



EMMSB

2014

LNCC-Petrópolis, 18 - 22 de Agosto de



# Triagem Virtual em Larga Escala

Profa. Dra. Rafaela Ferreira

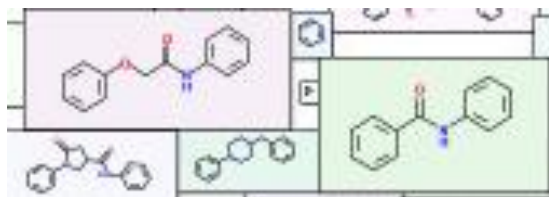
Dept. de Bioquímica e Imunologia

rafaelasf@gmail.com

21 de agosto de 2014



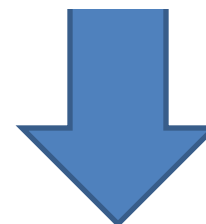
# O que é triagem virtual?



**Grande conjunto de compostos**



**Estratégia  
computacional**



**Compostos Seleccionados**



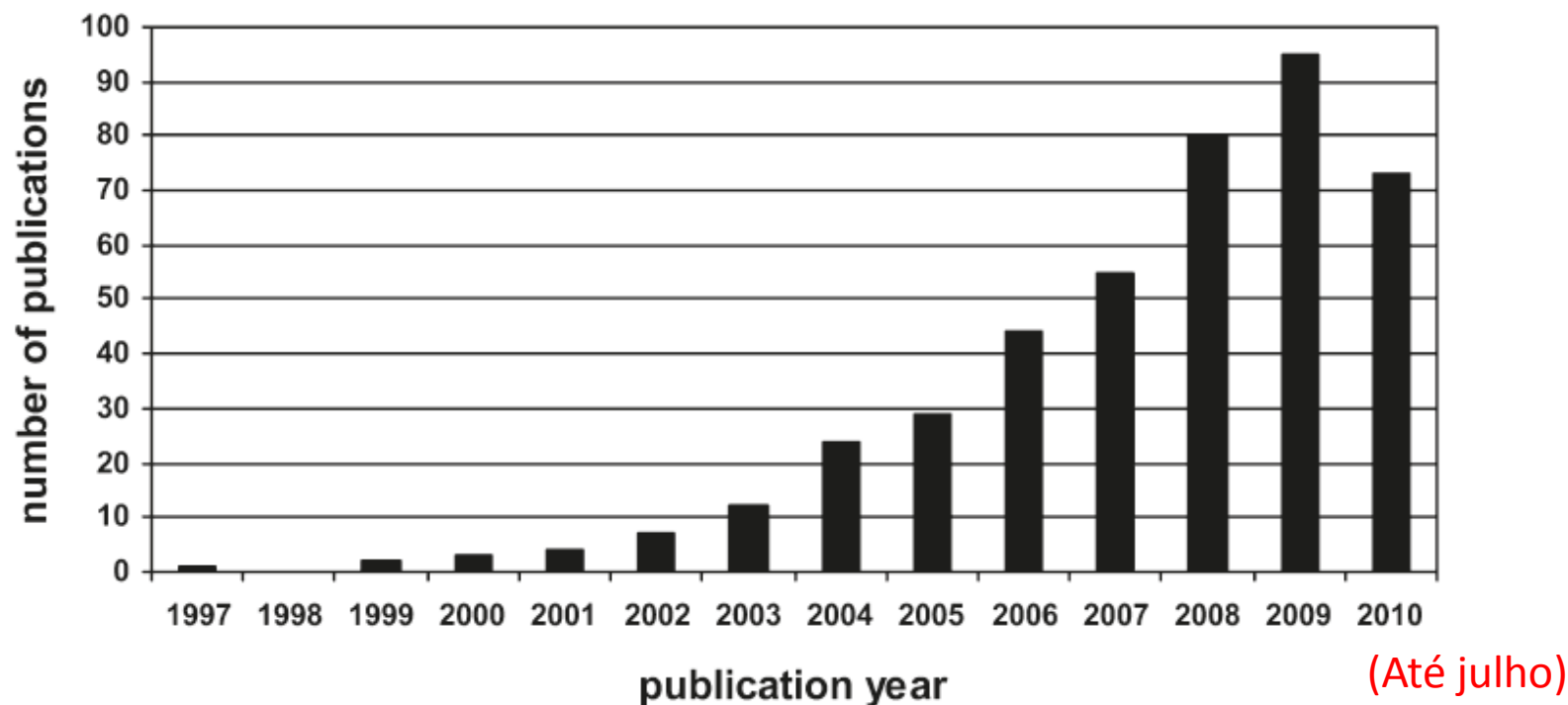
**Validação experimental**

## Geralmente apenas uma pequena porcentagem do conjunto de compostos é avaliada

|                        | Número<br>de artigos |              | Número<br>de artigos |
|------------------------|----------------------|--------------|----------------------|
| screening library size |                      | comps tested |                      |
| <1000                  | 16                   | 1–10         | 50                   |
| 1000–10000             | 30                   | 10–50        | 161                  |
| 10001–100000           | 89                   | 50–100       | 71                   |
| 100001–1000000         | 169                  | 100–500      | 95                   |
| 1000001–10000000       | 78                   | 500–1000     | 13                   |
| >10000001              | 13                   | ≥1000        | 16                   |
| not reported           | 26                   | not reported | 15                   |

**50 em 1.000.000 = 0.005%**

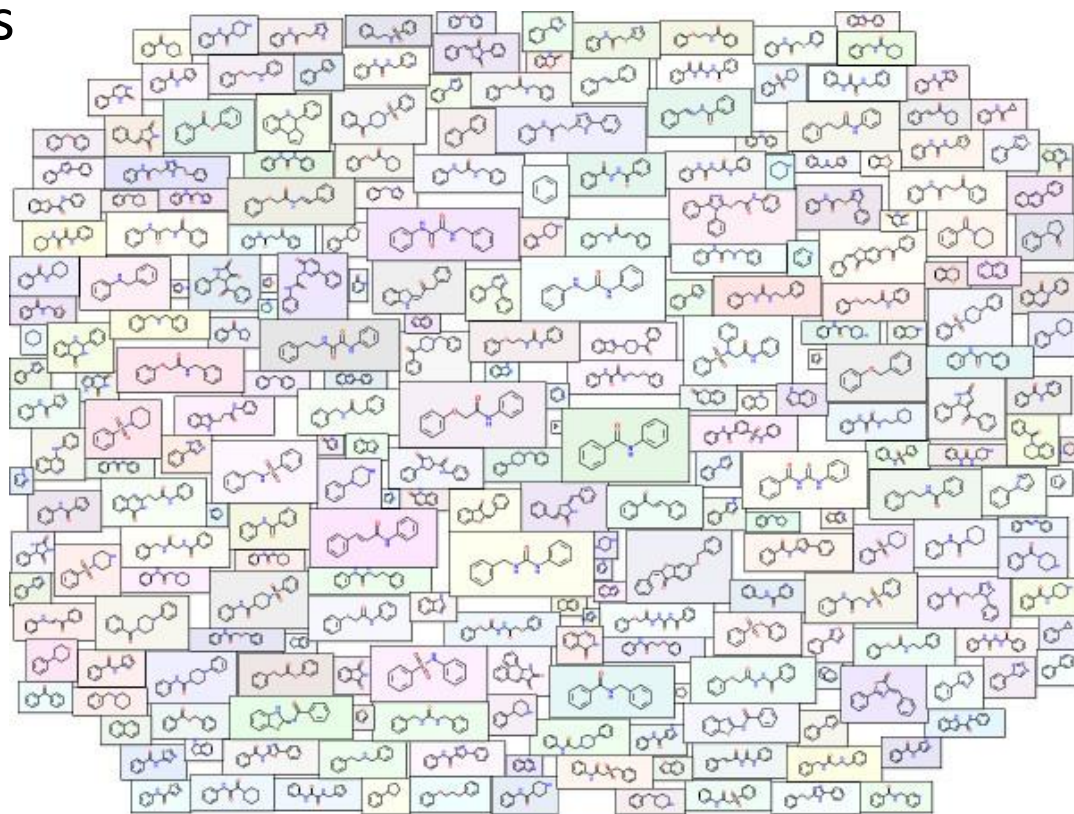
## O emprego de triagem virtual tem aumentado nos últimos anos



**Figure 1.** Prospective virtual screening applications. The total number of prospective VS publications that have appeared between 1997 and 2010 (through July 15th) in the 12 surveyed journals is reported.

# O espaço químico é muito amplo

- Espaço químico = conjunto de todas as estruturas químicas possíveis
- Estima-se cerca de  $10^{60}$  pequenas moléculas
  - GD-13 – 977 milhões - 99,9% nunca foi sintetizado
  - GB-17 – 166 bilhões



# Milhões de compostos são comercialmente disponíveis

- ZINC – maior base de dados de compostos comercialmente disponíveis (zinc.docking.org)
  - Mais de 35 milhões de compostos (agosto de 2014)

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

Shoichet Laboratorydocking.org

**ZINC**<sup>12</sup>

Not Authenticated – sign in

Active cart: Temporary Cart (0 items)

AboutSearchSubsetsHelpSocial+149

Quick Search Bar...Go

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 21 million purchasable compounds in ready-to-dock, 3D formats. ZINC is provided by the [Shoichet Laboratory](#) in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). To cite ZINC, please reference: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model.* 2012 DOI: [10.1021/ci3001277](#). The original publication is Irwin and Shoichet, *J. Chem. Inf. Model.* 2005;45(1):177-82 [PDF](#), [DOI](#). We thank [NIGMS](#) for financial support (GM71896).

ZINC ID, Drug Name, SMILES, Catalog, Vendor Code, TargGo

Structure/DrawPhysical PropertiesCatalogs & VendorsZINC IDSTargetsRingsCombination

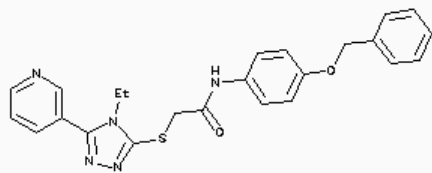
[What's NEW?](#) [Feedback](#) [Like us](#)  
[@chem4biology](#) [Blog](#) [RSS](#)  
[Video Walkthroughs](#)

 **ZINC Database**  
 278

**Quick Links**  
[Download](#)  
[Target focused](#)  
[Natural](#)  
[Products](#)  
[PBCs](#)  
[Carts](#)

**Your Carts**  
[Create an account](#) or [login](#) to have multiple carts.

**Molecule of the Day** [2375784](#)





# Bases de dados podem ser filtradas de acordo com propriedades físico-químicas

Relação com biodisponibilidade



| Propriedade                | Similar a<br>Fármaco<br>( <i>drug-like</i> ) | Similar a<br>protótipo<br>( <i>lead-like</i> ) | Fragmentos<br>( <i>fragment-like</i> ) |
|----------------------------|----------------------------------------------|------------------------------------------------|----------------------------------------|
| Massa molecular (g/mol)    | < 500                                        | < 350                                          | < 250                                  |
| cLogP                      | < 5                                          | 3,5                                            | 2,5                                    |
| No. doadores de lig. de H  | < 5                                          | < 4                                            | < 2                                    |
| No. aceptores de lig. de H | < 10                                         | < 8                                            | < 4                                    |



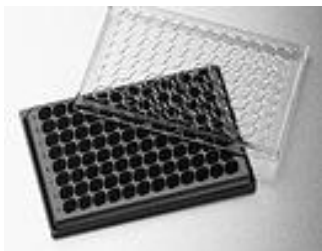
Permite otimização

[Lipinski, J Pharmacol Toxicol Methods. 2000 Jul-Aug;44\(1\):235-49.](#)  
[Teague, Davis, Leeson, Oprea, Angew Chem Int Ed Engl. 1999 Dec 16;38\(24\):3743-3748.](#)  
[Carr RA, Congreve M, Murray CW, Rees DC, Drug Discov Today. 2005 Jul 15;10\(14\):987](#)

# Abordagens para descoberta de ligantes para um alvo

## Experimental (HTS)

Todos compostos são experimentalmente avaliados



**Alto custo**

**Baixa % de ligantes**

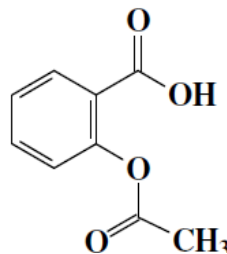
**Extensivo trabalho de validação**

## Triagem virtual (vHTS)

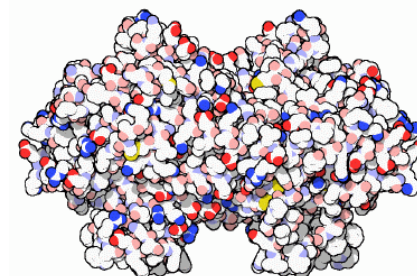
Apenas o conjunto selecionado de compostos é experimentalmente avaliado



Baseada na estrutura do ligante (LBVS)



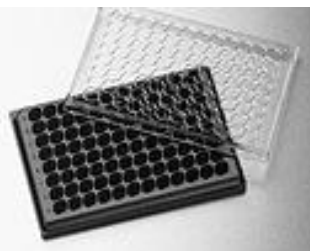
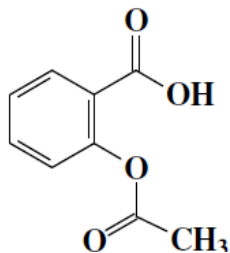
Baseada na estrutura do alvo (SBVS)



# Que tipo de informação podemos empregar para construir estratégias de triagem virtual?

## LBVS

Atividade biológica de pequenas moléculas

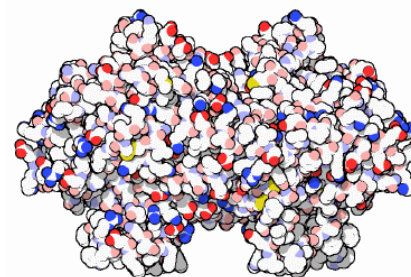


## SBVS

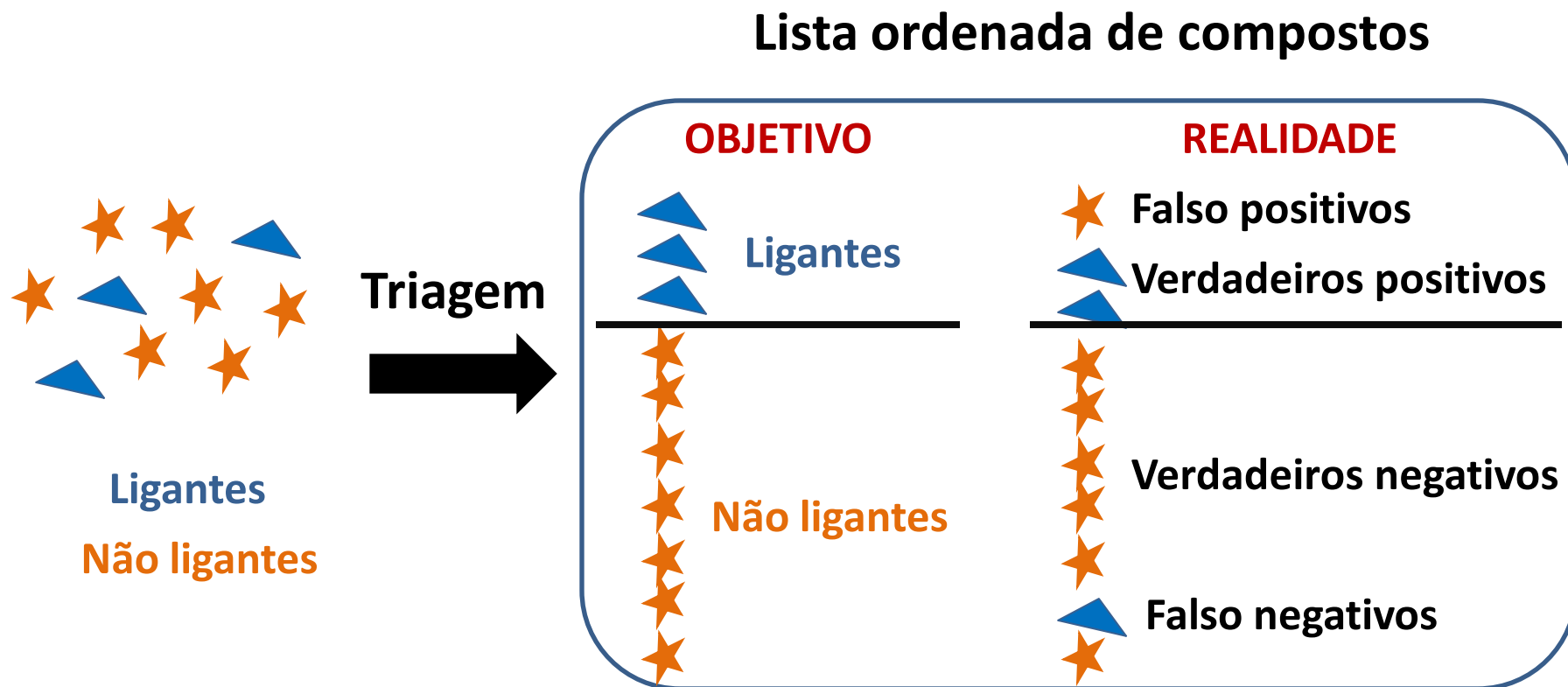
Estrutura do alvo



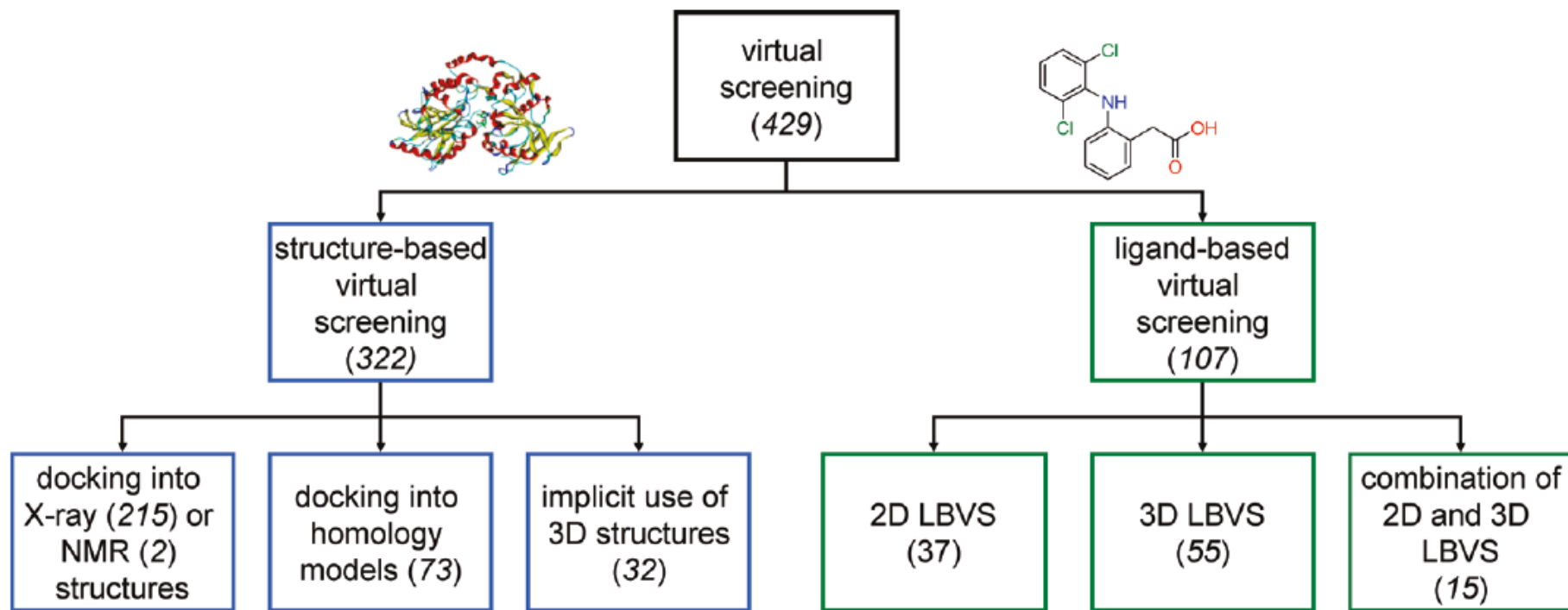
102.500 estruturas (12/08/14)



# O objetivo na triagem virtual é conseguir classificar compostos ligantes melhor do que não ligantes

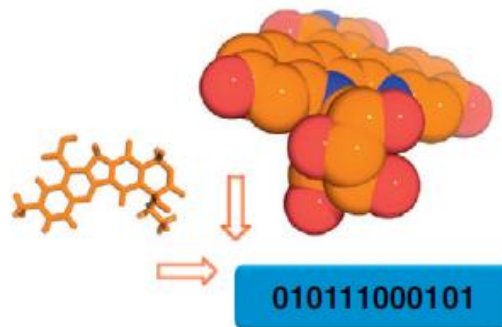


# A triagem virtual por docking é a estratégia mais comumente empregada



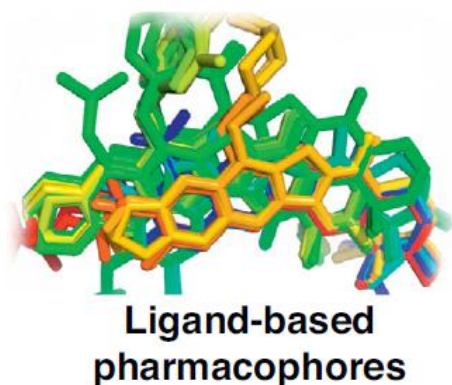
# Triagem virtual baseada na estrutura de ligantes

## Similaridade química

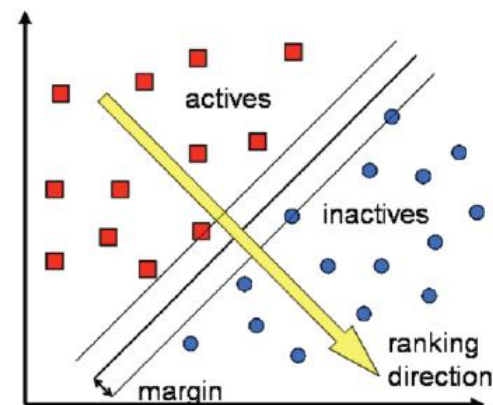


2D and 3D similarity

## Construção de farmacóforo



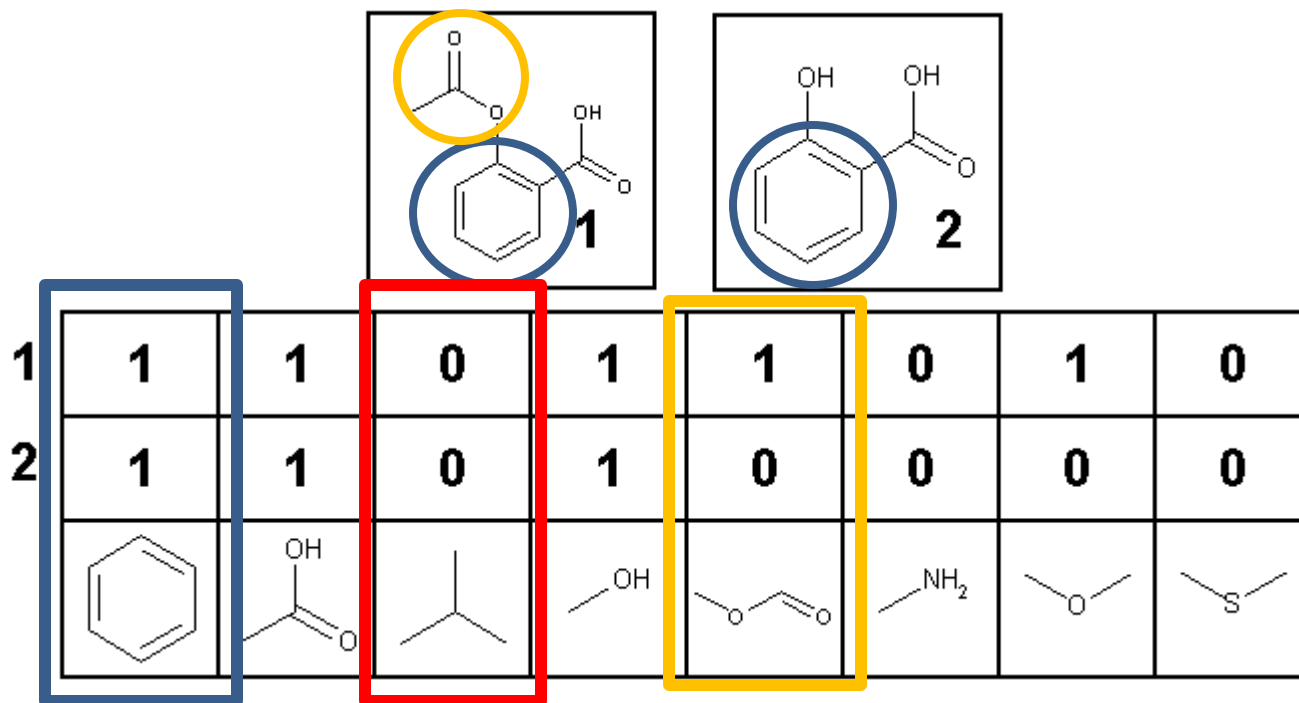
## Aprendizagem de máquinas



# Triagem virtual baseada na estrutura de ligantes

## 1. Comparação de similaridade 2D

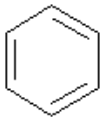
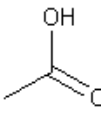
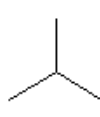
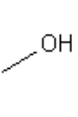
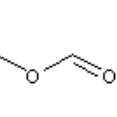
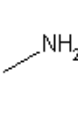
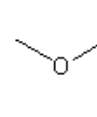
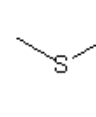
- 1.1 Determinação de digitais moleculares (*molecular fingerprints*)
  - Baseado na conectividade química



# Triagem virtual baseada na estrutura de ligantes

## 1. Comparação de similaridade 2D

- 1.2 Cálculo da similaridade através de uma métrica (ex: coeficiente de Tanimoto)

|   |                                                                                   |                                                                                   |                                                                                   |                                                                                   |                                                                                   |                                                                                    |                                                                                     |                                                                                     |
|---|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1 | 1                                                                                 | 1                                                                                 | 0                                                                                 | 1                                                                                 | 1                                                                                 | 0                                                                                  | 1                                                                                   | 0                                                                                   |
| 2 | 1                                                                                 | 1                                                                                 | 0                                                                                 | 1                                                                                 | 0                                                                                 | 0                                                                                  | 0                                                                                   | 0                                                                                   |
|   |  |  |  |  |  |  |  |  |

A = número de fragmentos em ambos conjuntos = 3

B = número de fragmentos em (1), mas não em (2) = 2

C = número de fragmentos em (2), mas não em (1) = 0

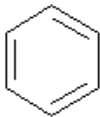
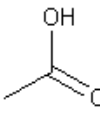
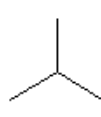
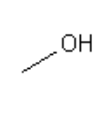
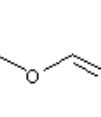
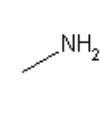
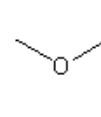
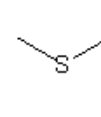
**Coeficiente de Tanimoto =  $A/(A+B+C)$**

$$= 3/(3 + 2 + 0) = 0.6 \text{ ou } 60\%$$

# Triagem virtual baseada na estrutura de ligantes

## 1. Comparação de similaridade 2D

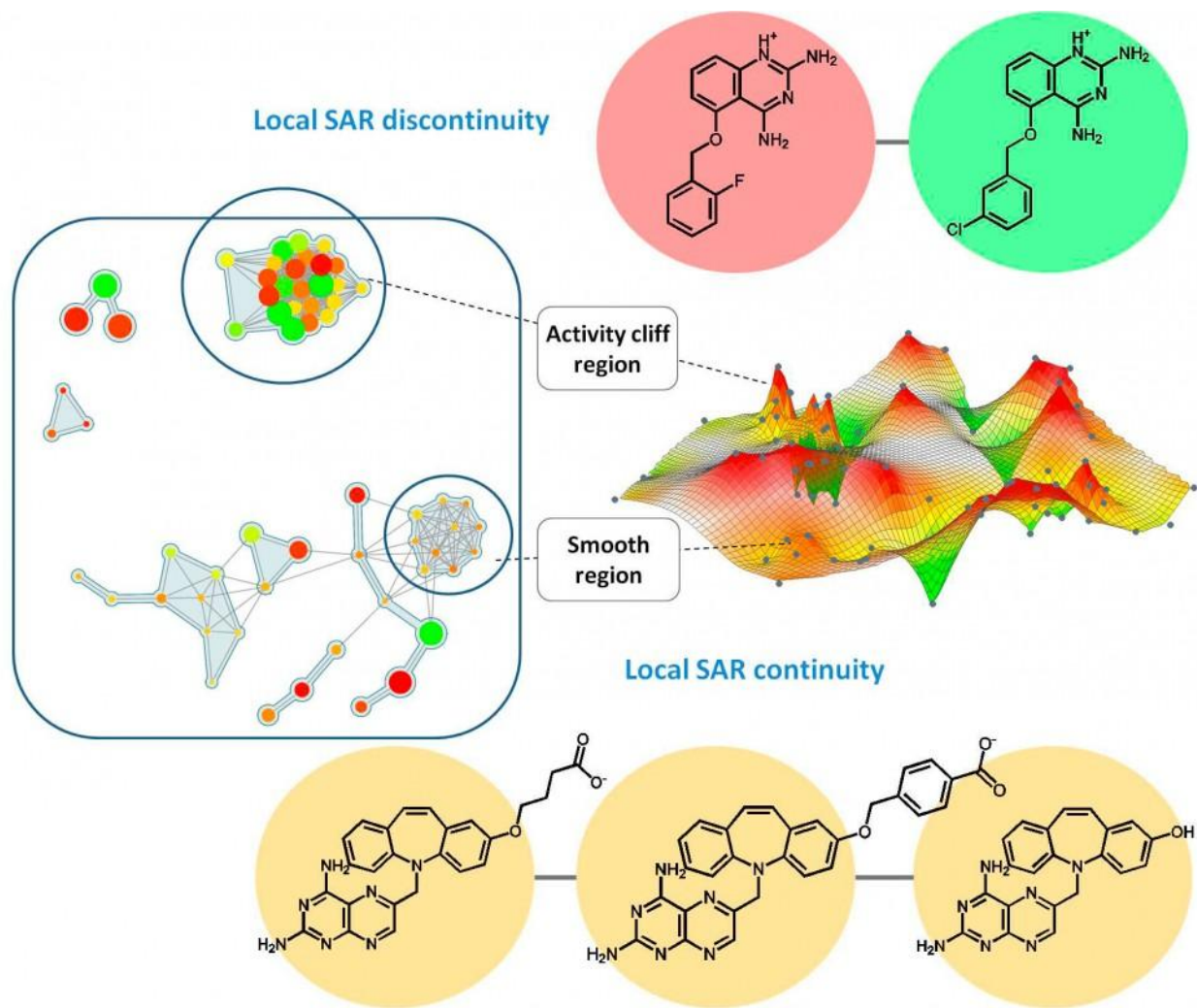
- 1.3 Triagem da base de dados por comparação com ligantes conhecidos

|   |                                                                                   |                                                                                   |                                                                                   |                                                                                   |                                                                                    |                                                                                     |                                                                                     |                                                                                     |
|---|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1 | 1                                                                                 | 1                                                                                 | 0                                                                                 | 1                                                                                 | 1                                                                                  | 0                                                                                   | 1                                                                                   | 0                                                                                   |
| 2 | 1                                                                                 | 1                                                                                 | 0                                                                                 | 1                                                                                 | 0                                                                                  | 0                                                                                   | 0                                                                                   | 0                                                                                   |
|   |  |  |  |  |  |  |  |  |

Definição de um valor do coeficiente de Tanimoto como cutoff

**Se o valor for muito alto – apenas busca de análogos**

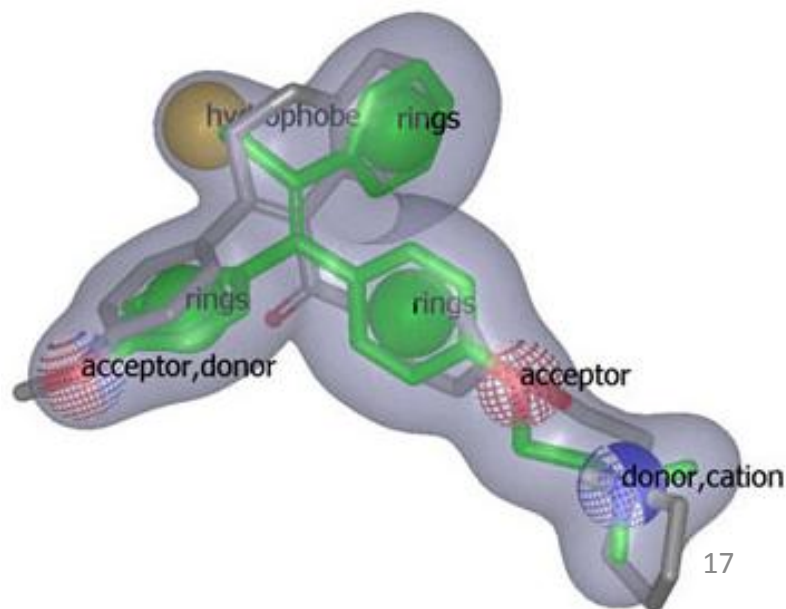
# Discontinuidades de relação estrutura-atividade são um desafio para métodos de similaridade 2D



# Triagem virtual baseada na estrutura de ligantes

## 2. Comparação de similaridade 3D

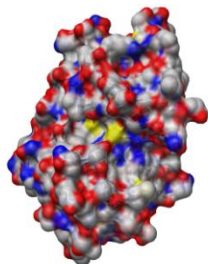
- 2.1 Cálculo de volume e de outras propriedades dos compostos ativos
- 2.2 Sobreposição de compostos ao padrão gerado
  - maximização de sobreposição (ex: algoritmo genético)



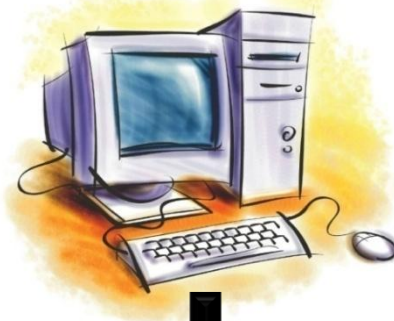
# Triagem virtual baseada na estrutura do receptor

## 1. Docking molecular

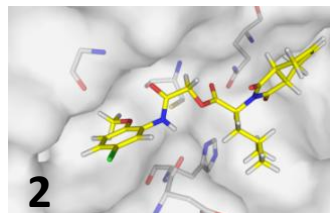
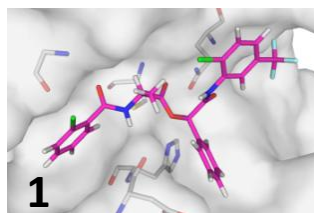
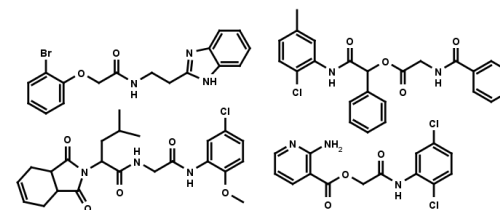
Estrutura do alvo



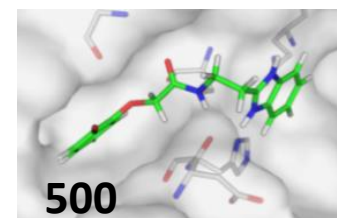
**DOCKING**



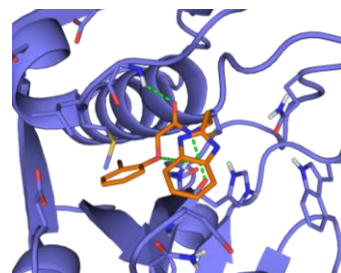
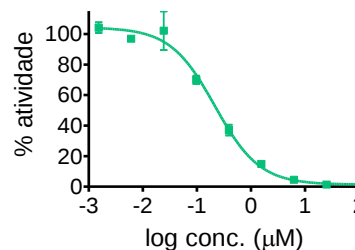
Base de dados  
 $10^3 - 10^6$  compostos



...



Inspeção visual e  
seleção de compostos

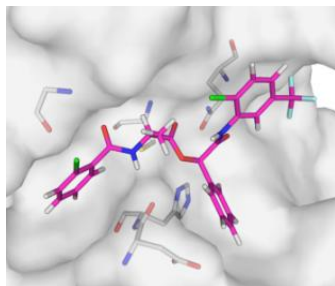


Validação experimental

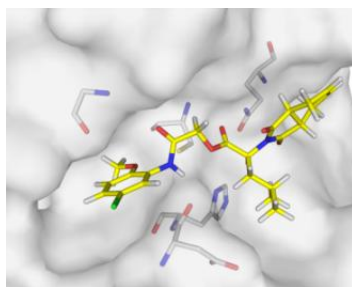
# Por que realizar uma seleção “manual” de compostos?

- Baixa precisão da função de pontuação
  - Presença de falsos positivos

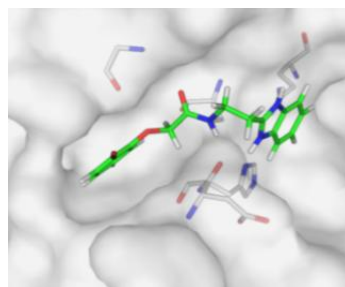
## CLASSIFICAÇÃO DA BASE DE DADOS



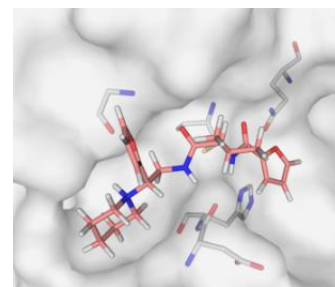
↓  
- 36 (1ª)



↓  
- 35 (2ª)



↓  
- 30 (3ª)



↓  
- 28 (4ª)

Pose melhor pontuada de cada molécula

Pontuação e classificação

- Possibilidade de incorporar conhecimento específico para uma classe de alvos

# Critérios para seleção de compostos

- **Complementaridade ao sítio**
- Facilidade de síntese de análogos

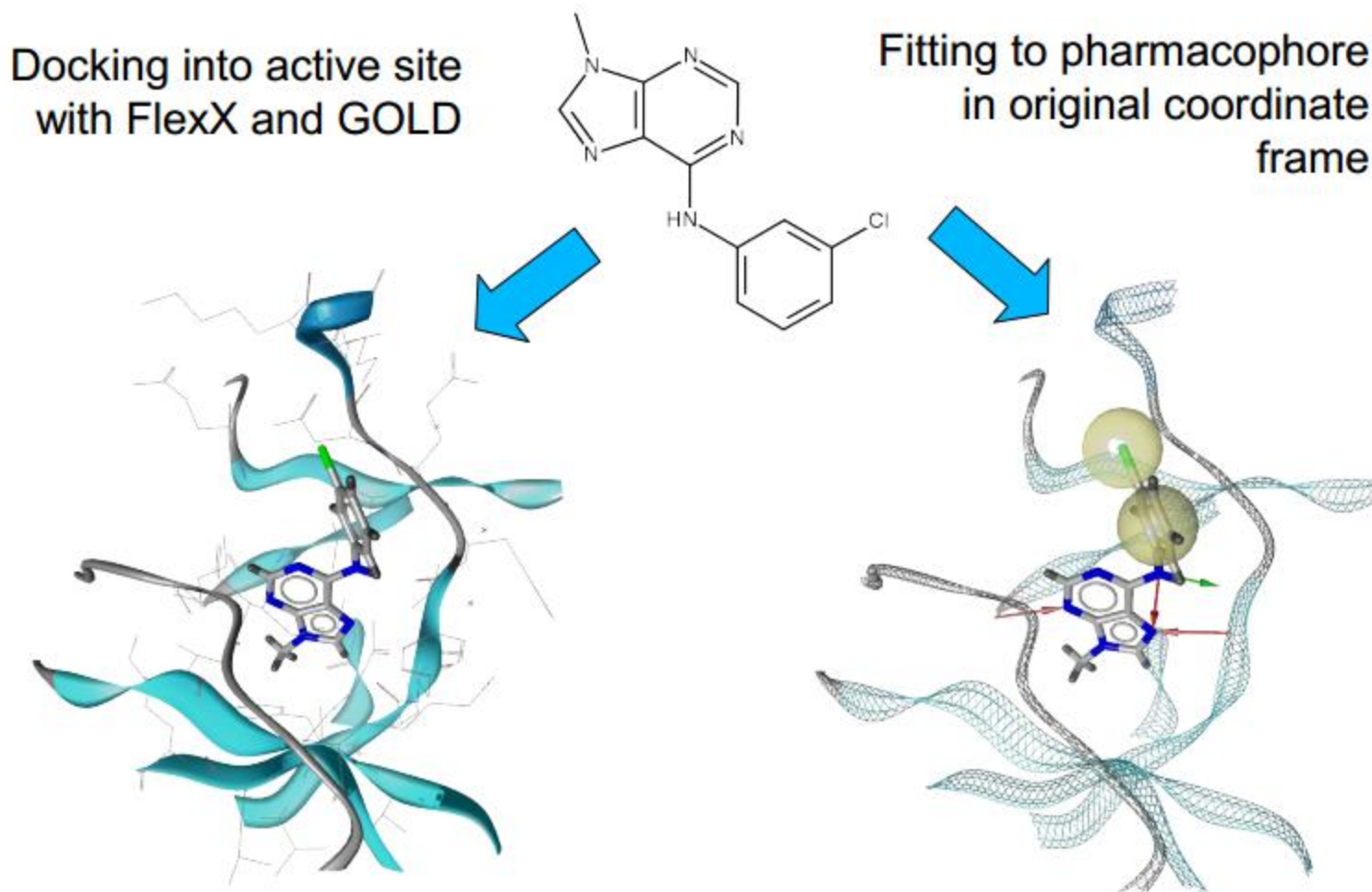
- **Diversidade química**

- Predição de ADMETox

Também podem ser  
aplicados antes da triagem

# Triagem virtual baseada na estrutura do receptor

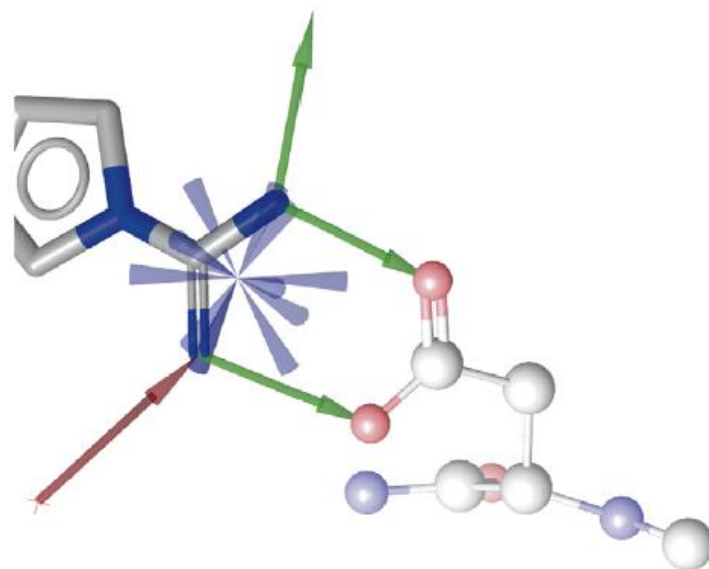
## 2. Determinação de farmacóforo



# Triagem virtual baseada na estrutura do receptor

## 2. Determinação de farmacóforo

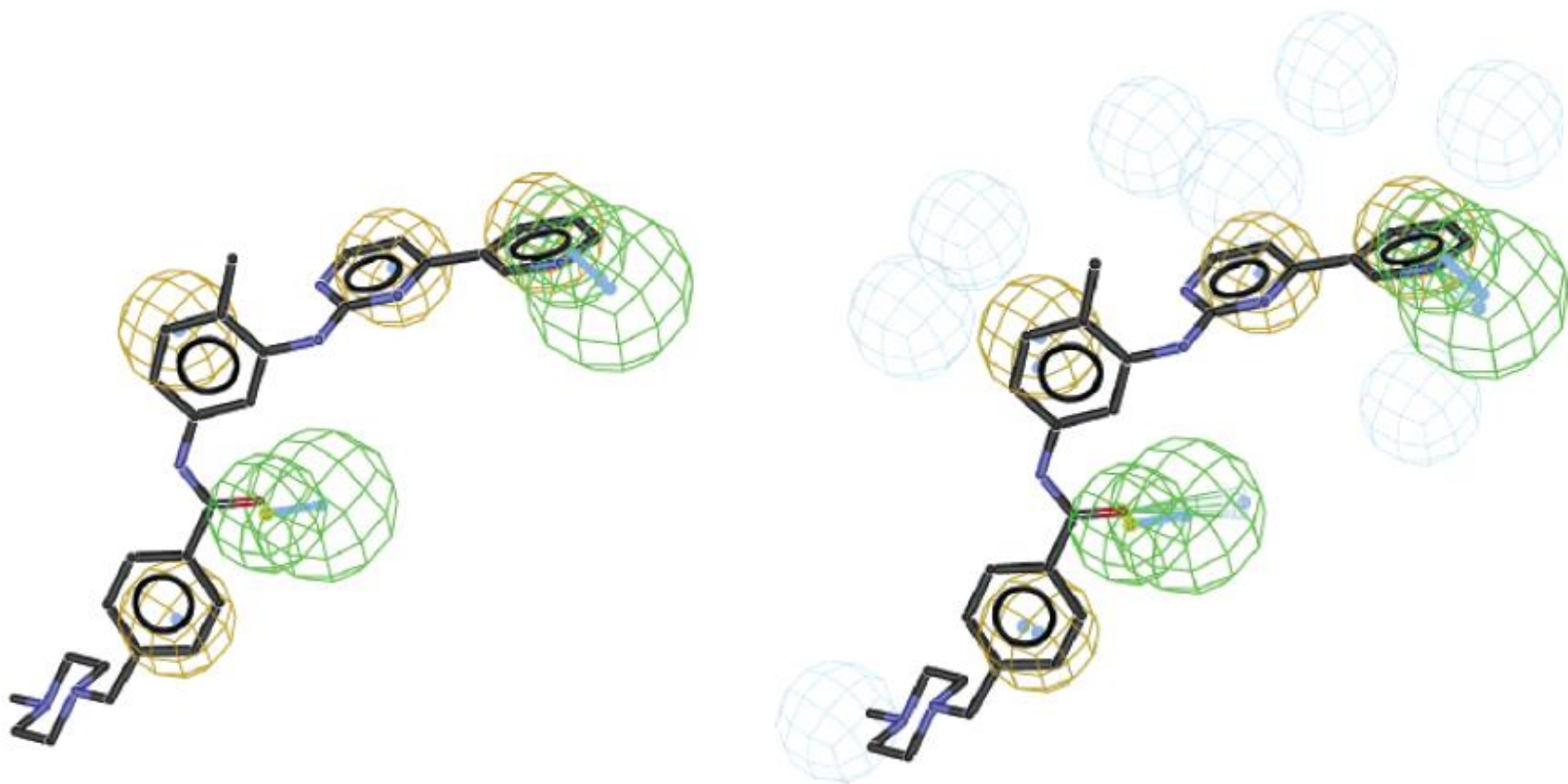
- Farmacóforo 3D
  - Conjunto de grupos funcionais essenciais para atividade, organizados espacialmente
- Farmacóforo extraído a partir de complexos
- Aceptores e doadores de ligação de hidrogênio – vetores
- Interações hidrofóbicas e eletrostáticas - esferas



*Drug Discovery Today*

# Triagem virtual baseada na estrutura do receptor

## 2. Determinação de farmacóforo



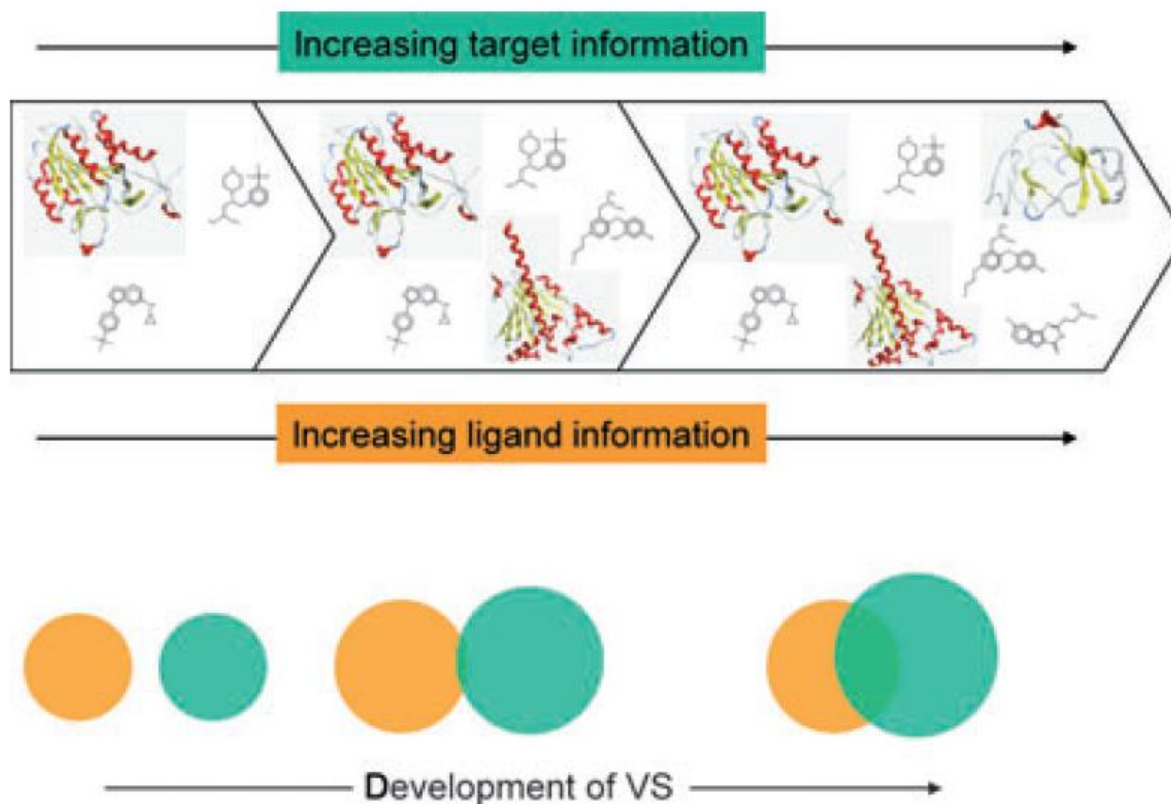
# Combinação de LBVS e SBVS

**ZINC**

ChEMBL

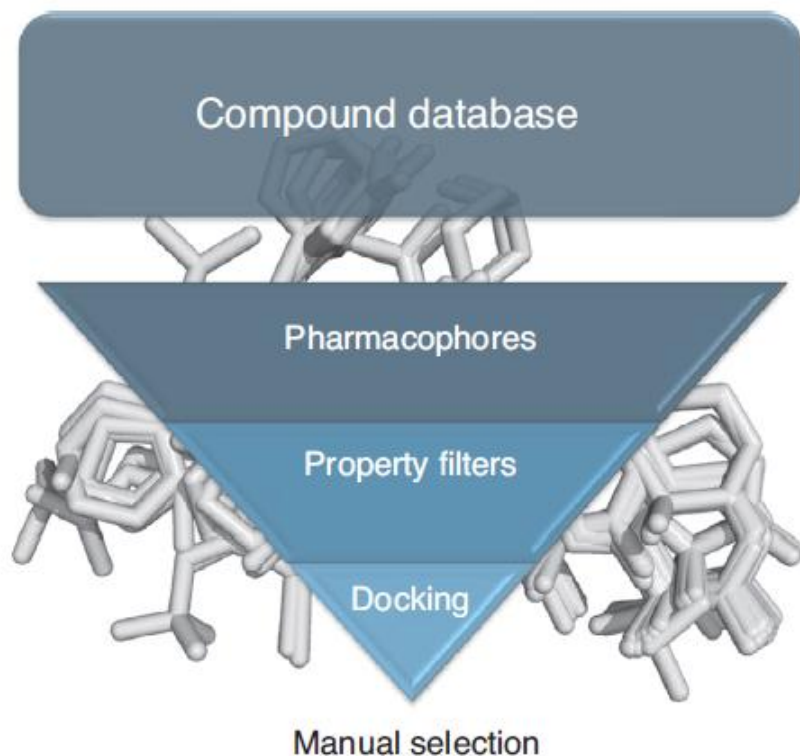


RCSB **PDB**  
PROTEIN DATA BANK

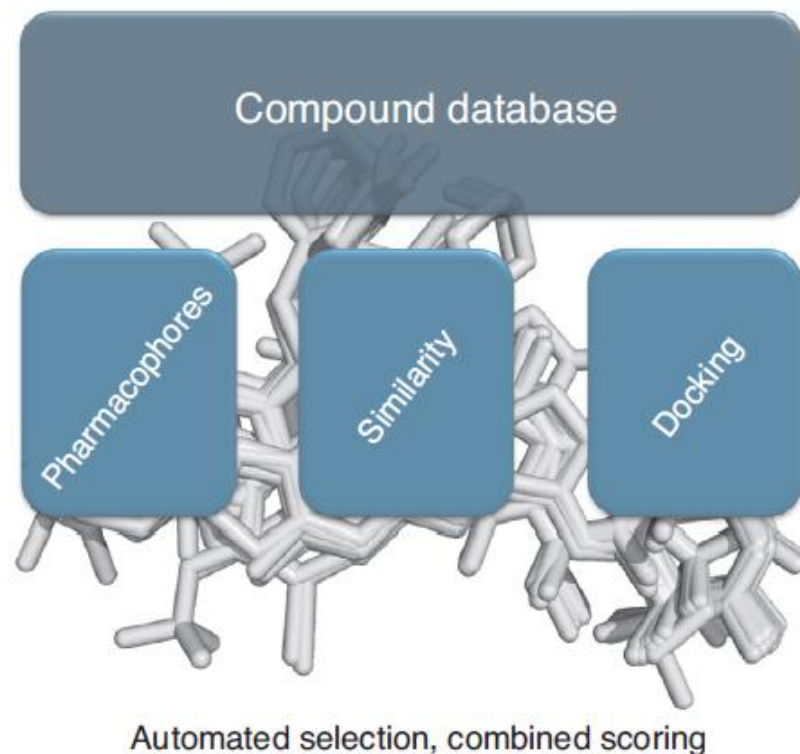


# Combinação de LBVS e SBVS

## Aplicação sequencial

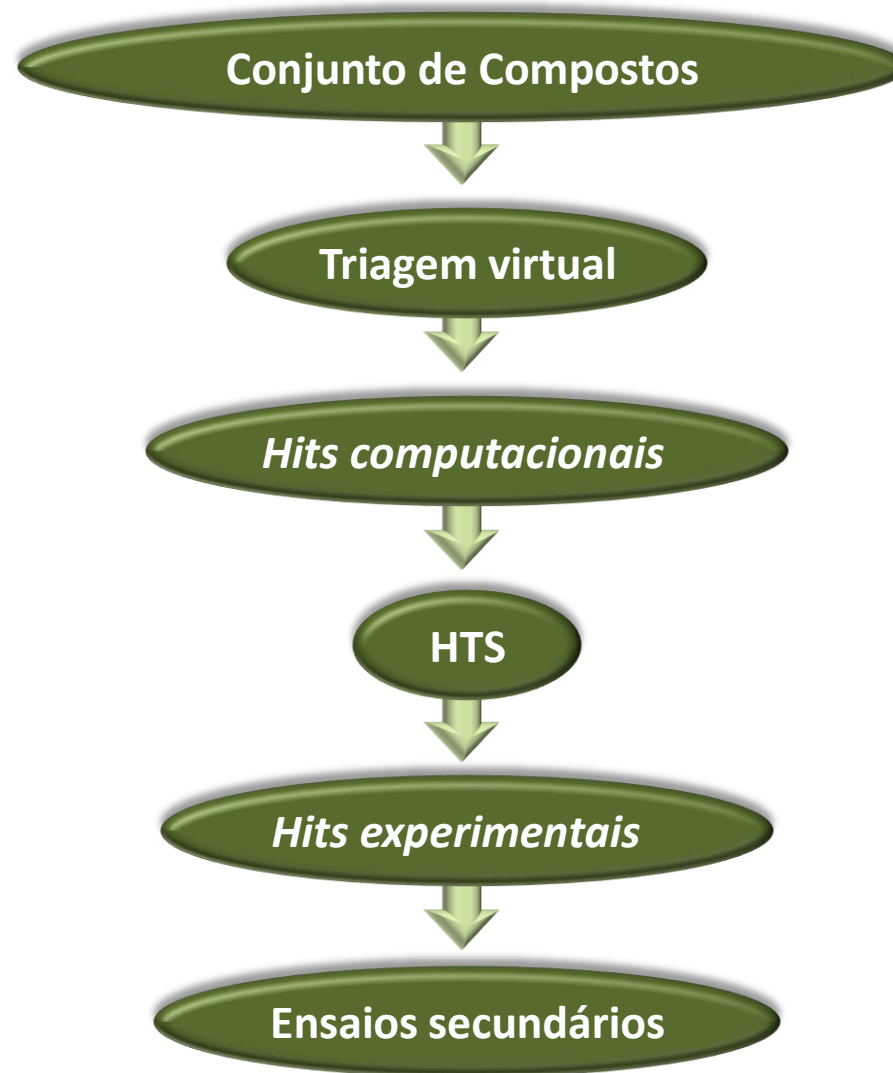


## Aplicação em paralelo

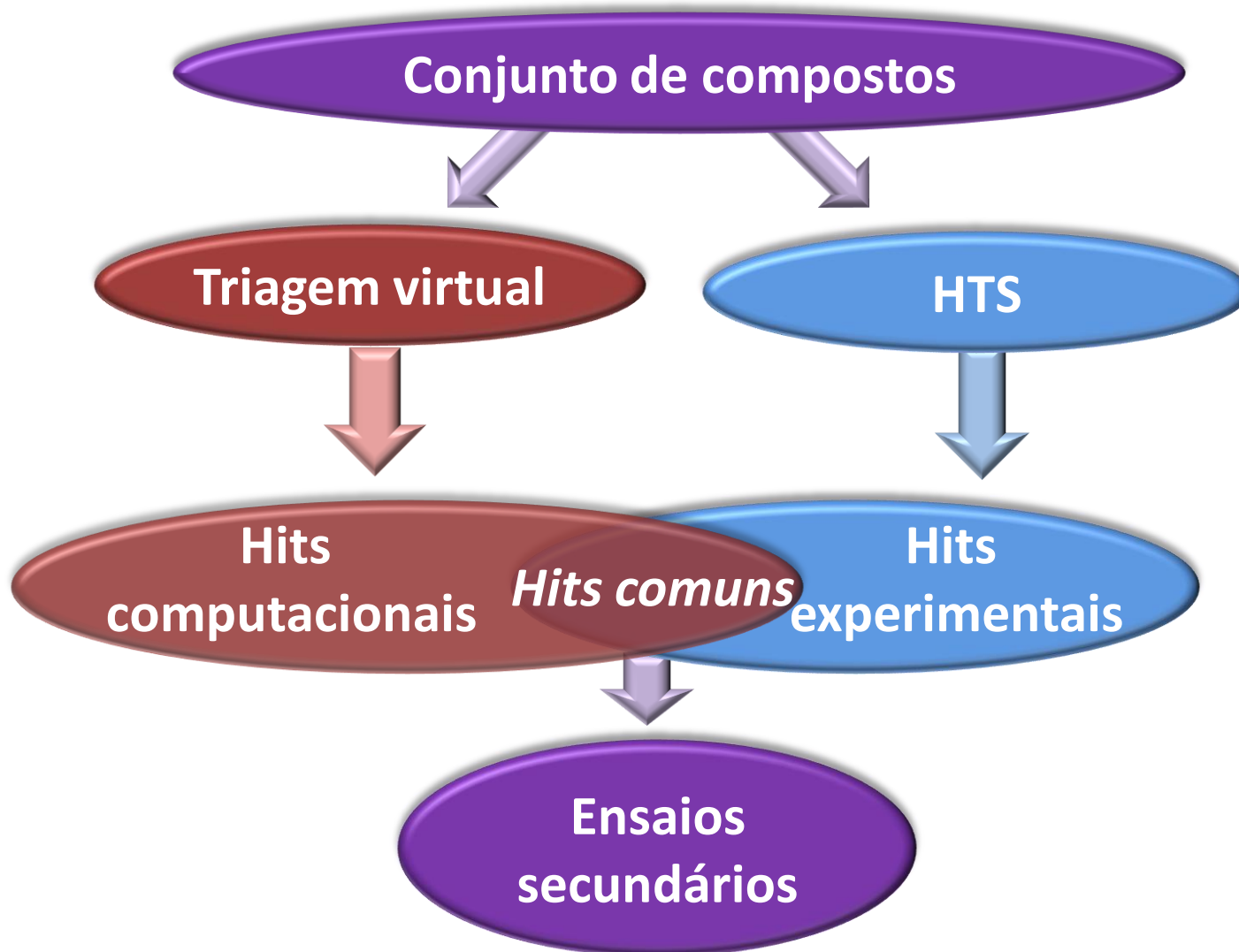


*Drug Discovery Today: Technologies*

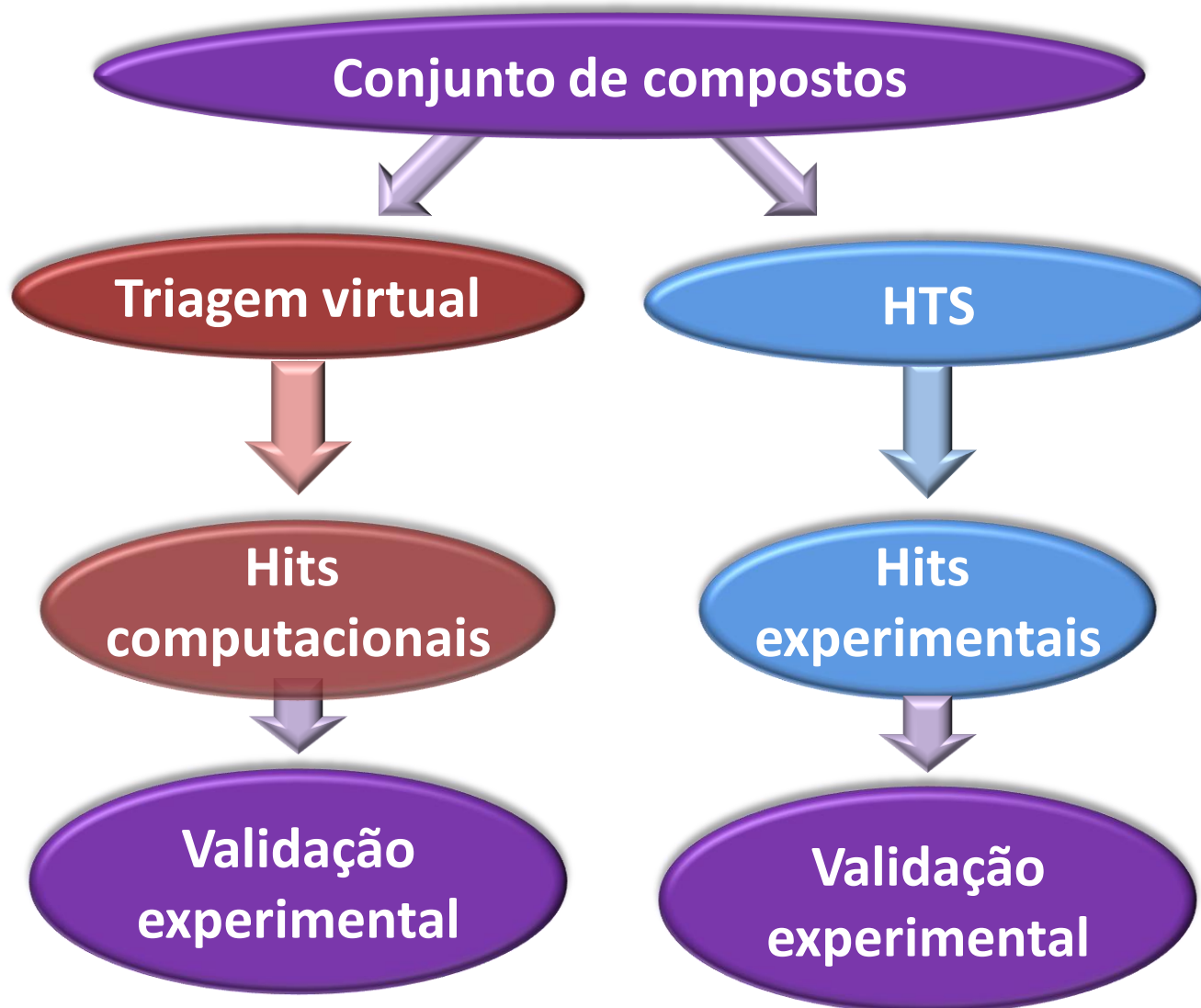
# Combinação de triagem virtual e HTS



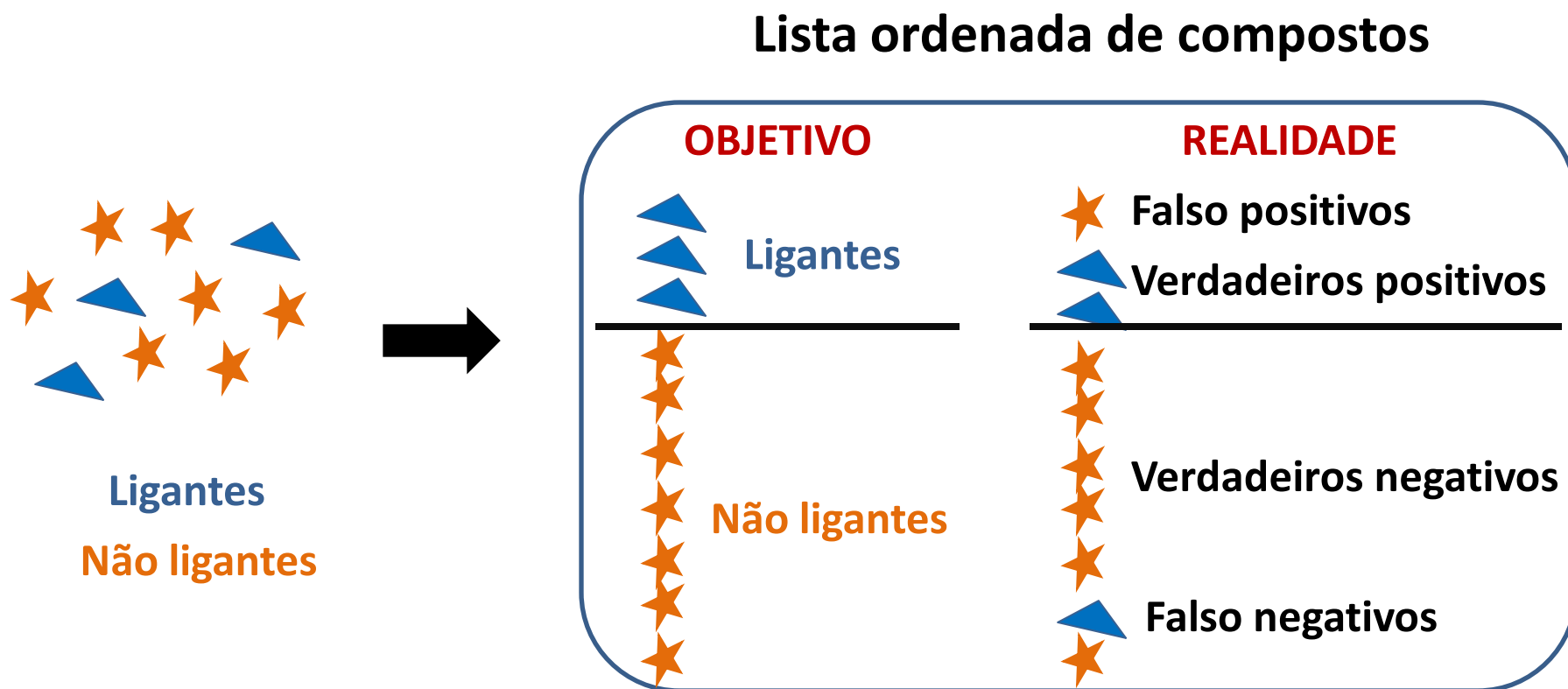
# Combinação de triagem virtual e HTS



# Combinação de triagem virtual e HTS



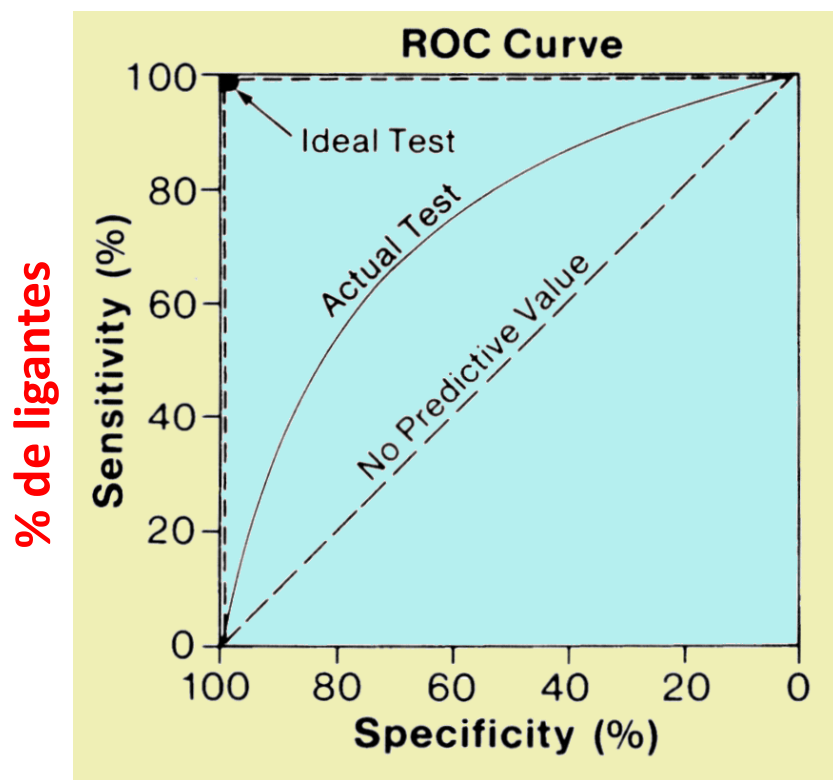
# Avaliação da performance na triagem virtual



**Pouca (ou nenhuma) informação sobre os negativos**

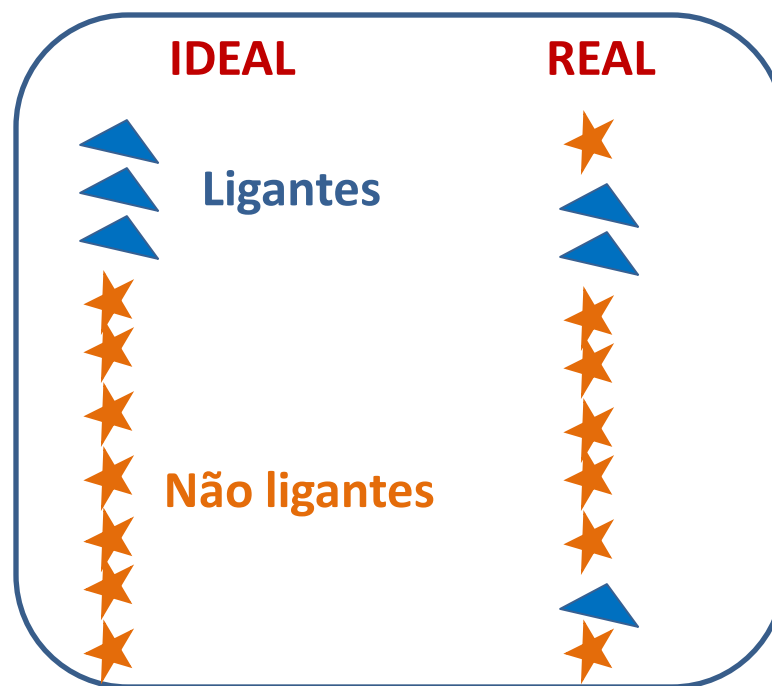
# Análise retrospectiva: Capacidade de classificar bem ligantes

- Curva ROC – Ligantes conhecidos são melhor classificados do que *decoys*?



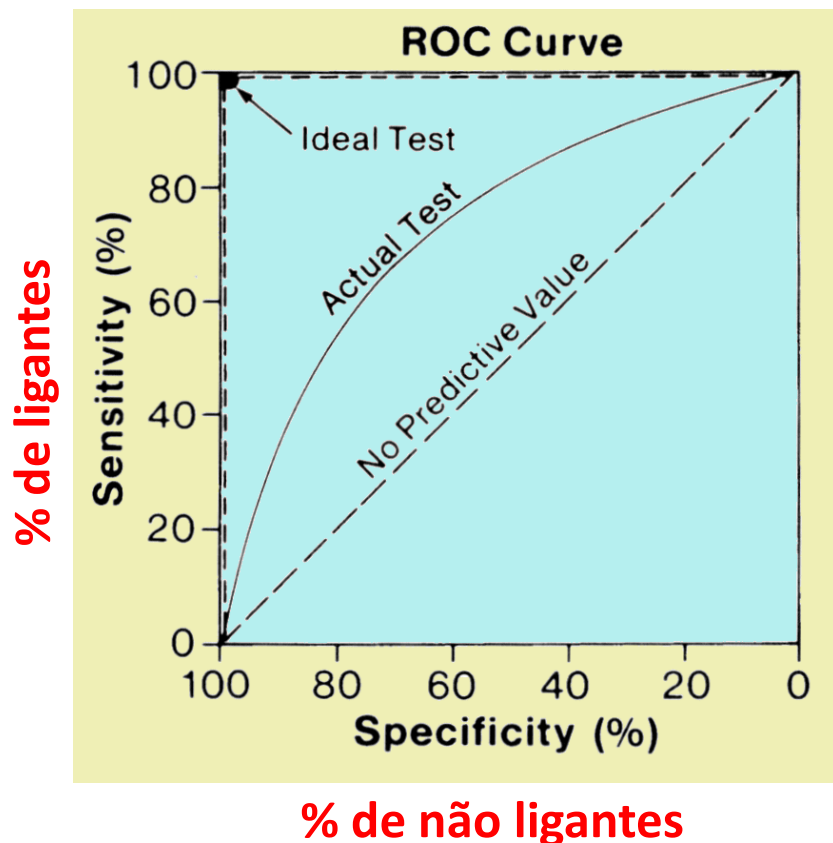
% de ligantes

% de não ligantes



# Análise retrospectiva: Capacidade de classificar bem ligantes

- Curva ROC – Ligantes conhecidos são melhor classificados do que *decoys*?

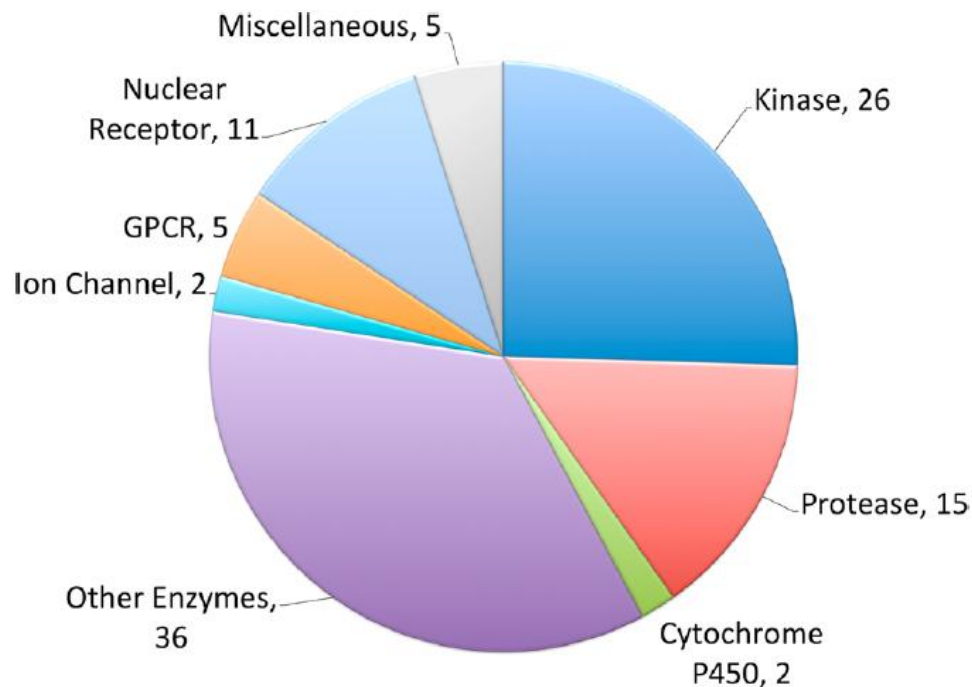


## Limitações:

- Atividade dos *decoys* é desconhecida
- Não garante desempenho prospectivo
- Resultado depende muito dos decoys utilizados

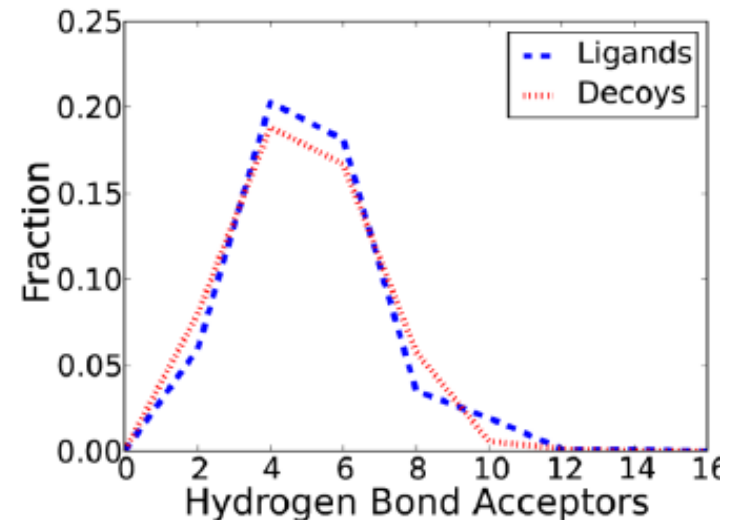
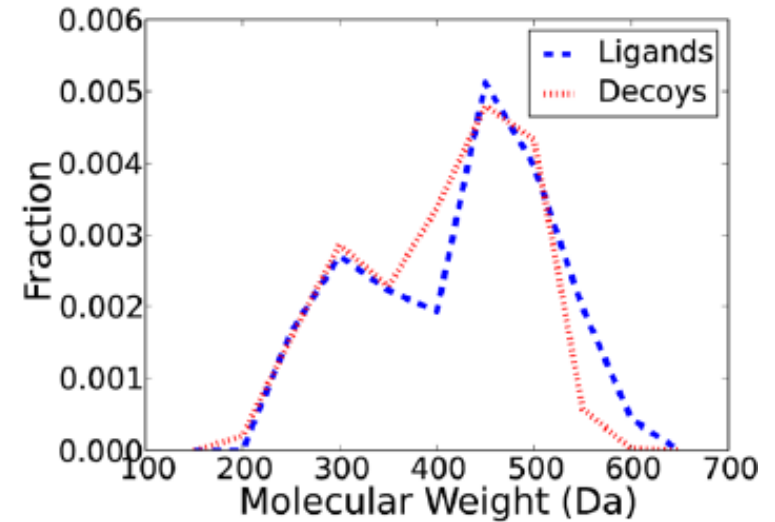
# É importante escolher criteriosamente os decoys

- DUD – 40 alvos e decoys para docking
- DUD-E:
  - Alvos mais diversos – 102 no total
  - 22886 ligantes obtidos do ChEMBL, cada um com 50 decoys gerados do ZINC

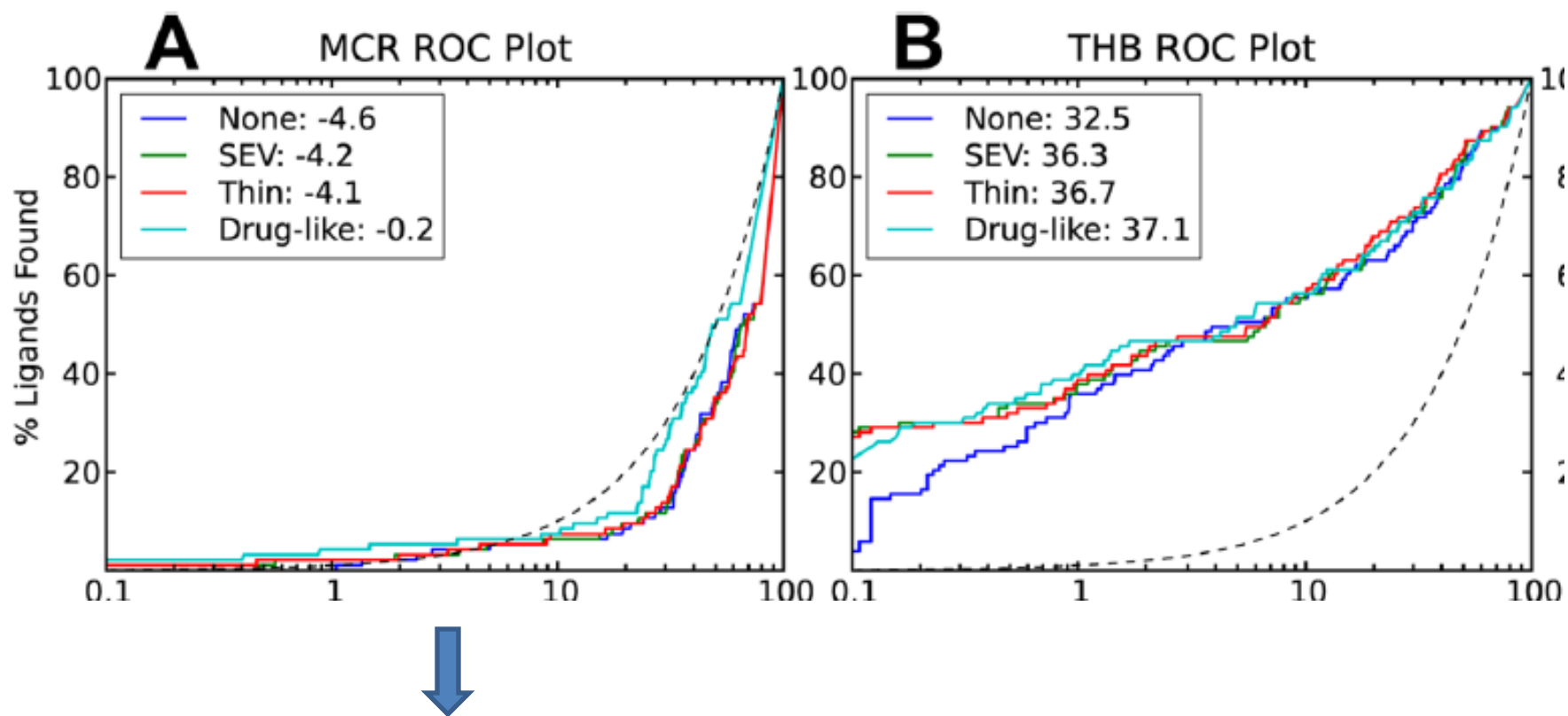


# DUD-E

- Propriedades consideradas:
  - Massa molecular
  - LogP
  - Numero de ligações rotacionáveis
  - Número de doadores de lig de H
  - Número de aceptores de lig de H
  - Carga total



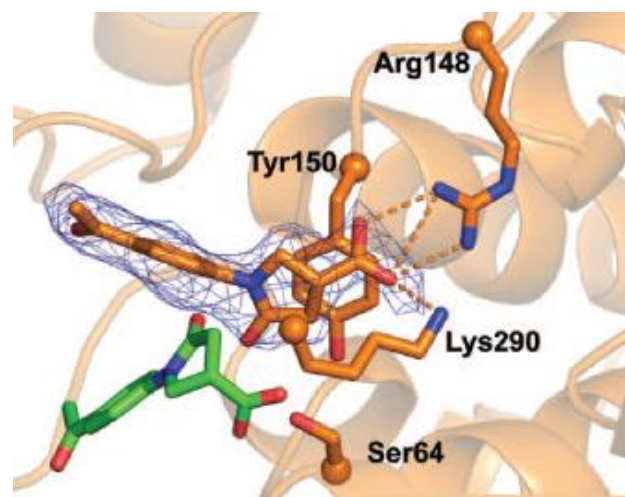
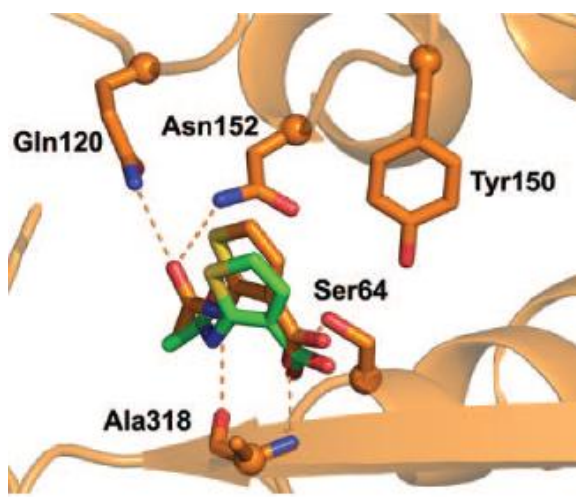
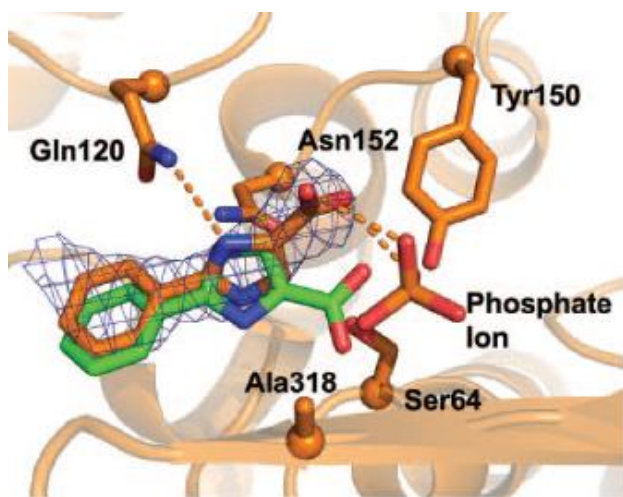
# Através de curvas ROC diversos protocolos podem ser comparados



Escala logartimica – enfatiza dados para o topo da lista

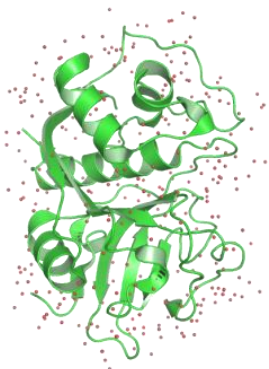
# Análise retrospectiva – Reprodução de dados de cristalografia

- Comparação entre estrutura cristalográfica e pose predita:
  - $Rmsd < 2 \text{ \AA}$
  - Interações intermoleculares mantidas?

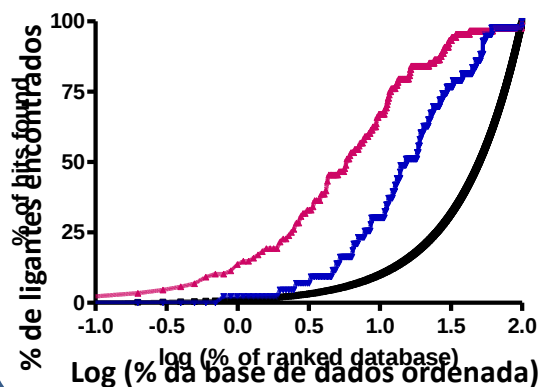
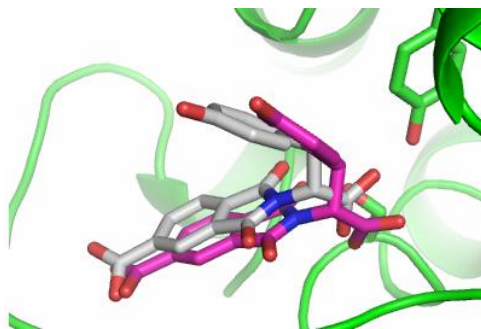


# Análises retrospectivas podem auxiliar a parametrização da triagem

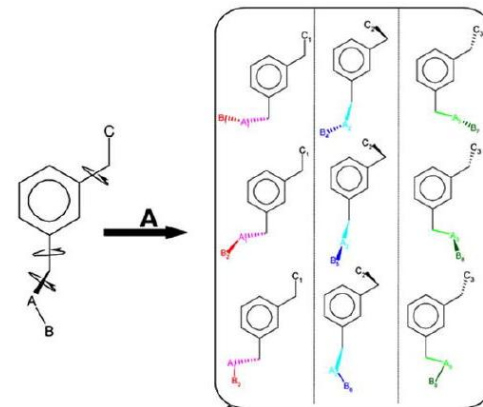
Preparação do receptor e da base de dados



Avaliação de diversos protocolos de docking



Triagem virtual utilizando o(s) melhor(es) protocolo(s)



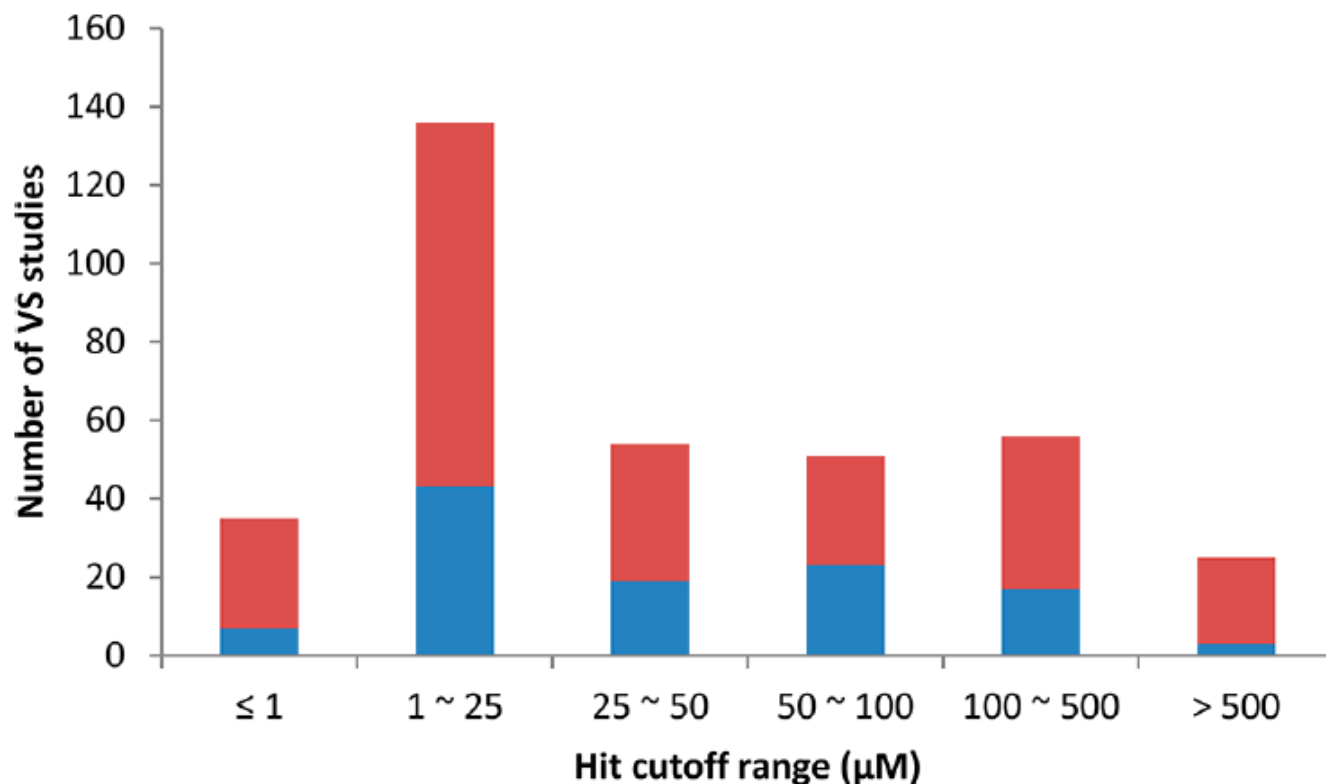
$$E_{score} = E_{vdW} + E_{elec} + \Delta G_{desolv}^L$$

Avaliação dos resultados e seleção dos compostos

Validação experimental

# Dentre os compostos selecionados após VS, quantos são confirmados experimentalmente como ligantes?

- “Hits” definidos arbitrariamente em diferentes estudos



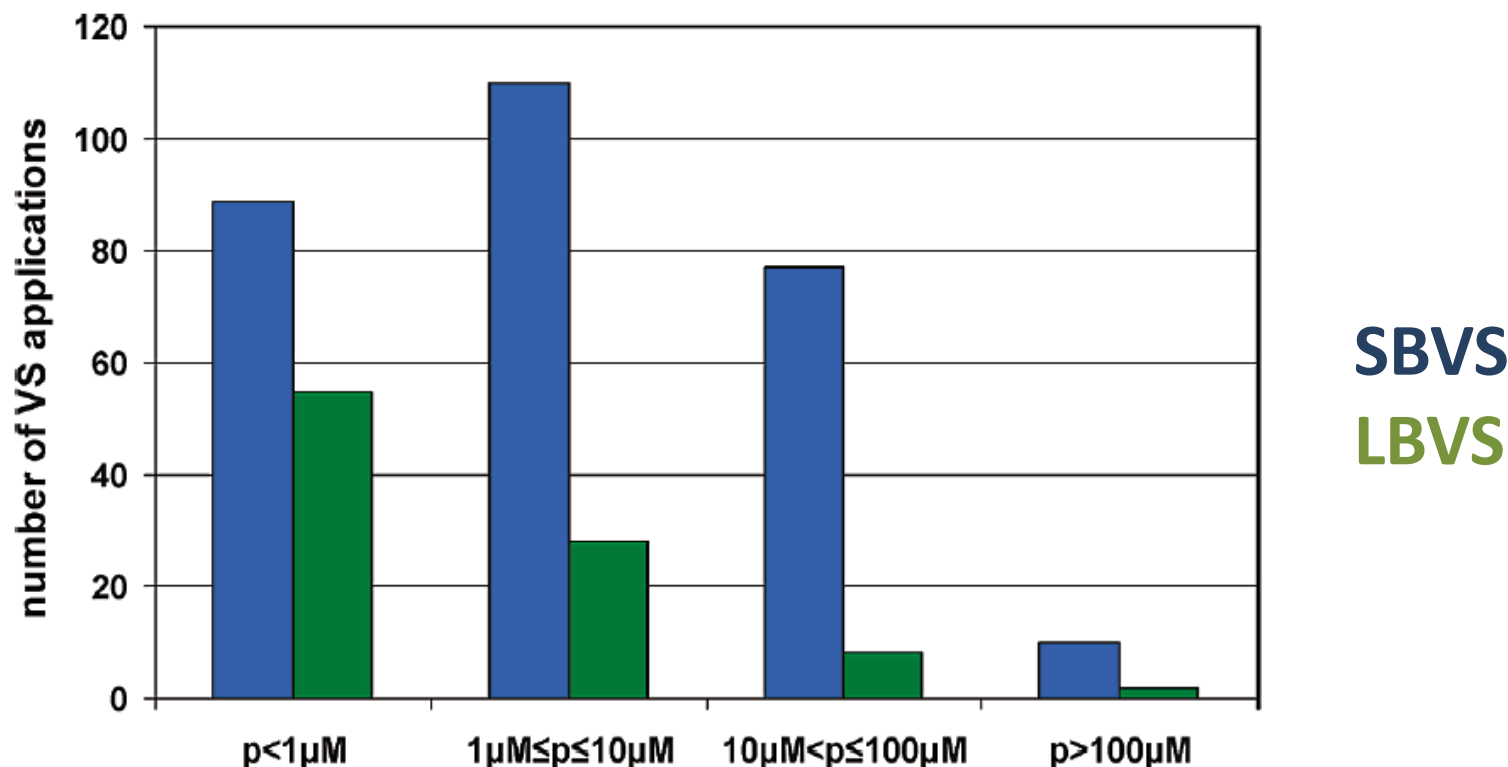
**Figure 1.** Hit cutoff ranges. Values in blue were obtained from studies with clearly defined hit cutoffs. Values in red are estimated from the lowest experimental activity reported for a hit.

## Dentre os compostos selecionados após VS, quantos são confirmados experimentalmente como ligantes?

- Valor mediano = 13 %
  - HTS experimental – valores entre 0.01 e 0.14%

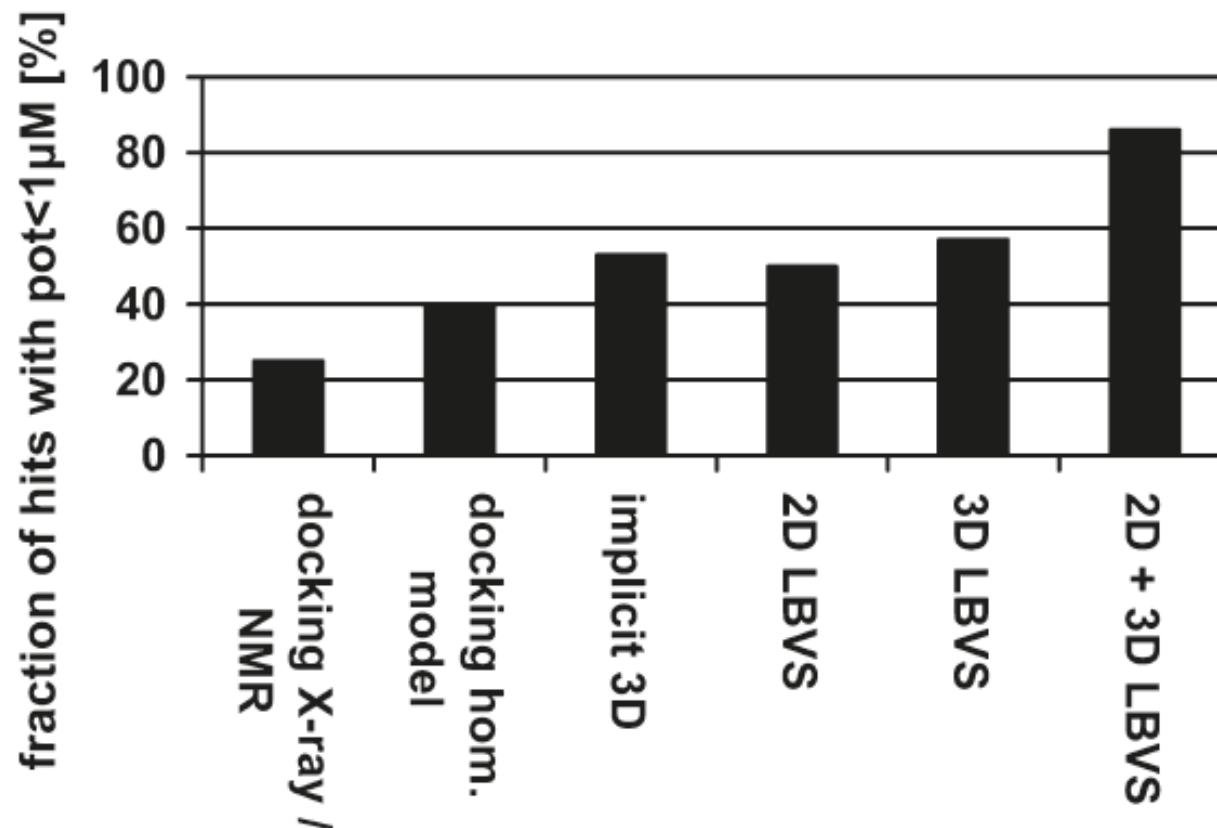
| calcd hit rate (%) |     | % dos estudos avaliados |
|--------------------|-----|-------------------------|
| <1                 | 8   | 2                       |
| 1–5                | 60  | 15.2                    |
| 6–10               | 65  | 16.5                    |
| 11–15              | 65  | 16.5                    |
| 16–20              | 25  | 6.3                     |
| 21–25              | 29  | 7.3                     |
| ≥25                | 103 | 26.1                    |
| ND                 | 40  | 10.1                    |

## Qual a afinidade dos compostos encontrados?

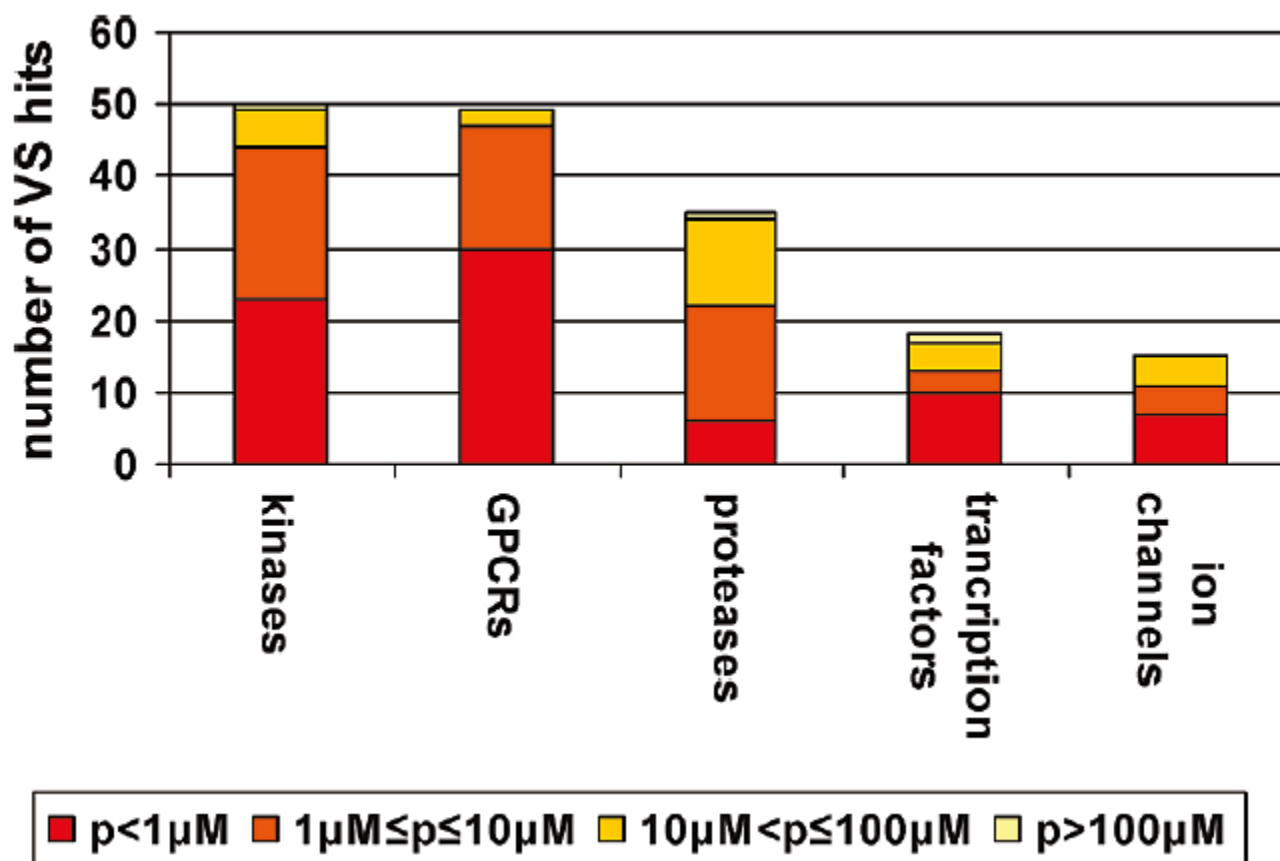


**Figure 4.** Potency distribution of virtual screening hits. Reported VS hits discovered by SBVS (blue) or LBVS (green) approaches were assigned to four potency ( $p$ ) ranges.

## Qual a afinidade dos compostos encontrados?



## A afinidade média varia de acordo com a classe do alvo

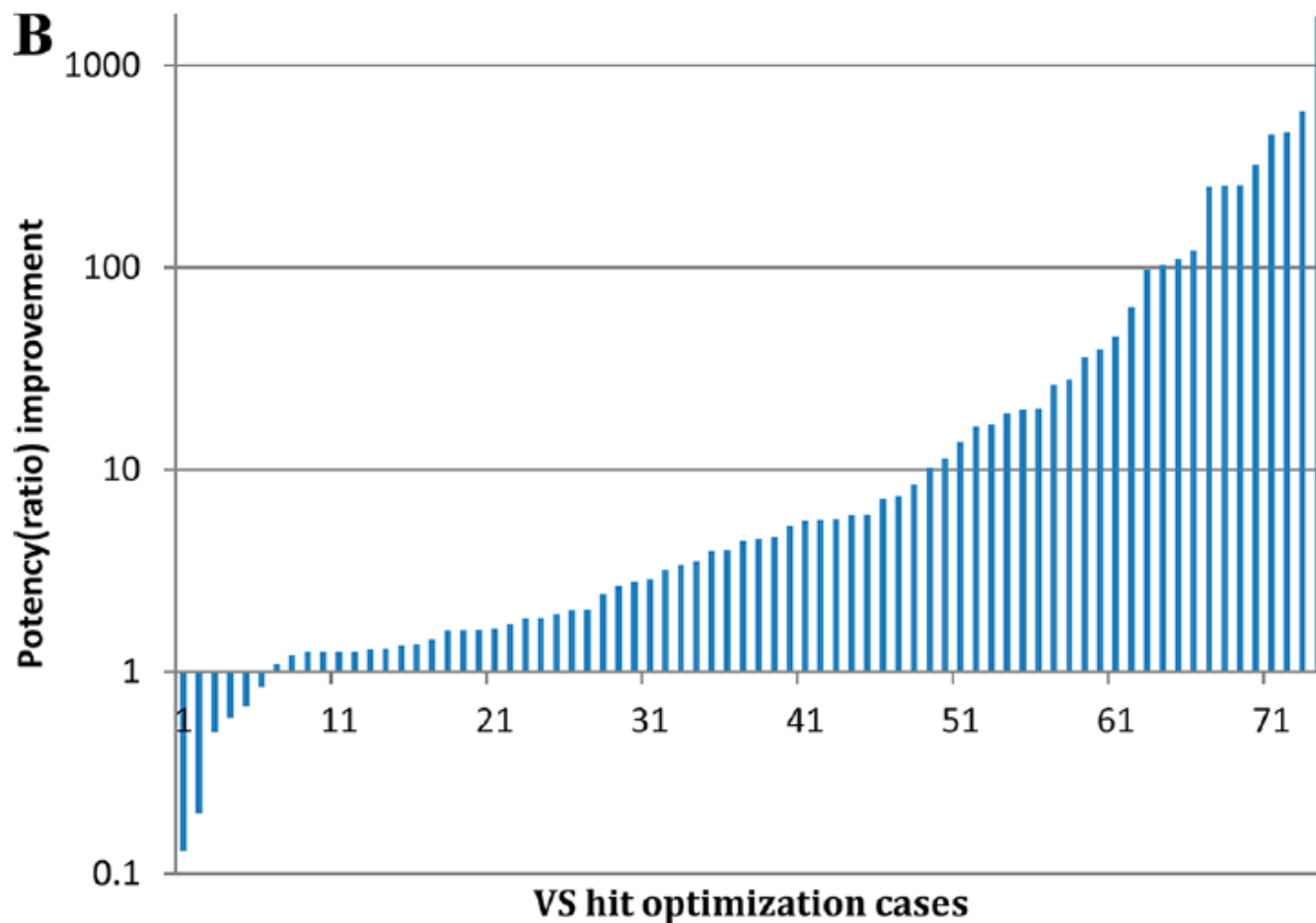


# Os hits descobertos por VS podem ser eficientemente otimizados

- Maior aumento de potência através de otimização por química sintética
- Em média, potência aumenta 70x

|                                  | optimization method |                 |                                   |
|----------------------------------|---------------------|-----------------|-----------------------------------|
|                                  | analogue search     | synthetic SAR   | analogue search and synthetic SAR |
| no. of cases                     | 31                  | 44              | 5                                 |
| LE(ratio) <sup>a</sup> mean      | 1.14 ± 0.22         | 1.10 ± 0.28     | 1.21 ± 0.31                       |
| potency(ratio) <sup>b</sup> mean | 12.30 ± 28.06       | 118.25 ± 297.82 | 14.20 ± 9.74                      |

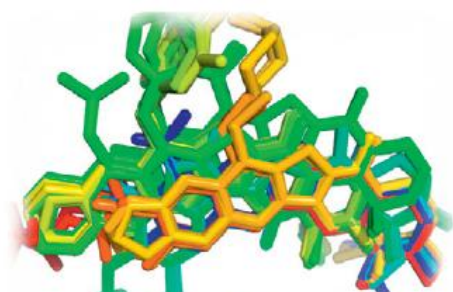
# Os hits descobertos por VS podem ser eficientemente otimizados



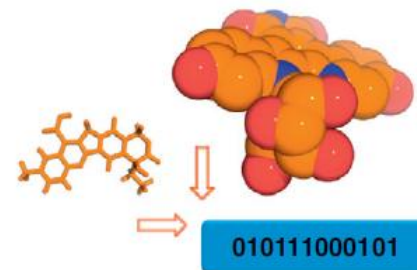
## Vantagens e desvantagens de HTS e VS

| Método | Vantagens                                                                             | Desvantagens                                         |
|--------|---------------------------------------------------------------------------------------|------------------------------------------------------|
| HTS    | Informação experimental sobre toda a base de dados                                    | Alta porcentagem de falso-positivos                  |
|        | Possibilidade de descoberta de inibidores que atuem em diversos sítios do receptor    | Grande exigência de infraestrutura                   |
|        | Independente de existência de outros ligantes e de conhecimento de estrutura do alvo  |                                                      |
| VS     | Restrição do número de compostos a serem testados                                     | Necessidade de modelo tridimensional do alvo (SBVS)  |
|        | Menor exigência de infraestrutura                                                     | Ligantes restritos a um determinado sítio de ligação |
|        | Maior porcentagem de ligantes específicos entre compostos avaliados experimentalmente | Falso-negativos                                      |

# Vantagens e desvantagens de LBVS



Ligand-based  
pharmacophores



2D and 3D similarity

Possível sem informação sobre a estrutura proteica

Simple e rápido

Possibilidade de scaffold hopping

Desconsideração da estrutura proteica

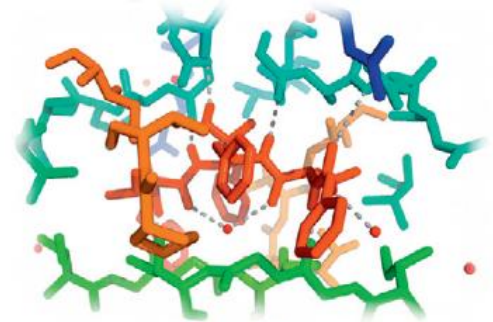
Viés para ligantes semelhantes aos já conhecidos

Descritores dependentes de conformação

# Vantagens e desvantagens de SBVS



Structure-based  
pharmacophores



Docking

Possível sem informação sobre ligantes  
Predição de modo de ligação

Necessidade de considerar múltiplas conformações da proteína

Todo o sítio de ligação é considerado  
Scaffold hopping

Não apresenta viés para ligantes  
conhecidos

Dificuldade em determinar  
características essenciais

Funções de pontuação super simplificadas  
Alto custo computacional



# Programa de Pós-Graduação em Bioinformática Universidade Federal de Minas Gerais

- Genômica, Bioinformática estrutural



CONCEITO  
NÍVEL



Programa de Pós-Graduação em

## BIOQUÍMICA E IMUNOLOGIA

Instituto de Ciências Biológicas  
Universidade Federal de Minas Gerais



CONCEITO  
NÍVEL 7

**IC**

Filipe Vilela



**MESTRADO**

Fabiana Póvoa



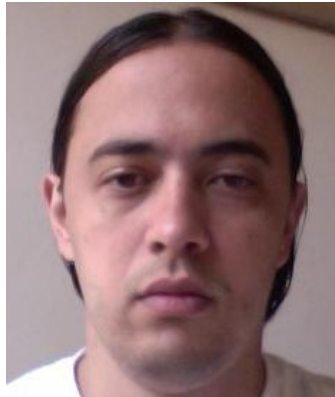
**CO-ORIENTAÇÃO**

Luan Carvalho Saulo Braga João Fernandes



**DOUTORADO**

Francielly Rodrigues Raoni Souza



Edson Folador Alberto Júnior

