Protein-protein interactions: evolution, prediction and regulation

Anna Panchenko,

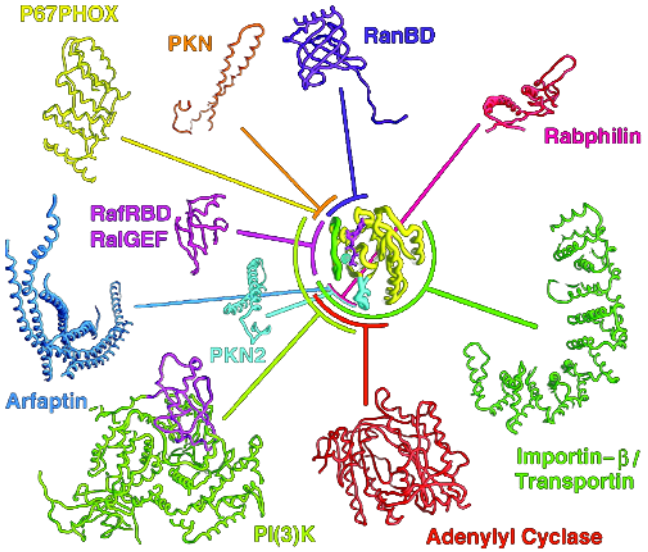
National Center for Biotechnology Information, NIH, USA

Outline

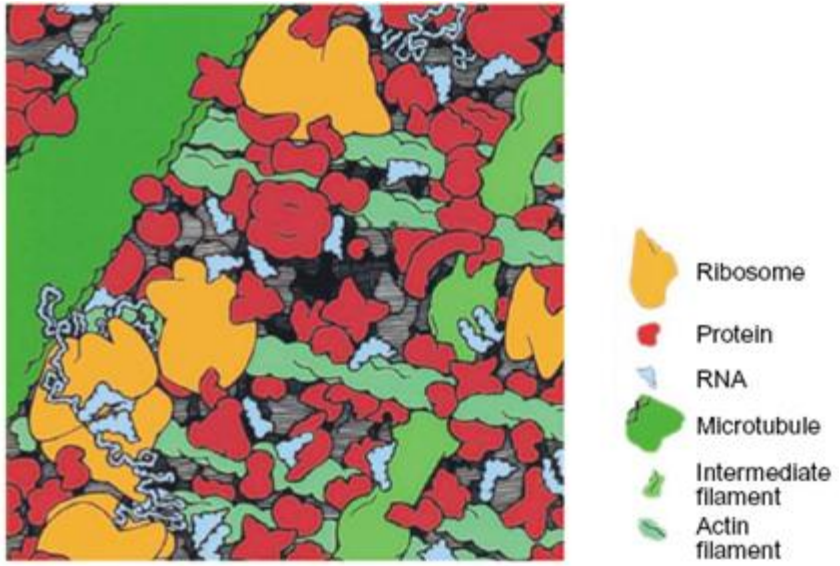
- Types of protein-protein interactions and their role in cell function.
- Physico-chemical properties of complexes and interfaces, binding hot spots.
- Experimental methods to identify interactions.
- Computational methods to predict PPIs.
- Evolution of protein interactions.
- Regulation of protein-protein binding.

Proteins function while interacting with other partners

- Many cellular processes are regulated through protein-protein interactions, distortions may cause diseases
- Proteins provide specific binding interfaces to interact with ligands.
- Binding selectivity and affinity is determined by physico-chemical properties of binding interfaces.
- Binding interfaces share common properties: conservation of certain amino acids, hot spots, geometry.



Vetter & Wittinghofer, Science 2001



Different types of protein-protein interactions.

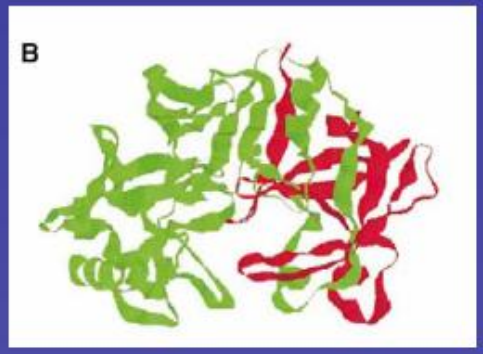
- Permanent/obligatory – subunits might not be stable in isolation and transient – subunits might fold independently.
- External are between different chains; internal are within the same chain.
- Homo- and hetero-oligomers depending on the similarity between interacting subunits.

Types of protein-protein interactions (PPI)

Obligate PPI

usually permanent

the protomers are **not** found as stable structures on their own *in vivo*



Obligate heterodimer

Human cathepsin D

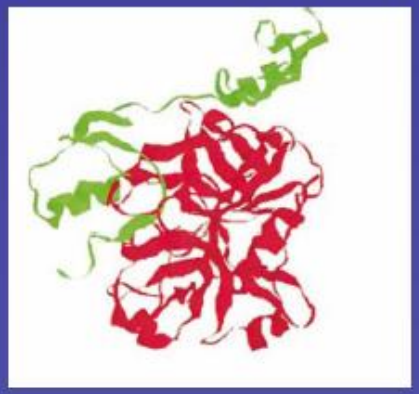
Non-obligate PPI

Permanent

(many enzyme-inhibitor complexes)

dissociation constant K_d=[A][B] / [AB]

10^{-7} - 10^{-13} M



Non-obligate permanent heterodimer

Thrombin and rodniin inhibitor

Transient

Weak

(electron transport complexes)

K_d mM-μM

Non-obligate transient homodimer, Sperm lysin (interaction is broken and formed continuously)

Intermediate

(antibody-antigen, TCR-MHC-peptide, signal transduction PPI), K_d μM-nM

Strong

(require a molecular trigger to shift the oligomeric equilibrium)

K_d nM-fM

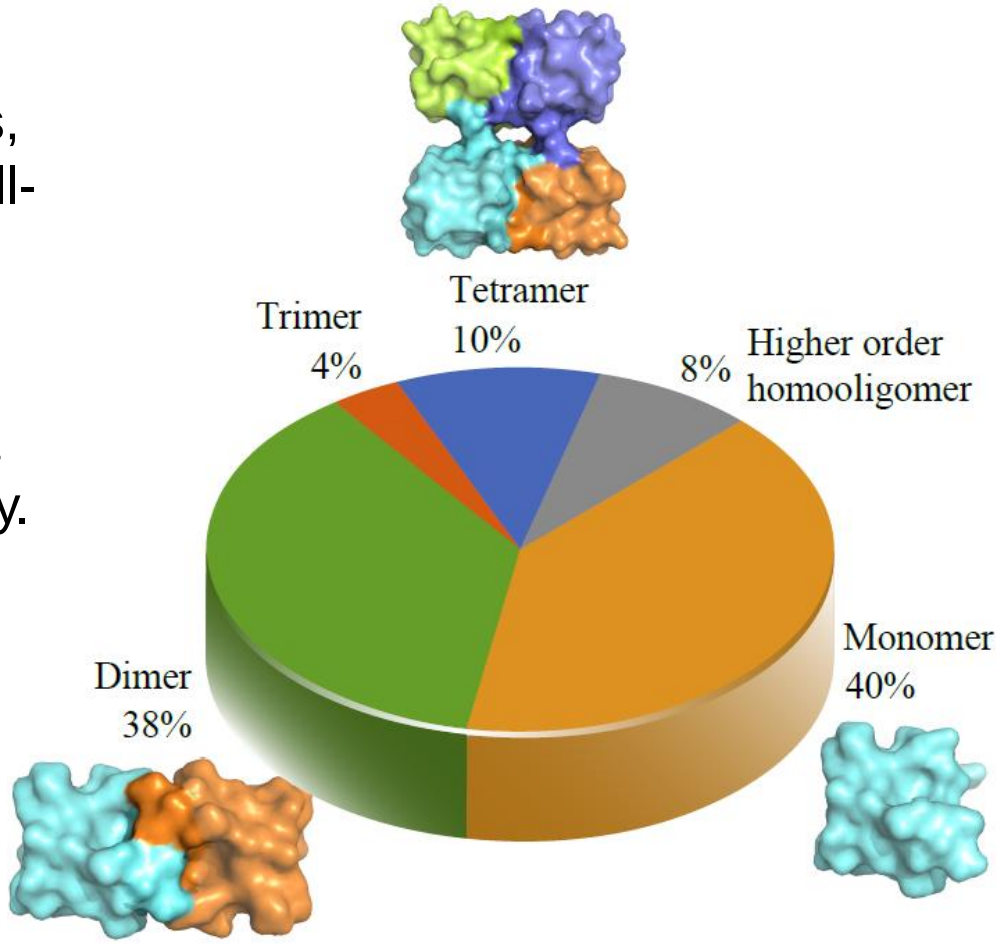
Bovine G protein dissociates into Gα and Gβγ subunits upon GTP, but forms a stable trimer upon GDP

Role of homooligomers in a cell

• Mediate and regulate gene expression, activity of enzymes, ion channels, receptors and cell-cell adhesion processes.

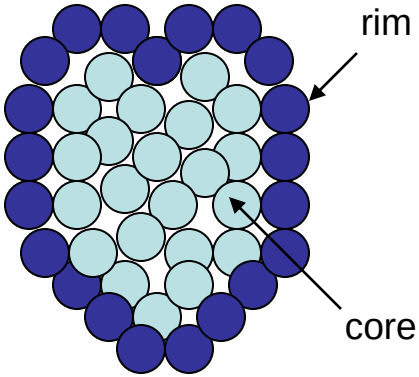
• Provide sites for allosteric regulation, new binding sites at interfaces to increase specificity.

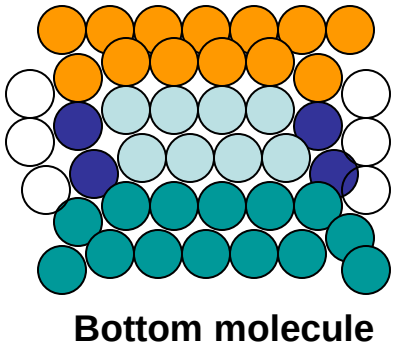
• Provides stability, protection against denaturation.



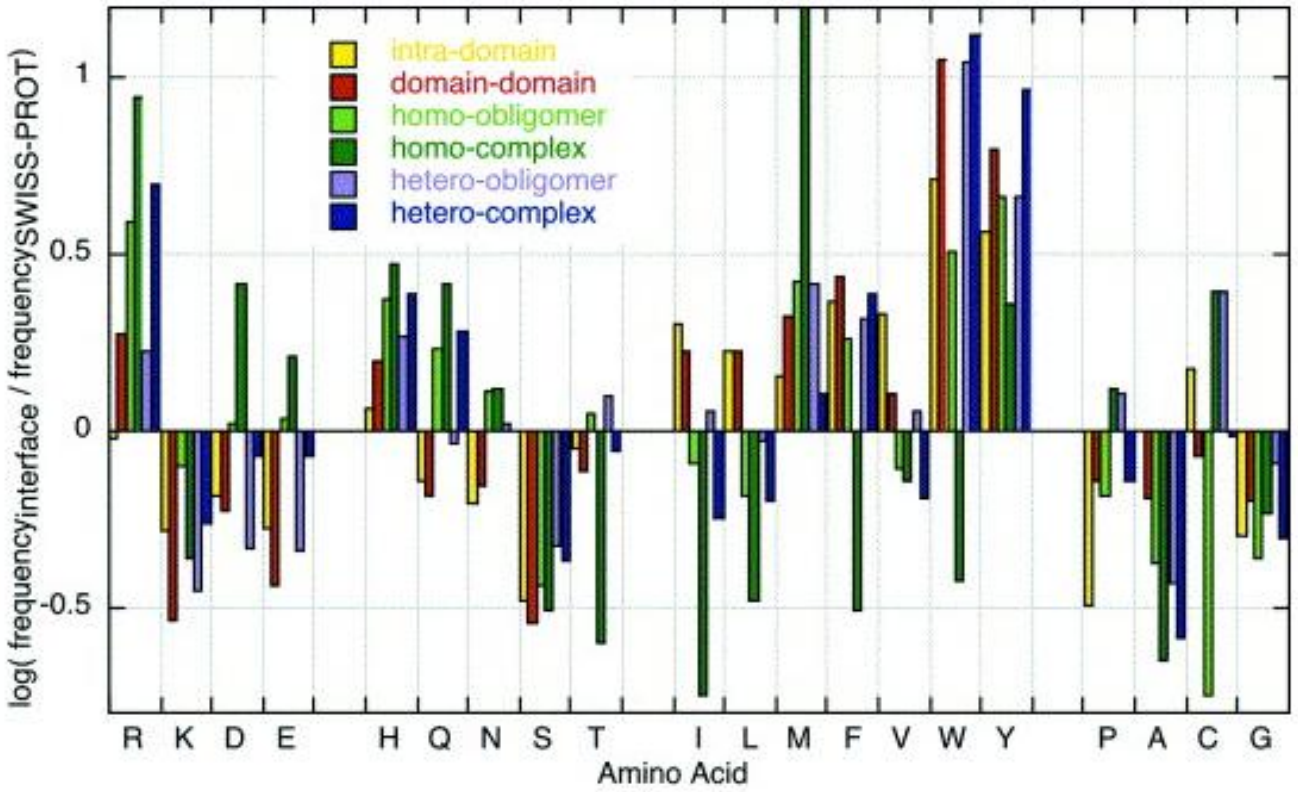
Common properties of protein-protein interactions.

- Majority of protein complexes have a buried surface area ~1600±400 Å^2 (“standard size” patch).
- Complexes of “standard size” do not involve large conformational changes while large complexes do.
- Protein recognition site consists of a completely buried core and a partially accessible rim.





Amino acid composition of different types of complexes



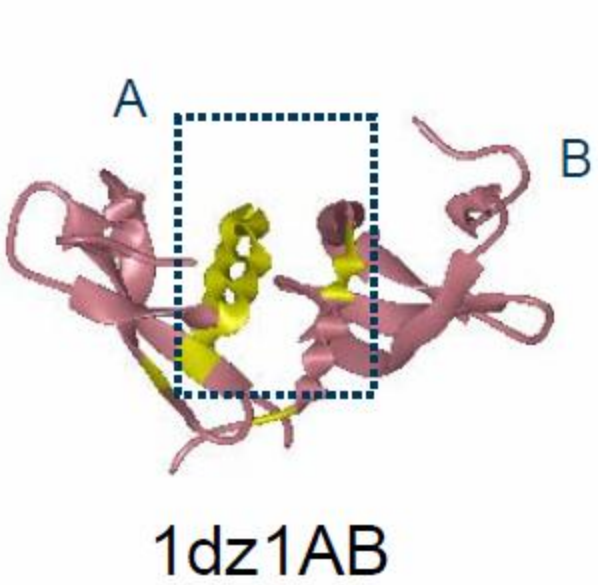
Ofran & Rost, JMB, 2003

Properties of different types of interfaces

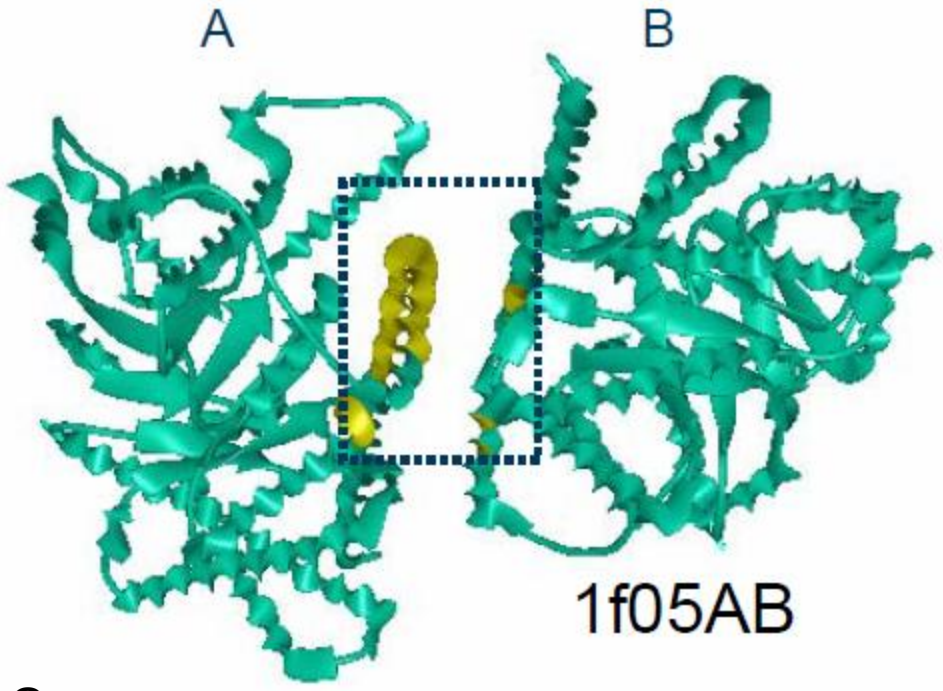
- Non-obligate complexes tend to be more hydrophilic.
- Hydrophobic groups tend to be burried upon complex formation.
- Electrostatics, hydrogen bonds, salt bridges confer specificity.
- Permanent interfaces tend to be larger, less planar, and tightly packed.

Classification of interfaces

Similar interfaces- dissimilar functions



Chromatin, Mouse hp1 (m31) C terminal domain

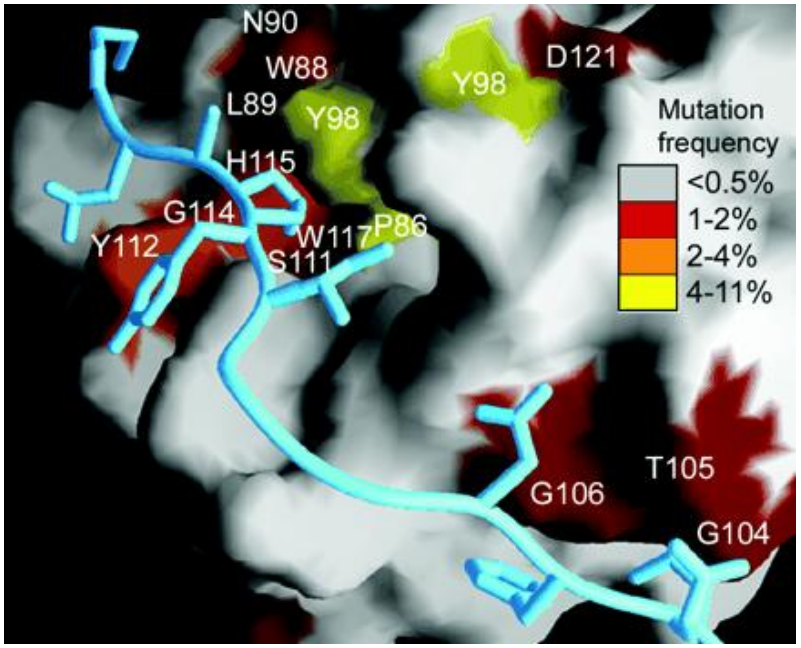


Human transaldolase

Keskin, Gursoy, Nussinov, PRISM

Binding hot spots

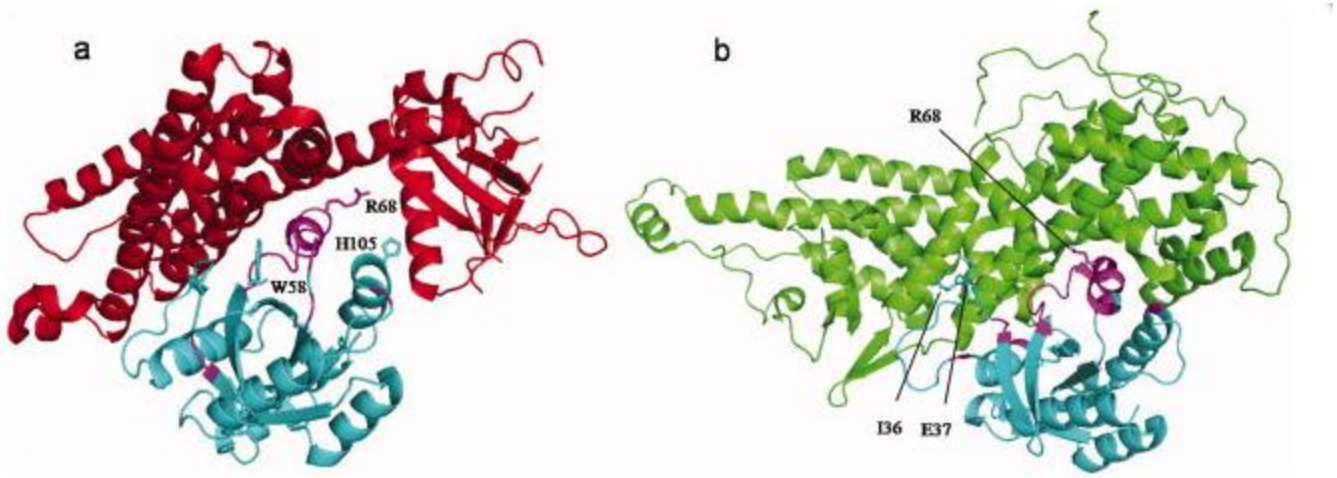
- Interface sites which contribute the most to binding energy (>2kcal/mol).
- Amino Acid composition: aromatic, Thr, Ser, Cys.
- Structurally and sequence conserved



Why do we need to identify binding hotspots?

- To understand how proteins bind to different partners – “binding promiscuity”

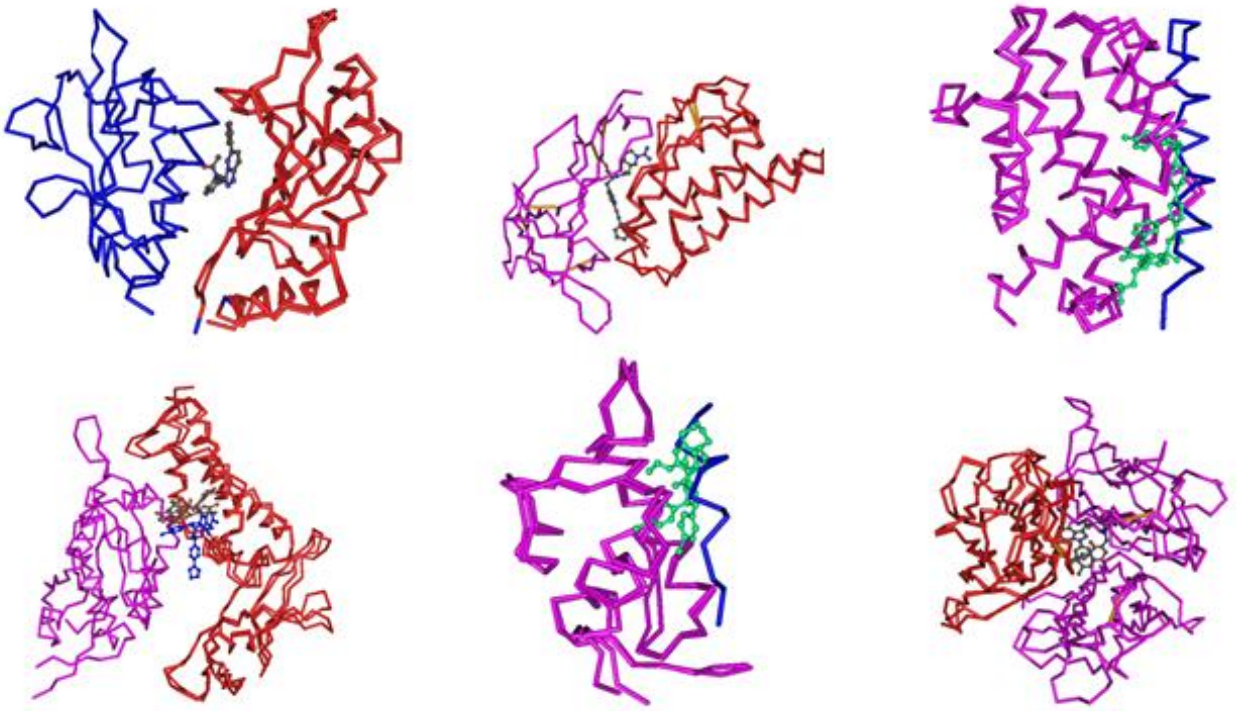
Interaction between GTPase domain and GEF



Tyagi et al, Protein Science 2009

Why do we need to identify binding hotspots?

- To target protein-protein interfaces by small molecule drugs

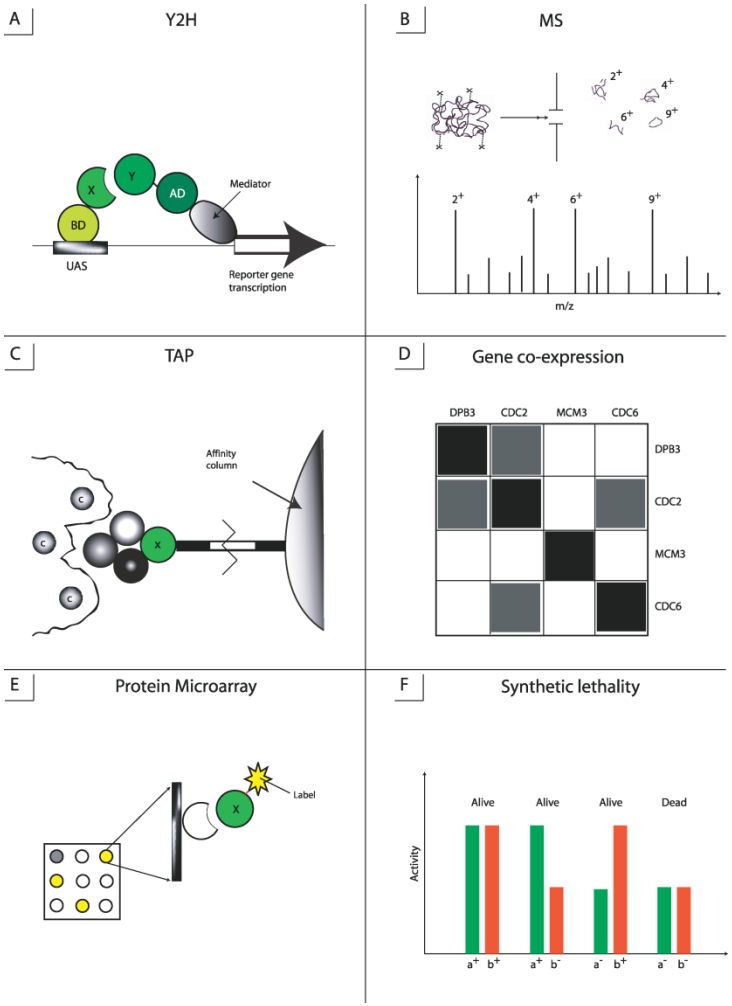


High-throughput methods to detect protein-protein interactions

Yeast two hybrid

Tandem Affinity Purification

Protein Microarray



Mass spectroscopy

Gene co-expression

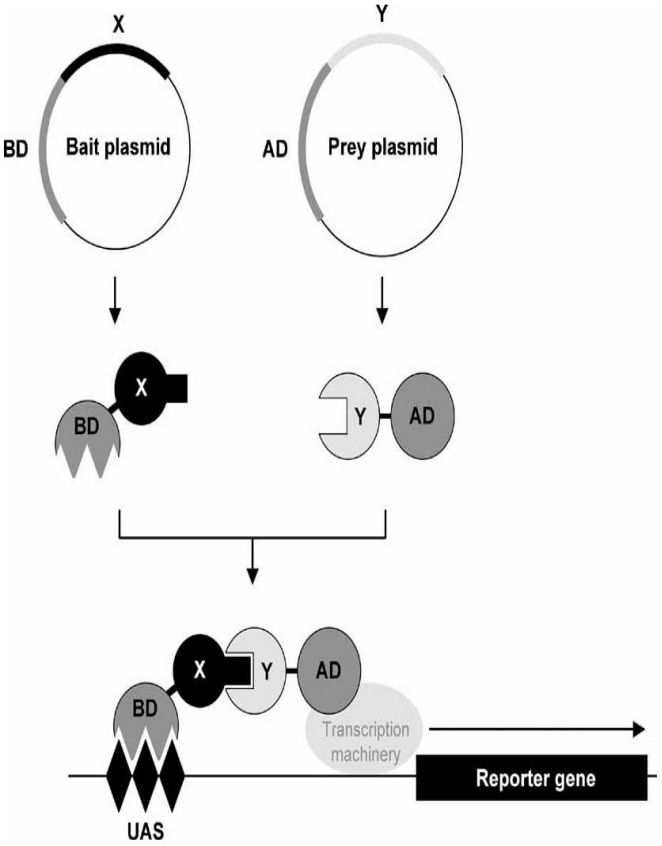
Synthetic lethality

Shoemaker & Panchenko, PloS Comp Biol, 2007

Yeast two-hybrid experiments.

- Many transcription factors (ex: Gal4, LexA) have two distinct domains; one that directs binding to a promoter DNA sequence (BD) and another that activates transcription (AD).

- Fields and Song (1989) demonstrated that DNA-binding domain can not activate transcription at a promoter unless physically (not necessarily covalently) associated with an activating domain.



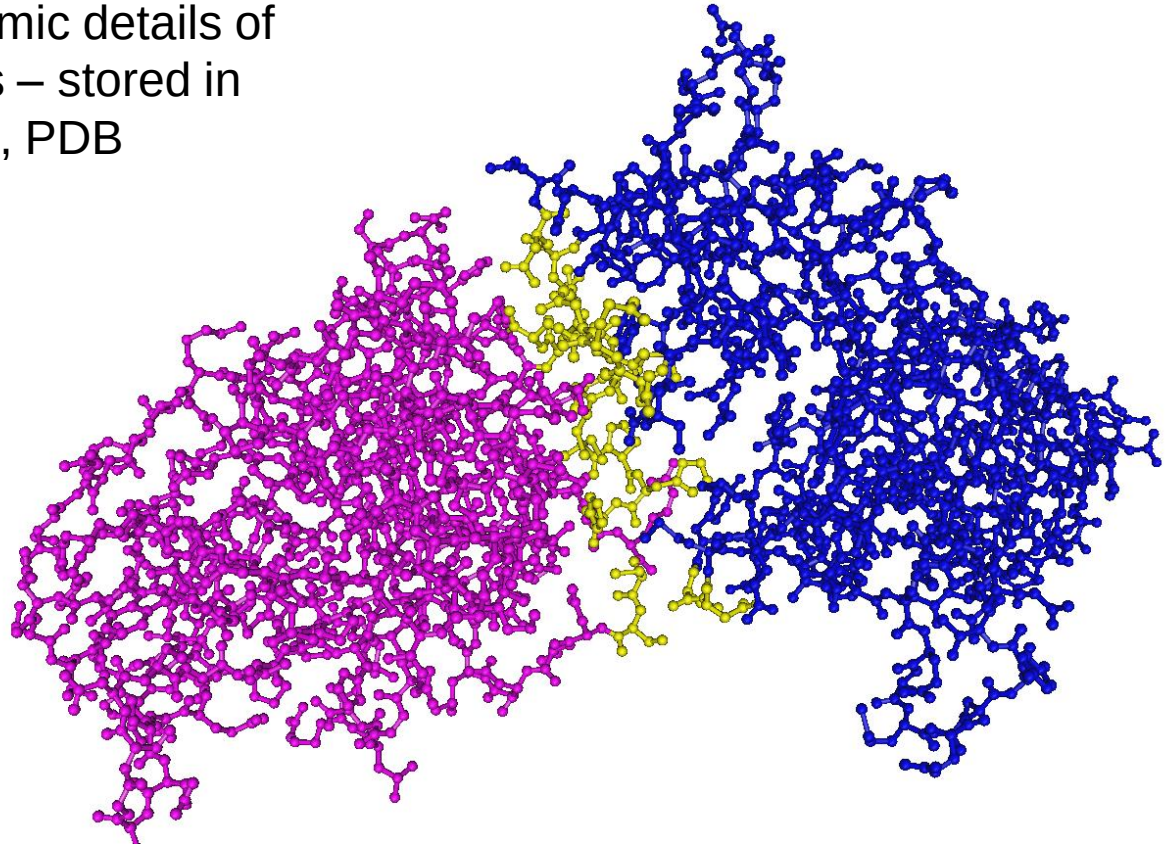
Causier, Mass Spectroscopy Reviews, 2004

Low-throuput biophysical methods

- X-ray crystallography, NMR
- Fluorescence resonance energy transfer (FRET)
- Surface plasmon resonance (SPR)
- Isothermal titration calorimetry (ITC)
- Atomic force microscopy

Resolving atomic details of interaction interfaces

X-ray, NMR – atomic details of binding interfaces – stored in Protein Databank, PDB



Prediction of protein-protein interactions

Methods of prediction of functional associations and protein interactions

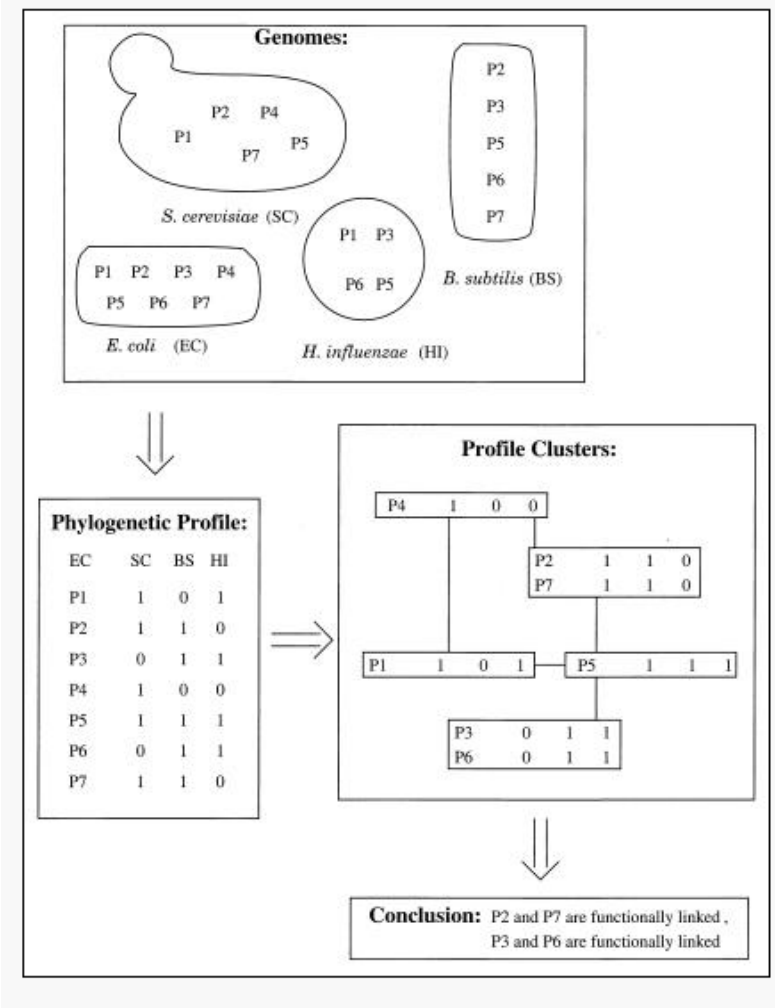
Method Name	Protein/Domain Interaction	Physical Interaction/ Functional Association
Gene co-expression	P	F
Synthetic lethality	P	F
Gene cluster and gene neighbor	P	F
Phylogenetic profile	P, D	F
Rosetta Stone	P	F
Sequence co-evolution	P, D	F
Classification	P, D	P
Integrative	P, D	P
Domain association	D	P
Bayesian networks	P, D	F, P
Domain pair exclusion	D	P
p-Value	D	P

Phylogenetic profile method.

Functionally linked and putative interacting proteins should have orthologs in the same subset of fully sequenced organisms (Pellegrini et al, *PNAS* 1999).

Drawbacks:

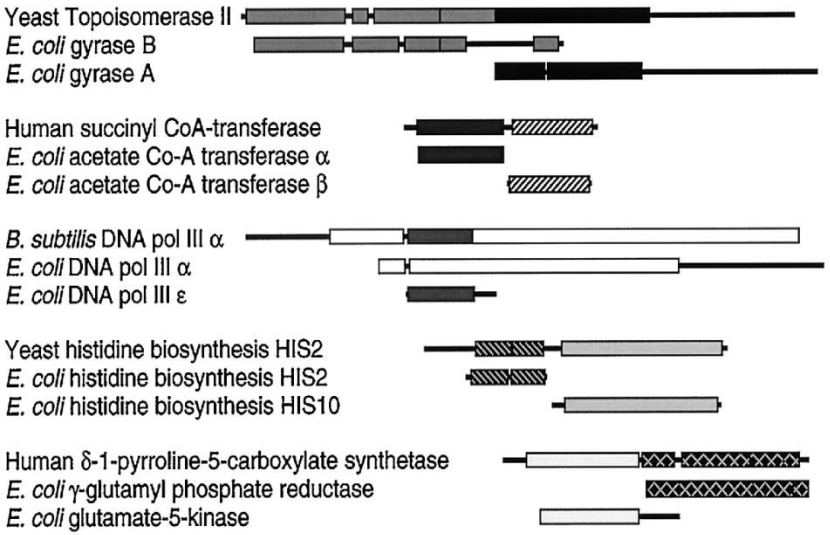
- high computational cost;
- dependence on homology detection between distant organisms;
- ubiquitous unlinked proteins present in all genomes – false positives;
- shared phylogenetic history between two proteins – false positives.



Rosetta Stone approach.

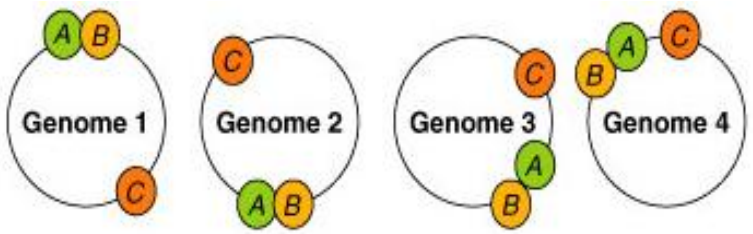
• Some pairs of interacting domains have homologs which are fused into one protein chain – “Rosetta Stone” protein (Marcotte et al, *Science*, 1999).

• In *E.coli* ~ 6809 pairs of non-homologous proteins; both proteins from each pair could be mapped to a single protein from some other genome.



Gene neighborhood method.

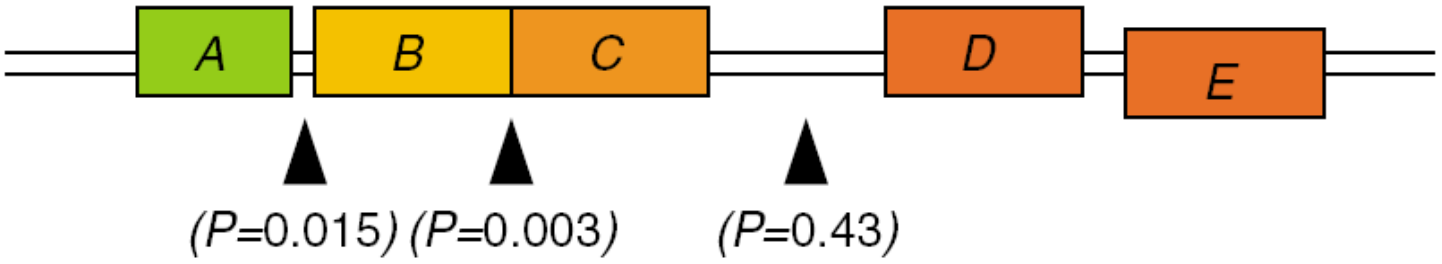
- Gene pairs from conserved gene clusters encode proteins which are functionally related and possibly interact.
- Conservation of gene order can be used to predict gene function.
- Analysis of gene order conservation : 65%–75% of co-regulated genes interact physically (prediction of archael exosome by comparing GN in archaea, *Koonin et al, Genome Res 2001)*



Bowers et al, Genome Biology, 2004

Gene cluster method.

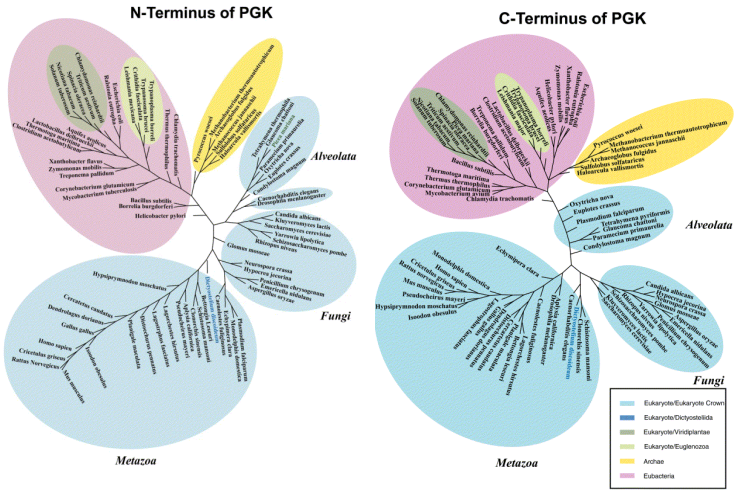
- Bacterial genes of related function are often transcribed simultaneously – operon.
- Identification of operons is based on intergenic distances.

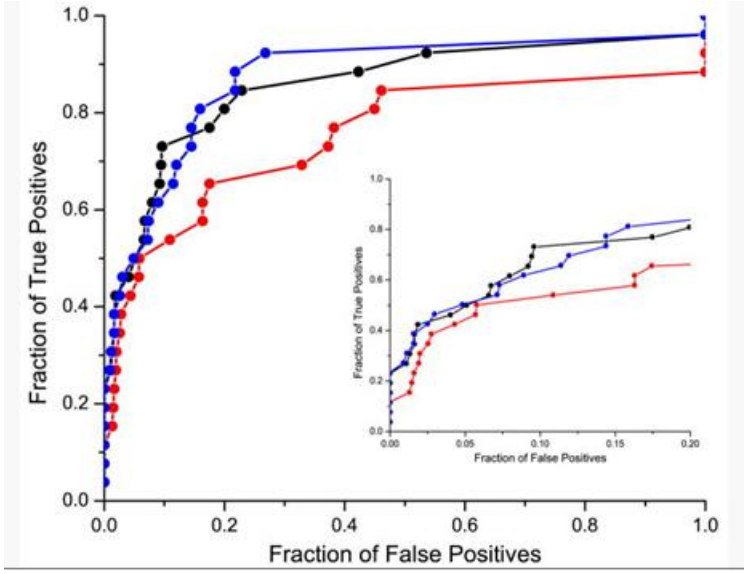


Bowers et al, Genome Biology, 2004

Coevolution of interacting proteins – “mirrortree” methods.

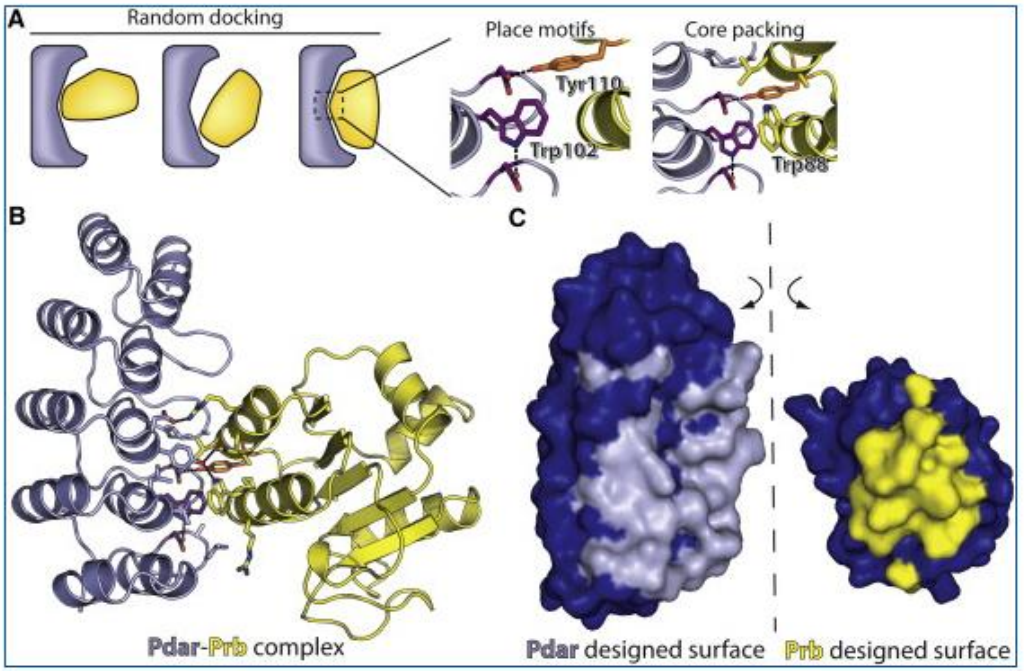
- Interacting proteins may co-evolve and their phylogenetic trees show similarity (*Goh et al, J.Mol.Biol.,2000).*
- Similarity between phylogenetic trees is measured by correlation coefficient between distance matrices.
- Signal comes from both correlated evolution of binding sites and whole protein sequence (*Kann et al, JMB 2009*).





Interface design

- Computationally alter interface to modify function
- Create useful properties
- Alter oligomeric state
- Alter specificity
- Novel interactions

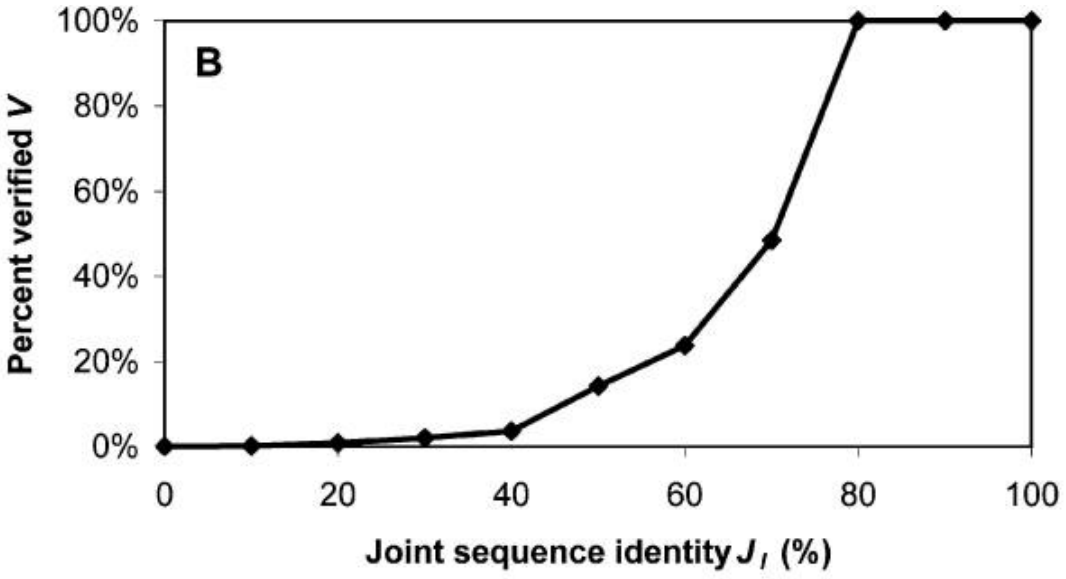


Karanicolas et al, Mol Cell, 2011

Evolution of protein interactions

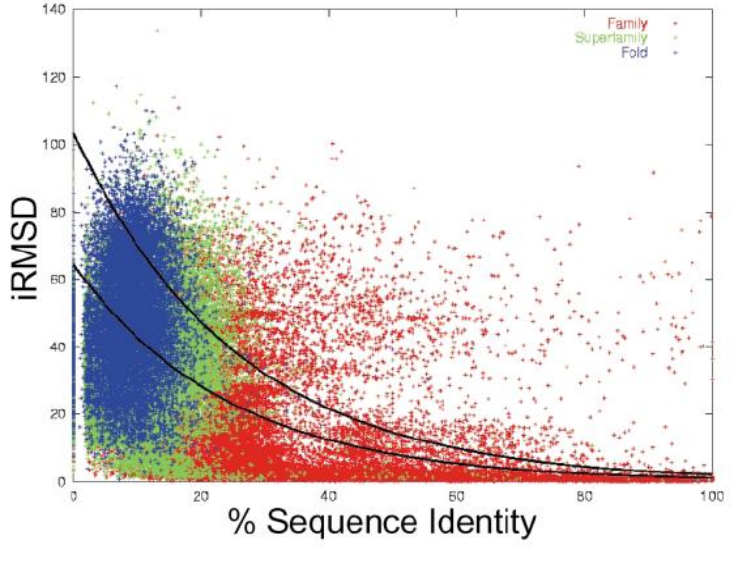
Conservation of protein-protein interactions.

Conservation of interologs



Yu et al, Genome Res, 2004

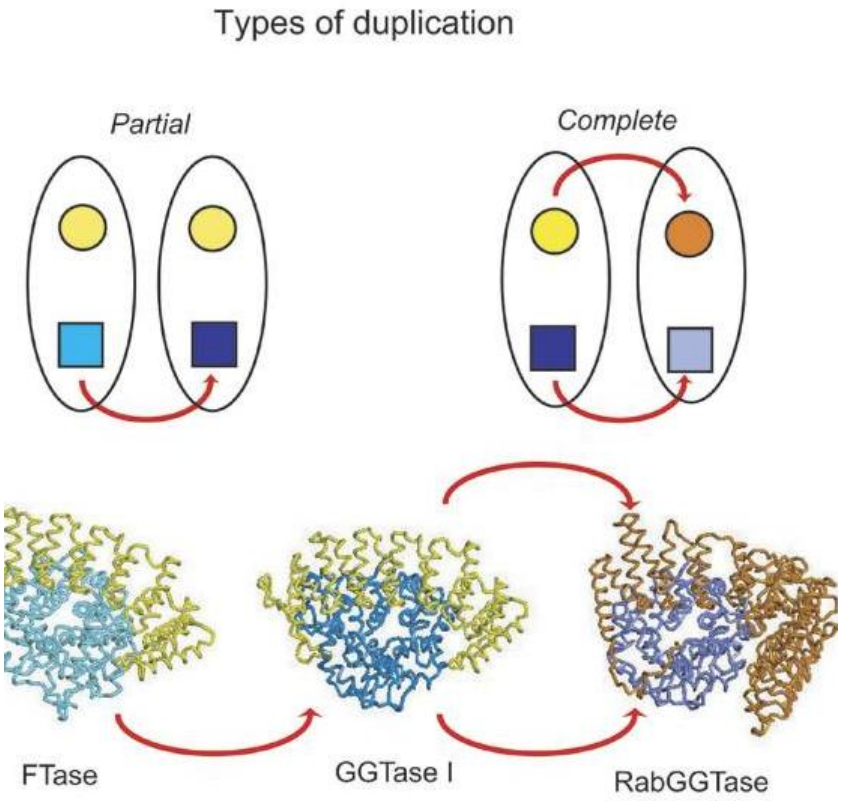
Conservation of binding interfaces



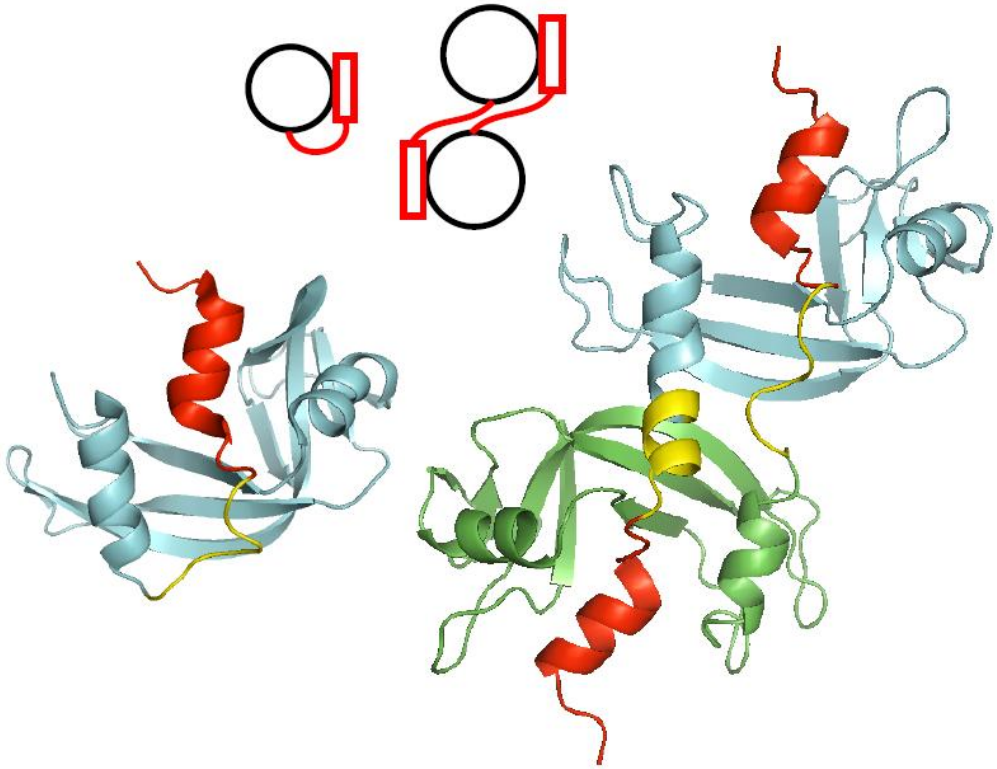
Aloy et al, *J. Mol. Biol*., 2003

Mechanisms of evolution of novel interfaces

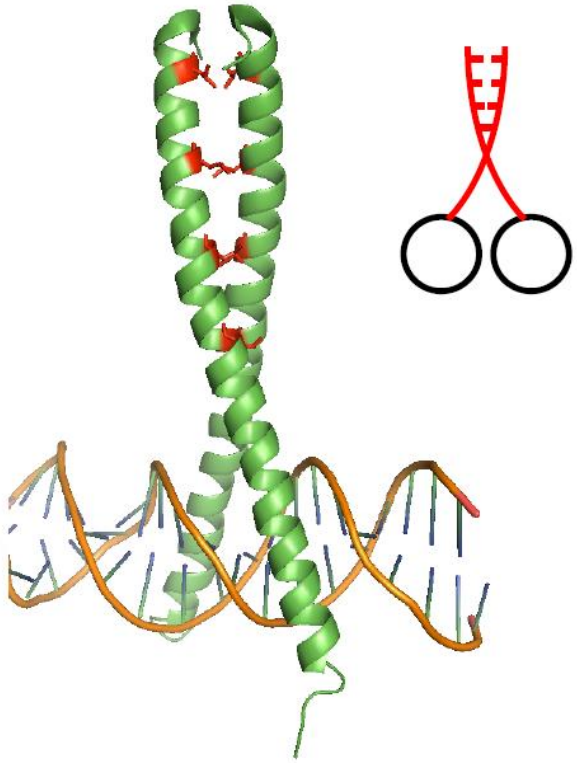
- Gene duplication with subsequent diversification (Pereira-Leal and Teichmann, Genome Res, 2005; Reid et al, BMC Genomics, 2010)
- Domain shuffling
- Point mutations on interfaces
- Insertions and deletions
- Other



Evolutionary mechanisms to form oligomers

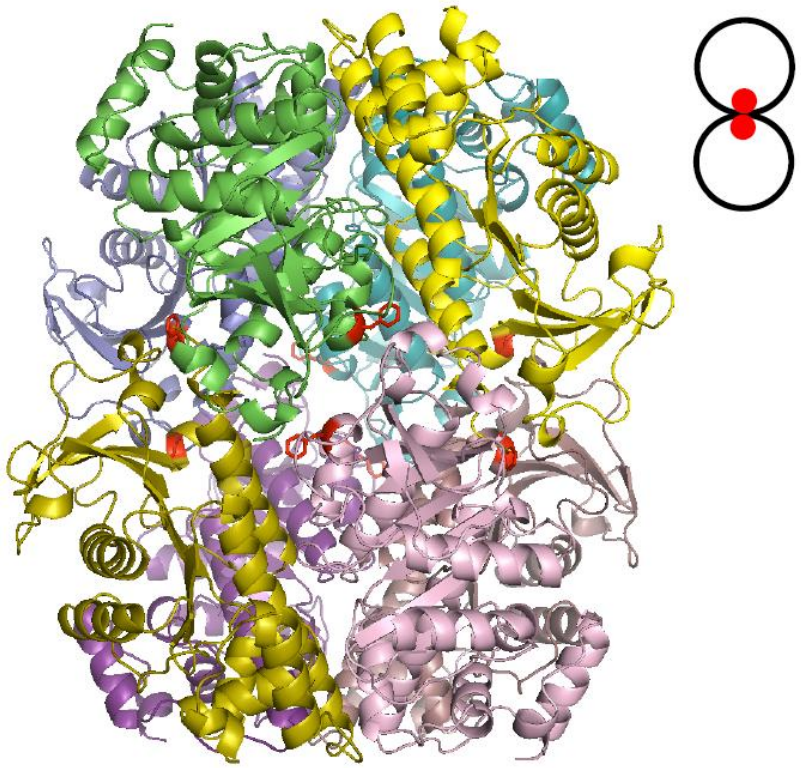


Domain swapping

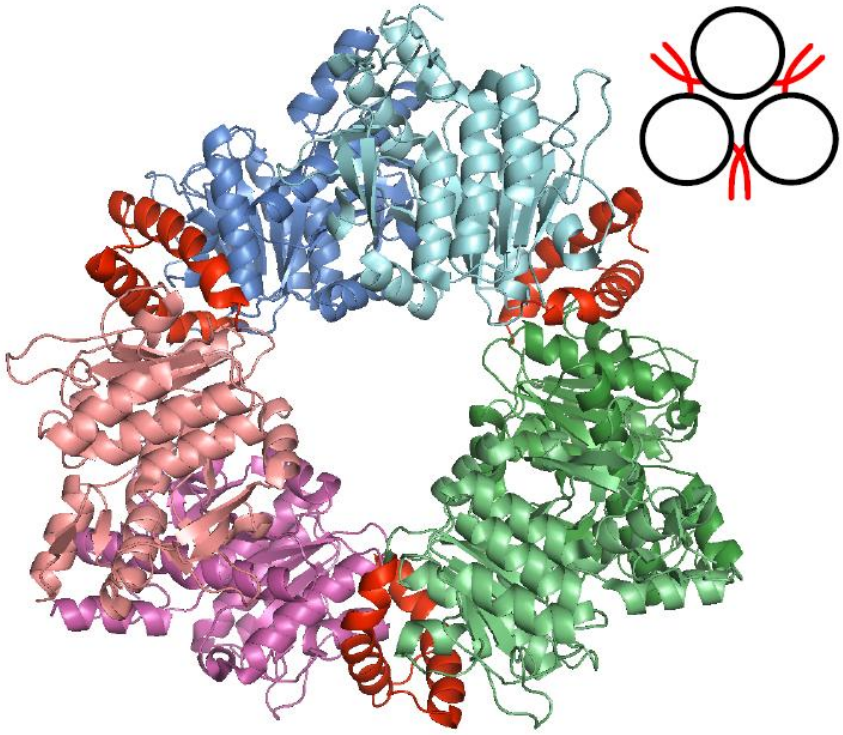


Leu zipper

Evolutionary mechanisms to form oligomers



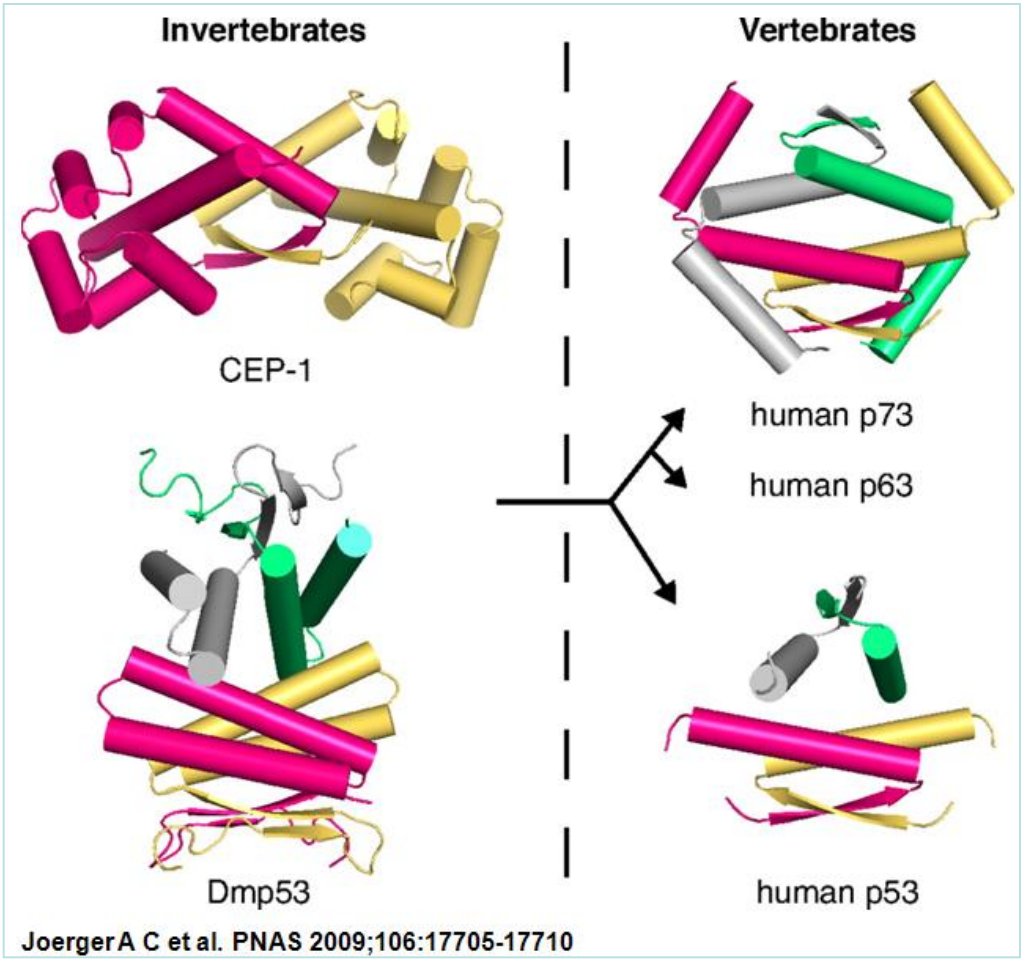
Mutations on interface



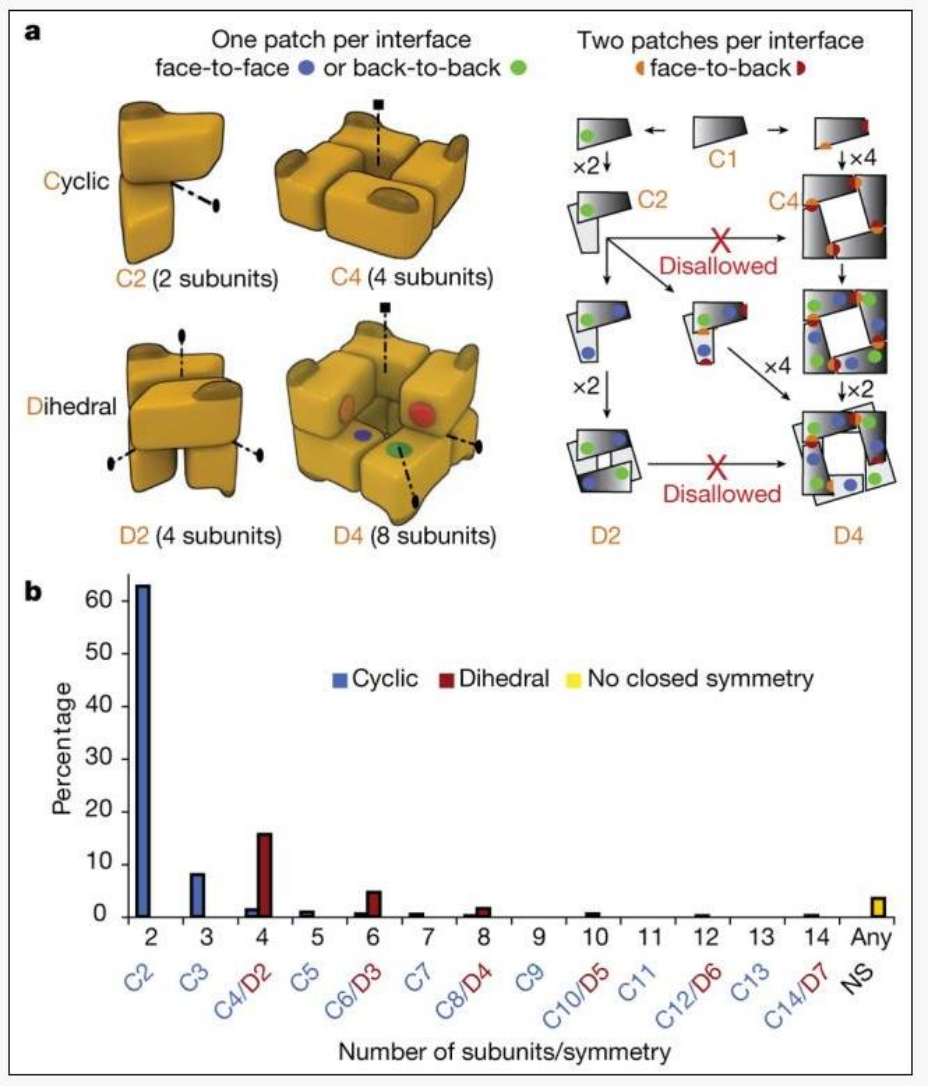
Insertions/deletions

Evolution of new specificity through oligomerization

Stabilization of p63/p73 tetramer leads to separation of their pathways from p53 pathway

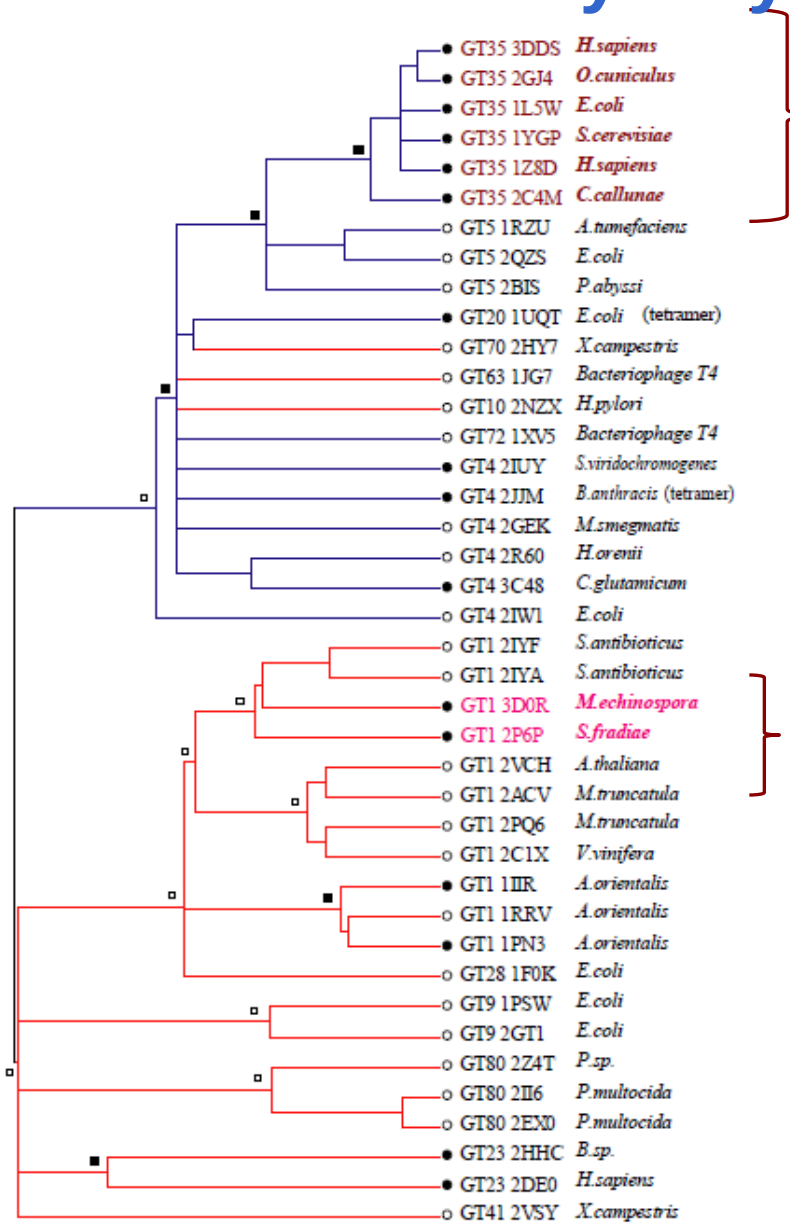


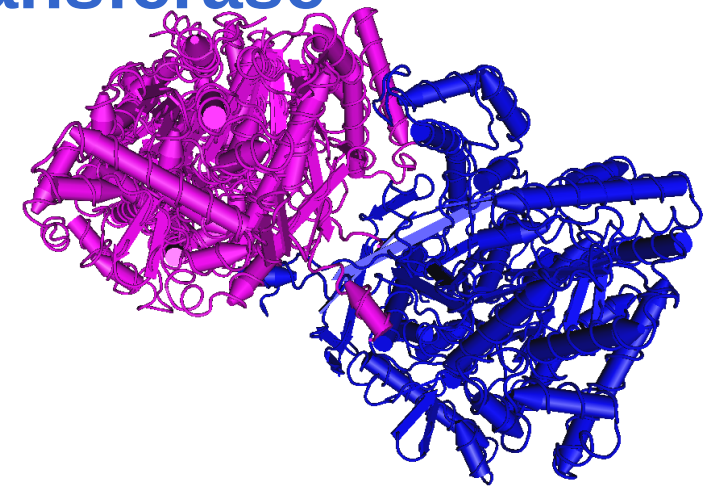
Assembly pathway mimics the evolutionary pathway

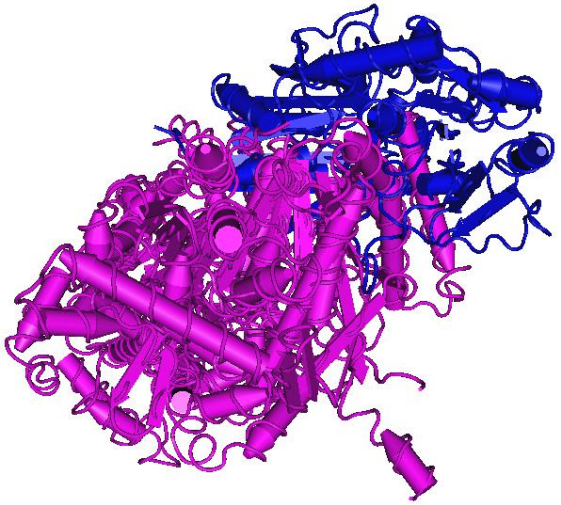


Levy et al, Nature, 2008

Evolution of homooligomeric binding modes: Glycosyltransferase



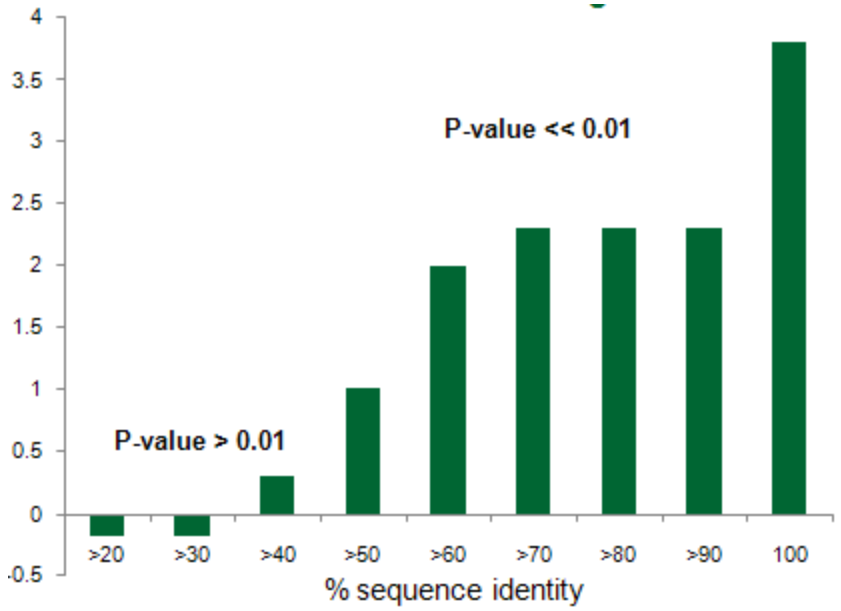




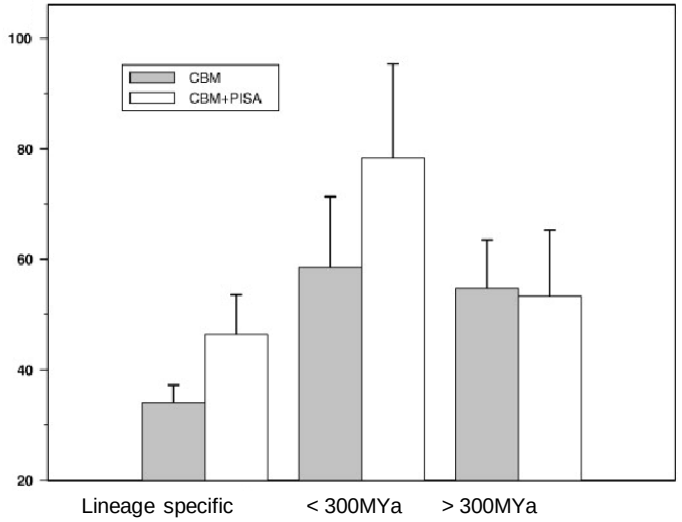
Hashimoto et al, J Mol Biol, 2010

Conservation of binding modes in evolution

Logarithm of probability ratio for finding the same or different binding modes on phylogenetic tree



Interface size



- binding modes are well conserved within phylogenetic clades sharing more than 50% sequence identity
- lineage-specific binding modes are smaller, less stable. Newer interactions are weaker

Dayhoff et al, JMB 2010

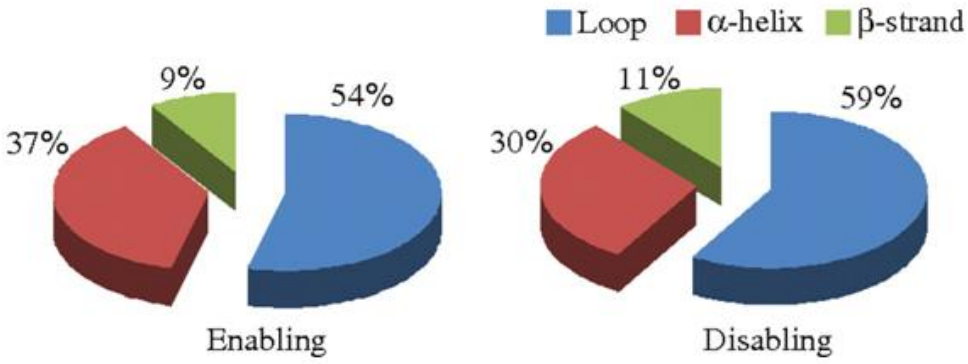
Role of Insertions and deletions in formation of oligomers

- "enabling" and "disabling" loops (*Akiva et al, PNAS 2008*)

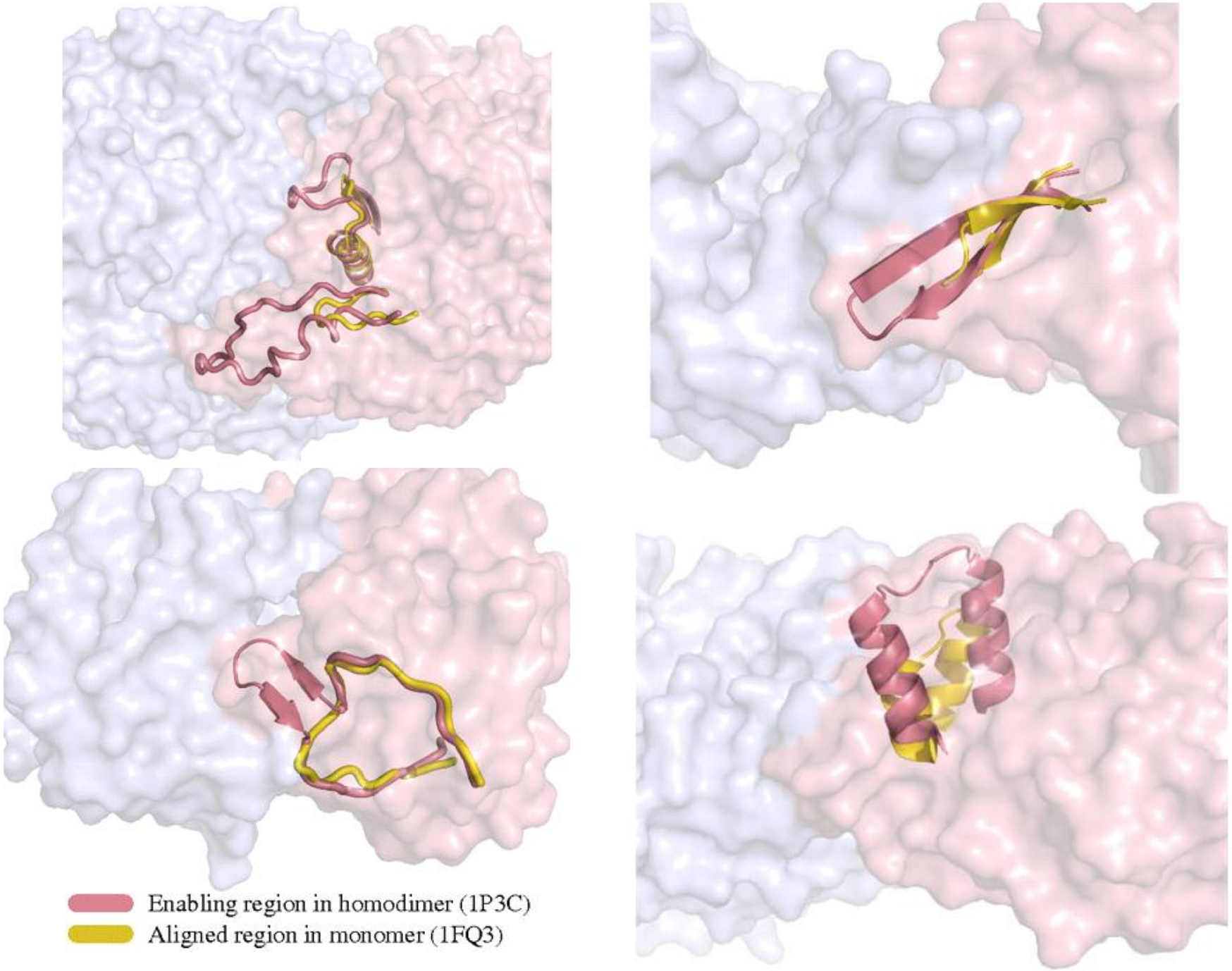
- Insertions/deletions occur more frequently or the interface than on the surface (*P*-value ≪ 10e-7) – "enabling" and "disabling regions"

- 25% homooligomers have enabling and disabling regions;

- they contain more polar and charged residues, Gly and Pro than "conventional interfaces"

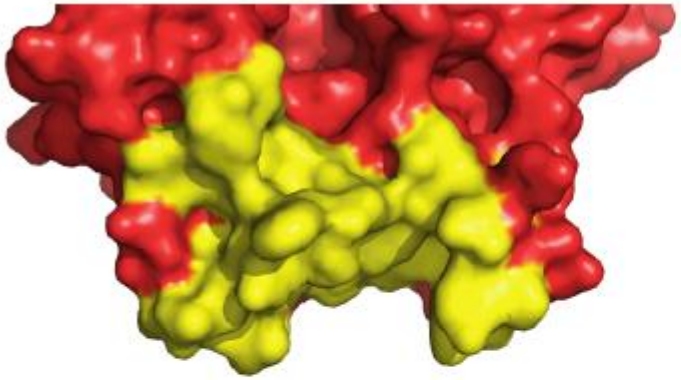


Hashimoto & Panchenko, PNAS 2010

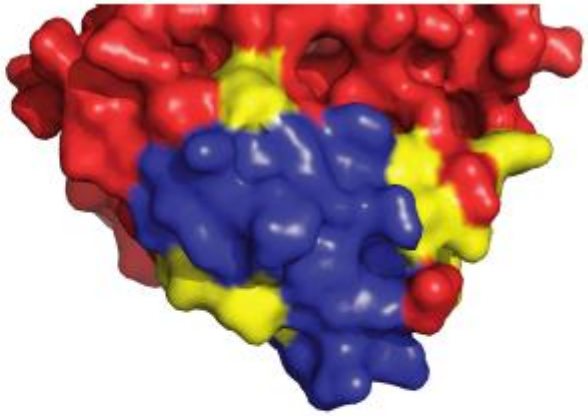

Enabling region in homodimer (1P3C)
Aligned region in monomer (1FQ3)

Disabling regions

Glycogenin glucosyltransferase, disrupting features



Eukaryotes, dimer



Bacteria, monomer

```
1LL2: 122 aap--------------------dpgwpDCFNSGVFVYQPSVETyNQLLHVASEQGs----- 158
1G9R: 123 nwlgasidlfverqegykqkigxadgeYYFNAGVLLINLKKWRrHDIFKXSSEWVeqykdv 183

1LL2: 159 FDGGDQGLLNTFFNSWattdirKHLPFIYNLSSISIYSYlpafk------------afgaNAK 209
1G9R: 184 XQYQDQDILNGLFKGGv-----CYANSRFNFXPTNYAFXanwfasrhtdplyrdrtntvxPVA 241
```

Prediction of oligomeric states from sequence

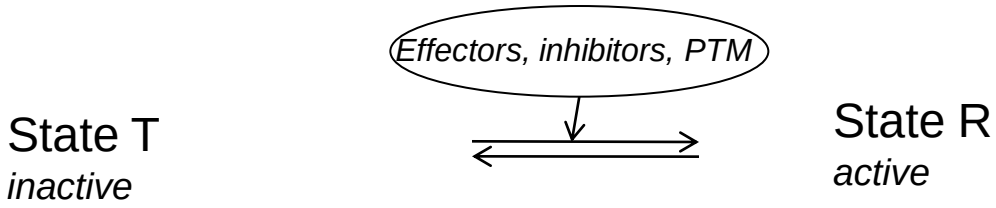
	Sensitivity	Specificity	Precision	Error rate
Enabling/disabling features	0.70	0.74	0.94	0.36
% identity	0.71	0.62	0.91	0.38
RMSD	0.72	0.60	0.90	0.40
GSAS	0.81	0.57	0.91	0.43
BLAST	0.74	0.53	0.89	0.47

Regulation of protein-protein binding

Mechanisms of regulation

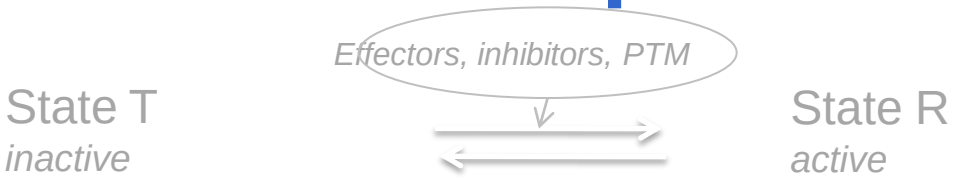
- Availability/abundance
  - Gene expression, translation
  - Translocation of proteins or substrates
  - Turnover
- Proteolytic activation
- Inteins

Mechanisms of protein regulation



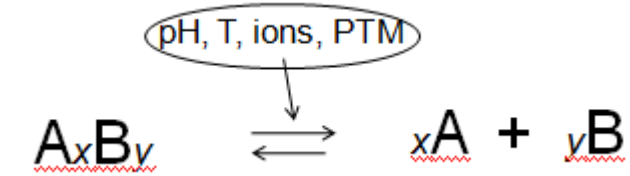
- Regulation by another protein or small molecule
- Reversible covalent post-translational modifications
- Allosteric activation and inhibition

Mechanisms of protein regulation



- Regulation by another protein or small molecule
- Reversible covalent post-translational modifications
- Allosteric activation and inhibition

Transitions between different oligomeric states



active *inactive*

$$K_d = \frac{[A]^x \times [B]^y}{[A_x B_y]} = e^{-\Delta G_{diss}/RT}$$

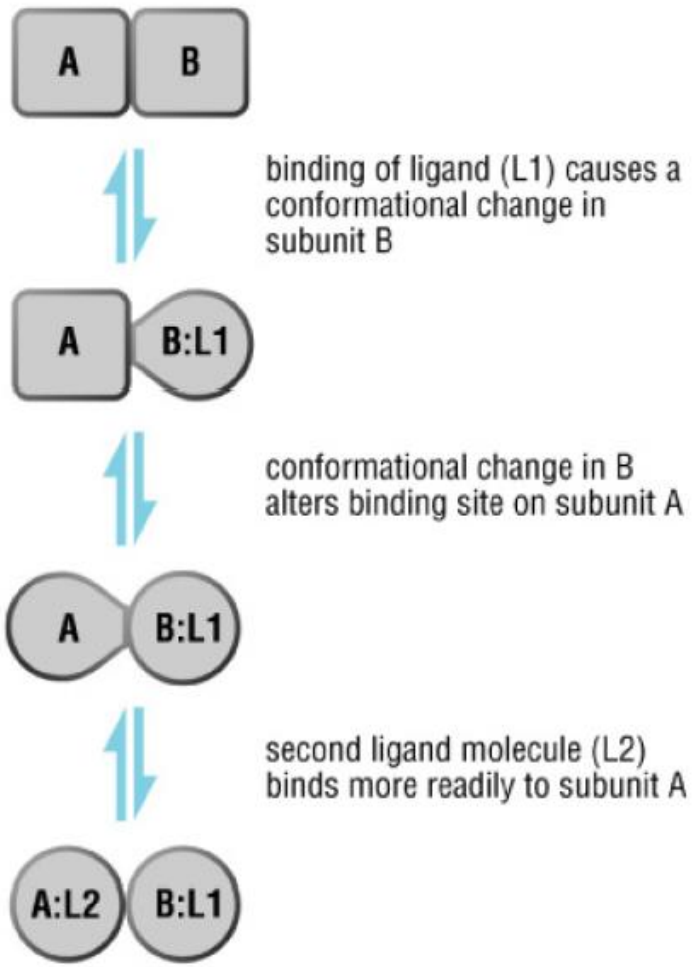
Dissociation constant

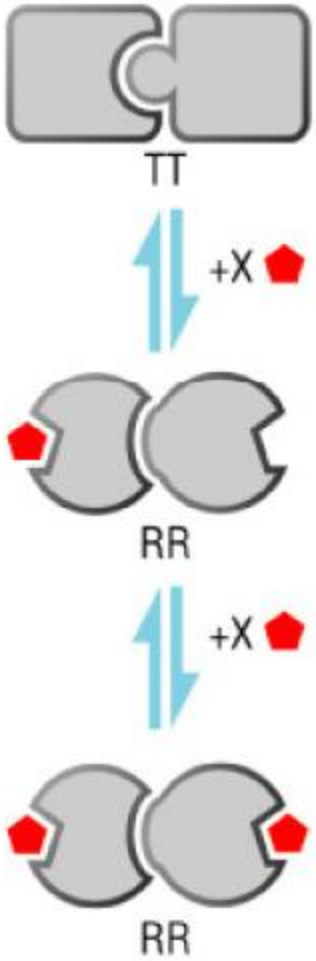
AB + C $\rightleftharpoons$ AC + B

$$K_{BC} = \frac{K_{AC}}{K_{AB}}$$

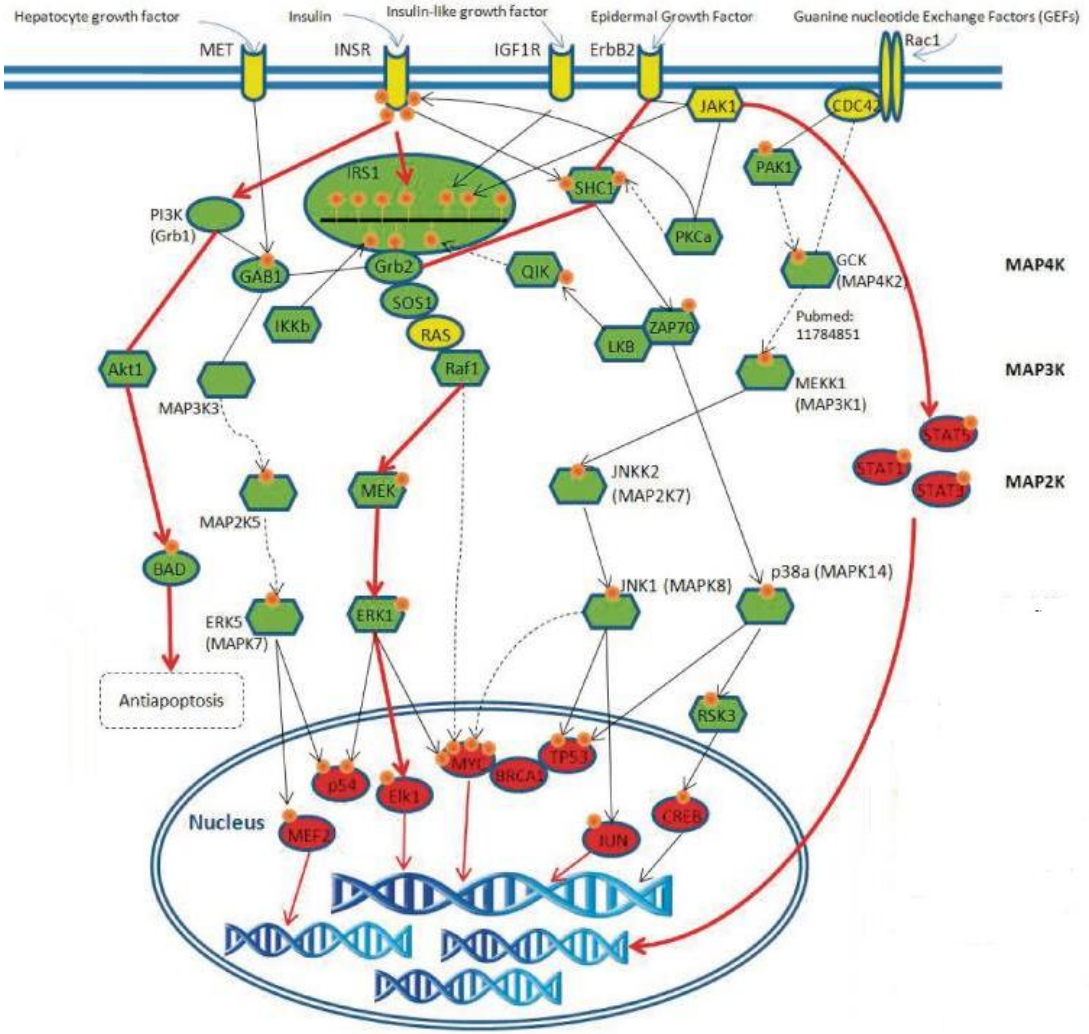
Binding selectivity constant

Allosteric regulation



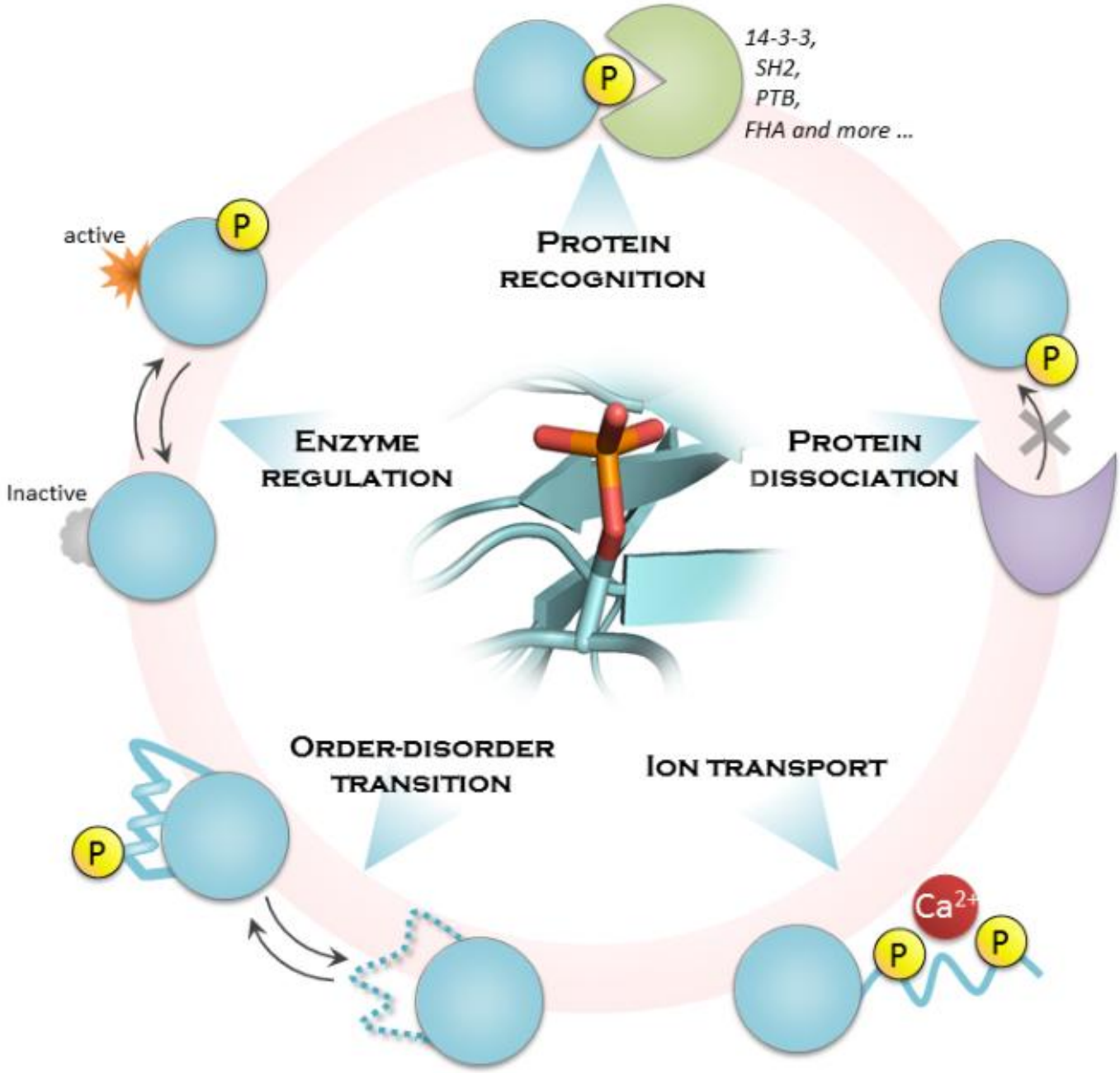


Signal transduction through protein-protein interactions and post-translational modifications

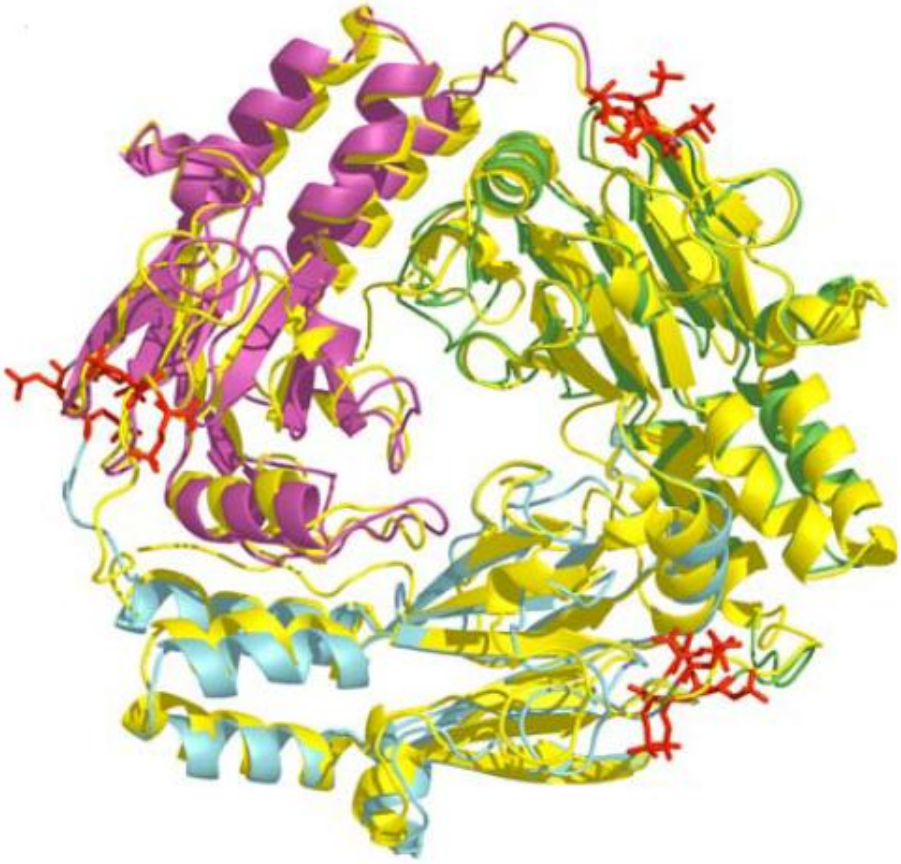


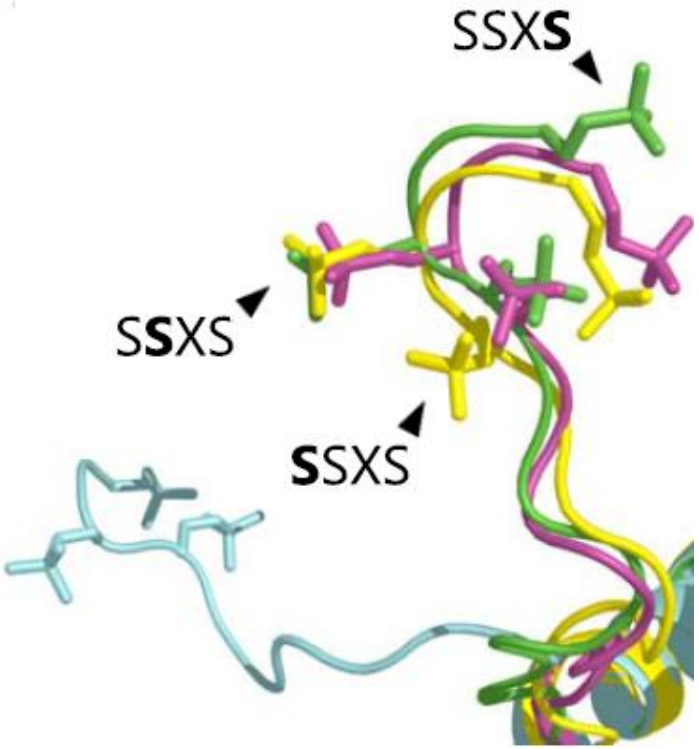
Protein control through covalent modifications

- 50-90% of human proteins is post-translationaly modified
- over 40 different modifications have been described
- most important: phosphorylation, glicosylation, lipidation, methylation, N-acetylation, S-nitrosylation, SUMOilation


14-3-3,
SH2,
PTB,
FHA and more ...
P
PROTEIN RECOGNITION
active
Inactive
ENZYME REGULATION
PROTEIN DISSOCIATION
ORDER-DISORDER TRANSITION
ION TRANSPORT
Ca2+

Smad2-MH2 trimer: phospho-group promotes the trimer formation





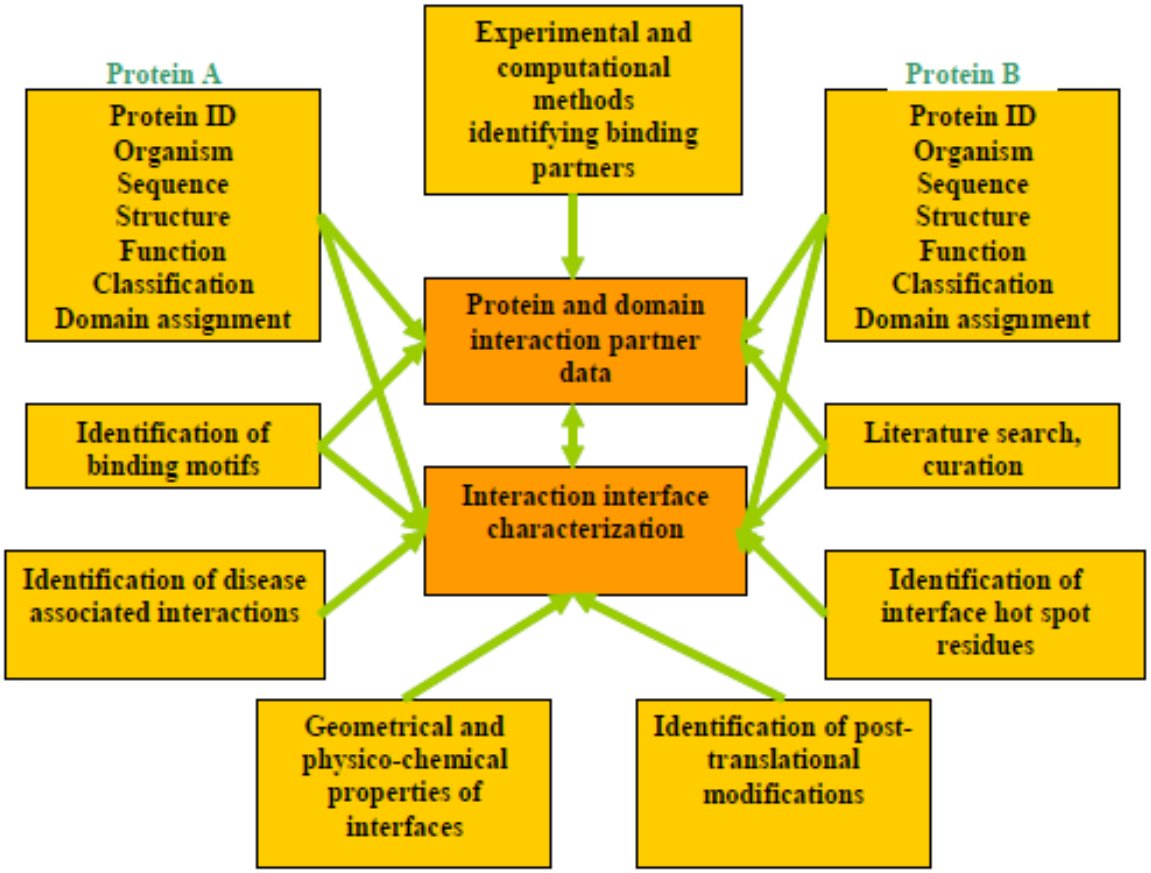
Nishi et al, submitted

Effect of phosphorylation/dephosphorylation on Smad complex formation

Protein	Site	pSite	AB	BC	AC	Average ΔΔΔG, kcal/mol
Smad2 (1khx)	SSXS	S->pS	0.59	0.59	0.59	0.59
	SSXS	pS->S	0.88	0.88	0.88	0.88
	SSXS	pS->S	1.53	1.53	2.86	1.97
Smad1 (1khu)	SSXS	S->pS	-2.11	1.58	-0.22	-0.25
	SSXS	S->pS	-0.9	1.49	-1.74	-0.38
	SSXS	S->pS	-1.45	-1.87	-1.08	-1.47

Protein-protein interaction databases

Data flow in protein interaction databases



Protein interaction databases

Database	Proteins/Domains	Type	Number of Interactions
DIP[a], LiveDIP	P	E,S	55,733
BIND[a]	P	E,C,S	83,517
MPact/MIPS[a]	P	E,C,F	15,488 (4,300)[b]
STRING	P	E,P,F	730,000 (proteins)
MINT[a]	P	E,C	71,854
IntAct[a]	P	E,C	68,165
BioGRID[a]	P	E,C	116,000 (30,000)[b]
HPRD	P	E,C	33,710
ProtCom	P,D	S,H	1,770
3did, Interprets	D	S,H	3,304
Pibase, ModBase	D	S,H	2,387
CBM	D	S	2,784
SCOPPI	D	S	3,358
iPfam	D	S	3,019
InterDom	D	P	30,037
DIMA	D	F,S	—
Prolinks	P	F	—

BioGRID, Stark et al, NAR 2011

Organism	Experiment Type	Raw Interactions	Non-Redundant Interactions	Unique Proteins	Unique Publications
Human	PHYSICAL	60570	39635	10259	12411
	GENETIC	513	489	525	198
	COMBINED	61083	39938	10448	12470
All	PHYSICAL	204613	140813	31754	20369
	GENETIC	184715	132714	9420	8827
	COMBINED	389328	267879	33563	26894

IBIS – NCBI server to analyze and infer interactions and binding sites

http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi

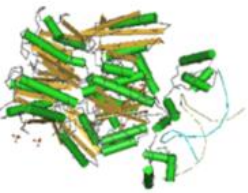
"Observed" interactions – from structures

"Inferred" interactions – from homologous structures with observed interactions

Biological relevance of binding sites:

- occurs in several non-redundant homologs;
- structurally and sequence conserved;
- binds biologically active molecules;
- validated by PISA algorithm (Krissinel & Henrick,*2007*);
- overlaps with the curated binding site

Shoemaker et al, NAR 2010
Thangudu et al, BMC Bioinformatics, 2010



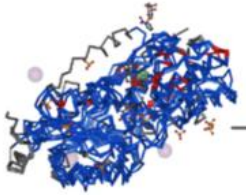
Protein Structure

BLAST

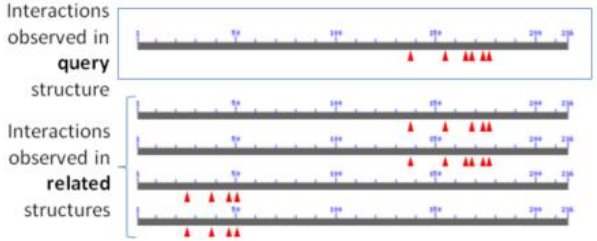
Protein Sequence

>Transcriptional Regulator- Phosphoprotein-
XNVTIYDVAREASVSXATVSRVVNGNPNVK
PSTRKKVLETIERLGYRPNAVARGLASKKT
TTVGVIIPDISNIFYAELARGIEDIATXYKYNII
LSNSDQNQDKELHLLNNXLGKQVDGIIFXS
GNVTEEHVEELKKSPV.....

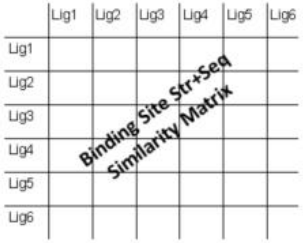
VAST

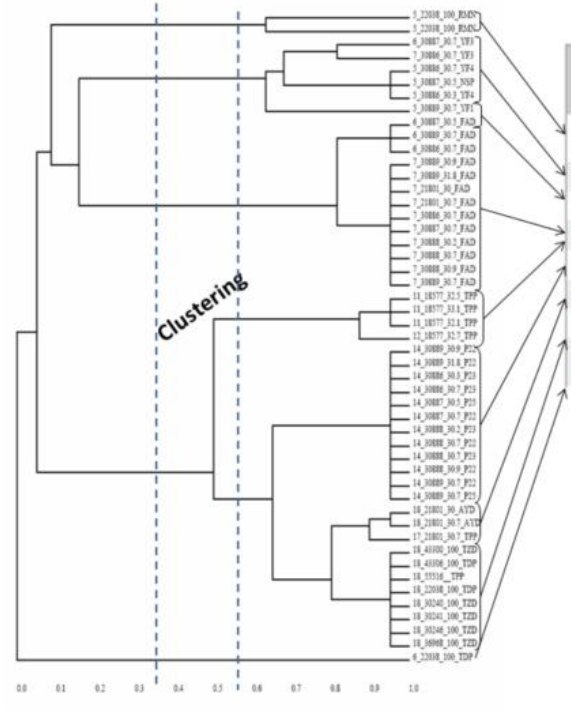


Related structures



Extracting structurally aligned binding sites





Clustering of binding sites and choosing a cutoff to define binding site clusters

Inferred binding site clusters

Structure of complex	Ranking Score	Number of Cluster Members	Average %Identity to Query	Number of Binding Site Residues	Curator Annotation	Taxonomic Diversity
SH2	2.6	4	39	11	SH3/SH2 domain interface	Homo sapiens
PTKc_Syk	5.2	3	47	19	-	Homo sapiens
PTKc_Src_like	2.6			23	-	Amniota
PTKc_Abl	2.5			22	-	Euarchontoglires
PTKc_Ack_like	2.4			19	-	Euarchontoglires
PTKc_Csk	2.1			28	-	Amniota
S_TKc	2.0			15	-	Homo sapiens
No Domain Assigned	Singleto			28	-	Homo sapiens

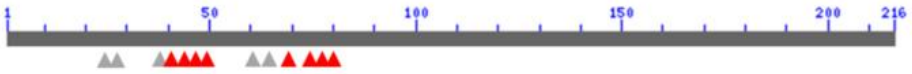
Binding site alignment

27	28	44	45	46	47	48	49	60	62	64	75	78	79
V	K	G	V	E	V	H	P	K	N	W	L	G	Y
L	H	G	V	E	F	G	S	K	Q	W	V	S	Y
-	-	-	A	A	F	L	T	-	-	W	L	M	Y
-	-	-	A	A	F	-	T	-	-	W	L	M	Y
-	-	-	A	-	F	L	T	-	-	W	L	M	Y
-	-	-	A	-	F	L	T	-	-	W	-	M	Y

Alignment of binding site residues

Scoring of binding sites for biological relevance using PSSMs, conservation, contacts, and sequence similarity

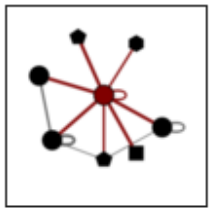
Transfer of the inferred binding site on to query sequence



Protein PDB ID or GI | New Search

Query 2OCJ_B

P53 Tumor Suppressor



All interactions for query sequence

Search 2OCJ B interactions

Similarity to query

- Sequence Identity: %Id list
- Structure RMSD: RMSD list

Interaction partner type

- PDB Code: PDB code lis
- Taxonomy: Taxonomies list

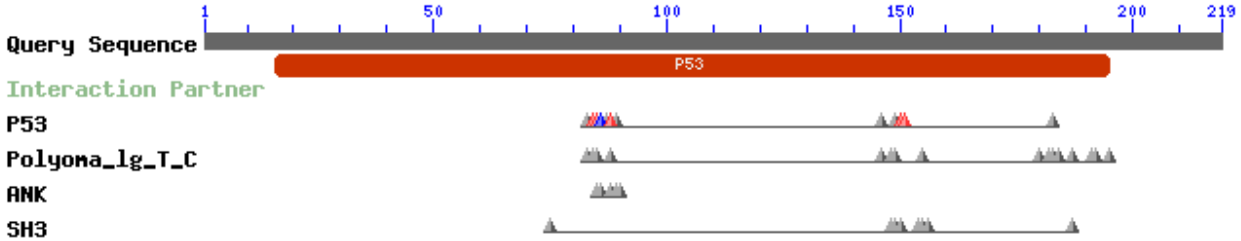
Reset

PISA Validation: ☑

Protein-Protein (4) | Protein-Chemical (1) | Protein-DNA/RNA (3) | Protein-Ion (1)

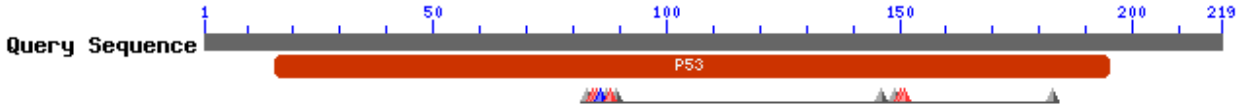
P53 Tumor Suppressor (sequence B)

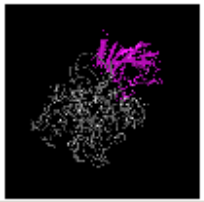
Other 2OCJ sequences ▼



List of protein interaction partners and binding sites. Similar binding sites of homologs of the query are grouped into clusters. To view the cluster members click on the plus sign. "**O**" denoted observed interactios. Note: singletons might not provide enough evidence for biological relevance of binding site.

Interaction Partner	Ranking Score	Number of Cluster Members	Average %Identity to Query	Number of Binding Site Residues	Taxonomic Diversity
⊟ P53	6.1	29	98	12	Euarchontoglires





View Binding Sites

Download Cn3D

Homologous complex	Homolog	Interaction partner	%Identity to query	Binding site (83 84 85 86 87 88 89 146 149 150 151 183)
ry	-	-	-	C P H H E R C N C M G A
2AC0	A	B	100	- P H H - R - - - M G -
3EXJ	B	A	88	C P H H - R - - - M G -

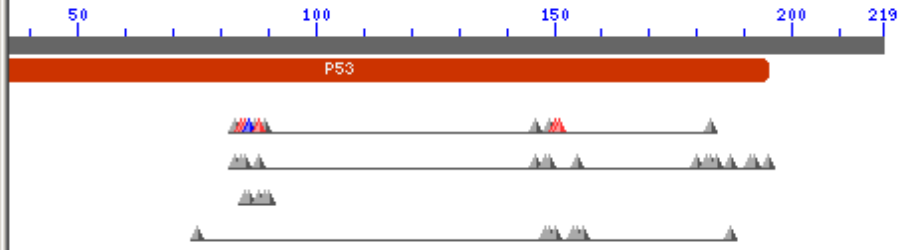
* click structure accession or sequence letter to explor structure and sequence information.

⊞ Polyoma_lg_T_C	singleton	3	100	16	Homo sapiens
⊞ ANK	singleton	1	100	5	Homo sapiens
⊞ SH3	singleton	1	100	8	Homo sapiens

Protein-Chemical (1) | Protein-DNA/RNA (3) | Protein-Ion (1)

P53 Tumor Suppressor (sequence B)

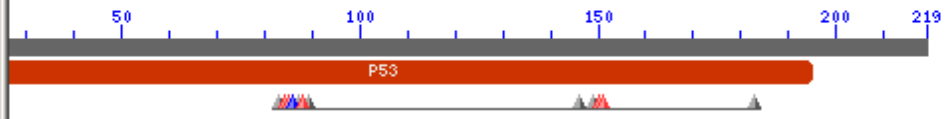
Other 2OCJ sequences



...hers and binding sites. Similar binding sites of homologs of the query are grouped into

...ers click on the plus sign. "O" denoted observed interactios. Note: singletons might not provide ...ance of binding site.

Ranking Score	Number of Cluster Members	Average %Identity to Query	Number of Binding Site Residues	Taxonomic Diversity
6.1	29	98	12	Euarchontoglires



...logous ...mplex	Homolog	Interaction partner	%Identity to query	Binding site
	-	-	-	83 84 85 86 87 88 89 146 149 150 151 183 C P H H E R C N C M G A
...AC0	A	B	100	- P H H - R - - - M G -
...EXJ	B	A	88	C P H H - R - - - M G -

Download Cn3D

* click structure accession or sequence letter to explor structure and sequence information.

⊞ Polyoma_lg_T_C	singleton	3	100	16	Homo sapiens
⊞ ANK	singleton	1	100	5	Homo sapiens
⊞ SH3	singleton	1	100	8	Homo sapiens

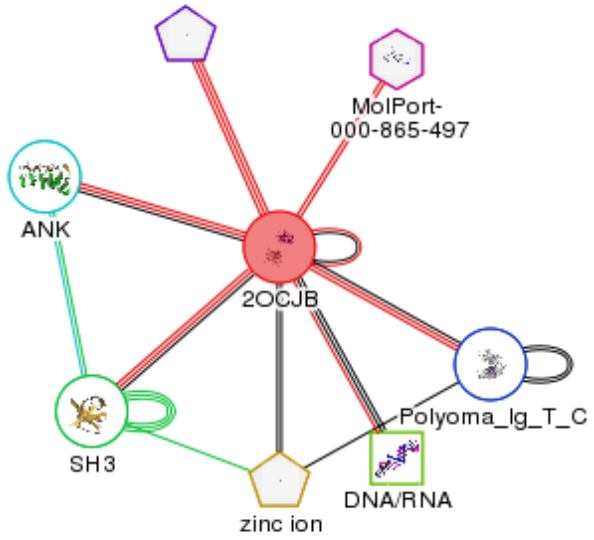
http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi?ngtype=10&...

Close
Help

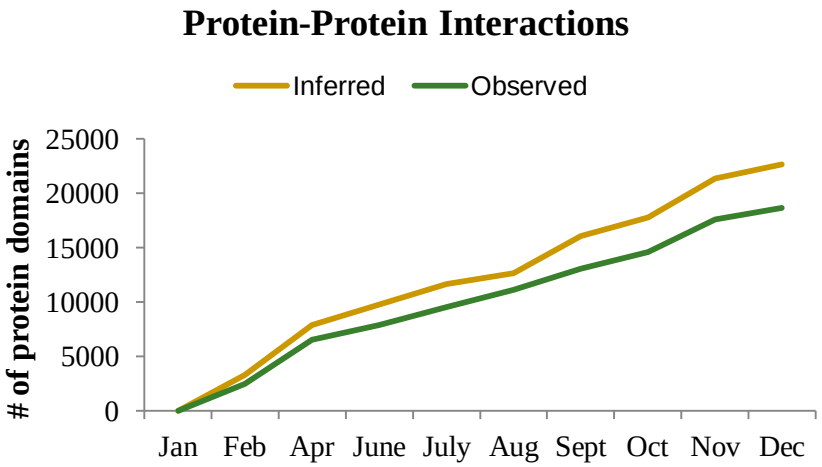
Shapes
- Protein/domain
- Peptide
- Chemical
- Ion
- DNA/RNA
- Self-interaction

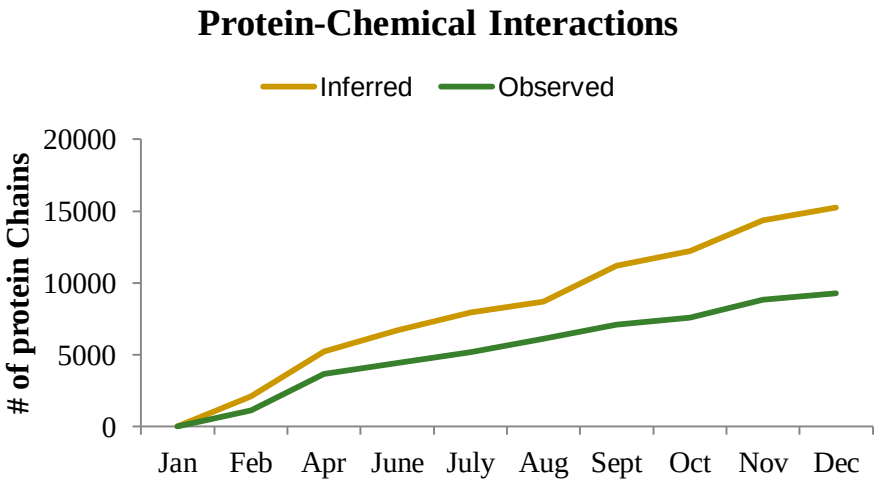
Colors
- Observed interactions
- Query protein & query's inferred interactions
- Other nodes & inferred interactions

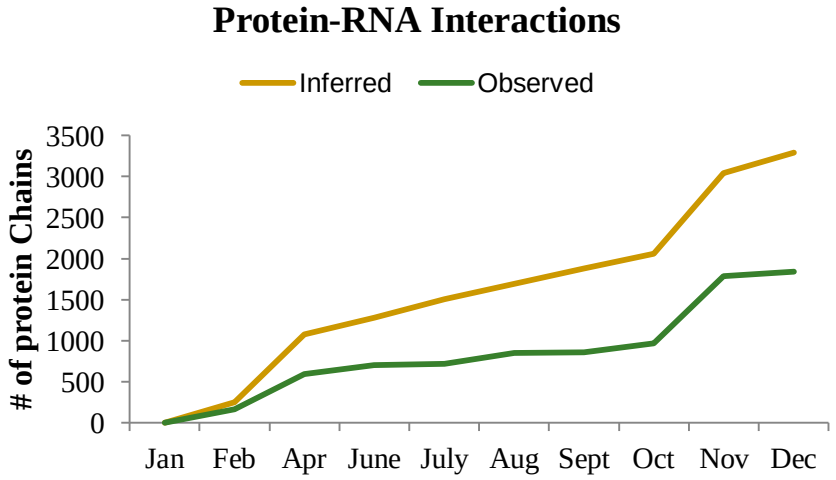
Query 2OCJB: P53 Tumor Suppressor
Interaction network (using default parameters)

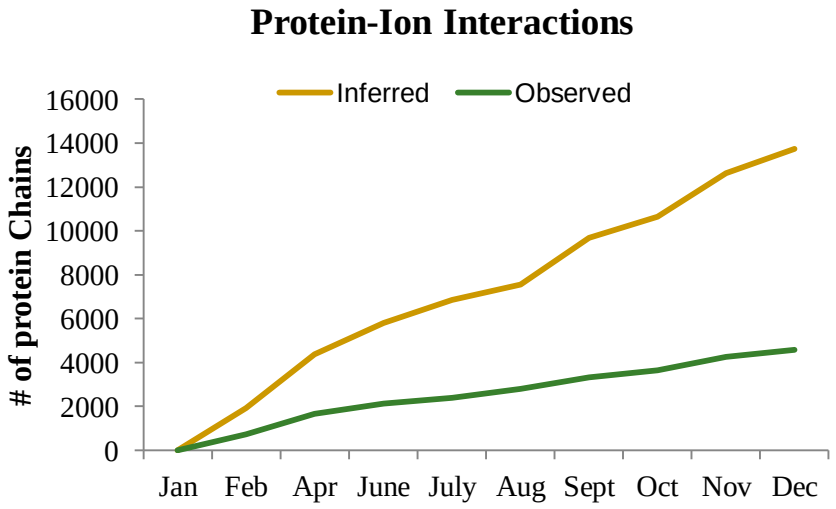


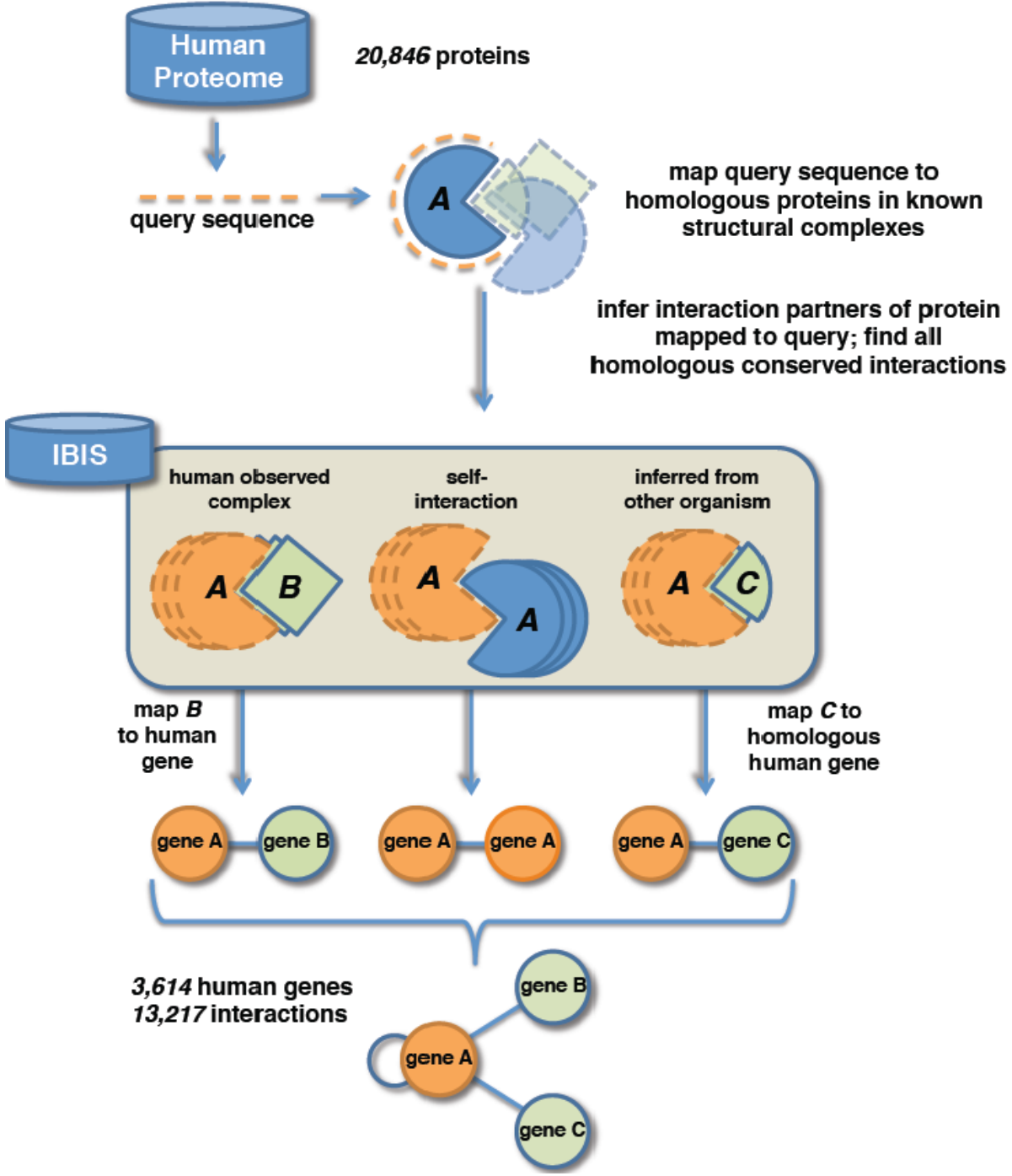
Growth of IBIS data over 2010



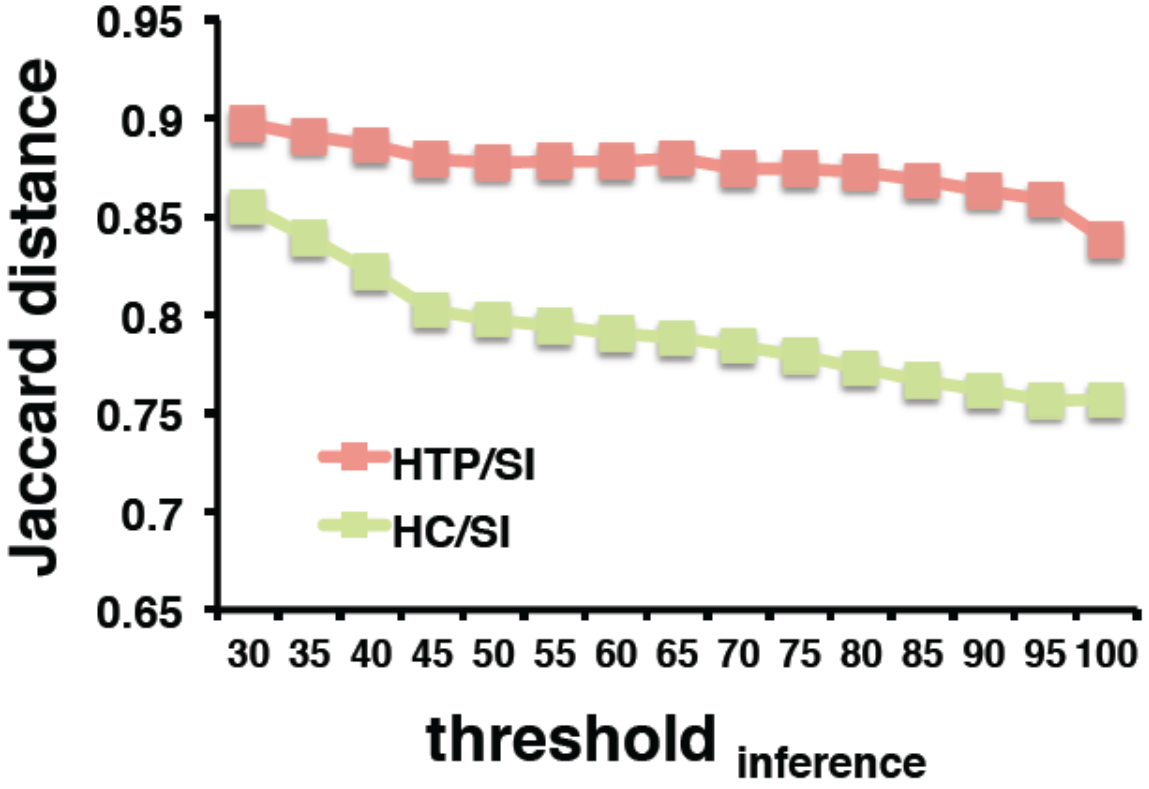







Human Proteome
20,846 proteins
query sequence
A
map query sequence to homologous proteins in known structural complexes
infer interaction partners of protein mapped to query; find all homologous conserved interactions
IBIS
human observed complex
self-interaction
inferred from other organism
A
B
A
A
A
C
map B to human gene
map C to homologous human gene
gene A
gene B
gene A
gene A
gene A
gene C
3,614 human genes
13,217 interactions
gene B
gene A
gene C

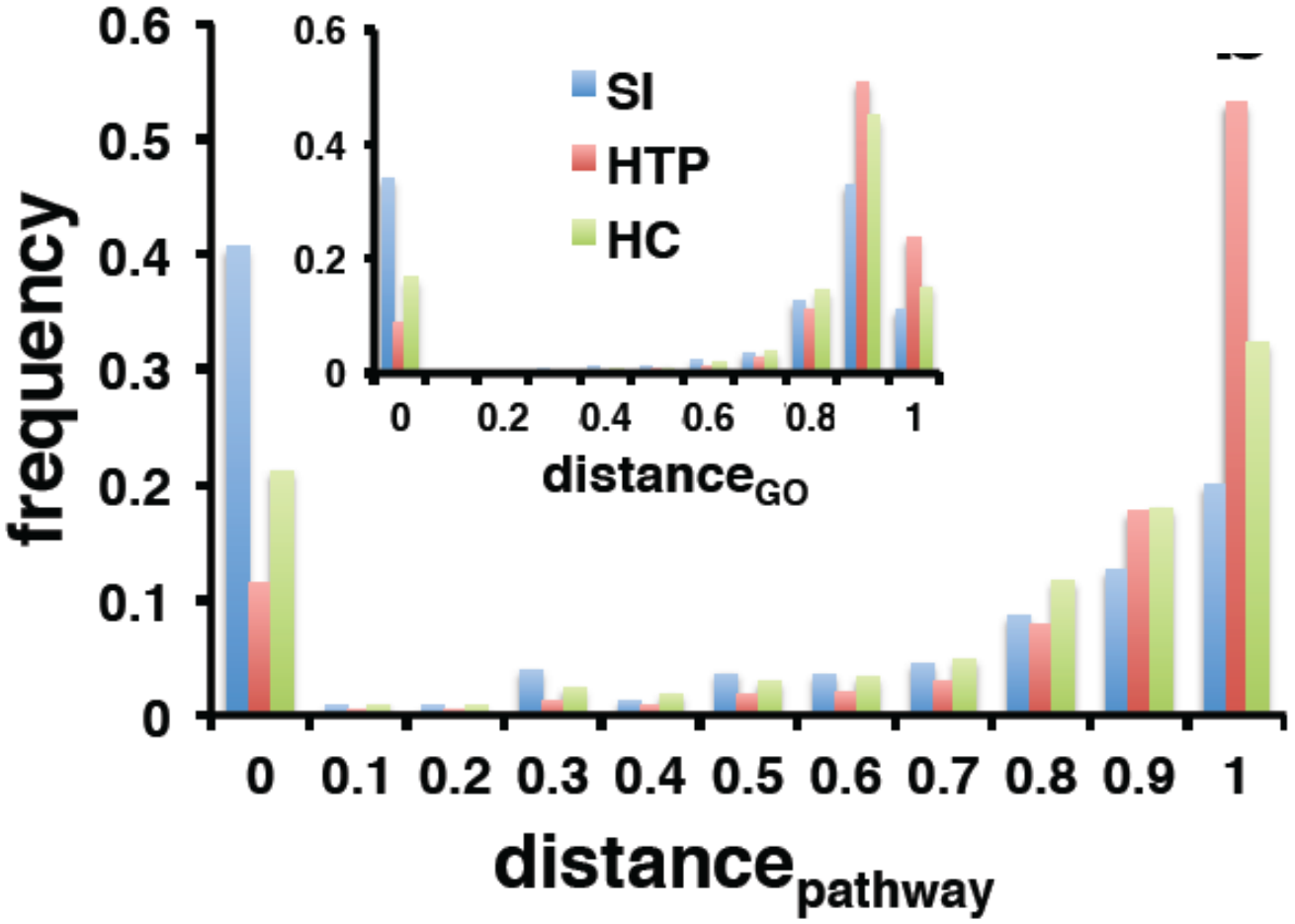
Comparison of structurally inferred (SI), high-throughput (HTP) and high confidence HTP (HC) networks

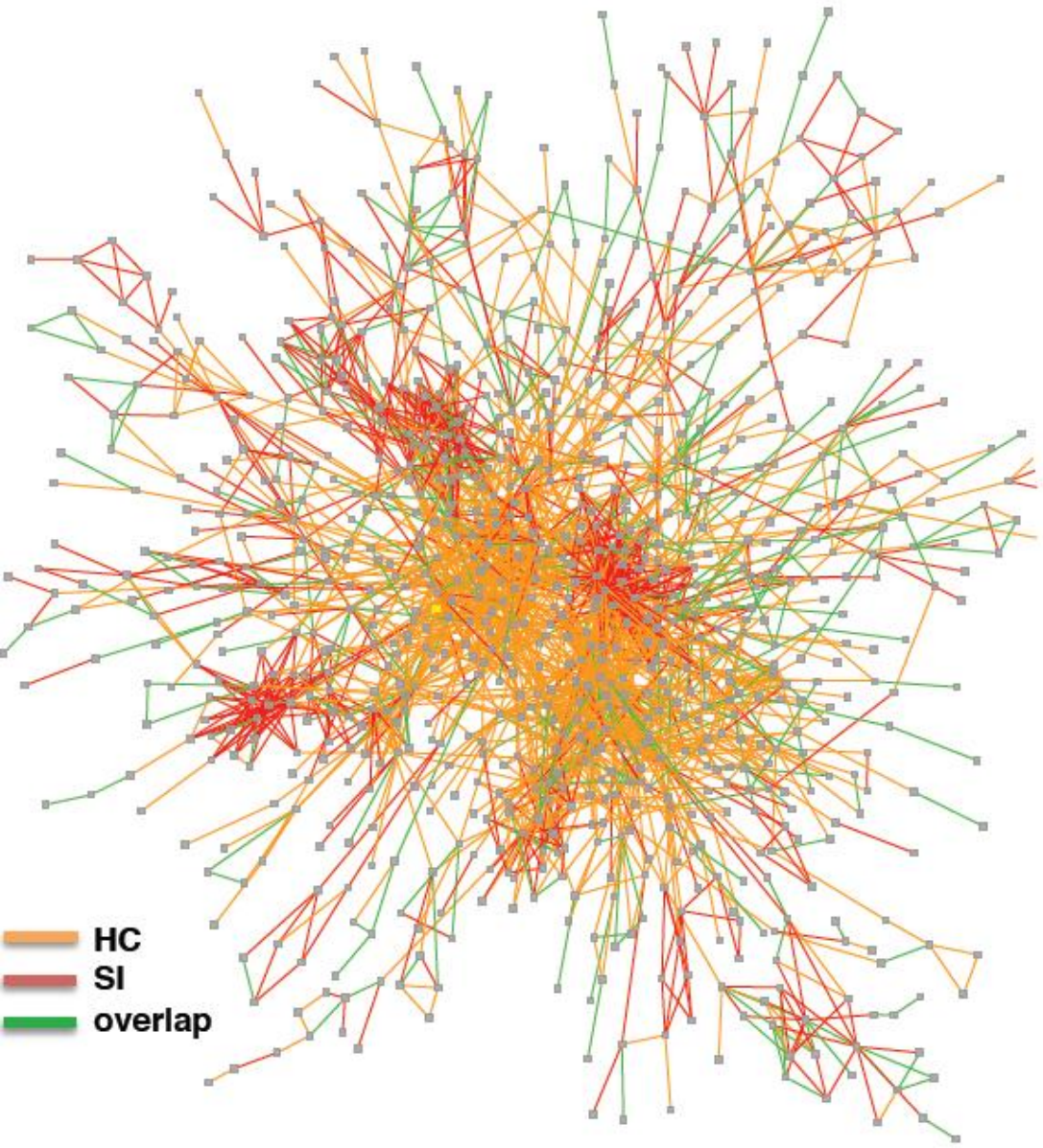


Inference threshold – similarity between query protein and closest homolog with known complex

Tyagi et al, submitted

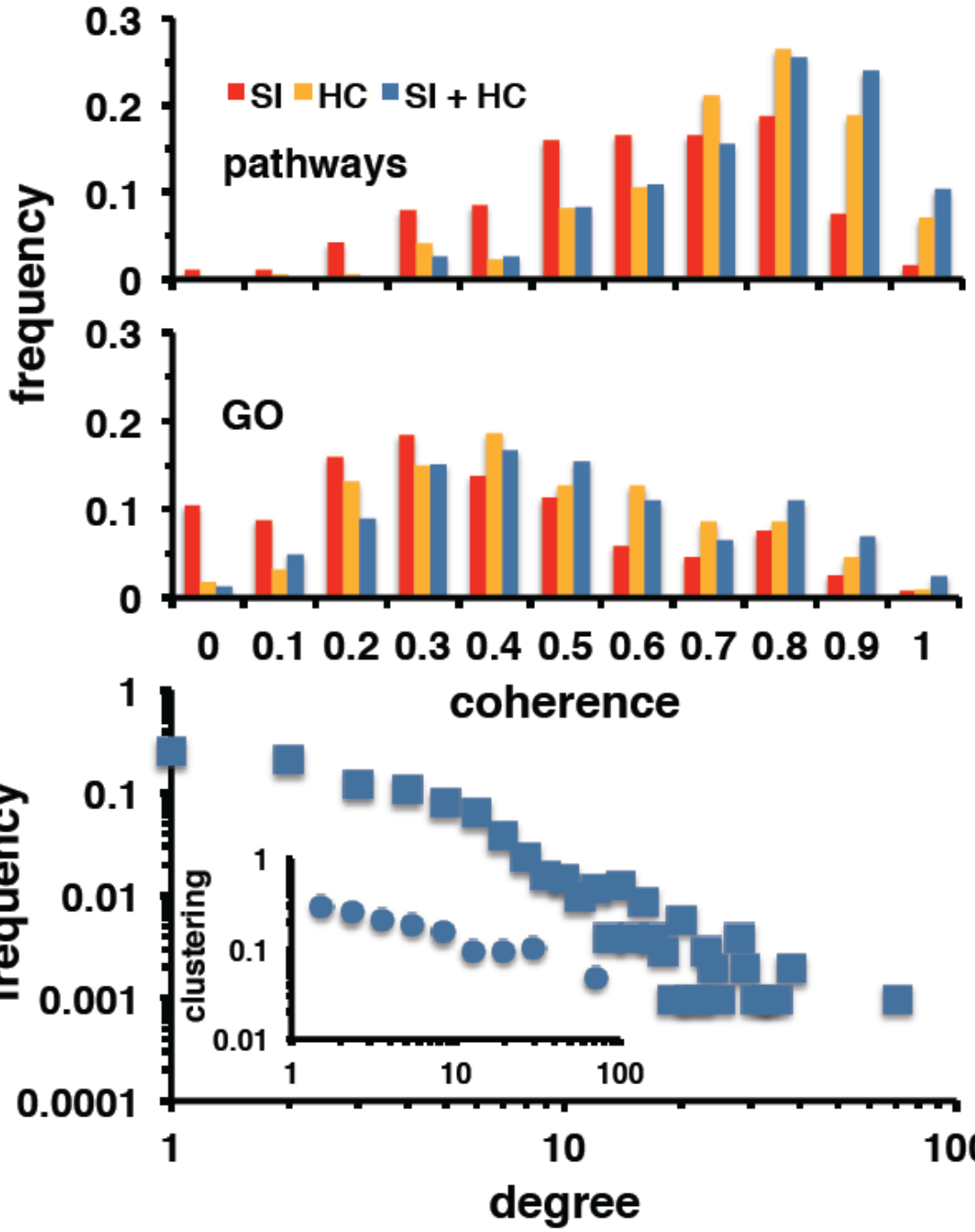
Structurally inferred networks are more functionally coherent than high-throughput networks





“Merged” networks = structurally inferred (SI) + high confidence high-throughput (HC), ~5500 proteins and ~17000 interactions

SI and HC complement each other; ~20% of HC interactions are observed in SI and ~50% SI interactions are observed in HC



Coherence – fraction of proteins in a network composed by proteins from a given pathway

Scale-free, modularity, “small-world” properties` of merged network

Acknowledgments

Kosuke Hashimoto

Manoj Tyagi

Hafumi Nishi

Stefen Wuchty

IBIS:

Stephen Bryant

Jessica Fong

Thomas Madej

Aron Marchler-Bauer

Benjamin Shoemaker

Rajesh Thangudu

Dachuan Zhang