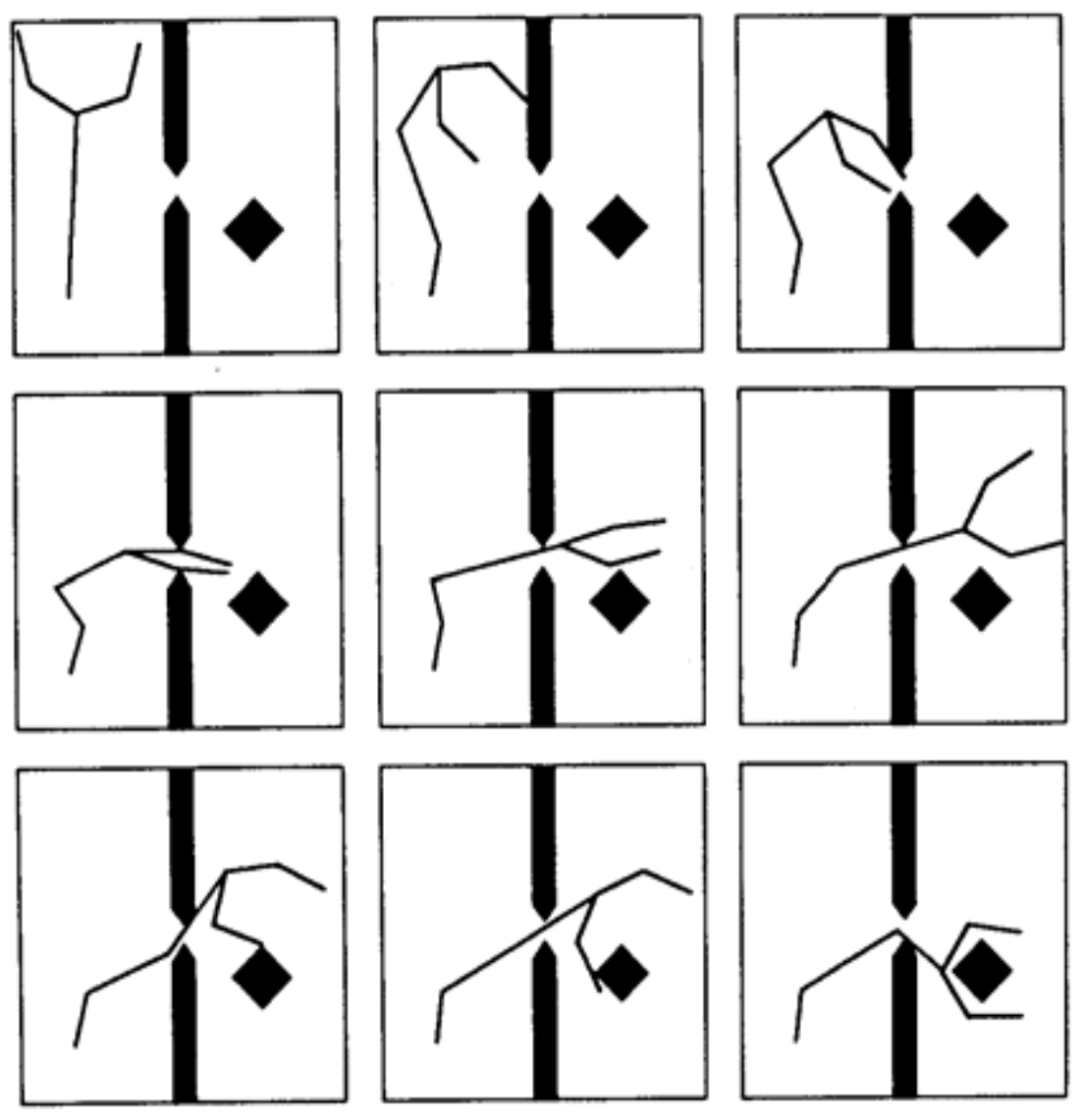
A Motion Planning Approach to Flexible Ligand Binding

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

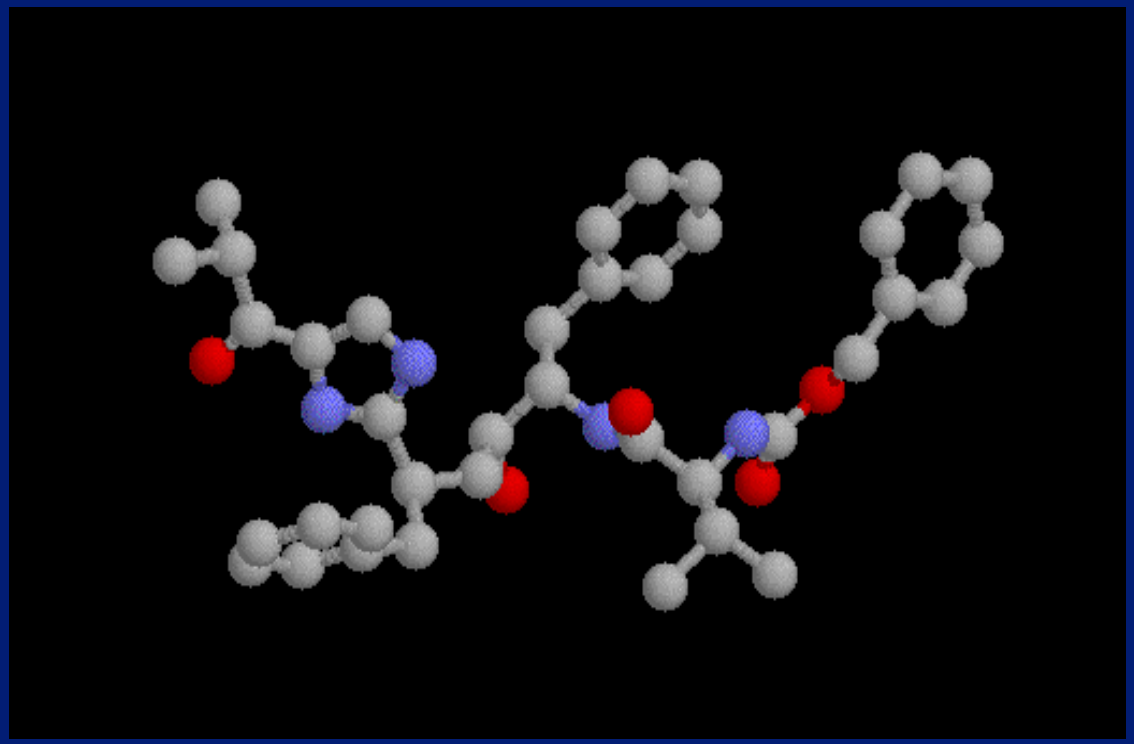
What is Motion Planning?



Why Motion Planning?

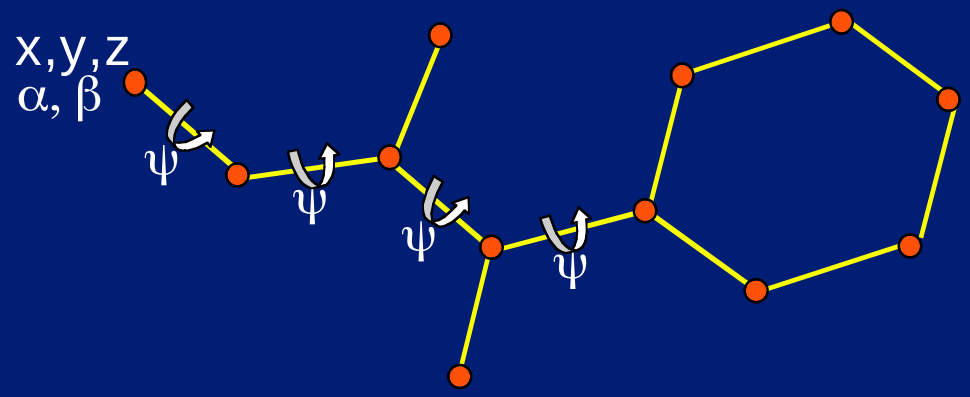
- Motion planning can:
  - Sample the space of possible paths taken by the ligand as it approaches and binds to the receptor
  - Examine the energetics of the ligand along each of these paths
  - Make estimates of the relative rates of binding and dissociation
  - Identify regions of the protein that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

Why Robotics?



Ligand $\stackrel{?}{=}$ Articulated Robot

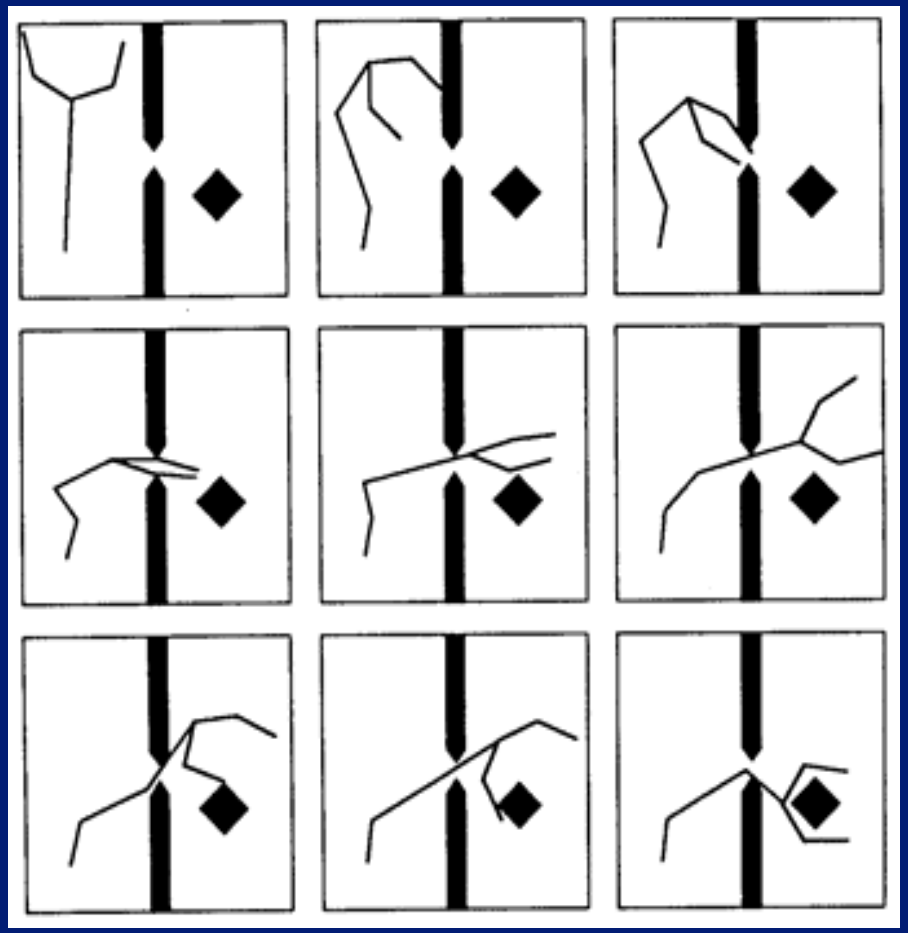
Ligand Modeling



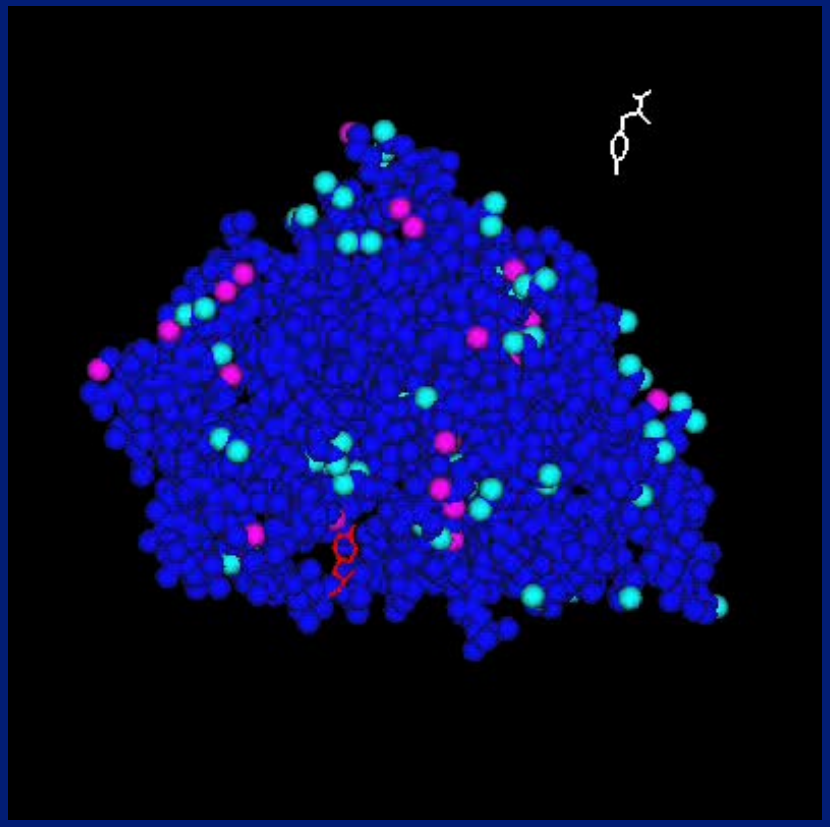
- Degrees of Freedom (DOF) = 9
  - 3 coordinates to position root atom (x,y,z)
  - 2 angles to specify first bond (α, β)
  - Torsional angles for all remaining non-terminal atoms (ψ)
  - Bond angles are assumed constant
  - Terminal hydrogens are modeled by increasing radius of terminal atoms

Motion Planning

Articulated Robot



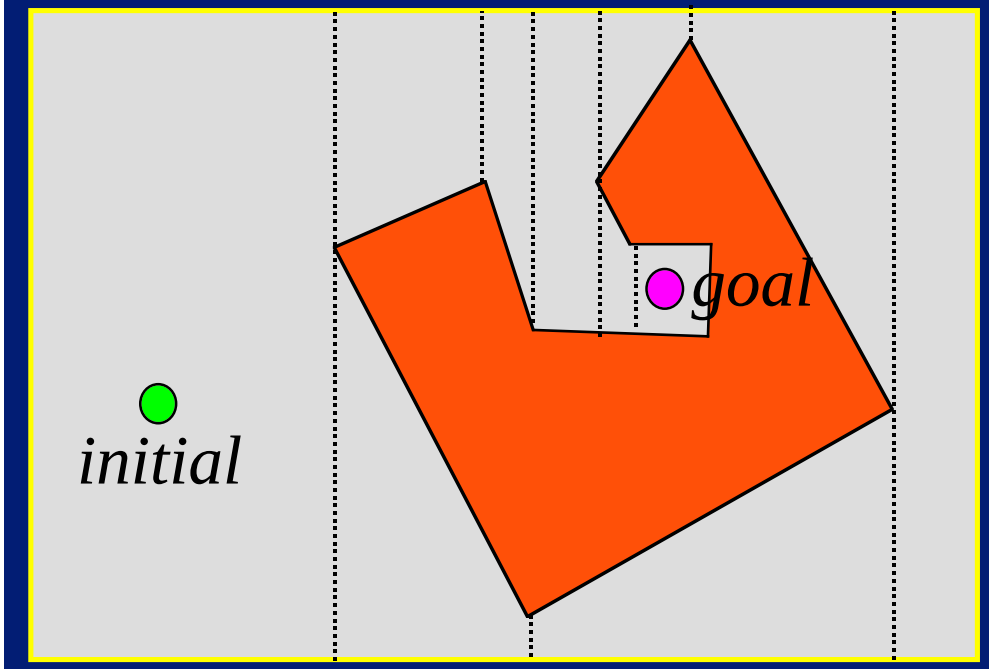
Ligand



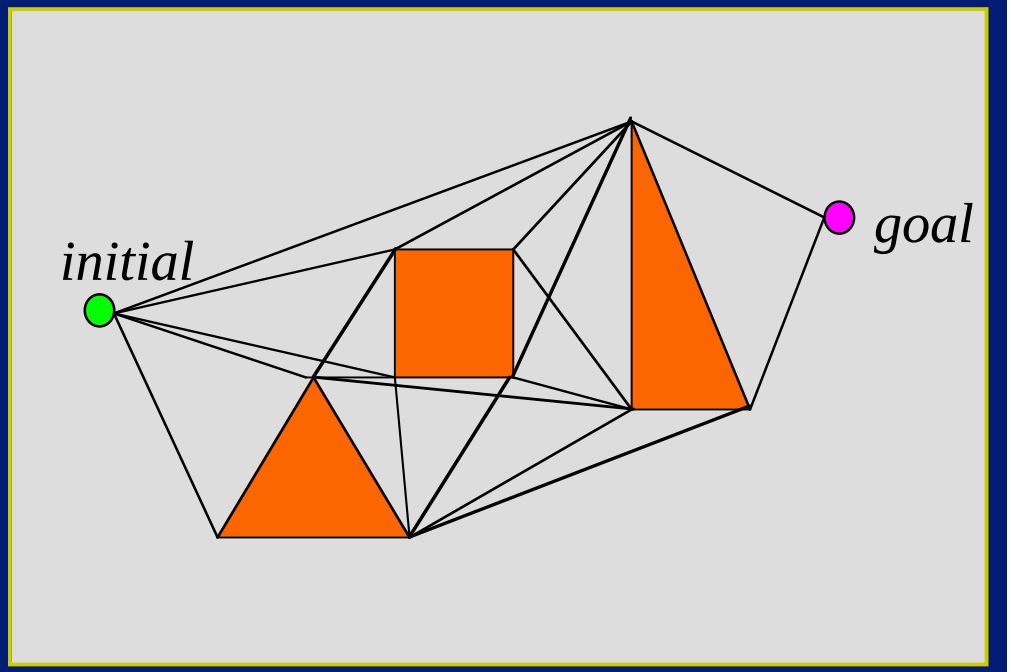
Motion Planning Algorithms

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

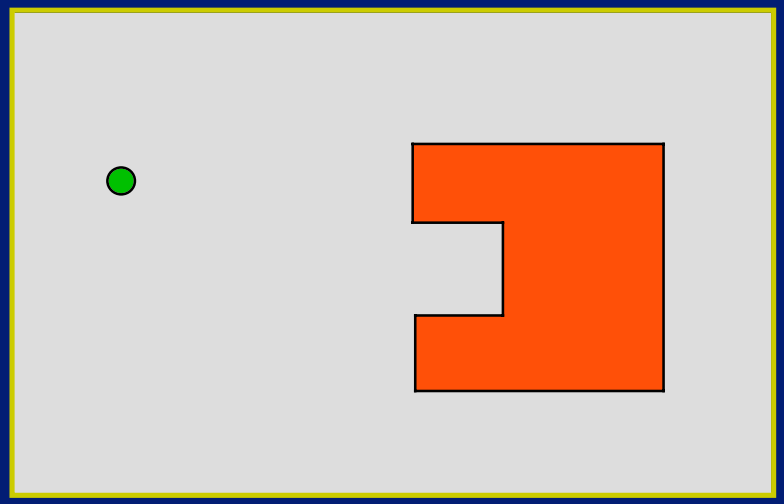
Cell Decomposition



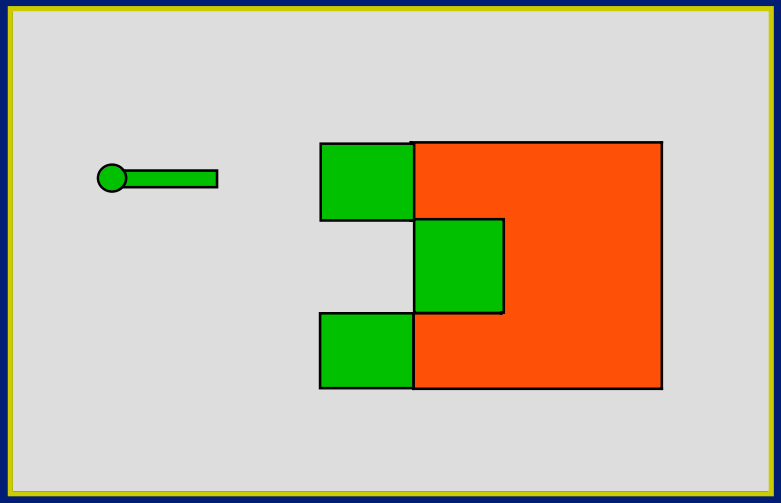
Visibility Roadmaps

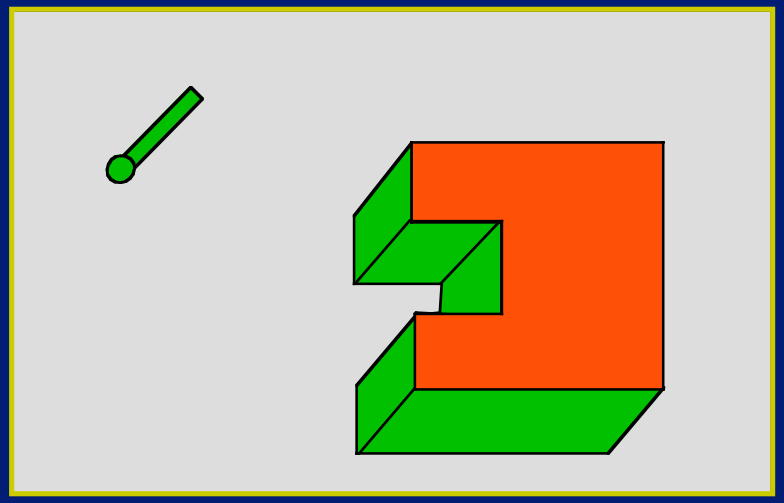


Obstacles in a Workspace



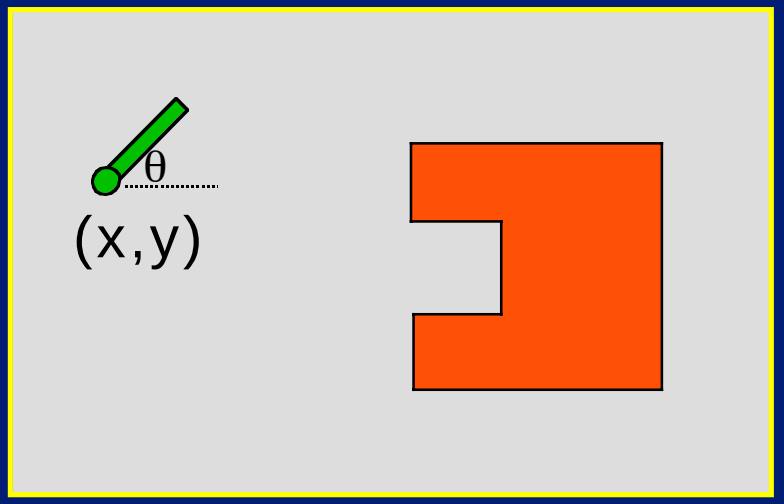
Obstacle seen by
a 0-D robot



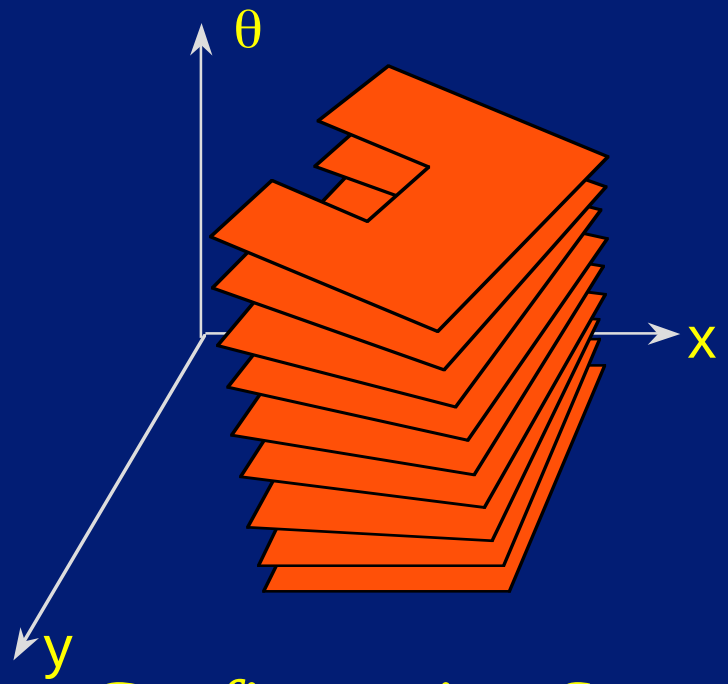


Obstacles seen by fixed orientation 1-D robots

Space



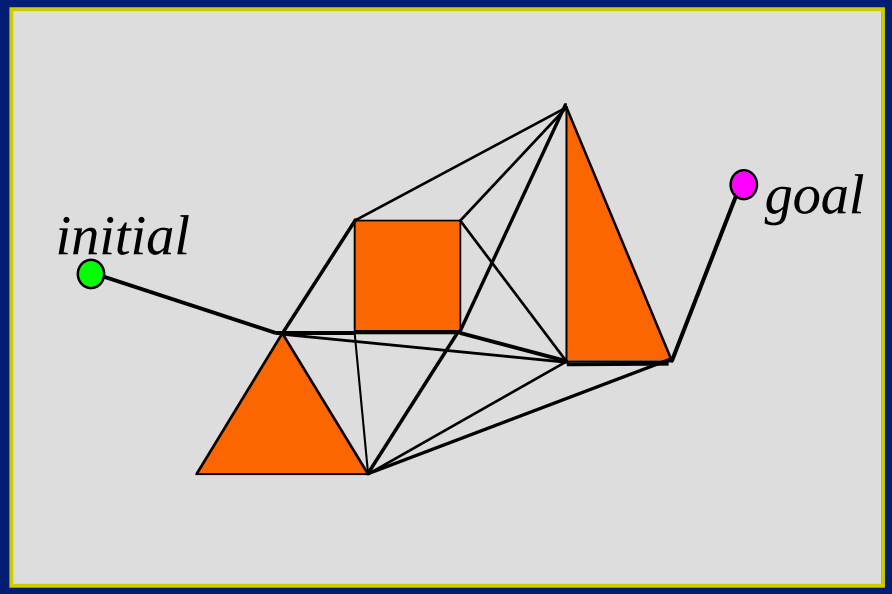
Work Space



Configuration Space

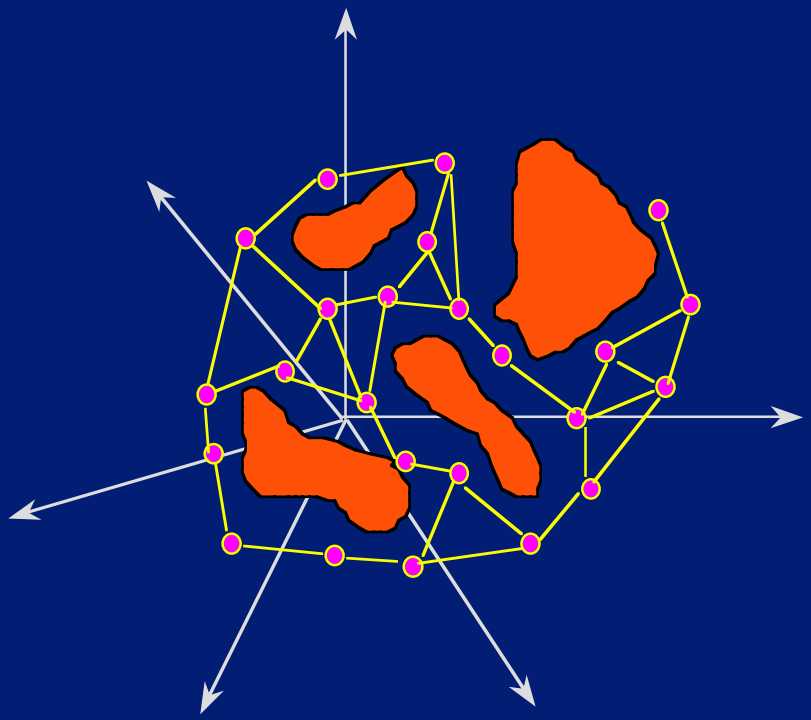
- DOF = 3 : x, y, θ
- 1-D robot in 2-D workspace = 0-D robot in 3-D configuration space
- Problem is representing the obstacle in Configuration Space

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

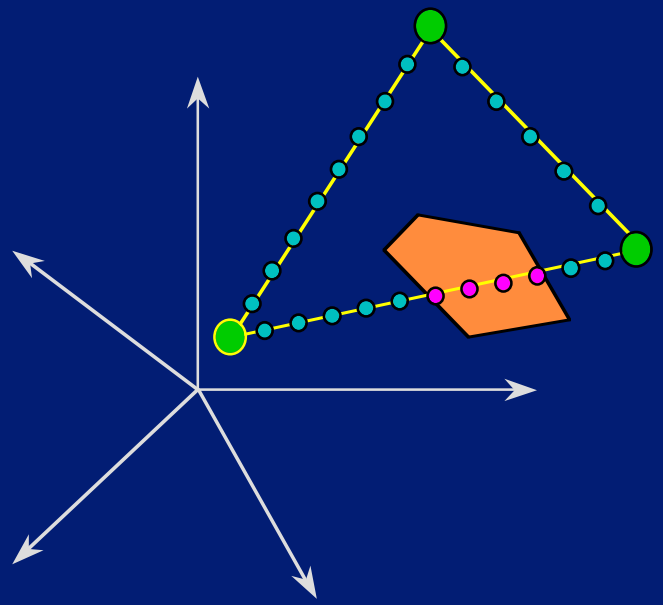
Probabilistic Roadmap Planner



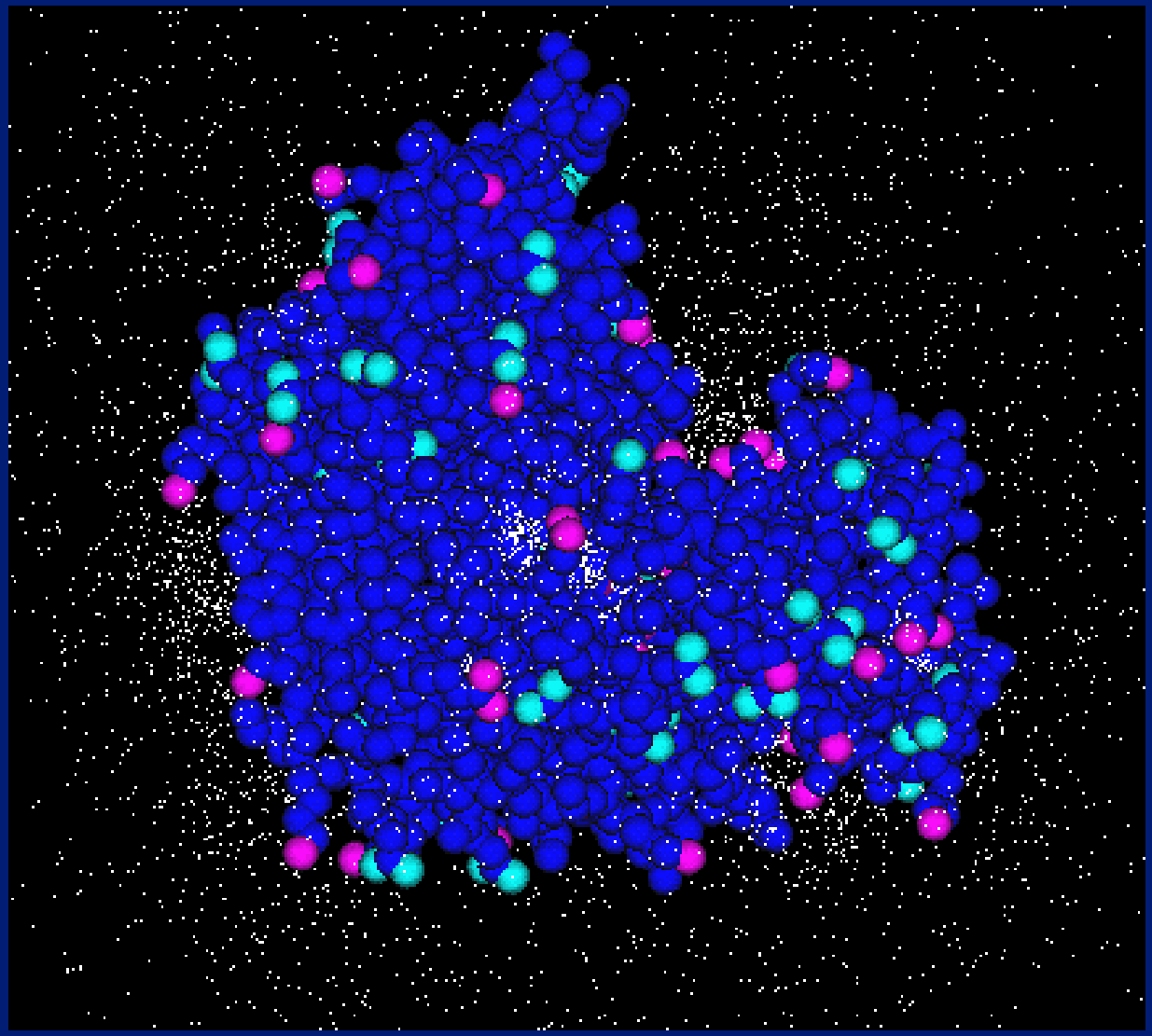
- Complete representation of obstacles in high dimensional configuration space is very difficult
- Hence milestones are generated by sampling randomly from C-space and only accepting samples that are collision free
- Connect milestones to their nearest neighbors with a **local path planner**

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

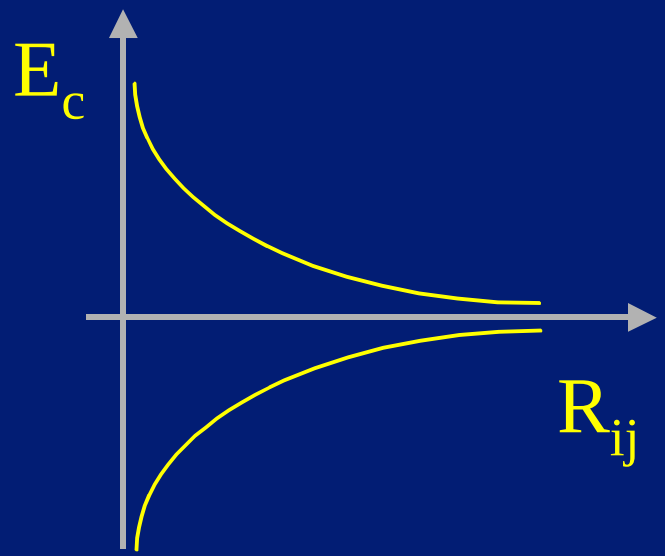
Energy = electrostatic interaction (E_c)

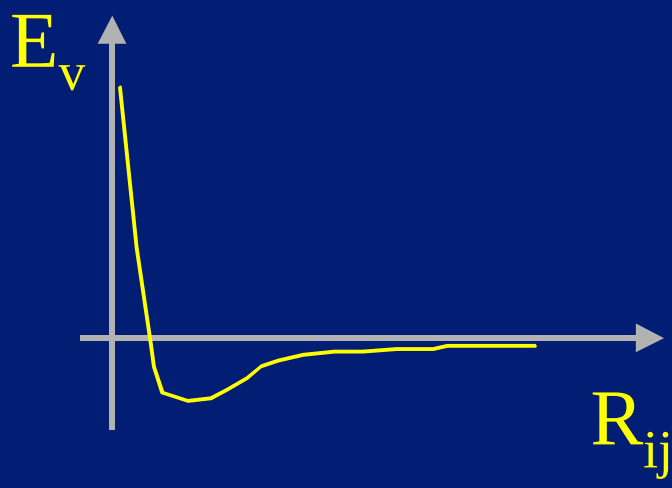
+

van der Waals interaction (E_v)

$E_c = 332\ Q_iQ_j/(\varepsilon R_{ij})$

$E_v = 0.2[(R_0/R_{ij})^{12} - 2(R_0/R_{ij})^6]$





Solvent Effects

$$E_c = 332\, Q_i Q_j / (\varepsilon R_{ij})$$

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

- Solution: Poisson-Boltzman equation

$$\nabla\ [\varepsilon(r)\nabla\ .\ \phi(r)] - \varepsilon(r)k(r)^2\sinh([\phi(r)] + 4\pi r^f(r)/kT = 0$$

- 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

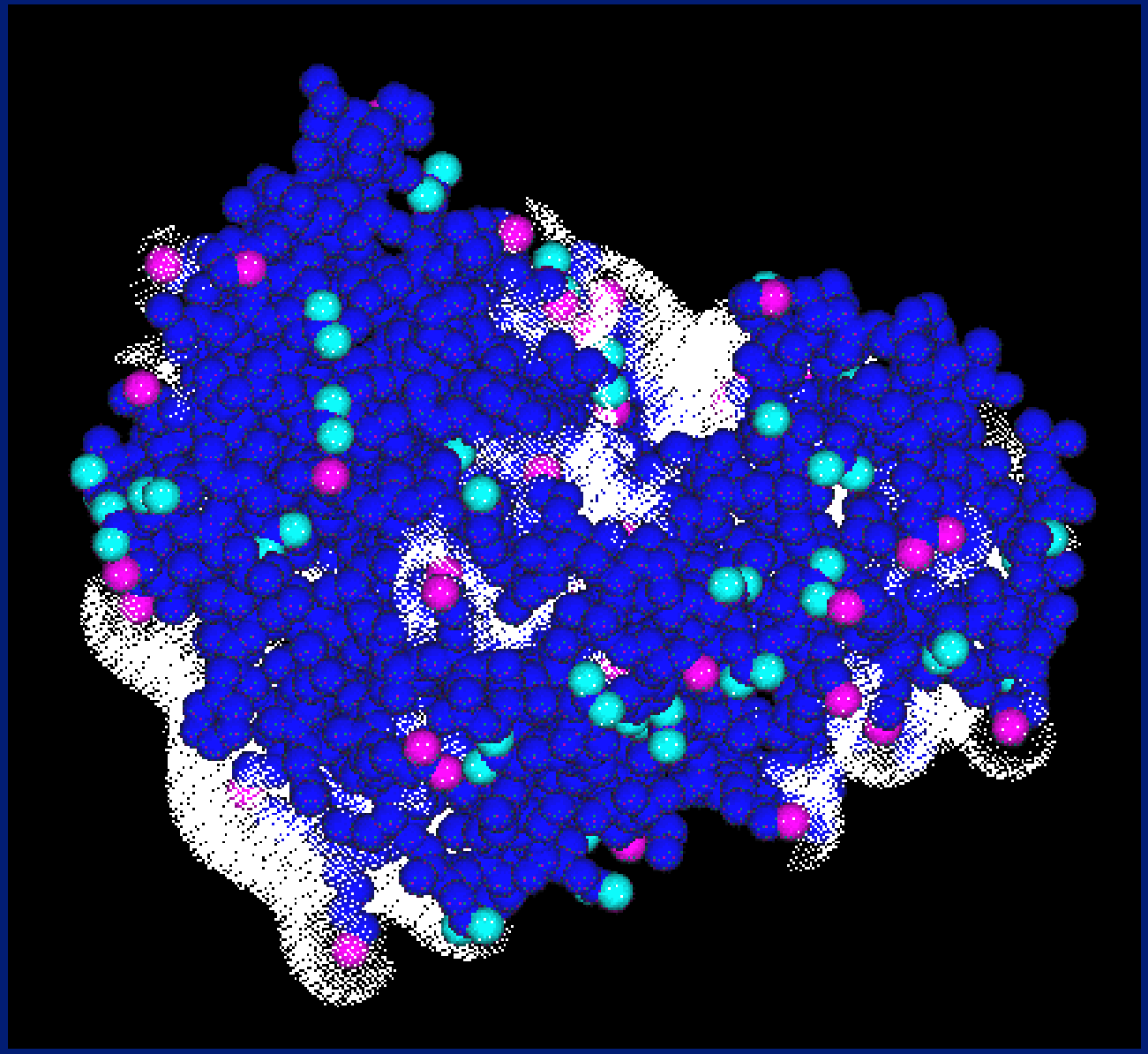
Computing Energy

- 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

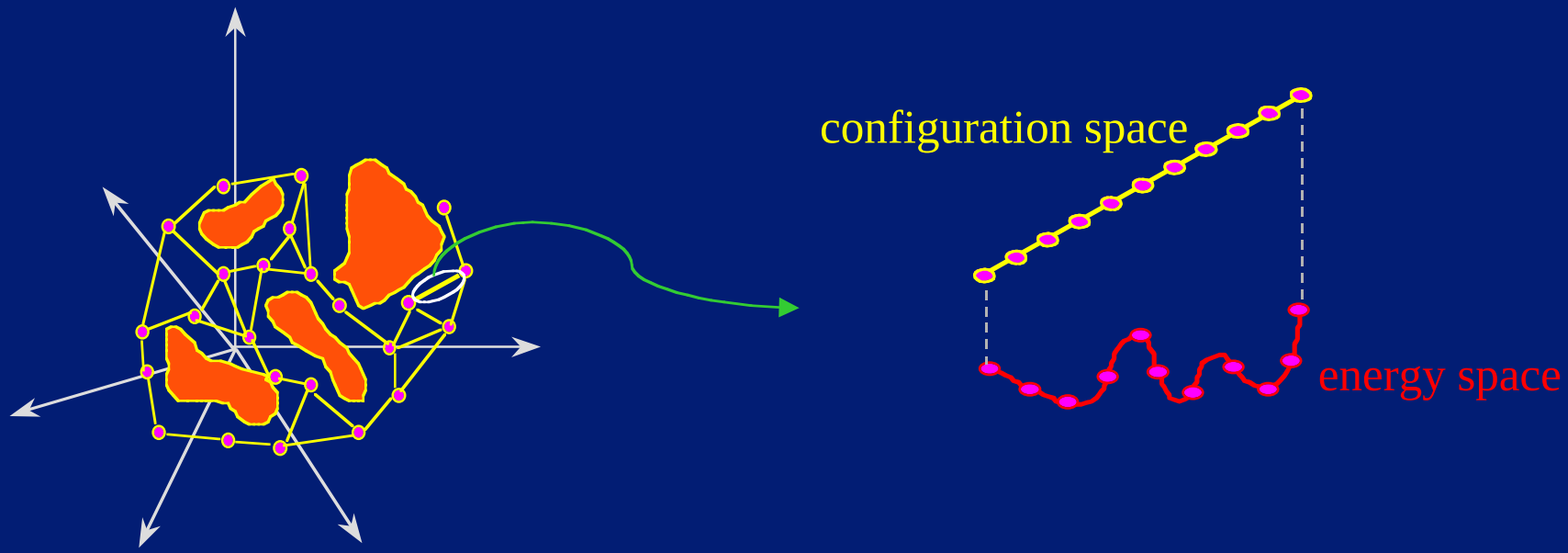
- Total energy of ligand:
  - Energy of interaction of the ligand with the receptor
    - » Two lookups into precomputed arrays for E_c and E_v
  - Internal energy of the ligand
    - » Standard van der Waal's and Coulombic equations

Grid Points with energy <= -3 kCal/Mol

(For a single negatively charged Oxygen atom)



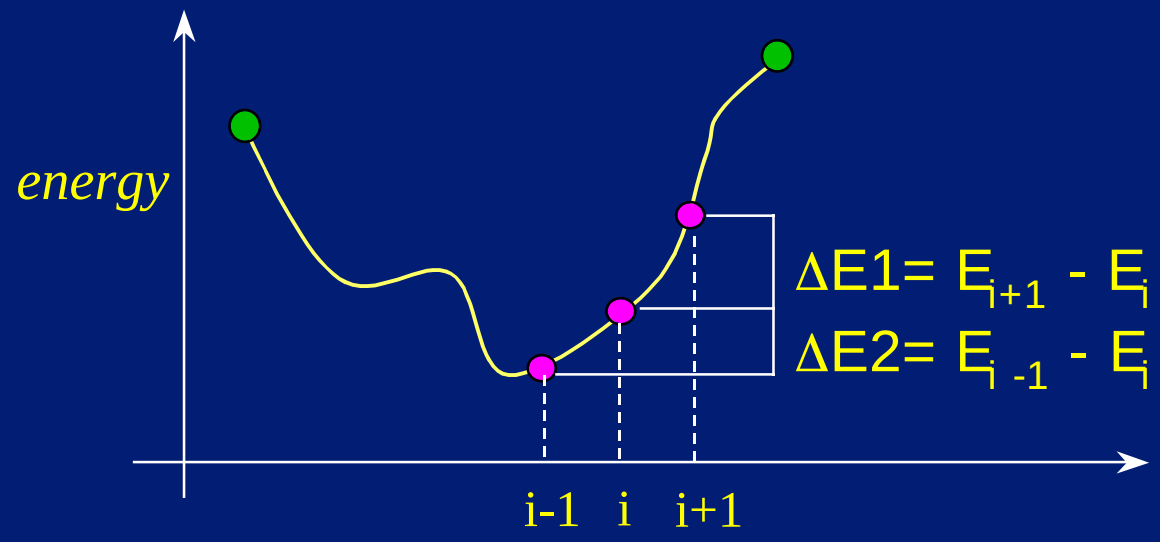
Energy-Based Probabilistic Roadmaps



- 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

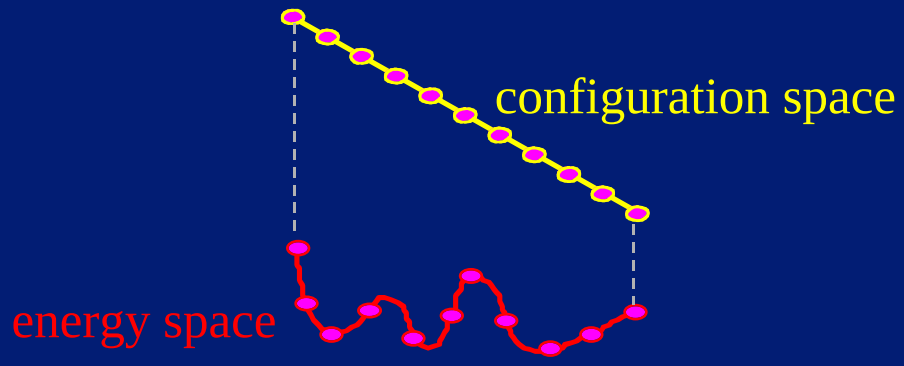
- Need to assign weights to each link in the graph such that the minimum path weight between any two nodes corresponds to energetically favourable motion



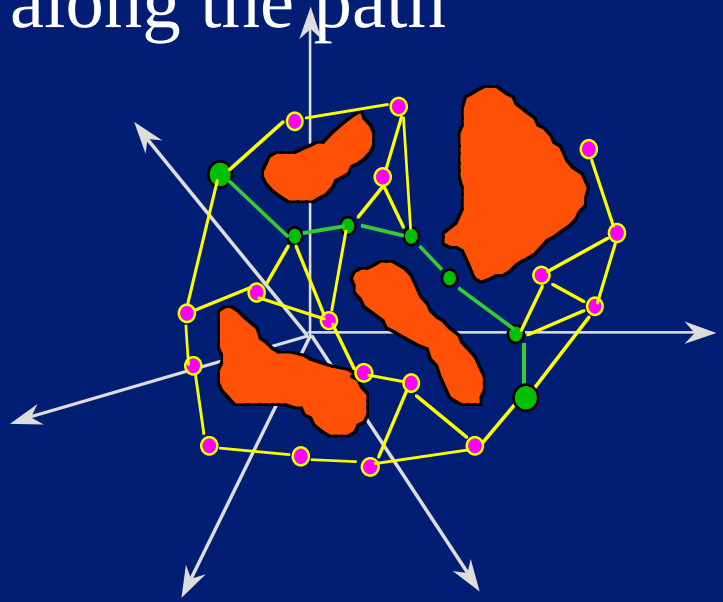
$$P(\text{going from i to i+1}) = \frac{e^{-\Delta E1/kT}}{e^{-\Delta E1/kT} + e^{-\Delta E2/kT}}$$

Local Path Planning

- Edge Weight = Σ - log (Probability going from i to i+1)



- "Difficulty score" of a given path = sum of individual edge weights along the path

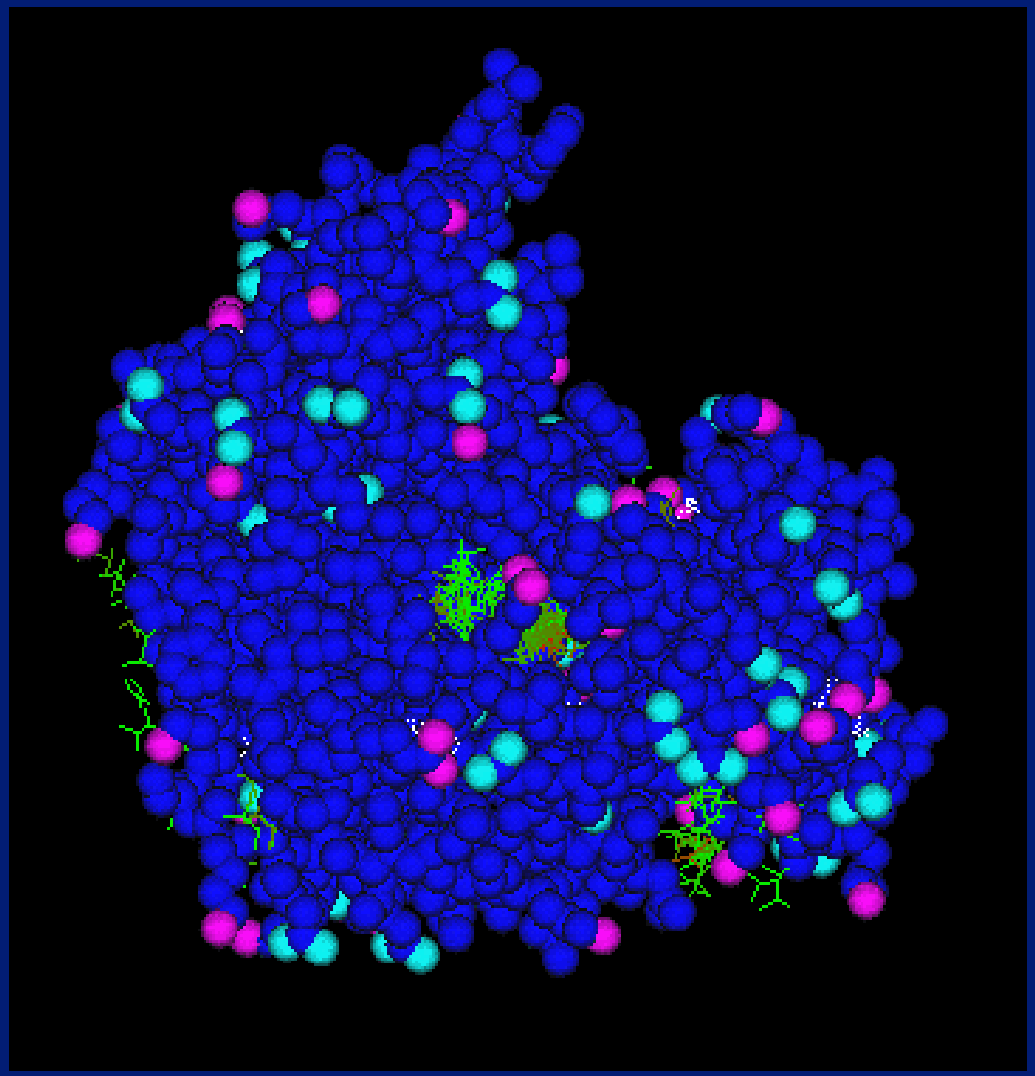


Finding binding sites

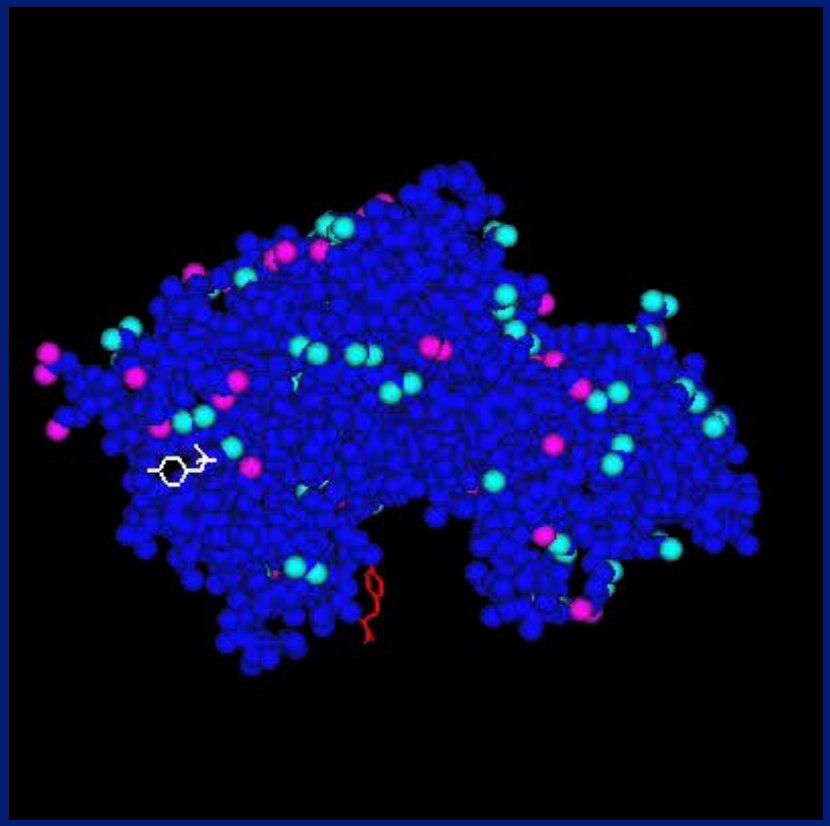
- Sample low energy regions of configuration space
- Select best N samples (i.e. with lowest energies)
- Create new samples around these N samples
- Select new lowest energy samples and iterate

- 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

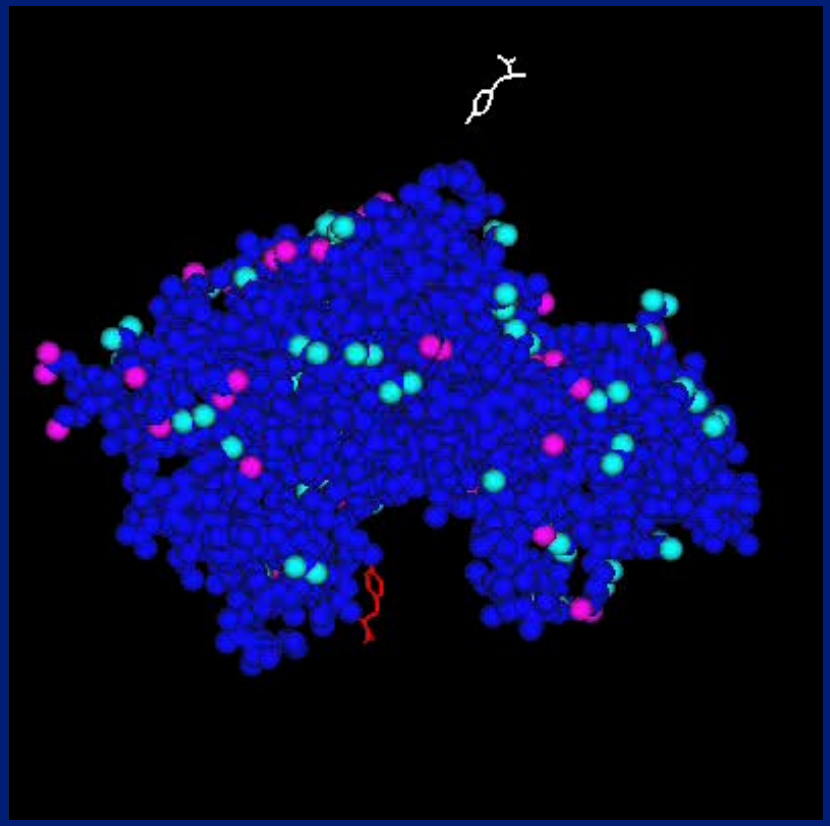
Lowest Energy Configurations



Results



Results



Results - 1ldm

Row number	RMSD from catalytic configuration (Å)	Configuration energy (kcal/mol)	Avg path weight entering configuration	Avg path weight leaving configuration
0	0.00	-11.79	112.98	134.54
1	31.04	-13.65	85.07	109.94
2	27.49	-12.66	90.48	111.98
3	1.73	-11.72	113.81	137.28
4	28.99	-11.54	85.32	105.19
5	24.67	-11.31	86.26	103.95
6	29.84	-11.27	86.49	107.53
7	29.32	-11.04	85.24	104.64
8	27.07	-10.96	81.70	102.28
9	31.00	-10.13	87.69	104.50
10	28.24	-9.97	86.36	98.89

Receptor: Lactate Dehydrogenase (2386 atoms, 309 residues)
Ligand: Oxamate (6 atoms, 7 degrees of freedom)

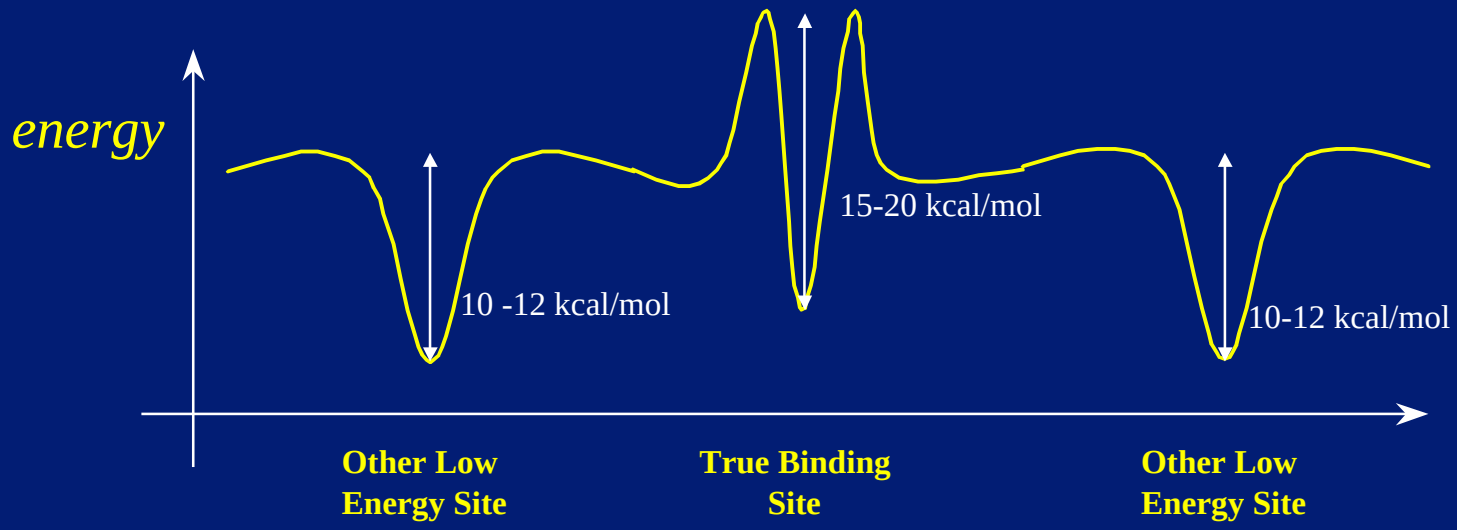
Results - 4ts1

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

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

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

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

Results

	DOF	Sampling Time (6000 nodes)	Linking Time	Final Nodes	Connected Components
1ldm	7	9 s	57 sec	6129	2
4ts1	9	27 s	4 min 13 sec	6530	4
1stp	11	39 s	4 min 43 sec	6635	5