3D Structural Alignment
and
Tunnel Computation

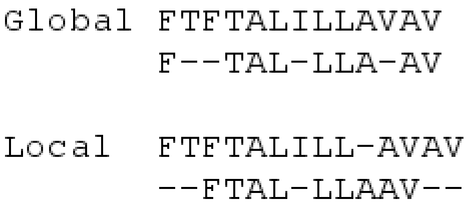
David Sehnal

Contents

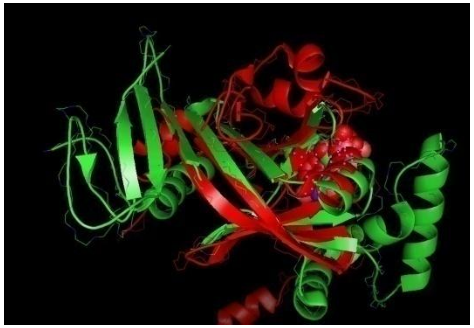
- Two Worlds
  - 3D Structural Alignment (on the small scale)
  - Tunnel Computation
- The Worlds Collide?

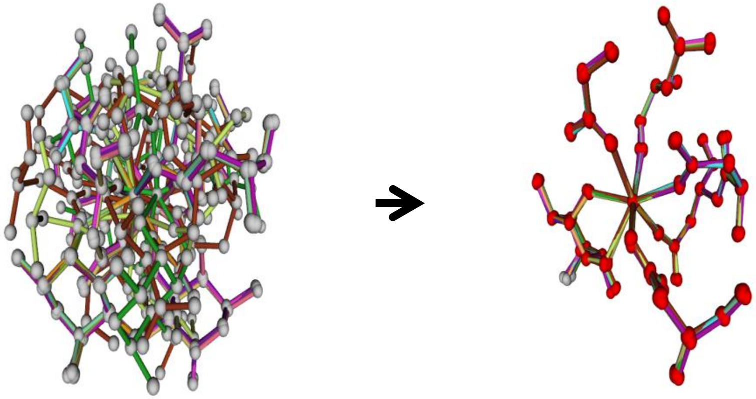
3D Alignment

- **Large scale** (proteins)
  - Global (classification, evolutionary links)
  - Local (compare smaller sub-structures)
- **Small scale**



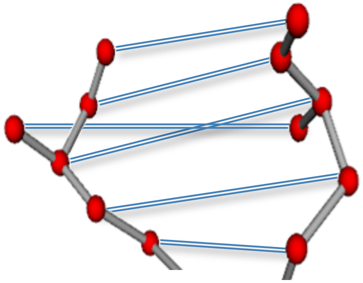
```
Global FTFTALILLAVAV
       F--TAL-LLA-AV

Local  FTFTALILL-AVAV
       --FTAL-LLAAV--
```





3D Alignment

- What we need?
  - Know which atoms/amino acids/… belong together



  - Find a transformation that fits the corresponding pairs

$$\mathrm{RMSD}(A,B) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|A_i - B_i\|^2}$$

The Transformation Part

- Horn K.P.: Closed-form solution of absolute orientation using unit quaternions, Journal of the Optical Society of America, 1986.
- Karney C.F.F.: *Quaternions in molecular modeling*. Journal of Molecular Graphics and Modelling, 2007.

$$\mathrm{RMSD}(T(A),B) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|R(A_i)+t-B_i\|^2}$$

Translation component

Rotation component

$$\sum_{i=1}^{n}\|R(A_i)-B_i\|^2 = \overset{\text{Stays the same}}{\boxed{\sum_{i=1}^{n}\|R(A_i)\|^2+\sum_{i=1}^{n}\|B_i\|^2}} - 2\overset{\text{Maximize this}}{\boxed{\sum_{i=1}^{n}R(A_i)\cdot B_i}}$$

$$\boxed{q^T\left(\sum_{i=1}^{n}\hat{\mathbf{A}}_i^T\mathbf{B}_i\right)q}$$

$$R(x) = qxq^*$$

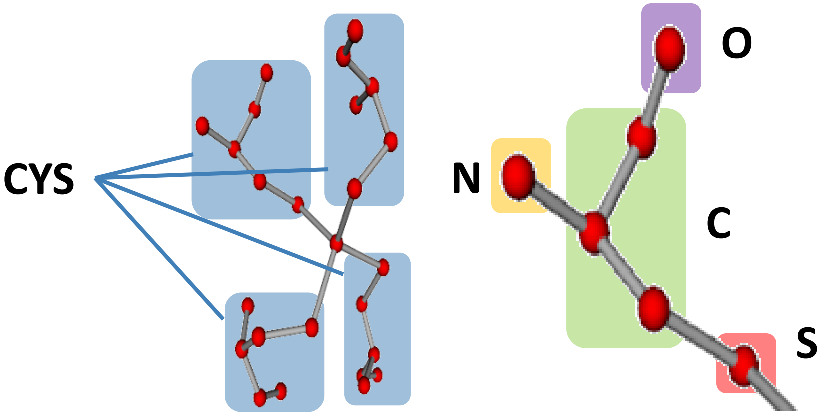
Pairing the Elements

- Large Scale
  - Usually work on the secondary structure (amino acids)
  - Combinatorial extensions, subgraph isomorphism, dynamic programming, …
- **Small Scale**
  - Usually work on 3D structure
  - Exhaustive search, pairing based on distance, subgraph isomorphism, …

(My) Simple Algorithm

1. For each bijection f: Atoms(A) -> Atoms(B)
   - Align sequences f(A) and B using the quaternion method
2. Return the alignment that yields the lowest RMSD

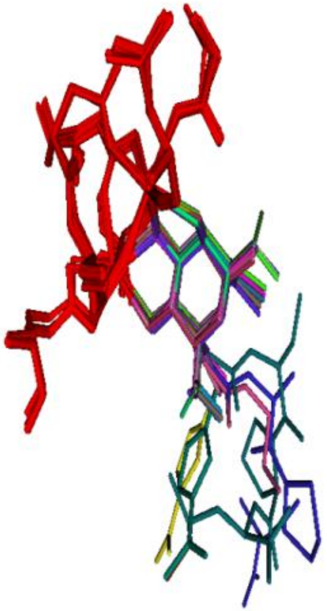
- Improvement for small protein substructures:
  - Grouping of atoms



Applications

Reverse Proteins

Classification



Lectines

66 structures

56 atom on each

Apoptosis

Analysis of the

model structure

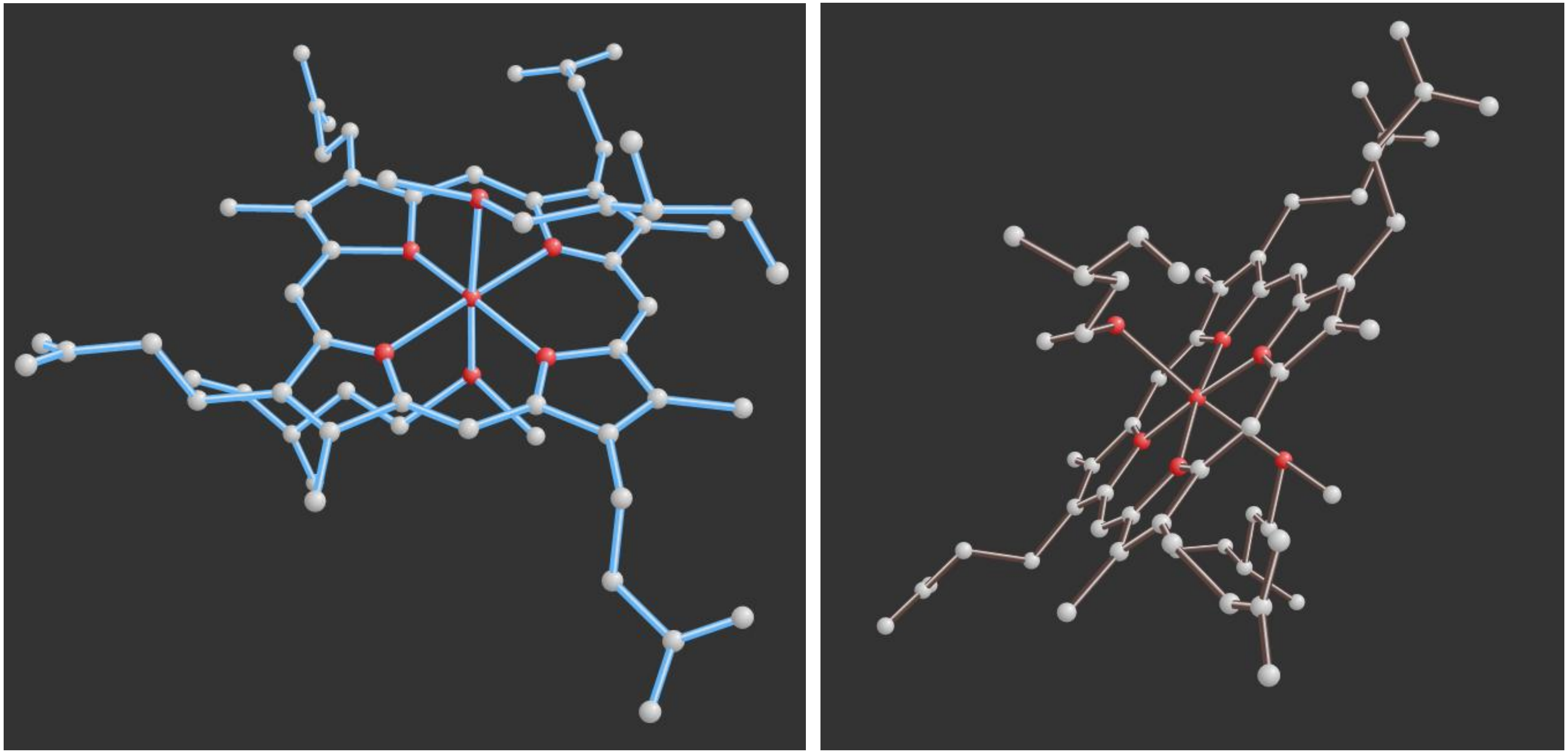
(Sort of Mine) Better Approach

- Hoppe A., Frommel C.: *NeedleHaystack: a program for the rapid recognition of local structures in large sets of atomic coordinates,* Journal of applied crystallography, 2003.

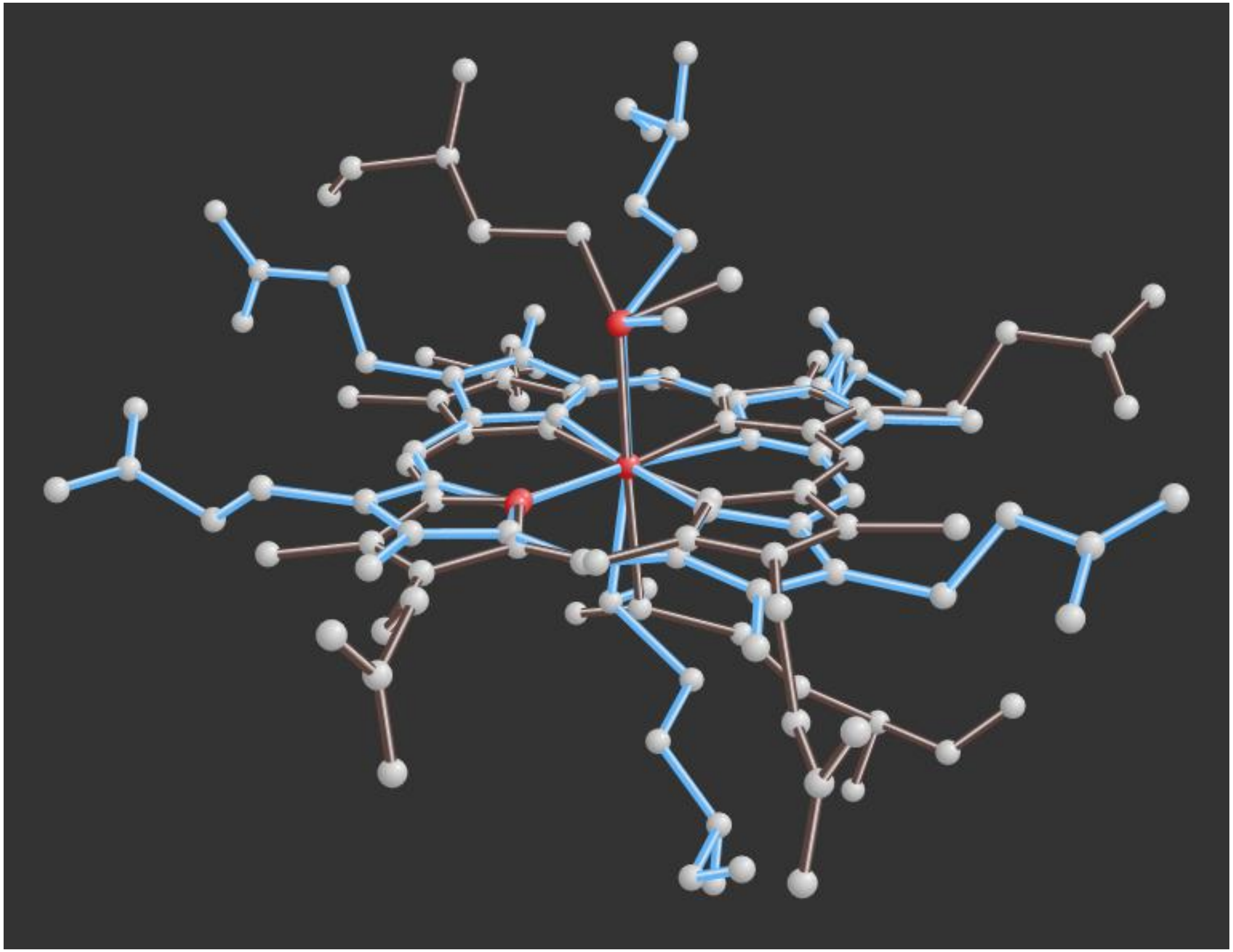
- Find sets of 3 suitable pivots in each structure
- For each pair of pivot sets
  - Align the structures using these pivots (quaternions)
  1. Match atoms based on distance (with a threshold)
  2. Align again (quaternions)
  3. (Goto 1 until a stable configuration is reached)
  - Score = f(A, B, RMSD_M)

  M is the set of matched atoms, N is the total number of atoms, and RMSD_M is the RMSD of matched atoms.
- Return the alignment with the best score

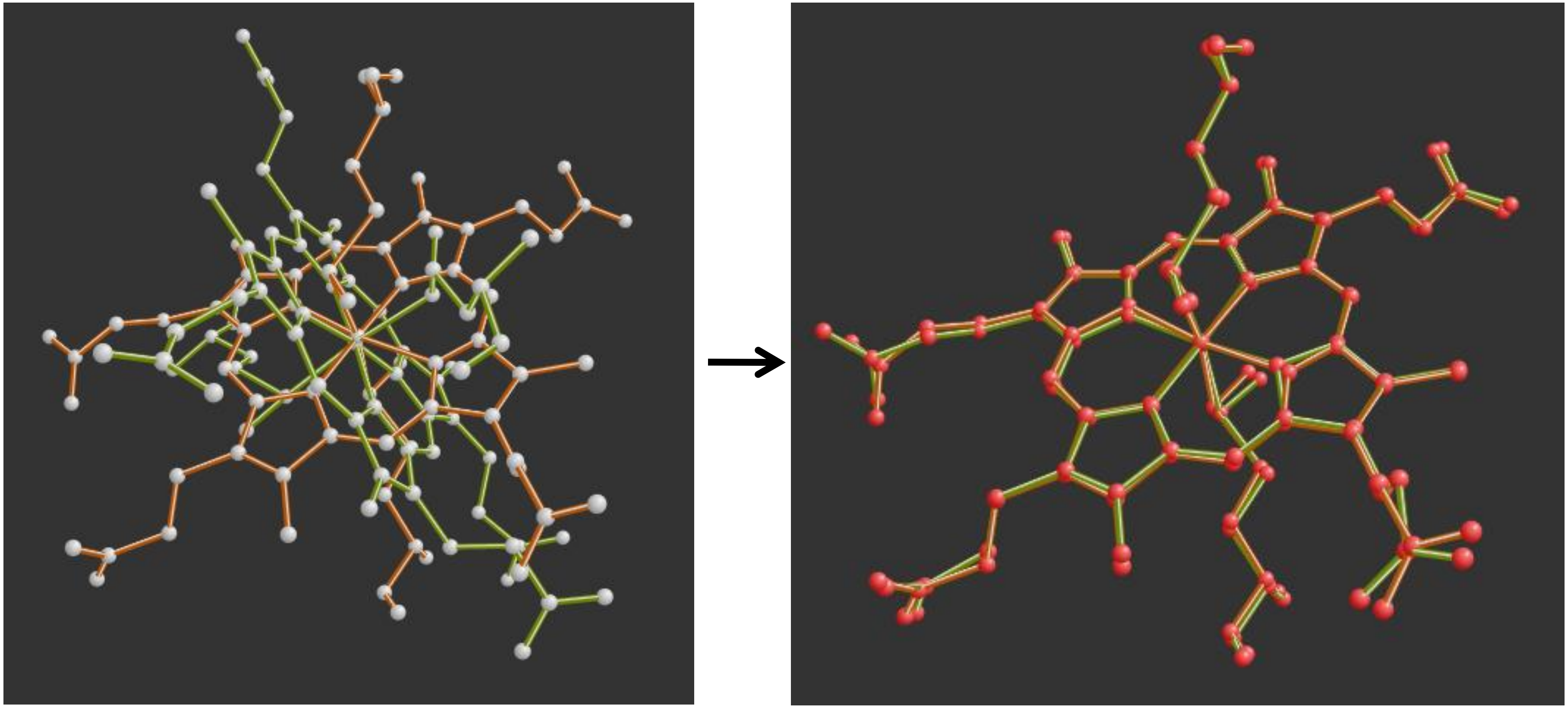
Example (FeN_4S_2)



Example (FeN_4S_2)

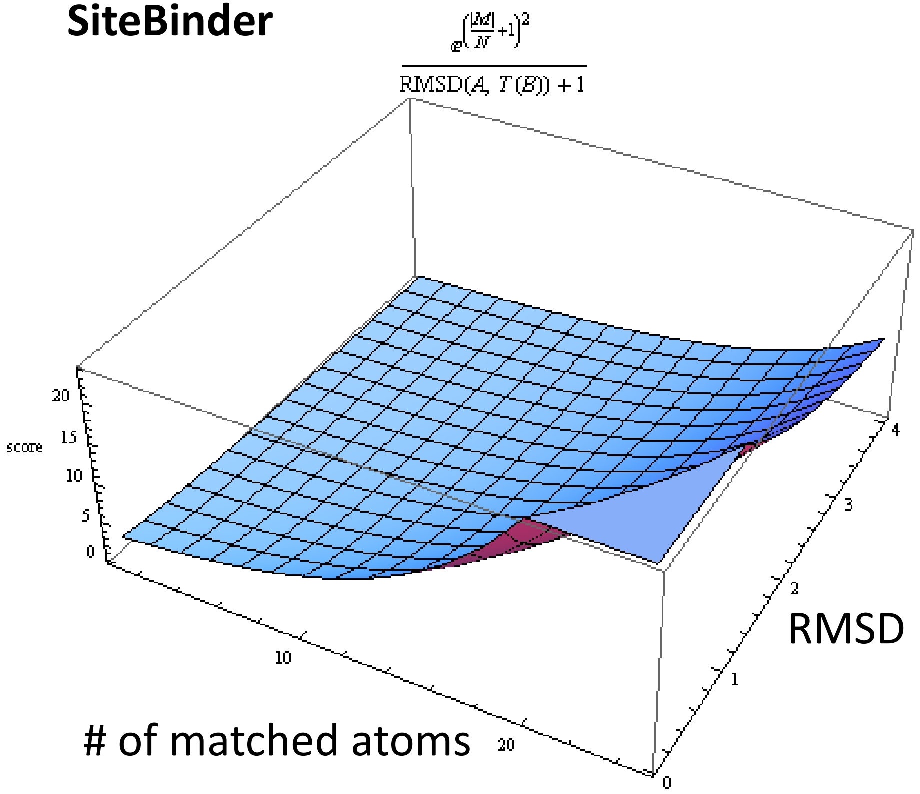


Example (FeN_4S_2)

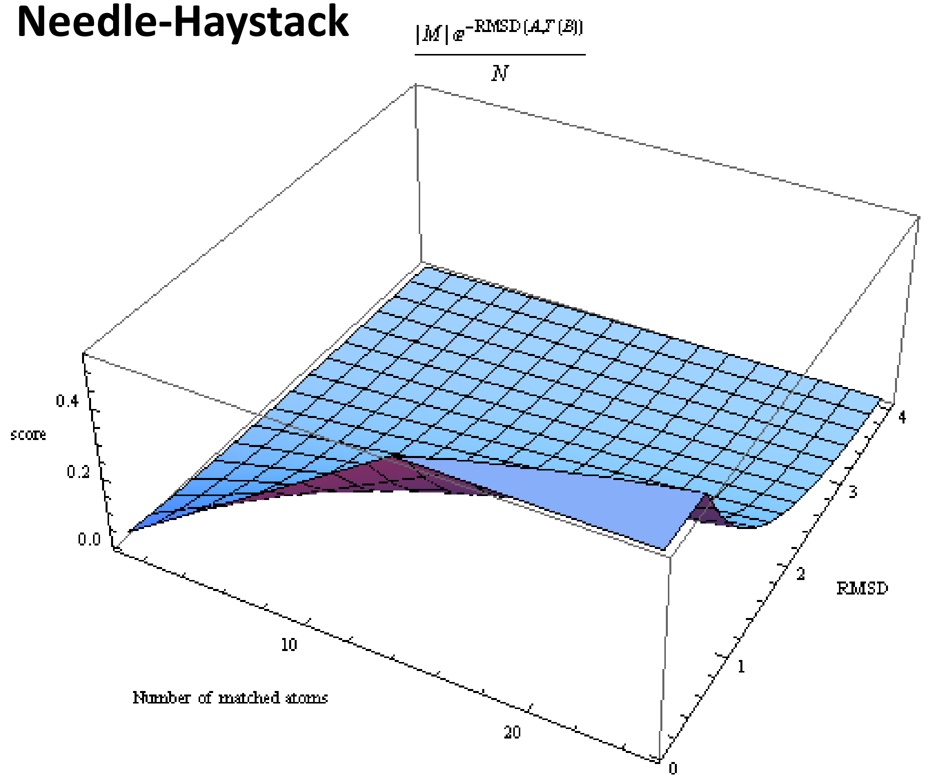


Scoring

SiteBinder



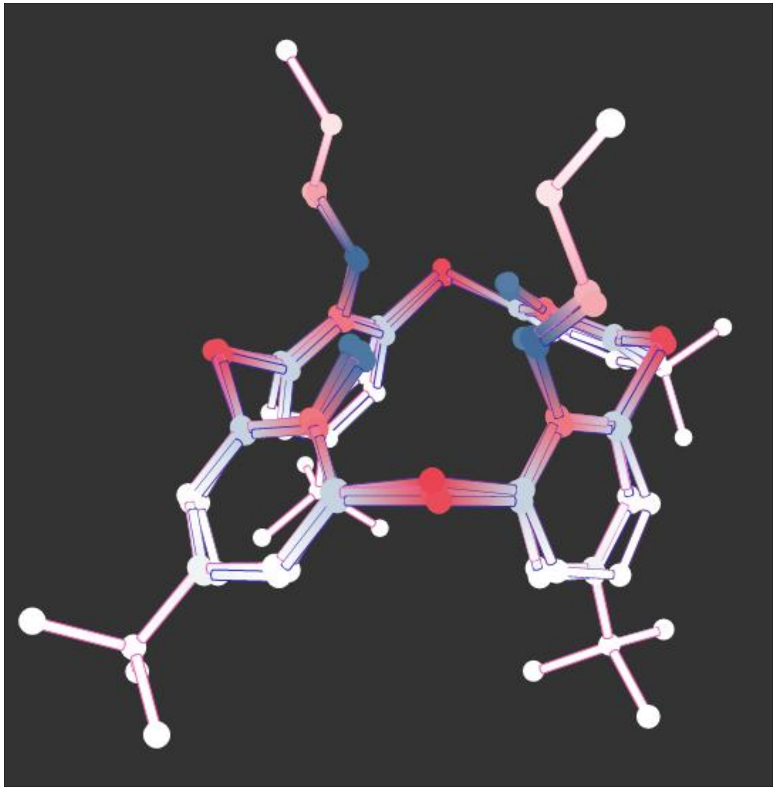
Needle-Haystack



$$\frac{e^{(1+|M|/N)^2}}{1+\mathrm{RMSD}_M(A,B)}$$

M is the set of matched atoms, N is the total number of atoms, and RMSD_M is the RMSD of matched atoms.

Different Pivot Criteria (Charges)



$$\mathrm{score}^*(M,N,A,B) = R_M^2 \cdot \mathrm{score}'(M,N,A,B) = R_M^2 \cdot \frac{e^{(1+|M|/N)^2}}{1+\mathrm{RMSD}_M(A,B)}$$

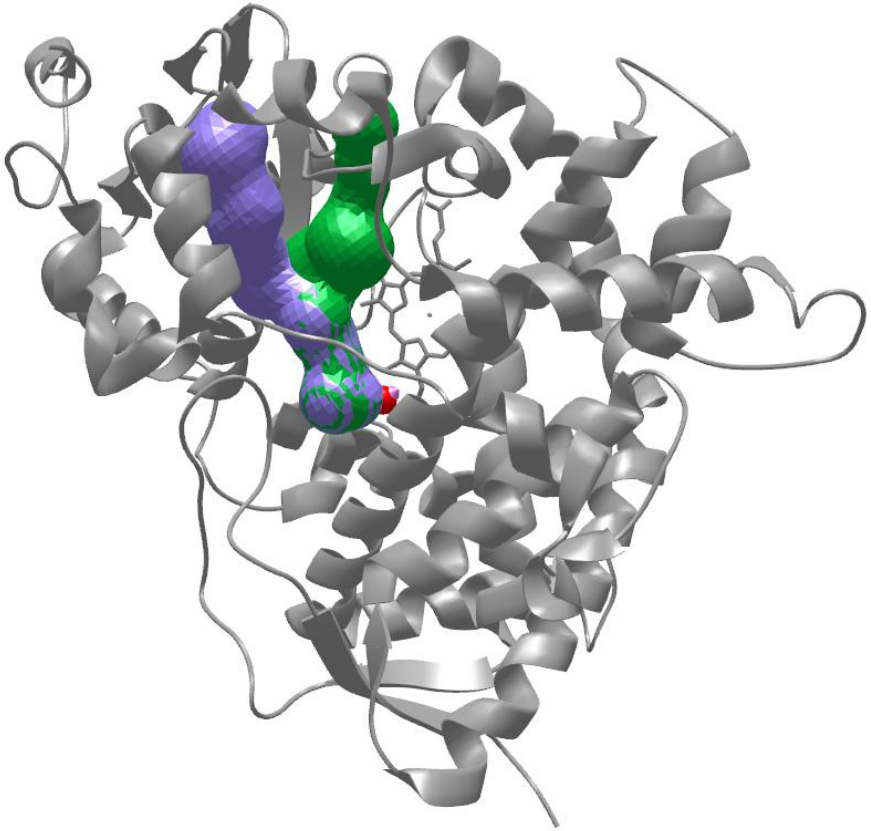
R_M^2 says how well the matched charges correlate.

3D Alignment Summary

3D structural alignment using pivot elements and improved scoring function to identify the correct result.

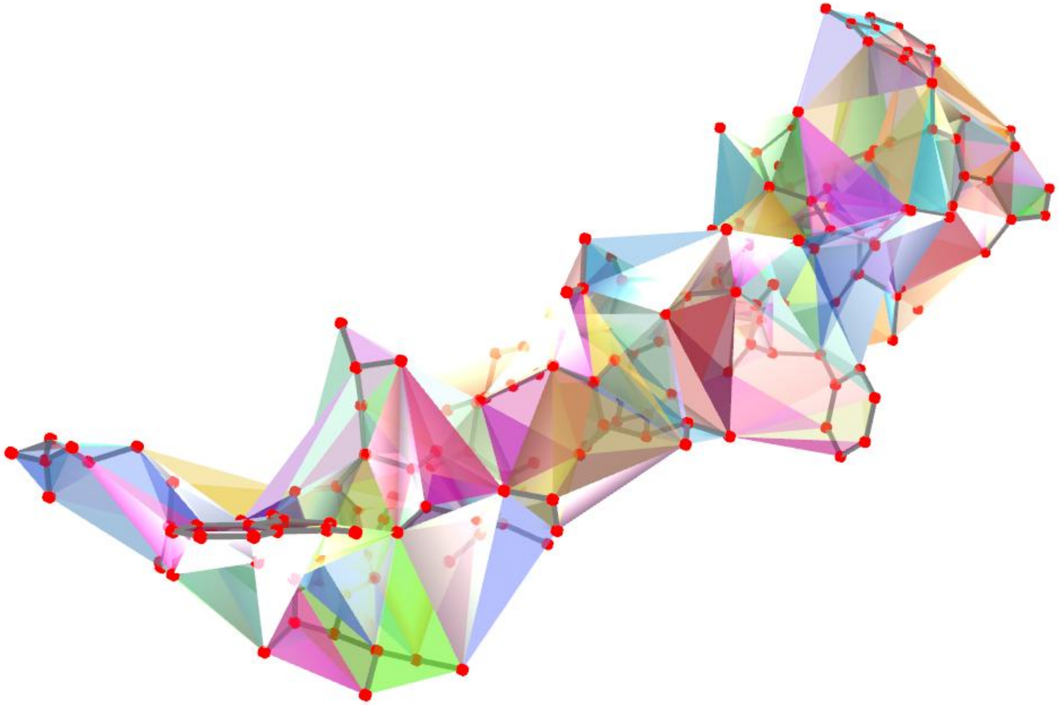
Tunnels

- **MOLE**, Caver, Hollow, MolAxis
- Static analysis vs. dynamics
- Grid/Triangulation Based

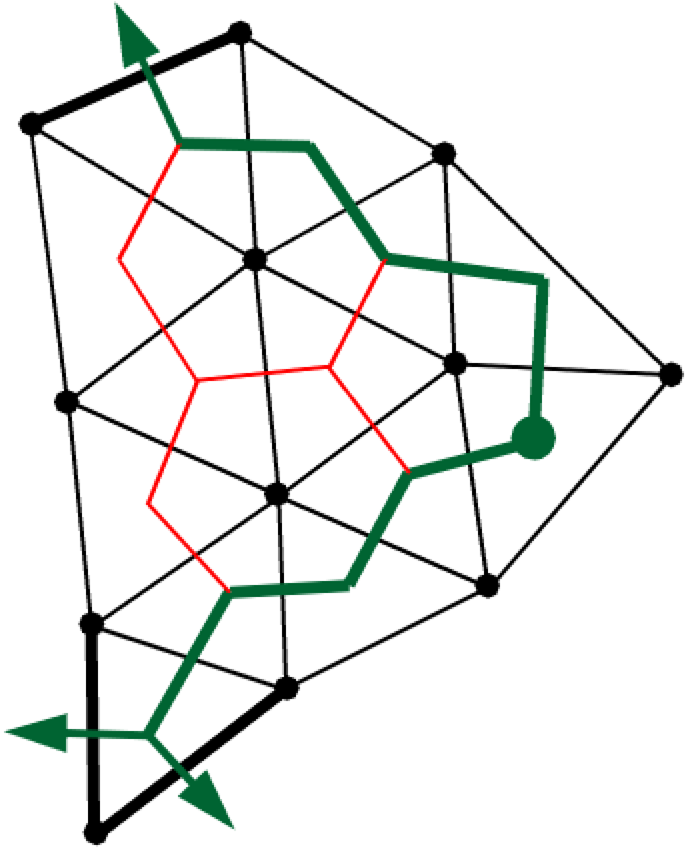


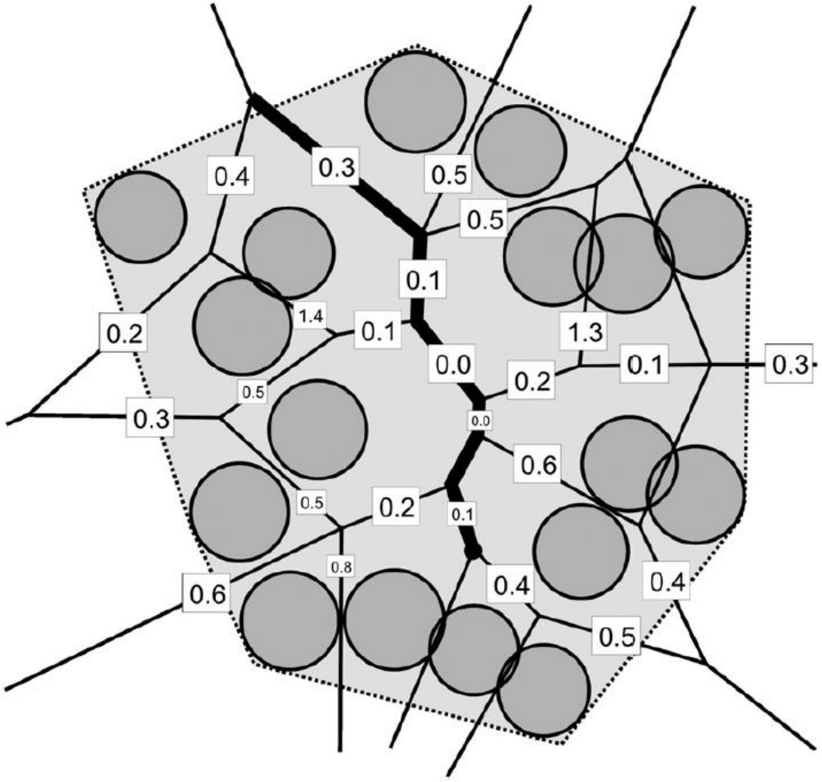
Protein (Delaunay) Triangulation

- Edelsbrunner H., Liang J.: Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design, Protein Science, 1998.



Finding the Tunnels – The Idea





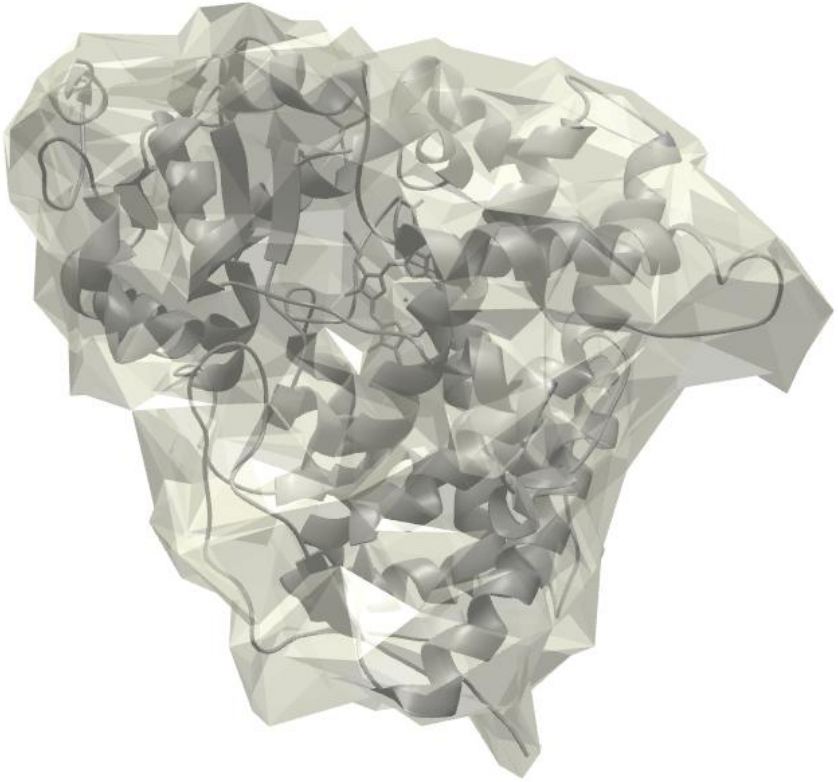
Tunnel Finding Algorithm

1. Triangulate the molecule (Voronoi diagram)
2. Specify starting point (from a database/user)
3. Do depth-first (A*) search
   - Each time a boundary tetrahedron is visited, report a tunnel
   - Edge weight ~ amount off space around the edge
4. Post-process the tunnels (remove duplicate tunnels, …)

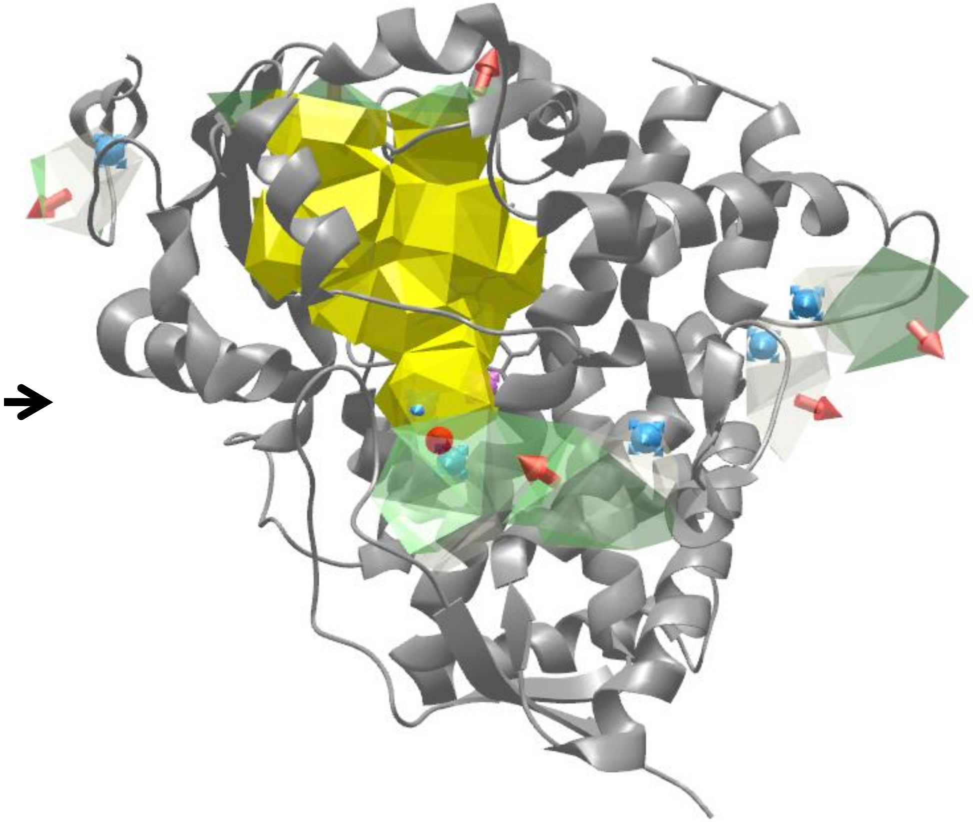
(My) Modified Approach

1. Triangulate the molecule (Voronoi diagram)
2. Identify cavities ("empty space")
   - Remove tetrahedrons that are too big or too small
   - Find connected components in the new graph
3. Identify start points within cavities
   - Deep point(s) in each cavity
   - Database/user
4. Identify end points within cavities
   - Connected components of the "boundary"
5. For each pair of start and end points in the same cavity use Dijkstra's algorithm to find the tunnel
   1. Edge weight ~ amount of space around the edge
6. Post-process the tunnels (remove duplicate tunnels, …)

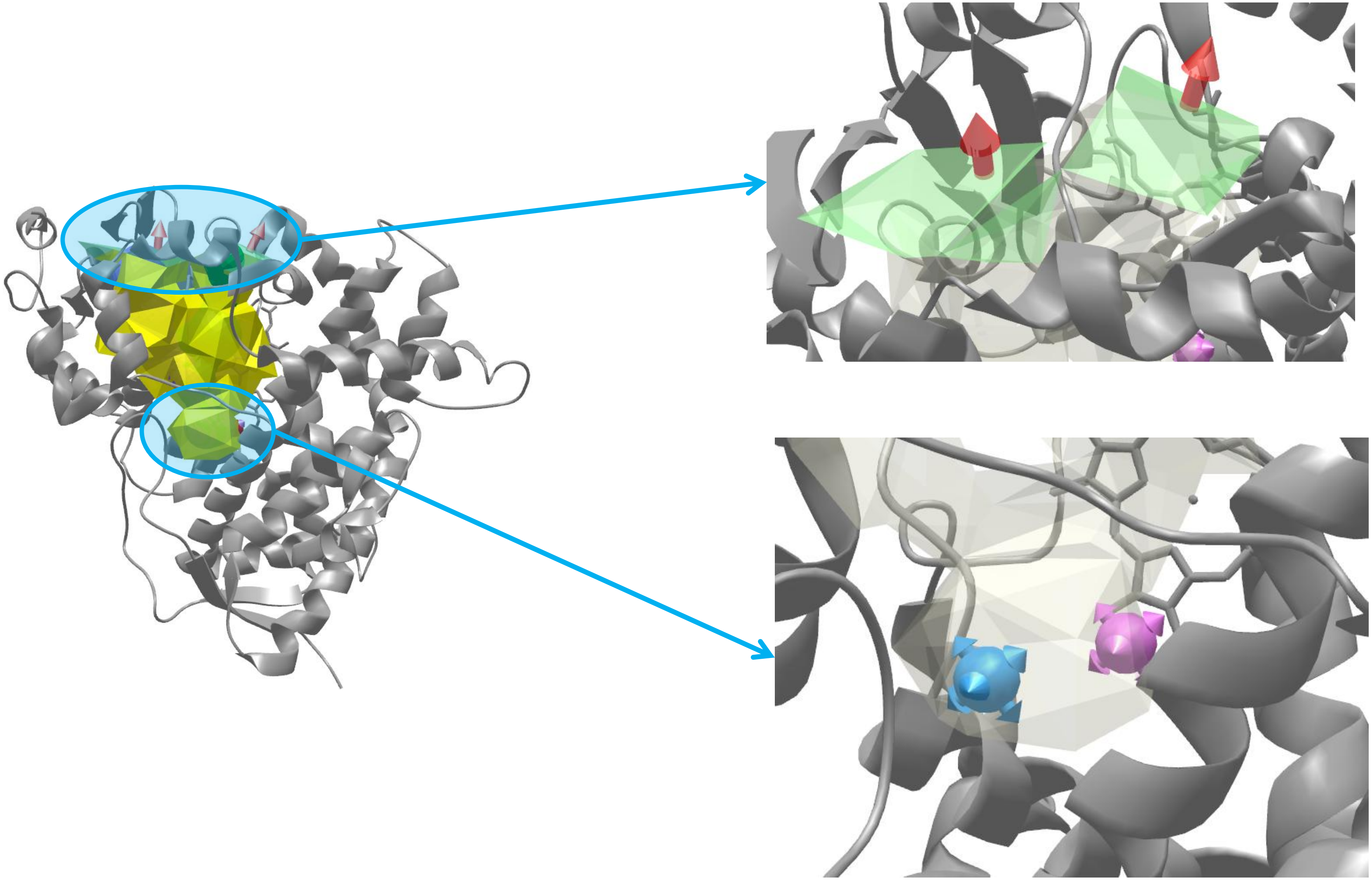
The Cavities



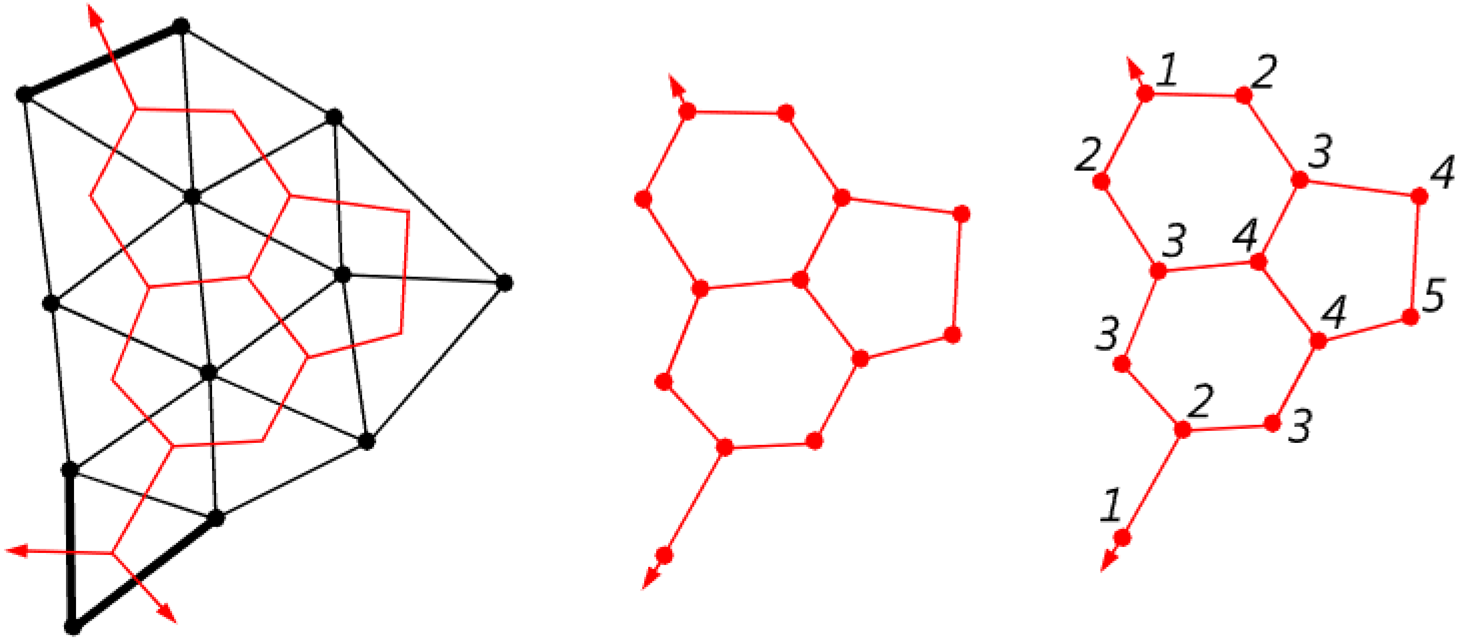
➔



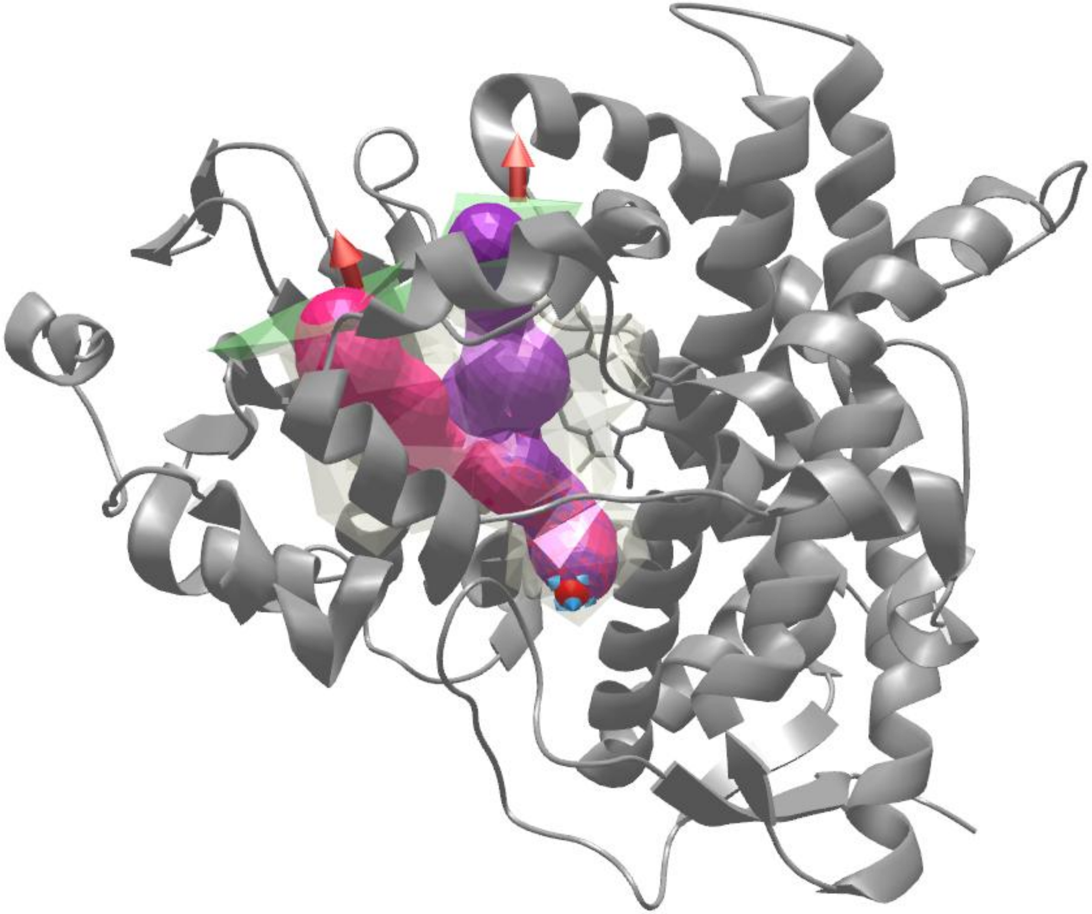
Start and End Points



Depth Computation



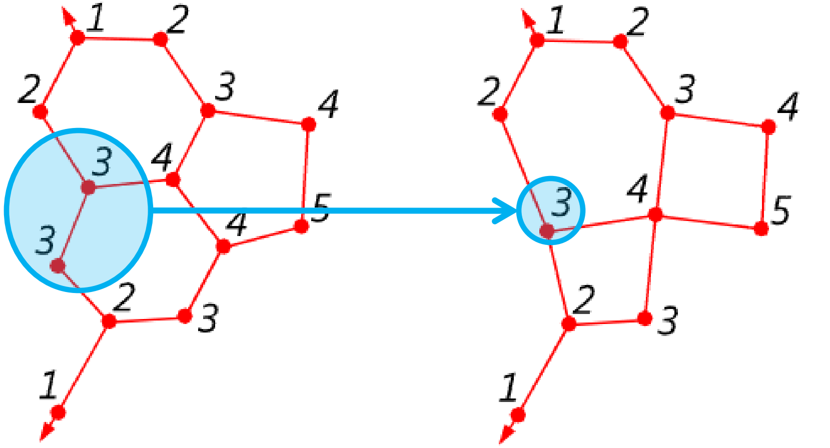
The Tunnels

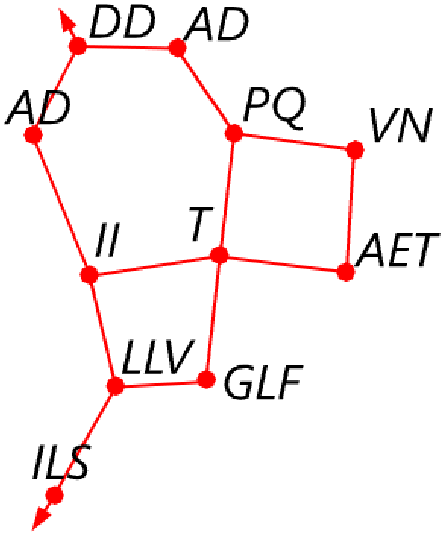


Tunnel Computation Summary

Triangulate the protein, split it into smaller graphs and use Dijkstra’s algorithm to find interesting paths (= tunnels) in them.

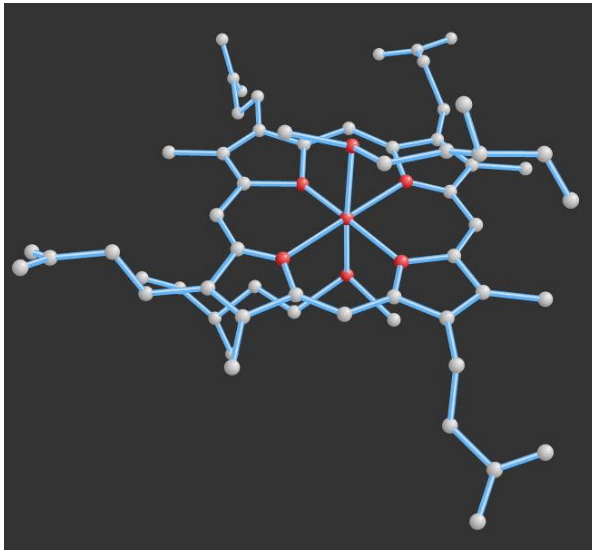
Aligning the Tunnels/Cavities





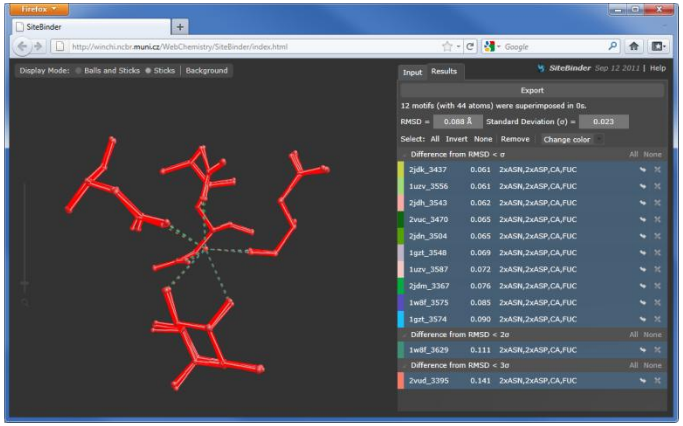
Reduced Cavity Graph

?



Summary

- 3D Alignment Algorithm

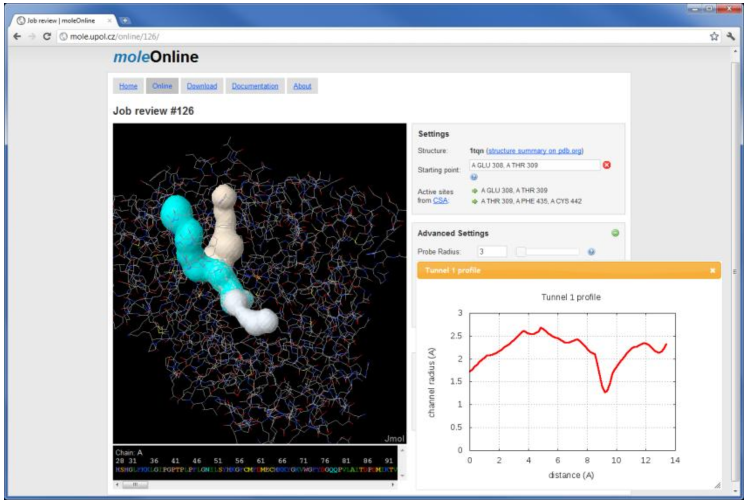

Firefox
SiteBinder
http://winchi.ncbr.muni.cz/WebChemistry/SiteBinder/index.html
Google
Display Mode: Balls and Sticks Sticks Background
Input Results
SiteBinder Sep 12 2011 | Help
Export
Select: All Invert None Remove Change color

12 motifs (with 44 atoms) were superimposed in 0s.

RMSD = 0.088 Å Standard Deviation (σ) = 0.023

Difference from RMSD < σ		
2jdk_3437	0.061	2xASN,2xASP,CA,FUC
1uzv_3556	0.061	2xASN,2xASP,CA,FUC
2jdh_3543	0.062	2xASN,2xASP,CA,FUC
2vuc_3470	0.065	2xASN,2xASP,CA,FUC
2jdn_3504	0.065	2xASN,2xASP,CA,FUC
1gzt_3548	0.069	2xASN,2xASP,CA,FUC
1uzv_3587	0.072	2xASN,2xASP,CA,FUC
2jdm_3367	0.076	2xASN,2xASP,CA,FUC
1w8f_3575	0.085	2xASN,2xASP,CA,FUC
1gzt_3574	0.090	2xASN,2xASP,CA,FUC
Difference from RMSD < 2σ		
1w8f_3629	0.111	2xASN,2xASP,CA,FUC
Difference from RMSD < 3σ		
2vud_3395	0.141	2xASN,2xASP,CA,FUC

- Tunnel Finding Algorithm



Job review #126

Settings

Structure: 1tqn (structure summary on pdb.org)

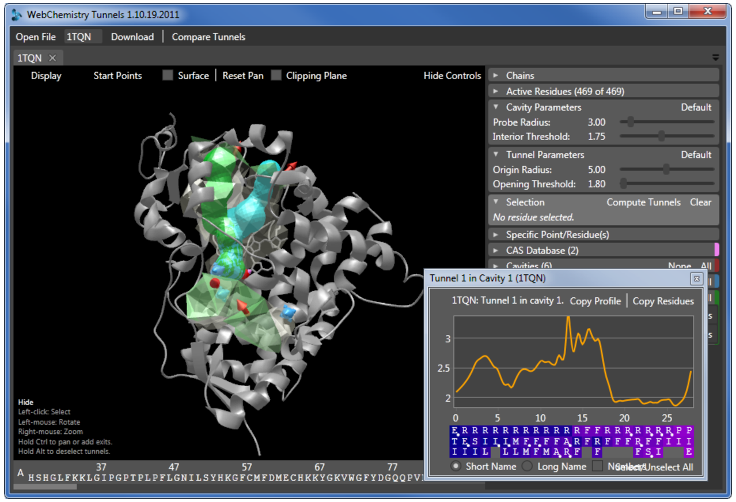
Starting point: A GLU 308, A THR 309

Active sites from CSA: A GLU 308, A THR 309; A THR 309, A PHE 435, A CYS 442

Advanced Settings

Probe Radius: 3

Chain: A



Cavity Parameters: Probe Radius: 3.00; Interior Threshold: 1.75

Tunnel Parameters: Origin Radius: 5.00; Opening Threshold: 1.80

Selection: No residue selected.

1TQN: Tunnel 1 in cavity 1.

Hide
Left-click: Select
Left-mouse: Rotate
Right-mouse: Zoom
Hold Ctrl to pan or add exits.
Hold Alt to deselect tunnels.

Future Work

- Combine 3D structural alignment and tunnel computation

- (Write PhD thesis about it)

Thank you for your attention.