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”.

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.

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]

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

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

HOLE

[Smart et al., 1993, Biophys. J., 65:2455-2460]

User specifies

- initial point p within channel
- vector v (approximately) in the direction of the channel

Uses Metropolis Monte Carlo simulated annealing to find the centre and radius of the largest sphere that can be placed within the channel with its centre on the plane perpendicular to v that passes through p.

Repeat for the next plane, at a fixed displacement from previous plane.

Stop when the accommodated sphere has a radius > 5 $\mathring{\text{A}}$.

Repeat from p in the direction $-v$.

Tilton et al., 1986, J. Mol. Biol., 192:443-456

Method to find channels within proteins

Atomic radii reduced by Δr, where:

$$(\Delta r)^2 = \frac{B}{8\pi^2}$$

Uses iterative procedure to find the largest sphere that can be inscribed by four atoms.

Intersections of these spheres define connections, the size of which is given by the circle of intersection.