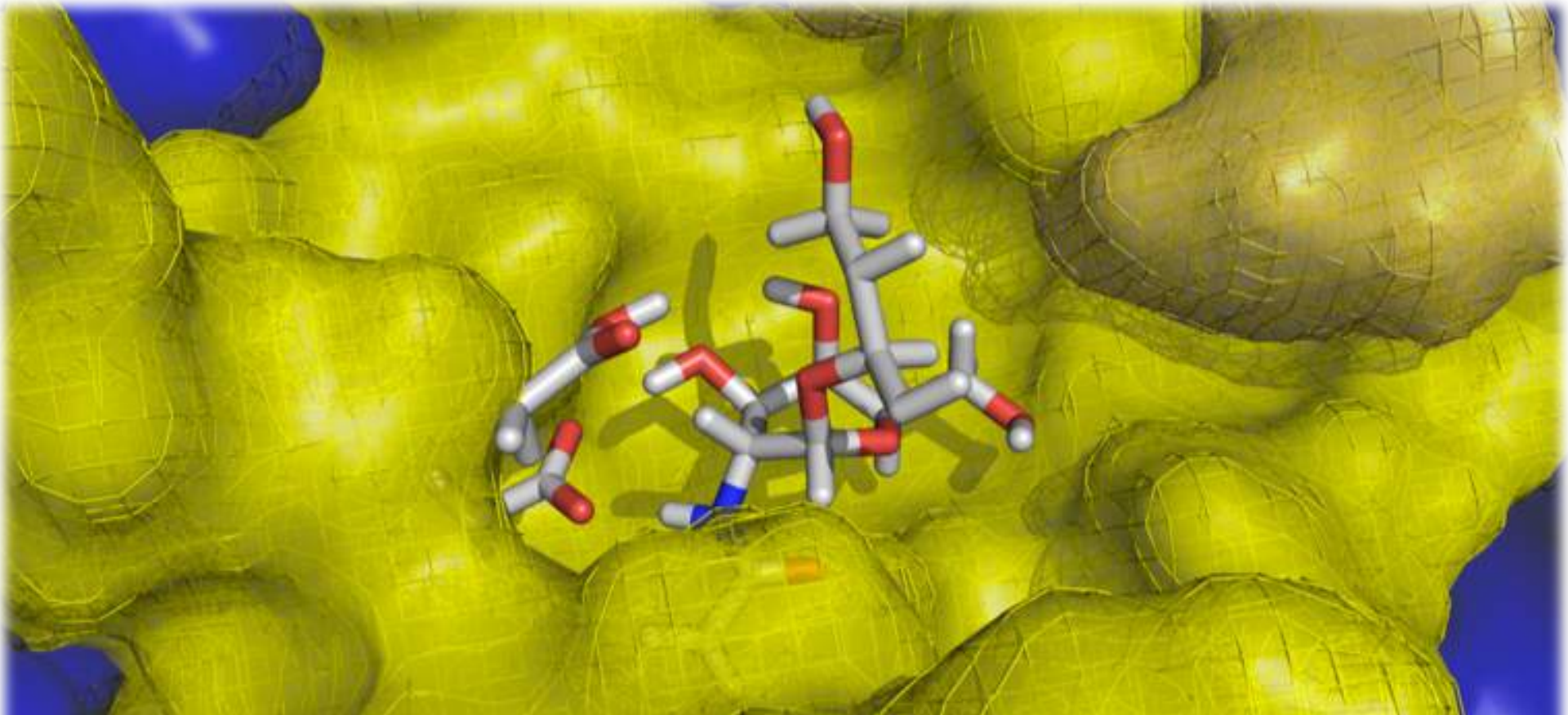
Molecular Docking

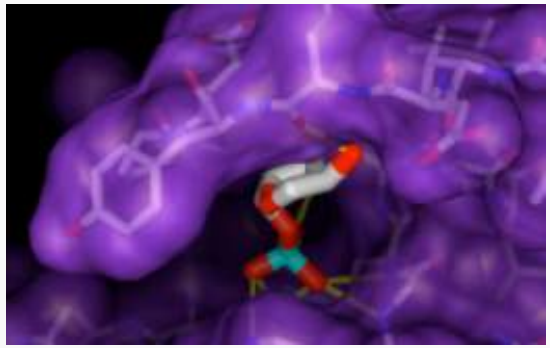
Pedro Alexandrino Fernandes,

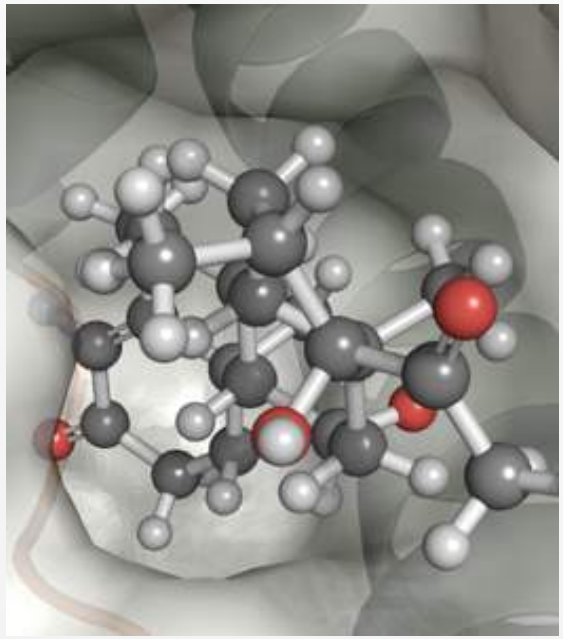
Dep. Chemistry & Biochemistry, University of Porto, Portugal

pedro.fernandes@fc.up.pt

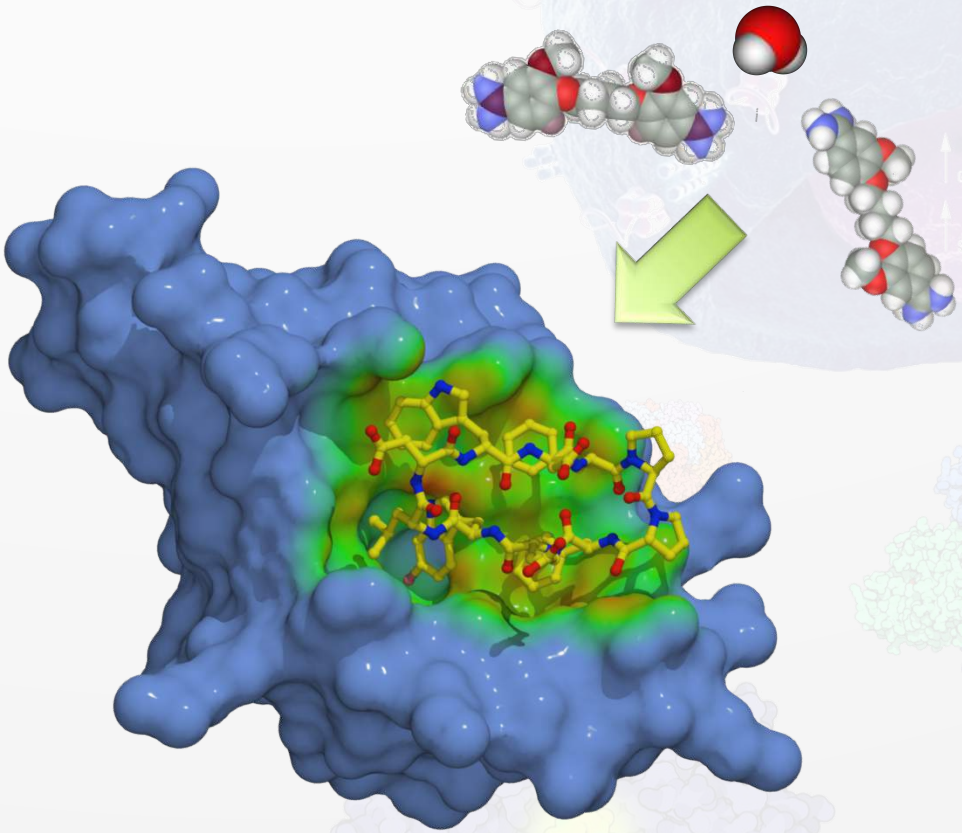


1. Introduction





Intermolecular Associations

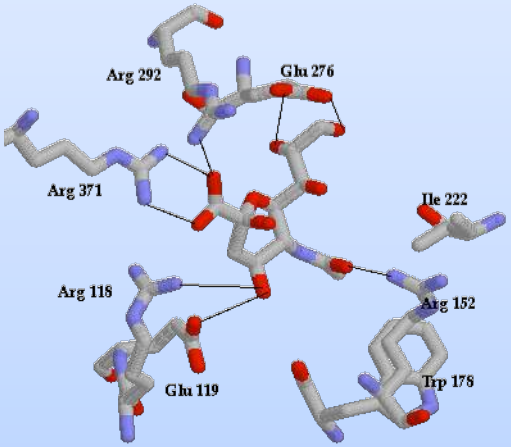


1. Introduction

What type of forces govern these type of interactions ?

Hydrogen-Bonds

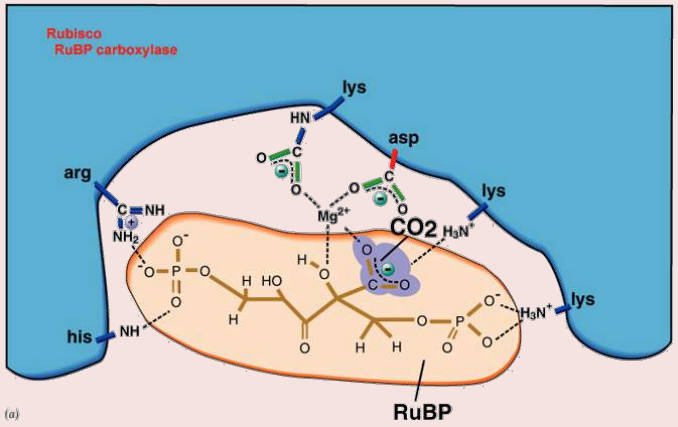
Electrostatic attractions between two dipoles and are often referred to as dipole-–dipole interactions.



Contribution for ΔG_{bind} in solution:
1-2 kcal/mol (≈ 5 kcal/mol in water)

Ionic Interactions

Electrostatic attractions between oppositely charged atoms



Ionic interactions present in RuBisCO – an enzyme involved in the Calvin cycle, that catalyzes the first major step of carbon fixation, a process by which the atoms of atmospheric carbon dioxide are made available to organisms in the form of energy-rich molecules such as glucose

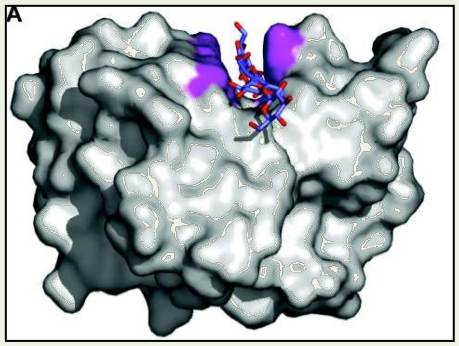
Contribution for ΔG_{bind} in solution:
1-5 kcal/mol

1. Introduction

What type of forces govern these type of interactions ?

Van der Waals

Van der Waals (VDW) or London dispersion forces are weak electrostatic attractive forces between atoms or non-polar molecules caused by the attraction between temporary and induced dipoles produced by the time-dependent fluctuations of the electron density.

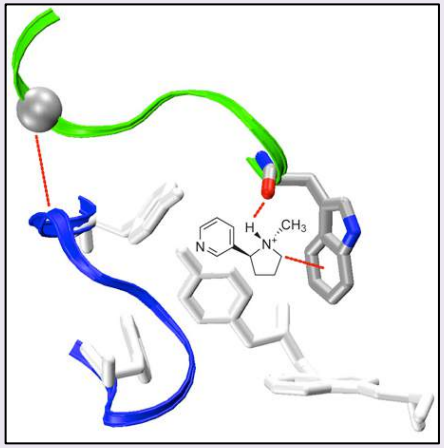


Carbohydrates binding to CBms (carbohydrate binding modules) are largely driven by van der Waals interactions.

Contribution for ΔG_{bind} in solution ≈ 0.5 - 1 kcal/mol per heavy atom pair

Cation-π and Hb-π

Cation-π and Hydrogen bond-π interactions, are noncovalent molecular interactions between the face of an electron-rich π system (e.g. benzene, ethylene) andan adjacent cation (e.g. Li+, Na+) or a hydrogen bond donor (e.g. OH).



Dougherty found that nicotine makes a strong hydrogen bond in the brain's acetylcholine receptors. This same hydrogen bond, in the receptors in muscle cells, is weak. The cause of this difference in binding potency is a single point mutation that occurs in the receptor near the key tryptophan amino acid that makes the cation-π interaction (lysine vs glycine). Nature 458, 534-537.

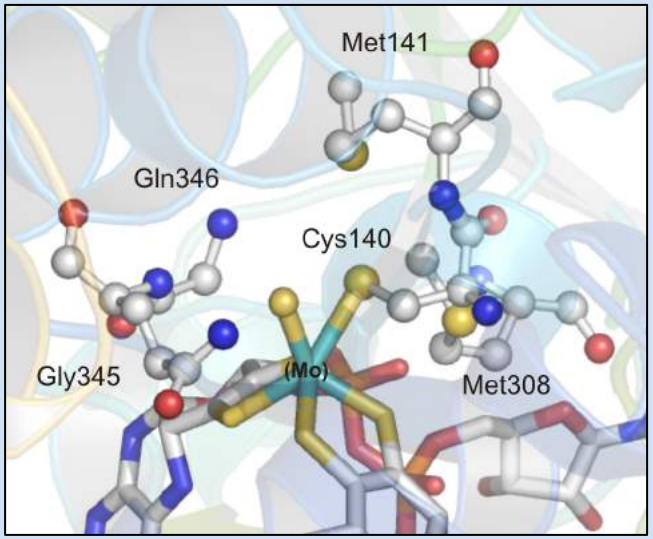
Contribution for ΔG_{bind} in solution ≈ 1-2 kcal/mol

1. Introduction

What type of forces govern these type of interactions ?

Metal chelation

These types of interactions are found when metal ions interact with small molecules or proteins .

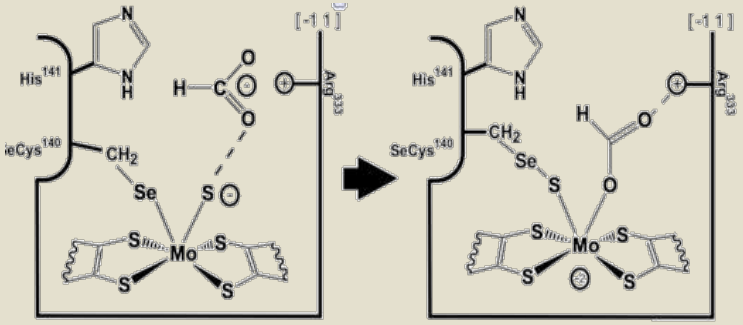


Contribution for ΔG_{bind} in solution ≈ 3-5 kcal/mol

Covalent interactions

Covalent interactions occur when two ligand and receptor react and form chemical bonds.

The sulfur Shift



Contribution for ΔG_{bind} in solution ≈ 30-200 kcal/mol

1. Introduction

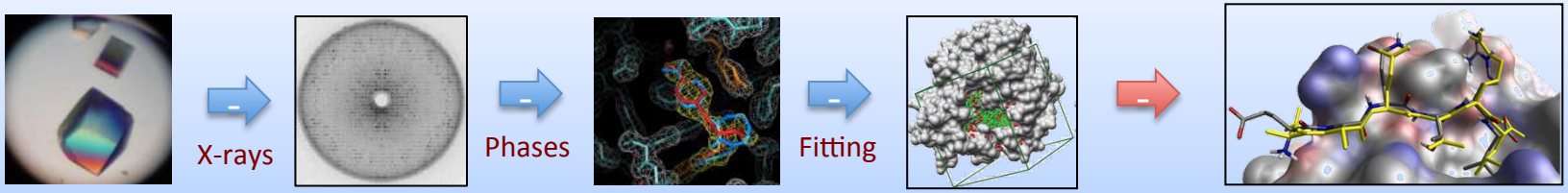
How can we study Protein-Ligand and Protein-Protein Interactions ?

1. Introduction

Available methodologies to study chemical entities interactions

Experimental Techniques

Method	High-Throughput Approach	Living Cell Assay	Type of Interactions	Type of Characterization
Y2H [47,48]	+	In vivo	Physical interactions (binary)	Identification
Affinity purification–MS [61]	+	In vitro	Physical interactions (complex)	Identification
DNA microarrays/Gene coexpression [113]	+	In vitro	Functional association	Identification
Protein microarrays [114–116]	+	In vitro	Physical interaction (complex)	Identification
Synthetic lethality [85,86]	+	In vivo	Functional association	Identification
Phage display [117]	+	In vitro	Physical interaction (complex)	Identification
X-ray crystallography, NMR spectroscopy [84]	–	In vitro	Physical interactions (complex)	Structural and biological characterization
Fluorescence resonance energy transfer [89]	–	In vivo	Physical interaction (binary)	Biological characterization
Surface plasmon resonance [91]	–	In vitro	Physical interaction (complex)	Kinetic, dynamic characterization
Atomic force microscopy [93]	–	In vitro	Physical interaction (binary)	Mechanical, dynamic characterization
Electron microscopy [118]	–	In vitro	Physical interaction (complex)	Structural and biological characterization



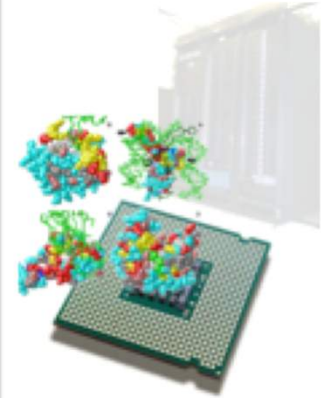
2.1 How to study molecular interactions

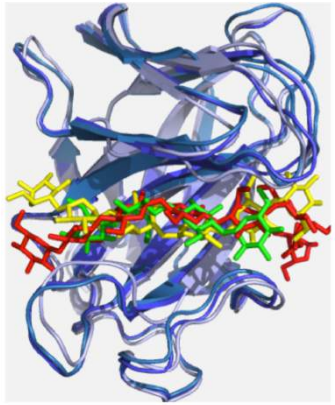
Available methodologies to study chemical entities interactions

X-ray crystallography and NMR spectroscopy retrieve detailed structural information but are time and resource intensive, if not infeasible…

Computational Techniques

To overcome these limitations, several computational methodologies and algorithms have emerged in the last 20 years endeavoring to foresee and improve the understanding of this difficult-to-obtain structural information.





The holy grail of molecular docking is to replace experimental studies of protein-ligand or protein-protein complexes by modeling their structures and binding affinities *in silico*

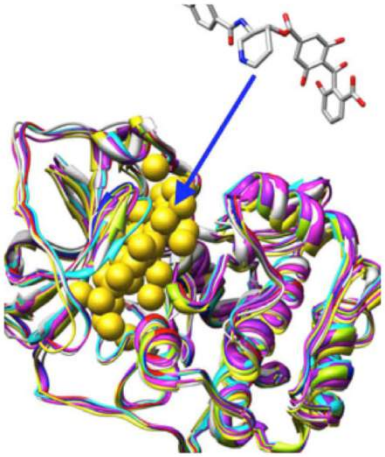
2.2 Molecular Docking

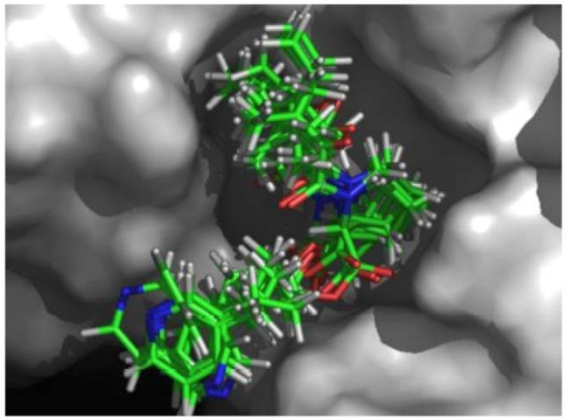
a) Molecular Docking Definition.

Molecular Docking Definition

Molecular docking tries to predict the **structure of the intermolecular complex** formed between two or more constituent molecules, and tries to distinguish, from **the energy point of view**, complexes and/or forms of coordination that two molecules can adopt.

Docking attempts to find the “best” matching between two molecules.





2.2. Molecular Docking

b) Goals and Applications

Goal of Molecular Docking:

Given two biological molecules determine:

- Whether the two molecules "interact"
- If so, what is the orientation that maximizes the interaction free energy (**Binding Pose**)

Applications of Molecular Docking

Virtual screening (hit identification)	***Protein engineering***
Drug Discovery (lead optimisation)	***Enzymatic reactions mechanisms***
Binding-site identification (blind docking)	***Protein – Protein (or Protein – Nucleic Acid) interactions***
De-orphaning of a receptor	***Etc…***

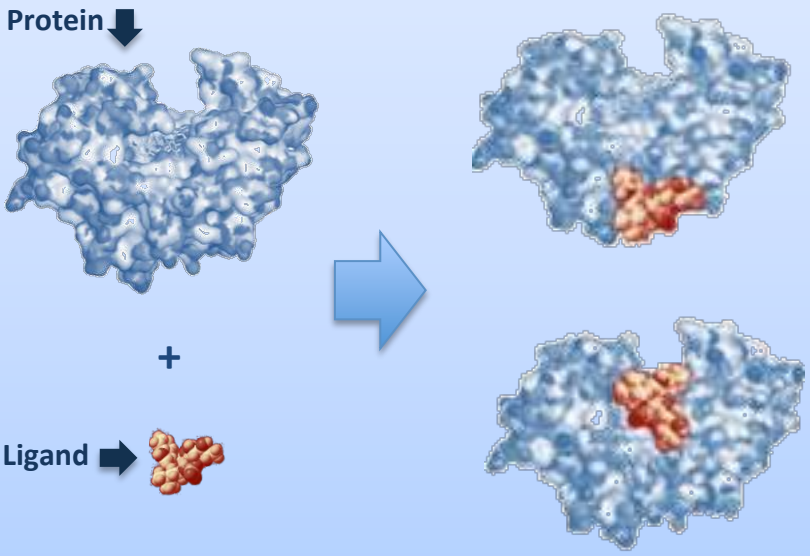
2.2. Molecular Docking

c) General types of Molecular Docking

Protein – Ligand Docking

(Protein-Ligand Docking)

The protein-ligand docking tries to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein receptor or enzyme.



The holy grail of protein-ligand docking is to replace experimental studies of protein-ligand complexes by modeling their structures and binding affinities *in silico*

Pharmaceutical research employs docking techniques for a variety of purposes, most notably in the virtual screening of large databases of available chemicals in order to select likely drug candidates.

Characteristics:

- Generally involves the docking of a small molecule to a macromolecule.
- Typically the ligand is not fixed in shape (as opposed to macromolecular docking)
- The ligand has often 5-12 rotatable bonds

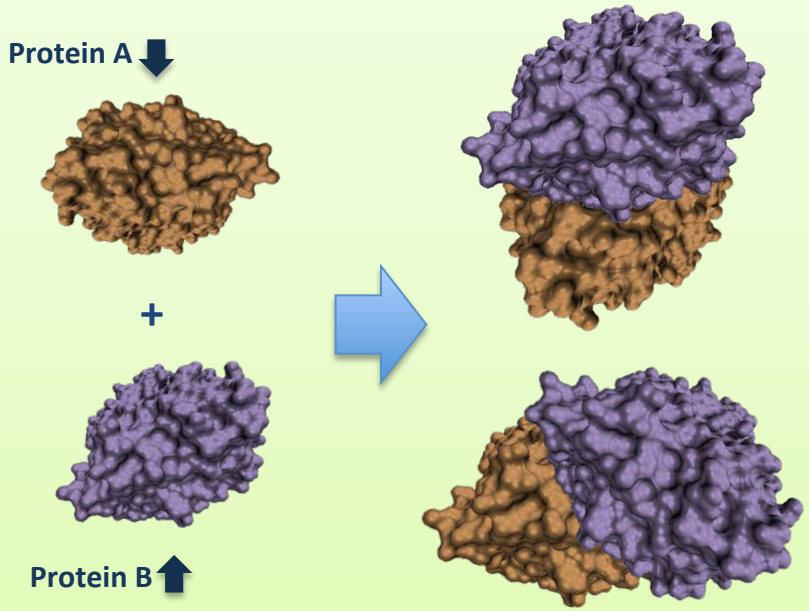
2.2. Molecular Docking

c) General types of Molecular Docking

Protein – Protein Docking

(Protein-Protein Docking)

Protein-protein docking encompasses the determination of the molecular structure of complexes formed by two or more proteins without the need for experimental measurement.



Try to answer the following questions:

- These proteins bind in vivo?
- If they do bind: What is the spatial configuration which they adopt in their bound state?
- How strong or weak is their interaction?

Characteristics:

- Generally, two macromolecules are docked, such as protein-DNA, protein-protein, protein-membrane, etc.
- The docking involves large contact areas.
- In a few cases the shape of the molecules are fixed.
- The energetically favorable complexes are created based on geometric properties like shape complementarities and on the energetics of specific residue interactions.

2.2. Molecular Docking

d) Tasks involved in a Molecular Docking Algorithm

Two main tasks of the Molecular Docking Tools

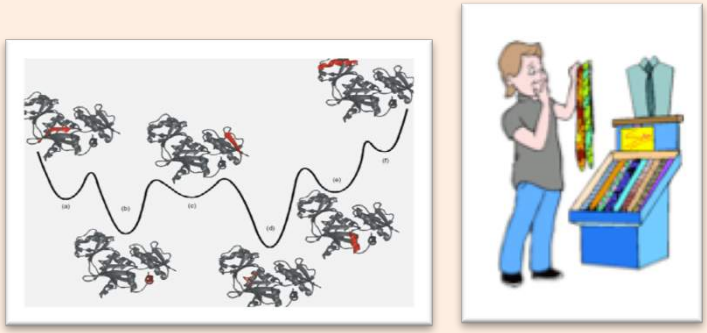
Search Algorithm

The search algorithm is a process where all possible conformations and orientations of the complex (the paired ligand and protein) in a space (the binding site of interest) is being searched.

Sampling of conformational space

Scoring Function

The pose score is a measure of the fit of a ligand into the active site. Scoring during the posing phase usually involves simple energy calculations (electrostatic, van der Waals, ligand strain).

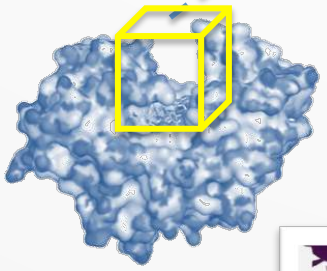


Being capable of generating the right conformation is not enough.

Definition: Molecular docking tries to predict the structure of the intermolecular complex formed between two or more constituent molecules, and to distinguish, from the energy point of view, complexes and/or forms of coordination that two molecules can adopt.

2.2. Molecular Docking

e) General Molecular Docking Algorithm

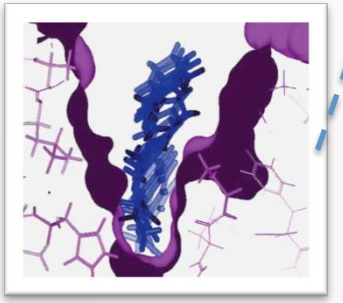


Protein

Definition of the active site

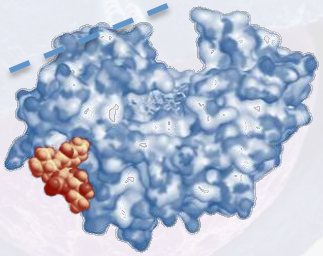
Ligand

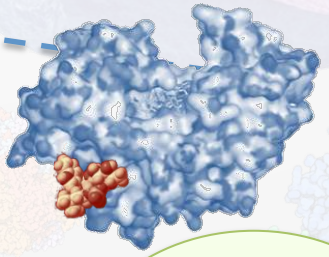
Search Algorithm



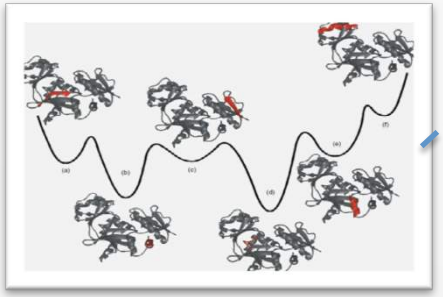
Complex

N cycles

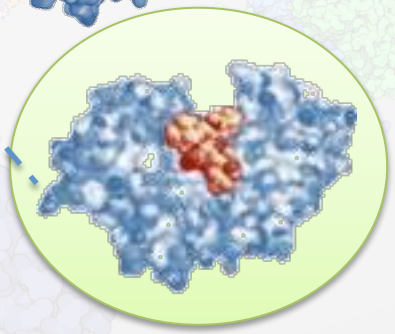




Scoring Function



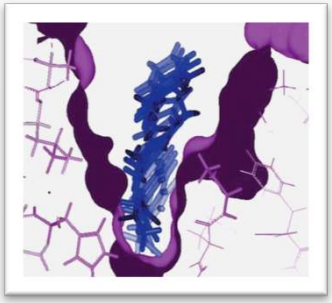
Final Solution



2.2. Molecular Docking

f) Search Algorithms

Search Algorithm



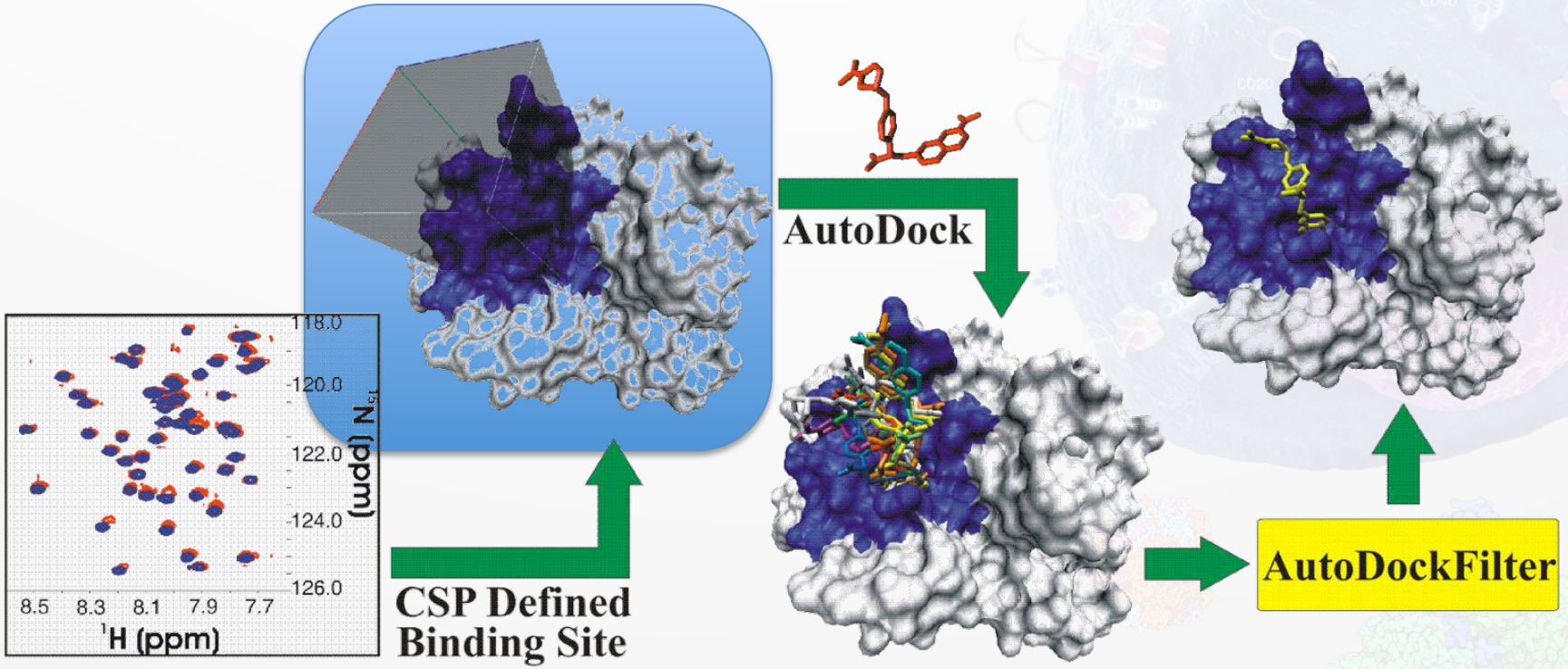
"How to explorer the binding mode…"

"How to explore the conformational space…"

2.2. Molecular Docking

f) Search Algorithms

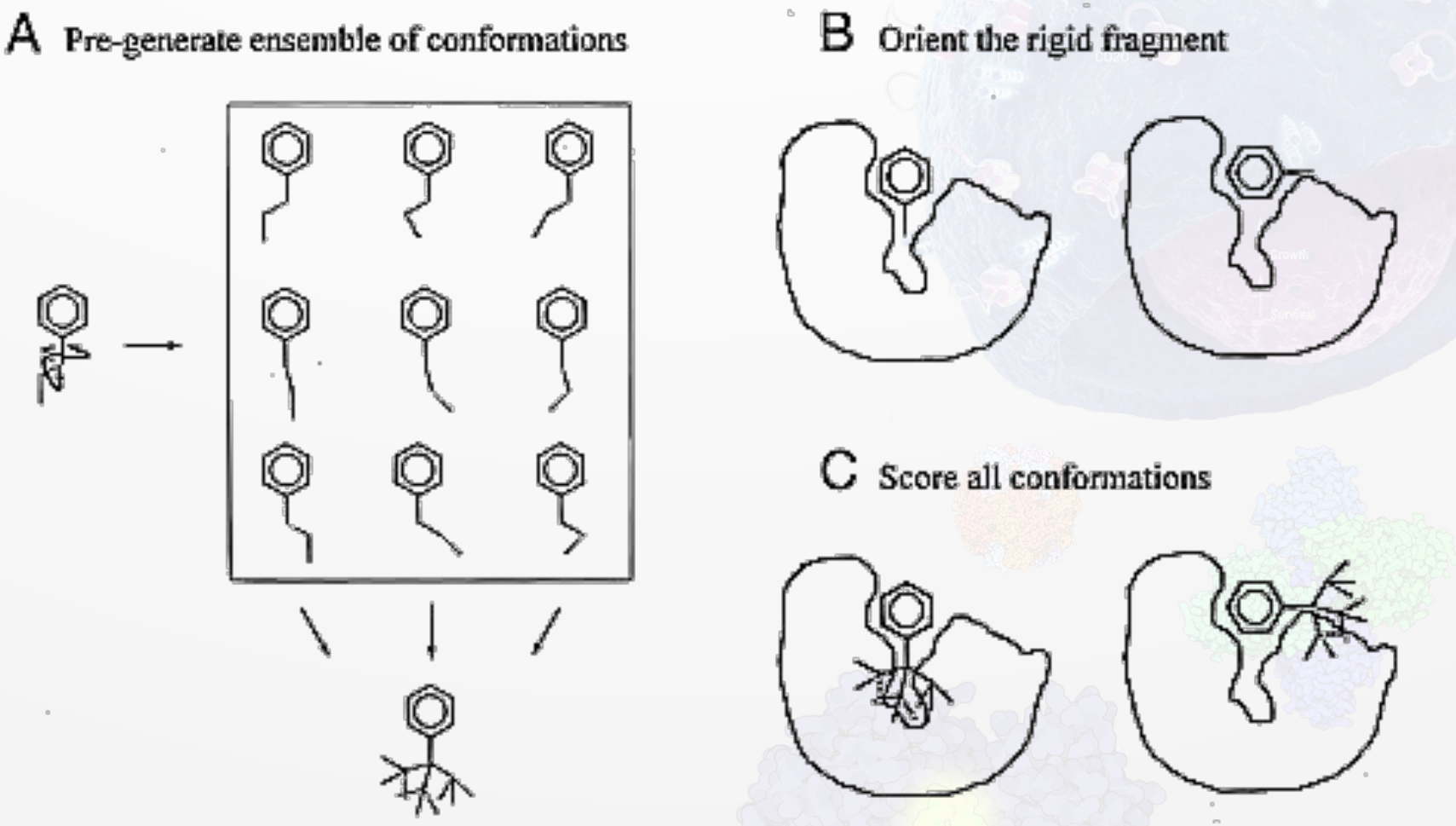
Definition of the active site



2.2. Molecular Docking

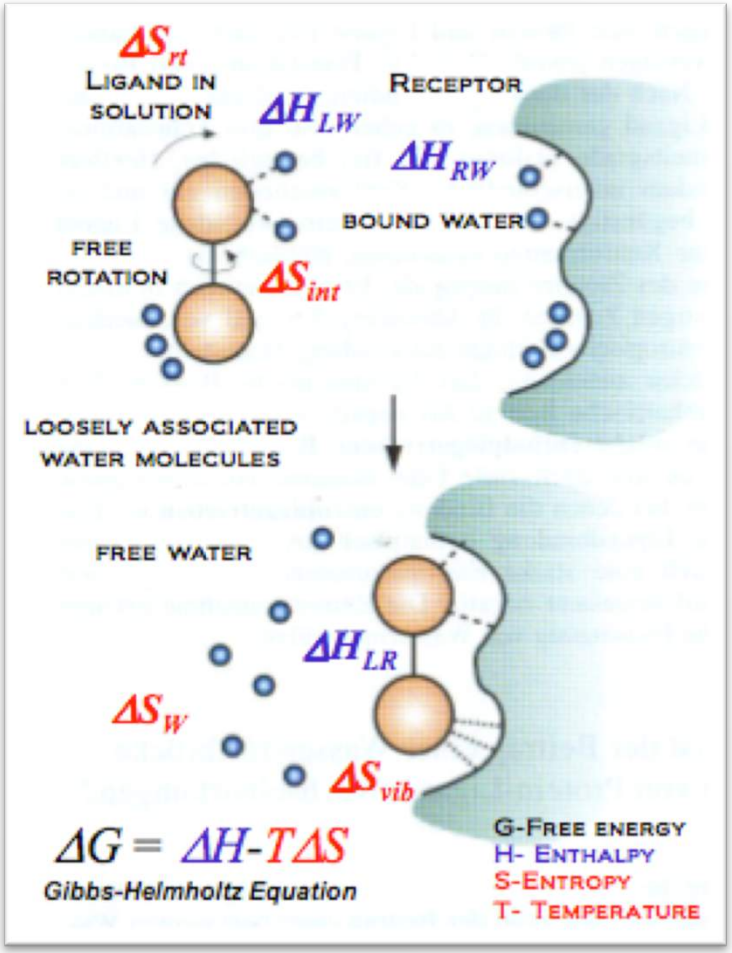
f) Search Algorithms

The search algorithm is a process where all possible conformations and orientations of the complex (the paired ligand and protein) in a space (the binding site of interest) is being searched.



2.2. Molecular Docking

f) Search Algorithms

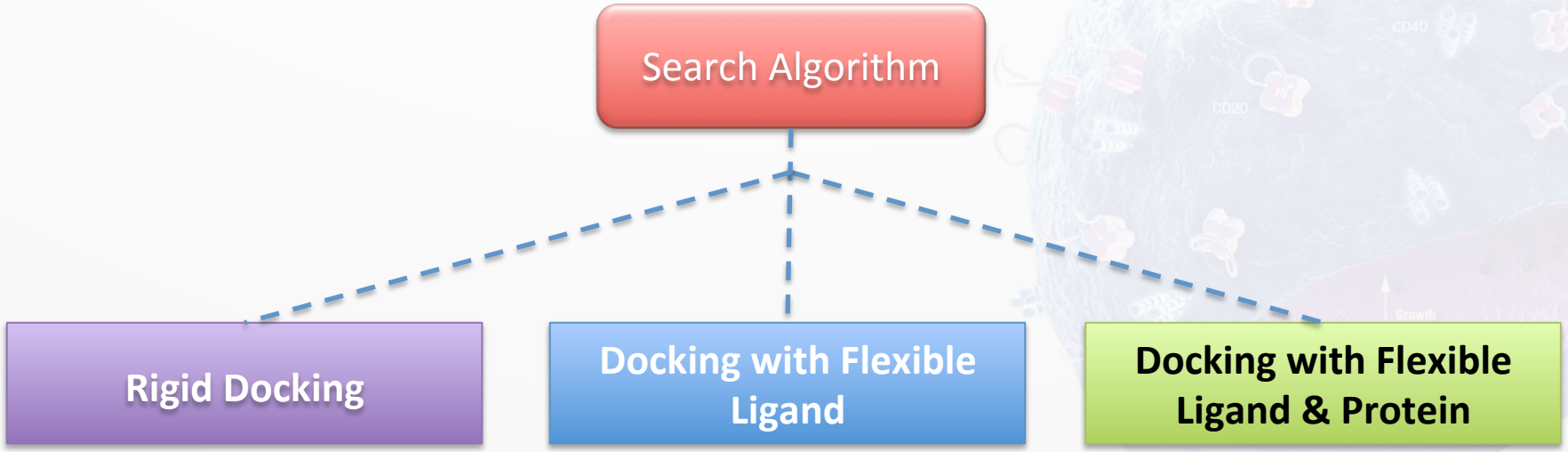


Conceptual steps in receptor-ligand binding:

- Approach
- Desolvation of the ligand and of the binding site of a protein
- Binding of the ligand in the protein cavity
- Change of the ligand orientation
- Adoption and “freezing” of the correct “active” conformation.
- Establishing of new interactions : Coulombic and dispersive.

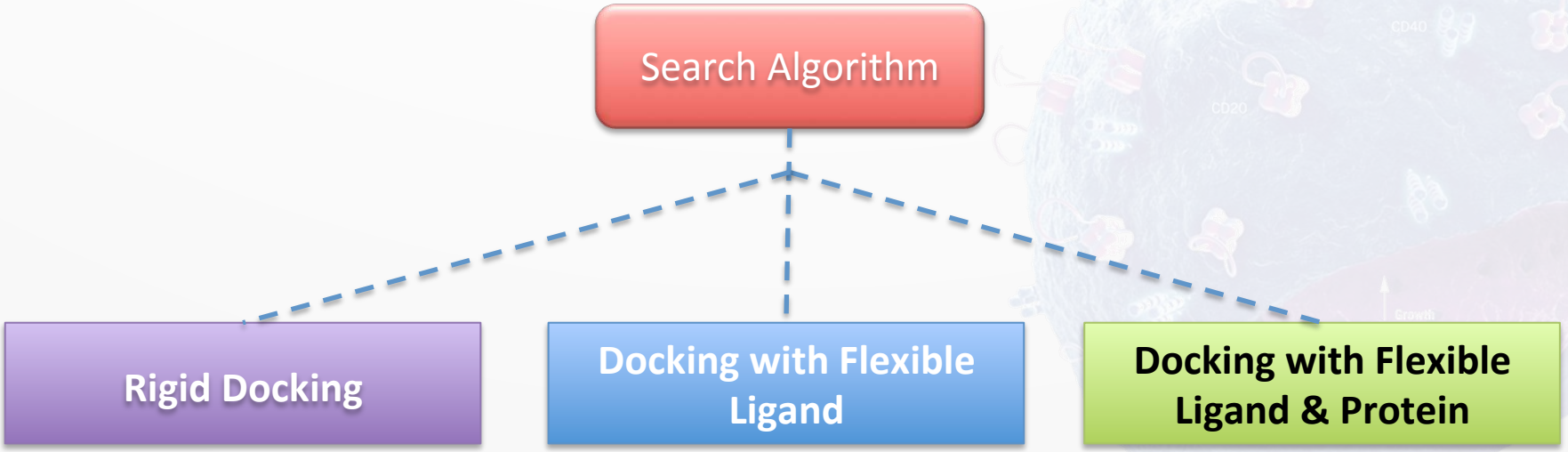
2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

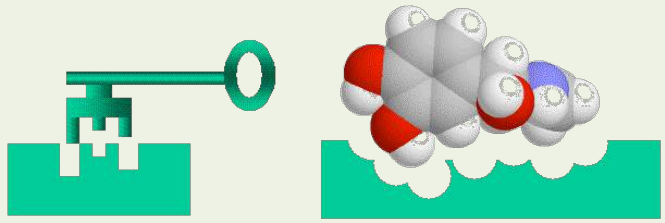
SEARCH ALGORITHM : *Rigid Docking*

- Historically the first approaches to simulate Protein-Ligand Docking.
- **Conformational Space:** Protein and ligand are held fixed in conformational space which reduces the problem to the search for the relative orientation for the two molecules with lowest energy.
- **DOF (Degrees of Freedom)**= 6 (3 for rotation and 3 for translation)
- **Goal:** Search for the relative orientation of the two molecules with lowest energy.
- **Advantages:** Very Fast
- **Disadvantages**: Not very accurate

Based on the Lock and Key model

(Emil Fischer, 1894)

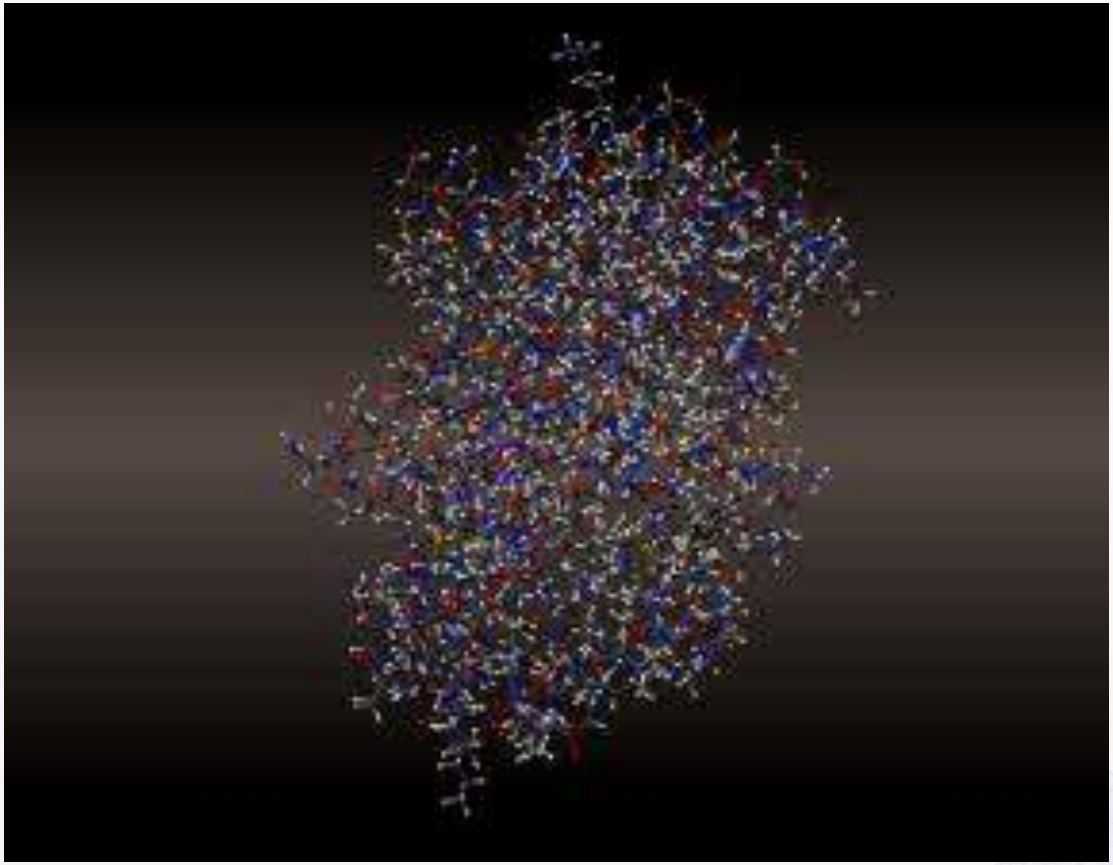
"The ligand binds to the protein only if their shapes are complementary."

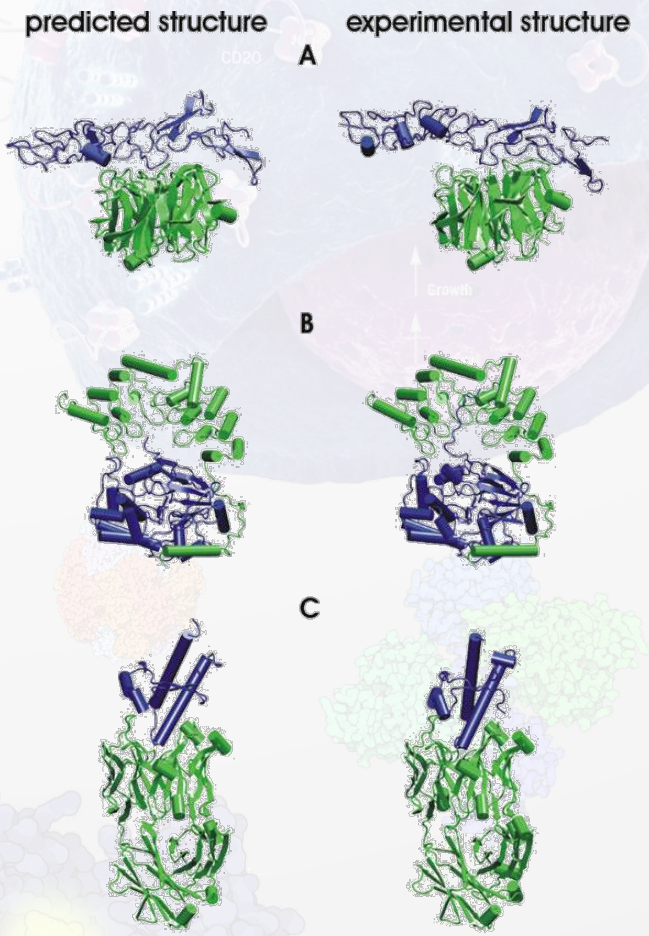


2.2. Molecular Docking

f) Search Algorithms

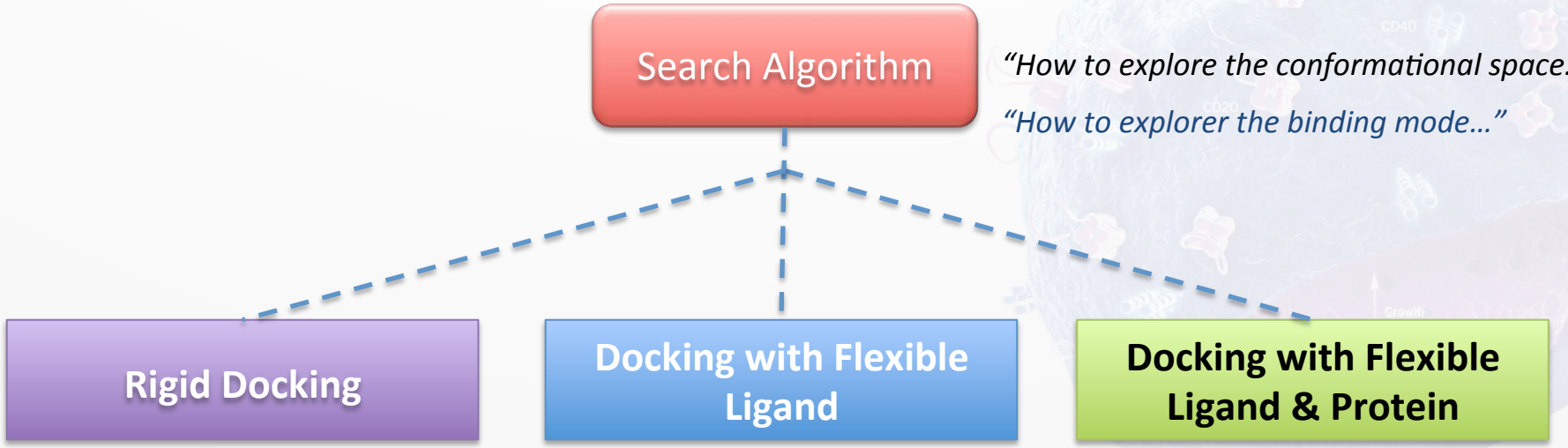
SEARCH ALGORITHM : *Rigid Docking*





2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

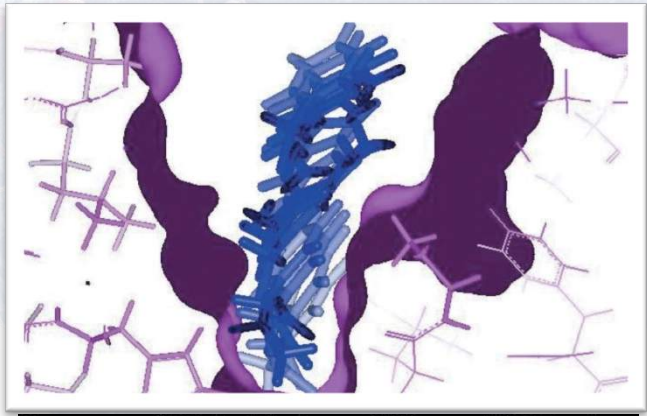
f) Search Algorithms

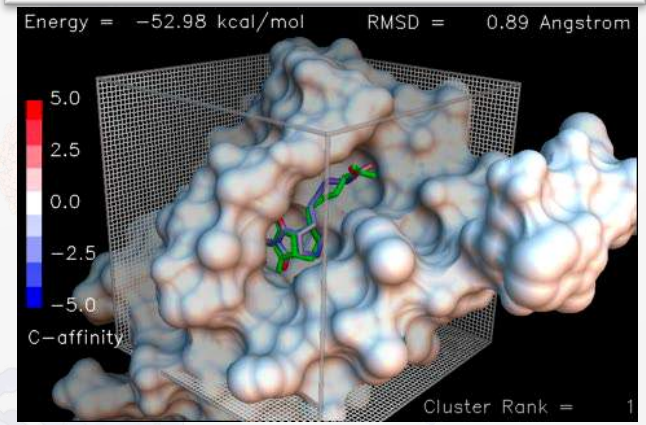
SEARCH ALGORITHM : *Flexible Ligand Docking*

- Presently, is the most popular approximation that is used.
- **Conformational Space:** Protein is held fixed but the ligand is flexibilized.
- **DOF (Degrees of Freedom)**= *= 3 + 3 + n*

  - Position or Translation – *(x,y,z) = 3*
  - *Orientation (Euler angles) – (Φ,Θ,Ψ) = 3*
  - *Rotatable Bonds or Torsions (Flexible Ligands) –(tor_1, tor_2, … tor_n) = n*

- **Advantages:** Moderately Fast and fairly good results
- **Disadvantages**: Protein is fixed. Inappropriate when large conformation rearrangements are involved in the docking process

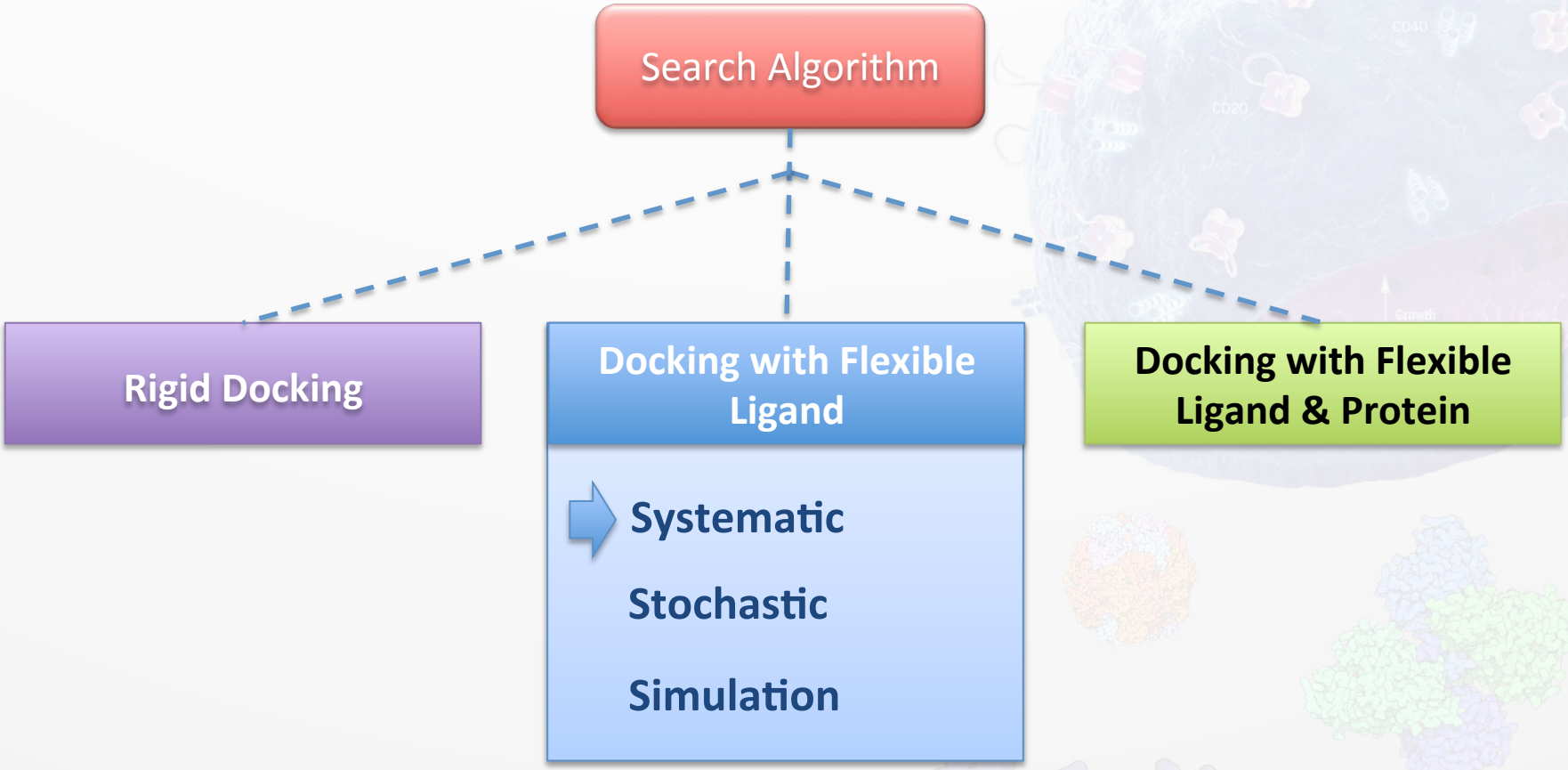




By rough count, about 30 docking programs have appeared in a public venue. The best of these predict the experimental pose about 70% of the time, although selecting the program that will give the best result for any given target is not straightforward.

2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

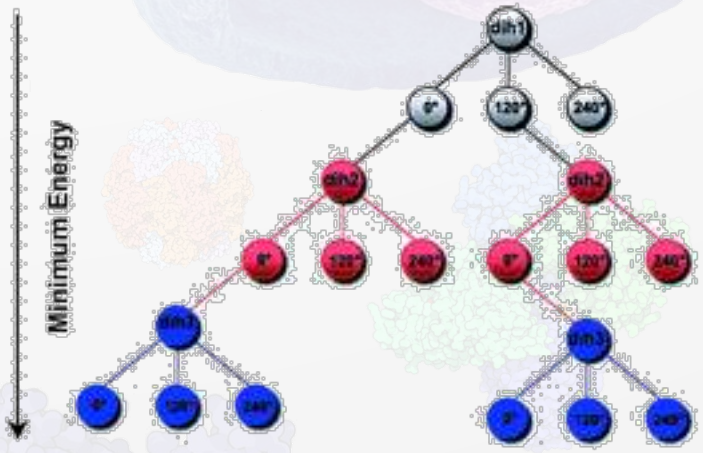
SEARCH ALGORITHM : *Ligand Flexible Docking*

Systematic

search Algorithms

- **Conformational Search**
- **Fragment Based**
- **Databases**

"These algorithms try to explore all the degrees of freedom in a molecule. "



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Ligand Flexible Docking*

Systematic

search Algorithms

Conformational Search

Description :

- "Brute Force" solution to explore the flexibility of the ligand.
- All rotatable bonds in the ligand are systematically rotated through 360° (or the ones that were chosen) using a fixed increment, until all possible combinations have been generated and evaluated
- The number of structures generated increases immensely with the number of rotatable bonds, a phenomenon known as the combinatorial explosion
- Application is very limited

2.2. Molecular Docking

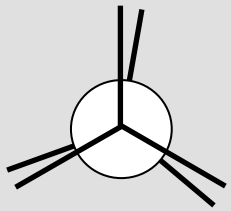
f) Search Algorithms

SEARCH ALGORITHM : *Ligand Flexible Docking*

Systematic

search Algorithms

Conformational Search



$$N_{Total}(Confórmeros) = \prod_{i}^{N} \frac{360}{\theta_i}$$

3 Dihedrals | 360º increments of 60º = **125 Conformations**

3 Dihedrals | 360º increments of 30º = **1728 Conformations**

6 Dihedrals | 360º increments of 30º = **248832 Conformations**

2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM : *Ligand Flexible Docking*

Systematic

search Algorithms

Conformational Search

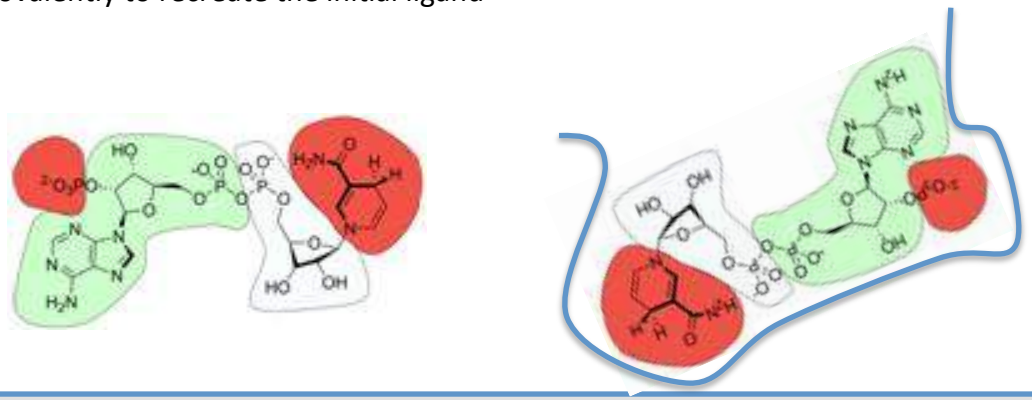
Fragment Based

Description :

- Incrementally grow the ligands into the active site, either by docking the several fragments into the active-site (LUDI, FlexX, DOCK, ADAM).

"The place-and-join approach"

docks several fragments of the ligand into the active-site and linking them covalently to recreate the initial ligand



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM : *Ligand Flexible Docking*

Systematic

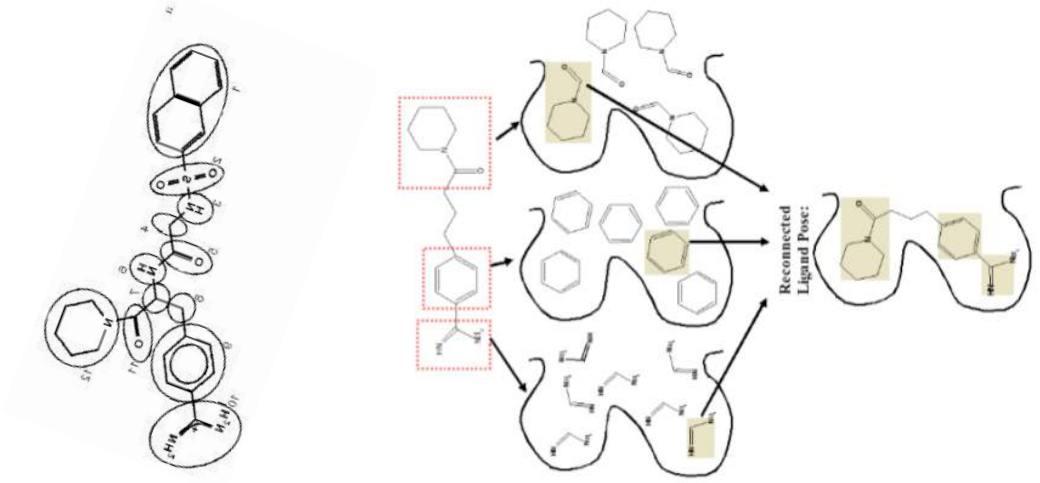
search Algorithms

Conformational Search

Fragment Based

"The incremental approach"

The ligand is divided until a rigid core-fragments is obtained. Afterwards, the flexible fragments are subsequently and successively added to recreate the initial ligand.



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Ligand Flexible Docking*

Systematic

search Algorithms

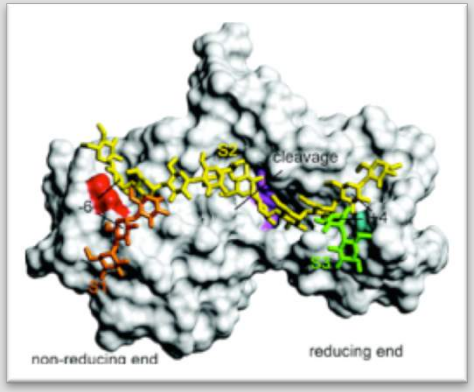
Conformational Search

Fragment Based

Databases

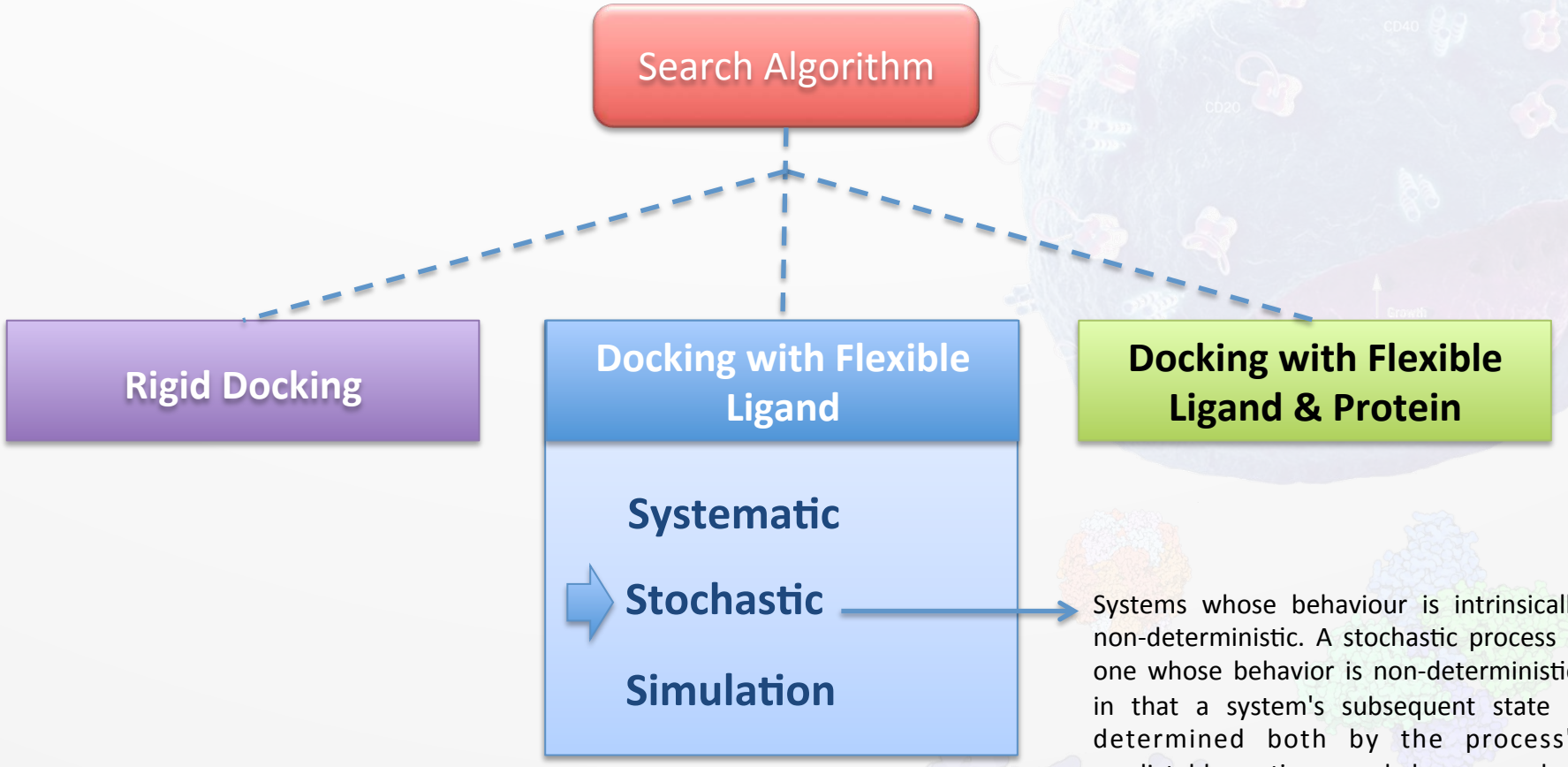
Description :

- These methods overcome the combinatorial explosion problem of DOF (degrees of freedom) by using libraries of pre-generated conformations (conformational ensembles) to deal with the ligand flexibilty issue.
- From this database, it is create a library of the most relevant and possible conformations of the ligand.
- Each of conformation of the ligand is then docked to receptor (rigid docking)



2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

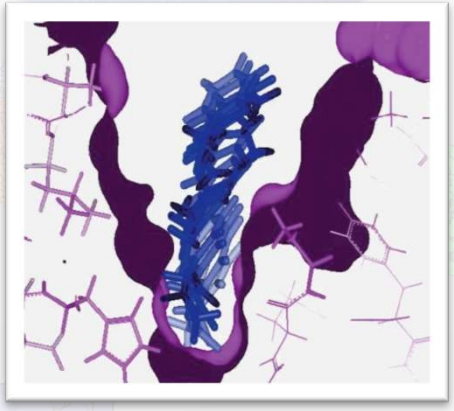
SEARCH ALGORITHM : *Flexible Ligand Docking*

Random or stochastic algorithms

search Algorithms

- Monte Carlo
- Genetic Algorithms
- TABU Search

Random search algorithms sample the conformational space by performing random changes to a single ligand or a population of ligands. The alteration performed is at each step accepted or rejected based on a predefined probability function.



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Flexible Ligand Docking*

Stochastic

search Algorithms

Monte Carlo

Description :

- In this approach an initial configuration is refined by taking random steps which are accepted or rejected based on their induced improvement in score, until a certain number of steps have been tried.
- In this approach the acceptance criteria for a newly obtained pose is based on a Boltzmann probability function.
- does not require any sort of derivative information.
- Very efficient method to step energy barriers, allowing to perform a complete searches of the conformation space of the ligand.
- Programs that use this method: Prodock, ICM, MCDOCK, DockVision, and QXP.

2.2. Molecular Docking

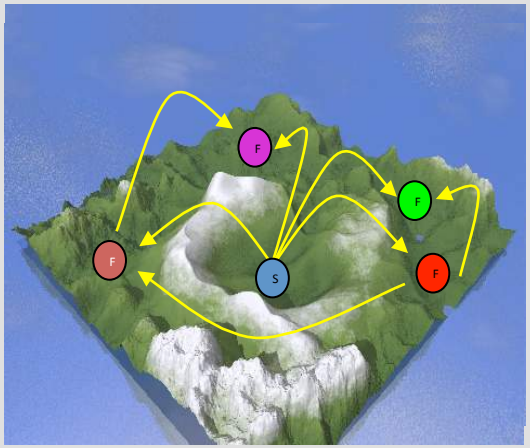
f) Search Algorithms

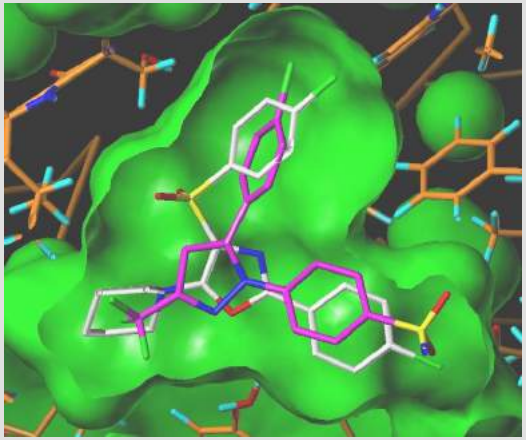
SEARCH ALGORITHM : *Flexible Ligand Docking*

Stochastic

search Algorithms

Monte Carlo





Monte Carlo Process:

- Start with configuration A (energy E_A)
- Make random move to configuration B (energy E_B)
- Accept move when:

  $E_B < E_A$ or if

  $E_B > E_A$ but just with probability P:

2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM : *Flexible Ligand Docking*

Stochastic

search Algorithms

Monte Carlo

Genetic Algorithm

Description :

- Method derived from genetics and the theory of biological evolution.
- Start from an initial population of different conformations of the ligand with respect to the protein (different from the previous methods)
- Each conformation is defined by a set of state variables (defined as genes) that describe aspects like the translation, orientation, and conformation of the ligand in relation to the protein receptor.
- The full set of the ligands state variables is defined as the genotype (Gtype), whereas the atomic coordinates refer to the phenotype (Ptype).
- Genetic operators (mutations, crossovers, and migrations) are applied to the population to sample the conformational space, until a final population that optimizes a predefined fitness function is reached.

2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM : *Flexible Ligand Docking*

Stochastic

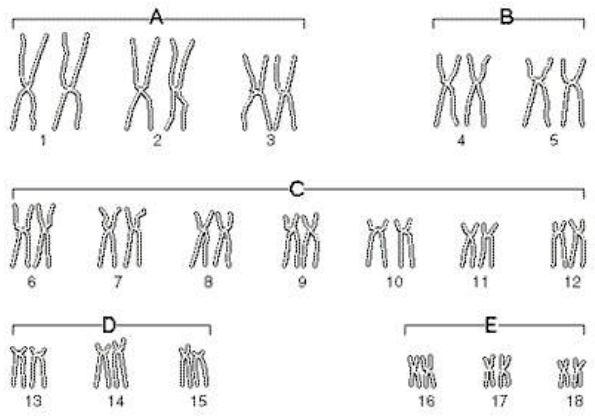
search Algorithms

Monte Carlo

Genetic Algorithm

Description :

- The conformations that do not meet a positive development are excluded and the process repeated.
- The final population contains the best conformations of the ligand.
- Programs: GOLD, AutoDock, DIVALI and DARWIN.



1. Select pairs of GTYPE according to their PTYPE fitness.

2. Apply the genetic operators (crossover, mutation...) to create new GTYPE.

3. Develop GTYPE to get the PTYPE of a new generation and start again from 1.

2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Flexible Ligand Docking*

Stochastic

search Algorithms

Monte Carlo

Genetic Algorithm

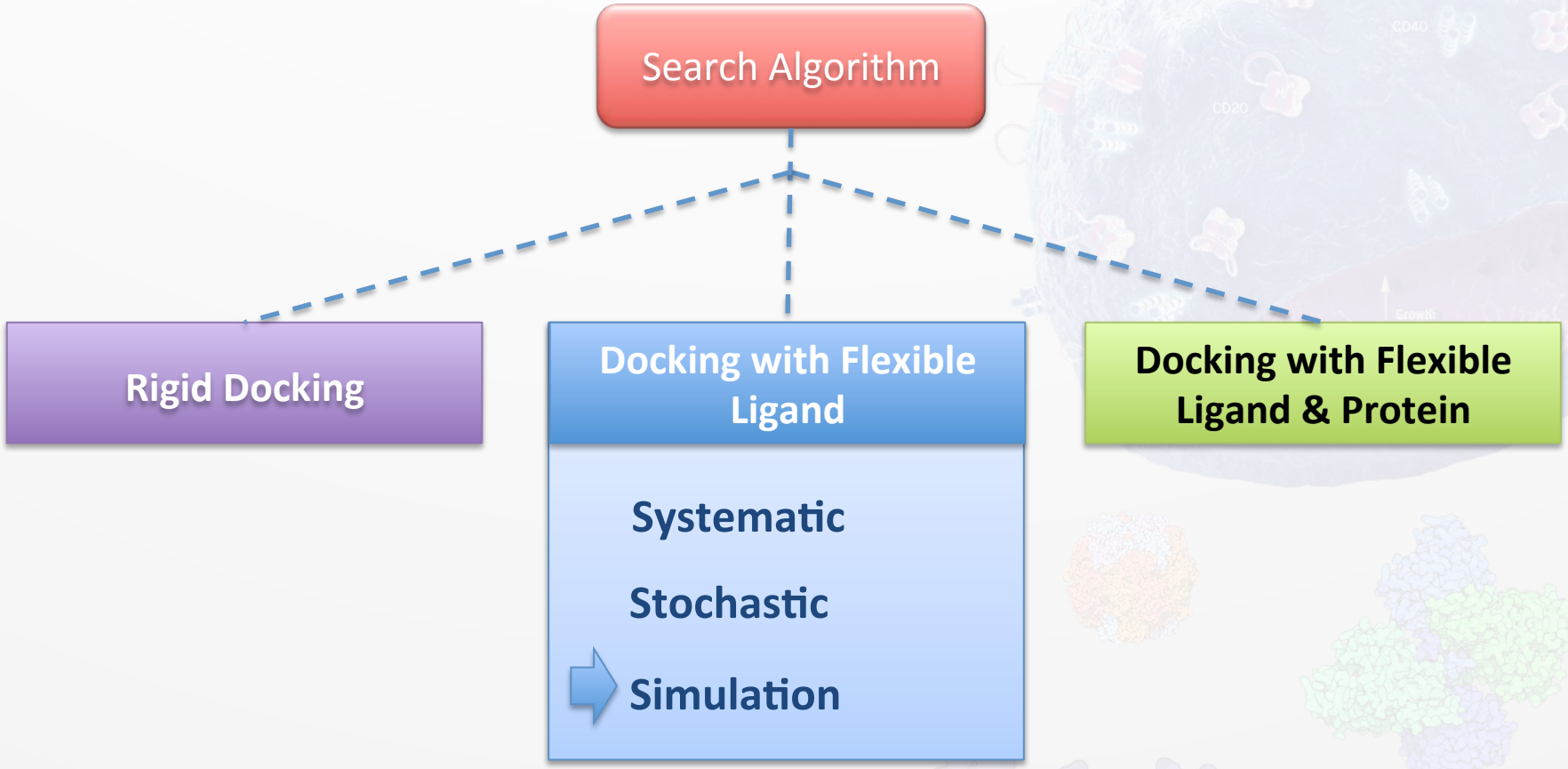
TABU Search

Description :

- Impose restrictions that prevent the search from revisiting already explored areas of the conformational space, promoting the analysis of new regions.
- This is accomplished through a list that stores previously visited solutions.

2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Ligand Flexible Docking*

Simulation

search Algorithms

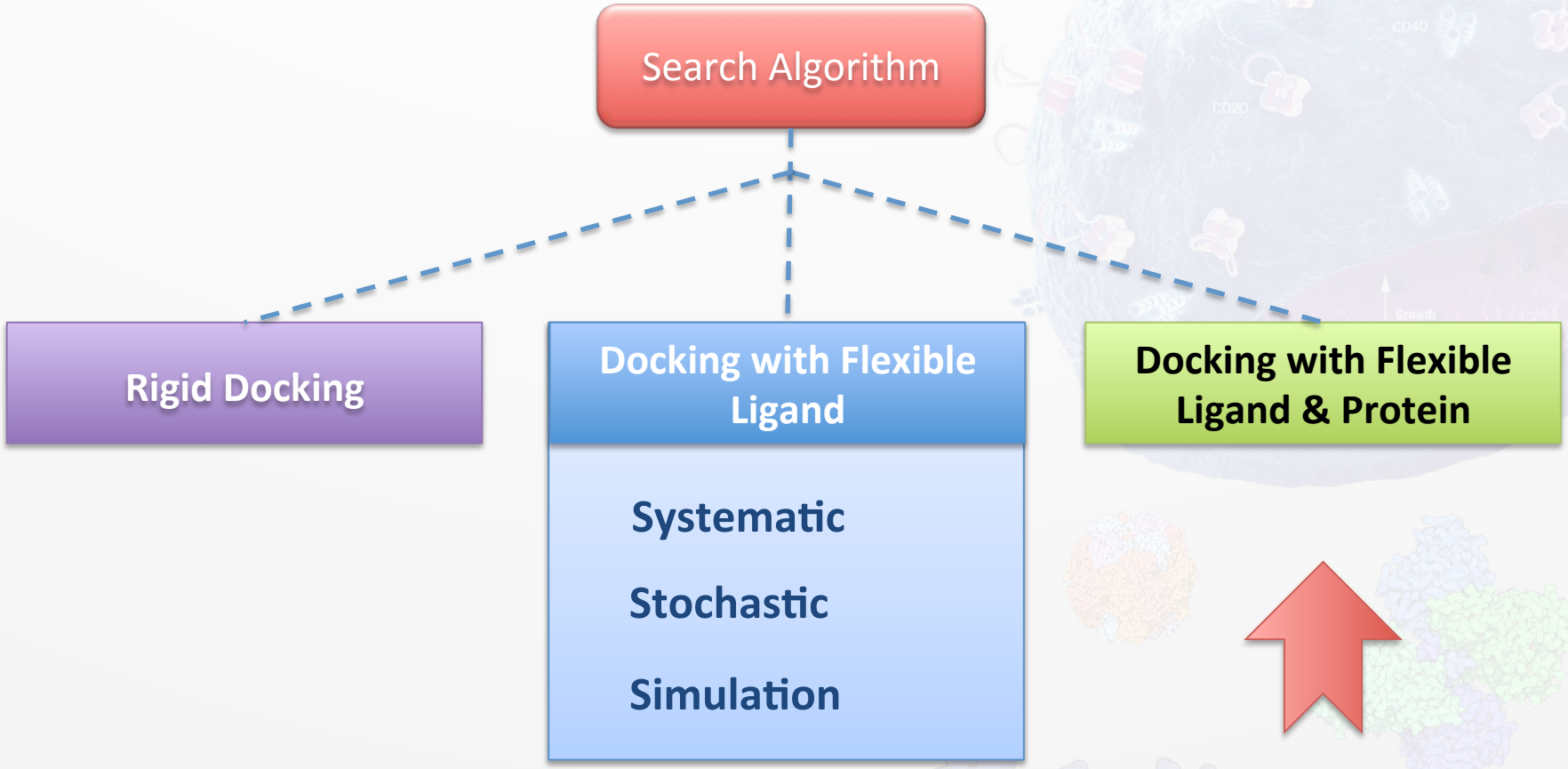
Description :

- Based on the calculation of the solutions to Newton's equations of motion.
- Generally they are not used as a standalone approximation.

Energy Minimizations	Molecular Dynamics
• Direct searches (e.g., simplex) • Gradient methods (e.g., steepest descend) •Conjugate-gradient methods (e.g., Fletcher-Reeves) • Second- derivative methods (e.g., Newton-Raphson) • Least- squares methods (e.g., Marquardt)	▪ Explores most favourable conformations. ▪ Difficult to navigate a rugged hypersurface of a biological system and crossing high-energy barriers ▪ Difficult to sample the conformational space within a feasible simulation period.

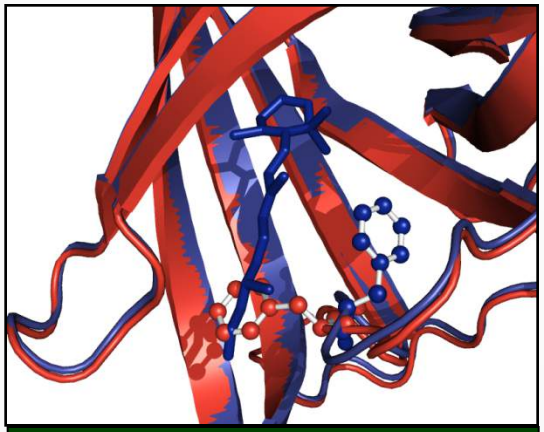
2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

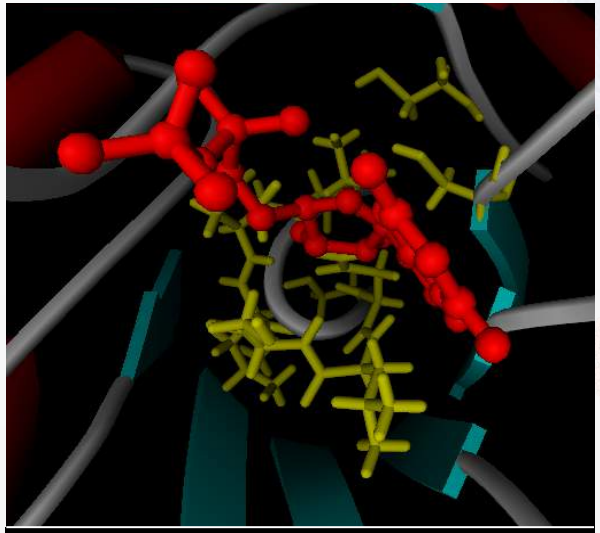


In several enzymes the same active site binds several and different ligands and substrates sequentially.

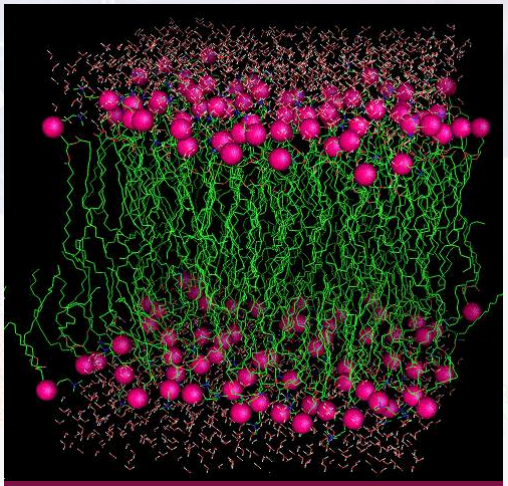
NMR studies reveals different conformations of the same enzyme:ligand complex.

Despite its importance, the intrinsic mobility of the proteins has often been ignored in drug design, (...).

S. Teague, *Nature Reviews*, 2, 2003.



In **enzymes** the substrates must be absorbed and the transition state orientations enforced and products desorbed.



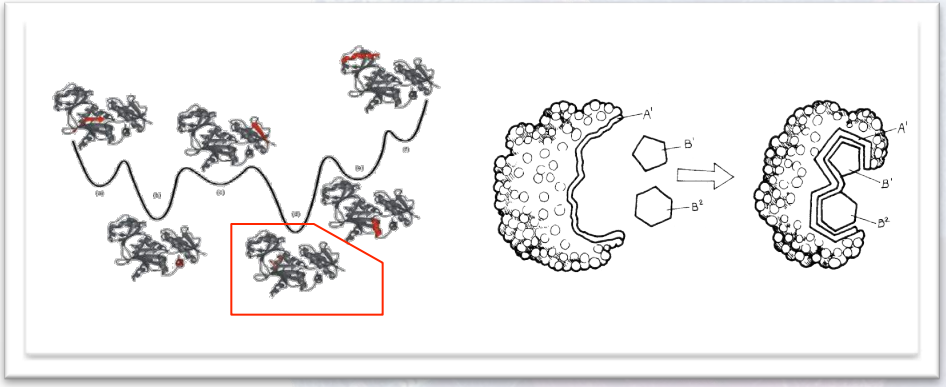
In, **membrane**-bound receptors , the message of the ligand must be transmitted from outside of the cell to the inside.

2.2. Molecular Docking

f) Search Algorithms

Induced fit Model
(Koshland , 1956)

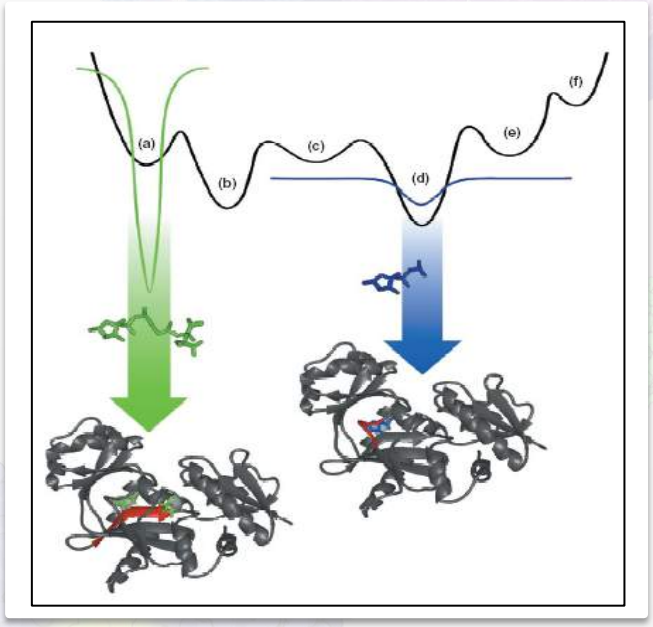
A model for enzyme-substrate interaction to describe that only the proper substrate is capable of inducing the proper alignment of the active site that will enable the enzyme to perform its catalytic function. It suggests that the active site continues to change until the substrate is completely bound to it, at which point the final shape and charge is determined.



Ensemble of conformational states Model
(Carlson, 2000)

It is assumed that the protein coexist in solution within different conformations of lower energy. As the ligand approaches the protein that interacts selectively with a particular conformation, increasing its proportion in relation to others, favoring by this way the process of docking.

In the end the ligand induces a conformational rearrangement in the protein that would not be adopted in the unbounded state.

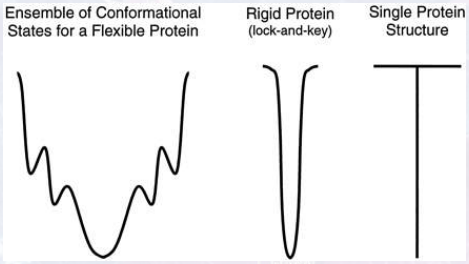


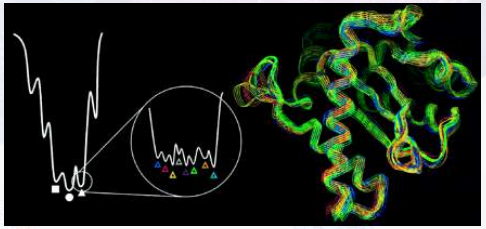
2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Docking with Flexible Protein and Ligand*

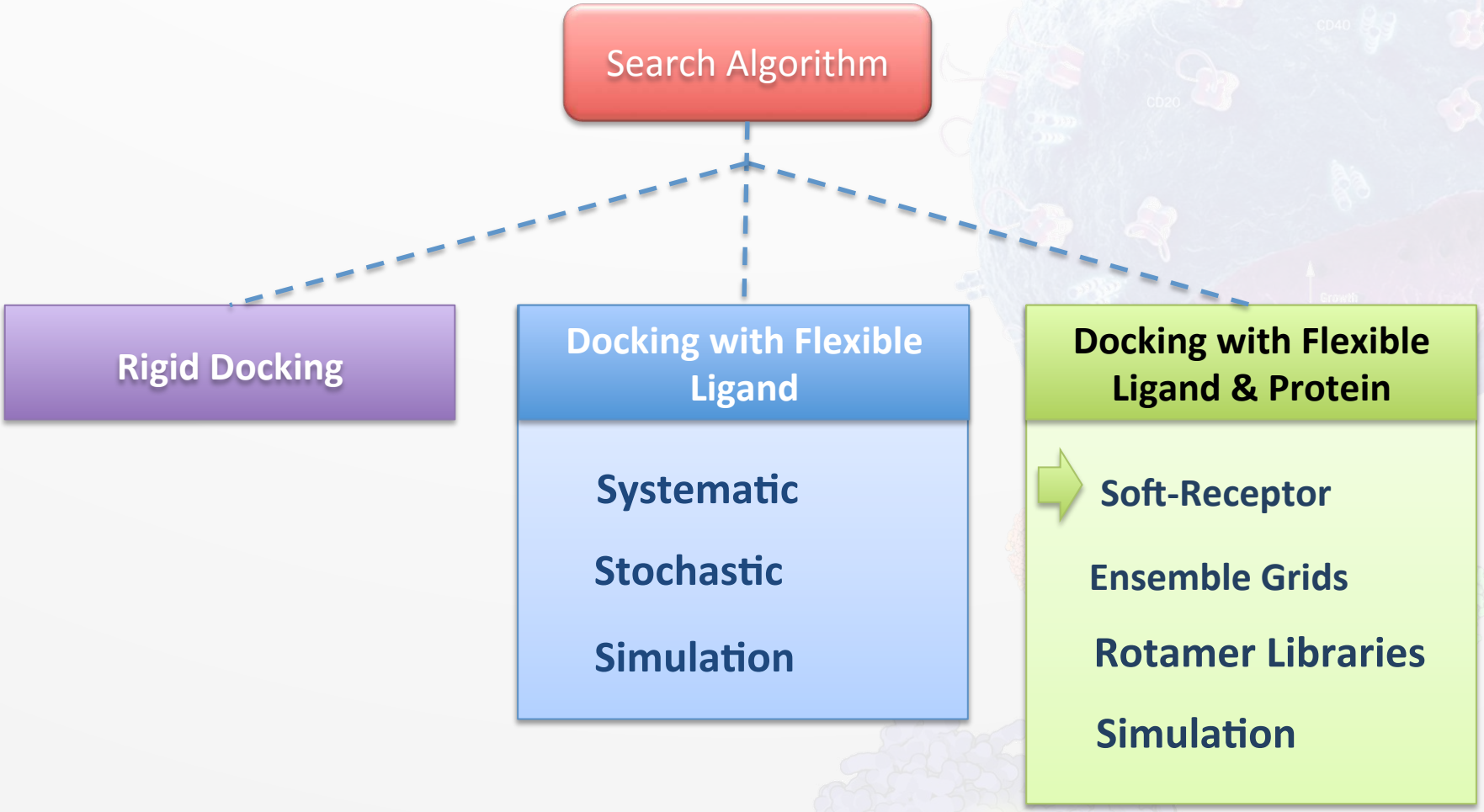
- It is the most accurate approximation and is now becoming used.
- **Conformational Space:** Protein and the Ligand are flexibilized.
- **DOF (Degrees of Freedom)**= *= 3 + 3 + n (n = DOF of protein and Ligand)*
- Many biological systems experiencing significant structural changes and are important to the coordination of the ligands.
- **Advantages:** The most accurate approximation
- **Disadvantages**: Time demanding





2.2. Molecular Docking

f) Search Algorithms



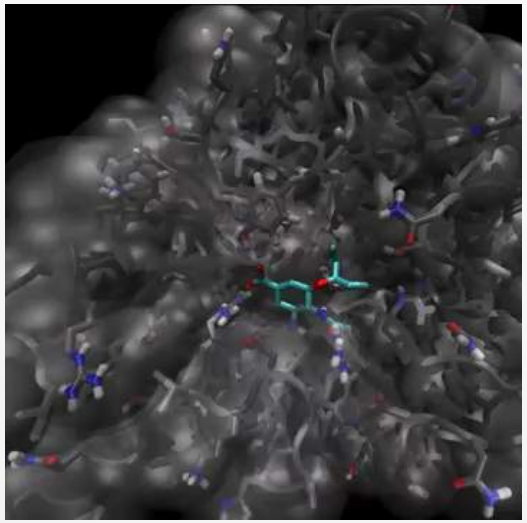
2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Docking with Flexible Protein and Ligand*

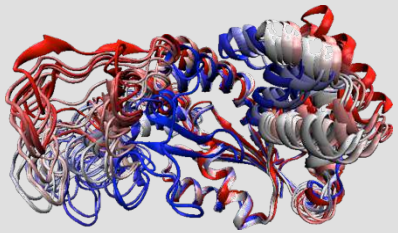
Soft-Receptor

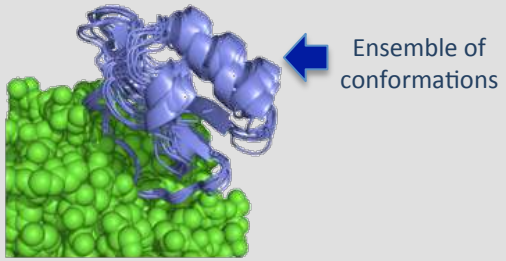
search Algorithms



Description :

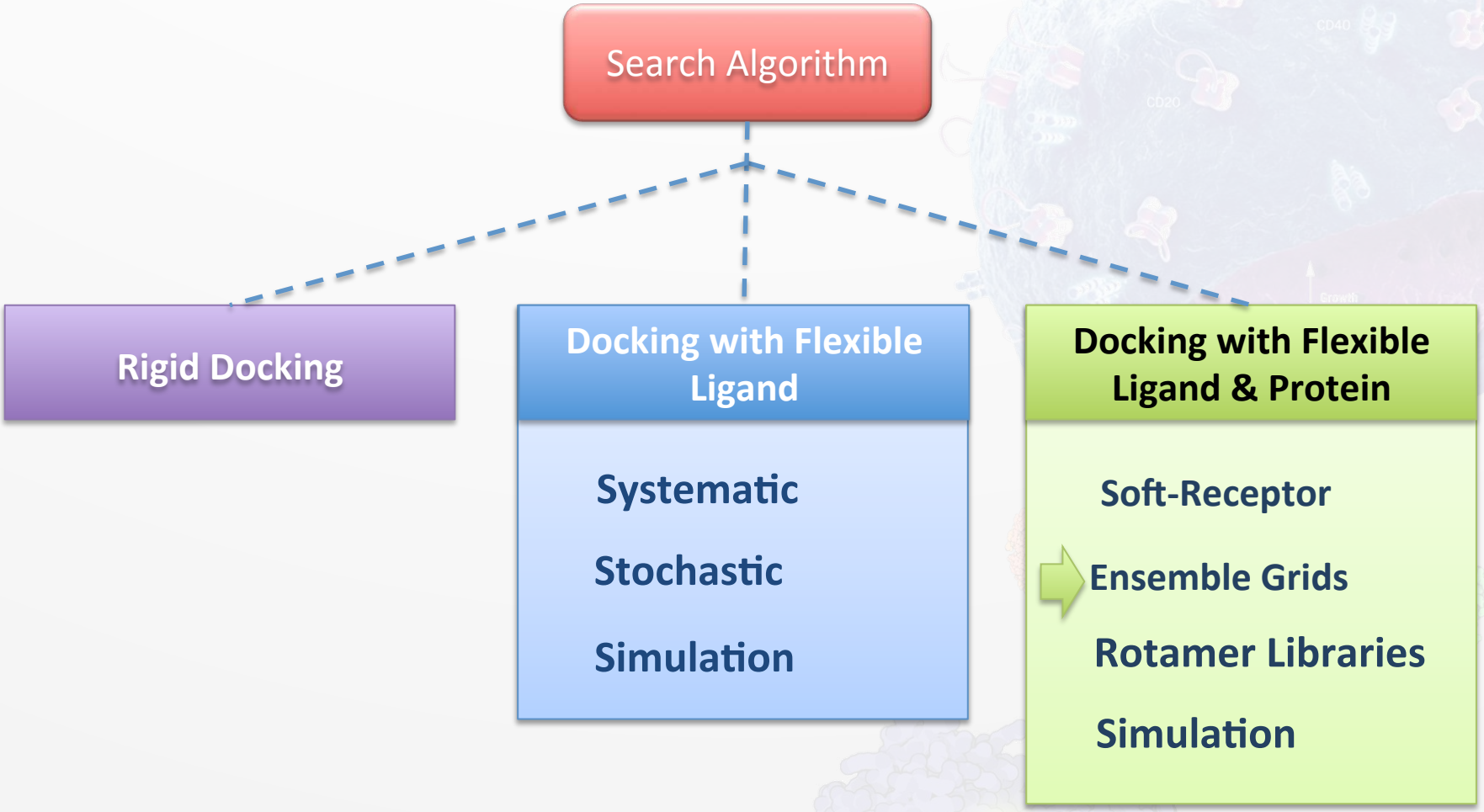
- These methods use an ensemble of protein conformations as a target for docking instead of a single structure.
- How to obtain multiple protein structures? (NMR, X-ray, calculations, …)
- The flexible ligand is then docked to each of these conformations





2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Docking with Flexible Protein and Ligand*

Soft-Receptor

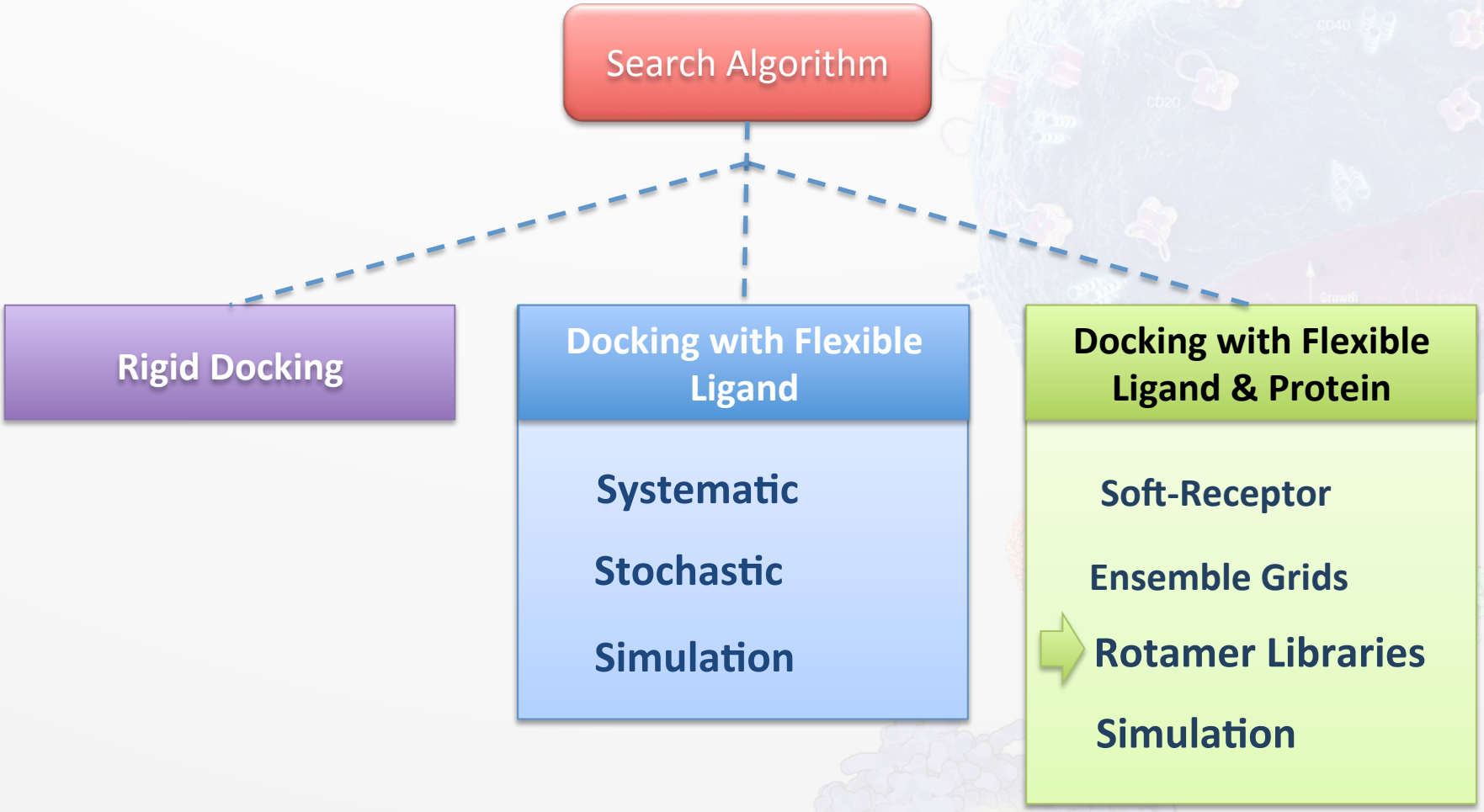
search Algorithms

Description :

- These methods combine the information of several different protein conformations (experimental or computationally derived) to generate an “energy weighted average” grid.
- The flexible ligand is then docked to this structure.
- This approach is relatively cheap as a single energy grid is used as a target for ligand docking.
- Cannot handle large scale motion.

2.2. Molecular Docking

f) Search Algorithms



2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Docking with Flexible Protein and Ligand*

Rotamer Libraries

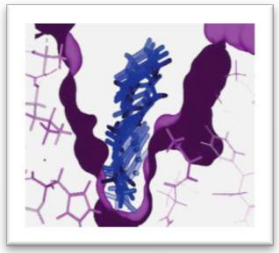
search Algorithms

Description :

- These method try to represent the protein conformational space as a set of experimentally observed and preferred rotameric states for each side chain (Rotamer Libraries).
- Generally, only applicable to 5 or 10 amino acids
- It neglects the flexibility of the backbone
- However, focusing solely on the side chains neglects any real change in the backbone of the protein, and therefore to give a reasonable account of protein flexibilityit is required to go beyond simple-side chain reorientation is needed.

2.2. Molecular Docking

f) Search Algorithms



Search Algorithm

"How to explore the conformational space…"

"How to explorer the binding mode…"

Rigid Docking

Docking with Flexible Ligand

- Systematic
- Stochastic
- Simulation

Docking with Flexible Ligand & Protein

- Soft-Receptor
- Ensemble Grids
- Rotamer Libraries
- Simulation

2.2. Molecular Docking

f) Search Algorithms

SEARCH ALGORITHM *: Docking with Flexible Protein and Ligand*

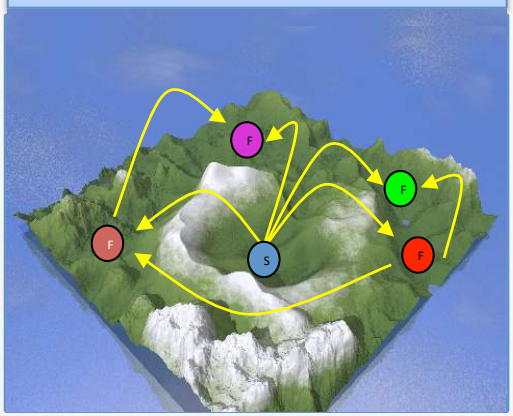
Simulation

search Algorithms

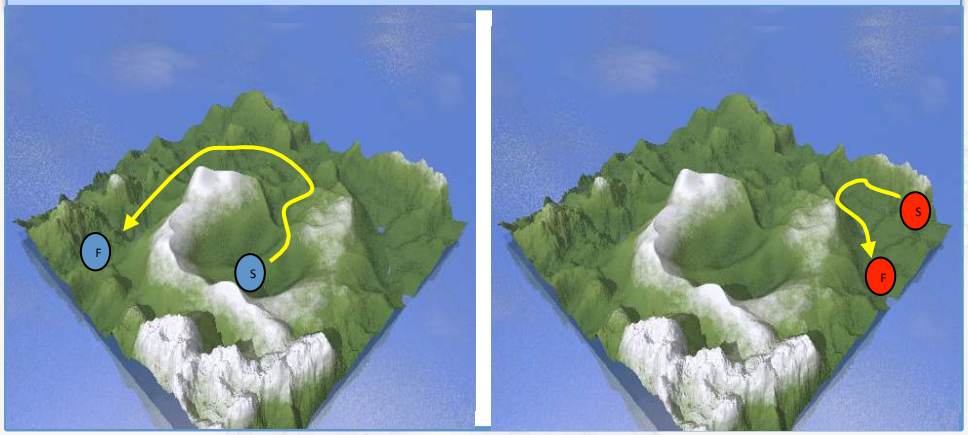
Description :

- Involves the creation of an ensemble of conformation of the protein and calculating thermodynamic properties such as free energy calculations with MC or MD simulations
- Generally they are not used as a standalone approximation. (long time requirements)

Monte Carlo



Molecular Dynamics

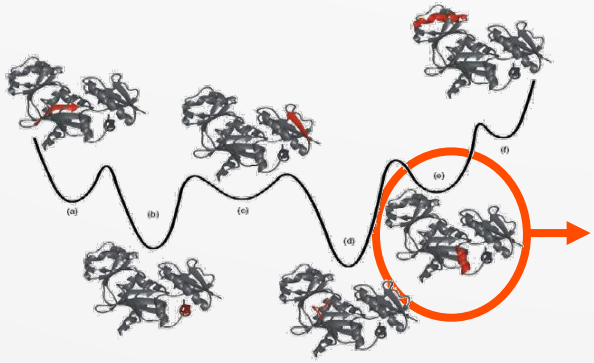


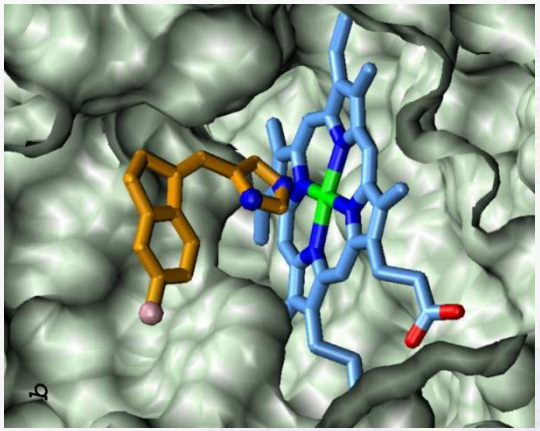
These methods are highly reliable and are typically comparable with experimental data within 1-2 kcal/mol. They are the best choice for accurately assessing chemical modifications that can be made to existing inhibitors to improve their binding affinity

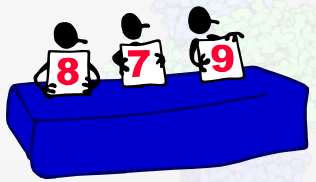
2.2. Molecular Docking

g) Scoring Functions

Scoring Function

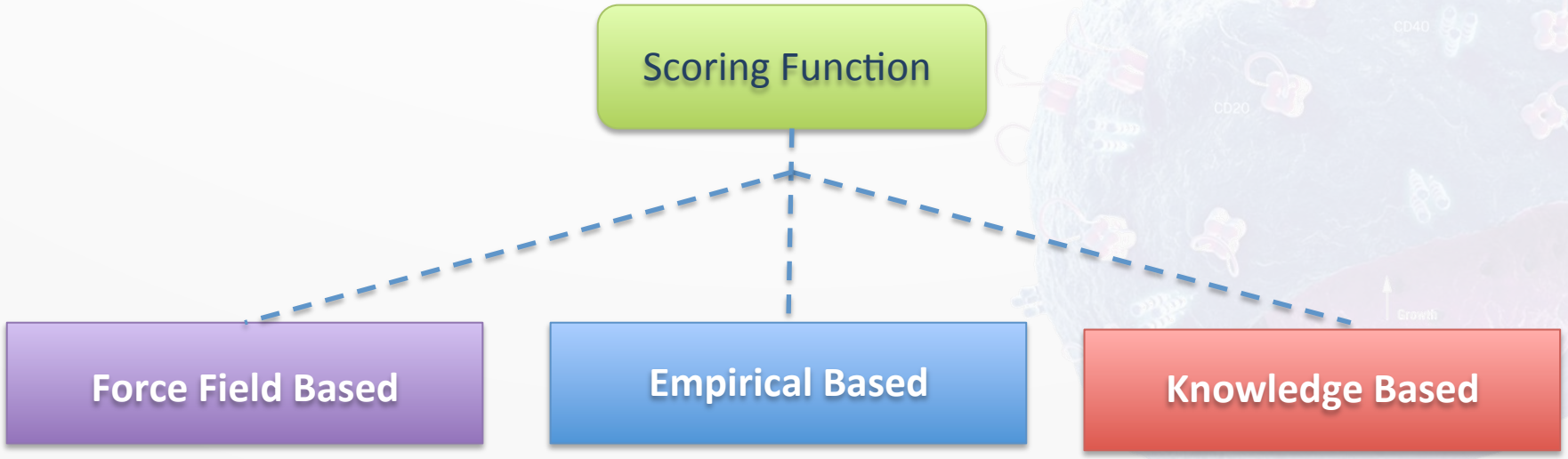






2.2. Molecular Docking

g) Scoring Functions



Being capable of **generating the right conformation is not enough**. It is also necessary to be able to recognize it.

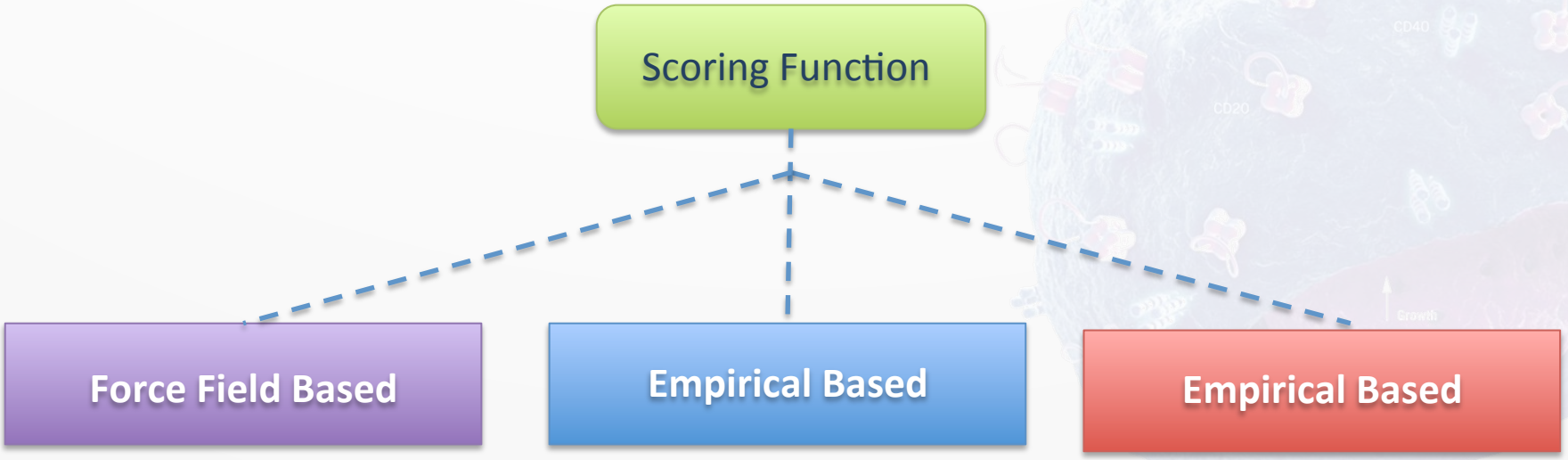
The evaluation and **ranking of the ligand conformations** predicted on the basis of the search algorithm is a critical aspect of every docking protocol.

The scoring function should enable **the distinction between the true binding modes** and all the other alternative modes explored, or between active and random compounds.

However, a very rigorous scoring function would be computationally too expensive.

2.2. Molecular Docking

g) Scoring Functions



2.2. Molecular Docking

g) Scoring Functions

Force Field Based

Scoring Functions

Description:

- Standard force fields that quantify the sum of two energies:

  The interaction energy between the receptor and the ligand, and the internal energy of the ligand

- The energies are normally accounted through a combination of a van der Waals with an electrostatic energy terms:

  A Lennard-Jones potential is used to describe the van der Waals energy term, whereas the electrostatic term is given by a Coulombic term.

- Disadvantages:

  Absence of solvation and entropic terms – these are treated empirically by non – forcefield methods.
  Inaccurate treatment of the long-range effects involved in binding.

- Software: AutoDock, GOLD (Gold Score), G-Score, D-Score

2.2. Molecular Docking

g) Scoring Functions

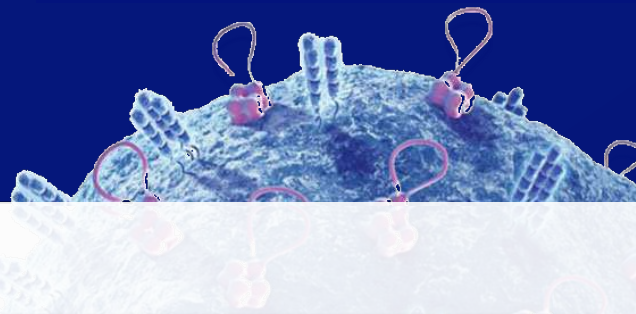
Force Field Based

Scoring Functions

	Protein–ligand	Internal ligand
G-Score	$E_{vdW} + E_{H-bond} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{8}} - \frac{B_{ij}}{d_{ij}^{4}}\right) + (E_{da} + E_{ww}) - (E_{dw} + E_{aw})\right]$	$E_{vdw} + E_{torsion} = \sum_{lig}\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{6}}\right) + \sum_{lig}\frac{1}{2}V\left[1 + \frac{n}{\lvert n\rvert}\cos(\lvert n\rvert\omega)\right]$
D-Score	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} + \frac{B_{ij}}{d_{ij}^{6}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$	
Gold	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$	$E_{vdW} + E_{electrostatic} = \sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$ + optional E_{H-bond}
AutoDock	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\times\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$ $E(t) =$ angular weight factor	$E_{vdW} + E_{H-bond} + E_{electrostatic} = \sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^{6}}\right) + E(t)\left(\frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}}\right) + 332.0\frac{q_i q_j}{4(d_{ij}) d_{ij}}\right]$ $E(t) =$ angular weight factor
DOCK (v4.0)	$E_{vdW} + E_{electrostatic} = \sum_{prot}\sum_{lig}\left[\left(\frac{A_{ij}}{d_{ij}^{a}} + \frac{B_{ij}}{d_{ij}^{b}}\right) + 332.0\frac{q_i q_j}{\in (d_{ij}) d_{ij}}\right]$	

2.2. Molecular Docking

g) Scoring Functions



Force Field Based

Scoring Functions

AutoDock Scoring Function

$$\Delta G = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i_C,j} S_i V_j e^{(-r_{ij}^2/2\sigma^2)}$$

SCORING FUNCTION IN AUTODOCK

Molecular Mechanics Terms

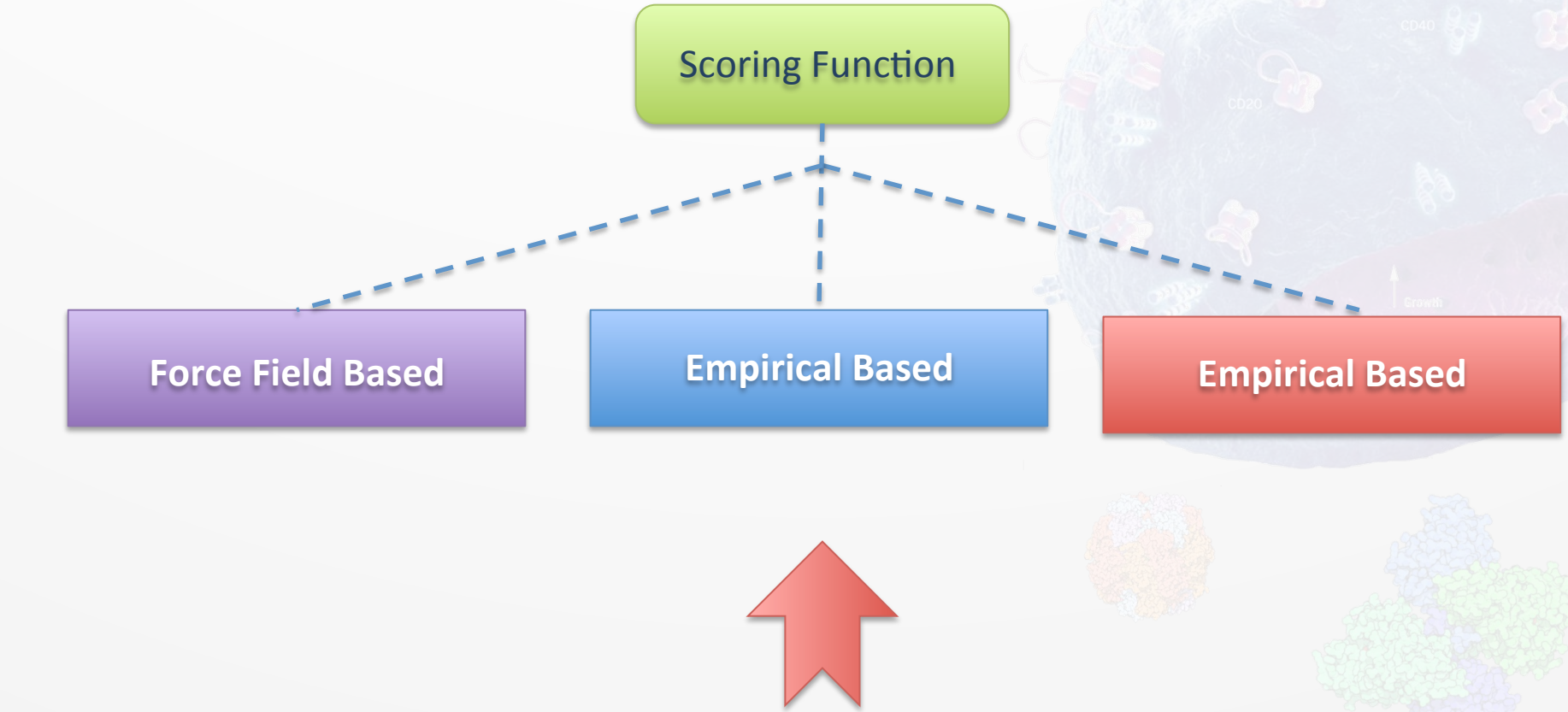
van der Waals	$\Delta G_{vdW} = W_{vdW} \sum_{i,j} (A_{ij} / r_{ij}^{12} - B_{ij} / r_{ij}^{6})$
Hydrogen Bonding	$\Delta G_{H\text{-}bond} = W_{H\text{-}bond} \sum_{i,j} E(t) * (C_{ij} / r_{ij}^{12} - D_{ij} / r_{ij}^{10} + E_{hbond})$
Electrostatics	$\Delta G_{elec} = W_{elec} \sum_{i,j} (q_i * q_j) / (\varepsilon(r_{ij}) * r_{ij})$
Desolvation	$\Delta G_{desolv} = W_{desolv} \sum_{i\,(C),j} (S_i * V_j * \exp (-r_{ij}^2 / (2 * \sigma^2))$

Change in Torsional Free Energy when the Ligand goes from Unbound to Bound

Torsional	$\Delta G_{tor} = W_{tor} N_{tor}$

2.2. Molecular Docking

g) Scoring Functions



2.2. Molecular Docking

g) Scoring Functions

Empirical

Scoring Functions

Description:

- Algorithms that try to reproduce experimental data and are based on the idea that binding energies can be approximated by a sum of several individual and uncorrelated terms.

- Experimentally determined binding energies and sometimes a training set of experimentally resolved receptor–ligand complexes, and are used to determine the coefficients for the various terms by means of a regression analysis.

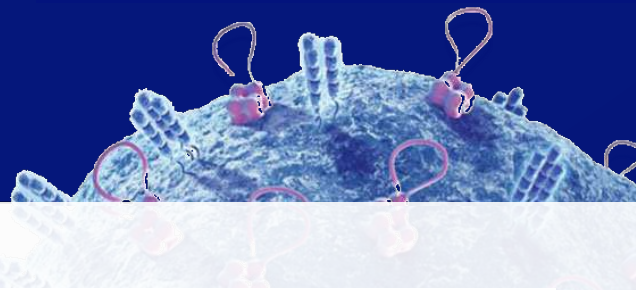
Disadvantages:

Dependent on available experimental data (non versatile and non transferable).

- Software: GOLD (Chem Score), Bo¨ hm's scoring Function (LUDI), SCORE, F-SCORE,etc.

2.2. Molecular Docking

g) Scoring Functions



Empirical

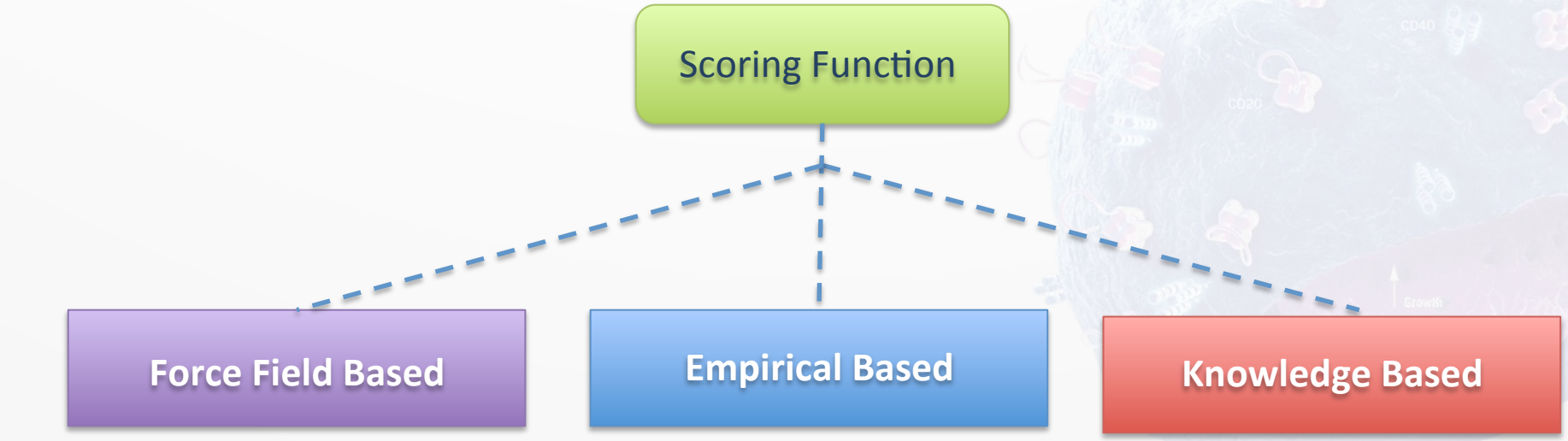
Scoring Functions

	Functional form
LUDI	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{hydrophobic} \sum_{hydrophobic} \lvert A_{hydrophobic} \rvert + \Delta G_{rotor} N_{rotor} + \Delta G_0$ $A_{hydrophobic}$ = molecular surface area
F-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta\alpha) + \Delta G_{aromatic} \sum_{aromatic} f(\Delta R, \Delta\alpha) + \Delta G_{contact} \sum_{contact} f(\Delta R, \Delta\alpha) + \Delta G_{rotor} N_{rotor} + \Delta G_0$
Chem-Score	$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta\alpha) + \Delta G_{metal} \sum_{metal} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo} f(\Delta R) + \Delta G_{rotor} \sum_{rotor} f(P_{nl}, P_{nl}') + \Delta G_0$

The free energy of binding, ΔG_{bind}, is approximated as a sum of contributing free energy terms of hydrogen bonding (*H-bond*), ionic (*ionic*), hydrophobic (*hydrophobic*), ligand rotational entropy (*rotor*), contact (*contact*), lipophilic (*lipo*) and metal (*metal*) components. The scoring functions differ in the terms included and the functional forms of the contributing free energy terms. $\Delta G_{H\text{-}bond}$, ΔG_{ionic}, $\Delta G_{hydrophobic}$, ΔG_{rotor}, $\Delta G_{aromatic}$, $\Delta G_{contact}$, ΔG_{metal}, ΔG_{lipo} are regression coefficients for each corresponding free energy term. ΔG_0 is a regression constant. The free energy terms are calculated with a function, *f*, which can depend on an angular ($\Delta\alpha$) and/or a b distance (ΔR) term.

2.2. Molecular Docking

g) Scoring Functions



2.2. Molecular Docking

g) Scoring Functions

Knowledge Based

Scoring Functions

Description:

- Focused on following the rules and general principles statistically derived that aim to reproduce experimentally determined structures, instead of binding energies
- Try to implicitly capture binding effects that are difficult to model explicitly.
- These potentials are based on the frequency of occurrence of different atom–atom pair contacts and other typical interactions in large datasets of protein–ligand complexes of known structure.

Disadvantages:

Dependent on the information available in limited sets of structures.

- Software: Muegges's Potential of Mean Force (PMF), DrugScore, and SMoG score.

2.2. Molecular Docking

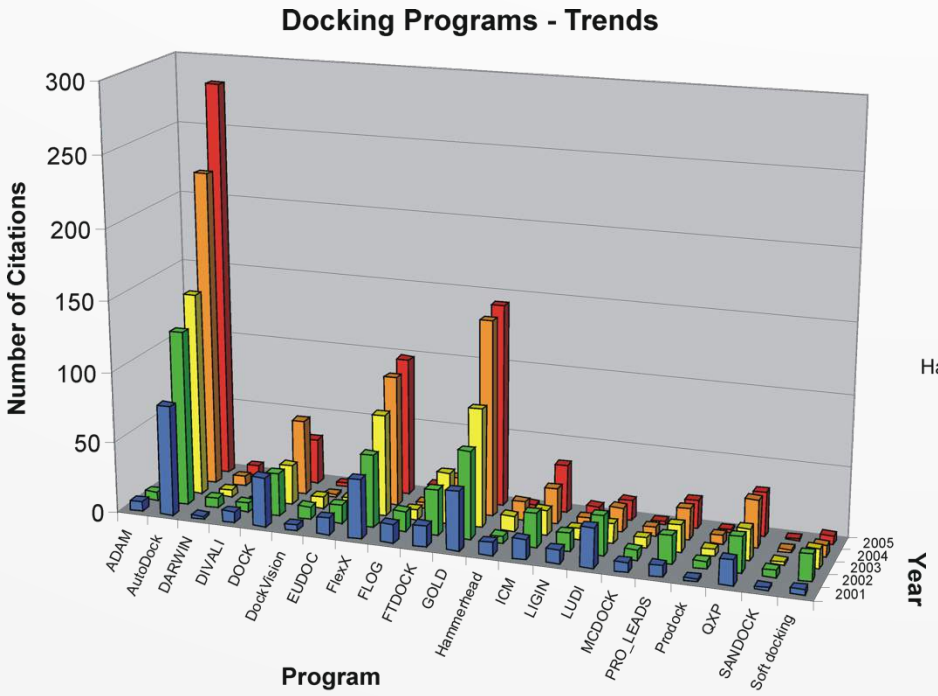
g) Scoring Functions

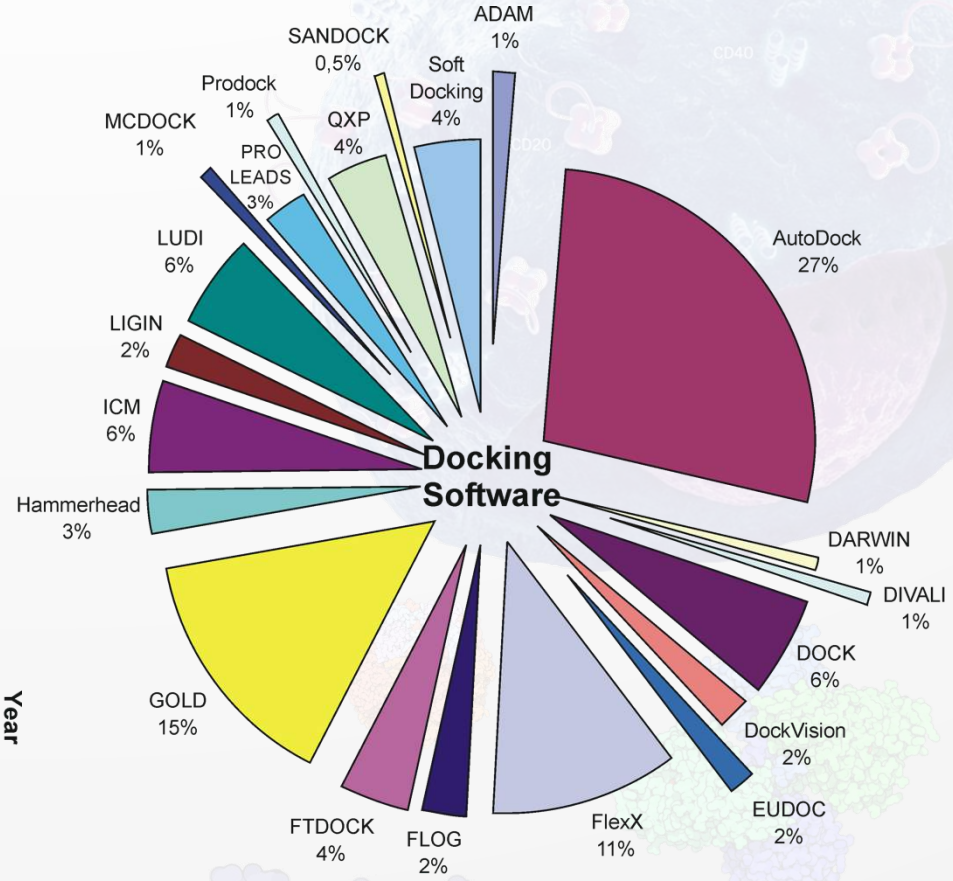
Knowledge Based

Scoring Functions

	Functional form
PMF	Parametrized pairwise potential PMF score : $PMF = \sum_{prot}\sum_{lig} A_{ij}(d_{ij}) \quad A_{ij}(d_{ij}) = -k_B T \ln\left[f^{j}_{Vol\_corr}(r) \frac{\rho^{ij}_{seg}(r)}{\rho^{ij}_{bulk}} \right]$ where k_B is the Boltzmann constant, $f^{j}_{Vol\_corr}(r)$ is a ligand volume correction factor and $\frac{\rho^{ij}_{seg}(r)}{\rho^{ij}_{bulk}}$ indicates a radial distribution function for a protein atom i and a ligand atom j.
DrugScore (v1.2)	$\Delta W = \gamma \sum_{prot}\sum_{lig} \Delta W_{ij}(r) + (1-\gamma) \times \left[\sum_{lig} \Delta W_i(SAS, SAS_0) + \sum_{prot} \Delta W_j(SAS, SAS_0) \right]$ SAS = Solvent accessible surface area terms, W_{ij} = distance dependent pairwise potential
SMoG	$G = \sum_{ij} g_{ij} \Delta_{ij}; \qquad \Delta_{ij} = \begin{cases} 0 & \text{(i, j more than 5 Å)} \\ 1 & \text{(i, j within 5 Å)} \end{cases}; \qquad g_{ij} = -kT \log\left[\frac{p_{ij}}{\overline{p}} \right];$ p_{ij} an $\overline{p}$ are interatomic and averaged interactomic interactions

2.3. Available Docking Software





Protein-ligand docking: Current status and future challenges, Proteins, 2006, 65, 15.

2.4. Conclusions

Most docking programs are normally able to predict known protein- bound poses with averaged accuracies of about 1.5–2 Å with reported success rates in the range of 70 – 80%

Improvement beyond this range seems for now unachievable, even with the inclusion of receptor flexibility.

Major bottlenecks:

Scoring functions are the Achilles' heel of docking programs.

The solvent effect and the direct participation of water molecules in protein–ligand interactions.

The limited resolutions of most crystallographic targets

Protein flexibility, both in terms of intrinsic structural flexibility and in terms of conformational alterations upon ligand binding.

Molecular Docking

Pedro Alexandrino Fernandes,

Dep. Chemistry & Biochemistry, University of Porto, Portugal

pedro.fernandes@fc.up.pt