


Detection of Protein Binding Sites I

Thomas Funkhouser
Princeton University
CS597A, Fall 2007



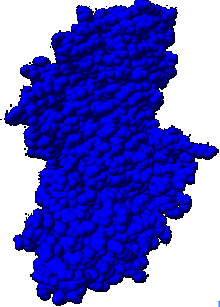
Introduction

Goal:


- Given a protein structure, predict where ligands bind

Applications:

- Function prediction
- Drug discovery
- etc.




lhd




Outline

- Introduction
- Binding site representations
- Binding site prediction
- Evaluation methods
- Discussion



Outline

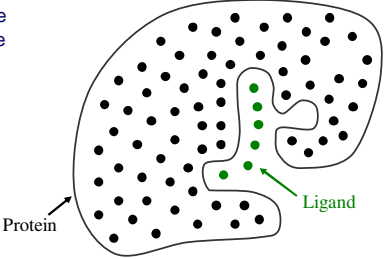
- Introduction
- Binding site representations ←
- Binding site prediction
- Evaluation methods
- Discussion




Binding Site Representations

Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

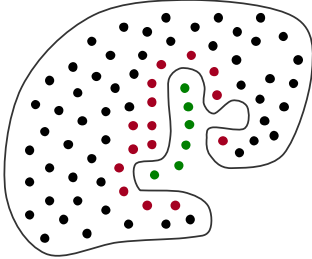




Binding Site Representations

Possible descriptions:

- ∅ Set of atoms/residues
- Surface
- Volume

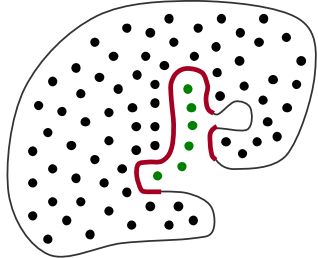


Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

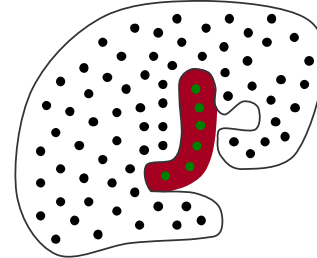


Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume

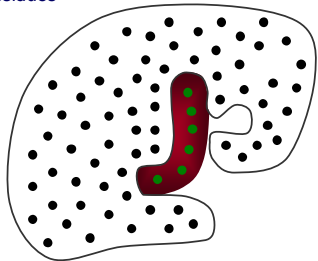


Binding Site Representations



Possible descriptions:

- Set of atoms/residues
- Surface
- Volume



Outline



Introduction

Binding site representations

Binding site prediction ←

Evaluation methods

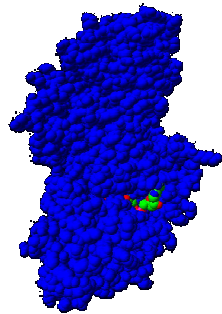
Discussion

Binding Site Prediction



Possible predictions:

- Set of atoms/residues
- Surface
- Volume



thid

Binding Site Prediction

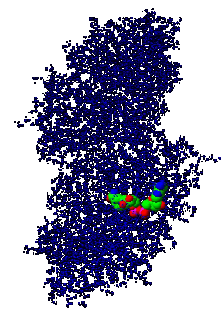


Possible predictions:

- Set of atoms/residues
- Surface
- Volume

Sequence

```
STAGKYKCKAAVLEWEKKPFSIHVEVAPPKADHVEKMFVATGICRSDD  
HVVSGTLVTFPLFVLAGHEAAGVVEISGIEVTVVPPGDKVPLFPQCGKC  
RVECKAFPGNCLKNDLSMPGDTMQGQTSRFTCRGKPHHPLGTTSTGQVT  
VYDERSYAKEDAAAPFLKRYLGGQSTGYSAAKPKAKTQKSTAVLGL  
GGVGLSYMEGCKAAGAARHGVDNKIKDFAKAKVIGATLGVNPDYKRF  
QVLTENSGGQWDFKIKRITKMTALACQQLAVGSSVGVPPPSGN  
LSMNPMLLSGRTWKGAFGGPKSKDSVPLKVADEFMAKFALEDFLTHV  
PFRINSGFDLLSGSHETLTP
```



thid

Binding Site Prediction

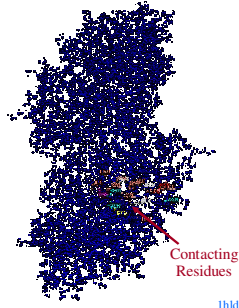


Possible predictions:

- ∅ Set of atoms/residues
- Surface
- Volume

Sequence

```
STAGKVIKCKAAVLWEEKRPFSEEVEVAPPKAHEVRKMMVATGCRSDSD  
HYVSGITLTPFPVLAHREKAGEVSEISGIVTFVPEDEKVFLETPQDCK  
RVCKIFEGNFCLENDLSMPRGTMQGTSTRETCRGRPHFGLSTESQVY  
VVDHESVAKIDAAAPLEKVLGCGFSTGYGSANVKAQVYQSTCAVFGLE  
GGVGLSYMGCAAGAAARHGDHINQDFKACAEVATEVYVQDYKRFY  
QEVLTENNGVDSSEVIGRLDMYVIALSCQELATGVNVMVGVPPISQK  
LSDNMLLESQFTWGHAFYGFKSDISVPELVADPMAKRFALDPLTIVL  
PFRKNEGFDLRSIGSRITLTF
```



ihhd

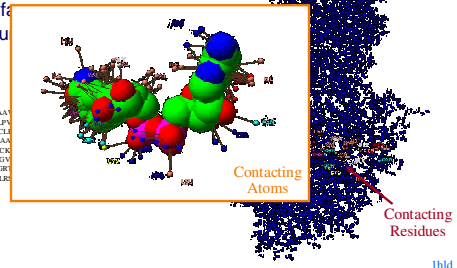
Binding Site Prediction



Possible predictions:

- ∅ Set of atoms/residues
- Surface
- Volume

```
STAGKVIKCKAAVLWEEKRPFSEEVEVAPPKAHEVRKMMVATGCRSDSD  
HYVSGITLTPFPVLAHREKAGEVSEISGIVTFVPEDEKVFLETPQDCK  
RVCKIFEGNFCLENDLSMPRGTMQGTSTRETCRGRPHFGLSTESQVY  
VVDHESVAKIDAAAPLEKVLGCGFSTGYGSANVKAQVYQSTCAVFGLE  
GGVGLSYMGCAAGAAARHGDHINQDFKACAEVATEVYVQDYKRFY  
QEVLTENNGVDSSEVIGRLDMYVIALSCQELATGVNVMVGVPPISQK  
LSDNMLLESQFTWGHAFYGFKSDISVPELVADPMAKRFALDPLTIVL  
PFRKNEGFDLRSIGSRITLTF
```



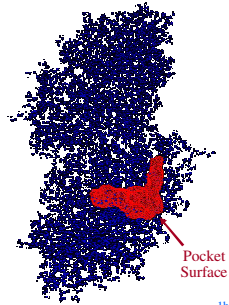
ihhd

Binding Site Prediction



Possible predictions:

- Set of atoms/residues
- ∅ Surface
- Volume



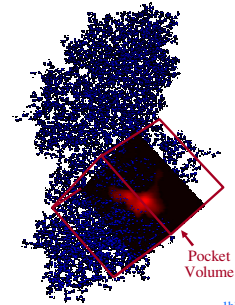
ihhd

Binding Site Prediction



Possible predictions:

- Set of atoms/residues
- Surface
- ∅ Volume



ihhd

Binding Site Prediction Methods



Homology-based methods

- Alignment with known sites
- Conservation

Sequence-based methods

- Motifs

Structure-based methods

- Geometric
- Chemical

Hybrid methods

- Machine learning

Binding Site Prediction Methods



Homology-based methods

- Alignment with known sites
- Conservation

Sequence-based methods

- Motifs

Structure-based methods

- ∅ Geometric ←
- Chemical

Hybrid methods

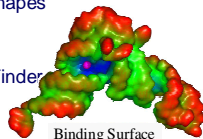
- Machine learning

Geometric Methods

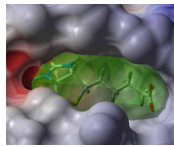


Find residues/surfaces/volumes inside large cavities

- LIGSITE
- PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnets
- PASS
- PocketFinder
- etc.



Binding Surface Prediction
[Coleman06]



Binding Volume Prediction
[Ant04]

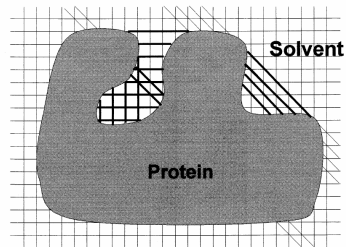
Geometric Methods



Find residues/surfaces/**volumes** inside large cavities

- ~~LIGSITE~~
- PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnets
- PASS
- PocketFinder
- etc.

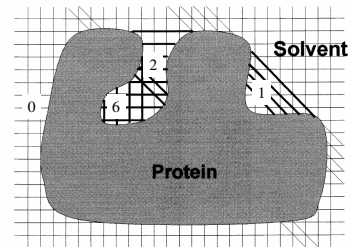
LIGSITE



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]

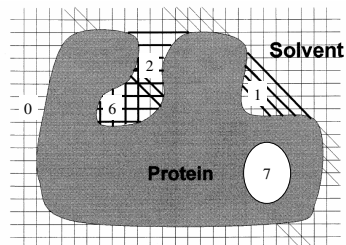
LIGSITE



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]

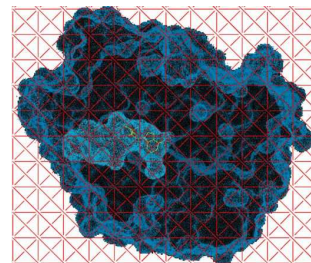
LIGSITE



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)

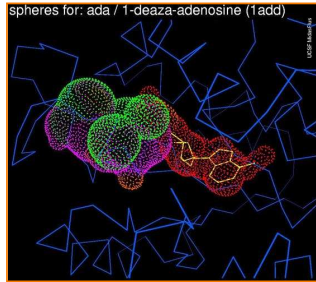
[Hendlich97]

LIGSITE



Metaphorics

LIGSITE



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

Metaphorics

Geometric Methods



Find residues/surfaces/volumes inside large cavities

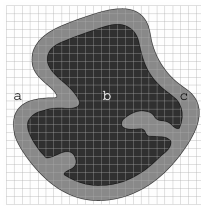
- LIGSITE
- ~~Ø~~ PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnet
- PASS
- PocketFinder
- etc.

PocketPicker



Buriedness index

- At every grid cell within some distance from protein surface



Grid

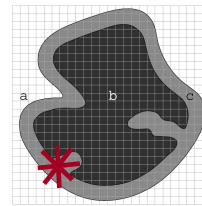
[Weisel07]

PocketPicker

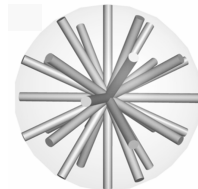


Buriedness index

- At every grid cell within some distance from protein surface, sample 30 directions



Grid



Search
Directions

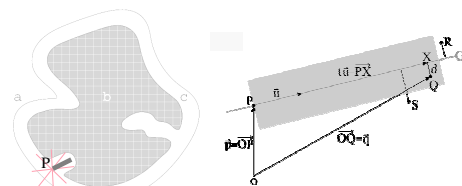
[Weisel07]

PocketPicker



Buriedness index

- At every grid cell within some distance from protein surface, sample 30 directions, and count how many sample directions find a protein atom within radius 10\AA



Grid

[Weisel07]

PocketPicker

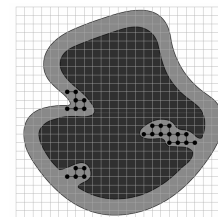


Buriedness index

- At every grid cell within some distance from protein surface, sample 30 directions, and count how many sample directions find a protein atom within radius 10\AA

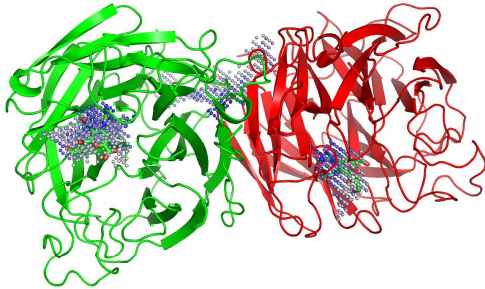
Site prediction

- Cluster grid positions with buriedness indices between 16 and 26



[Weisel07]

PocketPicker



[Weisel07]

Geometric Methods



Find residues/surfaces/volumes inside large cavities

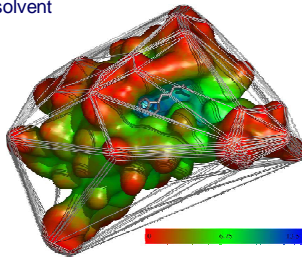
- LIGSITE
- PocketPicker
- \emptyset Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnet
- PASS
- PocketFinder
- etc.

Travel Depth



Definition:

- Distance from point on protein surface to convex hull through solvent

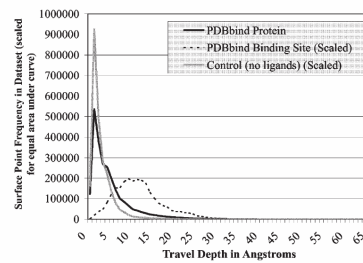


[Coleman06]

Travel Depth



Distributions of travel depths:

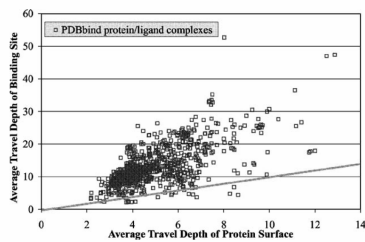


[Coleman06]

Travel Depth



Distributions of travel depths:

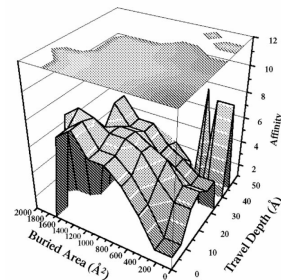


[Coleman06]

Travel Depth



Correlation with buried area:

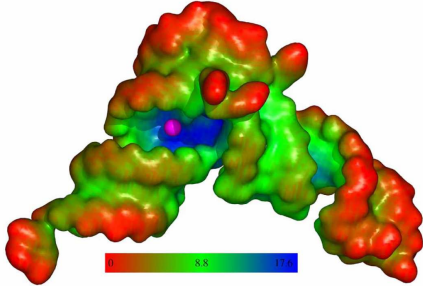


[Coleman06]

Travel Depth



Results:



[Coleman06]

Geometric Methods



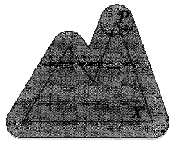
Find residues/surfaces/volumes inside large cavities

- LIGSITE
- PocketPicker
- Travel Depth
- \emptyset Mathematical morphology
- Alpha shapes
- Surfnet
- PASS
- PocketFinder
- etc.

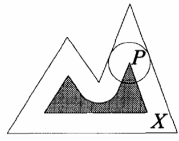
Mathematical Morphology



Basic operators



Dilate



Erode

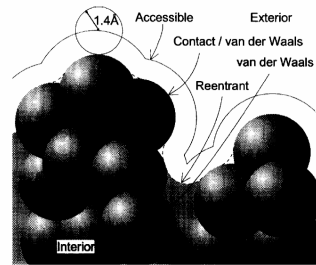


Close

Mathematical Morphology



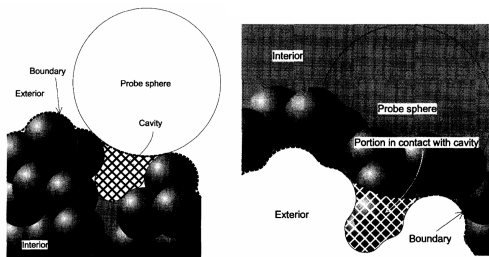
Finding cavities



Mathematical Morphology



Finding cavities



Geometric Methods



Find residues/surfaces/volumes inside large cavities

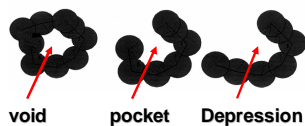
- LIGSITE
- PocketPicker
- Travel Depth
- Mathematical morphology
- \emptyset Alpha shapes
- Surfnet
- PASS
- PocketFinder
- etc.

Alpha Shapes



Aim to provide rigorous method for finding ...

- Void – completely surrounded
- Pocket – connected to outside through bottleneck
- Depression – connected to outside, without bottleneck



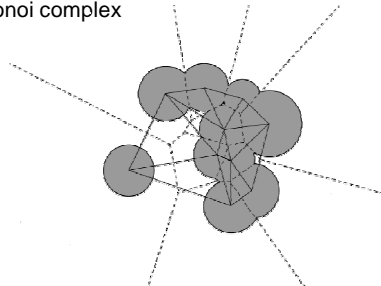
void pocket Depression

[Liang98]

Alpha Shapes



Atom positions define a Delaunay triangulation and Voronoi complex

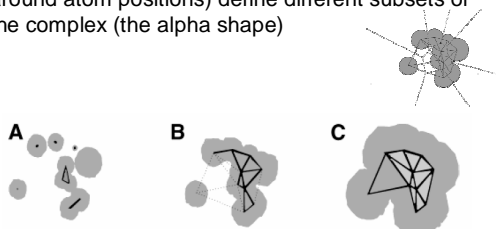


[Liang98]

Alpha Shapes



Different values of alpha (radius of expansion around atom positions) define different subsets of the complex (the alpha shape)

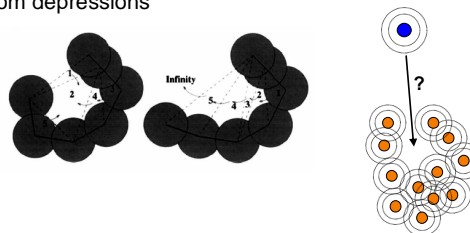


[Liang98]

Alpha Shapes



Analysis of triangles in alpha shape provides way to find sizes of bottlenecks and to distinguish pockets from depressions



[Liang98]

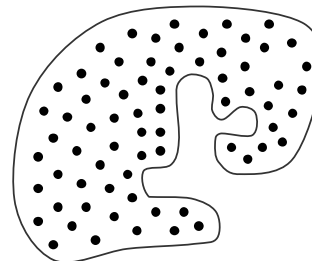
Geometric Methods



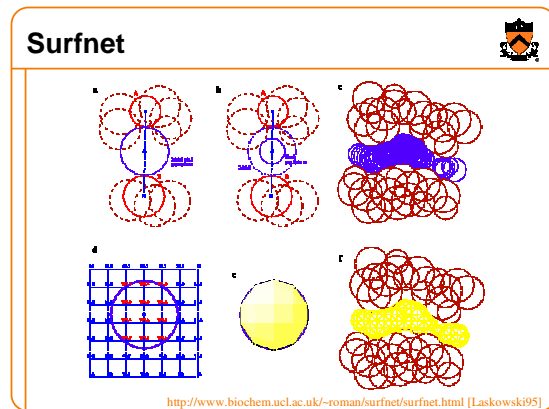
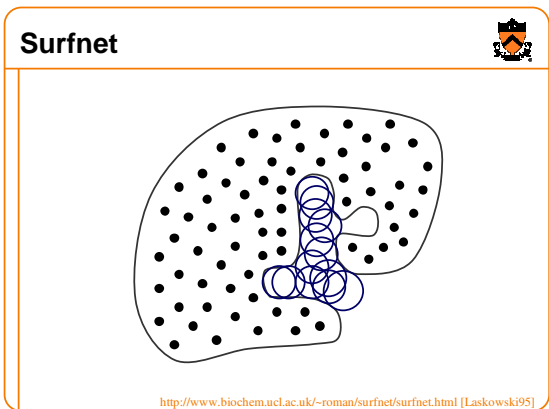
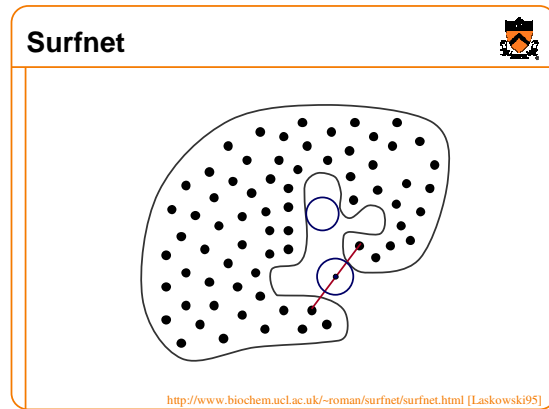
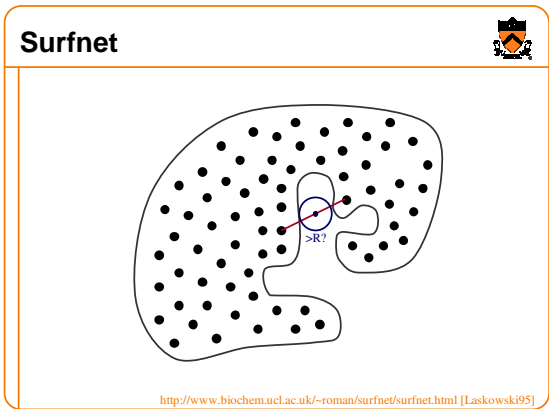
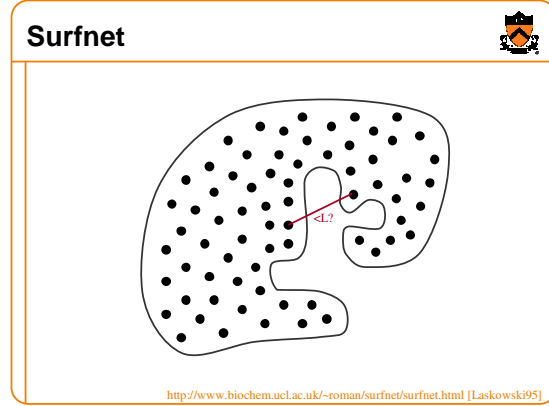
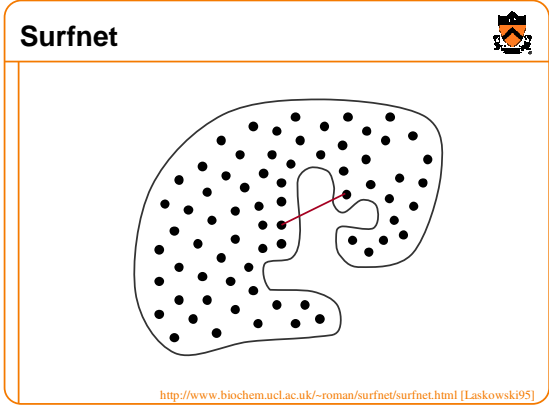
Find residues/surfaces/volumes inside large cavities

- LIGSITE
- PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- ~~Surfnet~~
- PASS
- PocketFinder
- etc.

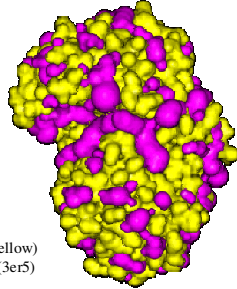
Surfnet



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



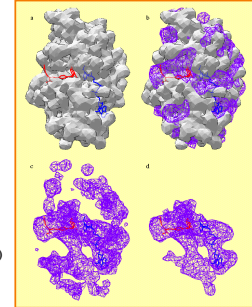
Surfnet



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

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

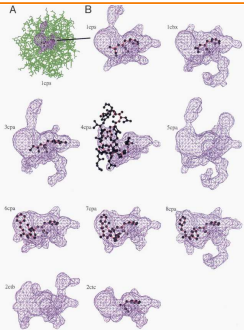
Surfnet



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

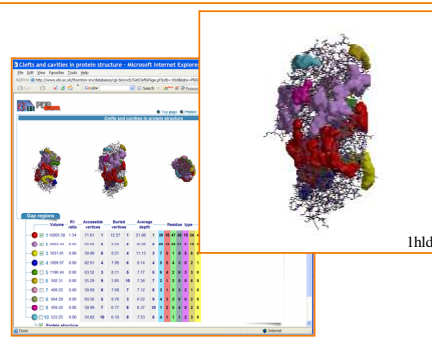
<http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html> [Laskowski95]

Surfnet



[Laskowski96]

Surfnet



PDBsum

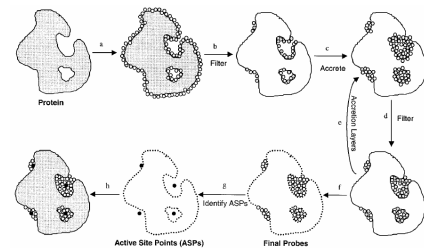
Geometric Methods



Find residues/surfaces/volumes inside large cavities

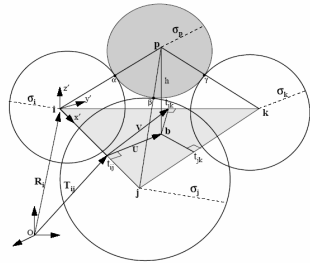
- LIGSITE
- PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnet
- ~~PASS~~
- PocketFinder
- etc.

PASS



[Brady00]

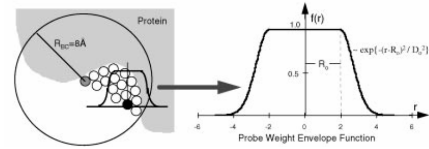
PASS



Sphere Placement

[Brady00]

PASS



Sphere Weighting
(Blur)

[Brady00]

Geometric Methods



Find residues/surfaces/volumes inside large cavities

- LIGSITE
- PocketPicker
- Travel Depth
- Mathematical morphology
- Alpha shapes
- Surfnet
- PASS
- ~~PocketFinder~~
- etc.

PocketFinder

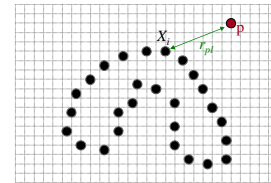


Steps:

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)

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

Lenert-Jones
Potential



PocketFinder

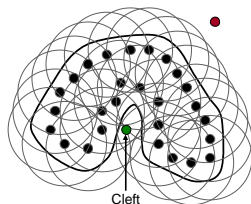


Steps:

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)

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

Lenert-Jones
Potential



PocketFinder

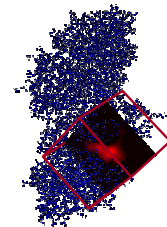


Steps:

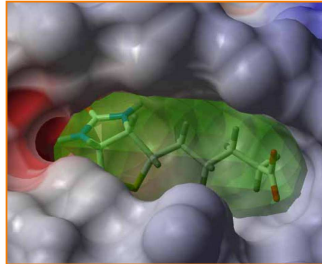
1. Create grid potential map of van der Waals force field
2. Apply threshold to keep grid cells with high values
3. Eliminate small pockets (<100 Å)

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

Lenert-Jones
Potential



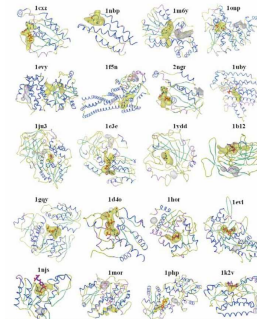
PocketFinder



Biotin-streptavidin binding site predicted with PocketFinder

[An04]

PocketFinder



Two largest predicted envelopes (1st:yellow, 2nd:gray)

[An04]

Outline



Introduction

Binding site representations

Binding site prediction

ØEvaluation methods Next Time!

Discussion

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.
- [Hendlich97] M. Hendlich, F. Rippman, G. Barricelli, "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. Sudhakar, 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. Sudhakar, 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.