Docking
Benzamidine Docking to Trypsin

Energy = -51.01 kcal/mol  Cluster Rank = 1
RMSD = 0.23 Angstrom
Relationship to Drug Design

- **Ligand-based design**
  - QSAR
  - Pharmacophore modeling
  - Can be done without 3-D structure of protein

- **Receptor/Structure-based design**
  - Molecular docking
    - Hit identification
    - Lead identification and optimization
    - ADMET (e.g., cytochrome P450 docking)
  - *De novo* design
  - Requires 3-D model of protein
Overview

- Protein-small molecule docking
- Protein-protein docking
Basic Principles

\[ \Delta G = -RT \ln K_A \]

\[ K_A = K_i^{-1} = \frac{[EI]}{[E][I]} \]

Aspects of Docking

• Docking algorithms need to generate poses of ligands in binding site
• Then score the poses - simple energy function
• Then rank the poses - more complex energy function -> $\Delta G$
  – Desolvation
  – Entropy
Additional Challenges in Docking

• Limited resolution of receptor structures
• Flexibility of receptor
• Conformational changes in ligand and/or receptor upon binding
• Role of water molecules in binding
  – Desolvation of ligand and binding site
  – Waters bridging protein and ligand
Receptor Representations for Docking

- Atomic - usually only in final ranking in conjunction with potential energy function
- Surface - Connolly’s MS program
- Grid

Ligand Flexibility

- Systematic search
- Random (Stochastic) search
- Simulation methods
Systematic Search

- Conformational search methods (combinatorial explosion)
- Fragmentation methods
  - Dock molecular fragments and link them covalently
- Database methods
  - Libraries of pre-generated conformers
  - e.g., OMEGA - eliminates those that are close in RMSD or have high internal strain
Random (Stochastic) Search

- Monte Carlo methods
- Tabu search - take into consideration areas of conformational space already explored (using RMSD)
- Genetic algorithms - adapt principles of biological competition and population dynamics
Genetic Algorithm in Docking

<table>
<thead>
<tr>
<th>Rotation 1</th>
<th>Translation 1</th>
<th>Conformation 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation 2</td>
<td>Translation 2</td>
<td>Conformation 2</td>
</tr>
<tr>
<td>Rotation n</td>
<td>Translation n</td>
<td>Conformation n</td>
</tr>
</tbody>
</table>

Evaluate energies and take best scoring

Crossover →

R1  T1  C50 → Mutation

R1  T5  C50
Simulation Methods

- Molecular dynamics
- Elevated temperatures
- Different starting positions for ligand
- Generate ensemble of conformations
Protein Flexibility

- MD/MC methods - time-consuming
- Rotamer libraries - neglects changes in backbone
- Protein ensemble grids - multiple structures from crystallography, NMR, or simulation
- Soft-receptor modeling - “energy-weighted” average grid based on multiple conformations
Scoring

• Need to differentiate correct poses from “incorrect” ones
  – RMSD between experimentally observed and docked positions
• Calculates an “energy”
• Ideal is to have lowest energy pose be the one with the lowest RMSD relative to experimental pose
Scoring Functions

• Force field-based
• Empirical scoring
• Knowledge-based scoring
• Consensus scoring
Force Field-Based Scoring

• Calculates interaction energy between protein and ligand and the internal energy of ligand
• Pairwise interactions for van der Waals and electrostatics
• May also have explicit term for H-bonds
• Does not inherently include solvation and entropy
Empirical Scoring Functions

- Designed to reproduce experimental data (e.g., binding energies)
- Sum of several uncorrelated terms (e.g., H-bond, ionic, hydrophobic, aromatic)
- Coefficients of terms calculated from regression analysis using experimental data => dependent on quality and quantity of data
- Can include terms for entropy and solvation
Knowledge-Based Scoring Functions

• Designed to reproduce experimental structures rather than binding energies
• Simple atomic interaction-pair potentials based on frequency of occurrence of different pair contacts in dataset
• Simple, fast to calculate
• Limited by quantity and quality of data
Consensus Scoring

• Combine information from different scoring methods
• Need to consider correlation of different scoring functions
Evaluating Scoring Schemes

- Several recent studies
- Ideal result is correct redocking of ligand (< 2Å RMSD) and having lowest energy pose as best docked
- Most programs able to predict accuracy of ~1.5-2Å with 70-80% success
- No single scoring function superior to others
- Scoring functions respond differently to features in binding sites
- Consensus scoring gaining favor
Posing vs Scoring

- Where are calculation errors greater?
- No comprehensive study yet available
- Consensus is that scoring represents the major limiting factor
Improving Scoring Functions

- Include solvation and rotational entropy contributions
- Tune scoring schemes on case-by-case basis
  - Example: cAMP-dependent kinase

Staurosporine binds at Mg-ATP site => had to balance energy functions for hydrophobic, H-bond, ionic interactions

Examples of Docking Programs

De Novo Design