Overview

- Applications
- Docking Algorithms
- Scoring Functions
- Results
- Demonstration
Docking Applications

Drug Design
• Lead Generation
• Lead Optimisation
• Library Design
• Compound Purchases

Academic
• Predicting Crystal Structure Complexes
Lead Generation

• Alternative to Experimental HighThroughput Screening

• Issues
  – Availability of Crystal Structures
  – False Positives

• Often Difficult to Demonstrate Cost Benefit
Lead Optimisation

• Ranking Derivatives of Known Active Molecules
• Subject to Systematic Failures
• Not Sufficiently Discriminating for Many Optimisation Series
• Changing Chemical Series is really Lead Generation
Library Design

- Aim to Eliminates Molecules which not Credible
- Receptor Shape is Important
- Can be Lead Optimisation when only Substituents on an Active core vary
Compound Purchases

- Millions of Catalogue Molecules
  - Diverse
  - Approximately 0.5 Million Drug-like
- Issues
  - False Negatives
  - Delivery Timescales
- Cost Benefit is Demonstratable
Drug-like Filtering

- **Properties**
- **Substructures**
  - Unstable
  - Reactive
  - Undesirable (eg toxic)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centres</td>
<td>2:9</td>
</tr>
<tr>
<td>Mass</td>
<td>150:800</td>
</tr>
<tr>
<td>XSA</td>
<td>20:240</td>
</tr>
<tr>
<td>Rot-Bonds</td>
<td>&lt;=10</td>
</tr>
<tr>
<td>Conformers</td>
<td>&lt;=1000000</td>
</tr>
</tbody>
</table>
Classifying Docking Algorithms

• Ligand Conformations
  – Rigid
  – Fixed Sample ~ 100
  – Flexible

• Constraints
  – Residues and Pockets
  – Pharmacophores

• Charges and Tautomers

• Water Molecules
Conformations

- Small Molecules are NOT Rigid
- Literature on Conformational Generation
  - Use of Constraints
  - Speed Enhancements
- Often 000’s Representative Conformations
- Suitable Problem for Parallel Computing
Constraints: Binding Site

- Identified by Crystal Structure
  - Activity associated with another site
  - Different binding mode at same site
- Confirmed by Mutagenesis Studies
- Suggested by Software
  - Size of cavity/cleft
  - Binding Potential
Binding Site Identification

- Place Protein in 3-D Grid
- Remove Grid Points Inside Protein
- Remove Grid Points Inside 8 Å Sphere Positioned on the Grid Outside the Protein
- Binding Site
  - Minimum of 3 Grid Points
  - Minimum of 3 Residues
Constraints: Pharmacophores

- Centre Types
  - H-bond Donor
  - H-bond Acceptor
  - Acid
  - Base
  - Positive Charge
  - Negative Charge
  - Aromatic Ring
  - Lipophile
  - Lewis Base
  - 2 User Definable

- Options
  - 3 or 4 Centre Types
  - User Definable Bins
  - Occurrence Frequency

- Output to file
Centre Positions

3.0 H-bond distance
1.0 Tolerance
Charges and Tautomers

• Docking Doesn’t Need Hydrogen Positions
• Charges Important for Energy Functions
  – Sometimes Alternative Charge Models
  – Rarely Multiple Same Charges Within Site
• Tautomer State
  – Sometimes Too Many Alternatives
• Charges and Tautomeric State Must Complement Ligand
Water Molecules

• Essential
  – Mediates binding for all Ligands

• Optional
  – Presence required by some Ligands
  – Inhibits binding of other Ligands

• Solvation & Desolvation Free Energy
  Critical for Scoring Function
Structure-based Virtual Screening

Suppliers Catalogues

Combinatorial Libraries

Property & Substructure Filtering

Conformer Generation

Dock & Score

Saved Hits

Protein Active Site

Generation of Queries

Pharmacophores

CAN-DDO Project
• 3.5 Billion Molecules
• 12 Proteins
• 1.7 Million PCs
2D -> 3D Coordinates

Pharm list by permuting centres

Centre & distance constraints

Match Pharmacophore?

Yes

No

Generate next conformer

Saved Solution

Dock by Fitting to Pharm
2D -> 3D Coordinates → Pharm list by permuting centres → Centre & distance constraints

Generate next conformer

VdW or CPK Contacts?
- No
- Yes → Deduce bond to relieve distance constraint

Skip intervening conformers

Match Pharmacophore?
- No
- Yes → Dock by Fitting to Pharm

Saved Solution

Refine & Score
Limitations of Current Methods

- Rigid Protein Side-Chains
- Binding Relevance for Biological Activity
- Scoring Functions
Terms in Scoring Functions

- Time Required
  - Entropy
  - (De)Solvation

- Level of Approximation
  - Hydrophobic
  - Electrostatics
  - H-bonding
  - VdW
  - Contacts
Classifying Scoring Functions

- Knowledge-based (Atom pairs in contact)
  - DrugScore, PMF
- Energy
  - GOLD, DOCK, LigandFit, MOE
- Energy + Parameterised Solvation
  - ChemScore (Glide, THINK)
- Free Energy Perturbation
Knowledge-Based Functions

Score = $\sum_{r<\text{cutoff}} A_{ij} (r)$

- Potentials of Mean Force (PMF)
- DrugScore
- Less Confused by Crystal Structure Precision
Energy

• Lennard Jones
  \[ \sum_{ij} \left( \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) \]

• Torsion Term
  \[ \sum (1 - \cos^2 \omega) \text{ Conjugated} \]
  \[ \sum (1 + \cos^3 \omega) \text{ Non-conjugated} \]

• Electrostatics
  \[ \sum_{ij} q_i q_j / \varepsilon r_{ij} \]
Enhanced ChemScore

\[ \Delta G = \Delta G_0 + \Delta G_{\text{hbond}} \times N_{\text{hbond}} + \Delta G_{\text{lipo}} \times N_{\text{lipo}} + \Delta G_{\text{bad}} \times N_{\text{bad}} + \Delta G_{\text{rot}} \times N_{\text{rot}} + E \]

where

\[ \Delta G_0, \Delta G_{\text{hbond}}, \Delta G_{\text{lipo}}, \Delta G_{\text{bad}}, \Delta G_{\text{rot}} \]
are constants

\(-5.48; -3.34; -0.117; 0.058; 2.56\)

\[ N_{\text{hbond}} \]
is the number of interactions (using geometric criteria)

\[ N_{\text{lipo}} \]
is the number of lipophilic contacts (cf PMF, DrugScore)

\[ N_{\text{bad}} \]
is the number of lipophilic-hydrophilic contacts (extension)

\[ N_{\text{rot}} \]
is the number of frozen rotatable bonds in the ligand

\[ E \]
is the VdW interaction energy and ligand torsional energy (extension)
Free Energy Perturbation

$$\Delta G_{\text{sol}} = - \Delta G_{\text{sol}} (\text{ligand}) - \Delta G_{\text{sol}} (\text{protein})$$
$$+ \Delta G_{\text{gas}} + \Delta G_{\text{sol}} (\text{complex})$$

- Error Prone due to Subtraction of Large Numbers
- Solvent Accessible Surface Area (SASA) approximation (cf ChemScore)
General Observations

- Single Electrostatic and Tautomer models have Systematic Failures
- Energy Inadequate for Ranking Series of Diverse Molecules
- Free Energy Perturbation Methods Slow
Excuses for Inaccuracies

- Rigid Side-Chains
- Estimation of Solvation Effects
- Precision of Force Fields
- Biologically Non-relevant Binding
- Kinetics vs Thermodynamics
Ostriches

- Ignoring Fundamental Theory
  - PMF
  - DrugScore
- Omitting Geometry Refinement
- Rigid and Semi-Rigid Ligands
  - Dock
  - LigandFit
- Biased Validations
  - J Med Chem (2005)
Performance

- **THINK 1.03** (used for CAN-DDO)
  - 42,000,000,000 molecules
  - 126,000 years
  - 900 molecules per CPU day (excluding redundancy)

- **THINK 1.30** (current release)
  - Optimised with assistance from Intel
  - Up to **100** times faster
  - More centres useful for larger sites
  - Refinement of docked geometry
  - About **500,000** molecules per 2GHz CPU day
Validation and Results

• Reproduce Ligand-Protein Crystal Structures
  – RMS Deviation of non-H Atoms
  – Docking Score

• Dock Actives
  – Used for Developing Scoring Functions

• Prediction
  – Enrichment over Random
  – Percentage of False Positives
Selection Criteria

- Possible Kinases Cancer Targets
- X-ray Crystal Structures in PDB
- Resolution (1.9-2.8)
- All Atoms (1IAN Cα only)
- Ligand Flexibility <=10 Rotatable Bonds (excludes 1GAG, 1IR3, 2FGI, 5TMP, 1LCK)
- 21 Structures Processed
<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Score</th>
<th>RMS</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2KI5</td>
<td>-52.2 (-39.8)</td>
<td>5.88 (5.32)</td>
<td>3 A S</td>
</tr>
<tr>
<td>1QHI</td>
<td>-78.4 (-73.5)</td>
<td>1.03 (0.98)</td>
<td>3 A</td>
</tr>
<tr>
<td>1STC</td>
<td>-105.7</td>
<td>0.59</td>
<td>2</td>
</tr>
<tr>
<td>1E8Z</td>
<td>-86.8</td>
<td>1.77</td>
<td>2</td>
</tr>
<tr>
<td>1AQ1</td>
<td>-87.0</td>
<td>0.89</td>
<td>2</td>
</tr>
<tr>
<td>1AGW</td>
<td>-60.0 (-53.9)</td>
<td>3.45 (0.65)</td>
<td>2</td>
</tr>
<tr>
<td>1FGI</td>
<td>-83.8 (-69.4)</td>
<td>0.56 (0.52)</td>
<td>2</td>
</tr>
<tr>
<td>1FVV</td>
<td>-81.3 (-73.3)</td>
<td>0.61 (0.56)</td>
<td>3</td>
</tr>
<tr>
<td>1FVT</td>
<td>-69.8 (-58.5)</td>
<td>1.44 (0.72)</td>
<td>3</td>
</tr>
</tbody>
</table>

A  All Site Points  1  Two Centre Fit
S  Single Bond Increment  3  Three Centre Fit
C  Conjugated Bond Increment  W  Water Site Point
<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Score</th>
<th>RMS</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DI8</td>
<td>-55.2 (-52.9)</td>
<td>0.98 (0.70)</td>
<td>2 W</td>
</tr>
<tr>
<td>1DI9</td>
<td>-39.5</td>
<td>1.97</td>
<td>2 W C</td>
</tr>
<tr>
<td>1YDR</td>
<td>-60.1 (-43.9)</td>
<td>3.00 (2.33)</td>
<td>2</td>
</tr>
<tr>
<td>1YDS</td>
<td>-55.5 (-48.1)</td>
<td>2.80 (0.76)</td>
<td>2 W</td>
</tr>
<tr>
<td>1YDT</td>
<td>-36.8</td>
<td>2.86</td>
<td>2</td>
</tr>
<tr>
<td>1PME</td>
<td>-28.4</td>
<td>1.30</td>
<td>3 W*2 C</td>
</tr>
<tr>
<td>4ERK</td>
<td>-17.9 (-11.8)</td>
<td>7.83 (3.33)</td>
<td>2</td>
</tr>
<tr>
<td>1IEP</td>
<td>-90.6 (-84.8)</td>
<td>0.73 (0.69)</td>
<td>3</td>
</tr>
<tr>
<td>1FPU</td>
<td>-47.1</td>
<td>0.75</td>
<td>3 C</td>
</tr>
<tr>
<td>1DM2</td>
<td>-54.0</td>
<td>0.61</td>
<td>3</td>
</tr>
<tr>
<td>1CKP</td>
<td>-43.2 (-43.0)</td>
<td>1.99 (1.10)</td>
<td>2 C</td>
</tr>
<tr>
<td>1QCF</td>
<td>-58.8</td>
<td>1.33</td>
<td>2</td>
</tr>
</tbody>
</table>
## Find-a-Drug Cancer Results

<table>
<thead>
<tr>
<th>PDB Code</th>
<th>Protein</th>
<th>Number Tested</th>
<th>Number Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>1FLT</td>
<td>VEGFr</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>821P</td>
<td>RAS</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>1C1Y</td>
<td>RAF</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1FGI</td>
<td>FGFr-1</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>1E7U</td>
<td>PI3K</td>
<td>47</td>
<td>5</td>
</tr>
</tbody>
</table>
Find-a-Drug Project

• Processed
  – 270+ protein targets
  – 60+ billion molecules (40-500 million per query)

• Validation using NCI Cancer Data
  – 20% true positive
  – >10x enrichment
• Download THINK software
  www.treweren.com

• Virtual Screening
  Keith.Davies@treweren.com