Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

• **The energy evaluation and ranking** is used as a sieve

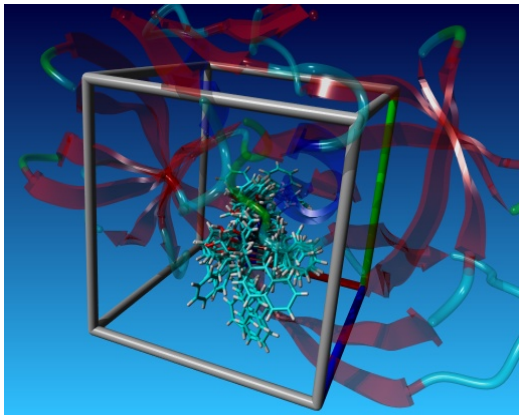
to filter the numerous positions of one single ligand,
i.e. conformations yielded by the molecular docking process

and to evaluate the scoring of different ligands

• The interactions receptor-ligand depend on the ligand, receptor and ion concentrations in solution :

$$RL \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} R + L$$

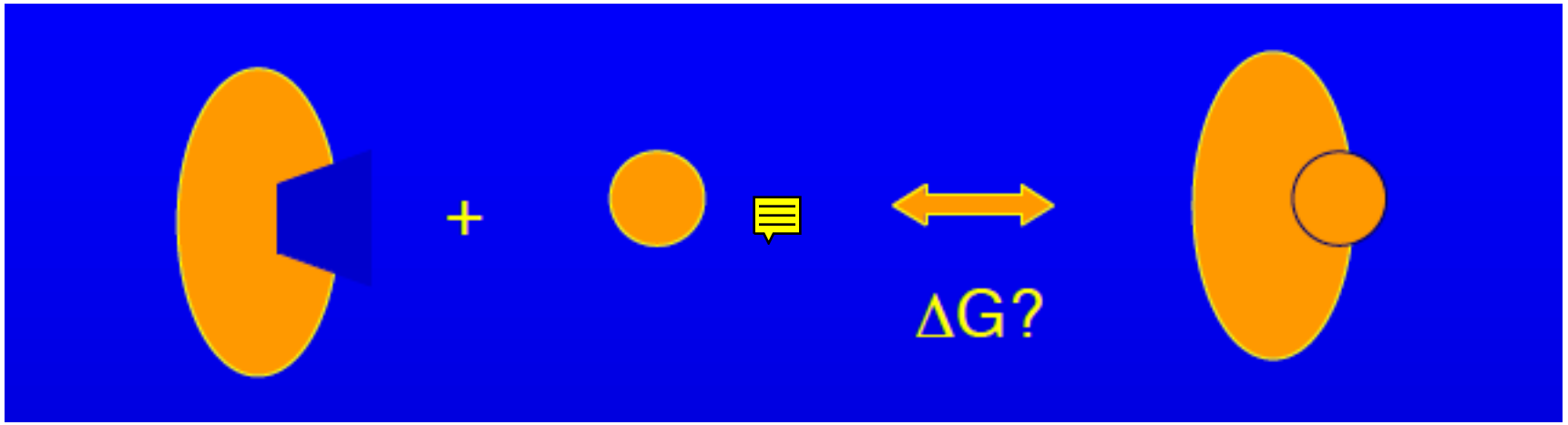
The dissociation constant is : $K_d = \frac{[R][L]}{[RL]}$



Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

$$K_d = \frac{[R][L]}{[RL]}$$

- K_d is the constant which is usually used to describe the affinity of the ligand for the receptor. The smaller K_d the higher the affinity.

- The dissociation constant is related to the free energy change of the complex formation by the following expression :

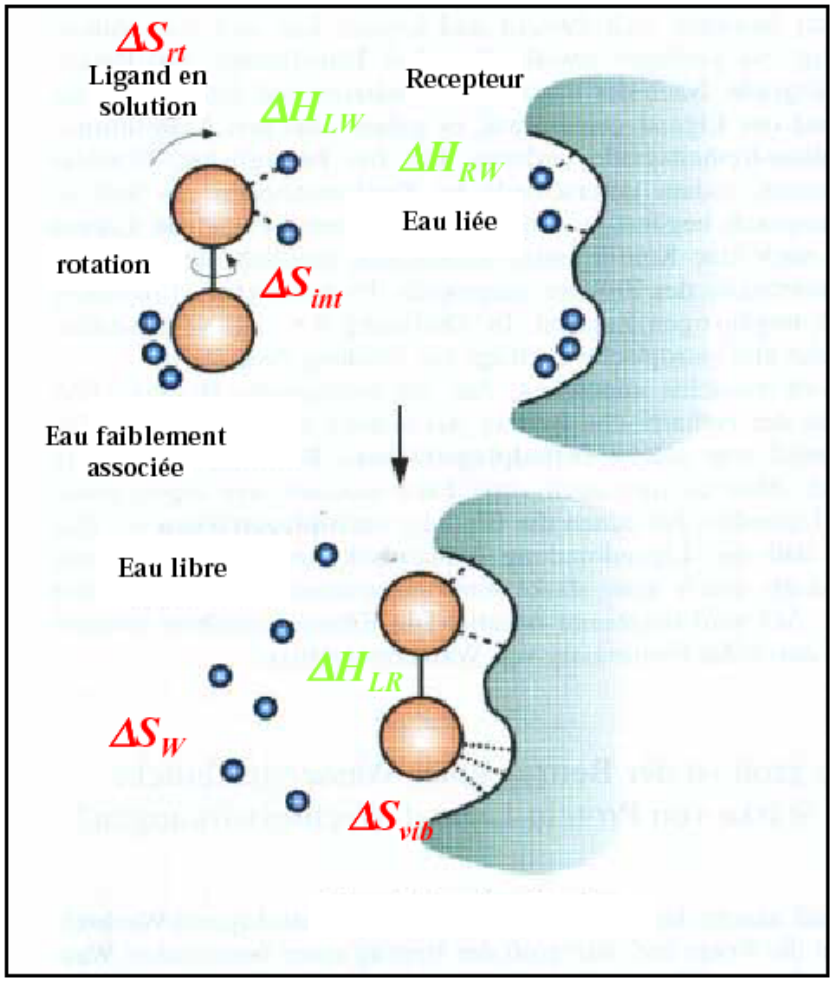
$$RL \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} R + L$$

$$\Delta G = \Delta G^0 - RTlnK_d$$

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



Energie libre

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

Entropie

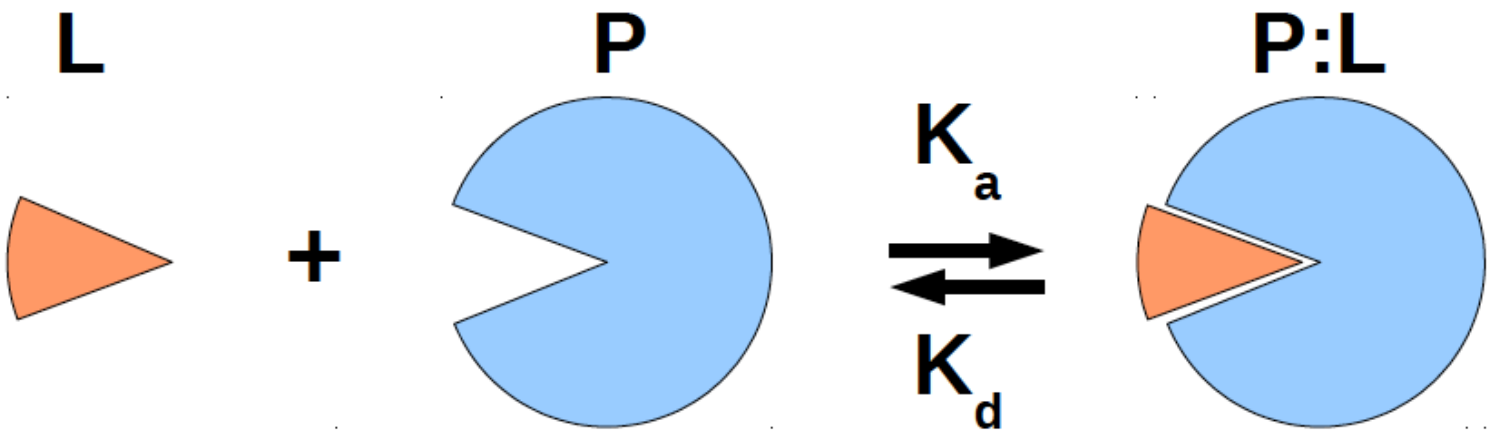
Enthalpie

ΔH = variation of internal energy (enthalpy)
(bonds, interactions, deformations, ...)
ΔS = variation of entropy
(disorder, degrees of freedom, ...)

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



Ratio of concentrations at equilibrium

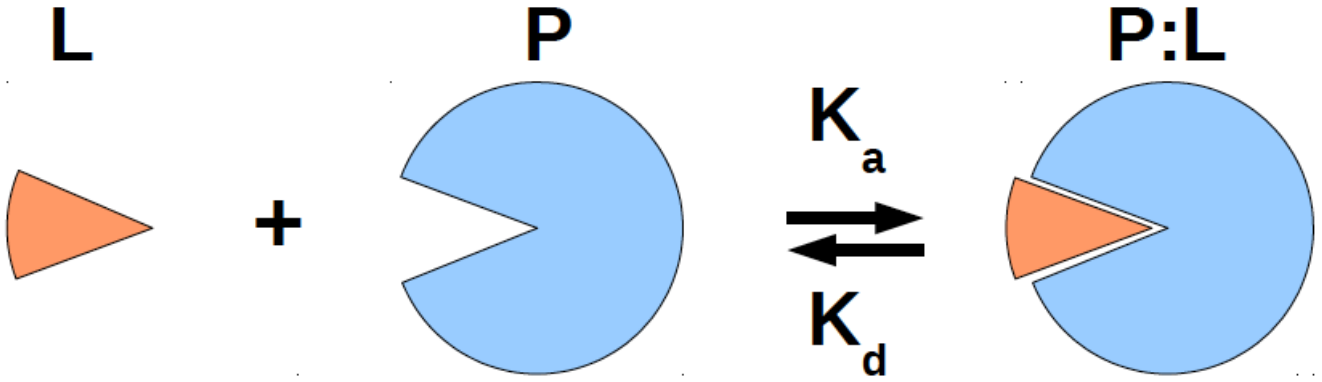
$K_a = [P:L]/[P][L] = 1/K_d$

The higher Ka the stronger is the association

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



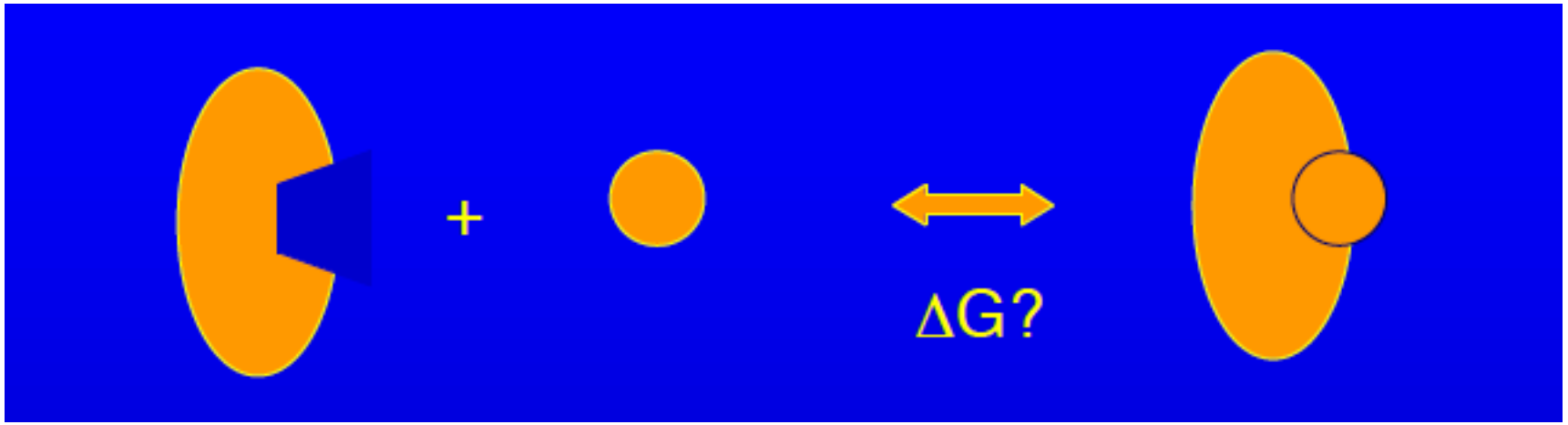
$$\Delta G^\circ = - RT \ln K \quad \text{ou} \quad K = \exp(-\Delta G^\circ / RT)$$

- $K>1$ et $\Delta G^\circ<0$: Equilibrium in favor of P:L
- $K<1$ et $\Delta G^\circ>0$: Equilibrium in favor of P and L
- $K=1$ et $\Delta G^\circ=0$: [P:L] = [P][L]

Structure-based

Scoring Functions

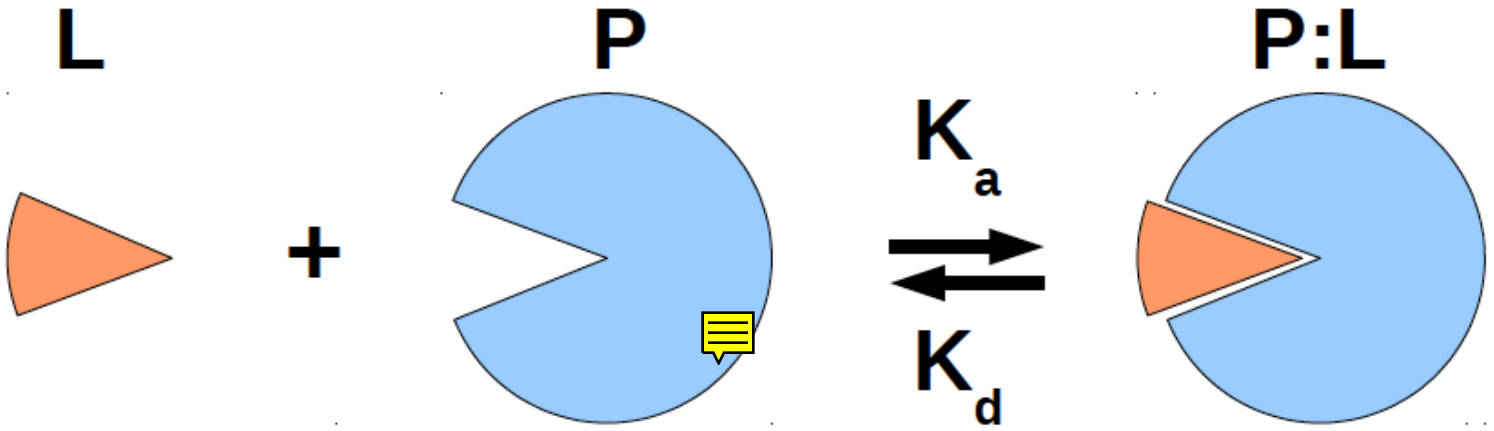
Energy evaluations and ranking of the generated conformations



Structure-based

Scoring functions

Energy evaluations and ranking of the generated conformations



K (mol^{-1})	1E+9	1E+6	1E+3	1E+0	1E-3	1E-6	1E-9
ΔG° (kJ/mol)	-51.35	-34.23	-17.12	0	17.12	34.23	51.35

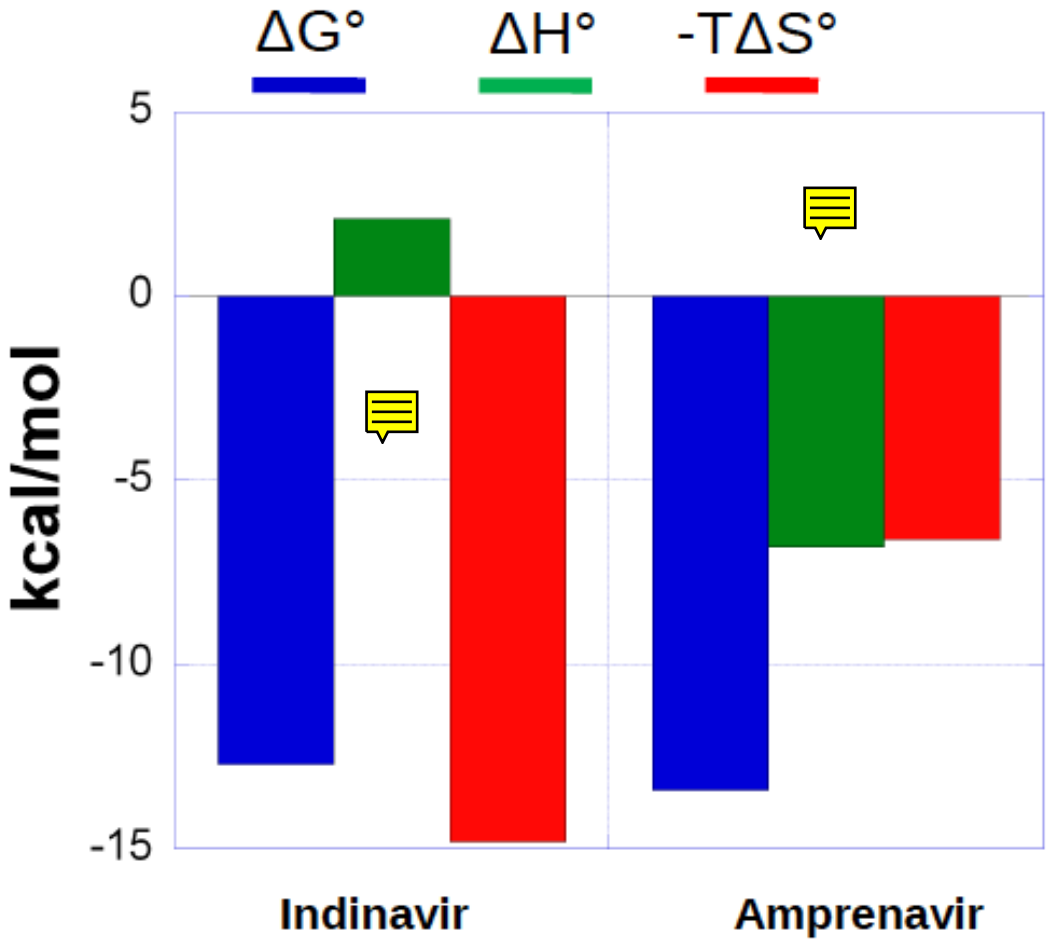
Pharmaceutical ligands K ≈ 10^9 mol^{-1}

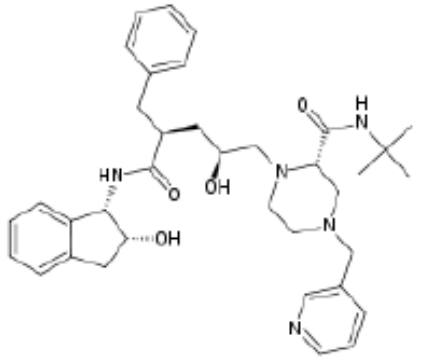
Structure-based

Scoring Functions

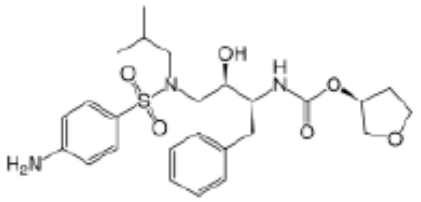
Energy evaluations and ranking of the generated conformations

example : HIV protease inhibitors





Indinavir: entropie



Amprenavir: enthalpie et entropie

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Experimental measures

- Different biophysical techniques (potentiometry, spectrophotometry, mass spectrometry, NMR, calorimetry) to measure K or ΔG

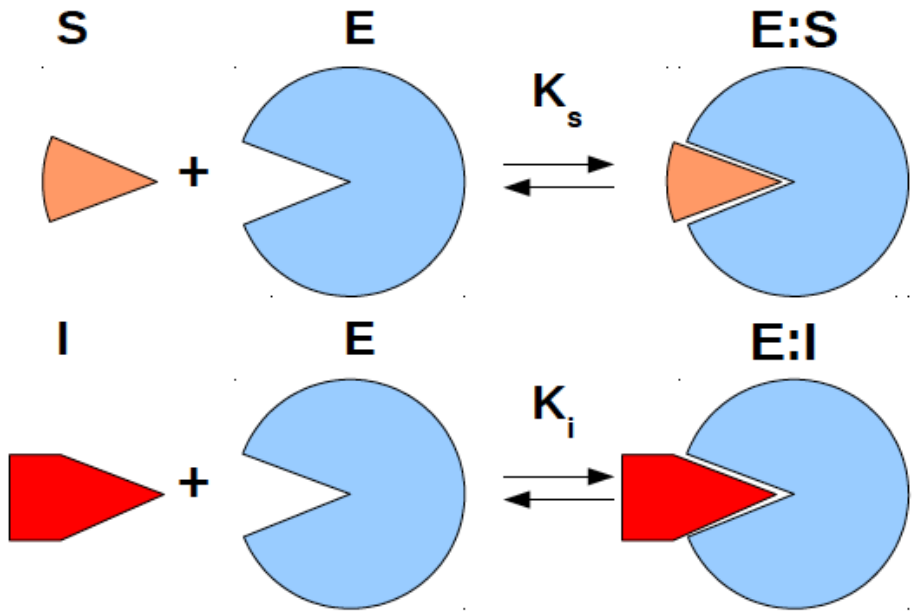
- Calorimetry also gives ΔH and ΔS

- allows quantification of the protein-ligand interaction

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



Case of competitive inhibition

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_s}}$$

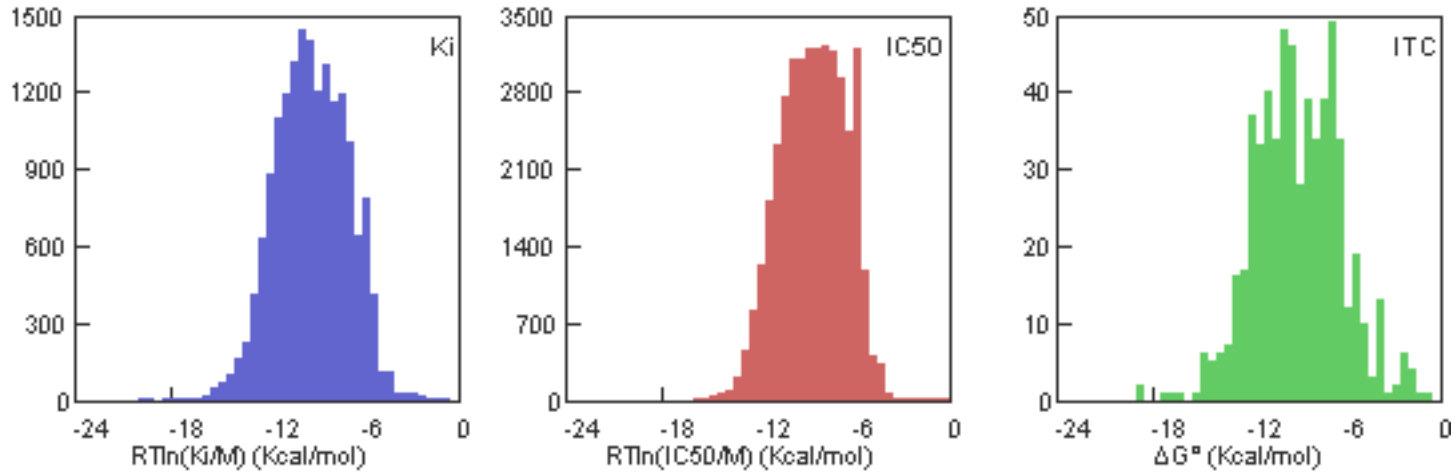
- K_s = dissociation constant of substrate S
- K_i = dissociation constant of inhibitor I
- IC_{50} = Inhibitor concentration for 50% inhibition

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Public data bases of inhibition constants



K_i from 10^{-1} to 10^{-15} M

ΔG from -5 to - 85 kJ/mol

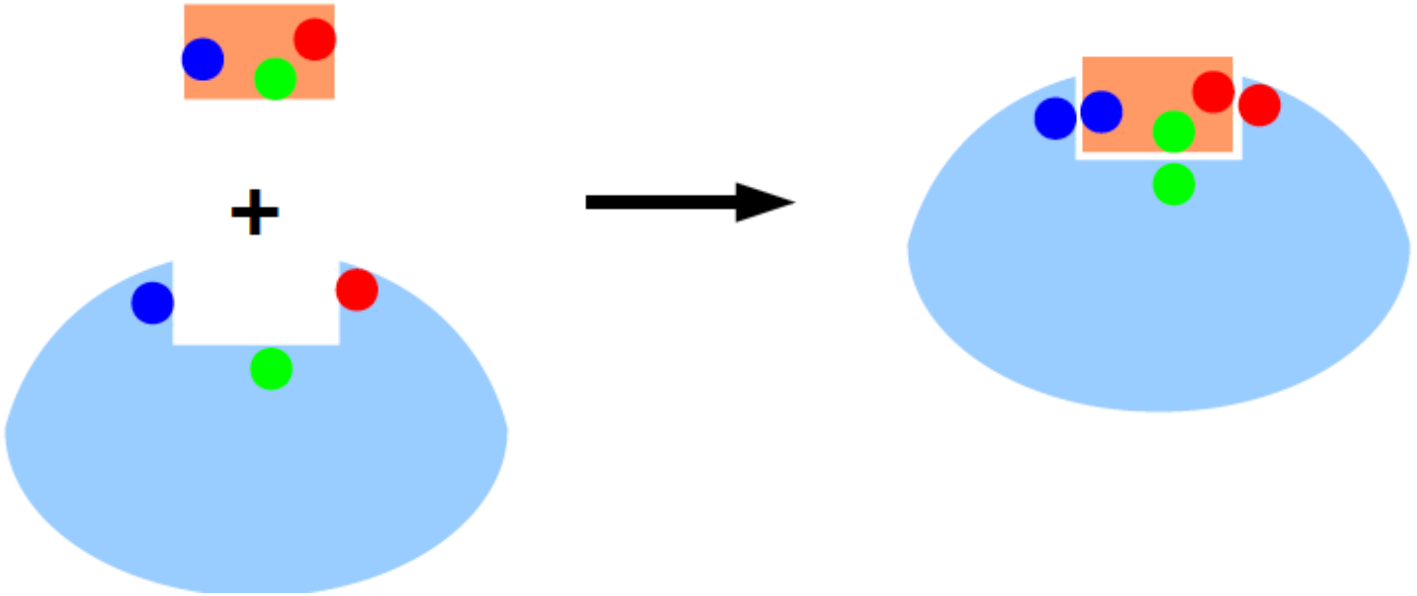
valeurs typiques : $K_i = 10^{-8}$ M ; ΔG = - 45 kJ/mol

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Molecular basis of protein-ligand interactions



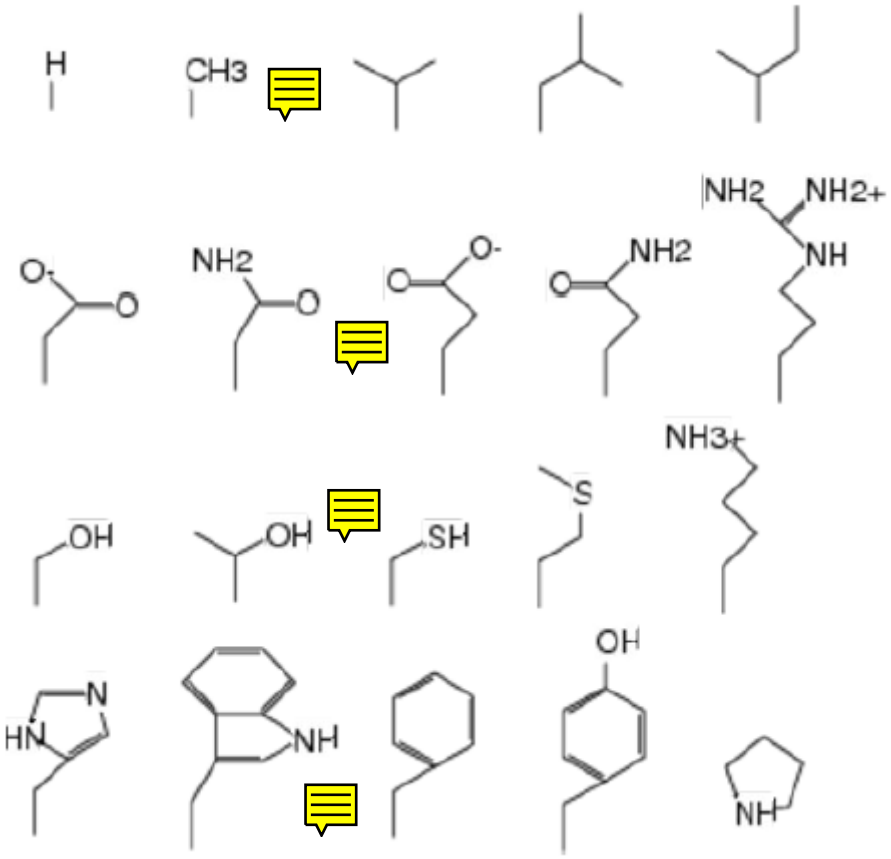
Non-covalent Interactions between complementary chemical groups

Structure-based

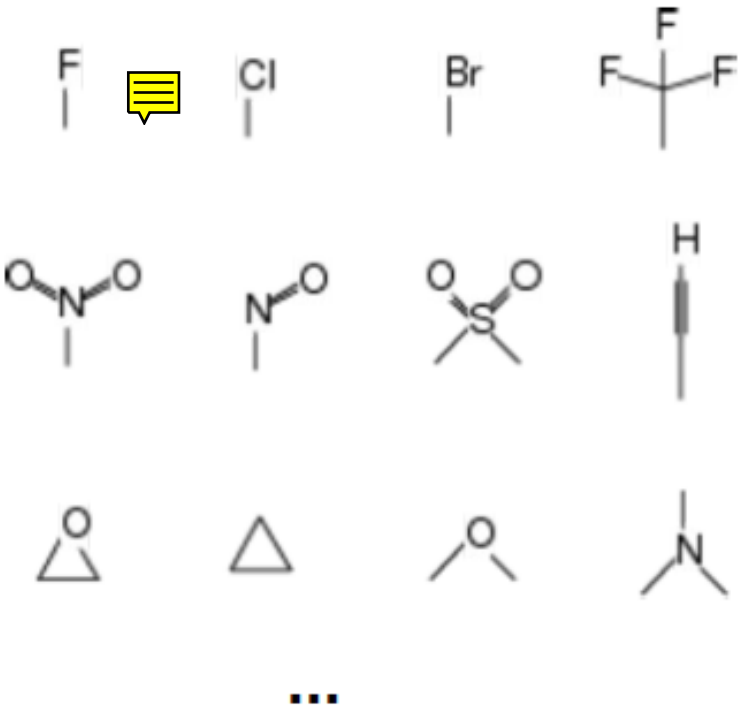
Scoring Functions

Energy evaluations and ranking of the generated conformations

Chemical groups in proteins



Chemical groups in ligands (those found in proteins + …



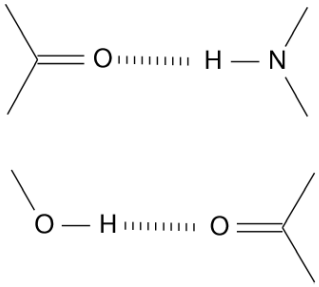
Structure-based

Scoring Functions

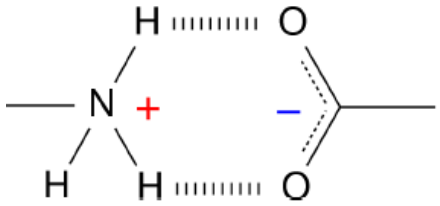
Energy evaluations and ranking of the generated conformations

Non covalent interactions

1. Hydrogen bonds

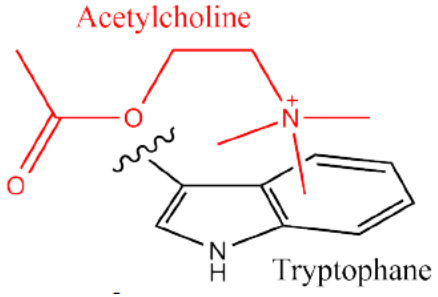


2. Ionic interactions



3. Apolar interactions

4. Other (π–π, cation-π)

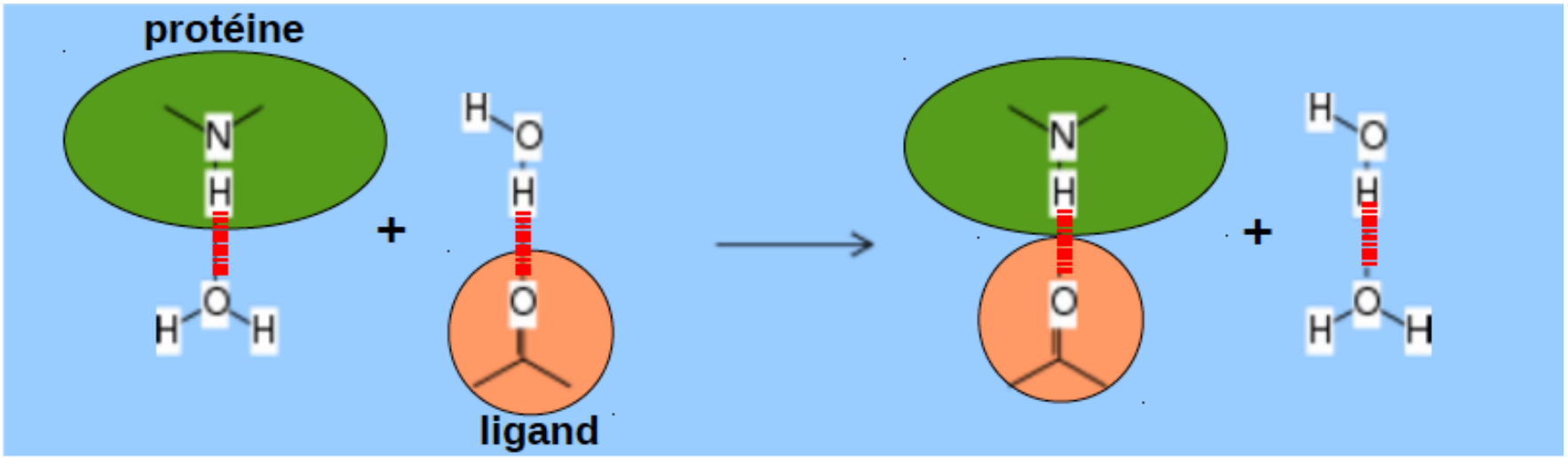


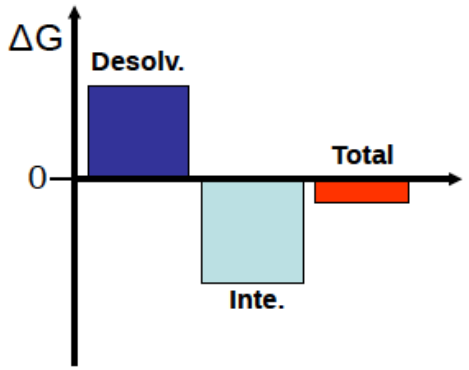
Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Role of solvent: polar interactions





- desolvation : unfavorable
- Protein-ligand polar interactions: favorable

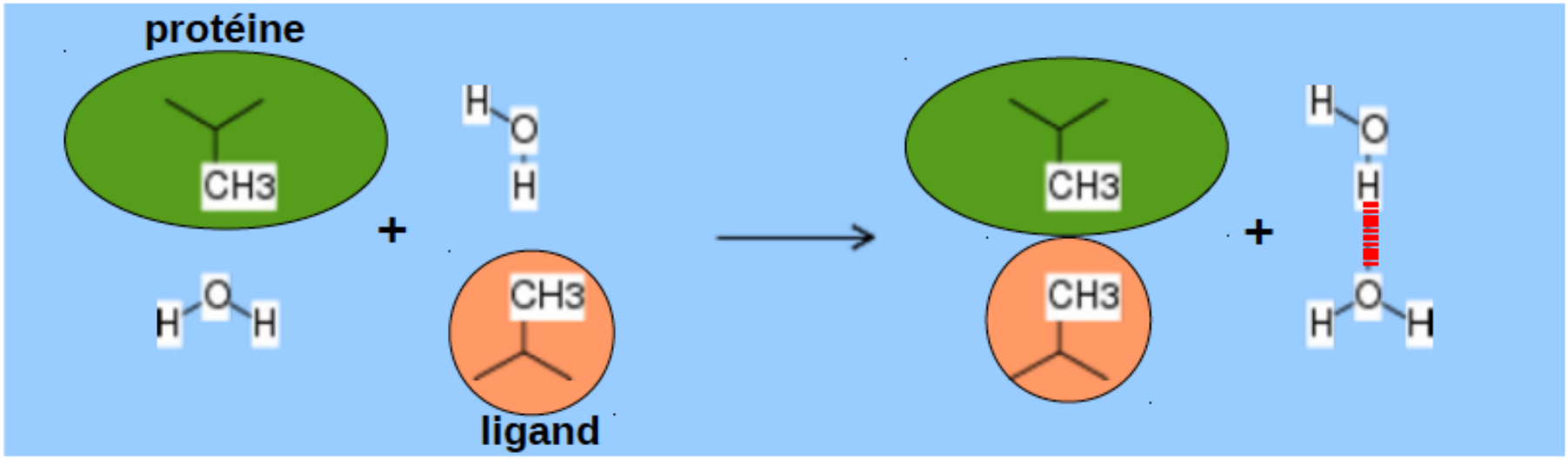
Balance close to zero

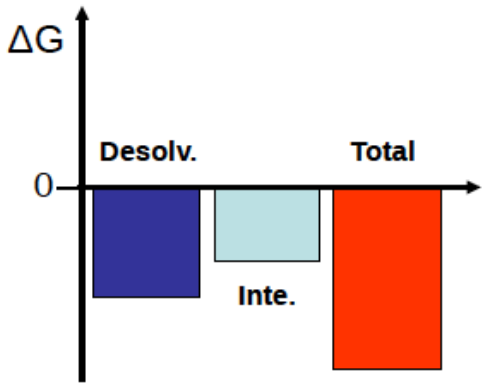
Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Role of solvent: hydrophobic interactions





- desolvation : favorable
- Protein-ligand polar interactions: favorable

Balance favorable

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

Role of solvent

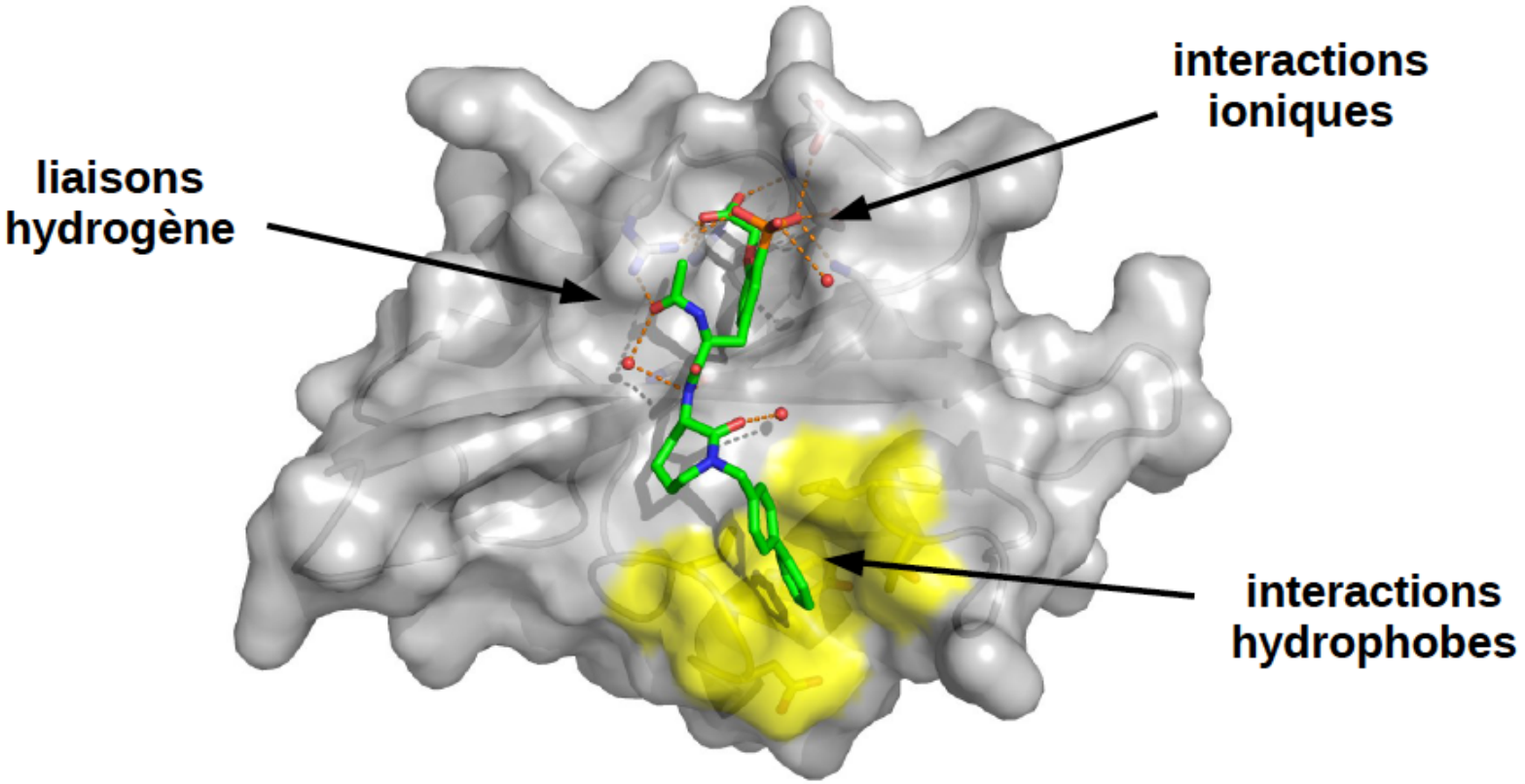
- Solvent contribution to the association of polar groups is unfavorable and can compensate the formed interactions
- Solvent contribution to the association of apolar groups is favorable
- Apolar interactions make a important contribution in an indirect manner to the stability of the protein-ligand complex
- However apolar interactions are less specific than polar interactions

Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations

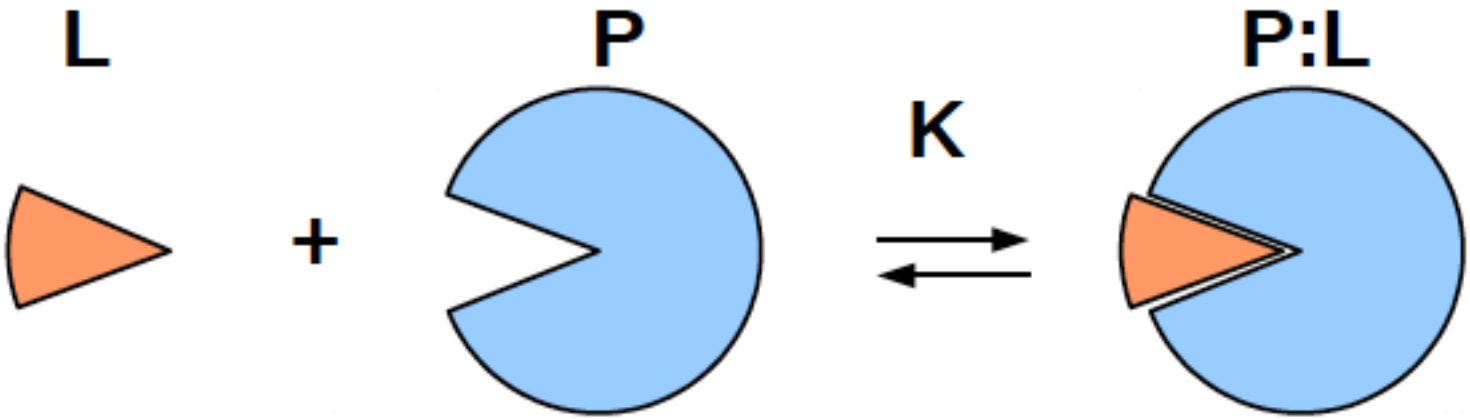
SH2domain of SRC kinase bound to RU82197 inhibitor



Structure-based

Scoring Functions

Energy evaluations and ranking of the generated conformations



$$\Delta G^\circ = - RT \ln K \approx ?$$

Scoring: approximation of the free energy of binding of the protein-ligand complex

Structure-based

Scoring Functions

Empirical scoring functions

There are three families of scoring functions which are used in the receptor-based methods.

- **Empirical scoring functions** obtained by multiple regression and approaches based on a knowledge base

- Approaches estimating the ligand/receptor interaction with an **approximation of the free energy change based on a force field**

- **Linear methods and methods of perturbation** of the free energy

Structure-based

Scoring Functions

Empirical scoring functions

• **Empirical functions and functions based on a knowledge base** rest on the **use of a data base of structures** of ligand/receptor complexes.

• **In the case of empirical functions** the data base of structural data is used during the development of the function so as to calibrate it.

• **In the knowledge base approaches** the data base is used to search the frequencies of distribution observed for certain interactions in the experimental structures. The data base is the centre of the evaluation method.

Structure-based

Scoring Functions

Empirical scoring functions

• **Empirical functions and functions based on a knowledge base** rest on the **use of a data base of structures** of ligand/receptor complexes.

• **In the case of empirical functions** the data base of structural data is used during the development of the function so as to calibrate it.

• **In the knowledge base approaches** the data base is used to search the frequencies of distribution observed for certain interactions in the experimental structures. The data base is the centre of the evaluation method.

Structure-based

Scoring Functions

Empirical scoring functions

$$\Delta G \approx S = \Sigma_i k_i F_i$$

- F_i : Function ofligand/protein coordinates describing one type of interactions (H bonds, ionic, hydrophobic, …)

- k_i : constant weighing the interaction type F_i by adjustment with experimental data

– structure of the complex
– K_a

Structure-based

Scoring Functions

Empirical scoring functions

$$\Delta G \approx S = \Sigma_i k_i F_i$$

Interaction	k_i	F_i
Liaison hydrogène	ΔG_{hb}	$f(\Delta R, \Delta\alpha)$
Interaction ionique	ΔG_{ionic}	$f(\Delta R, \Delta\alpha)$
Interaction hydrophobe	ΔG_{lipo}	A_{lipo}
Entropie translation/rotation	ΔG_0	
Entropie conformation	ΔG_{rot}	N_{rot}

Structure-based

Scoring Functions

Empirical scoring functions

$$\Delta G_{liaison} = \Delta G_0 + \Delta G_{lh}\Sigma_{lh} f(\Delta R, \Delta\alpha) + \Delta G_{ion}\Sigma_{ion} f(\Delta R, \Delta\alpha) + \Delta G_{lipo}|A_{lipo}| + \Delta G_{rot} NROT$$

- ΔG_0 is a constant term which does not depend directly on the specific interactions between the receptor and the ligand such as the translational and rotational entropy loss of the molecules upon complex formation.
- ΔG_{hb} is a contribution to the ideal hydrogen bond. f(ΔR,$\Delta\alpha$) is a penalty function taking into account the deviation from the ideal geometry of a hydrogen bond.
- ΔG_{ion} is the contribution of an ideal ionic interaction
- ΔG_{lipo} represents the contribution of the hydrophobic interactions which are supposed to be proportional to A_{lipo}. A_{lipo} is the contact surface between the ligand and the receptor.
- ΔG_{rot} represents the entropic penalty related to the loss of internal degrees of freedom of the ligand. NROT is the number of rotatable bonds in the ligands.

Structure-based

Scoring Functions

Empirical scoring functions

$$\Delta G_{liaison} = \Delta G_0 + \Delta G_{lh} \Sigma_{lh} f(\Delta R, \Delta\alpha) + \Delta G_{ion} \Sigma_{ion} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} |A_{lipo}| + \Delta G_{rot} NROT$$

• The 5 adjustable parameters ΔG_0, ΔG_{hb}, ΔG_{ion}, ΔG_{lipo} and ΔG_{rot} have been obtained by fitting on a data base of 45 protein/ligand complexes and for which a K_d is known.

• This function is at fault in cases when the desolvation offsets the protein/ligand interactions, when several aromatic group contacts are formed.

$$\Delta G_{liaison} = \Delta G_0 + \Delta G_{polaire} + \Delta G_{apolaire} + \Delta G_{solv} + \Delta G_{rot} NROT$$

Structure-based

Scoring Functions

Knowledge-based functions

Principle:

– Force of an interaction is related to its frequency of observation
In a data base of structures
– No need of calibration with experimental dtaa

$$\Delta G \approx S = \Sigma_i \Sigma_j w_{xy}(r_{ij})^2$$

$$w_{xy}(r)^2 = - R\,T\,\ln g_{xy}(r)^2$$

- g_{xy} = distribution of the distances between atome of type x and y
- w_{xy} = Inter atom potential

Structure-based

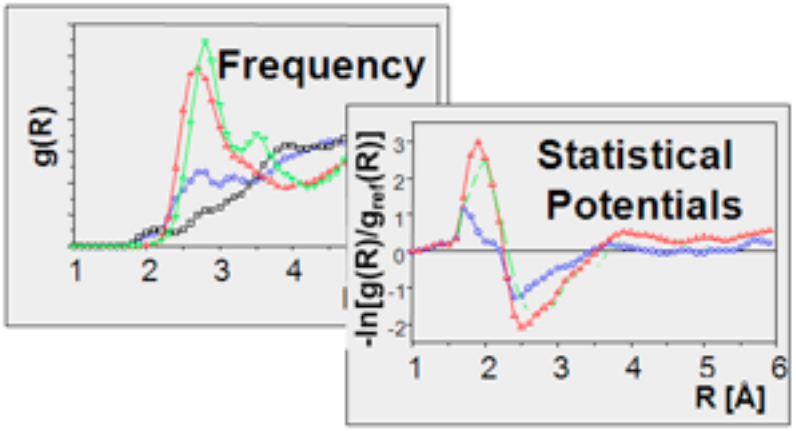
Scoring Functions

Knowledge-based functions

• These methods rest on the belief that a structural sample large enough enables to deduce rules and general principles which are implicitly included in the database.

• Only the interactions which are observed with a high frequency are considered as favourable.

• The frequency distributions are converted into energy by an inverse Boltzmann law :



$$\Delta W_{AB}(R_C) = -RTln[p_{AB}(r \leq R_C)/p_{XX}(r \leq R_C)]$$

Structure-based

Scoring Functions

Force field based Functions

- Based on the physics principles
- Depend on coordinates (ligand/protein) and on the force field parameters

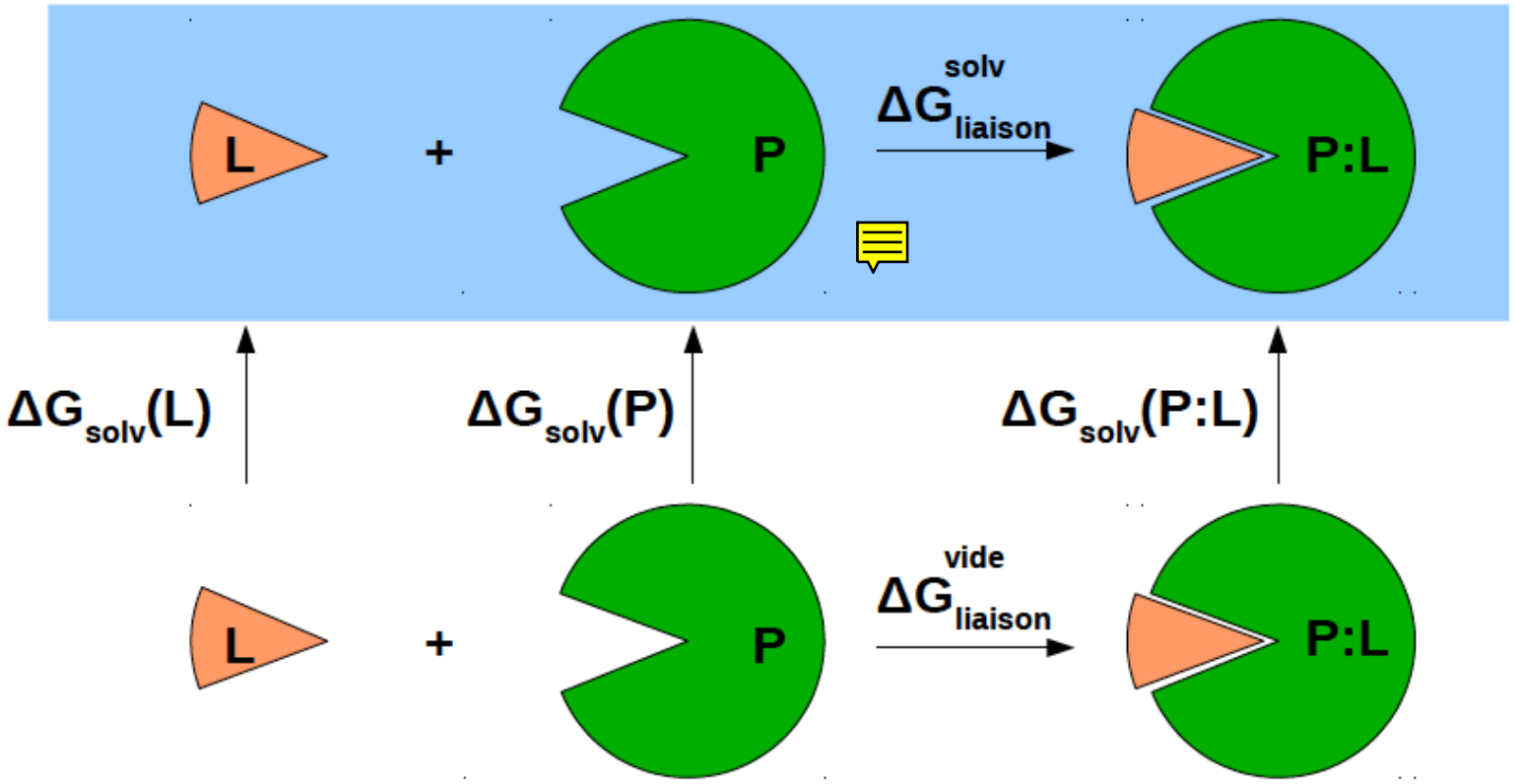
$$E = \sum_{liaisons} k_l (l-l_0)^2 + \sum_{angles} k_a (a-a_0)^2 + \sum_{torsions} k_t (1+\cos(nt-\tau))$$
$$+ \sum_{i,j\ non\ liés} q_i q_j / (4\pi\varepsilon_0 r_{ij}) + \varepsilon_{ij} \left[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^{6} \right]$$

Structure-based

Scoring Functions

Force field based Functions

Thermodynamic cycle to compute the free energy of binding



$$\Delta G^{solv}_{liaison} = \Delta G^{vide}_{liaison} + \Delta G_{solv}(P:L) - \Delta G_{solv}(P) - \Delta G_{solv}(L)$$

Structure-based

Scoring Functions

Force field based Functions

méthode MMGBSA

$$\Delta G^{solv}_{liaison} = \Delta G^{vide}_{liaison} + \Delta G_{solv}(P:L) - \Delta G_{solv}(P) - \Delta G_{solv}(L)$$

$$\rightarrow \Delta G^{solv}_{liaison} \approx \Delta E^{vide}_{MM} - T\Delta S^{vide}_{MM} + \Delta G^{elec}_{solv} + \Delta G^{hydroph}_{solv}$$

- E^{vide}_{MM} : énergie ➔champ de force (**M**écanique **M**oléculaire)
- S^{vide}_{MM} : entropie ➔champ de force **(MM)**
- G^{elec}_{solv} : effet électrostatique de la solvatation

  ➔ électrostatique des milieux continus (**G**eneralized-**B**orn)
- $G^{hydroph}_{solv}$: effet hydrophobe de la solvatation

  ➔ terme surfacique (**S**urface **A**rea)

Structure-based

Scoring Functions

Force field based Functions

• These scoring functions rest on the decomposition of the free energy change of the complex formation into different contributions with a physical meaning.

• There are several ways of decomposing the free energy change. The most frequent terms are:

$$\Delta G_{liaison} = \Delta E^{ligand} + \Delta E^{interm}_{vdW} + \Delta G^{complex}_{np} + \Delta G_{elect}$$

• These terms represent the change in internal energy of the ligand upon complex formation, the intermolecular van der Waals interactions, the non polar contribution of the hydrophobic effect and the change of the electrostatic interactions.

Structure-based

Scoring Functions

Comparison

- **Force field based**

+ physical nature of the interactions
– costly, requires knowledge of ff. parameters

- **Empirical based**

+ rapid, implicit inclusion of several effects
– depend on the ensemble of data to parametrize the scoring equation, impose the different types of interactions

- **Knowledge-based**

+ rapid, implicit inclusion of several effects
– possible bias from the data base, define the different types of interactions