Structure Based Drug Design

- Pose Prediction
- Binding Discrimination
- Affinity Prediction

Virtual Screening

Lead Optimization
Structure Based Drug Design

- Pose Prediction
- Binding Discrimination
- Affinity Prediction

Virtual Screening

Lead Optimization
Drug Discovery Funnel

http://pharmit.csb.pitt.edu
Drug Discovery Funnel

\[
gaussian_1(d) = \frac{w_{gauss1}}{\sigma_{gauss1}^2} e^{-\frac{(d-\mu_{gauss1})^2}{2\sigma_{gauss1}^2}}
\]
\[
gaussian_2(d) = \frac{w_{gauss2}}{\sigma_{gauss2}^2} e^{-\frac{(d-\mu_{gauss2})^2}{2\sigma_{gauss2}^2}}
\]
\[
repulsion(d) = \begin{cases} 
  w_{repulsion} d^2 & d < 0 \\
  0 & d \geq 0
\end{cases}
\]

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

\[
h\text{bond}(d) = \begin{cases} 
  w_{hbond} & d < -0.7 \\
  0 & d > 0 \\
  w_{hbond}(1.5 - d) & \text{otherwise}
\end{cases}
\]

Scoring

\[ R = 0.53 \]
\[ \text{RMSE} = 0.89 \]

\[
\begin{align*}
\text{gauss}(a_1, a_2, d) &= e^{\frac{-((a_1 - d)^2 + (a_2 - d)^2)}{\sigma^2}} \\
\text{repulsion}(a_1, a_2, d) &= \begin{cases} 
(d_{\text{dir}}(a_1, a_2) - a)^2, & d_{\text{dir}}(a_1, a_2) < a \\
0, & \text{otherwise}
\end{cases} \\
\text{vdw}(a_1, a_2, d) &= \frac{d_{\text{dir}}(a_1, a_2)}{d} - 2 \left( \frac{d_{\text{dir}}(a_1, a_2)}{d} \right)^4 \\
\text{electric}(a_1, a_2, d) &= \frac{\text{partial}_\text{charge}(a_1) \cdot \text{partial}_\text{charge}(a_2)}{d^2}
\end{align*}
\]

\[
\begin{align*}
\text{hydrogen\_bond}(a_1, a_2, d) &= \begin{cases} 
0, & (a_1, a_2) \text{ do not form hydrogen bond} \\
1, & d_{\text{dir}}(a_1, a_2) < -0.7 \\
0, & d_{\text{dir}}(a_1, a_2) \geq b \\
\end{cases}, \text{ otherwise}
\end{align*}
\]

\[
\begin{align*}
\text{hydrophobic}(a_1, a_2, d) &= \begin{cases} 
0, & \text{not\_hydrophobic}(a_1) \text{ or not\_hydrophobic}(a_2) \\
1, & d_{\text{dir}}(a_1, a_2) < 0.5 \\
0, & d_{\text{dir}}(a_1, a_2) \geq b \\
\end{cases}, \text{ otherwise}
\end{align*}
\]

\[
\begin{align*}
\text{non\_hydrophobic}(a_1, a_2, d) &= \begin{cases} 
0, & \text{is\_hydrophobic}(a_1) \text{ or is\_hydrophobic}(a_2) \\
1, & d_{\text{dir}}(a_1, a_2) < 0.5 \\
0, & d_{\text{dir}}(a_1, a_2) \geq 1.5 \\
1.5 - d_{\text{dir}}(a_1, a_2), & \text{otherwise}
\end{cases}
\end{align*}
\]

\[
\text{ad4\_solvation}(a_1, a_2, d) = \left( (\text{solv}(a_1) + q \cdot \text{partial\_charge}(a_1)) \cdot \text{volume}(a_1) + (\text{solv}(a_2) + q \cdot \text{partial\_charge}(a_2)) \cdot \text{volume}(a_2) \right) e^{-\left(\frac{d_{\text{dir}}(a_1, a_2)}{\sigma_{\text{lennard\_jones}}} \right)^2} \\
\text{counts} : \#\text{heavy\_atoms}, \text{ligand\_length}, \#\text{hydrophobic\_atoms}, \#\text{torsions}, \#\text{torsions}^2, \sqrt{\text{v}\text{\_torsions}}
\]
Scoring

R = 0.53
RMSE = 0.89

Vina Docking

CSAR
343 Complexes

Vina Features:
- gauss (31)
- repulsion (8)
- vdw (1)
- electrostatic (2)
- hydrogen bond (3)
- hydrophobic (4)
- nonhydrophobic (1)
- ad4_solvation (2)
- counts (6)

CSAR
293 Complexes

Vina Features:
- gauss (31)
- repulsion (8)
- vdw (1)
- electrostatic (2)
- hydrogen bond (3)
- hydrophobic (4)
- nonhydrophobic (1)
- ad4_solvation (2)
- counts (6)

Feature Selection

Select Features

Linear Regression

Cross-Validation (100 trials)

Predicted pK

\[ R = 0.53 \]
\[ \text{RMSE} = 0.89 \]


smina.sf.net

Linear Regression

Cross-Validation (100 trials)

Select Features

Crystal Structures

Docked Structures

Regression Model

vdw hydrogen bond
ad4_solvation
num tors squared

Train on full set

Crystal Structures

Docked Structures

Regression Model

\[ \text{Vina Docking} \]

\[ \text{RMSD} \leq 2 \text{ to crystal pose} \]

\[ \text{CSAR} \]

\[ 343 \text{ Complexes} \]

\[ \text{CSAR} \]

\[ 293 \text{ Complexes} \]

\[ \text{Vina Features:} \]

\[ \text{gauss (31)} \]

\[ \text{repulsion (8)} \]

\[ \text{vdw (1)} \]

\[ \text{electrostatic (2)} \]

\[ \text{hydrogen bond (3)} \]

\[ \text{hydrophobic (4)} \]

\[ \text{nonhydrophobic (1)} \]

\[ \text{ad4_solvation (2)} \]

\[ \text{counts (6)} \]
Protein-Ligand Scoring

Model

Pose Prediction
Binding Discrimination
Affinity Prediction
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
Deep Learning

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

\[ \frac{\partial L}{\partial w_{jk}^l} = a_{k}^{l-1} \delta_j^l \quad \text{and} \quad \frac{\partial L}{\partial b_j^l} = \delta_j^l \]
Convolutional Neural Networks

Convolution Feature Maps

Fully Connected

Dog: 0.99
Cat: 0.02

Convolution

Fully-connected

weight 1
weight 2
weight 3
weight 4
weight 5
Convolutional Filters

-1  -1  -1
0    0    0
1    1    1

-1  0   1

-1  0   1

-1 -1 -1
-1  8  -1
-1 -1 -1
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.
Why Grids?

Cons

• coordinate frame dependent
• pairwise interactions not explicit

Pros

• clear spatial relationships
• amazingly parallel
• easy to interpret
Data Augmentation

![Graphs showing data augmentation results](image-url)
Data Augmentation

![Graphs showing data augmentation results](image)
Training

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

Crossdockaged 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

Crossdockked 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

Clustered Cross Validation
Affinity Results

- Clumped Split
  - Spearman = 0.570, RMSE = 1.686

- Random Split
  - Spearman = 0.690, RMSE = 1.496

- PDBbind Core Set
  - Spearman = 0.789, RMSE = 1.336

- Spearman Correlation
  - Clumped
  - Random
  - Core Set
Flexible Docking Scoring

Top Pose RMSD Distribution (Ligand)

Top Pose RMSD Distribution (Flexible Residues)
Virtual Screening

Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data

Fergus Imrie1, Anthony R. Bradley2, Misaela van der Schaar3, and Charlotte M. Dean4
1 Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3QU, U.K.
2 Structural Genomics Consortium, University of Oxford, Oxford OX3 7DD, U.K.
3 Department of Chemistry, University of Oxford, Oxford OX1 3TA, U.K.
4 Diamond Light Source Ltd., Didcot OX11 ORO, U.K.

Department of Engineering, University of Oxford, Oxford OX1 3PJ, U.K.
Allen Turing Institute, London NW1 2DB, U.K.
Virtual Screening

In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening

Jochen Sleg, Florian Flachsenberg, and Matthias Rarey*  
Universität Hamburg, ZIH - Center for Bioinformatics, Research Group for Computational Molecular Design, Bundesstraße 43, 20146 Hamburg, Germany

Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening

Preprint submitted on 24.03.2019, 15:39 and posted on 25.03.2019, 12:58 by Liyang Chen, Anthony Cruz, Steven Ramsey, Callum J. Dickson, José S. Duca, Viktor Hornak, David R. Koes, Tom Kurtzman
Virtual Screening

In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening

Jochen Sleg, Florian Flachsenberg, and Matthias Rarey*
Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Bundesstraße 43, 20146 Hamburg, Germany

Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening

Preprint submitted on 24.03.2019, 15:39 and posted on 25.03.2019, 12:58 by Lieyang Chen, Anthony Cruz, Steven Ramsey, Callum J. Dickson, José S. Duca, Viktor Hornak, David R. Koes, Tom Kurtzman
Visualization
Anatomy of a deep learning paper

- Strong empirical results
- Post hoc theoretical explanation
Visualizing with Atomistic Probes

Redocked Training Set
Visualizing with Atomistic Probes

Crossdocked Training Set
Hydrogen Bonds... or Not

Receptor Atom Type

Ligand Atom Type
- NitrogenDonor
- NitrogenAcceptor

Ligand Atom Type
- OxygenDonorAcceptor
- OxygenAcceptor

Redocked Training Set
Hydrogen Bonds... or Not

Redocked Training Set
Hydrogen Bonds... or Not

Redocked Training Set
Visualizing with Atomistic Probes

Oxygen Acceptor

Nitrogen Acceptor
Visualizing with Atomistic Probes

Aliphatic Carbon

Aromatic Carbon
Visualizing with Atomistic Probes

Oxygen Donor/Acceptor

Nitrogen Donor
Visualizing with Atomistic Probes

Oxygen Donor/Acceptor  Nitrogen Donor
Visualizing Network Decisions

masking

layer-wise relevance

gradients

1UGX
Score: 0.62
Visualizations

Masking

LRP

Gradients

2IDZ
Score: 0.04

3EJT
Score: 0.92
Masking

Delete single ligand atoms

Delete ligand fragments

Delete single residues

Score

Average

0.86 - 0.84 = 0.02
Masking: Enzyme Mutants

- **PDB ID:** 1YZ3
  - Wild Type $K_i$: 1.55 nM
  - E219A $K_i$: 1375 nM
  - D267A $K_i$: 999 nM

- **PDB ID:** 3C3U
  - Wild Type $K_i$: 5.9 nM
  - L308A $K_i$: 2800 nM

- **PDB ID:** 2DOR
  - Wild Type $K_i$: $2.9 \times 10^4$ nM
  - N193A $K_i$: $1.16 \times 10^7$ nM
Pose Sensitivity

Partially Aligned Poses
Layer-wise Relevance

\[ f(x) = \ldots = \sum_{d \in l+1} R_d^{(l+1)} = \sum_{d \in l} R_d^{(l)} = \ldots = \sum_d R_d^{(1)} \]

\[ R_{i \leftarrow j}^{(l,l+1)} = R_j^{(l+1)} \cdot \left( \alpha \cdot \frac{z_{i,j}^+}{z_j^+} + \beta \cdot \frac{z_{i,j}^-}{z_j^-} \right) \]
Gradients

2x2x2 Max Pooling
3x3x3 Convolution
2x2x2 Max Pooling
3x3x3 Convolution
2x2x2 Max Pooling
12x12x12x64
12x12x12x32
24x24x24x32
24x24x24x35
48x48x48x35

Affinity
Pose Score

Fully Connected
Fully Connected
Rectified Linear Unit
Rectified Linear Unit
Rectified Linear Unit
Rectified Linear Unit
Pseudo-Huber Loss
Softmax+Logistic Loss

48x48x48x35
24x24x24x35
2x2x2 Max Pooling
3x3x3 Convolution
2x2x2 Max Pooling
3x3x3 Convolution
2x2x2 Max Pooling
3x3x3 Convolution
12x12x12x64
12x12x12x32
24x24x24x32
24x24x24x35
48x48x48x35

Gradients
Gradients

\[
\frac{\partial L}{\partial \theta^l_{jk}} = a_{(l-1)3}^{(l-1)2} \delta_j^{(l-1)} \quad \text{and} \quad \frac{\partial L}{\partial b_j^l} = \delta_j^{(l-1)}
\]

Affinity

Pose Score

Rectified Linear Unit

Softmax+Logistic Loss
Gradients

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html
Deep Dreams of Molecules
Deep Dreams of Molecules
Screening with Pseudo Ligands

- Overlap Mult
- Overlap Threshold
- Overlap L2
- Overlap L1
- CNNscore
- CNNaffinity
- Vina

AUC

CNN

L1

L2

Mult

Threshold

$\rho = -0.11$

$\rho = -0.23$

$\rho = 0.064$

$\rho = -0.13$
Gradients: Beyond Scoring

More Oxygen Here

Less Oxygen Here

$$\frac{\partial L}{\partial A} = \sum_{i \in G_A} \frac{\partial L}{\partial g_i} \frac{\partial g_i}{\partial D} \frac{\partial D}{\partial A}$$
Gradients: Beyond Scoring

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

More Oxygen Here

Less Oxygen Here

2Q89
Minimizing Low RMSD Poses

![Graph showing RMSD change with categories of better and worse. The x-axis represents RMSD change, while the y-axis shows the number of poses. The graph includes two bars representing Best and First Minimization. The bars indicate a distribution of poses across different RMSD change values.](image)
Iterative Refinement
Iterative Refinement

- **Best**
- **First Minimization**
- **Second Iteration**

![Graph showing Iterative Refinement](image)

- **X-axis**: RMSD Change
- **Y-axis**: # Poses

The histogram illustrates the distribution of RMSD changes across different iterations, showing the improvement in poses from the best, first minimization, and second iteration.
Iterative Refinement

![Graph showing iterative refinement process with different iterations and RMSD change metrics.](image-url)
libmolgrid

Caffe Training

PyTorch Training

Keras Training

GPU Performance

GPU Memory Utilization

e = molgrid.ExampleProvider(balanced=True, shuffle=True)
e.populate('examples.txt')
gmaker = molgrid.GridMaker()
batch = e.next_batch(batch_size)
gmaker.forward(batch, input_tensor,
random_translation=0, random_rotation=True)
Case Studies
Case 1: Profilin-Actin
Profilin

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

• Whole protein docking of early hit
• Identified 5 sites
• Pharmacophore screen (Pharmit)
• Ranked with Vina and CNN
Results

57 compounds tested, 3 actives identified
Results

57 compounds tested, **3 actives** identified

1 (Vina) didn't work in cells
Results

57 compounds tested, 3 actives identified

1 (Vina) didn't work in cells

All predicted to bind to different sites

1 uM  10 uM  50 uM  100 uM

DMSO  C73  C74

CNN

Vina
Case 2: TIGIT
Can we block TIGIT/PVR interaction with a small molecule?
Does anything bind to this pocket?

Fragment Docking

Pharmacophore Search

Consensus Scoring (CNN and Vina)
### Screening

10 diverse compounds selected for screening
- top ranked by Vina
- top ranked by CNN

<table>
<thead>
<tr>
<th>Name</th>
<th>CNN Affinity</th>
<th>CNN Score</th>
<th>Vina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 1</td>
<td>7.69807</td>
<td>0.994763</td>
<td>85.95</td>
</tr>
<tr>
<td>Compound 2</td>
<td>5.57909</td>
<td>0.0180277</td>
<td>-8.12632</td>
</tr>
<tr>
<td>Compound 3</td>
<td>6.73692</td>
<td>0.0624742</td>
<td>-9.81935</td>
</tr>
<tr>
<td>Compound 4</td>
<td>6.87897</td>
<td>0.953488</td>
<td>-3.81378</td>
</tr>
<tr>
<td>Compound 5</td>
<td>6.32813</td>
<td>0.209807</td>
<td>-8.60293</td>
</tr>
<tr>
<td>Compound 6</td>
<td>5.689</td>
<td>0.0437</td>
<td>-8.991</td>
</tr>
<tr>
<td>Compound 7</td>
<td>4.368</td>
<td>0.022</td>
<td>-9.34722</td>
</tr>
<tr>
<td>Compound 8</td>
<td>4.81</td>
<td>0.072</td>
<td>-6.81787</td>
</tr>
<tr>
<td>Compound 9</td>
<td>5.22</td>
<td>0.032</td>
<td>-6.264</td>
</tr>
<tr>
<td>Compound 10</td>
<td>6.67</td>
<td>0.361</td>
<td>6.1053</td>
</tr>
</tbody>
</table>
Results

**TIGIT:CD155 Interaction**

Negative Control
Positive Control

Test Compound #
(all at 100 μM except #2 and #3 at 50 μM)

% Activity

EC_{50} = 11.6μM

Fold Induction

**PD-1:PD-L1 Interaction**

% Activity

IC50 > 100 μM
16% inhibition at 100 μM

**Compounds**

- Compound 1 (Log [μM])
- TIGIT:CD155 Interaction IC50 = 1 μM
- TIGIT:CD155 Interaction IC50 ~ 14 μM
- PD-1:PD-L1 Interaction IC50 = 14 μM

**Cellular assay**

- 65% inhibition at 100 μM
- 100 uM
But...

The first trial was promising, but the maximum does was limited by DMSO concentration. Future trials at appropriate dosages showed no response.
But...
Case 3: Mystery Target
Approach

- unbiased MD simulations
- pocket identification (mdpocket)
- select receptor using pocket volume/druggability
- pharmacophore query from fragment docking
- minimization w/Vina
- filters: energy minimized RMSD, Lipinski, PAINS
- hits
- consensus scoring
- CNN, Vina, GlideXP
- best min-rank, best max-rank
- sort by sum-rank
- diversity filter at most 2 cmpds w/ Tanimoto > 0.7

select ~20 final hits per pocket

HBD <= 5, HBA <= 10, MW <= 500, logP <= 5

select ~20 final hits per pocket
Screening Hits

- 50 compounds tested
- designed against 3 putative allosteric pockets
- 4 hits (3 from P2, 1 from P4)
- P2 was potentially a very desirable pocket to hit for target-specific reasons
Screening Hits

• 50 compounds tested
• designed against 3 putative allosteric pockets
• 4 hits (3 from P2, 1 from P4)
• P2 was potentially a very desirable pocket to hit for target-specific reasons
<table>
<thead>
<tr>
<th>MolPort ID</th>
<th>CNNScore</th>
<th>-Vina</th>
<th>-GlideXP</th>
<th>cnn_rank</th>
<th>vina_rank</th>
<th>glide_rank</th>
<th>sum_rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7.35853</td>
<td>9.73614</td>
<td>6.68</td>
<td>3.0</td>
<td>5.0</td>
<td>7.5</td>
<td>15.5</td>
</tr>
<tr>
<td>1</td>
<td>6.36562</td>
<td>9.25733</td>
<td>6.56</td>
<td>4.0</td>
<td>7.0</td>
<td>9.0</td>
<td>20.0</td>
</tr>
<tr>
<td>2</td>
<td>6.14087</td>
<td>8.73366</td>
<td>6.94</td>
<td>6.0</td>
<td>12.0</td>
<td>5.0</td>
<td>23.0</td>
</tr>
<tr>
<td>3</td>
<td>5.66284</td>
<td>9.40722</td>
<td>6.76</td>
<td>13.0</td>
<td>6.0</td>
<td>6.0</td>
<td>25.0</td>
</tr>
<tr>
<td>4</td>
<td>6.02615</td>
<td>8.82225</td>
<td>6.21</td>
<td>7.0</td>
<td>8.0</td>
<td>13.0</td>
<td>28.0</td>
</tr>
<tr>
<td>5</td>
<td>5.32736</td>
<td>10.18550</td>
<td>6.51</td>
<td>14.0</td>
<td>4.0</td>
<td>10.0</td>
<td>28.0</td>
</tr>
<tr>
<td>6</td>
<td>5.80743</td>
<td>7.26154</td>
<td>8.24</td>
<td>10.0</td>
<td>18.0</td>
<td>1.0</td>
<td>29.0</td>
</tr>
<tr>
<td>7</td>
<td>5.93622</td>
<td>8.75104</td>
<td>6.38</td>
<td>11.0</td>
<td>11.0</td>
<td>11.0</td>
<td>30.0</td>
</tr>
<tr>
<td>8</td>
<td>6.20744</td>
<td>6.18992</td>
<td>6.68</td>
<td>19.0</td>
<td>7.5</td>
<td>5.0</td>
<td>31.5</td>
</tr>
<tr>
<td>9</td>
<td>5.06070</td>
<td>10.32670</td>
<td>5.20</td>
<td>15.0</td>
<td>1.0</td>
<td>16.0</td>
<td>32.0</td>
</tr>
<tr>
<td>10</td>
<td>4.77866</td>
<td>10.32050</td>
<td>5.32</td>
<td>16.0</td>
<td>2.0</td>
<td>15.0</td>
<td>33.0</td>
</tr>
<tr>
<td>11</td>
<td>5.72094</td>
<td>8.76388</td>
<td>6.24</td>
<td>11.0</td>
<td>10.0</td>
<td>12.0</td>
<td>33.0</td>
</tr>
<tr>
<td>12</td>
<td>7.55155</td>
<td>8.26183</td>
<td>4.26</td>
<td>1.0</td>
<td>15.0</td>
<td>18.0</td>
<td>34.0</td>
</tr>
<tr>
<td>13</td>
<td>3.84854</td>
<td>10.19640</td>
<td>5.99</td>
<td>18.0</td>
<td>3.0</td>
<td>14.0</td>
<td>35.0</td>
</tr>
<tr>
<td>14</td>
<td>4.29961</td>
<td>7.43586</td>
<td>8.18</td>
<td>17.0</td>
<td>17.0</td>
<td>3.0</td>
<td>37.0</td>
</tr>
<tr>
<td>15</td>
<td>3.34137</td>
<td>7.79760</td>
<td>8.21</td>
<td>19.0</td>
<td>16.0</td>
<td>2.0</td>
<td>37.0</td>
</tr>
<tr>
<td>16</td>
<td>2.77562</td>
<td>8.28298</td>
<td>8.17</td>
<td>20.0</td>
<td>14.0</td>
<td>4.0</td>
<td>38.0</td>
</tr>
<tr>
<td>17</td>
<td>5.70688</td>
<td>8.80796</td>
<td>4.77</td>
<td>12.0</td>
<td>9.0</td>
<td>17.0</td>
<td>38.0</td>
</tr>
<tr>
<td>18</td>
<td>5.92500</td>
<td>8.53183</td>
<td>4.25</td>
<td>9.0</td>
<td>13.0</td>
<td>19.0</td>
<td>41.0</td>
</tr>
<tr>
<td>19</td>
<td>7.52880</td>
<td>6.05508</td>
<td>3.26</td>
<td>2.0</td>
<td>20.0</td>
<td>20.0</td>
<td>42.0</td>
</tr>
</tbody>
</table>
But...

*Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.*
But...

**Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.**

But but...

Thermal shift assays show binding

- **more interesting version of protein**
  - Positive control: 6µM
  - ΔTm: 4°C

- **less interesting version of protein**
  - Positive control: 6µM
  - ΔTm: 12°C

**positive control**

**6µM**

ΔTm (°C)

0 1 2 3 4

- *******

**less interesting control**

ΔTm (°C)

- ****

- **positive control**

- **6µM**

ΔTm (°C)

0 2 4 6 8 10 12 14 16

- ****
But...

Sorry to be the bearer of potentially bad news but ... it seems that there may have been some interference (quenching of the product fluorophore) with the compounds/samples.

But but...

Thermal shift assays show binding

But but but...

Those error bars
Case 4: DUSP6
A Tumor Cell-Selective Inhibitor of Mitogen-Activated Protein Kinase Phosphatases Sensitizes Breast Cancer Cells to Lymphokine-Activated Killer Cell Activity

Vassily N. Korotchenko, Manush Saydmohammed, Laura L. Vollmer, Ahmet Bakaran, Kyle Sheetz, Karl T. Debiec, Kristina A. Greene, Christine S. Agliori, Inet Bahar, Billy W. Day, Andreas Vogt, Michael Tsang... See fewer authors

A cell-active inhibitor of mitogen-activated protein kinase phosphatases restores paclitaxel-induced apoptosis in dexamethasone-protected cancer cells

Andreas Vogt, Peter R. McDonald, Andreea Tanema, Rafael P. Sikorski, Peter Wyp, John J. Skele, III, and John S. Lazo

DOI: 10.1159/10.002.008.065 Published February 2008
Results

34 compounds tested at the highest possible concentration (300 µM or 75 µM depending on solubility) and 1/10 that

24 hour exposure

Stain for phospho-ERK

Selected hits with >1.5-fold increase in pERK over DMSO

Of the **six visually possibilities**, three were from the Maya (BCI1) and three from the Ahmet site (BCI4)

**5** hits selected by Vina and **1** by the CNN
Possible positives

DMSO
25 µM BCI215
1 µg/ml TPA
300 µM C15
300 µM C17

PKC activator
Possible positives

DMSO

300 µM D11

300 µM J23

300 µM K03

25 µM BCI215
Possible positives

- DMSO
- 25 µM BCI215
- 1 µg/ml TPA
- 75 µM C23, enlarged
- 75 µM C23
- 7.5 µM C22

Vina
Generative Modeling
Discriminative Model

Features $X$ $\rightarrow$ Prediction $y$
Generative Model

Features $X$
Generative Model

Features $\mathbf{X}$
Generative Model

y?

Features X
Generative Adversarial Networks

True Examples

Generator

Discriminator

Loss

Is this a real dog picture?
Generative Adversarial Networks

Ian Goodfellow @goodfellow_ian · 2h
4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661
arxiv.org/abs/1812.04948

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.

noise sample  Generator  generated receptor & ligand grid
Autoencoding

Encoder

Latent Space

Generator

L2 Loss
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*†
Variational Autoencoding Examples

VAE

Atom Fitting

2BES
# Variational Autoencoding Examples

<table>
<thead>
<tr>
<th>PDB</th>
<th>True structure</th>
<th>True density</th>
<th>Gen. density</th>
<th>Fit density</th>
<th>Fit structure</th>
<th>Gen. L2 distance</th>
<th>Fit L2 distance</th>
<th>Fit RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3h78</td>
<td><img src="image" alt="3h78 True structure" /></td>
<td><img src="image" alt="3h78 True density" /></td>
<td><img src="image" alt="3h78 Gen. density" /></td>
<td><img src="image" alt="3h78 Fit density" /></td>
<td><img src="image" alt="3h78 Fit structure" /></td>
<td>9.4053</td>
<td>8.3141</td>
<td>0.6150</td>
</tr>
<tr>
<td>4jx9</td>
<td><img src="image" alt="4jx9 True structure" /></td>
<td><img src="image" alt="4jx9 True density" /></td>
<td><img src="image" alt="4jx9 Gen. density" /></td>
<td><img src="image" alt="4jx9 Fit density" /></td>
<td><img src="image" alt="4jx9 Fit structure" /></td>
<td>13.8545</td>
<td>9.7198</td>
<td>0.8820</td>
</tr>
<tr>
<td>3lgp</td>
<td><img src="image" alt="3lgp True structure" /></td>
<td><img src="image" alt="3lgp True density" /></td>
<td><img src="image" alt="3lgp Gen. density" /></td>
<td><img src="image" alt="3lgp Fit density" /></td>
<td><img src="image" alt="3lgp Fit structure" /></td>
<td>14.8525</td>
<td>12.5245</td>
<td>1.2056</td>
</tr>
<tr>
<td>4cwf</td>
<td><img src="image" alt="4cwf True structure" /></td>
<td><img src="image" alt="4cwf True density" /></td>
<td><img src="image" alt="4cwf Gen. density" /></td>
<td><img src="image" alt="4cwf Fit density" /></td>
<td><img src="image" alt="4cwf Fit structure" /></td>
<td>11.4730</td>
<td>9.0564</td>
<td>0.6725</td>
</tr>
</tbody>
</table>
Context Encoding

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

receptor grid

Generator

generated ligand grid

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

1m5w

Generated

Fit Densities

Fit Atoms
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

3bxg

Generated

Fit Densities

Fit Atoms
Context Encoding with Fully Convolutional Network

3ebp

Generated

Fit Densities

Fit Atoms
Context Encoding with Fully Convolutional Network

3ebp

Generated
Fit Densities
Fit Atoms
An Aside…
An Aside...
An Aside...

Area
An Aside...

![Area graphs](image)
Acknowledgements

Jocelyn Sunseri
Paul Francoeur
Josh Hochuli
Hunter Haaf
Rachel Rosenzweig
Keshavan Seshadri
Matt Ragoza
Alec Helbling

Department of Computational and Systems Biology

Google Cloud
NVIDIA
National Institute of General Medical Sciences R01GM108340
"RELAY THERAPEUTICS"
Do I have more time?
Do you care about chemistry education?
Molecular Active Learning

“Overall, we found small but significant effects of using ARS-based technologies on a number of desirable cognitive and non-cognitive learning outcomes.”

Go to this URL: http://3dmol.csb.pitt.edu/viewer.html
github.com/gnina

github.com/3dmol

http://bits.csb.pitt.edu

@david_koes