What is a drug?

According to the Food, Drug, and Cosmetic Act (1): a substance recognized in an official pharmacopoeia or formulary (2): a substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease (3): a substance other than food intended to affect the structure or function of the body (4): a substance intended for use as a component of a medicine but not a device or a component, part, or accessory of a device

http://www.merriam-webster.com/dictionary/drug
What is a drug?

According to the Food, Drug, and Cosmetic Act (1) : a substance recognized in an official pharmacopoeia or formulary (2) : a substance intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease (3) : a substance other than food intended to affect the structure or function of the body (4) : a substance intended for use as a component of a medicine but not a device or a component, part, or accessory of a device

http://www.merriam-webster.com/dictionary/drug

A small molecule intended to affect the structure/function of macromolecules
COX-1
THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS

PHASE I
IND SUBMITTED

PHASE II
NUMBER OF VOLUNTEERS
TENS

PHASE III
HUNDREDS

PHASE IV
THOUSANDS

FDA APPROVal
POST-APPROVAL RESEARCH & MONITORING

POTENTIAL NEW MEDICINES

1 FDA-APPROVED MEDICINE

$2.6 BILLION

Source: Pharmaceutical Research and Manufacturers of America (http://phrma.org)
THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS

PHASE I
IND SUBMITTED
PHASE II
NDA/BLA SUBMITTED
PHASE III
FDA APPROVAL
PHASE IV
TENS HUNDREDS THOUSANDS
NUMBER OF VOLUNTEERS

POTENTIAL NEW MEDICINES

Source: Pharmaceutical Research and Manufacturers of America (http://phrma.org)

$2.6 BILLION
1. Does the compound do what you want it to?
2. Does the compound not do what you don’t want it to?
3. Is what you want it to do the right thing?
Drug Discovery

Omics

Target Identification

Screening

Lead Identification

Lead Optimization

Compounds

Hits

Leads

Clinical Candidates

Cost
Computational Drug Discovery

Omics
- Target Identification

Virtual Screen
- Compounds
- Hits

Lead Identification
- Leads

Modeling
- Lead Optimization

Clinical Candidates
- Cost
Kinds of Virtual Screening

ADMET

Ligand Based
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based
- dock and score
- simulation
ADMET

Absorption
Distribution
Metabolism
Excretion
Toxicity

Will this be a usable drug?

Screening for ADMET:
Cytochrome P450 interaction
Lipinski’s Rule of Five
QSPR: Quantitative Structure Property Relationship
Kinds of Virtual Screening

ADMET

Ligand Based
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based
- dock and score
Ligand Based: Similarity

Fingerprint Methods

- map molecules to a descriptor space:
  1D: molecule weight, #h-bonds, etc.
  2D: paths, bond distances between atom-pairs

- similarity is “distance” between descriptors
- for bit vectors, Tanimoto distance used

\( T(A, B) = \frac{|A \cap B|}{|A \cup B|} \)
Topological Fingerprints

Daylight/FP2 Fingerprints
- all paths up to 7 bonds long
- each path corresponds to bit position (hashing)
- fast similarity checking (Tanimoto)
Topological Fingerprints

Daylight/FP2 Fingerprints

- all paths up to 7 bonds long
- each path corresponds to bit position (hashing)
- fast similarity checking (Tanimoto)
Ligand Based: Similarity

Superposition Methods
- compute “overlap” between molecules
- consider shape, electrostatics, pharmacophores

http://www.cresset-group.com/
Representing Compounds

Conformations

A single compound has many different shapes

Choices: Store sampling of explicit conformations, search for a good conformation, ignore conformations (2D only)
# Ligand Based: QSAR

Quantitative Structure/Activity Relationships

## Properties

<table>
<thead>
<tr>
<th>Cmpd Number</th>
<th>Cmpd Name</th>
<th>X</th>
<th>Log EC$_{50}$</th>
<th>n</th>
<th>Calculated Log EC$_{50}$</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6a</td>
<td>H</td>
<td>1.07</td>
<td>0</td>
<td>0.79</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td>6b</td>
<td>Cl</td>
<td>0.09</td>
<td>0.71</td>
<td>0.21</td>
<td>-0.12</td>
</tr>
<tr>
<td>3</td>
<td>6d</td>
<td>NO$_2$</td>
<td>0.66</td>
<td>-0.28</td>
<td>1.02</td>
<td>-0.36</td>
</tr>
<tr>
<td>4</td>
<td>6e</td>
<td>CN</td>
<td>1.42</td>
<td>-0.57</td>
<td>1.26</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>6f</td>
<td>C$_6$H$_5$</td>
<td>-0.62</td>
<td>1.96</td>
<td>-0.81</td>
<td>0.19</td>
</tr>
<tr>
<td>6</td>
<td>6g</td>
<td>N(CH$_3$)$_2$</td>
<td>0.64</td>
<td>0.18</td>
<td>0.65</td>
<td>-0.01</td>
</tr>
<tr>
<td>7</td>
<td>6h</td>
<td>I</td>
<td>-0.46</td>
<td>1.12</td>
<td>-0.12</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

Biological Activity = Learned linear function of properties

3D-QSAR: includes geometric/structural properties
Ligand/Receptor Based: Pharmacophore

**Pharmacophore:**

IUPAC: The ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.

**Common Features:**
aromatic ring
hydrophobic area
positive ionizable
negative ionizable
hydrogen bond donor
hydrogen bond acceptor
Pharmacophore Features

- Aromatic
- Hydrogen Donor
- Hydrogen Acceptor
- Positive
- Hydrophobic
- Negative
- Aromatic
**Charge-Charge**

\[ |\vec{F}_{Q-q}| = |\vec{F}_{q-Q}| = k \frac{|q| \times |Q|}{r^2} \]
Charge-Charge

2QWK.pdb

Salt Bridge

Inhibitor of the influenza virus neuraminidase (antiviral agent)
Hydrogen Bond
Hydrogen Bond
Hydrogen Bond
Hydrogen Bond
Hydrogen Bond

Distance:
D-A: 2.5Å – 3.5Å (4.0Å?)
H-A: 1.5Å – 2.5Å

Angle:
Depends on context

Turkey Ovomucoid Inhibitor
Hydrophobic
Hydrophobic

MDM2 (over expressed in >50% of cancers) down-regulates p53 (guardian of the genome)
Hydrophobic

MDM2 (over expressed in >50% of cancers) down-regulates p53 (guardian of the genome)
Hydrophobic

p53/MDM2

MDM2 (over expressed in >50% of cancers) down-regulates p53 (guardian of the genome)
Aromatic

http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch11/benzene-mo.jpg
Aromatic

http://www.chem.ucalgary.ca/courses/351/Carey5th/Ch11/benzene-mo.jpg
Aromatic

Rings offset
Interplanar distance: 3.3-3.8Å

Human liver glycogen phosphorylase a complexed with caffeine
http://pharmit.csb.pitt.edu
Kinds of Virtual Screening

ADMET

Ligand Based
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based
- dock and score
Pharmacophores Aren’t Enough
Pharmacophores Aren’t Enough

.2µM

50µM

n.i.
Docking

Determine the **conformation** and **pose** of a ligand at a docking site

Challenge is to find conformation and pose with the best **score**
Two Phase Docking

1. Global Pose Estimation

Stochastic

2. Local Refinement

Minimization
Two Phase Docking

1. Global Pose Estimation

Stochastic

Minimization
Scoring Goals

Affinity Prediction
  -how well does it bind?

Inactive/Active Discrimination
  -does it bind?

Pose Prediction
  -how does it bind?
Scoring Goals

Affinity Prediction
   - how well does it bind?

Inactive/Active Discrimination
   - does it bind?

Pose Prediction
   - how does it bind?

Speed
Scoring Goals

Affinity Prediction
- how well does it bind?

Inactive/Active Discrimination
- does it bind?

Pose Prediction
- how does it bind?

Speed

Approximations:
- Rigid or semi-rigid receptor
- Implicit water model
Scoring Types

Force-field based
- inter- and intra- molecular forces
  - van der Waals, electrostatic, torsional

Empirical
- parameterized function is fit to binding energy data

Knowledge based
- scoring function based on known structure, not physical principles

Consensus
Force Field: Dock 4.0

\[ E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left( \frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{Dr_{ij}} \right) \]

Coulomb’s Law
q: partial charges
D: dielectric constant

van der Waals
a = 12, b = 6
Lennard-Jones potential

\[ D = \begin{pmatrix} -0.41 & 0.205 \\ 0.205 & 205 \end{pmatrix} \]
Empirical: AutoDock Vina

\[
gauss_1(d) = w_{\text{guass}_1} e^{-(d/0.5)^2} \\
gauss_2(d) = w_{\text{guass}_2} e^{-((d-3)/2)^2} \\
\text{repulsion}(d) = \begin{cases} 
  w_{\text{repulsion}} d^2 & d < 0 \\
  0 & d \geq 0
\end{cases}
\]

\[
\text{hydrophobic}(d) = \begin{cases} 
  w_{\text{hydrophobic}} & d < 0.5 \\
  0 & d \geq 1.5 \\
  w_{\text{hydrophobic}} (1.5 - d) & \text{otherwise}
\end{cases}
\]

\[
h\text{bond}(d) = \begin{cases} 
  w_{\text{hbond}} & d < -0.7 \\
  0 & d > 0 \\
  w_{\text{hbond}} (-\frac{10}{7} d) & \text{otherwise}
\end{cases}
\]

<table>
<thead>
<tr>
<th>Weight</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0356</td>
<td>gauss_1</td>
</tr>
<tr>
<td>-0.00516</td>
<td>gauss_2</td>
</tr>
<tr>
<td>0.810</td>
<td>Repulsion</td>
</tr>
<tr>
<td>-0.0251</td>
<td>Hydrophobic</td>
</tr>
<tr>
<td>-0.587</td>
<td>Hydrogen Bonding</td>
</tr>
<tr>
<td>0.0585</td>
<td>N_{tot}</td>
</tr>
</tbody>
</table>
Knowledge Based: RF-Score

Pairwise Distance Counts (<12Å)

<table>
<thead>
<tr>
<th>Protein</th>
<th>C</th>
<th>N</th>
<th>O</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Br</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ligand
Can we do better?

Accurate pose prediction, binding discrimination, and affinity prediction without sacrificing performance?
Can we do better?

Accurate pose prediction, binding discrimination, and affinity prediction without sacrificing performance?

Key Idea: Leverage “big data”

231,655,275 bioactivities in PubCher
125,526 structures in the PDB
16,179 annotated complexes in PDBbind
Machine Learning

Features $X \rightarrow$ Model $\rightarrow y$ Prediction
Neural Networks

\[ \text{output} = \sigma \left( \sum_i w_i x_i + b \right) \]
Neural Networks

The universal approximation theorem states that, under reasonable assumptions, a feedforward neural network with a finite number of nodes can approximate any continuous function to within a given error over a bounded input domain.
Deep Learning
Deep Learning

\[ \delta^l = \left( (w^{l+1})^T \delta^{l+1} \right) \odot \sigma'(z^l) \]

\[ \frac{\partial L}{\partial w^l_{jk}} = \alpha_k^{l-1} \delta^l_j \text{ and } \frac{\partial L}{\partial b^l_j} = \delta^l_j \]
Image Recognition

![Image Recognition Diagram](https://devblogs.nvidia.com)

- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

**ILSVRC top-5 error on ImageNet**

- 2010
- 2011
- 2012
- 2013
- 2014
- Human
- ArXiv 2015

Convolutional Neural Networks

https://devblogs.nvidia.com
Convolutional Neural Networks

Convolution

Convolution Feature Maps

Fully Connected

Traditional NN

Dog: 0.99
Cat: 0.02

Convolution

Fully-connected

weight 1
weight 2
weight 3
weight 4
weight 5
CNNs for Protein-Ligand Scoring

Pose Prediction
Binding Discrimination
Affinity Prediction
Protein-Ligand Representation

(R,G,B) pixel
Protein-Ligand Representation

(R,G,B) pixel → (Carbon, Nitrogen, Oxygen,...) voxel

The only parameters for this representation are the choice of grid resolution, atom density, and atom types.
Model Optimization

Atom Types
- Vina (34)
- element-only (18)
- ligand-protein (2)

Atom Density Type
- Boolean
- Gaussian

Radius Multiple

Resolution

Pooling

Depth

Width

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers

Fully Connected Layers
Pose Prediction (PDBbind)

inter-target ranking

intra-target ranking
Affinity Prediction

**Data**

- unit1_pool
- unit1_conv1: 32 x 24^3
- unit2_pool
- unit2_conv1: 64 x 12^3
- unit3_pool
- unit3_conv1: 128 x 6^3

**Labels**

- output_fc
- output_af

**Loss**

- label

**Output**

- output
- loss

**Affinity**

- 48^3

**Comparison**

- CNN: R=0.687, RMS=2.186
- Vina: R=0.497, RMS=8.956
Beyond Scoring

data
48^3

unit1_pool
32 x 24^3

unit2_pool
64 x 12^3

unit3_pool
128 x 6^3

output_fc
2

output

loss
Beyond Scoring

\[
\frac{\partial L}{\partial w_{jk}} = a^{l-1}_k \delta^l_j \quad \text{and} \quad \frac{\partial L}{\partial b^l_j} = \delta^l_j
\]
Beyond Scoring
Beyond Scoring

Deep Dreams

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
https://deepdreamgenerator.com/#gallery
Beyond Scoring

2Q89

More Oxygen Here

Less Oxygen Here
Beyond Scoring

2Q89

More Oxygen Here

Less Oxygen Here

$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \left( \frac{\partial L}{\partial G_i} \right) \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}$$

unit1_pool

label
Related Work

MoleculLeNet: A continuous-filter convolutional neural network for modeling quantum interactions
Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, Klaus-Robert Müller
(Submitted on 26 Jan 2017)

Automatic chemical design using a data-driven continuous representation of molecules
Rafael Gómez-Bombarelli, David Duvenaud, José Miguel Hernández-Lobato, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, Alan Aspuru-Guzik
(Submitted on 7 Oct 2016; v1), last revised 6 Jan 2017 (this version, v2)

AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery
Izhak Wallach, Michael Dzamba, Abraham Heifets
(Submitted on 19 Oct 2015)

ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost
Justin S. Smith, Oleksandr Isayev, Adrian E. Roitberg
(Submitted on 27 Oct 2015; v1), last revised 6 Feb 2017 (this version, v4)

Convolutional Networks on Graphs for Learning Molecular Fingerprints
David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alan Aspuru-Guzik, Ryan P. Adams
(Submitted on 30 Sep 2015; v1), last revised 3 Nov 2015 (this version, v2)

Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity
Joseph Gomes, Bharath Ramsundar, Evan N. Feinberg, Vijay S. Pande
(Submitted on 30 Mar 2017)

Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules
Alessandro Lucchi1,2, 3, Gianluca Pola2, and Pierre Baldi2
1 School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland
2 Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States
DOI: 10.1021/ji500461y
Publication Date (Web): June 24, 2015

Low Data Drug Discovery with One-shot Learning
Han Altae-Tran, Bharath Ramsundar, Aneesh S. Pappu, Vijay Pande
(Submitted on 19 Nov 2016)

Massively Multitask Networks for Drug Discovery
Bharath Ramsundar, Steven Kearnes, Patrick Riley, Dale Webster, David Konerding, Vijay Pande
(Submitted on 6 Feb 2017)

Protein–Ligand Scoring with Convolutional Neural Networks
Matthew Regoes1, Joshua Hochhauser, Eldad Idron2, Joydlyn Sunerness, and David Ryan Koes1
1 Department of Neuroscience, 2 Department of Computer Science, 3 Department of Biological Sciences, and Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States
4 Department of Computer Science, The College of New Jersey, Ewing, New Jersey 08628, United States
DOI: 10.1021/acs.jcim.7b00140
Publication Date (Web): April 13, 2017
Copyright © 2017 American Chemical Society
Key Concepts

Ligand-Based Virtual Screening
Identifying new active compounds based on similarity to known active compounds

Pharmacophore
A spatial arrangement of molecular features essential for biological activity - hydrogen bonding, hydrophobic, charged, etc.

Docking
Predict the position, pose and affinity of a molecule using the receptor structure

Scoring
force field … empirical … knowledge based
Acknowledgements

Group Members
Jocelyn Sunseri
Alec Helbling
Matt Ragoza
Josh Hochuli
Pulkit Mittal
Aaron Zheng
Sharanya Bandla
Faiha Khan
Lily Turner
Dale Erikson

Matt Ragoza
Josh Hochuli
Elisa Idrobo
Jocelyn Sunseri
@david_koes
github.com/gnina
http://bits.csb.pitt.edu