

# Chemo-informatics and Medicinal Chemistry: Rules, Filters and Common Sense

Christopher A. Lipinski  
Scientific Advisor, Melior Discovery  
[clipinski@meliordiscovery.com](mailto:clipinski@meliordiscovery.com)

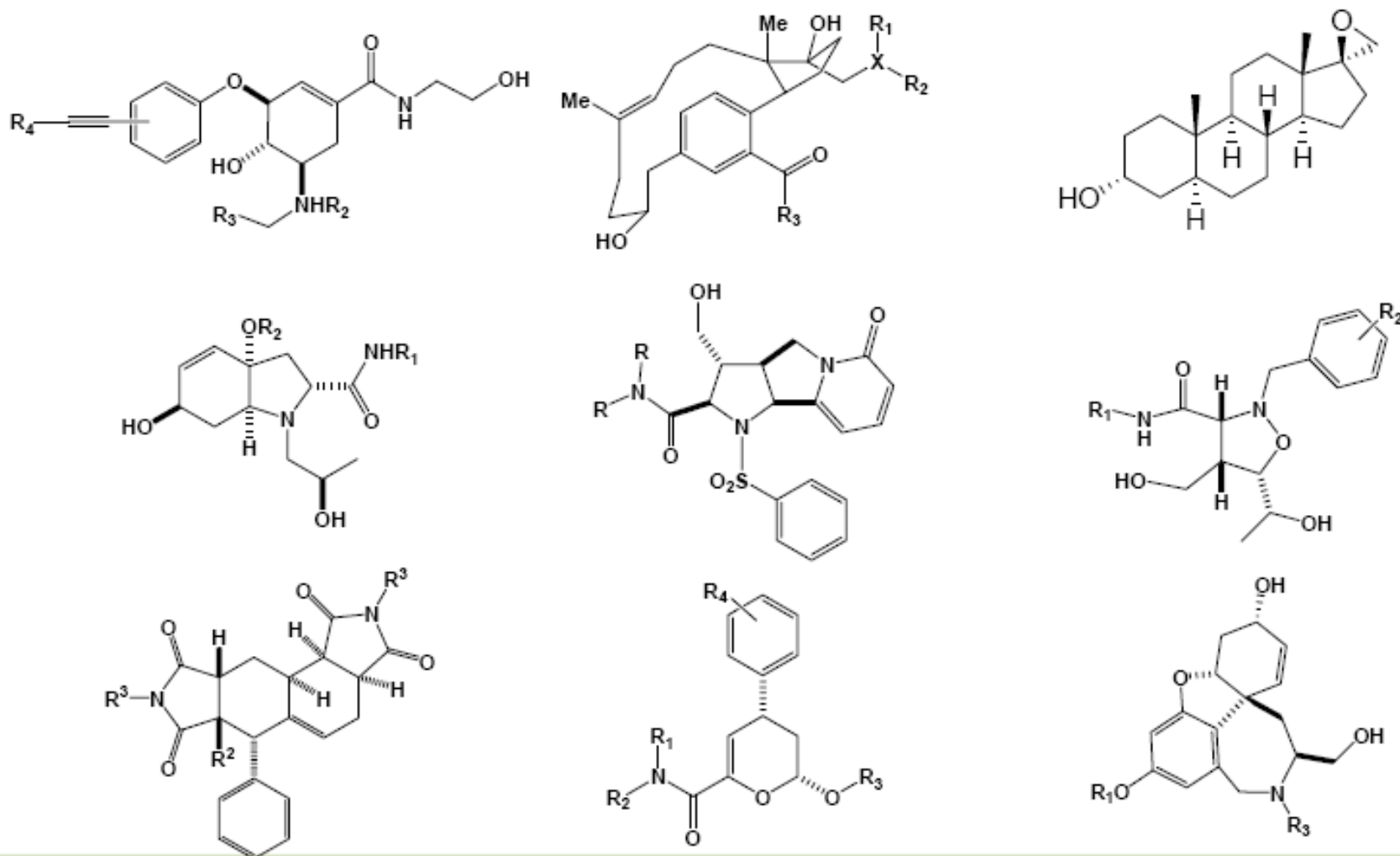
# Outline

- What is a medicinal chemist
  - How a medicinal chemist looks at structures
  - Medicinal chemistry pattern recognition
  - How to interact with medicinal chemists
  - Software for medicinal chemists
- 
- Beautiful biology ruined by bad chemistry
  - Annotation to uncover a chemistry problem
  - Beware the commercial CAS with no references
  - Beware the reverse cyclization chemistry
- 
- Conclusions

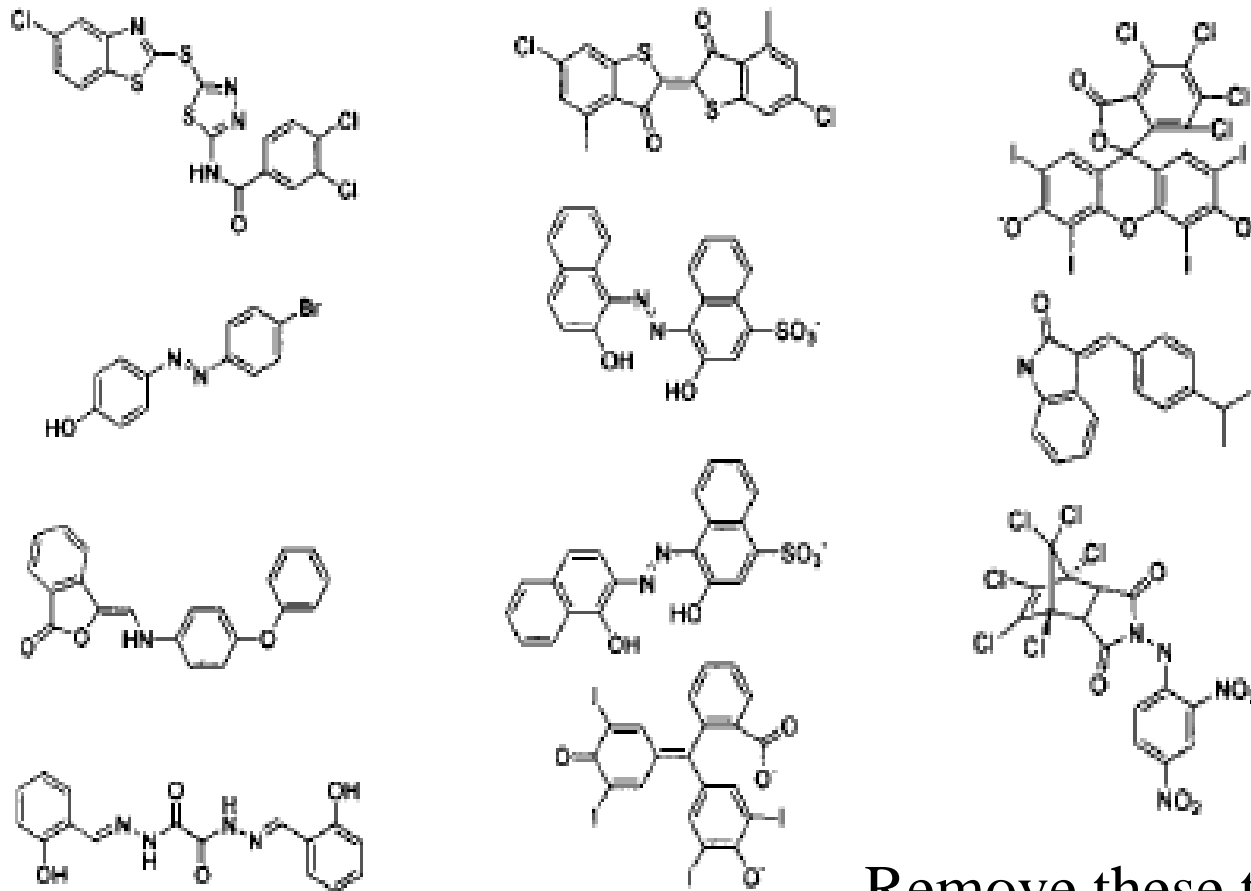
# What is a medicinal chemist?

- Solid synthetic chemistry background
- Chemical structures associated with biomedical information pattern recognition
- Chemical structures are the language
- Expert status requires time to learn patterns
  - 10,000 hour rule (10 years)
- Medicinal chemistry is a pattern recognition discipline

# Nice chemistry: topology in DOS libraries



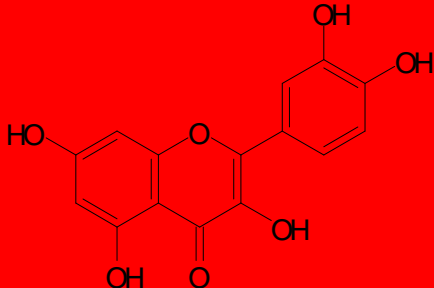
# Bad chemistry: aggregation false positives in HTS assays



These looked good to pharma  
screeners

Remove these types of  
compounds from any  
assays

# Chemistry “quality” and biology effort

CHEMISTRY	CAS	NAME	Promiscuous	Refs
				
117-39-5	117-39-5	Quercetin	1	<b>9644</b>
482-89-3	482-89-3	Indigo	1	<b>1170</b>
479-41-4	479-41-4	Indirubin	1	<b>263</b>
133052-90-1	133052-90-1	Bisindolylmaleimide I	1	<b>155</b>
109511-58-2	109511-58-2	U0126	1	<b>127</b>
125314-64-9	125314-64-9	Bisindolylmaleimide IX	1	<b>99</b>
82-08-6	82-08-6	Rottlerin	1	<b>95</b>
85753-43-1	85753-43-1	K-252c	1	<b>74</b>
145-63-1	145-63-1	Suramin	0	<b>1349</b>
19545-26-7	19545-26-7	Wortmannin	0	<b>510</b>
103745-39-7	103745-39-7	HA-1077	0	<b>140</b>
146986-50-7	146986-50-7	Y-27632	0	<b>113</b>
152121-47-6	152121-47-6	SB203580	0	<b>262</b>
167869-21-8	167869-21-8	PD 98059	0	<b>310</b>
318480-82-9	318480-82-9	SC68376	0	<b>8</b>

> 80%

flawed  
compounds?

# Structure Activity Relationships (SAR)

- Pattern of chemistry structures
- Depiction of numerical biology data
- Linguistic (text) description of SAR
  
- A medicinal chemist discerns some SAR just from the pattern of chemical structures
- A computational chemist needs data
- A lawyer needs a statement of the SAR

# The amygdala and emotional memory

- Pattern recognition is evolutionarily selected
- Chemists are particularly social
- Chemists are superb at pattern recognition
- Structure of “emotionally significant” compounds is locked into the amygdala
- Biologists just do not understand this

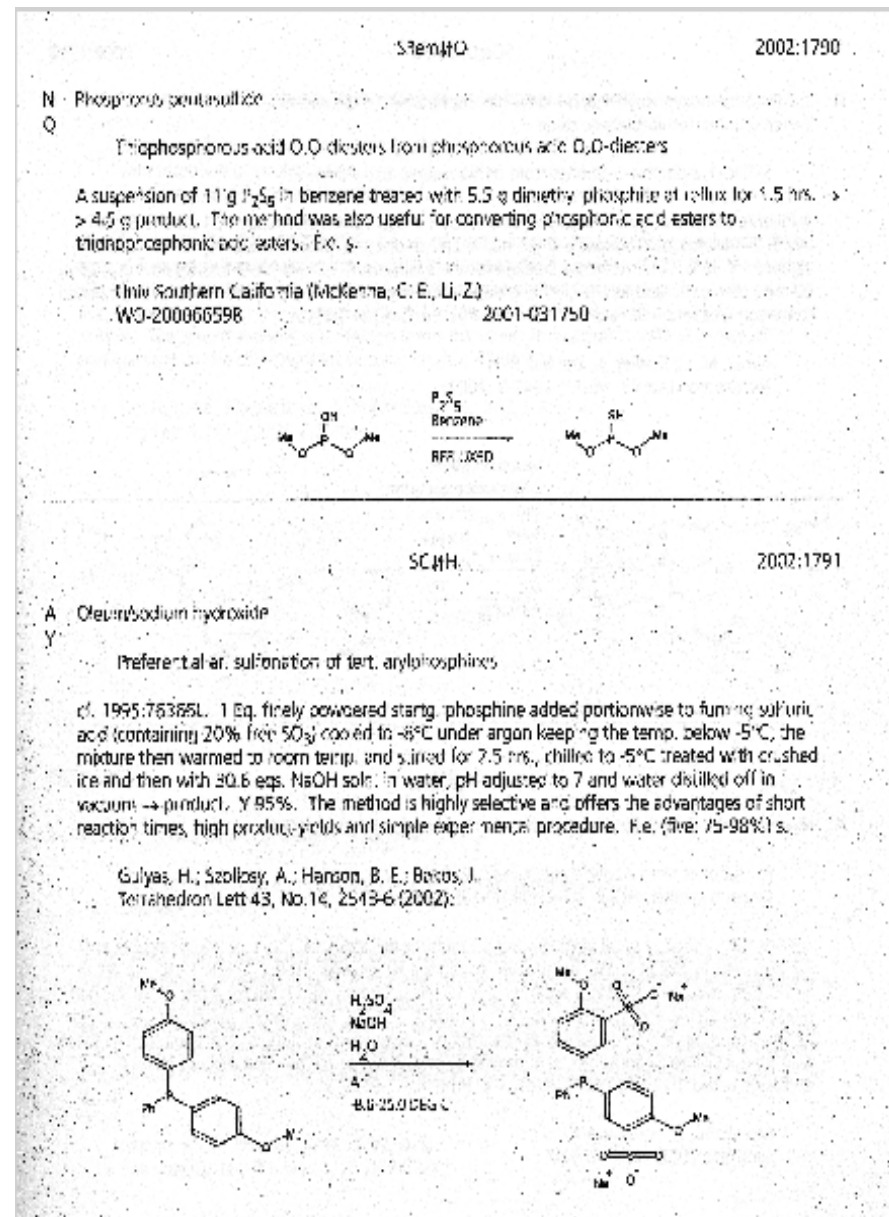


# Incomprehensible

“Reality” depiction in chemistry is incomprehensible to biology / genomics and vice versa.

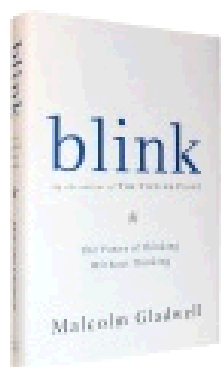
Chemistry much closer to pathology than to biology / genomics.

Graphical pattern recognition is the forte of chemists



# Chemistry pattern recognition

## What is Blink about?



BLINK

---

### 1. What is "Blink" about?

It's a book about rapid cognition, about the kind of thinking that happens in a blink of an eye. When you meet someone for the first time, or walk into a house you are thinking of buying, or read the first few sentences of a book, your mind takes about two seconds to jump to a series of conclusions. Well, "Blink" is a book about those two seconds, because I think those instant conclusions that we reach are really powerful and really important and, occasionally, really good.

# Influencing medicinal chemists

- Show a beautiful chemistry structure
- Avoid showing equations
- Avoid acronyms for compounds
- Use “real” descriptors
  - good eg. “rule of 5”
  - bad eg. topological indices
- Get the medicinal chemist to take ownership of your idea

# Where medicinal chemists are weaker

- Dislike equations and math
  - math is not required to synthesize compounds
  - use a graphic to illustrate your equation
- Dislike uncertainty, need SAR guidance
  - just tell me is it active or inactive!
  - tend to ignore error limits or annotations
  - Markush structures don't exist in real chemistry
- Influenced by past SAR history
  - fresh computational perspective is very valuable

# Chemistry- biology differences

- Chemistry

- certainty as to chemistry structures

- accepted proofs of structure are routine
- synthesis reproducibility is good
- data reported “as is” with no error limits

- Biology

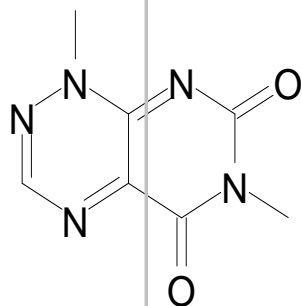
- uncertainty as to biological data

- reproducibility is poor, biology is typically complex
- data is always reported with error limits

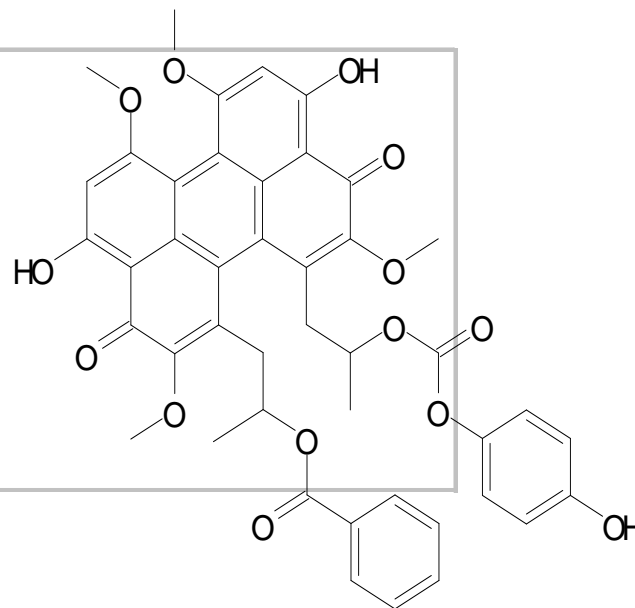
# Do medicinal chemists agree?

- How is the question posed?
- Binary
  - good or bad?
  - agreement may not be great
- Stepwise graded questions
  - do you see a potential problem?
  - can you fix the problem?
  - much better agreement
    - eg. consensus in medchem panel scoring

# Protein protein interaction garbage



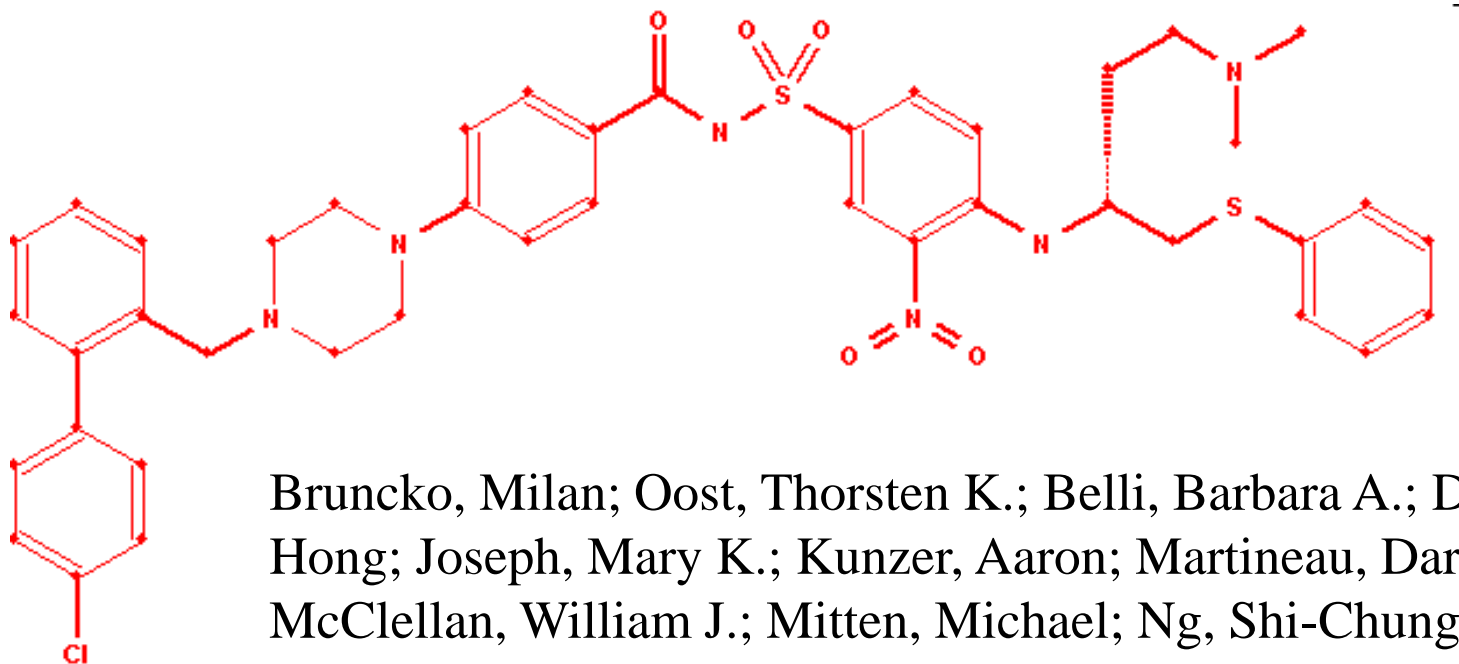
84-82-2



121263-19-2

Lepourcelet, Maina; Chen, Ying-Nan P.; France, Dennis S.; Wang, Huisheng; Crews, Phillip; Petersen, Frank; Bruseo, Charles; Wood, Alexander W.; Shivdasani, Ramesh A. Small-molecule antagonists of the oncogenic Tcf/ $\beta$ -catenin protein complex. *Cancer Cell* (2004), 5(1), 91-102.

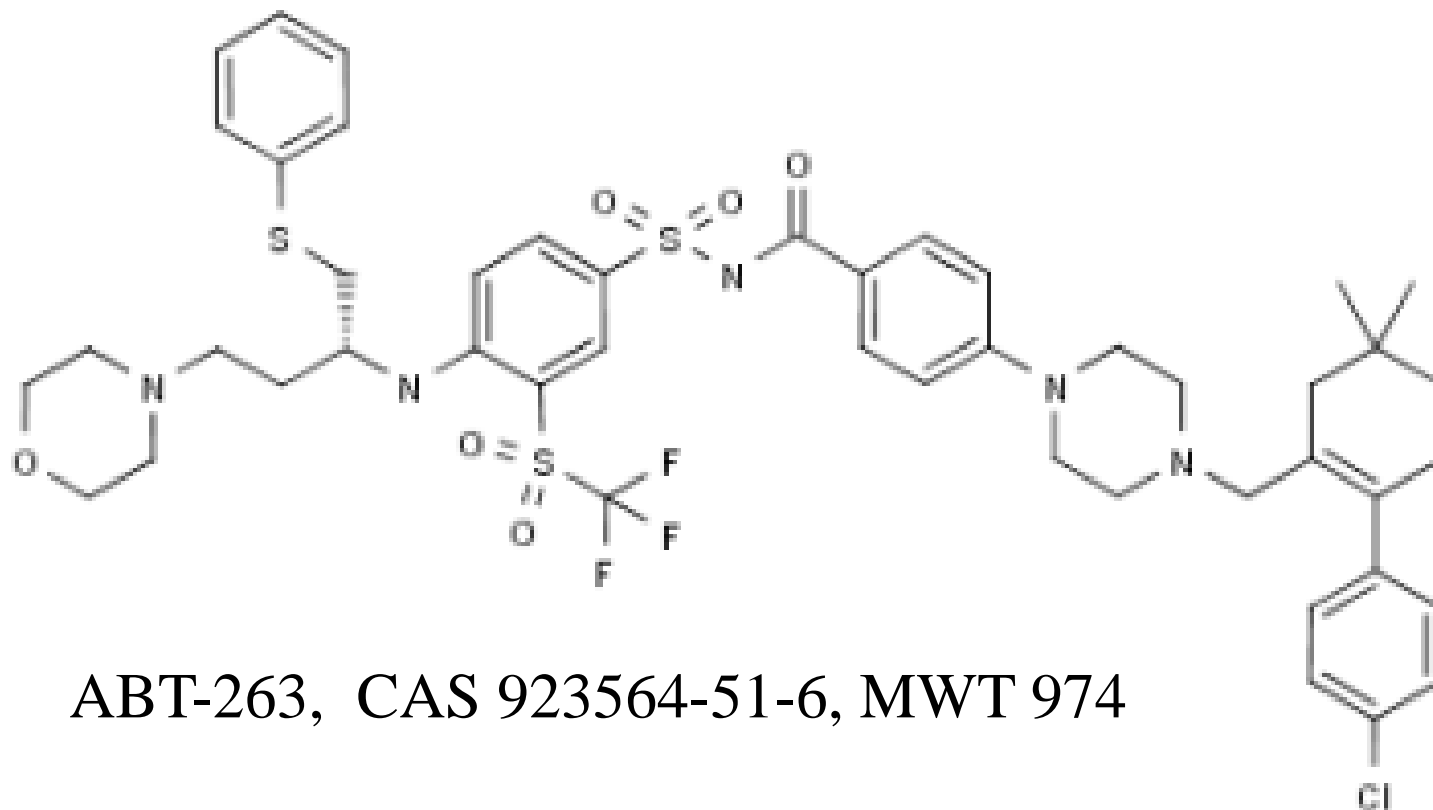
# Protein protein ligand ABT-737 lead



Bruncko, Milan; Oost, Thorsten K.; Belli, Barbara A.; Ding, Hong; Joseph, Mary K.; Kunzer, Aaron; Martineau, Darlene; McClellan, William J.; Mitten, Michael; Ng, Shi-Chung; Nimmer, Paul M.; Oltersdorf, Tilman; Park, Cheol-Min; Petros, Andrew M.; Shoemaker, Alexander R.; Song, Xiaohong; Wang, Xilu; Wendt, Michael D.; Zhang, Haichao; Fesik, Stephen W.; Rosenberg, Saul H.; Elmore, Steven W. Studies Leading to Potent, Dual Inhibitors of Bcl-2 and Bcl-xL. *Journal of Medicinal Chemistry* (2007), 50(4), 641-662.



# BCL-2 inhibitor compound in phase II

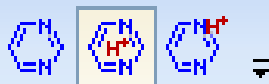


# Why is ABT-263 orally active???

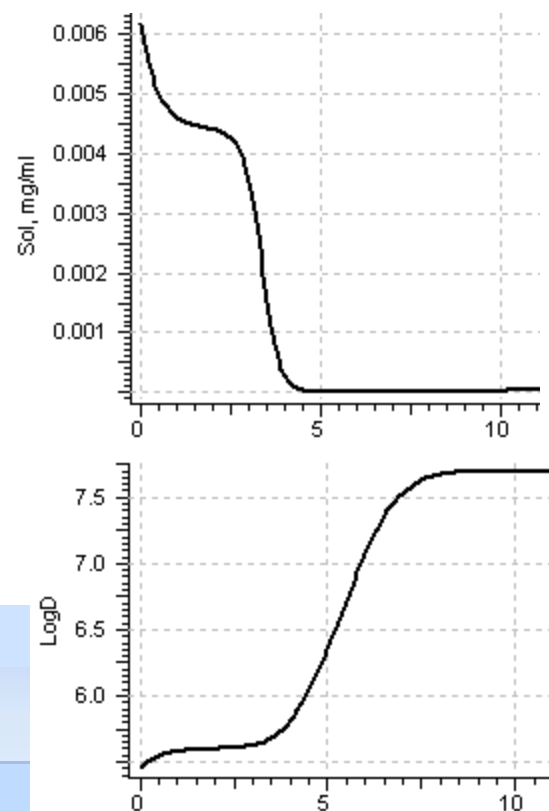
## Single-valued Properties

Name	Value	Error
LogP	9.7	1.21
MW	974.61	-
PSA	170.42	-
FRB	15	-
HDonors	2	-
HAcceptors	11	-
Rule Of 5	3	-

## pKa Results



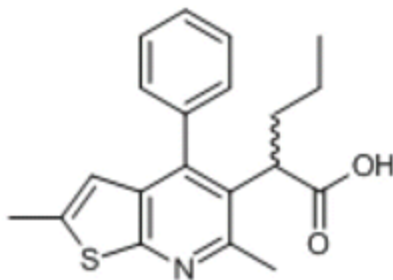
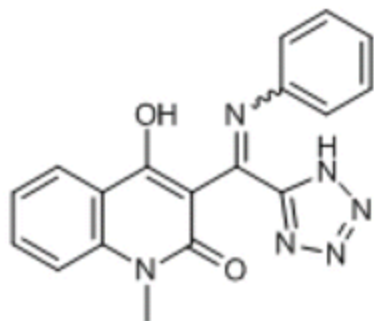
Diss. Atom	Acidic/Basic	Exact Apparent pKa V...	Error
54	MB	7.24	0.1
3	B	6.32	0.7
30	MA	4.64	0.1



Bioavailability is 20-50% depending on formulation

# Filtering and HTS common sense

- Filter enough to avoid HTS false positives
- Allow “flawed” compounds in screen if:
  - compound is not an HTS false positive
  - compound flaw is fixable in chemistry
  - people discipline exists to fix the flaw



**Rational design of small -  
molecule inhibitors of the  
LEDGF/p75 - integrase  
interaction and HIV  
replication. *Nature Chemical Bi  
ology* 6, 442–448 (2010)**

# Software for medicinal chemists

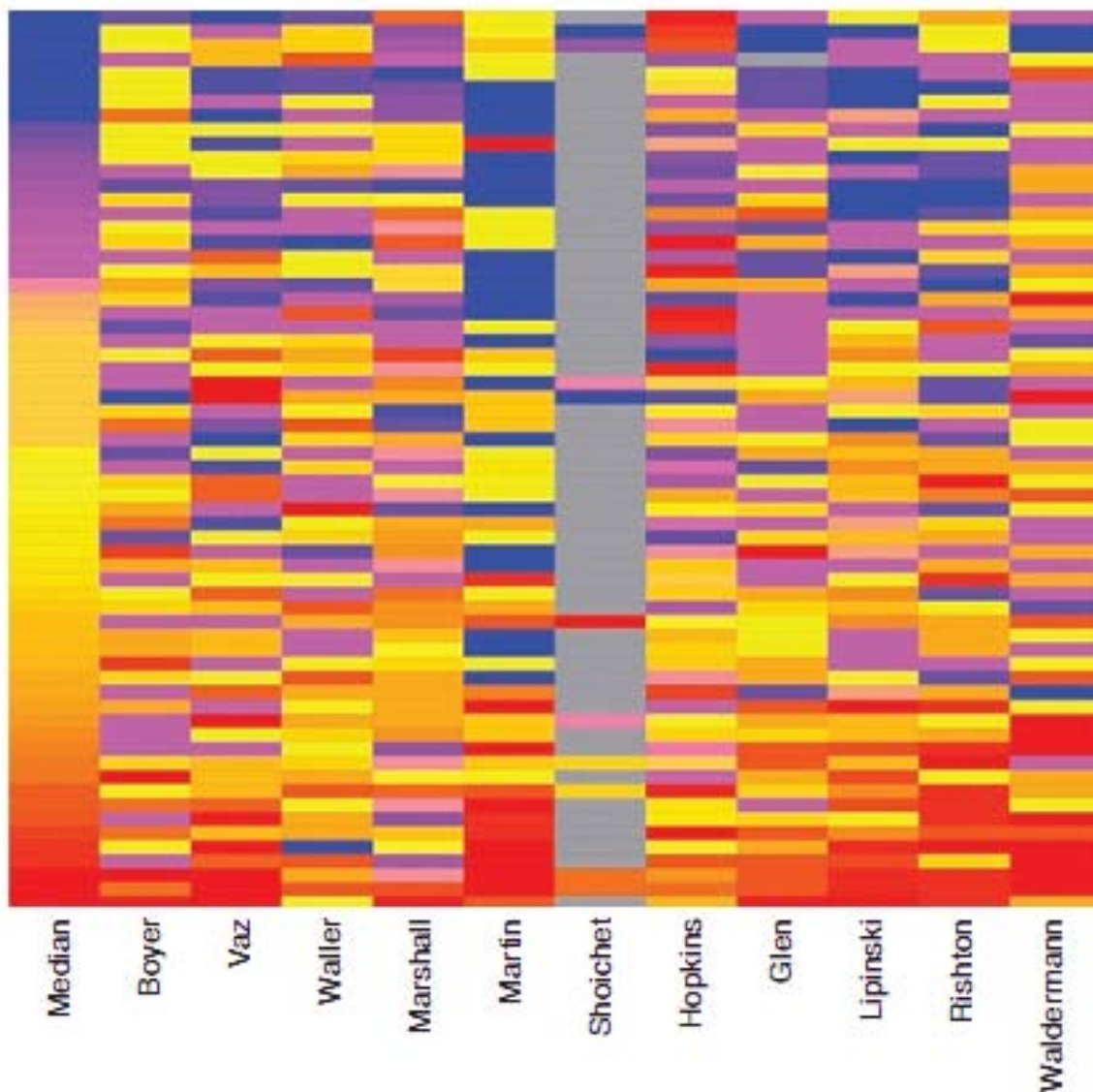
- 10% are true believers
  - attend the software workshops
  - ask for new features
- 90% are skeptics / slow adopters
  - intolerant of software intricacies
  - willing to spend an hour learning – maybe!
  - may need to be lead by “first adopters”
  - may need “dumbed down” software

# Strategies for looking at putative HTS hits

- Order compounds by “core ring or scaffold”
  - correlate core ring to activity
  - is core ring a frequent hitter?
- Put compounds through rules / filters
  - identify undesirable / reactive functionality

-----
- Medicinal chemistry annotation
  - examine compounds similar to the “hit”
  - look at chemistry, biology references

# Order compounds by “core ring or scaffold”



Oprea et al.  
Nature Chemical  
Biology 2009,  
5(7), 441-447.

Red is high  
dubiousity (low  
confidence), blue  
is low dubiousity  
(high confidence)

**NIH 64 chemical  
biology tools probes**

# Medicinal chemistry annotation

- Start with the structure of a hit. Is it known?
- What do you see in a substructure search?
- Try to understand the chemistry. How were the compounds made and how might they react?
- What is the pattern in the literature for compounds at about 85% similarity
- Look at 10 – 20 compounds and references.
- This type of annotation is best done with CAS SciFinder<sup>®</sup>
  - almost impossible to do using public domain tools

# How do we judge biology value?

- New biology appears in the literature
- Initially the biology looks interesting
- Chemistry in the biology has problems
- How to judge value if the chemistry tools illustrating the biology have potential flaws



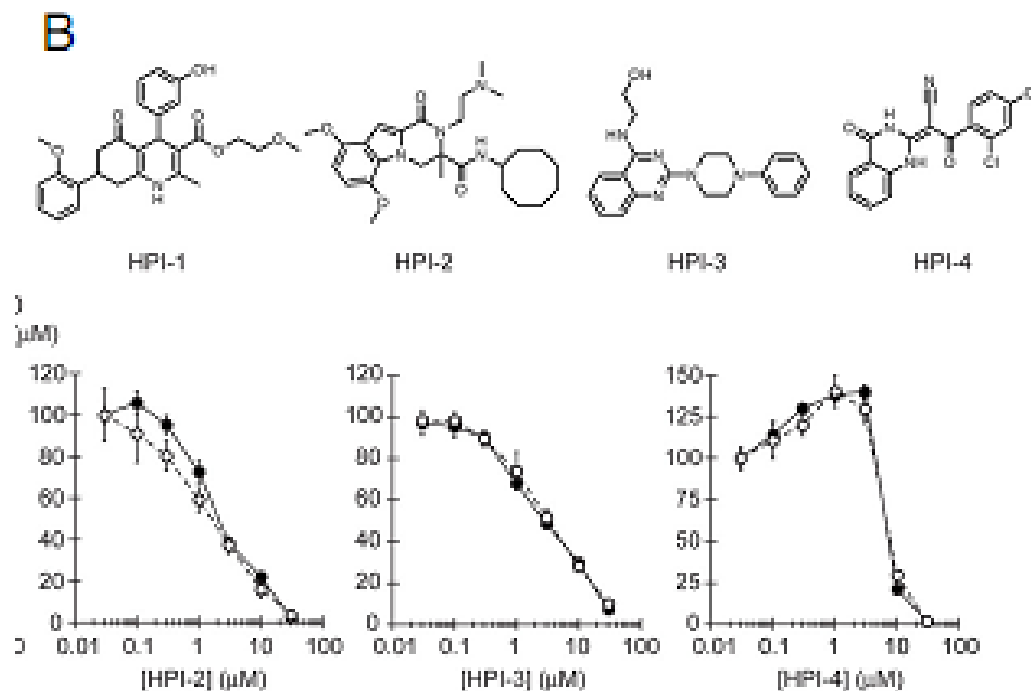
# Biology enthusiasm, but chemistry questions

## Small-molecule inhibitors reveal multiple strategies for Hedgehog pathway blockade

Joel M. Hyman<sup>a,1</sup>, Ari J. Firestone<sup>a,1</sup>, Vivi M. Heine<sup>b</sup>, Yun Zhao<sup>c,d</sup>, Cory A. Ocasio<sup>a</sup>, Kyuho Han<sup>a</sup>, Mark Sun<sup>a</sup>, Paul G. Rack<sup>a</sup>, Surajit Sinha<sup>a,2</sup>, Jason J. Wu<sup>e</sup>, David E. Solow-Cordero<sup>e</sup>, Jin Jiang<sup>c</sup>, David H. Rowitch<sup>b</sup>, and James K. Chen<sup>a,3</sup>

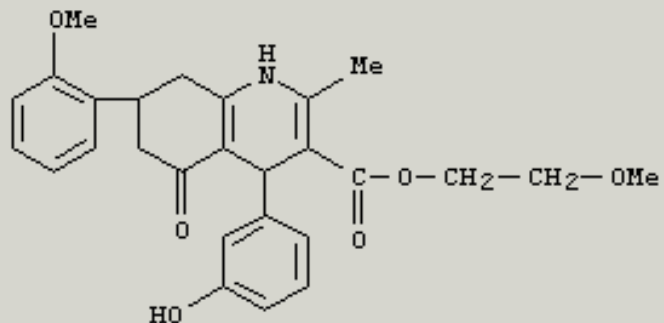
<sup>a</sup>Department of Chemical and Systems Biology and <sup>4</sup>Stanford High-Throughput Bioscience Center, Stanford University School of Medicine, Stanford, CA 94305; <sup>b</sup>Institute for Regenerative Medicine, Howard Hughes Medical Institute, University of California, San Francisco, CA 94143; <sup>c</sup>Department of Developmental Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390; and <sup>d</sup>Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Communicated by Matthew P. Scott, Stanford University School of Medicine, Stanford, CA, June 29, 2009 (received for review January 9, 2009)

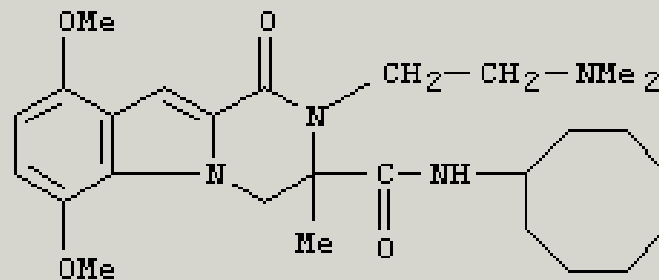


# The four compounds

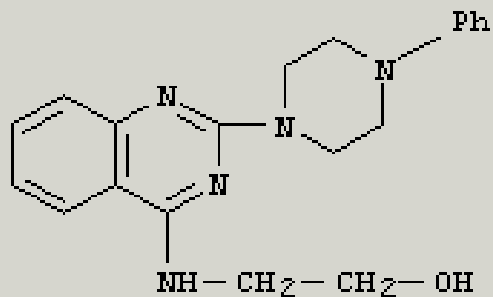
599150-20-6



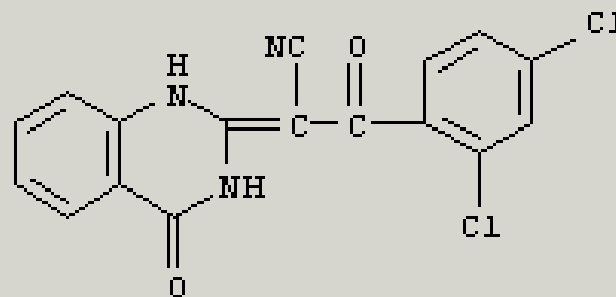
868881-36-1



796887-98-4



302803-72-1



# Hedgehog screening - my comment

## 4. [Chris](#) on August 12, 2009 2:18 AM writes...

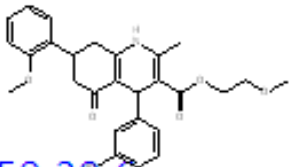
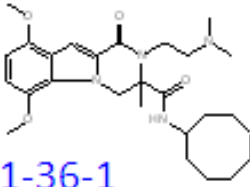
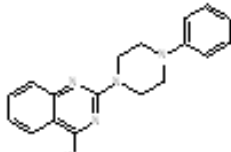
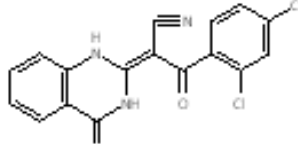
There is a common theme to the four "actives" identified in this paper. They are all commercially available compounds with a CAS registry number and (almost) no literature references. In each case there are commercially available analogs at high similarity again with CAS registry numbers and again no references. I frequently see this pattern in "actives" and it makes me deeply suspicious. What do you think is the probability that a vendor would make a totally novel series just to hit in my new screen? If I were suspicious I might think that the origin of each series was a compound with a flaw that hit enough screens to warrant preparing a flawed analog series. I particularly do not like HPI-4 with a push pull polarized double bond crying out "I am a Michael acceptor please interact with me".

**“actives” all commercially available compounds  
.... no literature references ..... suspicious**

## A profile to avoid

- The structure of a hit appears in CAS SciFinder<sup>®</sup>
- It is a commercial compound with a CAS Registry Number but no references
- There are multiple compounds at 85% or better similarity
- All the similar compounds are commercially available with no literature references
- **WARNING FLAG**
- This could be a problematic series that proliferates because it is a flawed HTS hit series

# Hedgehog screening – thiol trap filters

CHEMISTRY	NAME	Alarm	smartsfilter_matches
 599150-20-6	HPI-1	<b>failed</b>	<chem>C=CC(=O)O[c,C] ()</chem> <chem>C=CC(=O)[c,C] ()</chem> <chem>Oc1[a;R1][a;R1]a[a;R1][a;R1]1 ()</chem>
 868881-36-1	HPI-2	<b>passed</b>	
 796887-98-4	HPI-3	<b>failed</b>	<chem>[N;!\$([N+]);!\$(NC=[O,N])]c1[a;R1][a;R1]a[a;R1][a;R1]1 ()</chem>
 302803-72-1	HPI-4	<b>failed</b>	<chem>C=CC(=O)[c,C] ()</chem>

# Excellent recent paper

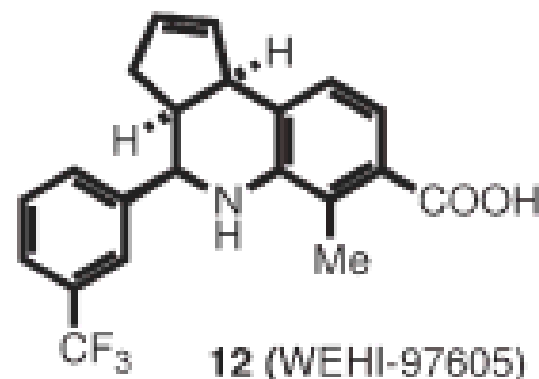
## New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays

Jonathan B. Baell and Georgina A. Holloway

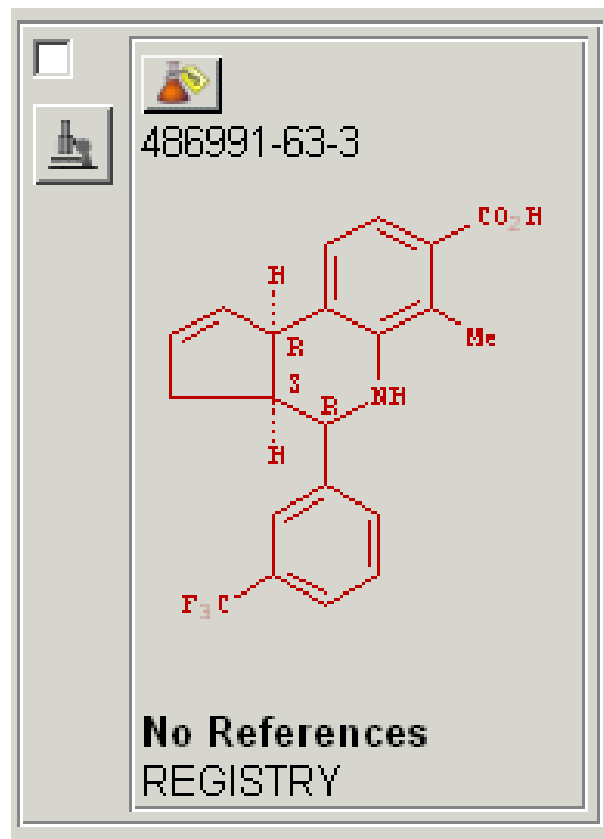
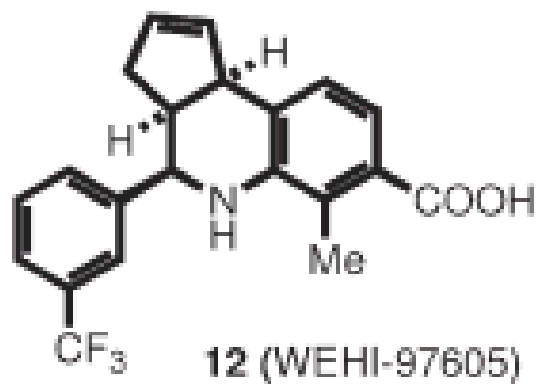
*J. Med. Chem.*, Articles ASAP (As Soon As Publishable)

**Publication Date (Web):** February 4, 2010 ([Article](#))

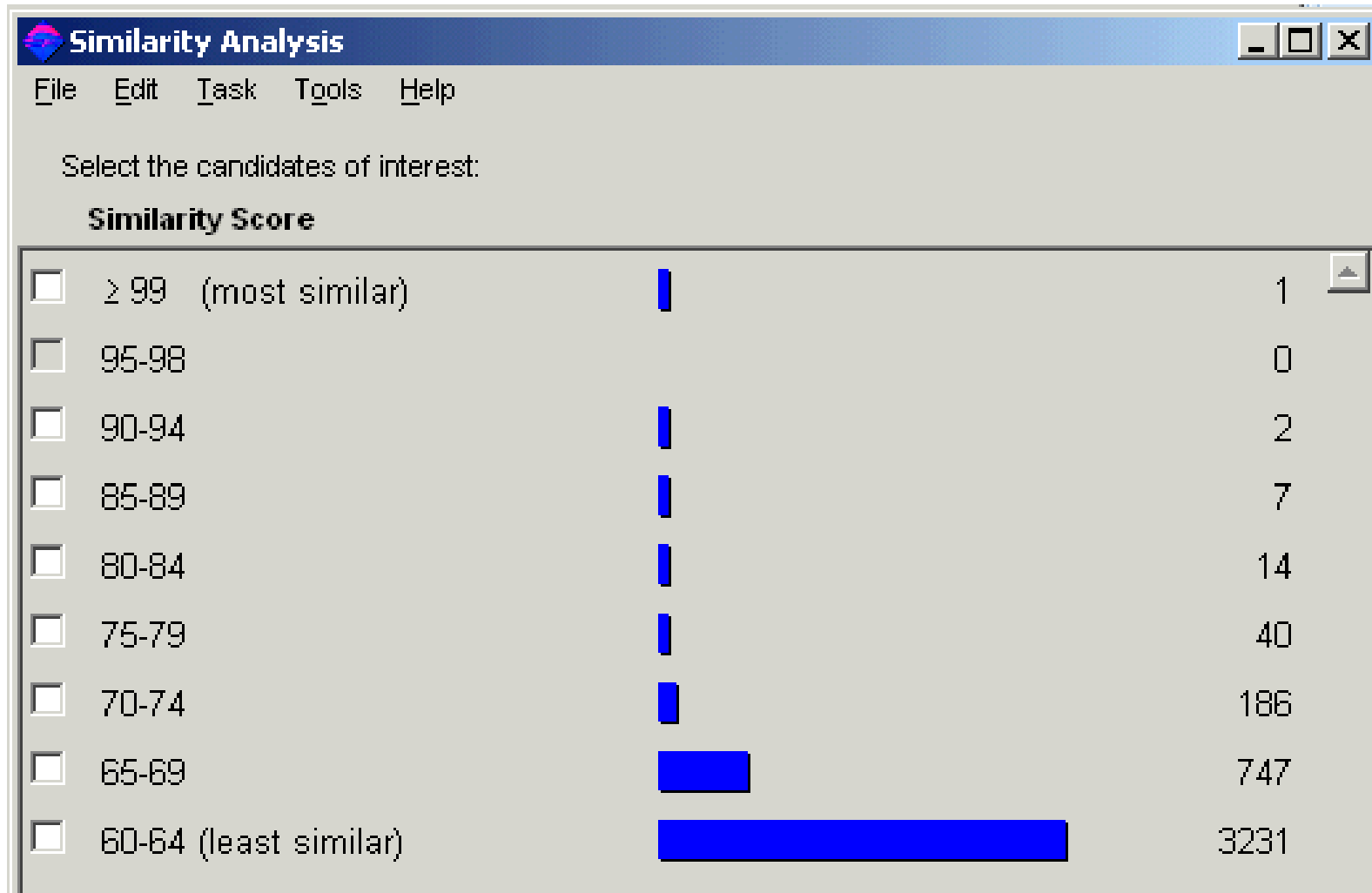
**DOI:** 10.1021/jm901137j



# Warning flag

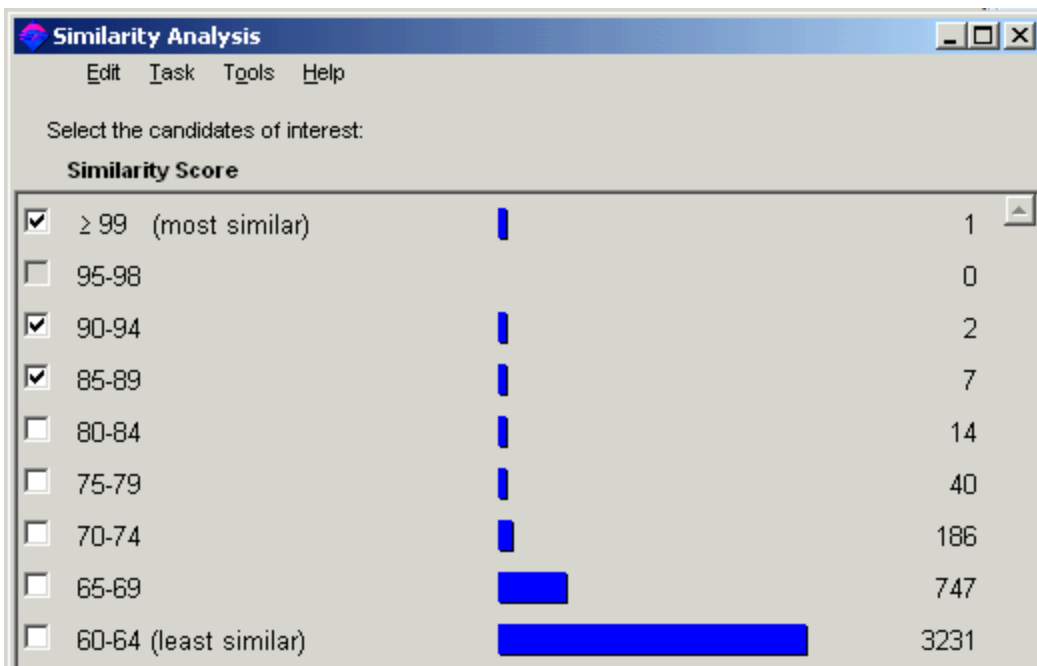


# Wow! Lots of analogs



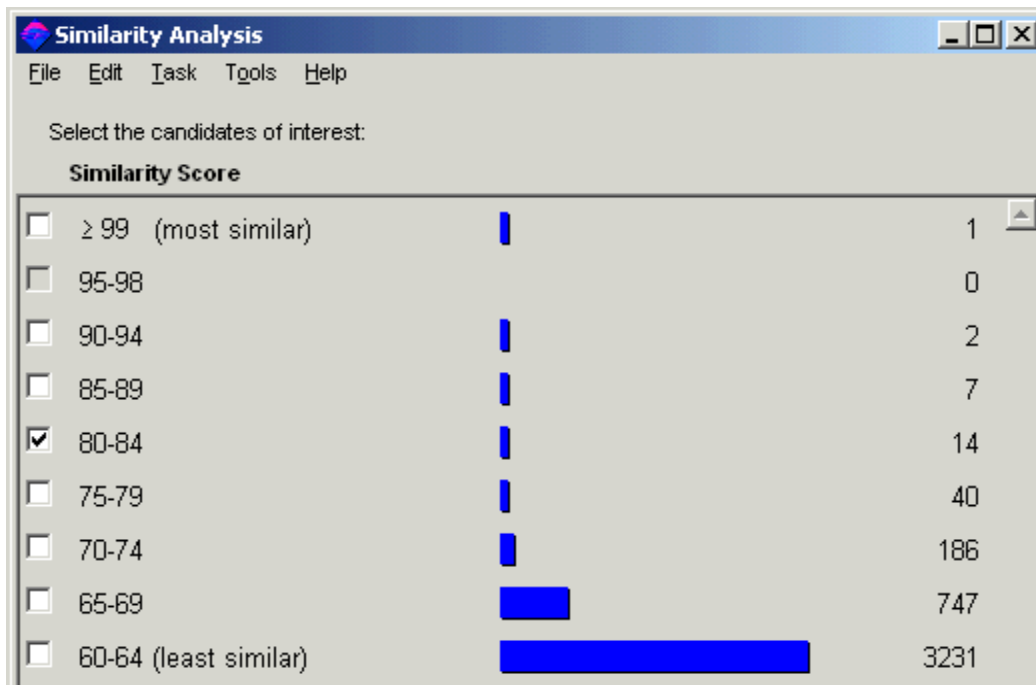


# 9 compounds at 85% or better similarity



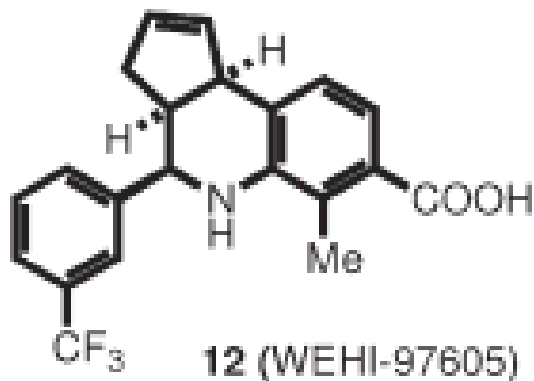
All 9 are commercially available with no literature references. **More warning flags**

# 14 compounds at 80 – 84% similarity

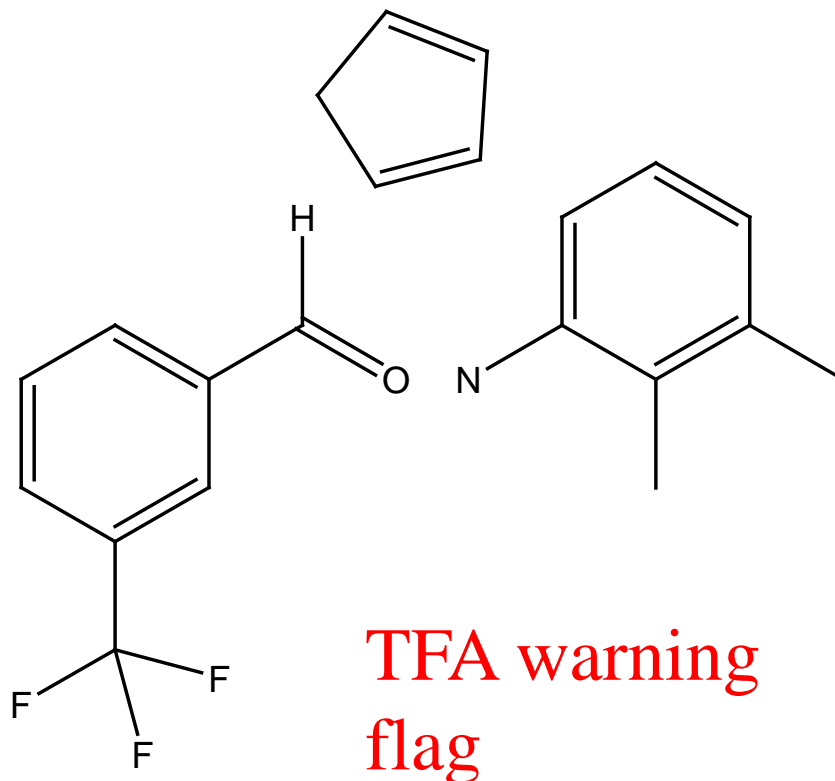


13 of 14 are commercially available with no literature references. One reference is to an HTS based patent. **Even more warning flags**

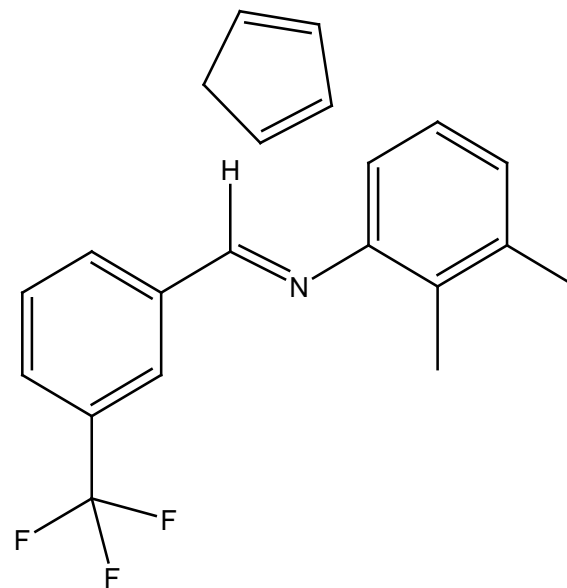
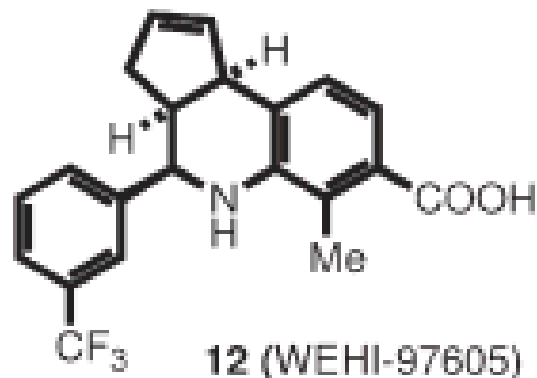
# Three component acid catalyzed cyclization



WEHI-97605 is made by TFA cyclization of cyclopentadiene, m-trifluorobenzaldehyde and an aniline (Synthesis (2004) 6, 949-959; Synthesis (2004) 1, 69-74)



# More clues from the chemistry



The forward direction  
aromatic imine alkene cyclization  
is catalyzed by organic or Lewis  
acids, eg. Tet Letters (2001) 42  
(42) 7405-7407.

## A warning note

- Structures with reverse cyclization (eg. Diels Alder or azadiene chemistry possibilities may be problems
- How many one pot multicomponent reactions touted as great HTS probes are HTS promiscuous because the mode of synthesis might lead to assay problems under certain conditions

# Conclusions

- A medicinal chemist superimposes the association of biomedical information to chemical structure
- Computational people need to understand the “personality” of medicinal chemists
- A rational dialog between medicinal chemists and computational chemists can be very productive