

Computational Methods in Drug Discovery

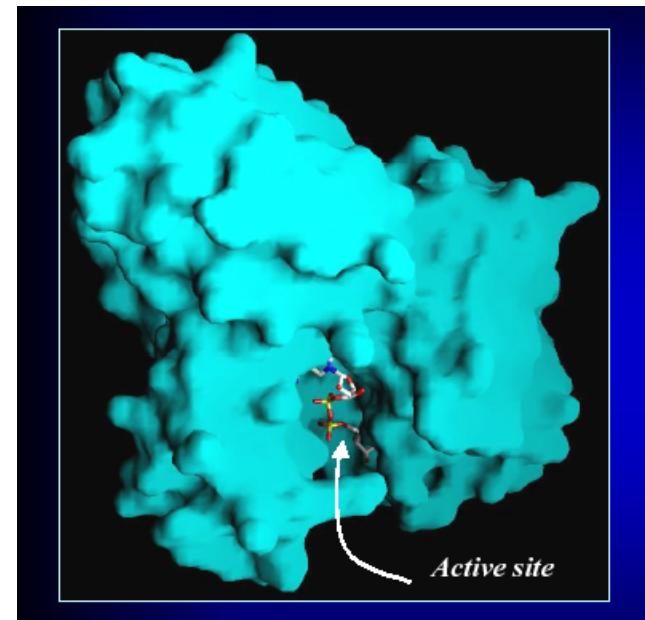
Structure-based

Structure-based computer-aided drug discovery (SB-CADD) approach:

helps to design and evaluate the quality, **in terms of affinity**, of series of ligands.

SB-CADD:

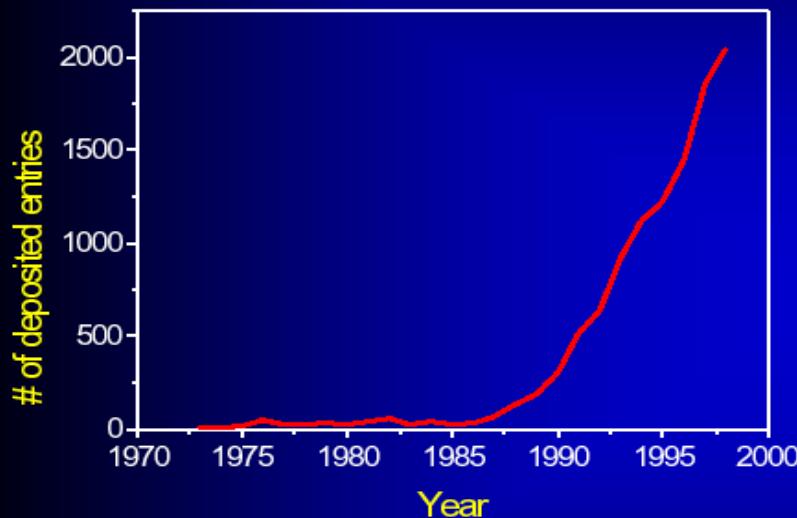
- **Structural information** about the target is a prerequisite
- Search for the **favorable interactions formed between ligands and a specific protein**
- **Novel compounds with those favorable interactions** can be designed through the careful analysis of the protein's binding site.



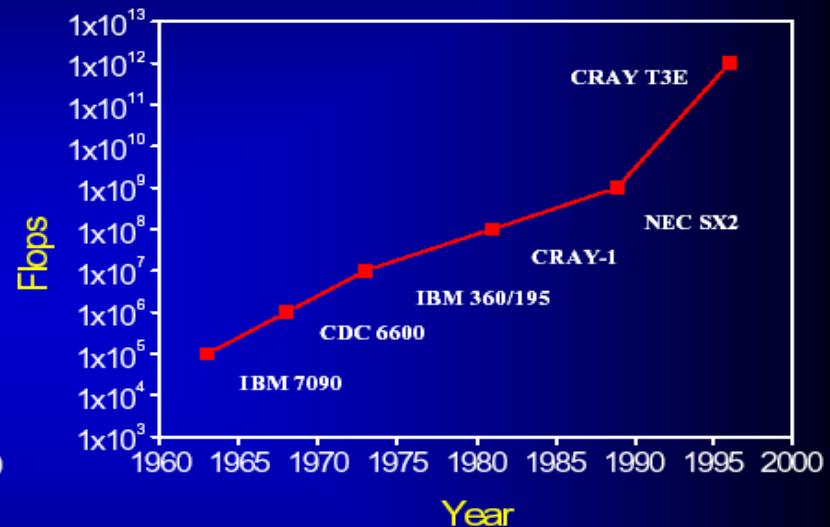
The large number of protein 3D structures that is available from the Brookhaven Protein Database (85,000 entries; 2012 but about 50,000 non redundant (100% sequence identity) enables scientists to perform, in principle, a *de novo* construction of ligands or a virtual screening of ligands that fit a certain binding site, in shape and in all other properties.

In Silico Screening: Why ?

Protein Data Bank Status



Computing Power



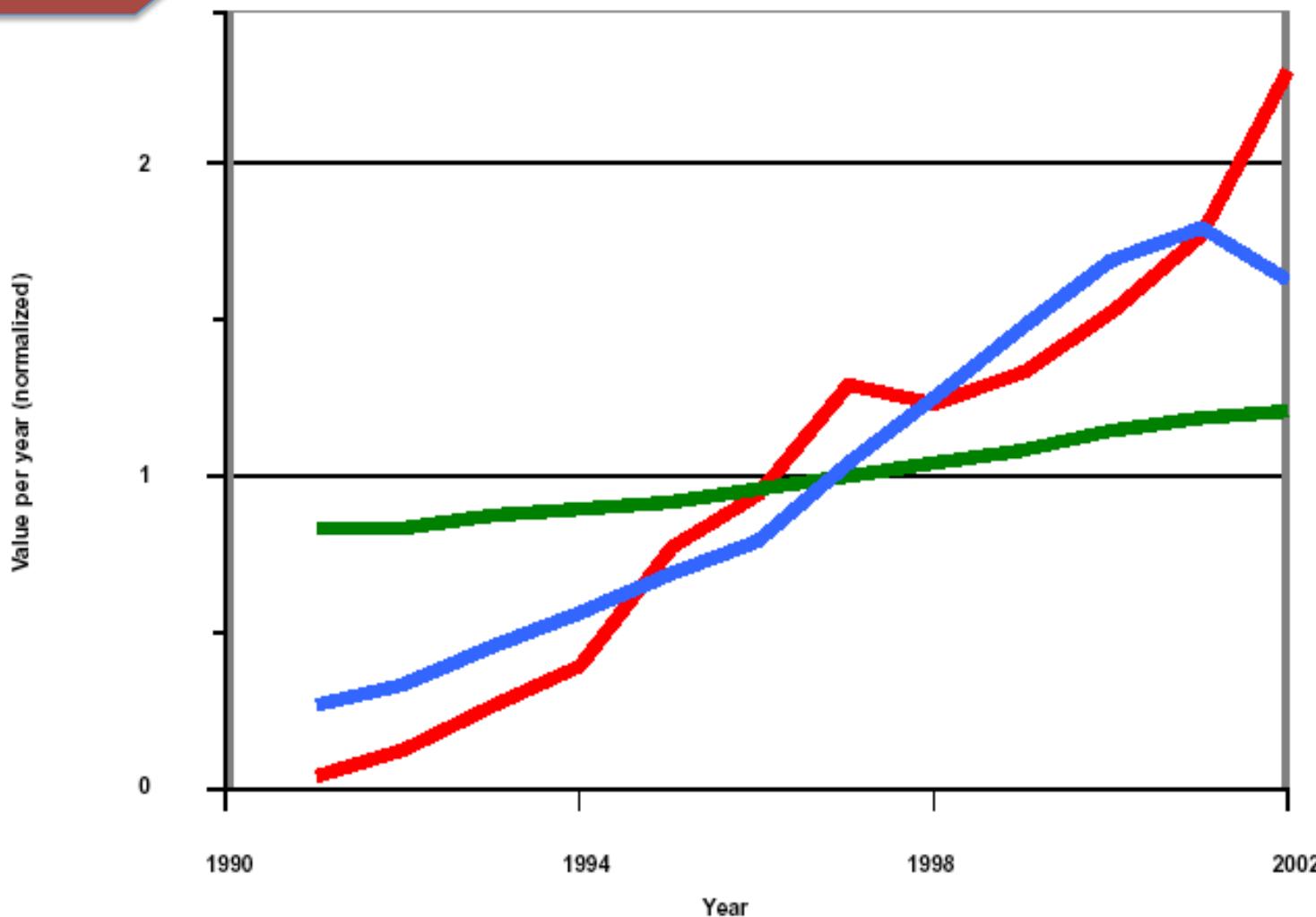


Figure 2. The plots show the number of publications within the PubMed database that used the term 'structure-based design' (in red) or 'therapeutic' (in green), and the number of X-ray structures (deposited within the indicated year) in the Protein Data Bank (in blue). The values were normalized by dividing the number of citations (structures) in each year by the mean annual value over the 12 years analyzed.

Structure-based ligand design started about 25 years ago

- *Dorzolamide* is used to treat glaucoma, a condition in which increased pressure in the eye can lead to gradual loss of vision.
- Dorzolamide was derived from a lead structure, which was modelled from the 3D-structure of the enzyme carbonic anhydrase.
- Carbonic anhydrase controls secretion of fluid within the eye and thereby determines the pressure within the eye
- Captopril is an angiotensin-converting enzyme (ace) inhibitor used for the treatment of hypertension and some types of congestive heart failure.
- Angiotensin I is converted to angiotensin II (a powerful vasoconstrictor) through removal of two c-terminal residues by the angiotensin-converting enzyme.
- Angiotensin II is a peptide hormone that causes vasoconstriction and a subsequent increase in blood pressure.

- Structure-based design is now an important technique in cases, where the target 3D structure is known or accessible.

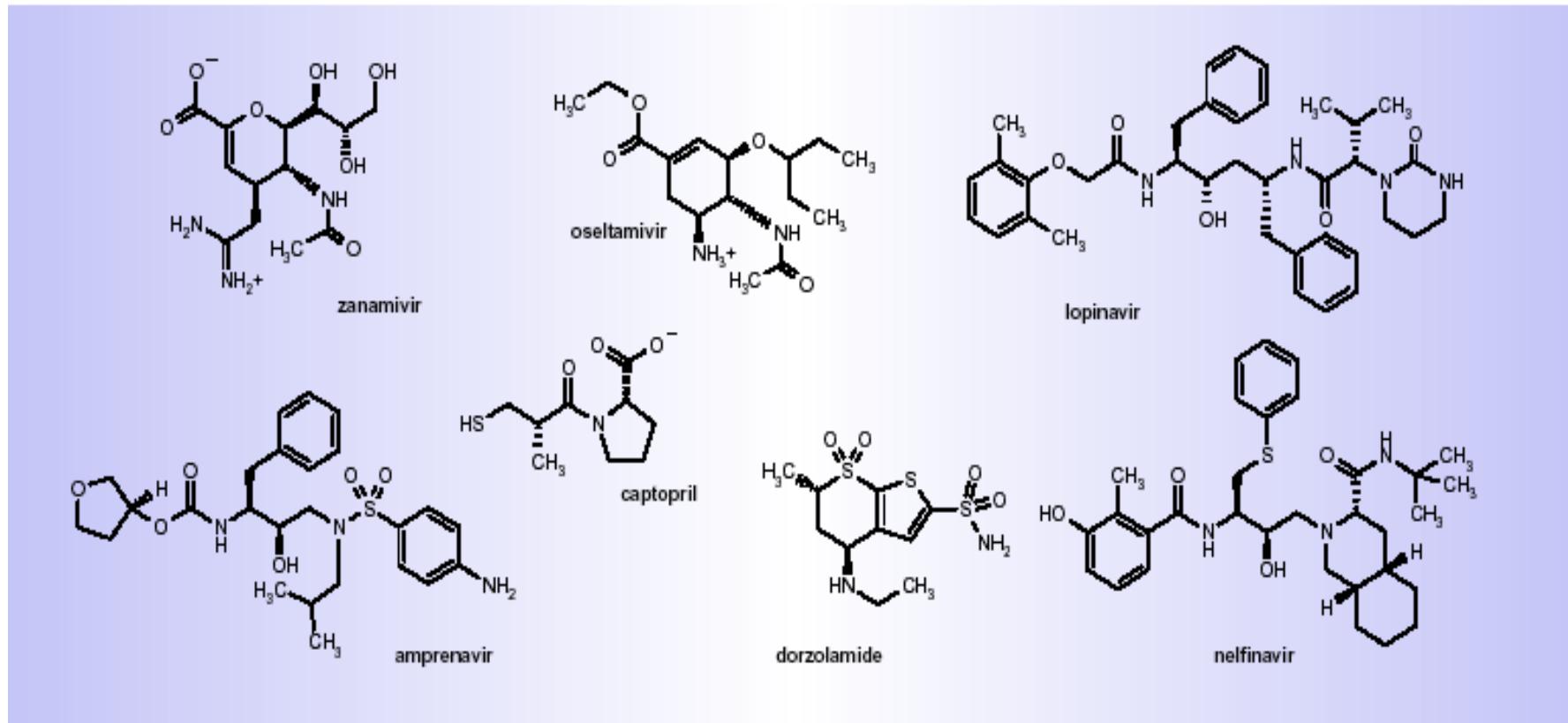


Figure 1. Approved and marketed drugs whose discovery has been aided by structure-guided design methods.

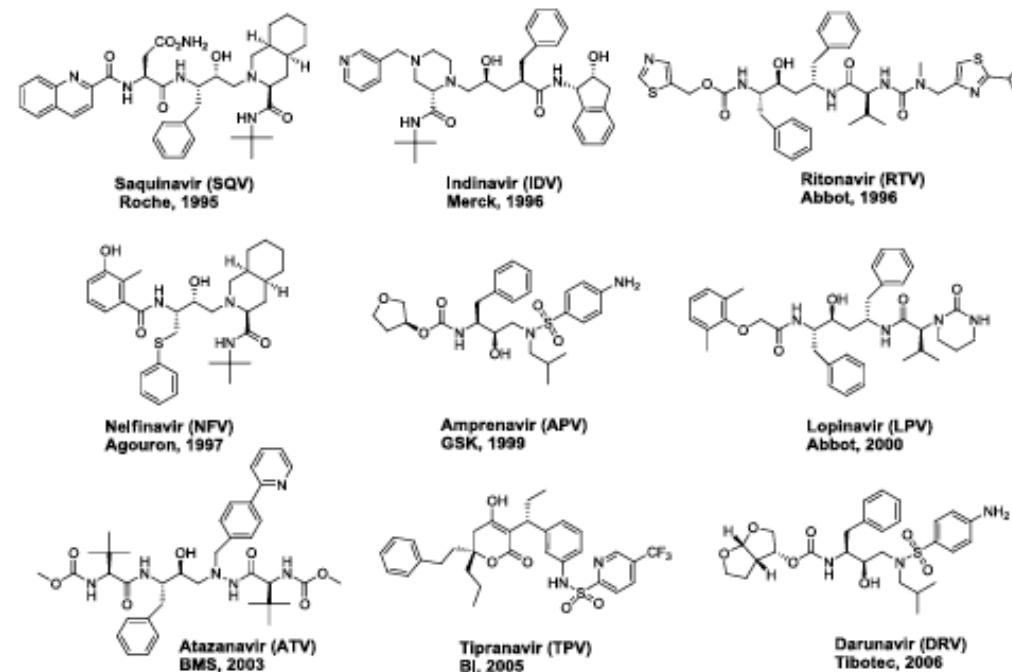
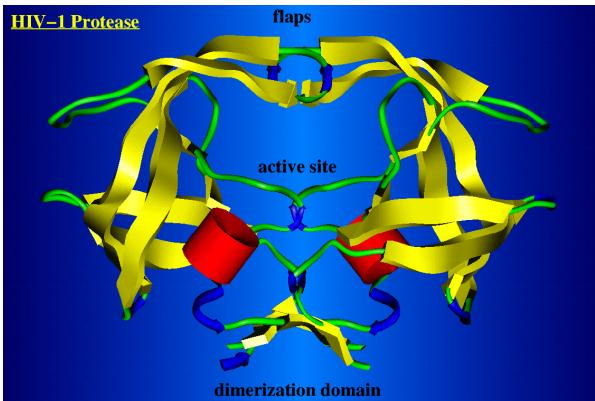
Other drugs followed with the HIV protease inhibitors nelfinavir and amprenavir.

Drug discovery efforts aided by structure-based design have led to the development of **nine FDA-approved protease inhibitors**.

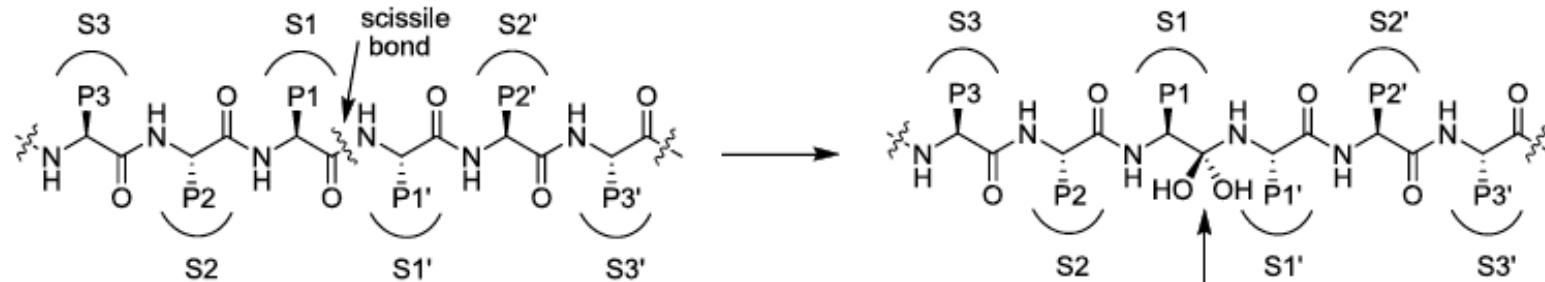
These inhibitors represent **the most potent anti-AIDS drugs** reported to date and are essential components of the highly active antiretroviral therapy (HAART).

HAART is credited with significantly reducing AIDS-related mortality and is currently implemented throughout the world as the standard of care for HIV-AIDS.

Figure 1. FDA-approved HIV-1 protease inhibitors.

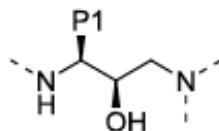


AIDS drugs

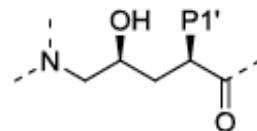


Polyprotein substrate

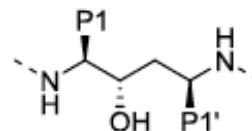
Transition state intermediate



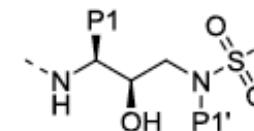
Hydroxyethylamine I



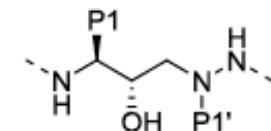
Hydroxyaminopentane II



Hydroxyethylene III



Hydroxyethylamino-sulfonamide IV



Aza-hydroxyethylamine V

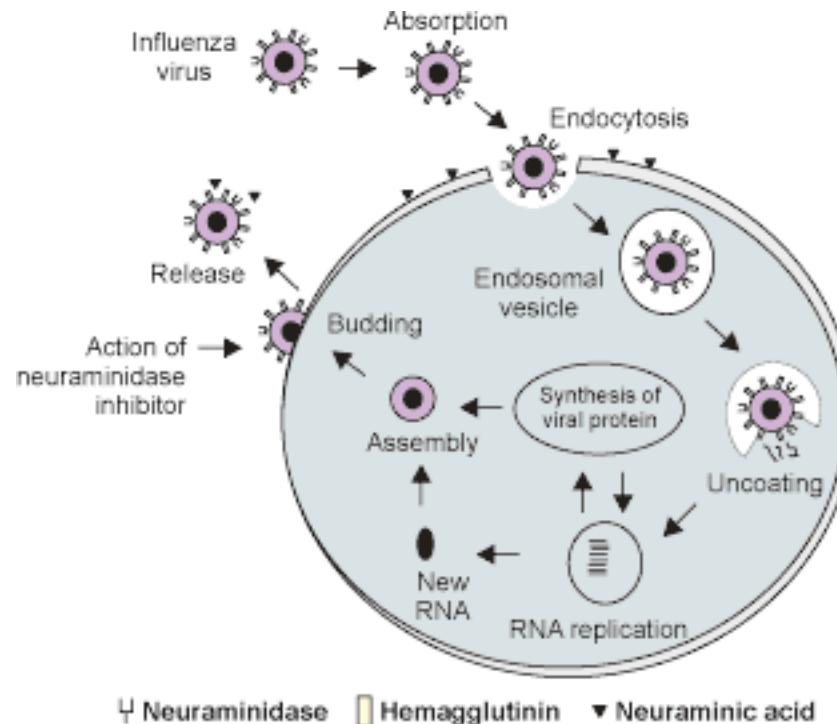
The scissile bond in polyprotein substrate is hydrolyzed by protease through the transition state intermediate (substrate amino acid residues are marked as...P3, P2, P1, P1', P2', P3'...and the corresponding enzyme binding sites as...S3, S2, S1, S1', S2', S3'...). Transition state mimics I–V used in the design of currently approved drugs.

Flu drugs

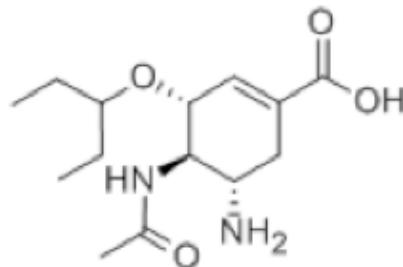
Hemagglutinin is a protein located at the virus external layer. It recognizes and connects to sialic acid of P-glycoprotein of the cells.

Neuraminidase is an enzyme that recognizes the same molecule as hemagglutinin but its role is to help liberating the virus.

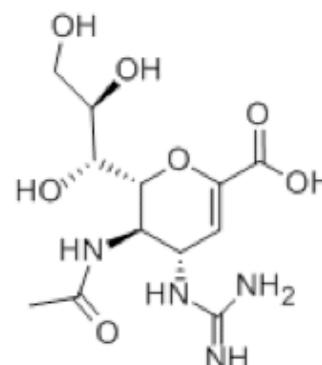
Neuraminidase recognizes the sialic acids and cleaves them, liberating the virus.



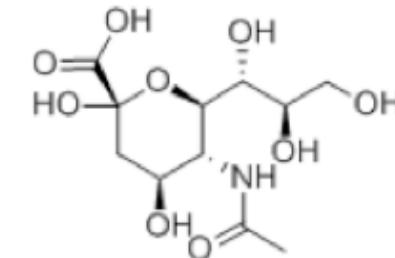
Flu drugs



Oseltamivir

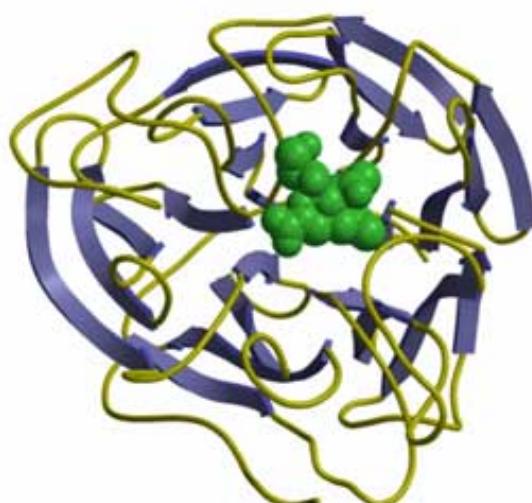


Zanamivir



Sialic acid

Tamiflu (Oseltamivir) and a related drug, Relenza (Zanamivir)



Structure of the influenza virus neuraminidase showing how influenza drug relenza blocks the function.

Relenza was designed to fit the structure

Computational Methods in Drug Discovery

Structure-based

Many more 3D structures of proteins and protein-ligand complexes will become available in the near future, due to high-throughput techniques in protein crystallisation and crystallography.

Once the major part of all protein folds will be known, homology modelling and molecular replacement in crystallography will gain further importance.

Two fundamental approaches of SB-CADD are

Ligand docking

Fragment-based Design

Ligand docking

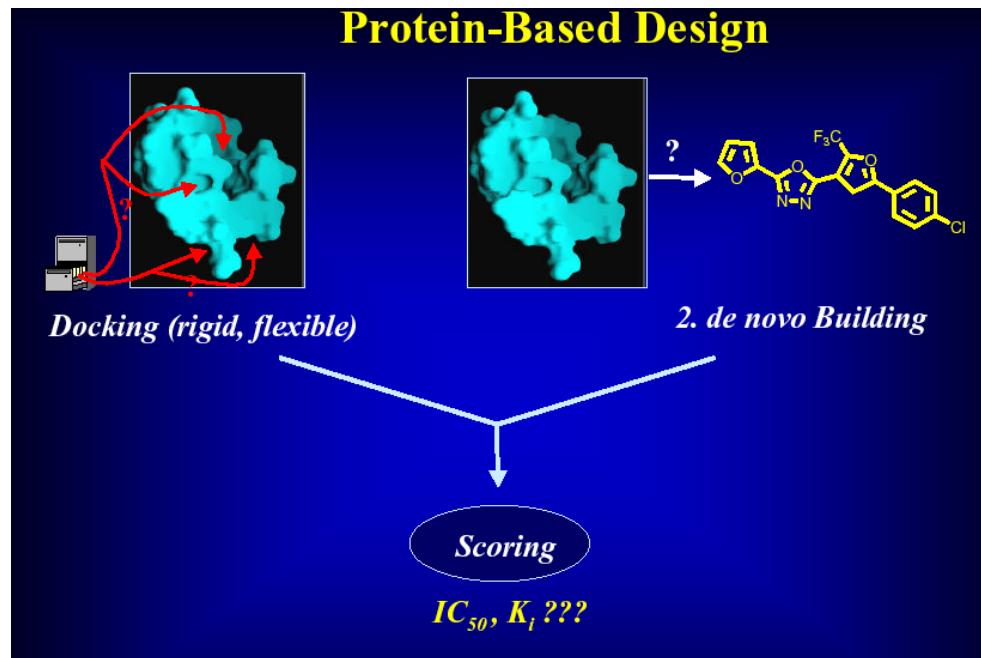
- positions in a receptor a series of molecules (“docking”)
- evaluates the “affinity” of the different generated complexes (“scoring”)

de novo rational or fragment-based design

- aims at building a complete molecule from molecular bricks (“building blocks”) positioned in the binding site of the receptors.
- Evaluation of affinity is then performed.

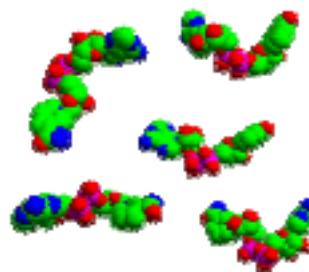
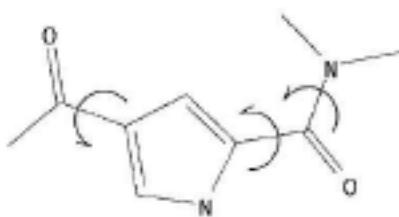
The ligand docking and the fragment-based design have in common

- the use of algorithms of molecular docking to position molecules or fragments in the binding site of the receptors generating a series of complexes
- as well as the use of scoring functions to evaluate the energy interactions and order the complexes.

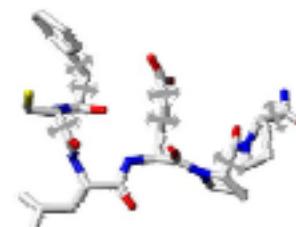


- Ligand binding is a dynamic process in which **both the ligand and protein may undergo conformational changes**

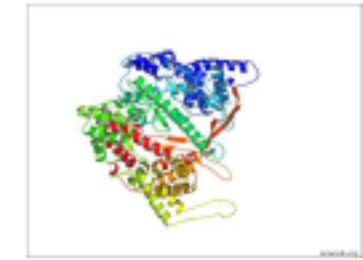
Ligand flexibility



Protein flexibility



Side chain movements



Large-scale movements

Ligand flexibility is accounted for

EXCEPT

when a huge number of compounds are to be docked

Incorporating the **whole protein flexibility** becomes quickly **impractical**.

BUT

Various approaches to model receptor flexibility.

- Molecular dynamics and Monte Carlo
- rotamer libraries and protein conformations

Structure-based

Ligand docking

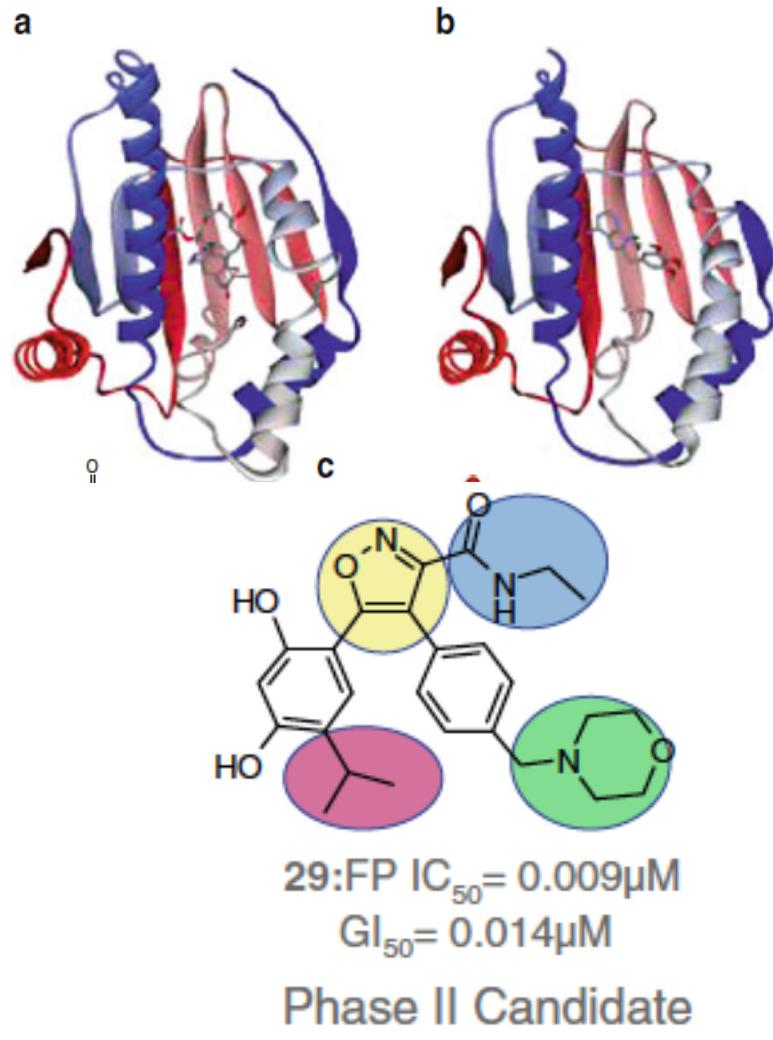
Structure-based virtual high-throughput screening (SB-vHTS)

➔ To identify putative hits out of hundreds of thousands of compounds

Opposed to **the 3D virtual screening ligand-based virtual screening** (without using the 3D structure) that are **more biased** by the properties of the known ligands used for the request.

Hsp90, a molecular chaperone, is an important therapeutic target for oncology.

- **0.7 million compounds screened** using crystal structures of Hsp90 bound to previously known inhibitors (Roughley et al. (2012)).
- 719 compounds from over 9000 non redundant hits tested
- 13 compounds with $IC_{50} < 100 \text{ mM}$ and 7 with $IC_{50} < 10 \text{ mM}$ were identified.
- After lead optimization **one compound (AUY922)** was evaluated for multiple myeloma, breast, lung, and gastric cancers.

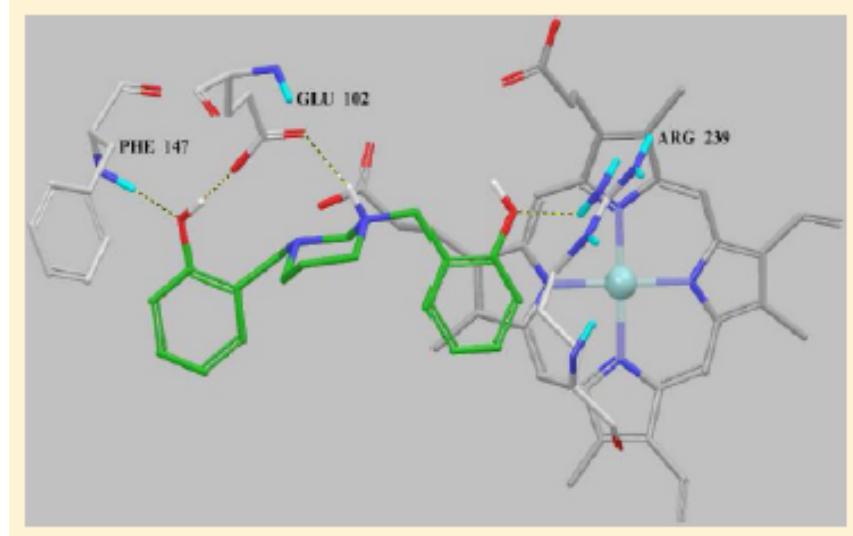
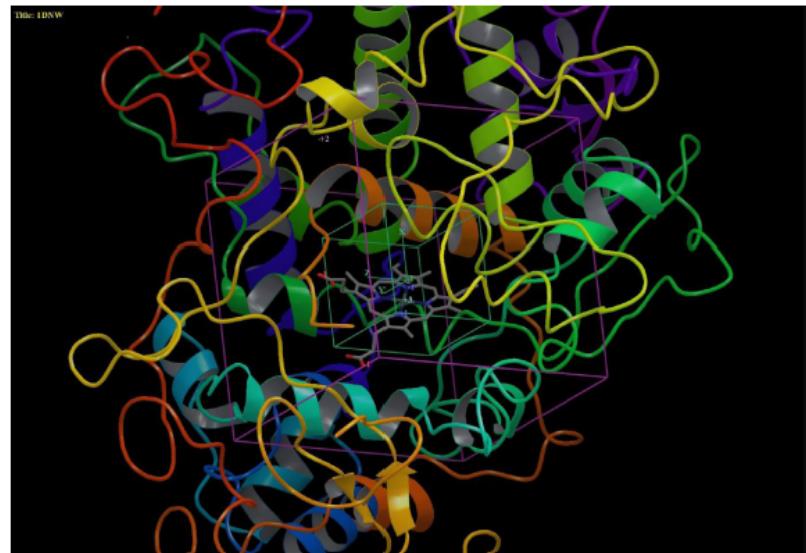


Structure-based

Ligand docking

Myeloperoxidase (MPO)

- a major player of the innate immune defense system but involved in many (inflammatory) diseases
- High-throughput molecular docking of 1350000 using the X-ray structure of MPO
- 81 tested for inhibition of the chlorination activity of MPO
- Eight inhibiting candidates of different chemical structures with two of the selected compounds having a submicromolar activity (Aldib et al., J. Med. Chem. 2010)



Structure-based

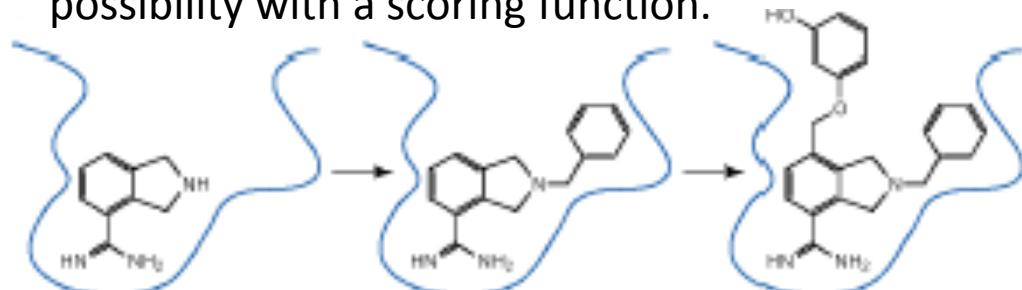
Fragment-based Design



consists in building *in situ* ligands **fragment by fragment** as in jigsaw puzzle.

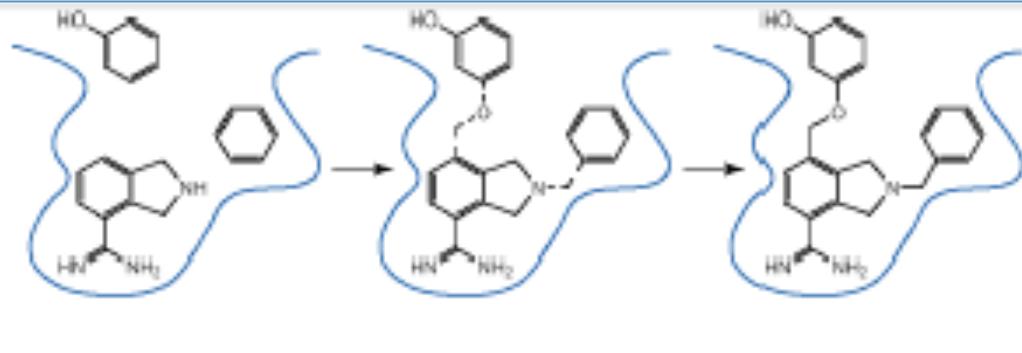
- *sequential growth technique:*

the molecule grows in the binding site controlled by a search algorithm which evaluates each growing possibility with a scoring function.



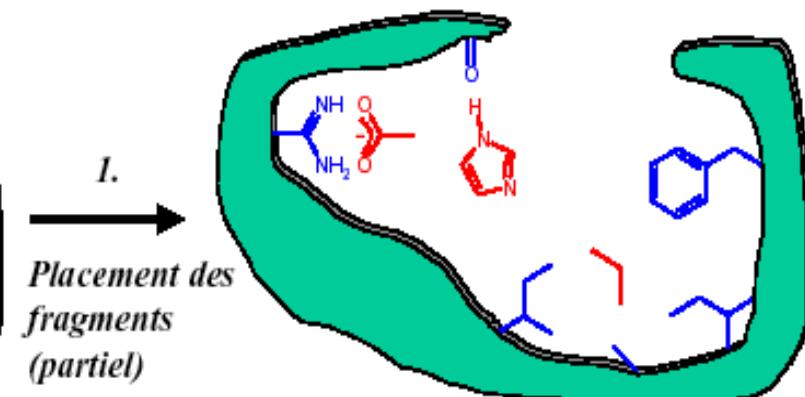
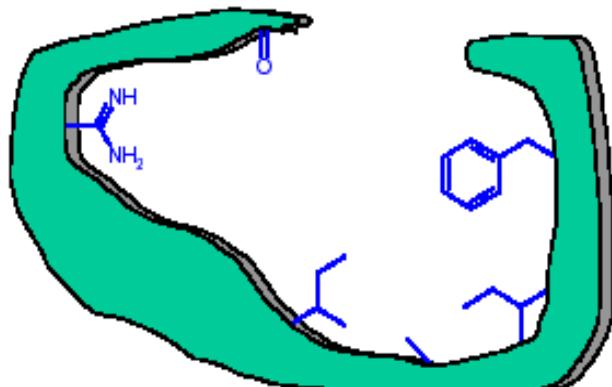
- *fragment placing and linking:*

the binding site is mapped to identify the possible anchor points for functional groups. These groups are then linked together and form a complete molecule.

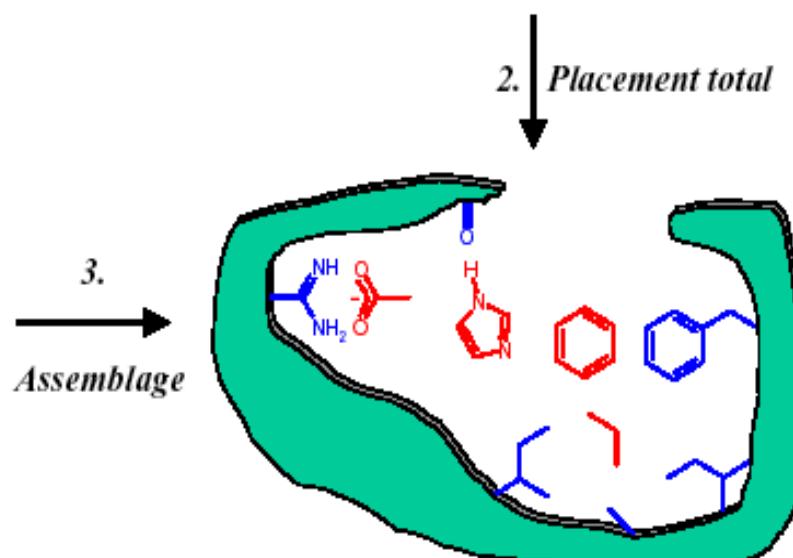
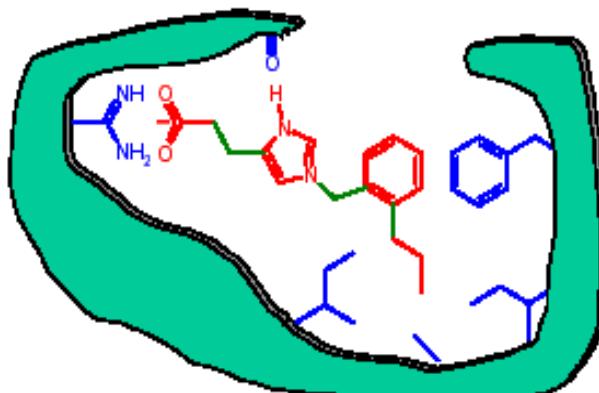


Structure-based

Fragment-based Design

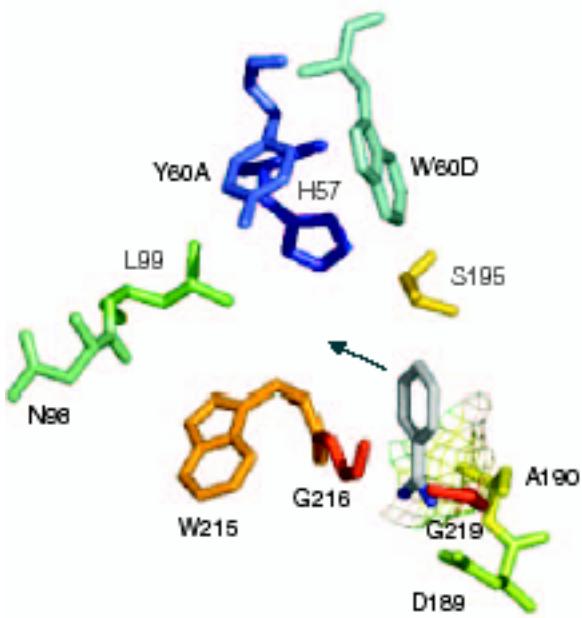


site actif

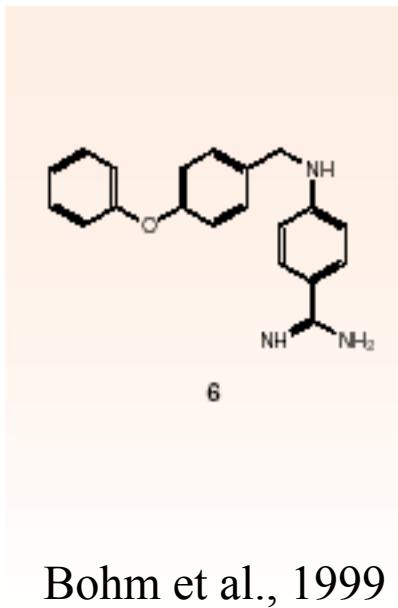


Structure-based

Fragment-based Design



Potency 10 nM to be compared to benzamidine (250 μ M) or p-aminobenzamidine (34 μ M)



Bohm et al., 1999

Thrombin:
involved in blood coagulation

Thrombosis:
a common cause of death in the
industrialized world

- Started from experimentally determined binding mode of **benzamidine** within the pocket of thrombin
- Arrow indicates the direction of fragment growing by *de novo* design.

Need of a molecular docking software having

an efficient research method

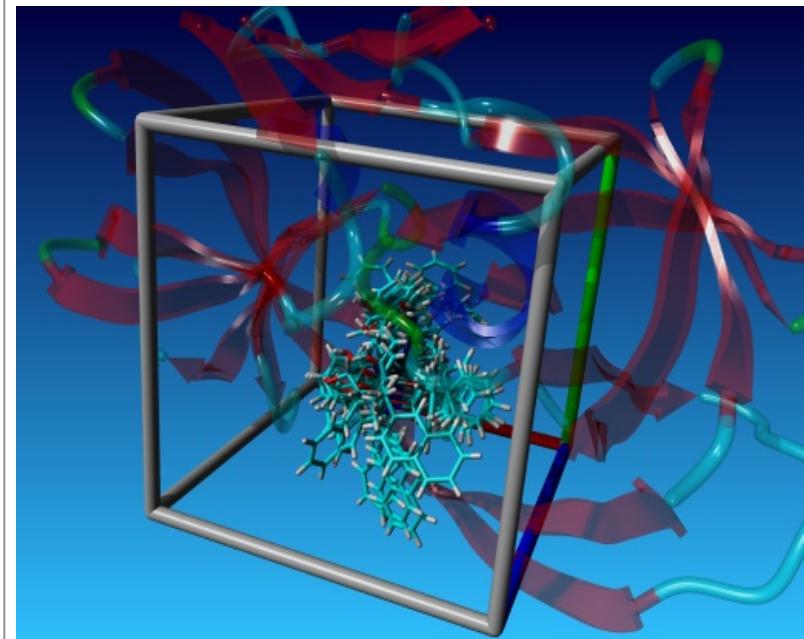
and

an adequate scoring function.

Molecular docking methods



- The molecular docking methods are used to **predict the structures of protein-ligand complexes.**
- The software produces **different binding modes for one ligand.**
 - The **docking method is combined to evaluation methods** to rank the complexes as a function of their predicted affinities.
 - Docking methods use **scoring functions in two ways.**
 - to rank the different positions and orientations of the same ligand
 - to rank different ligands.



Structure-based

Molecular docking methods



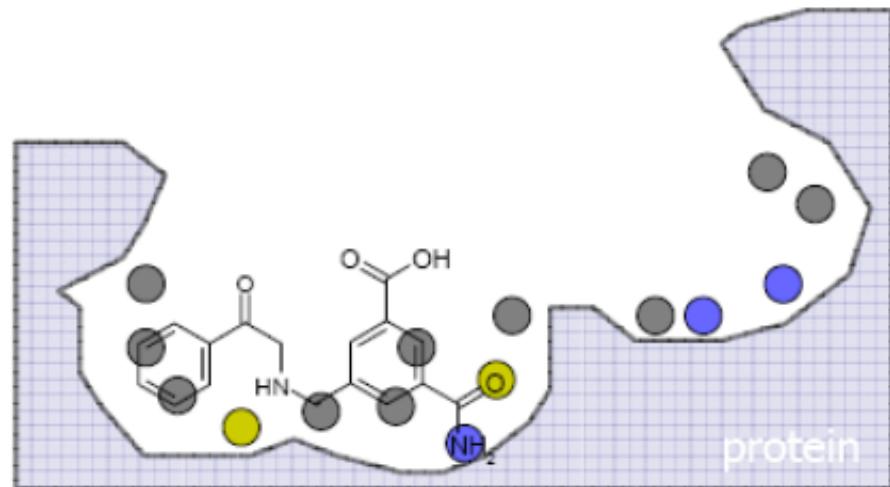
Two types:

- **Geometric methods** based on the shape complementarity (spatial) and on the chemical functionality between the receptor and the ligand (hydrogen bond donor and acceptor, hydrophobic sites)
- Docking methods based on algorithms of **optimisation of functions modelling the molecular interactions.**

Structure-based

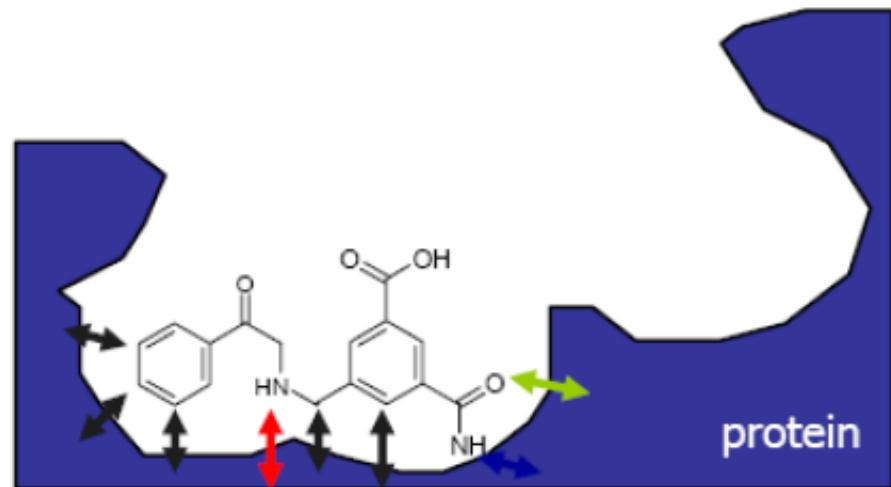
Molecular docking methods

Geometric methods



- Points of interest placed on the ligand and on the protein
- Superposition of the points matching both the ligand and the protein

Optimisation of functions modelling the molecular interactions



- Definition of an energy function for the system
- Optimisation of the interaction energy of the protein-ligand complex

Structure-based

Molecular docking methods

Geometric methods



The ligand is positioned in the binding site of the receptor so as to **make its geometry correspond to that of the receptor**

Binding site may have been previously mapped to identify the chemical sites of interest as well as the zones sterically accessible.

Structure-based

Molecular docking methods

Optimisation of an objective function

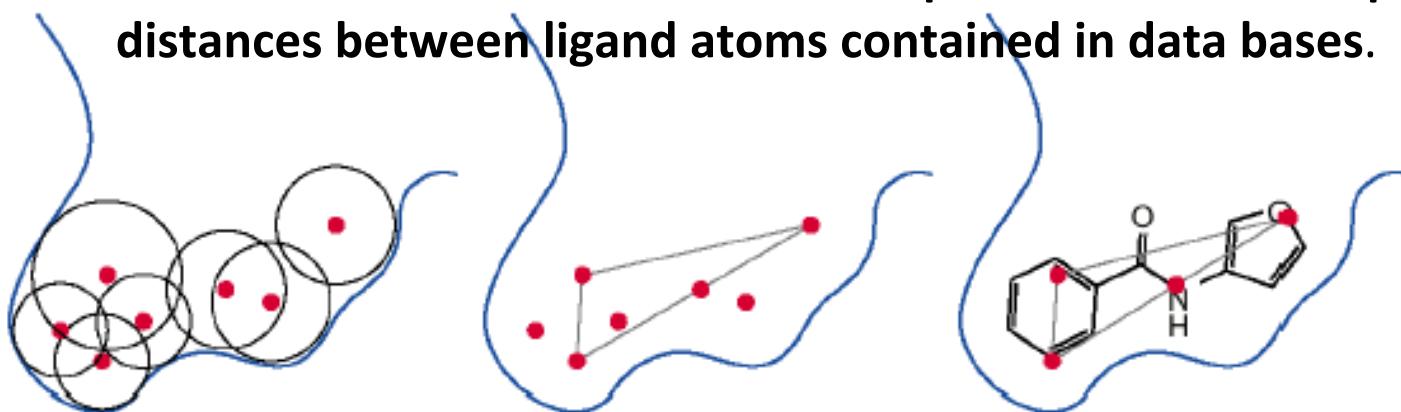
- based on the **optimisation of an objective function with multiple minima**
 - **Search the conformational space** using **a molecular mechanics force field** to evaluate the energy of the complexes.
 - The modelling of ligand/receptor is more detailed



Geometric methods

DOCK computes a “negative” image of the binding site from the molecular surface of the target.

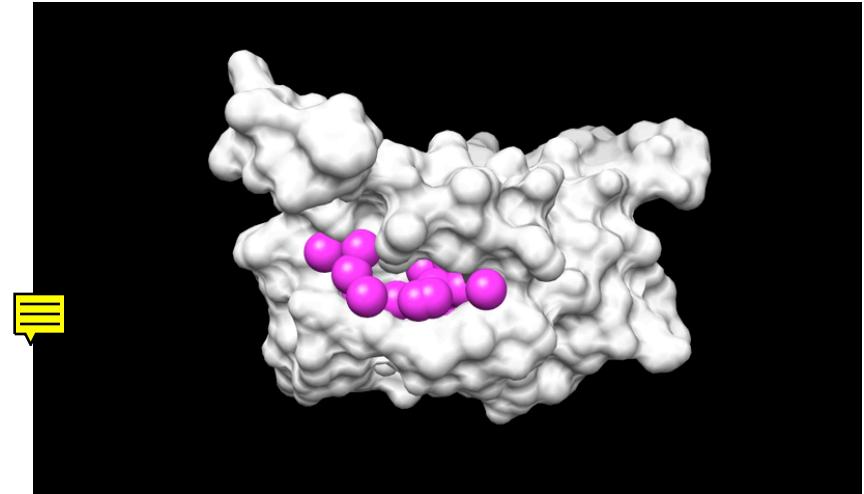
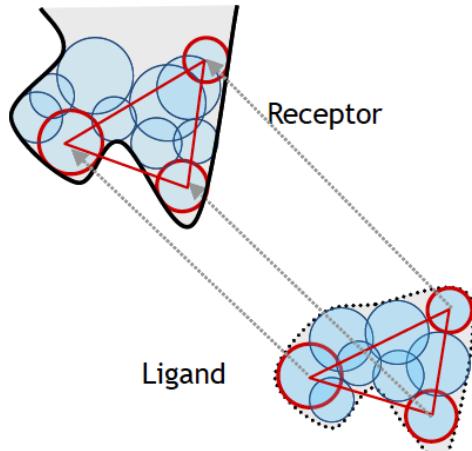
- This image is composed of superimposed spheres of different radii which are in contact with the receptor surface so as to be located between two points of this surface.
- The centres of the spheres are then linked to so as to generate families of positions for which **all distances between sphere centres are equal to distances between ligand atoms contained in data bases.**



Structure-based

Molecular docking methods

Geometric methods



- The spheres represent the volume which could be occupied by the ligand
- The orientation of the ligand is then refined by **a least square fit of the atomic positions relative to the centres of the spheres.**

Structure-based

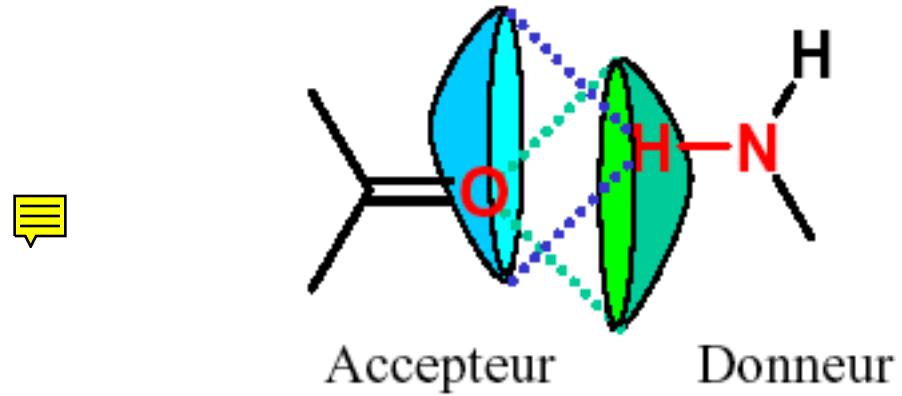
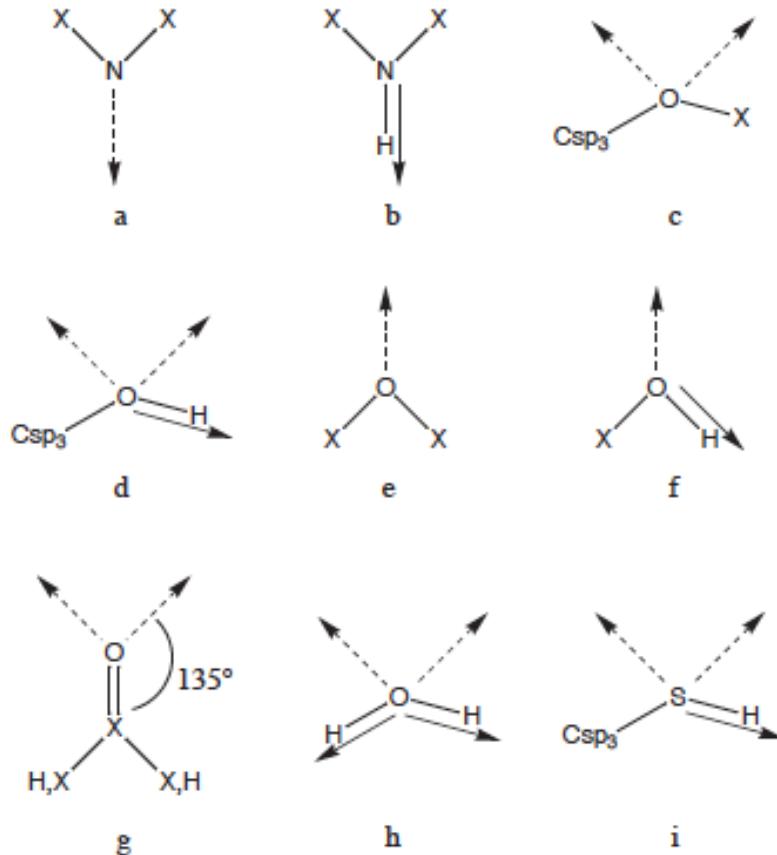
Molecular docking methods

Geometric methods

- Solvation Energy for Exhaustive Docking (SEED)
positions polar and apolar groups or fragments
on the surface of a receptor.

Geometric methods

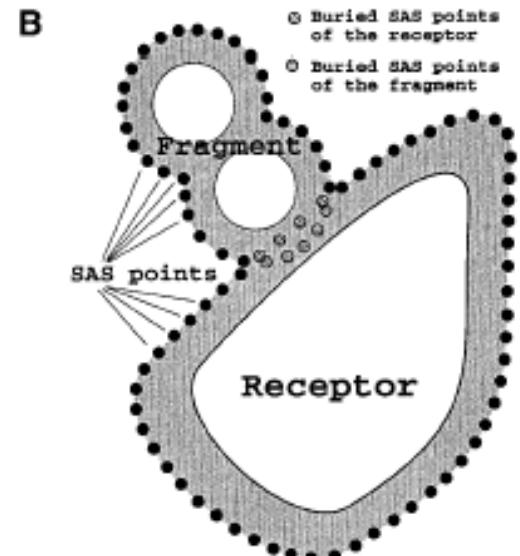
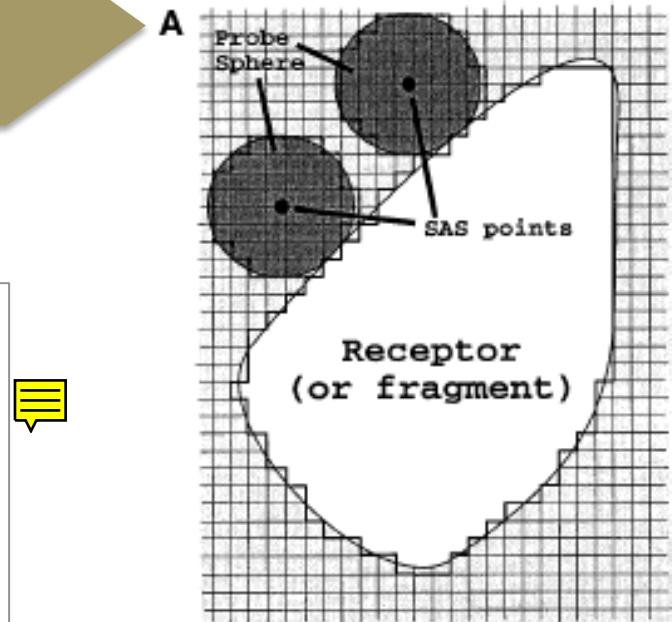
- Vectors are ascribed to **groups of the fragment and of the receptor able to form hydrogen bonds.**



- The docking is performed by **pairing vectors of the fragment and of the receptor** and by positioning the fragment at a distance depending on the type of atoms of the donor and acceptor groups.

Geometric methods

- For **apolar groups** points are uniformly distributed on the **solvent-accessible surface (SAS)** of the fragment and the receptor.
- **Hydrophobic zones** are determined by computing a **desolvation energy** as well as the **van der Waals interaction** between the probe and the ligand/receptor.
- The **vectors** for the apolar interactions for the fragment and the receptor are defined by **joining the best points of the SAS with the centre of the corresponding atom.**
- The docking is performed as for polar groups by pairing vectors of the fragment and the receptor.





Structure-based

Molecular
docking methods

Optimisation of an objective function

Two classes:

depending whether they require a continuous and derivable force field or not.

- ✓ Methods using the **energy minimisation in a force field and/or using molecular dynamics.**
 - ✓ Methods based on **Monte Carlo type algorithm and genetic algorithms** require only energy evaluations.

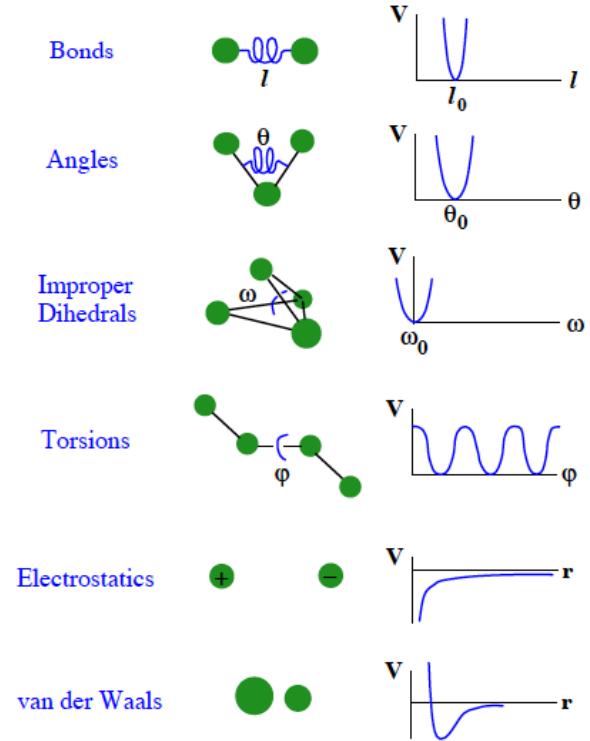
Optimisation of an objective function

Potential energy function (mathematical equation)

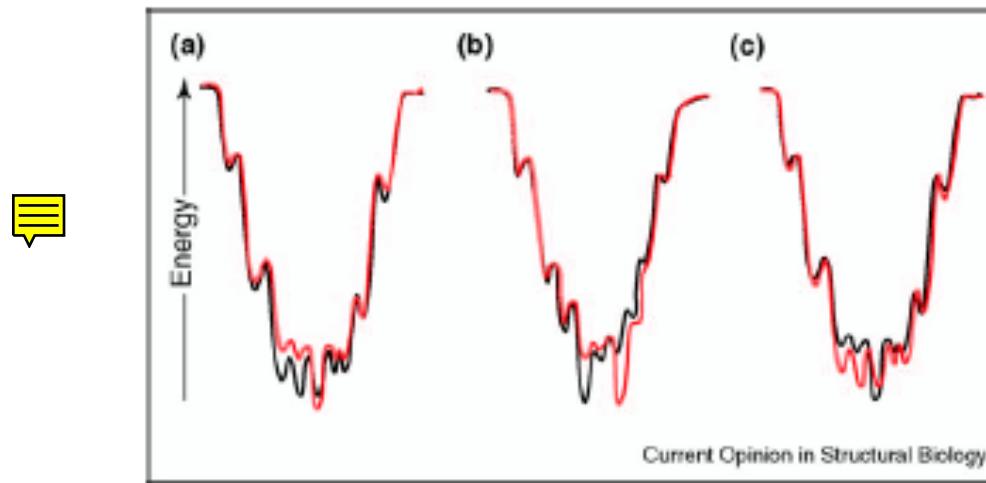
Empirical force field equations and parameters relate chemical structure and conformation to energy

$$\begin{aligned}
 E_{\text{total}} = & \sum_{\text{bonds}} K_r (r - r_{eq})^2 && \text{harmonic bond vibration} \\
 & + \sum_{\text{angles}} K_\varphi (\varphi - \varphi_{eq})^2 && \text{harmonic bond bending} \\
 & + \sum_{\text{dihedrals}} \sum_{\text{all_n}} \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)] && \text{bond rotation (torsions)} \\
 & + \sum_{i < j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{e R_{ij}} \right) && \text{non-bonded interactions}
 \end{aligned}$$

Empirical Potential Energy Function



Optimisation of an objective function

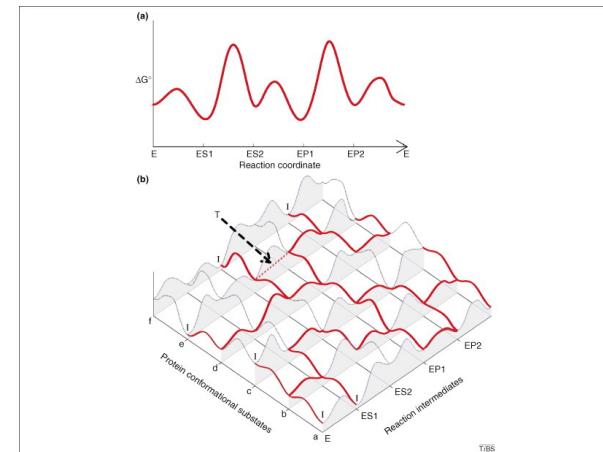


Allosterically reshaping the energy landscape of a protein. In all panels, black represents the landscape in the absence of ligand and red represents the ligand-bound state. (a) Ligand binding to a native state with several energetically comparable states may stabilize one of those states at the expense of others, reducing conformational heterogeneity. (b) A native protein may have two major conformational states and ligand binding may perturb their relative energies, leading to a discrete conformational change. This landscape depicts the paradigmatic T \leftrightarrow R equilibrium of Monod-Wyman-Changeux allosteric theory. (c) A stable native state with a narrow conformational distribution may be perturbed by ligand binding in such a way that a more heterogeneous set of possible conformations becomes accessible. This occurs in proteins that are autoinhibited, such as many kinases and WASP (see text); allosteric ligand binding relieves the inhibiting interaction. These three scenarios are of course only a sampling of the growing array of possible allosteric mechanisms.

Optimisation of an objective function

Methods based on a continuous and derivable force field

- **Multiple Copy Simultaneous search (MCSS)** is a method of stochastic sampling which determines the optimal positions and orientations of functional groups on the surface of a 3D structure of a protein.

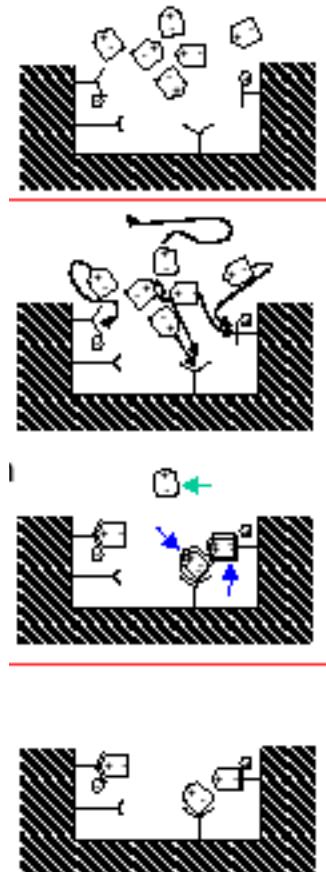




Optimisation of an objective function

Methods based on a continuous and derivable force field

- Several hundreds or thousands of copies are randomly distributed in a sphere or a box whose dimensions encompass the binding site of the receptor.
- Copies are subjected to a minimisation in the force field of the protein and the interactions between group copies are omitted.
 - The positions and orientations of the fragments are regularly compared to eliminate duplicated fragments. These are fragments converging to the same minimum.



Structure-based

Molecular docking methods

Optimisation of an objective function

Methods based on a continuous and derivable force field

Molecular Dynamics (solves the Newton's equation of motion)

- The core problem consists **in finding the global energy minimum** for a ligand-receptor complex.
- A molecular dynamics trajectory can be trapped in an energy well of this surface.



Optimisation of an objective function

Equations du mouvement de Newton

$F_i = m_i \mathbf{a} = m_i d^2 \mathbf{r}_i/dt^2 \rightarrow$ résolues numériquement

À $t = t_0$ position initiale $\mathbf{r}(t_0)$ + vitesse $\mathbf{v}(t_0)$ + Force $\mathbf{F}(t_0)$

Calcul des positions $\mathbf{r}(t_1), \mathbf{r}(t_2) \dots$ et de la trajectoire $\mathbf{r}(t_i)$

$\mathbf{F} \rightarrow$ dérive de l'énergie potentielle

$$m\mathbf{a} = -\nabla U$$

U = interactions effectives entre atomes

- We integrate Newton equation over a time step Δt . Integration means that we generate the trajectory by increments of Δt .

Optimisation of an objective function

- How do we integrate for a given U ?
- A simple way is to write a Taylor series expansion

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \mathbf{v}(t_0) \Delta t + 1/2 \mathbf{a}(t_0) \Delta t^2 + \dots$$

with a similar expression for $\mathbf{r}(t - \Delta t)$



$$\mathbf{r}(t_0 - \Delta t) = \mathbf{r}(t_0) - \mathbf{v}(t_0) \Delta t + 1/2 \mathbf{a}(t_0) \Delta t^2 - \dots$$

- Then add the two expansions

$$\mathbf{r}(t_0 + \Delta t) = 2\mathbf{r}(t_0) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t_0) \Delta t^2 + \dots$$

- This is the Verlet algorithm

- For the acceleration \mathbf{a} we use \mathbf{F}/m

- We evaluate all the terms on the right hand side and obtain the positions at the next step and then repeat the process
- Calculation of the positions $\mathbf{r}(t_2), \mathbf{r}(t_3) \dots$ and of the trajectory $\mathbf{r}(t_i)$

Optimisation of an objective function



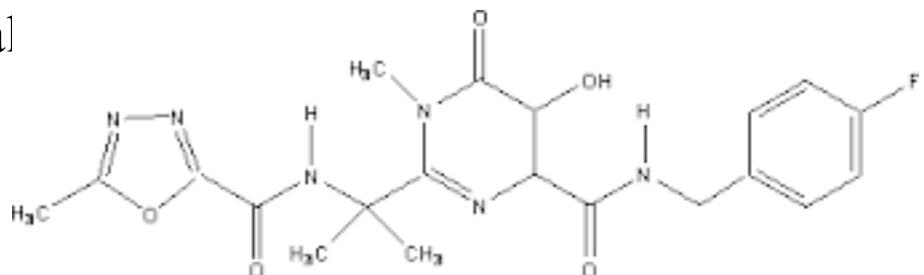
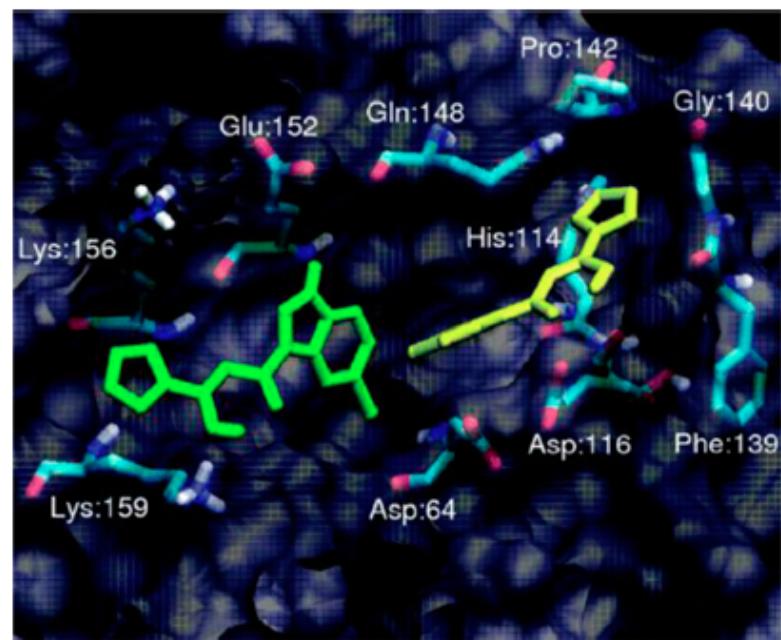
Methods based on a continuous and derivable force field

Relaxed complex scheme (RCS):

a hybrid method of combining docking method with dynamic approach provided by MD simulations

RCS was used to describe a **novel trench** in HIV integrase (Schames et al., 2004)

➔ led to the discovery of the integrase inhibitor **raltegravir** (Merck: Summa et al 2008)



Structure-based

Molecular docking methods

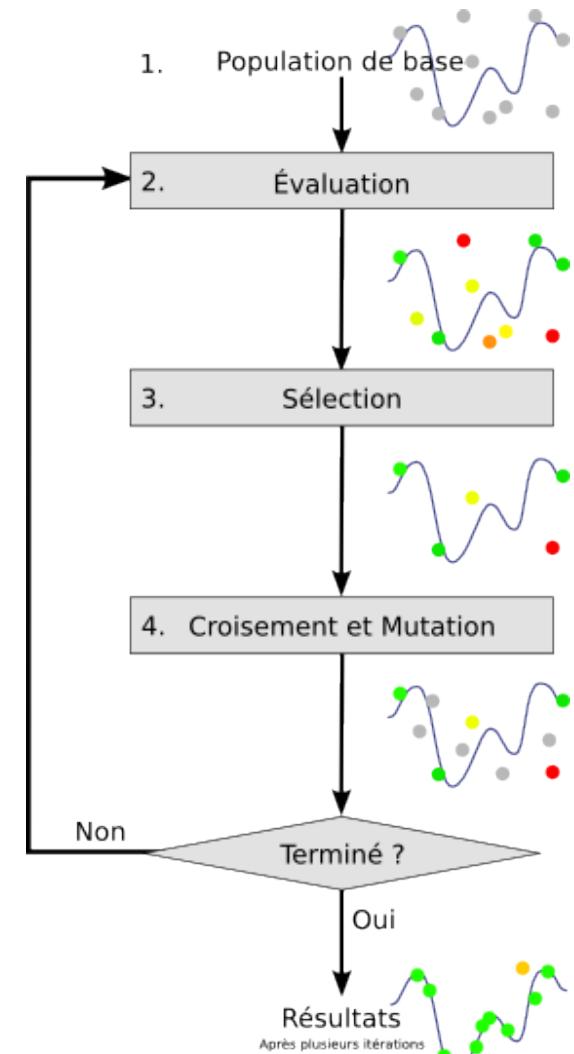
Optimisation of an objective function

Methods requiring only energy evaluations

Genetic algorithms

- A genetic algorithm consists in **making changes to an initial population** by proceeding in a first step in the **reproduction of selected individuals and in transformations of obtained “children”**.

- In a second stage the selection of a certain number of children based on a score is performed.
- This process is repeated iteratively until a score criterion is reached.



Structure-based

Molecular docking methods

Optimisation of an objective function

Genetic algorithms

- Analogy with biological evolution
- Gene = code for the position/conformation of the ligand in the protein binding site



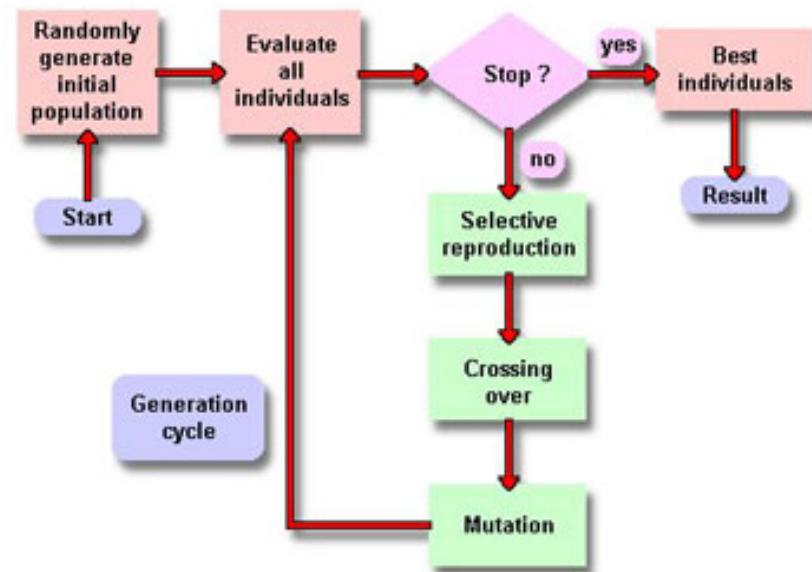
- Genome = ensemble of genes
- Simulation= evolution of the genome by different operations (reproduction, selection pressure)

Optimisation of an objective function

A genetic algorithm is characterised by several properties :

- the way of **representing the solutions** of the problem (“individuals”)
- the criterion of **selection of individuals** so they become parents
- the **genetic operators** applied to individuals of the population **to create new individuals**
 - the **evaluation** of the individuals

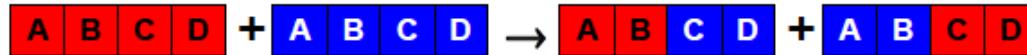
Genetic algorithms



The genetic algorithms are particularly interesting for molecular docking.

Optimisation of an objective function

- Initialization : choice of a population of “genes”
- Selection with a scoring function
- Reproduction:
 - cross-over

 $\boxed{\text{A | B | C | D}} + \boxed{\text{A | B | C | D}} \rightarrow \boxed{\text{A | B | C | D}} + \boxed{\text{A | B | C | D}}$

- mutations

 $\boxed{\text{A | B | C | D}} \rightarrow \boxed{\text{A | B | C | D}}$

- Stops with convergence of the population or the number of generations

Optimisation of an objective function

Genetic algorithms

- Genetic Optimisation Ligand Docking (GOLD) uses a genetic algorithm to search the conformational space.

the way of representing the solutions of the problem (“individuals”)

- A solution corresponds to one  conformation of a ligand/receptor complex.
- An individual is represented by two binary chains containing the information of the ligand and the receptor and two chains of integers for the interactions of hydrogen bond type between the ligand and receptor.

1	0	1	1	0	1	0	0	1	1	1	0	0	1	1	0
2	0	4	7	5	1	6	3	5	4	1	2	0	3	0	7

LIGAND PROTEINE

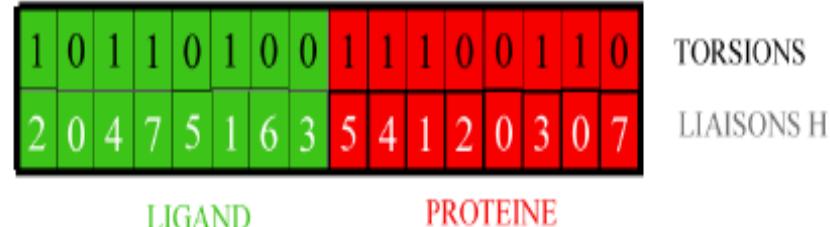
TORSIONS

LIAISONS H

Optimisation of an objective function



Genetic algorithms



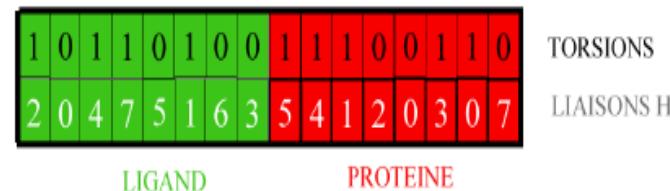
- For the binary chains coding for the conformation each byte codes for a rotation angle about a single bond.
- The two chains of integers code for the correspondence between the hydrogen bond acceptors and donors in the receptor and ligand. The first chain codes for the correspondences between the acceptor sites of the ligand with the donor sites of the receptor and vice-versa for the second chain.
- V being the integer value at position P in the first chain means that the Pth ligand acceptor is in interaction with the Vth donor in the protein.
- If V is zero then the acceptor site P of the ligand is in interaction with no donor site of the receptor.

Structure-based

Molecular docking methods

Optimisation of an objective function

Genetic algorithms



There are **different genetic operators used to create new individuals** after each iteration.

Optimisation of an objective function

- An operator of chromosomal crossing creating two new individuals from two parents.

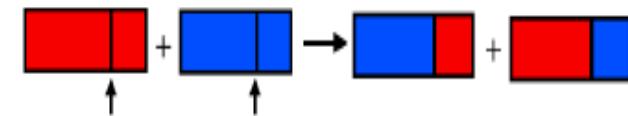
If A and B are parents and AB and BA the children, AB will have its genotype composed of the first part of A genotype and the second part of B genotype.

- A mutation operator consists in randomly modifying elements of the genotype. In a chain of binaries it consists in transforming a bit from 0 to 1.

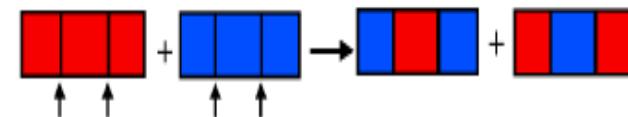
The choice of the operator to be applied on an individual is randomly

Genetic algorithms

Croisement à un point



Croisement à deux points



Opérateur de mutation

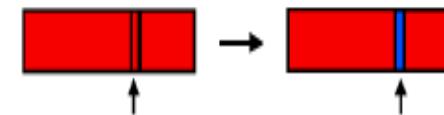
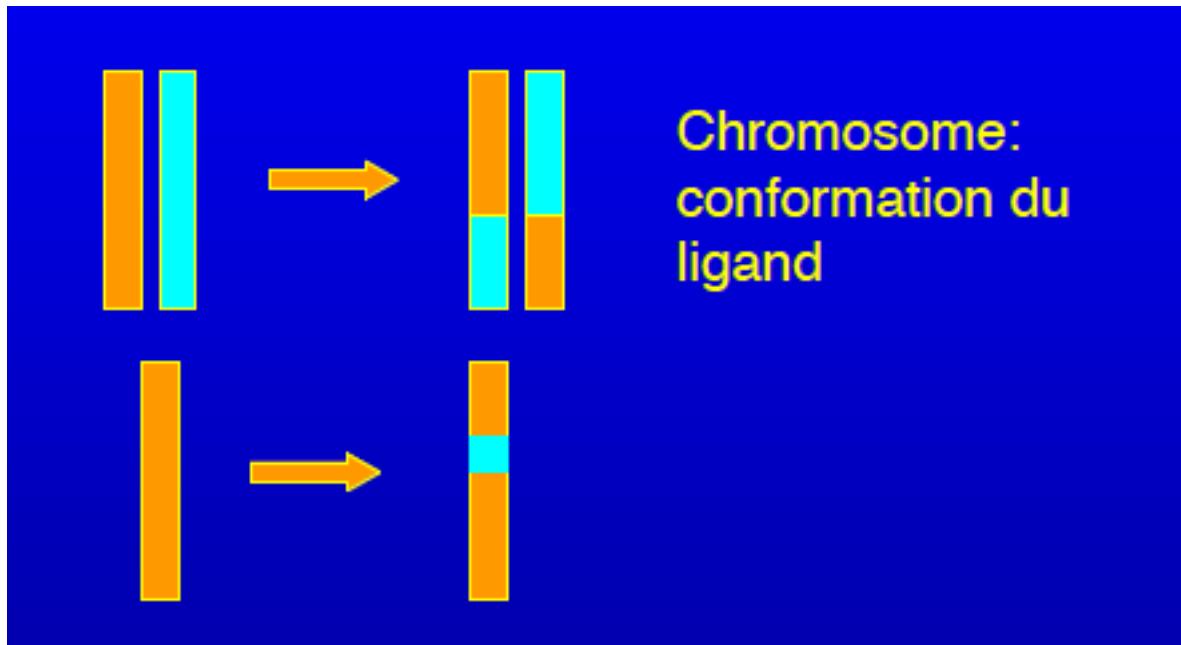


Figure 17: Opérateurs génétiques : opérateurs de croisement et de mutation. Pour le croisement à un point le programme détermine un point où les chaînes des parents seront coupées et crée deux enfants portant chacun le début de la chaîne d'un parent et la fin de celle de l'autre. Le croisement à deux points est identique mais coupe les chaînes de parents en trois parties. Enfin l'opérateur de mutation consiste à changer la valeur d'une position prise au hasard dans un chromosome.

Structure-based

Molecular docking methods

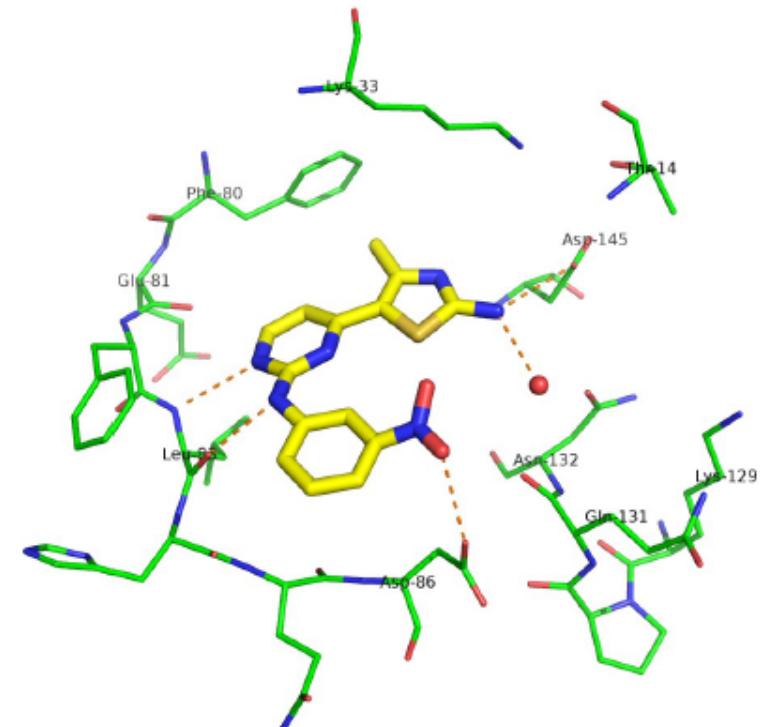
Optimisation of an objective function



Optimisation of an objective function

- The two chains of integers code for the correspondence between the hydrogen bond acceptors and donors in the receptor and ligand. The first chain codes for the correspondences between the acceptor sites of the ligand with the donor sites of the receptor and vice-versa for the second chain.
- V being the integer value at position P in the first chain means that the Pth ligand acceptor is in interaction with the Vth donor in the protein.
- If V is zero then the acceptor site P of the ligand is in interaction with no donor site of the receptor.

cyclin-dependent kinase 2 (CDK2)



CK7

Optimisation of an objective function

- par algorithme génétique (e.g. Gold)

