

In silico Screening in drug discovery

Matthieu Montes

PhD Student

University Paris 5

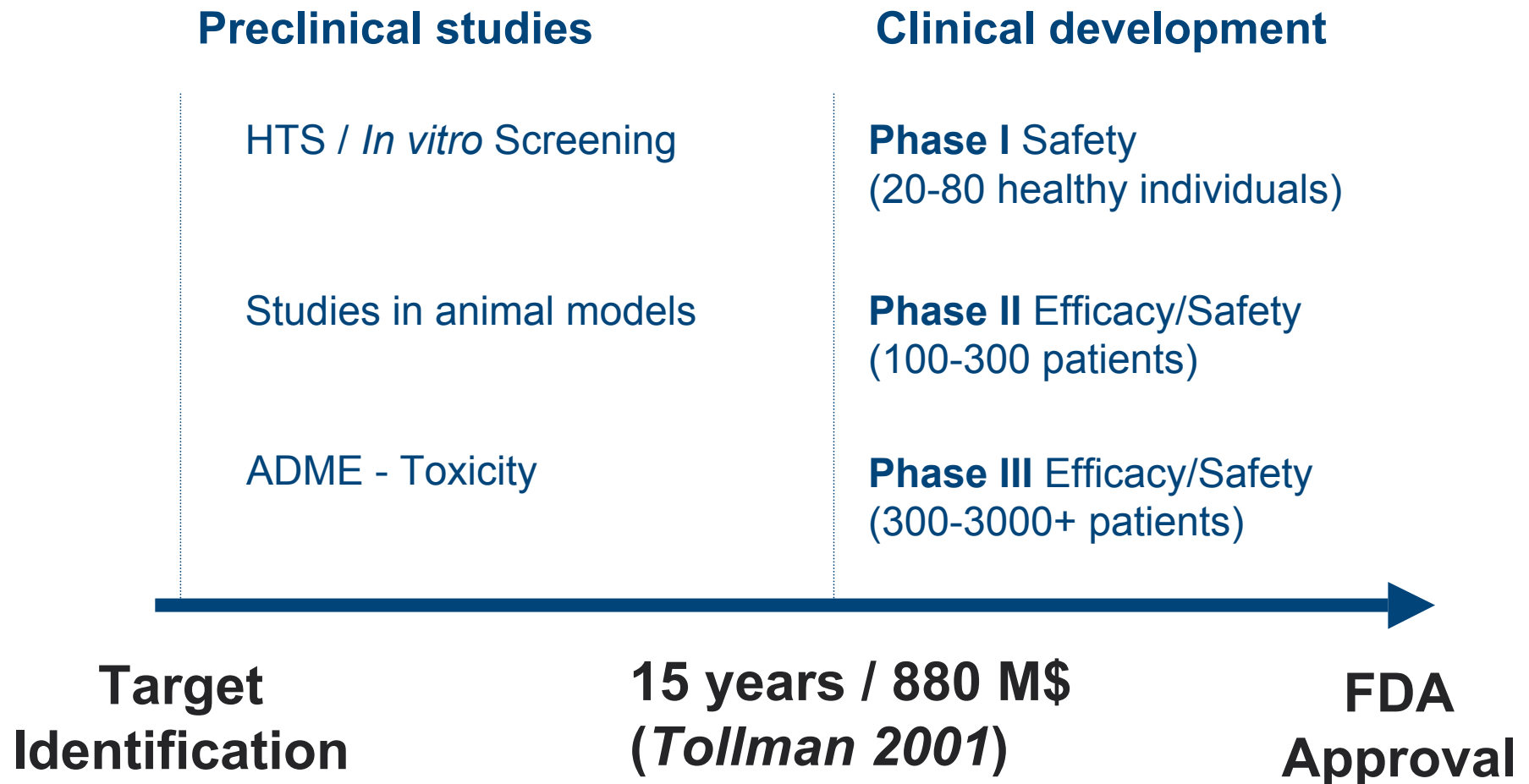
Inserm U648



Intro: why *in silico* methods ?

- Huge amount of data to be processed:
 - Growing number of identified proteins of therapeutical interest
 - Growing number of protein 3D structures
 - Around 7 million chemicals easily available in 2006
- Huge cost of Experimental High Throughput screening
- Growing computational resources

Intro: drug discovery



Intro: drug discovery

- In the early stage of research:

Compound collection



Hit generation *HTS*

Hits



Lead generation

Leads

Intro: drug discovery

Compound collection



In silico screening

Compound collection



Hit generation *HTS*

Hits



Lead generation

Leads

In silico screening

- Use of physico-chemical filters in order to reduce the number of compounds to be tested experimentally for hit/lead generation



As the prospector and his sieves :)

In silico screening

- *In silico* ADME-tox prediction filters can be applied to keep in the compound collection only drug-like/lead-like compounds

Nb: ADME-Tox = Absorption, distribution, metabolism, excretion and toxicity

=> Preliminary prediction of the *in vivo* behaviour of a compound to assess its potential to become a drug

ADME-Tox filtering

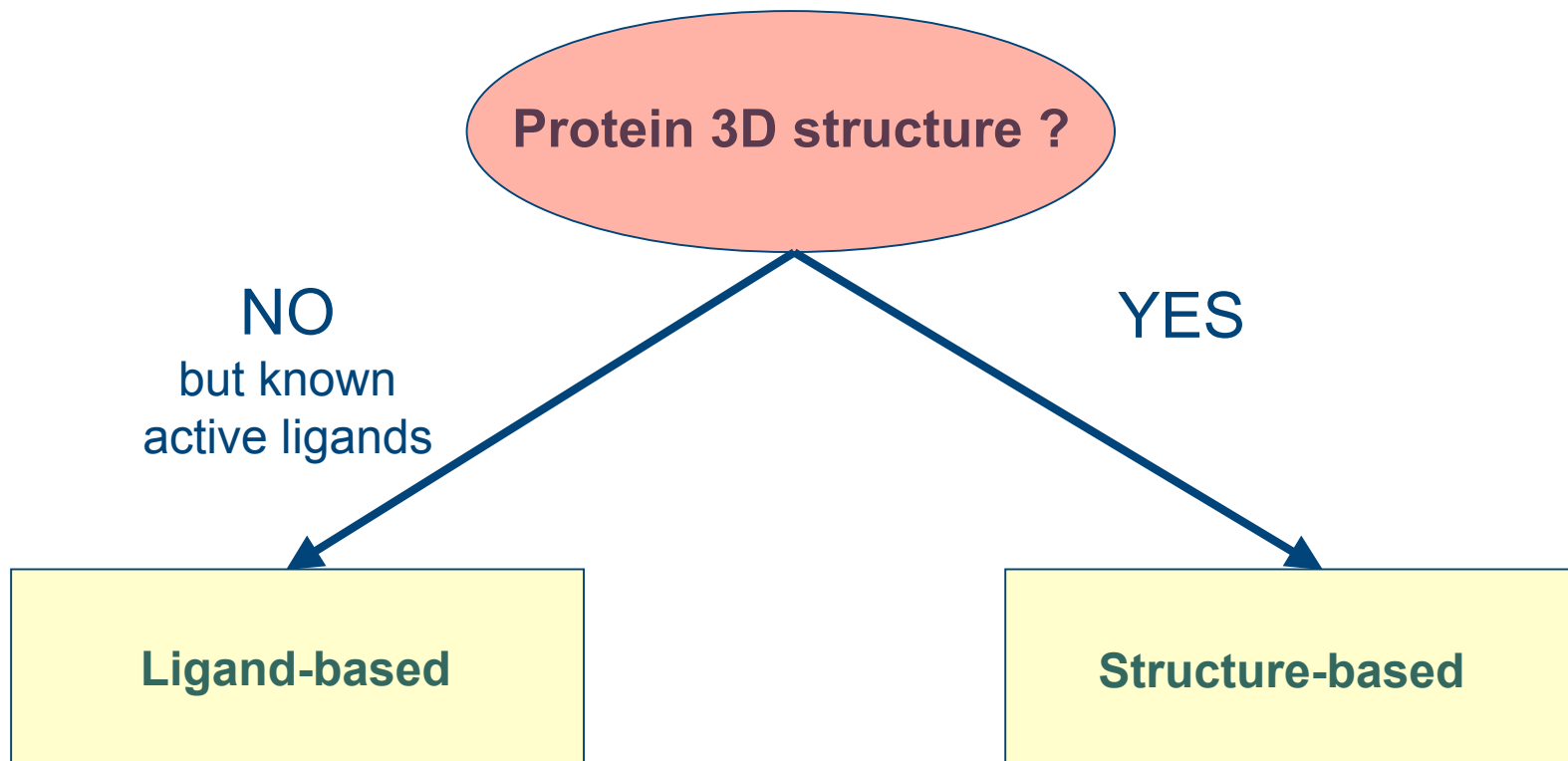
- Drug-like/ lead-like properties can be predicted using simple descriptors like #Hbond donors/acceptors, #rotatable bonds, logP, MW, PSA...(Lipinski, Veber)
- reactive functional groups (Oprea)
- frequent hitters (Shoichet)
- toxic groups, metals

ADME-Tox filtering

FAF-drugs,
Free ADME-tox
Filtering of compound
collections
Miteva et al, NAR, 2006

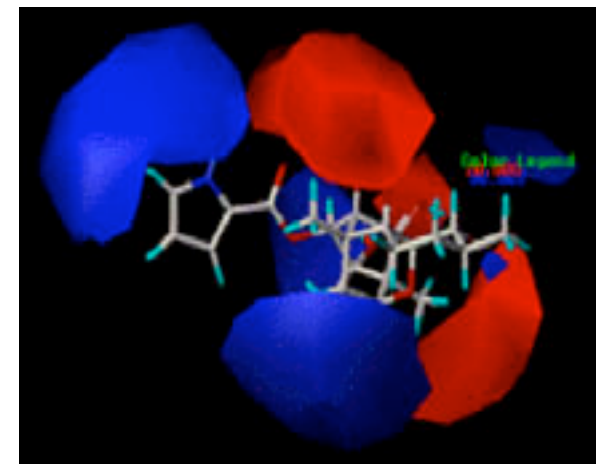
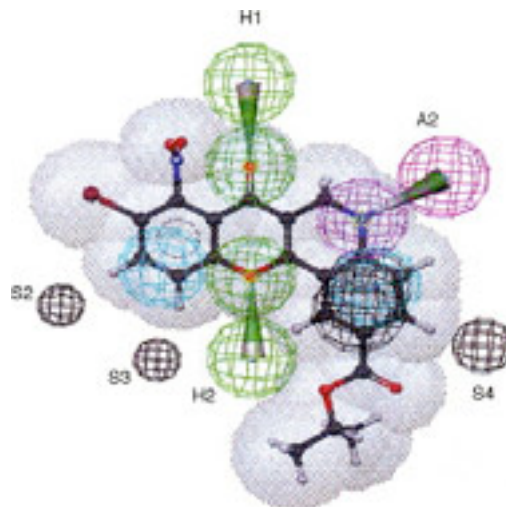
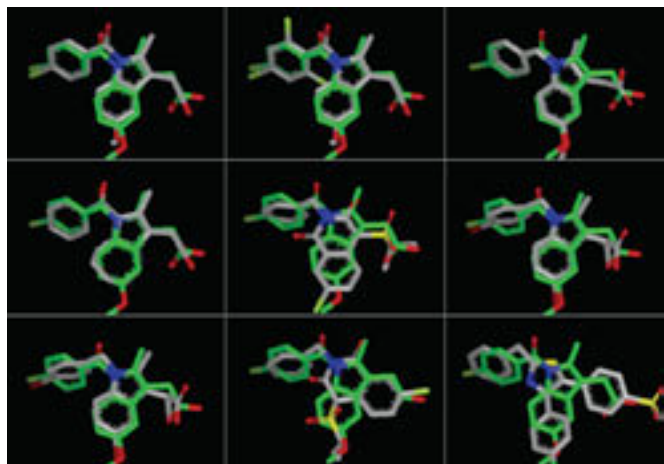
<http://bioserv.rpbs.jussieu.fr>

In silico screening



Ligand-based virtual screening

- Compound selection on the basis of known bioactive molecules
- Different 2D/3D methods can be performed:
 - Structure / Shape similarity search
 - Pharmacophore matching
 - QSAR



Structure-based virtual screening

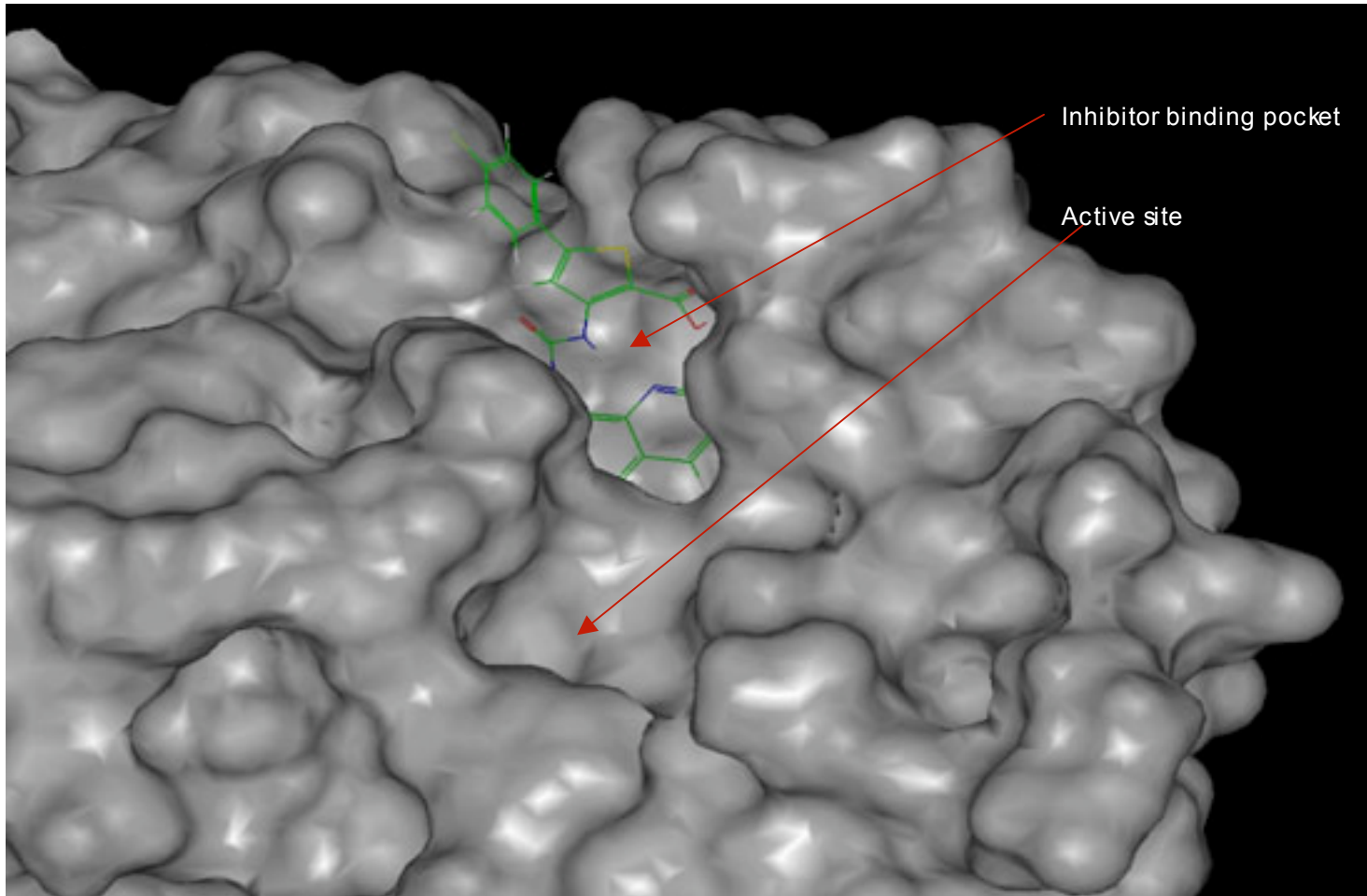
- Compound selection regarding their predicted affinity with the target 3D structure
- **Docking:** modeling of the ligand positioning into the binding pocket
- **Scoring/ranking:** ligand pose evaluation, ranking of the ligands regarding their predicted affinities for the target

Structure-based virtual screening

- Receptor 3D structure:
- Identification and definition of the binding site
- Docking method
- Scoring function

Binding site identification

- Active site or exosite ?



Binding site definition

- Do we have experimental data ?
 - X-ray crystallography, NMR, site-directed mutagenesis
- Binding site definition using known residues or bound ligand
- If not, binding sites can be predicted theoretically

Binding site identification

- In general: cavities at the surface of the macromolecule with special physicochemical properties compatible with ligand binding
 - Volume, apolar surface, roughness...
- Binding site prediction algorithms
 - Geometric and/or energetic criteria

Docking

- Rigid body docking:
 - Ligands multi-conf / receptor rigid
 - Ligands multi-conf / receptor multi-conf
- Flexible docking
 - Ligands flexible / receptor rigid
 - Ligands flexible / receptor flexible (side chains only or all atoms)

Rigid body docking

- Fast docking procedure to remove compounds that simply cannot fit into the binding pocket
- Ligand multi-conformer generation
 - Multi-conformer compound collection (OMEGA, Corina...)
 - Multi-conformer generation « on the fly » (DOCK)

<http://www2.chemie.uni-erlangen.de/software/corina/index.html>
<http://www.eyesopen.com/products/applications/omega.html>

Flexible docking

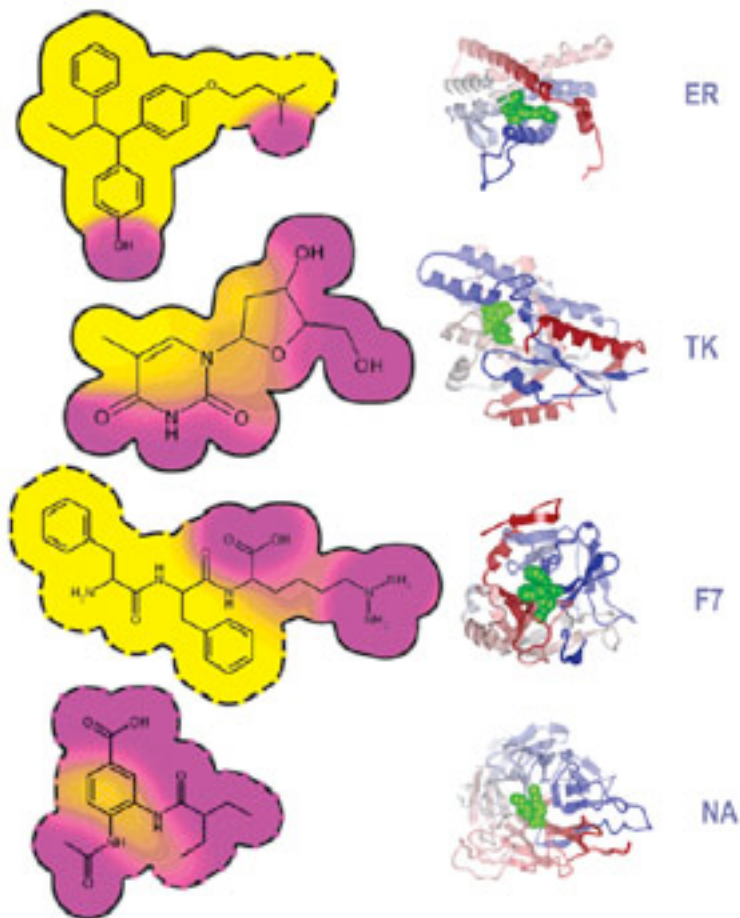
- Several different algorithms to explore ligands degrees of freedom:
 - Systematic search (fragmentation/reconstruction: DOCK, FlexX, Surflex...)
 - Stochastic search (GA/MC: GOLD, ICM, LigandFit)
 - Deterministic (MM, MD)
- In general, receptor remains rigid

Scoring functions

- Approximate models to determine relative binding affinity
- Different categories:
 - Force field based (DOCK, GOLD, ICM)
 - Empirical (PLP, FlexX, Surflex)
 - Knowledge based (PMF, DrugScore)
 - Mixed (LigScore)
- Consensus scoring ?

Virtual screening performance

- Virtual screening protocol developed in our lab using



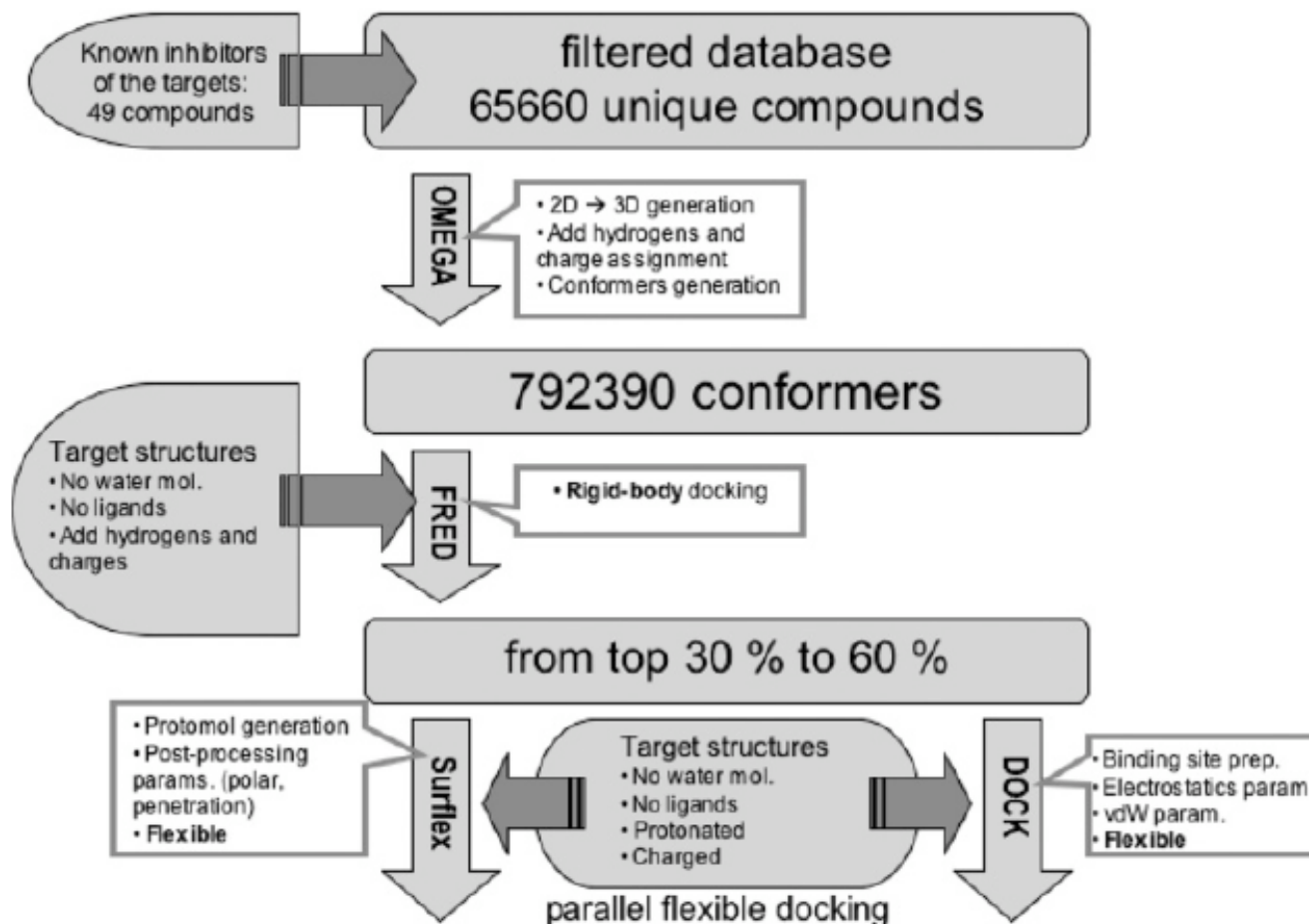
4 different targets with different binding site properties

65560 drug-like compound collection including known actives

Docking and scoring softwares freely available to academics

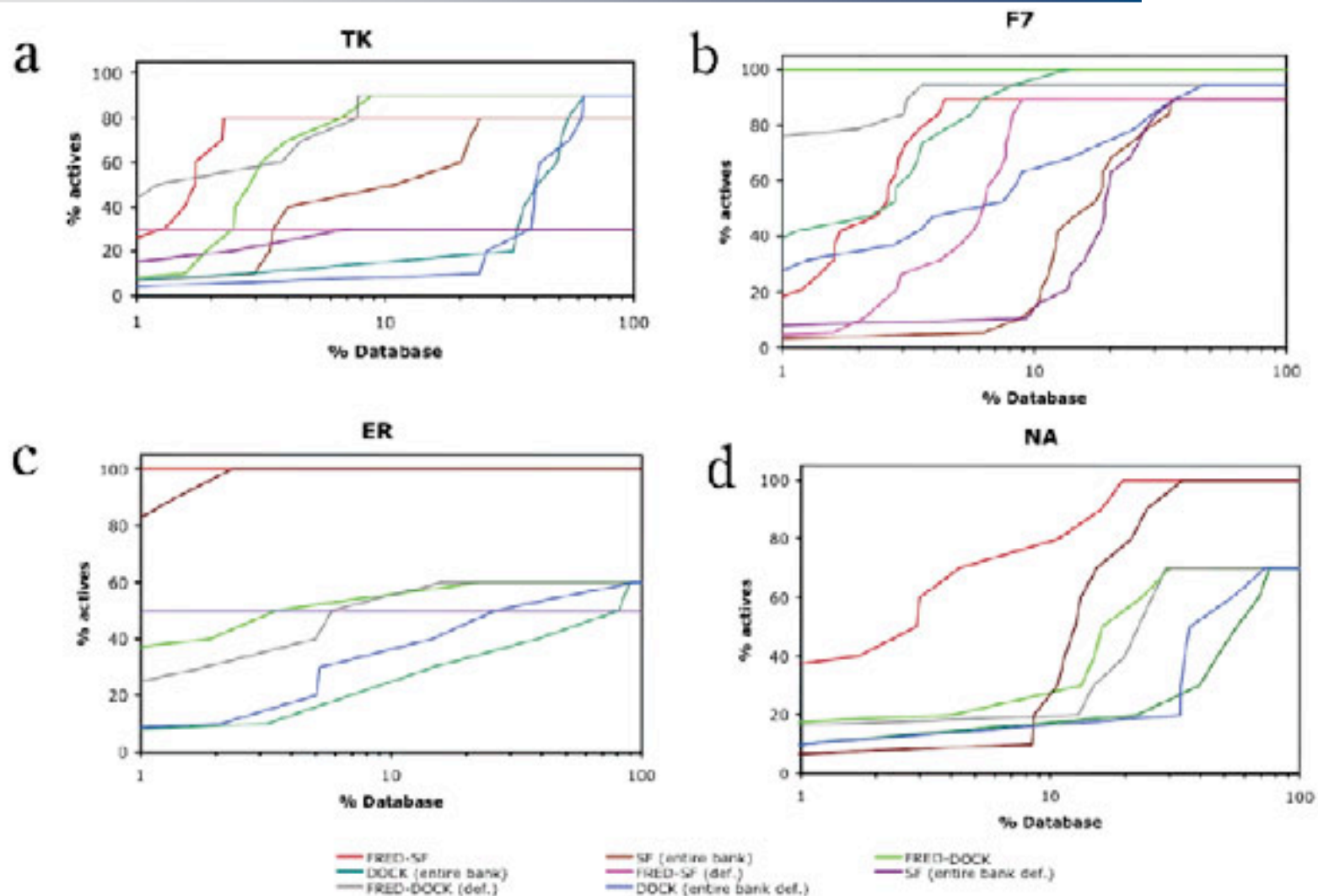
Virtual screening performance

- Virtual screening protocol developed in our lab



Miteva MA*, Lee WH*, Montes M, Villoutreix BO, *J Med Chem*, 2005

Virtual screening performance



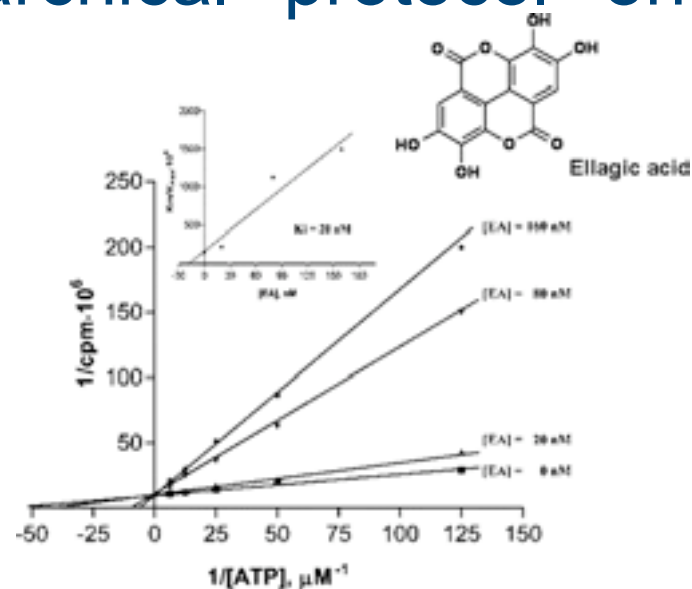
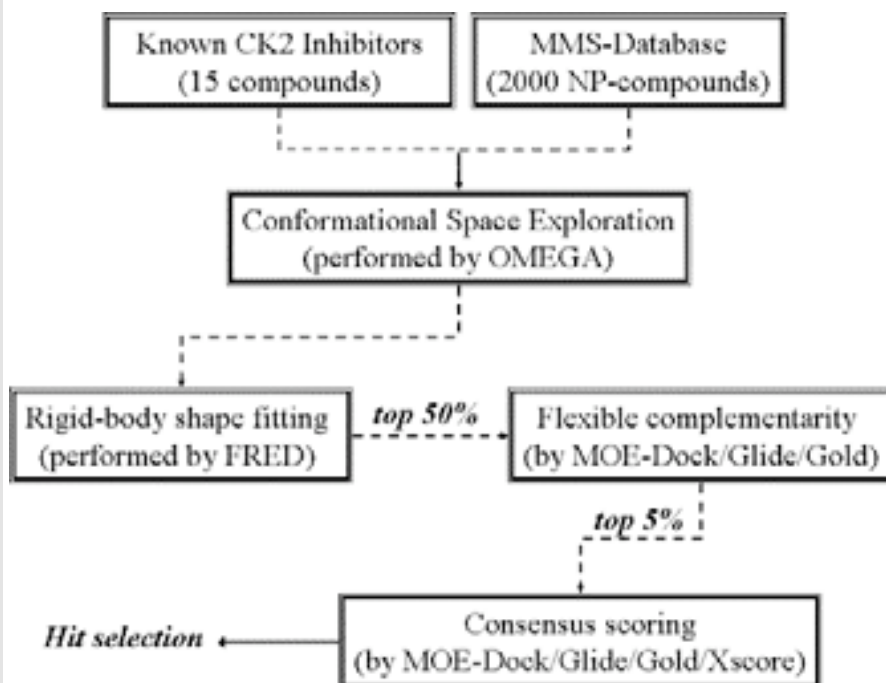
Miteva MA*, Lee WH*, Montes M, Villoutreix BO, *J Med Chem*, 2005

Virtual screening performance

- Docking and scoring parameters need to be tuned according to the binding pocket properties
- Better performance using a hierarchical protocol (chaining tools) than using a single docking/scoring software
- Virtual screening of a large compound collection on a drug binding pocket is possible even with limited computer resources

Applications of Structure-based VLS

- Application of our hierarchical protocol on Casein Kinase 2



Discovered Ellagic acid as a new potent and selective inhibitor of casein Kinase 2
=> $K_i = 20\text{nM}$

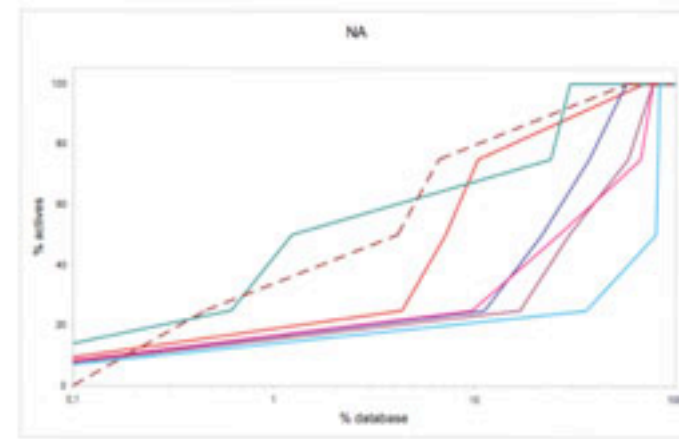
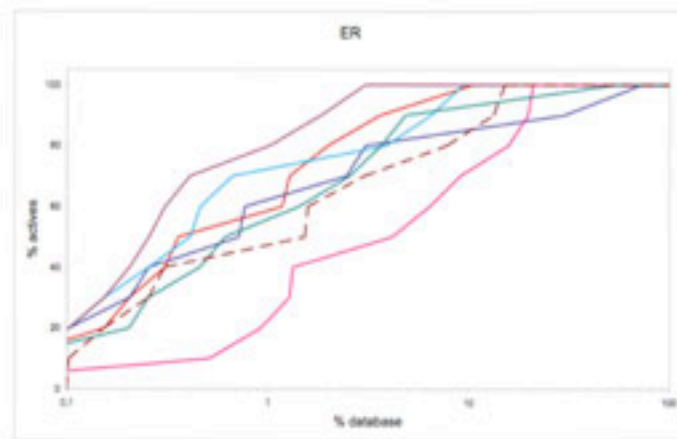
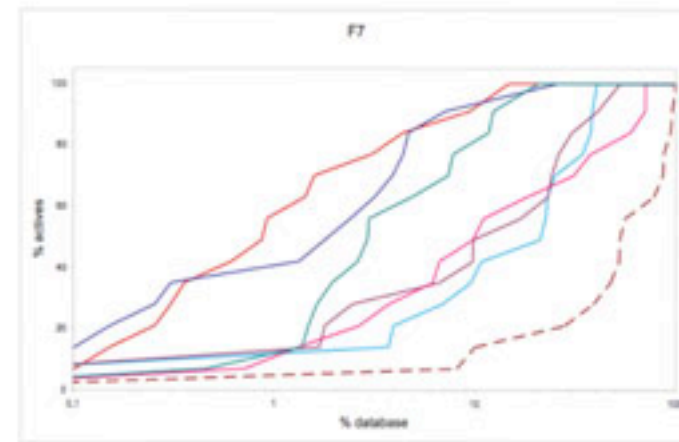
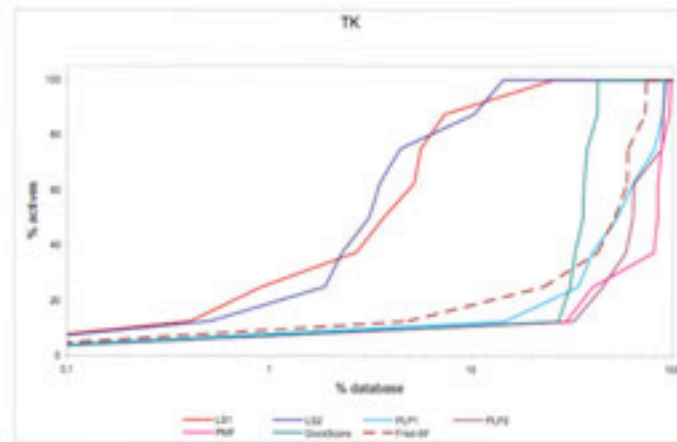
Cozza et al, *J Med Chem*, 2006

Virtual screening optimization

- Evaluation of LigandFit on ER, F7, TK and NA
- Docking procedure: Monte Carlo, 5 scoring functions
- Focused libraries* on ER, F7, TK and NA

**top 2000 compounds after our FRED/Surflex hierarchical protocol*

Virtual screening optimization



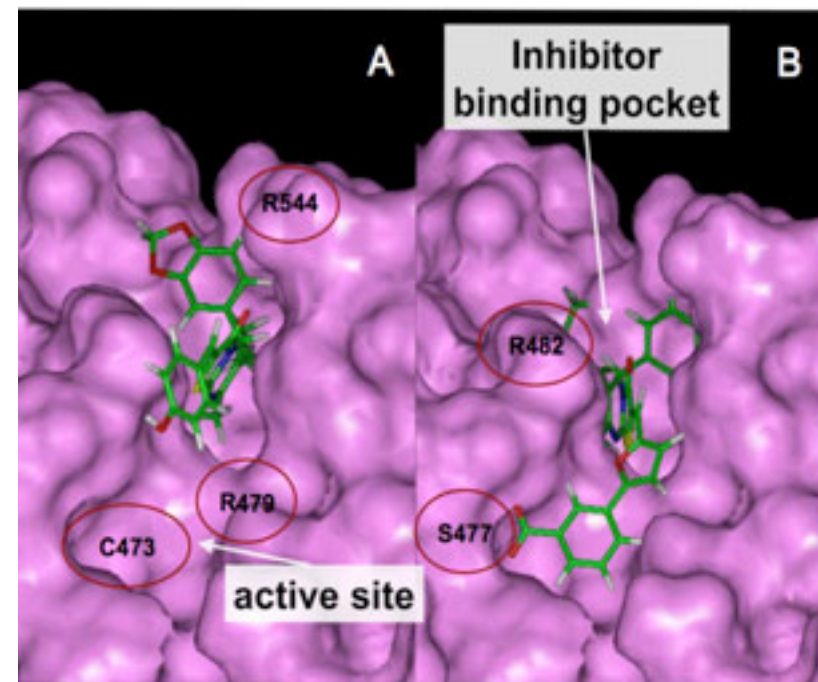
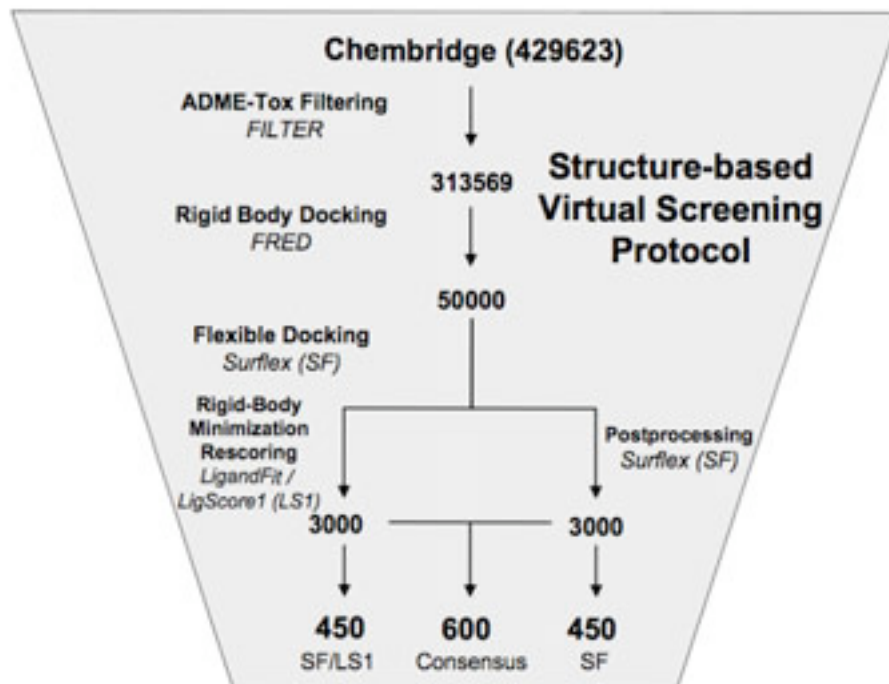
Montes M, Miteva MA, Villoutreix BO, *minor revision in Proteins* (2006)

Virtual screening optimization

- Good performance of the LigScore scoring function
- Improved enrichment on 3/4 proteins tested
- Improvement of our hierarchical protocol using LigandFit as a final step

Applications of Structure-based VLS

Target: Key regulator of the cell cycle; Application: cancer



Applications of Structure-based VLS

Target: Key regulator of the cell cycle; Application: cancer

% inhibition at 100 μ M	21-40	41-50	51-60	61-90	91-100
Number of compounds	76	13	7	1	2

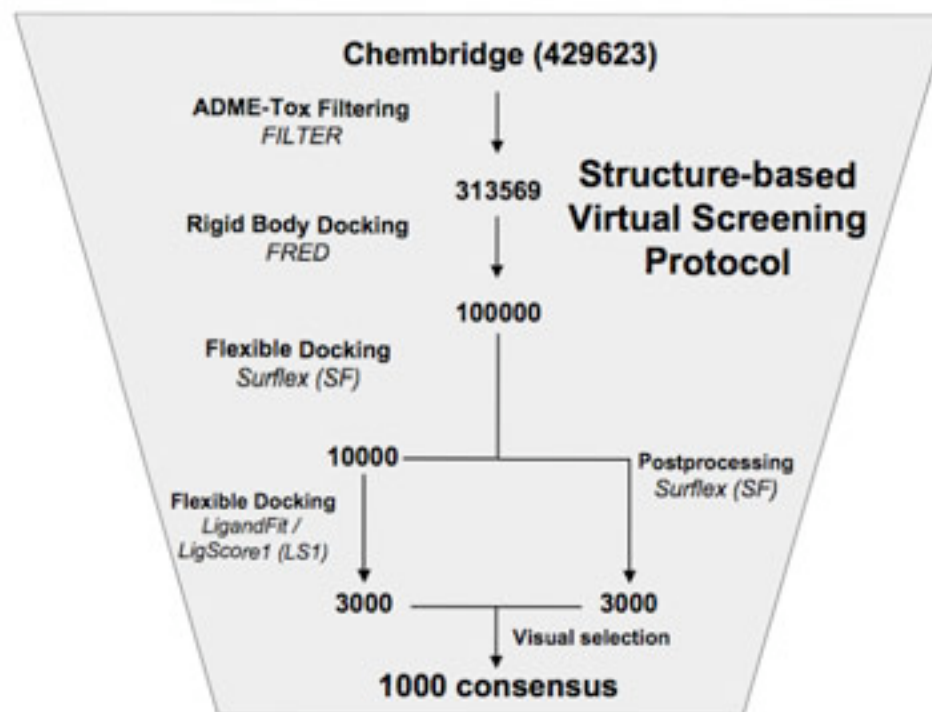
=> 5 compounds with IC_{50} from 10 to 80 μ M, several inhibit clonal proliferation (non-quinone derivatives with same activity as the best known inhibitors, less reactivity)

=> New chemical scaffolds, possible optimization via chemistry / ligand based approaches

Applications of Structure-based VLS

Target:
Key regulator of protein
stability

Application: cancer



⇒ Tests in progress with our experimentalists collaborators: several specific non-covalent compounds already identified, IC_{50} around the μM range

⇒ New chemical scaffolds, possible optimization via chemistry / ligand based approaches

Applications of Structure-based VLS

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
PfDHFR [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

Alvarez, *Curr Opin Chem Biol*, 2004

Conclusion



- Possible to use homology models
- Possible to predict druggable pockets
- Prioritize compounds, targets, pockets
- Propose interesting candidates, growing number of success stories
- Fast and possible for academic groups with limited computer resource

- Deciding about ADME-Tox filters
- Problems with receptor flexibility
- Problems with docking accuracy
- Problems with ligands scoring / ranking
- Problems with selectivity



Acknowledgements

- Structural Bioinformatics group, Inserm U648

- Dr Bruno Villoutreix

- Dr Maria Miteva

- Olivier Sperandio

- David Lagorce



- All the experimentalists in our group and other universities

- Dr Gerry Nicolaes

Thank you for your attention



matthieu.montes@univ-paris5.fr

www.VLS3D.com



Any questions ?