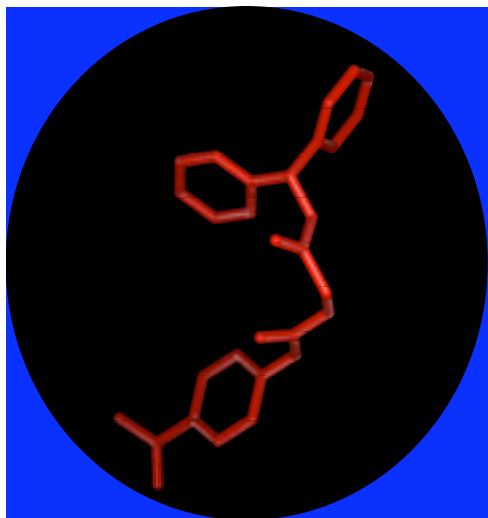


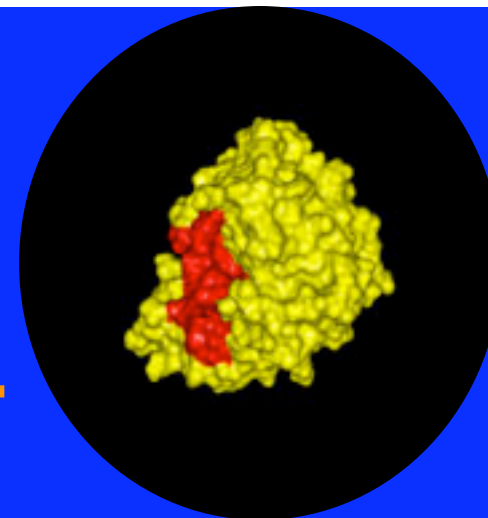
High-Throughput in silico Screening of Compound Collections



Bruno Villoutreix, PhD. Inserm U648
www.vls3d.com



Outline



- **Introduction: Virtual screening**
 - **Compound collections**
- **In silico ADME/tox prediction**
- **In silico screening: ligand-based & structure-based**
 - **Applications**
 - **Conclusions**

Intro: therapeutic targets

- **Human Genome Project**

--> 30,000 genes, with ~10,000 believed to be involved in the pathogenesis of disease,

--> thus around 10,000 potential drug targets

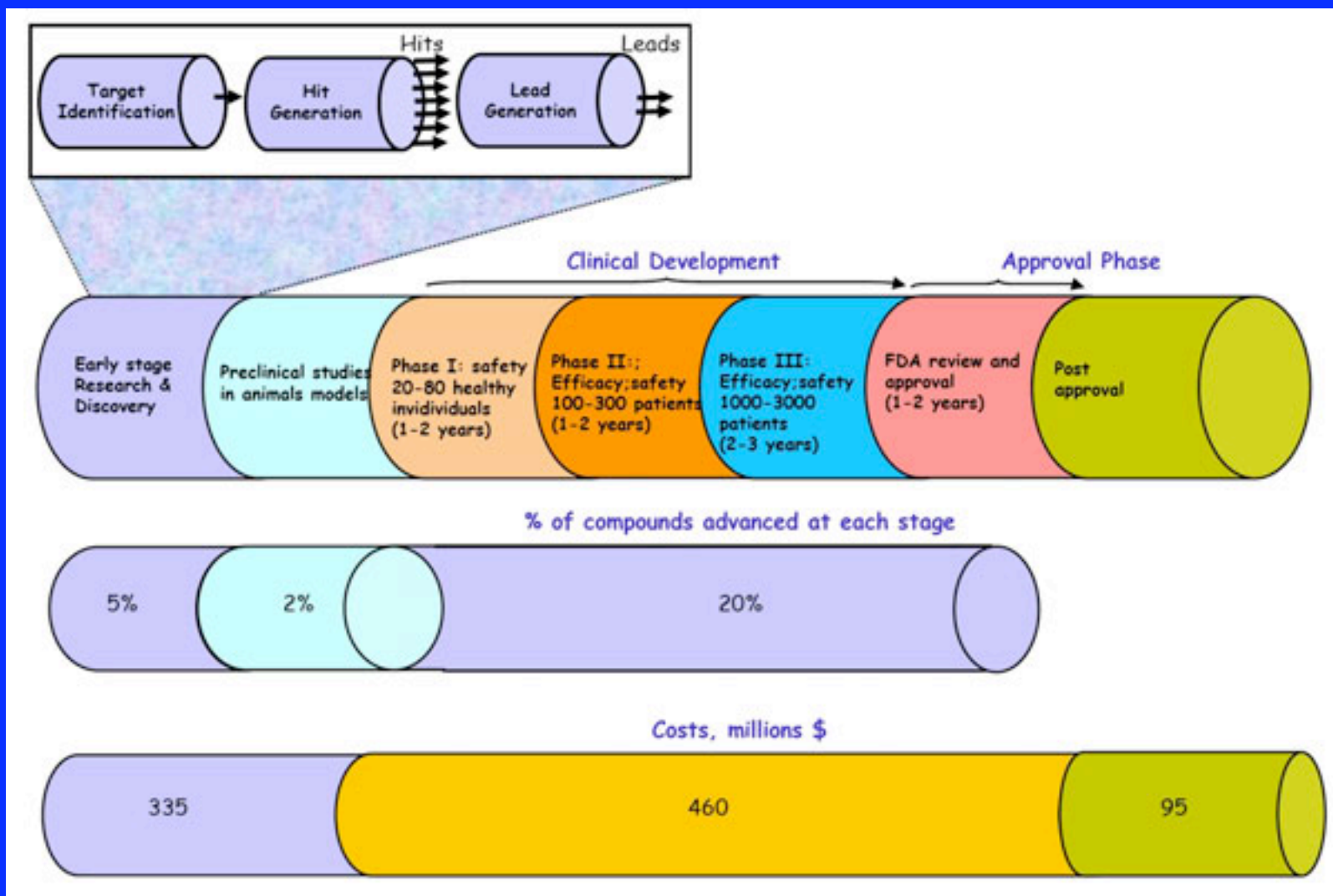
Intro: small molecules

In 2006

**Small molecules easily available from
chemical vendors**

Total = 7 million molecules

Intro: Drug Discovery



Intro: Drug Discovery

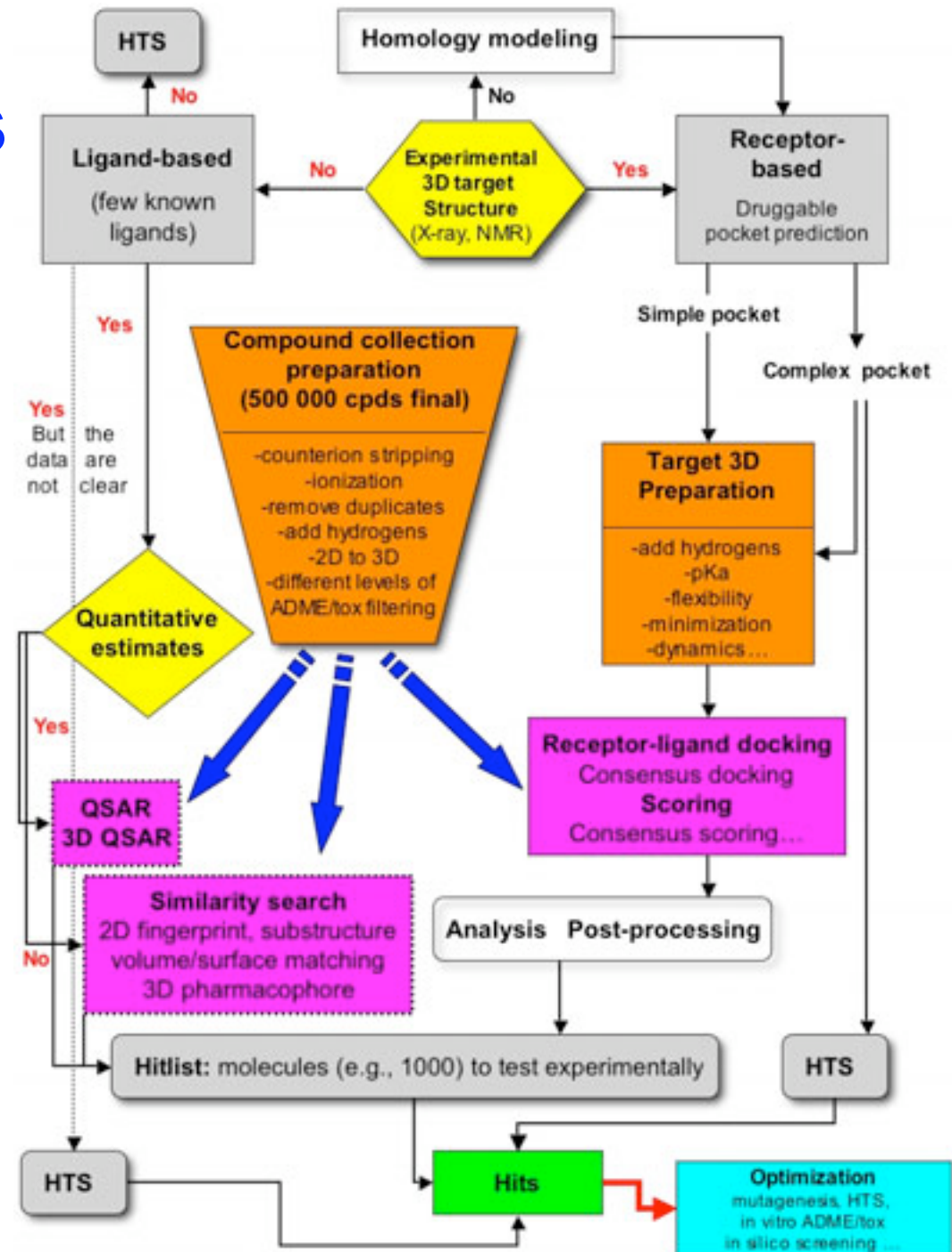
In silico methods ?

They can play many roles including:

- **Help to design compound collections**
- **Search for new hits and facilitate lead optimization**

Intro: the process

in silico
&
in vitro
screening



Compound collections

Compound collections

- At least 36 databases worldwide available to academic groups
- At least 7 million compounds in SMILES or SDF
- Many are listed at:
<http://www.bioscreening.com/>

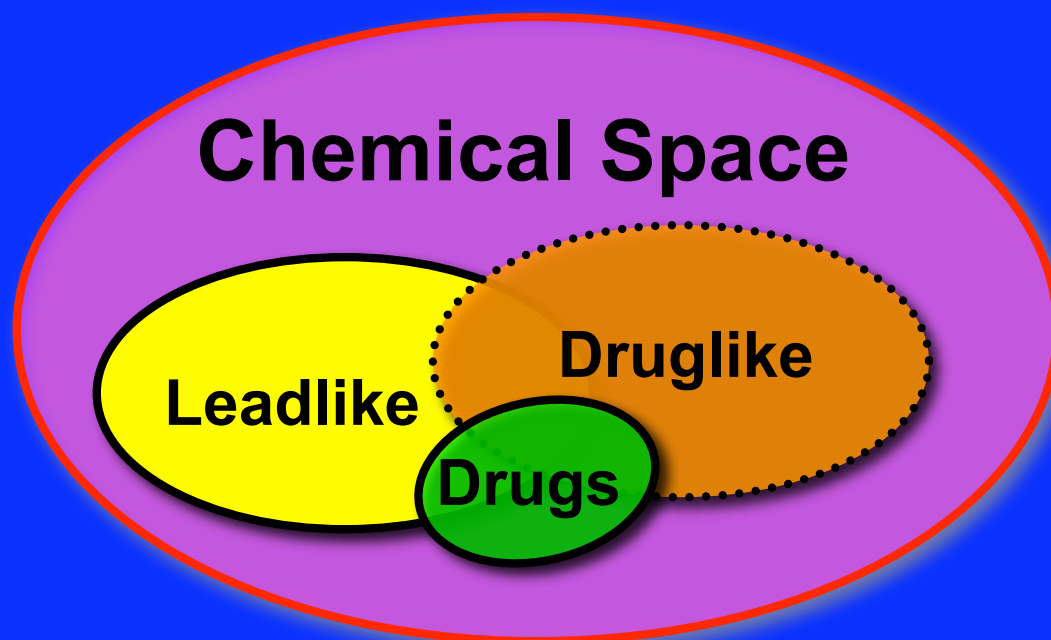
Compound collections

The compound collections available from chemical vendors are :

- Compound libraries for cherry-picking
- Building blocks
- Natural products
- Diversity libraries
- Targeted libraries
- Combinatorial chemistry libraries

Compound collections

How to design a reasonable starting library ?



Compound collections

**It is possible to apply “drug-likeness filters”
also called in silico ADME/tox prediction
(absorption, distribution, metabolism,
excretion and toxicity)**

ADME/tox filtering ?

In silico ADME/tox goal:

Provides a preliminary prediction of the in-vivo behavior of a compound to assess its potential to become a drug

ADME/tox filtering ?

Often, the drug/lead-like space is defined by some descriptors like: logP, MW, H-bond donors, and H-bond acceptors, rotatable bonds.....

Simple counting methods...neural nets

ADME/tox filtering ?

E.g.: simple rules for oral bio-availability

- **Lipinski** rules: HBD, HBA, MW, log P
- **Oprea...** rules: number of rigid bonds, ring bonds, rotatable bonds, number of rings
- **Veber** rules: polar surface area, rotatable bonds
- **Oprea** reactive functional groups: thiols, azides, esters...
 - **Shoichet** frequent hitters
 - **Toxic agents**: acting via tight binding to specific receptors: hERG, aryl hydrocarbon receptor, Cyt P450...

ADME/tox filtering ?

Property	Launched Compounds	Top Selling Compounds
Entries	882	138
ClogP	2.4	2.1
%ClogP > 5	13.5	8.7
%Rule-of-5 Violations	10.4	5.8
Polar Surface Area (Å ²)	136.0	134.0
%PSA > 200 Å ²	18.6	16.7
MW (AMU)	359.7	357.0
%MW > 500	11.2	10.1
No. H-Bond Acceptors	4.3	4.3
No. H-Bond Donors	2.0	2.0
No. Rotatable Bonds	6.9	6.7
%Rotatable Bonds > 10	16.2	14.5

**Launched =
Gained FDA
approval**

**Top selling
drugs in
2003-2004**

***James Blake
Array Biopharma***

Address: http://www.rpb.puc.br/REDS/cgi-bin/Access.cgi?index.php?id=admtox

[Home](#)
[P-Server](#)
[Specific Services](#)
[Software](#)
[Tutorials](#)
[Links to other servers](#)

[Sequences](#)
[Structure](#)
[Protein Structure Modeling](#)
[Protein Interactions](#)
[Small Compounds](#)

version: 0.0.0.0

ADMETox

Original server Map

AMDE-Tox (poor absorption, distribution, metabolism, elimination (ADME) or toxicity) filtering for small compounds. Based on a set of elementary rules.

Specify data format:

Choose a small compound file (smiles format):

or paste a small compound description (smiles format):

Toxic atoms filter: Toxic atoms:

Molecular weight	<input type="button" value="Yes"/>	Min	<input type="text" value="200"/>	Max	<input type="text" value="600"/>
Hydrogen donor number	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="8"/>
Hydrogen acceptors	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="12"/>
Flexible bonds	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="15"/>
Rigid bonds	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="50"/>
Ring Number	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="7"/>
Ring size	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="12"/>
Number of atoms	<input type="button" value="Yes"/>	Min #Carbons	<input type="text" value="0"/>	Min #Hetero	<input type="text" value="0"/>
Carbon/Hetero ratio	<input type="button" value="Yes"/>	Min	<input type="text" value="0.1"/>	Max	<input type="text" value="1.0"/>
Charge number	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="0"/>
Total Charge	<input type="button" value="Yes"/>	Min	<input type="text" value="2"/>	Max	<input type="text" value="0"/>
logP	<input type="button" value="Yes"/>	Min	<input type="text" value="2"/>	Max	<input type="text" value="6"/>
PSA	<input type="button" value="Yes"/>	Min	<input type="text" value="0"/>	Max	<input type="text" value="150"/>

Authors: S. Violas, M. Miteva, M. Montes, D. Gomes, P. Tuffery, B. Viloutreix.

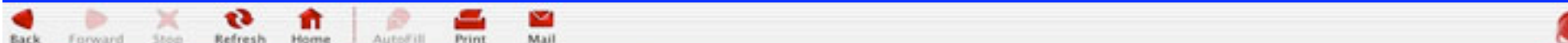
ADME/tox

F-A-F-drugs

(Flexible-ADME/T-Filtering)

Miteva et al., NAR 2006

Compound collections



Address: http://bioserv.rpbs.jussieu.fr/RPBS/cgi-bin/Resource.cgi?chzn_igran&chzn_rsrc=banks

Live Home Page Apple Apple Support Apple Store Mac Mac OS X Microsoft MacTopic Office for Macintosh MSN



Home

P-Server

Specific Services

Software

Tutorials

links to other servers

Sequences

Structure

Protein Structure Modeling

Protein Interactions

Small Compounds

version française

pe-server services map

Access to services :

- ▶ Banks
- ▶ ADME-Tox
- ▶ Pockets
- ▶ Screening
- ▶ Tools

Last Update: December 2005

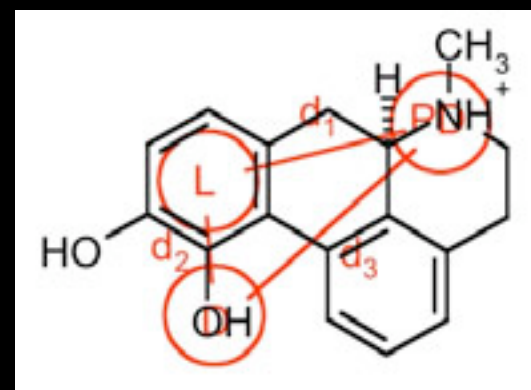
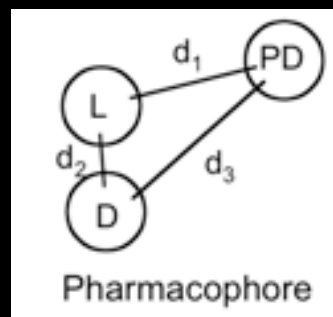
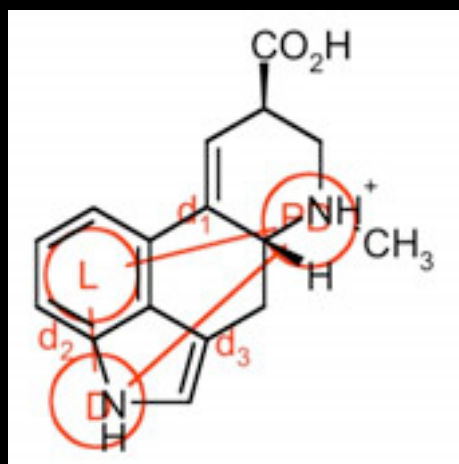
Vendor	Catalogue	Version	#raw	#verylight	#light	#heavy
Ambinter		January 2004	525000	303099 smiles: s mol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Mmol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	200169 smiles: s mol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 Mmol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14	131752 smiles: s mol2: 1 2 3 4 5 6 7 8 9 Mmol2: 1 2 3 4 5 6 7 8 9
	Asinex_Gold	October 2005	234411	150055 smiles: s mol2: 1 2 3 4 5 6 7 8 9 10 11 Mmol2: 1 2 3 4 5 6 7 8 9 10 11	113725 smiles: s mol2: 1 2 3 4 5 6 7 8 Mmol2: 1 2 3 4 5 6 7 8	87094 smiles: s mol2: 1 2 3 4 5 6 Mmol2: 1 2 3 4 5 6
Asinex	Asinex_Plat	October 2005	131097	95169 smiles: s mol2: 1 2 3 4 5 6 7 Mmol2: 1 2 3 4 5 6 7	86426 smiles: s mol2: 1 2 3 4 5 6 Mmol2: 1 2 3 4 5 6	64716 smiles: s mol2: 1 2 3 4 5 Mmol2: 1 2 3 4 5
ChemBridge			429623	304205 smiles: s mol2: 1 2 Mmol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	216211 smiles: s mol2: 1 2 Mmol2: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	154351 smiles: s mol2: 1 Mmol2: 1 2 3 4 5 6 7 8 9 10 11

In silico screening

- **Ligand-based**
- **Structure-based**

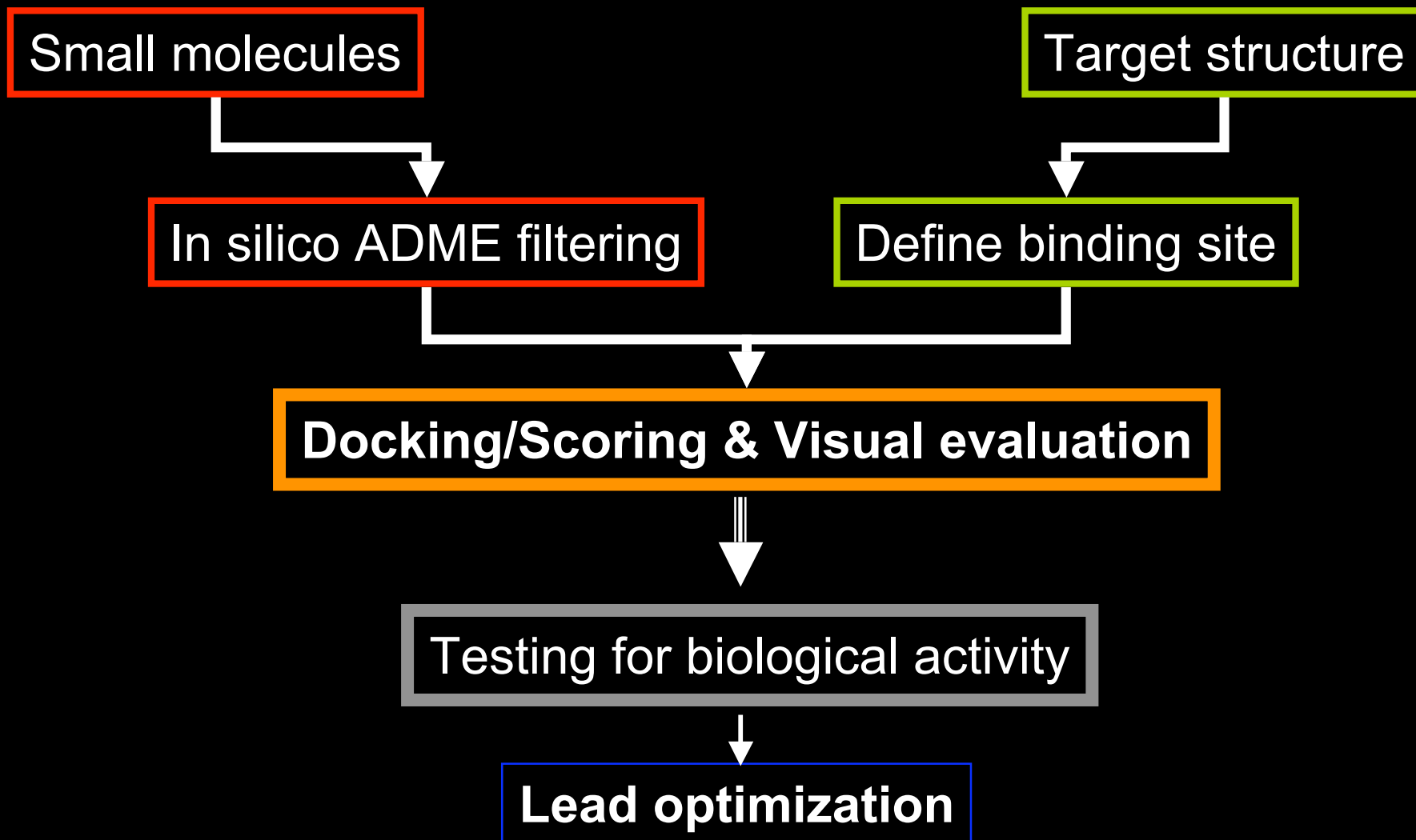
Ligand-based screening

- ✓ Given a bioactive conformer, it is possible to run 2D/3D similarity search or pharmacophore matching



Ex: L = site lipophile; D = Donneur H; PD = Charge +

Structure-based screening



Docking

- **“Rigid”**: main concepts

- Ligands: multi-conf. / receptor: rigid
- Ligands: multi-conf. / receptor: multiconf.

- **“Flexible”**: main concepts

- Ligands: systematic (fragmentation) search / receptor: rigid
- Ligands: stochastic search (MC, GA) search / receptor: rigid
- Ligands: deterministic (MM, MD) search / receptor: rigid

- Ligands: flexible / receptor: flexible



Side-chains only

All atoms, usually
pregenerated multiconf.
or full MD during docking...

Sperandio et al., Curr Prot Pep Science 7:369-393 (2006)
Villoutreix et al., Curr Prot Pep Science 2007

Scoring

A. Force Field-Based Scores

B. Empirical Scoring Functions

C. Knowledge-Based (*contact preferences*)

D. Consensus Scoring ?

Often, scoring functions:

- **Describe only enthalpic contributions (ΔH)**
- **No estimate of ΔG unless some terms are present**
- **Non-bonded interactions are evaluated**

Applications of structure-based screening

Structure-based screening

Novel ligands recently identified through structure-based VS.

Target [reference]	Target class	Target structure	Approximate database size	Method(s) used	Lead potency (μM)	Supporting data
AmpC β -lactamase [16]	Hydrolase	X-ray	200 k	NWU DOCK	26	X-ray complex
BCR-ABL [49]	Kinase	X-ray	200 k	DOCK	25	Cell-based inhibition
Anthrax EF [17*]	Adenylyl cyclase	X-ray	200 k	NWU DOCK	20	Enzyme kinetics
IMPDH [23]	Dehydrogenase	X-ray	3500 k	FlexX	30	Enzyme kinetics
Casein kinase II [13]	Kinase	Homology	400 k	DOCK	0.08	Inhibition, selectivity, SAR
K ⁺ Channel [50]	Ion channel	Homology	50 k	DOCK	10	Cell-based inhibition
Thyroid hormone receptor [51]	Nuclear receptor	Homology	250 k	ICM	0.75	Inhibition
CDK2 [15]	Kinase	X-ray	50 k	LIDAEUS	2	X-ray complex
TGF β RK [30]	Kinase	X-ray	200 k	Catalyst	0.005	X-ray complex
Cyclophilin [28]	Immunophilin	X-ray		Unity/FlexX	6	Cell-based inhibition
tRNA guanine transglycosylase [26]		X-ray	800 k	Unity/FlexX	0.25	Enzyme kinetics
PDHFR [29]	Reductase	Homology	230 k	Catalyst/DOCK	0.9	Enzyme kinetics
α -Amylase [27]	Hydrolase	X-ray	200 k	Unity/FlexX	None	NMR, SPR, affinity Chromatography

High-throughput docking as a source of novel drug leads

Juan C Alvarez

Current Opinion in Chemical Biology 2004, 8:365–370

Structure-based screening

**Another example
G protein-coupled receptor**

Homology modeling + virtual screening

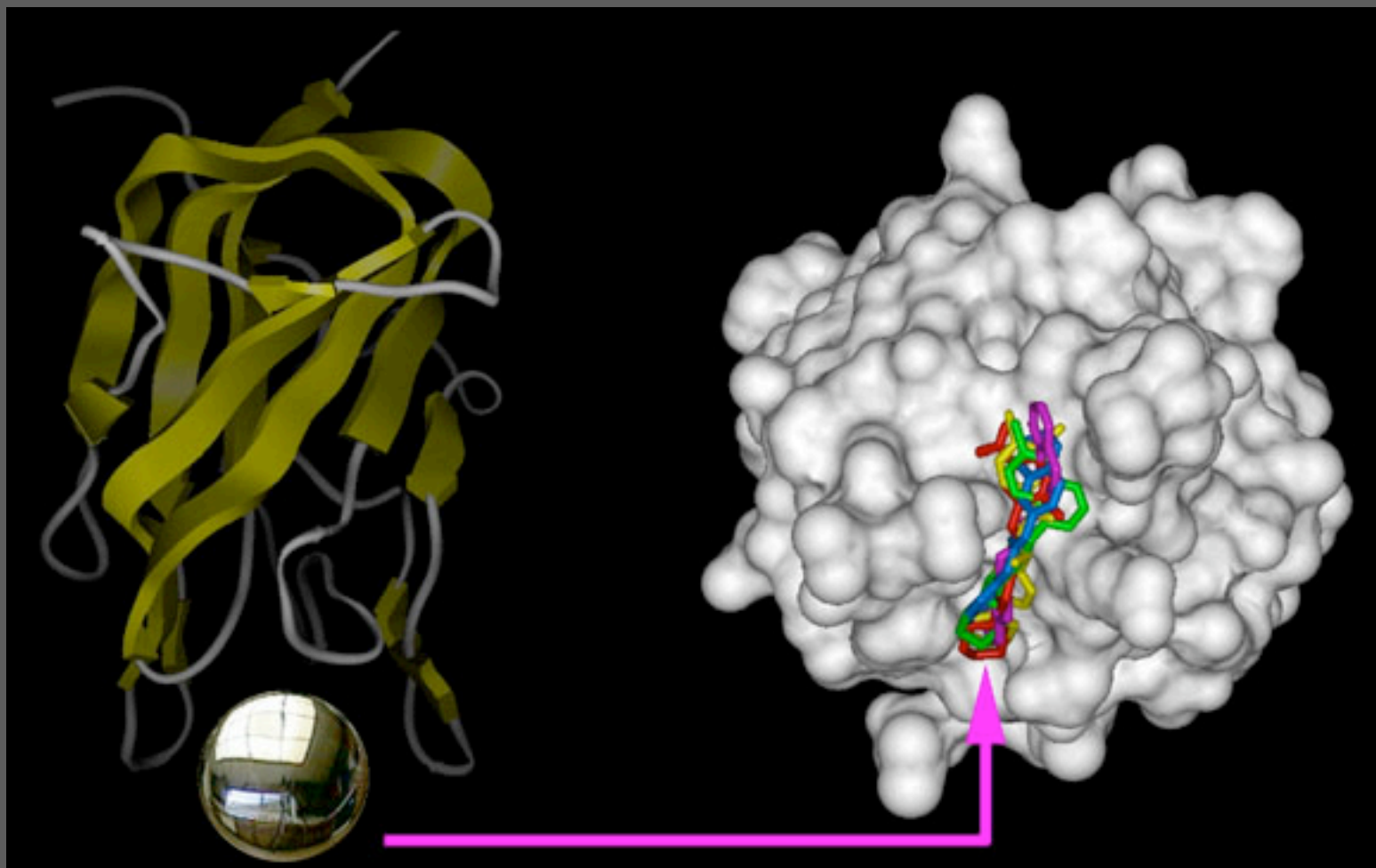
Molecules successful so far

Clinical trial: phase I to phase III

Becker OM'group USA

Structure-based screening in our group

Inhibiting protein-membrane interaction via structure-based VLS



Structure-based screening in our group

protein-membrane interaction:

400000 --->ADME/T ---> 300000 --->

docking/scoring ---> 500 selected and tested

---> over 10 hits

Hit rate = (10/500)x100 = 2%
About 100 higher than via HTS

Conclusions

Challenges generating/maintaining compound collections & deciding about ADME/tox filters

Problems dealing with flexibility

Problems with docking/scoring accuracy

Problems with selectivity



Possible to use homology models for SB-VLS

Possible to predict druggable pockets

Prioritize compounds, targets and pockets

Relatively fast

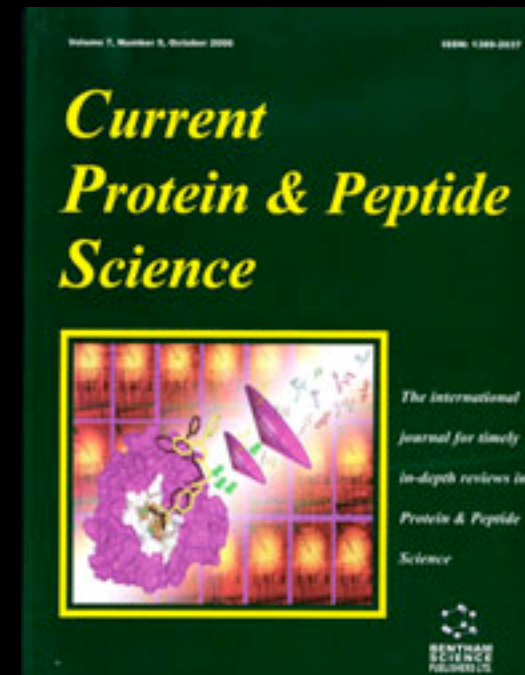
Propose interesting candidates, number of success stories is growing rapidly



Acknowledgements

- M. Montes
- N. Renault
- O. Sperandio
- D. Lagorce
- Dr. M. Miteva

- Dr. G. Nicolaes (Maastricht)
- K. Segers (Maastricht)
- Dr. P. Tuffery, Inserm U726



www.vls3D.com