Protein-Ligand Complex:

Prediction of Possible Binding Sites

Based on : **"Methods for the Prediction of Protein-Ligand Binding Sites for Structure- Based Drug Design and Virtual Ligand Screening"**, A.T.R. Laurie and R.M. Jackson, Current Protein and Peptide Science, 2006, 7, 395-406

Noël Malod-Dognin

South-West University, Blagoevgrad, Bulgaria.

The 4^{th} of March

Protein ligand complex

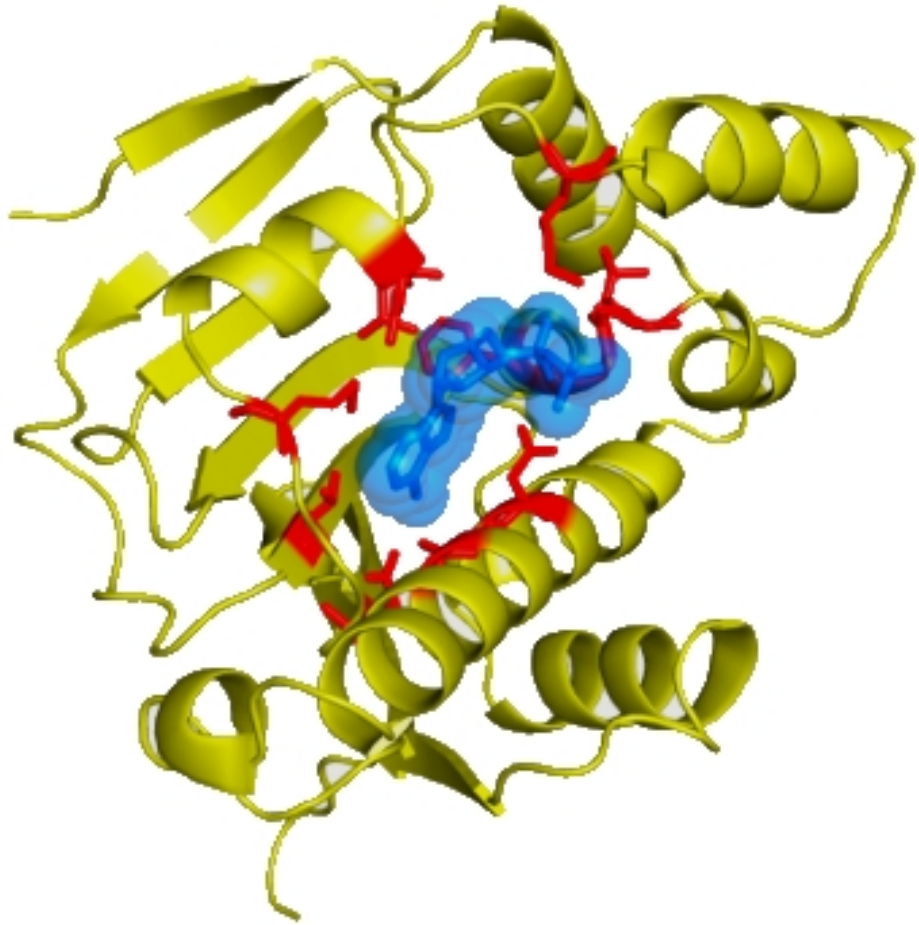


FIGURE: A ligand is bound on a particular part of the protein surface called binding site

Related problems

Protein docking :

To determine the 3D structure of a protein ligand complex starting from the individual structures of its constituents.

- How do the protein and the ligand interact ?

1. **Binding site prediction** :

   To locate the possible binding sites on the protein 3D structure

2. Molecular complementarity :

   To estimate the compatibility between the ligand and the possible binding sites

3. Optimal protein/ligand binding mode :

   To find how the ligand attach to the binding site

Binding site properties (1/2)

- Largest cleft in the protein ($>$ 83% of the cases, Laskowski in 1996)

Binding site location	No. of proteins (as %)
Largest cleft	141 = 67%
2^{nd} largest cleft	28 = 13%
3^{rd} largest cleft	14 = 7%
In none of above	27 = 13%

TABLE: Huang et al., 2006, over 210 proteins from the Protein Ligand Database

- Significantly larger than other clefts ($> 100 Å^3$)
  $\rightarrow$ **Size $=$ Functional requirement**

$\neq$ Protein-Protein binding sites (planar surfaces)

Binding site properties (2/2)

- Energetically unstable regions (Elcock, 2001)
  - Contain charged residues located in unfavorable environments
  - On a set of 216 proteins, Elcock shows a strong relationship between electrostatic energy and conservation
    - Functional residues $\simeq$ most destabilizing ones
- Protein-Protein binding sites tend to contain proline brackets (Kini & Evans, 1995)
- A protein has in average 9 partners (Aloy & Russell, 2004)

Binding site prediction methods

Geometry, cartesian 3D grid

- CAVITY SEARCH (Ho & Marshall, 1990)
- **POCKET** (Levitt & Banaszak, 1992)
- **LIGSITE** (Hendlich et al., 1997)
- **POCKETPICKER** (Weasel et al., 2007)

Geometry, volume

- Delaney (1992)
- PASS (Brady & Stouten, 2000)
- SURFNET (Laskowski, 1995)

Other geometrical approaches

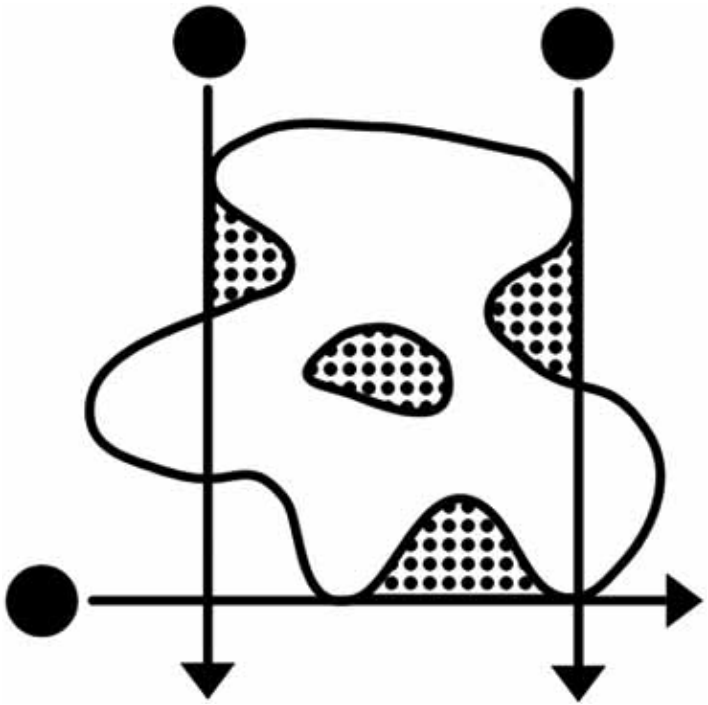
- Del Carpio (1993)
- APROPOS (Peters et al., 1996)

Energy based approaches

- POCKETOME (Totrov & Abagyan, 2005)
- **Q-SITE-FINDER** (Laurie & Jackson, 2005)

Cleft detection, geometry / grid methods

POCKET, Levitt & Banaszak 1992



Principle :

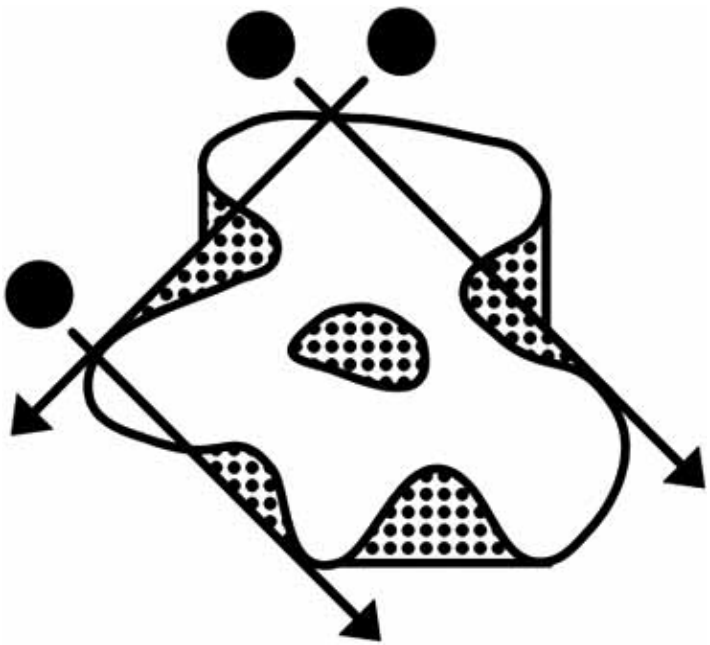
- Probe spheres of 3Å radius are passed along each line of a cartesian 3D grid
  - 3 scan directions
- Interaction = one atom is found inside a sphere
- Pockets are identified by sequences : interaction, no interaction, interaction

Problem :

- Found pockets depend on the cartesian basis used (of the grid orientation)

Cleft detection, geometry / grid methods

LIGSITE, Hendlich, Rippmann & Barnickel, 1997

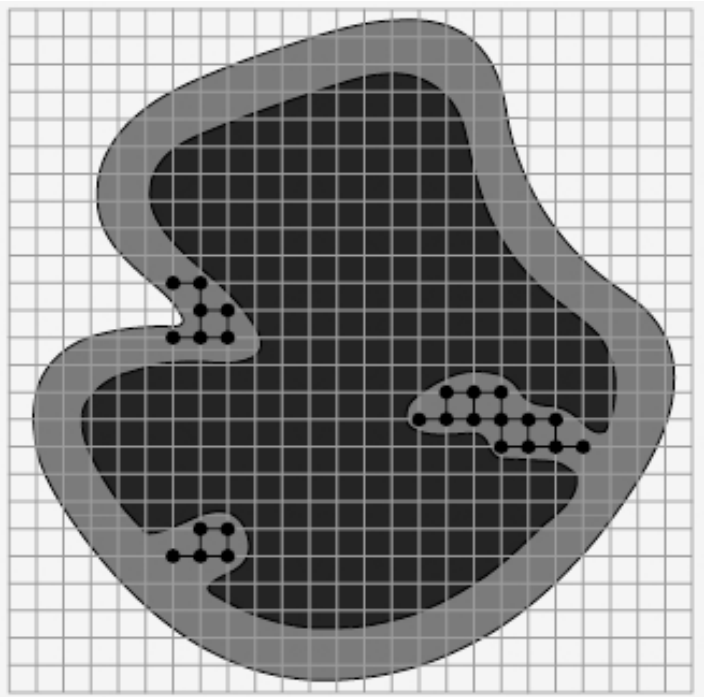


Principle :

- Similar to POCKET, but also scan the cubic diagonals
  - 7 scan directions, less dependant of cartesian basis
- PSP event = each time a point of the grid is inside a pocket (max 7 times)
  - A minimum PSP threshold can be used to filter pockets

Cleft detection, geometry / grid methods

POCKETPICKER, Weasel et al., 2007



Principle :

- Buriedness of a grid point is determined by the protein surface intersecting a 10Å sphere
- Heuristically evaluated

- Buried grid points are clustered

Success rate comparison

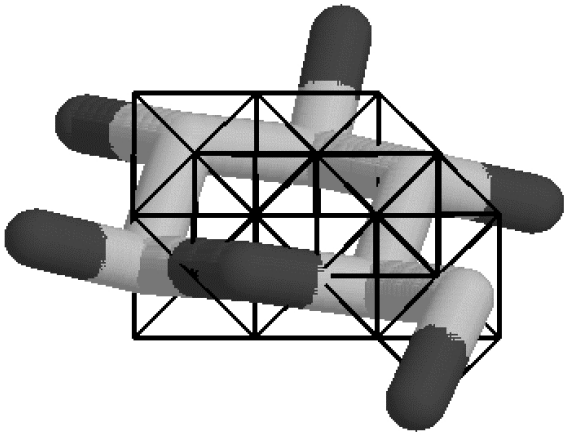
	Top 1		Top 3	
Method	unbound	bound	unbound	bound
PocketPicker	69%	72%	85%	85%
LIGSITE	58%	69%	75%	87%
CAST	58%	67%	75%	83%
PASS	60%	63%	71%	81%
SURFNET	52%	54%	75%	78%

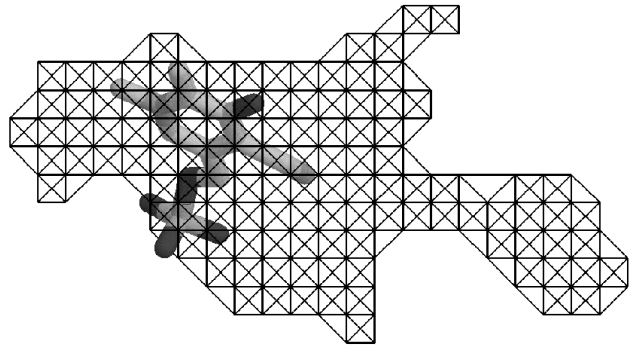
TABLE: Weisel et al., 2007, Success rate of different methods on a set of 48 binding sites

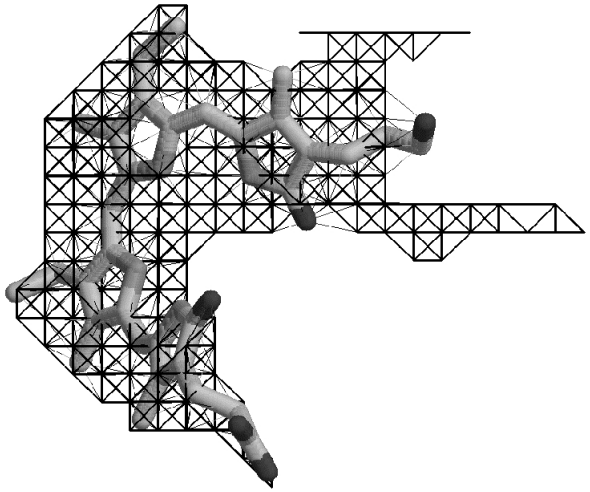
From 69% to 87% of success rate, i.e., when the binding sites are located inside the retrieved pockets

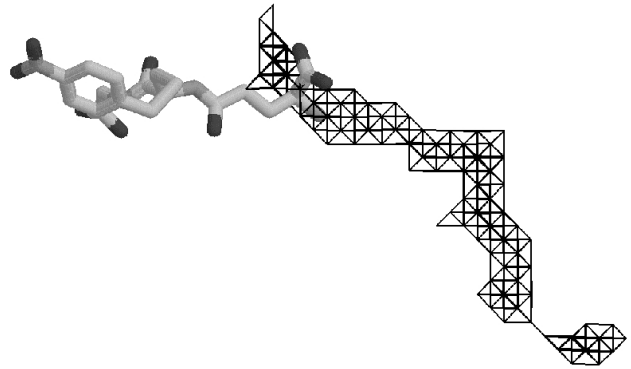
What about precision ?

I.e., The location of the real binding sites vs the predicted ones.



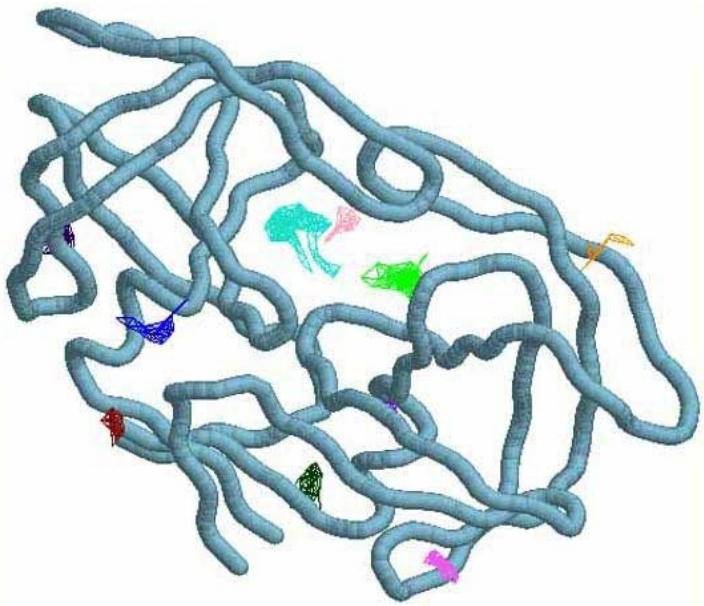






Energy based method

Q-SITE-FINDER, Laurie & Jackson, 2005



Principle :

- 3D cartesian grid
- For each grid point is computed the interaction energy with a methyl probe
- Favourable methyl binding grid points are clustered

Same success rate as LIGSITE, but more precise localization of the binding site (not enought statistic about this...)

Possible research direction

Precise statistics are needed :

- Where are the missing 10% of binding sites ?
- About the precision of the methods

Binding site prediction :

Hybridize methods like :

1. Geometry based methods
   - POCKETPICKER with exact / more precise buriedness computation
2. Energy based methods
3. Structure comparison methods
   - Pocket vs Known binding site

In order to :

$\rightarrow$ Refine binding site location inside a pocket by using energy methods

$\rightarrow$ Obtain a consensus between the different methods