CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Docking and Post-Docking strategies

Didier Rognan

Bioinformatics of the Drug
National Center for Scientific Research (CNRS)

didier.rognan@pharma.u-strasbg.fr

A few starting points

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

- J Comput-Aided Mol Des. Special Issue on 'Evaluation of Computational Methods', 2008, 131-265

- Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. (2008) *Br J Pharmacol* , **153**: S7-26.

- Sousa SF, Fernandes PA, Ramos MJ. (2006) *Proteins*, **65**:15-26.

- Kitchen DB, Decornez H, Furr JR, Bajorath J. (2004) *Nat Rev Drug Discov,* **3**: 935-49.

- Brooijmans N, Kuntz ID. (2003) *Annu Rev Biophys Biomol Struct* , **32**: 335-73.

- Lengauer T. (2007) Bioinformatics - From Genomes to Therapies Volume 2 : Molecular Interactions (WILEY-VCH)

- Kubinyi H. (2006) in *Computer Applications in Pharmaceutical Research and Development,* S. Ekins, Ed.; Wiley-Interscience, New York, pp. 377-424.

Strategic importance of docking

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Scientific reasons

1. Increasing number of interesting macromolecular targets (500 → 6,000)
2. Increasing number of protein 3-D structures (X-ray, NMR)
3. Better knowledge of protein-ligand interactions
4. Development of chem- and bio-informatic methods
5. Increasing computing facilities

Economic reasons

1. High cost of high-througput screening (HTS): 0.2-1 € /molecule

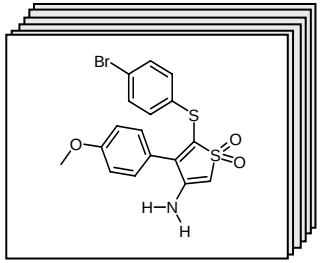
2. Increase the ratio $\frac{\#\ \text{of active molecules (hits)}}{\#\ \text{of tested molecules}}$

Applications

1. Identifying/optimize ligands for a given target
2. Identifying target(s) for a given ligand

Docking Flowchart

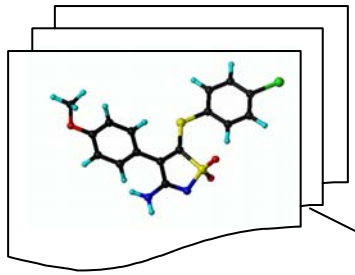
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



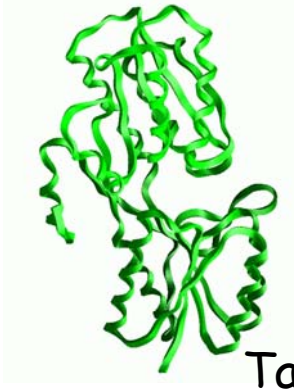
2-D Library

Filtering

Preparation



3-D Library



Target

Docking

Scoring

Mol #	ΔG_{bind}
11121	-44.51
222	-42.21
3563	-41.50
6578	-40.31
25639	-40.28
.	
.	
.	
.	
.	
100000	22.54

Post-processing

Hitlist

Which Compound Library ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Commercially-available screening collections are important sources for identifying hits by virtual screening (VS)

Supplier	Size	Website
ACD Stock	87657	www.acdlabschem.com
Akos Lithuania	156140	www.akosgmbh.com
Akos Resynthetic	134465	www.akosgmbh.com
Akos OWH	28600	www.akosgmbh.com
Asinex Platinum	129860	www.asinex.com
Asinex Preformated	14240	www.asinex.com
Asinex Gold	194032	www.asinex.com
Bionet (Ryan)	35080	www.ryansci.com
ChemStar	85981	www.chemstaronline.com
ChemStar Plated	40000	www.chemstaronline.com
Chembridge DVS	30000	www.chembridge.com
Chembridge Express	321716	www.chembridge.com
Chemical Diversity Clab	175551	www.chemdiv.com
Chemical Diversity IDC	110484	www.chemdiv.com
Comgenex	119247	www.comgenex.com
Enamine RS	109695	www.enamine.net
Enamine S	124420	www.enamine.net
Evotec Lead Discovery	76817	www.evotecaoi.com
Evotec Screening	122998	www.evotecaoi.com
G&J (Ryan)	1867	www.ryansci.com
IBS n	195899	www.ibscreen.com
IBS s	172153	www.ibscreen.com
IF Lab ltd	109400	iflab@iflab.kiev.ua
Key Organics	34736	www.keyorganics.ltd.uk
Maybridge (Ryan)	75142	www.maybridge.com
MDPI	9367	www.mdpi.net
Menai (Ryan)	4512	www.ryansci.com
Nanosyn Explore	18613	www.nanosyn.com
Nanosyn Pharma	44000	www.nanosyn.com
NCI DTP	126705	dtp.nci.nih.gov/docs/3d%5Fdatabase/dis3d.html
Otava Stock	104258	http://www.otava.com.ua
Peakdale	3770	www.peakdale.com
Pharmeks	84143	www.pharmeks.com
Princeton BioMolecular Research Gold	204215	www.princetonbio.com
Ryan Intermediate	1136	www.ryansci.com
Ryan Labo	59214	www.ryansci.com
Sigma-Aldrich L	13929	www.sigma-aldrich.com
Sigma-Aldrich R	69238	www.sigma-aldrich.com
Sigma-Aldrich S	46284	www.sigma-aldrich.com
Sigma-Aldrich	58298	www.sigma-aldrich.com
Specs Biospecs 10T	172469	www.specsnet.com
Timtec Stock	100276	www.timtec.com
Timtec MaxiVerse	9600	www.timtec.com
Toslab collection + ugi	11743	www.toslab.com
WorldMolecules	20401	www.worldmolecules.com

> 3,000,000 million compounds available

Sirois et al. *Comput. Biol. Chem.* (2005) **29**, 55-67.

Commercially-available screening collections

- They cannot be directly used as such for VS because of some issues:
  - ✓ redundancy (intra and inter-duplicates)
  - ✓ diversity
  - ✓ unknown drug- or lead-likeness
  - ✓ unsuitable format (non-ionized, counter-ions, racemates)
- 'Unified' and filtered screening collections are available

ChemNavigator

http://www.chemnavigator.com

i-Research chemical Library

21 million samples
not ready to screen
not 'clean'
not free

ELSEVIER MDL

http://www.mdli.com

MDL screening Compounds Directory

3.5 million structures
not ready to screen
± clean
not free

A free database for virtual screening
ZINC
is not commercial

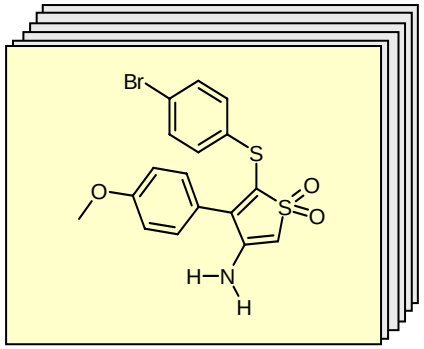
http://blaster.docking.org/zinc/

Zinc

3.3 million structures
ready to screen
relatively clean
free

Library set-up

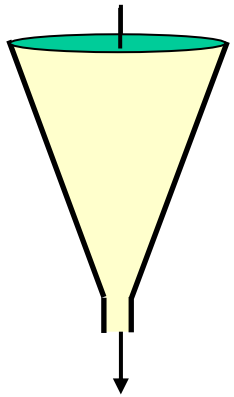
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



2-D Lib.

C[1](=C(C(=CS@1(=O)=O)SC[9]:C:C:C(:C:C:@9)Br)C[16]:C:C:C:C:C@16)N

1-D Lib. (Full database)



Filtering

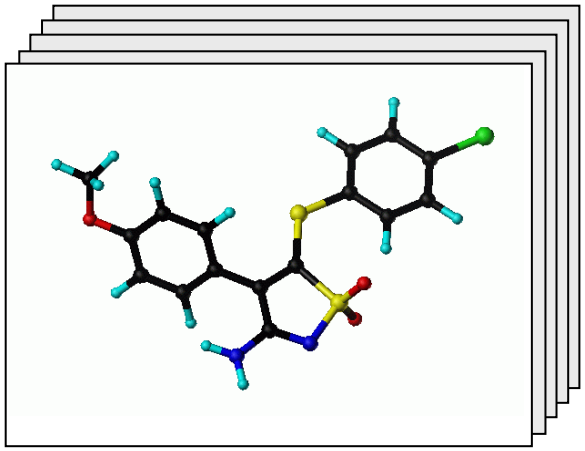
- ✓ *Chemical reactivty*
- ✓ *Physicochemical properties*
- ✓ *Pharmacokinetics*
- ✓ *Lead-likeness*
- ✓ *Drug-likeness*

C[1](=C(C(=CS@1(=O)=O)SC[9]:C:C:C(:C:C:@9)Br)C[16]:C:C:C:C:C@16)N

1-D Lib. (Filtered database)

2D → 3D

Hydrogens
Stereochemistry
Tautomerism
Ionisation



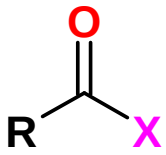
3-D Lib.

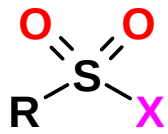
Charifson et al. *JCAMD*, 2002, **16**, 311-23

Chemical Filters

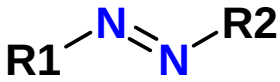
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

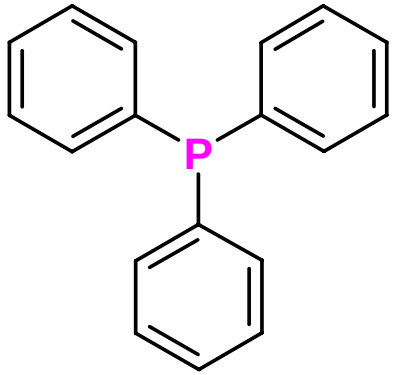
✓Remove any molecule bearing a chemically reactive moiety



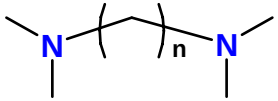


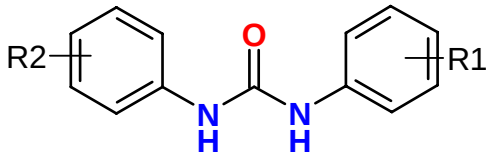
$R\text{-}NO_2$





Remove promiscuous binders (e.g. bis cations)





✓Known fluorescent molecules (dyes)

McGovern et al. *J. Med. Chem.* 2003, **46**, 4265-4272.

Pharmacokinetical filters

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

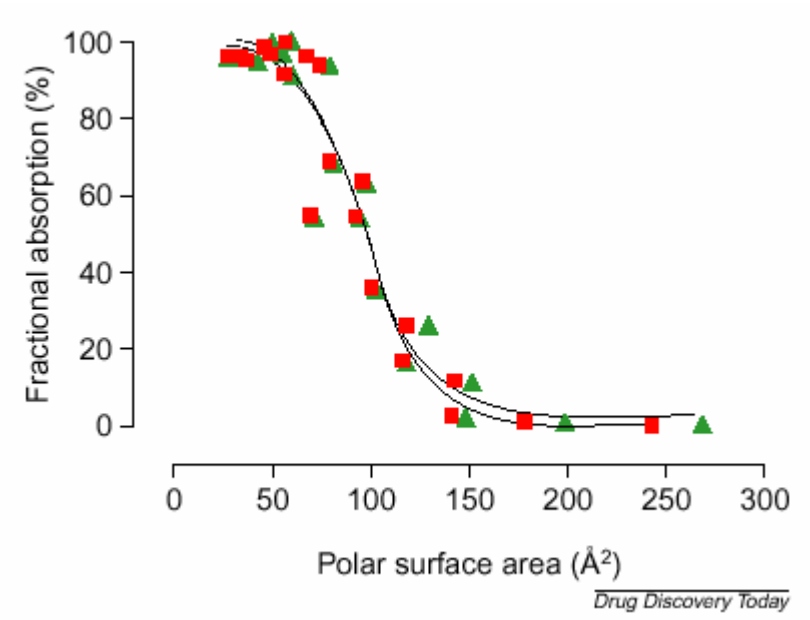
Physicochemical descriptors favouring membrane permeation

- ✓ Molecular weight < 500
- ✓ $logP_{calc}$ <5
- ✓ H-bond donors <5
- ✓ H-bond acceptors <10

Rule of 5 (Lipinski)

< 2 violations !!

✓Polar surface area <150 Å^2



Pickett et al. Drug Discovery Today, Feb.2000

Lead-likeness

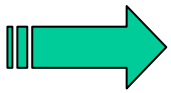
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

What are the differences between a lead and a drug ?

Drugs are generally …

- larger (ΔMW= 100)
- more hydrophobic (ΔclogP = 0.5)
- more complex (ΔRTB = 2, ΔRNG=ΔHAC=1)
- less soluble (ΔlogSw= -2)

…than their corresponding leads.

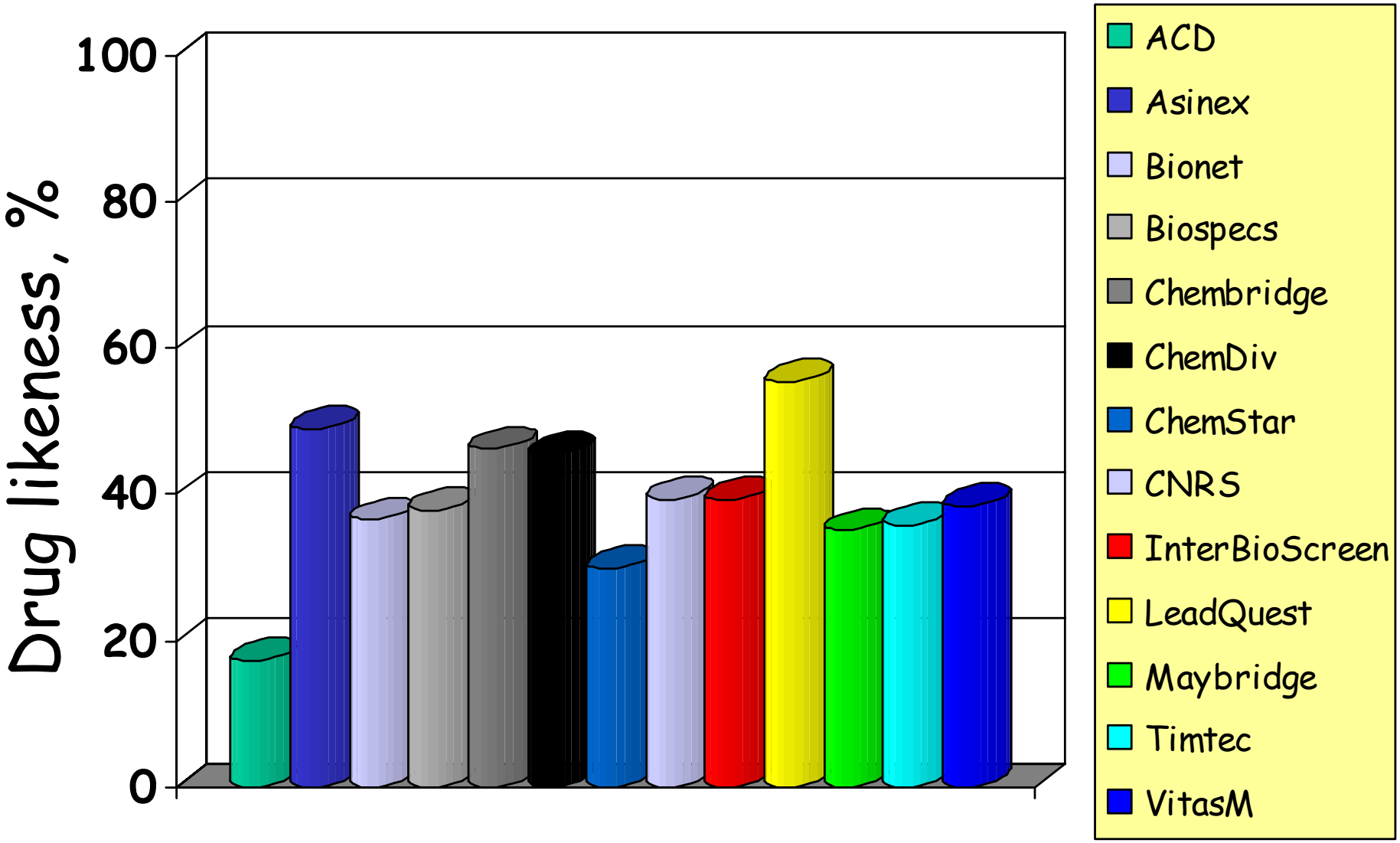


Needle screening (Roche)
SAR by NMR (Abott)
Scaffold co-crystallisation (Astex, Plexxikon)

Hann et al., *J. Chem. Inf. Comput. Sci.* **41**, 856-864 (2001)
Oprea et al. *J. Chem. Inf. Comput. Sci.* **41**, 1308-1315 (2001)
Oprea et al. *JCAMD,* **16**, 325-334 (2002)

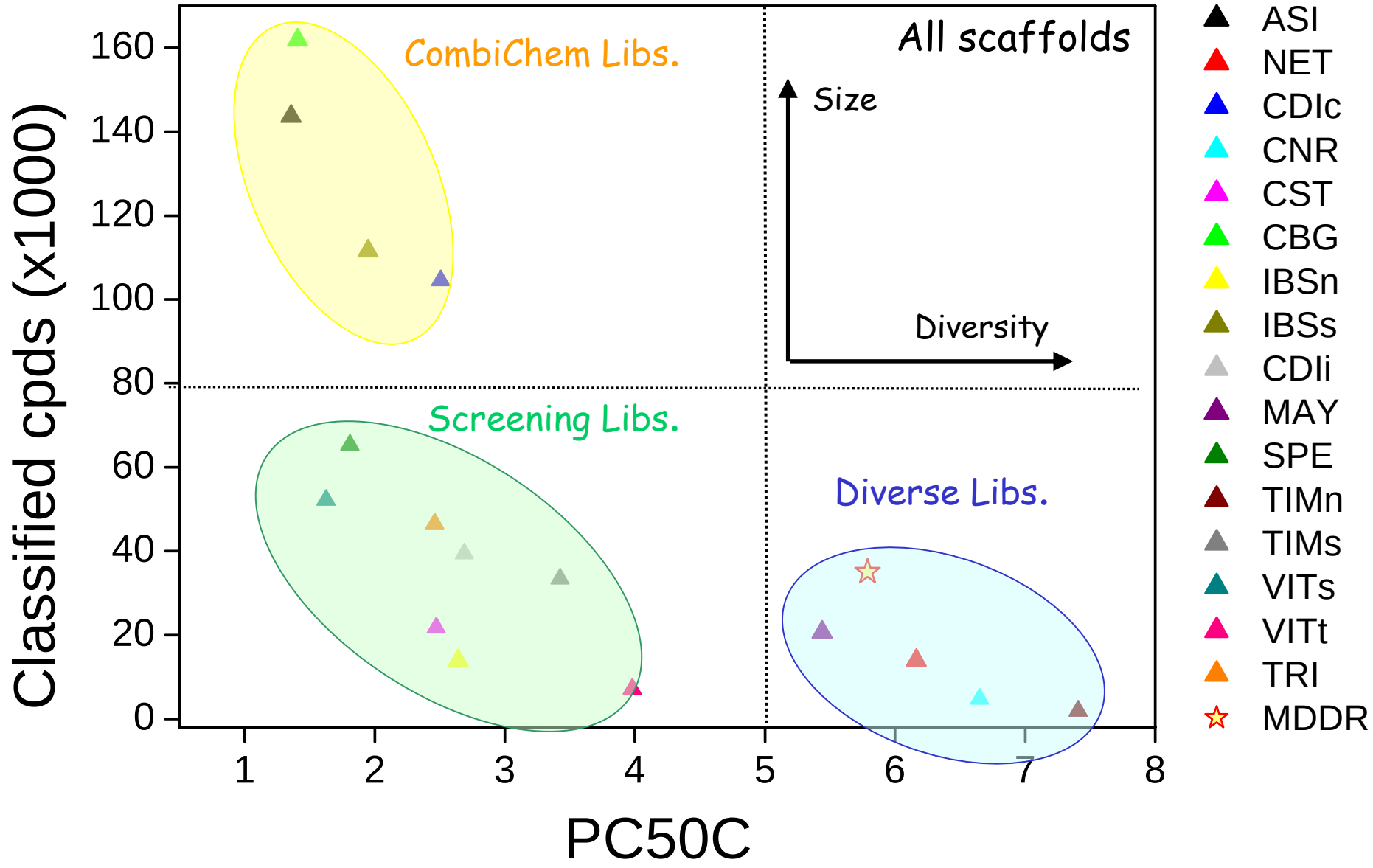
Commercially-available databases are not 'drug-like'

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



Krier et al. (2006) *J. Chem. Inf. Model.*, **46**, 512-524.

Commercially-available databases are not 'diverse'



Krier et al. (2006) *J. Chem. Inf. Model.*, **46**, 512-524.

Which Ligand Conformation ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

- Any low-energy conformer (1/2D→3D: Corina, Concord, Omega)

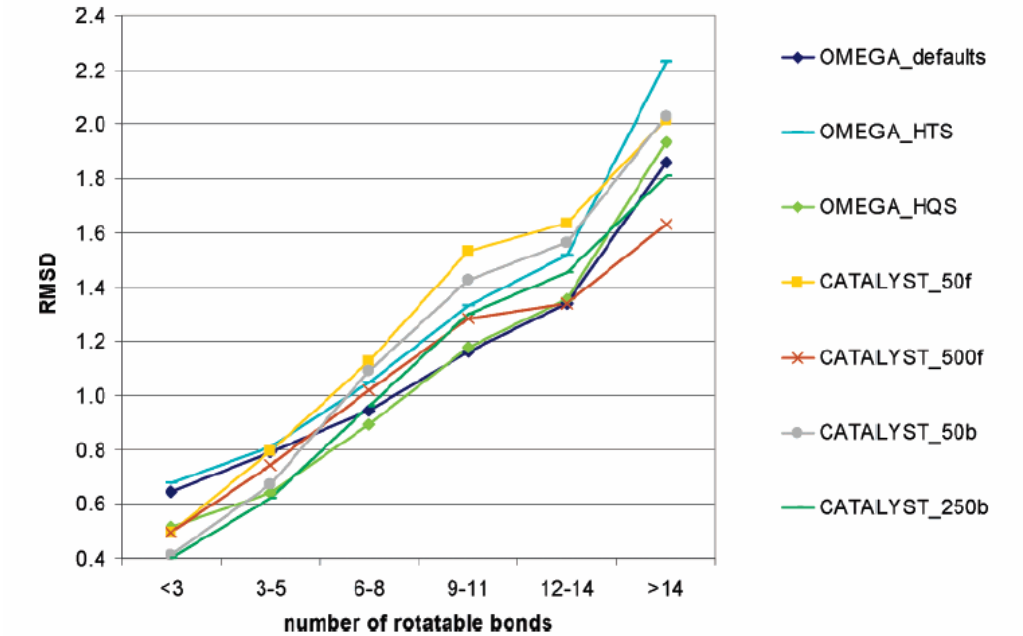
Most docking tools handle ligand flexibility on-the-fly (incremental building)

Issues : Seed structure, SMILES strings, Energy refinement

- Conformers database, for rigid dockers (e.g. Fred)

- Many tools (Catalyst, Omega) predict bioactive-like conformers for most drug-like compounds

- Only has to be done once



rmsd to protein-bound X-ray coord. (n =778)

Kirchmair et al. *J Chem Inf. Model* (2006) **46**, 1848-1861

Which Protein coordinates ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

- X-ray >> Homology models

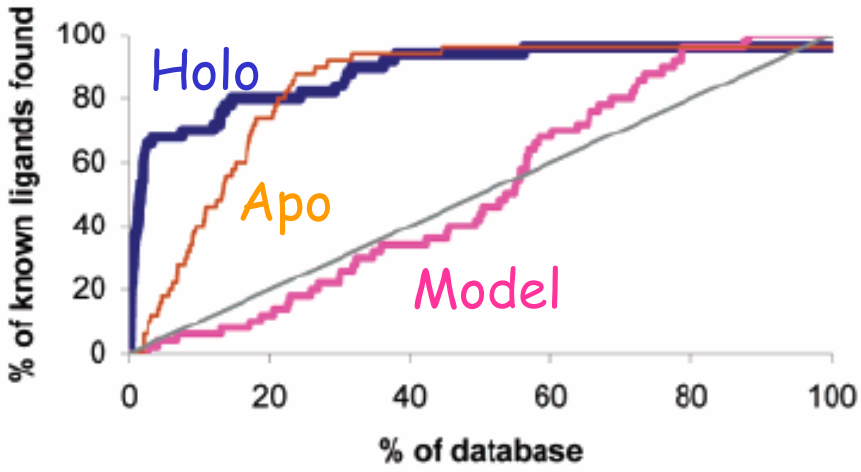
Table 1. Average Enrichment of Actives for Homology Models[a]

target	crystal structure	sequence id >50%	sequence id <50%
factor VIIa	5.7	6.6	1.2[a]
CDK2	4.5	3.1	3.5
both	5.1 ± 2.3	5.4 ± 3.2	2.1[a] ± 1.3

Homology models acceptable if close to X-ray templates

Oshiro et al. *J Med Chem* (2004), **47**, 764-767

- Holo >> Apo structures



Docking of 50 known GART Inhibitors
+ 95 000 MDDR decoys

McGovern et al. *J Med Chem* (2003), **46**, 2895-2905

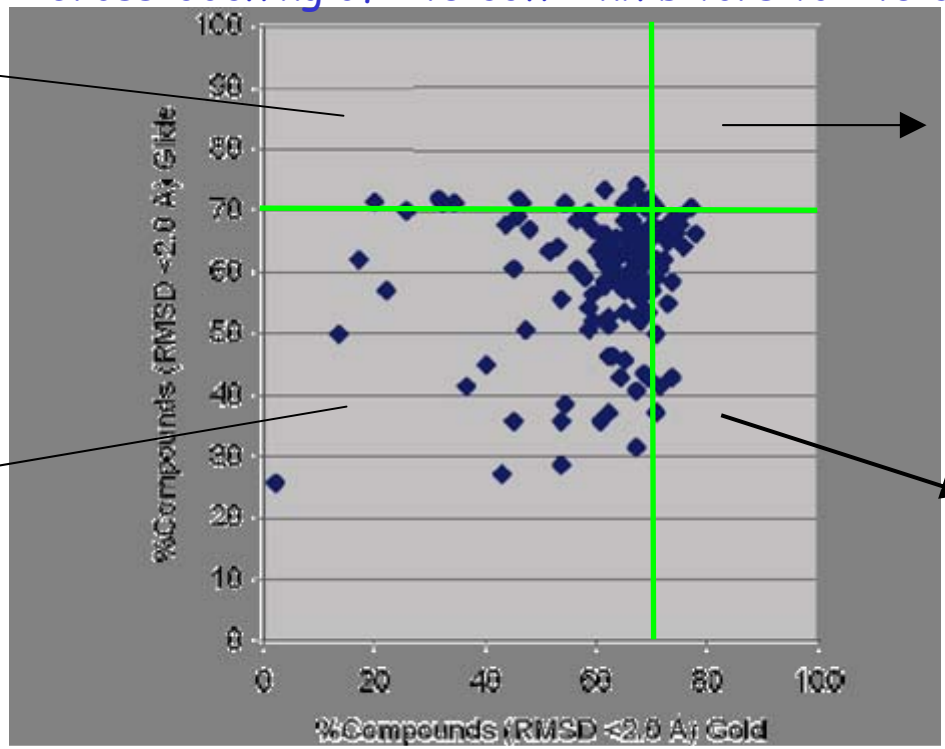
Which Protein coordinates ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

What to do if numerous X-ray structures of protein-ligand are available ?

Cross-docking of 140 cdk2 inhibitors to 140 cdk2 PDB entries

✓Glide
× Gold

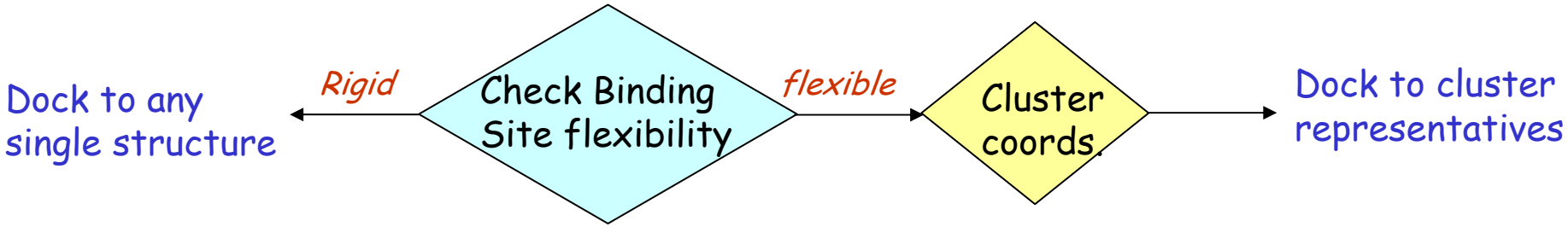


✓Glide
✓ Gold

× Glide
× Gold

× Glide
✓Gold

Duca et al. *J Chem Info Model* (2008), **48**, 659-668).



Which Docking Tool ?

Over 65 docking tools around

Which one to take ?

Which Docking Tool ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Program	Reference(s)	Website	Search algorithm[a]
Affinity	N/A	www.accelrys.com/products/datasheets/i2_affinity_data.pdf	MC
ADAM	(Yamada and Itai, 1993; Mizutani *et al.*, 1994, 2006)	www.immd.co.jp/en/product_2.html	RBD
Autodock 4.0	(Morris *et al.*, 1998; Osterberg *et al.*, 2002)	autodock.scripps.edu/	GA
CDOCKER	(Wu *et al.*, 2003a)	N/A	MD-SA
CHARMM	(Vieth *et al.*, 1998b, a)	www.charmm.org/	GA and MC
CLIX	(Lawrence and Davis, 1992)	N/A	RBD
DARWIN	(Taylor and Burnett, 2000)	N/A	GA
DIVALI	(Clark and Ajay, 1995) ,	N/A	GA
DOCK 6	(Oshiro *et al.*, 1995; Knegtel *et al.*, 1997; Kang *et al.*, 2004; Moustakas *et al.*, 2006)	dock.compbio.ucsf.edu/	IC
DockIt	N/A	www.metaphorics.com/products/dockit.html	N/A
DockVision 1.0.3	(Hart and Read, 1992) ,	www.dockvision.com/	MC
DoMCoSAR	(Vieth and Cummins, 2000)	N/A	SA
DragHOME	(Schafferhans and Klebe, 2001)	N/A	RBD
EADock	(Grosdidier *et al.*, 2007)	N/A	EA
eHiTs	(Zsoldos *et al.*, 2006)	www.simbiosys.ca/ehits/index.html	RBD of fragments followed by reconstruction
EUDOC	(Pang *et al.*, 2001) ,	N/A	RBD
FDS	(Taylor *et al.*, 2003)	N/A	MC
DAIM-SEED-FFLD	(Majeux *et al.*, 1999; Budin *et al.*, 2001; Kolb and Caflisch, 2006)	biocroma.uzh.ch/SDL/DAIM-SEED-FFLD_agreement.pdf	RBD
FITTED	(Corbeil *et al.*, 2007)	www.fitted.ca	GA
FlexX 2.2	(Rarey *et al.*, 1996)	www.biosolveit.de/	IC
FlexE	(Rarey *et al.*, 1999; Claußen *et al.*, 2001)		
FlipDock (formerly pyDock)	(Zhao and Sanner, 2007)	www.scripps.edu/~yongzhao/FLIPDock/	GA
FLOG	(Miller *et al.*, 1994)	N/A	IC
FRED	(McGann *et al.*, 2003)	www.eyesopen.com/products/applications/fred.html	RBD
FTDock	(Gabb *et al.*, 1997)	www.bmm.icnet.uk/docking/ftdock.html	RBD
GAMBLER	(Charifson *et al.*, 1999)	N/A	GA
GasDock	(Li *et al.*, 2004)	N/A	GA
GEMDOCK	(Yang and Chen, 2004)	N/A	EA
GlamDock	(Tietze and Apostolakis, 2007)	N/A	MC/SA
Glide 4.0	(Friesner *et al.*, 2004; Sherman *et al.*, 2006)	www.schrodinger.com/ProductDescription.php?mID=6&sID=6&cID=0	Hierarchical filters and MC

Abbreviation: N/A, not available.
[a]DE, differential evolution; EA, evolutionary algorithm; GA, genetic algorithm; IC, incremental construction; MA, matching algorithm; MD, molecular dynamics; RBD, rigid-body docking; SA, simulated annealing; TS, Tabu search.

Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

Which Docking Tool ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Program	Reference(s)	Website	Search algorithm[a]
GOLD 3.1	(Verdonk *et al.*, 2003, 2005)	www.ccdc.cam.ac.uk/products/life_sciences/gold/	GA
Hammerhead	(Welch *et al.*, 1996)	N/A	IC
HADDOCK	(Dominguez *et al.*, 2003)	www.nmr.chem.uu.nl/haddock/	SA
HierDOCK/HierVLS	(Floriano *et al.*, 2004; Trabanino *et al.*, 2004)	N/A	IC
ICM	(Abagyan *et al.*, 1994b; Totrov and Abagyan, 1997)	www.molsoft.com/docking.html	MC
LibDock	(Diller and Merz, 2001)	N/A	MA
LIDAEUS	(Wu *et al.*, 2003b)	N/A	MA
LIGIN	(Sobolev *et al.*, 1996)	N/A	RBD
MacDOCK	(Fradera *et al.*, 2004)	N/A	IC
MCDOCK	(Liu and Wang, 1999)	N/A	MC
MolDock	(Thomsen and Christensen, 2006)	N/A	DE
MVP	N/A	N/A	IC
PatchDOCK	(Schneidman-Duhovny *et al.*, 2005)	N/A	Shape complementarity
PAS-Dock	(Tøndel *et al.*, 2006)	N/A	TS
PhDOCK	(Joseph-McCarthy *et al.*, 2003)	N/A	MA
Ph4DOCK	(Goto *et al.*, 2004)	N/A	MA
PIPER	(Kozakov *et al.*, 2006)	N/A	Fast Fourier transform
Plants	(Korb *et al.*, 2006)	N/A	Ant colony optimisation
Prodock	(Trosset and Scheraga, 1999)	N/A	MC
PRO_LEADS	(Murray *et al.*, 1999)	N/A	TS
ProPose	(Seifert, 2005)	N/A	IC
PSI_DOCK	(Pei *et al.*, 2006)	N/A	GA with TS
Q-fit	(Jackson, 2002)	N/A	MA
Quantum	N/A	www.q-lead.com/cnt/quantum	N/A
QXP/Flo +	(McMartin and Bohacek, 1997)	N/A	MC
RiboDock	(Morley and Afshar, 2004)	N/A	MC
ROSETTALIGAND	(Meiler and Baker, 2006)	N/A	MC
SANDOCK	(Burkhard *et al.*, 1998)	N/A	MA
SDOCKER	(Wu and Vieth, 2004)	N/A	Random walk
SLIDE	(Schnecke and Kuhn, 2000; Zavodszky and Kuhn, 2005)	N/A	IC
SKELGEN	(Alberts *et al.*, 2005)	N/A	IC
SODOCK	(Chen *et al.*, 2007)	N/A	Swarm Optimisation
SG-DOCK/SP-DOCK	(Fradera *et al.*, 2000)	N/A	IC
Surflex 2.1	(Jain, 2003, 2007)	N/A	IC with MA
Yucca	(Choi, 2005)	N/A	MC

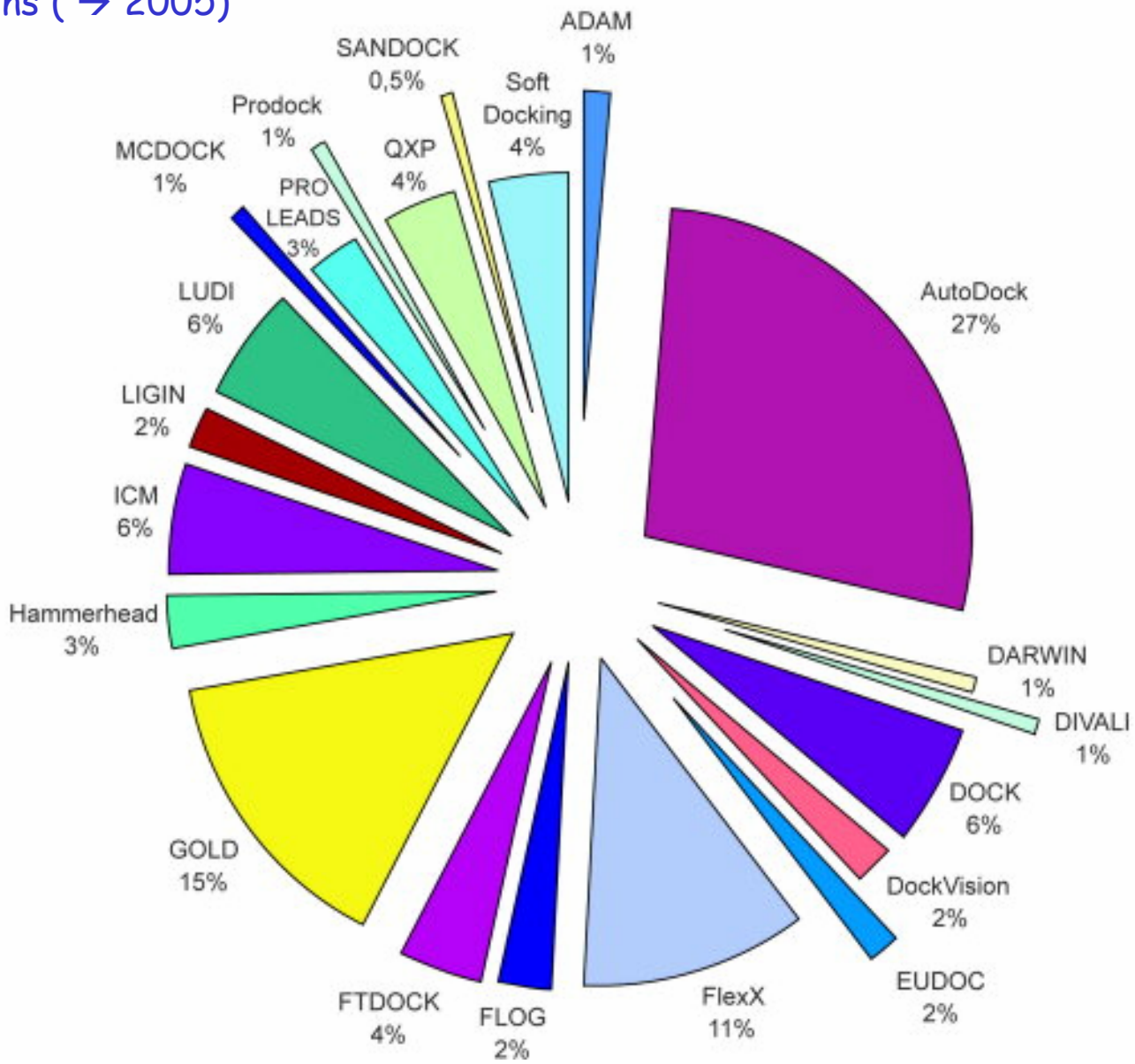
Abbreviation: N/A, not available.
[a]DE, differential evolution; EA, evolutionary algorithm; GA, genetic algorithm; IC, incremental construction; MA, matching algorithm; MD, molecular dynamics; RBD, rigid-body docking; SA, simulated annealing; TS, Tabu search.

Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

Which Docking tool ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

ISI citations (→ 2005)



Sousa et al (2006) *Proteins*, **65**:15-26.

Docking methods

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Docking means : Find **quickly** (< 1 min) :

- the bound conformation of the ligand
- the respective orientation of the ligand v. protein

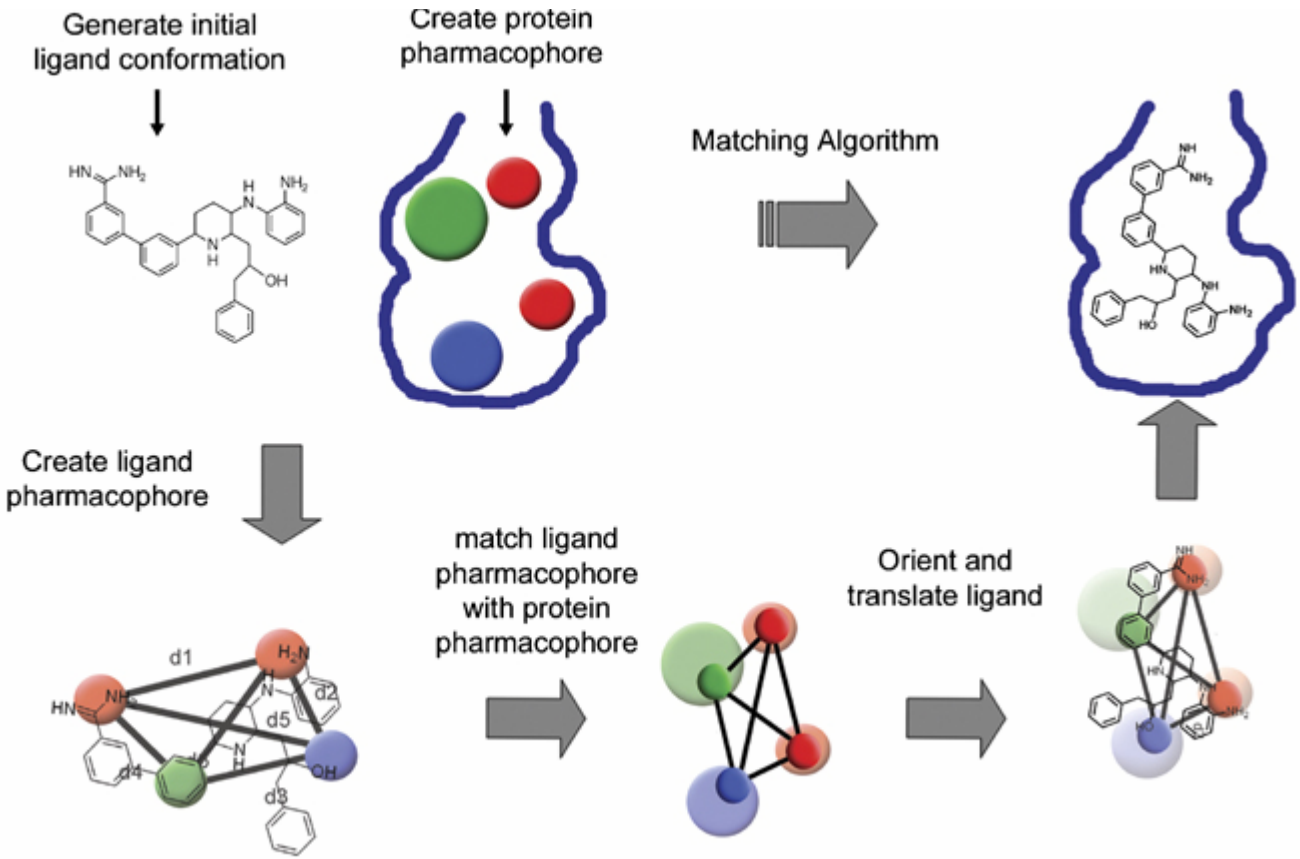
Pose

Step 1. Conformational sampling of the ligand in the protein, active site

Step 2. Scoring each pose with a scoring function

Conformational sampling methods

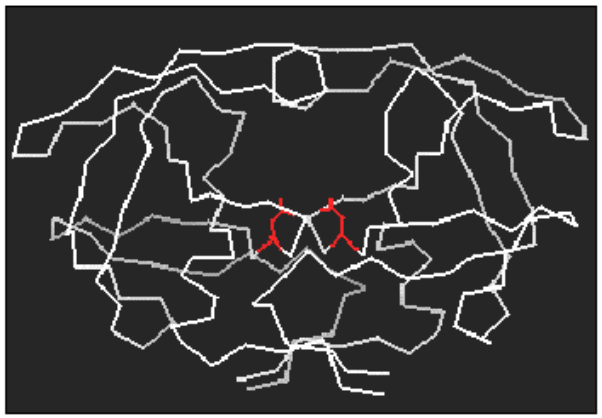
- 1. Rigid body docking (DOCK, FRED)



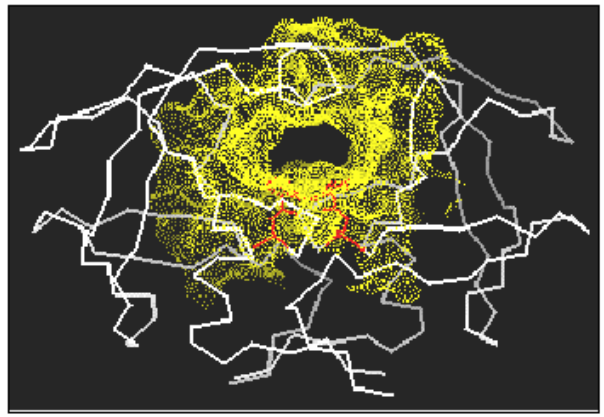
Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

Conformational sampling methods

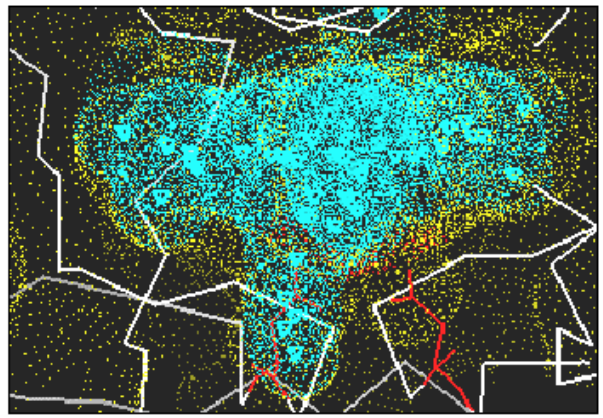
Surface-based orientation (e.g. DOCK)



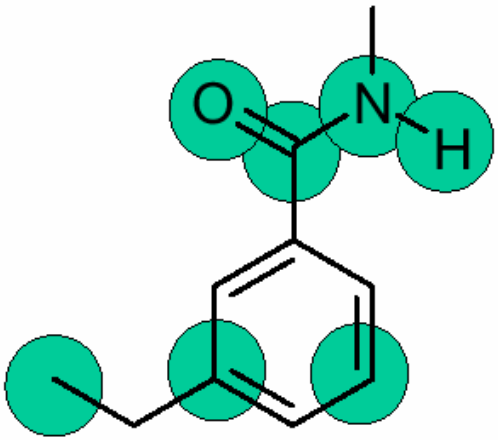
1. 3D structure



2. Molecular surface (active site)



3. Filling the surface by overlapping spheres

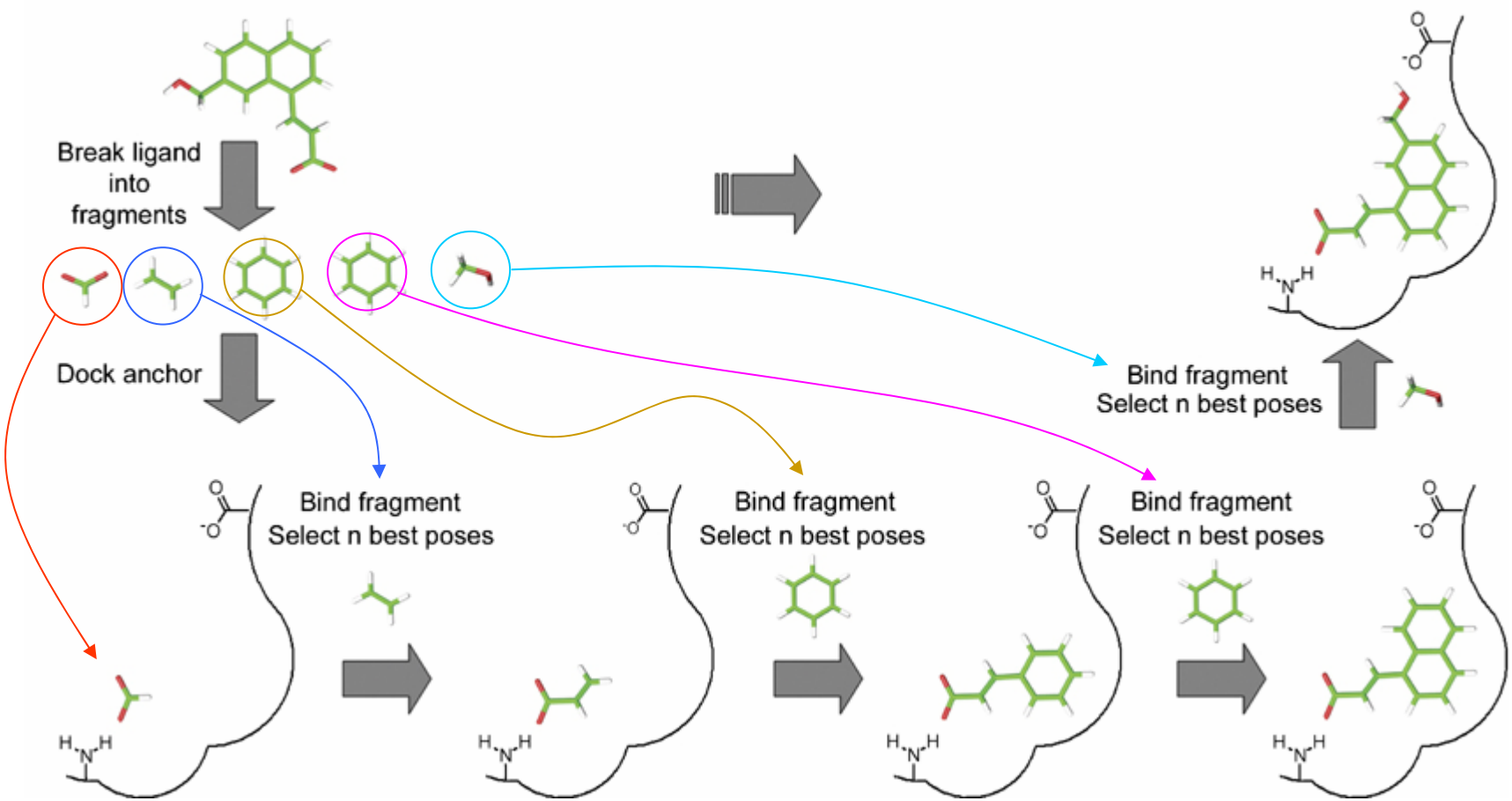


4. Matching sphere centers with atoms

Conformational sampling methods

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

2. Incremental construction (Dock, FlexX, Surflex, eHITS)

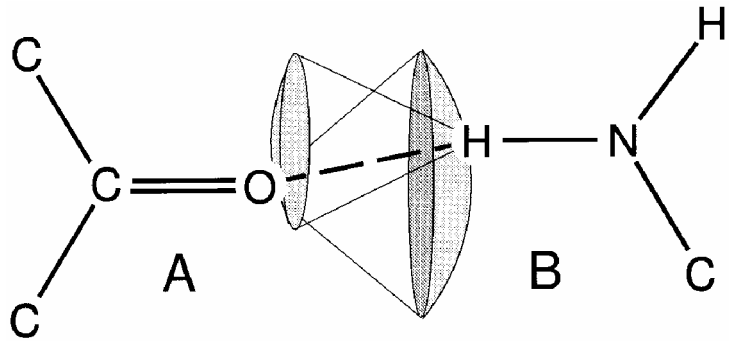


Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

Conformational sampling methods

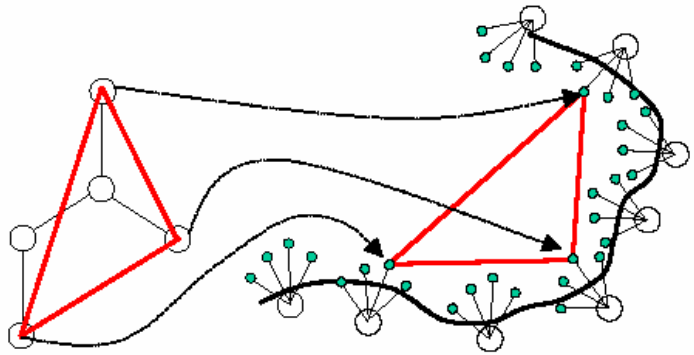
FlexX: Rarey et al. J. Comp.-Aid. Mol. Design (1996) **10**, 41-54

1. Interaction centers and interaction surfaces identified on both receptor (A) and ligand (B)
   - H bond
   - Salt bridges
   - Aromatics
   - methyl-aromatics
   - amide-aromatics



2. Match if overlap of complementary interaction centers/surfaces

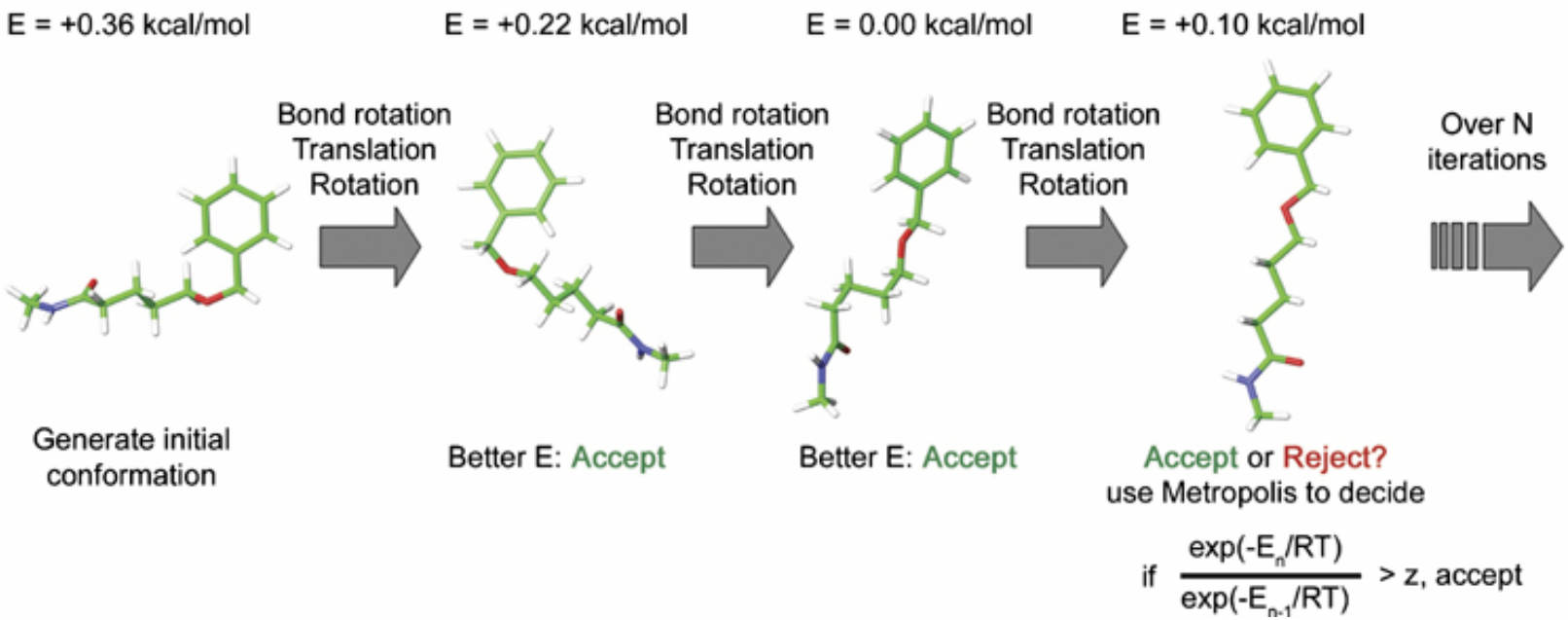
3. Ligand placement using triangulation



Conformational sampling methods

3. Stochastic methods : MC, MD, GA

Monte Carlo (MC): Glide



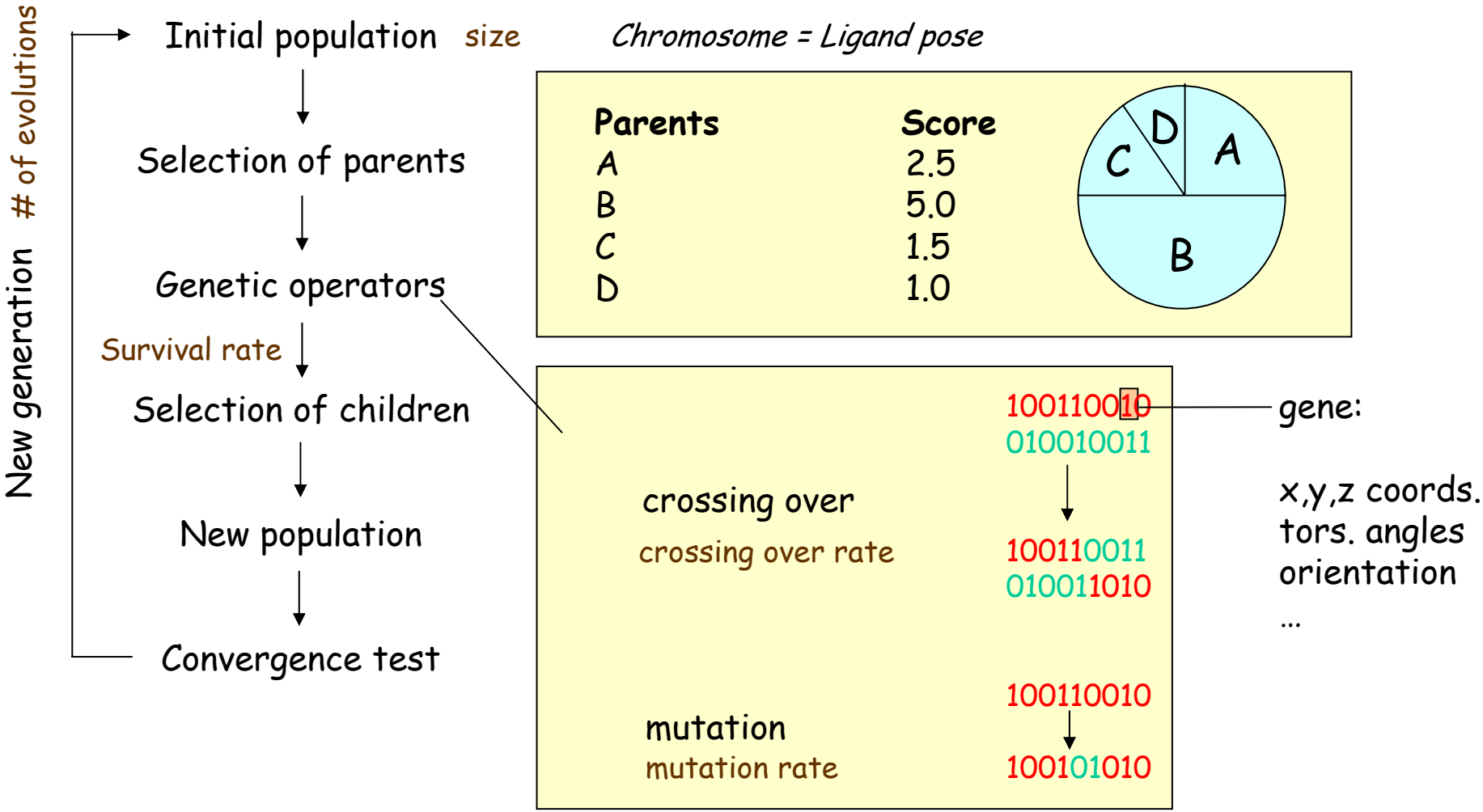
Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

Conformational sampling methods

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

3. Stochastic methods : MC, MD, GA (Gold, AutoDock, Glide)

Genetic algorithm (GA): Gold, AutoDock



Which Scoring function ?

Over 30 scoring functions around

Which one to take ?

Which Scoring function ?

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Scoring function	*Software implementations*	*Class*	*Reference(s)*
ChemScore	GOLD, FRED, CScore, PRO_LEADS	Empirical	(Eldridge *et al.*, 1997)
eHiTS SF	eHiTS	Empirical	(Zsoldos *et al.*, 2003)
FlexX SF	FlexX	Empirical	(Rarey *et al.*, 1996)
Fresno	*Standalone*	Empirical	(Rognan *et al.*, 1999)
GlideScore	Glide	Empirical	(Friesner *et al.*, 2004, 2006)
Hammerhead SF	Hammerhead, Surflex, LigandFit	Empirical	(Jain, 1996)
HINT	*Standalone*	Empirical	(Cozzini *et al.*, 2002)
LigScore	LigandFit	Empirical	(Krammer *et al.*, 2005)
PLP	LigandFit, FRED, DockIt	Empirical	(Gehlhaar *et al.*, 1995)
RankScore	FITTED	Empirical/FF	(Moitessier *et al.*, 2006a)
SAFE_p	*None*	Empirical/FF	(Sussman *et al.*, 2002)
SCORE	*Standalone*	Empirical	(Wang *et al.*, 1998)
SCORE 3.0	PSI-DOCK	Empirical	(Pei *et al.*, 2006)
SCORE1	LUDI	Empirical	(Böhm, 1994)
SCORE2	LUDI	Empirical	(Böhm, 1998)
ScreenScore	FRED	Empirical/consensus	(Stahl and Rarey, 2001)
SIE	*Standalone*	Empirical/FF	(Naim *et al.*, 2007)
SLIDE SCORE	SLIDE	Empirical	(Schnecke and Kuhn, 2000)
VALIDATE	*Standalone*	Empirical/FF	(Head *et al.*, 1996)
X-Score	*Standalone*	Empirical/consensus	(Wang *et al.*, 2002)
AutoDock SF	AutoDock, SODOCK	FF/empirical	(Morris *et al.*, 1998)
DockScore	DOCK, CScore	FF	(Meng *et al.*, 1992)
GoldScore	GOLD, CScore	FF	(Jones *et al.*, 1997)
HADDOCK Score	HADDOCK	FF	(van Dijk *et al.*, 2006b)
ICM SF	ICM	FF	(Abagyan *et al.*, 1994b)
QXP SF	QXP/MCDOCK	FF	(McMartin and Bohacek, 1997)
BLEEP	*Standalone*	Knowledge based	(Mitchell *et al.*, 1999)
$DrugScore^{CSD}$	*Standalone*	Knowledge based	(Velec *et al.*, 2005)
$DrugScore^{PDB}$	*Standalone*	Knowledge based	(Gohlke *et al.*, 2000)
M-Score	*Standalone*	Knowledge based	(Yang *et al.*, 2006)
PMF	CScore, LigandFit, BioMedCAChe, DockIt	Knowledge based	(Muegge and Martin, 1999; Muegge, 2006)
SMoG	SMoG	Knowledge based	(DeWitte and Shakhnovich, 1996; Ishchenko and Shakhnovich, 2002)

Abbreviations: FF, force field; SF, scoring function.

Moitessier et al. (2008) *Br J Pharmacol* , **153**: S7-26.

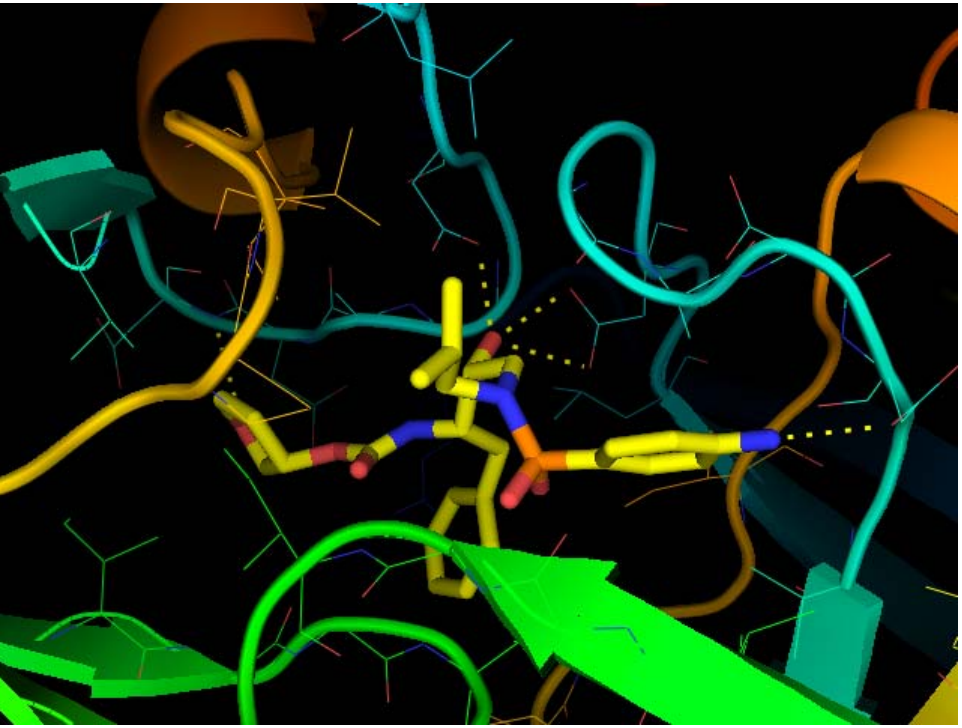
Scoring functions

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Two aims:

- ✓ Rank docking poses by decreasing interaction energy
- ✓ Predict the absolute binding free energy (affinity) of compounds

?



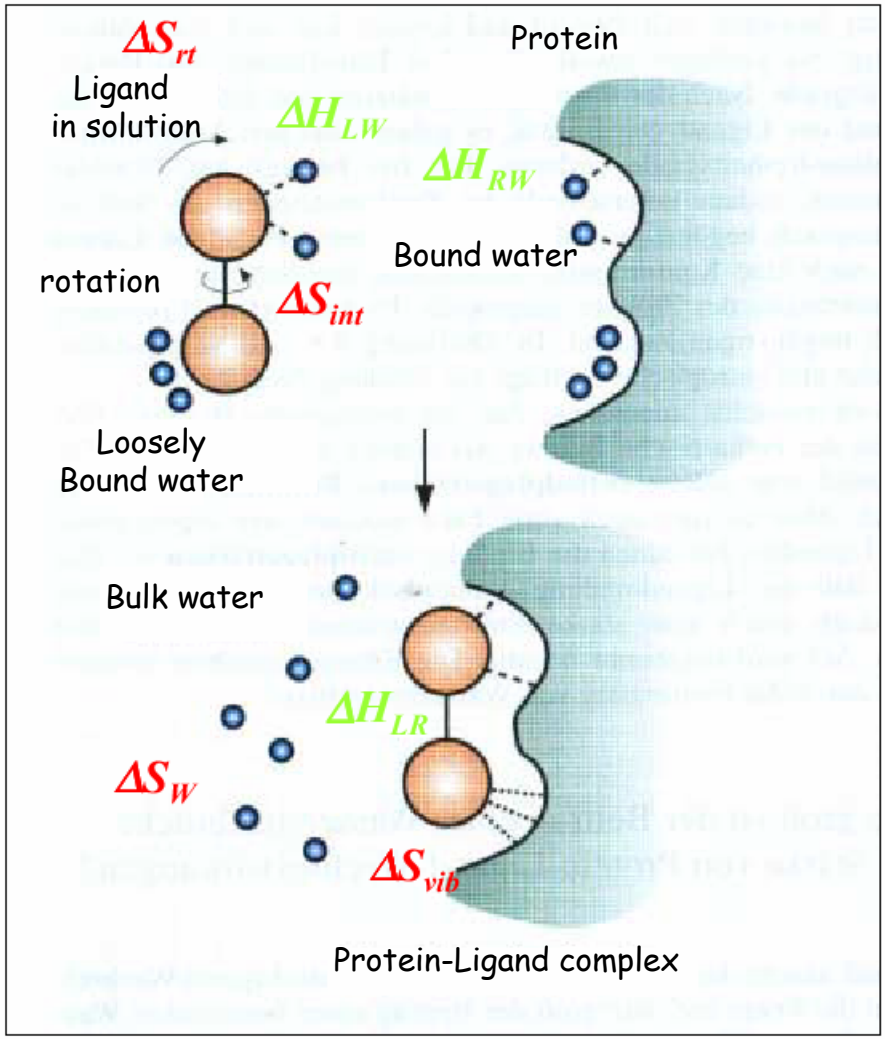
$$\Delta G_{binding} = RT.\log K_D$$

Binding free energy

Equilibrium dissociation constant

$\Delta G_{binding}$ = f (Interactions)

Scoring ΔG_{bind} is extraordinarily difficult



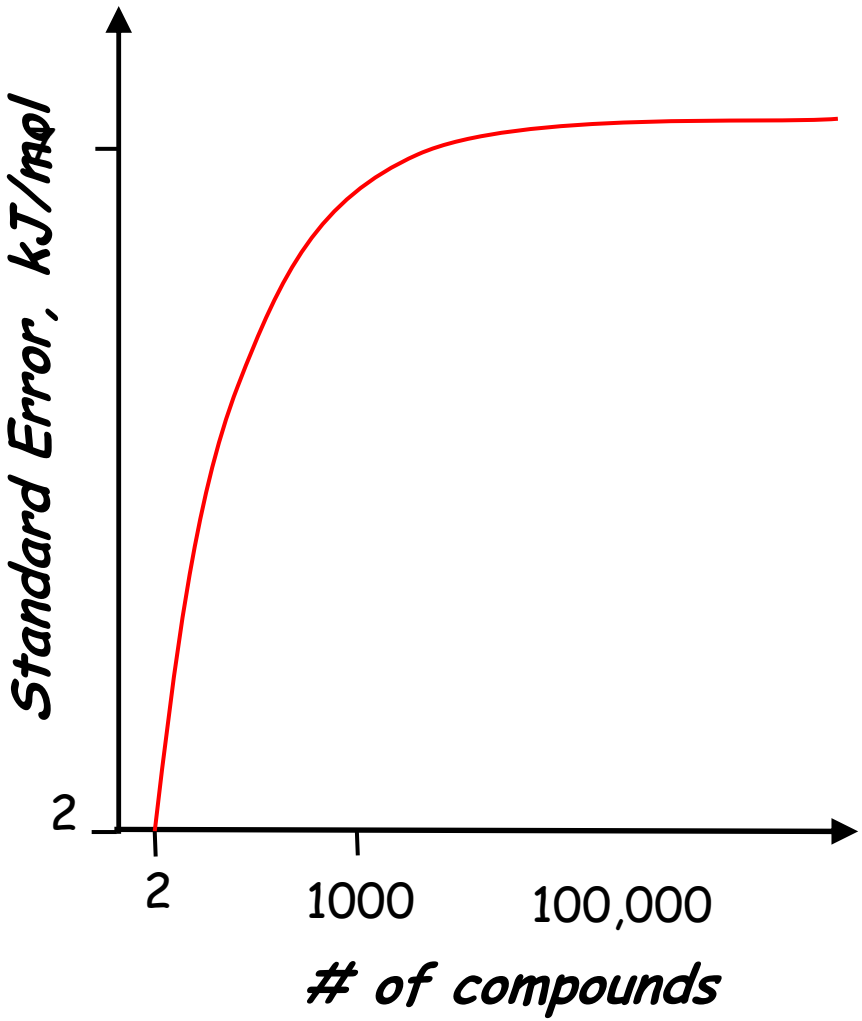
Binding free energy

$$\Delta G = \Delta H - T\Delta S$$

Entropy

Enthaly

Predicting binding free energies



Précision

Thermodynamic methods (2)

MM-PBSA, GBSA (100)

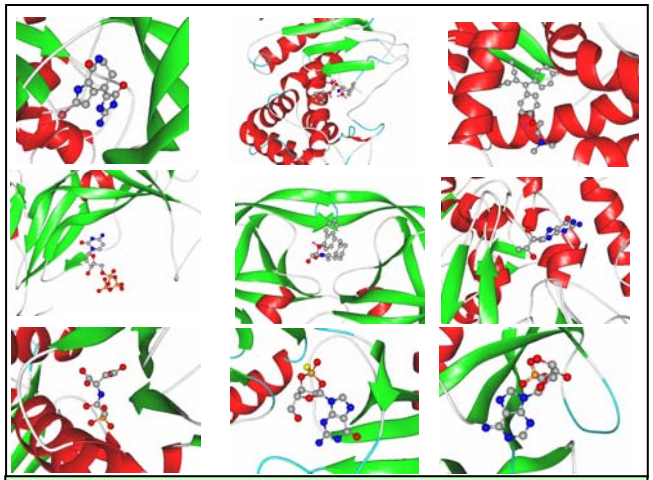
QSAR (<1000)

Empirical functions (>100,000)

Empirical Scoring fucntions

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Ludi, FlexX, Glidescore, Chemscore, Fresno, PLP



Protein-Ligand complexes (PDB)

$$\Delta G_{bind} =$$
$$\Delta G_0$$
$$+ \Delta G_{hb} \sum_{hb} f(\Delta R, \Delta\alpha)$$
$$+ \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha)$$
$$+ \Delta G_{lipo} \sum_{hb} |A_{lipo}|$$
$$+ \Delta G_{rot} NROT$$

H-bond

Ionic

Lipophilic

Rotation

ΔG_o: constant term ⇔ entropy lost
ΔG_{hb}: contribution of one perfect H-bond
$f(\Delta R, \Delta\alpha)$: penalty for bad geometry
ΔG_{ionic} : ionic contribution
ΔG_{lipo} : lipophilic contribution
A_{lipo} : lipophilic contact surface
ΔG_{rot} : lost of free energy due to internal rotation
Nrot : number of rotatable bond

- ✓ fast
- ✓ Implicit inclusion of many effects

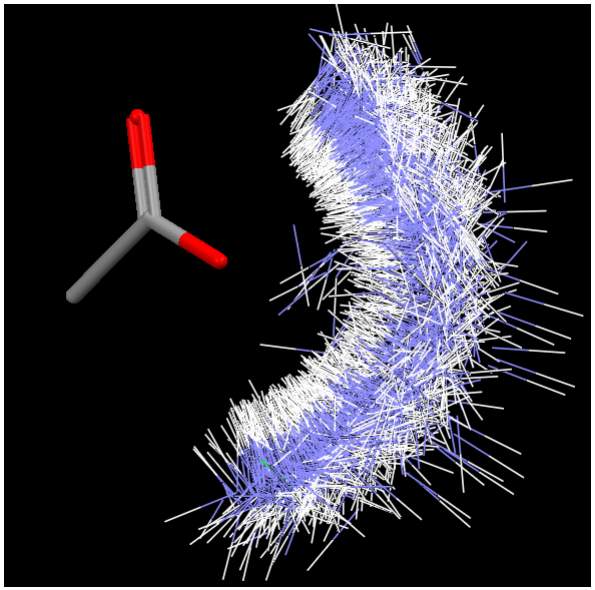
- ✓ Strongly depends on training set
- ✓ No penalty for repulsive interactions

Böhm: *J Comp Aided Mol Design* (1994), **8**, 243-256

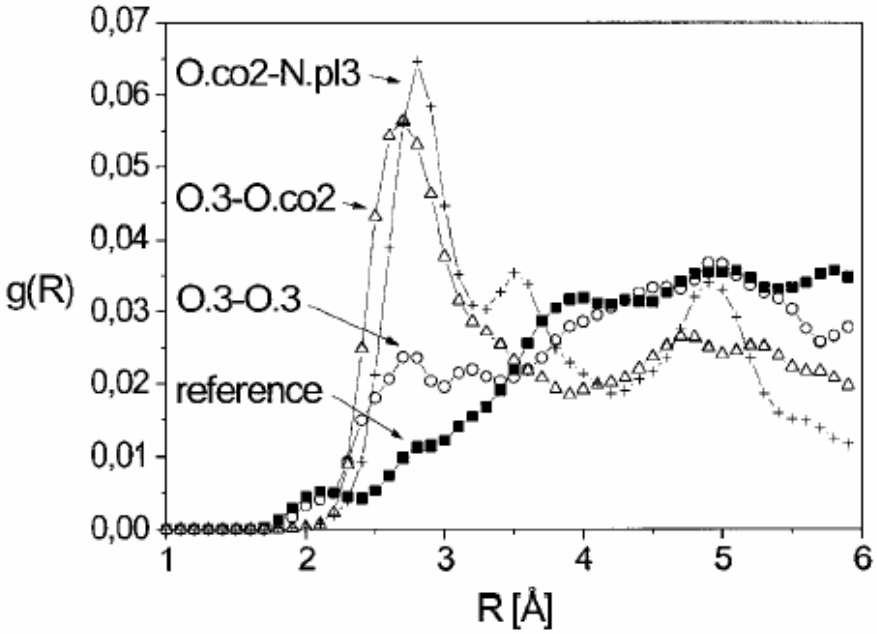
Potentials of mean force

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

PMF, Drugscore, Bleep, Smog



Pairwise atomic contacts
(O.co2, NH^+)



No parametrization/fit to experimental values

No intra-molecular contribution

$$A(r) = -K_B T \ln g_{ij}(r)$$

Helmhotz free energy (atom pair)

Radial distribution for distance r

$$Score = \sum_i \sum_j A(r_{ij})$$

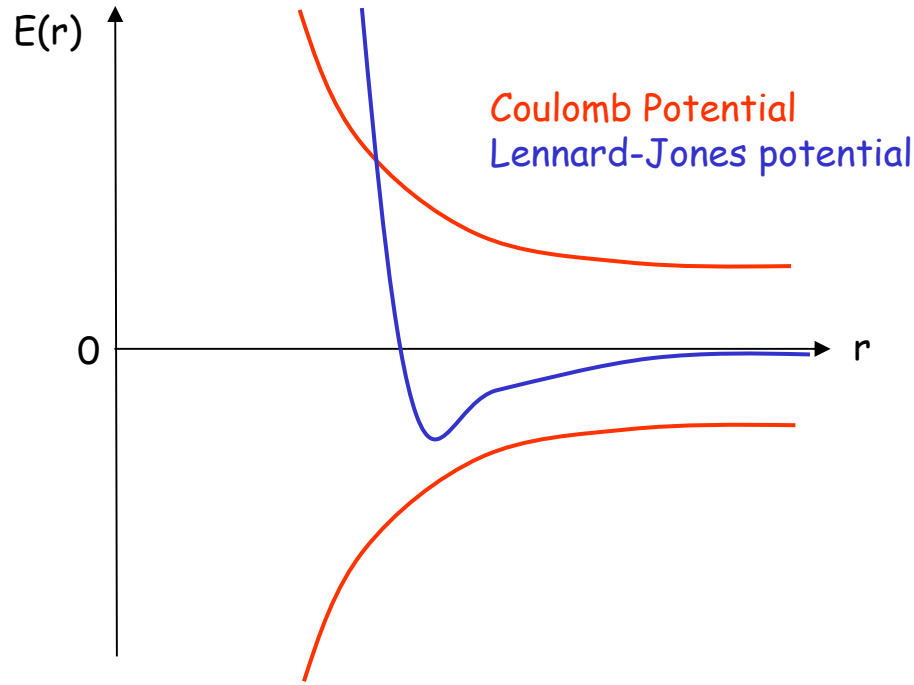
Force-fields

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Dock, AutoDock, Goldscore

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} [\frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^{6}} + 332 \frac{q_i q_j}{Dr_{ij}}]$$

Describes only non-covalent interactions (Coulomb, Lennard-Jones)



✓ independent of training set

✓ stringly depends on ligand size (increasing number of interactions)

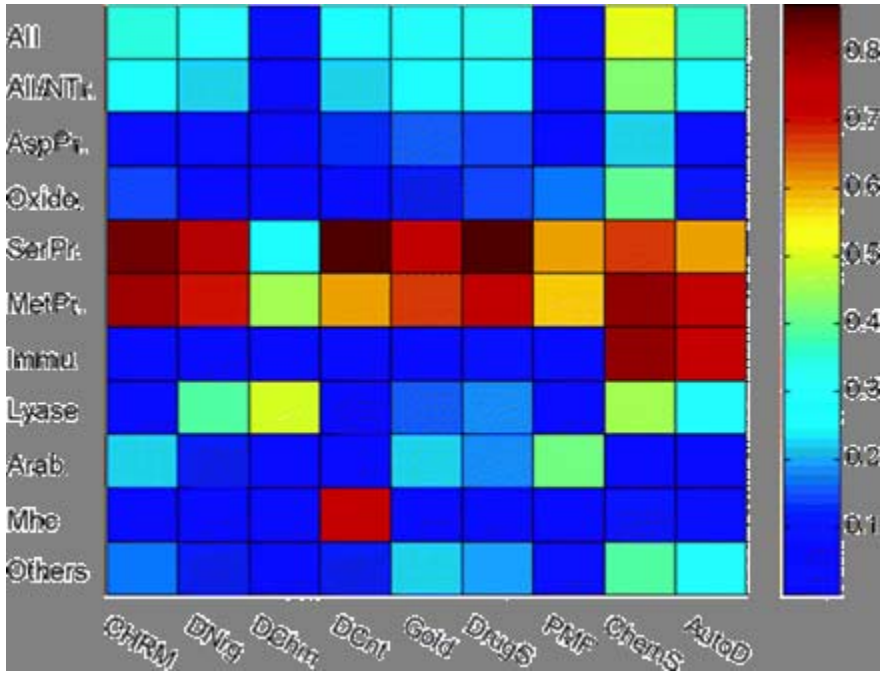
✓ force field accuracy, difficulty to set parameters

✓ no account for entropic contributions

Predicting ΔG bind is very difficult

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Prediction of ΔG_{bind} from 189 high-resolution PDB structures



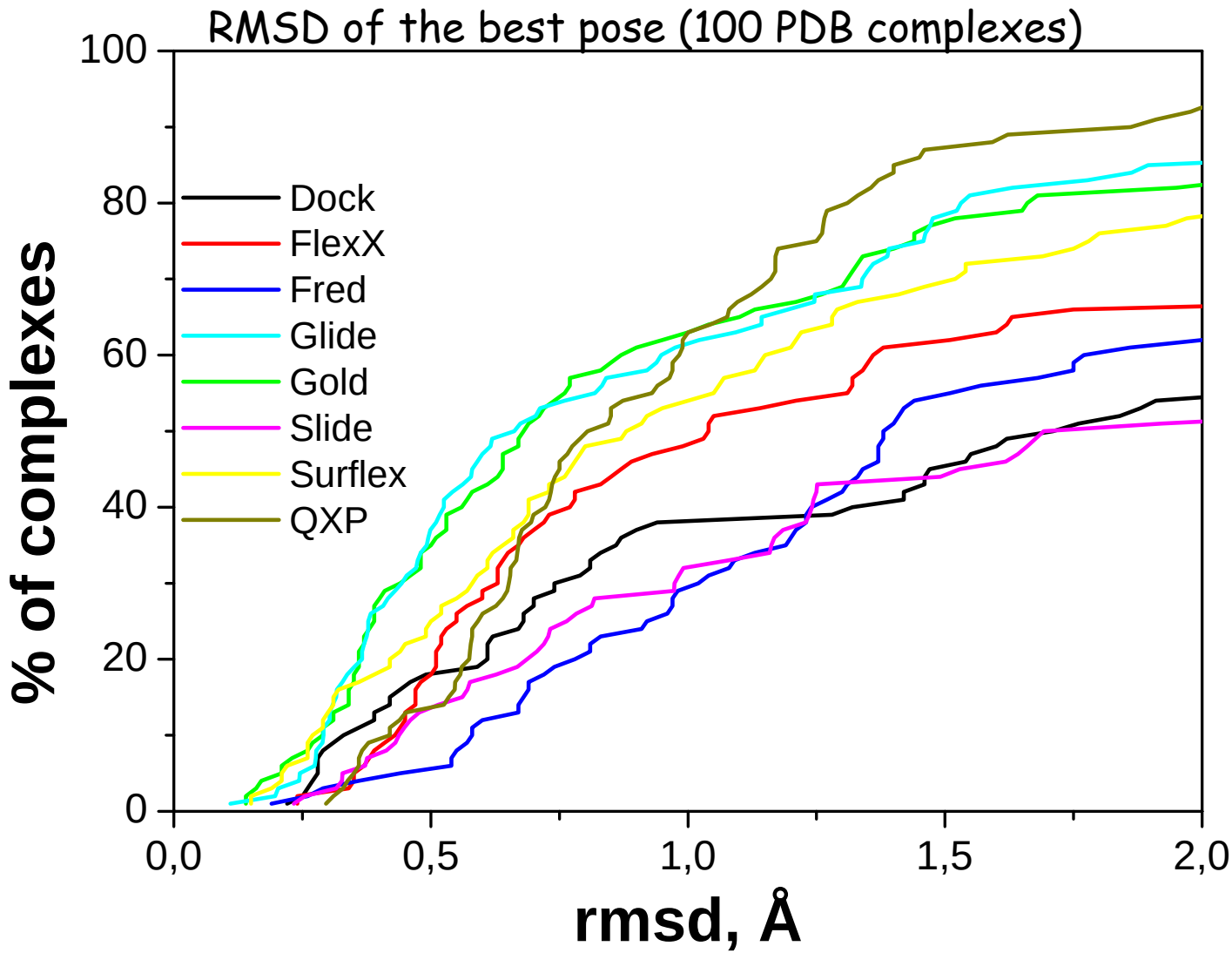
R^2

- ✓High-throughput prediction of ΔG_{bind} is still a challenge
- ✓Current accuracy : 7-10 kJ/mol (nM - μM - μM)
- ✓Tends to work better for some target families (e.g. proteases) and a set of congeneric ligands

Ferrara et al. *J. Med. Chem.* 2004, **47**:3032-3047

Docking Accuracy

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

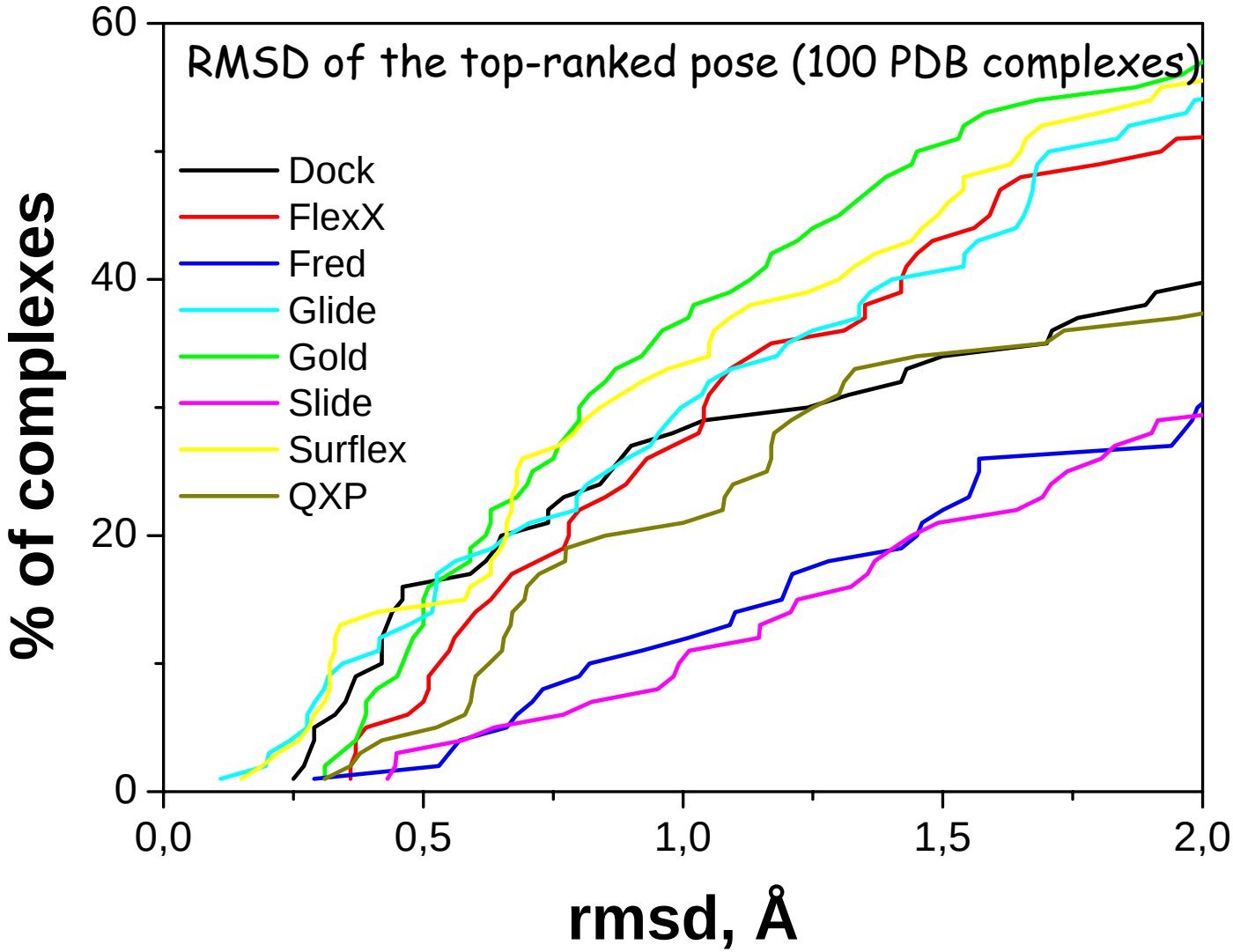


Finding a reliable pose out of a set of 30 solutions is feasible !

Kellenberger et al. *Proteins* (2004), **57**, 224-242

Ranking accuracy

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



Ranking the most reliable solution at the top of the list is still an issue !

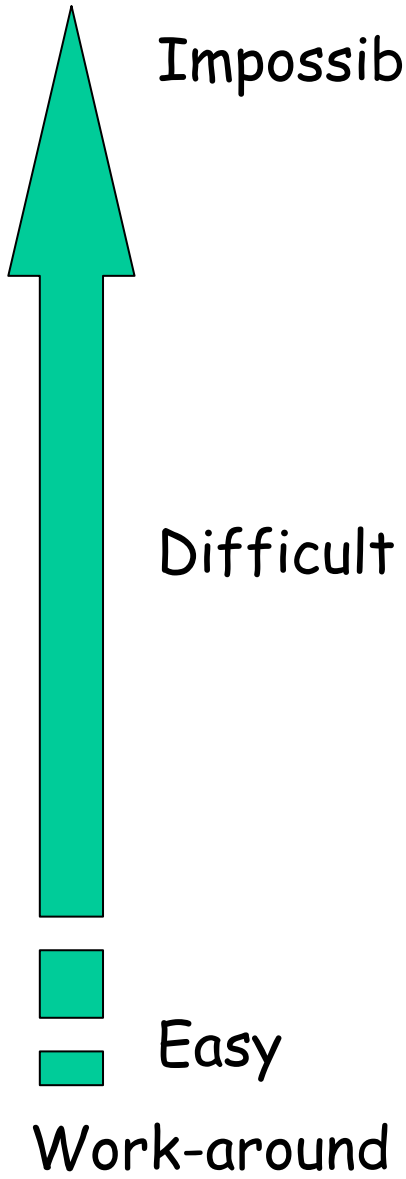
Kellenberger et al. *Proteins* (2004), **57**, 224-242

Source of Docking Errors

Nature of the active site
(flat vs. cavity)

Protein flexibility
Missed influence of water
Inaccuracy of the scoring function
Unusual binding mode/interactions
Ligand flexibility
Ligand symmetry

Inadequate set of protein coordinates
Wrong atom typing



Pros and Cons of docking codes

Program	Pros	Cons
DOCK	small binding sites opened cavities small hydrophobic ligands	flexible ligands highly polar ligands
FLEXX	small binding sites small hydrophobic ligands	very flexible ligands
FRED	large binding sites flexible ligands small hydrophobic ligands high speed	small polar buried ligands
GLIDE	flexible ligands small hydrophobic ligands	ranking very polar ligands low speed
GOLD	small binding sites small hydrophobic ligands	ranking very polar ligands ranking ligands in large cavities
SLIDE	sidechain flexibility	sensitivity to input coordinates
SURFLEX	large and opened cavities small binding sites very flexible ligands	low speed for large ligands
QXP	optimising known binding modes	Sensitivity to input coordinates

Kellenberger et al. *Proteins* (2004), **57**, 224-242

Post-processing

Simplify output

Automated selection of hits (by rank, by score)

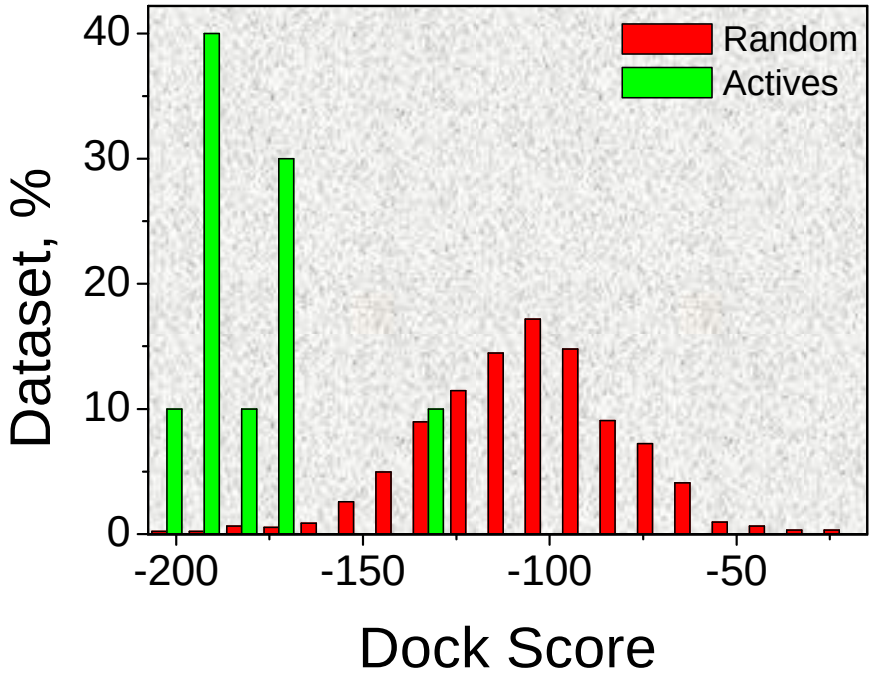
Remove false positives

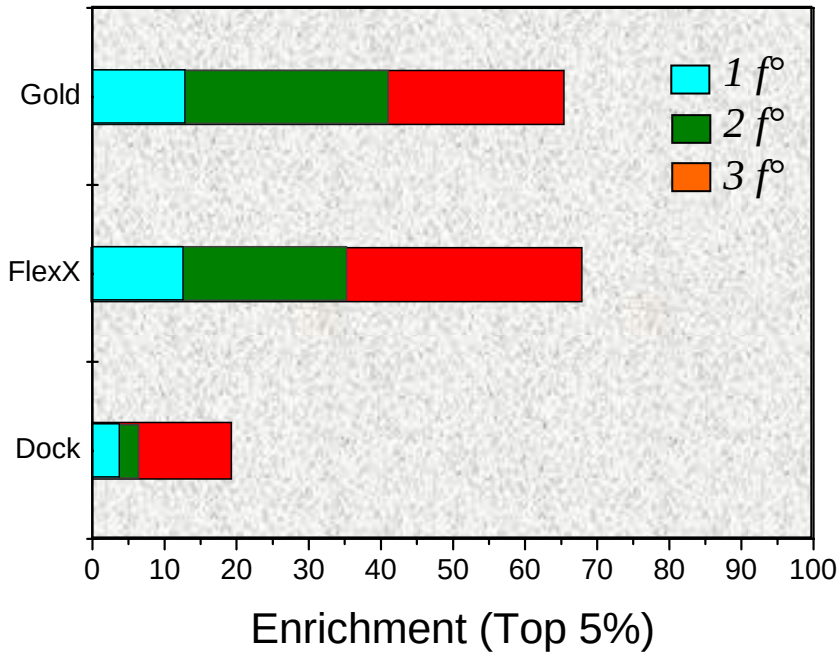
1. by consensus scoring (Charifson et al. *J. Med. Chem.*, **42**, 5100 (1999)
2. by efficient post-processing
3. by analysis of molecular diversity (descriptors, scaffolds)
4. by human inspection (3-D visualisation)

Consensus scoring

Target: LBD Domain of the Erα receptor

Library: 10 known antagonists + 990 randomly-chosen drug-like cpds





Bissantz et al. *J. Med. Chem.*, **43**, 4759 (2000)

The mean value of repeated samplings tends to be closer to the reality

(Wang et al., JCICS, 2001, 41, 422-426.)

Post-processing

Simplify output

Automated selection of hits (by rank, by score)

Remove false positives

1. by consensus scoring
2. by efficient post-processing
3. by analysis of molecular diversity (descriptors, scaffolds)
4. by human inspection (3-D visualisation)

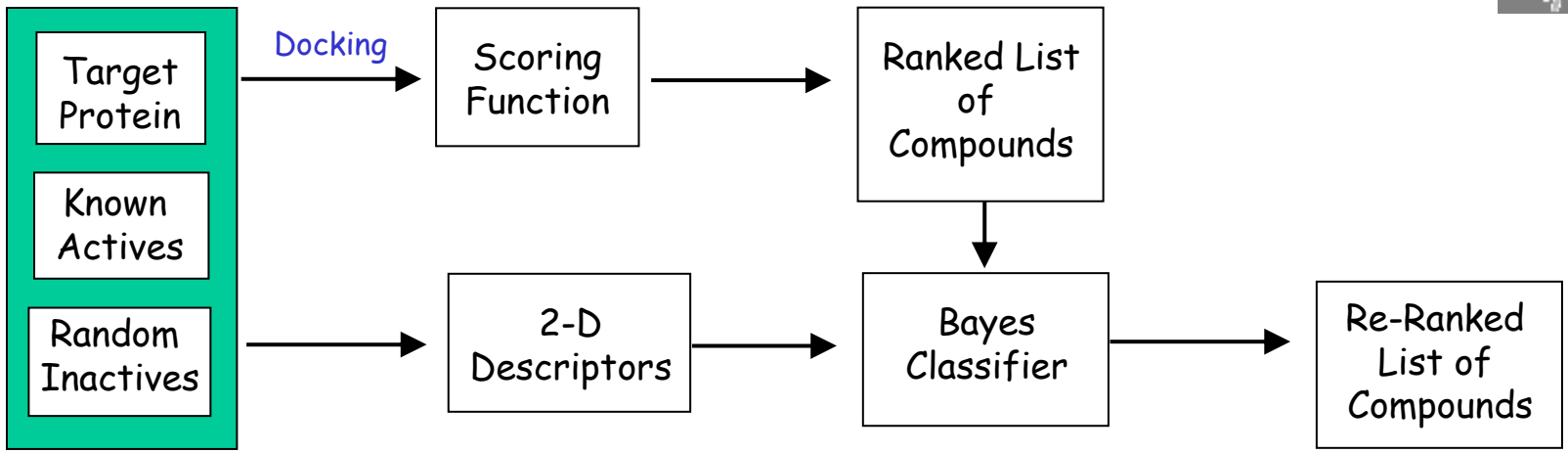
Efficient Post-processing

the most investigated area for HTVS:

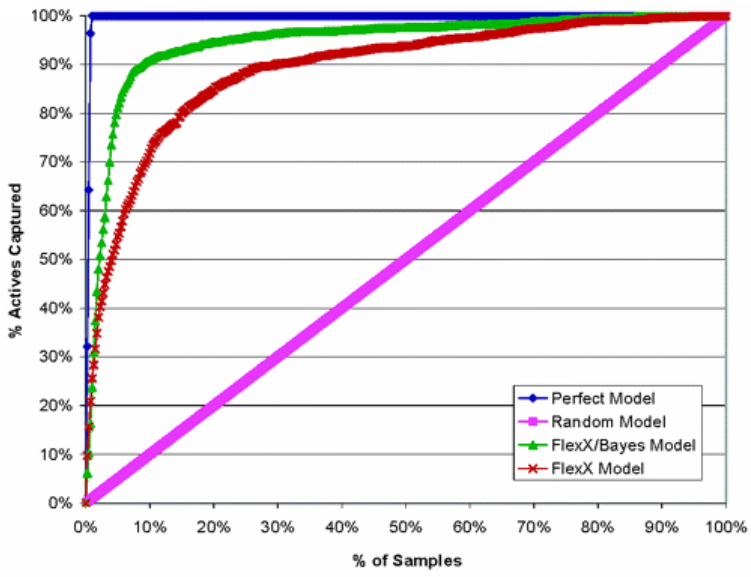
- challenging virtual hits for consistency of topological and energical descriptors in order to remove false positives
  - buriedness of ligand (Stahl et al. *J Mol Graph Model*. 1998)
  - holes along the protein-ligand complex (Stahl et al. *J Mol Graph Model*. 1998)
  - Protein side chain entropy (Giordanetto et al. *JCICS*, 2004)
- combining computational methods
  - Docking + QSAR (Klon et al. *J. Med. Chem*. 2004)
  - Refining docking poses by MM-PB(GB)/SA (Page et al. *J Comput Chem* 2006)
- using consensus docking (Paul et al. *Proteins*, 2002)
- using a multi-conformers description of the protein
  (Vigers et al. *J Med Chem*, 2004; Yoon et al. *JCICS*, 2004)
- using a topological scoring function (Marcou et al. JCIM 2007)

Docking + QSAR

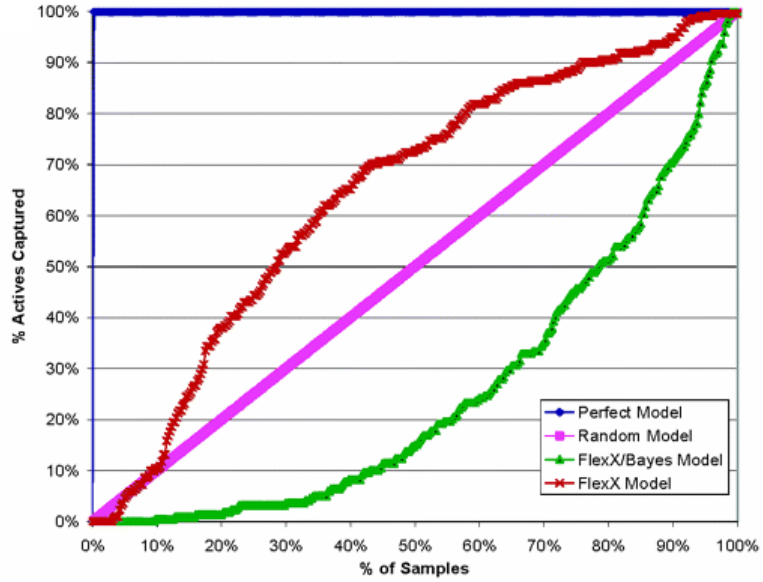
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



PTP1B docking with FlexX



PKB docking with FlexX



QSAR post-processing is efficient if the docking scores already provide some enrichment !

Klon et al. *J Med Chem*, 2004, **47**, 2743-2749

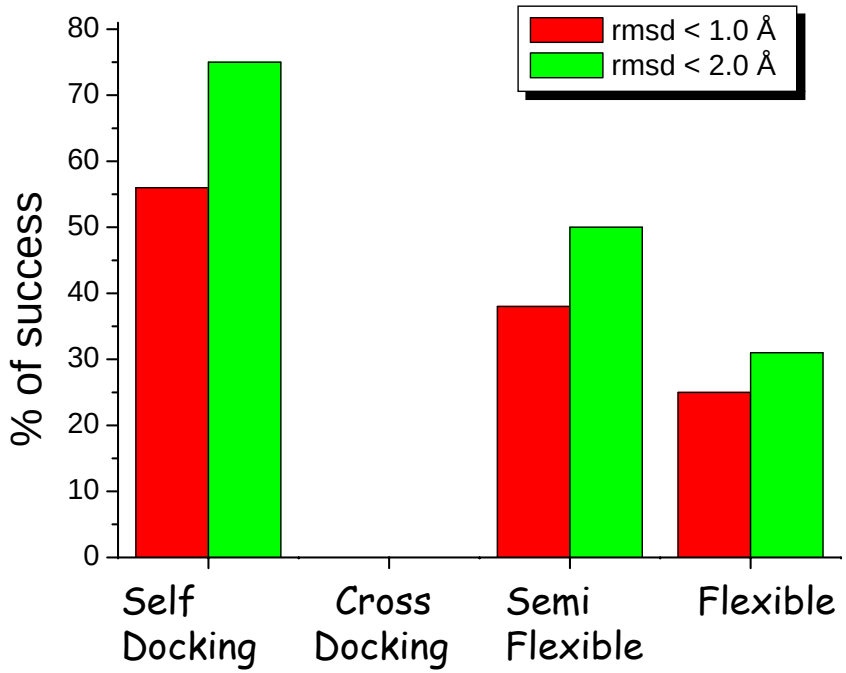
Flexible Protein-Flexible ligand docking

- Use of soft potentials (e.g. Glide)
- Generate/use several coordinates of the same target (Ensemble docking) & Merge results

e.g. FlexE, AutoDock, FITTED

- Full flexibility (MD, MC, GA) e.g. FLIPDock

Docking of HCV Polymerase inhibitors



Full flexible docking enhance false positives rate

Ensemble docking is sensitive to the choice of templates

A minimum of three templates seem to be mandatory

Moitessier et al. *J Chem Inf Model* 2008, **48**, 902-909

Refining docking poses par MM-PB/SA, MM-GB/SA

Alternative treatment of electrostatic interactions

$$\Delta G_{binding} = G_{complex} - G_{protein} - G_{ligand}$$

$$G = E_{MM} + G_{polar} + G_{nonpolar} - TS_{MM}$$

Poisson-Boltzman Eqn
Generalised Born Eqn

~ SASA

G: average free energy

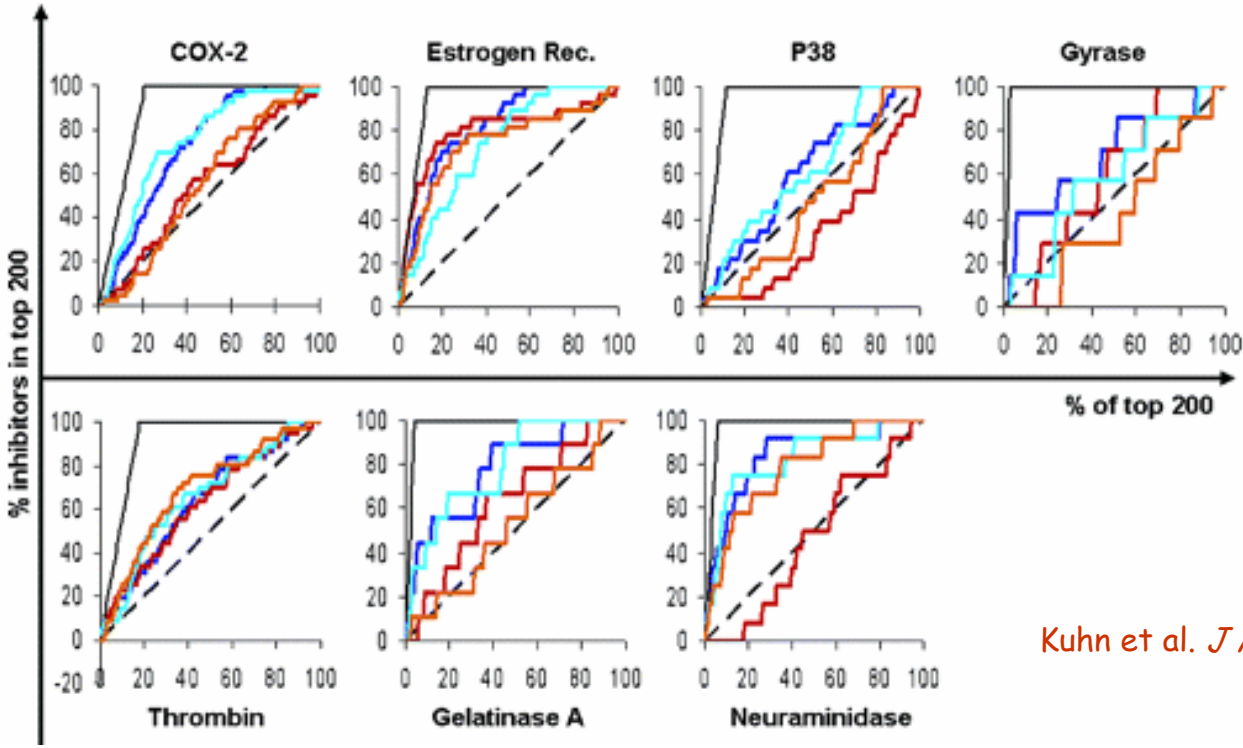
E_{MM}: average sum of molecular mechanical energy terms

(E_{bond} + E_{angle} + $E_{torsion}$ + E_{vdW} + $E_{electrostatic}$)

G_{polar} + $G_{nonpolar}$: free energy of the solvent continuum

TS_{MM}: Entropy of the solute (MD, NM)

Refining docking poses par MM-PB/SA, MM-GB/SA



Ranking
Fred/Chemscore
MM (MAB*)
MM(MAB*)-PB/SA
MM(GAFF)-PB/SA

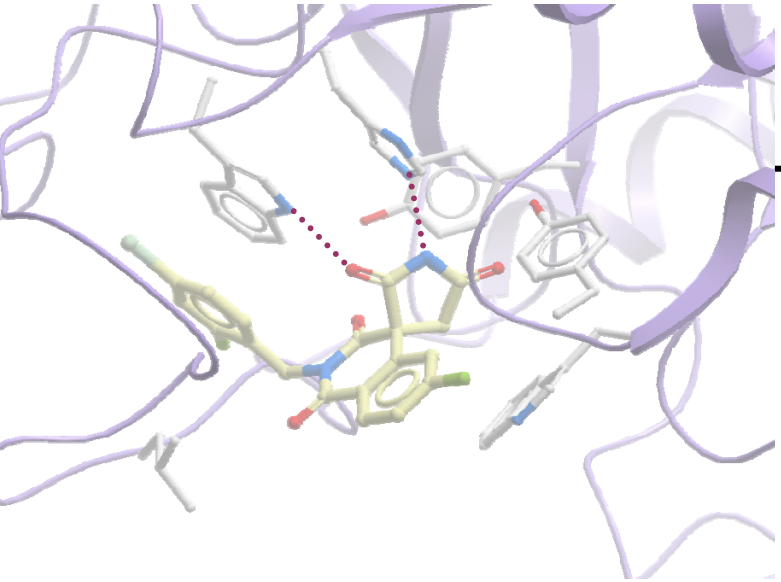
Kuhn et al. *J Med Chem*, 2004, **4**8,4040-4048

Better treatment of solvation effects
Penalize polar-apolar mismatches
Fast enough for a few hundreds of ligands
ok for a set of highly congeneric cpds

MM-PBSA benefit is target-dependent
Still unsuitable for chemically diverse ligands

Use of topological scoring functions

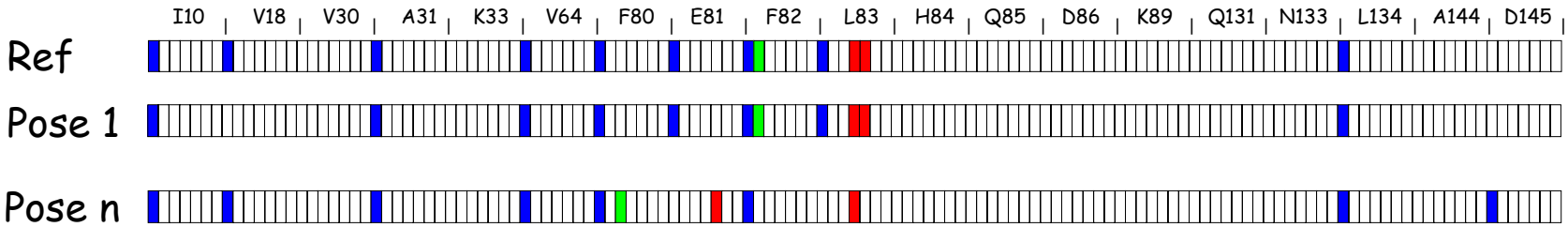
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



OEChem

Molecular Interactions

1. Hydrophobic
2. Aromatic/aromatic (face to face)
3. Aromatic/aromatic: (edge to face)
4. H-bond (acceptor → donor)
5. H-bond (donor →acceptor)
6. Ionic interaction (neg→ pos)
7. Ionic interaction (pos → neg)
8. Weak H-bonds (acceptor → donor)
9. Weak H-bonds (donor → acceptor)
10. Pi-cation interactions
11. Metal complexation



Rank poses by decreasing IFP similarity (IFP-Tc)

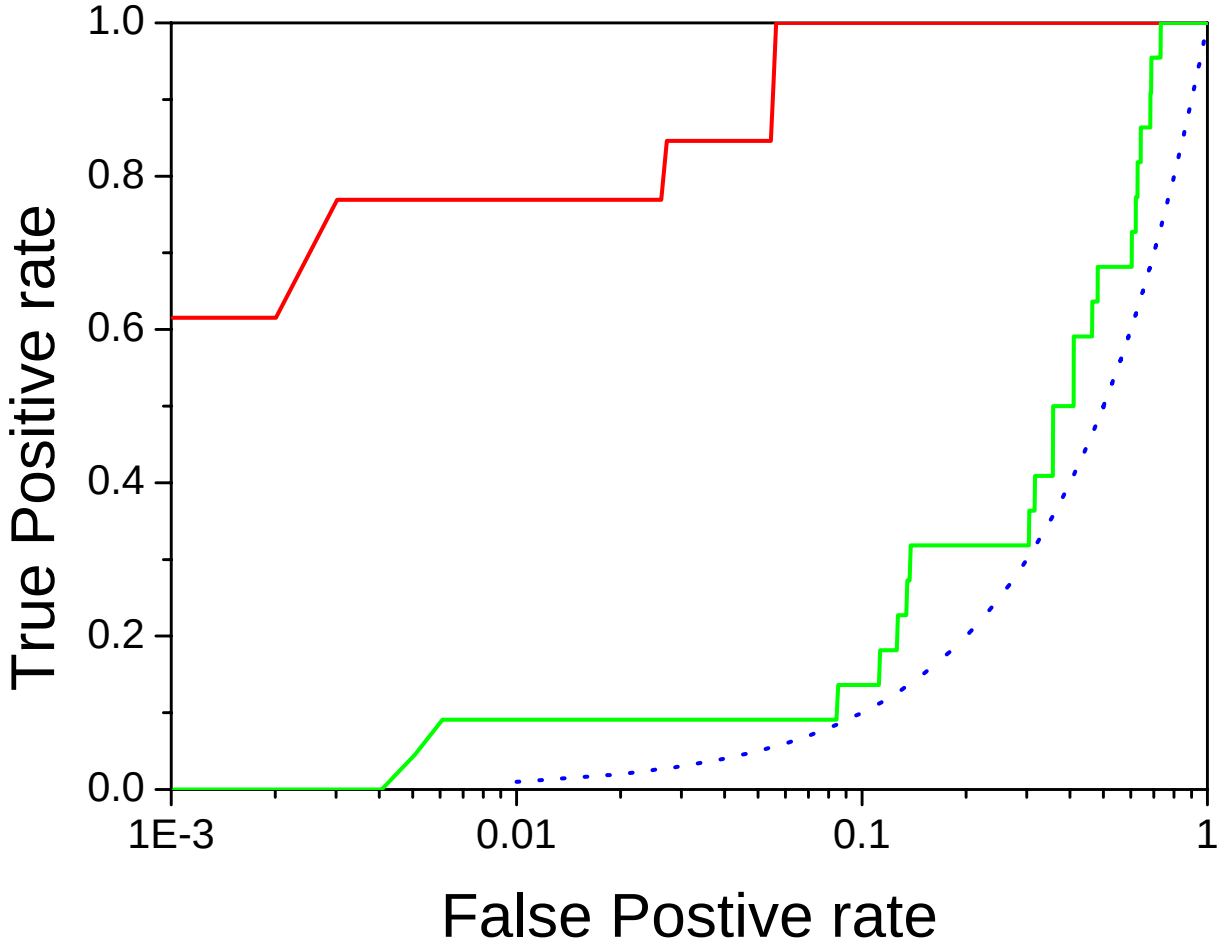
Cluster poses by IFP diversity

Marcou et al *J. Chem. Inf. Model.*, **2007**, *47*, 195-207

Use of topological scoring functions

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

GOLD docking of β2 full/partial agonists in an 'early active state model of the β2 receptor



IFP score (ROC=0.99)
Goldscore (ROC=0.61)
Random picking

De Graaf et al. *Unpublished data*

Post-processing

Simplify output

Automated selection of hits (by rank, by score)

Remove false positives

1. by consensus scoring
2. by efficient post-processing
3. by molecular diversity (descriptors, scaffolds)
4. by human inspection (3-D visualisation)

Post-processing by molecular diversity

- Virtual hits sharing the same scaffold and the same pose tends to be true positives

- Individual numerical values are less important than their distribution in ligand space

→scaffold enrichment in virtual hits

→Distribution of docking scores among classes

→Applicable to any dataset (vHTS, HTS) independently of prior knowledge

Post-processing by molecular diversity

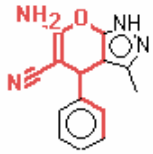
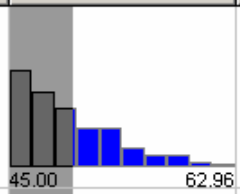
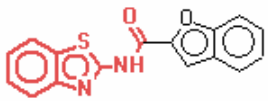
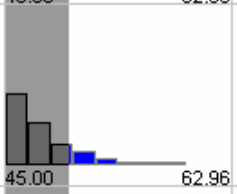
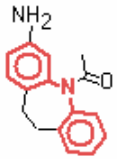
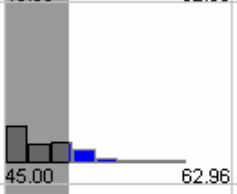
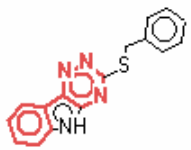
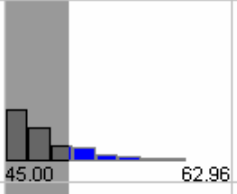
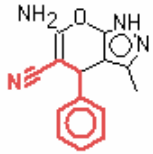
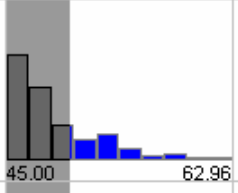
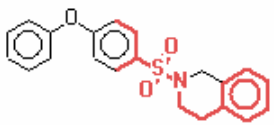
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Bioreason ClassPharmer

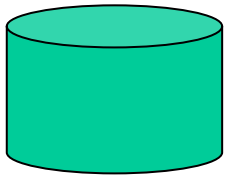
File Edit List View Window Help

Find Compound ID: Search... Find Substructure: Search...

Classes | Compounds | Attributes

Class	ID	Scaffold	Compounds	Mdn(Gold_Score)	Dst(Gold_Score)	Avg(Gold_Score)	PrbDst(Gold_Score)	%Rng(Gold_Score)	Max(Gold_Score)
Class	1		429	48.9	 45.00 62.96	49.95	1.0	40	62.96
Class	6		196	47.08	 45.00 62.96	48.08	0.59	22	59.24
Class	11		128	48.1	 45.00 62.96	48.61	1.0	28	58.53
Class	16		161	47.68	 45.00 62.96	48.45	0.97	26	59.2
Class	17		348	47.82	 45.00 62.96	49.11	1.0	31	62.91
Class	38		57	47.99		48.89	1.0	29	59.59

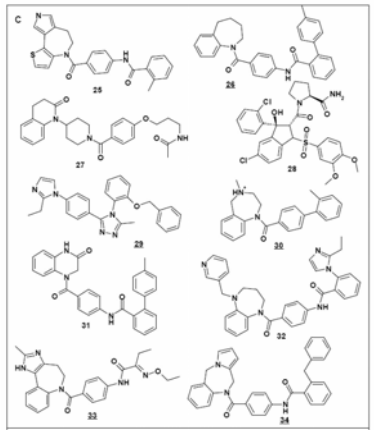
Post-processing by molecular diversity



Bioinfo Lib.
(550K drug-like cpds)

Random selection

(990 cpds)

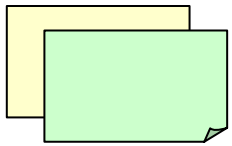


True V1aR antagonists.
(10 cpds)

Bissantz et al. *Proteins*, **50**, 5-25 (2003)

Test Dataset
(1,000 cpds)

Docking to the V1a receptor (FlexX, GOLD)

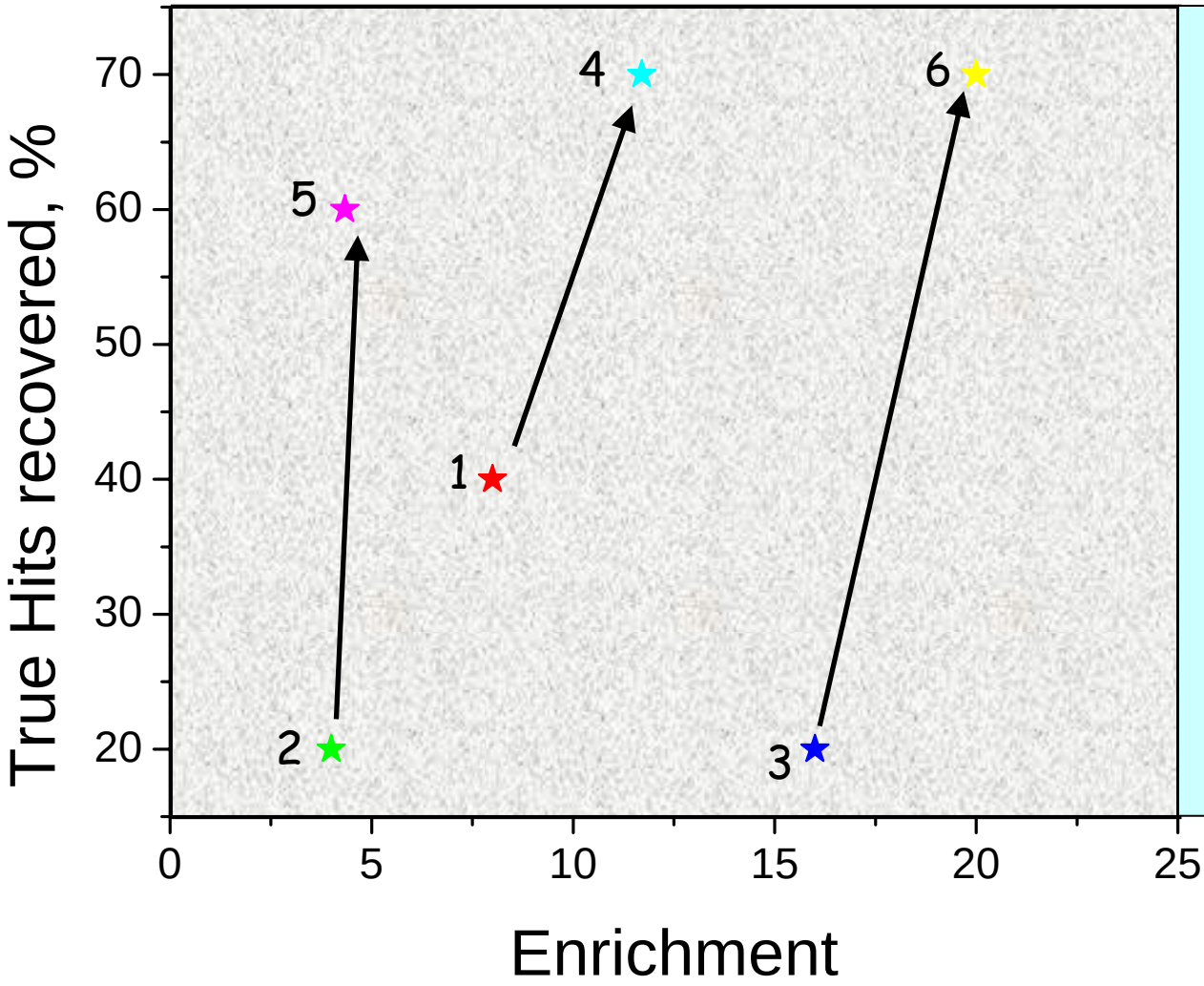


Docking scores

How to prioritise the selection of true hits ?

Post-processing vHTS results

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE



1. FlexX
top 5%

2. Gold
top 5%

3. FlexX + Gold
Common top 5%

4. FlexX
ClassPharmer processing

5. Gold
ClassPharmer processing

6. Gold + FlexX
ClassPharmer processing

Enrichment: Enrichment in true hits over random picking

Post-processing

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Simplify output

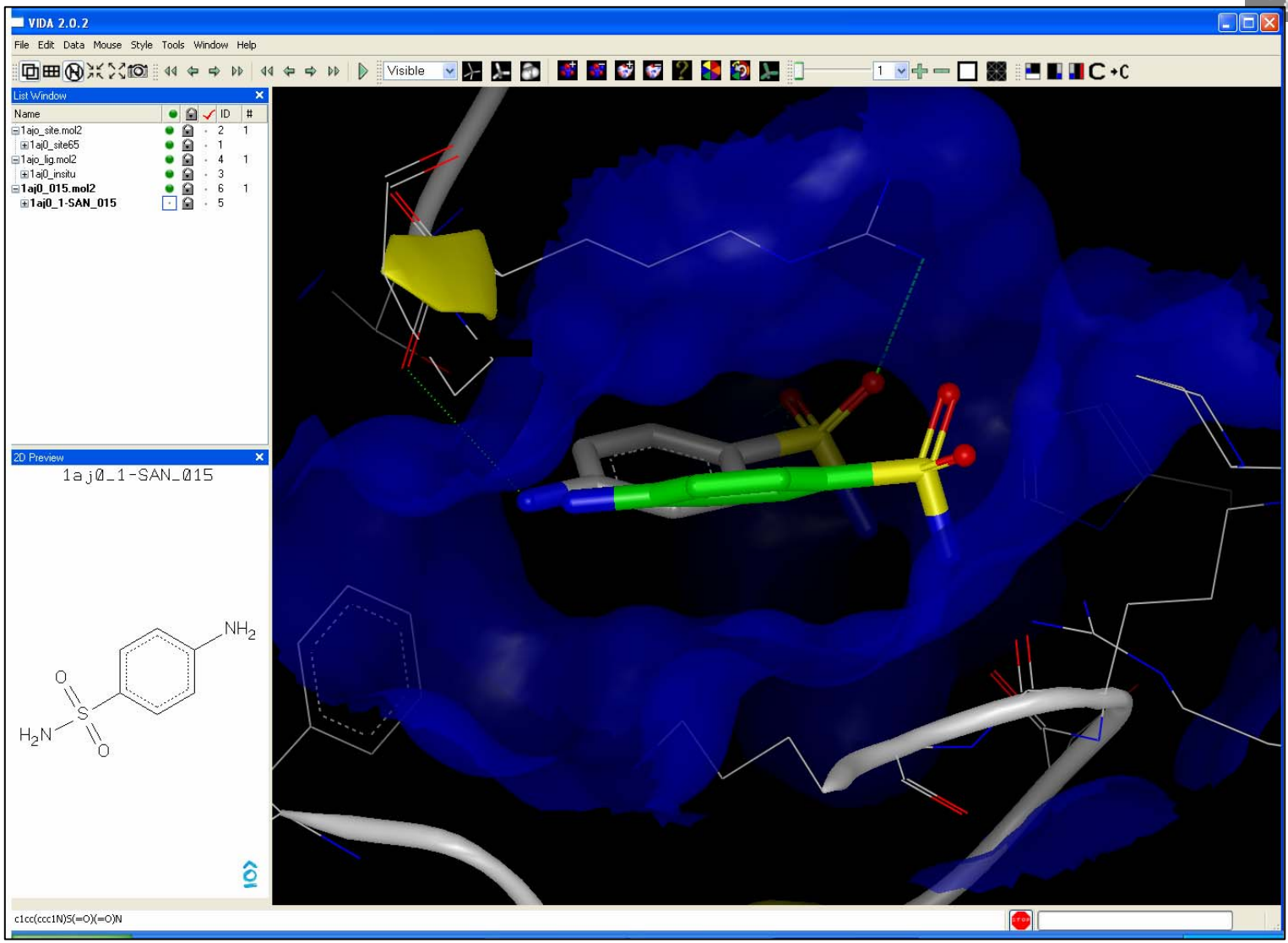
Automated selection of hits (by rank, by score)

Remove false positives

1. by consensus scoring
2. by efficient post-processing
3. by molecular diversity (descriptors, scaffolds)
4. by human inspection (3-D visualisation)

Post-processing: Visual Inspection

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE



No post-processing tool outperforms the brain of an experienced computational chemist !

Protein-based virtual screening: Current status

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

What is possible ?

- ✓ Discriminate true hits from randomly-chosen drug-like ligands
- ✓ Achieve true hit rates of ca. 10%
- ✓ Retrieving about 50% of all true hits
- ✓ Prioritizing ligands for synthesis and experimental screening
- ✓ Using virtual screening for hit finding

What remains to improve ?

- ✓ Use of homology models
- ✓ Predicting the exact orientation (Δ rmsd :2Å)
- ✓ Predicting the absolute binding free energy (ΔG= 7-10 kJ/mol)
- ✓ Discriminating true hits from "similar inactives"
- ✓ Catching all hits
- ✓ Protein flexibility, screening multiple targets
- ✓ Using virtual screening for lead optimization
- ✓ Pre and post-processing of vHTS