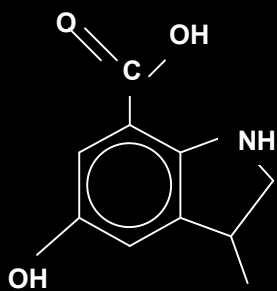


Drug Design Bioinformatics



Content:

- Drug discovery process
- Drug design and virtual screening
- Screening based on the structure of the ligand
- Screening based on the structure of the receptor
 - ✓ Protein structure prediction
 - ✓ *Defining druggable pocket - protein docking*
 - ✓ *Small compound libraries - ADME-Tox*
 - ✓ *Docking/scoring*
 - ✓ *Lead optimization*

Bring one molecule to market

- 15-20 years
 - 500 M €
- In Silico approaches could save 2-3 years and 200 M €

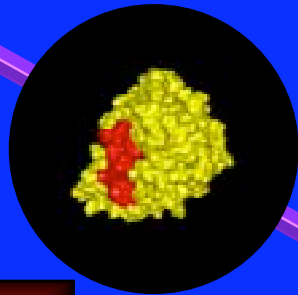
Cheng A et al., **J Comput Chem** 23:172-183 (2002)

Michelson S and Scherrer, www.currentdrugdiscovery.com, April 2003

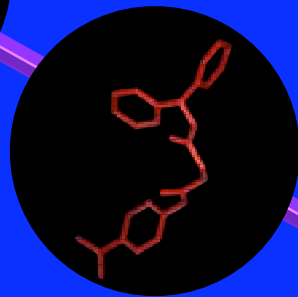
Drug Discovery & Development



Identify disease



Isolate protein involved in disease (2-5 years)



Find a drug effective against disease protein (2-5 years)

Preclinical testing (1-3 years)

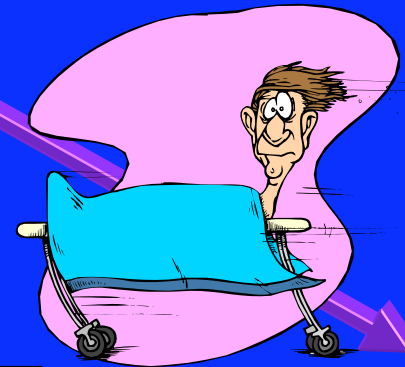


Animal studies

Human clinical trials (2-10 years)



Scale-up



FDA approval (2-3 years)



Formulation

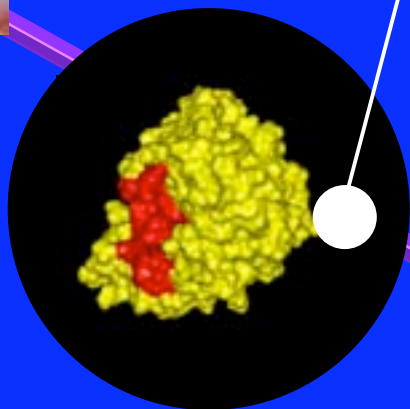


Technology and the drug discovery process



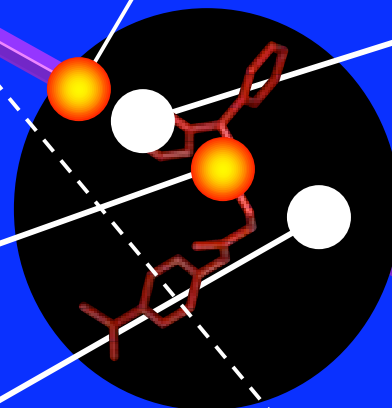
GENOMICS, PROTEOMICS & BIOPHARM. (Xray, NMR...)

Produce many more targets and "personalized" targets



HIGH THROUGHPUT SCREENING

Screening compounds to find activity against a target



VIRTUAL SCREENING

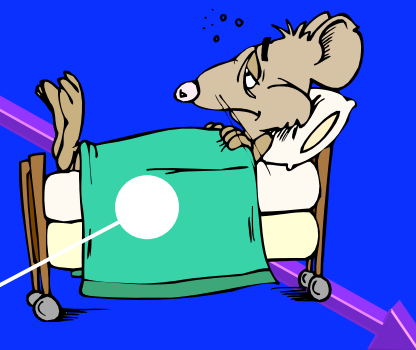
Identify hits on a computer

COMBINATORIAL CHEMISTRY

Produce vast numbers of compounds

MOLECULAR MODELING

Computer graphics & modeling to help the process



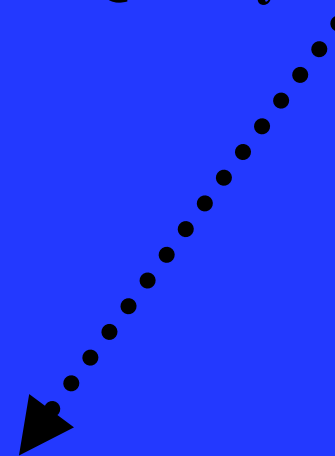
IN VITRO & IN SILICO ADME MODELS

These models start to replace animal testing

Drug discovery



Quality vs quantity



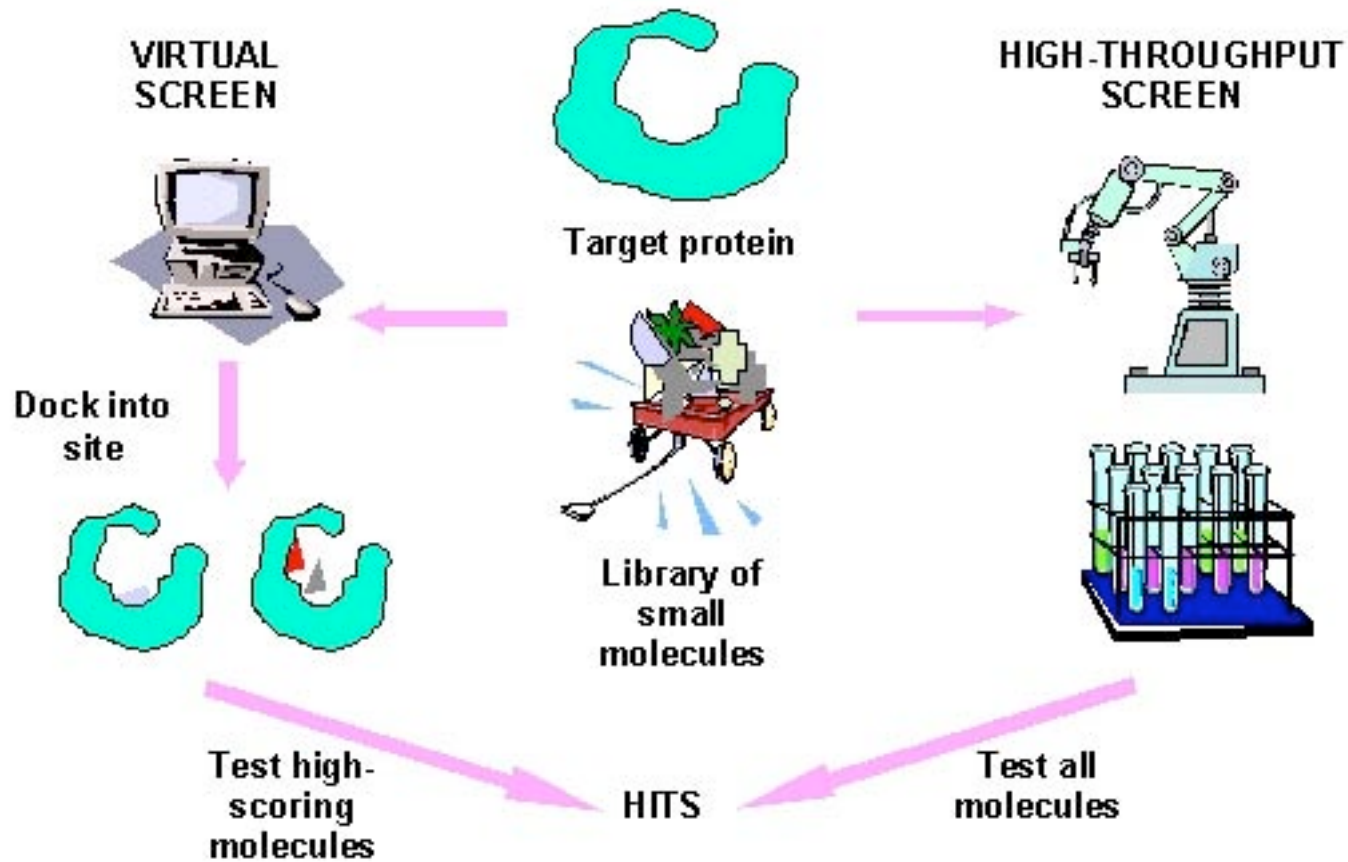
1970 : Random screening - Chemistry

1980 : Structure-based drug design

1990 : High-throughput screening (HTS) + Chemistry

2000 : Structure-based design (VLS) + HTS + Chemistry

Protein (3D known or unknown)



In Silico = 1-3M
compounds/week

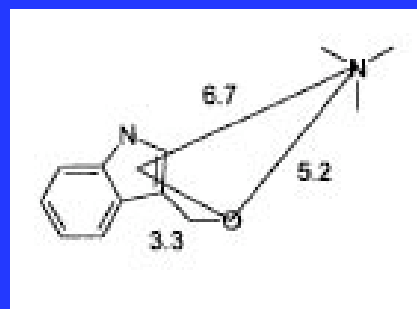
Experimental = 3000
compounds/week

Drug Design

(3D structure protein unknown)

Ligand-based drug design: possible methods

- Ligand-based lead finding



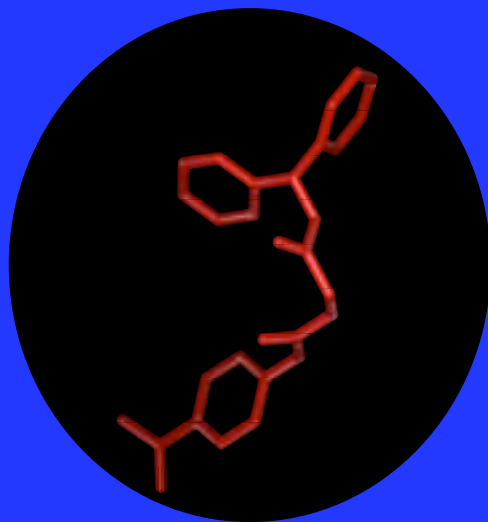
Klabunde and Hessler, *ChemBioChem* 3:928-944 (2002)

Drug Design 2D/3D: lead finding

- ✓ Protein 3D structure not well defined
- ✓ Natural ligand can provide good starting point
- ✓ Structure activity relationships (SAR) can be derived from natural ligand and analogues
- ✓ The resulting pharmacophore models can be used for virtual screening

Pharmacophore here = a set of features that is common to a series of active molecules

Structure-based (receptor)



Reviews: The process of Structure-based drug design

Anderson AC.

Chem and Biology, 10:787-797 (2003)

**Docking and scoring..Nature Drug Discovery review, vol 3, november
2004.**

Kitchen DB et al.

S-B VLS



Receptor



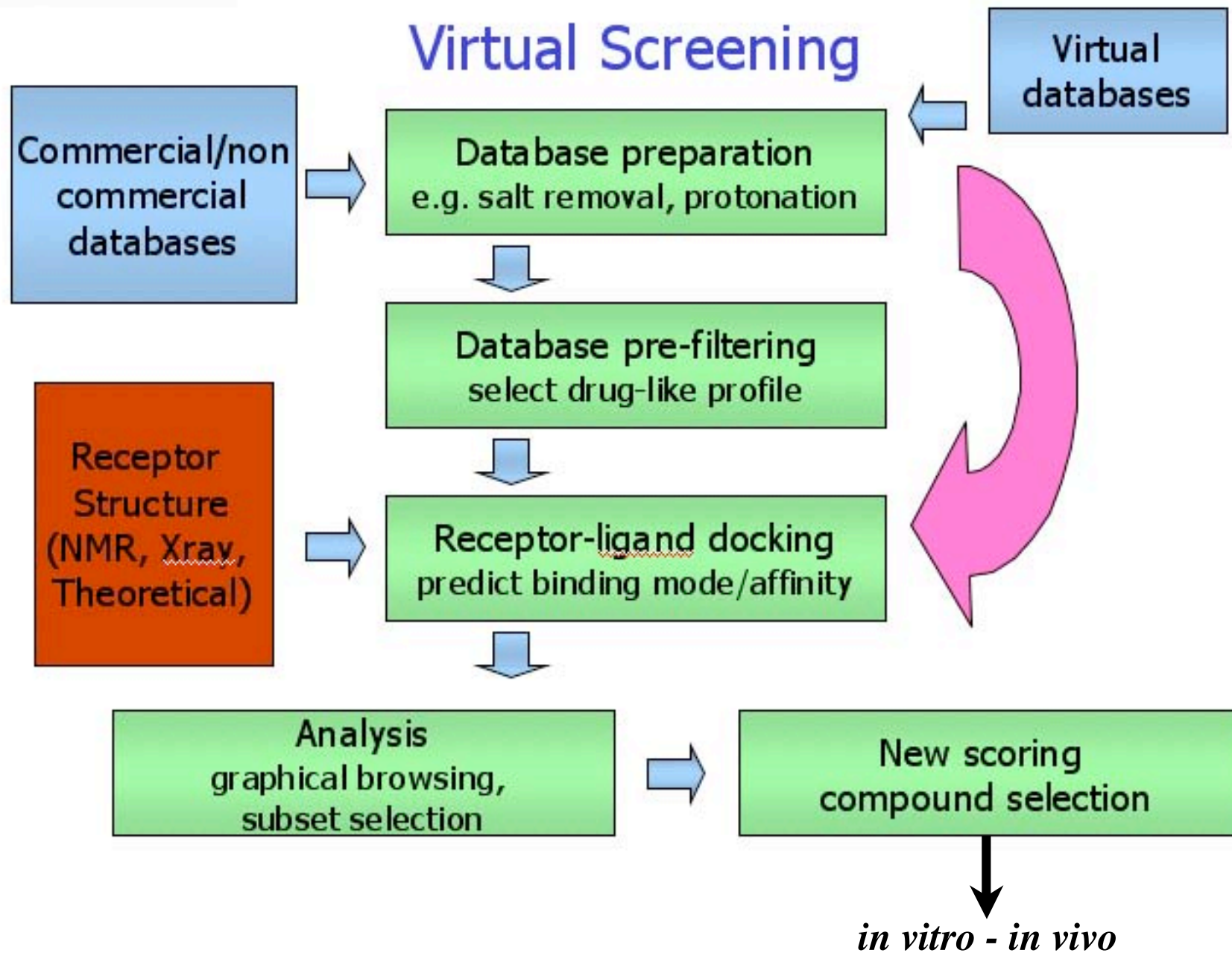
Compounds
libraries
proprietary
ACD...

Ligands

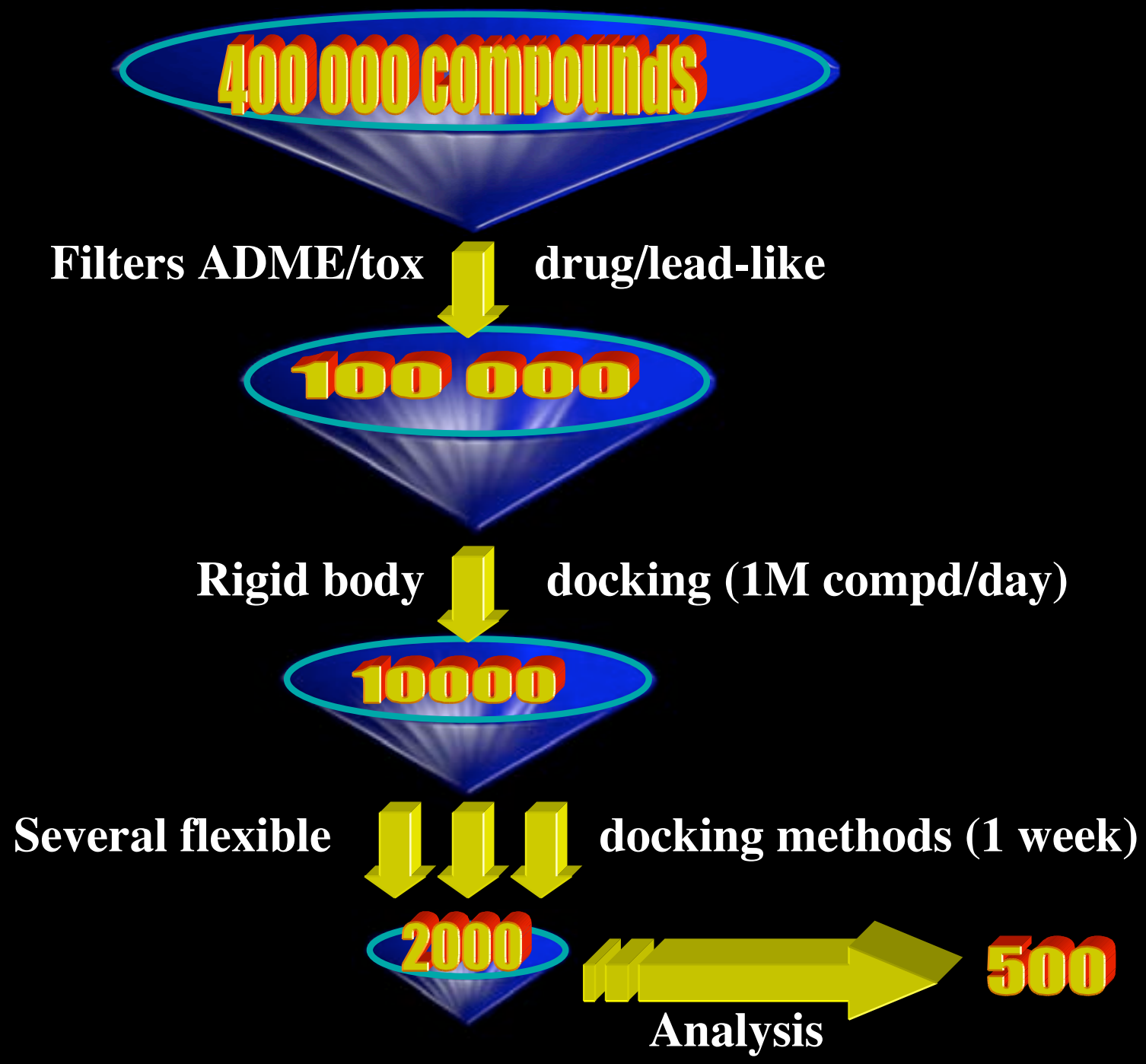
Match receptor and ligand (Docking)
Accuracy & speed

Score the ligands

Waskowycz et al., *IBM Systems Journal* 40:360-376 (2001)
Langer and Hoffmann, *Current Pharmaceutical Design* 7:125-133 (2001)



S
—
B
V
L
S



Proteins

Need a 3D structure, X-ray, NMR...homology model
Loop simulation and/or full MD simulation...
Analysis of the PDB file.....missing residues...
Run structure validation methods.....

Structural Predictions

- **Comparative Model Building (Homology)**
- **Threading (not for drug design)**
- ***ab initio* or *de novo* (not for drug design)**

Utility of homology models in the drug discovery process

Alexander Hillisch, Luis Felipe Pineda and Rolf Hilgenfeld

DDT Vol. 9, No. 15 August 2004

The paper shows that homology models can be used for in silico screening

Explore “flexibility of the receptor”

Molecular dynamics

Solve Newton's equations with the help of an energy function and forcefield parameters.

$$\vec{F}_i = m_i \vec{a}_i$$

$$\vec{F}_i = - \frac{d\vec{E}(\mathbf{r}_i, \dots, \mathbf{r}_n)}{d\mathbf{r}_i(t)}$$

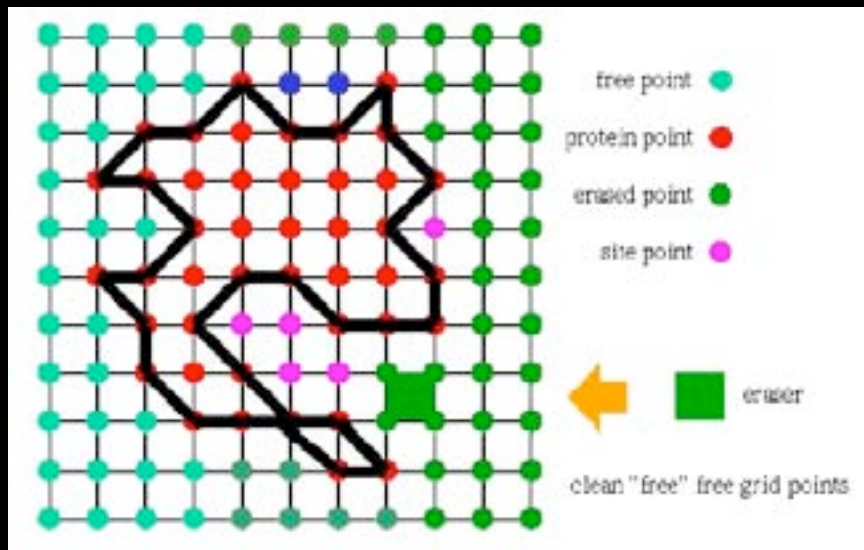
$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2}$$

$$\vec{r}_i(t + \Delta t) - \vec{r}_i(t) = \frac{d\vec{r}_i(t^*)}{dt} \Delta t = \vec{v}_i(t^*) \Delta t$$

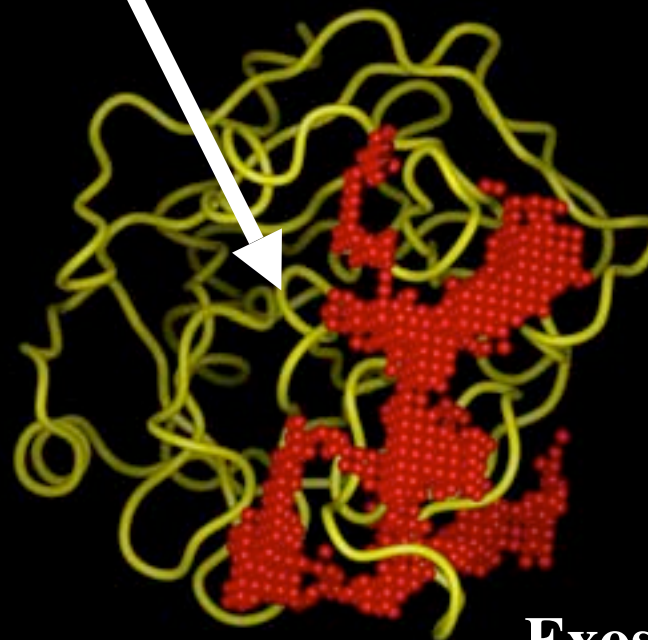
Proteins

3D OK, definition of druggable pocket

- Defining the binding pocket



Active site



Exosite

Accelrys

- Defining the binding pocket

Interaction site:

Coat one region of the protein with polar and steric probes

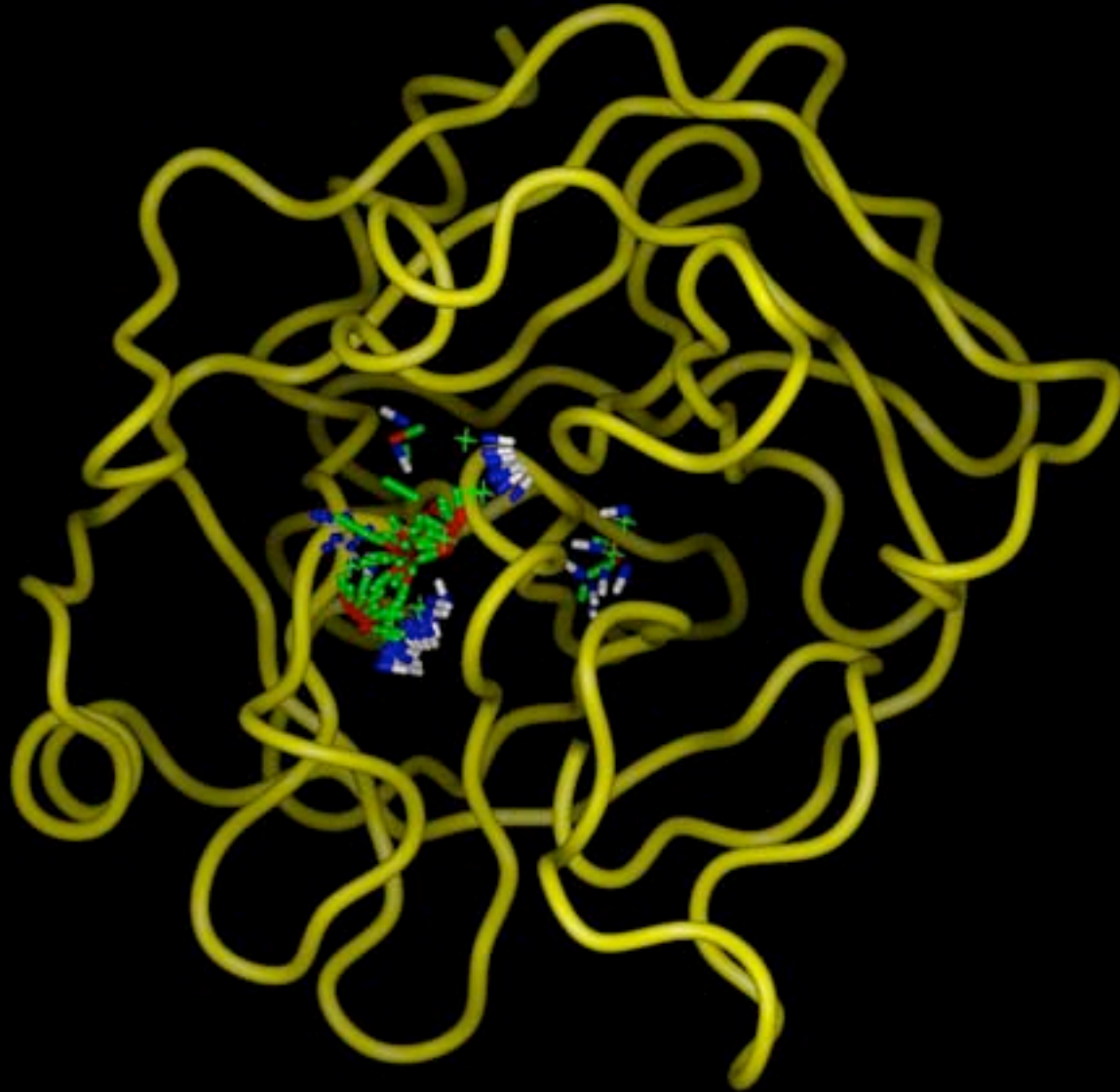
Define positions for:

H-donor

H-acceptor

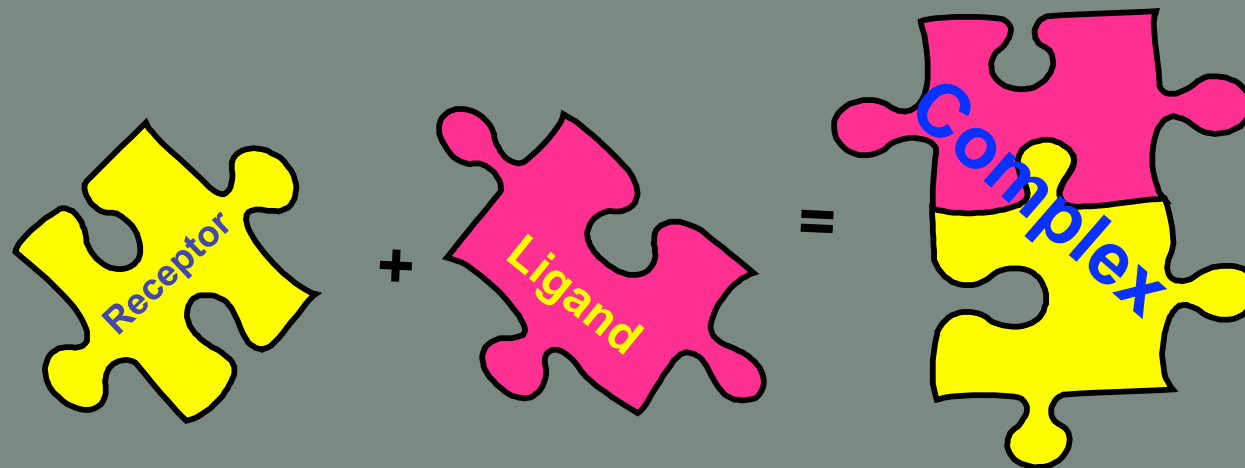
Lipophilic-aliphatic

Lipophilic-aromatic



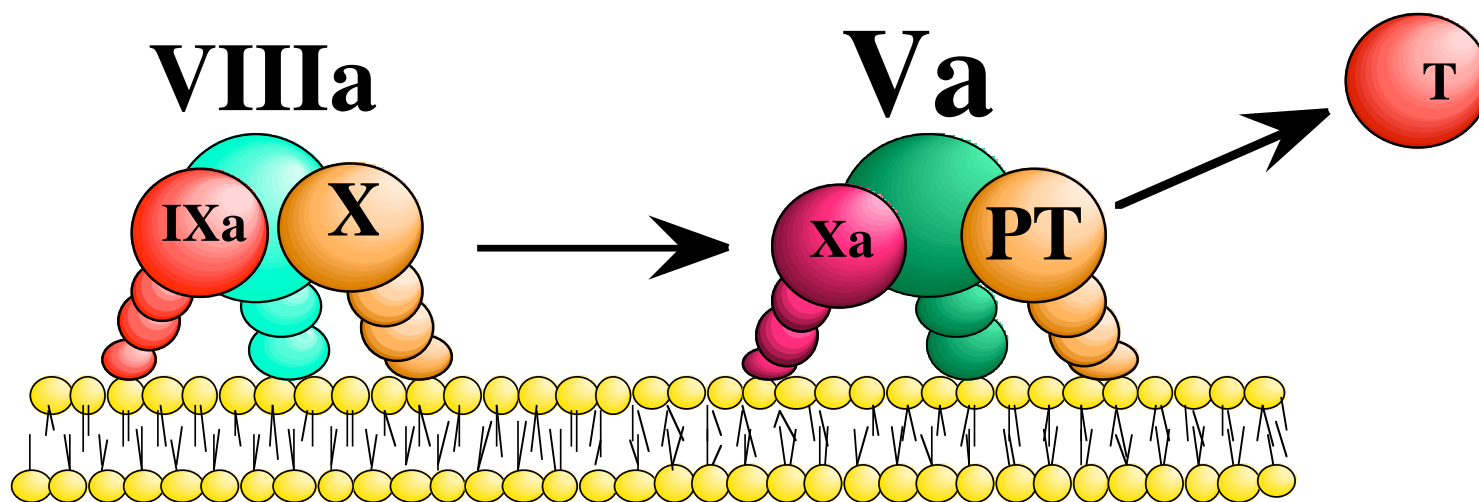
In some cases, to inhibit protein-protein interactions, it is important to have a structure of the complex

As such, it can be valuable to predict protein-protein interaction via docking



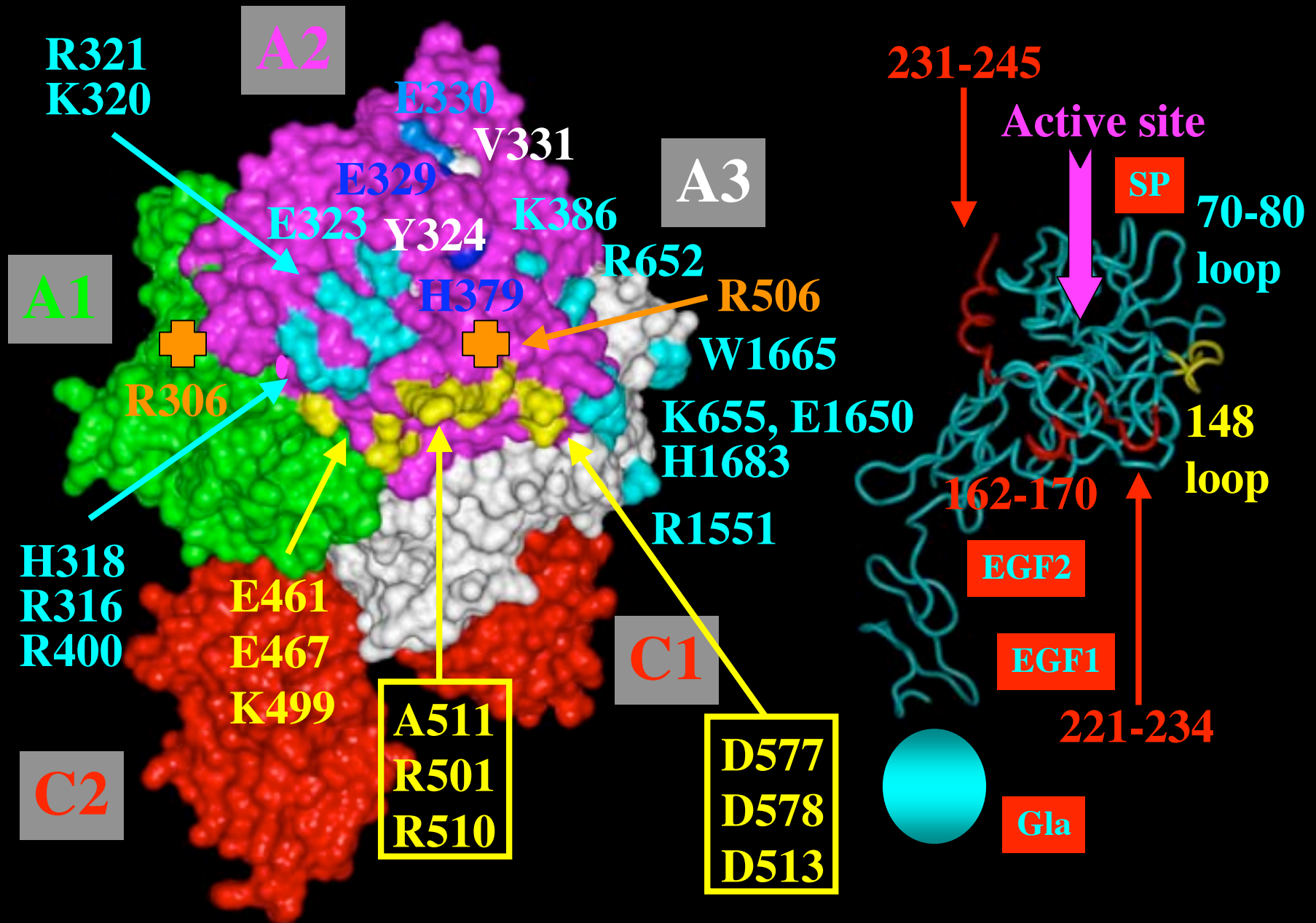
Search for new anticoagulant compounds:

Homology Modeling + Mutagenesis
+ Protein Docking + VLS + Chemistry



Protein-protein interaction and drug design

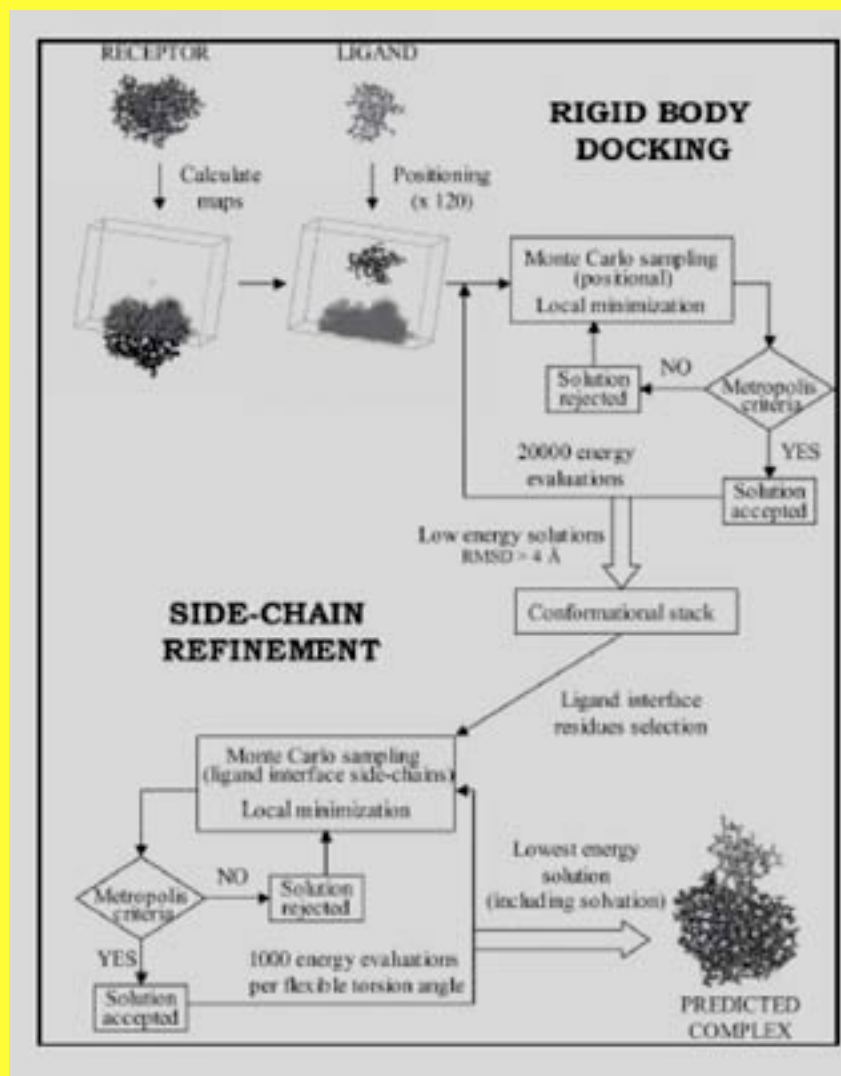
- J. Med Chem. 45: 1543-1558 (2002)
- J. Mol. Recognition vol 15: 405-422 (2002)



FVa (homology)

FXa (X-ray)

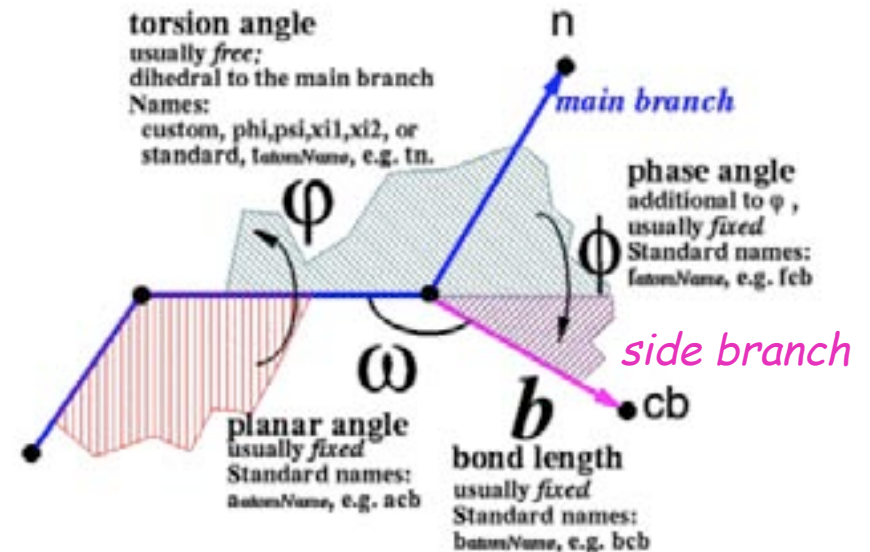
Soft ligand-
protein docking
in internal
coordinates
(ICM)
(Juan - Rubben)



Internal Coordinate Mechanics (ICM)

✓ Types of variables:

- b - bond length
- ω - bond angles
- φ - torsion angles
- ϕ - phase angles



Cartesian space, large number of degrees of freedom. If one is switching to Internal coordinate system operating in torsion space and fixing bond lengths and angles, then the number of degrees of freedom is reduced.

Cartesian description requires 3 variables (x,y,z) per atom. Internal coordinates description uses bond lengths, planar angles and torsion angles instead. Since bond lengths and planar angles are essentially rigid, one can consider them as constants and only allow torsion angle changes (rotation around the bonds). This reduces the dimensionality of the conformation space at least threefold...(in practice even more)...

Rigid body docking step

✓ The ligand atomic tree is randomly positioned inside the grid (common origin, random orientation of the ligand).

✓ The algorithm:

- 120 starting orientations of the ligand (or full search):



- An imaginary dodecahedron is created around the ligand
- 20 vertices of the dodecahedron are sequentially oriented toward the receptor
- Six 60° rotations are made around the axis defined by the centers of mass of receptor and ligand

- For each starting conformation

- Sampling only 6 positional variables of ligand by pseudo-Brownian method (with random translation and rotation)
- Local "rigid" energy minimization (up to 200 steps of conjugate gradient)
- Acceptance via Metropolis criteria on the total energy (T= 300K for bound or 5000K for unbound)
- Stop after 20000 energy evaluation (for each 120 starting points if the docking is not global..if global much more)

- In the set of all the accumulated conformations

- Remove geometrically similar (RMSD < 4 Å)

Pseudo-Brownian step:

Calculate new values of 6 variables from Cartesian coordinates of the first three atoms and rebuild the molecule according to these new variables.

Internal coordinates are not natural for intermolecular relative positions (so we need to move three representative atoms in Cartesian coord.)

Metropolis Monte Carlo

1. Get initial conformation C_{old} and calculate $E_{old}(C_{old})$
2. $C_{new} \leftarrow$ new variables (new trial conformation)
3. calculate $E_{new}(C_{new})$
4. if $E_{new} < E_{old}$
 accept C_{new}
 else (if energy is higher)
 compute

$$\exp\left(\frac{-\Delta E}{kT}\right) > A$$

if this number is greater than a random number A such that $0 < A < 1$, accept the new state even though it is higher in energy, if less, discard C_{new}

5. Repeat for N steps

Higher $T \rightarrow$
Samples more widely,
Larger acceptance ratio
(if T high one
Accepts more
Higher-energy solutions

The interaction energy function used for this rigid docking

$$E = E_{Hvw} + E_{Cvw} + E_{el}^{solv} + E_{hb} + E_{hp}$$

The interaction energy between the **ligand atoms** and the **grid points** are calculated and selection of the move according to metropolis.

When the receptor and the ligand are in complex, re-ranking can be done from ASP method, where the solvation of both, the receptor alone and the ligand alone is compared to the solvation of the complex. ASP method assumes some solvation values for each atom type following:

$$E_{solv} = \sum_i \sigma_i a_i$$

a_i - solvent - accessible surface areas

σ_i - solvation energy densities derived from
the water - vacuum transfer energies

Refinement of ligand side-chain conformation: introducing flexibility on the ligand side

- ✓ For each final conformation from rigid body docking step
 - Select side-chain torsion angles of the ligand surface residues in the vicinity of 4 Å of the receptor
 - Change these angles using Biased Probability Monte Carlo in each random step and change 6 positional variables in each random step but with loose restraints are imposed on the positional variables of the ligand to keep it close to the starting conformation

Biased Probability Monte Carlo (BPMC)

- The idea:
To sample with larger probability those regions of the conformational space which we know a-priori (on the average) are highly populated regions.
- Makes a step to a new random position independent of the previous position, but according to the predefined continuous probability distribution
- The local conformational preferences are represented by multidimensional ellipsoidal zones.
- The positions, sizes and probabilities of preferred zones in ϕ - φ and χ subspaces (for example) were calculated statistically. ("Ramachandran maps")

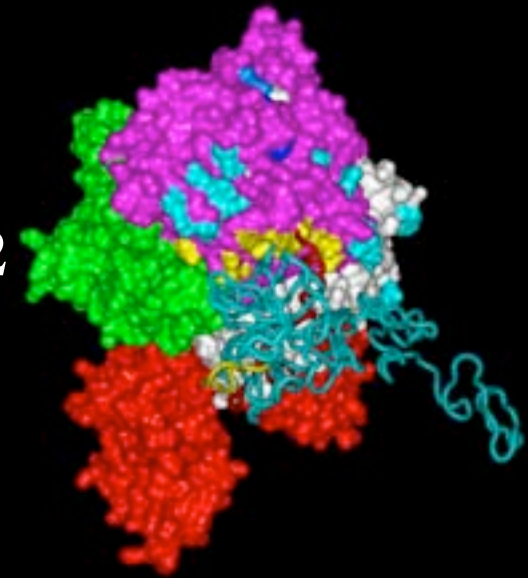
The energy function optimized is again:
between the receptor (grid)
but computed differently (not soft potentials)

$$E = E_{\text{int}}(\text{ligand}) + E_{H_{vw}} + E_{C_{vw}} + E_{el}^{\text{solv}} + E_{hb} + E_{hp} (+E(\text{solv}) + E_{\text{entropy_side_chains}})$$

Side chain entropy incorporated into energy calculations by relating the entropy with accessible surface. The accessibility of some reference atoms at the tip of a side-chain may reflect the number of reachable states



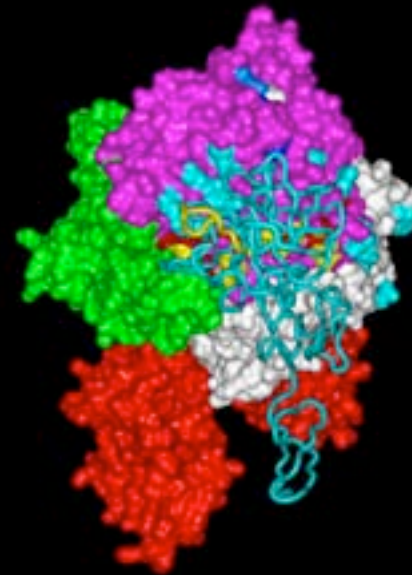
Model 1



Model 2



Model 3



Model 4

**Receptor
Pocket...
END**

next...small molecules

Small compound databases

Drug-like/lead-like

Focused

1D or 2D to 3D

Small molecule databases:

French, provided by UMR7081 (Pr. Hibbert, Strasbourg): 7000 ? (available ?)

MDL/ACD: 450000 (less than 200000 drug-like) (in 2002-2003)

See info at: <http://www.inist.fr/titanesciences/>

NCI: 250000 (in 2003)

<http://cactus.nci.nih.gov/ncidb2/download.html>

GenomeNetJapan: 8000 (in 2003)

http://bioserv.rpbs.jussieu.fr/RPBS/html/fr/T0_Home.html

<http://www.genome.ad.jp/dbget/ligand.html>

Cambridge (www.ccdc.cam.ac.uk): 250000 (in 2003), many are not drug like and the format is not convenient

Drug companies: several millions (usually not available)

Other place: ChemDiv: <http://www.chemdiv.com/> (600000 drug-like compounds), about 10000 new compounds/month).....

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

Prediction of drug-like properties !

Observation	Reference
Newer approved drugs may be more likely to have unrecognized ADRs than established drugs	Lasser et al (2002) <i>JAMA</i> 287 :2215.
Twenty drugs pulled from the market for safety reasons in the past 20 years; most were because of hepatotoxicity, followed by blood disorders or drug interactions	Thomas (2002) <i>Toxicol & App Pharmacol</i> 183 :81.
Adverse drug reactions reported to be the leading cause of death in the US	Lazarou et al (1998) <i>JAMA</i> 279 :1200.

Table 1. The growing problem of adverse drug reactions.

- **Many drug candidates fail in clinical trials due to pharmacological and toxicity issues**
- **Choosing 'drug-like' compounds. Drug-like molecules exhibit favorable absorption, distribution, metabolism, excretion, and toxicological (ADMET) parameters.**
- **There are several filtering levels (high, low) depending the stage of the project (research of hit, lead.....)**
- **Methods to select drug-like compounds: They include simple counting methods such as the rule of five, functional group filters (e.g. for reactive groups).. More elaborated filters for specific ADME properties are being developed, such as filters for prediction of membrane permeation, metabolic clearance.....**
- **Computational techniques used to identify druglikeness include neural networks, genetic algorithms... The aim of the GA or other method is to identify sets of weights that will maximize the discrimination between active and inactive molecules when the molecules are represented by more than one feature.**

Is There a Difference between Leads and Drugs? A Historical Perspective

J. Chem. Inf. Comput. Sci. 2001, 41, 1308-1315

Lead structures are ligands that typically exhibit suboptimal target binding affinity. Leads should display the following properties:

- relatively simple chemical features
- membership to a well-established SAR (structure-activity relationship) series
- favorable patent situation
- good ADME properties

Often, the druglike space is defined by LogP, MW, H-bond donors, and H-bond acceptors

Many library design programs include filters for the “Lipinski rule of five”

Should the aim of such programs be to identify drugs (not leads), then the use of such filters is appropriate. However, most lead discovery projects applied these filters ad litteram, i.e., using MW < 500 and CLogP < 5, regardless of the fact that these values had been obtained from analyzing drugs, not leads.

Thus, for instance, leadlike libraries should be designed with lower MW and lower LogP profiles, as opposed to druglike libraries.

When designing leadlike libraries care should be exercised not to exceed the following property values: 450 Dalton in MW, CLogP lower than +4.5 but higher than -3.5, no more than 4 rings, no more than 10 nonterminal single bonds, no more than 5 hydrogen-bond donors, and no more than 8 hydrogen-bond acceptors.

(this does not apply to drugs for the central nervous system (LogP and Hb-donors and acceptors should be different))

Several authors have analyzed all small molecule drugs launched in the year 2000. They have found that most drugs have been derived from lead structures **surprisingly closely** related to the final drug compound. Comparing popular properties of the lead compounds to the final marketed drugs the authors have found that most of the lead-drug pairs analyzed are within 25% of Mw for instance.... This finding suggests that **successfully optimized leads may need to be much more drug-like than one commonly assumes.**

Thus how to generate a good starting library is still under debate

ADME/tox goal: provides a preliminary prediction of the in-vivo behavior of a compound to assess its potential to become a drug.

Clearly, many factors contribute.

Can be performed *in vitro* but also *in silico*

***In silico*, the predictions are usually based on the chemical structure alone**

Simple filters remove counter ions, compounds with Zn, Al, Hg...or some toxic groups

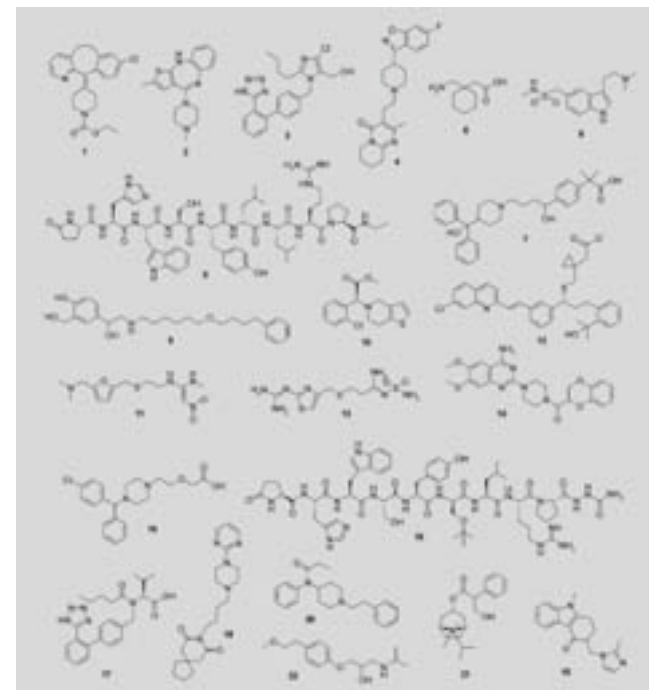
Filters for oral absorption (Lipinski rules)

drug-like filters

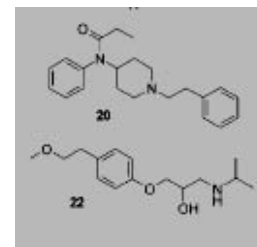
Remove poor
Compounds with
ADME/Tox models

(ADME/Tox=absorption, distribution, metabolism,
Elimination and toxicity)

Li AP., **DDT** 6: 357-366 (2001)
Muegge I, **Med Res Rev.** 23:302-21 (2003)



Human intestinal
Absorption
Serum protein binding
Solubility, 'rule of 5'
Blood brain barrier



Lipinski rule-of-five (see [Lipinski et al, Advanced drug delivery reviews, 46:3-26 \(2001\)](#))
4 parameters (guidelines for oral bioavailability)

Poor absorption or permeation is more likely when

1) the molecular weight is over 500 (but can be higher)
2) the calculated octanol/water partition coefficient $clogP$ is higher than 5 (e.g., lipophilicity is high)

3) when there are more than 10 hydrogen bond acceptors (expressed as sum of Ns and Os)

4) more than 5 hydrogen bond donors (expressed as the sum of OHs and NHs)

(compound classes that are substrates for biological transporters or derived from natural products are exceptions to the rule)

**Molecular descriptor: log P (total lipophilicity of a compound)
The logarithm of the partition coefficient between octanol and water (the log converts the value onto a free energy scale)**

ClogP breaks a molecule into fragments, ClogP is computed by adding appropriate values for the fragments...

<http://www.molinspiration.com/jme/index.html>

(to compute properties and drug-likeness)

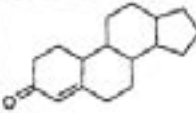

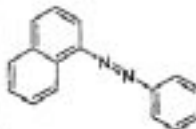
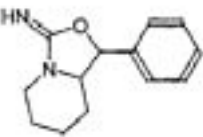

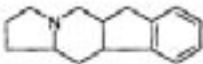
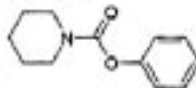
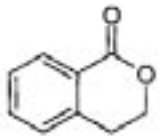
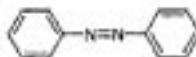

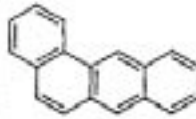
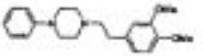
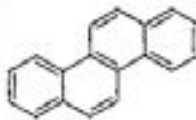
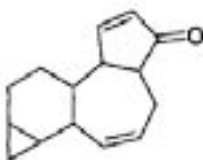
JME Editor courtesy of Peter Ertl, Novartis

Simple filters obtained after analysis of drugs used in human:

Table 1. Typical Ranges for Parameters Related to Drug-Likeness^a

<i>Parameter</i>	<i>Minimum</i>	<i>Maximum</i>
LogP	-2	5
Molecular weight (MW)	200	500
Hydrogen bond acceptors	0	10
Hydrogen bond donors	0	5
Molar refractivity	40	130
Rotatable bonds	0	8
Heavy atoms	20	70
Polar surface area (PSA) [Å ²]	0	120
Net charge	-2	+2

Simple filters obtained after analysis of drugs used in human:

framework	toxicity abnormality	framework	toxicity
	abnormality		autonomic nervous system
	blood		brain and covering/CNS
	cardiac		gastrointestinal
	glands		kidney, ureter, bladder
	liver		lung
	tumor		vascular
	mutation		skin

Experimental ADME: intestinal absorption
There are several mechanisms of intestinal drug absorption, a major one is via passive diffusion.

Caco-2 cells are used to model intestinal absorption.

The transport is evaluated by adding a drug above the caco-2 cells. Uptake is monitored by quantifying the amount of drug in the medium on the opposite side of the membrane (basolateral compartment)

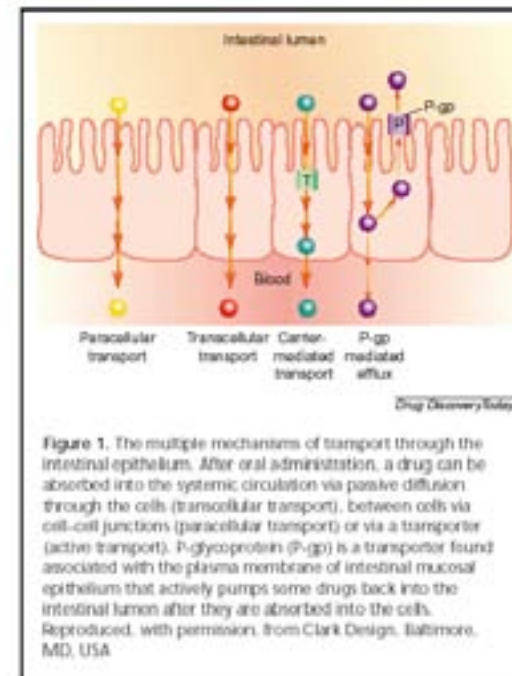


Figure 1. The multiple mechanisms of transport through the intestinal epithelium. After oral administration, a drug can be absorbed into the systemic circulation via passive diffusion through the cells (transcellular transport), between cells via cell-cell junctions (paracellular transport) or via a transporter (active transport). P-glycoprotein (P-gp) is a transporter found associated with the plasma membrane of intestinal mucosal epithelium that actively pumps some drugs back into the intestinal lumen after they are absorbed into the cells. Reproduced, with permission, from Clark Design, Baltimore, MD, USA

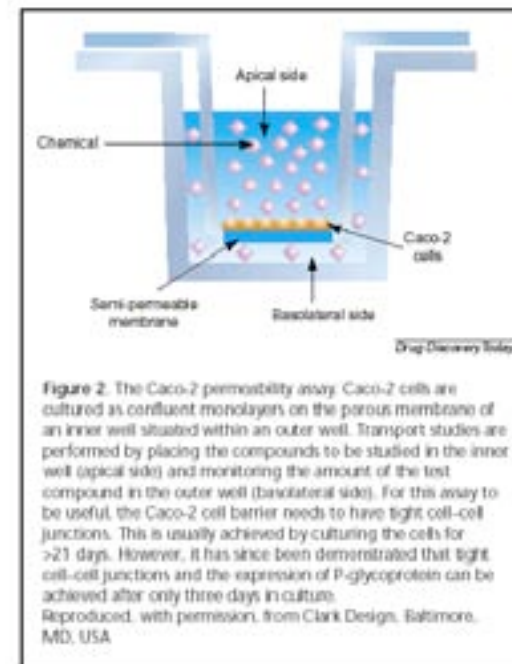
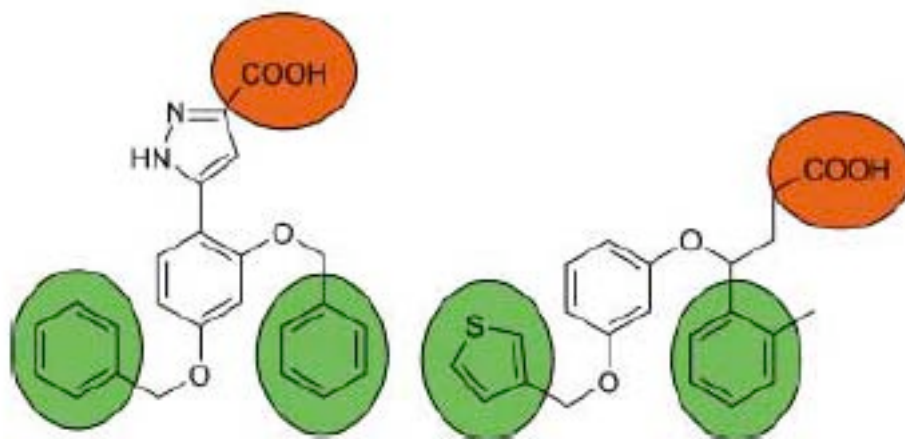


Figure 2. The Caco-2 permeability assay. Caco-2 cells are cultured as confluent monolayers on the porous membrane of an inner well situated within an outer well. Transport studies are performed by placing the compounds to be studied in the inner well (apical side) and monitoring the amount of the test compound in the outer well (basolateral side). For this assay to be useful, the Caco-2 cell barrier needs to have tight cell-cell junctions. This is usually achieved by culturing the cells for >21 days. However, it has since been demonstrated that tight cell-cell junctions and the expression of P-glycoprotein can be achieved after only three days in culture. Reproduced, with permission, from Clark Design, Baltimore, MD, USA

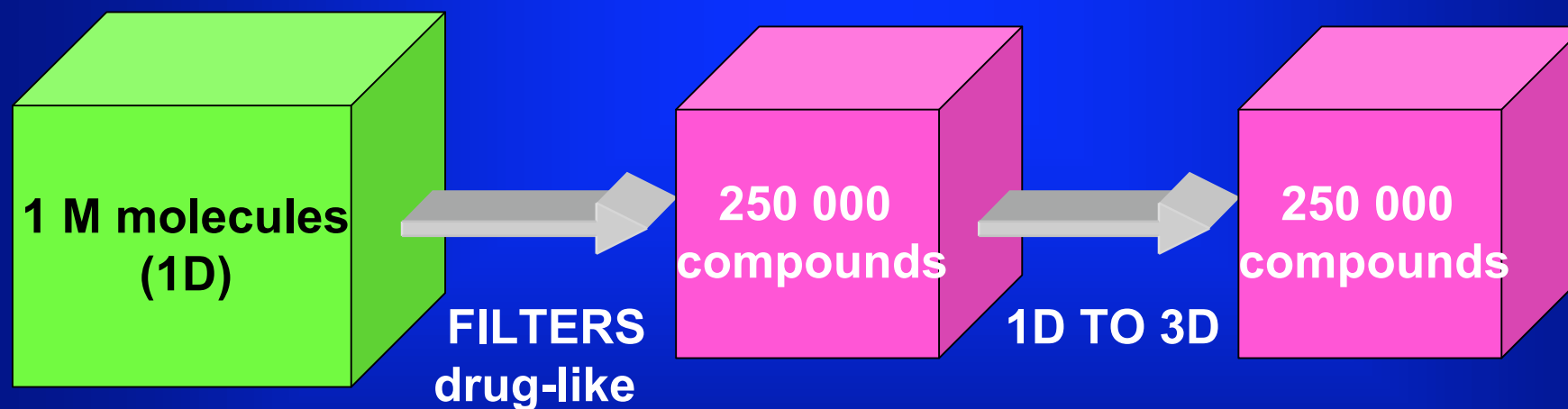
On the computer, polar surface area can be correlated to passive intestinal absorption... one could also try to simulate membrane penetration, but highly time consuming

Focused libraries: nowadays, small compound collections that are often designed and directed against target families are often used instead of enormous databank of small compounds.

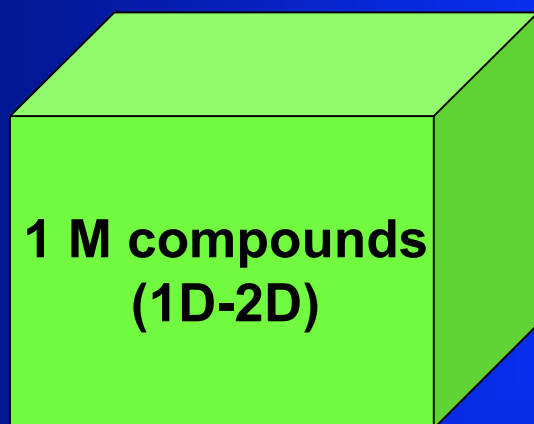


The balance between chemical diversity and drug-likeness is difficult

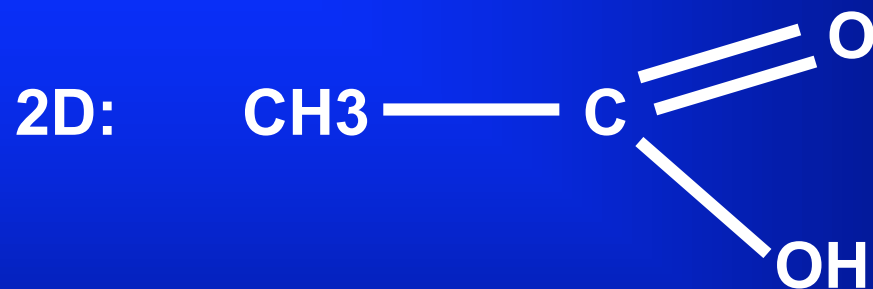
Small molecule database: 1D to 3D



E.g., Starting database: 1M



1D: CC(=O)O ACETIC ACID
(SMILES notation, usually hydrogens are not explicitly represented, all the compounds can be in the same file)



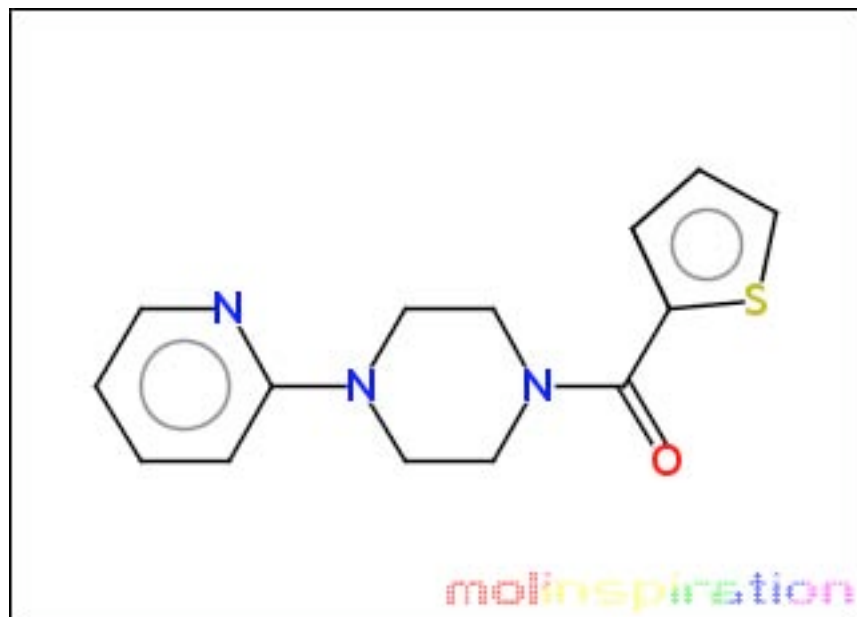
SMILES: Simplified Molecular Input Line Entry Specification

http://www.daylight.com/smiles/f_smiles.html

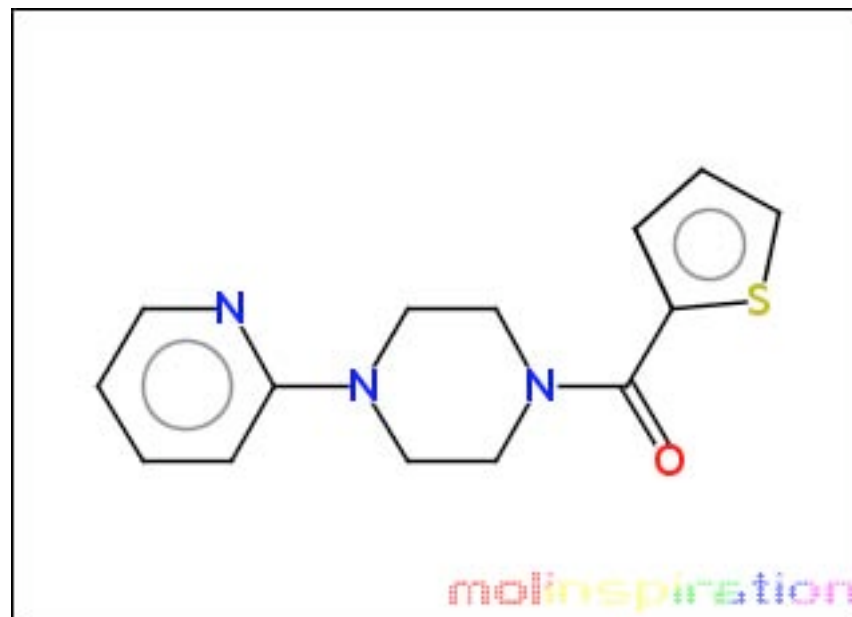
Tutorial: The official guide to the complete SMILES language

Different ways of writing same compound

SMILES C(N1CCN(CC1)c1cccn1)(c1cccs1)=O



SMILES O=C(N1CCN(CC1)c2cccn2)c3cccs3



1D/2D to 3D

**Generate 1 or several conformers for each molecule
(interesting for rigid body docking)**

Some tools to do that: Omega

<http://www.eyesopen.com/about/>

Corina...

<http://www2.ccc.uni-erlangen.de/software/corina/corina.html>

Often, the compounds are stored in a multi-mol2 file (3D)

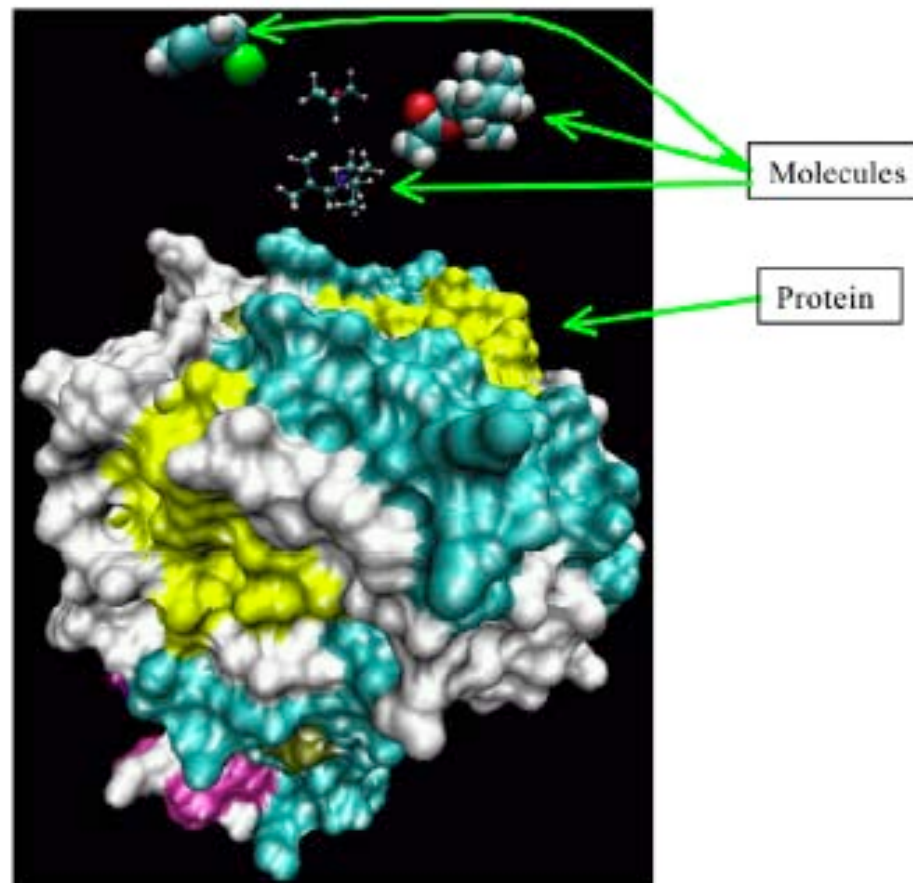
Information about the Tripos mol2 format:

<http://www.tripos.com/custResources/mol2Files/index.html>

SDF or Structures Data File is a common file format developed by [Molecular Design Limited](#) to handle a list of molecular structures with associated properties. The file format has been published (Dalby et al., *J. Chem. Inf. Comput. Sci.* 1992, 32, 244-255).

For 2D or 3D structures

Docking



A) First a computer program is used to place the small molecules in the receptor pocket. A simple scoring function allows for selection of the poses

This is **docking**

B) Second the binding enthalpies of the docked molecules are estimated by evaluating their complementarity to the target in terms of shape and properties such as electrostatics.... Often entropic effects of binding are also assessed. This prediction of the “binding free energy” (affinity) is called “scoring**” (different units, some try to reproduce K_d or K_i ...). A molecule with a good score is potentially a good binder.**

Docking/scoring

1) rigid (but with tricks)

2) “flexible”

Note: the size and diversity of the bank is critical to be able to really evaluate VLS tools

Docking Methodology

Various methods for docking/scoring have been developed:

- Fast shape matching (Dock, Eudock, Fred)
- Incremental construction (FlexX, Hammerhead, Surflex, Dock)
- (Lamarckian) Genetic algorithm (Gold, Autodock)
- Simulated annealing (Affinity...)
- Monte Carlo simulations (MCDock, QXP, ICM, LigandFit...Affinity)
- others

Usually $\text{RMSD} < 2 \text{ \AA}$ is considered acceptable, but often RMSD can be very confusing in judging such modeling packages (see *J. Chem. Inf. Comput. Sci.* 2004, 44, 871-881)

Docking Methodology

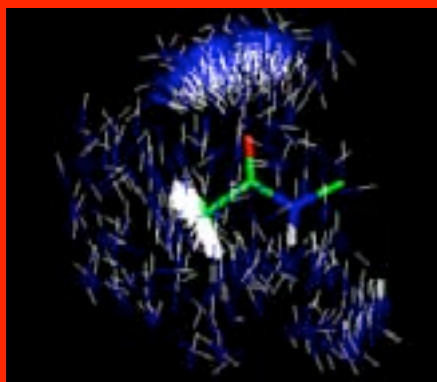
H.J. Böhm, J. Comput. Aided Molec. Des. 8, 623-632 (1994)
M.D. Miller, R.P. Sheridan, S.K. Kearsley, J. Med. Chem. 1999, 42, 1505-1514

Incremental construction:

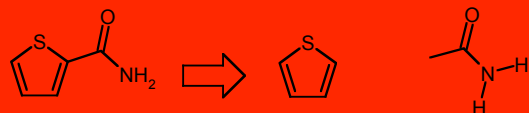
Modelling receptor ligand interactions:

Receptor interaction surface from crystallographic information etc.

Approximation by a finite set of interaction centers



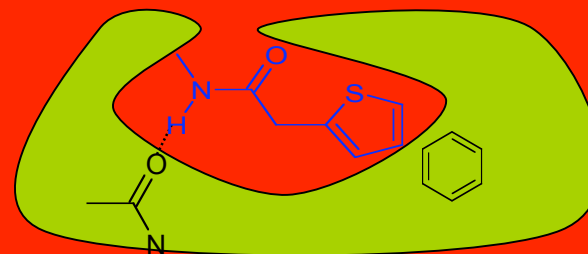
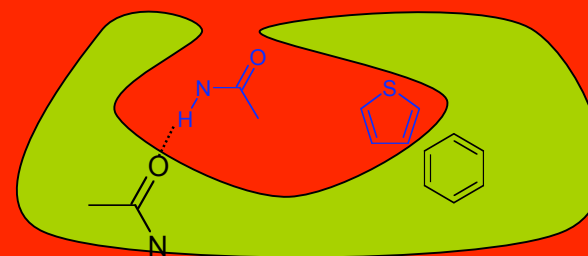
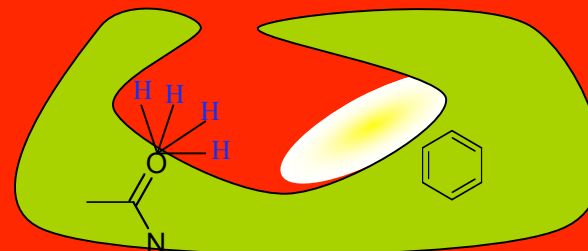
Fragmentation of ligand into base fragments



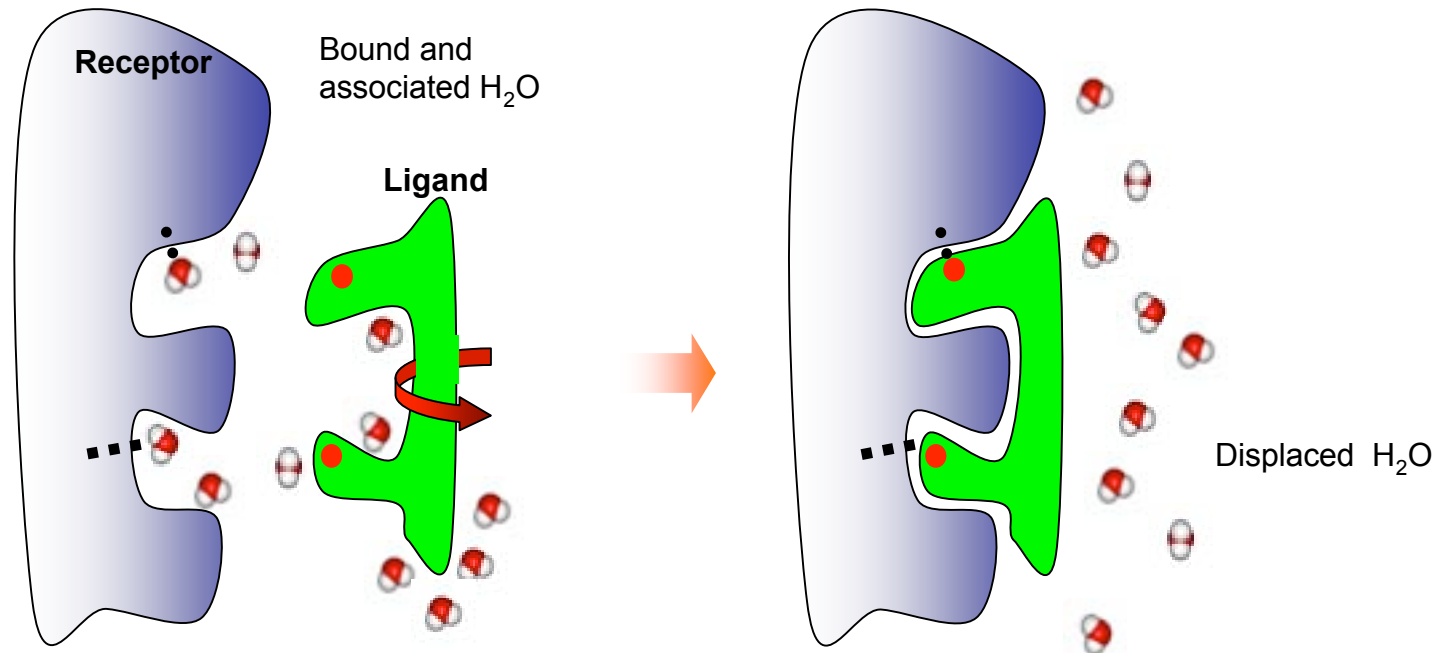
Place ligands into active site by matching interaction centers

Reduction of number of solutions by clash test and clustering

Link base fragments in compliance with a torsional database or a forcefield



• Scoring



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

Scoring= we would like to compute real free energy of binding...but not really possible for VLS

Upon complex formation:

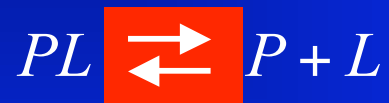
- water molecules are released
- receptor and ligand lose degrees of freedom
- interactions between ligand and receptor

complication: mutual compensation of enthalpy and entropy

•Scoring

Binding Affinities

- Ligand receptor binding affinity can be experimentally determined.
- Experimental errors lie in the range of 0.1-0.25 kcal/mol.



$$K_D = \frac{[P][L]}{[PL]}$$

$$\Delta G_{binding} = RT \ln K_D$$

$$\Delta G_{binding} = -RT \ln K_A$$

- 1-2 H-bonds \cong 1.5 kcal/mol \cong 1 order in K_D

•Scoring

Problems in Calculating Binding Energies

- Some contributions (*e.g.*, entropy) can only be roughly estimated: $\Delta G = \Delta H - T\Delta S$ where ΔS is hard to approximate.
- Free energy perturbations, thermodynamic integration too time consuming.
- Possible only Linear Interaction Energy

$$\Delta G_{bind} = \alpha\Delta\langle V_{l-s}^{vdw} \rangle + \beta\Delta\langle V_{l-s}^{el} \rangle + \gamma$$

Scoring functions

- Simple functions designed to rank protein-ligand complexes according to their binding affinity.
- **No single scoring function works for all cases.**

Docking Methodology: Scoring

Scoring functions are used:

- during docking for optimization of ligand orientation and conformation
- for docked ligands to estimate affinity relative to other compounds

Various criteria for the quality of a docking function:

- ability to find the correct binding mode out of alternative docking solutions
- ability to rank related ligands with respect to their binding affinity
- ability to select (however weak) inhibitors from a large database of inactive compounds
- they should be fast and error tolerant

The inaccuracy of functions used to estimate the affinity between receptor and ligand is considered to be the major weakness of docking programs

•Scoring

A: Force Field-Based Scores

- Describe only enthalpic contributions (ΔH). No estimate of ΔG
- Use non-bonded interactions

$$E_{non_bonded} = \sum_i^{lig} \sum_j^{prot} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + 332 \frac{q_i q_j}{D r_{ij}} \right)$$

- Augment force field terms with solvation and entropy terms

•Scoring

B: Empirical Scoring Functions

- Use regression to fit coefficients to a set of physically motivated terms in order to reproduce the experimental binding affinity of a training set of known protein-ligand complexes
- Data: A set of protein-ligand complexes with known 3D structures and binding affinities (ΔG)
 - Requires K_i values

•Scoring

- force field based methods (Dock, Gold)

separate contributions from hydrogen bonds, ionic and lipophilic interactions, clashes and entropy (nr of rotatable bonds)

- potentials of mean force (PMF, Drug Score)

description of observed interatomic distances and/or frequencies implying that these describe favorable/unfavorable interactions

- consensus scoring

combination of multiple scoring functions increases hit rates by reducing the number of false positives (**WARNING!!!!!!!!!!!!!!!!!!!!!!!!!!!!**)

- two stage ranking

first filter to limit the number of docked conformations,
second filter to reject false positives

Docking Methodology: Empirical Scoring function like in Ludi

$$\Delta G = \overset{1}{\Delta G_0} + \overset{2}{\Delta G_{hb}} \sum_{h-bonds} f(\Delta R)f(\Delta\alpha) + \overset{3}{\Delta G_{ion}} \sum_{ionic} f(\Delta R)f(\Delta\alpha) + \overset{4}{\Delta G_{lipo}} A_{lipo} + \overset{5}{\Delta G_{rot}} NR + \overset{6}{\Delta G_{aro/aro}} \Delta N_{aro/aro}$$

Term 1 = constant

Term 2 = optimal Hbond

Term 3 = salt bridge

(they have scaling function to penalize deviations from ideal interaction geometry)

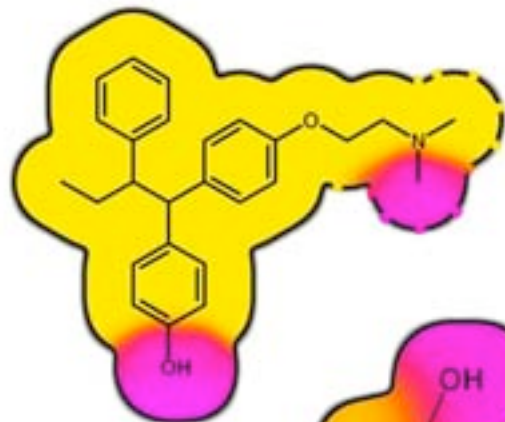
Term 4 = sum over all atom-atom contact (lipophilic)

Term 5 = a term taking into account loss of entropy upon ligand binding via accounting burial of rotatable bonds

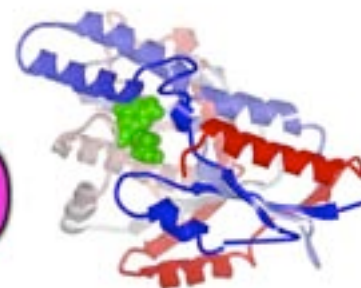
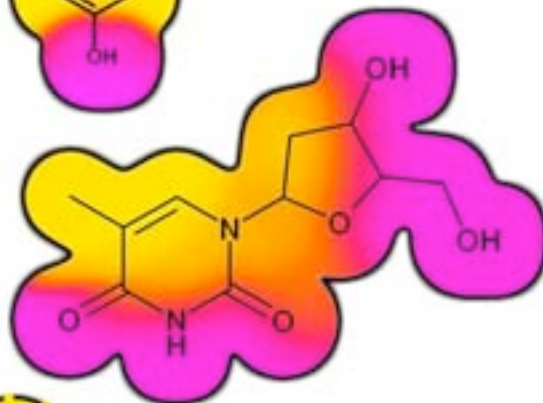
Term 6 = aromatic contact

Problems are the weight factors of the individual energy terms

Estrogen R



Thymidine K



Factor 7a



Neuraminidase



Is it possible to have a “golden” scoring function? When one looks the way these functions are generated and the different types of binding pocket, clearly this does not possible.

Submitted:
Fast structure-based virtual ligand screening by combining FRED, DOCK and Surflex

M A. Miteva, W. H. Lee, M. O. Montes, B. O. Villoutreix

Evaluation of different docking/scoring combinations

No scoring function predicts absolute free binding energies

If two sets of protein coordinates are used with crystal water, then some changes.....

Ligand flexibility

For the ligand: several conformations for each molecule

Incremental reconstruction

Real simulation of all atoms...

soft-potentials can help in the initial phase

Receptor flexibility

1. Use a static crystal structure of a receptor complexed with another ligand, or use an unliganded structure, the first is usually preferable.
2. Build a receptor model that tolerates ligand binding with some clashes without explicit repacking of the receptor sidechains (the model is static but permissive and implies multiple binding modes).
3. Use several alternative receptor binding site conformations for docking and merge the docking results.
4. Include partial receptor flexibility by allowing receptor relaxation for different trial conformations of the ligand.
5. Perform joint global optimization of molecular dynamics simulations of ligand and the receptor binding site.

**Despite many
problems...
VLS works**

VLS

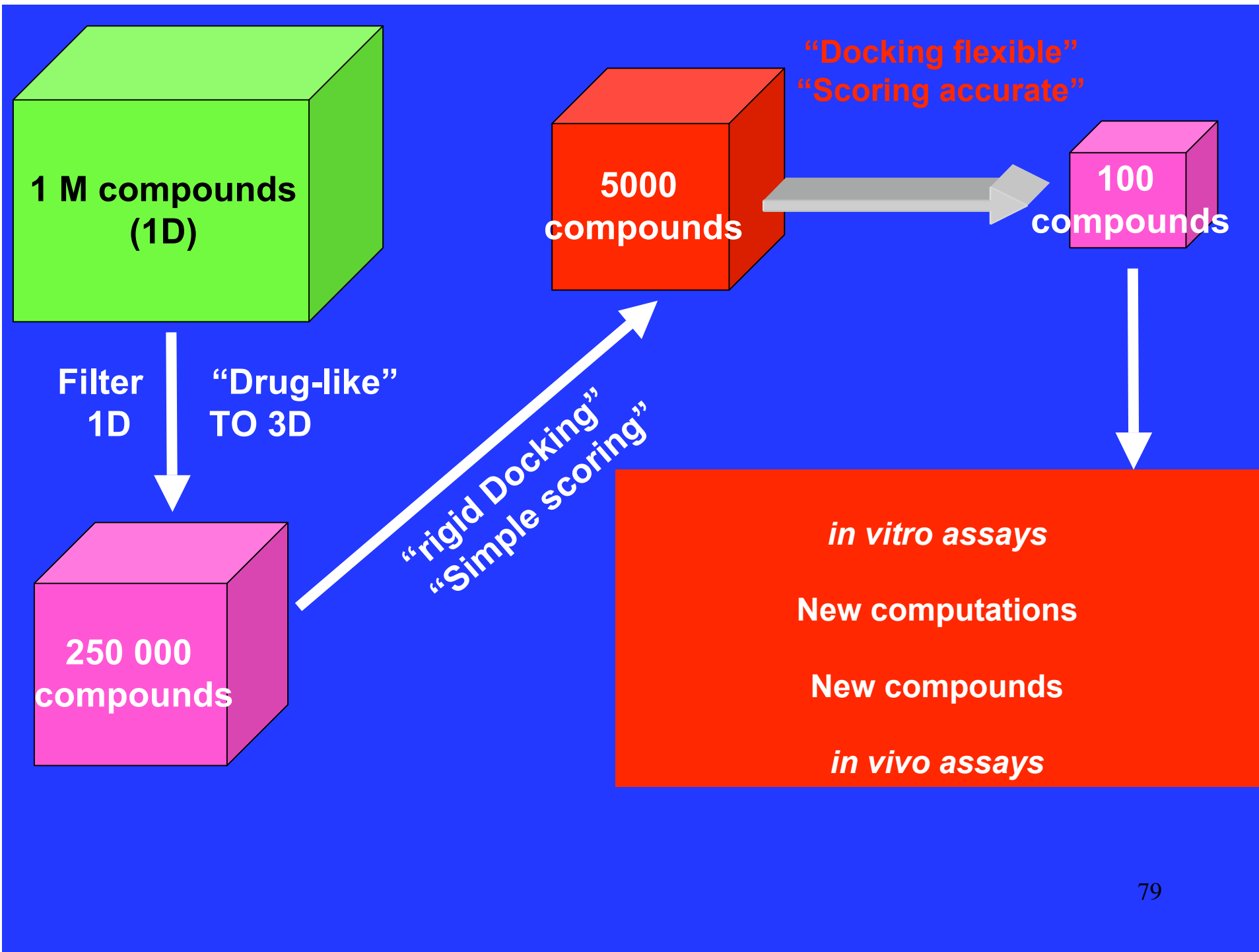
Novel ligands recently identified through structure-based VS.

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
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

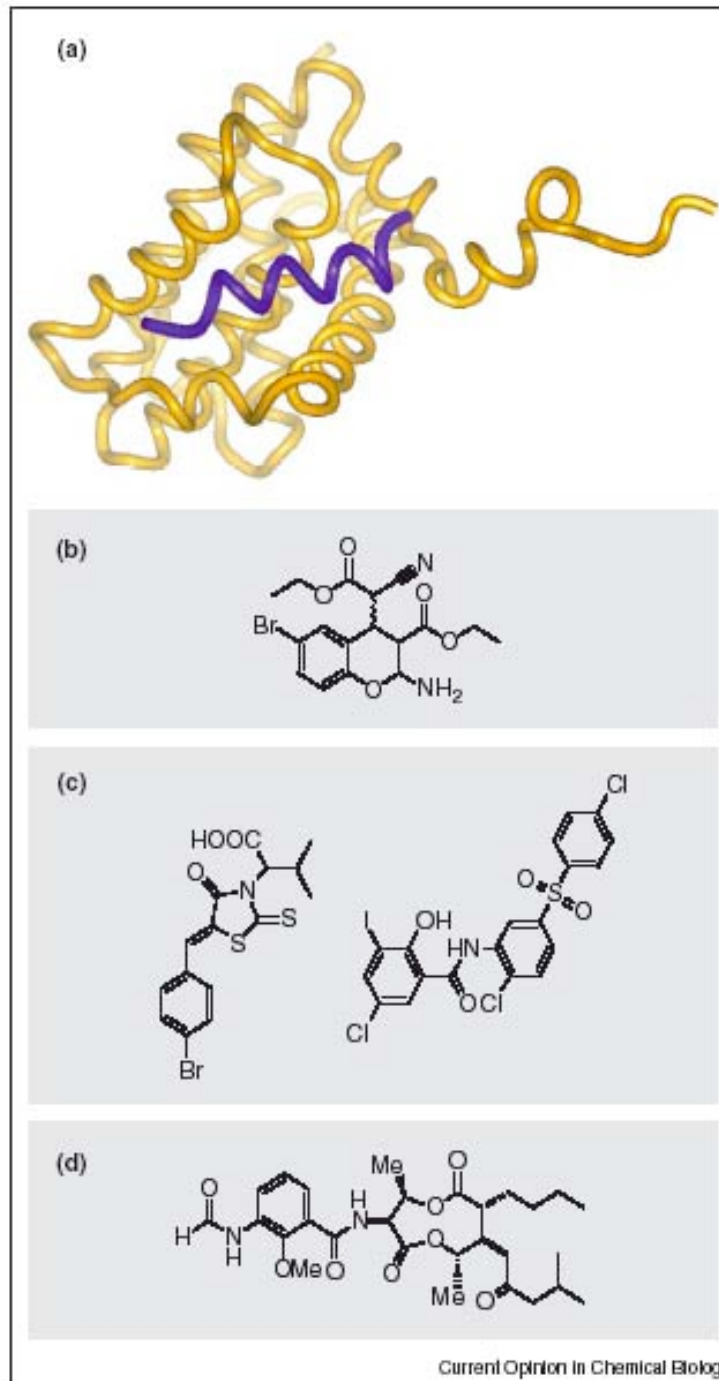
High-throughput docking as a source of novel drug leads

J. C. Alvarez

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



Wang et al
PNAS
97;7124-7129 (2000)
(DOCK)



Lead Optimization

- **Try to plan mutants to check predictions**
- **Define key interactions and add new groups to increase affinity or specificity**
- **Fill free space**

Conclusion: Drug-design

Protein

- 3D (most likely ok in the future with Structural genomics)
- Druggable pocket possible to identify
- Problems if conformational changes

Small compounds

- Databases 1D/2D/3D difficult to maintain
- Drug-lead-like filters need some thinking

Docking/scoring Rigid / Flexible

- Need tuning according to the target
- Eventually run consensus docking/scoring on several conformation of the target
- Problems with flexibility, docking and scoring