Analysis of Protein Binding Sites

Thomas Funkhouser

Princeton University

CS597A, Fall 2007

Protein Binding Site Properties

Residue properties:
- Amino acid type
- Surface accessibility
- Conservation
- Charge
- Hydrophobicity
- Secondary structure type
- Flexibility / Destabilization

Surface/volume properties:
- Cavity size
- Cavity depth
- Cavity shape
- Surface curvature
- Electrostatic potential

Others

> Which properties are favored in binding sites?

Protein Binding Site Types

Site types:
- Protein-ligand
- Protein-protein
- Protein-DNA
- etc.

Protein Binding Site Types

Site types:
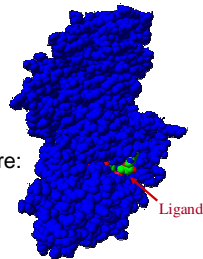- Ø Protein-ligand
- Protein-protein
- Protein-DNA
- etc.

Protein-Ligand Site Data

Databases derived from PDB:
- PDBLIG [Chalk04]
- Ligand Depot [Feng04]
- PLD [Puvanendrampillai03]
- MSDsite [Golovin05]
- Relibase [Hendlich98]
- etc.

Databases derived from literature:
- Catalytic Site Atlas [Porter04]

Ligand

1hld

Protein-Ligand Site Analysis

Example study: [Bartlett et al., 2002]

Data set:
- X-ray structures from PDB
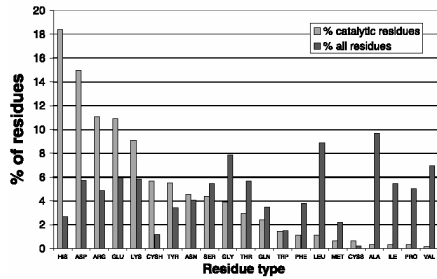- 178 non-homologous proteins
- Catalytic residues

Residue properties:
- Amino acid type
- Secondary structure
- Solvent accessibility
- Flexibility
- Conservation
- etc.

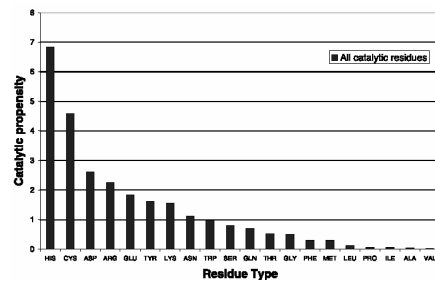## Protein-Ligand Site Analysis

Amino acid type

[Bartlett02]

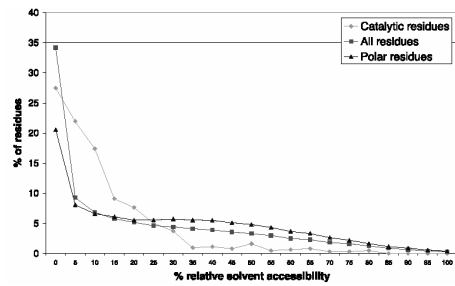## Protein-Ligand Site Analysis

Amino acid type

[Bartlett02]

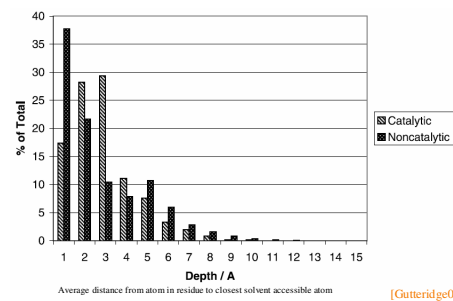## Protein-Ligand Site Analysis

Solvent accessibility

[Bartlett02]
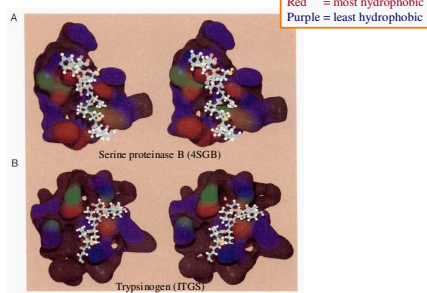
Protein-Ligand Site Analysis

Depth from surface

Average distance from atom in residue to closest solvent accessible atom

[Gutteridge03]

Protein-Ligand Site Analysis

Hydrophobicity

Red = most hydrophobic
Purple = least hydrophobic

Serine proteinase B (4SGB)

Trypsinogen (ITGS)

[Young94]

Protein-Ligand Site Analysis

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]

Protein-Ligand Site Analysis

Secondary structure type

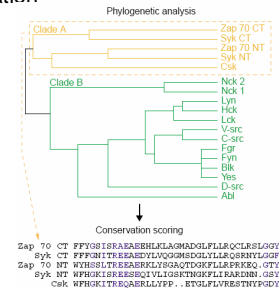|	Catalytic Residues	All Residues
Alpha helix	28%	47%
Beta sheet	22%	23%
Coil	50%	30%

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

[Bartlett02]

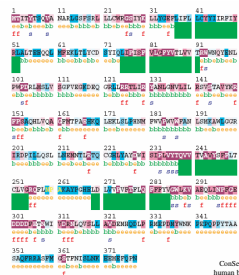## Protein-Ligand Site Analysis

Conservation

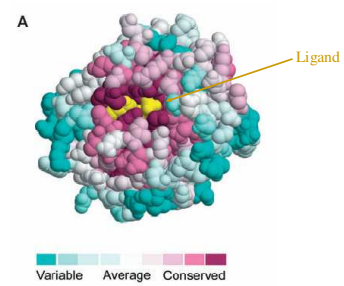[Campbell03]

Protein-Ligand Site Analysis

Conservation

ConSeq predictions demonstrated on
human bestrophin using 43 homologues
obtained from the Pfam database
(SWISS-PROT: VMD2_HUMAN)
(family code: DUF289)
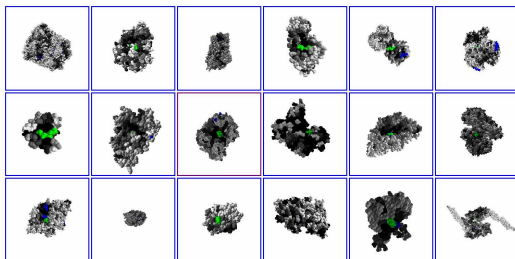
[Berezin04]

Protein-Ligand Site Analysis

Conservation

[Nimrod05]

Protein-Ligand Site Analysis
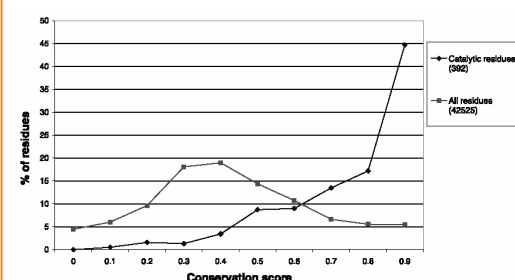
Conservation

Less Conserved   More Conserved
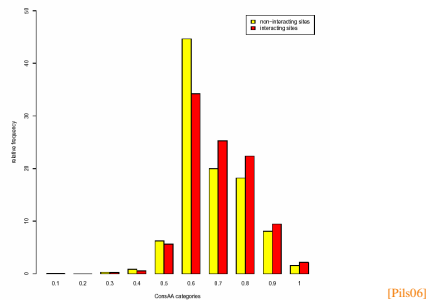
Protein-Ligand Site Analysis

Conservation

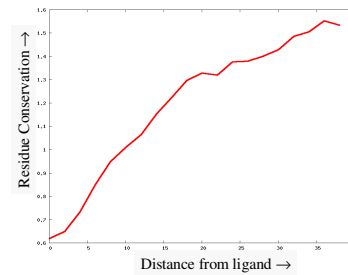[Bartlett02]

Protein-Ligand Site Analysis

Conservation

[Pils06]

Protein-Ligand Site Analysis
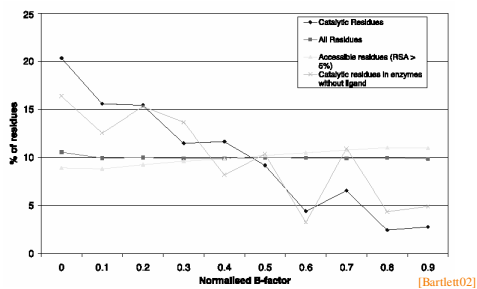
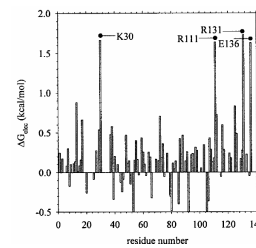Conservation

Protein-Ligand Site Analysis

Flexibility

[Bartlett02]

Protein-Ligand Site Analysis
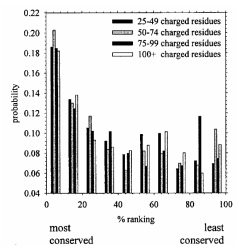
Contribution to stability

Electrostatic free energies for side-chains of residues in CRABP.
(Positive values indicate residues that destabilize protein) [Elcock01]

Protein-Ligand Site Analysis

Contribution to stability

Histogram showing the distribution of sequence entropy ranks for the top
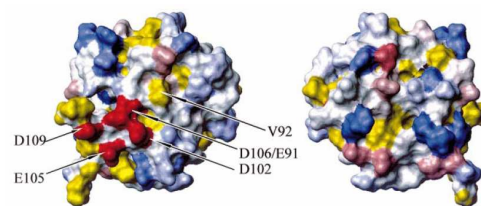10% most destabilizing charged residues in proteins of varying sizes.
[Elcock01]

Protein-Ligand Site Analysis

Contribution to stability

Red   = strongly destabilizing
White = near-zero effect.
Blue   = strongly stabilizing
Yellow = hydrophobic

ΔG_{elec} values of the residue side-chains for MTH538
[Elcock01]

Protein-Ligand Site Analysis

Residue properties:
- Amino acid type
- Surface accessibility
- Conservation
- Charge
- Hydrophobicity
- Secondary structure type
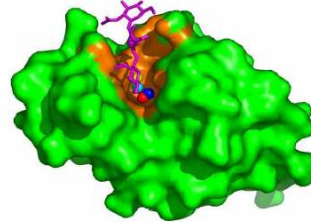- Flexibility / Destabilization

Surface/volume properites:
- Cavity size
- Cavity depth
- Cavity shape
- Surface curvature
- Electrostatic potential

Others

Protein-Ligand Site Analysis

Active sites are usually found in surface cavities

[Huang06]

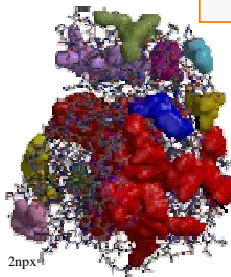## Protein-Ligand Site Analysis

Cavity size

Ligand found in largest cleft in ~80% of proteins

2npx

[Laskowski96]

Protein-Ligand Site Analysis

Cavity volume

Relation Between Cavity Volume and Drug Volume

— All surface cavities
• Drug-binding cavity
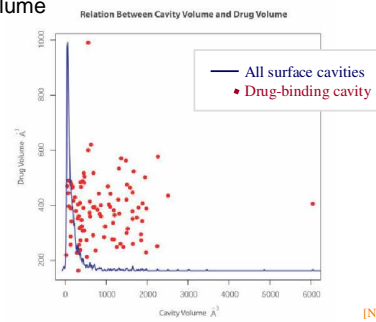
[Nayal06]

Protein-Ligand Site Analysis

Cavity volume

[Liang98]

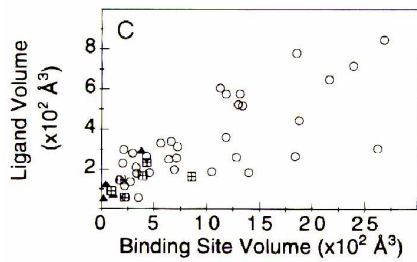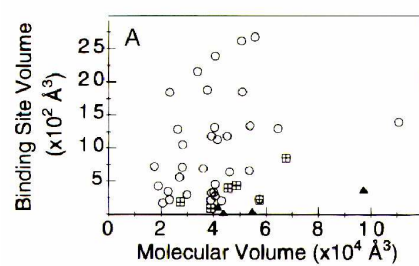
Protein-Ligand Site Analysis

Cavity volume
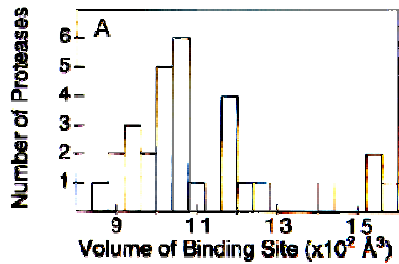

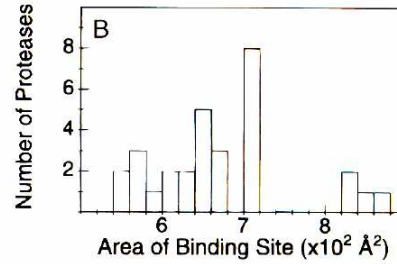[Liang98]

Protein-Ligand Site Analysis

Cavity volume

[Liang98]

Protein-Ligand Site Analysis

Cavity surface area

[Liang98]

Protein-Ligand Site Analysis

Cavity surface curvature

Ligand

http://honiglab.cpmc.columbia.edu/grasp/pictures.html
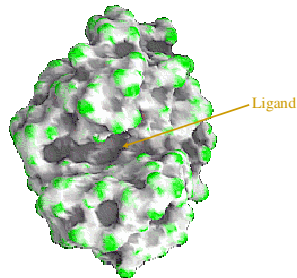
Protein-Ligand Site Analysis

Number of cavity openings
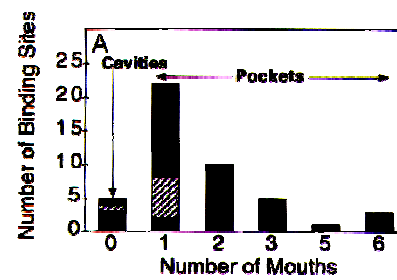
[Liang98]

Protein-Ligand Site Analysis
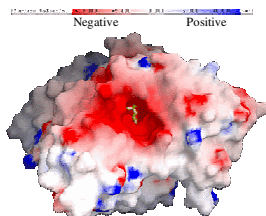
Electrostatic potential

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

Protein-Ligand Site Analysis

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

Protein-Ligand Site Analysis

Electrostatic potential

Negative | Positive

Lysozyme | Curvature | Electrostatic Potential

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

Protein-Ligand Site Analysis

Electrostatic potential:

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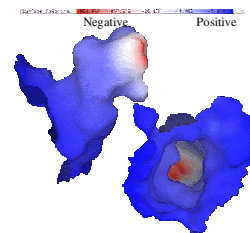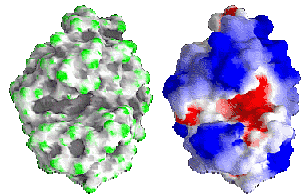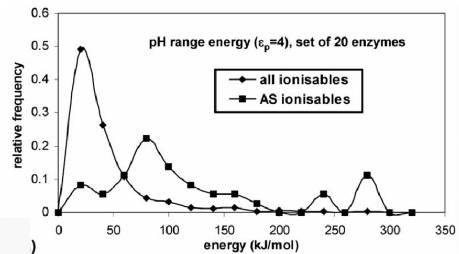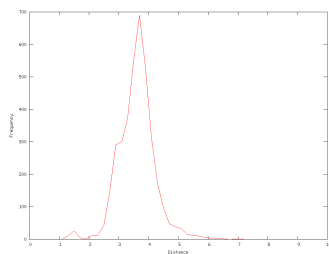all and active site (AS) residues

[Bate04]

Protein-Ligand Site Analysis

Distance from protein surface

Distance from ligand atom
to closest protein atom

Protein-Ligand Site Summary

Distributions of properties:

Surface cavity property	Category	Drug-binding cavities	Non drug-binding cavities
Cavity rank[a]	Size	1.89 ± 2.07	8.88 ± 5.4
Number of residues[b]	Size	22.8 ± 14.3	7.31 ± 5.4
Number of atoms[c]	Size	85.0 ± 62.4	18.7 ± 21.2
Smallest moment of inertia[d]	Size/shape	$1.7 \times 10^4 \pm 2.5 \times 10^4$	$1.2 \times 10^3 \pm 8.3 \times 10^3$
Depth standard deviation[e]	Size/shape	2.3 ± 1.1 (Å[3])	0.75 ± 0.45
Maximum depth[f]	Size/shape	10.5 ± 4.0 (Å)	4.75 ± 1.67
Average depth[g]	Size/shape	5.3 ± 1.9 (Å)	3.2 ± 0.7
Normalized smallest moment of inertia[h]	Shape	17.0 ± 11.7	3.9 ± 5.3
Proportion of cavity at depth between [6.5, 6.75][i]	Shape	0.02 ± 0.013	0.003 ± 0.0001
Largest moment of inertia[j]	Size/shape	$1.6 \times 10^4 \pm 8.4 \times 10^4$	$2.8 \times 10^3 \pm 1.6 \times 10^4$
Average side-chains residual entropy[k]	Rigidity	-0.41 ± 0.18 (kcal)	-0.55 ± 0.25
Average curvature[l]	Shape	-49.0 ± 8.3	-57.0 ± 13.1
Maximum curvedness[m]	Shape	6.4 ± 2.9	4.0 ± 4.9
Maximum mean curvature[n]	Shape	5.3 ± 2.6	3.5 ± 4.2
Curvedness < 0.5[o]	Shape	0.35 ± 0.04	0.29 ± 0.08
Proportion of proline[p]	Amino acid composition	0.019 ± 0.028	0.04 ± 0.09
Proportion of cavity with logP between $[-1, 0]$[q]	Hydrophobicity	0.09 ± 0.07	0.15 ± 0.16
Side-chain residual entropy standard deviation[r]	Rigidity	0.43 ± 0.18 (kcal)	0.55 ± 0.17

[Nayal06]

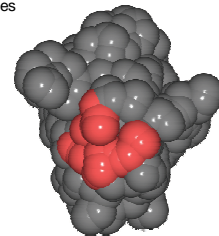## Protein Binding Site Types

Site types:
- Protein-ligand
- ØProtein-protein
- Protein-DNA
- etc.

Protein-Protein Site Analysis
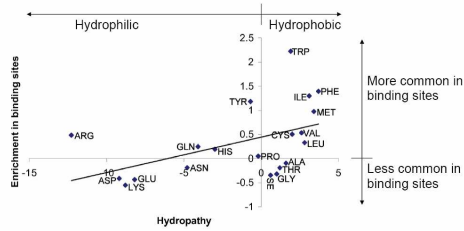
Example study: [Boas & Altman, 2000]
- 5.5×10^5 solvent accessible atoms in 4,800 chains
  - § 1.2×105 are in protein-protein binding sites
  - § 4.3×105 are non-binding sites
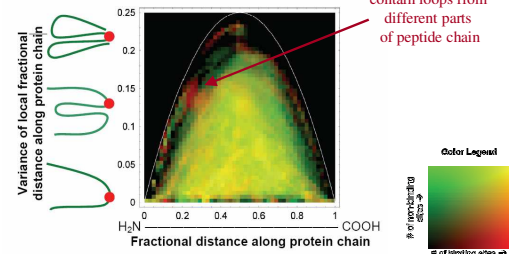
Protein-Protein Site Analysis

Hydrophobicity

Hydrophobic residues are slightly more common in binding sites

[Boas00]

Protein-Protein Site Analysis

Primary structure proximity

Binding sites often contain loops from different parts of peptide chain

[Boas00]

Protein-Protein Site Analysis

Secondary structure

Antiparallel β sheet

Parallel β sheet

Right-handed γ helix

Left-handed α helix

Right-handed α helix
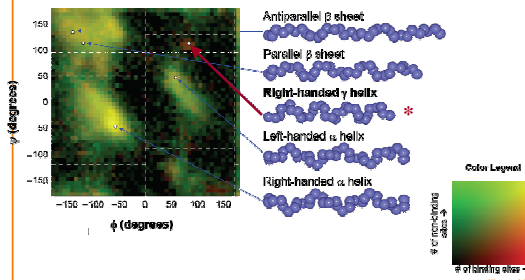
[Boas00]

Protein-Protein Site Analysis

Surface curvature

Concavities are less common in binding sites

Saddle surfaces are more common in binding sites

[Boas00]

Protein-Protein Site Analysis

Electrostatic potential

Binding sites often have large potentials
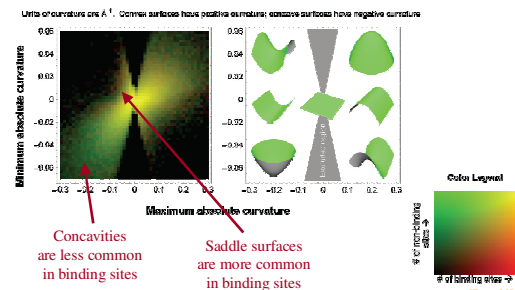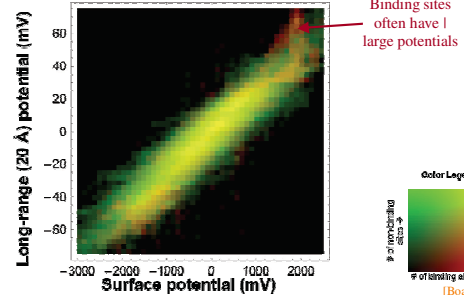
[Boas00]

Protein Binding Site Types

Site types:
- Protein-ligand
- Protein-protein
Ø Protein-DNA
- etc.

Protein-DNA Site Analysis

Example study: [Jones et al., 2003]

Data set:
- 427 protein-DNA complexes

Properties:
- Accessible surface area
- Electrostatics
- Amino acid type
- Hydrophobicity
- Conservation

[Jones03]

Protein-DNA Site Analysis

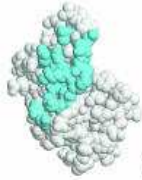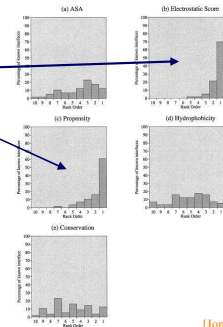Most distinctive properties for DNA binding sites:
- Electrostatics
- Amino acid type

1mjo

[Jones03]

Summary

Residue properties:
- Amino acid type
- Surface accessibility
- Conservation
- Charge
- Hydrophobicity
- Secondary structure type
- Flexibility / Destabilization

Surface/volume properties:
- Cavity size
- Cavity depth
- Cavity shape
- Surface curvature
- Electrostatic potential

Different properties are favored for different type of binding sites

Others

Discussion

?

References

[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.

[Boas00] F.E. Boas and R. Altman, "Predicting protein binding sites", 2000, http://www.stanford.edu/~boas/science/predicted_binding_sites/binding_site.pdf

[Liang98] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, S. Subramaniam, "Analytical shape computing of macromolecules I: molecular area and volume through alpha shape,"
Proteins, 33, 1998, pp. 1-17.

[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.

[Jones03] Susan Jones, Hugh P. Shanahan, Helen M. Berman, and Janet M. Thornton, "Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins," Nucleic Acids Res. 2003 December 15; 31(24): 7189–7198.

[Nayal06] M. Nayal, B. Honig, "On the nature of cavities on protein surfaces: Application to the identification of drug-binding sites," Proteins: Structure, Function, and Bioinformatics, 63, 4, 2006, pp. 892-906.

[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. i328-i337.

[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.