A fast flexible docking method using an incremental construction algorithm. (M. Rarey et al, `96)

Jang Lee
1690 Structural bioinformatics
Mar 29th `06
Protein-ligand docking

- Target proteins are typically enzymes
- It has catalytic pocket where substrate is bound and catalyzed
- Compounds of small molecular weight with steric and chemical complementarity can bind into the active site and inhibit its enzymatic action
- Research and pharmaceutical interest
Features and synthesis of a number of algorithms

- Ligand flexibility – Dock, Kuntz 1982
- Incremental construction - Leach and Kuntz, 1996
- Protein-ligand interaction – LUDI, Bohm 1992
- Pose clustering - Linnainmaa et al., 1988
Confset tree

- Incremental construction: base + additional fragments
- Each node has coordinate information and list of interactions of newly added fragment
- Conformation can be retrieved by traversing the tree
- Interactions are stored using a linked list
- Tail of linked list point to parent node’s interaction list
Base fragment selection

- Ligand is sliced at each rotatable, acyclic bond
  - A set of rigid fragments to be assembled are produced
- Hydrogen bonding has larger directivity than lipophilic interactions
- Fragment rich in atoms capable of hydrogen bonding and with small number of alternative conformations is chosen as base
Base fragment placement

- Geometric hashing
- Pose clustering
  - Delta-compatible
- Clustering – complete linkage
  - Placements within a certain RMSD tolerance are merged into one cluster
- Overlap test
Incremental construction

- Geometrically more restricted fragments are added first – helps reduce search space
- Interactions for the fragment are searched
- Coarsened grid points are used in initial search, and solutions are optimized in subsequent stage
Placement optimization

- Weighted superposition – Kabsch, 1976
- Maximize weight contribution from each atom
- Weight is overlap of occupied sphere with the sphere of ideal interaction center
- Overlap test
- Remove atoms colliding with another
- Second round of superposition
- Include in solution set or reject by presence of steric clashes
Solution set management

- Retain k best solutions
  - Global conformation space can be probed
- Clustering to remove redundancy
- Priority queue is used to keep solutions
- Insertion and deletion are logarithmic in k
Overlap test

- Check if atoms clash with another
- Performed many times and is computationally intensive
- Box hashing and atom caching
- Box hashing: Given an atom’s coordinate and radius, all intersecting boxes are looked up to retrieve enclosed receptor atoms
Atom caching

- All atoms within a sum of van der Waals radius of ligand atom and the largest atom are retrieved and maintained for each query.
- In subsequent query, if the atom’s center has not moved by more than certain distance from the center of previous query, then previous query result can be reused.
Test data

- 19 protein-ligand complexes
- 5 Small ligands
- 2 Short peptides
- Streptavidin, thrombin, ribonuclease, thermolysin, carboxypeptidase, HIV protease, RAS protein
## Small ligands

<table>
<thead>
<tr>
<th>pdb</th>
<th># place.</th>
<th>RMSD(A)</th>
<th>Rank</th>
<th>Ecalc (kJ/mol)</th>
<th>Eexp (kJ/mol)</th>
<th>Error (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5tim</td>
<td>6</td>
<td>0.87</td>
<td>3</td>
<td>-16.3</td>
<td>-13.1</td>
<td>-24.427</td>
</tr>
<tr>
<td>1ldm</td>
<td>28</td>
<td>0.62</td>
<td>1</td>
<td>-32.8</td>
<td>-30.8</td>
<td>-6.494</td>
</tr>
<tr>
<td>2phh</td>
<td>179</td>
<td>0.29</td>
<td>2</td>
<td>-30</td>
<td>-26.7</td>
<td>-12.360</td>
</tr>
<tr>
<td>3ptb</td>
<td>14</td>
<td>0.48</td>
<td>1</td>
<td>-28.3</td>
<td>-27.2</td>
<td>-4.044</td>
</tr>
<tr>
<td>1ulb</td>
<td>134</td>
<td>0.65</td>
<td>1</td>
<td>-17.9</td>
<td>-30.2</td>
<td>40.728</td>
</tr>
</tbody>
</table>
Short peptides

- Interaction centers are scattered around molecule, and base fragment is hard to identify
- 3tpi: 0.58Å, -24 Ecalc vs -24.5 kJ/mol Eexp
- 4ts1: 14th solution. 0.71Å RMSD, -11.8 vs. -32 kJ/mol. Three water molecules
- Performance was reasonable
Dehydrofolate reductase

- Dehydrofolate reductase
- $3.9 \times 10^7$ possible conformations
- Base fragment placement showed 0.29Å RMSD (rank 9)
- Final complex constructed showed 0.9Å RMSD
Streptavidin

- Streptavidin
- Two water molecules form three hydrogen bondings
- Difference of 45.4 kJ/mol (-31.2 Ecalc and -76.6 Eexp)
- Calculated conformation is still close to experimental one despite negligence of water molecules: 0.81Å RMSD
Carboxypeptidase-A

- 3cpa
- Diverse conformations when ranked by energy
- Top two solutions are very distinct in conformation
- High RMSD – 3.08A and 2.62A
HIV protease

- 4phv
- Water molecule is conserved
- 69 potential interactions
- Base placement: 0.85Å RMSD
- 20 final placements are generated
- Final construction: 1.04Å RMSD
- Energy difference: 14kJ/mol
Discussion

- Rigid conformation assumption on receptor protein
- Water molecule cannot be handled unless it’s modeled to be a part of receptor
- Occasionally large energy difference is seen with no apparent reasons
- Possibly more interactions smaller in scale have to be considered