



# Detection of Protein Binding Sites I

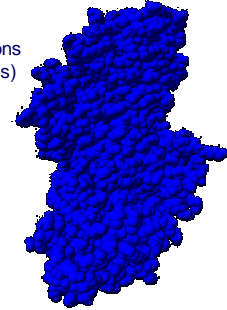
Thomas Funkhouser  
Princeton University  
CS597A, Fall 2005



## Introduction

Motivation:

- 3D structures are available for proteins whose interactions with small molecules (ligands) are not known



thld



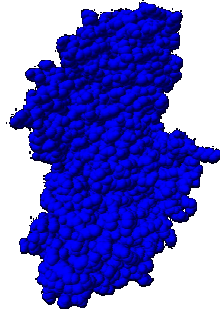
## Introduction

Goal:

- Given a protein structure, predict its ligand bindings

Applications:

- Function prediction
- Drug discovery
- etc.



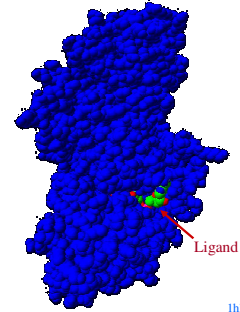
thld



## Introduction

Questions:

- Where will the ligand bind?
- Which ligand will bind?
- How will the ligand bind?
- When?
- Why?
- etc.



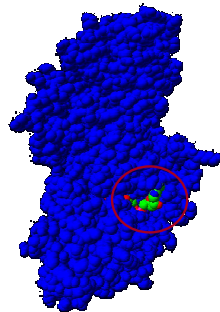
thld



## Introduction

Questions:

- $\emptyset$ Where will the ligand bind?
- Which ligand will bind?
- How will the ligand bind?
- When?
- Why?
- etc.



thld



## Outline

Introduction

Binding site representations

Binding site localization methods

- Geometric
- Evolutionary
- Chemical

Evaluation methods

Discussion

## Outline



Introduction

**Binding site representations** ←

Binding site localization methods

- Geometric
- Evolutionary
- Chemical

Evaluation methods

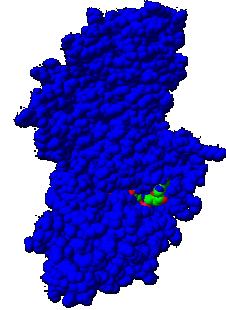
Discussion

## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume



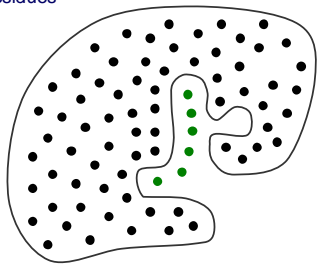
lhld

## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

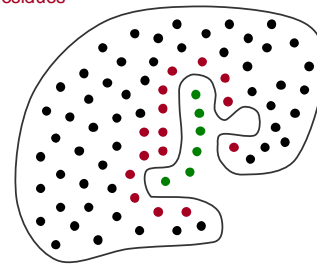


## Binding Site Representations



Possible descriptions:

- ~~Set of atoms/residues~~
- Surface
- Volume

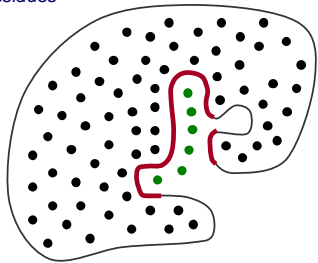


## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- ~~Surface~~
- Volume

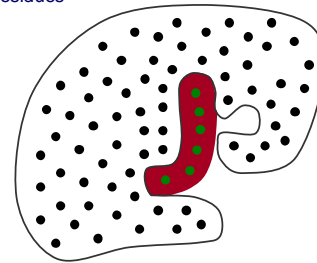


## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- ~~Volume~~

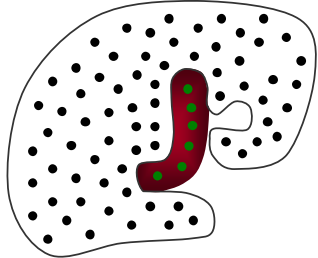


## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

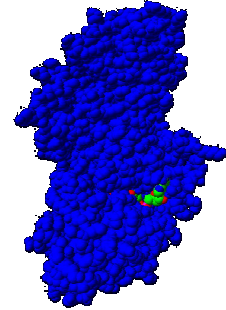


## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume



1hhd

## Binding Site Representations

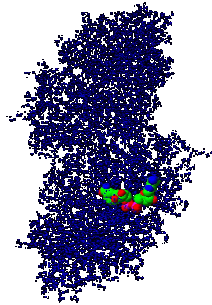


Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

Sequence

```
STAGKYKCKAAVLWEKKPSEIEVEVAPPKAEYRKMVATGICRSD
HYVSGILVTEPLFVAGHRAAGVYSHSGEYTFVPGDAVELETPQCKK
RVCKRFEGNFCLENDLSMPRTGMQDTSRFTCKGRPHFLGITSFSQYT
VYDEIVAKDDASRERKVELGGQRTYVGSAAVAVYQSTCAVWGL
GGVGLSYMGCKAAGAARIHGVINDKRFARAKAEVGTATCVNPDYKRF
QVLTETMNGGVDSFVIGRLDITMYALSCQAEAVGVSVVDPFDSGN
LSMNPMLLSGRVWGLRHFGRKSKDVPKLVADPMAKRFALDFLETVL
PFERNEGFDLRSGRSRTILTF
```



1hhd

## Binding Site Representations

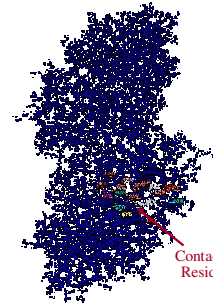


Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

Sequence

```
STAGKYKCKAAVLWEKKPSEIEVEVAPPKAEYRKMVATGICRSD
HYVSGILVTEPLFVAGHRAAGVYSHSGEYTFVPGDAVELETPQCKK
RVCKRFEGNFCLENDLSMPRTGMQDTSRFTCKGRPHFLGITSFSQYT
VYDEIVAKDDASRERKVELGGQRTYVGSAAVAVYQSTCAVWGL
GGVGLSYMGCKAAGAARIHGVINDKRFARAKAEVGTATCVNPDYKRF
QVLTETMNGGVDSFVIGRLDITMYALSCQAEAVGVSVVDPFDSGN
LSMNPMLLSGRVWGLRHFGRKSKDVPKLVADPMAKRFALDFLETVL
PFERNEGFDLRSGRSRTILTF
```



Contacting Residues

1hhd

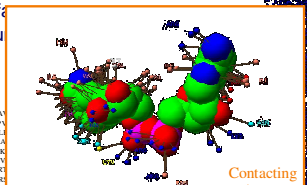
## Binding Site Representations



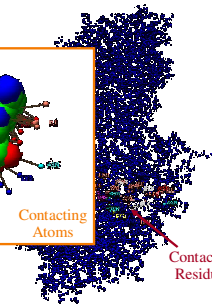
Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

```
STAGKYKCKAAV
HYVSGILVTEPLF
RVCKRFEGNFCLE
VYDEIVAKDDAA
GGVGLSYMGCKK
QVLTETMNGGV
LSMNPMLLSGR
PFERNEGFDLRS
```



Contacting Atoms



Contacting Residues

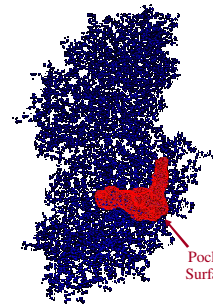
1hhd

## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume



Pocket Surface

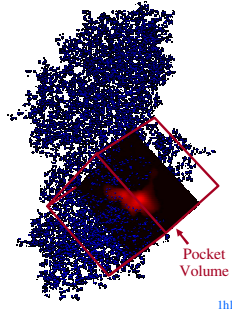
1hhd

## Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume



ihd

## Outline



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

## Outline

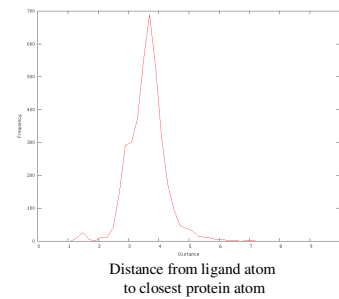


- Introduction
- Binding site representations
- Binding site localization methods
  - Geometric ← This Lecture
  - Evolutionary
  - Chemical
- Evaluation methods
- Discussion

## Binding Site Analysis



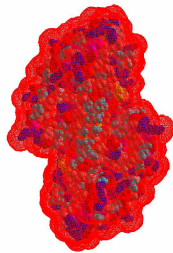
Analyze bound ligands to guide predictions



## Binding Site Analysis



Analyze bound ligands to guide predictions



Distance from ligand atom  
to closest protein atom < 5.5Å

## Outline



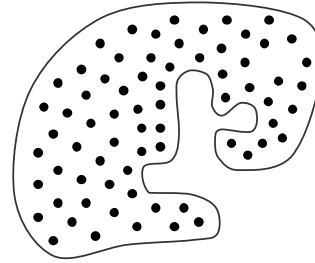
- Introduction
- Binding site representations
- Volumetric binding site localization methods
  - Surfnet
  - LIGSITE
  - PASS
  - CASTp/APROPOS
  - PocketFinder
- Evaluation methods
- Discussion

## Outline



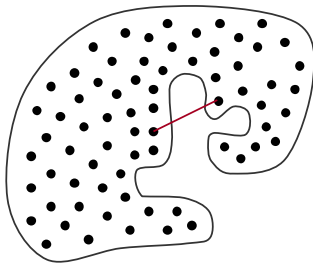
- Introduction
- Binding site representations
- Volumetric binding site localization methods
  - ØSurfnet
  - LIGSITE
  - PASS
  - CASTp/APROPOS
  - PocketFinder
- Evaluation methods
- Discussion

## Surfnet



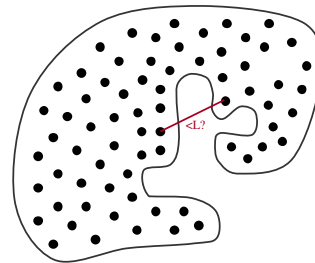
<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

## Surfnet



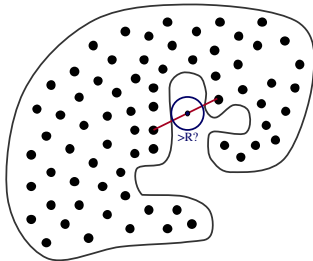
<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

## Surfnet



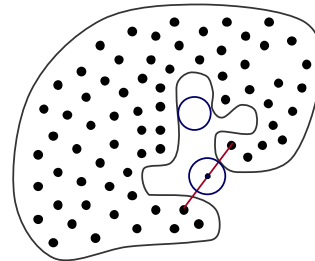
<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

## Surfnet



<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

## Surfnet



<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

### Surfnet

<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

### Surfnet

<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

### Surfnet

Gap-regions (purple) in the surface (yellow) of aspartic protease endothiapsin (3er5)

<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

### Surfnet

Gap-regions (purple) in the surface (gray)

<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

### Surfnet

[Laskowski96]

### Surfnet

Entity	ASU	Asymmetry	Residues	Average	Residue Type
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1

1hld

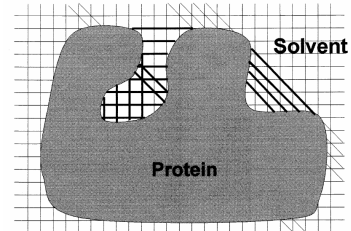
PDBsum

## Outline



- Introduction
- Binding site representations
- Volumetric binding site localization methods
  - Surfnets
  - **LIGSITE**
  - PASS
  - CASTp/APROPOS
  - PocketFinder
- Evaluation methods
- Discussion

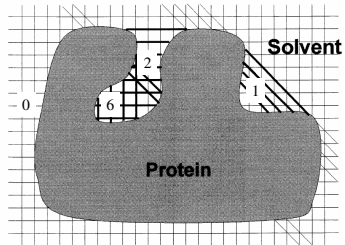
## LIGSITE



Traverse X, Y, Z + 4 cubic diagonal vectors over grid, adding 1 to each grid point lying in region between protein atoms. Scores range from 0 (completely open) to 7 (tightly buried or cavity)

[Hendlich97]

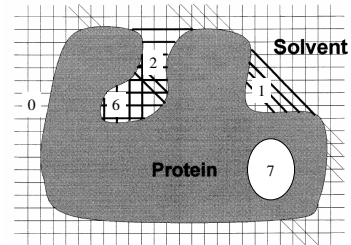
## LIGSITE



Traverse X, Y, Z + 4 cubic diagonal vectors over grid, adding 1 to each grid point lying in region between protein atoms. Scores range from 0 (completely open) to 7 (tightly buried or cavity)

[Hendlich97]

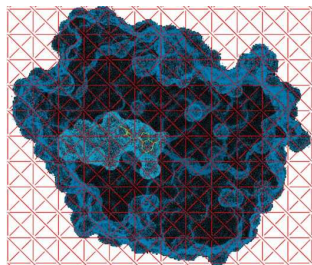
## LIGSITE



Traverse X, Y, Z + 4 cubic diagonal vectors over grid, adding 1 to each grid point lying in region between protein atoms. Scores range from 0 (completely open) to 7 (tightly buried or cavity)

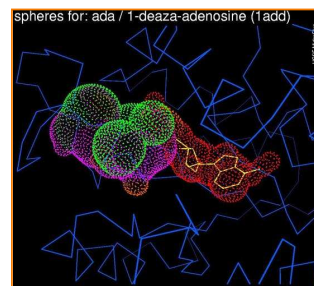
[Hendlich97]

## LIGSITE



Metaphorics

## LIGSITE



Spheres colored by LIGSITE score (red > orange > magenta > green > blue)

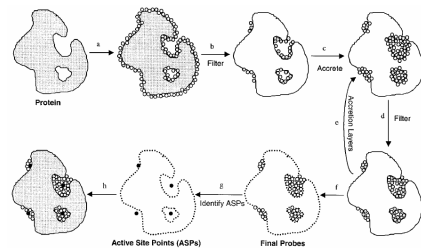
Metaphorics

## Outline



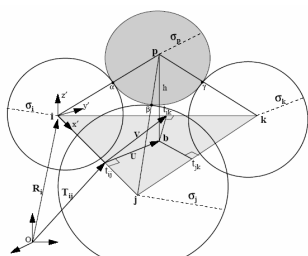
- Introduction
- Binding site representations
- Volumetric binding site localization methods
  - Surfnets
  - LIGSITE
  - PASS
  - CASTp/APROPOS
  - PocketFinder
- Evaluation methods
- Discussion

## PASS



[Brady00]

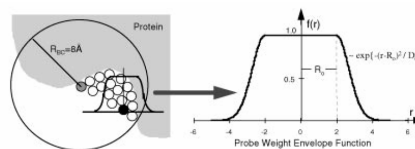
## PASS



Sphere Placement

[Brady00]

## PASS



Sphere Weighting  
(Blur)

[Brady00]

## Outline

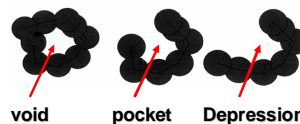


- Introduction
- Binding site representations
- Volumetric binding site localization methods
  - Surfnets
  - LIGSITE
  - PASS
  - CASTp/APROPOS
  - PocketFinder
- Evaluation methods
- Discussion

## Pockets and Voids



A void is a “missing” part of the molecule.  
 A pocket is a void that is connected to the outside through a bottleneck.  
 If there is no bottleneck, there is no pocket.



[Liang98]



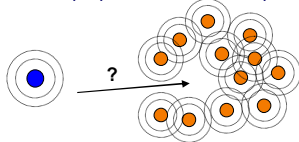
## Revised Problem Description



Given a list of positions and radii of balls in 2D/3D return a list of voids, with volume, position, surface information, etc.

Issues:

- How do we find voids?
- How do we determine voids from pockets?
- How to find the properties of the voids/pockets found?

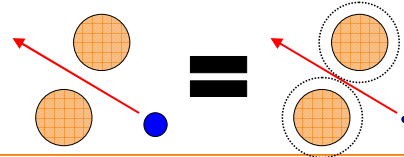


## Invariance to size of balls



To test if a ball  $B$  of radius  $R$  can be placed at position  $P$ , one only needs to check that the distance to every ball  $B_i$  is greater than  $R+R_i$ .

As a result, we can increase the size of  $B_i$  by a constant amount and reduce the size of  $B$  by the same amount without changing the solution.

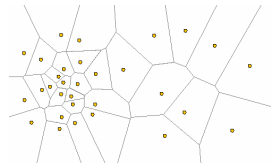


## Voronoi, Delaunay



From MathWorld:

*The partitioning of a plane with  $n$  points into convex polygons such that each polygon contains exactly one generating point and every point in a given polygon is closer to its generating point than to any other.*



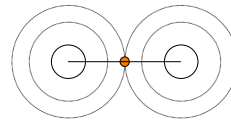
Applet:

<http://cg.cs.mcgill.ca/~godfried/teaching/projects97/delaunay/alpha.html>

## Edge covering.



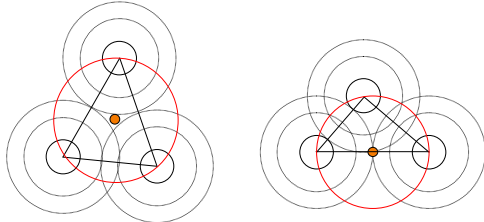
The last point on an edge to be covered is right in the middle.



## Triangle covering.



The last point in a triangle to be filled is the center of the circumscribing circle.



The last point to get filled in an obtuse triangle is along the edge.

## Overview of algorithm



For each edge, triangle, tetrahedron (in 3D):

- Calculate the minimum radius needed to cover.
- Insert into ordered list.

For a given radius (alpha), the triangles that can't be covered are each a void on their own.

If an edge between two voids cannot be covered, those two voids are combined.

If an edge between the outside and a void cannot be covered, the void is a pocket or a depression.

## Pocket or Not I



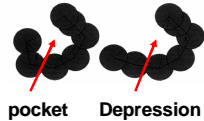
If making the balls larger disconnects a void from the outside before filling it, that void has a mouth.

Pocket:

- A void connected to the outside with a mouth.

Depression:

- A void connected to the outside without a mouth.



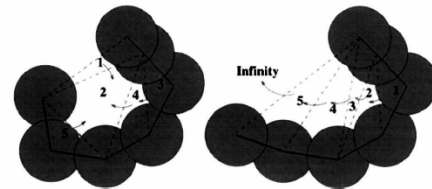
[Liang98]

## Pockets or Not II



A non-obtuse triangle connected to the outside is automatically a pocket.

An obtuse triangle is a pocket if the triangle it is connected to is a pocket.

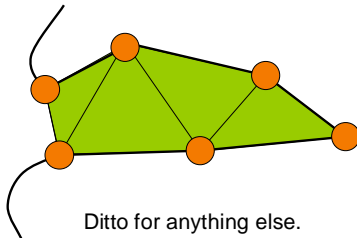


[Liang98]

## Volume, Area, etc.



To measure area of a void simply sum the areas of all the triangles comprising the void, then subtract the area of the sections of the balls within the void.



## Outline



Introduction

Binding site representations

Volumetric binding site localization methods

- Surfnets
- LIGSITE
- PASS
- CASTp/APROPOS
- **∅PocketFinder**

Evaluation methods

Discussion

## PocketFinder

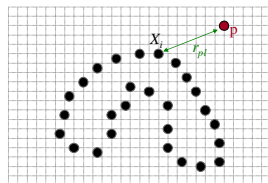


Steps:

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



## PocketFinder

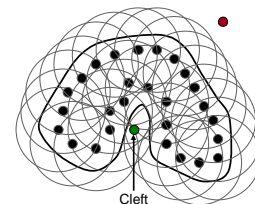


Steps:

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



## PocketFinder

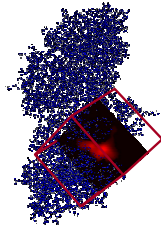


### Steps:

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)

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

Lenert-Jones  
Potential



## Outline



Introduction

Binding site representations

Binding site localization methods

- Geometric
- Evolutionary
- Chemical

Evaluation methods ←

Discussion

## General Evaluation Method



Gather a set of PDB files

- Both bound and unbound (with homologues)

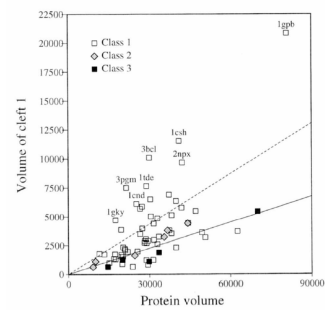
Predict binding sites (clefts, pockets)

- Output is usually grid, polyhedron, set of spheres

Report results

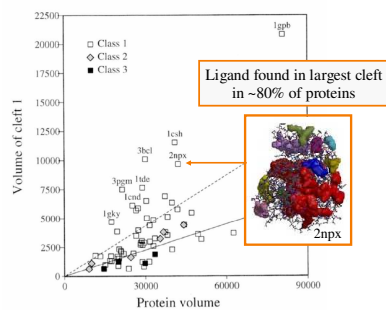
- Measure properties of predicted binding sites
- Test how well predictions match bound ligands

## Surfnet



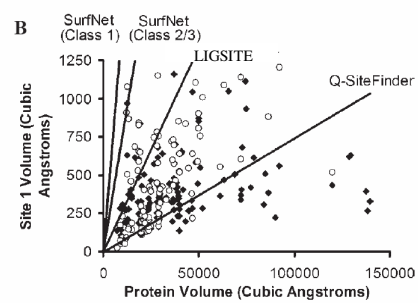
[Laskowski96]

## Surfnet



[Laskowski96]

## LIGSITE



[Laure05]

## CASTp



CAST has been compared to other pocket finding methods such as APROPOS and VOIDOO, and in general manages to find more “correct” binding sites.

Protein	# of cavities	
	VOLBL	Rashin
1eca	10	9
1rx0	3	0
2act	20	21
2cha	23	26
2lyz	12	8
2ptn	19	13
2su3	2	2
3cyt	8	5
3rn3	4	5
4pti	2	2
5nbn	17	23
8ln	42	30

## PocketFinder



### Liganded-pocket data set

- Consider all protein-ligand complexes from PDB
- Eliminate frequent co-factors (HEM, etc.)
- Eliminate ligands far from protein (>3.5Å)
- Eliminate ligands in seams between asymmetric units
- Eliminate “duplicates” (?)
- 50 < protein residues < 2000
- 6 < ligand atoms
- 2.5Å < resolution

5,616 bound binding sites

[An04]

## PocketFinder



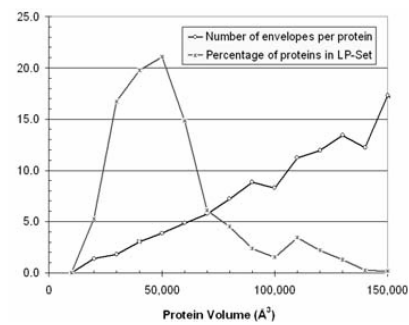
### Unliganded-pocket data set

- Align unliganded PDB files with liganded ones
- Single chain proteins
- 95% < sequence identity
- No mutations on surface within 8Å of ligand
- No other ligands within 8Å of ligand
- 2.5Å < resolution

11,510 unbound binding sites

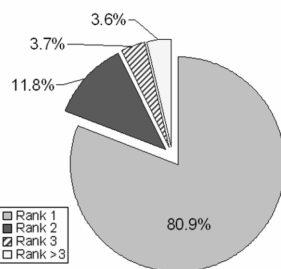
[An04]

## PocketFinder



[An04]

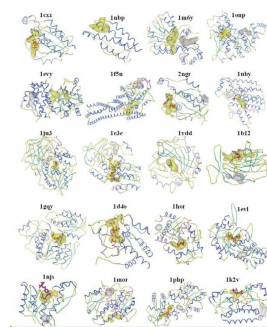
## PocketFinder



Rank of the real binding sites in the predicted putative binding site lists.

[An04]

## PocketFinder



Two largest predicted envelopes (1<sup>st</sup>:yellow, 2<sup>nd</sup>:gray)

[An04]

## PocketFinder

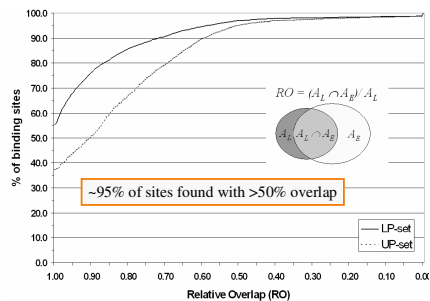


Accuracy measured by overlap of protein atoms in contact with ligand and protein atoms in contact with predicted envelope.

$$RO = (A_L \cap A_E) / A_L$$

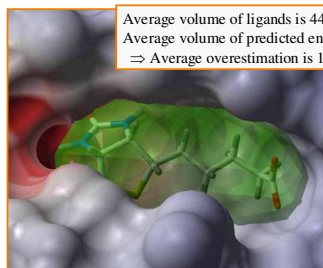
$A_L$  = solvent accessible surface area of protein atoms within 3.5Å of bound ligand  
 $A_E$  = solvent accessible surface area of protein atoms within 3.5Å of predicted envelope

## PocketFinder



[An04]

## PocketFinder

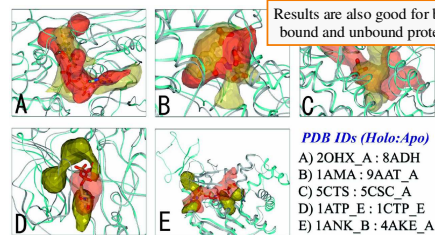


Average volume of ligands is 440Å<sup>3</sup>  
 Average volume of predicted envelopes is 611Å<sup>3</sup>  
 ⇒ Average overestimation is 1.4x (?)

Biotin-streptavidin binding site predicted with PocketFinder

[An04]

## PocketFinder



Effect of conformational changes on predictions for bound (holo) and unbound (apo) proteins  
 (enzymes: gray-holo, green-apo)  
 (envelopes: red-holo, yellow-apo)

[An04]

## 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.
- [Brady00] G.P. Brady, Jr., P.F.W. Stouten, "Fast prediction and visualization of protein binding pockets with PASS," *Journal of Computer-Aided Molecular Design*, 14, 4, 2000, pp. 383-401.
- [Hendlich07] M. Hendlich, F. Rippman, G. Barnickel, "LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins," *J. Mol. Graph.*, 15, 1997, pp. 359-363.
- [Laskowski95] R.A. Laskowski, "Surfnet: a program for visualizing molecular surfaces, cavities, and intermolecular interactions," *J Mol Graph*, 13, 1995, pp. 323-330.
- [Laskowski96a] R.A. Laskowski, N.M. Luscombe, M.B. Swindells, J.M. Thornton, "Protein clefts in molecular recognition and function," *Prot. Sci.*, 5, 12, 1996, pp. 2438-2452.
- [Laurie05] A.T.R. Laurie, R.M. Jackson, "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites," *Bioinformatics*, 2005.
- [Liang98a] J. Liang, H. Edelsbrunner, P. Fu, P.V. Sathhakar, S. Subramaniam, "Analytical shape computing of macromolecules I: molecular area and volume through alpha shape," *Proteins*, 33, 1998, pp. 1-17.
- [Liang98b] J. Liang, H. Edelsbrunner, P. Fu, P.V. Sathhakar, S. Subramaniam, "Analytical shape computing of macromolecules II: identification and computation of inaccessible cavities inside proteins," *Proteins*, 33, 1998, pp. 18-29.
- [Liang98c] J. Liang, H. Edelsbrunner, C. Woodward, "Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design," *Protein Science*, 7, 1998, pp. 1884-1897.
- [Peters96] K.P. Peters, J. Fauck, C. Frommel, "The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria," *J Mol Biol*, 256, 1996, pp. 201-213.