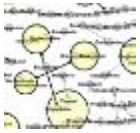
# Cheminformatics Introduction

Jeremy Yang

Associate Instructor
IU School of Informatics and Computing

*Instructor: Prof. Joanne Luciano*

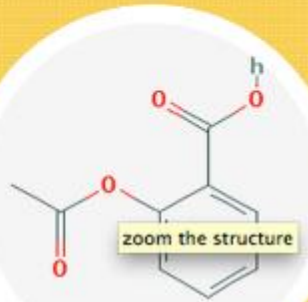


DATA SCIENCE FOR DRUG DISCOVERY, HEALTH AND TRANSLATIONAL MEDICINE (DSDHT) INFO I-590

# Cheminformatics: Introduction


FREE ONLINE COURSE
*introducing cheminformatics*

learncheminformatics.com

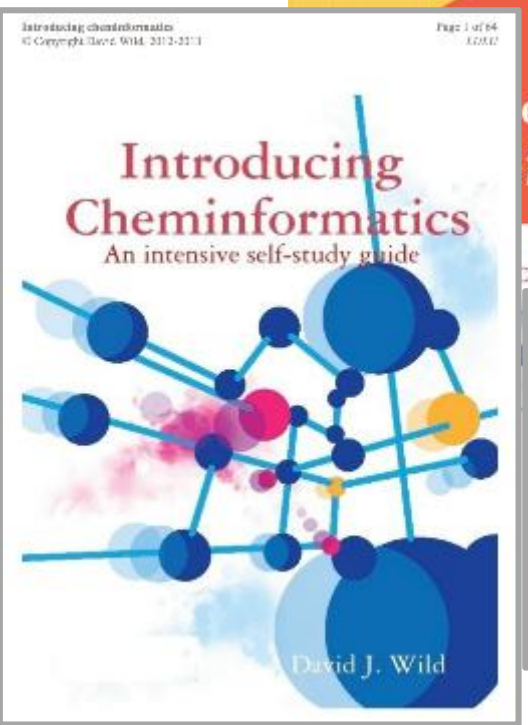



**learn cheminformatics!**

Cheminformatics is the study of all aspects of the representation and use of chemical and related biological information on computer. It has applications in drug discovery, health, data mining and many other areas

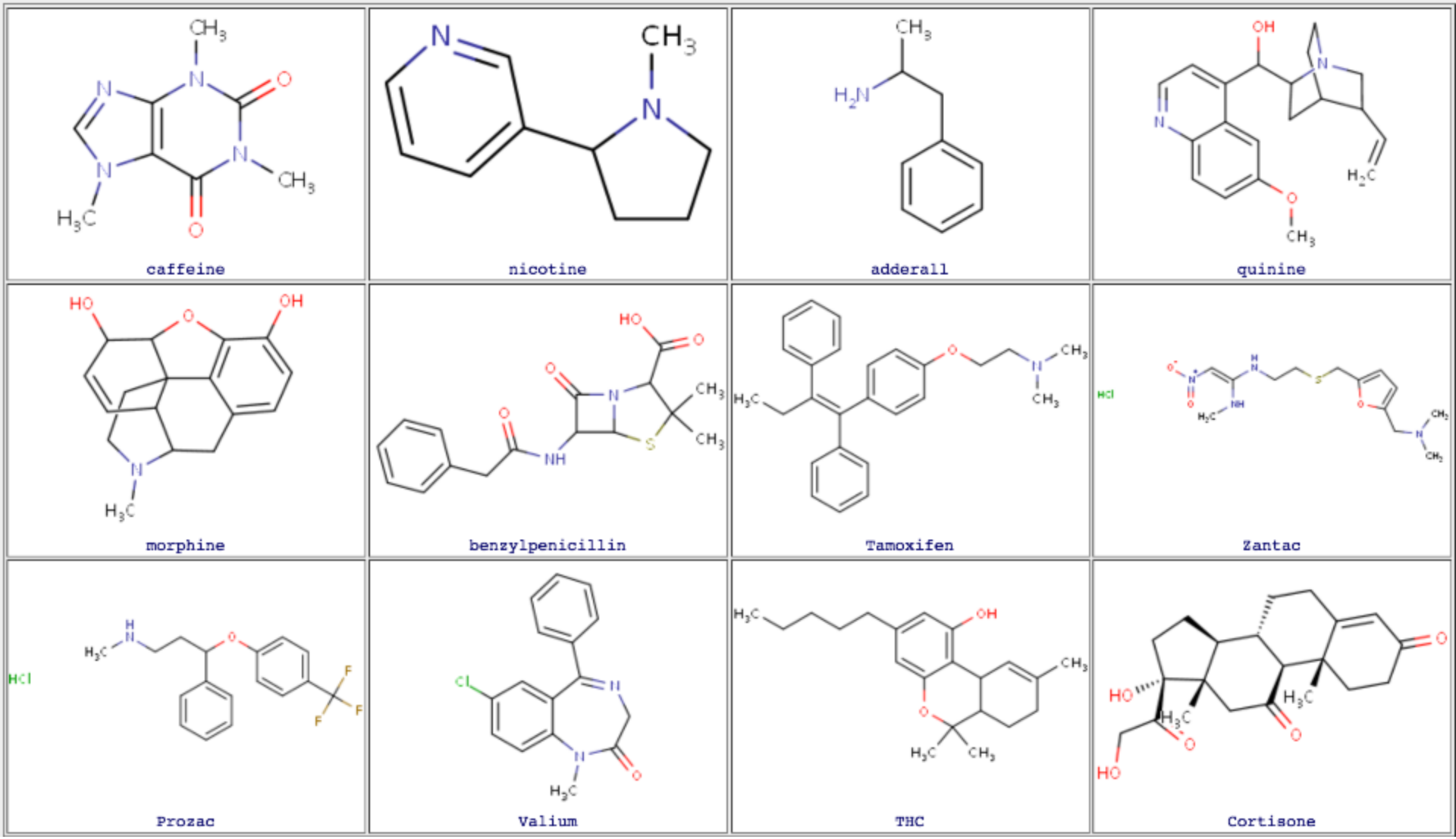
This site gives you links to some resources to get you started in learning about cheminformatics - a free online course, a low-cost eBook, an established open access journal and a

eBook just $9.95
*Kindle and PDF format*
*library versions available*



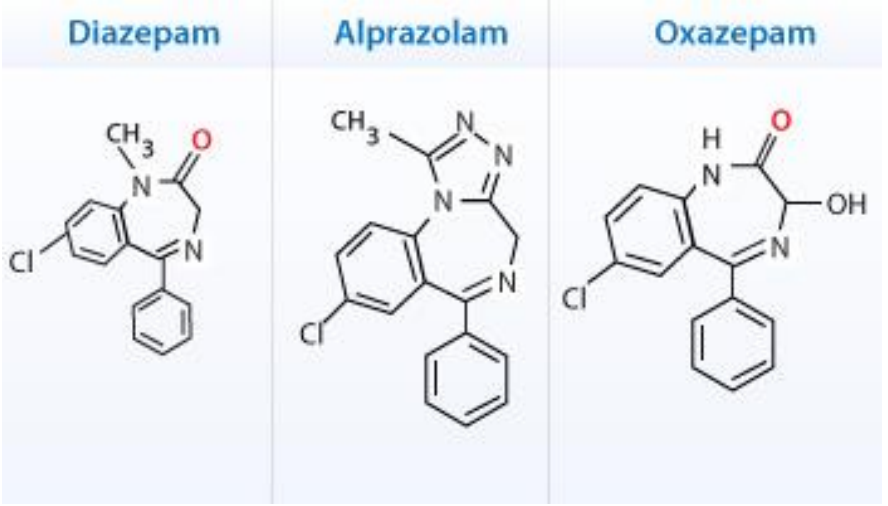

Cheminformatics is the study of all aspects of the representation and use of chemical and related biological information on computer. It has applications in drug discovery, health, data mining and many other areas

# Cheminformatics:
# What makes drugs so special?



# Cheminformatics: Chemotypes

- Benzodiazepines class of tranquilizers sharing a common ring system, a pattern, chemotype or scaffold.
- Similarity principle: similar chemicals, bioactivity.
- Representing chemicals as molecular graphs enables efficient storage and analysis.




| Drug name | Brand names |
|---|---|
| diazepam | Valium, Ducene, Antenex, Valpam |
| oxazepam | Serepax, Murelax, Alepam |
| nitrazepam | Mogadon, Alodorm |
| temazepam | Normison, Temaze, Temtabs |
| lorazepam | Ativan |
| flunitrazepam | Rohypnol, Hypnodorm |
| bromazepam | Lexotan |
| clonazepam | Rivotril, Paxam |

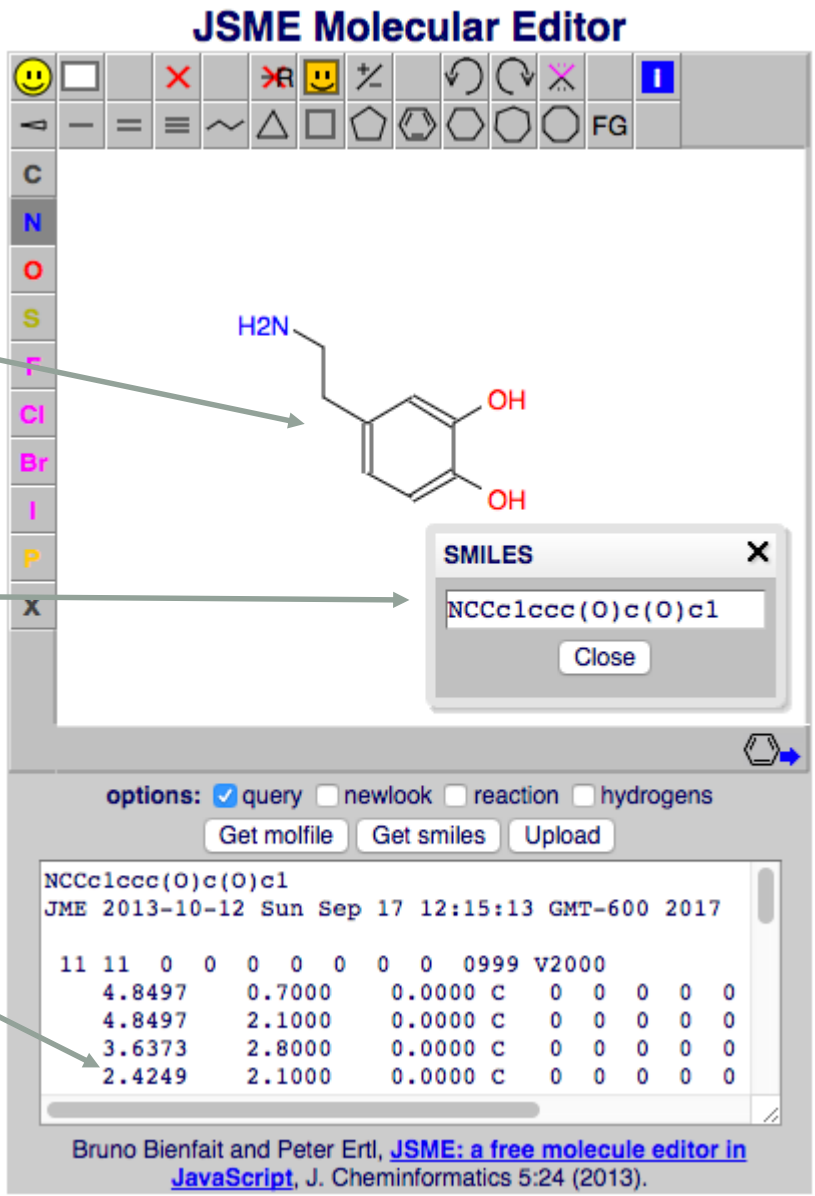
From https://www.wonderwhizkids.com/conceptmaps/Tranquilizers.html

# Cheminformatics: Representing molecules computationally

molecular depiction (2D)

SMILES (Simplified Molecular Input Line Entry System)

molfile (connection table)




http://pasilla.health.unm.edu/jsme.html

# Cheminformatics: Chemical databases

NIH › U.S. National Library of Medicine › National Center for Biotechnology Information

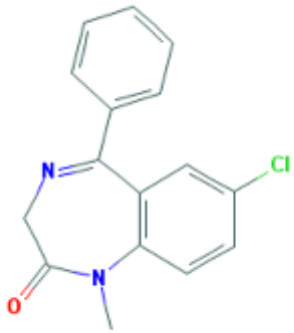
PubChem OPEN CHEMISTRY DATABASE

Search Compounds

Compound Summary for CID 3016 | Download | Share

## Diazepam

STRUCTURE | VENDORS | DRUG INFO | PHARMACOLOGY | LITERATURE | PATENTS | BIOACTIVITIES

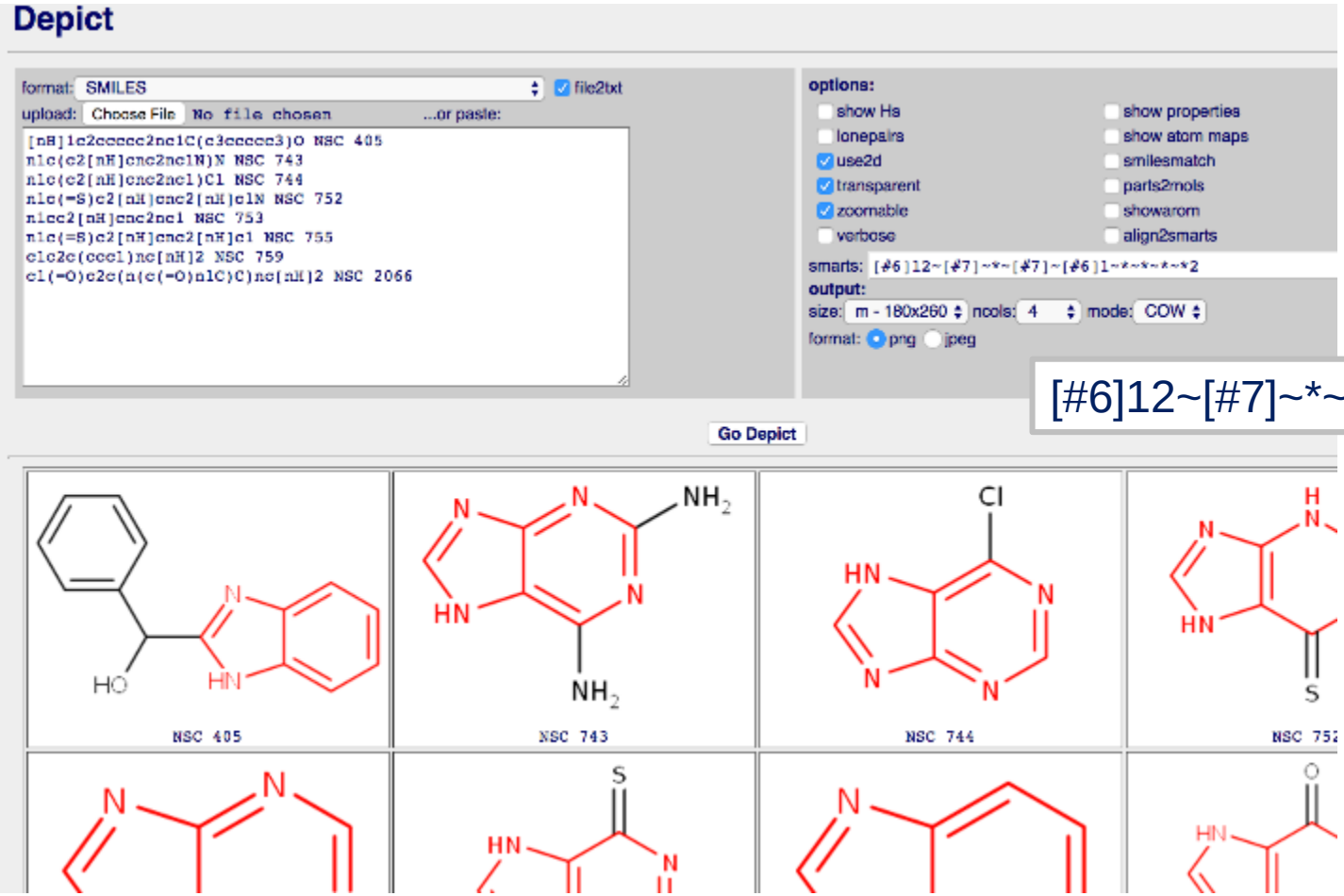



| | |
|---|---|
| **PubChem CID:** | 3016 |
| **Chemical Names:** | Diazepam; Valium; Diazemuls; Relanium; Ansiolisina; Stesolid Mor |
| **Molecular Formula:** | $C_{16}H_{13}ClN_2O$ |
| **Molecular Weight:** | 284.743 g/mol |
| **InChI Key:** | AAOVKJBEBIDNHE-UHFFFAOYSA-N |
| **Drug Information:** | Drug Indication \| Therapeutic Uses \| Clinical Trials \| FDA Orange Book \| FDA UNII |

"PubChem, released in 2004, provides information on the biological activities of small molecules."

https://pubchem.ncbi.nlm.nih.gov/

# Cheminformatics: Chemical patterns



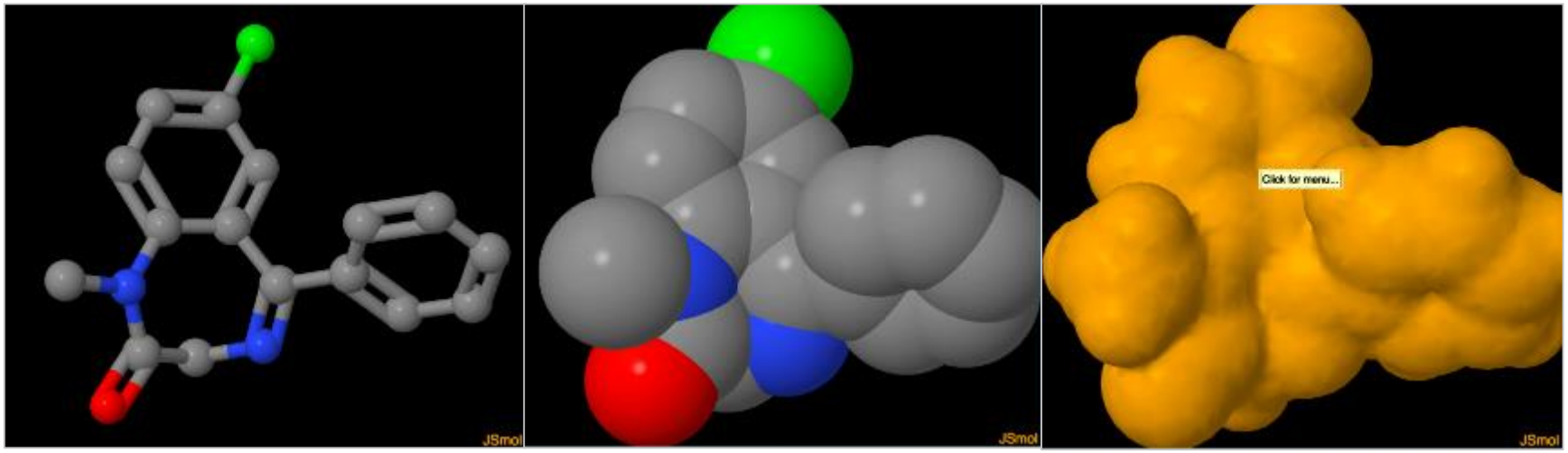

Imidazole pattern subgraph matching with SMARTS.

[#6]12~[#7]~*~[#7]~[#6]1~*~*~*~*2

SMARTS = Smiles Arbitrary Target Specification (akin to regexp)

http://pasilla.health.unm.edu/tomcat/biocomp/depict

# Cheminformatics: Molecules in 3D




ball and stick spacefill VDW surface

- Drug bioactivity is related to 3D properties and interactions.

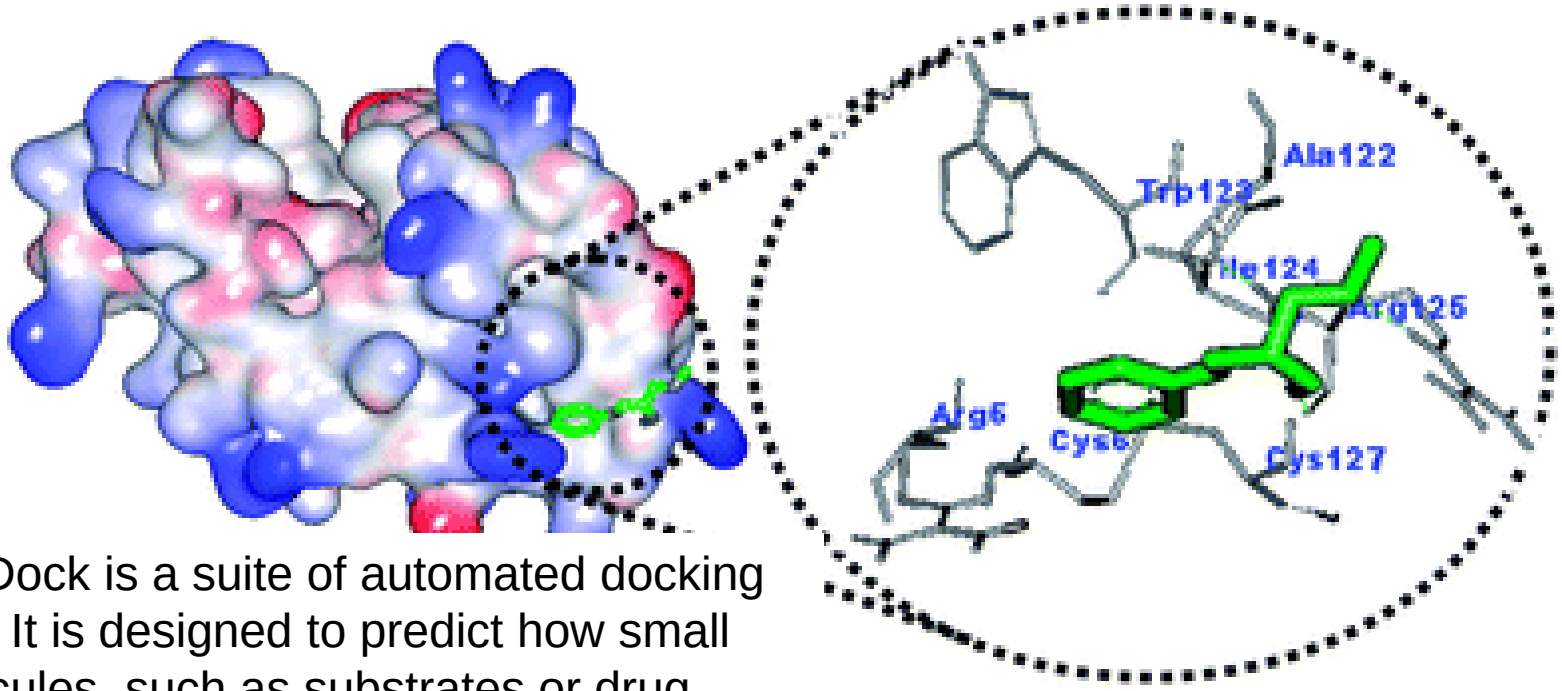
Display by JSMol http://jmol.sourceforge.net/

Conformation by RDKit, http://www.rdkit.org/

# Cheminformatics: Glossary

| | |
|---|---|
| small molecule | not a protein or nucleic acid (macromolecule) |
| pharmacology | study of uses, effects and modes of action of drugs |
| bioactivity | effect of a chemical on a biological target |
| IC50, EC50 | measures of bioactive potency |
| structure-activity relationship | bioactivity vs. structural features within a chemical series |
| molecular descriptor | measured, theoretical or computed property |
| substructure searching | searching for molecular subgraphs |
| similarity searching | searching for similar molecules |
| pharmacophore | 3D features pattern interacting with a binding site |
| molecular modeling | computational models, 2D, 3D and statistical, for explaining, visualizing and predicting |
| computational chemistry | connotes physics-based modeling, molecular mechanics and dynamics, etc. |
| virtual screening | *in silico* search (assay) for bioactive candidates |

# Cheminformatics:
# Modeling protein-ligand binding




AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.

AutoDock is one of several tools for docking and modeling protein-ligand binding.

http://autodock.scripps.edu/

# Cheminformatics:
# Drug-likeness and lead-likeness

The Lipinski Rule of Five (Ro5) was formulated by Christopher A. Lipinski and published in 1997, and is intended to guide library selection for drug discovery. The following molecular descriptors are used:

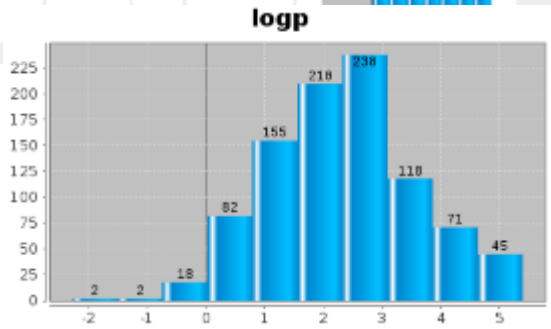
- mwt - avg molecular weight
- hbd - hydrogen bond donor count
- hba - hydrogen bond acceptor count
- LogP - octanol-water partition coefficient

The Ro5 specifies that compounds should violate no more than one of the following criteria:

- mwt <= 500
- hbd <= 5
- hba <= 10
- LogP < 5.0

| property | min | max | mean | median | std | violations | |
|---|---|---|---|---|---|---|---|
| MWT | 96.1 | 457.4 | 241.4 | 238.2 | 62.6 | 0 |   |
| HBA | 0 | 8 | 1.5 | 1 | 1.3 | 0 |  |
| HBD | 0 | 6 | 1.2 | 1 | 1.2 | 2 |  |
| LOGP | -2.6 | 7.1 | 2.3 | 2.3 | 1.3 | 78 |   |

| Ro5 violations | #mols |
|---|---|
| 0 | 921 |
| 1 | 78 |
| 2 | 1 |
| 3 | 0 |
| 4 | 0 |




http://pasilla.health.unm.edu/tomcat/biocomp/ro5

# Cheminformatics: Drug discovery applications

- Databases of relevant compounds and metadata
- 2D virtual screening (substructure, similarity)
- 3D virtual screening (shape, pharmacophores)
- Property prediction (e.g. solubility, logP)
- ADMET prediction (absorption, distribution, metabolism, excretion, toxicity)
- Clustering (unsupervised ML)
- Quantitative structure-activity relationship modeling (QSAR)
- Modeling/predicting protein-ligand binding
- *De novo* design (e.g. GAs)

# Cheminformatics: Take home messages

- Understanding the structure and properties of chemicals enables us to explore biological effects and discover new drugs.
- Chemical databases provide access to a wealth of biomedical and drug discovery data, searchable via cheminformatics.
- In silico predictive modeling accelerates and economizes early stage drug discovery.