverse generally neutral compounds of low molecular weight
3A4	Structurally diverse compounds of relatively high molecular weight

Reference: Lewis et al., 1999b.

What Governs CYP Specificity?

Hard question:

- Active site
  - Size or shape => Docking
  - Flexibility
- Reactivity => SMARTCyp
- Accessibility
  - Channels
  - Membrane attachment
- Interactions with partners



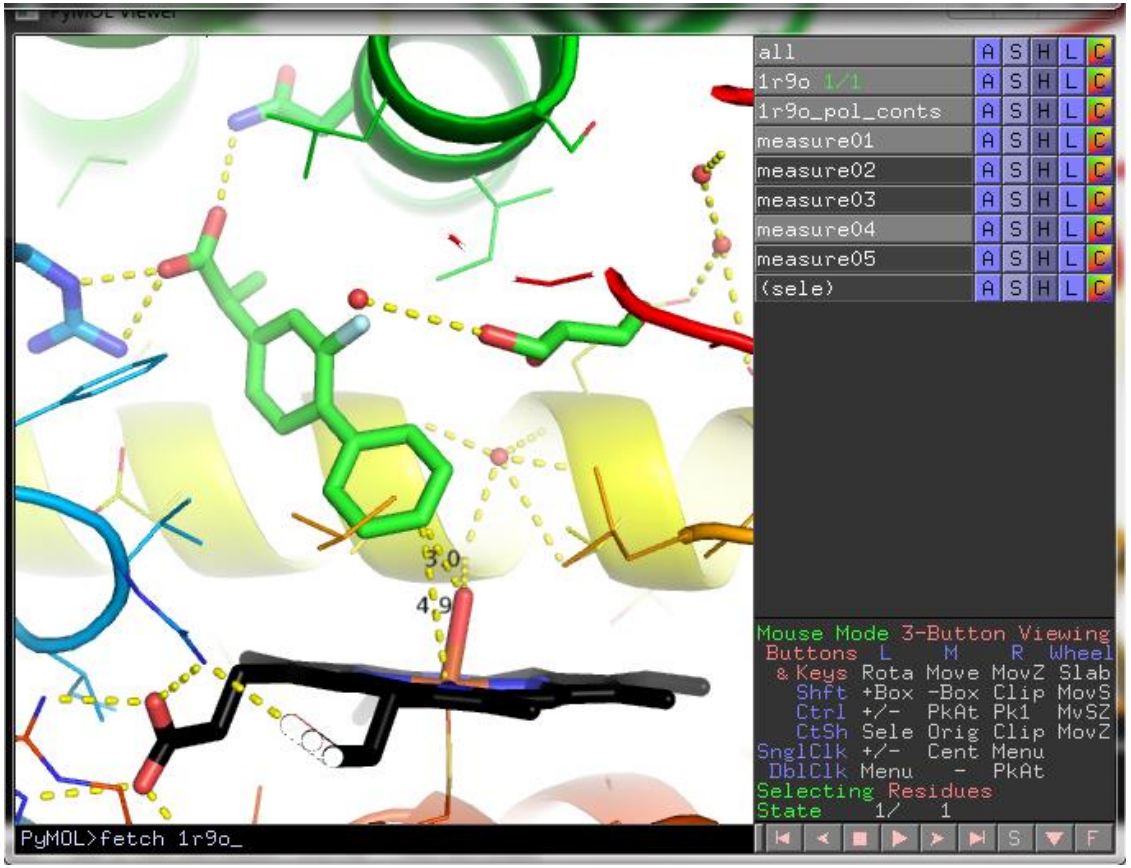
⇒ Identification of **Site-of-Metabolism (SOM)**

(weak point of drug - crucial in drug development)

CYP active site

- Contain **heme b** bound to Cys from one side and (sometimes) ligand from the other
- **Iron** atom in heme is catalytically active

⇒ **SOM** is the atom closest to heme iron



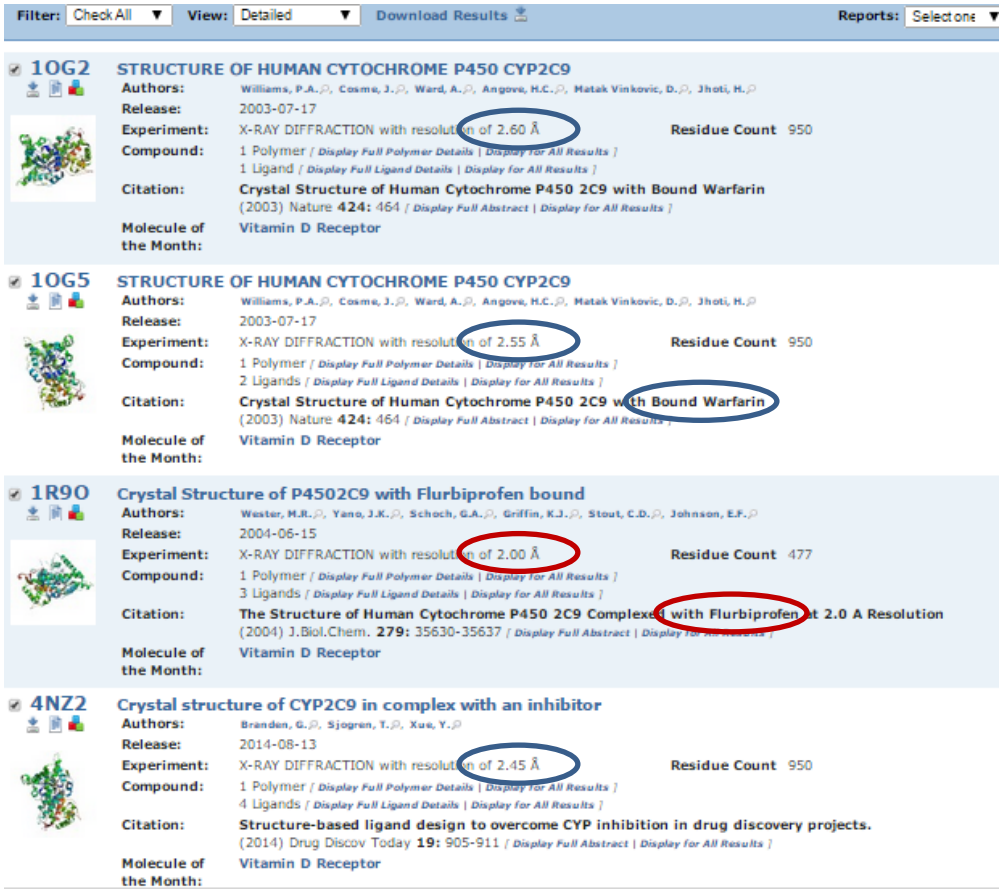
Flurbiprofen in active site of CYP2C9 – PDBID: **1r9o**

Pymol vizualization by all A>preset>ligand sites>cartoon

DOCKING

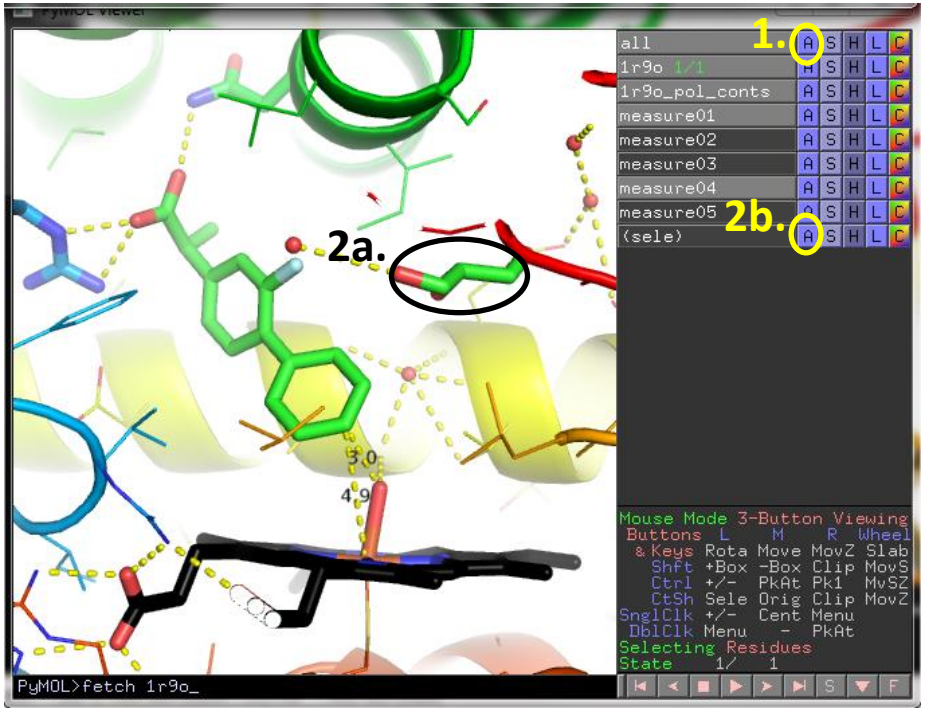
Protein Selection

- Find human cytochrome P450 2C9 at PDB database: www.rcsb.org
- Select those structures with some similar molecule in active site and with good resolution (<2.5Å)
- Open selected structure in Pymol



Protein and Ligand Preparation in Pymol

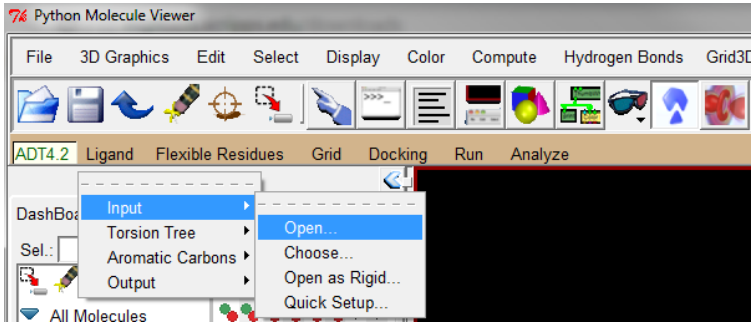
1. Pymol vizualization at line
   all>A>preset>ligand sites>cartoon
2. Delete glycol (crystalization agent)
   Select by mouse>(sele)>A>remove atoms
3. Delete waters
   A>remove waters
4. Extract ligand
   select by mouse>(sele)>A>extract object
5. Add hydrogens on ligand
   obj01>A>hydrogens>add
6. Save ligand
   File>Save Molecule>obj01>obj01.pdb
7. Add hydrogens on protein
   1r9o>A>hydrogens>add
8. Save protein
   File>Save Molecule>1r9o>1r9oH.pdb

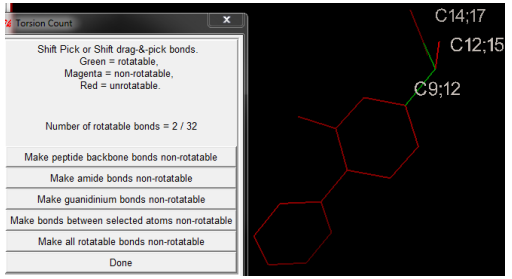


Flurbiprofen in active site of CYP2C9 –
PDBID: **1r9o**
Heme was colored black

Ligand Preparation – Autodock Tools

1. Start ADT or PMV program
   Line starting with ADT appears
2. Select ligand
   Ligand>Input>Open> *.pdb
   ligand will be prepared for docking
3. Check of aromatic atoms
   Ligand>Aromatic carbons>Set Names>
   (Shift+left mouse over cycle)
   aromatic atoms are shown in green
4. Check of rotatable torsions
   Ligand>Torsion Tree> Choose Torsions…
   rotatable bonds are shown in green, unrotatable in red, shift picking can select bond to be non-rotatable and vice versa.
   Flurbiprofen have 2 rotatable bond, bond between aromatics
5. Save pdbqt file
   Ligand>Output>Save as> *.pdbqt





Ligand PDBQT file

```
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: C9_10  and  C12_14
REMARK    2  A    between atoms: C12_14  and  C14_16
ROOT
HETATM    1  C   FLP A 501       7.165  28.800  -3.419  1.00 45.59     0.001 A
HETATM    2  C1  FLP A 501       7.996  29.908  -2.978  1.00 46.26     0.012 A
HETATM    3  C2  FLP A 501       7.405  31.131  -2.409  1.00 46.93    -0.023 A
HETATM    4  C3  FLP A 501       5.918  31.186  -2.355  1.00 46.14     0.012 A
HETATM    5  C4  FLP A 501       5.110  30.094  -2.794  1.00 43.88     0.001 A
HETATM    6  C5  FLP A 501       5.733  28.920  -3.347  1.00 44.98     0.000 A
HETATM    7  C6  FLP A 501       8.208  32.219  -1.923  1.00 49.58     0.013 A
HETATM    8  C7  FLP A 501       9.477  31.935  -1.184  1.00 49.56     0.016 A
HETATM    9  C8  FLP A 501      10.293  32.997  -0.670  1.00 50.29     0.009 A
HETATM   10  C9  FLP A 501       9.984  34.388  -0.968  1.00 50.23    -0.041 A
HETATM   11  C10 FLP A 501       8.772  34.689  -1.677  1.00 51.61     0.051 A
HETATM   12  C11 FLP A 501       7.940  33.653  -2.226  1.00 51.99     0.128 A
HETATM   13  F   FLP A 501       6.813  34.090  -2.912  1.00 55.17    -0.205 F
ENDROOT
BRANCH  10  14
HETATM   14  C12 FLP A 501      10.854  35.588  -0.436  1.00 50.59     0.115 C
HETATM   15  C13 FLP A 501      10.210  36.301   0.807  1.00 48.83     0.026 C
BRANCH  14  16
HETATM   16  C14 FLP A 501      11.120  36.626  -1.611  1.00 50.40     0.181 C
HETATM   17  O   FLP A 501      10.588  37.740  -1.596  1.00 51.19    -0.647 OA
HETATM   18  O1  FLP A 501      11.869  36.290  -2.555  1.00 50.57    -0.647 OA
ENDBRANCH  14  16
ENDBRANCH  10  14
TORSDOF 2
```

atom description x [A] y [A] z [A] occ.+Bfactor+charge+ atomic type

Atom lines with charges

Torsion definition> branching out of rigid ROOT part

Number of active torsions

Atomic types:

A – aromatic carbon

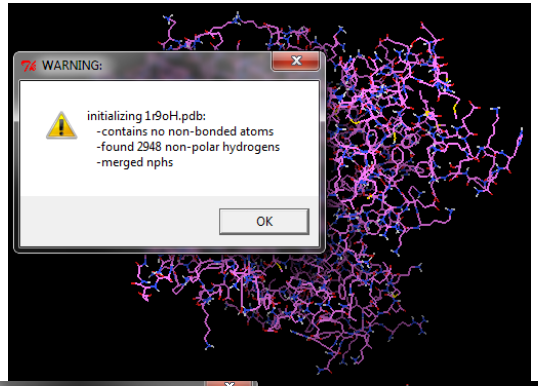
C – aliphatic carbon

F – fluorine

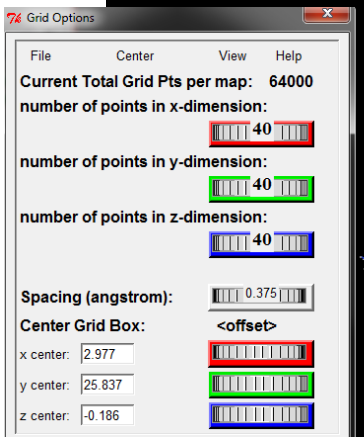
OA – hydrogen-bond acceptor oxygen

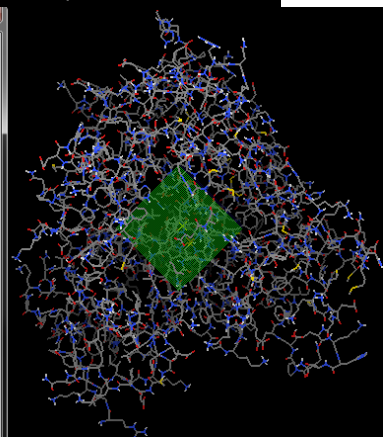
Protein Preparation – Autodock Tools

1. Select Protein Grid>Macromolecule>Open>
   *.pdb
   protein will be prepared for docking (nonpolar hydrogens merged with carbon, charges assigned) and saved as *.pdbqt (iron will have charge 0)



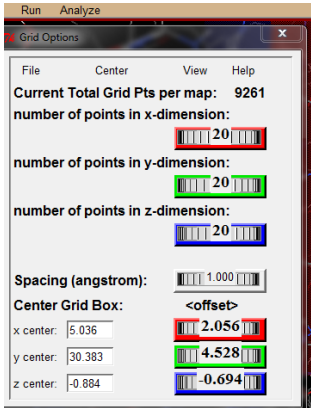
2. Analysis of necessary grid size
   Grid>Grid Box…
   cube with size will appear aromatic atoms are shown in red-green-blue sizes with yellow cross in its center
   - move yellow cross over the heme iron and then above the ring using <offset> settings
   - change spacing to 1 A (Vina use this settings)
   - adjust size of cube so it will cover active site, but not much more
   - note cube parameters

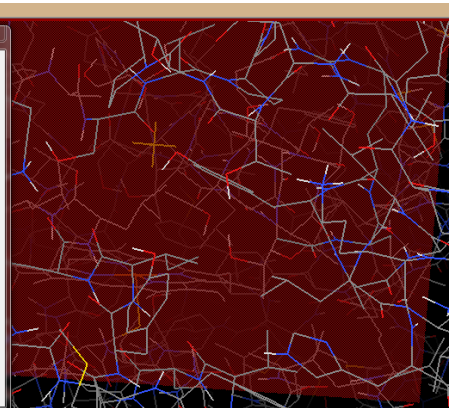




3. Edit protein PDBQT to charge iron(+2)

```
HETATM 4482 FE   HEM A 500       5.890  24.568  -1.058  1.00 18.80      2.000 Fe
```





Docking Setup – Vina

- Copy **`vina.exe`** into the same directory as both pdbqt (i.e. Desktop)
- Open **`cmd`** program
  (type cmd into the search field in Windows start panel)
- Go to directory with files
  see example on right
- Type **vina.exe** to see help
  (Tab key will fill up name of file)
- Either it is possible to fill in all parameters in one long line or within config file (next slide)

(cd Desktop) change directory
(cd ..) change directory up
(dir) list entries in directory

```
C:\windows\system32\cmd.exe
C:\Users\berka>cd Desktop

C:\Users\berka\Desktop>dir
 Svazek v jednotce C je TI312407008A.
 Sériové číslo svazku je 70CD-275A.

 Výpis adresáře C:\Users\berka\Desktop

28.11.2014  00:47    <DIR>          .
28.11.2014  00:47    <DIR>          ..
27.11.2014  23:27           616 284 1r9o.pdb
27.11.2014  23:28           616 284 1r9oH.pdb
28.11.2014  00:17           363 926 1r9oH.pdbqt
27.11.2014  23:28             5 015 obj01.pdb
28.11.2014  00:06             1 743 obj01.pdbqt
08.11.2014  11:14    <DIR>          Unused
11.05.2011  11:37           781 824 vina.exe
               Souborů:          6,   Bajtů:               2 385 076
               Adresářů:         3,   Volných bajtů: 22 923 960 320

C:\Users\berka\Desktop>vina.exe
Missing receptor.

Correct usage:

Input:
  --receptor arg        rigid part of the receptor (PDBQT)
  --flex arg            flexible side chains, if any (PDBQT)
  --ligand arg          ligand (PDBQT)

Search space (required):
  --center_x arg        X coordinate of the center
  --center_y arg        Y coordinate of the center
  --center_z arg        Z coordinate of the center
  --size_x arg          size in the X dimension (Angstroms)
  --size_y arg          size in the Y dimension (Angstroms)
  --size_z arg          size in the Z dimension (Angstroms)

Output (optional):
  --out arg             output models (PDBQT), the default is chosen based on
                        the ligand file name
  --log arg             optionally, write log file

Misc (optional):
  --cpu arg                 the number of CPUs to use (the default is to try to
                            detect the number of CPUs or, failing that, use 1)
  --seed arg                explicit random seed
  --exhaustiveness arg (=8) exhaustiveness of the global search (roughly
                            proportional to time): 1+
  --num_modes arg (=9)      maximum number of binding modes to generate
  --energy_range arg (=3)   maximum energy difference between the best binding
                            mode and the worst one displayed (kcal/mol)

Configuration file (optional):
  --config arg          the above options can be put here

Information (optional):
```

Docking parameters file - config.txt

Protein
Ligand

Output

Box center

Box size

Search exhaustiveness

Number of reported poses

```
receptor = 1r9oH.pdbqt
ligand   = obj01.pdbqt

out      = test_out.pdbqt
log      = test_out.log

center_x = 5
center_y = 30
center_z = -1
size_x   = 20
size_y   = 20
size_z   = 20

exhaustiveness = 8
num_modes      = 9
```

Docking Run - Vina

- Run

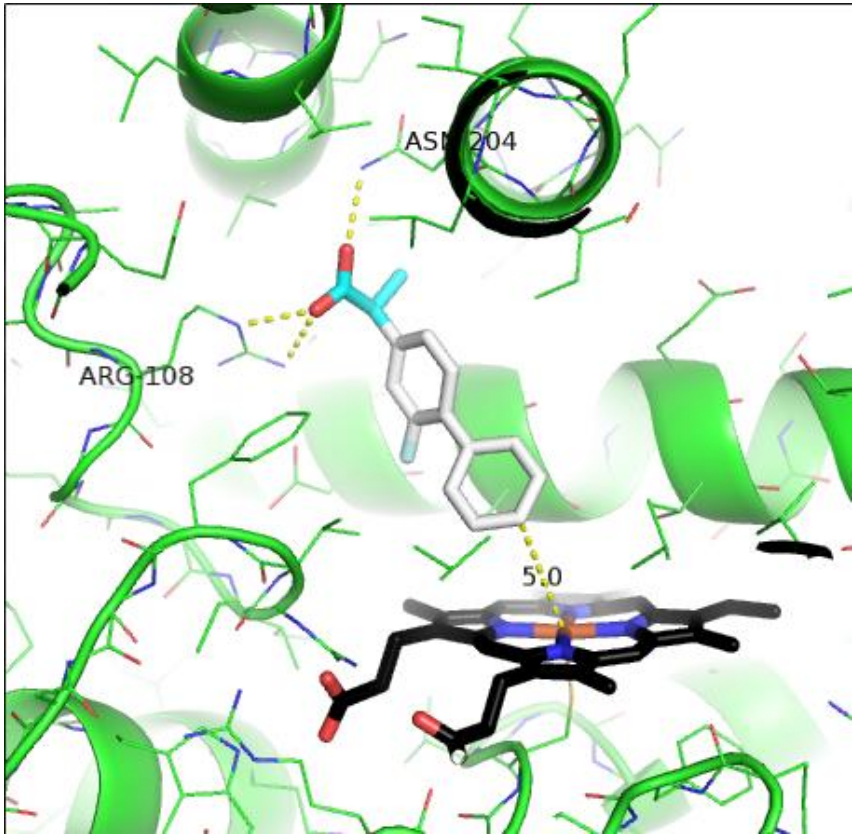
  `vina.exe --config config.txt`

- Vina detects available number of CPUs
- docking should take about 10 s
- progress is shown with stars
- Results is shown with predicted affinity in kcal/mol, and RMSD to the first predicted pose (same info is in log file)

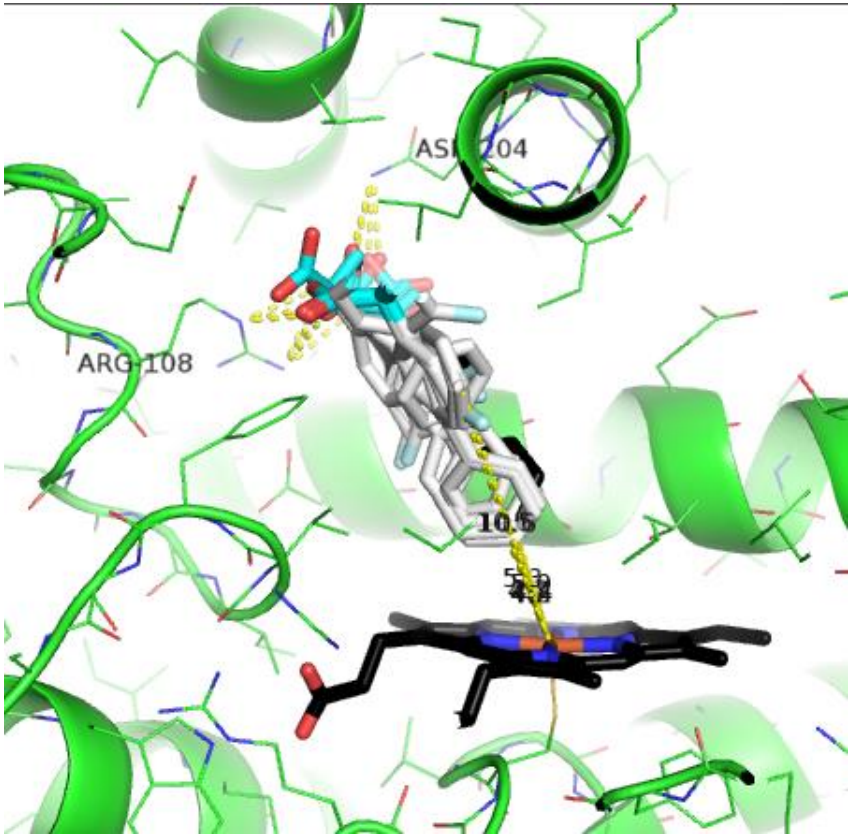
```
C:\Users\berka\Desktop>vina.exe --config config.txt
#################################################################
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
#################################################################

Detected 4 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 863478692
Performing search ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -9.4      0.000      0.000
   2         -9.3      1.214      2.065
   3         -8.5      3.379      6.863
   4         -8.3      3.419      6.868
   5         -8.2      3.305      6.892
   6         -8.2      3.343      6.941
   7         -7.8      3.930      4.646
   8         -7.5      2.660      3.349
   9         -7.1      4.351      5.263
Writing output ... done.
```

Docking analysis – Pymol

First pose – almost same as xray



All poses overlayed over xray



Redocking was succesful

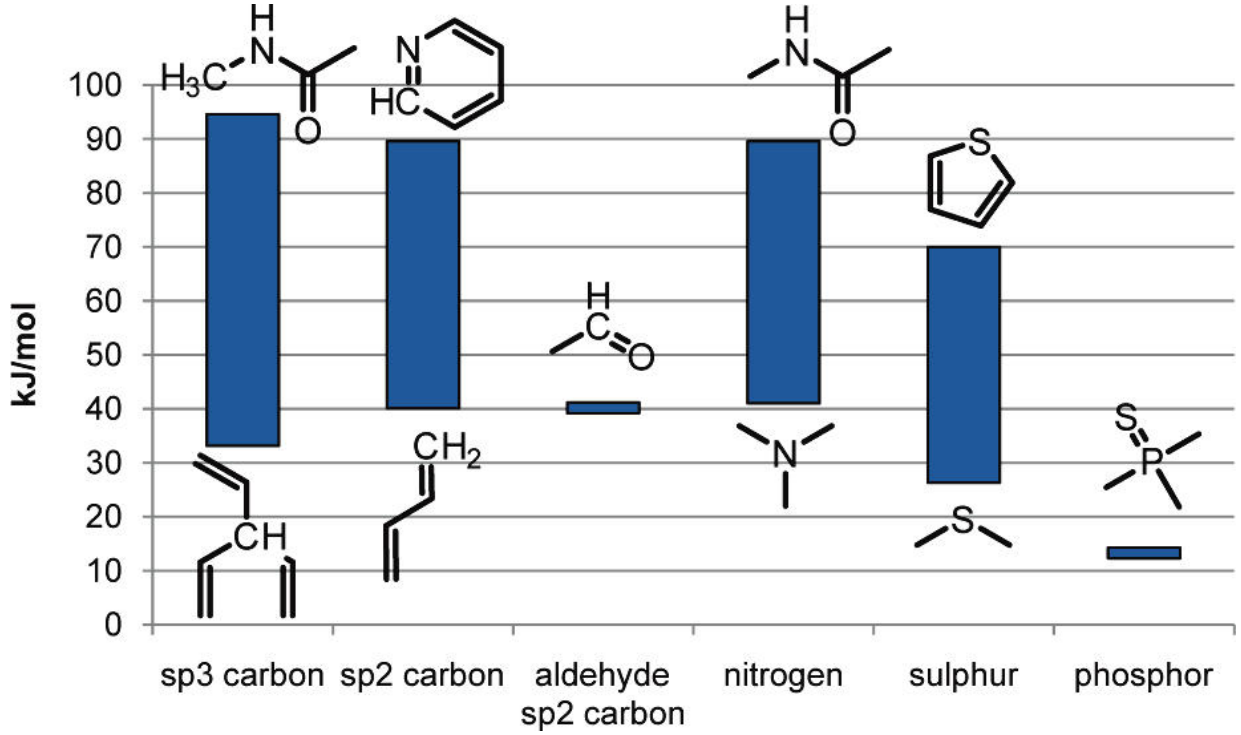
Docking – Further Exercises

- Play a little with docking parameters
  - Starting position of ligand – shift it in Pymol outside of protein
  - Exhaustiveness – as small as 1 and as large as 20
  - Charge on iron (0, 2, 3 or left there oxygen atom on iron (reactive intermediate compound I has configuration Fe=O)
  - Add free rotation between both phenyl rings
- Prepare and dock ibuprofen – will it bind similarly as flurbiprofen?
  - Build modul in Pymol
  - download sdf with major microspecies from PubChem,
  - download structure from www.rcsb.org

SMARTCYP

SMARTCyp

- Predictions of SOM (sites-of-metabolism) based on DFT calculations and cheminformatics
- Activation energies for individual CYP reactions on individual sites DFT (B3LYP/6-311++G(2d,2p))
- Comparison of HAT energies in individual fragments:



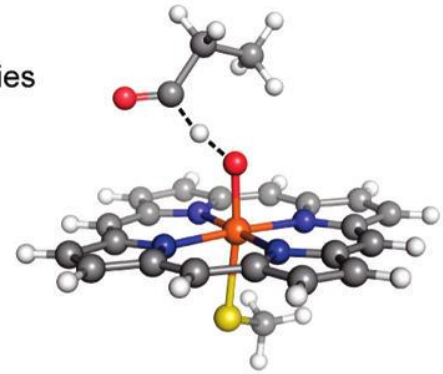
Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L (2010) SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. ACS Med Chem Lett 1: 96–100

SMARTCyp – principle II

Atom Reactivity Library

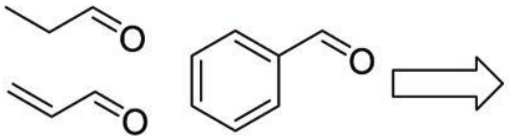
A. Calculate Quantum Chemical Reference Energies

Calculate transition state energies using density functional theory



B. Define SMARTS Rules

Group calculations by fragments and calculate average energies

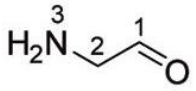


[CX3H1](=O)[#6]

40.2 kJ/mol

SMARTCyp

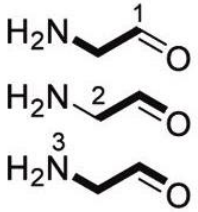
1. Assign Energies By SMARTS matching



Atom	SMARTS	Energy
1	[CX3H1](=O)[#6]	40.2
2	[CX4][N]	39.8
3	[N^3][H1,H2]	54.1

2. Compute Accessibility Descriptor

$A_i = Maxbonds_i / Maxbonds_{all}$



$A_1 = 2 / 3 = 0.67$

$A_2 = 2 / 3 = 0.67$

$A_3 = 3 / 3 = 1.00$

3. Compute Score and Rank Atoms

Score, S = E - 8A
Lowest score gets rank 1

$S_1 = 40.2 - 8*0.67 = 34.84$

$S_2 = 39.8 - 8*0.67 = 34.44$

$S_3 = 54.1 - 8*1.00 = 46.10$

Atom 1 - Rank 2

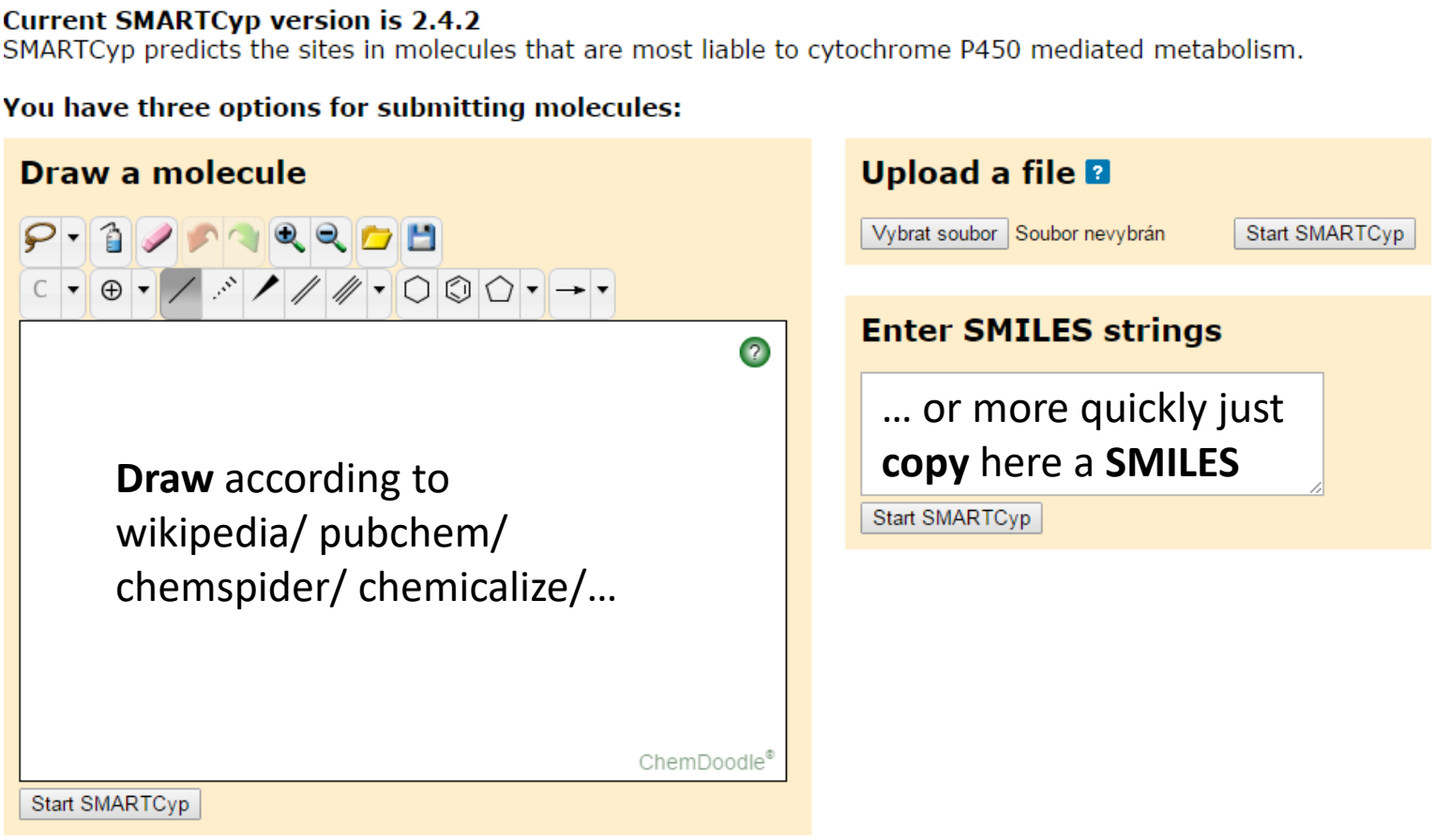
Atom 2 - Rank 1

Atom 3 - Rank 3

Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L (2010) SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. ACS Med Chem Lett 1: 96–100

SMARTCyp - input

- https://smartcyp.sund.ku.dk/



P. Rydberg et al ACS Med. Chem. Lett., 2010, 1, 96-100 , P. Rydberg et al Bioinformatics, 2010, 26, 2988-2989, P. Rydberg and L. Olsen, ACS Med. Ch Lett., 2012, 3, 69-73, P. Rydberg and L. Olsen, ChemMedChem, 2012, 7, 1202-1209 , P. Rydberg et al., Angew. Chem, Int. Ed. 2013, 52, 993-997, and Rydberg et al., Mol. Pharmaceutics 2013, 10, 1216-1223

SMARTCyp - results

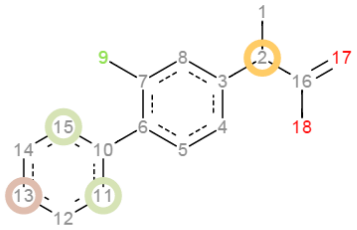
Flurbiprofen

CC(c1ccc(c(c1)F)c2ccccc2)C(=O)O

Results from SMARTCyp version 2.4.2

These results were produced: 2014-11-26_22-54-52. The infiles were: [smiles.smi].

To alternate between heteroatoms and atom numbers, move the mouse cursor over the figure.



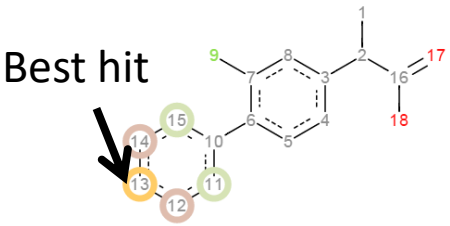
Generic –
good for CYP3A4

Standard | CYP2C | CYP2D6

1: null

Rank	Atom	Score	Energy	Accessibility	2DSASA
1	C.2	59.69	66.4	0.8	7.86
2	C.13	71.35	80.8	1	36.24
3	C.11	73.28	80.8	0.8	27.97
4	C.5	75.87	80.8	0.5	23.31
5	C.12	77.76	86.3	0.9	33.5
6	C.4	78.38	84.1	0.6	23.07
7	C.8	78.44	84.1	0.6	21.5
8	C.1	80.27	89.6	0.9	53.32
9	C.16	991.53	999	0.9	6.8
10	C.10	993.24	999	0.7	3.88
11	C.3	993.28	999	0.7	3.01
12	C.6	994.08	999	0.6	3.06
13	C.7	994.76	999	0.5	6.02

Best hit



With pharmacophore
for CYP2C9

Standard | CYP2C | CYP2D6

1: null

Rank	Atom	Score	Energy	S2End	COODist	2DSASA
1	C.13	79.35	80.8	0	10	36.24
2	C.12	90.86	86.3	1	9	33.5
3	C.11	91.48	80.8	2	8	27.97
4	C.2	113.29	66.4	2	2	7.86
5	C.5	121.22	80.8	5	5	23.31
6	C.1	122.87	89.6	1	3	53.32
7	C.4	130.42	84.1	4	4	23.07
8	C.8	130.48	84.1	4	4	21.5
9	C.10	1,022.44	999	3	7	3.88
10	C.6	1,034.32	999	4	6	3.06
11	C.7	1,040.11	999	5	5	6.02
12	C.16	1,045.93	999	1	1	6.8
13	C.3	1,046.08	999	3	3	3.01

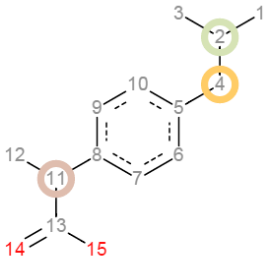
Ibuprofen

CC(C)Cc1ccc(cc1)C(C)C(=O)O

Results from SMARTCyp version 2.4.2

These results were produced: 2014-11-26_23-06-32. The infiles were: [smiles.smi].

To alternate between heteroatoms and atom numbers, move the mouse cursor over the figure.

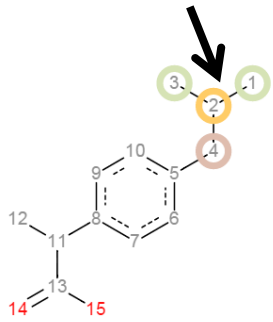


Standard | CYP2C | CYP2D6

1: null

Rank	Atom	Score	Energy	Accessibility	2DSASA
1	C.4	59.21	66.4	0.78	24.31
2	C.11	59.85	66.4	0.78	8.16
3	C.2	68.42	75.9	0.89	9.15
4	C.1	79.21	89.6	1	59.79
5	C.12	80.36	89.6	0.89	53.32
6	C.6	80.82	86.3	0.56	25.81
7	C.7	80.87	86.3	0.56	24.72
8	C.13	991.62	999	0.89	6.8
9	C.5	993.49	999	0.67	4.3
10	C.8	993.53	999	0.67	3.33

Best hit



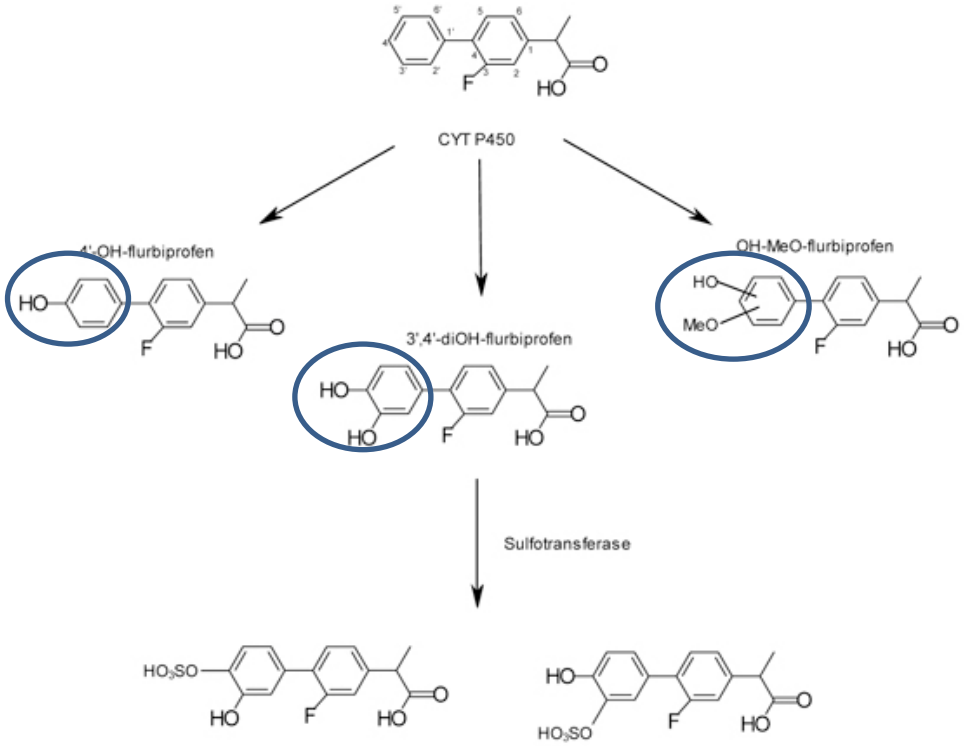
Standard | CYP2C | CYP2D6

1: null

Rank	Atom	Score	Energy	S2End	COODist	2DSASA
1	C.2	81.43	75.9	1	8	9.15
2	C.4	83.13	66.4	2	7	24.31
3	C.1	87.21	89.6	0	9	59.79
4	C.11	113.27	66.4	2	2	8.16
5	C.12	122.87	89.6	1	3	53.32
6	C.6	126.61	86.3	4	5	25.81
7	C.7	132.55	86.3	4	4	24.72
8	C.5	1,028.33	999	3	6	4.3
9	C.13	1,045.93	999	1	1	6.8
10	C.8	1,046.07	999	3	3	3.33

Reality check

Flurbiprofen



Ibuprofen

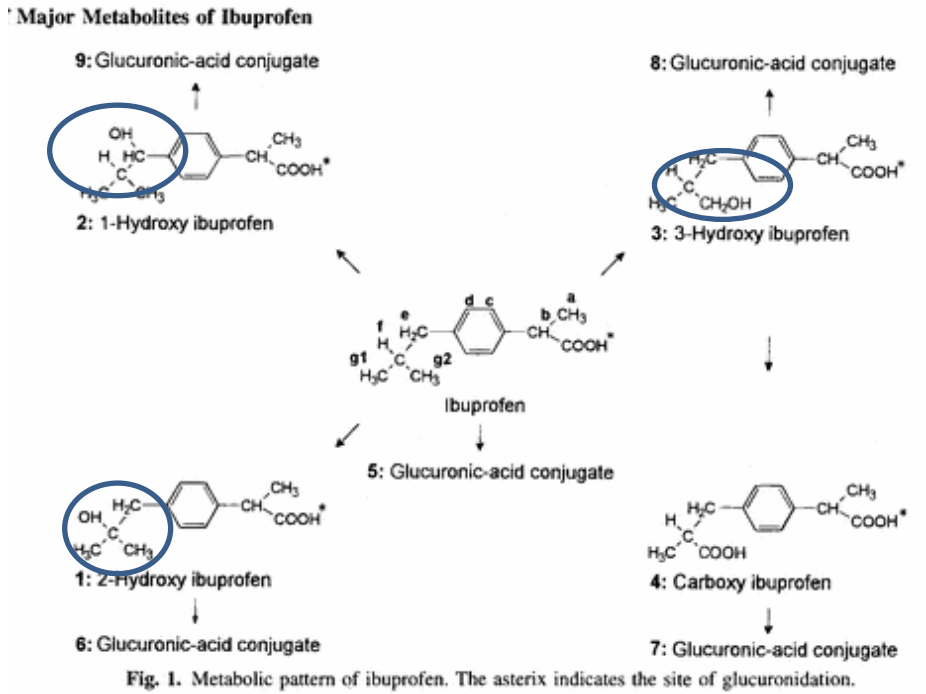


Fig. 1. Metabolic pattern of ibuprofen. The asterix indicates the site of glucuronidation.

Cormac D. Murphy - Fluorinated drug metabolism in microorganisms
Chimica Oggi-Chemistry Today (2012) 30(3)
http://www.teknoscienze.com/Articles/Chimica-Oggi-Chemistry-Today-Fluorinated-drug-metabolism-in-microorganisms.aspx

Dorte R. Kepp, et al: Isolation and Characterization of Major Phase I and II Metabolites of Ibuprofen. Pharm Res 1997, 14 (5), 676-680

Pymol installation on Windows

- Install miniconda or any other Python
  - Allows pip installation
- Download Whl files from Gohlke webpage
  - https://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol
- Installing PyMOL (and its Dependencies): open a Command Prompt:
  - pip install wheel
  - pip install pymol-2.3.0-cp37-cp37m-win_amd64.whl
  - pip install pymol_launcher-2.1-cp37-cp37m-win_amd64.whl
- Run Pymol
  - type pymol in the Command Prompt.
  - OR, browse your Python Scripts directory, right-click on the file pymol.exe and select the option Send to followed by Desktop (create shortcut).

https://bitsilla.com/wiki/doku.php?id=howto:pymol_install_on_windows