CombDock

[Inbar, Y., Benyamini H., Nussinov R. and Wolfson H.J. (2005) "Prediction of multimolecular assemblies by multiple docking". *J. Mol. Biol.*, **349**, 435-447]

- All pairs docking
  - $N(N-1)/2$ pairs
  - keep best K transformations for each pair
- Combinatorial assembly
  - find best spanning tree representing a valid complex
  - keep best D trees of size s starting at i
- Rescoring
  - cluster (to avoid redundancy in solution set)
  - geometric component
    - large interface area and small steric overlap
  - physico-chemical component
    - count number of buried non-polar atoms

How many spanning trees?

(i) If we have N vertices and 1 edge between each pair there are N^{N-2} spanning trees.

(ii) If we have K edges between each pair of vertices, then there are K^{N-1} graphs of type (i).

So there are $N^{N-2}K^{N-1}$ spanning trees.

Can't search the whole space!
So use a heuristic solution.

CombDock results (1)

Yeast RNA polymerase II elongation complex

10 protein chains

$K = 100$, 15 pairwise interactions predicted

10^{26} possible complexes

50188 complexes generated by combinatorial assembly

1113 complexes left after clustering

2nd ranked complex had RMSD of 1.37.sp 2 Human subunits modelled by homology
6th ranked complex had RMSD of 1.9.

CombDock results (2)

Bovine arp2/3 complex

7 protein chains

$K = 100$

1.68×10^{17} possible complexes

5488 complexes generated by combinatorial assembly

145 complexes left after clustering

3nd ranked complex had RMSD of 1.2.sp 2 *Drosophila melanogaster* subunits modelled by homology
10th ranked complex had RMSD of 1.9.

Structural biology, genomics and proteomics

Structural biology

— gives an understanding of biological function at the molecular level through determination of individual macromolecular structures and complexes.

Structural genomics

— aims to obtain 3D structures for all protein products of the genome.

Structural proteomics

— aims obtain structures of all macromolecular assemblies in the "complexome".

Protein-protein docking

- find sets of candidate docking orientations
- use a scoring function to evaluate these

"Although considerable improvement has been achieved, the scoring functions are still the weakest components of most docking algorithms."

[Inbar Y, Benyamini H, Nussinov R, Wolfson HJ.
Prediction of multimolecular assemblies by multiple docking.
J Mol Biol. 2005 Jun 3;349(2):435-447]

Computational challenges

- to predict atomic models of macromolecules (usually proteins) where experimentally determined structures are not available;
- to predict the structures of pairs of interacting macromolecules;
- to predict assemblies of many macromolecules by multiple docking;
- to fit high-resolution atomic models of macromolecules inside low-resolution envelopes obtained from cryo-EM experiments;
- to build a 3D atlas of a cell at atomic resolution.

Different resolutions

High-resolution

— X-ray, NMR

atomic detail

— atomic coordinates (x,y,z)

— Low-resolution

— (cryo)electron microscopy

— voxels with density values

3SOM: preprocessing step

- construct 3D grid with voxel size equal to that of density map
- estimate electron density value for each voxel in 3D grid
- smooth, threshold (produces volume matrix) and “edge detect” (produces surface matrix)

```
copy volume matrix to surface matrix

if ( volume[i][j][k] == "true" ) {
    if ( all 6 neighbours of volume[i][j][k] == "true" ) {
        set surface[i][j][k] = "false"
    }
}
```


3SOM: fast fitting round

- fit surface voxel of A onto surface voxel of B
- rotate B so that surface normals are aligned
- rotate B around surface normal in 9° steps
- compute surface overlap

$$\frac{number\ of\ overlapping\ surface\ voxels}{number\ of\ surface\ voxels\ in\ target}$$