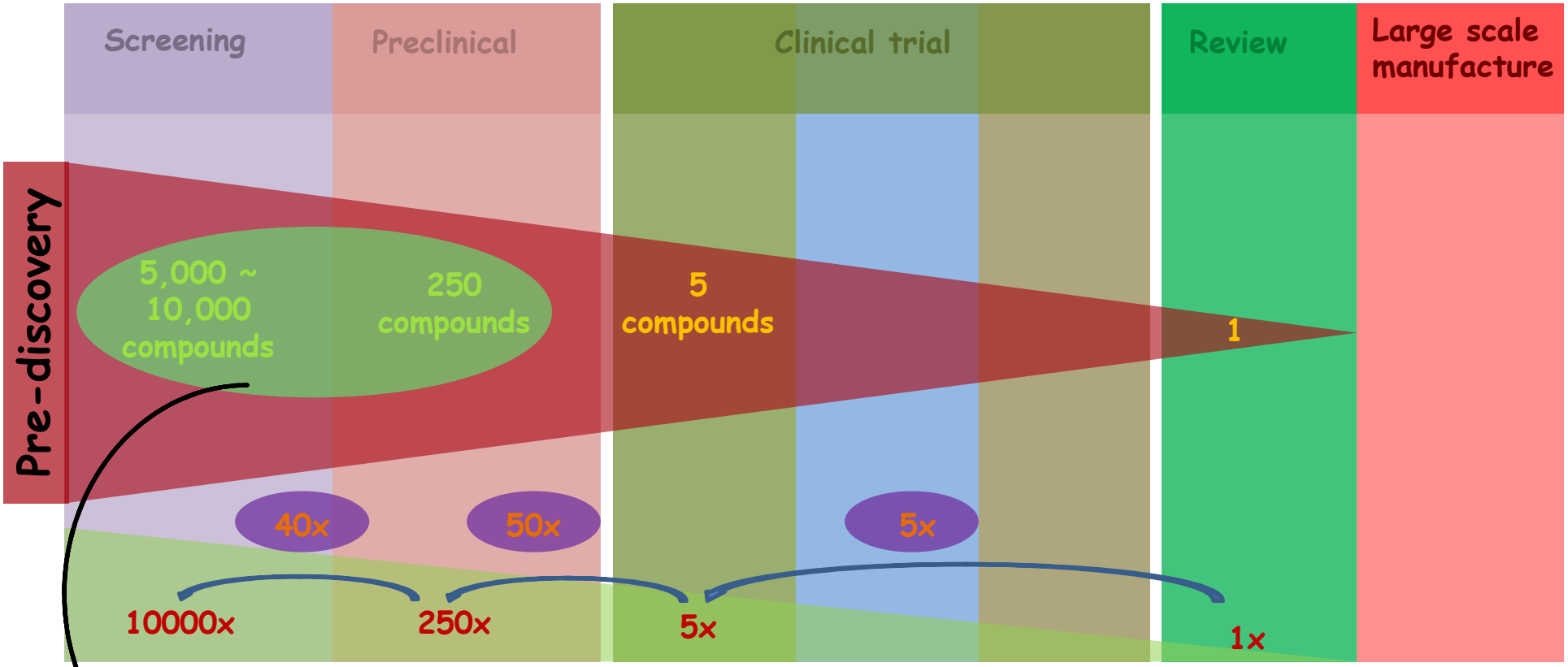
Molecular Docking

Molecular Docking In Drug Discovery

Screening specific candidates is challenging



40x reduction of non-specific candidates

Molecular Docking: General Concepts

- Molecular docking is a well established computational technique which predicts the interaction energy between two molecules.

- Molecular docking studies are used to determine the interaction of two molecules and to find the best orientation of ligand which would form a complex with overall minimum energy.

- The small molecule, known as ligand usually fits within protein's cavity which is predicted by the search algorithm. These protein cavities become active when come in contact with any external compounds and are thus called as active sites.

Different types of Interactions

Interactions between particles can be defined as a consequence of forces between the molecules contained by the particles. These forces are divided into four categories:

Electrostatic forces - Forces with electrostatic origin due to the charges residing in the matter. The most common interactions are charge-charge, charge-dipole and dipole-dipole.

Electrodynamics forces-most widely known is the Van der Waals interactions.

Steric forces - Steric forces are generated when atoms in different molecules come into very close contact with one another and start affecting the reactivity of each other. The resulting forces can affect chemical reactions and the free energy of a system.

Solvent-related forces -These are forces generated due to chemical reactions between the solvent and the protein or ligand. Examples are Hydrogen bonds (hydrophilic interactions) and hydrophobic interactions.

Search Algorithm

These algorithms determine all possible optimal conformations for a given complex (protein-protein, protein-ligand) in a environment i.e. the position and orientation of both molecules relative to each other. They can also calculate the energy of the resulting complex and of each individual interaction.

The different types of algorithms that can be used for docking analysis are given below:

- Molecular dynamics
- Monte Carlo methods
- Genetic algorithms
- Fragment-based methods
- Point complementary methods
- Distance geometry methods
- Systematic searches

Scoring function

These are mathematical methods used to predict the strength of the non-covalent interaction called as binding affinity, between two molecules after they have been docked.

Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, for example between two proteins or between protein and DNA or protein and drug.

These configurations are evaluated using scoring functions to distinguish the experimental binding modes from all other modes explored through the searching algorithm.

For example:

- Empirical scoring function of Igemdock

Fitness = vdW + Hbond + Elec

- Binding Energy

$\Delta G_{bind} = \Delta G_{vdw} + \Delta Gh_{bond} + \Delta G_{elect} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$

Screening

Define the problem

↓

Chalk out the possibilities to solve the problem

↓

Identify a target (protein/ enzyme)

↓

Screen available ligands (phytochemicals) against the target

Effective ligands can be screened

Types of docking

The following are majorly used type of docking are-

Lock and Key or Rigid Docking –In rigid docking, both the internal geometry of the receptor and ligand is kept fixed during docking

Induced fit or Flexible Docking - In this model, the ligand is kept flexible and the energy for different conformations of the ligand fitting into the protein is calculated. Though more time consuming, this method can evaluate many different possible conformations which make it more reliable.

Major steps in molecular docking

Step I – Building the Receptor
In this step the 3D structure of the receptor should be downloaded from PDB; and modified. This should include removal of the water molecules from the cavity, stabilizing charges, filling in the missing residues, generation the side chains etc according to the parameters available. After modification the receptor should be biological active and stable.

Step II – Identification of the Active Site
After the receptor is built, the active site within the receptor should be identified. The receptor may have many active sites but the one of the interest should be selected. Most of the water molecules and heteroatoms if present should be removed.

Major steps in molecular docking

Step III – Ligand Preparation

Ligands can be obtained from various databases like ZINC, PubChem or can be sketched using tools like Chemsketch. While selecting the ligand, the LIPINSKY'S RULE OF 5 should be applied. The rule is important for drug development where a pharmacologically active lead structure is optimized stepwise for increased activity and selectivity, as well as drug-like properties, as described.

For the selection of a ligand using LIPINSKY'S RULE:

- Not more than 5 –H bond donors.
- Molecular Weight NOT more than 500 Da.
- Log P not over 5 for octanol water partition coefficient.
- NOT more than 10 H bond acceptors.

Step IV- Docking

This is the last step, where the ligand is docked onto the receptor and the interactions are checked. The scoring function generates scores depending on which the ligand with the best fit is selected.

Molecular Docking: A Case Study

Identifying target molecule

ADP ribose phosphatase of NSP3
PDB Accession: 6VXS

ADP-ribose 1′-phosphate + H_2O → ADP-ribose + phosphate

Procedure

Discovery Studio

↓

File

↓

Open

↓

Select the file

↓

Tools

↓

Receptor-Ligand interaction

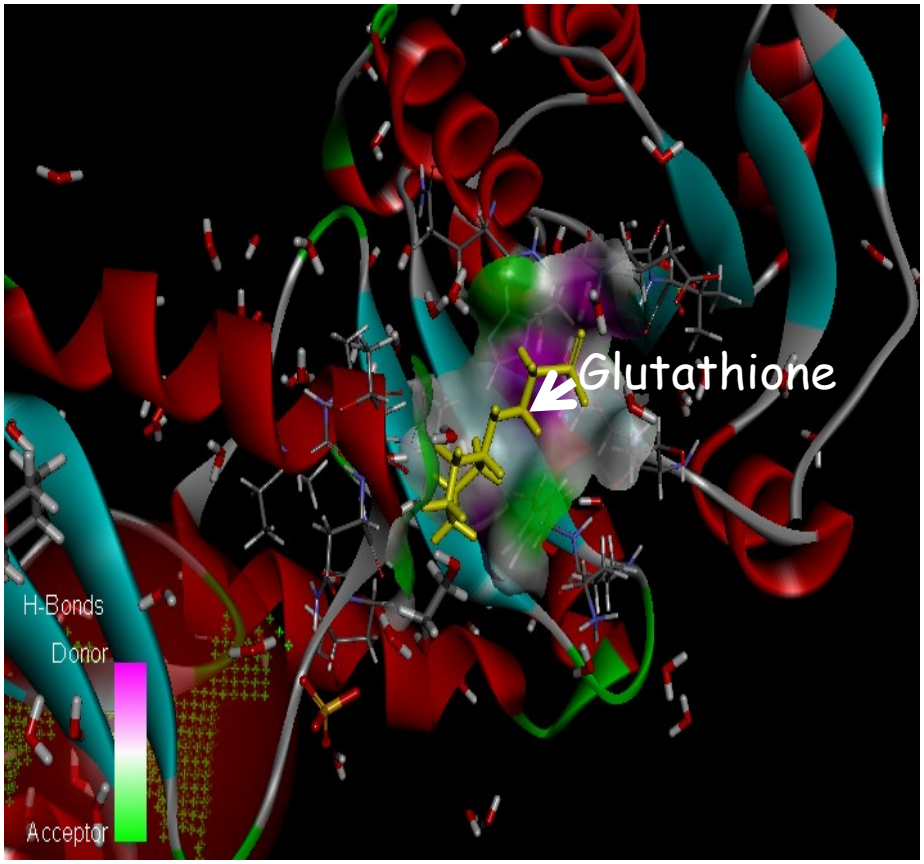
↓

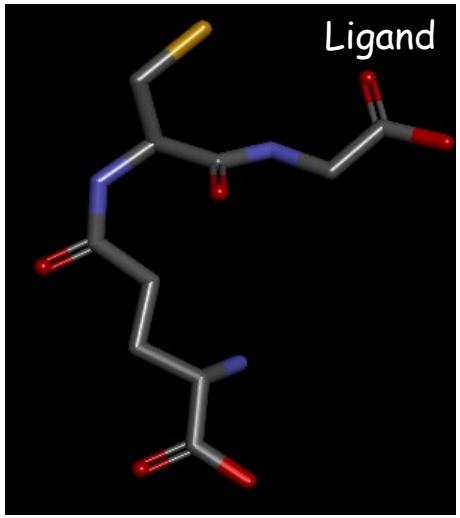
From receptor cavities

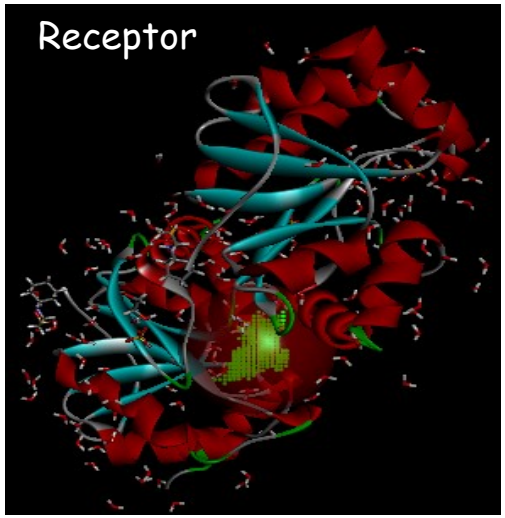
↓

Define and edit binding site

Screening of Phytochemicals







Ligand	Receptor (PDB Accession)
Glutathione	6VXS
-CDOCKER ENERGY	-CDOCKER INTERACTION ENERGY
39.2875	36.1134

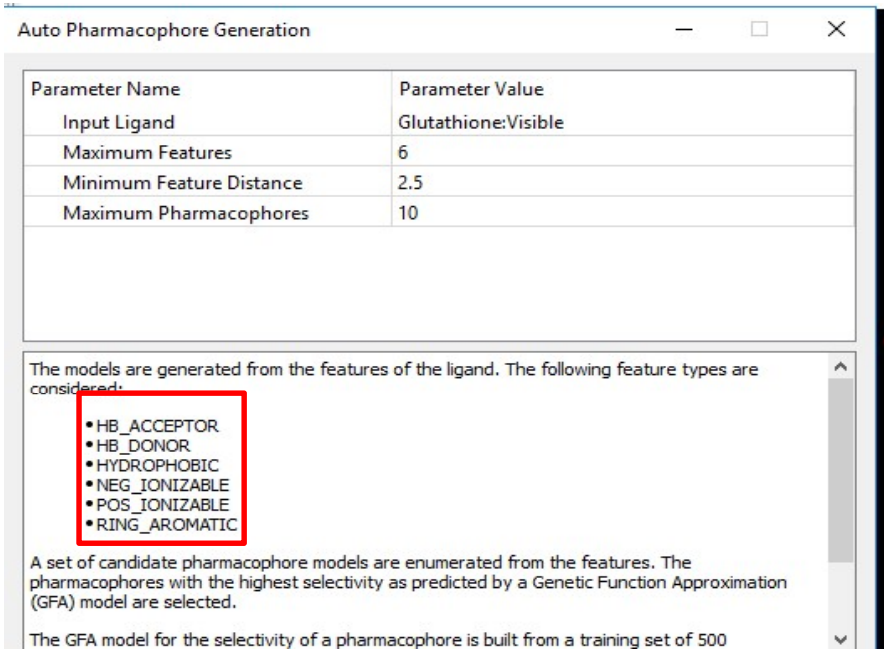
Procedure

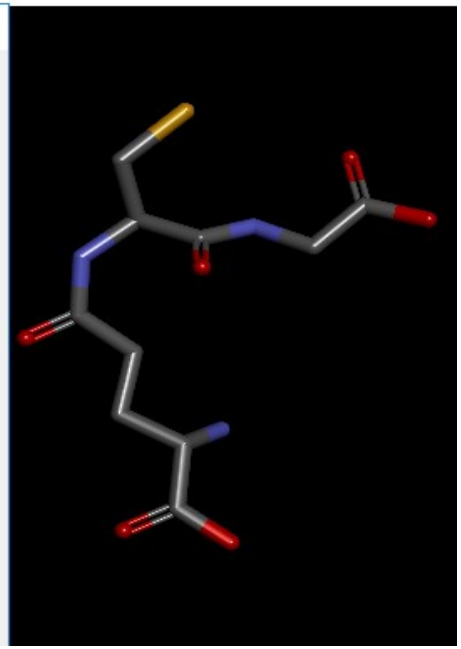
Receptor-Ligand interaction → Dock Ligands → CDOCKER

Glutathione effectively binds to the active site of the target protein

Possible pharmacophores

Automated generation of pharmacophores





Procedure

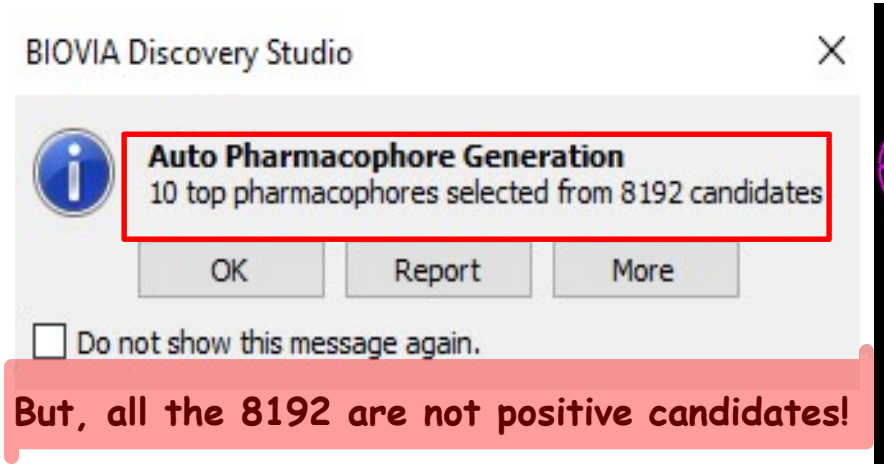
Tools

↓

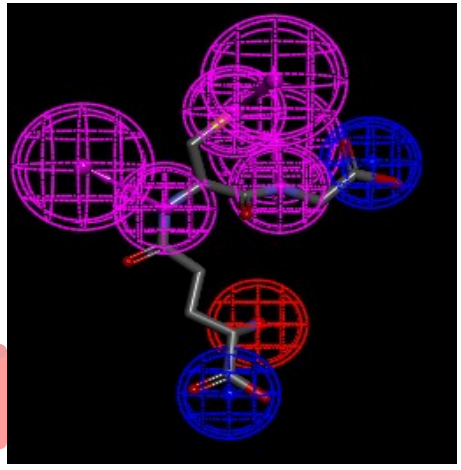
Pharmacaophore

↓

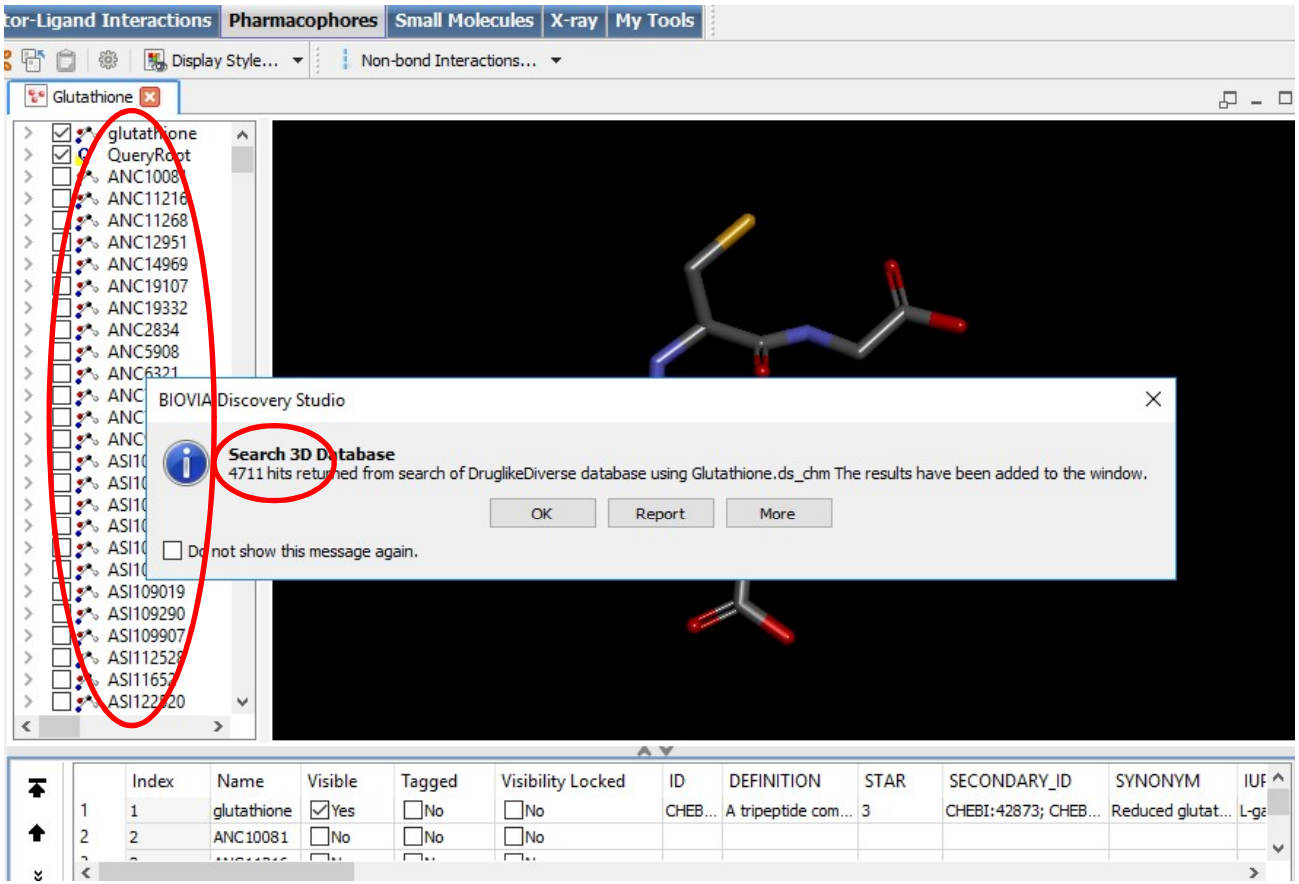
Auto pharmacophore generation



But, all the 8192 are not positive candidates!



Generation of pharmacophores depending on input query



Procedure

Tools

↓

Pharmacaophore

↓

Search, Screen and Profile

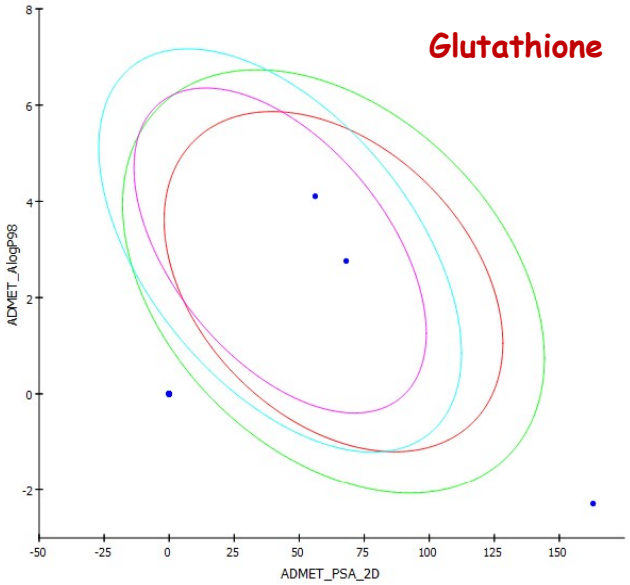
↓

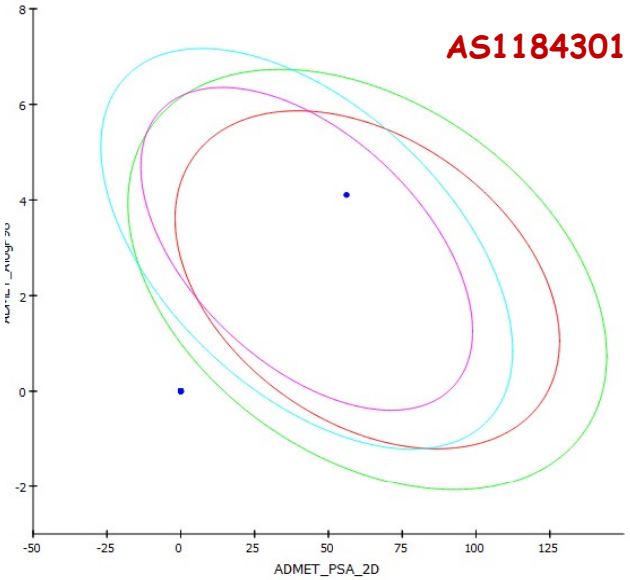
Search 3D Database

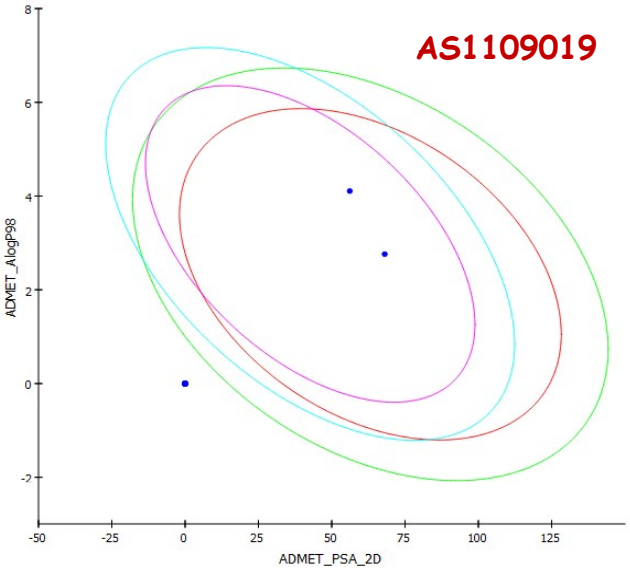
4711 hits from drug like diverse database

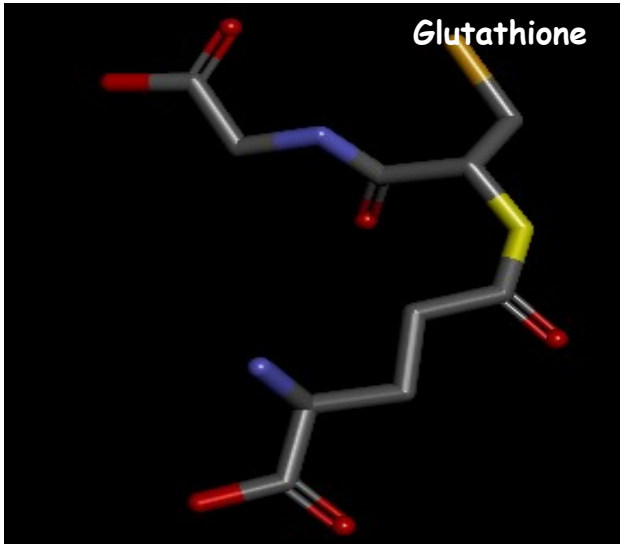
But, all the 4711 are not positive candidates!

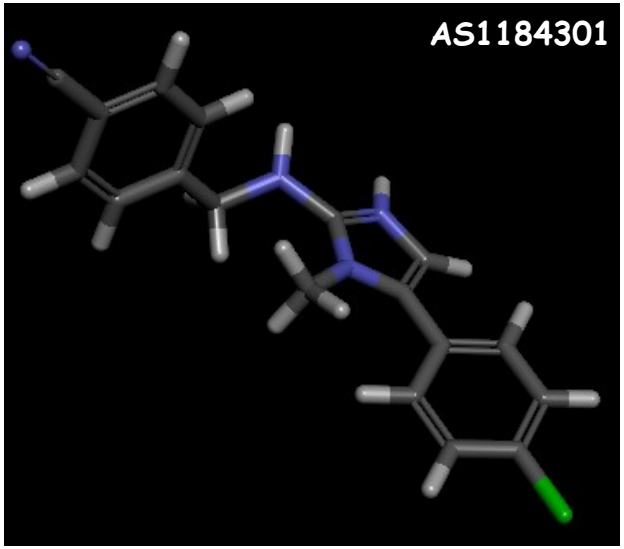
ADMET and Toxicity

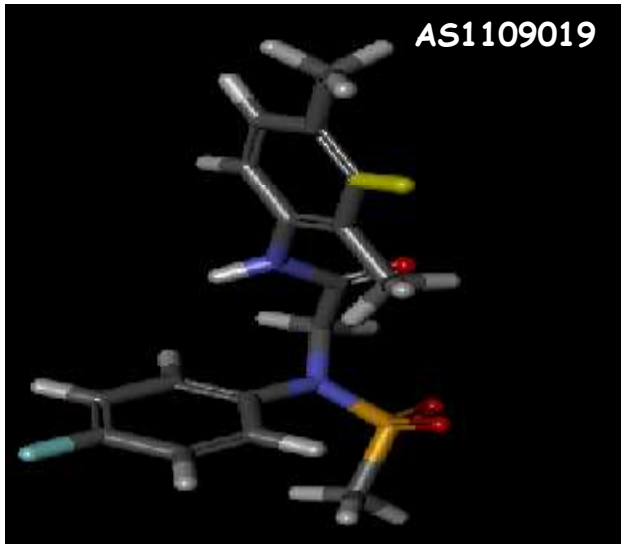












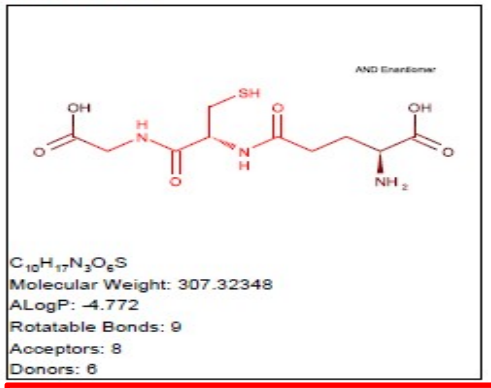
Mutagenic nature

Ligand	Mutagenic nature
Glutathione	Mutagen
AS1184301	Non-mutagen
AS1109019	Non-mutagen

Procedure

Tools → Small molecules → Screening → ADMET TOXICITY

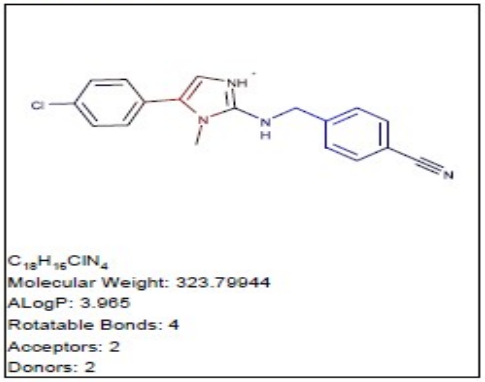
glutathione



$C_{10}H_{17}N_3O_6S$
Molecular Weight: 307.32348
ALogP: -4.772
Rotatable Bonds: 9
Acceptors: 8
Donors: 6

Model Prediction
Prediction: Mutagen
Bayesian Score: 1.35
Mahalanobis Distance: 13.4
Mahalanobis Distance p-value: 2.98e-07
Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

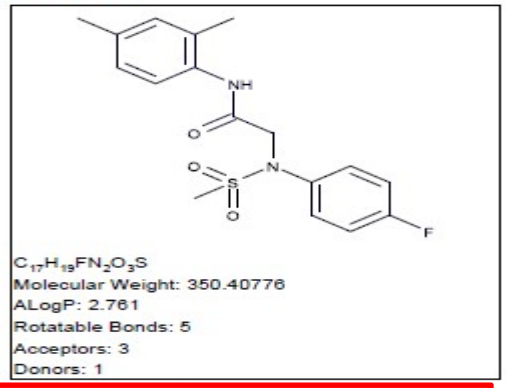
ASI184301



$C_{18}H_{16}ClN_4$
Molecular Weight: 323.79944
ALogP: 3.965
Rotatable Bonds: 4
Acceptors: 2
Donors: 2

Model Prediction
Prediction: Non-Mutagen
Bayesian Score: -3.13
Mahalanobis Distance: 9.61
Mahalanobis Distance p-value: 0.528
Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

ASI109019



$C_{17}H_{15}FN_2O_3S$
Molecular Weight: 350.40776
ALogP: 2.761
Rotatable Bonds: 5
Acceptors: 3
Donors: 1

Model Prediction
Prediction: Non-Mutagen
Bayesian Score: -3.27
Mahalanobis Distance: 11.3
Mahalanobis Distance p-value: 0.0134
Prediction: Positive if the Bayesian score is above the estimated best cutoff value from minimizing the false positive and false negative rate.
Bayesian Score: The standard Laplacian-modified Bayesian score.
Mahalanobis Distance: The Mahalanobis distance (MD) is the distance to the center of the training data. The larger the MD, the less trustworthy the prediction.
Mahalanobis Distance p-value: The p-value gives the fraction of training data with an MD greater than or equal to the one for the given sample, assuming normally distributed data. The smaller the p-value, the less trustworthy the prediciton. For highly non-normal X properties (e.g., fingerprints), the MD p-value is wildly inaccurate.

Receptor-Ligand interaction

Ligand	Receptor (PDB Accession)	-CDOCKER ENERGY	-CDOCKER INTERACTION ENERGY
Glutathione	6VXS	39.2875	36.1134
AS1184301	6VXS	8.083	22.690
AS1109019	6VXS	24.409	32.447