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

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

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

High Throughput Screening
Drug Discovery

High Throughput Screening

- Target and hit identification
- Hit refinement
- Lead refinement
- Regulatory development

- VHS
- LSI
- CCS

- Hit generation
- Lead generation
- Lead optimization

- Clinical candidate

- Handling of data points
- Generality
- Applications

- Evaluation of results
Is it true FDA is approving fewer new drugs lately? FDA sometimes hears concerns from the public that the Agency is not approving enough new drugs. Actually, the number of new drugs FDA approves each year has remained relatively steady over time. For instance, in 2010 FDA’s Center for Drug Evaluation and Research (CDER) approved 21 novel new drugs known as New Molecular Entities (NMEs). The chart below shows that this number is similar to NME approvals over the past five years.

<table>
<thead>
<tr>
<th>Calendar Year</th>
<th>NMEs Approved</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>22</td>
</tr>
<tr>
<td>2007</td>
<td>18</td>
</tr>
<tr>
<td>2008</td>
<td>24</td>
</tr>
<tr>
<td>2009</td>
<td>26</td>
</tr>
<tr>
<td>2010</td>
<td>21</td>
</tr>
</tbody>
</table>

FDAs Center for Drug Evaluation and Research (CDER) approved 21 New Molecular Entities (NMEs) in 2010. This is within an 18-26 range approved for the last several years.

In fact, 21 NME approvals in one year is about in line for yearly approvals for the past decade. The chart below shows a ten year period from 2001 through 2010, in which FDA averaged about 23 NME approvals per year (22.9).

<table>
<thead>
<tr>
<th>Calendar Year</th>
<th>NMEs Approved</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>24</td>
</tr>
<tr>
<td>2002</td>
<td>17</td>
</tr>
<tr>
<td>2003</td>
<td>21</td>
</tr>
<tr>
<td>2004</td>
<td>36</td>
</tr>
<tr>
<td>2005</td>
<td>20</td>
</tr>
<tr>
<td>2006</td>
<td>22</td>
</tr>
<tr>
<td>2007</td>
<td>18</td>
</tr>
<tr>
<td>2008</td>
<td>24</td>
</tr>
<tr>
<td>2009</td>
<td>26</td>
</tr>
<tr>
<td>2010</td>
<td>21</td>
</tr>
</tbody>
</table>

Since 2001 CDER has averaged slightly fewer than 23 NME approvals per year (22.9), similar to the 21 approved in 2010.

While 21 approvals in 2010 is typical of previous years, an increase in approvals would be the ideal scenario. A good part of the reason for this flat approval rate over time is that drug companies are not filing as many applications with FDA for new drug approval as they have in the past. The chart below shows that over the past five years, applications filed with FDA for NMEs have not been increasing. If applications do not increase, the likelihood of approvals increasing is reduced.

<table>
<thead>
<tr>
<th>Calendar Year</th>
<th>Applications Filed</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>26</td>
</tr>
<tr>
<td>2007</td>
<td>35</td>
</tr>
<tr>
<td>2008</td>
<td>34</td>
</tr>
<tr>
<td>2009</td>
<td>37</td>
</tr>
<tr>
<td>2010</td>
<td>23</td>
</tr>
</tbody>
</table>

NME applications to CDER are not increasing. If the number of applications does not increase, CDER does not expect to see much of a year-to-year increase in approvals.

The trend towards fewer filings for NMEs extends beyond the past five years. The chart below shows that except for 2002, in which 22 applications were filed with FDA for new NMEs, the 23 applications for NMEs in 2010 is the lowest number in over 15 years.

<table>
<thead>
<tr>
<th>Calendar Year</th>
<th>Applications Filed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>45</td>
</tr>
<tr>
<td>1997</td>
<td>41</td>
</tr>
<tr>
<td>1998</td>
<td>43</td>
</tr>
<tr>
<td>1999</td>
<td>36</td>
</tr>
<tr>
<td>2000</td>
<td>26</td>
</tr>
<tr>
<td>2001</td>
<td>30</td>
</tr>
<tr>
<td>2002</td>
<td>22</td>
</tr>
<tr>
<td>2003</td>
<td>26</td>
</tr>
<tr>
<td>2004</td>
<td>32</td>
</tr>
<tr>
<td>2005</td>
<td>38</td>
</tr>
<tr>
<td>2006</td>
<td>26</td>
</tr>
<tr>
<td>2007</td>
<td>35</td>
</tr>
<tr>
<td>2008</td>
<td>34</td>
</tr>
<tr>
<td>2009</td>
<td>37</td>
</tr>
<tr>
<td>2010</td>
<td>23</td>
</tr>
</tbody>
</table>

In fact, except for 2002, the 23 NME applications to CDER filed in 2010 is the lowest number filed in more than 15 years.

FDA has been taking action for some time to help drive new drug development and increase applications for novel new products. In 2004, noting a slowdown, FDA launched its Critical Path Initiative, FDA’s national strategy to help advance pharmaceutical innovation.

Our long-term efforts are showing positive signs and FDA will continue to support the scientific community to advance new drug development. For more information on FDA’s Critical Path Initiative visit:

http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm

Computational Drug Discovery

Virtual Screening

existing libraries

new compounds

Hit generation → Hits → Lead generation → Leads → Lead optimization → Clinical candidate
Computational Drug Discovery

Virtual Screening

existing libraries

new compounds

Hit generation

Hits

Lead generation

Leads

Lead optimization

Clinical candidate

\[
\text{HO}_2\text{C}-\text{R}_1\text{C}-\text{NH}_2 + \text{R}_2\text{CHO} + \text{R}_3\text{NC} \rightarrow \text{R}_1\text{N}=\text{R}_2\text{O} \text{C} = \text{N} \text{H} - \text{R}_3
\]
Computational Drug Discovery

Virtual Screening

existing libraries

new compounds

Hit generation

Hits

Lead generation

Lead optimization

Clinical candidate

\[ \text{HO}_2\text{C} \_ {\text{R}_1} \_ \text{NH}_2 + \text{R}_2\text{CHO} + \text{R}_3\text{NC} \rightarrow \text{R}_1 \_ \text{R}_2 \_ \text{N} = \text{O} \_ \text{R}_3 \]
Kinds of Virtual Screening

ADMET

Ligand Based
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based
- dock and score
- simulation
  MM/GBSA, MM/PBSA, thermodynamic integration, free energy perturbation, Jarzynski, umbrella sampling, Monte Carlo, weighted ensemble, metadynamics…
Kinds of Virtual Screening

ADMET

Ligand Based
- similarity to known binder
- QSAR
- pharmacophore

Receptor Based
- dock and score
- simulation
  MM/GBSA, MM/PBSA, thermodynamic integration, free energy perturbation, Jarzynski, umbrella sampling, Monte Carlo, weighted ensemble, metadynamics…

Not going to cover today 😞
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)

fingerprint:

O=C-N-C

O=C-O

O=N-O

N-C-C-C-C-C=0

C=C=C-C
Topological Fingerprints

ECFP4
- all substructures with diameter 4 around every atom
Ligand Based: QSAR

Quantitative Structure/Activity Relationships

\[ f(\vec{x}) = w_1 \vec{x}_1 + w_2 \vec{x}_2 + w_3 \vec{x}_3 + \ldots + b \]

Identifying Selectivity from Substructures

| ki unique SERT | 8.13 |
| ki all SERT | 9.65 |
| ki SERT | 9.69 |
| ki no SSRIS DAT | 5.83 |
| ki all DAT | 6.19 |
| ki DAT | 6.21 |

Karla Robles, Jerry Madura, David Koes
University of Pittsburgh, Duquesne University

QSAR Modeling: Predicting Ligand Binding Affinities and Substructures Key in Binding to SERT
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: 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
Charge-Charge

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

\[ \vec{F}_{q-Q} = \vec{F}_{Q-q} \]

\[ \vec{F}_{q-Q} = -\vec{F}_{Q-q} \]
Charge-Charge

2QWK.pdb

Salt Bridge
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

1HJA.pdb
Hydrophobic
Hydrophobic
Hydrophobic

p53/MDM2
Hydrophobic

p53/MDM2
Aromatic
Aromatic
**Aromatic**

Rings offset
**Interplanar distance: 3.3-3.8Å**
Pharmacophore Features

- Hydrogen Acceptor
- Hydrogen Donor
- Positive
- Hydrophobic
- Negative
- Aromatic
Pharmacophore

A spatial arrangement of molecular features essential for biological activity


Hydrophobic Features

Hydrogen Acceptor Feature

Hydrogen Donor Feature

Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search

Pharmer
Pharmer

Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search
Pharmer

Efficient and Exact Pharmacophore Search
Efficient and Exact Pharmacophore Search

Pharmer
Efficient and Exact Pharmacophore Search

![Graph showing the relationship between Tolerance Sphere Radii Multiplier and Time (s) for MOE HSP90, MOE FXIa, Pharmer HSP90, and Pharmer FXIa.](image-url)
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
2. Local Refinement

Stochastic

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 Scoring

<table>
<thead>
<tr>
<th></th>
<th>Protein–ligand</th>
<th>Internal ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G-Score</strong></td>
<td>$E_{vdw} + E_{H-bond} = \sum \sum_{\text{prot lig}} \left[ \left( A_{ij} - B_{ij} \right) + (E_{da} + E_{ww}) - \left( E_{dw} + E_{ww} \right) \right]$</td>
<td>$E_{vdw} + E_{\text{torsion}} = \sum_{\text{lig}} \left[ \left( C_{ij} + D_{ij} \right) + \sum \frac{1}{2} \left( 1 + \frac{1}{2} \cos(n</td>
</tr>
<tr>
<td><strong>D-Score</strong></td>
<td>$E_{vdw} + E_{\text{electrostatic}} = \sum \sum_{\text{prot lig}} \left[ \left( A_{ij} + B_{ij} \right) + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)} \right]$</td>
<td>$E_{vdw} + E_{\text{electrostatic}} = \sum_{\text{lig}} \left[ \left( A_{ij} + B_{ij} \right) + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)} \right]$ + optional $E_{H-bond}$</td>
</tr>
<tr>
<td><strong>Gold</strong></td>
<td>$E_{vdw} + E_{H-bond} + E_{\text{electrostatic}} = \sum \sum_{\text{prot lig}} \left[ \left( A_{ij} - B_{ij} \right) + E(t) \left( C_{ij} - D_{ij} \right) \right] + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)}$</td>
<td>$E_{vdw} + E_{H-bond} + E_{\text{electrostatic}} = \sum_{\text{lig}} \left[ \left( A_{ij} - B_{ij} \right) + E(t) \left( C_{ij} - D_{ij} \right) \right] + 332.0 \ \frac{q_i q_j}{4(d_{ij}^a d_{ij}^b)}$</td>
</tr>
<tr>
<td><strong>AutoDock</strong></td>
<td>$E_{vdw} + E_{H-bond} + E_{\text{electrostatic}} = \sum \sum_{\text{prot lig}} \left[ \left( A_{ij} - B_{ij} \right) + E(t) \left( C_{ij} - D_{ij} \right) \right] + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)}$</td>
<td>$E_{vdw} + E_{H-bond} + E_{\text{electrostatic}} = \sum_{\text{lig}} \left[ \left( A_{ij} - B_{ij} \right) + E(t) \left( C_{ij} - D_{ij} \right) \right] + 332.0 \ \frac{q_i q_j}{4(d_{ij}^a d_{ij}^b)}$</td>
</tr>
<tr>
<td><strong>DOCK (v4.0)</strong></td>
<td>$E_{vdw} + E_{\text{electrostatic}} = \sum \sum_{\text{prot lig}} \left[ \left( A_{ij} + B_{ij} \right) + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)} \right]$</td>
<td>$E_{vdw} + E_{\text{electrostatic}} = \sum_{\text{lig}} \left[ \left( A_{ij} + B_{ij} \right) + 332.0 \ \frac{q_i q_j}{\epsilon(d_{ij}^a d_{ij}^b)} \right]$</td>
</tr>
</tbody>
</table>
Dock 4.0

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

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

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

<table>
<thead>
<tr>
<th>Functional form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LUDI</strong></td>
</tr>
<tr>
<td>[ \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{hydrophobic}} \sum_{\text{hydrophobic}} A_{\text{hydrophobic}} + \Delta G_{\text{rotor}} N_{\text{rotor}} + \Delta G_0 ]</td>
</tr>
<tr>
<td>( A_{\text{hydrophobic}} = \text{molecular surface area} )</td>
</tr>
<tr>
<td><strong>F-Score</strong></td>
</tr>
<tr>
<td>[ \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{aromatic}} \sum_{\text{aromatic}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{contact}} \sum_{\text{contact}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{rotor}} N_{\text{rotor}} + \Delta G_0 ]</td>
</tr>
<tr>
<td><strong>Chem-Score</strong></td>
</tr>
<tr>
<td>[ \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{metal}} \sum_{\text{metal}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipo}} \sum_{\text{lipo}} f(\Delta R) + \Delta G_{\text{rotor}} \sum_{\text{rotor}} f(P_{nl}, P'_{nl}) + \Delta G_0 ]</td>
</tr>
</tbody>
</table>
### Empirical Scoring

**Functional form**

<table>
<thead>
<tr>
<th>Function</th>
<th>Expression</th>
</tr>
</thead>
</table>
| LUDI         | \( \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic}} f(\Delta R, \Delta \alpha) + \)  
|              | \( \Delta G_{\text{hydrophobic}} \sum_{\text{hydrophobic}} A_{\text{hydrophobic}} \)  
|              | + \( \Delta G_{\text{rotor}} N_{\text{rotor}} + \Delta G_0 \)  
|              | \( A_{\text{hydrophobic}} = \text{molecular surface area} \) |
| F-Score      | \( \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{aromatic}} \sum_{\text{aromatic}} f(\Delta R, \Delta \alpha) \)  
|              | + \( \Delta G_{\text{contact}} \sum_{\text{contact}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{rotor}} N_{\text{rotor}} + \Delta G_0 \) |
| Chem-Score   | \( \Delta G_{\text{bind}} = \Delta G_{\text{H-bond}} \sum_{\text{H-bond}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{metal}} \sum_{\text{metal}} f(\Delta R, \Delta \alpha) \)  
|              | + \( \Delta G_{\text{lipo}} \sum_{\text{lipo}} f(\Delta R) + \Delta G_{\text{rotor}} \sum_{\text{rotor}} f(P_{nl}, P'_{nl}) + \Delta G_0 \) |
AutoDock Vina

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

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

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

## Functional form

<table>
<thead>
<tr>
<th>PMF</th>
<th>Parametrized pairwise potential PMF score:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[ PMF = \sum_{prot} \sum_{lig} A_{ij}(d_{ij}) A_{ij}(d_{ij}) = -k_B T \ln \left( f_{Vol_corr}^j(r) \frac{\rho_{seg}^j(r)}{\rho_{bulk}^{ij}} \right) ]</td>
</tr>
<tr>
<td></td>
<td>where ( k_B ) is the Boltzmann constant, ( f_{Vol_corr}^j(r) ) is a ligand volume correction factor and ( \frac{\rho_{seg}^j(r)}{\rho_{bulk}^{ij}} ) indicates a radial distribution function for a protein atom ( i ) and a ligand atom ( j ).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DrugScore</th>
<th>( \Delta W = \gamma \sum_{prot} \sum_{lig} \Delta W_{ij}(r) + (1 - \gamma) \times \left[ \sum_{lig} \Delta W_i(SAS, SAS_0) + \sum_{prot} \Delta W_j(SAS, SAS_0) \right] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v1.2)</td>
<td>( SAS = ) Solvent accessible surface area terms, ( W_{ij} = ) distance dependent pairwise potential</td>
</tr>
</tbody>
</table>

| SMoG | \( G = \sum_{ij} g_{ij} \Delta_{ij} ; \quad \Delta_{ij} = \begin{cases} 0 & (i, j \text{ more than 5 Å}) \\ 1 & (i, j \text{ within 5 Å}) \end{cases} ; \quad g_{ij} = -kT \log \left( \frac{p_{ij}}{\bar{p}} \right) ; \) |
|       | \( p_{ij} \) and \( \bar{p} \) are interatomic and averaged interactomic interactions |


RF-Score

Pairwise Distance Counts (<12Å)

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

Random Forest
### RF-Score Output

**RMSE = 1.58**

![Graph showing predicted vs. measured binding affinities](image)

<table>
<thead>
<tr>
<th>Scoring function</th>
<th>$R$</th>
<th>$Rs$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF-Score</td>
<td>0.776</td>
<td>0.762</td>
<td>1.58</td>
</tr>
<tr>
<td>X-Score::HMScore</td>
<td>0.644</td>
<td>0.705</td>
<td>1.83</td>
</tr>
<tr>
<td>DrugScore&lt;sup&gt;CSD&lt;/sup&gt;</td>
<td>0.569</td>
<td>0.627</td>
<td>1.96</td>
</tr>
<tr>
<td>SYBYL::ChemScore</td>
<td>0.555</td>
<td>0.585</td>
<td>1.98</td>
</tr>
<tr>
<td>DS::PLP1</td>
<td>0.545</td>
<td>0.588</td>
<td>2</td>
</tr>
<tr>
<td>GOLD::ASP</td>
<td>0.534</td>
<td>0.577</td>
<td>2.02</td>
</tr>
<tr>
<td>SYBYL::G-Score</td>
<td>0.492</td>
<td>0.536</td>
<td>2.08</td>
</tr>
<tr>
<td>DS::LUDI3</td>
<td>0.487</td>
<td>0.478</td>
<td>2.09</td>
</tr>
<tr>
<td>DS::LigScore2</td>
<td>0.464</td>
<td>0.507</td>
<td>2.12</td>
</tr>
<tr>
<td>GlideScore-XP</td>
<td>0.457</td>
<td>0.435</td>
<td>2.14</td>
</tr>
<tr>
<td>DS::PMF</td>
<td>0.445</td>
<td>0.448</td>
<td>2.14</td>
</tr>
<tr>
<td>GOLD::ChemScore</td>
<td>0.441</td>
<td>0.452</td>
<td>2.15</td>
</tr>
<tr>
<td>SYBYL::D-Score</td>
<td>0.392</td>
<td>0.447</td>
<td>2.19</td>
</tr>
<tr>
<td>DS::Jain</td>
<td>0.316</td>
<td>0.346</td>
<td>2.24</td>
</tr>
<tr>
<td>GOLD::GoldScore</td>
<td>0.295</td>
<td>0.322</td>
<td>2.29</td>
</tr>
<tr>
<td>SYBYL::PMF-Score</td>
<td>0.268</td>
<td>0.273</td>
<td>2.29</td>
</tr>
<tr>
<td>SYBYL::F-Score</td>
<td>0.216</td>
<td>0.243</td>
<td>2.35</td>
</tr>
</tbody>
</table>
RF-Score Output

RMSE = 1.58

R = 0.46; RMSE = 1.6
Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

http://www.csardock.org/
Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

\[ R^2 = 0.28 \]
\[ \text{RMSE} = 1.9 \]

http://www.csardock.org/
Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

R² = 0.58
RMSE = 1.51

http://www.csardock.org/
Scoring

Ideally, score would equal affinity – but this is an unsolved problem.

\[ R^2 = 0.58 \]
\[ \text{RMSE} = 1.51 \]

http://www.csardock.org/
Scoring State of the Art

![Graph showing comparison between Vina and Vinardo]

**Pose Prediction**  **Binding Discrimination**  **Affinity Prediction**

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

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

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

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

ILSVRC top-5 error on ImageNet

Convolutional Neural Networks

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

Convolution Feature Maps

Fully Connected Traditional NN

Dog: 0.99
Cat: 0.02

Convolution

weight 1
weight 2
weight 3

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.
Optimized Models

Percent Targets with Low RMSD Top Pose

- Default2017
- Default2018
- HiRes Pose
- HiRes Affinity

Time (ms)

Pearson R

0.50 0.52 0.54 0.56 0.58 0.60

600

500 400 300 200 100

80

75

70

65

Default2018

HiRes Pose

HiRes Affinity

Default2018

48x48x48x28

2x2x2 Ave Pooling

3x3x3 Convolution

1x1x1 Convolution

8x8x8 Ave Pooling

5x5x5 Convolution

4x4x4 Max Pooling

3x3x3 Convolution

Rectified Linear Unit

3x3x3 Convolution

Rectified Linear Unit

3x3x3 Convolution

Rectified Linear Unit

2x2x2 Max Pooling

3x3x3 Convolution

Rectified Linear Unit

3x3x3 Convolution

Rectified Linear Unit

1x1x1 Convolution

Rectified Linear Unit
Pose Results

Redocked Pose

Percent Targets with Low RMSD Top Pose

Default2017  Default2018  HiRes Affinity  HiRes Pose  Vina
Pose Results

Crossdockged Pose

Percent Targets with Low RMSD Top Pose

Default2017  Default2018  HiRes Affinity  HiRes Pose  Vina
Affinity Results

HiRes Affinity
Spearman = 0.598, RMSE = 1.714

Default 2018
Spearman = 0.570, RMSE = 1.686

Vina
Spearman = 0.473, RMSE = 1.887
Affinity Results

HiRes Affinity

Spearman = 0.598, RMSE = 1.714

Vina

Spearman = 0.473, RMSE = 1.887

Default 2018

Spearman Correlation

Clustered       Random

Experiment  Experiment
Beyond Scoring
Beyond Scoring

\[
\frac{\partial L}{\partial w_{jk}^{l}} \Rightarrow a_{k}^{l-1} \delta_{j}^{l} \quad \text{and} \quad \frac{\partial L}{\partial b_{j}^{l}} = \delta_{j}^{l}
\]
Beyond Scoring

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

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

Data: 48 x 3

Unit 1 pool: 32 x 24 x 3

Unit 1 conv: 128 x 6 x 3

Unit 2 pool

Unit 3 conv: 128 x 6 x 3

Output fc:

Loss

Output
Beyond Scoring

Deep Dreams

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

2Q89

More Oxygen Here

Less Oxygen Here
Beyond Scoring

2Q89

More Oxygen Here

Less Oxygen Here

\[
\frac{\partial L}{\partial A} = \sum_{i \in G_A^{data}} \frac{\partial L}{\partial G_i} \frac{\partial G_i}{\partial D} \frac{\partial D}{\partial A}
\]
Minimizing Low RMSD Poses

![Histogram showing the distribution of RMSD changes between poses, with categories labeled as 'better' and 'worse'.]
3AO4
Iterative Refinement

![Graph](image)
Iterative Refinement

![Graph showing iterative refinement process](image-url)
Iterative Refinement

![Graph showing Iterative Refinement](image-url)
Related Work

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

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

**AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure–based Drug Discovery**
Izhar Wallach, Michael Dzamba, Abraham Heifets
(Submitted on 10 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 2016 (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, Alán Aspuru-Guzik, Ryan P. Adams
(Submitted on 30 Sep 2015 (v2), 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 Lusci, Gianluca Pillastriti, and Pierre Baldi
† School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland
‡ Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States
DOI: 10.1021/acs.jcim.5b00187
Publication Date (Web): June 24, 2013

**Low Data Drug Discovery with One-shot Learning**
Han Altae-Tran, Bharath Ramsundar, Aneesh S. Pappu, Vijay Pande
(Submitted on 10 Nov 2016)

**Massively Multitask Networks for Drug Discovery**
Bharath Ramsundar, Steven Kearnes, Patrick Riley, Dale Webster, David Konerding, Vijay Pande
(Submitted on 6 Feb 2015)

**Protein–Ligand Scoring with Convolutional Neural Networks**
Matthew Ragoze, Joshua Hochuli, Elias Idrobo, Jocelyn Sunseri, and David Ryan Koese
† Department of Neurosciences, ‡ Department of Computer Science, † Department of Biological Sciences, and † Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States
‡ Department of Computer Science, The College of New Jersey, Ewing, New Jersey 08628, United States
DOI: 10.1021/acs.jcim.6b00740
Publication Date (Web): April 3, 2017
Copyright © 2017 American Chemical Society
@david_koes
github.com/gnina
http://bits.csb.pitt.edu

OUR FIELD HAS BEEN STRUGGLING WITH THIS PROBLEM FOR YEARS.

STRUGGLE NO MORE! I'M HERE TO SOLVE IT WITH ALGORITHMS!
DEEP LEARNING!

SIX MONTHS LATER:
WOW, THIS PROBLEM IS REALLY HARD.
YOU DON'T SAY.