



# Detection of Protein Binding Sites II

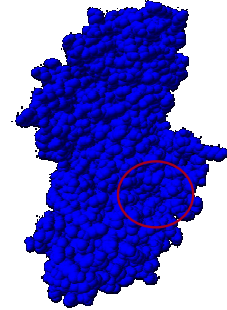
Thomas Funkhouser  
Princeton University  
CS597A, Fall 2005



## Introduction

Goal:

- Given a protein structure, predict where a ligand might bind



lhld



## Outline

- Introduction
- Binding site localization methods
  - Geometric ← Last Lecture
  - Evolutionary
  - Chemical
- Evaluation methods
- Discussion



## Outline

- Introduction
- Binding site localization methods
  - Geometric
  - Evolutionary } This Lecture
  - Chemical
- Evaluation methods
- Discussion



## Binding Site Properties

Residue properties:

- Depth
- Propensity
- Conservation
- Charge
- Hydrophobicity
- Secondary structure type
- Destabilization

Surface/volume properties:

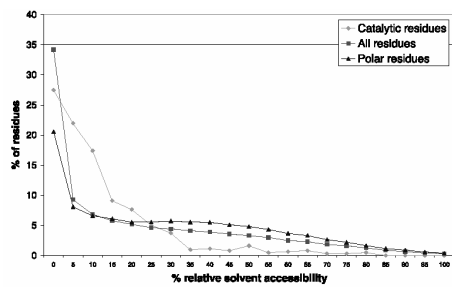
- Pocket size (last time)
- van der Waals force field (last time)
- Electrostatic potential

Other properties

- Microscopic titration curves

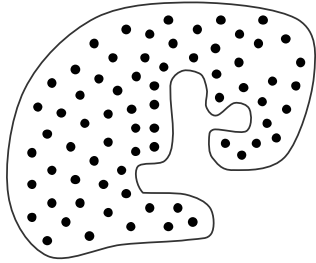


## Residue Solvent Accessibility

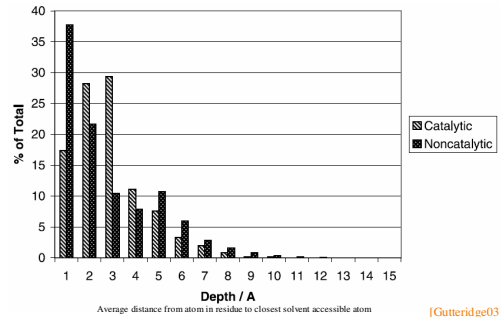


[Bartlett02]

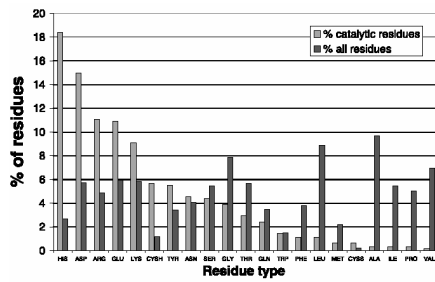
## Residue Depth



## Residue Depth

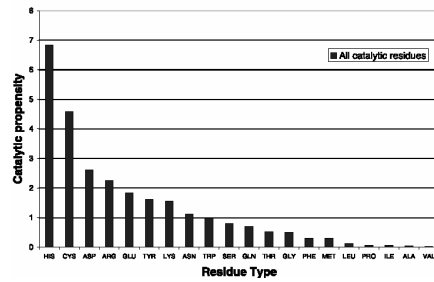


## Residue Propensity



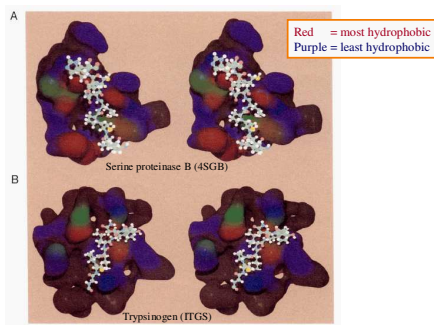
[Bartlett02]

## Residue Propensity



[Bartlett02]

## Residue Hydrophobicity



## Residue Hydrophobicity

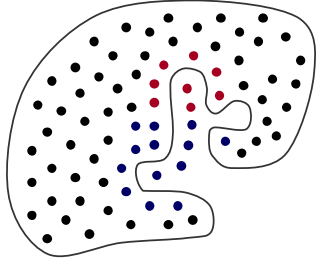


	Catalytic Residues	All Residues
Charged	65%	25%
Polar	27%	25%
Hydrophobic	8%	50%

% Catalytic residues (as compared to all residues) in data set with 178 enzymes

[Bartlett02]

## Residue Hydrophobicity



## Secondary Structure

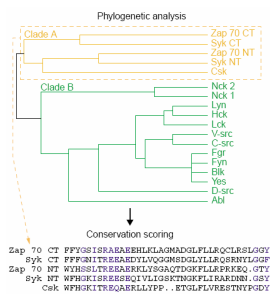


	Catalytic Residues	All Residues
Alpha helix	28%	47%
Beta sheet	22%	23%
Coil	50%	30%

% Catalytic residues (as compared to all residues)  
in data set with 178 enzymes

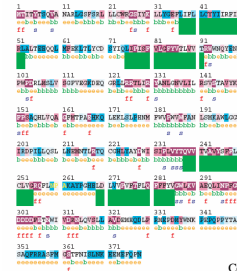
[Bartlett02]

## Residue Conservation



[Campbell03]

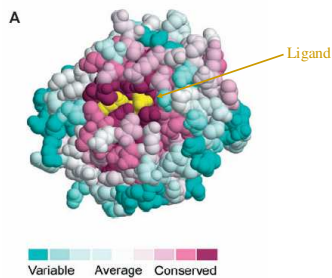
## Residue Conservation



ConSeq predictions demonstrated on human bestrophin  
using 43 homologues obtained from the Pfam database  
(SWISS-PROT: VMD2\_HUMAN) (family code: DUF289)

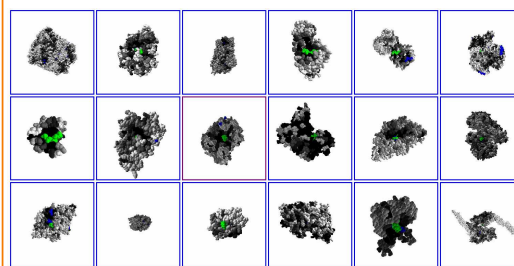
[Berezin04]

## Residue Conservation



[Nimrod05]

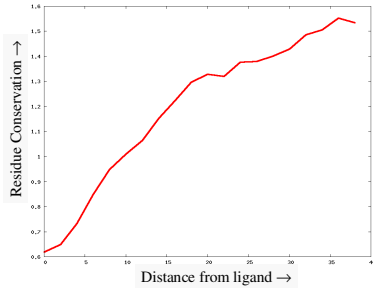
## Residue Conservation



Less Conserved

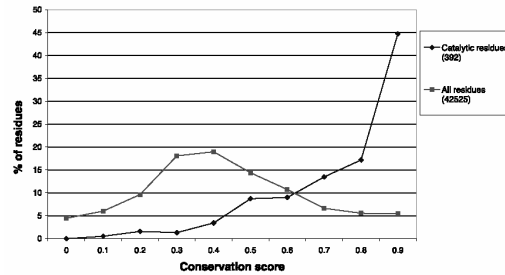
More Conserved

## Residue Conservation



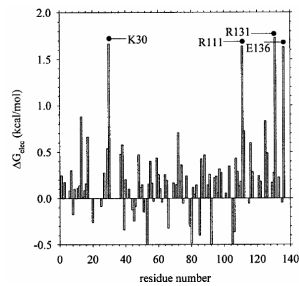
Average residue evolution rate as a function of distance from ligand for 55 proteins (lower evolution rates represent higher conservation)

## Residue Conservation



[Bartlett02]

## Residue Contribution to Stability

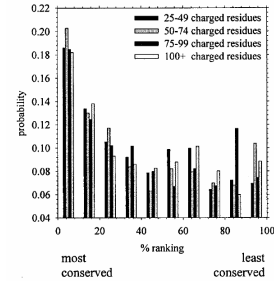


Electrostatic free energies for side-chains of residues in CRBP.

(Positive values indicate residues that destabilize protein)

[Elcock01]

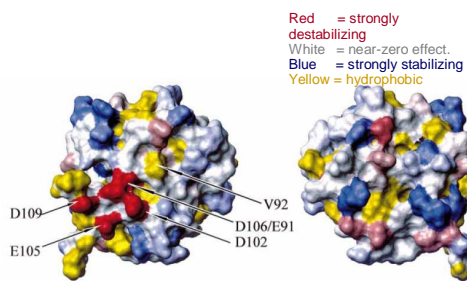
## Residue Contribution to Stability



Histogram showing the distribution of sequence entropy ranks for the top 10% most destabilizing charged residues in proteins of varying sizes.

[Elcock01]

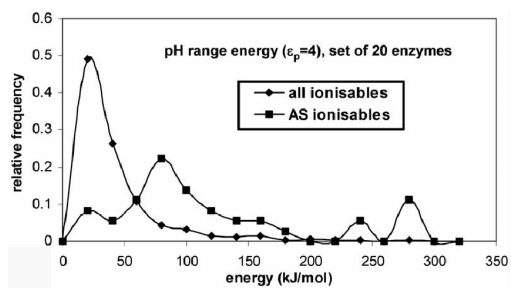
## Residue Contribution to Stability



$\Delta G_{elec}$  values of the residue side-chains for MTH538

[Elcock01]

## Electrostatic Potential



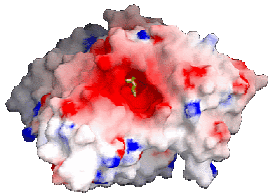
Relative frequencies of pH range energies for all and active site (AS) residues

[Bate04]

## Electrostatic Potential



Negative Positive



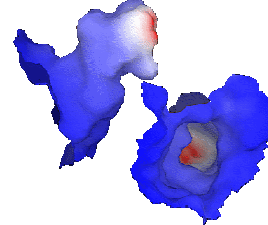
Acetyl choline esterase color coded by electrostatic potential.  
The negative charge in the pocket (red) corresponds to the positive charge on the ligand (acetyl choline)

<http://honiglab.cpmc.columbia.edu/grasp/pictures.html>

## Electrostatic Potential



Negative Positive



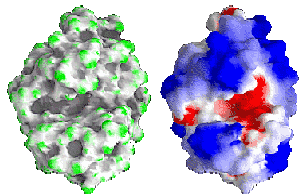
Acetyl choline esterase color coded by electrostatic potential.  
The negative charge in the pocket (red) corresponds to the positive charge on the ligand (acetyl choline)

<http://honiglab.cpmc.columbia.edu/grasp/pictures.html>

## Electrostatic Potential



Negative Positive



Lysozyme

Curvature Electrostatic Potential

<http://honiglab.cpmc.columbia.edu/grasp/pictures.html>

## van der Waals Force Field

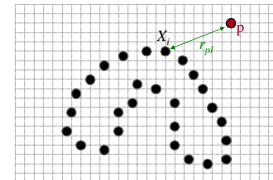


PocketFinder:

1. Create grid potential map of van der Waals force field
2. Apply threshold to keep grid cells with high values
3. Eliminate small pockets (<100 Å)

$$P_p^0 = \sum_{i=1}^N \left( \frac{A_{X_i,C}}{r_{pi}^{12}} - \frac{B_{X_i,C}}{r_{pi}^6} \right)$$

Lenert-Jones  
Potential



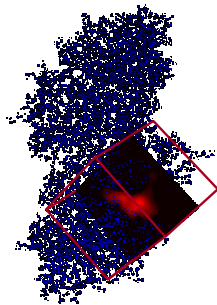
[An04]

## Probe Propensity



General strategy:

- Compute map with propensity for probe(s) of different types



## Probe Propensity

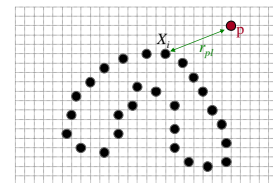


PocketFinder:

- Compute map with van der Waals force field

$$P_p^0 = \sum_{i=1}^N \left( \frac{A_{X_i,C}}{r_{pi}^{12}} - \frac{B_{X_i,C}}{r_{pi}^6} \right)$$

Lenert-Jones  
Potential



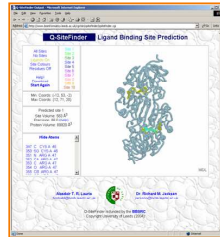
[An04]

## Probe Propensity



Q-SiteFinder:

- Compute map with binding energy of methyl probe



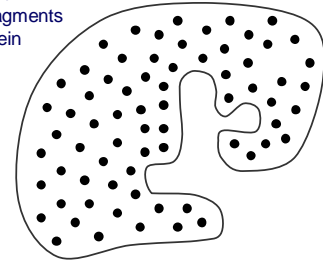
Q-SiteFinder [Laurie05]

## Probe Propensity



CS-Map:

- Compute map with distribution of fragments docked into protein



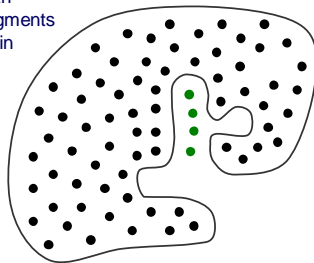
[Silberstein03]

## Probe Propensity



CS-Map:

- Compute map with distribution of fragments docked into protein



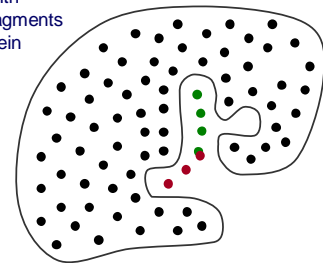
[Silberstein03]

## Probe Propensity



CS-Map:

- Compute map with distribution of fragments docked into protein



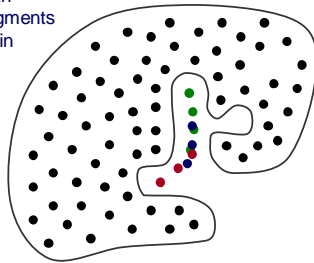
[Silberstein03]

## Probe Propensity



CS-Map:

- Compute map with distribution of fragments docked into protein



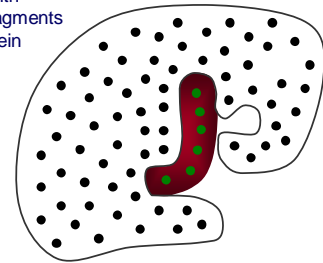
[Silberstein03]

## Probe Propensity



CS-Map:

- Compute map with distribution of fragments docked into protein



[Silberstein03]

## Machine Learning



Build classifier to recognize functional residues

- Depth
- Solvent accessibility
- Propensity
- Conservation
- Hydrophobicity
- Secondary structure type
- Pocket size
- Amino acid

[Gutteridge03]

## Machine Learning



Training set

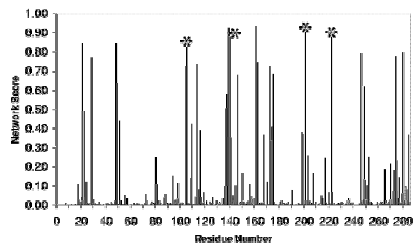
- 159 crystallized proteins
- 55,000 non-catalytic residues, 550 catalytic residues

[Gutteridge03]

## Machine Learning



Neural network classifier



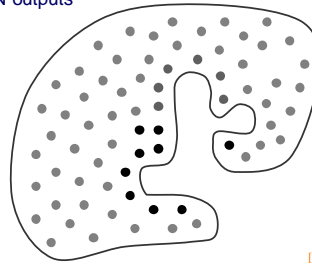
[Gutteridge03]

## Machine Learning



Output

- Spheres around clusters of nearby residues with high NN outputs



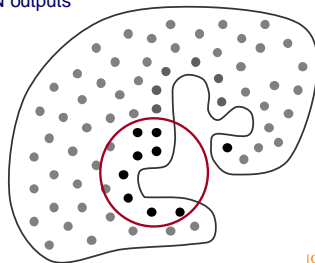
[Gutteridge03]

## Machine Learning



Output

- Spheres around clusters of nearby residues with high NN outputs



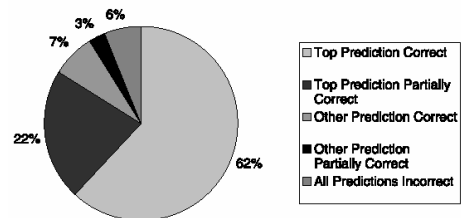
[Gutteridge03]

## Machine Learning



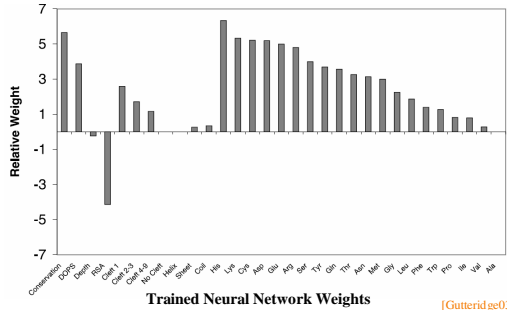
Evaluation

- Sphere overlaps known active site by > 50%



[Gutteridge03]

## Machine Learning



## Outline



Introduction

Binding site localization methods

- Geometric
- Evolutionary
- Chemical

Evaluation methods ←

Discussion

## Q-SiteFinder Evaluation



Test set:

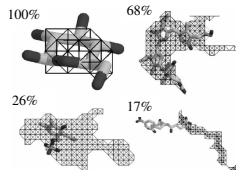
- 134 bound proteins (GOLD test set)
- 35 unbound proteins (homologues to bound proteins)

Metric:

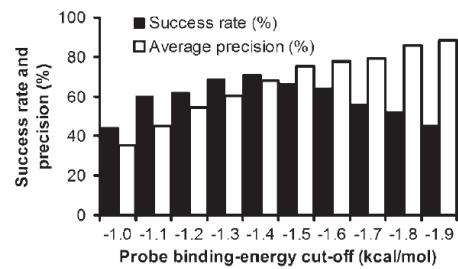
- Precision = % predicted site within 1.6Å of ligand

Success:

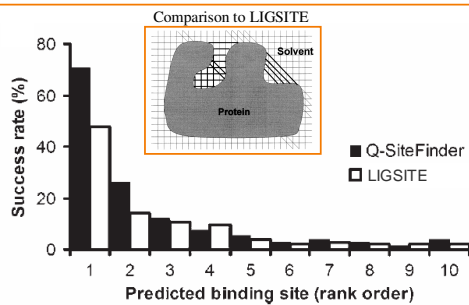
- Precision >25%



## Q-SiteFinder Evaluation

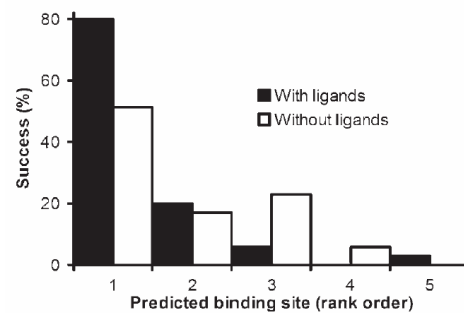


## Q-SiteFinder Evaluation



Q-SiteFinder cutoff = -1.4 kcal/mol, LIGSITE threshold = 5

## Q-SiteFinder Evaluation





## Discussion



?

## References



- [An04] J. An, M. Totrov, R. Abagyan, "Comprehensive Identification of "Druggable" Protein Ligand Binding Sites," *Genome Informatics*, 15, 2, 2004, pp. 31-41.
- [Bartlett02] G.J. Bartlett, C.T. Porter, N.Borkakoti, J.M. Thornton, "Analysis of catalytic residues in enzyme active sites," *J. Mol. Biol.*, 324, 1, 2002, pp. 105-121.
- [Bate04] P. Bate, J. Warwicker, "Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods," *J Mol Biol*, 340, 2, 2004, pp. 263-276.
- [Campbell03] S.J. Campbell, N.D. Gold, R.M. Jackson, D.R. Westhead, "Ligand binding functional site location, similarity and docking," *Curr Opin Struct Biol*, 13, 2003, pp. 389-395.
- [Elcock01] A.H. Elcock, "Prediction of functionally important residues based solely on the computed energetics of protein structure," *J. Mol. Biol.*, 312, 4, 2001, pp. 885-896.
- [Gutteridge03] A. Gutteridge, G.J. Bartlett, J.M. Thornton, "Using a neural network and spatial clustering to predict the location of active sites in enzymes," *J Mol Biol*, 330, 2003, pp. 719-734.
- [Laurie05] A.T.R. Laurie, R.M. Jackson, "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites," *Bioinformatics*, 2005.
- [Nimrod05] G. Nimrod, F. Glaser, D. Steinberg, N. Ben-Tal, T. Pupko, "In silico identification of functional regions in proteins," *Bioinformatics*, 21 Suppl., 2005, pp. 1328-1337.
- [Silberstein03] Michael Silberstein, Sheldon Dennis, Lawrence Brown III, Tamas Kovelyesi, Karl Chodfelter, Sandor Vajda, "Identification of Substrate Binding Sites in Enzymes by Computational Solvent Mapping," *J. Mol. Biol.*, 332, 2003, pp. 1095-1113.
- [Young94] L. Young, R.L. Jernigan, D.G. Covell, "A role for surface hydrophobicity in protein-protein recognition," *Protein Sci*, 3, 5, 1994, pp. 717-29.