



# **“Computational Strategies in Drug Discovery... myth and reality.”**

**Stefano Moro**



**Molecular Modeling Section (MMS)**

**Department of Pharmaceutical and Pharmacological Sciences  
University of Padova**

**©2014**



**The most insidious question that  
you can make me :**

**Is a drug designable?**



**Attending to answer this question...  
we try to find an intimate connection  
among these three concepts :**

**Design**

**Drug**

**Informatics**



... a bit of:

**Design**

**Drug**

**Informatics**





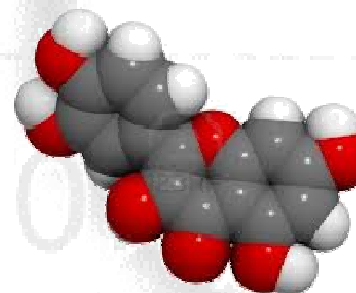
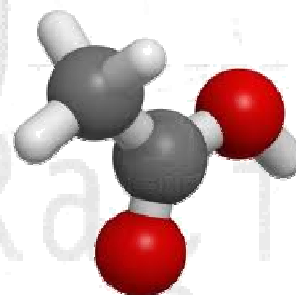
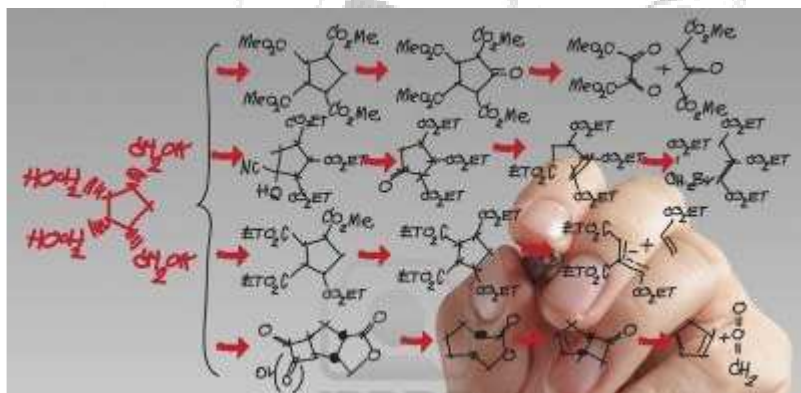
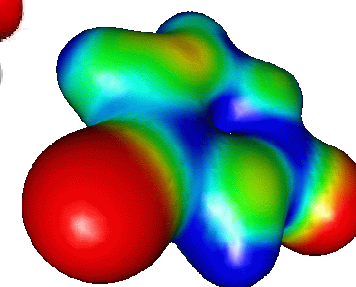
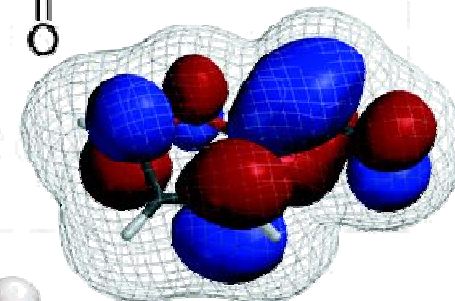
... be patient:

## Design:

set up a project of a work by making  
**drawings and calculations** necessary  
for its realization.



The chemical structure of 1,2-dihydroxyanthraquinone (DHA) is shown. It consists of a central benzene ring fused to two outer benzene rings. The central ring has two carbonyl groups (C=O) at the 9 and 10 positions. The right-hand outer ring has two hydroxyl groups (OH) at the 1 and 2 positions.



**... but what *chemical drawings* can be subjected to *calculation procedures*?**



... a bit of:

**Design**

**Drug**

**Informatics**



I know very well that you know what drug is... but reconsider its definition in this way:

## Drug:

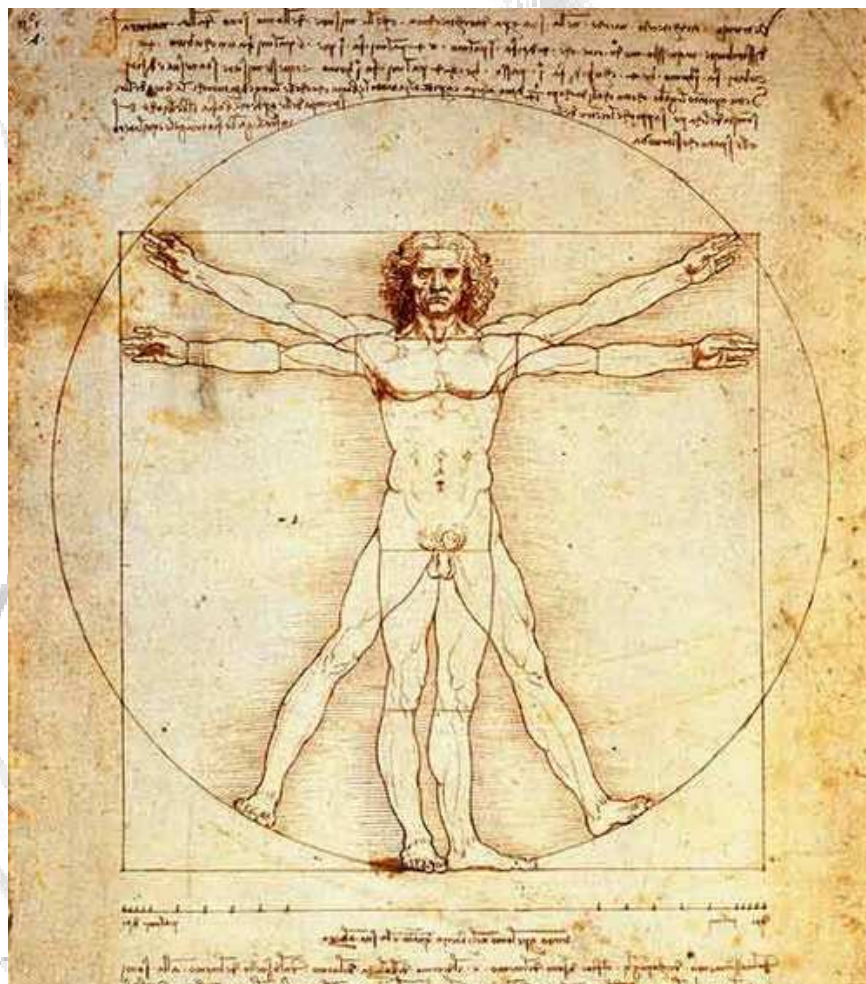
[from gr. φάρμακον] *Any substance, organic or inorganic, synthetic or natural, capable of producing in a **living organism** functional modifications, helpful or harmful, by chemical or physical action.*

*Dizionario Treccani*



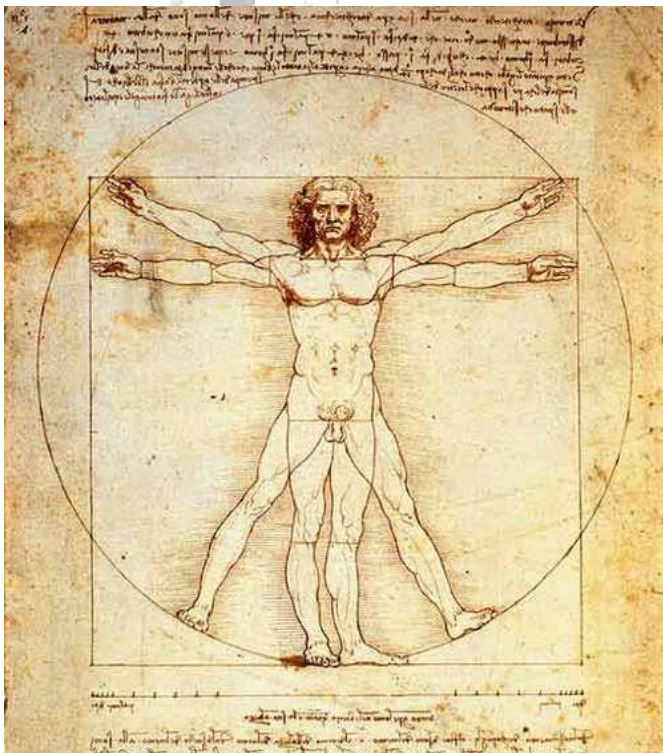


... and a '*living organism*' is difficult to accommodate in a design process (drawing and calculations), though ...





... living organism is still too complex to *draw*



**We will return later on this concept...**





... couple of bits of:

**Design**

**Drug**

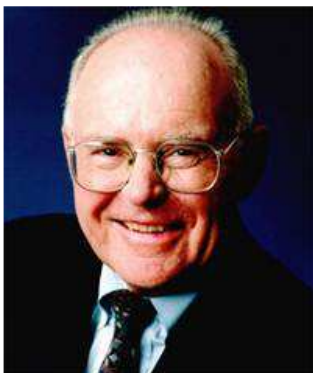
**Informatics**



**Now, in what informatics has influenced more in our daily life?**

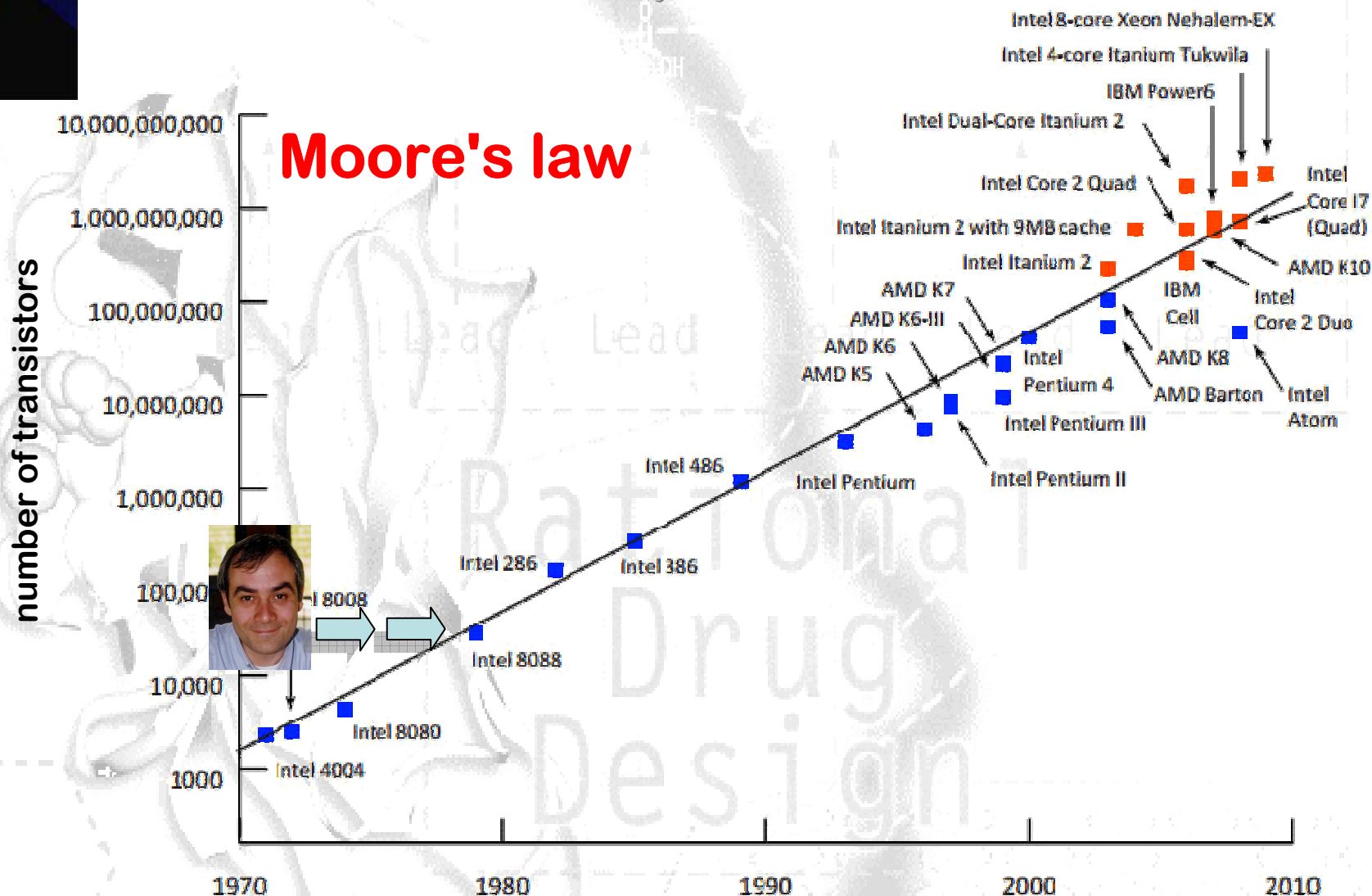
**Well, we probably summarize the answer in only one word:**

**Time**



# Informatics as synonymous of speed?

## Moore's law





## **couple of concrete example..**

**Informatics helps us to solve simple problems a number of times:**

**Calculate the molecular weight is trivial thing, calculate 4.5 million ... less!**

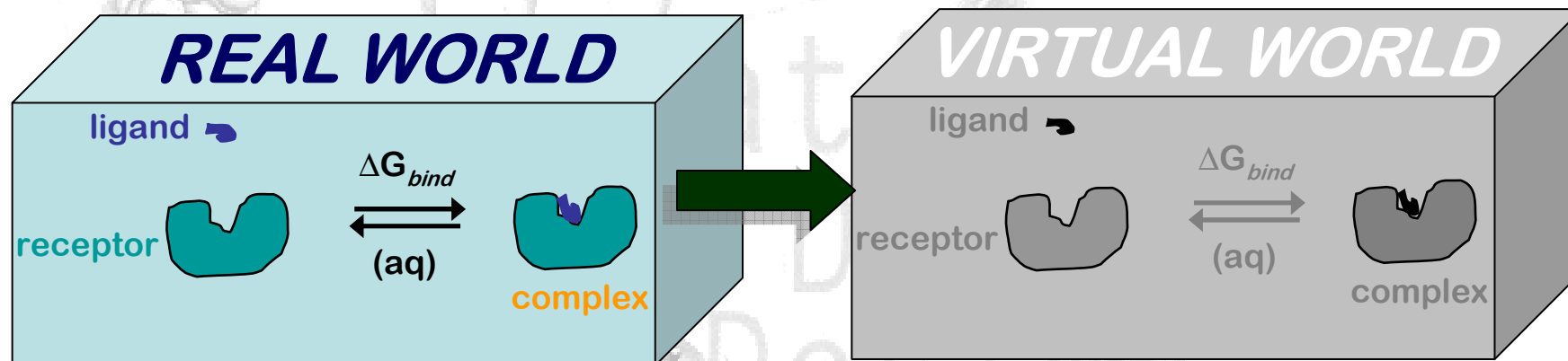
**or solve very complex problems:**

**Calculate, for example, the binding energy ( $\Delta G_{\text{bind}}$ , kcal x mol) between a ligand and its receptor!**



**Informatics is the basic science of any *virtualization* process:**

***virtualization process: the creation of a abstraction version of the real process.***

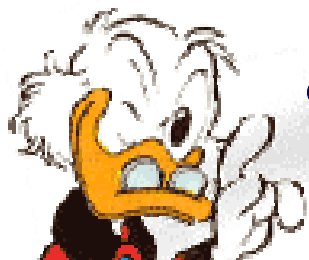




# The *accuracy* of this virtualization process is crucial:







**and remember... “Time is money!”**


**Bringing a new drug to market can take 8-14 years  
and costs between \$400 and \$900 million**

Rational  
Drug  
Design



## Some details about costs:

### *Experiment Typical Cost per Compound (€)*

<b>Computer modeling</b>	<b>7</b>	
Biochemical assay	270	
Cell culture assay	2.700	
Rat acute toxicity	8.100	
Protein crystal structure	68.000	
Animal efficacy trial	200.000	
Rat 2-year chronic oral toxicity	550.000	
Human clinical trial	3.500.000	

**You understand why it is so attractive to the pharmaceutical industry!**



# Back to “drug” design...

**Design**

**“Drug”**

**Informatics**



But unfortunately “drug” is not designable... yet!

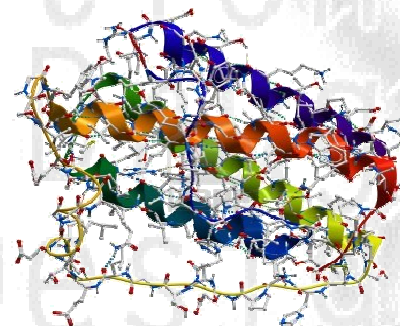
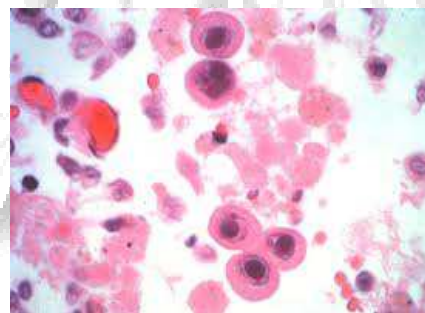
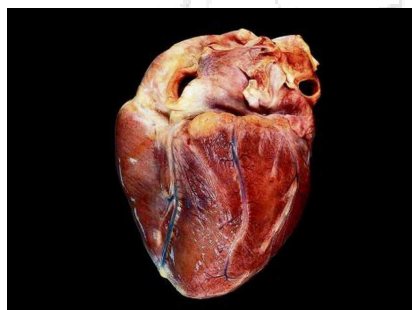
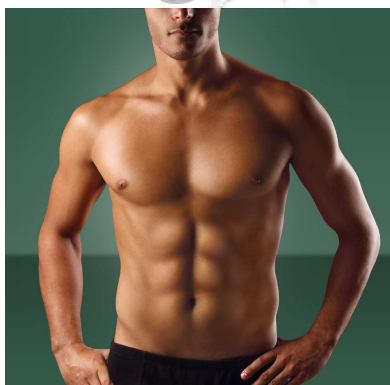
Could we suggest a possible substitute of “drug” that is more easily designable?

To do this we have to necessarily replace the concept of *living organism*!



# An egoistic solution:

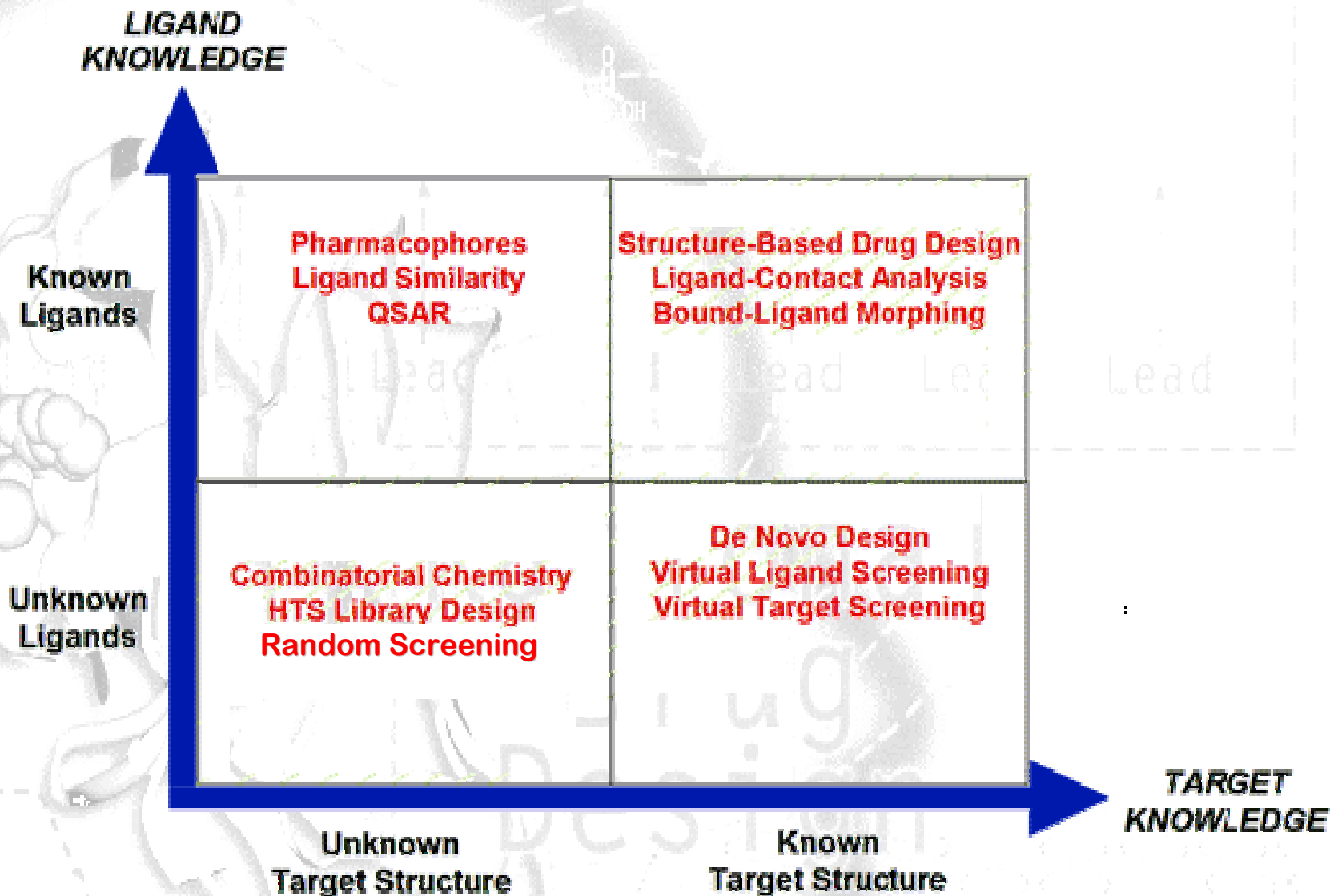
*drug*



*candidate*



# We surely need informatics but:





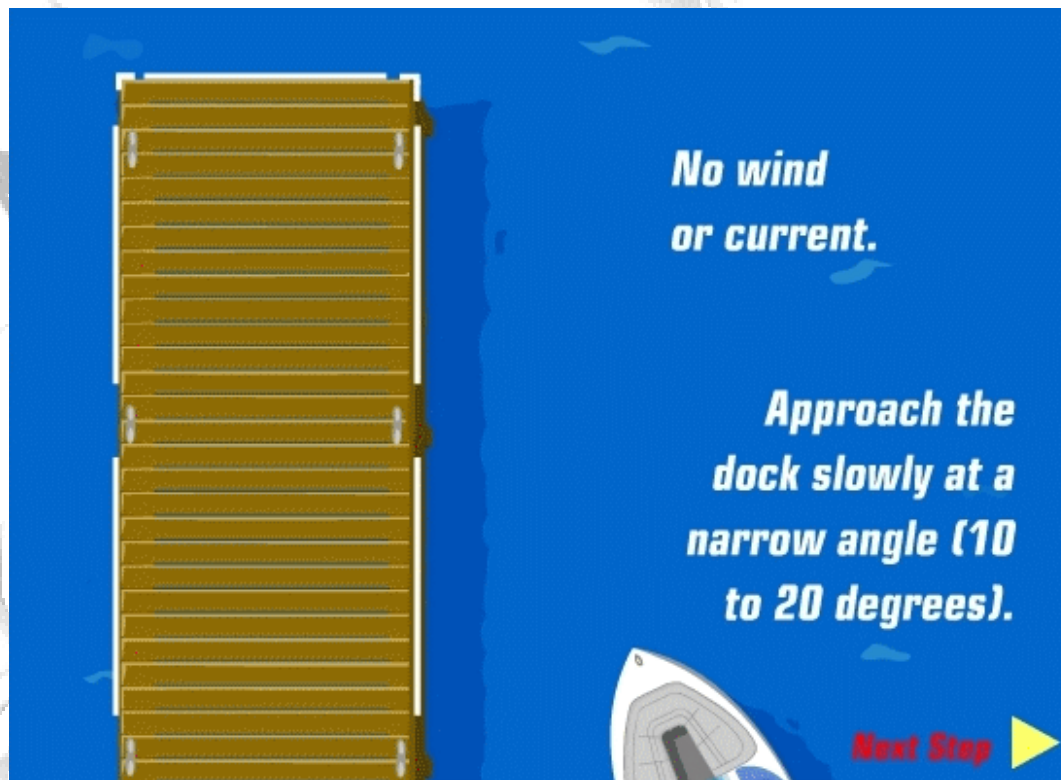
Back when I was young:

“It is generally accepted that receptor and substrate molecules recognize each other at their molecular surfaces. Therefore, the binding strength of a receptor-drug complex depends on the shape of the substrate surface and on the distribution of certain properties on this surface. Any method attempting to model biological activity should take into account this information and try to correlate it to biological activity...”

by Johann Gasteiger *et al J.A.C.S.* 1995, 117, 7769-7775



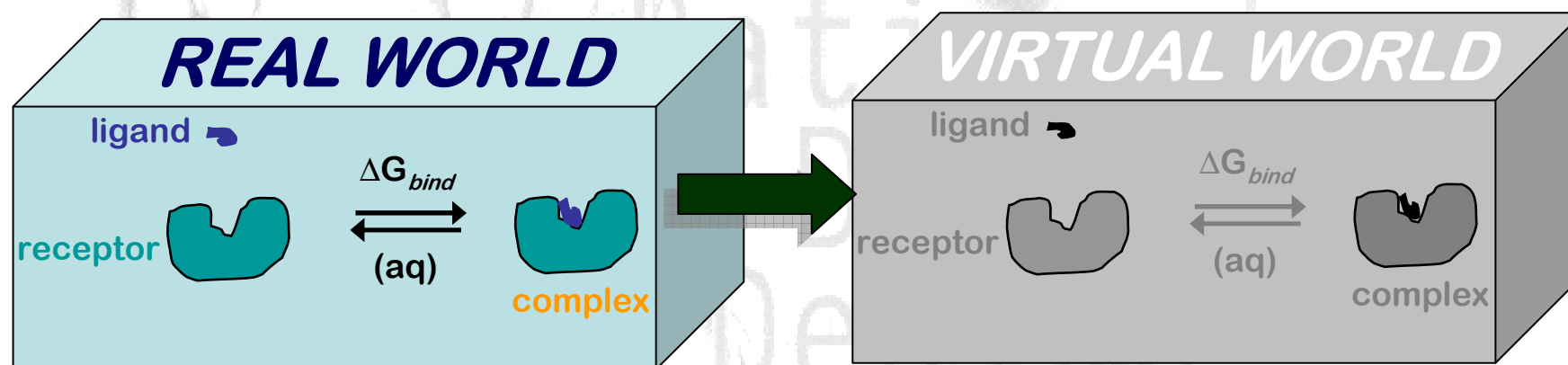
# Docking and Scoring

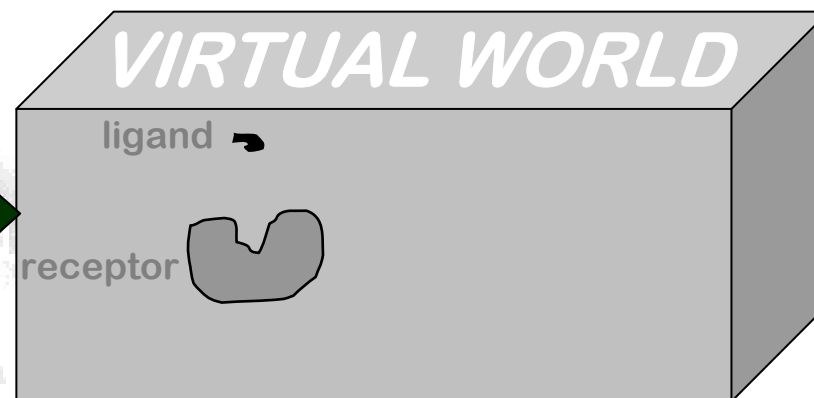
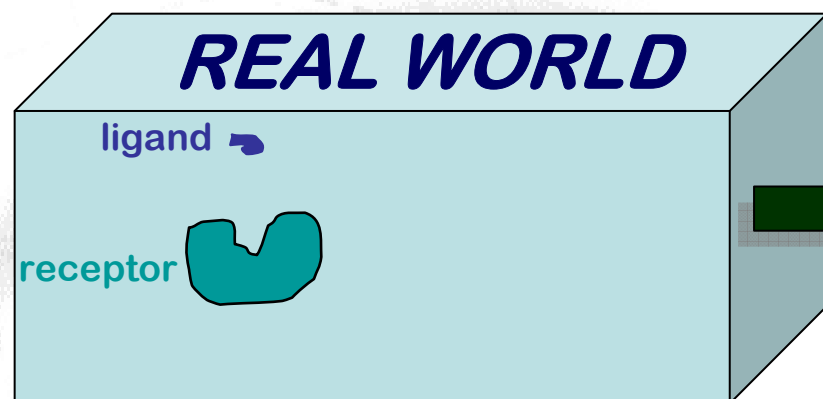




# What Docking really is?

From an informatics point of view, docking is a *virtualization process* dealing with the creation of a virtual (rather than real) version of the binding process.





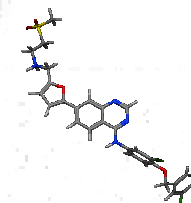
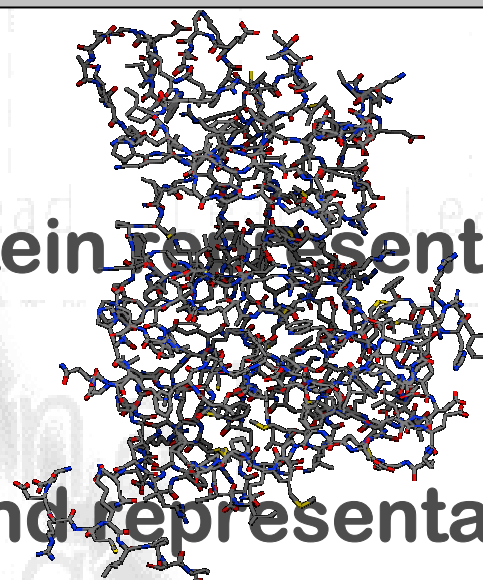
**Real protein**  
Epithelial Growth Factor Receptor  
(kinase domain)

**Real ligand**

**Lapatinib (Tykerb®, GSK)**

**3D protein representation**

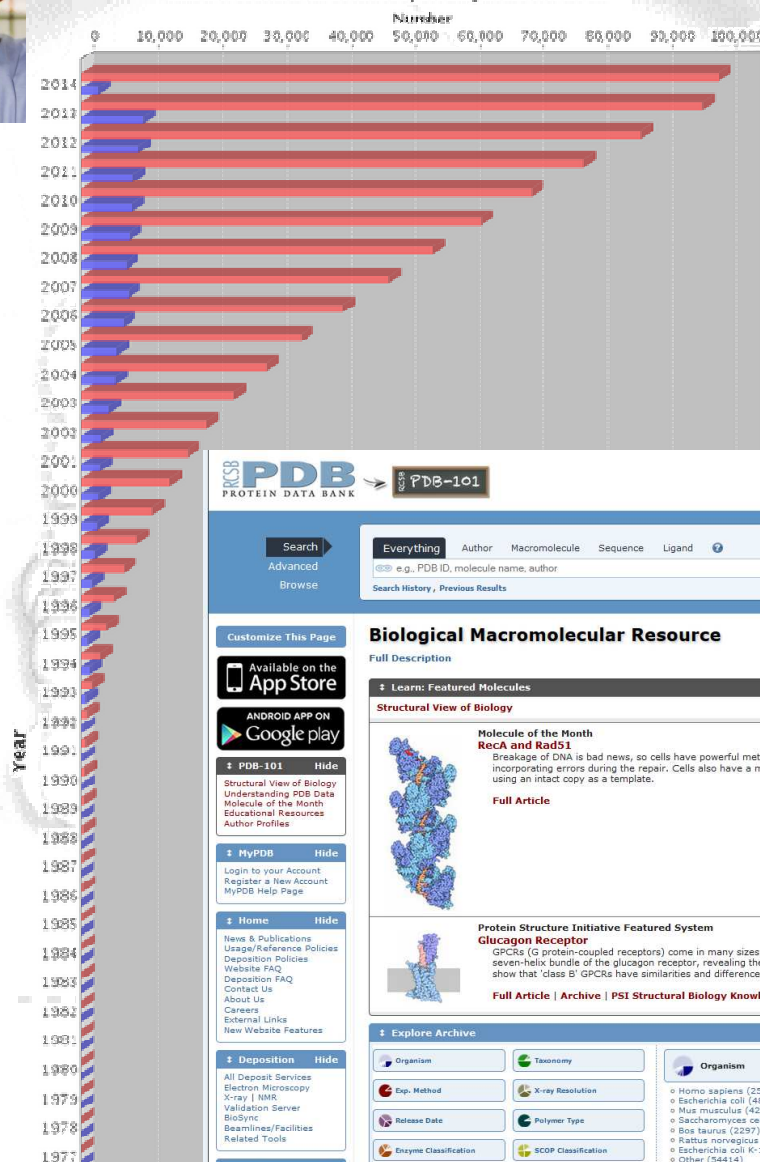
**3D ligand representation**





# Dockers are coordinates hunters!

Yearly Growth of Total Structures  
number of structures can be viewed by hovering mouse over the bar



99122, yesterday.

**Protein Data Bank**  
An Information Portal to Biological Macromolecular Structures  
As of Tuesday Apr 01, 2014 at 5 PM PDT there are 99122 Structures | PDB Statistics

**Search**  
Everything | Author | Macromolecule | Sequence | Ligand  
e.g., PDB ID, molecule name, author  
Search History | Previous Results

**Biological Macromolecular Resource**  
Full Description

**Learn: Featured Molecules**  
Structural View of Biology  
Molecule of the Month  
**RecA and Rad51**  
Breakage of DNA is bad news, so cells have powerful methods to fix damaged DNA. One method trims the broken ends and then reconnects them back together. This is fast and easy, but has the disadvantage of possibly incorporating errors during the repair. Cells also have a more accurate method to repair breaks that relies on duplicate copies of the genome. This process is called homologous recombination, and rebuilds the damaged areas using an intact copy as a template.  
Full Article

**Protein Structure Initiative Featured System**  
**Glucagon Receptor**  
GPCRs (G protein-coupled receptors) come in many sizes and shapes, each recognizing its own type of signaling molecule. PSI researchers at the GPCR Network have recently determined the structure of the signature seven-helix bundle of the glucagon receptor, revealing the atomic details of a class of GPCRs that recognize short peptide hormones. This structure, along with the related structure of the corticotropin-releasing factor receptor 1, show that 'class B' GPCRs have similarities and differences from their GPCR cousins.  
Full Article | Archive | PSI Structural Biology Knowledgebase

**Explore Archive**  
Organism | Taxonomy | Exp. Method | X-ray Resolution | Release Date | Polymer Type | Enzyme Classification | SCOP Classification

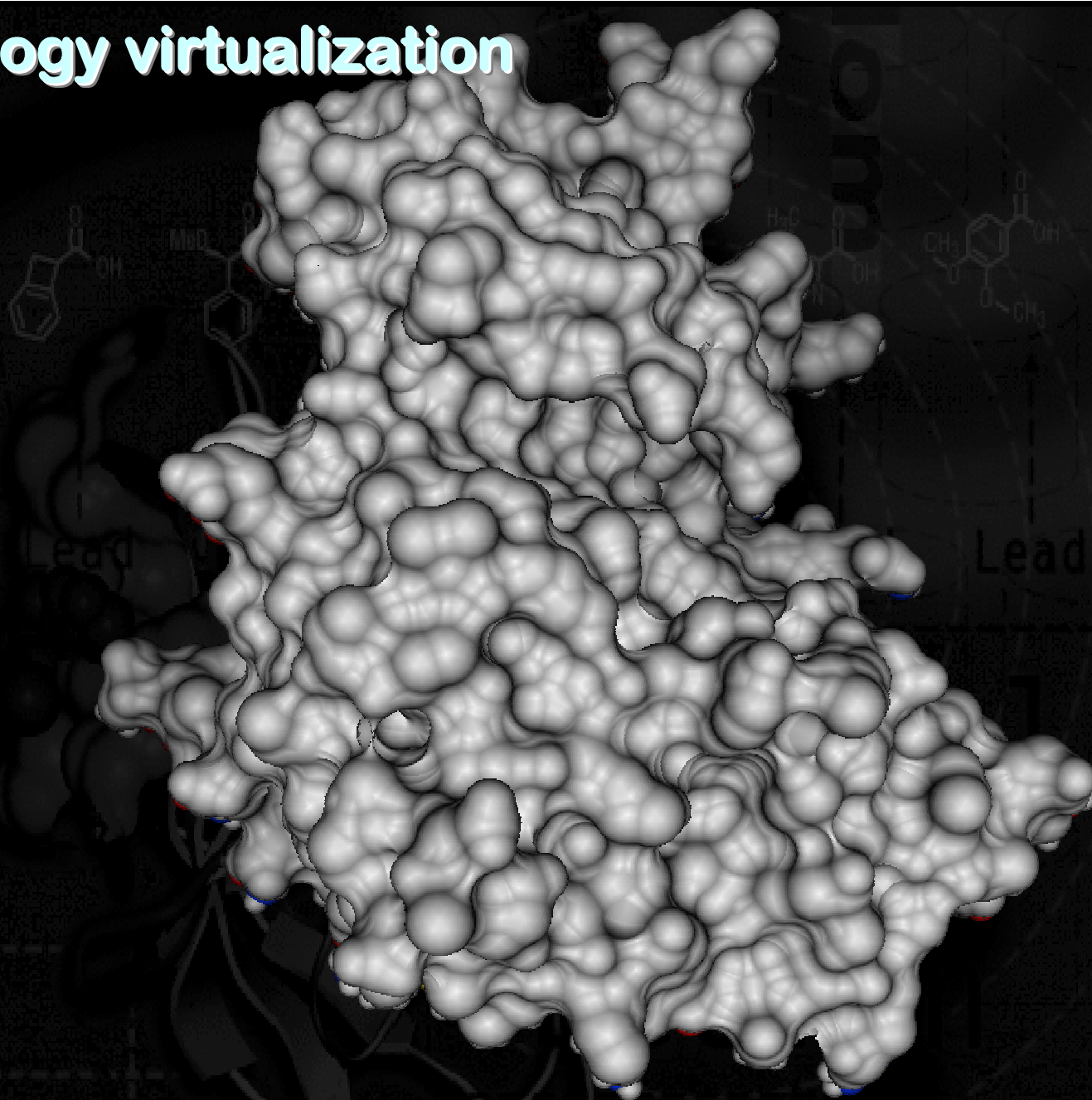
**Organism**  
Homo sapiens (25399)  
Escherichia coli (4818)  
Mus musculus (4296)  
Saccharomyces cerevisiae (2305)  
Bos taurus (2297)  
Rattus norvegicus (2108)  
Escherichia coli K-12 (1898)  
Other (34414)

**New Features**  
Latest release: December 2013  
Tabular Reports  
View Improved Tabular Reports  
Website Release Archive

**RCSB PDB News**  
2014-04-01  
Build a Model of GFP  
Visit PDB-101's Paper Models to create a Green Fluorescent Protein in 3D more  
Four Reel Wishes to Drib

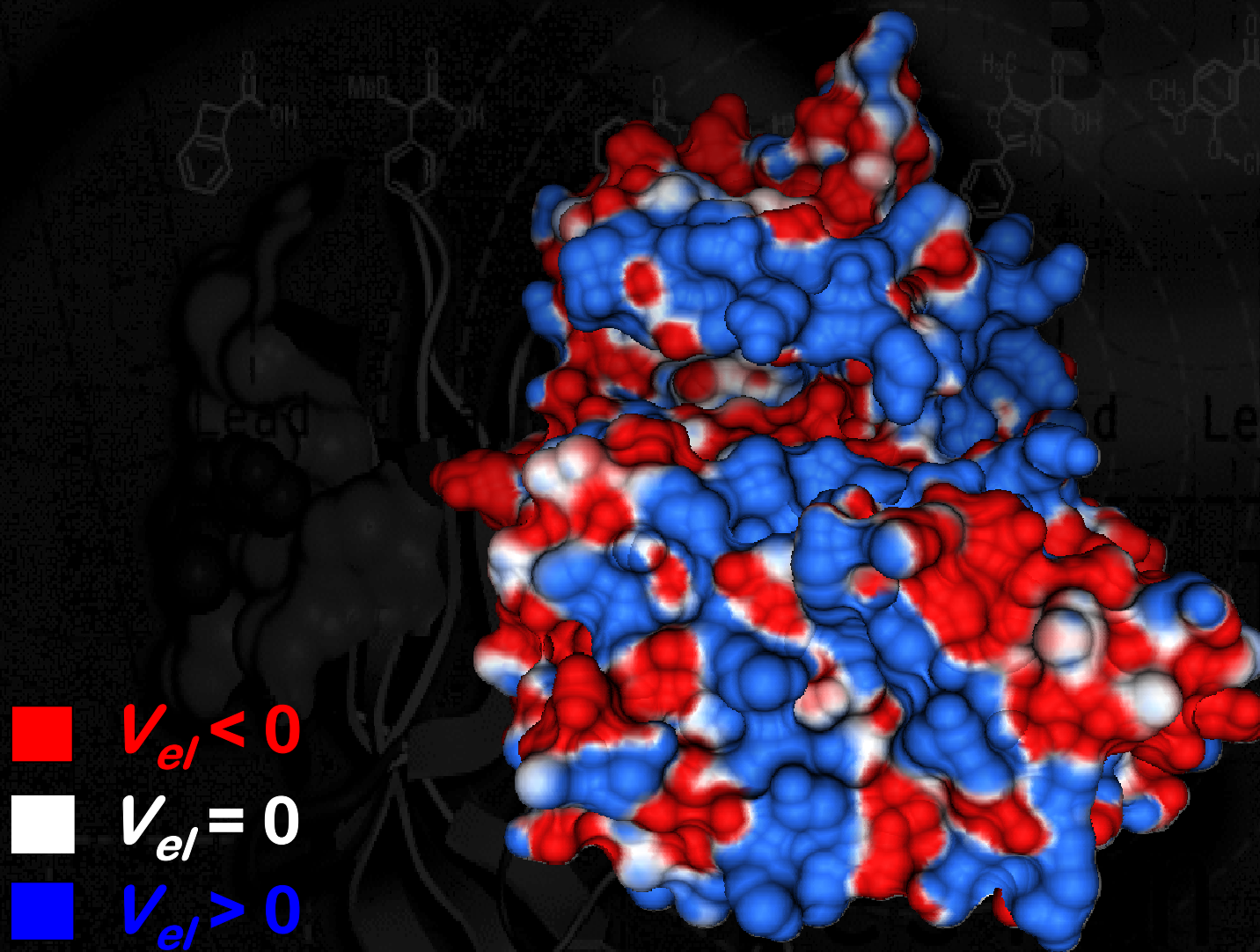


# Topology virtualization



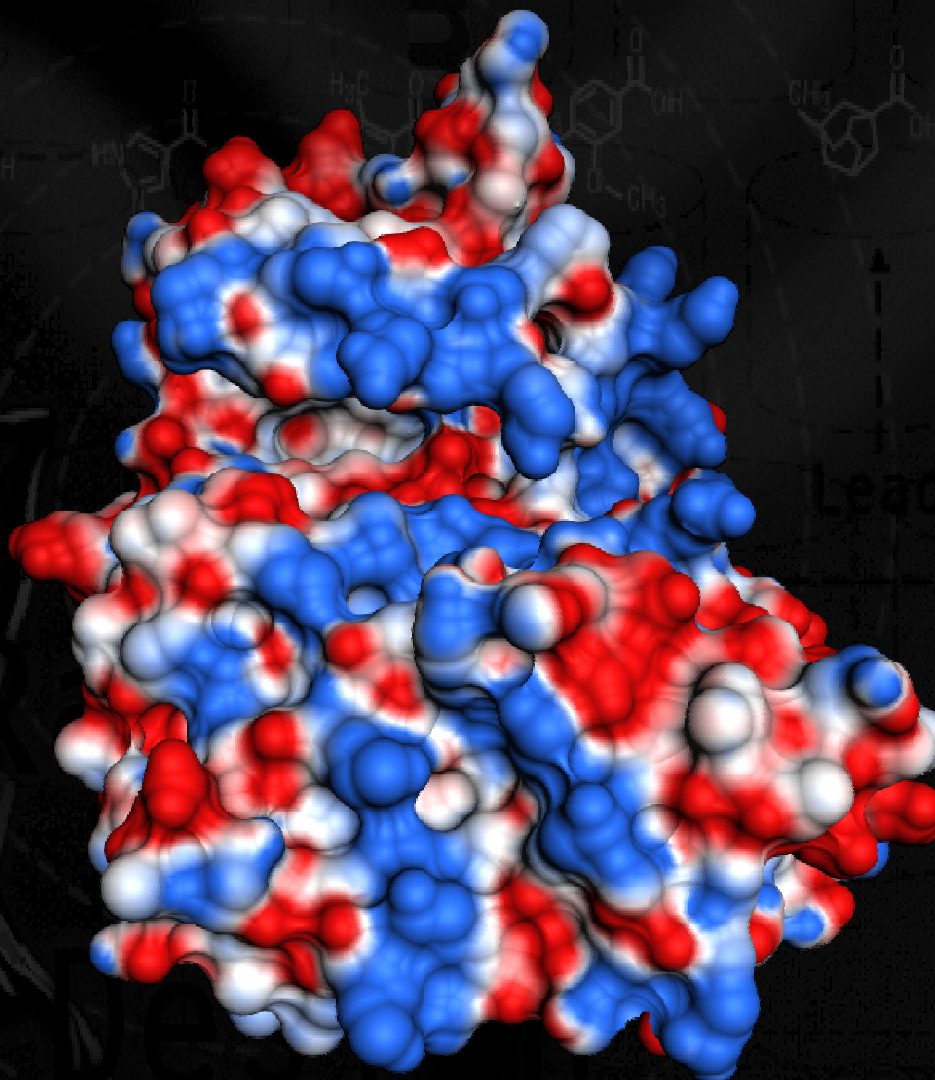
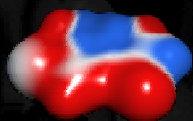


# Surface properties virtualization





# Here is the problem...



1. where?
2. how?
3. how long?



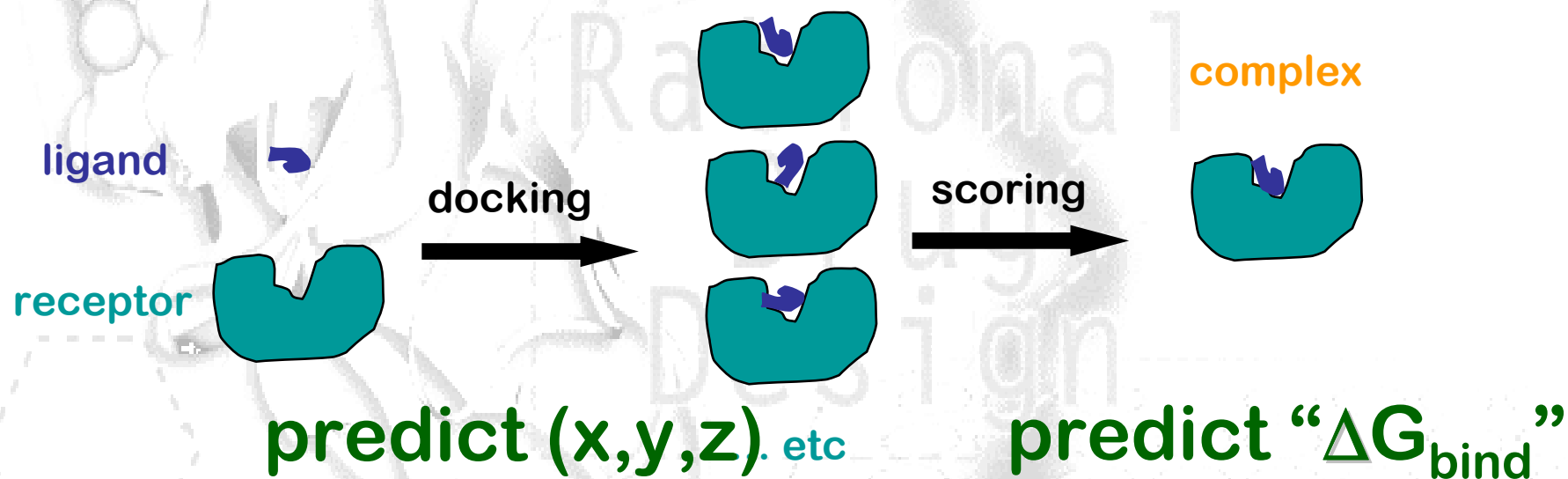
# Stretching time!

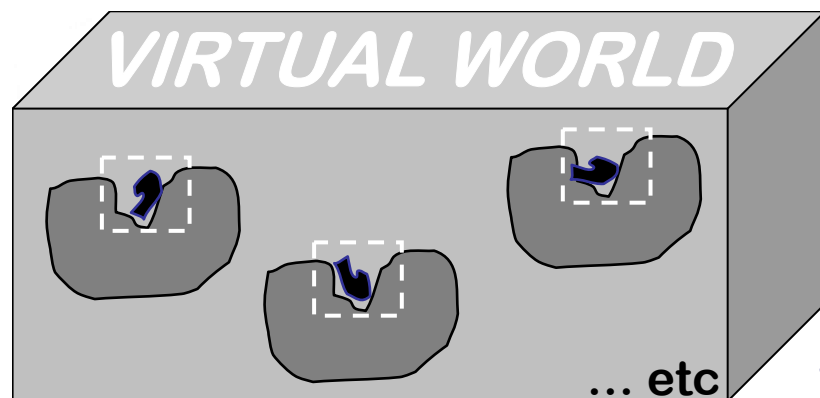


# The molecular docking problem:

To place a ligand (small molecule) into the binding site of a receptor in the manners appropriate for optimal interactions with a receptor (DOCKING).

To evaluate the ligand-receptor interactions in a way that may discriminate the experimentally observed mode from others and estimate the binding affinity (SCORING).





**Some definitions:**

**We define as POSE:**

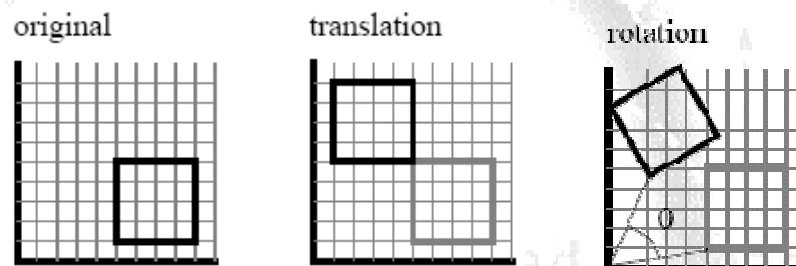
- a. the respective orientation of the ligand  
vs protein;**
- b. the bound conformation of the ligand.**





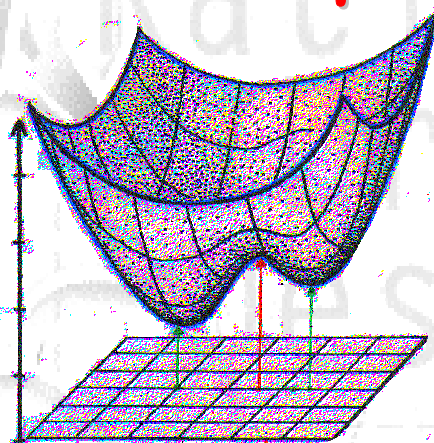
# What we need...

## 1. Translate and rotate a molecular object:



$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} d_x \\ d_y \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} \quad \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$$

## 2. Explore conformational space of a molecule:





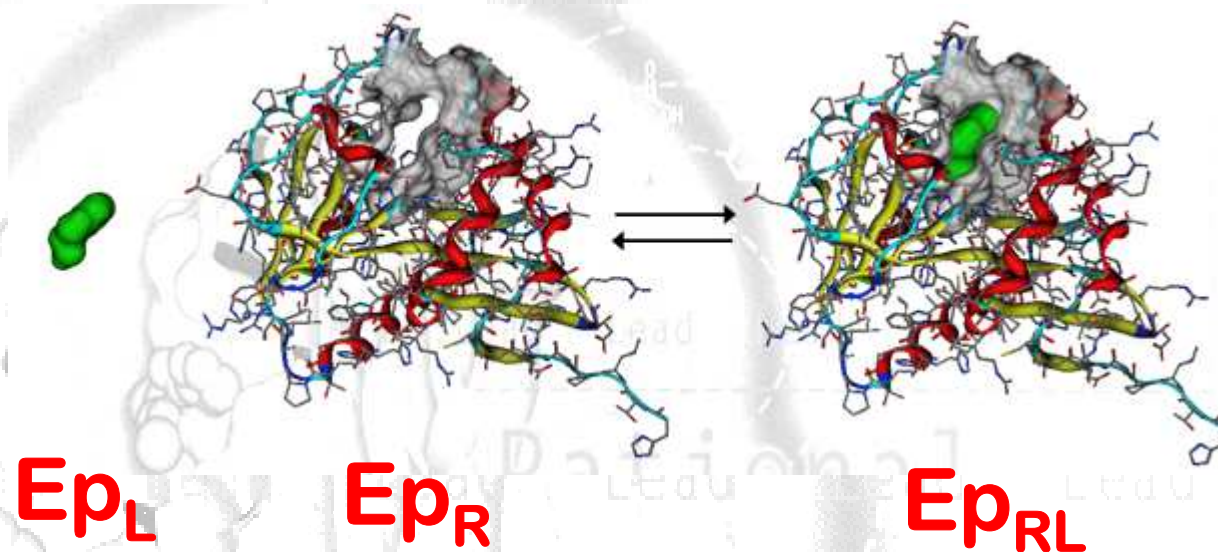
# What we need...

## 3. Assign the potential energy of each molecules:

$$E_p \cong \sum \left[ \frac{1}{2} k_{str} (r - r_e)^2 - D_e^{str} \right] + \sum \left[ \frac{1}{2} k_{ben} (\tau - \tau_e)^2 - D_e^{ben} \right] + \sum \left[ A (1 + \cos n\tau - \theta) \right] + \sum_{\text{nonbonded pairs}} \left( \frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^6} \right) + \sum_{\text{nonbonded pairs}} \frac{q_i q_k}{4\pi\epsilon_0 r_{ik}}$$

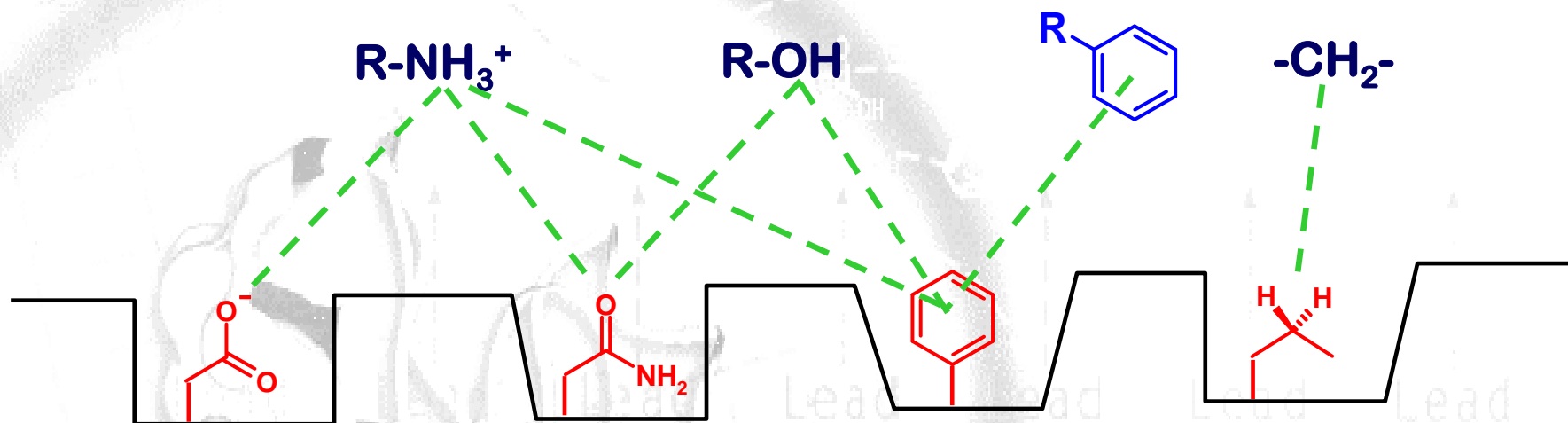
0.9760	0.5530	0.1180 C
2.1810	1.4240	0.1010 C
2.9110	1.3580	-1.2420 C
2.2460	1.6700	-2.0550 H
3.2180	0.3270	-1.4520 H
4.1000	2.2110	-1.2400 N
3.8190	3.1840	-1.1170 H
4.5460	2.1620	-2.1560 H
2.8610	1.1260	0.9100 H
1.8740	2.4560	0.3100 H
-0.1610	0.9170	-0.5670 N
-1.0220	-0.0510	-0.3510 C
-2.0310	-0.1010	-0.7380 H
-0.4920	-1.0270	0.4440 N
-0.9630	-1.8680	0.7470 H
0.7870	-0.6560	0.7530 C
1.4290	-1.2620	1.3760 H

# Force Field Based Scoring Functions



$$\Delta E_p = \sum_{\text{nonbonded pairs}} \left( \frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^6} \right) + \sum_{\text{nonbonded pairs}} \frac{q_i q_k}{4\pi\epsilon_0 r_{ik}}$$

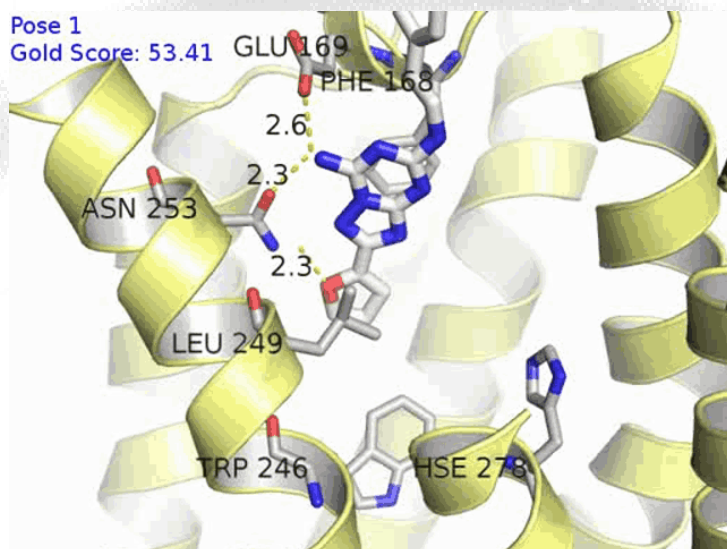
# Empirical Scoring Functions



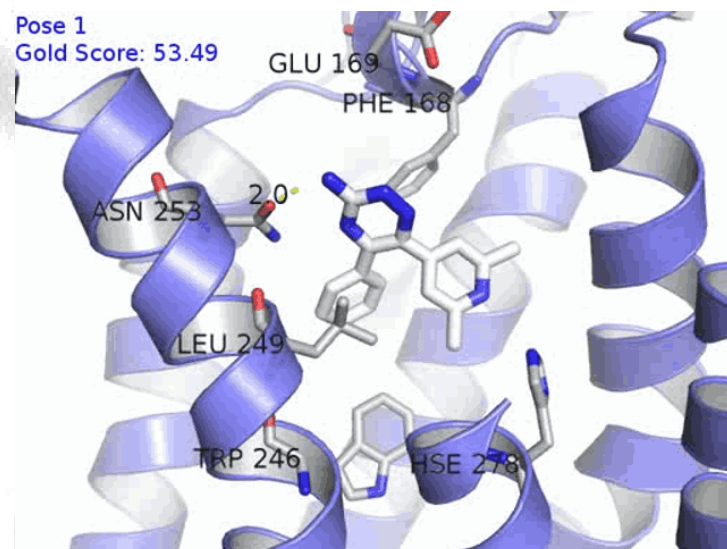
charge-charge interaction (*ionic bond*):  
 charge-dipole interaction:  
 charge- $\pi$  interaction:  
 hydrogen bond:  
 charge transfer interaction:  
 $\pi$ - $\pi$  interaction:  
 dipole-dipole interaction (van der Waals):

$-\Delta G^0 \cong$	5 ÷ 10
$-\Delta G^0 \cong$	1 ÷ 7
$-\Delta G^0 \cong$	8 ÷ 10
$-\Delta G^0 \cong$	1 ÷ 7
$-\Delta G^0 \cong$	1 ÷ 6
$-\Delta G^0 \cong$	1 ÷ 2
$-\Delta G^0 \cong$	0.5 ÷ 1

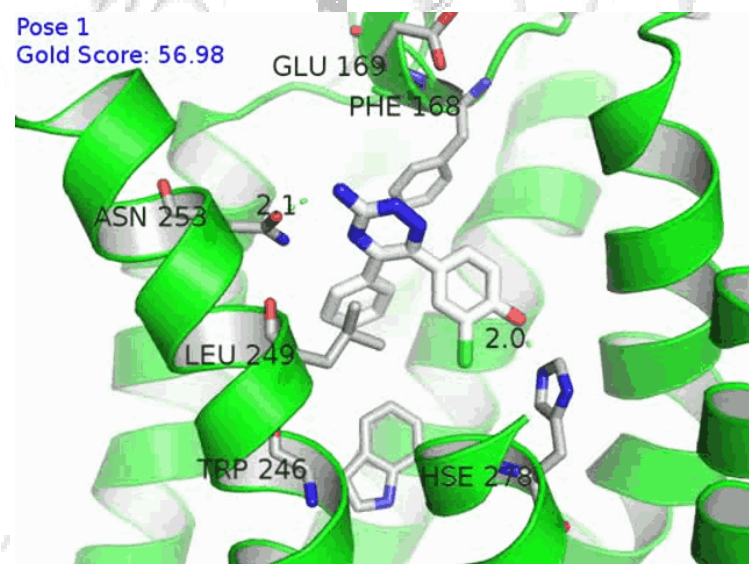
## Here is a real example:



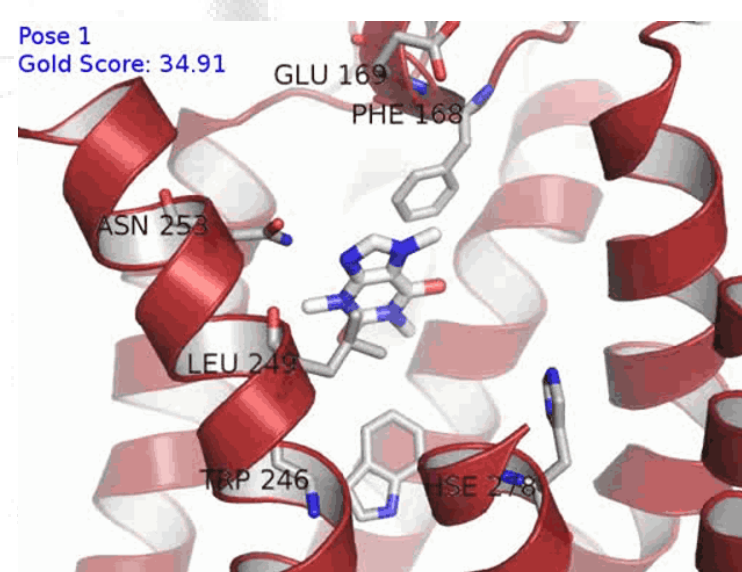
ZM



T4G



T4E



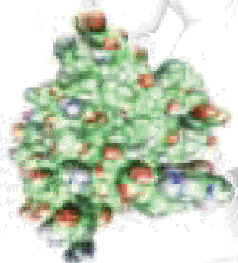
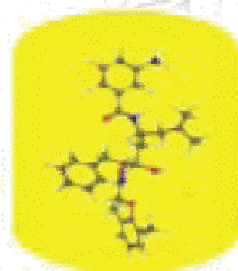
Caffeine



# Docking Applications:

- *High-Throughput Virtual Screening (HTVS);*

Chemical compounds  
ChemBridge ~ 500,000  
Drug like 500,000



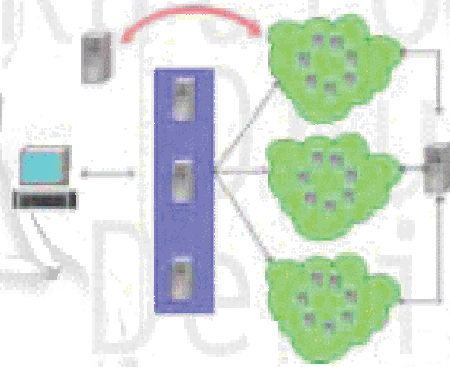
Targets  
Plasmeprin II (1lee, 1lf2,  
Plasmeprin IV (1ls5)



HTS Very expensive (1-10 \$ per compound, and nearly impossible)



~ 80 years of CPU time, 1 TB data



45 days on 1000 computers

Hits

Leads

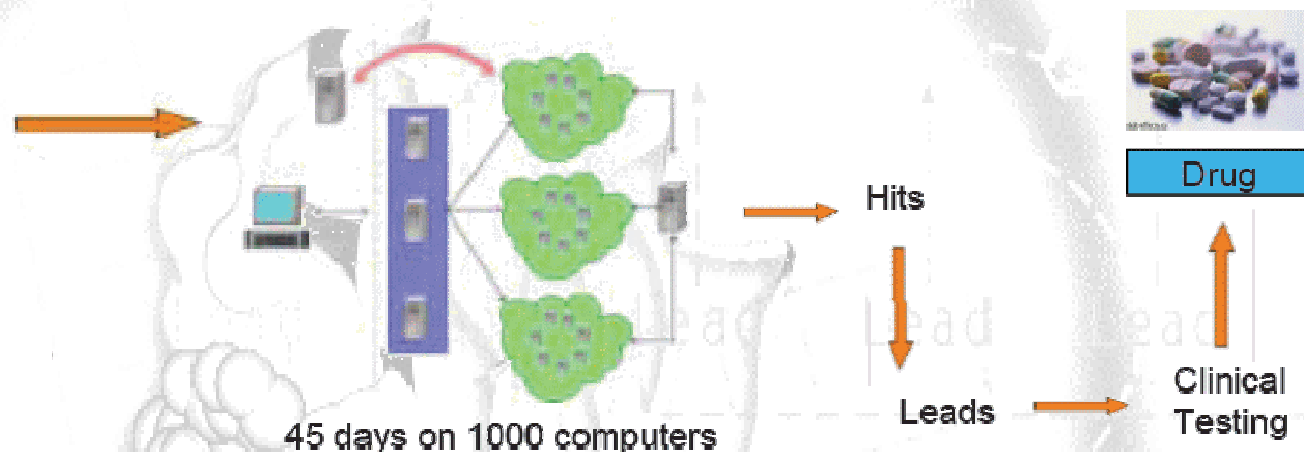


Drug

Clinical Testing

# Docking Applications:

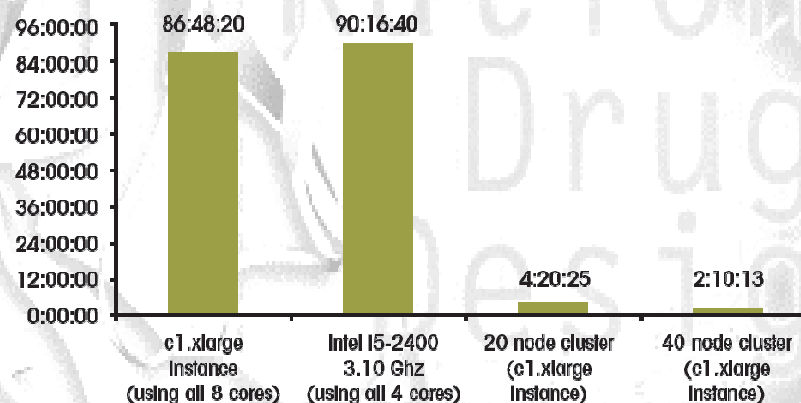
- *High-Throughput Virtual Screening (HTVS);*



45 days on 1000 computers

2005

Benchmarks for docking 100k structures



2013



# Beyond docking...

## SWOT analysis

### Molecular Docking

Fast and scalable pose sampling

Rigid protein

Difficult solvent treatment

Serious scoring problems

### Molecular Dynamics

Computational expensive

Investigate receptor full flexibility

Explicit solvent treatment

Accurate energy inspection

### Dock&MD

Fast and scalable pose sampling

Investigate receptor full flexibility

Explicit solvent treatment

Accurate energy inspection

Is it possible to discriminate the bioactive (XRAY) conformation from an ensemble of docked poses taking advantage of MD versatility?

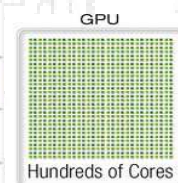
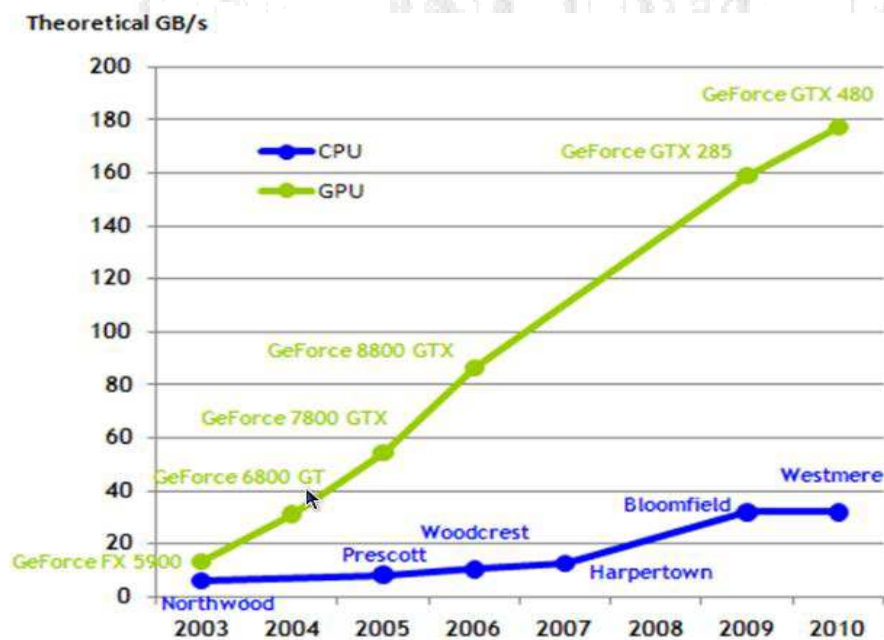
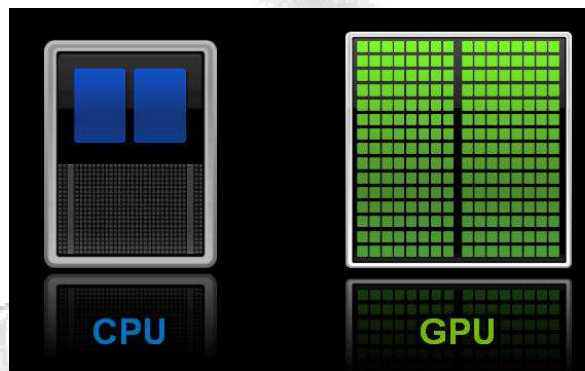


# Why we need *time* virtualization?

1. Several molecular properties are time-dependent
2. Conformational space is naturally explored following time coordinate
3. Any recognition process is time-dependent
4. Dynamics controls equilibrium position
5. ...

## Adenosiland – Bridging docking with MD

### My favorite C/G mutation



in collaboration with: 



# Bridging Docking and Molecular Dynamics

Molecular Docking

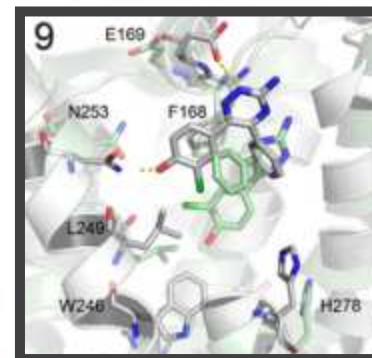
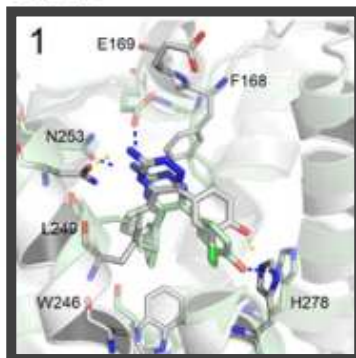


Pose 1

3UZC

.....

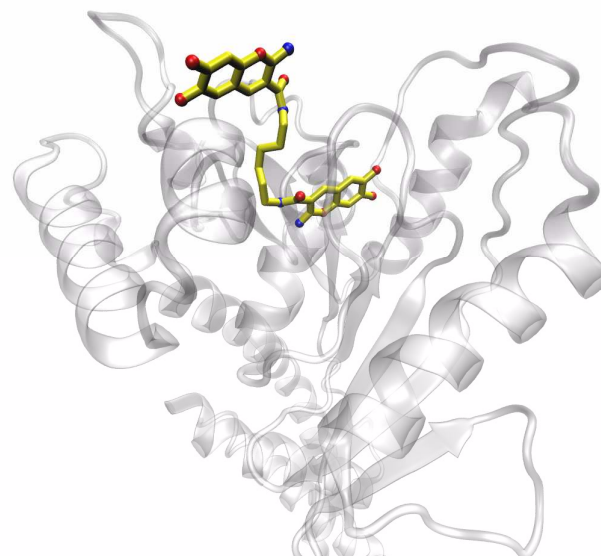
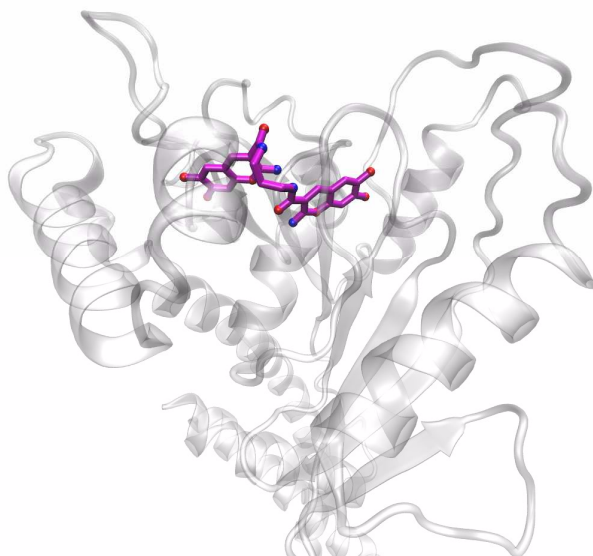
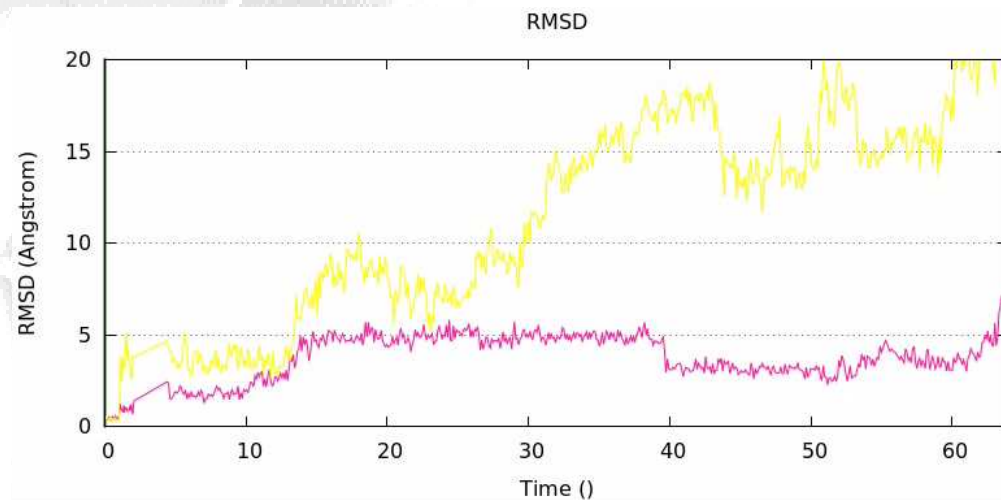
Pose n



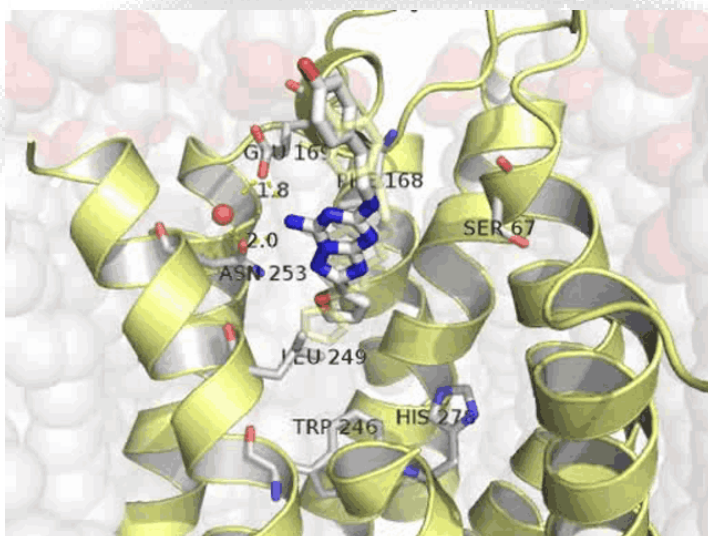
MD

MD

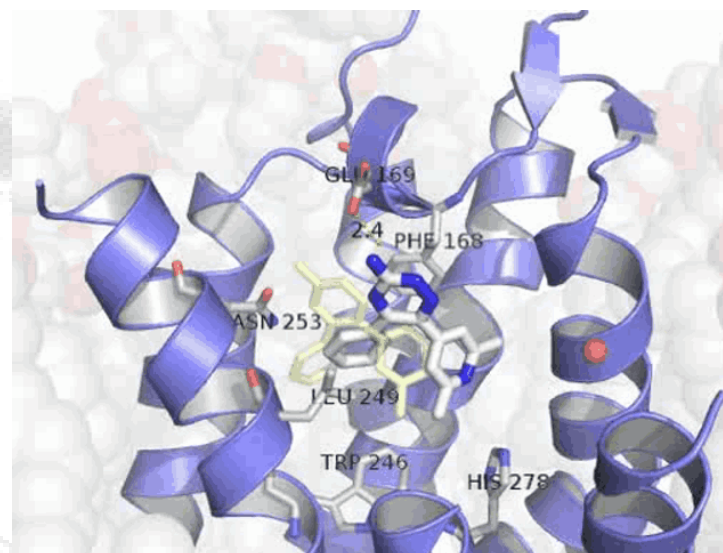
# Bridging Docking and Molecular Dynamics



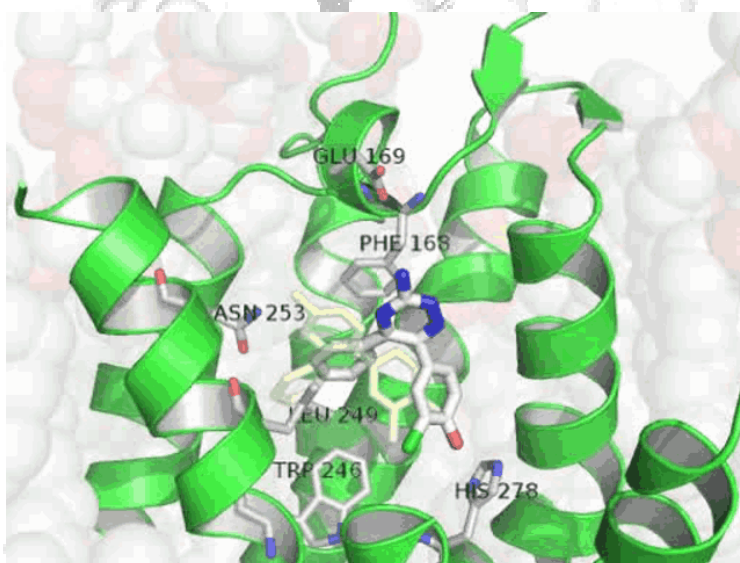
## Bridging Docking and Molecular Dynamics - A<sub>2A</sub> adenosine receptor



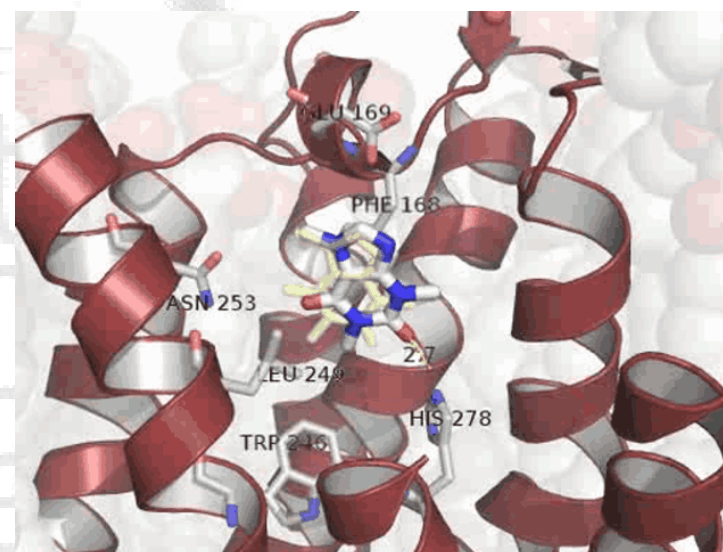
ZM



T4G



T4E

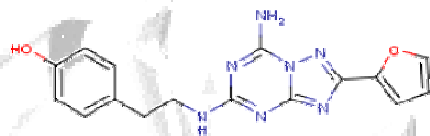


Caffeine

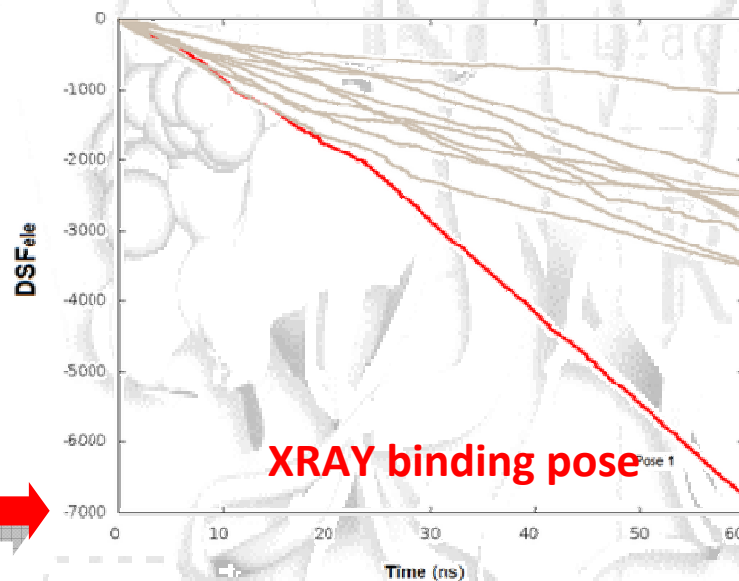
Sabbadin D., Ciancetta A., Moro S et al. J Chem Inf Mod (2014) in press

# Bridging Docking and Molecular Dynamics - A<sub>2A</sub> adenosine receptor

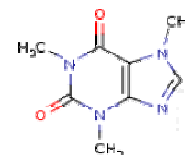
Comparison between a strong and a weak binders



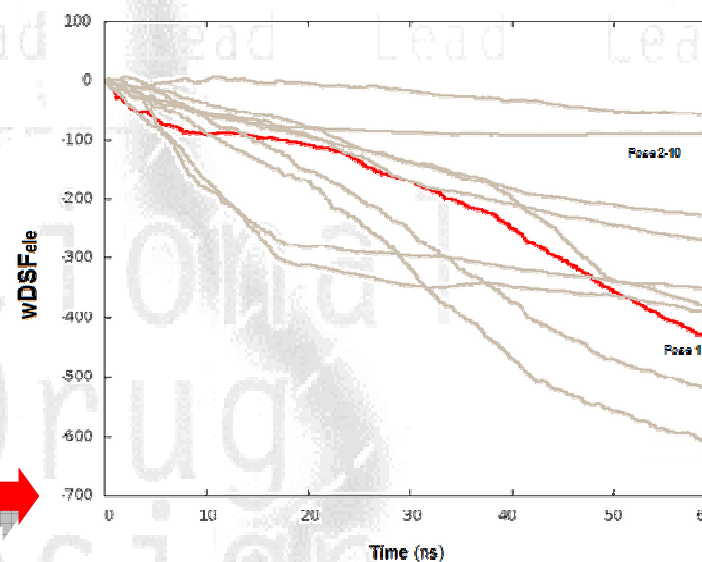
$K_i = 1.6 \text{ nM (h\_A}_{2A}\text{)}$



**XRAY binding pose**



$K_i = 23400 \text{ nM (h\_A}_{2A}\text{)}$

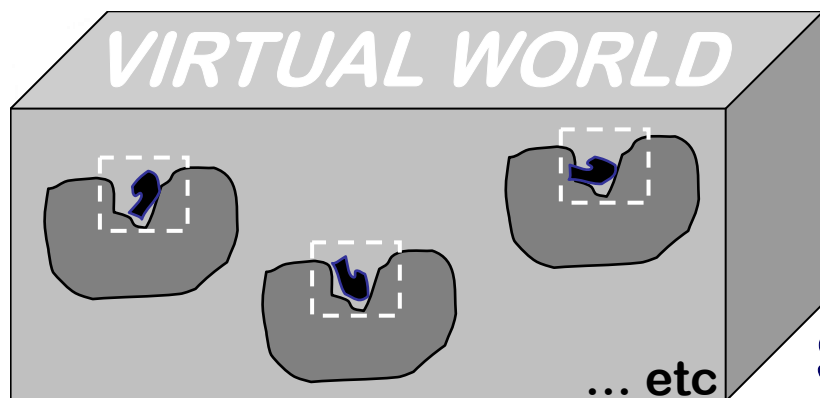




**GRAZIE  
PER LA PAZIENZA**

*Stefano Moro*





## Some definitions:

**RIGID**

**SEMI-FLEXIBLE**

**FLEXIBLE**

**Translations (3N)  
Rotations (3N)**

**Conf.  
LIG**

**Conf.  
REC**



**side-chains  
close to LIG**