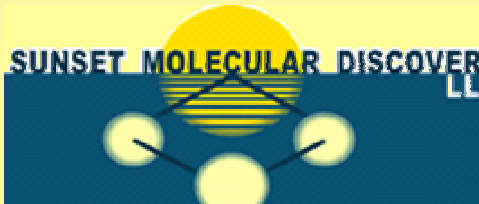
W O M B A T
WOrld of Molecular BioAcTivity

Tudor Oprea

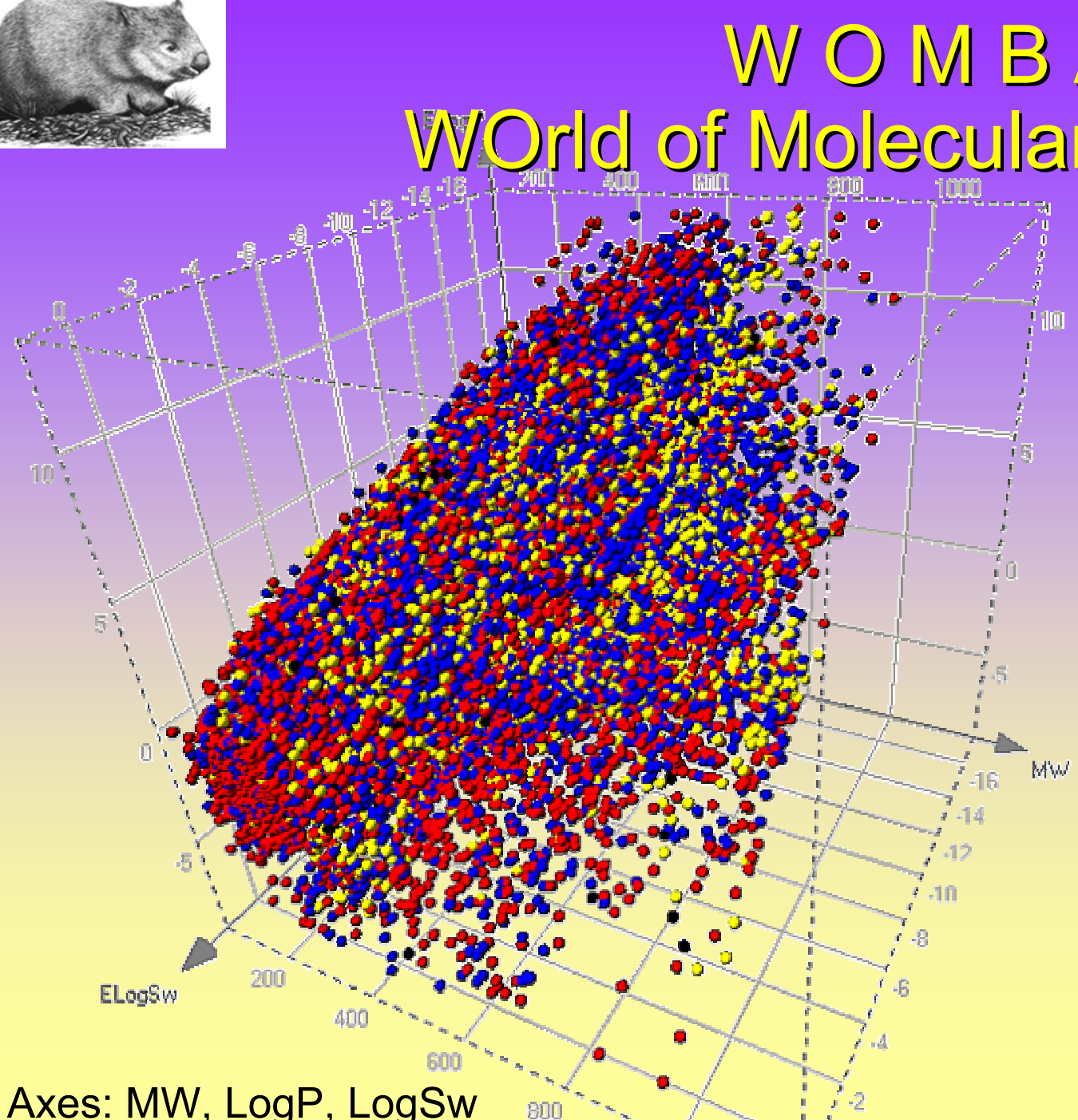
Sunset Molecular Discovery LLC

http://www.sunsetmolecular.com

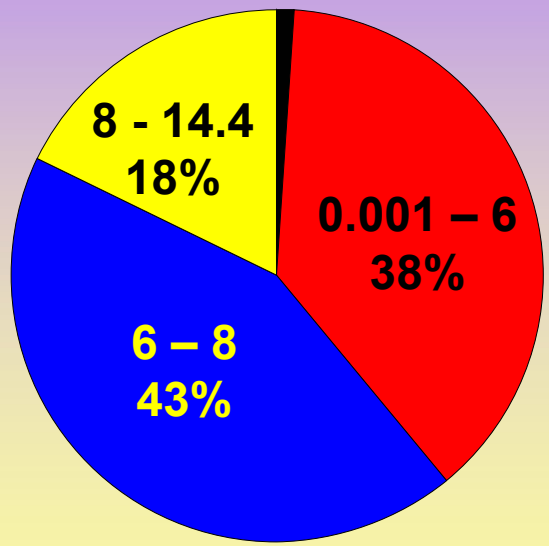
Daylight MUG 18
Santa Fe, NM, 02/24/04



W O M B A T
WOrld of Molecular BioAcTivity

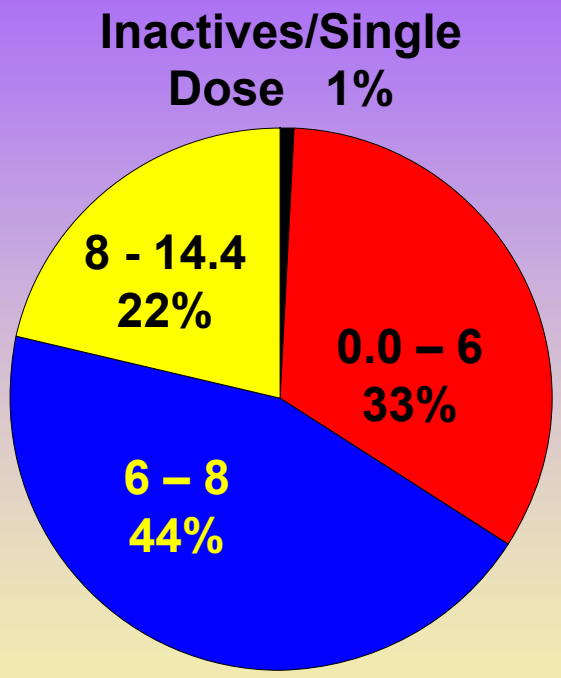


Axes: MW, LogP, LogSw

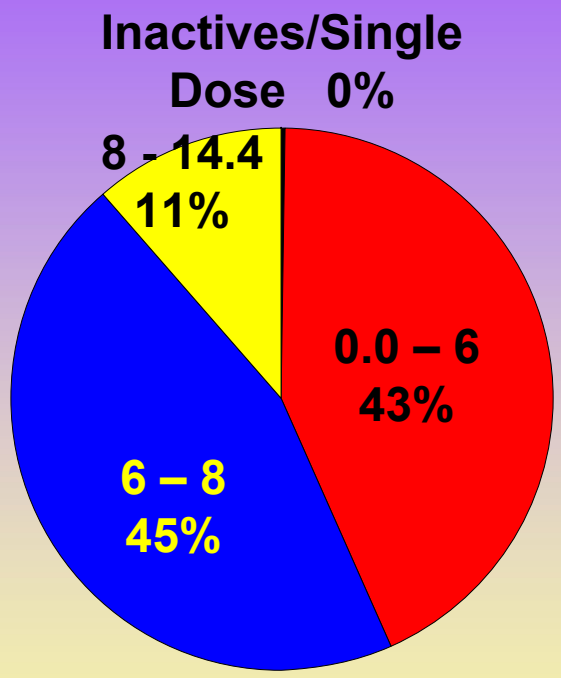


SUNSET MOLECULAR DISCOVERY LLC

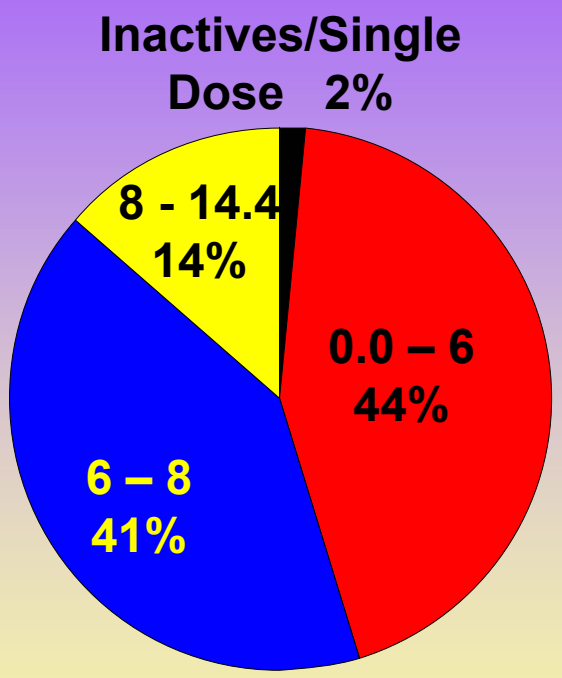
Bioactivity Distribution By Target Type



Receptors (56.2%)



Proteins (4.8%)



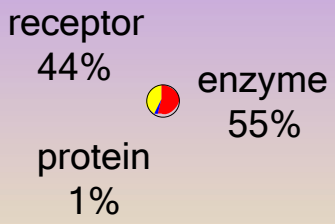
Enzymes (39%)

- Enzymes tend to have a higher rate of inactives/low actives
- Receptors tend to have more medium/high actives

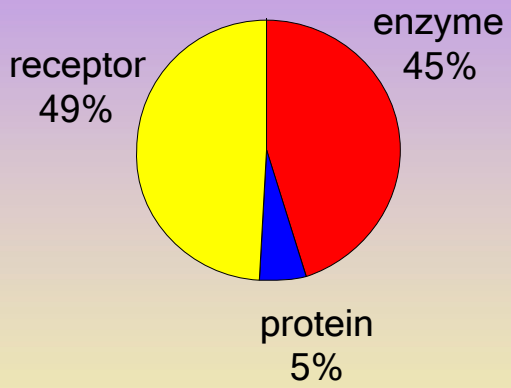
SUNSET MOLECULAR DISCOVER LL

Target Type Distribution By Activity

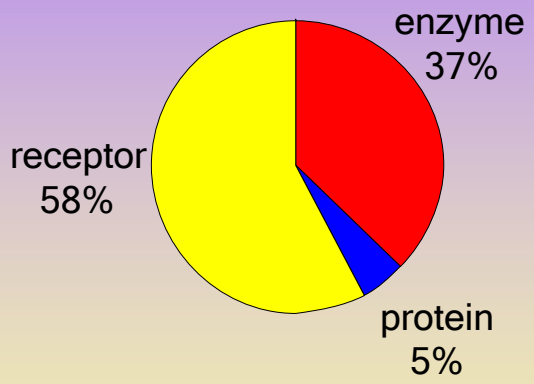
Inactives (1%)



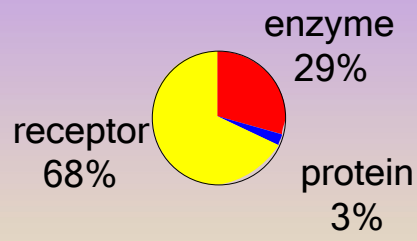
Low Act. (38%)



Medium Act. (43%)



High Act. (18%)



- Enzymes dominate the inactive/low activity bins
- Receptors clearly dominate the medium/high activity bins

WOMBAT History

- SADB5 (May 2002):
  - Project funded initially by AstraZeneca
  - 21700 structures (includes duplicates)
  - 36738 activities on 324 targets
  - 837 papers indexed (JMC 1996-1999)
  - 39.56% K_i, 53.52% IC_{50}
  - 5.54% D_2 or EC_{50}
- WOMBAT 2003.2 (September '03):
  - 53126 entries (47872 unique structures) 98662 activities on 506 unique targets, plus 236 inactives, 7982 'smaller than' & 159 'greater than' values
  - 2143 papers (2148 series) indexed (JMC 1994-1999)
  - 35.5% K_i, 56.6% IC_{50}, 4.85% D_2 or EC_{50}
  - literature coverage included BMCL (2002), QSAR (2000-2001)

Romanian Academy
Institute of Chemistry Timisoara

AstraZeneca

SUNSET MOLECULAR DISCOVER

WOMBAT 2004.1

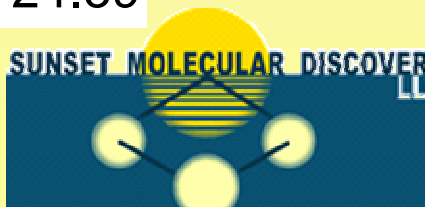
- 76,165 entries (68,543 unique SMILES) covering 3039 series (over 3000 papers) and ~143,000 activities on ~630 targets
- Activities now include inactives (635), < (8916), > (259), @ (578 – single dose)
- 37.1% K_i (& variations), 55.85% IC_{50} (& variations), 4.44% D_2 or EC_{50}, 0.9% K_b and K_d , 0.1% MIC, 0.04% ED_{50}
- *Biochem. Pharmacol.* 2001 [partial coverage], *Bioorg. Med. Chem. Lett.* 2002 [1-24], *Chembiochem* 2002 [partial], *Eur. J. Med. Chem.* 2001 [partial], *J. Amer. Chem. Soc.* 1975,1992,1993 [partial], *J. Healt. Sci.* 2003 [partial], *J. Med. Chem.* 1991 [partial], 1992-2000 [complete], 2003 [partial], *Quant. Struct.-Act. Relat.* 1998-2000 [partial]
- Fully integrated FEDORA server (Metaphorics LLC)
- New features include SwissProt IDs for most Targets and links (via the DOI format for 1737 entries) to PDF files for all literature entries

http://www.sunsetmolecular.com/products/?id=4

Activity Profile for WOMBAT 2004.1

Target Class	Compounds [(a)]	Percent
G-Protein Coupled Receptors	28973	38.04
Nuclear Hormone Receptors	688	0.90
Integrins	1772	2.33
Ion Channels	9008	11.83
Aspartyl Proteases	3351	4.40
Serine Proteases	1459	1.92
Kinases	2842	3.73
Cysteine Proteases	704	0.92
Phosphodiesterases	1689	2.22
Oxidoreductases	2010	2.64
Oxygenases	2829	3.71
Transporters	2264	2.97
Others	18576	24.39

(a) number of structures active at least once/target, % of total entries



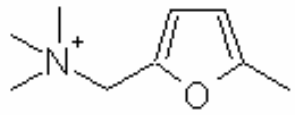
< Main > Keywords

Sunset Molecular Discovery, LLC
www.sunsetmolecular.com

WOMBAT

SMDL-00005985 | 249

structure



Cc1ccc(C[N+](C)(C)C)o1

SMDL_ID	molname
5985	MFT
generic_name	
reference: J. Med. Chem. 35(1)-1992 15-27	
formula: $C_9H_{16}NO$	mol_weight: 154.2340
est_LogKow: -2.5600	exp_LogKow
est_LogWSol: 0.8120	exp_LogWSol

structure_keywords
nc

act#	targ_type	target_name	act_type	act_val	act_min	act_max	inhib%
5	receptor	m1	IC50	8.2518	8.2518	8.2518	

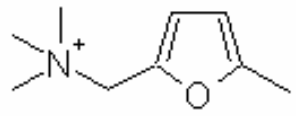
Main | < Keywords >

Sunset Molecular Discovery, LLC
www.sunsetmolecular.com

WOMBAT

SMDL-00005985 | 249

structure



activity # and value	5	8.2518

Biology

species	rat
tissue_type	membrane
tissue_source	cortex; brain
cell	
intracell_struct	
stimulus	
effect	antagonist

Enzyme

family	
name	
type	
EC	
substrate	
reactant	
cofactor	
radiosub	
actdet	

SwissProt

ID	P08482
species	rat

Receptor

family	GPCR; ion channel
name	muscarinic acetylcholine receptor
type	

Protein

family	
name	
type	

Binding

endogen_lig	
radio_lig	[3H]CD
nonspecific	

Other

References Are Stored Separately

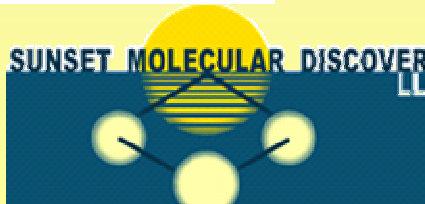
Sunset Molecular Discovery, LLC
www.sunsetmolecular.com

WOMBAT

References Database

1

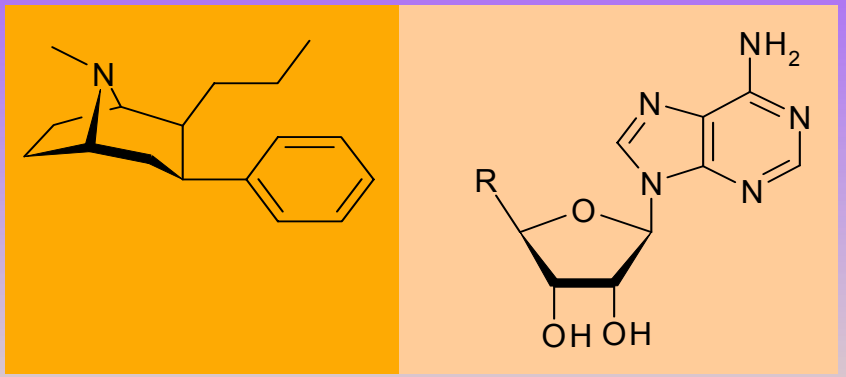
author_list	title
Yu,Y.; Deck,J.A.; Hunsaker,L.A.; Deck,L.M.; Royer,R.E.; Goldberg,E.; Vander Jagt,D.L.	Selective active site inhibitors of human lactate dehydrogenases A4, B4, and C4
DOI 10.1016/S0006-2952(01)00636-0	Web Link http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T4P-433NTS7-9&_coverDate=07%2F01%2F2001&_alid=143739897&_rdoc=1&_fmt=&_orig=search&_qd=1&_cdi=4980&_sort=d&view=c&_acct=C000053660&_version=1&_urlVersion=0&_userid=1550512&md5=766cd9c999c2080



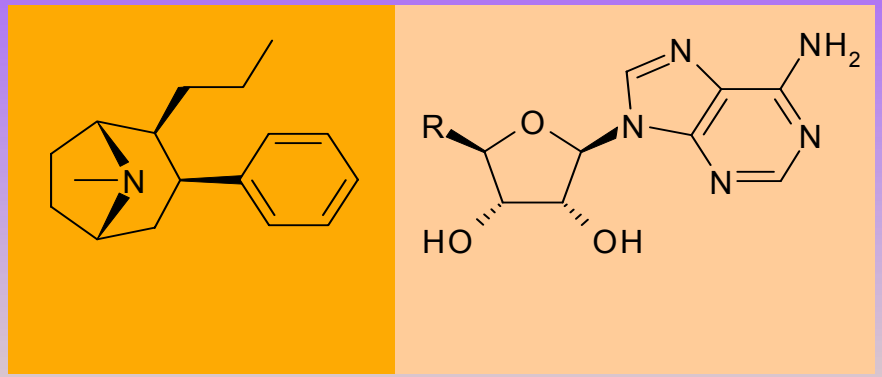
WOMBAT Quality Control

- Chirality: What chemists can interpret, computers are not always able (the "above/below the plane" must be strictly enforced)

Not machine-readable



Machine-readable



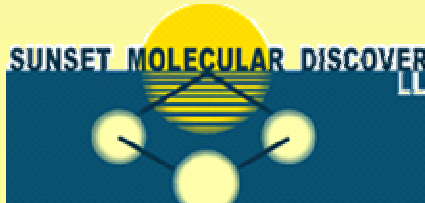
- Missing/altered atoms/substituents – overall error rate above 9%
  - Incorrectly drawn or written structures (3.4%); incorrect molecular formula or molecular weight (3.4%);
  - Unspecified binding position for substituents or ambiguous numbering scheme for the heterocyclic backbone (0.91%);
  - Structures with the incorrect backbone (0.71%);
  - Incorrect generic names or chemical names (0.24%);
  - Incorrect biological activity (0.34%);
  - Incorrect references (0.2%).

SUNSET MOLECULAR DISCOVER LL

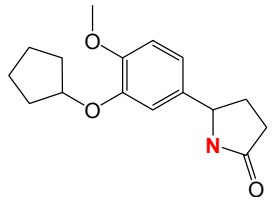
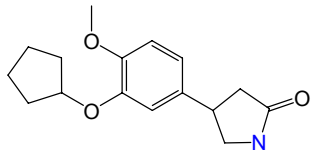
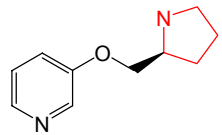
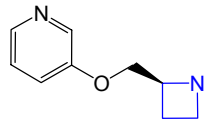
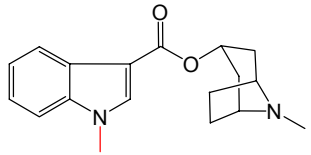
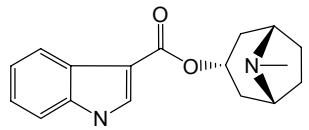
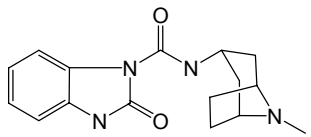
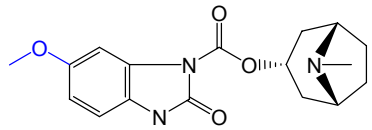
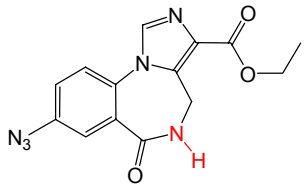
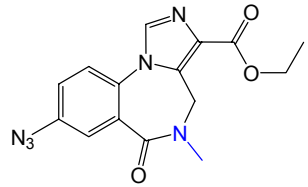
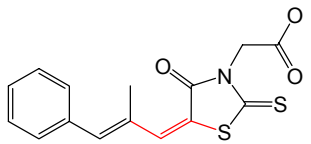
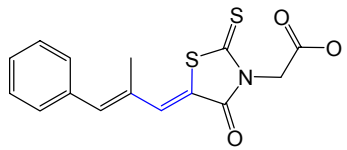
WOMBAT Quality Control…

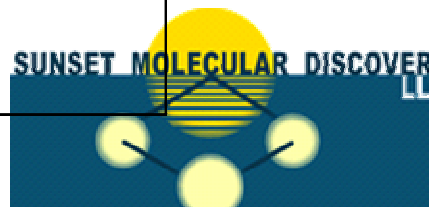
The number of known errors is finite.

The number of unkown errors is infinite.

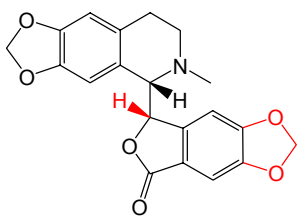
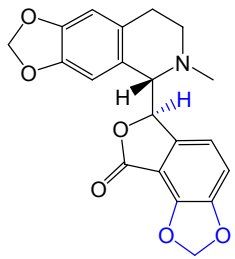
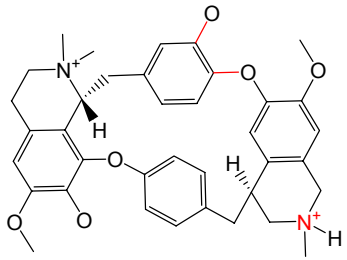
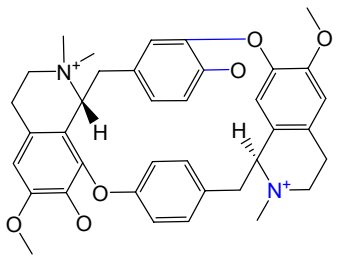
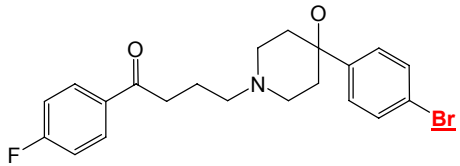
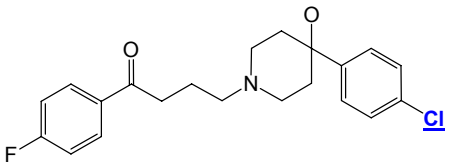
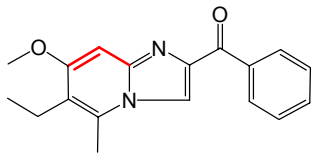
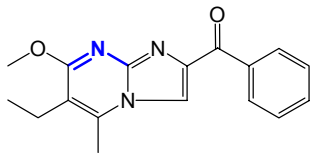
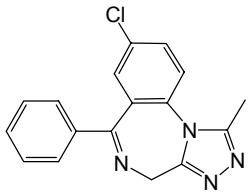
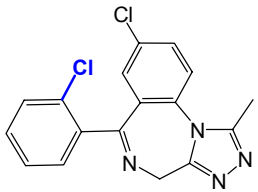


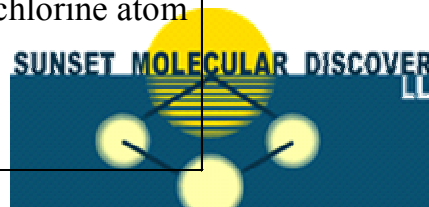
JMC Errors… 1

Reference	Published Structure	Corrected Structure	Comment
JMC 37-476 chart 1			*rolipram*: incorrect N atom position
JMC 43-2217 chart 1			*A-85380*: incorrect ring size
-||- & JMC 36-2645			*tropisetron*: methyl group in plus
-||-			*DAU-6285*: missing methoxy; N instead O
JMC 37-758 chart 1			*Ro-15-4513*: methyl group missing
JMC 37-787 figure 1			*epalrestat*: E/Z config: E instead Z

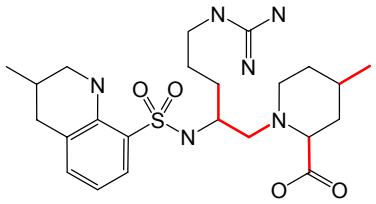
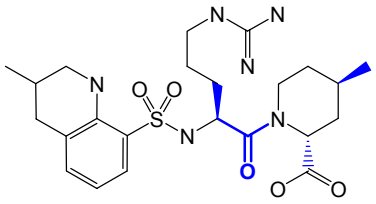
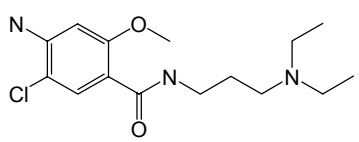
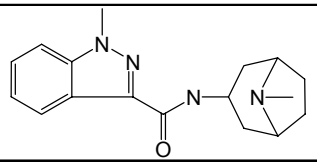
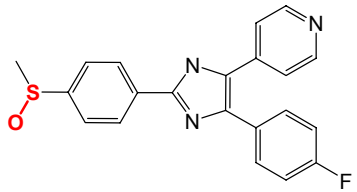
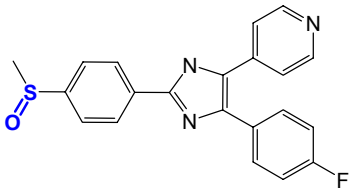
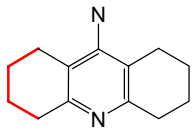
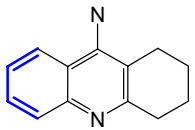
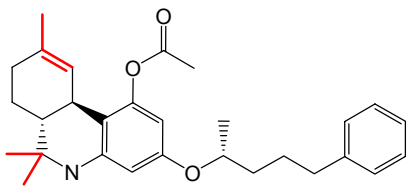
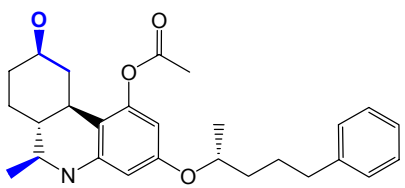


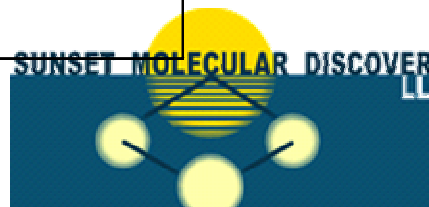
JMC Errors… 2

Reference	Published Structure	Corrected Structure	Comment
JMC 35-1969 chart 1			*bicuculline*: incorrect chirality; incorrect ring fusion
-‖-			*(+)-tubocurarine*: incorrect N atom position; substitution position
JMC 37-1769 chart 1			*haloperidol*: Br instead of Cl
JMC 38-16 scheme 1			*divaplon*: missing nitrogen atom
JMC 43-71 figure 2			*triazolam*: missing chlorine atom

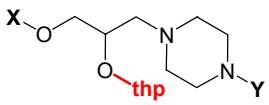
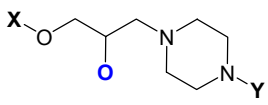
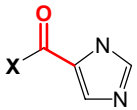
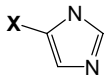
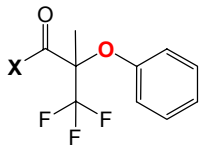
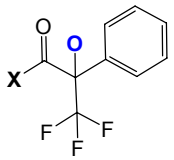
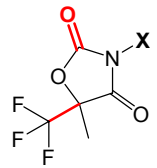
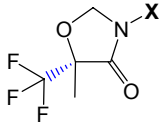
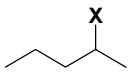
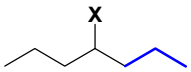
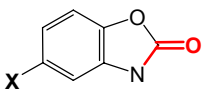
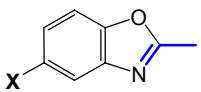


JMC Errors… 3

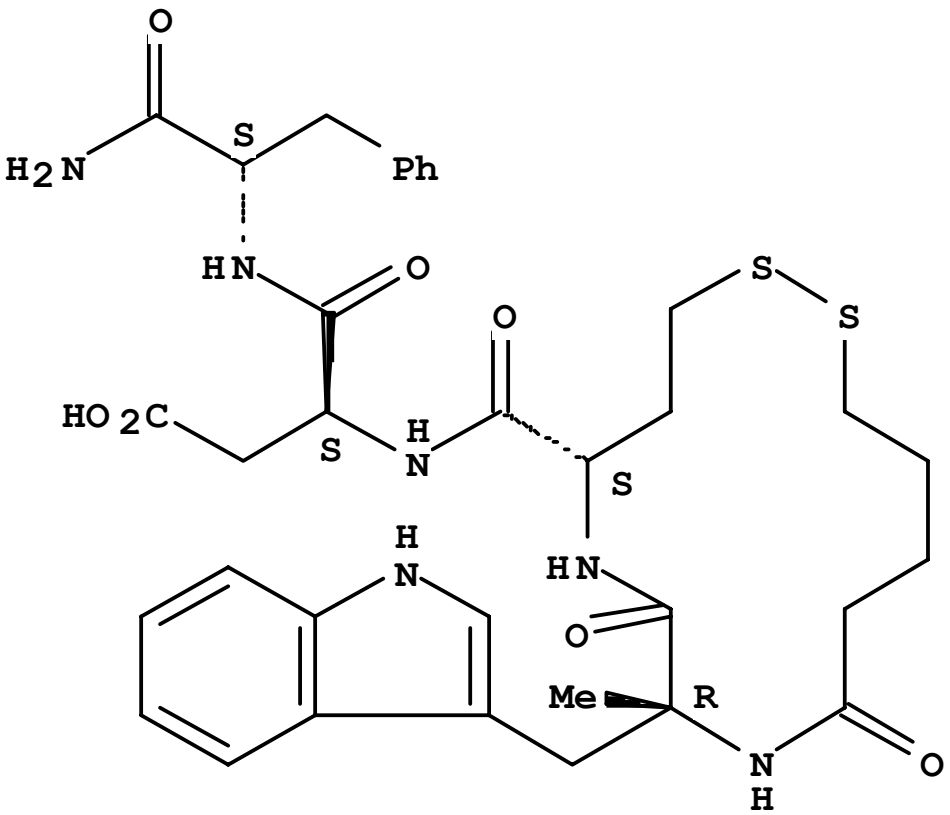
Reference	Published Structure	Corrected Structure	Comment
JMC 43-1793			*argatroban*: missing double bonded oxygen atom; missing chirality
JMC 41-1943 chart 1			*LY-297524*: completely different structures
JMC 41-4196			*SB-203580*: missing S=O double bond
JMC 38-3645 figure 1			*tacrine*: missing two double bonds
JMC 38-3094 figure 1			*levonantradol*: methyl instead hydroxy, methyl & plus an extra double bond

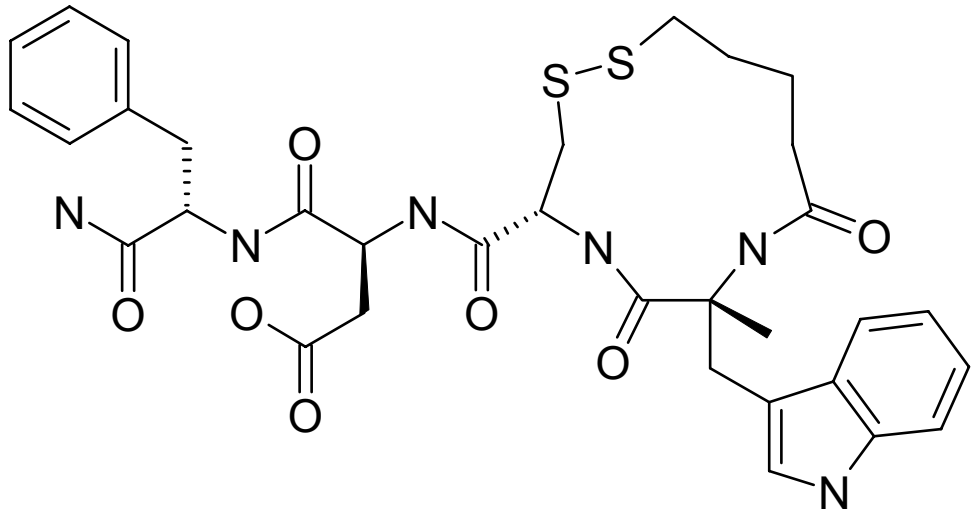


JMC Errors… 4

JMC 35-4509 table II			*40, 41*: THP in plus
JMC 35-3858 table IV			*53*/R_2: imidazoyl instead imidazolyl
JMC 43-236 table 1			*6b*: incorrect substituent
-‖-			*9*: double bonded O atom in plus
JMC 39-3636 table 1			*28xiii*/R_3: pent-4-yl instead hept-4-yl; confirmed from chemical name
JMC 40-1049 table 6			*69*/R: wrong substituent; confirmed from chemical name

Other Errors… SciFinder

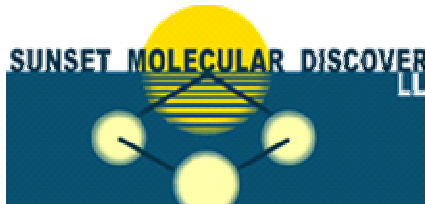




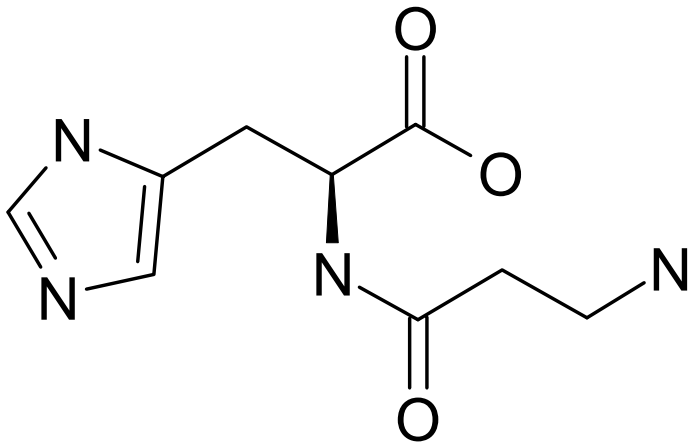
WOMBAT: RB-370

Registry Number: 187454-94-0

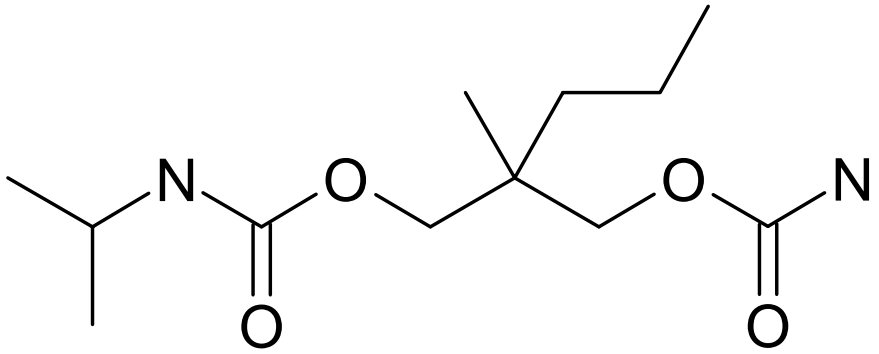
The correct structure has a 13-member ring



Other Errors… Merck Index



"Carisoprodol"
Merck Index 13th ed #1854

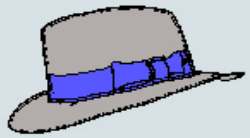


Carisoprodol - correct structure
Merck Index 13th ed has correct name

F E D O R A

fedora: home page
federation of research assets

FEDORA status help



Welcome to fedora

This a collection of research databases which collaborate via weblike interfaces. These servers have been developed by Metaphorics LLC, Daylight CIS and their partners.

Most of these servers require you to log in as a user. You are not currently logged in. If you are a registered user, you may **login here** for access to fedora services. If not (or if you've forgotten your password), you may **register here** for a demo account.

Note: your e-mail address is now your fedora **User Name**.

Note: Netscape (4.7 or later) is the preferred browser; current versions of Mozilla and IE also work.

Normal services:

- **dcm** -- **d**ictionary of chinese **m**edicine *(production, Login required)*
- **empath** -- metabolic pathway chart *(production, Login required)*
- **park** -- photo **ark**ive *(beta, Login required)*
- **planet** -- protein-ligand association network *(production, Login required)*
- **tcm** -- traditional chinese **m**edicines *(production, Login required)*
- **wdi** -- world **d**rug **i**ndex *(production, Login required)*
- **wdidemo** -- world **d**rug **i**ndex (demo) *(demo, Login required)*
- **wombat** -- world **o**f **m**olecular **b**ioactivity *(beta, Login required)*
- **wombatdemo** -- world **o**f **m**olecular **b**ioactivity *(demo, Login required)*

Public services:

- **ecbook** -- Enzyme Commission code book *(production)*
- **medusa** -- molecules in everyday usage *(production)*

Utilities:

- **fedora** -- **fed**eration **o**f research **a**ssets *(production, Login required)*
- **sandman** -- server **and** daemon **man**ager *(production, Login required)*
- **zi4** -- chinese character server: traditional, simplified and pinyin *(production, Login required)*

http://www.metaphorics.com

WOMBAT@FEDORA

wombat: **home page**

world of molecular bioactivity

Beta WOMBAT series authors help fedora

World of Molecular Bioactivity

The **WOMBAT dataset** is a digital curation of published biological activities of small molecules. Almost all such data are reported as mesurements of a series of molecules using a common protocol.The WOMBAT dataset is curated by Sunset Molecular Discovery LLC and is updated twice a year. This dataset is version **2003.1**.

The **wombat service** is a fedora server which provides flexible and convenient access to the wombat dataset using popular web browsers. It also provides communication with other fedora servers provided by Metaphorics, LLC.

Search WOMBAT

Search for almost anything here:

Find JME Submit

Other starting points

- **authors**: A-Z index of **9558** authors, authors.txt, first, last, most prolific
- **enzymes**: list all **648** enzymes, wombat_enzymes.txt, show top 10 enzymes
- **series**: list all **1721** series, references.txt, oldest, newest, biggest
- **keywords**: list all **2267** keywords, wombat_keywords.txt, keyword search
- WOMBAT status

WOMBAT Patterns

- Dave Weininger wrote a SMARTS generator starting from a SMILES that was hand-picked by Vera Povolna to match a *specific* (not the maximum common) substructure for each WOMBAT series
- These SMARTS are intended to capture the unique biological profile for each series – on occasion 2 such SMARTS were defined; note that hydrogens are matched exactly as defined in the series

[CH3]-[OH0]-[cH0]:1:[cH1,cH0]:[cH0]:2-[CH2]-[NH0](-[NH0]=[CH0](-[cH0]:2:[cH1]:[cH1]:1)-[CH2]-[cH0]:3:[cH0](:[cH1]:[nH0]:[cH1]:[cH0]:3-[ClH0])-[ClH0])-[CH0,SH0,CH1]=[OH0]

[CH2]-[CH2]-[NH0](-[CH2]-[CH2])-[CH2]-[CH2]-[OH0,SH0]-[cH0]:1:[cH1]:[cH1]:[cH0](:[cH1] :[cH1]:1)-[CH1]-2-[CH1](-[CH0,CH2]-[OH0]-[cH0]:3:[cH1]:[cH0](:[cH1]:[cH1]:[cH0]-2:3)-[OH0,OH1])-[cH0]:4:[cH1]:[cH1]:[cH1]:[cH1]:[cH1]:4

[OH1]-[CH0](=[OH0])-[CH2]-[CH1,CH2]-[NH1]-[CH0](=[OH0])-[CH2]-[NH1,NH0]-[CH0] (=[OH0])-[CH2,CH1,NH0]-[CH2]-[CH2]-[cH0]:1:[nH0]:[cH0]:2-[NH1]-[CH2]-[CH2]-[CH2]-[cH0]:2:[cH1]:[cH1]:1

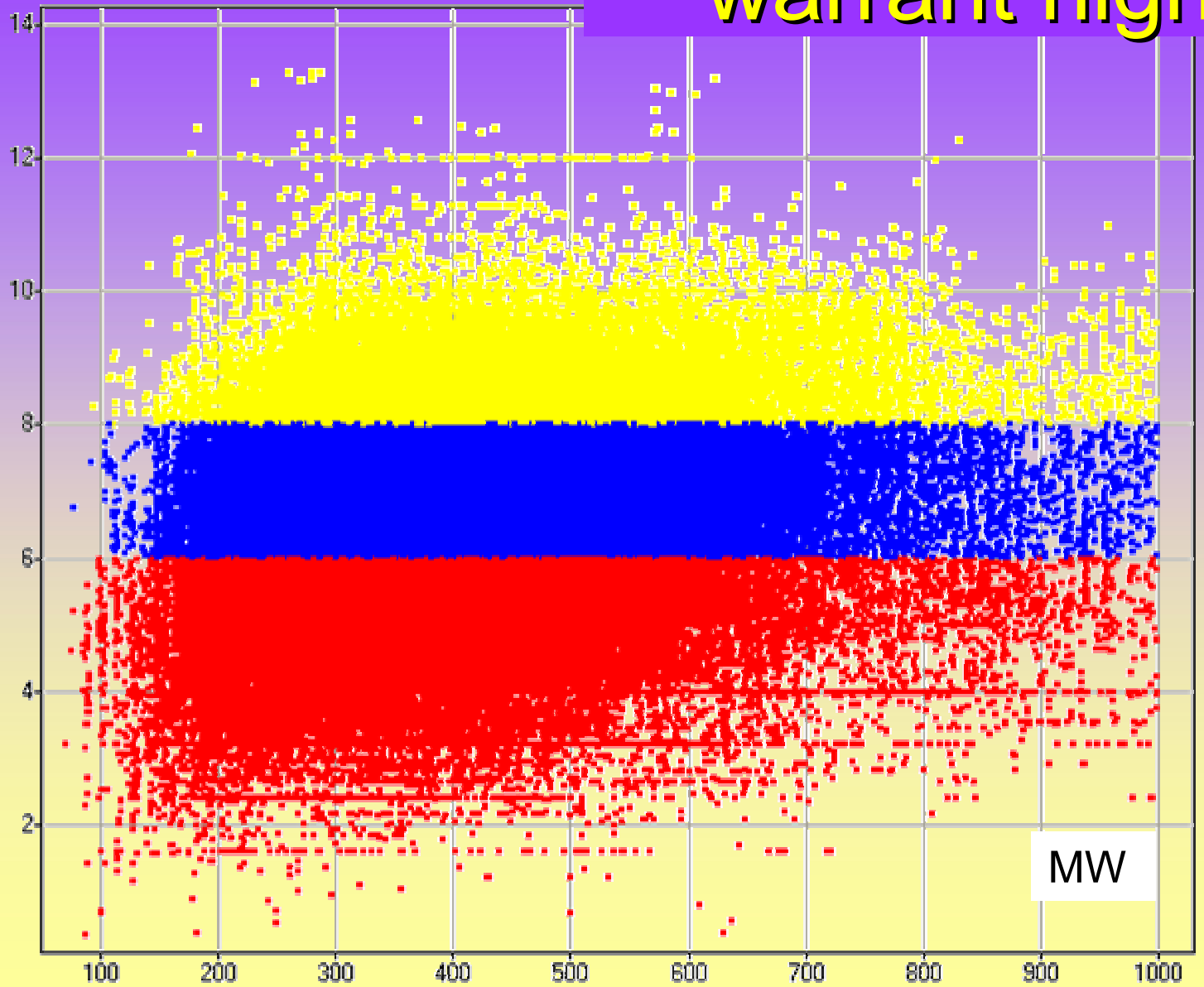
[OH1]-[CH0](=[OH0])-[CH2]-[CH1,CH2]-[NH1]-[CH0](=[OH0])-[CH2]-[NH0]-1-[CH0](-[CH1](-[CH2]-[CH2]-1)-[CH2]-[CH2]-[cH0]:2:[nH0]:[cH0]:3-[NH1]-[CH2]-[CH2]-[CH2]-[cH0]:3:[cH1]:[cH1]:2)=[OH0]

[NH2]-[CH2]-[CH2]-[CH2]-[NH1]-[CH2]-[CH2]-[CH2]-[CH2]-[NH1]-[CH2]-[CH2]-[CH2]-[NH1]-[CH0,SH0]=[OH0]

- Provides interesting associations in FEDORA

SUNSET MOLECULAR DISCOVER LL

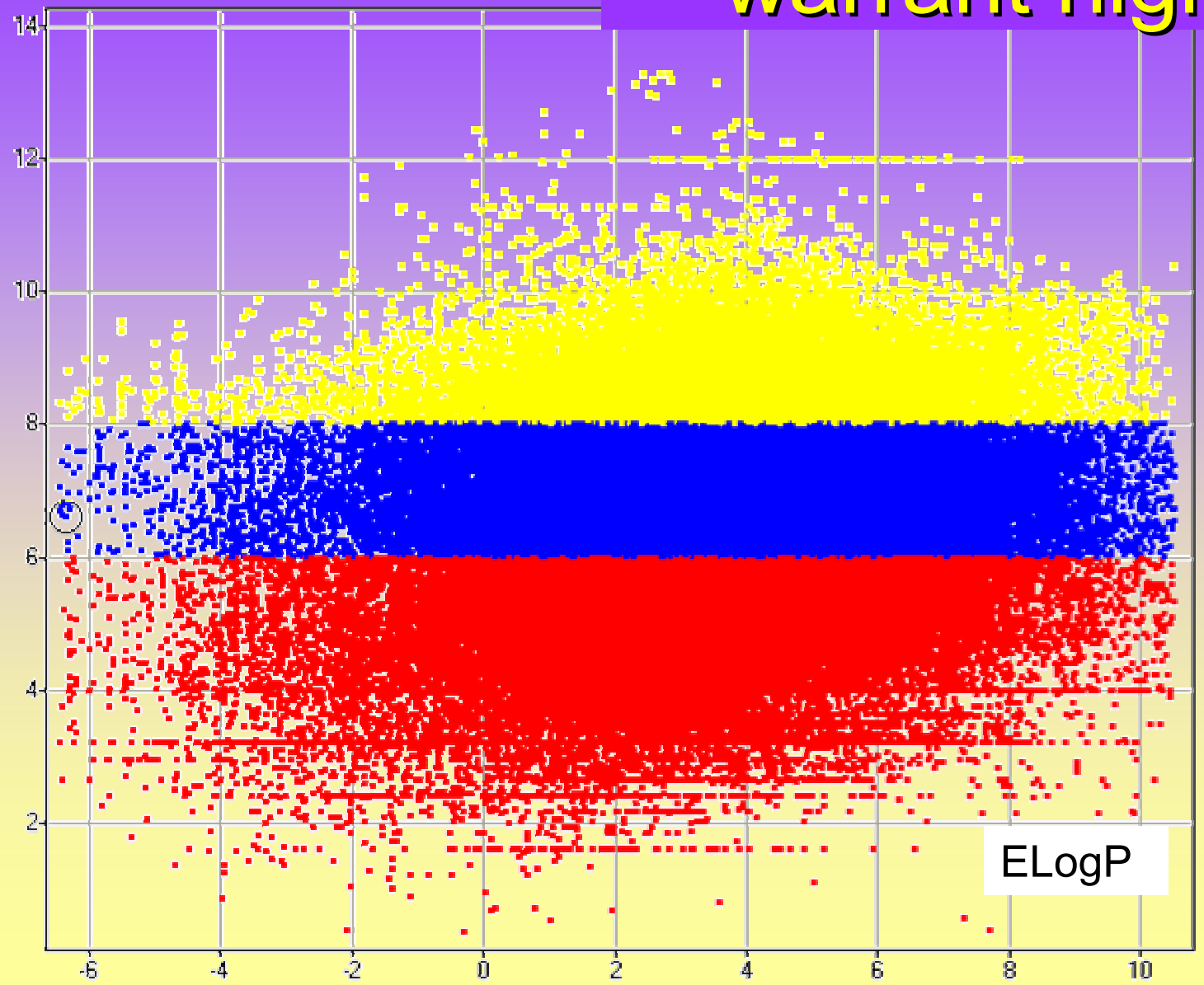
Increased MW does not warrant higher activity



67210 structures
138401 activities

SUNSET MOLECULAR DISCOVER

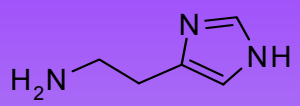
Increased LogP does not warrant higher activity



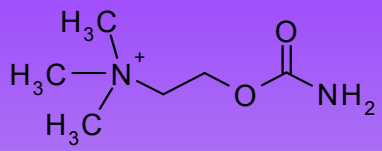
66824 structures
137766 activities

SUNSET MOLECULAR DISCOVER

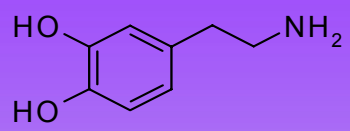
How Small Can Active Compounds Be?



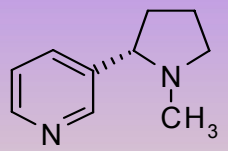
Histamine
MW = 111
LogP = -0.7
K_i = 8.2 (H_3)



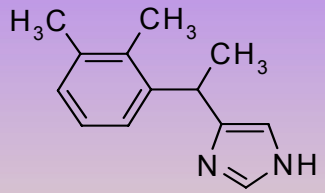
Carbachol
MW = 143
LogP = -3.8
IC_{50} = 8.2 (m)



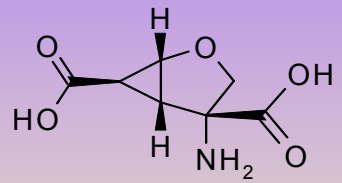
Dopamine
MW = 153
LogP = -1.0
IC_{50} = 8.7 (D_2)



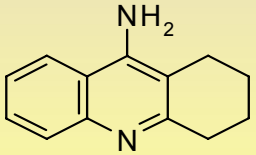
Nicotine
MW = 162
LogP = 1.2
K_i = 9.0 (nACh)



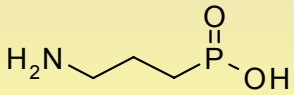
Medetomidine
MW = 200
LogP = 3.8
EC_{50} = 8.5 (α_2)



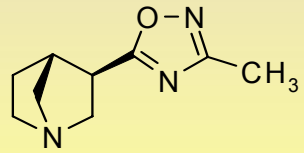
LY-379268
MW = 187
LogP = -4.6
EC_{50} = 8.6 ($mGLU_2$)



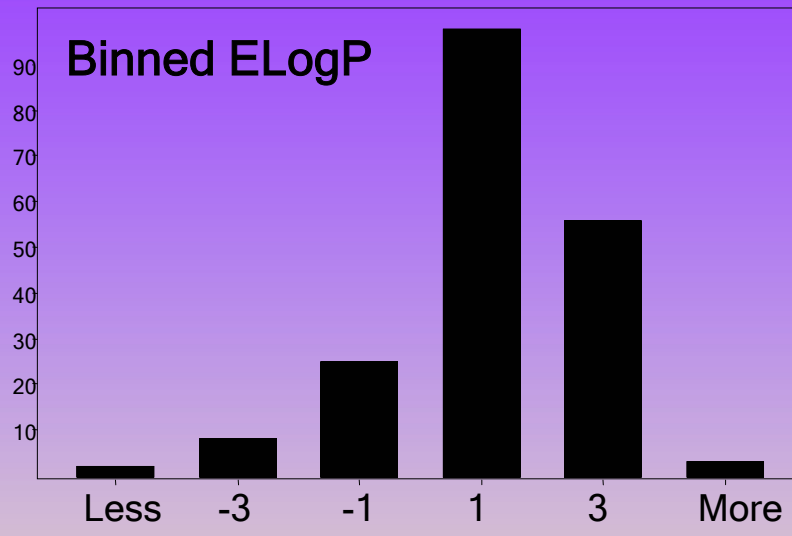
Tacrine
MW = 198
LogP = 2.7
IC_{50} = 8.2 (BChE)



CGP-27492
MW = 123
LogP = -1.7
IC_{50} = 8.6 (GABA-B)



L-670548
MW = 179
LogP = 0.77
K_i = 9.7 (m1)

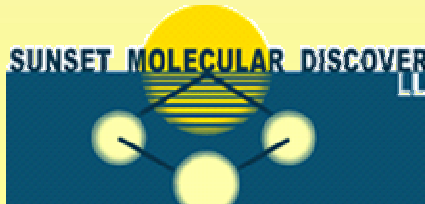


192 unique structures
46 targets
252 activities ≤ 10 nM
MW ≤ 200 amu
176 are likely to be charged at pH 7.4

SUNSET MOLECULAR DISCOVERY LLC

Acknowledgments

- Maria Mracec, Liliana Ostopovici, Ramona Rad, Alina Bora, Ionela Olah, Marius Olah, Magdalena Banda (Timisoara Institute of Chemistry of the Romanian Academy) and TIO introduced data in WOMBAT
- Marius Olah wrote the database interfaces
- Maria Mracec, Marius Olah and TIO did the keyword characterization
- Marius Olah, Maria Mracec, Cristian Bologa and TIO performed structural error checking
- Vera Povolna and David Weininger (Metaphorics) for implementing WOMBAT@FEDORA



EuroQSAR 2004

Chair: Prof. Dr. Esin AKI ŞENER
sener@pharmacy.ankara.edu.tr
Co-Chair: Prof. Dr. İsmail YALÇIN
yalcin@pharmacy.ankara.edu.tr
Address for Correspondence:
Armoria Congress
armoria@euro-qsar2004.org

15th European Symposium on Quantitative Structure-Activity Relationships
Istanbul / Turkey 05-10 September 2004
http://www.euro-qsar2004.org