title: IV Protein interactions - sites 2
short title: pp2_ppi_sites2
lecture: Protein Prediction 2 (for Computational Biology) - Protein function
TUM winter semester
Videos:  YouTube / www.rostlab.org

THANKS:

Special lectures:
  • (TBC)

No lecture:
  • 11/07 Department event
  • 11/09 Department event
  • 11/16 Department event
  • 11/23 Thanksgiving
  • 12/19-01/06 - no lecture Xmas

LAST lecture: Jan 18 (followed by 3 x wrap-up sessions)

Examen: TBC: Feb 01 most likely 12:30-14:00

Makeup: TBC: Apr 10 & Apr 12, 2018 - lecture time
PP2 Class 2018/19

Kayalvizi Kaya
Navarasa

Nabil Jabareen

Elisabeth Vrancea

Nathalie Gerstner

Vanessa Schmoll

Amrei Menzel

Daniel Strobl

Corinna Holetschek

Felix Offensperger

Michaela Müller

Issar Arab

© Burkhard Rost
IV. Predict protein interactions
IV.1 protein interactions

Protein-protein interactions (PPI): terminology
Different interfaces = different physics?

HIV gp120 / CD4 / FAB

Protein association

A activates
B activates
C activates
D activates ....

ABCD are associated
Physical interaction NOT association

HIV gp120 / CD4 / FAB

Yeast-2-Hybrid (Y2H) Method

Most common method to obtain binary protein-protein interaction data (Does X bind to Y?)
Original system (GAL4 system) developed by Fields & Song in 1989

Transcription Factor
BD=binding domain
AD=activation domain

BD and AD only function if they are physically linked with each other
IV.2 protein interactions

PPI - homology-based inference
Homology-based inference of PPI

A-B known experimentally
A’-B’ inferred by homology
Can we transfer binding through homology?

Obviously, otherwise no value in model organisms ...

\[ \text{similarity} > X \]

\[ \text{similarity} > X \]

\[ \text{A'} \quad \text{B'} \]

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Homology-based inference for PPI

Worm (C. Elegans)

Worm

Human

more similar

less similar

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Homology-based inference for PPI

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Homology-based inference for PPI, Structure, Localization, EC-numbers

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Homology-based inference of PPI: btw. species?

Typical assumption: corresponding pair of proteins/genes in different species. What about lateral gene transfer?
Homology

Human

Dog

Bird

Whale

© Wikipedia
Genome evolution

duplication

Species C1

Species B1

Species A

Species B2

Species C2
Genome evolution

Orthologs
Paralogs

Species A
Species B1
Species B2
Species C1
duplication
Species C2
Species A
Horizontal gene transfer

The sea slug *Elysia chlorotica* incorporates chloroplasts from the algae that it ingests via a process called kleptoplasty. Photosynthesis continues for up to 12 months using genes within the chloroplast, which are directed by algal nuclear genes that were transferred to the nuclei of the slug.

Inter and Intra-species the same?

similarity > X

Worm

similarity > X

Human

© Sven Mika & Burkhard Rost (Columbia New York)
Worm (C. Elegans)

- worm vs non-worm

Homology-Inference vs HSSP

- less similar
- more similar

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Much better intra-species

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Much better intra-species

S Mika & B Rost 2006 PLoS Genetics, Vol 2, e29
Genome evolution

Orthologs

Paralogs

Species A

Species B1

Species B2

Species C1

Species C2

duplication

© Burkhard Rost
Why?
“Paralogs” conserve interactions
“orthologs” don’t?
Model organisms pose problems for protein-protein interactions
INSERT: measuring the same interaction twice
Measuring the interaction between A-B twice, results in the same interface?
NOT homology-based

\[ \text{similarity} > X \]

Typical assumption: corresponding pair of proteins/gens in different species.
Set up: one PPI = same interface?

- **X-Y** in experiment 1
  - = X-Y
  - in experiment 2

- **X-Y** in family 1
  - = X’-Y’
  - in family 2

---

T Hamp & B Rost 2012 *PLoS Comp Biol* 8:e1002623
Not homology-based inference, but details!

identical

A

interface 1

B

identical

A

interface 2

B

A-B¹ experimental structure 1
A-B² experimental structure 2

interfaces 1 and 2 identical?
Mostly the same but many differ
Many examples for alternative interfaces
IV.3a protein interactions

PPI de novo?
Can we predict PPIs from sequence alone?
Simple method failed fully to do this, problem: too many false positives
Road to predicting protein-protein partners

- Implement simple method to do this failed entirely: too many false positives

- Reduce false positives:
  - predict surface residues (PROFacc, 1999)
  - note: 1/2 of residues -> 1/4 of false positives!
Prediction of solvent accessibility

- 50% of residues somehow accessible to solvent
- 10% not at all
Road to predicting protein-protein partners

- Implement simple method to do this failed entirely: too many false positives

- Reduce false positives:
  - predict surface residues (PROFacc, 1999)
    - note: 1/2 of residues -> 1/4 of false positives!
Road to predicting protein-protein partners

- Implement simple method to do this failed entirely: too many false positives

- Reduce false positives:
  - predict surface residues (PROFacc, 1999)
    note: 1/2 of residues -> 1/4 of false positives!
  - predict residues in external interfaces (InteractionSites, 2004)
Predict protein-protein binding partners

Reducing false positives:

- predict surface residues (PROFacc, 1999)
- predict residues in external interfaces (InteractionSites, 2004)
- predict residues saturated internally (PROFcon, 2004)
- localization (e.g. only all nuclear, LOCtree, 2004)
**Reduction by localization**

<table>
<thead>
<tr>
<th></th>
<th>Extra-cellular</th>
<th>Cytoplasm</th>
<th>Organelles</th>
<th>Mitochondria</th>
<th>Nuclear</th>
<th>TM transmembrane</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extra-cellular</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cytoplasm</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Organelles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mitochondria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nuclear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TM transmembrane</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E.g. nuclear:
=15% nuclear + 15% cytoplasm
=>
- come in with nuclear protein:
  maximally 30% of proteins to test
Predict subcellular localization: LOCtree 2: 18 classes!

T Goldberg, T Hamp & B Rost (2012) submitted k-mer profile kernel SVM
Reducing false positives:

- predict surface residues (PROFacc, 1999)
- predict residues in external interfaces (IntSites, 2004)
- predict residues saturated internally (PROFcon, 2004)
- localization (e.g. only all nuclear, LOCTree, 2004)

- predict residues in protein-substrate interfaces (active)
Predict protein-protein binding partners

Reducing false positives:

- ✔ predict surface residues (PROFacc, 1999)
- ✔ predict residues in external interfaces (InteractionSites, 2004)
- ✔ predict residues saturated internally (PROFcon, 2004)
- ✔ localization (e.g. only all nuclear, LOCtree, 2004)
- ❌ predict residues in protein-substrate interfaces (active)
- ✔ predict protein domains/improve alignments
Predict protein-protein binding partners

Reducing false positives:
- predict surface residues (PROFacc, 1999)
- predict residues in external interfaces (Interaction Sites, 2004)
- predict residues saturated internally (PROFcon, 2004)
- localization (e.g. only all nuclear, LOCtree)
- predict residues in protein-substrate interfaces (active)
- predict protein domains/improve alignments

Put it all together and predict binding partners
IV.4 protein interactions

PPI - data collection
Different interfaces = different physics?

At least 6 types of interfaces differ in sequence!

Internal (inter-domain and intra-domain)
External homomers (permanent/transient)
External heteromers (permanent/transient)

Molecules of experimentally determined structure (3D co-ordinates)

www.pdb.org
  • check out: Molecule of the Month

Stat 2010/04:
  • ~65,000 structures
  • 60K proteins
  • 2K DNA/RNA
  • 3K complexes
  • 56K X-ray
  • 8K NMR
  • 0.3K Electron microscopy
extract interactions how?
sequence unique
Different interfaces = different physics?

HIV gp120 / CD4 / FAB

Remove redundancy

All proteins

CD-Hit / UniqueProt, e.g. 70% PIDE
What does “in contact” mean?
Different interfaces, different physics?

At least 6 types of interfaces differ in sequence!

- Internal (inter-domain and intra-domain)
- External homomers (permanent/transient)
- External heteromers (permanent/transient)


Structure: Hendrickson lab
Develop method

1. PDB->Unique
2. parse heavy atoms <6.5 Ångstrøm (0.65 nm)
Each atom in the PDB belongs to a residue
Each residue belongs to a chain
Chains may have “breaks”
Map PDB to annotations

BLAST 10-10, >90% PIDE over >90% of length

chain A maps to SPa, chain B maps to SPb
if (SPa=SPb) assume A and B from same protein
else A and B from two different proteins
Develop method

- 1. parse heavy atoms <6.5 Ångström (0.65 nm)
- 2. map chains to SWISS-PROT, distinguish transient protein-protein interactions from others
- 3. PDB sub(PP)->Unique

NOW we have a data set and can apply machine learning
PPI interfaces use local segments

INSERT:
Problem with choosing thresholds:
protein flexibility prediction
PROFbval
Flexibility of proteins

superposition of 44 hen-white lysozyme structures

© Wikipedia
Danielkeedy

© Burkhard Rost
ROSTLAB. TUM
Backbone flexibility: B-value

![Graph showing frequency of B-values for proteins at different resolutions.](image)

- Black line: resolution $< 2.5$ (1513 proteins)
- Dark grey line: resolution $< 2$ (926 proteins)
- Light grey line: resolution $< 1.5$ (142 proteins)

A Schlessinger & B Rost *Proteins* 2005 61: 115-126

© Burkhard Rost

[Image: Back cover of ROSTLAB]
Backbone flexibility: B-value

where to threshold?

A Schlessinger & B Rost 2005 *Proteins* 61: 115-126
Backbone flexibility: B-value

The graph shows the distribution of normalized B-values for proteins, categorized by resolution. The x-axis represents the normalized B-value, while the y-axis shows the frequency. Three curves are depicted, each representing different resolution ranges:

- Green line: resolution < 2.5 (1513 proteins)
- Black line: resolution < 2 (926 proteins)
- Gray line: resolution < 1.5 (142 proteins)

The graph highlights the frequency of B-values across different levels of rigidity, with labels indicating increasing rigidity and increasing flexibility.
B-values imprinted onto sequence

![Graph showing log(frequency) vs amino acids with categories: Rigid, Intermediate, Flexible]
PROFbval reliability correlates with accuracy

non-strict mode
1 if $B_{\text{norm}} \geq 0.03$
0 else

strict mode
1 if $B_{\text{norm}} \geq -0.30$
0 else
PROFbval: predict flexibility/rigidity

© COVER of Proteins

[Diagram showing a model of proteins with colors indicating flexibility (red) and rigidity (blue).]

red = flexible  blue = rigid

A Schlessinger & B Rost 2005 Proteins 61: 115-126
PROFbval somehow separates active sites

A Schlessinger & B Rost unpublished
Backbone flexibility: B-value

A Schlessinger & B Rost 2005 *Proteins* 61: 115-126
differences in compositions
Different interfaces = different physics?

HIV gp120 / CD4 / FAB

Interface types differ in composition

They obviously differ! But, are these differences meaningful?
How to answer the *meaningful* question?
Are these differences statistically significant?

☑ Chi-square test:
  known problem: small data sets
  here millions of points

☑ all differences $< 10^{-300}$
  -> SIGNIFICANT

☑ ... unfortunately also:
  proteins [a-b] vs [c-d]
  1 vs 2 authors
  random subsets ...
Find-self test (statistical significance)

- procedure for P1:
  - randomly draw S

- report pair with minimal JS

- perform procedure for P2 and P3

Y Ofran & B Rost 2005 unpublished
# Find-self test on six types of interfaces

<table>
<thead>
<tr>
<th></th>
<th>internal</th>
<th>domain-domain</th>
<th>homo-obligomer</th>
<th>homo-oligomer</th>
<th>hetero-obligomer</th>
<th>hetero-oligomer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>909</td>
<td>65</td>
<td>3</td>
<td>-</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>812</td>
<td>2</td>
<td>-</td>
<td>128</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>2</td>
<td>925</td>
<td>-</td>
<td>12</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>130</td>
<td>7</td>
<td>-</td>
<td>811</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>8</td>
<td>58</td>
<td>-</td>
<td>38</td>
<td>896</td>
</tr>
</tbody>
</table>

Different contact preferences

- **Internal**
  - Intra-domain
  - Domain-domain

- **External**
  - Homomers (permanent): more often than expected
  - Heteromers (transient): less often than expected

Y Ofran & B Rost (2003) *J Mol Biol* **325**, 377-87; Fig. 2
Different contact preferences

intra-domain

transient hetero-oligomers

red more often than expected
blue less often than expected

Y Ofran & B Rost (2003) J Mol Biol 325, 377-87; Fig. 2
Different interfaces = different physics!

At least 6 types of interfaces differ in sequence!
- Internal (inter-domain and intra-domain)
- External homomers (permanent/transient)
- External heteromers (permanent/transient)

Machine learning
how to choose the input features?
ask your friend (ideally in the group)
Strength of prediction reflects reliability?
More complex system to predict structure

Sequence → PSI-BLAST → Filter

PROFsec

PROFacc

PHD→PROF

SPLIT again

JURY over 20

secondary structure solvent accessibility
Alignment information

<table>
<thead>
<tr>
<th>Protein</th>
<th>Alignments</th>
<th>Profile table</th>
</tr>
</thead>
<tbody>
<tr>
<td>: : : :</td>
<td>GSAPD NTEKQ CVHIR LMYFW</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>GGGG</td>
<td>5 . . . . . . .</td>
</tr>
<tr>
<td>Y</td>
<td>YYYY</td>
<td>. . 2 . 3 . . .</td>
</tr>
<tr>
<td>I</td>
<td>IEE</td>
<td>. . . . 5 . . .</td>
</tr>
<tr>
<td>Y</td>
<td>YYYY</td>
<td>. . . . 5 . . .</td>
</tr>
<tr>
<td>D</td>
<td>DDDD</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>P</td>
<td>PPPP</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>E</td>
<td>EA A</td>
<td>. 3 . 2 . . . .</td>
</tr>
<tr>
<td>D</td>
<td>VVEE</td>
<td>. 1 . 2 . 2 . .</td>
</tr>
<tr>
<td>G</td>
<td>GGGG</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>D</td>
<td>DDDD</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>P</td>
<td>PPPP</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>D</td>
<td>DTDD</td>
<td>. 4 . 1 . . . .</td>
</tr>
<tr>
<td>D</td>
<td>NQNN</td>
<td>. 1 . 3 . 1 . .</td>
</tr>
<tr>
<td>G</td>
<td>GNGG</td>
<td>. 4 . 1 . . . .</td>
</tr>
<tr>
<td>V</td>
<td>VIVV</td>
<td>. . . . 4 . . .</td>
</tr>
<tr>
<td>N</td>
<td>EPPK</td>
<td>. 1 . 1 . 2 . .</td>
</tr>
<tr>
<td>P</td>
<td>PPPP</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>G</td>
<td>GGGG</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>T</td>
<td>TTTT</td>
<td>. 5 . . . . . .</td>
</tr>
<tr>
<td>D</td>
<td>EKSA</td>
<td>. 1 . 1 . 1 . .</td>
</tr>
<tr>
<td>F</td>
<td>FFFF</td>
<td>. . . . 5 . . .</td>
</tr>
<tr>
<td>:</td>
<td>: : : :</td>
<td></td>
</tr>
</tbody>
</table>

corresponds to the the 21*3 bits coding for the profile of one residue

B Rost & C Sander (1993) PNAS 90:7558-62
What it takes to realize this as a server

PDB SWISS-PROT TrEMBL HSSP

Sequence ➔ PSI-BLAST Filter
Much more complex system for function

<table>
<thead>
<tr>
<th>Database</th>
<th>Institution</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB</td>
<td>UCSD</td>
<td>daily</td>
</tr>
<tr>
<td>SWISS-PROT</td>
<td>Geneva</td>
<td>weekly</td>
</tr>
<tr>
<td>TrEMBL</td>
<td>EBI Hinxton UK</td>
<td>weekly</td>
</tr>
<tr>
<td>HSSP</td>
<td>EBI</td>
<td>monthly</td>
</tr>
<tr>
<td>FSSP</td>
<td>EBI</td>
<td>monthly</td>
</tr>
<tr>
<td>DIP</td>
<td>UCLA</td>
<td>manually</td>
</tr>
<tr>
<td>Pfam</td>
<td>Sanger Hinxton UK</td>
<td>manually</td>
</tr>
<tr>
<td>EVAdb</td>
<td>internal</td>
<td>daily</td>
</tr>
<tr>
<td>BIG</td>
<td>internal</td>
<td>weekly</td>
</tr>
<tr>
<td>PEP</td>
<td>internal</td>
<td>weekly</td>
</tr>
</tbody>
</table>
Few features

- Profile
- Predicted 1D structure
  - Secondary structure
  - Solvent accessibility
  - Membrane regions
  - Disorder
- Predicted aspects of function
Let neural networks figure it out ...
Random split not enough
avoid overlap
training/cross-training
vs. testing
IV.5 protein interactions

PPI predict binding sites
Performance in predicting PPI sites
Method: neural network
Strength of prediction reflects reliability?

**strong**

0.9

**weak**

0.6

0.4
PP interfaces predicted from sequence

Y Ofran & B Rost 2007 Bioinformatics e13-16
PP interfaces predicted from sequence

Accuracy:
>94% for 1 in 10
>70% for 2 in 10

Successful prediction: skp1-skp2

Uniquitin ligase skp1-skp2 complex

Green: 2 correctly predicted residues
(pocket binding TRP109 of SKP-2 F-box protein)

Accuracy:
>94% for 1 in 10
>70% for 2 in 10
Prediction system

- **Level 1:**
  - Neural networks
  - **input:** alignment profile/predicted secondary structure + accessibility (PROF)/predicted sequence complexity/overall features (protein length, amino acid composition, asf.)
  - **output:** 2 units: is or is not P=P

- **Level 2:**
  - Neural networks using input from previous level

- **Level 3:**
  - simple clustering
PPI interfaces use local segments

PP interfaces predicted from sequence

Y Ofran & B Rost 2007 Bioinformatics e13-16
### Alignment information

<table>
<thead>
<tr>
<th>Protein</th>
<th>Alignments</th>
<th>profile table</th>
</tr>
</thead>
<tbody>
<tr>
<td>: : : :</td>
<td>GSAPD_NTEKQ CVHIR LMYFW</td>
<td></td>
</tr>
<tr>
<td>G G G G</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>Y Y Y Y</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>I I E E</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>Y Y Y Y</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
<tr>
<td>D D D D</td>
<td>5 . . . . . . . . . . . . 5 .</td>
<td></td>
</tr>
</tbody>
</table>

- The red dot corresponds to the 21×3 bits coding for the profile of one residue.

PP interfaces predicted from sequence
PPI hot spots?
Interaction HOT SPOTS

- residues that are essential for protein-protein interactions

- operational:
  - 1. residue in the interface
  - 2. mutation of the residue knocks out interaction
PP interfaces predicted from sequence

Very strong = hot spots?
Prediction of hot spots for CD4

- alanine scan for V1 domain of CD4 (bound to gp120) (A Ashkenazi et al. & DJ Capon (1990) *PNAS* 87, 7150)
  - red: observed
  - purple: predicted


Y Ofran & B Rost 2007 *PLoS CB* 3:e119
enough to publish?
Hot spots reliably predicted from sequence!

hottest of hot = no error!

worst: ~60% right

Accuracy/Specificity (correctly predicted/predicted)

Coverage/Sensitivity (correctly predicted/observed)


© Burkhard Rost ROSTLAB. TUM
What makes it work?

- Evolutionary information:
  - Optimally choosing profile
  - Explicitly using conserved residues

- (Predicted) 1D Structure
  important: good prediction + used correctly
  - Surface residues
  - Secondary structure

- Mark low-complexity and sticky

- Filtering “isolated predictions”
Hot spots prediction requires full information

- Sequence + Evolution + Exp. structure: 89
- Sequence + Evolution + Pred. structure: 85
- Evolution only: 36
- Sequence only: 35
- Hydrophobic Moment: 12

Functionally important residues - interactions sites

..LNDRA. ➔ ..LNDRA. ➔ ..- -P-. 


© Marco Punta & Yanay Ofran & Burkhard Rost (Columbia New York)
Find non-homologous competitive binder
IV.6 protein interactions

PPI - hubs
Network level distribution of PPIs
Will all proteins have a similar number of interactions on average, or will have some more than others?
What do you expect of the following?

Number of PPIs of one protein

Histogram

(number of proteins with that number of PPIs)

Number of PPIs of one protein
Which distribution do you expect?
If you plotted the histogram of settlement sizes, how would that look?
How to answer the question?
Sizes of metropolitan areas in the USA


FIGURE I
Log Size versus Log Rank of the 135 largest U. S. Metropolitan Areas in 1991
Source: Statistical Abstract of the United States [1993].
Zipf’s law

$y = \frac{1}{x}$
What do you expect of the following?

Number of PPIs of one protein

Histogram

(number of proteins with that number of PPIs)

Number of PPIs of one protein
Pick points at random: then what?
Half a Zipf is a Zipf
Connect micro- and macro-level

**Macro level:** networks
UP: more partners

**Micro level:** residues
RIGHT: more hotspots
Hubs: promiscuous proteins

Date/Party hubs
Notation introduced by Marc Vidal
JD Han et al. & M Vidal 2004 *Nature* 430:88-93

- **Date hubs** interactions at *different times/same location*?
- **Party hubs** interactions at *same time/different location*
More hotspots -> more party-hub like!

Y Ofran, A Schlessinger & B Rost submitted
More unstructured -> more date-hub like!

macro: more partners
micro: more hotspots

- Non-hubs
- Party hubs
- Date hubs

Y Ofran, A Schlessinger & B Rost submitted
Examples for Date & Party hubs

FUS3 MAP kinase - date hub (PDB 2b9f)
right complex with MSG5 binding motif (light blue)

ABC10-beta subunit of RNA polymerase - party hub (PDB 1r9sJ)
right: RNA Polymerase II elongation complex (ABC10-beta in red)
**preliminary Lecture plan (PP2 function)**

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>01: 2018/10/16</td>
<td>No lecture (makeup examen; PP last year)</td>
</tr>
<tr>
<td>02: 2018/10/18</td>
<td>No lecture (makeup)</td>
</tr>
<tr>
<td>03: 2018/10/23</td>
<td>Welcome: who we are</td>
</tr>
<tr>
<td>04: 2018/10/25</td>
<td>Intro function 1: concept of protein function</td>
</tr>
<tr>
<td>05: 2018/10/30</td>
<td>No lecture</td>
</tr>
<tr>
<td>06: 2018/11/01</td>
<td>No lecture (holiday, All Saints)</td>
</tr>
<tr>
<td>07: 2018/11/06</td>
<td>Intro function 2: contd.</td>
</tr>
<tr>
<td>08: 2018/11/08</td>
<td>Localization 1 (chalk talk)</td>
</tr>
<tr>
<td>09: 2018/11/13</td>
<td>No lecture (SVV)</td>
</tr>
<tr>
<td>10: 2018/11/15</td>
<td>Localization 2 (homology, motifs)</td>
</tr>
<tr>
<td>11: 2018/11/20</td>
<td>Localization 3 (motifs, machine learning)</td>
</tr>
<tr>
<td>12: 2018/11/22</td>
<td>No lecture (Thanksgiving)</td>
</tr>
<tr>
<td>14: 2018/11/29</td>
<td>Localization 5 (machine learning 2)</td>
</tr>
<tr>
<td>15: 2018/12/04</td>
<td>Localization 6</td>
</tr>
<tr>
<td>16: 2018/12/06</td>
<td>No lecture (<em>Dies Academicus</em>)</td>
</tr>
<tr>
<td>17: 2018/12/11</td>
<td>PPI 1 - sites / pairing (chalk)</td>
</tr>
<tr>
<td>18: 2018/12/13</td>
<td>PPI 2 - sites / PPI pairing (chalk)</td>
</tr>
<tr>
<td>19: 2018/12/18</td>
<td>PPI 3 - sites / DNA / RNA (Jia Jun Qiu)</td>
</tr>
<tr>
<td>20: 2018/12/20</td>
<td>No lecture</td>
</tr>
<tr>
<td>21-24:</td>
<td>no lectures - winter break (2018/12/24 - 2019/01/06)</td>
</tr>
<tr>
<td>25: 2019/01/08</td>
<td>PPI 4 - PPI pairing</td>
</tr>
<tr>
<td>28: 2019/01/10</td>
<td>PPI 5 - PPI pairing 2</td>
</tr>
<tr>
<td>29: 2019/01/15</td>
<td>SNP effect 1 (chalk talk)</td>
</tr>
<tr>
<td>30: 2019/01/17</td>
<td>SNP effect 2</td>
</tr>
<tr>
<td>31: 2019/01/22</td>
<td>SNP effect 3</td>
</tr>
<tr>
<td>32: 2019/01/24</td>
<td>WRAP up 1</td>
</tr>
<tr>
<td>33: 2019/01/29</td>
<td>WRAP up 2</td>
</tr>
<tr>
<td>34: 2019/01/31</td>
<td>?</td>
</tr>
<tr>
<td>35: 2019/02/05</td>
<td>?</td>
</tr>
<tr>
<td>36: 2019/02/07</td>
<td>Examen (10:00-13:00, lecture room LMU physics)</td>
</tr>
</tbody>
</table>
CONTINUE HERE