

Protein-Protein Docking

Pankaj K. Agarwal

BioGeometry

Department of Computer Science
Duke University

BioGeometry

Project Structure

A large on-going project

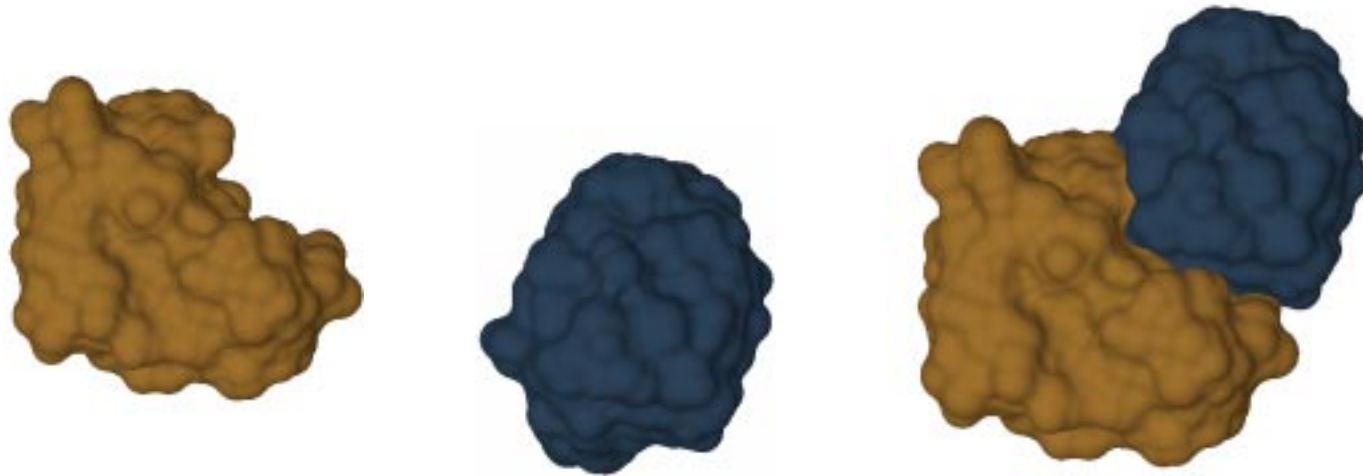
- ☆ Herbert Edelsbrunner
- ☆ Johannes Rudolph
- ☆ Sergei Bespamyatnikh
- ☆ Vicky Choi
- ☆ Yusu Wang
- ☆ Andrew Ban
- ⋮

Protein-Protein Docking: Geometric Issues

Several algorithmic and geometric questions

- ★ Sampling of 3D rotation space
- ★ Identifying critical points
- ★ Shape matching
- ★ Hierarchical representation of proteins
- ★ Kinetic data structures

Protein-Protein Docking



- ☆ Protein-protein docking different from protein-ligand docking
- ☆ Geometry plays an important role

Motivation

- ☆ Protein-protein interaction is key to understanding cellular events
- ☆ Experimental methods to study protein-protein interaction
 - X-ray crystallography
 - Site directed Mutagenesis
 - Electron microscopy, fluorescence spectroscopy, NMR
- ☆ Need new computational methods for understanding protein-protein interaction

Previous Work

☆ [Connolly, 1986]

- Identify a few points on each protein surface
- Match a 4-tuple of selected points on one protein surface with a 4-tuple on the other
- Works well in some cases

☆ Numerous new methods proposed in the last fifteen years

☆ [Halperin et al., 2002]

- Comprehensive survey of known docking algorithms

Previous Approaches

- ☆ Earlier methods simplify geometry
 - Results into false positive matches
 - Geometry is not sufficient to find the correct binding
- ☆ Bias search toward preselected binding sites
- ☆ Incorporate energy functions in search procedure
 - hydrogen bonds, electrostatics
 - solvation, hydrophobicity
- ☆ Energy functions are hard to compute, prone to artifacts, difficult to model.

Our Approach

- ☆ A more precise representation of the geometry of proteins
 - Avoid false positives; may have false negatives
 - Geometry is sufficient in most cases
- ☆ Use clever geometric techniques to speed up the algorithms

Assumptions

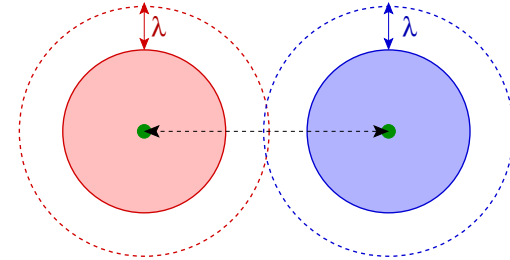
- ☆ Model protein as a rigid body
- ☆ Rigid motion: six degrees of freedom
- ☆ Allow a few collisions between atoms of two proteins

Three Approaches

- ★ Exhaustive search of the 6-dimensional motion space
 - Dense uniform sampling of the motion space
 - Proof of concept
- ★ Adaptive sampling + local optimization
 - Hierarchical sampling schemes
 - Hierarchical representation of proteins
 - Iterative methods for local improvement
- ★ Identify and match *feature* points on protein surfaces + local optimization
 - Algorithms for computing and matching feature points

Score Functions

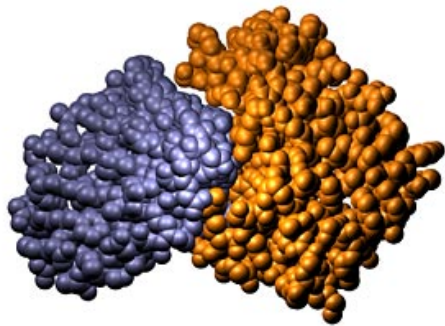
- ★ $A = \{A_1, \dots, A_m\}, B = \{B_1, \dots, B_n\}$: Two proteins
- ★ A is fixed, B allowed to translate and rotate
- ★ \mathcal{C} : configurations of B
- ★ $\pi \in \mathcal{C}, \mu(A, \pi(B))$: Score function
- b : # collision pairs ($b \approx 5$)
- λ : Threshold distance ($\lambda \approx 0.75A$)
- ★ A configuration π *valid* if $|\{(i, j) \mid A_i \cap B_j \neq \emptyset\}| \leq b$
- ★ For a valid configuration of B ,



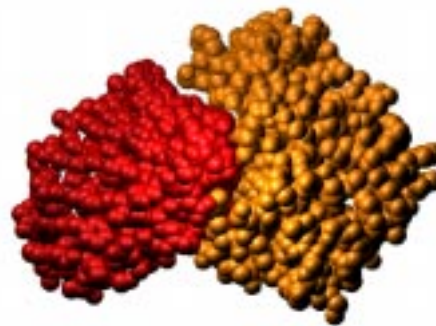
$$\mu(A, \pi(B)) = |\{(i, j) \mid d(A_i, B_j) \leq 2\lambda\}|$$

Exhaustive Search: Sampling

- ☆ A is fixed, B allowed to translate and rotate
- ☆ Rotation is represented by unit quaternions
- ☆ Rotation space: \mathbb{S}^3 ; Translation space \mathbb{R}^3
- ☆ $\mathcal{C} = \mathbb{S}^3 \times \mathbb{R}^3$: Set of all configurations of B .

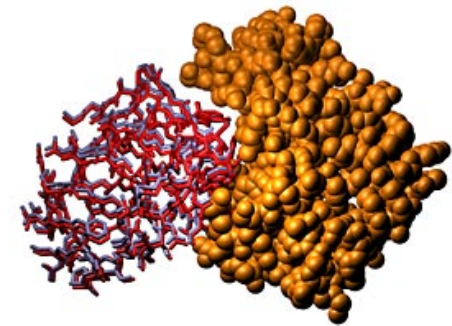


Original binding



Top score

1BRS (Barnase/barstar)

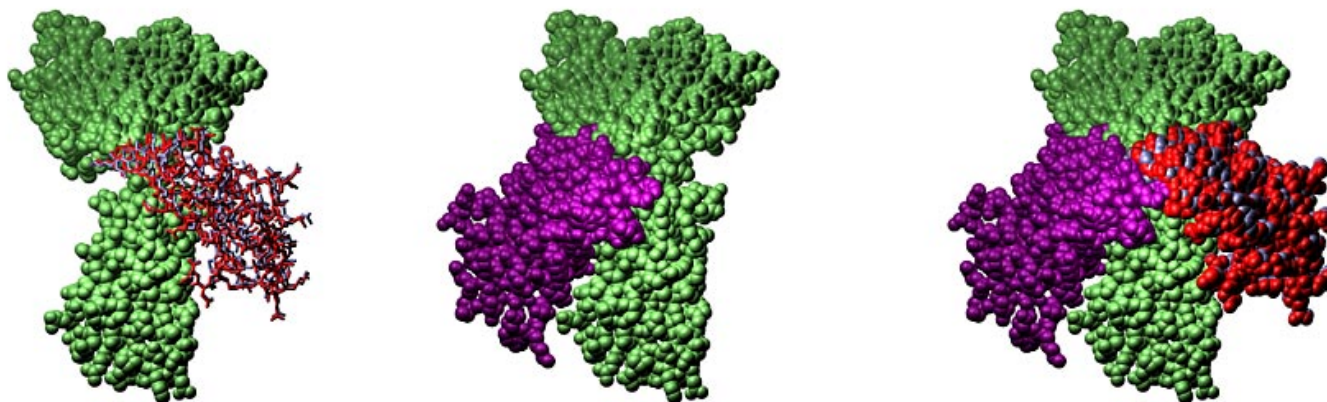


Original + Docked

Chain A (110 AA, 864 atoms), Chain D (89 AA, 693 atoms)

Exhaustive Search: Sampling

Sparse sampling may yield top score for a wrong site

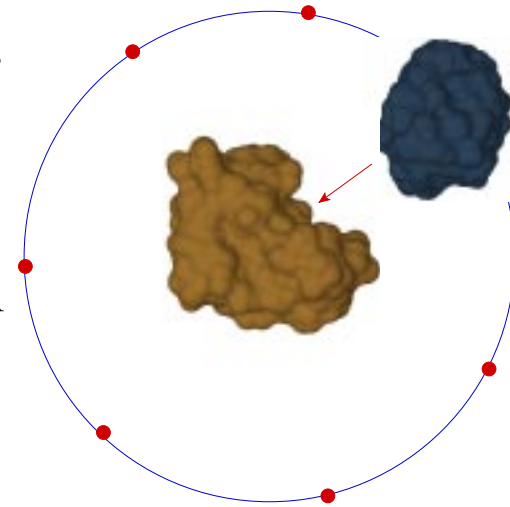


3HLA Histocompatibility antigen

Chain A (270 AA, 2189 atoms) Chain D (99 AA, 829 atoms)

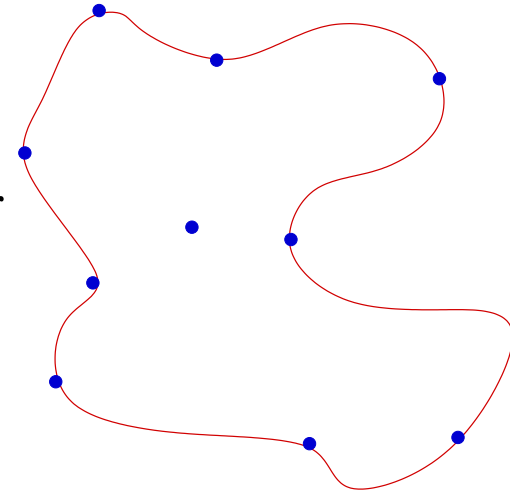
Adaptive Sampling

- ☆ Combine sampling and optimization methods
- ☆ Sparse sampling Π of \mathcal{C}
- ☆ For $\pi \in \Pi$
 - Rotate and translate B , starting from $\pi(B)$
 - Find a locally optimal configuration
- ☆ Detect collision and maintain score function during motion
 - Hierarchical representation
 - Kinetic data structures

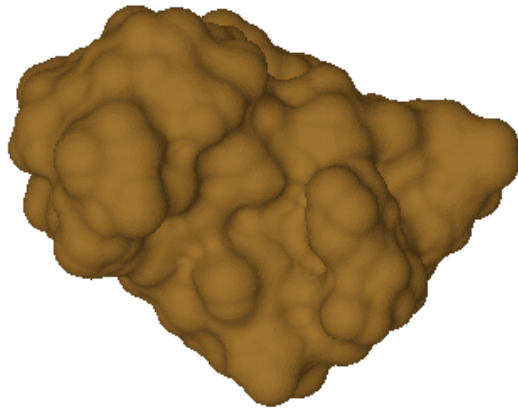


Critical-Point Based Approach

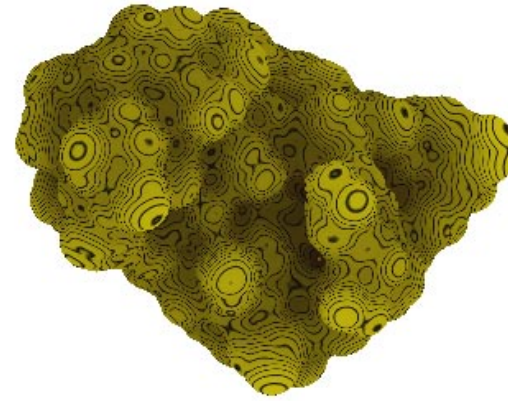
- ☆ Identify cavities and protrusions on A and B
- ☆ Define Morse function f on the molecular surface
- ☆ Compute critical points of f
- ☆ Use topological simplification to identify *main* critical points
- ☆ Find an initial matching between the critical points
- ☆ Use local optimization to refine initial matching



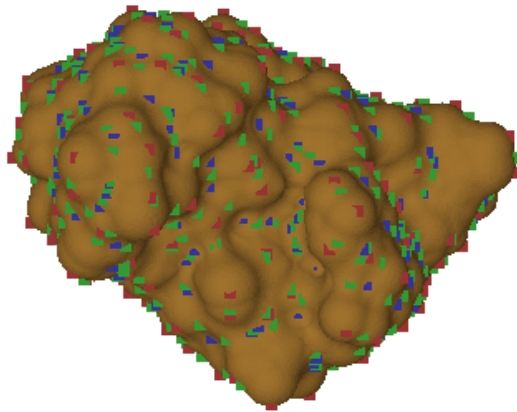
Morse Function: An Example



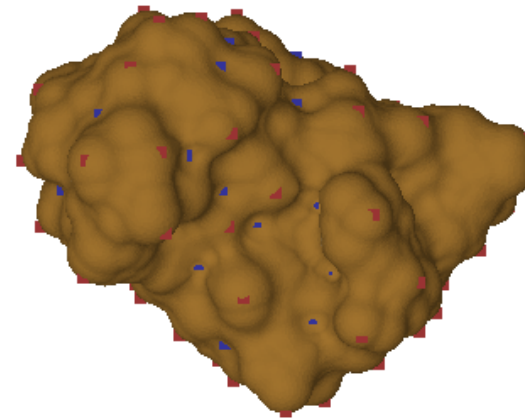
Original protein



Morse function



417 Critical points



100 Critical points

Conclusions & Research Directions

- ☆ Detailed geometry avoids false positives
- ☆ Better sampling, local optimization techniques needed
- ☆ Implementation of critical-point based approach in 3D
- ☆ Hierarchical shape matching
- ☆ Improved models for collision
- ☆ Modeling flexibility