V2 Protein Networks and Complexes

Connected graph $\iff$ there is a path between all pairs of nodes

In large (random) networks: complete $\{V\}$ is often not connected
$\rightarrow$ identify connected subsets $\{V_i\}$ with $\{V\} = \bigcup \{V_i\}$
$\rightarrow$ connected components (CC)

$\#CC = 5$
$N_{max} = 15$
$N_{min} = 1$
Basic Types: (1) Random Network

Given: $N$ vertices connected by $L$ edges
where the edges are randomly distributed between the vertices

Maximal number of links between $N$ vertices:

$$L_{max} = \frac{N(N - 1)}{2}$$

$=>$ probability $p$ for an edge between two randomly selected nodes:

$$p = \frac{L}{L_{max}} = \frac{2L}{N(N - 1)}$$

$=>$ average degree $\lambda$

$$\lambda = \frac{2L}{N} = p(N - 1)$$

Path lengths in a random network grow with $\ln(N)$ $=>$ “small world”
Random Network: $P(k)$

Network with $N$ vertices, $L$ edges

$=>$ probability for a random link:

$$p = \frac{2L}{N(N-1)}$$

Probability that random node has links to $k$ other particular nodes:

$$W_k = p^k (1 - p)^{N-k-1}$$

Probability that random node has links to any $k$ other nodes:

$$P(k) = \binom{N-1}{k} W_k = \frac{(N-1)!}{(N-k-1)!k!} W_k$$

Limit of large graph: $N \to \infty, p = \frac{\lambda}{N}$

$$\lim_{N \to \infty} P(k) = \lim_{N \to \infty} \frac{\frac{N!}{(N-k)!k!} p^k (1 - p)^{N-k}}{\frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k}}$$

$$= \frac{\lambda^k}{k!} \frac{1}{e^{-\lambda}}$$

3. Lecture WS 2019/20
Random Network: $P(k)$

Many independently placed edges $\Rightarrow$ Poisson statistics

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P(k \mid \lambda = 2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.14</td>
</tr>
<tr>
<td>1</td>
<td>0.27</td>
</tr>
<tr>
<td>2</td>
<td>0.27</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>4</td>
<td>0.090</td>
</tr>
<tr>
<td>5</td>
<td>0.036</td>
</tr>
<tr>
<td>6</td>
<td>0.012</td>
</tr>
<tr>
<td>7</td>
<td>0.0034</td>
</tr>
<tr>
<td>8</td>
<td>0.00086</td>
</tr>
<tr>
<td>9</td>
<td>0.00019</td>
</tr>
<tr>
<td>10</td>
<td>3.82e-05</td>
</tr>
</tbody>
</table>

$\Rightarrow$ Small probability for $k >> \lambda$

4. Lecture WS 2019/20
C(k) for a Random Network

Clustering coefficient when \( m \) edges exist between \( k \) neighbors

\[
C(k, m) = \frac{2m}{k(k-1)}
\]

Probability to have exactly \( m \) edges between the \( k \) neighbors

\[
W(m) = \binom{k}{m} p^m (1-p)^{\frac{k(k-1)}{2} - m}
\]

# possibilities of picking the \( m \) start nodes for the \( m \) edges from the \( k \) nodes.

Average \( C(k) \) for degree \( k \):

\[
C(k) = \frac{\sum_{m=0}^{\frac{k(k-1)}{2}} W(m) C(k, m)}{\sum_{m=0}^{\frac{k(k-1)}{2}} W(m)} = \ldots = p
\]

\( \rightarrow \) \( C(k) \) is independent of \( k \)

\(<=>\) same local connectivity throughout the network
Basic Types: (2) Scale-Free

Growing network a la Barabasi and Albert (1999):
- start from a small "nucleus" of $m_0$ connected nodes
- in each iteration step, add new node with $n$ links
- connect new links to existing nodes with probability $p_i$ proportional to degree $k_i$ of each existing node ("preferential attachment");

\[ p_i = \left( \frac{k_i}{\sum k_i} \right)^\beta \]

in BA-model $\beta = 1$

Properties:
- this leads to a power-law degree distribution:
  \[ P(k) \propto k^{-\gamma} \]
  with $\gamma = 3$ for the BA model
- self-similar structure with highly connected hubs (no intrinsic length scale)

=> average path length grows with $\ln(N) / \ln(\ln(N))$
=> this grows much slower than for random graphs
  => “very small world”
The Power-Law Signature

Power law \( P(k) \propto k^{-\gamma} \)

Take log on both sides:
\[
\log(P(k)) = -\gamma \log(k)
\]

Plot \( \log(P) \) vs. \( \log(k) \) \( \Rightarrow \) straight line

Note: for fitting \( \gamma \) against experimental data it is often better to use the integrated \( P(k) \)
\( \Rightarrow \) integral smoothens the data

\[
\int_{k_0}^{k} P(k) \, dk = \left[ -\frac{k^{-(\gamma-1)}}{\gamma} \right]_{k_0}^{k}
\]
Scale-Free: Examples

The World-Wide-Web:
=> growth via links to portal sites

Flight connections between airports
=> large international hubs, small local airports

Protein interaction networks
=> some central, ubiquitous proteins

http://a.parsons.edu/~limam240/blogimages/16_full.jpg
Algorithms on Graphs

How to represent a graph in the computer?

1. **Adjacency list**

   => list of neighbors for each node

   1: (3)  
   2: (3)  
   3: (1, 2, 4, 5)  
   4: (3, 5, 6)  
   5: (3, 4, 6, 7)  
   6: (4, 5)  
   7: (5)

   + minimal memory requirement  
   + vertices can easily be added or removed  
   – requires $O(\lambda)$ time to determine whether a certain edge exists

Note: for weighted graphs store pairs of (neighbor label, edge weight)
2. **Adjacency matrix** (see V1)

→ $N \times N$ matrix with entries $M_{uv}$

\[ M_{uv} = \text{weight when edge between } u \text{ and } v \text{ exists,} \]
\[ 0 \text{ otherwise} \]

→ symmetric for undirected graphs

+ fast $O(1)$ lookup of edges

– large memory requirements

– adding or removing nodes is expensive

\[ M = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0
\end{bmatrix} \]

Note: very convenient in programming languages that support sparse multi-dimensional arrays

=> Perl
3. Incidence matrix

→ $N \times M$ matrix with entries $M_{nm}$

\[ M_{nm} = \text{weight when edge } m \text{ ends at node } n \]

0 otherwise

→ for a plain graph there are two entries per column

→ directed graph:

  indicate direction via sign (in/out)

The incidence matrix is a special form of the **stoichiometric matrix** of reaction networks.
V2(b): Structures of Protein Complexes and Subcellular Structures

(1) We normally assume that various enzymes of a biochemical pathway „swim“ in the cytosol and randomly meet the substrate molecules one after another.

Yet, sometimes **multiple enzymes** of a biochemical pathway associate into **large complexes** and „hand over“ the substrates from one active site to the next one. Advantage: this avoids free diffusion, increases local substrate density.

(2) Membrane transporters and receptors often form **oligomers** in the **membrane**. Advantage:
(i) large structures are built from small building blocks (simplicity)
(ii) Oligomer formation can be regulated separately from transcription.

(3) Also: complicated structural components of the cell (e.g. cytoskeleton) are built from many small components (e.g. actin)
2.1 RNA Polymerase II

RNA polymerase II is the central enzyme of gene expression and synthesizes all messenger RNA in eukaryotes.

Structure of a **cellular editor** that "cuts and pastes" the first draft of RNA straight after it is formed from its DNA template.

It has two distinct, unequal halves surrounding a tunnel.

Larger part: appears to contain proteins and the short segments of RNA,

smaller half: is made up of proteins alone.

On one side, the tunnel opens up into a cavity, which is believed to function as a holding space for the fragile RNA waiting to be processed in the tunnel.

2.1 Protein synthesis: ribosome

The ribosome is a complex subcellular particle composed of protein and RNA. It is the site of protein synthesis, http://www.millerandlevine.com/chapter/12/cryo-em.html

Components of ribosome assemble spontaneously in vitro: no helper proteins (assembly chaperones) needed

Model of a ribosome with a newly manufactured protein (multicolored beads) exiting on the right.

large ribosomal subunit from *Haloarcula marismortui*. RNA is shown in gray and the protein backbone in yellow.

Ban *et al.* (2000)
2.1 Nuclear Pore Complex (NPC)

NPC is a 50-100 MDa protein assembly that regulates and controls trafficking of macromolecules through the nuclear envelope.

Three-dimensional image of the NPC obtained by electron microscopy.
A-B The NPC in yeast.
Figure A shows the NPC seen from the cytoplasm while figure B displays a side view.

C-D The NPC in vertebrate (Xenopus).

http://www.nobel.se/medicine/educational/dna/a/transport/ncp_em1.html


Molecular structure: lecture V20
2.1 Arp2/3 complex

The seven-subunit Arp2/3 complex choreographs the formation of branched actin networks at the leading edge of migrating cells.

(A) Model of actin filament branches mediated by *Acanthamoeba* Arp2/3 complex. (D) Density representations of the models of actin-bound (green) and the free, WA-activated (as shown in Fig. 1D, gray) Arp2/3 complex.

Volkmann *et al.*, Science 293, 2456 (2001)
2.1 icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine

Model for active-site coupling in the E1E2 complex. 3 E1 tetramers (purple) are shown located above the corresponding trimer of E2 catalytic domains in the icosahedral core. Three full-length E2 molecules are shown, colored red, green and yellow. The lipoic domain of each E2 molecule shuttles between the active sites of E1 and those of E2. The lipoic domain of the red E2 is shown attached to an E1 active site. The yellow and green lipoic domains of the other E2 molecules are shown in intermediate positions in the annular region between the core and the outer E1 layer. Selected E1 and E2 active sites are shown as white ovals, although the lipoic domain can reach additional sites in the complex.

Milne et al., EMBO J. 21, 5587 (2002)
2.1 Apoptosome

Apoptosis is the dominant form of programmed cell death during embryonic development and normal tissue turnover. In addition, apoptosis is upregulated in diseases such as AIDS, and neurodegenerative disorders, while it is downregulated in certain cancers. In apoptosis, death signals are transduced by biochemical pathways to activate caspases, a group of proteases that utilize cysteine at their active sites to cleave specific proteins at aspartate residues. The proteolysis of these critical proteins then initiates cellular events that include chromatin degradation into nucleosomes and organelle destruction. These steps prepare apoptotic cells for phagocytosis and result in the efficient recycling of biochemical resources.

In many cases, apoptotic signals are transmitted to mitochondria, which act as integrators of cell death because both effector and regulatory molecules converge at this organelle. Apoptosis mediated by mitochondria requires the release of cytochrome c into the cytosol through a process that may involve the formation of specific pores or rupture of the outer membrane. Cytochrome c binds to Apaf-1 and in the presence of dATP/ATP promotes assembly of the apoptosome. This large protein complex then binds and activates procaspase-9.
2.1.2 Categories of Protein Complexes

Complexes can be classified e.g. by function / size / involvement of other components (nucleic acids, carbohydrates, lipids).

Alternatively: mechanistic classification:

(1) transient vs. permanent
(2) obligate vs. non-obligate

Obligate: components function only when in the bound state.

Non-obligate: unbound components can also exist as monomers. Examples of non-obligate complexes: antibodies, signalling complexes, complexes of RNA polymerase with different initiation and elongation factors.
2.3 Determining molecular 3D structures: X-ray crystallography

Beam of photons (no mass) with high energy, method needs relatively large samples

Crystallize proteins, record diffraction patterns of X-rays (scatter at the electron clouds)
=> reconstruction

1 keV $\approx 10^5$ kJ/mol

http://www.molbio1.princeton.edu/macro/about.html
X-ray reconstruction

Reconstruct electron density maps from diffraction patterns at multiple wavelengths and orientations, refine structure computationally.

Main problem: proteins do not like to crystallize (especially membrane proteins)

PDB: experimental technique == X-Ray: 43138 results
X-Ray && "in the membrane": 1232 results

http://www.molbio1.princeton.edu/macro/about.html
2.3.2 Nuclear magnetic resonance

Place sample with the proteins into strong magnetic field
=> nuclei with non-zero magnetic moment
(\textsuperscript{1}H, \textsuperscript{13}C, \textsuperscript{15}N) align

Apply electromagnetic RF field
=> resonances of the nuclear spins depending on
  • atom type
  • chemical environment
  => extract distances between close by \textsuperscript{1}H atoms
   (distance constraints)

+ no crystallization required, proteins in physiological environment
+ atomic resolution
– too much overlap for larger proteins (≥15 kDa)
2.3.3 Electron microscopy

Like X-ray crystallography, but with electrons

=> much stronger interaction with electron clouds
  => works already on 2D crystals (membrane proteins!)
  => strong radiation damage of the sample
    => resolution limited (keep electron energy low)

For 3D tomography: rotate sample

Cryo-EM images of the LH1-RC core complexes of *Rhodospirillum rubrum* at 8.5 Å resolution

Atomic force microscopy

Scan membrane with proteins (in physiological conditions) => protein arrangement (coarse view) => protein shape (high resolution)

Shapes and sizes of monomeric LH1 from purple bacteria

AFM pulling

Mechanical Unfolding of a Titin Ig Domain: Structure of Unfolding Intermediate Revealed by Combining AFM, Molecular Dynamics Simulations, NMR and Protein Engineering

Susan B. Fowler¹, Robert B. Best¹, José L. Toca Herrera¹
Trevor J. Rutherford¹, Annette Steward¹, Emanuele Paci²
Martin Karplus²,³ and Jane Clarke¹*

Can also be applied to protein complexes

(a)

\[
\text{Force (pN)} \quad \text{Wild type} \quad \text{mutant} \quad \text{V4A}
\]

\[
\text{Extension (nm)}
\]

\[
\]
2.3.6 Fluorescence energy transfer

Chromophore of the cyan fluorescent protein (CFP) absorbs at 436 nm and emits at 480 nm, YFP absorbs at 480 nm and emits at 535 nm. => resonant (non-radiative) energy transfer from CFP onto YFP when both are close enough

Resonant Förster transfer $\propto d^{-6}$

Tag two potential complex partners with CFP and YFP and measure fluorescence spectrum:

- Observed when CFP and YFP are far away
- Observed when CFP and YFP are close
Lumier-based mammalian interactome mapping

LUMIER assay is based on co-immunoprecipitation. Protein A is fused to Renilla luciferase, while Protein B is linked to an affinity tag. Tagged constructs are transfected into appropriate cell lines where they are overexpressed. Cells are then lysed and protein B is immunoprecipitated using an appropriate antibody against the affinity tag.

+ Pro
+ Easy to perform, can be used in a HT screening format.
+ Can be used in different cell lines.
+ Well suited for binary interactions, indirect interactions can also be detected

- Con
- **Cells need to be lysed** prior to immunoprecipitation. This can result in the disruption of weak and transient PPIs, as well as the introduction of potential artifacts (e.g., by bringing together proteins in the lysate, which might not normally interact with one another in the cell, destabilizing proteins and exposing previously concealed non-native binding surfaces).

### Structural techniques - overview

<table>
<thead>
<tr>
<th></th>
<th>X-ray crystallography</th>
<th>NMR</th>
<th>EM/tomography</th>
<th>AFM</th>
<th>FRET</th>
<th>Y2H</th>
<th>TAP</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structure ≤ 3Å</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>structure ≥ 3Å</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>contacts</strong></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>proximity</strong></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>stoichiometry</strong></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>complex symmetry</strong></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thanks to improvements in EM detectors
Predicting Structures of Protein Complexes from Connectivities: CombDock

CombDock: automated approach for predicting 3D structure of heterogenous multimolecular assemblies.

Input: structures of $N$ individual proteins

Problem appears more difficult than the pairwise docking problem.

Idea: exploit additional geometric constraints that are part of the combinatorial problem.

Haim Wolfson
Tel Aviv University
http://www.cs.tau.ac.il/~wolfson/

Review: pairwise docking: Katchalski-Kazir algorithm

Discretize proteins A and B on a grid. Every node is assigned a value:

\[ f_{A_i,m,n} = \begin{cases} 1 : & \text{surface of molecule} \\ \rho : & \text{core of molecule} \\ 0 : & \text{open space} \end{cases} \]

and

\[ f_{B_i,m,n} = \begin{cases} 1 : & \text{inside molecule} \\ 0 : & \text{open space} \end{cases} \]

The correlation function of \( f_A \) and \( f_B \) is:

\[ f_{C_{x,y}} = \sum_{i=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} f_{A_{i,m,n}} \times f_{B_{i+a,m+b,n+y}} \]

Use FFT to compute correlation efficiently (see V3).

Output: solutions with best surface complementarity.

(1) All pairs docking module

Aim: predict putative pairwise interactions

Based on the $N$ individual protein structures perform pairwise docking for each of the $N (N - 1) / 2$ pairs of proteins

Since the correct scoring of pairwise-docking is difficult, the correct solution may be among the first few hundred solutions.

→ keep $K$ best solutions for each pair of proteins.

Inbal et al. varied $K$ from dozens to hundreds.

Spanning tree = a graph that connects all vertices and has no circles

Review: Spanning trees

Given a connected and undirected graph, a **spanning tree** of that graph is a subgraph that is a tree and connects all the vertices together.

A single graph can have many different spanning trees.

A **minimum spanning tree (MST)** or minimum weight spanning tree for a weighted, connected and undirected graph is a spanning tree with weight less than or equal to the weight of every other spanning tree. The weight of a spanning tree is the sum of weights given to each edge of the spanning tree.

For a graph with \( V \) vertices, a minimum spanning tree has \( (V - 1) \) edges.

*Kruskal’s algorithm* for finding a minimum spanning tree.

1. **Sort all the edges in non-decreasing order of their weight.**
2. **Pick the edge with smallest weight. Check if it forms a cycle with the spanning tree formed so far. If cycle is not formed, include this edge. Else, discard it.**
3. **Repeat step#2 until there are \((V-1)\) edges in the spanning tree.**
Example: Spanning tree – algorithm of Kruskal

Avoid constructing cycles

Algorithm stops when MST contains V-1 edges (here 7).
(2) Combinatorial assembly module

Input: $N$ subunits and $N (N - 1) / 2$ sets of $K$ scored transformations. These are the candidate interactions.

Reduction to a spanning tree

Build weighted graph representing the input:
- each protein structure = vertex
- each transformation (docking orientation) = edge connecting the corresponding vertices
- edge weight = docking score of the transformation

→ Since the input contains $K$ transformations for each pair of subunits, we get a complete graph with $K$ parallel edges between each pair of vertices.

www.wikipedia.org
(2) Combinatorial assembly module

For 2 subunits, each candidate binary docking complex is represented by an edge and the 2 vertices.

For the full complex, a candidate complex is represented by a spanning tree. Each spanning tree of the input graph represents a particular 3D structure for the complex of all input structures.

→ Problem of finding 3D structures of complexes is equivalent to finding spanning trees.

The number of spanning trees in a complete graph with \( N \) nodes and no parallel edges is \( N^{N-2} \) (Cayley’s formula).

Here, the input graph has \( K \) parallel edges between each pair of vertices. → the number of spanning trees is \( N^{N-2} K^{N-1} \).

→ Exhaustive searches are infeasible!

Cayley’s formula (the number of different trees on \( n \) vertices is \( n^{n-2} \), graphically demonstrated for graphs with 2, 3 and 4 nodes.

(2) Combinatorial assembly module: algorithm

CombDock algorithm uses 2 basic principles:
(1) hierarchical construction of the spanning tree
(2) greedy selection of subtrees

→ 2 subtrees of smaller size (that were previously generated) are connected with an input edge to generate trees with \( i \) vertices

In this way, the common parts of different trees are generated only once.

When connecting subtrees, check whether there are severe penetrations between pairs of subunits that are represented by different subtrees.

(2) Combinatorial assembly module: algorithm

Stage 1: algorithm start with trees of size 1.
Each tree contains a single vertex that represents a subunit.

Stage $i$: the tree complexes that consist of exactly $i$ vertices (subunits) are generated by connecting 2 trees generated at a lower stage with an input edge transformation.

Tree complexes that fulfil the penetration constraint are kept for the next stages.

Because it is impractical to search all valid spanning trees, the algorithm performs a greedy selection of subtrees.

For each subset of vertices, the algorithm keeps only the $D$ best-scoring valid trees that connect them.

The tree score is the sum of its edge weights (pairwise docking scores).

Example: arp2/3 complex

The arp2/3 complex consists of 7 subunits (top).

Shown are only the complexes of the different stages that were relevant to the construction of the third-best scoring solution with RMSD 1.2 Å (bottom).

**Red** edge: transformation of the current stage,

**Blue** edges: transformations of previous stages.

Final scoring

A geometric score evaluates the shape complementarity between the subunits:

- check distances between surface points on adjacent subunits.

- close surface points increase score,

- penetrating surface points decrease score.

Physico-chemical component of the final score counts all surface points that belong to non-polar atoms = this gives an estimate of the hydrophobic effect.

Clustering of solutions

Clustering of solutions:

(1) compute **contact maps** between subunits: array of $N \times (N - 1)$ bins.

If two subunits are in contact within the complex, set the corresponding bit to 1, and to 0 otherwise.

(2) superimpose complexes that have the same contact map and compute RMSD between $C^\alpha$ atoms.

If this distance is less than a threshold, consider complexes as members of a **cluster**.

From each cluster, keep only the complex with the highest score.

## Performance for known complexes

**Table 1. CombDock multimolecular assembly test cases**

<table>
<thead>
<tr>
<th>Target complex (PDB)</th>
<th>Bound/unbound</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nf-kappa-b p65</td>
<td>Bound</td>
<td>3</td>
<td>698</td>
</tr>
<tr>
<td></td>
<td>Unbound</td>
<td>3</td>
<td>698</td>
</tr>
<tr>
<td>Vhl/ElonginC/</td>
<td>Bound</td>
<td>3</td>
<td>328</td>
</tr>
<tr>
<td>ElonginB (1vcb)</td>
<td>Unbound</td>
<td>3</td>
<td>272</td>
</tr>
<tr>
<td>Arp2/3 complex</td>
<td>Bound</td>
<td>7</td>
<td>1709</td>
</tr>
<tr>
<td>(1k8k)</td>
<td>Unbound</td>
<td>7</td>
<td>1728</td>
</tr>
<tr>
<td>(1l0h)</td>
<td>Unbound</td>
<td>10</td>
<td>3576</td>
</tr>
<tr>
<td>MHCII/TCR/Sep3</td>
<td>Unbound</td>
<td>3</td>
<td>1030</td>
</tr>
</tbody>
</table>

SU, subunit; avg., average; the run time refers to the time of the combinatorial assembly module, running on a Linux machine with a 1 GHz single processor. For the unbound cases, the RMSD distances were calculated between all the Cα atoms of the predicted complex and a reference complex that was generated by superimposing the input unbound subunits on the corresponding bound subunits of the determined structure.

---

Examples of large complexes

(a) the best-ranked complex of the 10 subunits of RNA polymerase II, RMSD 1.4 Å.

(b) the third-best scoring assembly of the 7 subunits of the arp2/3 complex, RMSD 1.2 Å.

CombDock is not as successful for docking „unbound“ subunit structures that structurally differ from „bound“ conformations.

Summary

Today:
- Scale-free vs. random graphs
- Examples of important protein complexes
- Exp. methods to determine protein interactions
- Combinatorial assembly of protein complexes (CombDock)

Next lecture V3:
- Further computational methods to assemble higher-order protein complexes
- Docking into EM maps (FFT)