Structure Based Drug Design

Pose Prediction  Binding Discrimination  Affinity Prediction

Virtual Screening  Lead Optimization
Drug Discovery Funnel

http://pharmit.csb.pitt.edu
Pharmer

Pharmacophore
A spatial arrangement of molecular features essential for biological activity

Pharmer

Hydrophobic Features

Hydrogen Donor Feature

Hydrogen Acceptor Feature

Hydrogen Features
Drug Discovery Funnel

\[
\text{gauss}_1(d) = u_{\text{gauss}} e^{-d^2/(0.5)^2}
\]
\[
\text{gauss}_2(d) = u_{\text{gauss}} e^{-d-(3)/2^2}
\]
\[
\text{repulsion}(d) = \begin{cases} u_{\text{repulsion}} d^2 & d < 0 \\ 0 & d \geq 0 \end{cases}
\]
\[
\text{hydrophobic}(d) = \begin{cases} w_{\text{hydrophobic}} & d < 0.5 \\ 0 & d > 1.5 \\ w_{\text{hydrophobic}}(1.5 - d) & \text{otherwise} \end{cases}
\]
\[
\text{hbond}(d) = \begin{cases} w_{\text{hbond}} & d < -0.7 \\ 0 & d > 0 \\ w_{\text{hbond}}(1 - d^2) & \text{otherwise} \end{cases}
\]

Protein-Ligand Scoring

Model

Pose Prediction
Binding Discrimination
Affinity 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

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

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

Convolution

Feature Maps

Fully Connected

Traditional NN

Dog: 0.99
Cat: 0.02
**Convolutional Filters**

![Convolutional Filters](image)

<table>
<thead>
<tr>
<th>-1</th>
<th>-1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-1</th>
<th>-1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>8</td>
<td>-1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>
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.
PDBbind 2016 refined set
- 4056 protein-ligand complexes
- diverse targets
- wide range of affinities
- generate poses with AutoDock Vina
- include minimized crystal pose

Pocketome
- 2923 distinct pockets
- 27,142 receptor structures
- 4,138,117 non-redundant poses
- generate poses with AutoDock Vina
- include minimized crystal pose

Redocked Training Set

Crossdocked Training Set
Training

Clustered Cross-validation

Target sequence similarity < 0.5
AND
Ligand similarity < 0.9
Optimized Models

Default2018

HiRes Pose

HiRes Affinity
Pose Results

Redocked Pose

Percent Targets with Low RMSD Top Pose

- Default2017
- Default2018
- HiRes Affinity
- HiRes Pose
- Vina
Pose Results

Crossdockded Pose

Percent Targets with Low RMSD Top Pose

Default2017  Default2018  HiRes Affinity  HiRes Pose  Vina
Affinity Results

Clustered Split

Spearman = 0.570, RMSE = 1.686

Random Split

Spearman = 0.690, RMSE = 1.496

PDBbind Core Set

Spearman = 0.789, RMSE = 1.336
Gradients

48x48x48x35
2x2x2 Max Pooling
4x4x4x35
3x3x3 Convolution
24x24x24x32
2x2x2 Max Pooling
24x24x24x32
3x3x3 Convolution
12x12x12x32
2x2x2 Max Pooling
12x12x12x32
3x3x3 Convolution
6x6x6x32
2x2x2 Max Pooling
6x6x6x32
3x3x3 Convolution
6x6x6x32
Fully Connected

Affinity
Pose Score

Gradients

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
Deep Dreams of Molecules
Gradients: Beyond Scoring

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

![Graph showing the distribution of poses based on RMSD change. The x-axis represents RMSD change, ranging from better to worse, and the y-axis shows the number of poses. Two categories are depicted: Best and First Minimization.](image-url)
Iterative Refinement

![Graph showing Iterative Refinement](image)

- **Best**
- **First Minimization**
- **Second Iteration**
- **Third Iteration**
GNINA 1.0
https://github.com/gnina/gnina

Monte Carlo Chain
Monte Carlo Chain
Monte Carlo Chain

exhaustiveness
num_mc_steps
merge
num_mc_saved
num_modes
out

Monte Carlo Chain
fast refinement + Metropolis

SOFTWARE

GNINA 1.0: molecular docking with deep learning
Andrew T. McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes

Open Access

cnn_scoring=rescore
cnn_scoring=refinement
cnn_scoring=all
Virtual Screening

DUD-E Virtual Screening Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>DUD-E</th>
<th>LIT-PCBA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>NEF1%</td>
</tr>
<tr>
<td>RFscore-4</td>
<td>0.683</td>
<td>0.0514</td>
</tr>
<tr>
<td>RFscore-VS</td>
<td>0.963</td>
<td>0.857</td>
</tr>
<tr>
<td>Vina</td>
<td>0.745</td>
<td>0.118</td>
</tr>
<tr>
<td>Vinardo</td>
<td>0.764</td>
<td>0.187</td>
</tr>
<tr>
<td>General (Affinity)</td>
<td>0.756</td>
<td>0.179</td>
</tr>
<tr>
<td>General (Pose)</td>
<td>0.702</td>
<td>0.156</td>
</tr>
<tr>
<td>Dense (Affinity)</td>
<td>0.795</td>
<td>0.27</td>
</tr>
<tr>
<td>Dense (Pose)</td>
<td>0.767</td>
<td>0.313</td>
</tr>
<tr>
<td>Default (Affinity)</td>
<td>0.795</td>
<td>0.258</td>
</tr>
<tr>
<td>Default (Pose)</td>
<td>0.744</td>
<td>0.241</td>
</tr>
</tbody>
</table>

Default (Affinity)

- No enrichment
Profilin

- Actin-binding protein
- Accelerates actin polymerization in presence of proline-rich proteins (e.g. formin, WASP, VASP)
- Sequesters actin otherwise

Dave Gau
Partha Roy
Virtual Screen

- Scaffold Hop Early Hit
- Pharmacophore screen (Pharmit)
- Ranked with Vina and CNN
Results

C74
DMSO
C2

Choroidal explant angiogenesis (ex vivo)

CNV assay (In vivo)

Oxygen-induced retinopathy (OIR) - In vivo

Isolectin-B4
Pfn1

DMSO
C2

vascular area

vascular outgrowth (%)
Generative Modeling
Discriminative Model
Generative Model

Features $X \rightarrow y$?
Generative Adversarial Networks

True Examples

Generator

Discriminator

Loss

Is this a real dog picture?
Generative Adversarial Networks

https://thispersondoesnotexist.com
Generative Models

Generative models approximate a data distribution directly. They can map samples from one distribution (noise or input data) to realistic samples from an output distribution of interest.
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Rafael Gómez-Bombarelli†, Jennifer N. Wei‡, David Duvenaud†, José Miguel Hernández-Lobato§, Benjamín Sánchez-Lengeling†, Dennis Sheberla†, Jorge Aguilera-Iparraguirre†, Timothy D. Hirzel†, Ryan P. Adams‡, and Alán Aspuru-Guzik††

SMILES input
ENCODER Neural Network
CONTINUOUS MOLECULAR REPRESENTATION (Latent Space)
DECODER Neural Network
SMILES output
L2 Loss
Model
Variational Autoencoding Examples

VAE

Atom Fitting

2BES
Variational Autoencoding Examples

<table>
<thead>
<tr>
<th>Around a target molecule (3 samples per each mol)</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>mol_1</td>
<td>mol_2</td>
</tr>
<tr>
<td><img src="image1" alt="mol_1" /></td>
<td><img src="image2" alt="mol_2" /></td>
</tr>
<tr>
<td>(0.45, 0.80, 0.25)</td>
<td></td>
</tr>
<tr>
<td><img src="image3" alt="mol_3" /></td>
<td><img src="image4" alt="mol_1.2.3" /></td>
</tr>
<tr>
<td>(0.31, 0.56, 0.87)</td>
<td></td>
</tr>
<tr>
<td><img src="image5" alt="mol_4.5.6" /></td>
<td></td>
</tr>
<tr>
<td><img src="image6" alt="mol_1" /></td>
<td><img src="image7" alt="mol_2" /></td>
</tr>
<tr>
<td>(0.39, 0.80, 0.51)</td>
<td></td>
</tr>
<tr>
<td><img src="image8" alt="mol_3" /></td>
<td><img src="image9" alt="mol_1.2.3" /></td>
</tr>
<tr>
<td>(0.35, 0.56, 0.44)</td>
<td></td>
</tr>
<tr>
<td><img src="image10" alt="mol_4.5.6" /></td>
<td></td>
</tr>
<tr>
<td><img src="image11" alt="mol_1" /></td>
<td><img src="image12" alt="mol_2" /></td>
</tr>
<tr>
<td>(0.37, 0.80, 0.34)</td>
<td></td>
</tr>
<tr>
<td><img src="image13" alt="mol_3" /></td>
<td><img src="image14" alt="mol_1.2.3" /></td>
</tr>
<tr>
<td>(0.27, 0.56, 0.78)</td>
<td></td>
</tr>
<tr>
<td><img src="image15" alt="mol_4.5.6" /></td>
<td></td>
</tr>
</tbody>
</table>

pink : true ligand  green : generated molecule  rigid body alignment
Context Encoding

http://people.eecs.berkeley.edu/~pathak/context_encoder/
Context Encoding

receptor grid

Generator

generated ligand grid

GAN loss
Conditioning on the Receptor
Context Encoding with Fully Convolutional Network

1m5w

Generated

Fit Densities

Fit Atoms
Context Encoding with Fully Convolutional Network

3bxg

Generated

Fit Densities

Fit Atoms
Context Encoding with Fully Convolutional Network

3ebp

Generated
Fit Densities
Fit Atoms
Acknowledgements