A Motion Planning Approach to Flexible Ligand Binding

Amit P. Singh
Medical Informatics and Biochemistry, Stanford University

Jean-Claude Latombe
Computer Science, Stanford University

Douglas L. Brutlag
Biochemistry and Medical Informatics, Stanford University
What is Motion Planning?
Why Motion Planning?

- Motion planning can:
  - Sample the space of possible paths taken by the ligand as it approaches and binds to the receptor
  - Examine the energetics of the ligand along each of these paths
  - Make estimates of the relative rates of binding and dissociation
  - Identify regions of the protein that are responsible for affecting these rates (i.e. transition states, energy barriers)
Why Robotics?

Ligand \approx \text{Articulated Robot}
Ligand Modeling

- Degrees of Freedom (DOF) = 9
  - 3 coordinates to position root atom (x,y,z)
  - 2 angles to specify first bond (α, β)
  - Torsional angles for all remaining non-terminal atoms (ψ)
  - Bond angles are assumed constant
  - Terminal hydrogens are modeled by increasing radius of terminal atoms
Motion Planning

Articulated Robot

Ligand
Motion Planning Algorithms

- 0-D robot in 2-D workspace
- Degrees of Freedom (dof) of robot = 2 \((x, y)\)

Cell Decomposition

Visibility Roadmaps
Obstacles in a Workspace

Obstacle seen by a 0-D robot

Obstacles seen by fixed orientation 1-D robots
**Configuration Space**

- DOF = 3 : x, y, \(\theta\)
- 1-D robot in 2-D workspace = 0-D robot in 3-D configuration space
- Problem is representing the obstacle in Configuration Space
Roadmap Planner

- Select milestones
  - Usually determined by characteristics of the obstacle (e.g. vertices)
- Connect pairs of milestones with simple local paths
  - Pairs selected based on distance or visibility
- Navigate by finding closest milestone and then follow pre-computed paths
Probabilistic Roadmap Planner

- Complete representation of obstacles in high dimensional configuration space is very difficult.
- Hence milestones are generated by sampling randomly from C-space and only accepting samples that are collision free.
- Connect milestones to their nearest neighbors with a local path planner.
Local Path Planner

- Connect the two milestones in C-space with a straight line
- Discretize the line into small segments such that likelihood of a collision within a segment is very small
- Check for collision at each discretized point along the straight line path
- If there is no collision then a path exists
Distribution of Samples
Energy-Based Path Planning

- Finding whether a path exists is only part of the problem
- We need to find the energetically most favourable path

Energy:
- Interaction of the ligand with the receptor
  - The receptor is represented as a potential field that occupies the entire work-space
- Internal energy of the ligand
  - Interaction of ligand atoms with each other
Energy of Interaction

Energy = electrostatic interaction \( (E_c) \) 
+ 
van der Waals interaction \( (E_v) \)

\[
E_c = 332 \frac{Q_i Q_j}{\varepsilon R_{ij}} \quad \quad \quad E_v = 0.2\left[\left(\frac{R_0}{R_{ij}}\right)^{12} - 2\left(\frac{R_0}{R_{ij}}\right)^6\right]
\]
Solvent Effects

\[ E_c = 332 \frac{Q_i Q_j}{(\varepsilon R_{ij})} \]

- Is only valid for an infinite medium of uniform dielectric
- Dielectric discontinuities result in induced surface charges

- Solution: Poisson-Boltzmann equation
  \[ \nabla \left[ \varepsilon(r) \nabla \cdot \phi(r) \right] - \varepsilon(r)k(r)^2\sinh(\phi(r)) + 4\pi r^f(r)/kT = 0 \]

- Can only be solved analytically for simple dielectric boundaries like spheres and planes
- Finite difference solution by Delphi [Sharp and Honig, 1990] is based on discretizing the workspace into a uniform grid
Computing Energy

- Both $E_c$ and $E_v$ are pre-computed on a uniform grid of resolution 0.5 Å
- van der Waals interactions are cutoff after 10 Å

- Total energy of ligand:
  - Energy of interaction of the ligand with the receptor
    » Two lookups into precomputed arrays for $E_c$ and $E_v$
  - Internal energy of the ligand
    » Standard van der Waal’s and Coulombic equations
Grid Points with energy $\leq -3$ kCal/Mol
(For a single negatively charged Oxygen atom)
Key Differences:

- Each point in configuration space has an associated energy
- Randomly generated landmarks are probabilistically accepted based on energy of the configuration
- Local path planner is energy based such that paths are weighted proportional to difficulty of motion
Computing Path Weights

- Need to assign weights to each link in the graph such that the minimum path weight between any two nodes corresponds to energetically favourable motion.

\[
P(\text{going from } i \text{ to } i+1) = \frac{e^{-\Delta E_1/kT}}{e^{-\Delta E_1/kT} + e^{-\Delta E_2/kT}}
\]

\[
\Delta E_1 = E_{i+1} - E_i
\]

\[
\Delta E_2 = E_{i-1} - E_i
\]
Local Path Planning

- Edge Weight = $\sum - \log (\text{Probability going from } i \text{ to } i+1)$

- “Difficulty score” of a given path = sum of individual edge weights along the path
Finding binding sites

- Sample low energy regions of configuration space
- Select best N samples (i.e. with lowest energies)
- Create new samples around these N samples
- Select new lowest energy samples and iterate

- Able to find binding sites that are in a broad low energy valley
- Binding sites in narrow passages (deep valleys) are difficult
- Difficulties could be due to the energy function as well
Lowest Energy Configurations
Results
Results
<table>
<thead>
<tr>
<th>Row number</th>
<th>RMSD from catalytic configuration (Å)</th>
<th>Configuration energy (kcal/mol)</th>
<th>Avg path weight entering configuration</th>
<th>Avg path weight leaving configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
<td>-11.79</td>
<td>112.98</td>
<td>134.54</td>
</tr>
<tr>
<td>1</td>
<td>31.04</td>
<td>-13.65</td>
<td>85.07</td>
<td>109.94</td>
</tr>
<tr>
<td>2</td>
<td>27.49</td>
<td>-12.66</td>
<td>90.48</td>
<td>111.98</td>
</tr>
<tr>
<td>3</td>
<td>1.73</td>
<td>-11.72</td>
<td>113.81</td>
<td>137.28</td>
</tr>
<tr>
<td>4</td>
<td>28.99</td>
<td>-11.54</td>
<td>85.32</td>
<td>105.19</td>
</tr>
<tr>
<td>5</td>
<td>24.67</td>
<td>-11.31</td>
<td>86.26</td>
<td>103.95</td>
</tr>
<tr>
<td>6</td>
<td>29.84</td>
<td>-11.27</td>
<td>86.49</td>
<td>107.53</td>
</tr>
<tr>
<td>7</td>
<td>29.32</td>
<td>-11.04</td>
<td>85.24</td>
<td>104.64</td>
</tr>
<tr>
<td>8</td>
<td>27.07</td>
<td>-10.96</td>
<td>81.70</td>
<td>102.28</td>
</tr>
<tr>
<td>9</td>
<td>31.00</td>
<td>-10.13</td>
<td>87.69</td>
<td>104.50</td>
</tr>
<tr>
<td>10</td>
<td>28.24</td>
<td>-9.97</td>
<td>86.36</td>
<td>98.89</td>
</tr>
</tbody>
</table>

Receptor: Lactate Dehydrogenase (2386 atoms, 309 residues)
Ligand: Oxamate (6 atoms, 7 degrees of freedom)
### Results - 4ts1

<table>
<thead>
<tr>
<th>Row number</th>
<th>RMSD from catalytic configuration (Å)</th>
<th>Configuration energy (kcal/mol)</th>
<th>Avg path weight entering configuration</th>
<th>Avg path weight leaving configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
<td>-19.44</td>
<td>130.73</td>
<td>173.76</td>
</tr>
<tr>
<td>1</td>
<td>1.91</td>
<td>-20.31</td>
<td>128.61</td>
<td>166.73</td>
</tr>
<tr>
<td>2</td>
<td>21.59</td>
<td>-15.92</td>
<td>105.65</td>
<td>118.72</td>
</tr>
<tr>
<td>3</td>
<td>15.16</td>
<td>-14.53</td>
<td>109.82</td>
<td>129.15</td>
</tr>
<tr>
<td>4</td>
<td>23.55</td>
<td>-14.39</td>
<td>111.87</td>
<td>134.96</td>
</tr>
<tr>
<td>5</td>
<td>20.59</td>
<td>-14.30</td>
<td>114.13</td>
<td>133.87</td>
</tr>
<tr>
<td>6</td>
<td>22.19</td>
<td>-13.97</td>
<td>113.84</td>
<td>135.90</td>
</tr>
<tr>
<td>7</td>
<td>24.62</td>
<td>-12.89</td>
<td>118.82</td>
<td>138.15</td>
</tr>
<tr>
<td>8</td>
<td>19.13</td>
<td>-12.74</td>
<td>115.45</td>
<td>136.72</td>
</tr>
<tr>
<td>9</td>
<td>17.05</td>
<td>-12.31</td>
<td>120.24</td>
<td>142.72</td>
</tr>
<tr>
<td>10</td>
<td>36.81</td>
<td>-11.81</td>
<td>115.48</td>
<td>131.98</td>
</tr>
</tbody>
</table>

**Receptor:** Tyrosyl-transfer-RNA synthetase (2423 atoms, 319 residues)

**Ligand:** Tyrosine (13 atoms, 9 degrees of freedom)
Results - Characterizing the Binding Site

- Preliminary results indicate the following:
  - The best binding site is not necessarily the one with the lowest ligand energy
  - The true binding site is instead characterized by a distinct energy barrier around the site
  - The difficulty of leaving the true binding site is higher than other potential sites. The difficulty of entering the true site is also correspondingly higher.
# Results - 1stp

<table>
<thead>
<tr>
<th>Row number</th>
<th>RMSD from true binding configuration (Å)</th>
<th>Configuration energy (kcal/mol)</th>
<th>Avg path weight entering configuration</th>
<th>Avg path weight leaving configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
<td>-15.06</td>
<td>110.80</td>
<td>146.87</td>
</tr>
<tr>
<td>1</td>
<td>21.76</td>
<td>-15.79</td>
<td>80.78</td>
<td>108.42</td>
</tr>
<tr>
<td>2</td>
<td>27.14</td>
<td>-12.83</td>
<td>96.29</td>
<td>117.67</td>
</tr>
<tr>
<td>3</td>
<td>18.59</td>
<td>-12.82</td>
<td>85.84</td>
<td>101.24</td>
</tr>
<tr>
<td>4</td>
<td>23.52</td>
<td>-11.45</td>
<td>96.45</td>
<td>122.01</td>
</tr>
<tr>
<td>5</td>
<td>13.67</td>
<td>-11.36</td>
<td>86.51</td>
<td>106.05</td>
</tr>
<tr>
<td>6</td>
<td>15.18</td>
<td>-10.79</td>
<td>88.22</td>
<td>96.89</td>
</tr>
<tr>
<td>7</td>
<td>13.93</td>
<td>-10.68</td>
<td>95.14</td>
<td>116.92</td>
</tr>
<tr>
<td>8</td>
<td>14.63</td>
<td>-10.42</td>
<td>85.61</td>
<td>105.16</td>
</tr>
<tr>
<td>9</td>
<td>24.64</td>
<td>-9.96</td>
<td>85.71</td>
<td>105.17</td>
</tr>
<tr>
<td>10</td>
<td>20.43</td>
<td>-9.87</td>
<td>83.81</td>
<td>102.54</td>
</tr>
</tbody>
</table>

**Receptor:** Streptavidin (901 atoms, 121 residues)  
**Ligand:** Biotin (16 atoms, 11 degrees of freedom)
## Results

<table>
<thead>
<tr>
<th></th>
<th>DOF</th>
<th>Sampling Time (6000 nodes)</th>
<th>Linking Time</th>
<th>Final Nodes</th>
<th>Connected Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ldm</td>
<td>7</td>
<td>9 s</td>
<td>57 sec</td>
<td>6129</td>
<td>2</td>
</tr>
<tr>
<td>4ts1</td>
<td>9</td>
<td>27 s</td>
<td>4 min 13 sec</td>
<td>6530</td>
<td>4</td>
</tr>
<tr>
<td>1stp</td>
<td>11</td>
<td>39 s</td>
<td>4 min 43 sec</td>
<td>6635</td>
<td>5</td>
</tr>
</tbody>
</table>