Protein Docking

Amit P. Singh

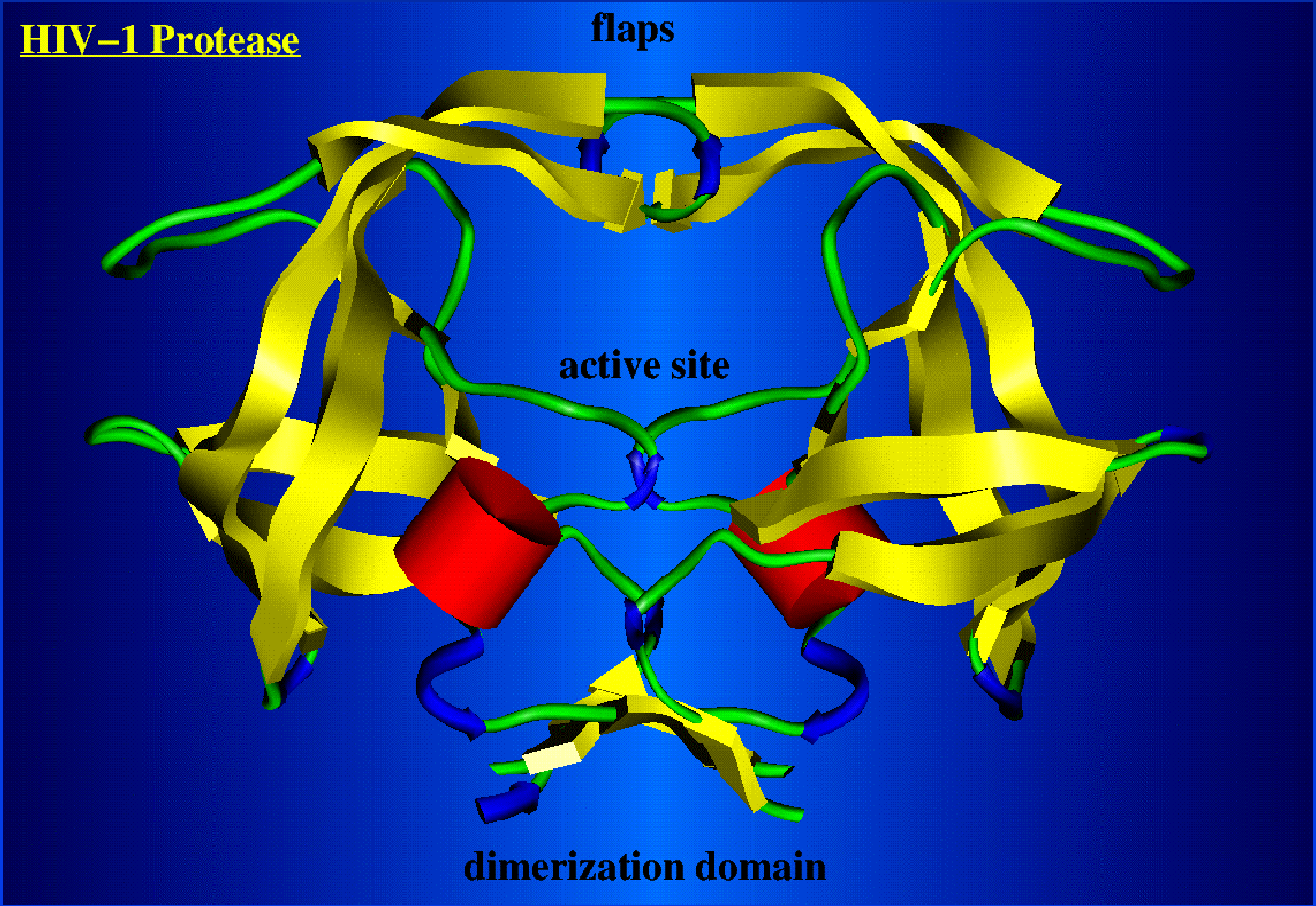
Biochemistry 218/MIS 231

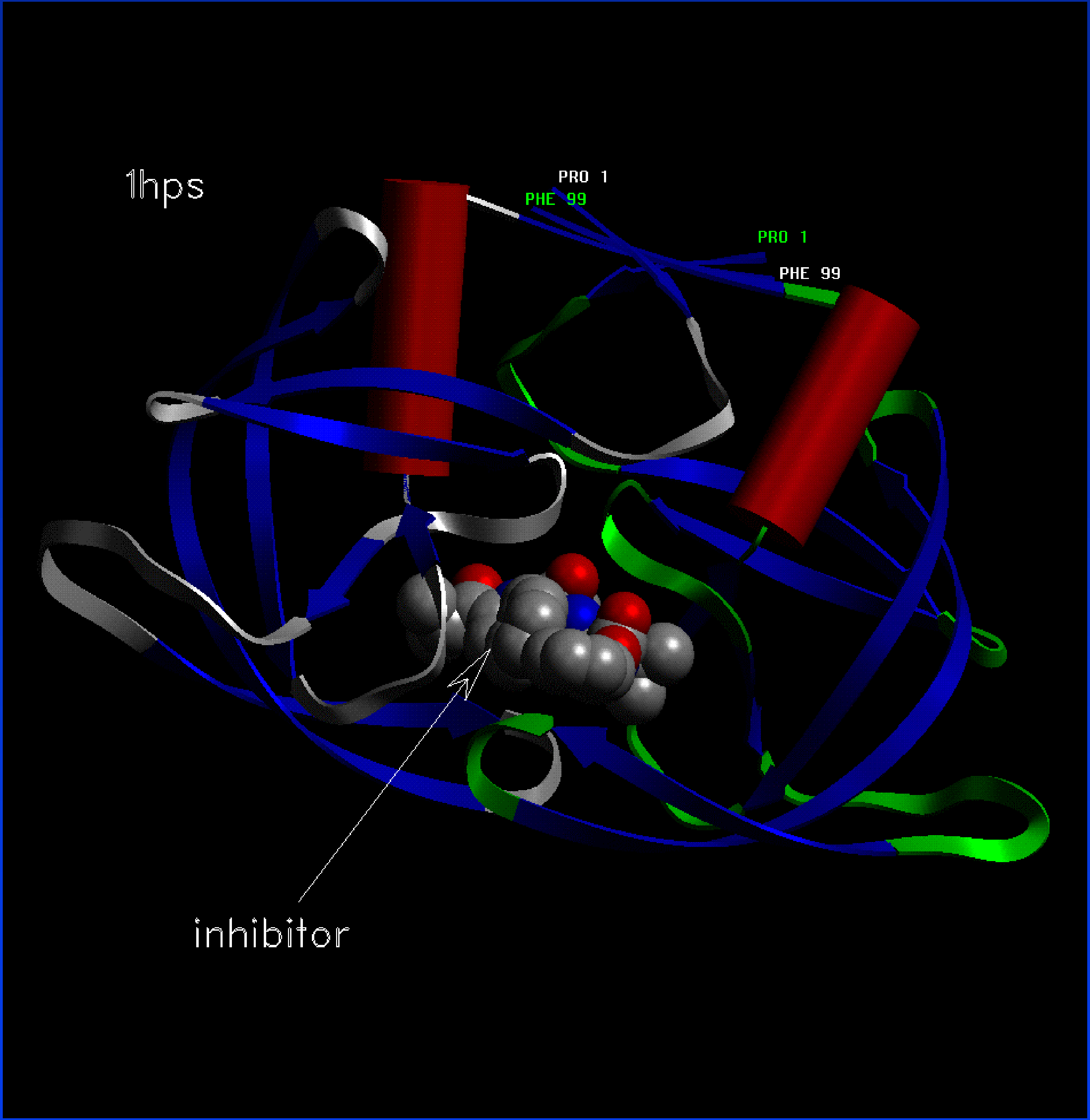
November 30, 1998

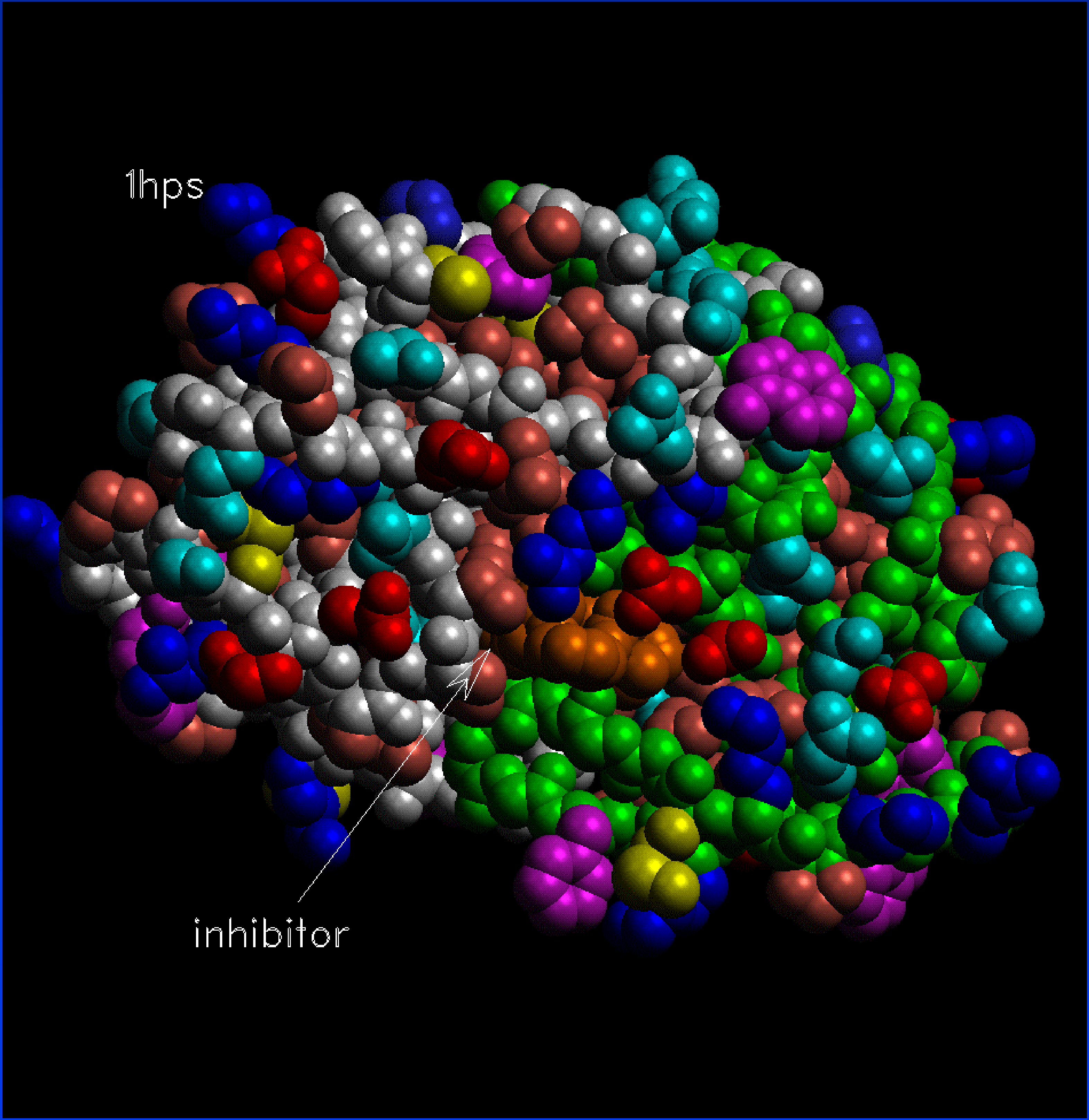
Why is Docking Important?

- Biomolecular interactions are the core of all the regulatory and metabolic processes that together constitute the process of life
- Computer-aided analysis of these interactions is becoming increasingly important as the database of known biomolecular structures continues to grow
- Increasing processing power makes the analysis and prediction of molecular interaction more tractable
- **AUTOMATED PREDICTION OF MOLECULAR INTERACTIONS IS THE KEY TO RATIONAL DRUG DESIGN**

An example: HIV-1 Protease




1hps
PRO 1
PHE 99
PRO 1
PHE 99
inhibitor

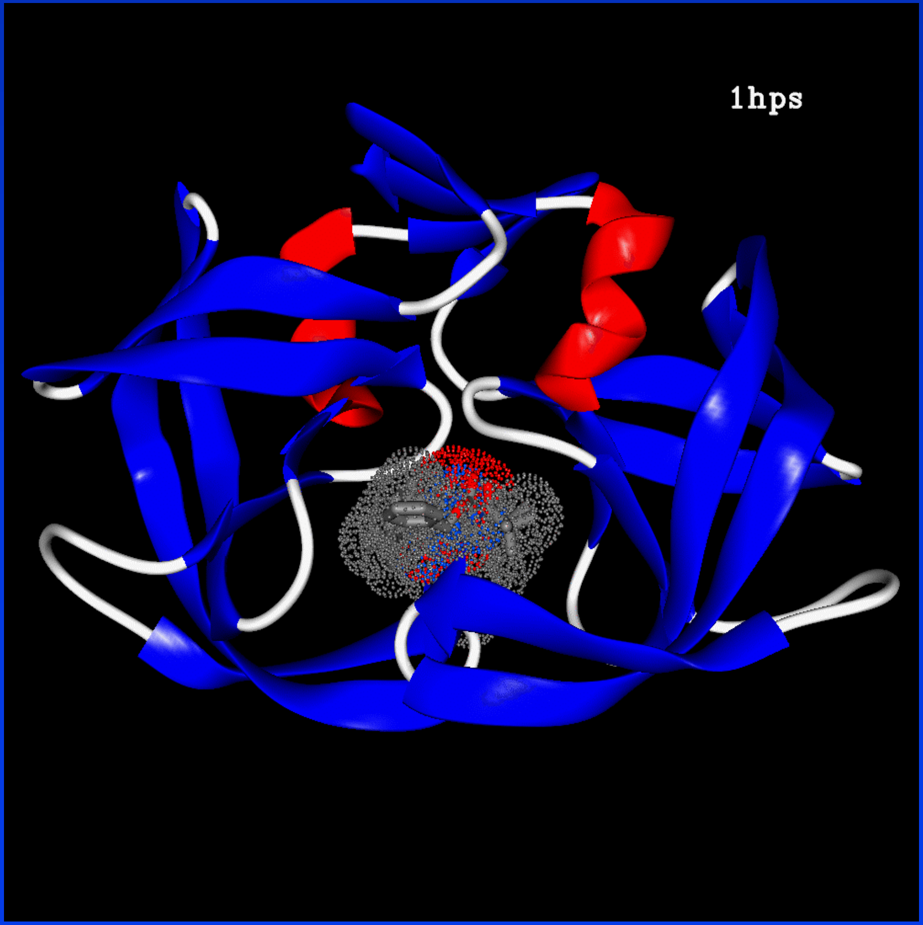

1hps
inhibitor

The Problem

- Given two biological molecules determine:
  - Whether the two molecules “interact”
    - » ie. is there an energetically favorable orientation of the two molecules such that one may modify the other’s function
    - » ie. do the two molecules fit together in any energetically favorable way
  - If so, what is the orientation that maximizes the “interaction” while minimizing the total “energy” of the complex
- GOAL: To be able to search a database of molecular structures and retrieve all molecules that can interact with the query structure

Why is this difficult?

- Both molecules are flexible and may alter each other’s structure as they interact:
  - Hundreds to thousands of degrees of freedom
  - Total possible conformations are astronomical



Classes of Docking Studies

- Protein-Protein docking
  - both molecules usually considered rigid
  - 6 degrees of freedom, 3 for rotation, 3 for translation
  - first apply only steric constraints to limit search space
  - then examine energetics of possible binding conformations

- Protein-Ligand docking
  - Flexible ligand, rigid-receptor
  - Search space much larger
  - Either reduce flexible ligand to rigid fragments connected by one or several hinges (reduces conformational space
  - Or search the conformational space using monte-carlo methods or molecular dynamics

Classes of Docking Studies

- Rough Docking
  - Search a database of potential ligands to select lead compounds for drug design
  - Often based on quick geometrical algorithms combined with heuristic functions to predict binding energy

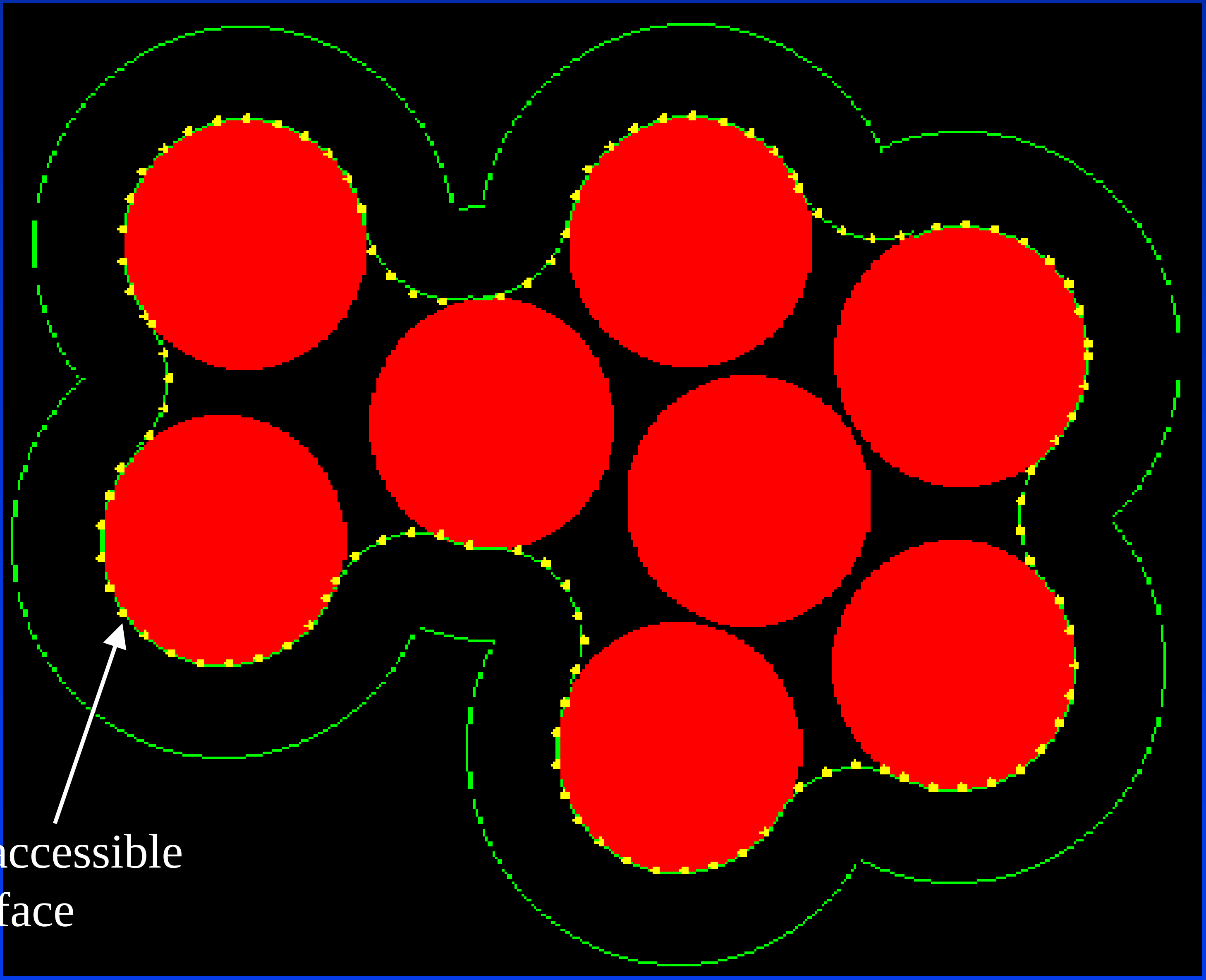
- Detailed Docking
  - Accurate analysis of a single instance of docking
  - To compute thermodynamic and kinetic properties of binding (free energy, rates of binding and dissociation)
  - Computing free energy of binding requires models of both enthalpic and entropic contributions
  - Large amount of conformational sampling required to compute the entropy of the ligand in the binding site

Protein-Protein Docking

- Surface representation
  - efficiently represent the docking surface
  - identify regions of interest
    - » cavities (binding site) and protrusions
- Surface matching
  - match corresponding surfaces to optimize binding score
- Current techniques:
  - Lenhoff, Nussinov and Wolfson, Kuntz et al., Singh and Brutlag

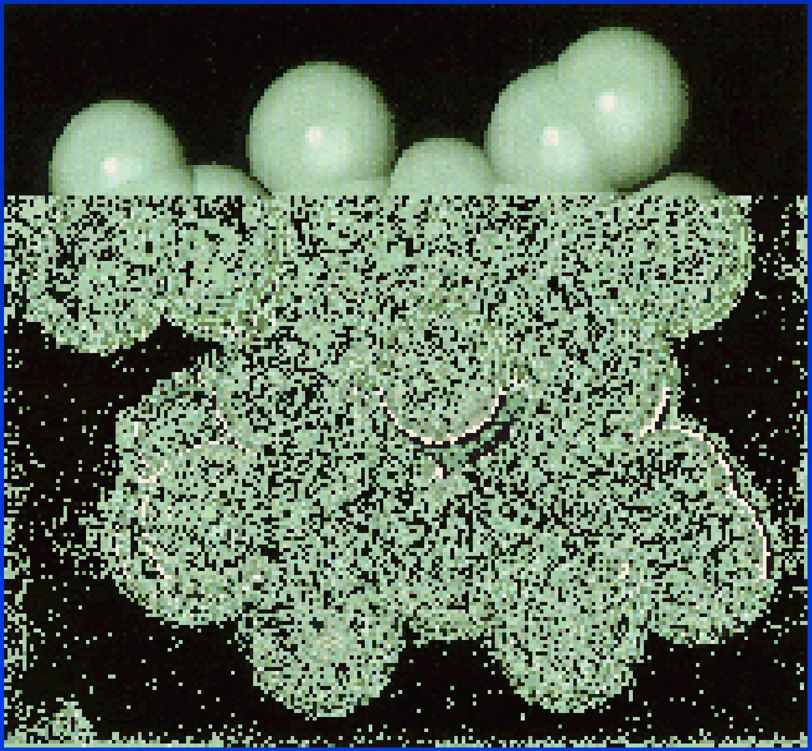
Surface Representation

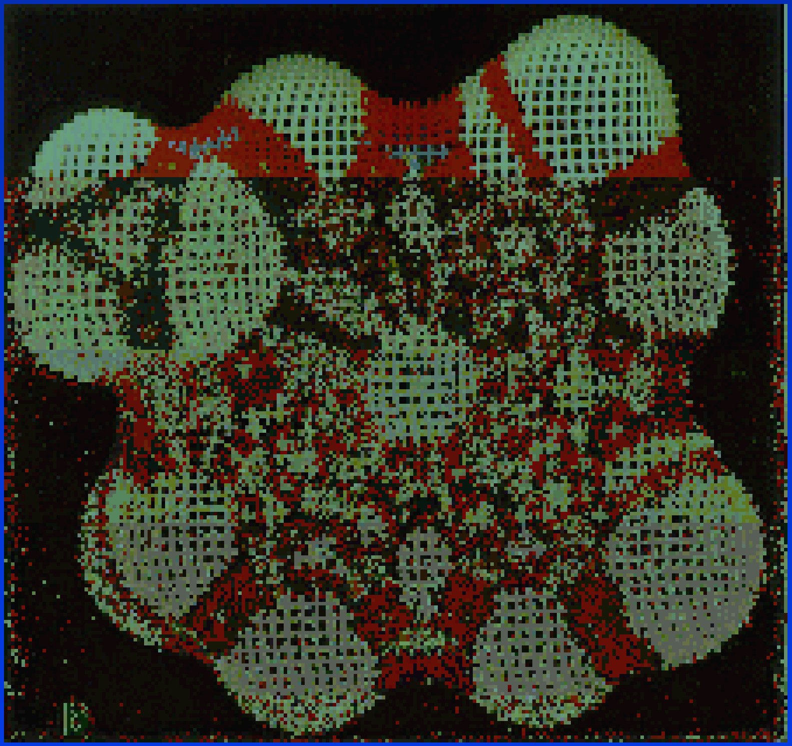
Connolly Surface



Solvent accessible
surface

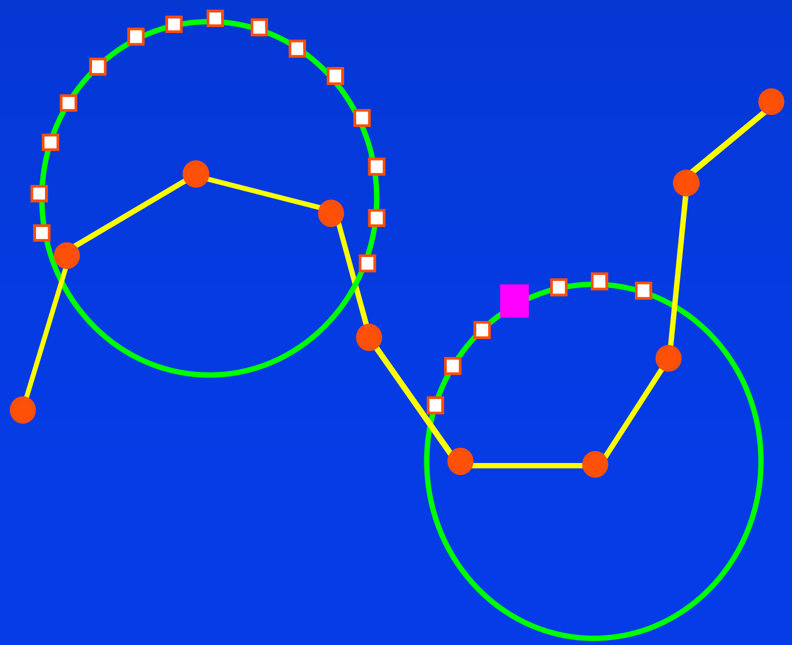
Surface Representation



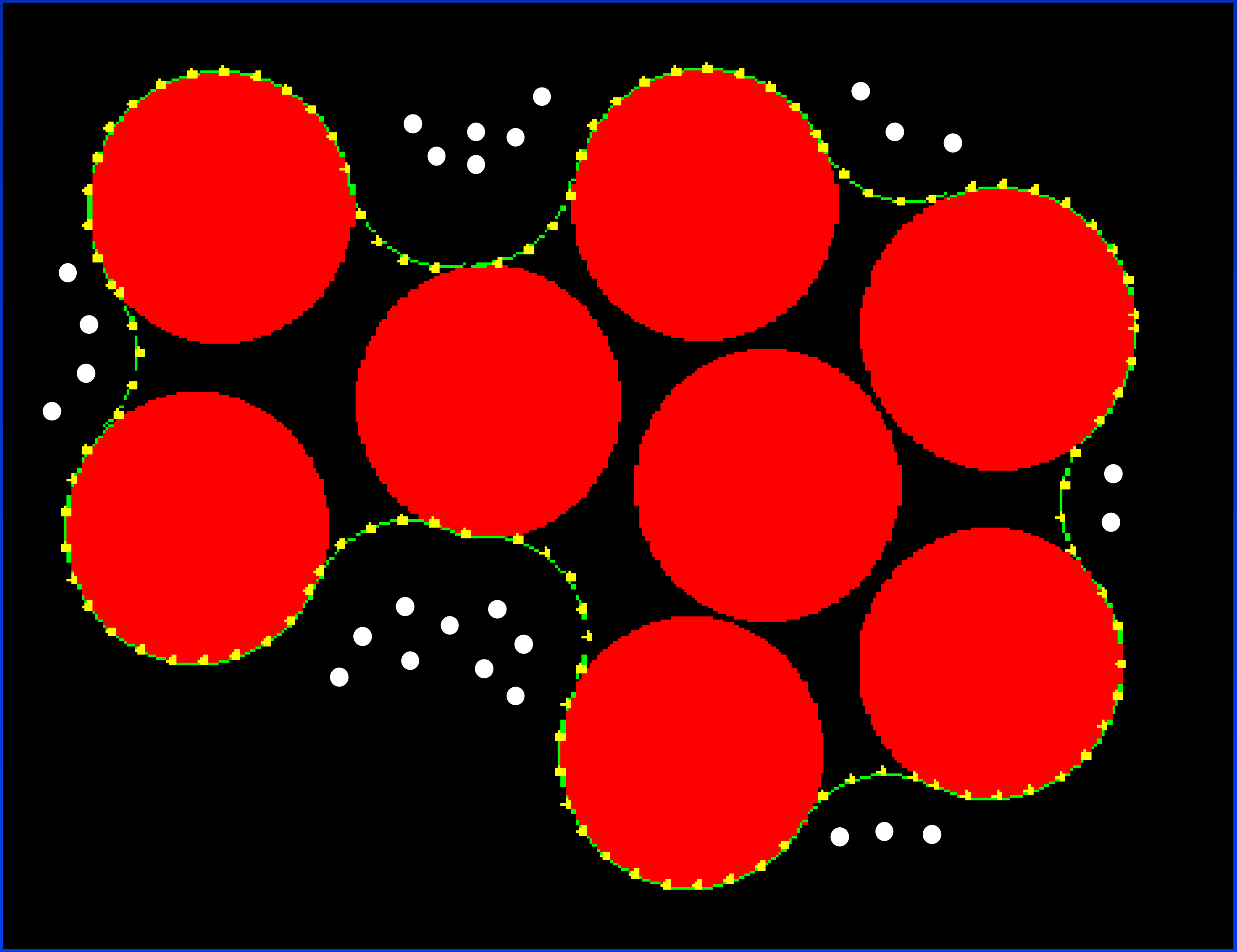


Lenhoff

- Computes a “complementary” surface for the receptor instead of the Connolly surface
- ie. Computes possible positions (near the surface of the receptor) for the atom centers of the ligand
- Based on the contact-score of uniformly distributed points on probe spheres

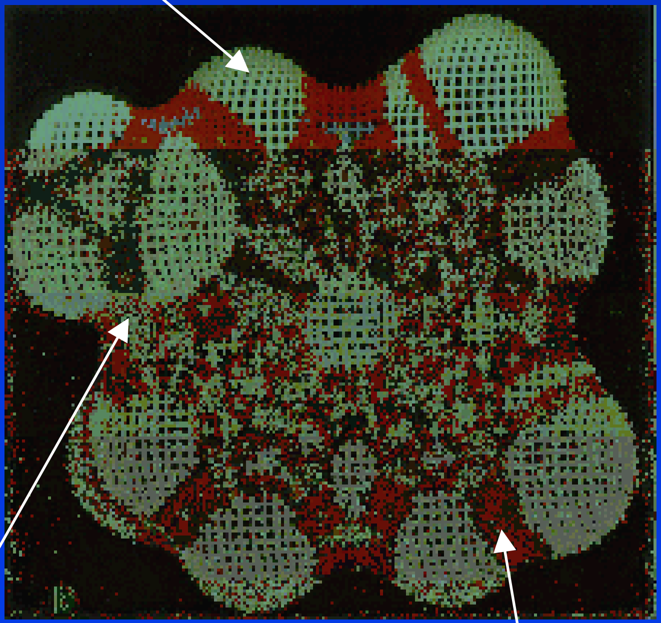


Lenhoff



Nussinov and Wolfson

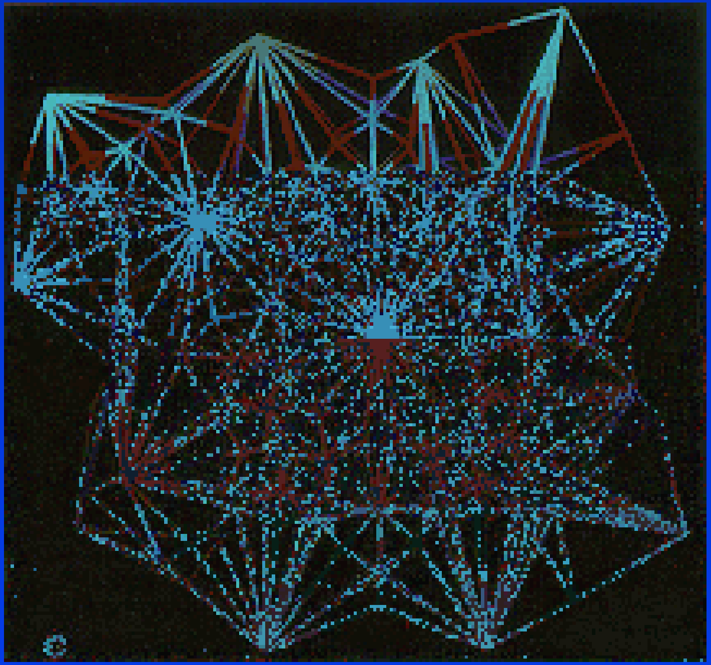
- Computes critical points on the Connolly surface
- Each concave, convex, and saddle face of the Connolly surface is replaced by a single “critical point”
- 44 atoms -> 5,355 Connolly Points -> 326 critical points



Convex (white)

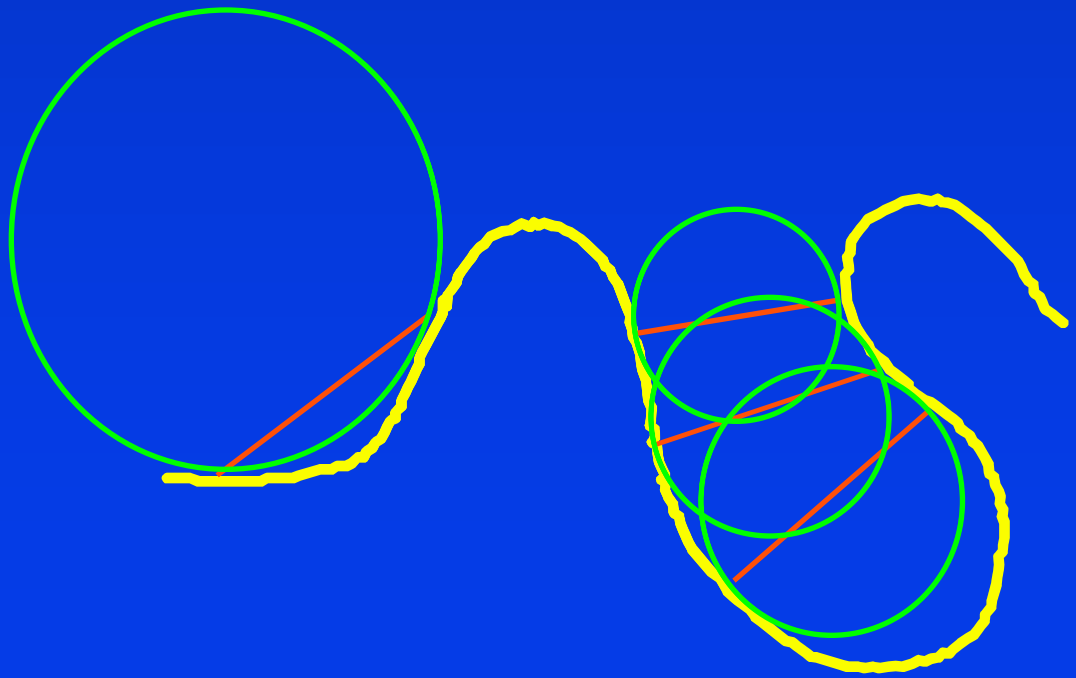
Concave (blue)

Saddle (red)



Kuntz

- Uses clustered-spheres to identify cavities on the receptor and protrusions on the ligand
- Compute a sphere for every pair of surface points, i and j, with the sphere center on the normal from point i
- Number of spheres is reduced by only retaining the smallest sphere for every surface point
- Regions where many spheres overlap are either cavities (on the receptor) or protrusions (on the ligand)



Surface Matching

- First satisfy steric constraints
  - Find the best fit of the receptor and ligand using only geometrical constraints
  - Compute scores based on RMSD (or number of contact points) instead of E_v
- Then use energy calculations to refine the docking
  - Compute the energy of interaction for each geometrically feasible docking pattern
  - Select the fit that has the minimum energy

Surface Matching

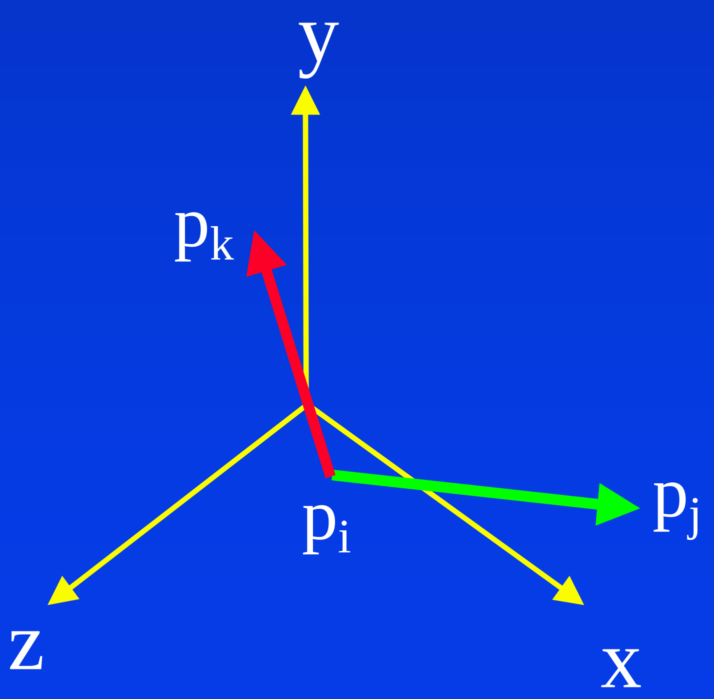
- The problem:
  - Find the transformation (rotation + translation) that will maximize the number of matching surface points from the receptor and ligand
- A Solution: Geometric Hashing
  - Compute all possible triangles formed by selecting triplets of atoms from the ligand and from the receptor
  - Compare all receptor triangles to all ligand triangles using a hash table
  - Use the set of triangles with the maximum number of matches to find the transformation matrix

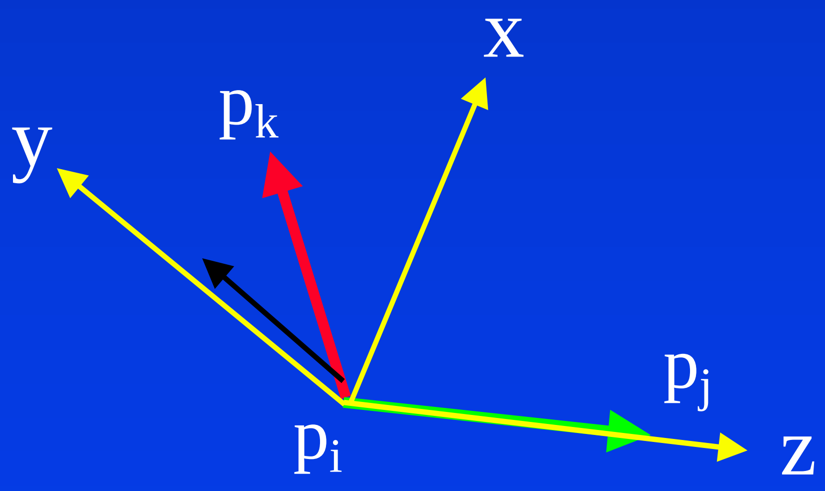
Geometric Hashing

- Building the table:
  - For each triplet of points from the ligand, generate a unique coordinate system
  - Record the position and orientation of all remaining points in this coordinate system in an index table
- Searching the table:
  - For each triplet of points from the receptor, generate a unique coordinate system
  - Search the table of ligand points to find the receptor coordinate system that results in the maximum number of similar points

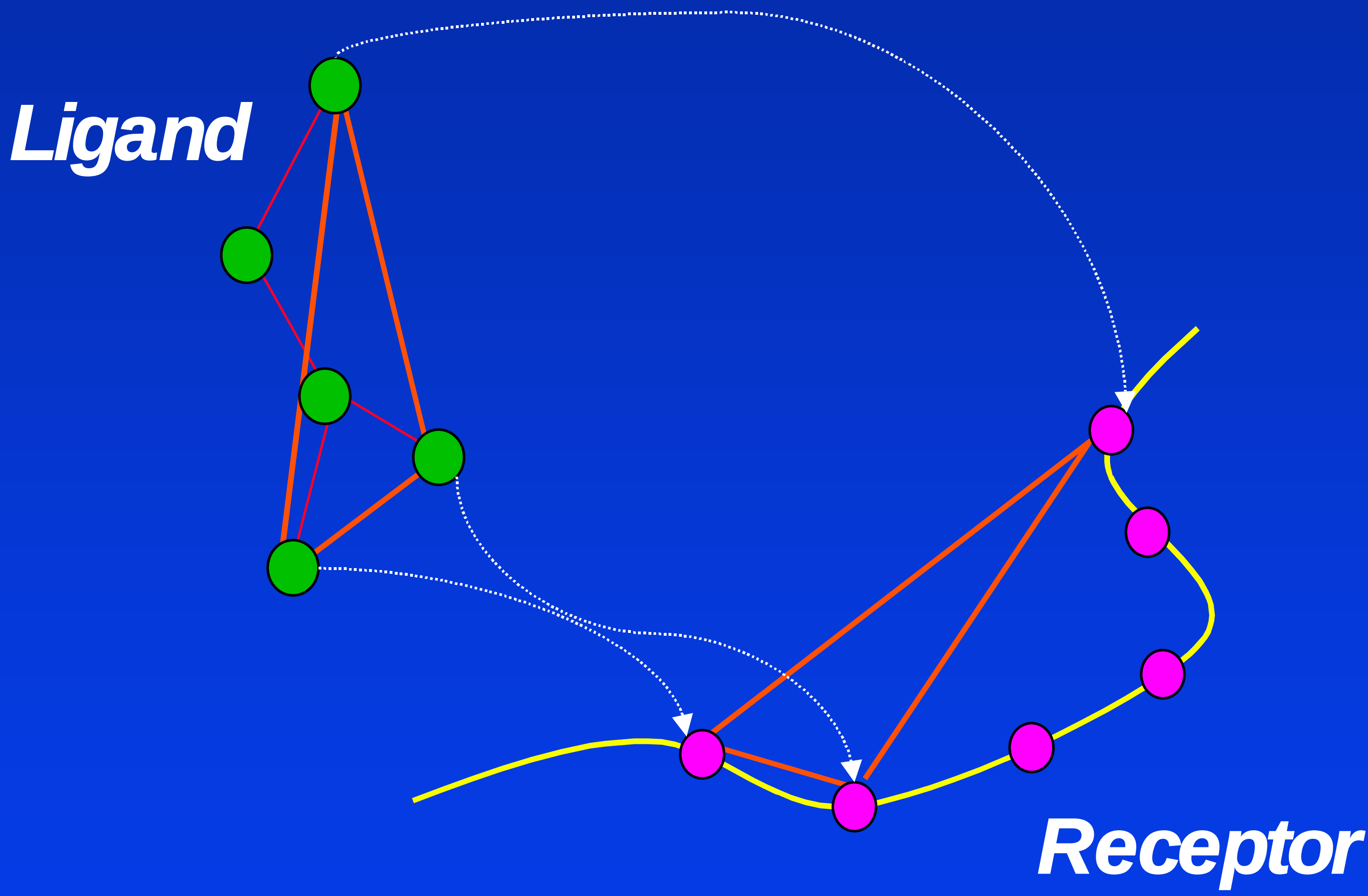
Generating a Coordinate System

- For each triplet of points (p_i, p_j, p_k)
  - Transform the coordinates such that vector(p_i p_i) lies on the Z-axis and the projection of vector(p_j p_k) on to the X-Y plane is parallel to the Y-axis





Matching Surfaces

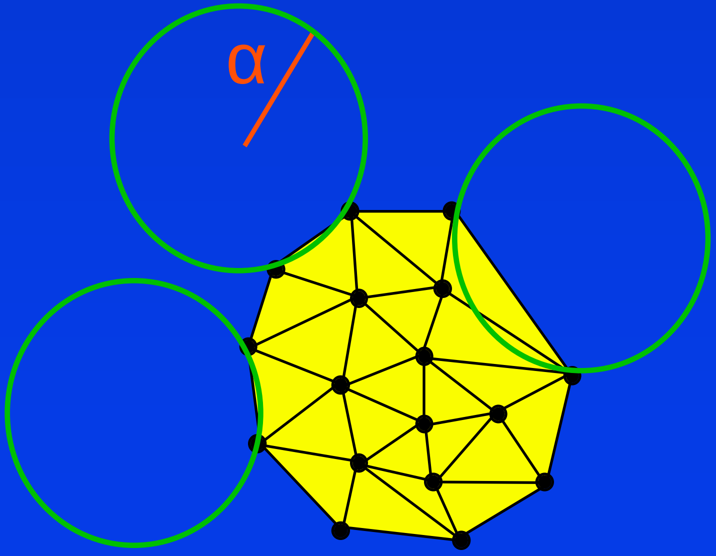


Our Approach

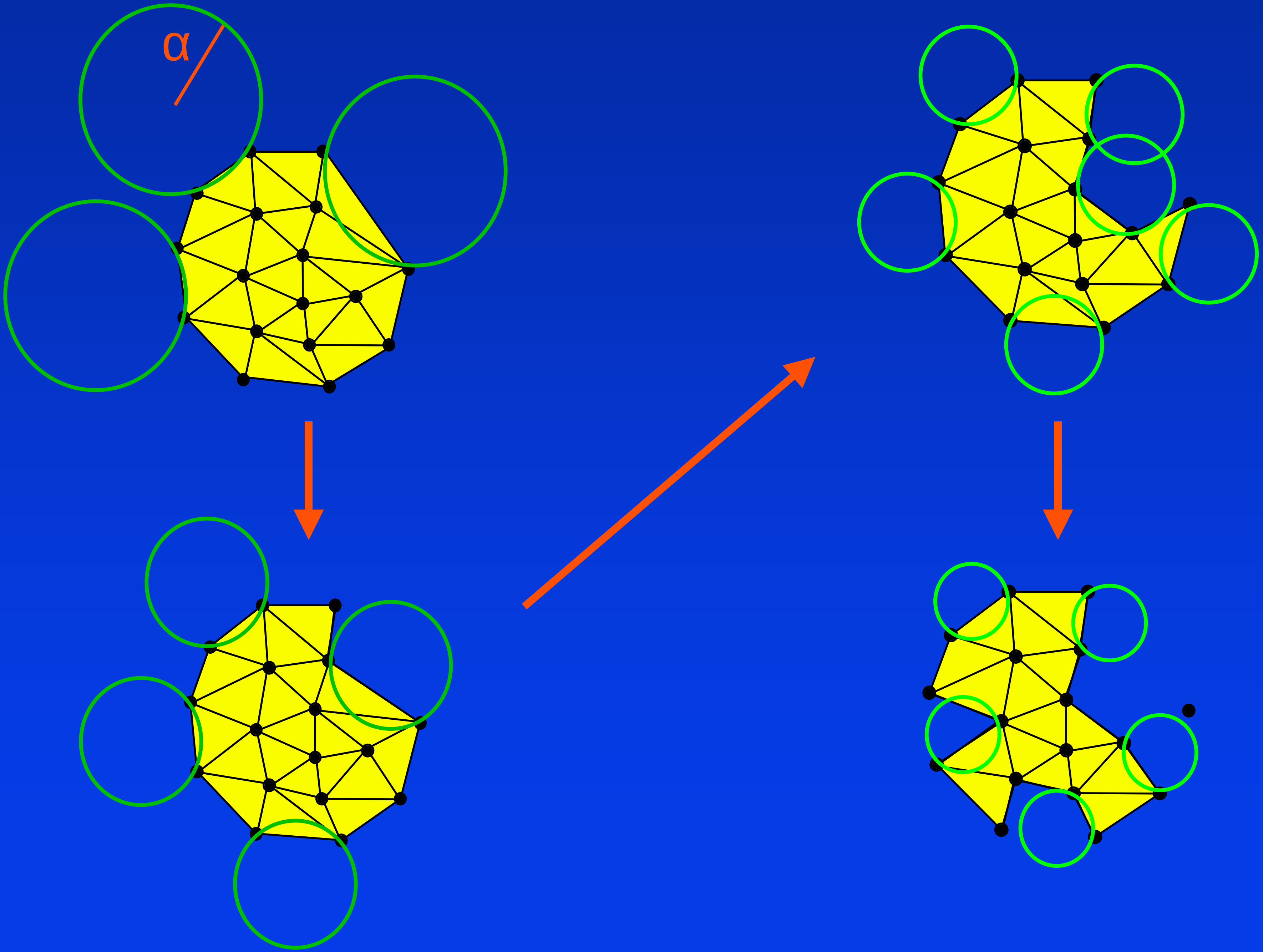
- Surface representation
  - Alpha-Shapes
    - » to obtain a triangulated protein surface
    - » to identify cavities and protrusions on the protein surface
- Surface matching
  - Geometric Hashing
    - » Hierarchical matching at varying resolution
    - » Matching of contiguous patches which have similar curvature and accessibility

What is an Alpha-Shape

- An Alpha-shape:
  - Formalizes the idea of “shape”
  - Captures the entire range of “crude” to “fine” shape representations of a point set
- In 2-dimensions:
  - An edge between two points is “alpha-exposed” if there exists a circle of radius alpha such that the two points lie on the surface of the circle and the circle contains no other points from the point set.

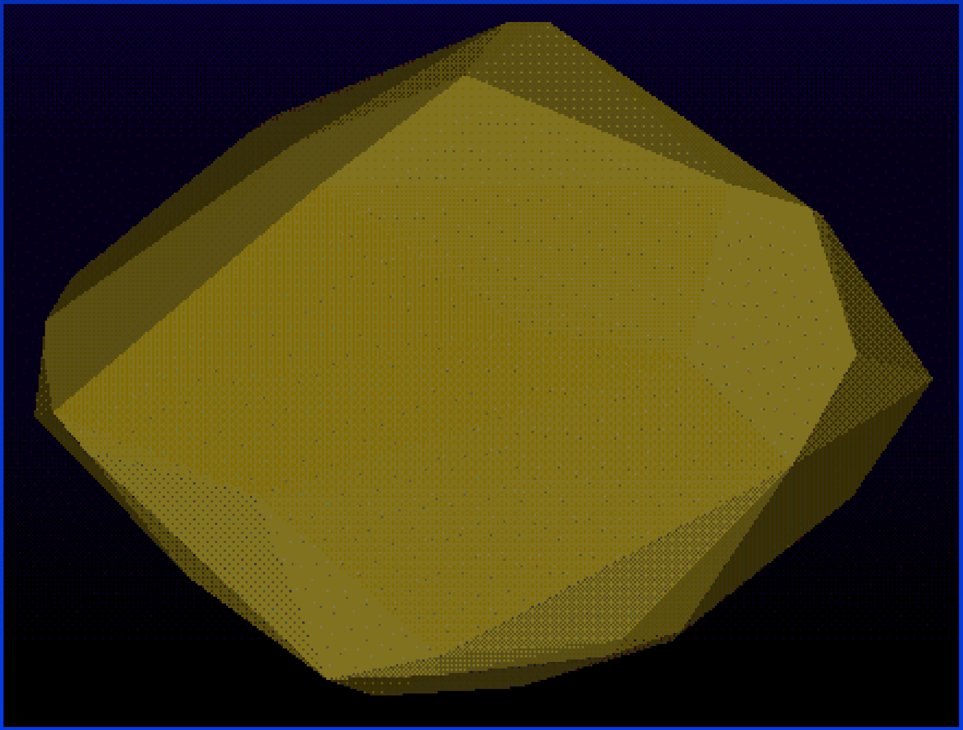


As Alpha decreases ...

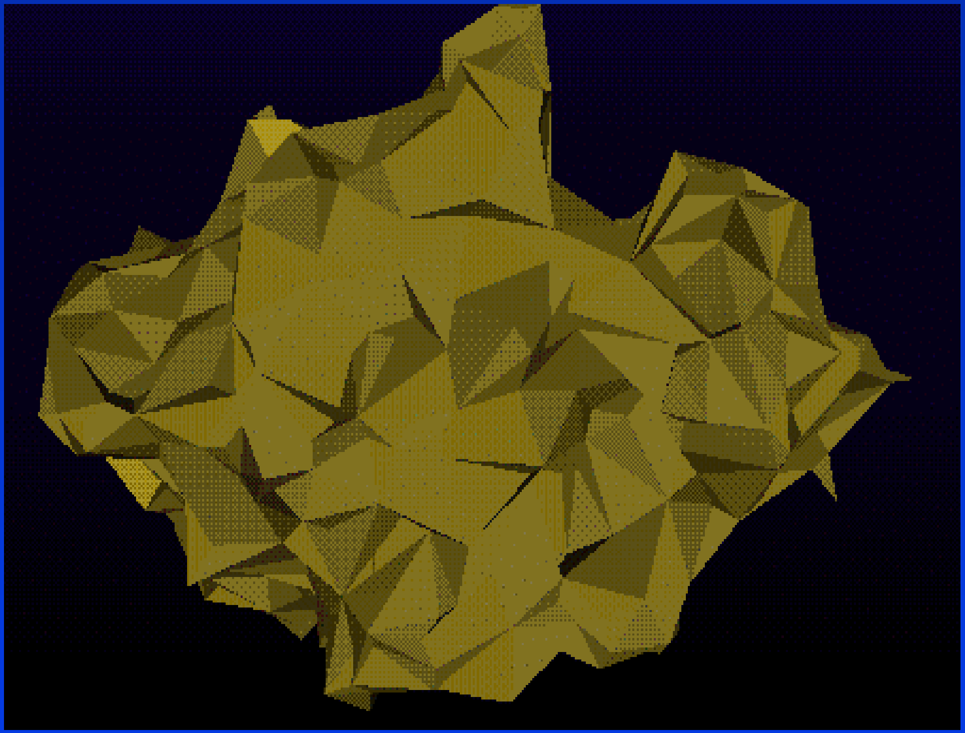


For example ...

Trypsin

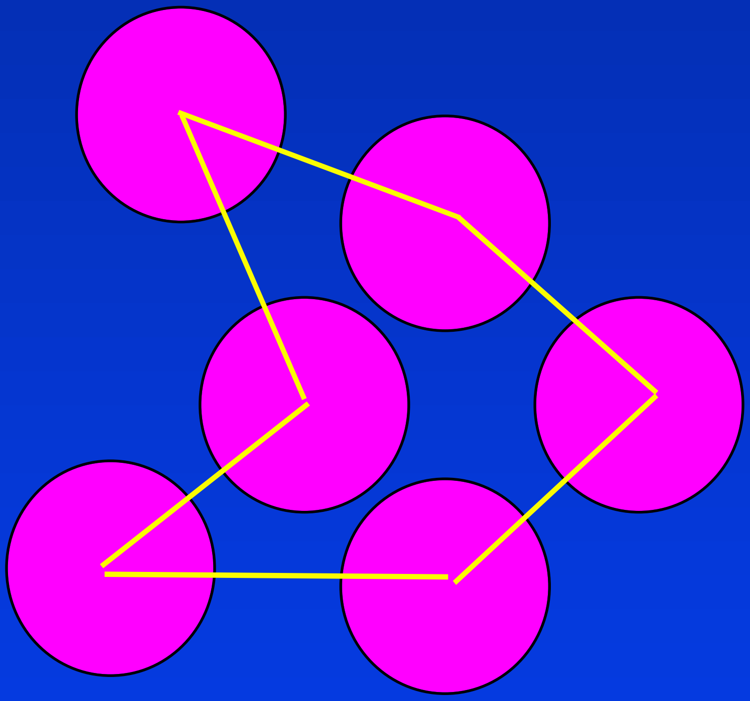


alpha = infinity

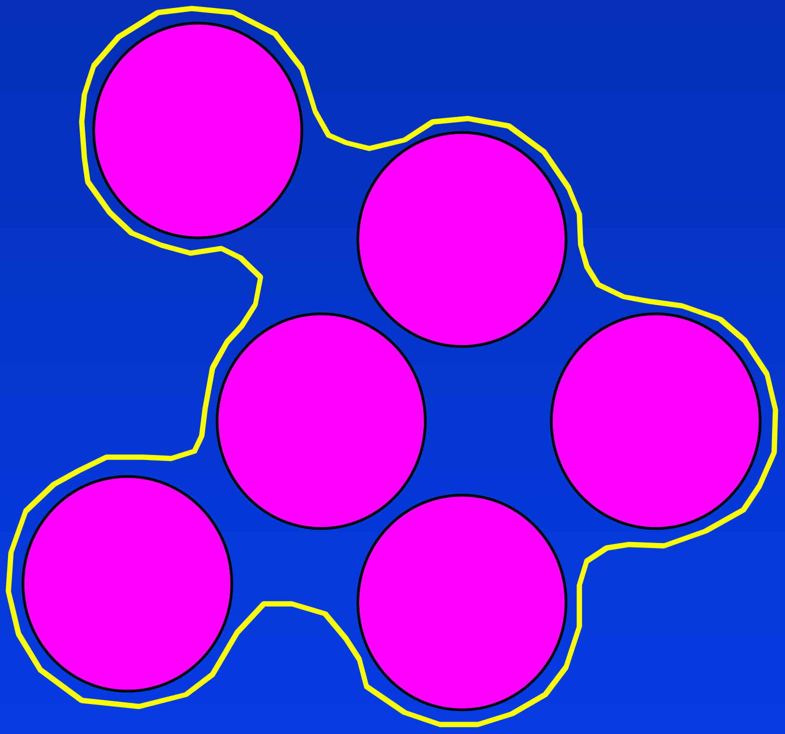


alpha = 3.0 Å

Alpha shape vs. Connolly surface

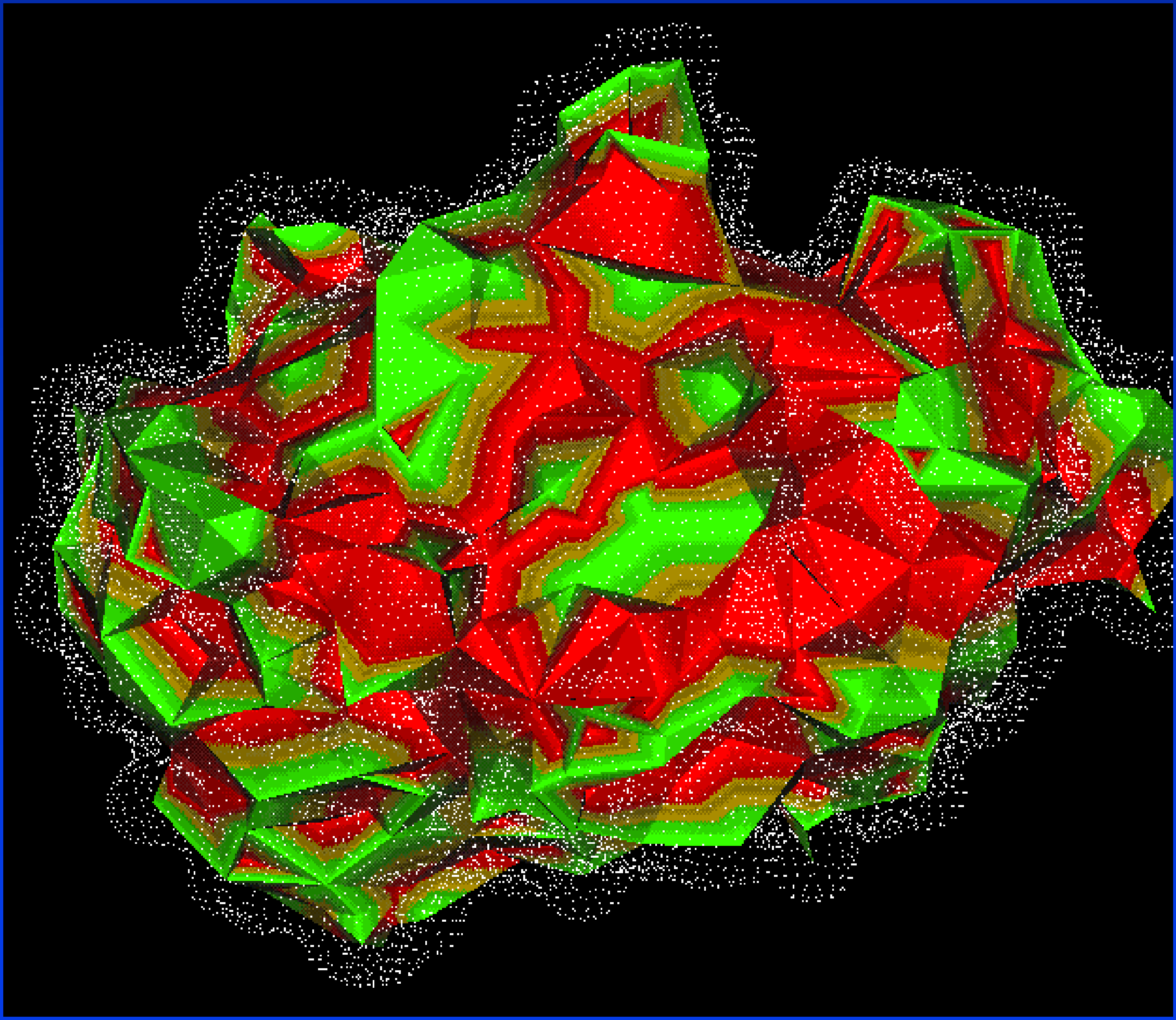


Alpha-shape



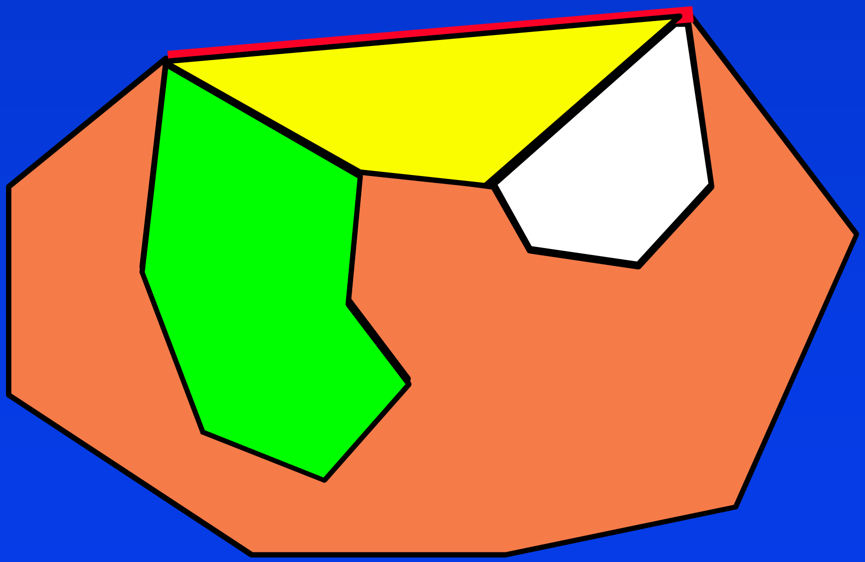
Connolly Surface

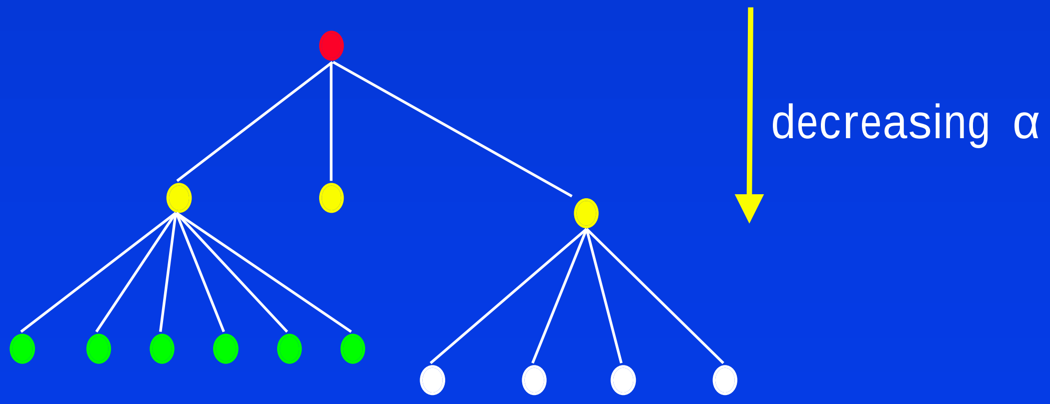
Alpha shape vs. Connolly surface



Identifying Cavities

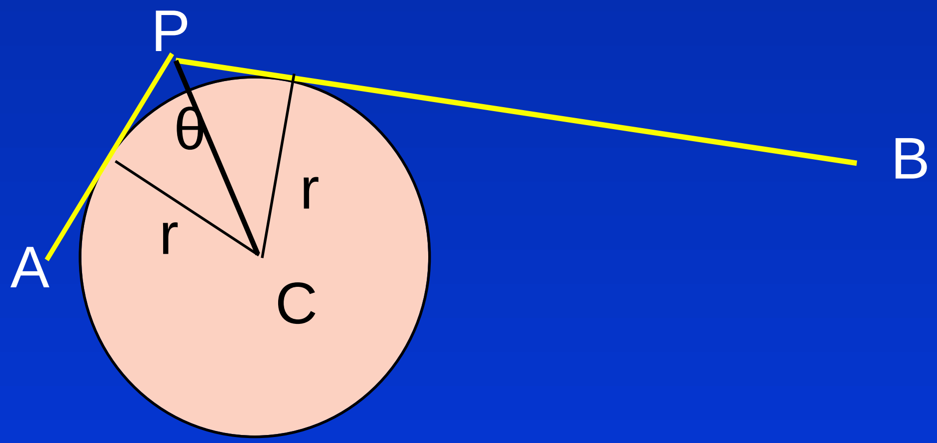
- As alpha decreases, edges appear on the surface and then disappear (as alpha gets even smaller)
- We can compute a hierarchy of cavities by following edges as the appear and then disappear





Curvature and Accessibility

- Curvature can be approximated at each vertex of the surface:



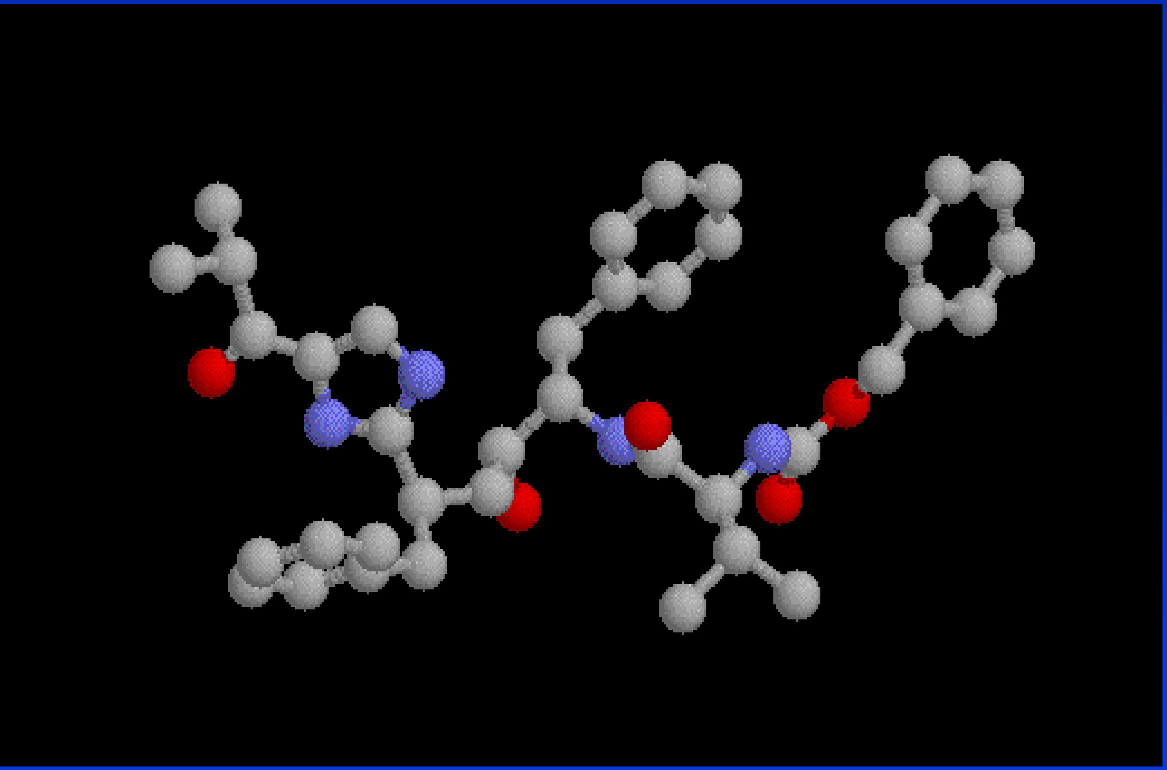
r = [(P-A)/2] * [tan(θ)]

Accessibility of atom *i* is the maximum sized sphere that can touch atom *i* without enclosing any other atoms within the sphere

Comparison

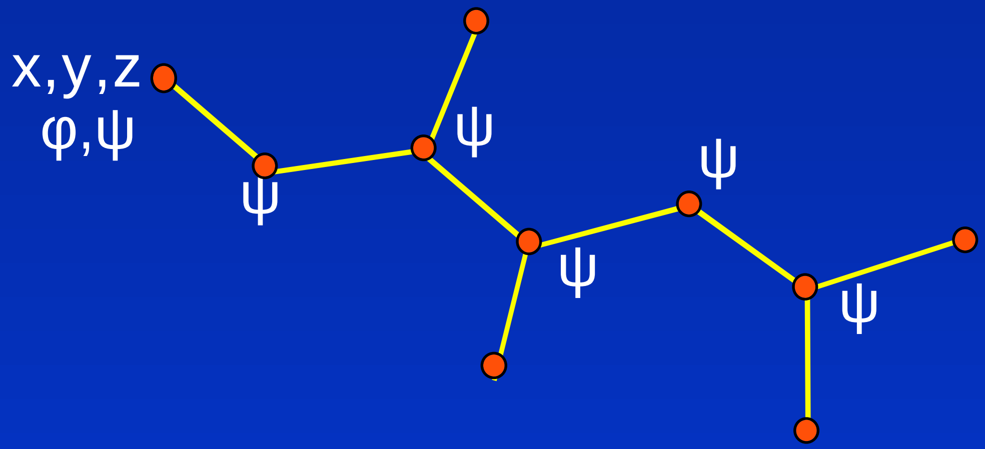
- Disadvantages of using Alpha-Shapes
  - Coarser approximation of the Connolly Surface
- Advantages of using Alpha-Shapes
  - Fewer points to be considered -> faster
  - Allows “fine” and “crude” matching
    - » This may automatically model partial flexibility
  - Additional use of curvature and accessibility to obtain surface patches
  - Matching patches individually may indicate possible hinge sites for flexible docking

Ligand Docking using Robotic Path Planning



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

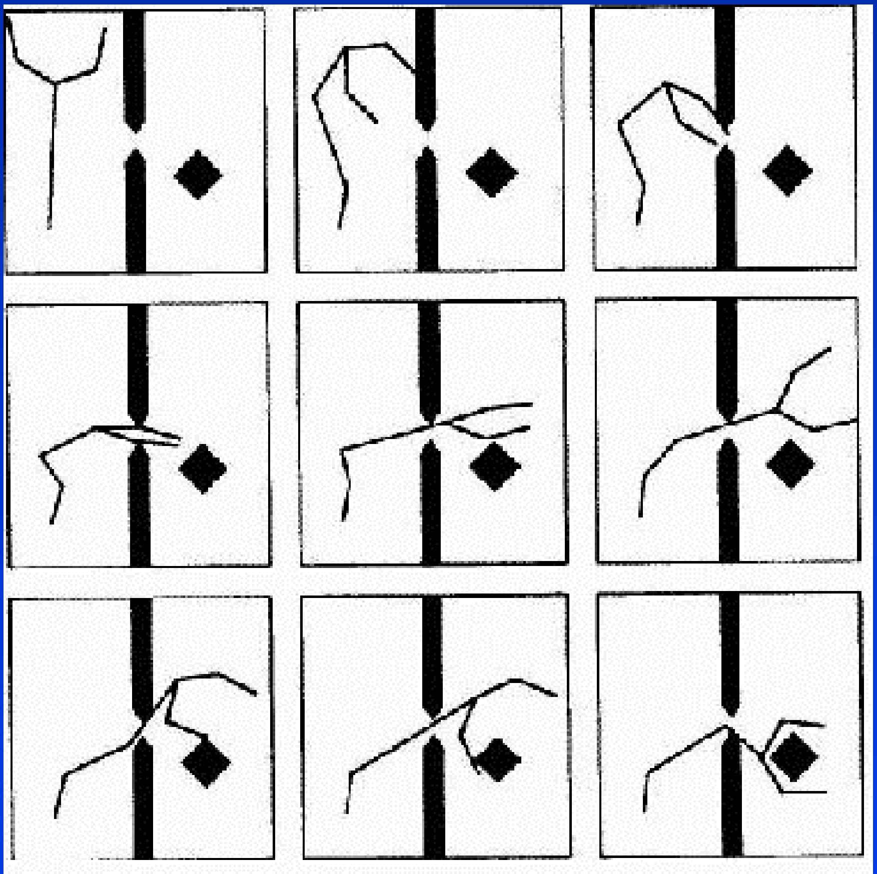
Ligand Modeling



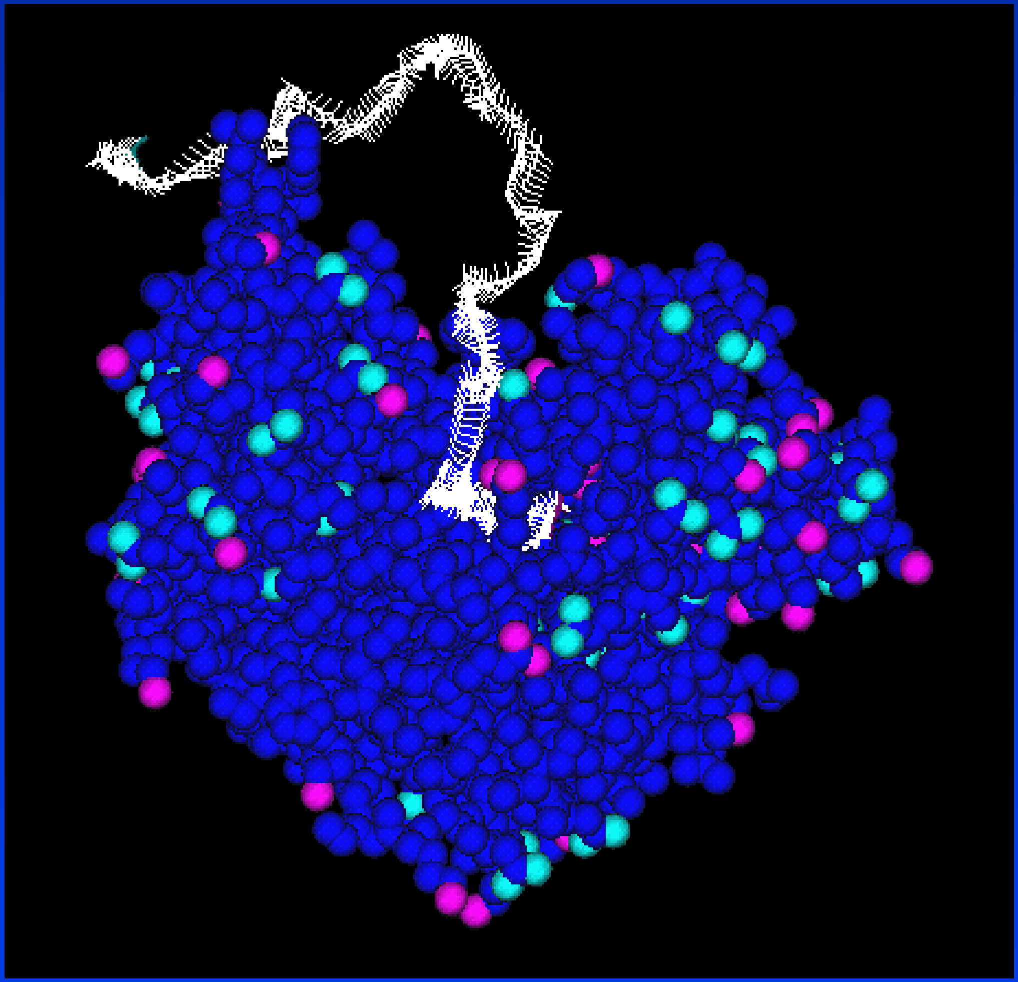
- DOF = 10
  - 3 coordinates to position root atom
  - 2 angles to specify first bond
  - Torsional angles for all remaining non-terminal atoms
  - Bond angles are assumed constant
  - Terminal hydrogens are modeled by increasing radius of terminal atoms

Path Planning

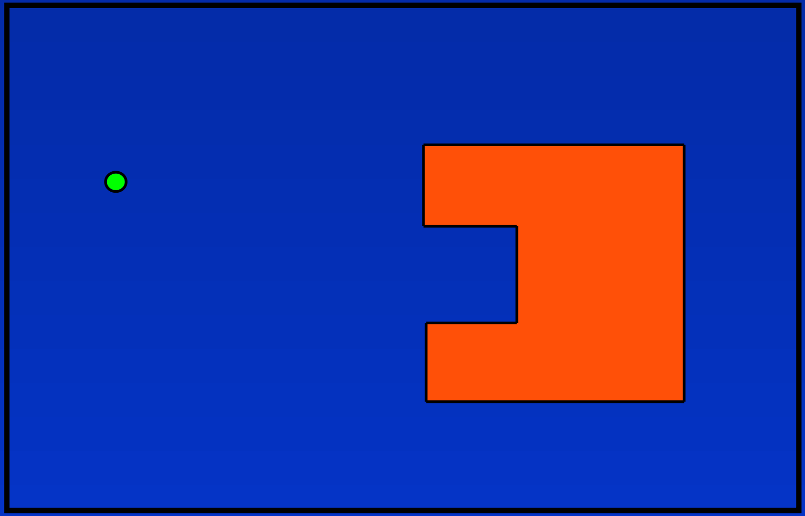
Articulated Robot



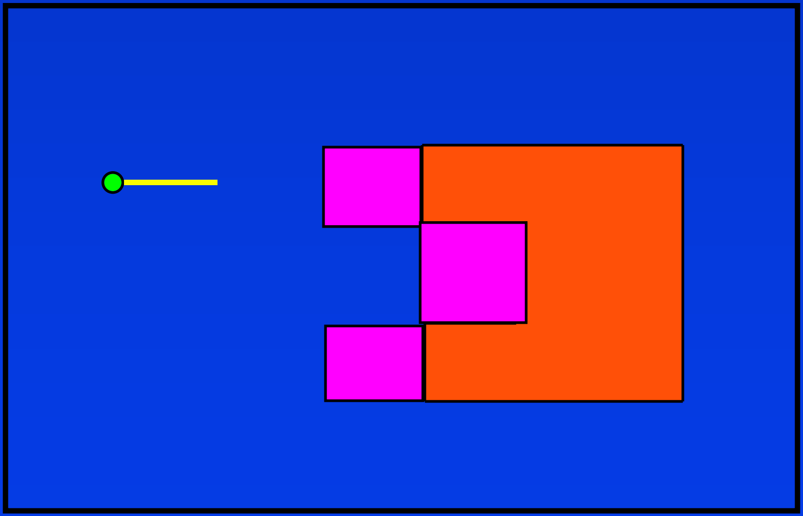
Ligand

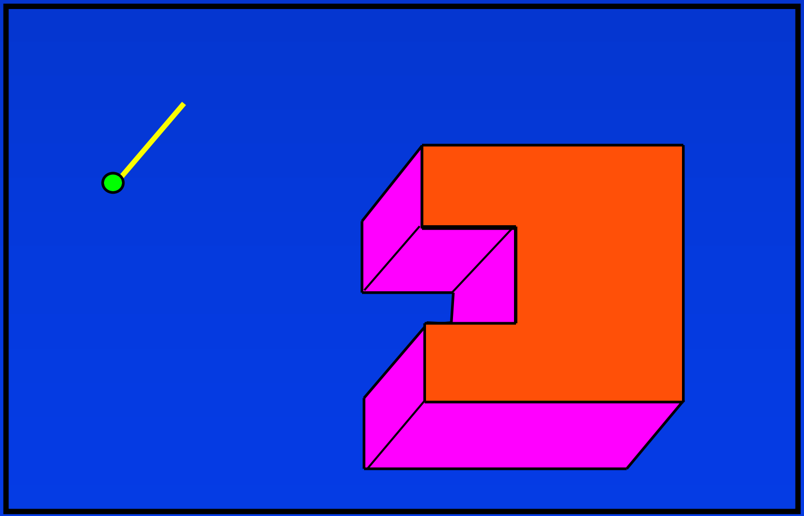


Obstacles in a Workspace



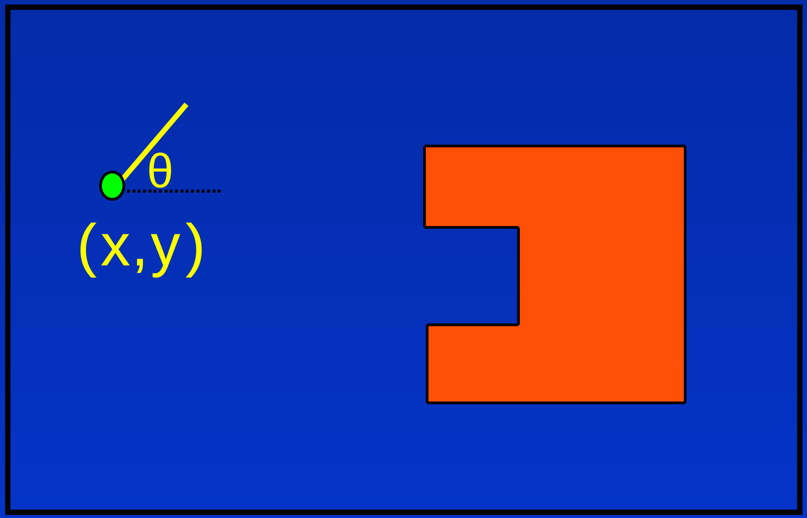
Obstacle seen by a 0-D robot



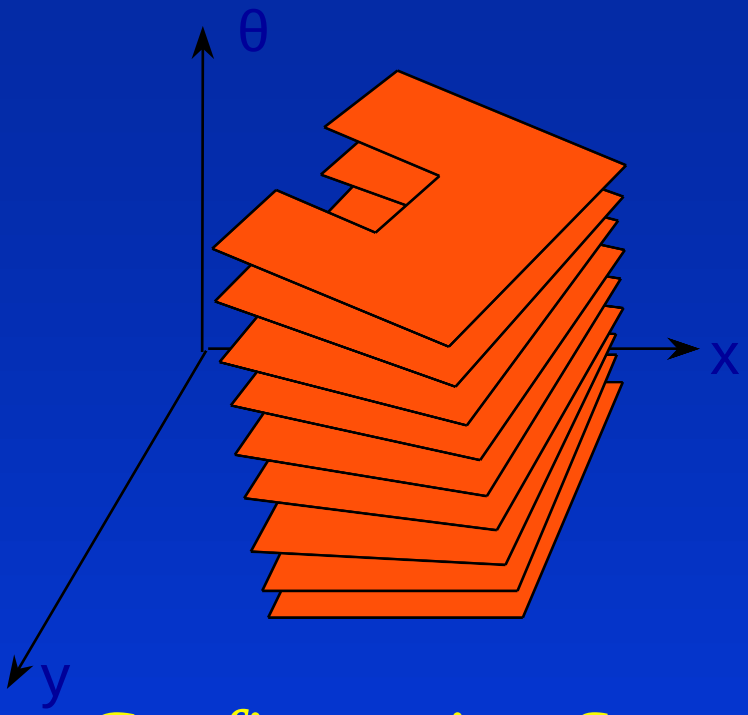


Obstacles seen by fixed orientation 1-D robots

Workspace vs. Configuration Space



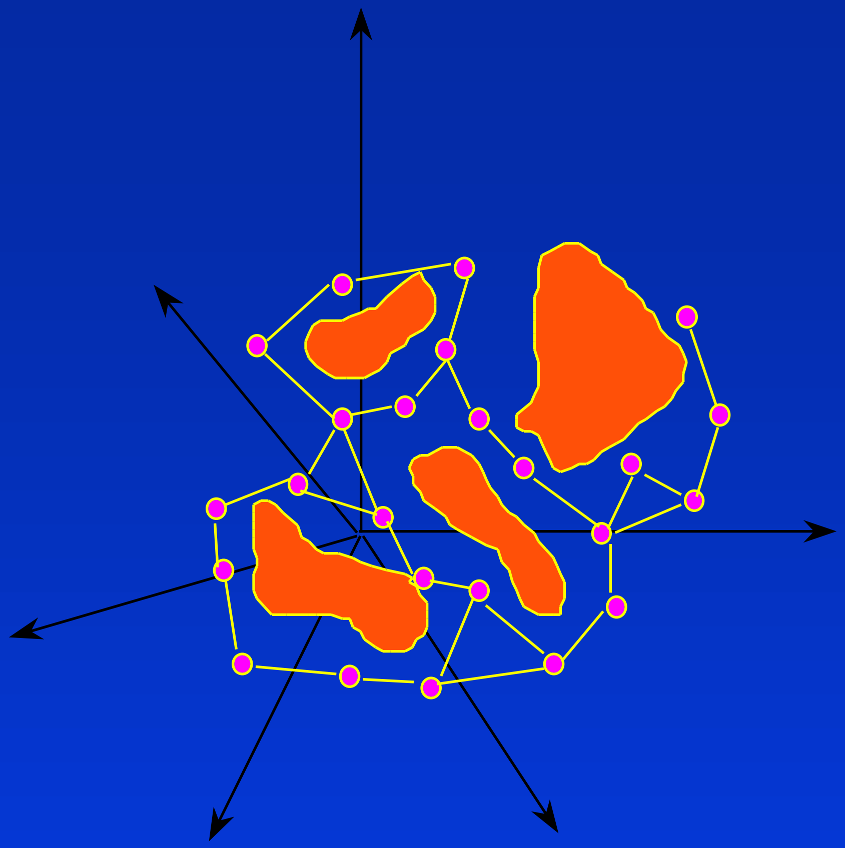
Work Space



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

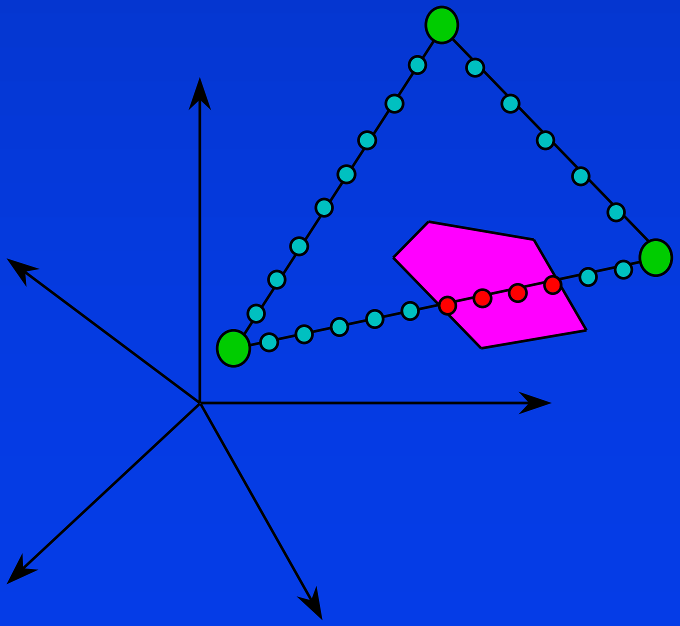
High Degree of Freedom Robots



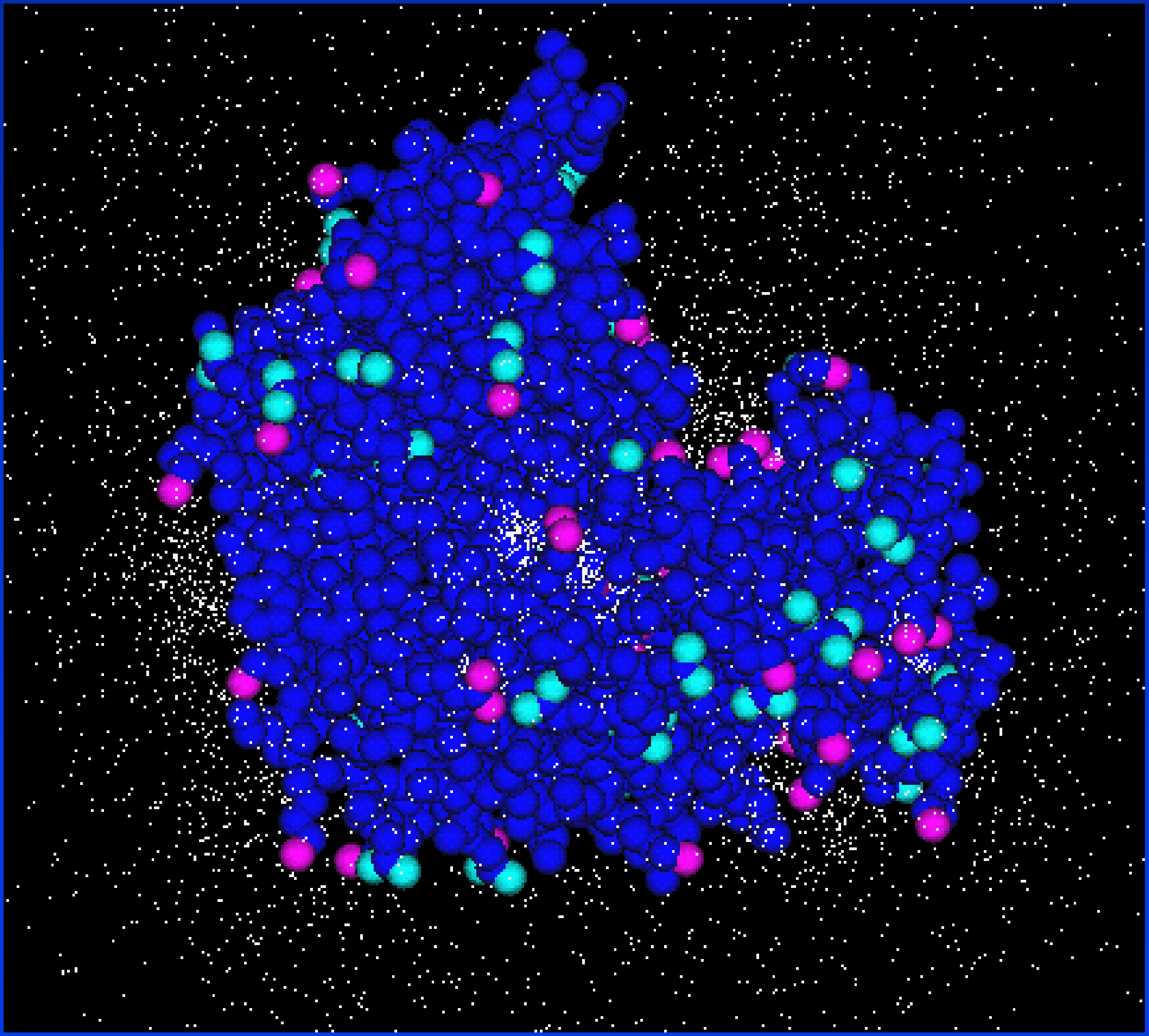
- Complete representation of obstacles in high dimensional configuration space is very difficult
- Hence sample randomly from C-space and only accept samples that are collision free
- Connect nearest nodes with a **local path planner**

Local Path Planner

- Connect any two points 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 of Interaction

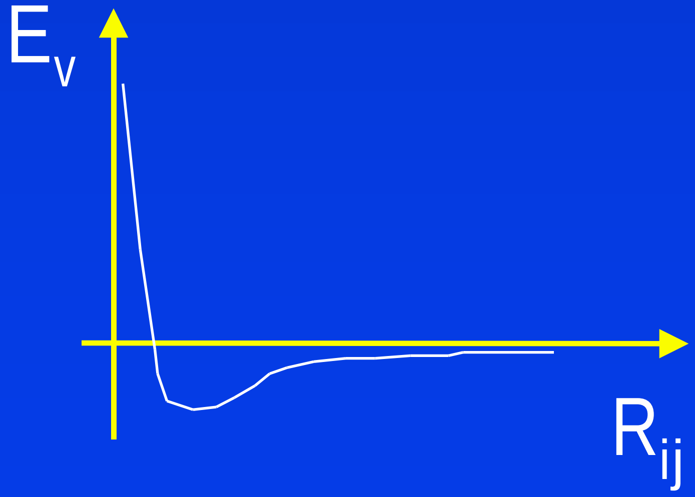
Energy = van der Waals interaction (E_v)

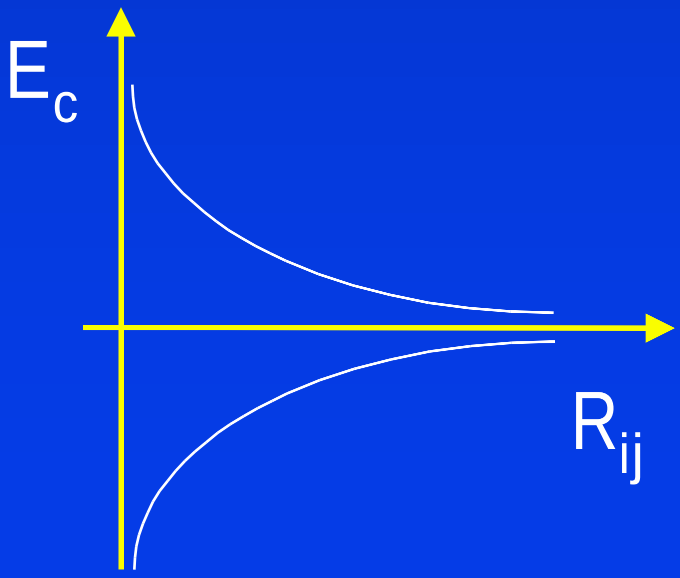
+

electrostatic interaction (E_c)

$$E_v = A/(R_{ij})^{12} - B/(R_{ij})^6$$

$$E_c = Q_iQ_j/(eR_{ij})$$





Solvent Effects

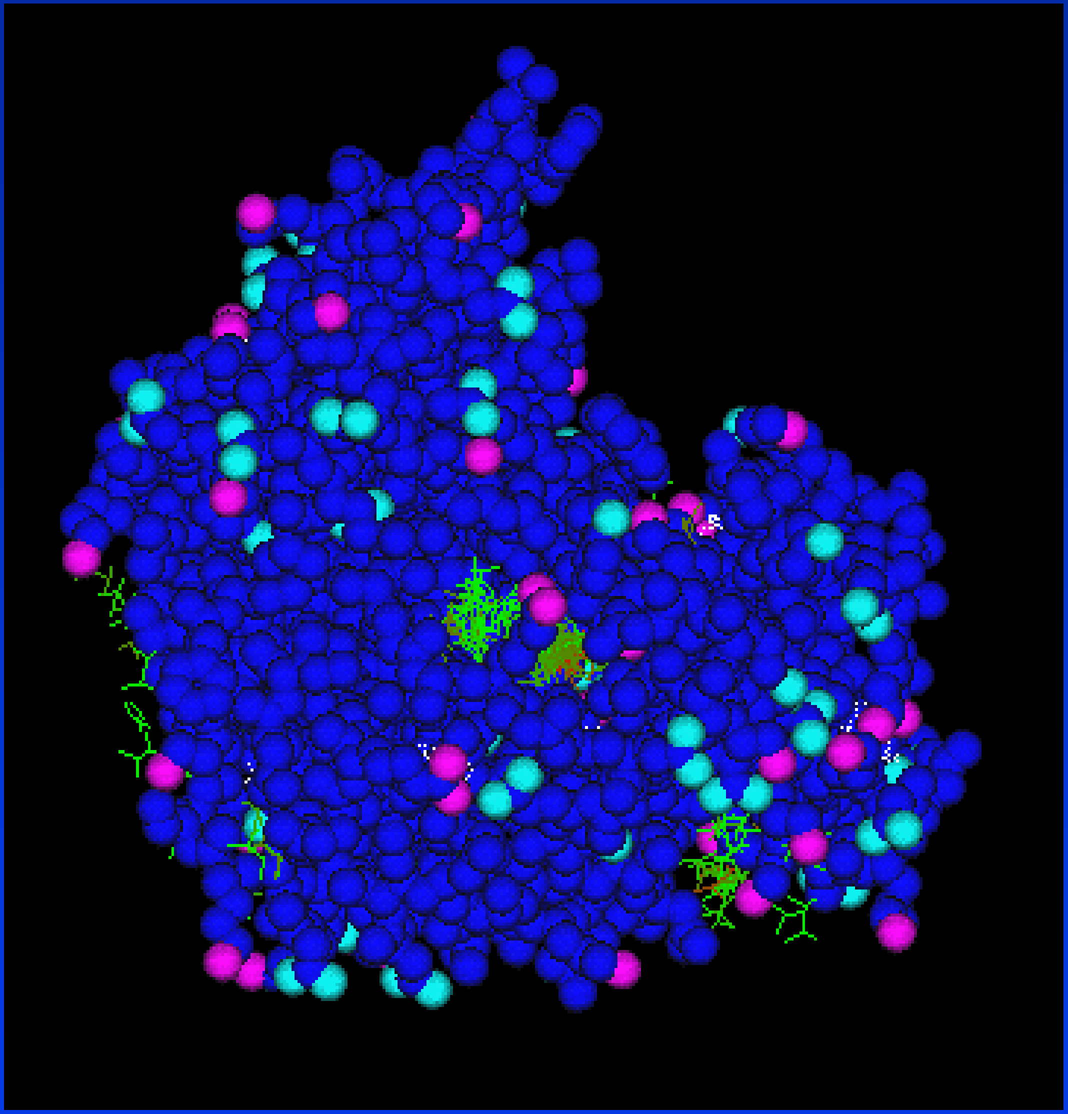
$$E_c = 332\ Q_iQ_j/(\varepsilon R_{ij})$$

- Is only valid for an infinite medium of uniform dielectric
- Dielectric discontinuities result in induced surface charges
- Solution: Poisson-Boltzman equation

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

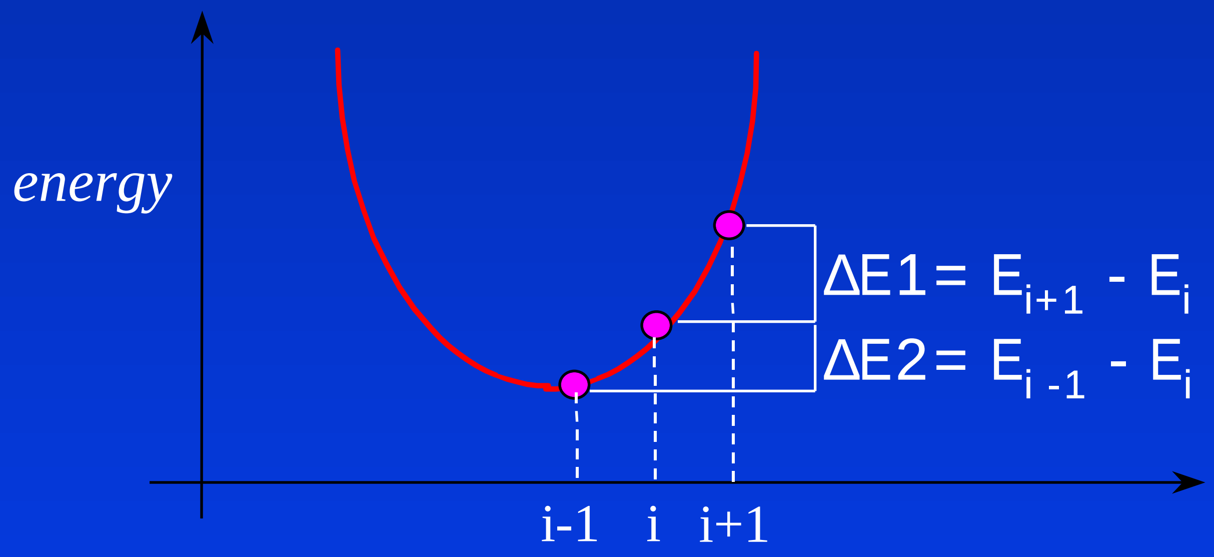
- Models effect of dielectric and ionic strength
- Can only be solved analytically for simple dielectric boundaries like spheres and planes
- Finite Difference solution is based on discretizing the workspace into a uniform grid

Lowest Energy Configurations



Local Path Planning

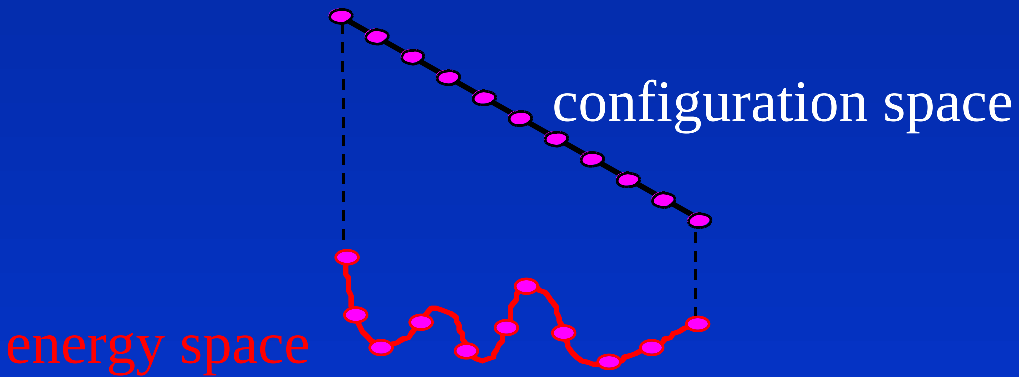
- Need to assign weights to each link in the graph such that the minimum weight path between two nodes corresponds to energetically favourable motion



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

Local Path Planning

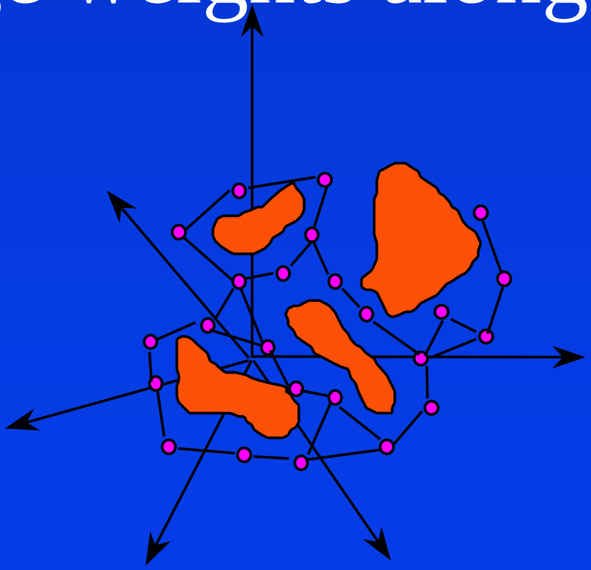
- Edge Weight = Σ - log (Probability of going forward)



configuration space

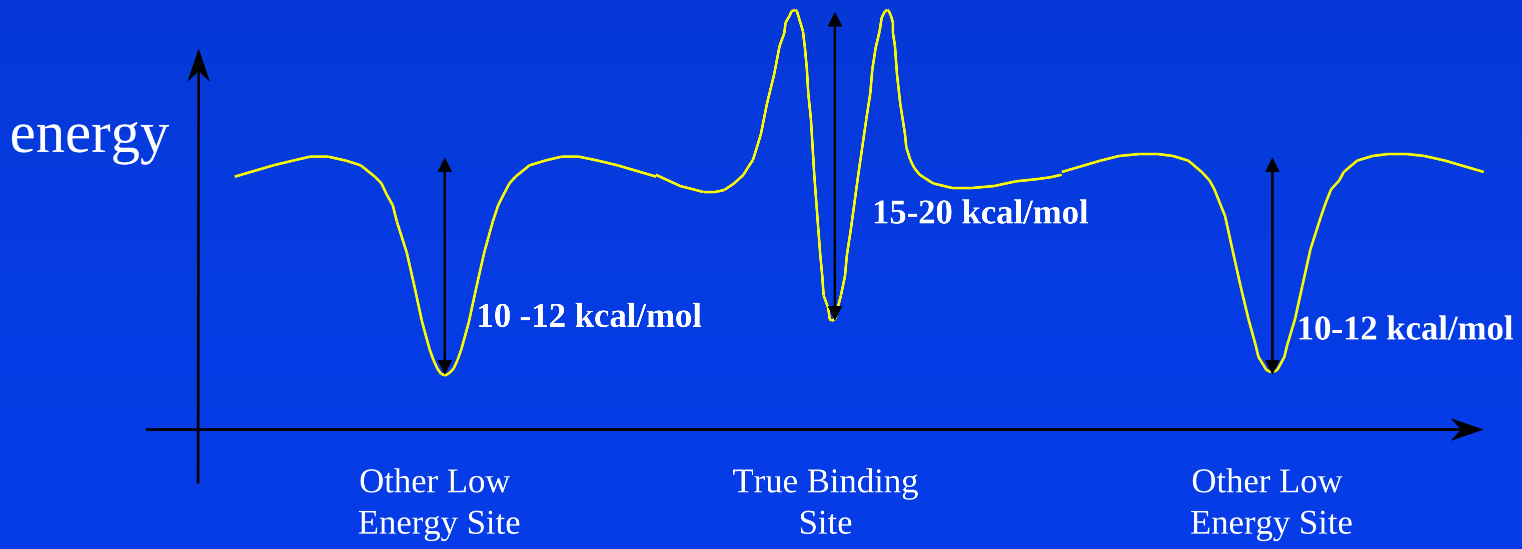
energy space

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



Results - Characterizing the Binding Site

- Tentative 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.



Flexible Ligand Docking