Structure-based drug design

Pavel Polishchuk

Institute of Molecular and Translational Medicine
Faculty of Medicine and Dentistry
Palacky University

pavlo.polishchuk@upol.cz
qsar4u.com
Biological networks

Lu Han et al. International Journal of Molecular medicine, 2014, 33, 581-588
Drug targets

Target identification & validation

Hit identification

Lead identification

Lead optimization

Preclinical studies

Clinical studies

Drug development stages
Computer-aided drug design (CADD)

<table>
<thead>
<tr>
<th>Protein structure</th>
<th>Unknown</th>
<th>Known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>CADD of no use</td>
<td>De novo design</td>
</tr>
<tr>
<td>Known</td>
<td>Ligand-based design</td>
<td>Structure-based design</td>
</tr>
<tr>
<td>Ligand structure</td>
<td>QSAR, pharmacophore modeling, similarity searching</td>
<td>Docking, molecular dynamics, pharmacophore modeling</td>
</tr>
</tbody>
</table>
Protein Data Bank (PDB)

http://www.pdb.org/pdb/home/home.do

<table>
<thead>
<tr>
<th>Exp. Method</th>
<th>Proteins</th>
<th>Nucleic Acids</th>
<th>Protein/NA Complexes</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-RAY</td>
<td>102721</td>
<td>1771</td>
<td>5255</td>
<td>4</td>
<td>109761</td>
</tr>
<tr>
<td>NMR</td>
<td>10163</td>
<td>1177</td>
<td>236</td>
<td>8</td>
<td>11584</td>
</tr>
<tr>
<td>ELECTRON MICROSCOPY</td>
<td>845</td>
<td>30</td>
<td>280</td>
<td>0</td>
<td>1155</td>
</tr>
<tr>
<td>HYBRID</td>
<td>92</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>other</td>
<td>178</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>201</td>
</tr>
<tr>
<td>Total</td>
<td>113999</td>
<td>2985</td>
<td>5789</td>
<td>26</td>
<td>122799</td>
</tr>
</tbody>
</table>

Yearly Growth of Total Structures

number of structures can be viewed by hovering mouse over the bar
Pharmacophores
A **pharmacophore** is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interaction with a specific biological target structure and to trigger (or block) its biological response.

Atom- and pharmacophore-based alignment

Methotrexate

Dihydrofolate

Hydrogen bonding patterns

Atom-based alignment

Pharmacophore alignment
GRID interaction fields: Convert regions of high interaction energy into pharmacophore point locations & constraints
[S. Alcaro et al., Bioinformatics 22, 1456-1463, 2006]

Start from target-ligand complex: Convert interaction pattern into pharmacophore point locations & constraints
Grid-based pharmacophore modeling

Step 1
- 3D complex

Step 2
- Map C
  - Sub α
  - Sub β
  - Sub α + β

Step 3
- Map E
  - Map D
  - Map A
  - Map B

Step 4
- Single feature
  - Step 5
  - Multiple features
  - Step 6

Step 5
- Single feature

Step 6
- Multiple features

Feature-based pharmacophores (LigandScout)

PDB code: 2VDM

H-bonds formed by the ligand
Hydrophobic interaction

Pharmacophore features
- H-bond donor
- H-bond acceptor
- Positive ionizable
- Negative ionizable
- Hydrophobic
Shared consensus pharmacophore (LigandScout)

RET Kinase Inhibitors
Merged consensus pharmacophores (LigandScout)

RET Kinase Inhibitors

2IVV

2IVU
Pharmacophore applications

Virtual screening
Compound profiling
Library design
Compound profiling

Pharmacophore models

Ligands

Fragment-based library design

Antagonists of thromboxane A2 receptor

Khrustova T., et al., Reports of NAS of Ukraine, 2014, 103-08
Molecular dynamics & pharmacophores (LigandScout)
Virtual screening with dynamic pharmacophores

Capture interactions at different points along the trajectory

Model Cluster 1

Model Cluster 2

Model Cluster 3

Virtual Screening/Hit Finding

Alignment for Lead Optimization

provided by Prof. T. Langer
Pharmacophore vector

Table:

| AR2 | AR1 | H1  | H5  | H6  | H4  | H3  | H2  | HBA1 | HBA9 | HBA8 | HBA6 | HBA5 | HBA4 | HBA3 | HBA2 | HBD1 | PI1 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
## Pharmacophore vectors

<table>
<thead>
<tr>
<th>AR2</th>
<th>AR1</th>
<th>H1</th>
<th>H5</th>
<th>H6</th>
<th>H4</th>
<th>H3</th>
<th>H2</th>
<th>HBA1</th>
<th>HBA9</th>
<th>HBA8</th>
<th>HBA6</th>
<th>HBA7</th>
<th>HBA5</th>
<th>HBA4</th>
<th>HBA3</th>
<th>HBA2</th>
<th>HBD1</th>
<th>PI1</th>
<th>#p4</th>
<th>% tf</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1818</td>
<td>9,1%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>826</td>
<td>4,1%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>805</td>
<td>4,0%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>683</td>
<td>3,4%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>455</td>
<td>2,3%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>451</td>
<td>2,3%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>424</td>
<td>2,1%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>384</td>
<td>1,9%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>363</td>
<td>1,8%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>353</td>
<td>1,8%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>332</td>
<td>1,7%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>325</td>
<td>1,6%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>259</td>
<td>1,3%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>236</td>
<td>1,2%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>217</td>
<td>1,1%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>205</td>
<td>1,0%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>180</td>
<td>0,9%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>176</td>
<td>0,9%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>169</td>
<td>0,8%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>169</td>
<td>0,8%</td>
</tr>
</tbody>
</table>

---

**Pharmacophore vectors provided by Prof. T. Langer**

### MD pharmacophore feature
- **AR2**
- **H5**
- **H6**
- **HBA1**
- **HBA2**
- **HBA3**
- **HBA4**
- **HBA5**
- **HBA6**
- **HBA7**
- **HBA8**
- **HBA9**
- **HBD1**
- **PI1**

### PDB pharmacophore feature
- **AR1**
- **H1**
- **H4**
- **H2**

### PDB initial pharmacophore
- **#p4**
- **% tf**

---

*provided by Prof. T. Langer*
Pharmacophores: conclusion

+ Universal representation of binding pattern
+ Qualitative output
+ Very fast screening
+ Scaffold hopping

- Structure-based models can be very specific
Molecular docking
Docking is an *in silico* tool which predicts

**Pose** – a possible relative orientation of a ligand and a receptor as well as conformation of a ligand and a receptor when they are form complex

**Score** – the strength of binding of the ligand and the receptor.

---

Complex 3D jigsaw puzzle
Conformational flexibility
Mutual adaptation (“induced fit”)
Solvation in aqueous media
Complexity of thermodynamic contribution
No easy route to evaluation of $\Delta G$

Simplification and heuristic approaches are necessary

“At its simplest level, this is a problem of subtraction of large numbers, inaccurately calculated, to arrive at a small number.”

Protein-ligand docking software consists of two main components which work together:

1. **Search algorithm (sampling)** - generates a large number of poses of a molecule in the binding site.

2. **Scoring function** - calculates a score or binding affinity for a particular pose.
## Search algorithms (sampling)

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Receptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigid</td>
<td>Rigid</td>
</tr>
<tr>
<td>Flexible</td>
<td>Rigid</td>
</tr>
<tr>
<td>Flexible</td>
<td>Flexible</td>
</tr>
</tbody>
</table>

- **Fast & Simple** (Rigid, Rigid)
- **Slow & Complex** (Flexible, Flexible)
Rigid docking: DOCK

Binding site is filled with spheres ("negative image") and the spheres centers are matched to ligand atoms.
Flexible ligand docking

Systematic search (exhaustive)

- Genetic algorithm (GOLD)
- Monte Carlo (MOE)

Stochastic search

- Simulated annealing (AutoDock)
- Ant colony (PLANTS)

Incremental search (FlexX)

- Genetic algorithm (GOLD)
- Simulated annealing (AutoDock)
- Ant colony (PLANTS)

•••
Ligand flexibility: Incremental search
“Induced fit”
the protein may deform slightly to accommodate different ligands

Expand protein conformational space
generation of possible conformation and dock to them
The ultimate goals of an ideal scoring function:

accurate within less that 1 pK$_D$ unit (<1.4 kcal/mol)

generally valid (non system specific, large affinity range)

robust (tolerant with respect to the structural uncertainties)

physically meaningful (interpretable)

fast and easy to compute
Classes of scoring function

Forcefield-based
Based on terms from molecular mechanics forcefields
GoldScore, DOCK, AutoDock

Empirical
Parameterised against experimental binding affinities
ChemScore, PLP, Glide SP/XP

Knowledge-based potentials
Based on statistical analysis of observed pairwise distributions
PMF, DrugScore, ASP
Force field-based methods

Molecular Mechanics (MM):
- atoms $\rightarrow$ charged spheres
- bonds $\rightarrow$ springs
- classical potentials
- no electrons $\rightarrow$ no bond formation / cleavage
- typically parameterized to reproduce molecular potential energy surface ($\rightarrow$ conformational $\Delta H$ in the gas phase!)

Scoring protein-ligand complexes:
+ for pose prediction in docking
- for ligand ranking by affinity

Terms accounting for (de)solvation & entropic factors required (cf. MM-PBSA)
Typical force-field scoring function

\[ E = \sum_{i} \sum_{j} \left( \frac{q_i q_j}{\varepsilon_{ij} r_{ij}} + \frac{A_{ij}}{r_{ij}^6} - \frac{B_{ij}}{r_{ij}^{12}} \right) + \]

\[ + \Delta G_{HB} \sum \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^6} \right) + \]

\[ + \Delta G_{tor} N_{rot} + \]

\[ + \Delta G_{sol} \]
Empirical scoring function

Empirical scoring functions

Regression-based:

\[ pKi = \sum pKi_n f_n(\text{structure}) \]

- affinity
- weighting factors
- structure descriptors

determined via regression analysis (MLR, PLS)

Data:

- Experimental binding affinities
- Experimental structures
Böhm’s empirical scoring function

\[ \Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h–bonds}} f (\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic interactions}} f (\Delta R, \Delta \alpha) + \Delta G_{\text{lip}o} |A_{\text{lip}o}| + \Delta G_{\text{rot}}^{NROT} \]

**H-bonding** and **ionic terms** are dependent on geometry of interactions, large deviation in distance and angles are penalized.

**Lipophilic term** is proportional to the contact surface area involving non-polar atoms.

**Conformational entropy term** is proportional to the number of rotatable bonds.

Knowledge-based scoring function

Derivation from crystal-structure data

\[ P_{ij}(r) = - \ln \frac{g_{ij}(r)}{g_{ref}} \]

- \( P_{ij} \): distance-dependent pair potential
- \( g_{ij} \): frequency distribution of atom-atom contacts
- \( g_{ref} \): reference distribution

No experimental affinities used!

Statistical potential

Frequency of occurrence
Docking quality assessment

Pose reproducibility

Self-docking / re-docking

RMSD
Docking quality assessment

Energy reproducibility

Test set size = 800 molecules

HIV-1 protease

\[ r < 0.55 \]

**Docking quality assessment**

**Antagonists of $\alpha_{\text{IIb}\beta3}$**

<table>
<thead>
<tr>
<th>Activity, pIC$_{50}$</th>
<th>PLANTS</th>
<th>S4MPELE</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.77</td>
<td>-77</td>
<td>-62</td>
</tr>
<tr>
<td>6.82</td>
<td>-85</td>
<td>-69</td>
</tr>
<tr>
<td>7.96</td>
<td>-88</td>
<td>-74</td>
</tr>
<tr>
<td>5.85</td>
<td>-87</td>
<td>-76</td>
</tr>
<tr>
<td>5.89</td>
<td>-93</td>
<td>-81</td>
</tr>
</tbody>
</table>

## Docking validation

**MOE 2010.10**
- Stochastic search
- Force-field based scoring function

**FlexX**
- Incremental search
- Empirical scoring function

**PLANTS**
- Stochastic search
- Empirical scoring function

<table>
<thead>
<tr>
<th>Binding site</th>
<th>ligand</th>
<th>MOE</th>
<th>FlexX</th>
<th>PLANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tirofiban</td>
<td>1.15</td>
<td>3.55</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>L-739758</td>
<td>2.14</td>
<td>3.39</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td>Eptifibatide</td>
<td>3.20</td>
<td>5.81</td>
<td>5.74</td>
</tr>
<tr>
<td>2VDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2VC2</td>
<td></td>
<td>1.37</td>
<td>1.85</td>
<td>8.87</td>
</tr>
<tr>
<td>2VDN</td>
<td></td>
<td>1.72</td>
<td>2.52</td>
<td>4.46</td>
</tr>
</tbody>
</table>

Polishchuk, P. G. et al., *Journal of Medicinal Chemistry*, **2015**, 58, 7681-7694
Docking protocol

Check structures and their tautomeric state

Protonation states and H-bonding network

Energy minimization

Consider water molecules: keep or remove them

Validate your protocol and chosen program on known ligands to reproduce poses and binding energies
Conclusion

+ Relatively fast
+ Determine binding poses & energy
+ Good in ranking ligands for virtual screening

- Low accuracy of binding energy estimation
- Require knowledge about binding site

Literature:

**Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions**
I. Halperin, B. Ma, H. Wolfson, and R. Nussinov

**A review of protein-small molecule docking methods**
R. D. Taylor, P. J. Jewsbury & J. W. Essex
*Journal of Computer-Aided Molecular Design*, 16: 151–166, 2002

**A Critical Assessment of Docking Programs and Scoring Functions**
Binding site identification
## Druggable & non-druggable sites

<table>
<thead>
<tr>
<th></th>
<th>Druggable</th>
<th>Non-druggable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>bigger</td>
<td>smaller</td>
</tr>
<tr>
<td>Complexity</td>
<td>more complex</td>
<td>less complex</td>
</tr>
<tr>
<td>Shape</td>
<td>more compact</td>
<td>longer and narrower</td>
</tr>
<tr>
<td>Exposure</td>
<td>less expose</td>
<td>more expose</td>
</tr>
<tr>
<td>Proportion of hydrophobic surface</td>
<td>higher (&gt;70%)*</td>
<td>lower (&lt;50%)</td>
</tr>
</tbody>
</table>

* Hydrophobic interactions mainly contribute to the binding energy.
Druggability assessment methods

Empirical model trained on NMR-determined hit rates of 10000 compounds on 28 binding sites of 23 proteins

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>total surface area</td>
<td>2.78</td>
</tr>
<tr>
<td>polar contact area</td>
<td>-0.44</td>
</tr>
<tr>
<td>apolar contact area</td>
<td>2.98</td>
</tr>
<tr>
<td>first principal moment</td>
<td>-1.03</td>
</tr>
<tr>
<td>third principal moment</td>
<td>1.2</td>
</tr>
<tr>
<td>pocket compactness (volume/surface area)</td>
<td>13.6</td>
</tr>
</tbody>
</table>

$R^2 = 0.72$
$Q^2_{LOO} = 0.59$

Druggability assessment methods

**SiteMap (Dscore)**

$$D\text{score} = 0.094n^{1/2} + 0.60e - 0.324p$$

- $n$ - the number of site points found for the site, capped at 100,
- $e$ - the degree of enclosure of the site,
- $p$ – the hydrophilic score computed for the site.

$$\text{drugscore} = \frac{e^{-z}}{1 + e^{-z}}$$

$$z = \beta_0 + \beta_1 f_1(d_1) + \beta_2 f_2(d_2) + \beta_3 f_3(d_3)$$

**fpocket**

- $d_1$ - local hydrophobicity density
- $d_2$ - hydrophobicity score
- $d_3$ - normalized polarity score

**DoGSiteScorer**

SVM model based on volume, area, polarity, shape, depth, etc.
Similarity search based on nearest neighbors

---

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein class</td>
<td>Druggability is assumed from sequence or fold-similarity to known targets</td>
<td>Fast and simple</td>
<td>Druggability may be different between close homologues. Not applicable to novel targets</td>
</tr>
<tr>
<td>Empirical</td>
<td>Cavities on protein surfaces are detected and druggability is evaluated by a trained function</td>
<td>Fast, some programs are freely available</td>
<td>Applicability is limited by the training set</td>
</tr>
<tr>
<td>Simulation based</td>
<td>The affinity of probe molecules to the protein is calculated from molecular simulations</td>
<td>Can be applied to any system</td>
<td>Slower and more complex to use</td>
</tr>
</tbody>
</table>
Blind docking & site identification

Hetenyi C. and van der Spoel D., Protein Science, 2011, 20, 880-893

1) if BD and PS approaches result in the same pocket it is reliable;
2) if only BD found the site and not PS this may be an error;
3) if several PS approaches result in the same pocket but not BD, that can be a suggestion to repeat docking in identified pocket.

BD – blind docking
PS – pocket search
Homology modeling
Only small portion of protein structures were resolved and many more protein 3D structures are still unknown. It happens due to difficulties in protein crystallization for X-ray analysis, poor solubility of proteins for NMR experiments.

Assumptions:
1) 3D structure of a protein is determined by its amino acid sequence
2) proteins with high sequence identity/similarity will have similar 3D structures

What is homology modeling?

Predicts the three-dimensional structure of a given protein sequence (target) based on an alignment to one or more known protein structures (templates).
Steps:

1) Template recognition & initial alignment
2) Alignment correction
3) Backbone generation
4) Loop modeling
5) Side-chain modeling
6) Model optimization
7) Model validation
Homology modeling: prerequisites

- Comparable to medium resolution X-ray/NMR structure
- Structure based design
- Understanding substrate specificity

- Properties that do not occur in template (e.g. electrostatics, cavity volume)
- Protein engineering (e.g. stability)
- Molecular replacement in X-ray crystallography

- Assignment of function/protein fold
- Focus mutagenesis experiments
- Finding binding/active sites by 3D motif searching
- NMR structure refinement
 Alignment

Align & determine structurally conserved regions (ClustalW, T-Coffee, BLAST, FASTA, etc)

Check alignment:
Secondary structure elements preserved?
Motifs and key residues aligned?
Are gaps in acceptable locations (ie: loops)?
Try to improve alignment manually
Backbone generation

One simply copies the coordinates of those template residues that show up in the alignment with the model sequence.

If two aligned residues differ, only the backbone coordinates (N, Cα, C and O) can be copied.

If they are the same, one can also include the side chain.
1) **knowledge based**: one searches the PDB for known loops with endpoints that match the residues between which the loop has to be inserted and simply copies the loop conformation.

   Advantage: all loops found are guaranteed to have reasonable internal geometries and conformations

   Disadvantage: may not fit properly into the given model protein’s framework

2) **energy based**: as in true *ab initio* fold prediction, an energy function is used to judge the quality of a loop
Model refinement & validation

Molecular mechanics energy minimization

Molecular dynamics

Check the correctness:
overall fold/structure,
errors of localized regions and stereochemical parameters:
bond lengths, angles, geometries
Model validation tools

Molprobity  http://molprobity.biochem.duke.edu/

WHAT IF  http://www.cmbi.kun.nl/gv/servers/WIWWWI/

SOV  http://predictioncenter.llnl.gov/local/sov/sov.html

PROVE  http://www.ucmb.ulb.ac.be/UCMB/PROVE/

ANOLEA  http://www.fundp.ac.be/pub/ANOLEA.html

ERRAT  http://www.doe-mbi.ucla.edu/Services/ERRATv2/

VERIFY3D  http://shannon.mbi.ucla.edu/DOE/Services/Verify_3D/

BIOTECH  http://biotech.embl-ebi.ac.uk:8400/

ProsaII  http://www.came.sbg.ac.at

WHATCHECK  http://www.sander.embl-heidelberg.de/whatcheck/
Automated web-based homology modeling tools


WHAT IF: [http://www.cmbi.kun.nl/swift/servers/](http://www.cmbi.kun.nl/swift/servers/)

The CPHModels Server: [http://www.cbs.dtu.dk/services/CPHmodels/](http://www.cbs.dtu.dk/services/CPHmodels/)

3D Jigsaw: [http://www.bmm.icnet.uk/~3djigsaw/](http://www.bmm.icnet.uk/~3djigsaw/)

SDSC1: [http://cl.sdsc.edu/hm.html](http://cl.sdsc.edu/hm.html)

COMPOSER

MODELER http://salilab.org/modeler

InsightII http://www.msi.com/

SYBYL http://www.tripos.com/
Molecular dynamics
Molecular dynamics simulations permit the study of complex, dynamic processes that occur in biological systems. These include, for example,

- Protein stability
- Conformational changes
- Protein folding
- Molecular recognition: proteins, DNA, membranes, complexes
- Ion transport in biological systems

and provide the mean to carry out the following studies,

- Drug Design
- Structure determination: X-ray and NMR
Biological molecules exhibit a wide range of time scales over which specific processes occur; for example:

- Local Motions (0.01 to 5 Å, $10^{-15}$ to $10^{-1}$ s)
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- Rigid Body Motions (1 to 10 Å, $10^{-9}$ to 1 s)
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions
- Large-Scale Motions (> 5 Å, $10^{-7}$ to $10^{4}$ s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding
Classical mechanics

\[ x = x_0 + v_0 t + \frac{at^2}{2} \]

\[ a = -\frac{1}{m} \frac{dU}{dr} \]

\[
U(\vec{R}) = \sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2 + \sum_{\text{dihedrals}} k_i^{\text{dihedral}} [1 + \cos (n_i \phi_i + \delta_i)] + \sum_{i \neq j} 4 \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{i \neq j} \frac{q_i q_j}{\epsilon r_{ij}}
\]

\( U_{\text{bond}} \) = oscillations about the equilibrium bond length
\( U_{\text{angle}} \) = oscillations of 3 atoms about an equilibrium angle
\( U_{\text{dihedral}} \) = torsional rotation of 4 atoms about a central bond
\( U_{\text{nonbond}} \) = non-bonded energy terms (electrostatics and Lenard-Jones)
Typical workflow

to remove any strong van der Waals interactions

solvate system (periodic box) and
assign velocities from Gaussian distribution
Molecular dynamics: shortcomings

- Quality of the forcefield

- Size and Time – atomistic simulations can be performed only for systems of a few tenths of angstroms on the length scale and for a few nanoseconds on the time scale

- Conformational freedom of the molecule – the number of possible conformations a molecule can adopt is enormous, growing exponentially with the number or rotatable bonds.

- Only applicable to systems that have been parameterized

- Connectivity of atoms cannot change during dynamics – no chemical reactions
PhD in chemoinformatics