In Silico Drug Design and Development

Joseph Badillo
MacMillan Group Meeting
March 1, 2016
In Silico Drug Design and Development

Outline

General outline

1) Some vocabulary

2) Intro to docking and scoring functions

3) Examples of drug lead discovery using in silico methods
   a) Mycobacterium tuberculosis
   b) Type-II diabetes
   c) Cancer

4) Computational method used to find modifiers for siRNA
**Computer aided drug design (CADD):** is the use of computing power to streamline the drug discovery and development process.

**In silico vocabulary:**

**Virtual screening:** computational technique used to evaluate vast libraries chemical structures as potential new drug compounds (docking a library of compounds).

**Docking:** predicts the orientation (or pose) in which a ligand will bind to a host protein.

**Scoring:** a fast approximate mathematical method used to predict the strength of non-covalent interactions (binding affinity) between two molecules after they have been docked. The scoring function is one of the most important components in structure-based drug design.

**Bioinformatics:** Most important is structural information about potential biological targets

**Chemical informatics:** Design of *in silico* filters to eliminate compounds with undesirable properties. Prediction of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) based on criteria such as polar surface area (PSA), calculated log P, and the number of H-bond donors and accepters. How "druglike" are a set of compounds.
In Silico Drug Design and Development

**Traditional drug development:**

- Disease ID
- Target ID
- Target validation
- Lead discovery
- Lead optimization
- Preclinical tests
- Clinical trials

3-6 years and millions of dollars

**Computer-aided drug design (CADD) applications in various stages of drug development:**

**Target ID:** Bioinformatics, reverse docking, protein structure prediction

**Target validation:** Target drugability, tool compound design

**Lead discovery:** Library design, docking/scoring, virtual screening

**Lead optimization:** Quantitative structure-activity relationship, structure-based optimization

**Preclinical tests:** In silico ADMET (absorption, distribution, metabolism, excretion, toxicity), physiologically-based pharmacokinetic (PBPK) simulations
Molecular docking: is a tool used in structural molecular biology and computer-assisted drug design. The goal of molecular docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure.

The question:
"Given the structure of a protein and that of a potential ligand, can the two form a favorable complex? What are the bases for binding and specificity?" -Brian Shoichet (UCSF)

Three main challenges associated with docking:

Molecular flexibility: There are many states to consider in docking flexible molecules (both target and ligand).

Accurate scoring: Evaluating fit for the docking molecules from a database, and ranking them accordingly.

Specificity: Understanding important interactions.
In Silico Drug Design and Development

Docking

Increase in the number of docking of papers from 1990 to 2013 (PMC-NCBI database):

Keywords: 'dock' or 'docking'

In Silico Drug Design and Development

Docking

- All docking publications from 1990 to 2013:

Free energy of binding ($\Delta G$) is related to binding affinity ($K_i$):

$$\Delta G = -RT \ln K_A \quad \implies \quad K_A = K_i^{-1} = \frac{[EI]}{[E][I]}$$

Three common types of scoring functions:

**Force field**
Affinities are estimated by summing the strength of intermolecular interactions between all atoms using a force field.

**Empirical**
Based on counting the number of various types of interactions between two binding partners.

**Knowledge-based**
Based on statistical observations of intermolecular close contacts in large 3D databases (aka statistical potentials).
### In Silico Drug Design and Development

#### Scoring functions

**Common scoring functions:**

<table>
<thead>
<tr>
<th>Type</th>
<th>Scoring function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force field-based</td>
<td>DOCK, DOCK3.5 (PBSA), DOCK/GBSA(SDOCK), AutoDock, GOLD, SYBYL/D-Score, SYBYL/G-Score</td>
</tr>
<tr>
<td>Empirical</td>
<td>FlexX, Glide, ICM, LUDI, PLP, ChemScore, X-Score, Surflex, SYBYL/F-Score, LigScore, MedusaScore, AIScore, SFCscore</td>
</tr>
<tr>
<td>Knowledge-based</td>
<td>ITScore, PMF, DrugScore, DFIRE, SMoG, BLEEP, MScore, GOLD/ASP, KScore</td>
</tr>
</tbody>
</table>

**In Silico Drug Design and Development**

*Docking*

- **Success rates of 16 scoring functions for a test set of 100 diverse protein-ligand complexes, using the criterion of rmsd ≤ 2 Å**

<table>
<thead>
<tr>
<th>Scoring function</th>
<th>Type of scoring</th>
<th>Success rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITScore/SE</td>
<td>K</td>
<td>91</td>
</tr>
<tr>
<td>DrugScore</td>
<td>K</td>
<td>87</td>
</tr>
<tr>
<td>ITScore</td>
<td>K</td>
<td>82</td>
</tr>
<tr>
<td>Cerius2/PLP</td>
<td>E</td>
<td>76</td>
</tr>
<tr>
<td>SYBYL/F-Score</td>
<td>E</td>
<td>74</td>
</tr>
<tr>
<td>Cerius2/LigScore</td>
<td>E</td>
<td>74</td>
</tr>
<tr>
<td>DrugScore</td>
<td>K</td>
<td>72</td>
</tr>
<tr>
<td>Cerius2/LUDI</td>
<td>E</td>
<td>67</td>
</tr>
<tr>
<td>X-Score</td>
<td>E</td>
<td>66</td>
</tr>
<tr>
<td>AutoDock</td>
<td>F</td>
<td>62</td>
</tr>
<tr>
<td>DFIRE</td>
<td>K</td>
<td>58</td>
</tr>
<tr>
<td>DOCK/FF</td>
<td>F</td>
<td>58</td>
</tr>
<tr>
<td>Cerius2/PMF</td>
<td>K</td>
<td>52</td>
</tr>
<tr>
<td>SYBYL/G-Score</td>
<td>F</td>
<td>42</td>
</tr>
<tr>
<td>SYBYL/ChemScore</td>
<td>E</td>
<td>35</td>
</tr>
<tr>
<td>SYBYL/D-Score</td>
<td>F</td>
<td>26</td>
</tr>
</tbody>
</table>

K = knowledge-based, E = empirical, and F = force field

Common force field scoring functions: Structure of p38 mitogen-activated protein kinase with bound inhibitor BIRB796 (PDB code: 1KV2).

- Electrostatic potential
- Interaction between Glu71 and BIRB796
- Van der Waals
- H-bond (angle dependent)
- 2 like charges
- 2 opposite charges

In Silico Drug Design and Development

Scoring functions

- Lennard-Jones potential:

\[ V = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \]

- \( \varepsilon \) = well depth
- \( r \) = distance between particles
- \( \sigma \) = finite distance at which the inter-particle potential is zero
In simple force-field potentials the individual terms are summed to give $\Delta G_{\text{binding}}$:

$$\Delta G_{\text{binding}} = \Delta G_{vdw} + \Delta G_{H-Bond} + \Delta G_{\text{electrostatic}} + \Delta G_{\text{tortional}} + \Delta G_{\text{solvation}}$$

Three main force-field scoring function limitations:

Implicit solvation models are often inadequate and based on a continuous medium instead of "explicit" solvent molecules.

There is a universal set of weighting coefficients for different energy terms for diverse protein–ligand complexes.

Inaccurate treatment of entropic effects may easily render useless the accuracy of electrostatic calculations.
**Docking ensembles using OpenEye FREAD: Finding a good pose.**

- **Initial 1D structure**
- **Enumerate conformers**
- **Enumerate rotations**
- **Enumerate translations**
- **Enumerate poses that clash with the protein (green) or are too far from the site (white)**
- **Score poses and eliminate those with low score**
- **Optimize poses**

Typically evaluates $10^5$ to $10^9$ conformations for each molecule in seconds.

In Silico Drug Design and Development

Docking

Flexx mechanism for drug fragment docking into HIV protease:
In Silico Drug Design and Development

Docking

Docking vs. HTS for lead discovery

High throughput screening (HTS)

Protein target

Small molecule library

Test only high-scoring molecules

Hits

Test all molecules
**In Silico Drug Design and Development**

**Inhibitors of DHPR via in silico screening**

- **Mycobacterium tuberculosis (MTB):** a leading cause of death in developing countries, especially for people with compromised immune systems as a consequence of HIV infection.

**Current treatments:**

- rifampicin: disrupts cell wall bio-synthesis
- ethambutol: inhibits nucleic acid synthesis
- pyrazinamide: inhibits translation (pro-drug for pyrazinonic acid)

Resistance to these therapies have emerged so new new enzyme targets are needed
**In Silico Drug Design and Development**

*Inhibitors of DHPR via in silico screening*

- **Screening of compounds against DHPR from the Merck chemical collection by two approaches:**

  **DHPR (dihydrodipicolinate reductase):** has been found to play an essential role in bacterial cell wall synthesis and is a potential therapeutic target for MTB.

**Virtual screening approach:** Docking using FLOG (flexible ligands oriented on a grid) algorithm

  FLOG: at each point (on a grid) the potential energy is calculated and stored for each atom type.

  ![Diagram](image)

**Traditional screening approach:** Screen thousands of compounds *in vitro*

  DHPR catalyzed hydride transfer was monitored by following NADPH oxidation to NAD⁺ (via absorbance or fluorometric analysis)


In Silico Drug Design and Development
_Inhibitors of DHPR via in silico screening_

Active site of dihydridipicolinate reductase (DHPR):

X-ray stucture of *E.coli* DHPR with 2,6-PDC bound in active site stacked with NAD^+ cofactor

In Silico Drug Design and Development

Inhibitors of DHPR via in silico screening

View of L-613,517 docked into the DHPR active site:

1.6 X 10^6 compounds docked

500 scoring compounds chosen for in vitro assay

Lead compound:

L-613,517 (magenta)

IC_{50} = 7.2 \mu M (E. coli)

IC_{50} = 7.2 \mu M (M. tuberculosis)

The overall hit rate (IC_{50} values < 100 \mu M) for the virtually screened compounds was 6%
Compounds identified through traditional HTS:

Thousands of compounds screened from the Merck Chemical Repository

Lead compounds identified:

- L-298,878
  - $IC_{50} = 54 \mu M$ (E. coli)
  - $IC_{50} = 35 \mu M$ (M. tuberculosis)

- L-245,060
  - $IC_{50} = 20 \mu M$ (E. coli)

- L-273,552
  - $IC_{50} = 93 \mu M$ (E. coli)

The overall hit rate ($IC_{50}$ values < 100 $\mu M$) for the Merck HTS compounds was $\leq 2\%$.
**In Silico Drug Design and Development**

*Inhibitors of DHPR via in silico screening*

**Virtual screening vs. traditional high throughput screening:**

Lead identified using VS:

L-613,517

\[ IC_{50} = 7.2 \, \mu M \, (E. \, coli) \]
\[ IC_{50} = 7.2 \, \mu M \, (M. \, tuberculosis) \]

*over all hit rate was 6%*

Lead identified using HTS:

L-298,878

\[ IC_{50} = 54 \, \mu M \, (E. \, coli) \]
\[ IC_{50} = 35 \, \mu M \, (M. \, tuberculosis) \]

*over all hit rate was ≤2%*

Type-II diabetes: metabolic disorder characterized by high blood sugar due to inulin resistance or lack of inulin.

Long-term effects lead to heart disease, stroke, kidney failure, and nerve damage.

Protein tyrosine phosphate-1B (PTP1B): overproduction has been implicated in the onset of type-II diabetes, shown to deactivate the insulin receptor by hydrolyzing phosphotyrosines, and is therefore a potential drug target.

PTP1B complexed with two BPPM molecules
Size: 38 kD, 321 residues

X-ray structure
1.90 Å

Molecular docking and HTS for the discovery of novel PTP1B inhibitors:

HTS: a 400,000 "corporate" compound library was screened against PTP1B.
543 compounds inhibited the enzyme at 300 µM.
85 had IC₅₀ values ranging from 1-100 µM.
Hit rate of 0.021%.

Molecular docking: 250,000 commercial compounds from the ACD, Biospecs, and Maybridge databases were evaluated in DOCK 3.5.
An average of 350 conformations per compound (∼90 x 10⁶ conformers).
1000 compounds were considered for further evaluation (889 commercially available).
365 compounds (178 "spanners" and 187 "nonspanners") were tested \textit{in vitro}.
127 compounds had IC₅₀ < 100 µM
21 hits with IC₅₀ < 10 µM
Hit rate of 34.6%.
In Silico Drug Design and Development

Inhibitors of protein tyrosine phosphatase-1B

**Kinetic analysis of docking hits:** 4 out of the 127 hits behaved as simple competitive inhibitors.

![Compound 3](image)

<table>
<thead>
<tr>
<th>Compound 3</th>
<th>Docking rank: 39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docking score: -39.6 kcal/mol</td>
<td></td>
</tr>
<tr>
<td>$\text{IC}_{50} = 8.6 \mu\text{M}$ (vs. PTP1B)</td>
<td></td>
</tr>
</tbody>
</table>

Lineweaver-Burk analysis of competitive inhibitor compound 3 ($p\text{NPP} = p$-nitrophenyl phosphate).

The molecular surface of PTP1B docked with compound 3 ($K_i = 10.3 \mu\text{M}$):

Extensive shape complementarity

Aryl ring sandwiched in between Phe183 and Tyr46 (stacked with Tyr46)

Herringbone interaction (C–H - - $\pi$–interaction)

Comparision of the docked ligands to phosphotyrosine:

X-ray structure of the PTP1B catalytic site bound to phosphotyrosine.

Charged compounds featuring carboxylic or salicylic acid groups hydrogen bond with the phosphate recognition residues: Ser216, Gly220, and Arg221.

\[
\text{IC}_{50} = 21.6 \mu\text{M (vs. PTP1B)}
\]

**In Silico** Drug Design and Development

*Inhibitors of protein tyrosine phosphatase-1B*

**Chemical informatics**: Compare and contrast the "druglikeness" of the docked vs. HTS hit list.

**Lipinski's rule of five (RO5):**
1) molecular weight ≤ 500
2) calculated logP ≤ 5
3) ≤ 5 hydrogen bond donors
4) ≤ 10 hydrogen bond acceptors

<table>
<thead>
<tr>
<th>list</th>
<th># of compounds</th>
<th>passing 3/4</th>
<th>passing 4/4</th>
<th>average RO5 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTS hits</td>
<td>81</td>
<td>49</td>
<td>19</td>
<td>2.73</td>
</tr>
<tr>
<td>high-scoring docking molecules</td>
<td>889</td>
<td>773</td>
<td>577</td>
<td>3.47</td>
</tr>
<tr>
<td>docking molecules selected for testing</td>
<td>365</td>
<td>332</td>
<td>248</td>
<td>3.57</td>
</tr>
<tr>
<td>bioactive docking hits</td>
<td>127</td>
<td>116</td>
<td>73</td>
<td>3.49</td>
</tr>
</tbody>
</table>

Bioactive hits from docking are more drug like!
**Important things to consider:**

1) Although there is a 1700-fold enrichment in the docking hit rate vs. HTS. These libraries contain fundamentally different structures. A apples to apples comparison is needed.

2) Are the docking molecules binding at the active site as predicted?

3) Are the scoring functions accurate?

![Chemical structures](compound_1.png)

**compound 1**

docking score $-33.4$ kcal/mol

$IC_{50} = 4.1 \, \mu M \text{ (vs. PTP1B)}$

![Chemical structures](compound_8.png)

**compound 8**

docking score $-42.0$ kcal/mol

$IC_{50} = 21.6 \, \mu M \text{ (vs. PTP1B)}$

These inaccuracies may be attributed to failure to consider enzyme desolvation and conformational changes.

In Silico Drug Design and Development

p53-MDM2 inhibitor

**New potential cancer therapy:**

The tumor suppressor p53 is mutated in 50% of human cancers. In the remaining 50% it is rendered inactive due to interactions with the murine double minute 2 (MDM2) protein.

Inhibiting the p53-MDM2 interaction presents a novel strategy for the development of therapeutics against a variety of cancers.

X-ray structures of MDM2 bound to 3 different inhibitors:

Each ligand type induces different MDM2 conformers used in docking analysis.

MDM2/nutlin

MDM2/MI63

MDM2/K23

The binding modes of MDM2 crystal ligands predicted by the AutoDock Vina:

PDB: 4HG7
RMSD- 0.383 Å
nutlin

PDB: 1RV1
RMSD- 0.464 Å
benzodiazepinedione

PDB: 3LBL
RMSD- 0.405 Å
MI63

PDB: 3LBK
RMSD- 0.641 Å
K23

Excellent agreement with X-ray pose
**Repurposing drugs through virtual screening:**

7,800 approved drugs from the ZINC database were selected for virtual screening against MDM2.

6 hit compounds were identified. ZINCC00537755 (fluspiriline) had the highest binding energy.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Binding Free Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>fluspiriline</td>
<td>-122 kcal/mol</td>
</tr>
<tr>
<td>MI63</td>
<td>-126 kcal/mol</td>
</tr>
</tbody>
</table>

Fluspirilene

Discovered in 1963 at Janssen Pharmaceutica

Schizophrenia
**In Silico Drug Design and Development**

p53-MDM2 inhibitor

Fluspirilene showed comparable inhibition to known inhibitor nutlin:

![Bar chart showing cell growth comparison](image)

In vitro cell proliferation assay

Fluspirilene (10 μM) showed a broad-spectrum in the NCI60 human tumor cell lines:

In vitro cell proliferation assay
Argonaute proteins form complexes responsible for RNA silencing in eukaryotes:

X-ray structure of hAgo2 interactions with the RNA 5-guide strand

siRNAs are promising targets for potential therapeutics many of which have been considered to be "undrugable"
Predicted hAgo2 binding modes for 5'-end guide strand analogs purine analogs:

- 7-EAA
- 7-EAA triazole
- 2-AP triazole


Schirle, N. T.; MacRae, I. J. Science 2012, 336, 1037.
Predicted hAgo2 binding modes for 5'-end guide strand analogs purine analogs:

<table>
<thead>
<tr>
<th></th>
<th>predicted hAgo2 binding</th>
<th>siRNA activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>adenosine</td>
<td>1.0</td>
<td>++</td>
</tr>
<tr>
<td>7-EAA</td>
<td>8.8</td>
<td>+</td>
</tr>
<tr>
<td>7-EAA triazole</td>
<td>9.2</td>
<td>+</td>
</tr>
<tr>
<td>2-AP-triazole</td>
<td>9.5</td>
<td>+</td>
</tr>
</tbody>
</table>

^aLower number represents better-predicted binding

^b+++ = <10% luciferase activity remaining after knockdown; ++ = 10-40%; + = 41-70%; − = >70%

Luciferase knockdown activity in HeLa cells:

All siRNAs were prepared with a 5'-phosphorylated guide strand.

5'-p-XUAGGAUUCAUAAUAGGAGAU-3'
3'-ACUAUCCUAAGUAUAAUCCUC-5'

Sequence of siRNA used in this study. X indicates guide strand position 1.
Predicted hAgo2 binding modes for 5'-end guide strand 1-ethynyl ribose derivatives:

7-ER

7-ER triazole 1

7-ER triazole 2


Schirle, N. T.; MacRae, I. J. Science 2012, 336, 1037.
Synthesis of 1-ER Phosphoramidite:

1. BzO \rightarrow \text{TMSCCAl(Et)Cl} \rightarrow \text{CH}_2\text{Cl}_2 \rightarrow 44\% \quad \text{(single anomer)}

2. \text{NH}_4\text{OH, EtOH} \rightarrow \text{93\% (2 steps)}

3. \text{TBDMSCl, AgNO}_3 \rightarrow \text{pyridine, THF} \rightarrow 48\% (2'-TBDMS) \quad 27\% (3'-TBDMS)

4. \text{CIP(OCH}_2\text{CH}_2\text{CN}(\text{N(iPr})_2) \rightarrow \text{DIPEA, CH}_2\text{Cl}_2 \rightarrow \text{81\%}

**Predicted hAgo2 binding modes for 5'-end guide strand 1-ethynyl ribose derivatives:**

<table>
<thead>
<tr>
<th></th>
<th>predicted hAgo2 binding</th>
<th>siRNA activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>adenosine</td>
<td>1.0</td>
<td>++</td>
</tr>
<tr>
<td>1-ER</td>
<td>2.1</td>
<td>++</td>
</tr>
<tr>
<td>1-ER triazole I</td>
<td>1.2</td>
<td>+++</td>
</tr>
<tr>
<td>1-ER triazole II</td>
<td>1.3</td>
<td>+++</td>
</tr>
</tbody>
</table>

- **a**Lower number represents better-predicted binding
- **b**+++ = <10% luciferase activity remaining after knockdown; ++ = 10-40%; + = 41-70%; − = >70%

Luciferase knockdown activity in HeLa cells:

All siRNAs were prepared with a 5'-phosphorylated guide strand.
Luciferase knockdown activity in HeLa cells:

All siRNAs were prepared with a 12 position-phosphorylated guide strand.
**Luciferase knockdown activity in HeLa cells:**

All siRNAs were prepared with a 5'-phosphorylated guide strand. Modifications made at the 19-position of the passenger strand.

Sequence of siRNA used in this study:

5' -p-AUAGGAUUCAUAAUAGGAGAU-3'  
3' -ACXAUCCUAAGUAUAACUC-5'  

X = uridine or nucleoside analog

**In Silico Drug Design and Development**

*How easy is it?*

- **SMILES to Hits**

  2D Sketch Program

  ![2D Sketch](image)

  
  \[
  \text{Me} \quad \text{O} \\
  \text{iPr} \quad \text{OH}
  \]

  →

  Simplified Molecular Input Line Entry Specification (SMILES)

  \[
  \text{OC}([\text{C@H}(\text{C})\text{C}1=\text{CC}=\text{C}(\text{CC}(\text{C})\text{C}=\text{C}1)=\text{O}}
  \]

  convert to 3D structure

  Enzyme substrate complex binding score

  ![Enzyme substrate complex](image)

  →

  ![AutoDock](image)
**In Silico Drug Design and Development**

**ZINC database**

- **ZINC**: a free public resource for ligand discovery. The ZINC database contains 120 million commercially available "drug like" molecules in biologically relevant 3D representations that can be downloaded in popular ready-to-dock formats.

- Designed for investigators who are not computer specialists


http://zinc.docking.org/
**PubChem** is a database containing compounds and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information (NCBI).

- Contains over 7.7 million compounds
- 1.2 million bioactivity results
- Over 2.1 million compounds tested
- Over 10 thousand protein targets

In Silico Drug Design and Development

summary

**In silico drug design advantages and limitations:**

**Advantages:**

- Inexpensive
- Low waste generation
- Inconceivable amounts of chemical space can be investigated in seconds

**Disadvantages/Challenges:**

- Potential problems with synthetic accessibility
- Addressing receptor flexibility (reverse docking) is a major challenge
- Developing new filters for removing promiscuous binders and reactive inhibitors

"Given the structure of a protein and that of a potential ligand, can the two form a favorable complex? What are the bases for binding and specificity?"

Probably...Yes!