

# CS597A Structural Bioinformatics

Thomas Funkhouser  
Princeton University  
Fall 2007

## Outline

### Overview of structural bioinformatics

- Goals
- Challenges
- Applications

### Overview of course

- Goals
- Lectures
- Coursework
- Projects

## Bioinformatics

### Definition:

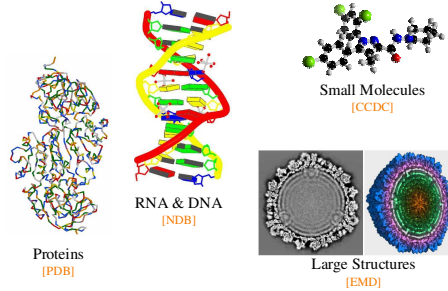
- "The collection, archiving, organization, and interpretation of biological data." [Orengo, 2003]

### Generally speaking:

- Large biological data sets
- Computer representations and algorithms
- Storage, analysis, visualization, prediction, and design

## Structural Bioinformatics

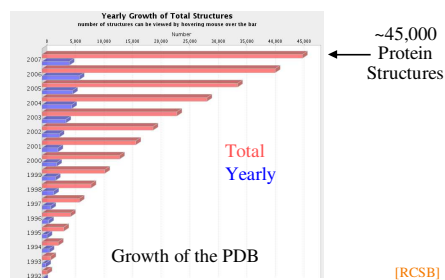
### Focus on data sets with molecular structure



## Structural Bioinformatics

### Motivation 1:

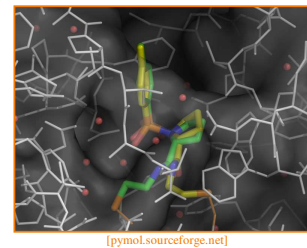
- Lots of structural data is becoming available



## Structural Bioinformatics

### Motivation 2:

- Structural data enables detailed understanding of molecules and their interactions





## Structural Bioinformatics

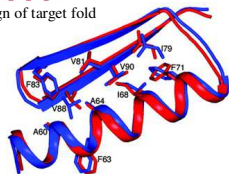


### Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Ø Design



Conceptual design of target fold



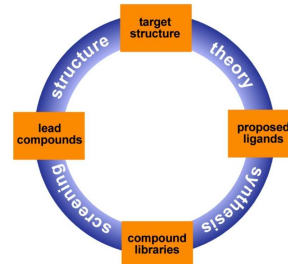
Computationally designed protein (blue) overlaid with solved x-ray structure (red)  
[Kuhlman03]

## Structural Bioinformatics



### Applications:

- Biology
- Medicine
- Chemistry
- Chem Eng
- Agriculture
- Material science
- Nanotechnology
- etc.



Example: Structure-Based Drug Design

## Structural Bioinformatics



### Research challenges:

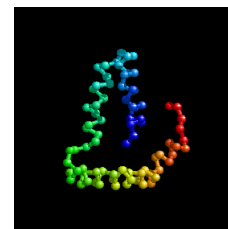
- Structure determination
- Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.

## Structural Bioinformatics



### Research challenges:

- Ø Structure determination
- Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



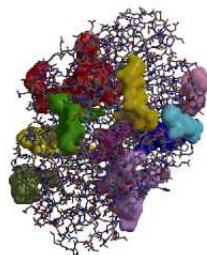
Protein Folding Simulation  
[David Jones]

## Structural Bioinformatics



### Research challenges:

- Structure determination
- Ø Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



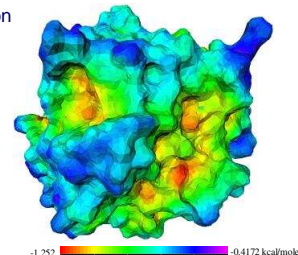
Surface Clefts detected in 1b14  
[Surfnet]

## Structural Bioinformatics



### Research challenges:

- Structure determination
- Binding site detection
- Ø Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



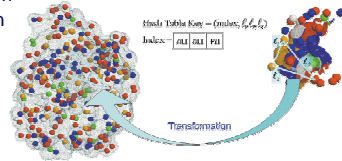
Surface potential for the Carboxypeptidase A  
[Continuum Dynamics]

## Structural Bioinformatics



Research challenges:

- Structure determination
- Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



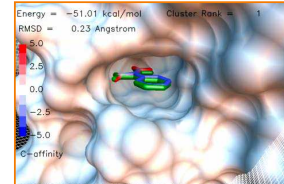
[Shulman-Peleg04]

## Structural Bioinformatics



Research challenges:

- Structure determination
- Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



Camphor binding to Cytochrome P-450 (2cpp)

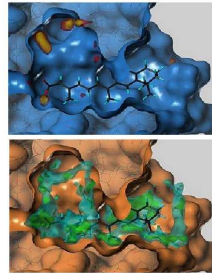
[AutoDock]

## Structural Bioinformatics



Research challenges:

- Structure determination
- Binding site detection
- Binding site modeling
- Binding site matching
- Binding prediction
- Molecular design
- etc.



Superficial FCG assay for the binding site of the protein-coupled receptor 1 (GPCR). The legend is not visible. Blue binding site shows map for carbonyl degree probe. Orange binding site shows map for urethane C=O surface probe.

[CCDC]

## Outline



Overview of structural bioinformatics

- Goals
- Challenges
- Applications

Overview of course ←

- Goals
- Lectures
- Coursework
- Projects

## CS597A



Goals

- Survey current methods in structural bioinformatics
- Investigate specific research problems in depth
- Build shared infrastructure for research
- Build links across disciplines

## Lectures



Class meetings:

- One topic per meeting
- 50% lecture
- 50% discussion

Speakers:

- Professors
- Guests
- Students

Day	Topic	Speaker	Notes	Prereq	Prereq
Mon	Intro				
Tue	Protein structure determination				
Wed	Protein structure determination				
Thu	Protein structure determination				
Fri	Protein structure determination				
Sat	Protein structure determination				
Sun	Protein structure determination				
Mon	Protein structure determination				
Tue	Protein structure determination				
Wed	Protein structure determination				
Thu	Protein structure determination				
Fri	Protein structure determination				
Sat	Protein structure determination				
Sun	Protein structure determination				
Mon	Protein structure determination				
Tue	Protein structure determination				
Wed	Protein structure determination				
Thu	Protein structure determination				
Fri	Protein structure determination				
Sat	Protein structure determination				
Sun	Protein structure determination				

## Lectures



### Student presentations:

- Choose topic from schedule
- Submit list of relevant papers/resources
- 15-20 minute presentation
- Lead discussion afterwards

### Related talks

- Seminars
- Colloquia

Tomorrow (Sep 19) at 12:30PM in CS 402:

Barry Honig, Columbia University  
"Relating Cellular to Molecular Specificity"

PICASso Series on Computation and Data Analysis  
in Biology and Information Sciences

## Readings



### Book

- P.E. Bourne, H. Weissig, "Structural Bioinformatics," Wiley-Liss, 2003.

### Papers

- Read ~1 paper for each lecture
- Be prepared for class discussion



## Course Projects



### Description

- Investigate research problem
- Any problem(s) related to course
- Multi-disciplinary teams (if possible)

### Dates

- Project 1: Analysis
  - § Presentation on Oct 25
  - § Report due Nov 15
- Project 2: Prediction/Design
  - § Presentation on Dec 17
  - § Report due ~Jan 15

## Course Projects

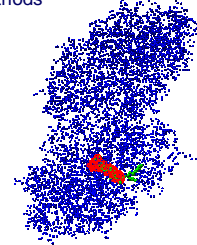


### Possible approaches

- Develop new method
- Evaluate/compare existing methods
- Combine methods in system
- Apply method

### Key elements

- Investigation
- Evaluation

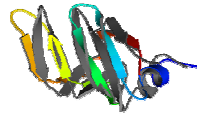
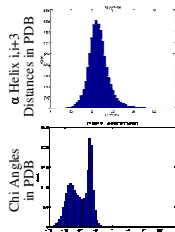


## Course Projects



### Example project 1:

- Scott McAllister: "Generating Likely Distance Bounds for Efficient Protein Structure Prediction"



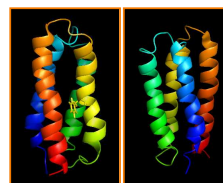
Rank	Original		Distance Bound		Both Bounds	
	Energy	RMSD	Energy	RMSD	Energy	RMSD
1	-412.7	10.11	-407.6	7.55	-365.4	11.21
2	-397.0	10.88	-378.3	3.88	-329.5	10.18
3	-388.5	7.37	-375.3	10.35	-310.2	10.24
4	-369.9	6.40	-363.6	7.41	-305.9	10.98
Best	-262.4	5.50	-378.3	3.88	-209.3	4.93

## Course Projects

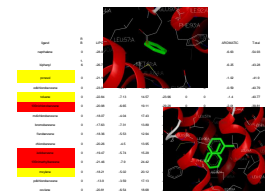


### Example project 2:

- Jermont Chen: "Binding Studies of de Novo Proteins"



De Novo Protein (824) Mutation Phe → Ala (F64A)



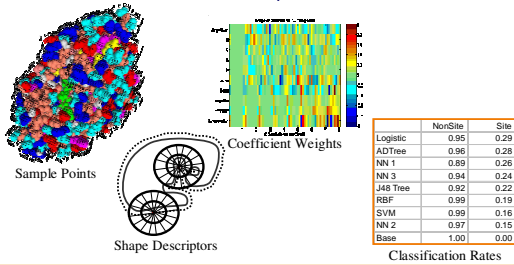
Binding affinities of small molecules computed with a docking simulation (yellow binds in NMR study, red does not)

## Course Projects



### Example project 3:

- Julie Chen and Phil Shilane: "Protein Analysis with Geometric Distribution of Properties"

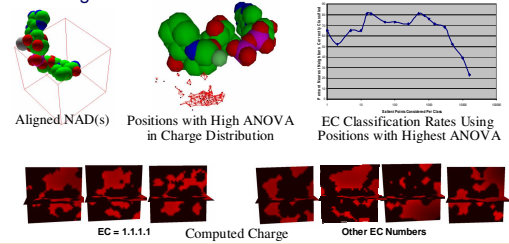


## Course Projects



### Example project 4:

- Josh Lees and Josh Podolak: "Intraclass and Interclass Electric Field Variance as a Tool for Analyzing NAD+-binding Sites"



## Wrap Up



### Students' to do list:

- Hand in survey questionnaire
- Send me a picture of yourself by email
- Sign up for in-class presentations
- Start thinking about project topics
- Attend Barry Honig's talk

## Questions



?