Cristina Smaranda Domnica Sisu

Journal Club
21st March 2011

Papers

- Tanguy Chouard, Breaking the Protein Rules, *Nature* **471,** 151-153 (2011)
- Michail Yu. Lobanov , Eugeniya I. Furletova, Natalya S. Bogatyreva, Michail A. Roytberg, Oxana V. Galzitskaya , Library of Disordered Patterns in 3D Protein Structures, *PLoS Comput Biol* **6 (10),** e1000958 (2010)
- Jessica H. Fong, Benjamin A. Shoemaker, Sergiy O. Garbuzynskiy, Michail Y. Lobanov, Oxana V. Galzitskaya, Anna R. Panchenko, Intrinsic Disorder in Protein Interactions: Insights From a Comprehensive Structural Analysis, *PLoS Comput Biol* 5**(3),** e1000316 (2009)

Breaking the Protein Rules

"If dogma dictates that proteins need a structure to function, then why do so many of them live in a state of disorder?"

- 40% of all human proteins contain at least one intrinsically disordered segment ≥ 30 AA
- 25% are completely disordered

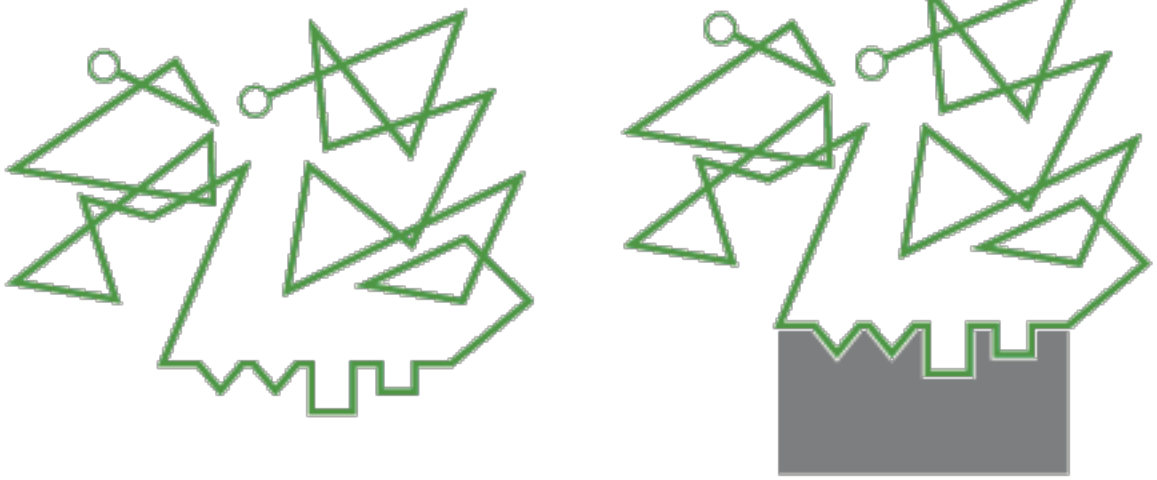
Intrinsically Disordered Proteins

- No unique 3D structure -> can fold upon binding.
- Large conformational changes.
- Conformation determined by AAs and binding partners.
- Can have many different functions.
- Can accommodate larger interfaces on smaller scaffolds.
- AAs: low aromatic content & high net charges.
- Rapid degradation of proteins.

Orders of Disorder

LOCK AND KEY

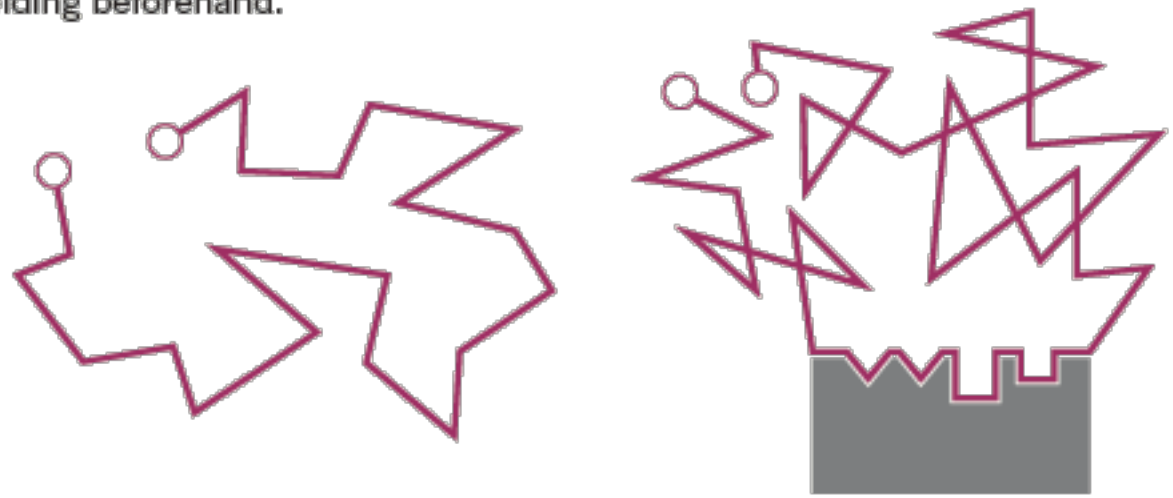
In the conventional view, an enzyme folds up immediately into a unique and stable 3D shape, the key (left). Its shape perfectly matches and allows it to bind its substrate, the lock (right).



Orders of Disorder

FOLD AS YOU BIND

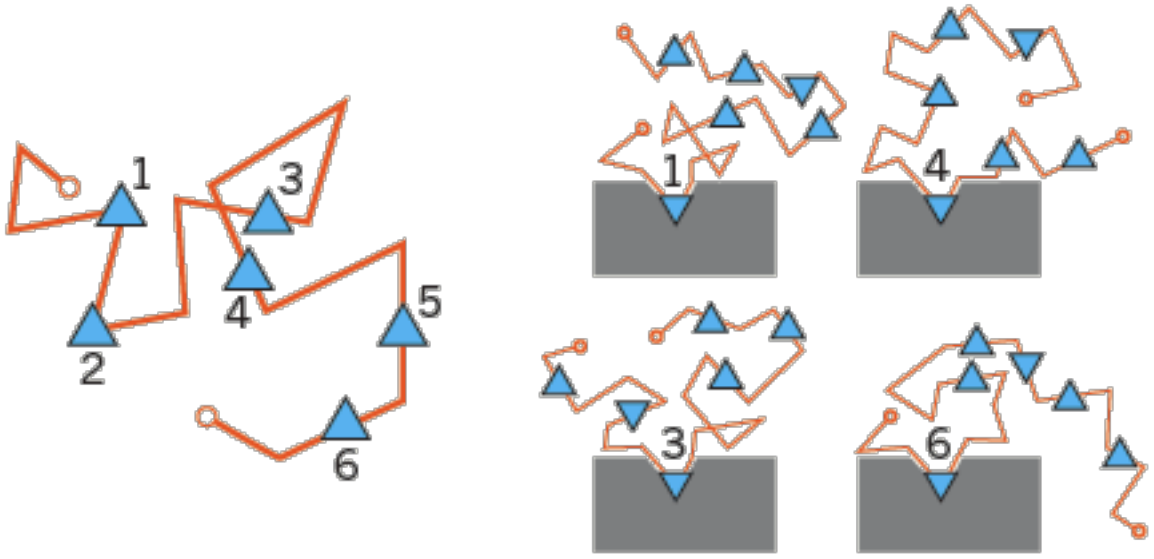
A disordered part of the gene-regulatory protein CREB (left) uses the lock to mould itself into the shape of the key when the two meet (right), rather than folding beforehand.



Orders of Disorder

SHAPE SHIFTING

The signalling protein Sic1 remains disordered in its bound state, and each of six phosphate groups occupies the binding site in turn. The protein is a mix of different conformations shifting around in constant dynamic equilibrium.



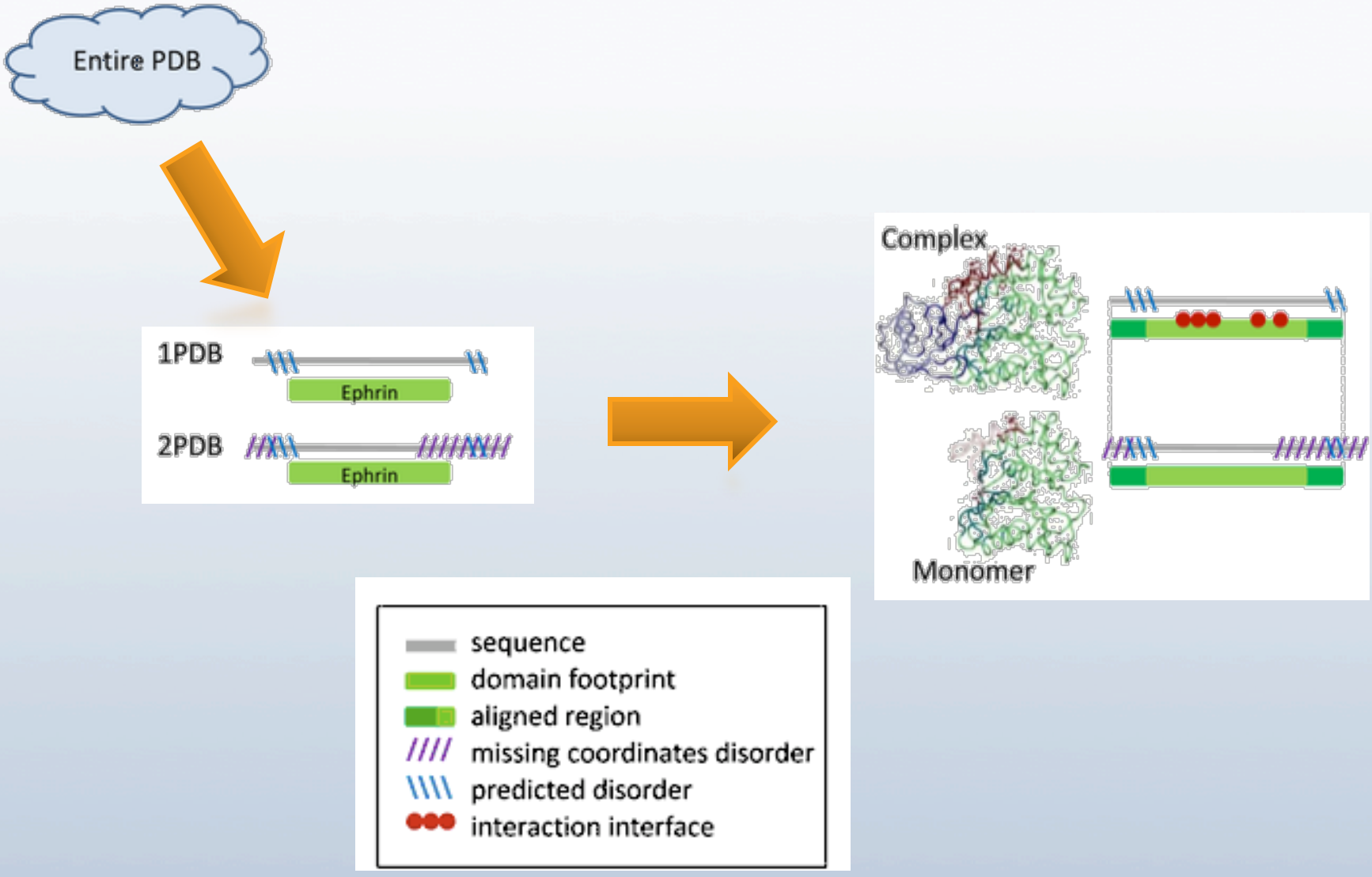
Ideas

- Construct a data set of intrinsically disordered regions in proteins and protein complexes.
- What is the functional importance of disordered regions?
- Classification of sequence patterns in disordered regions.

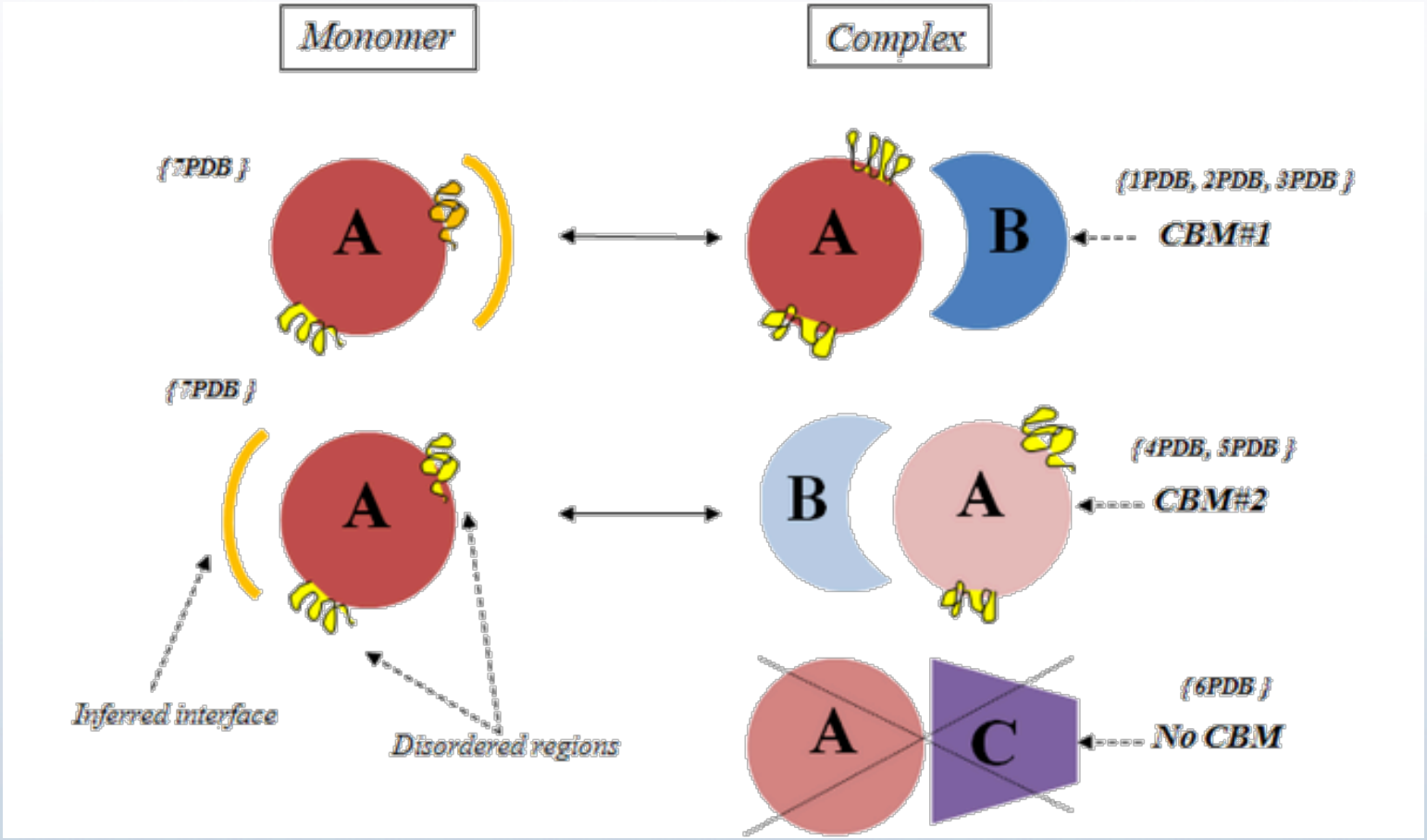
Defining Disordered Regions

- Regions with missing coordinates in X-ray solved structures.
- Regions with low packing density as predicted by FoldUnfold program.
- Confirmed disordered regions = missing coordinates + positive prediction.
- *Fraction Disorder* = #residue in disordered regions/ #residue in aligned regions.

Selection of the Data Set



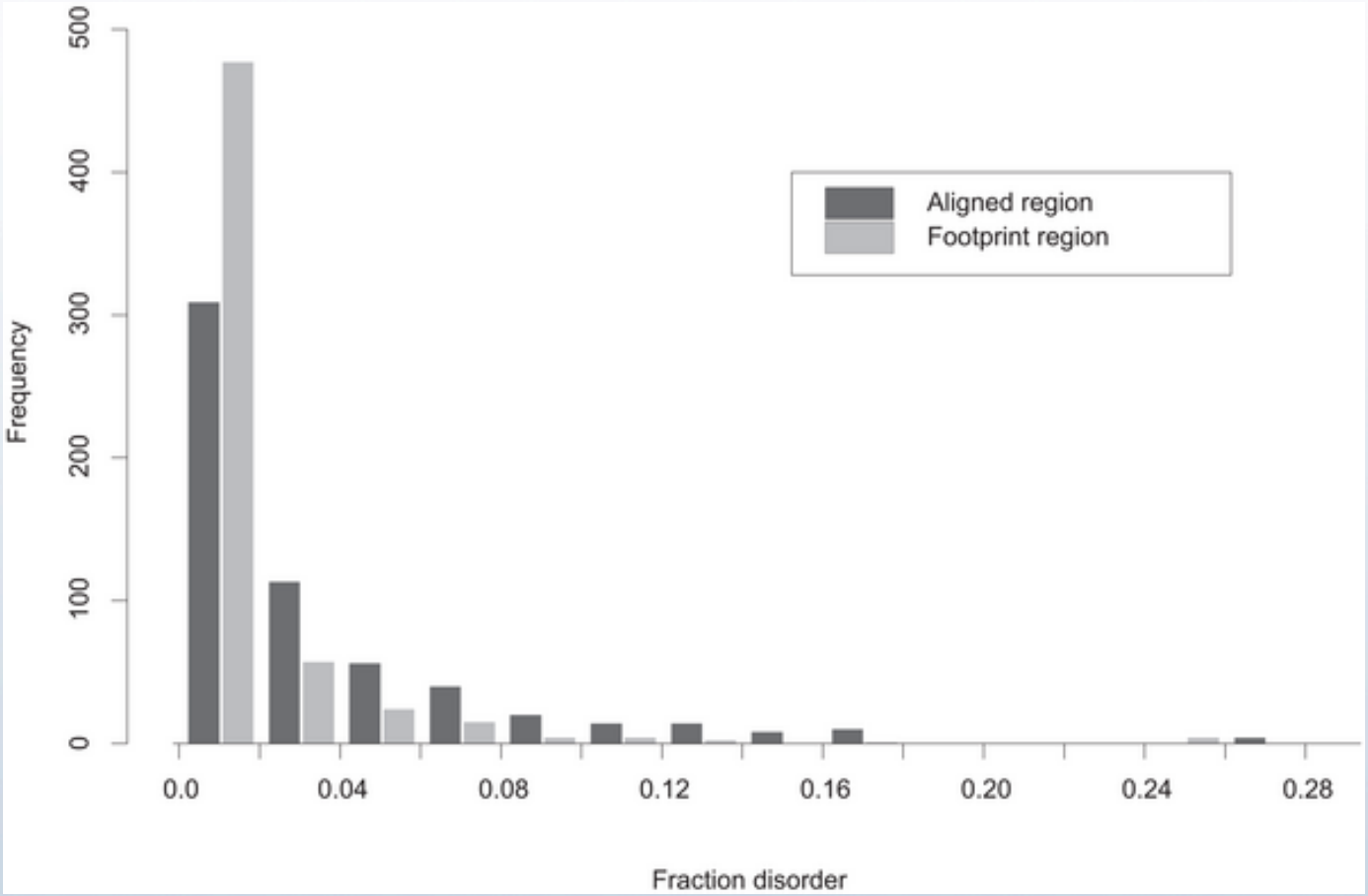
Biological Importance Refinement



Data Sets Statistics

- 4884 dimer complexes
- 418 unique monomers
- 588 conserved binding modes interactions
- 149 interactions used to analyse the disorder in mononmeric vs complexed state
- Each protein function is defined based on the GO functional annotation

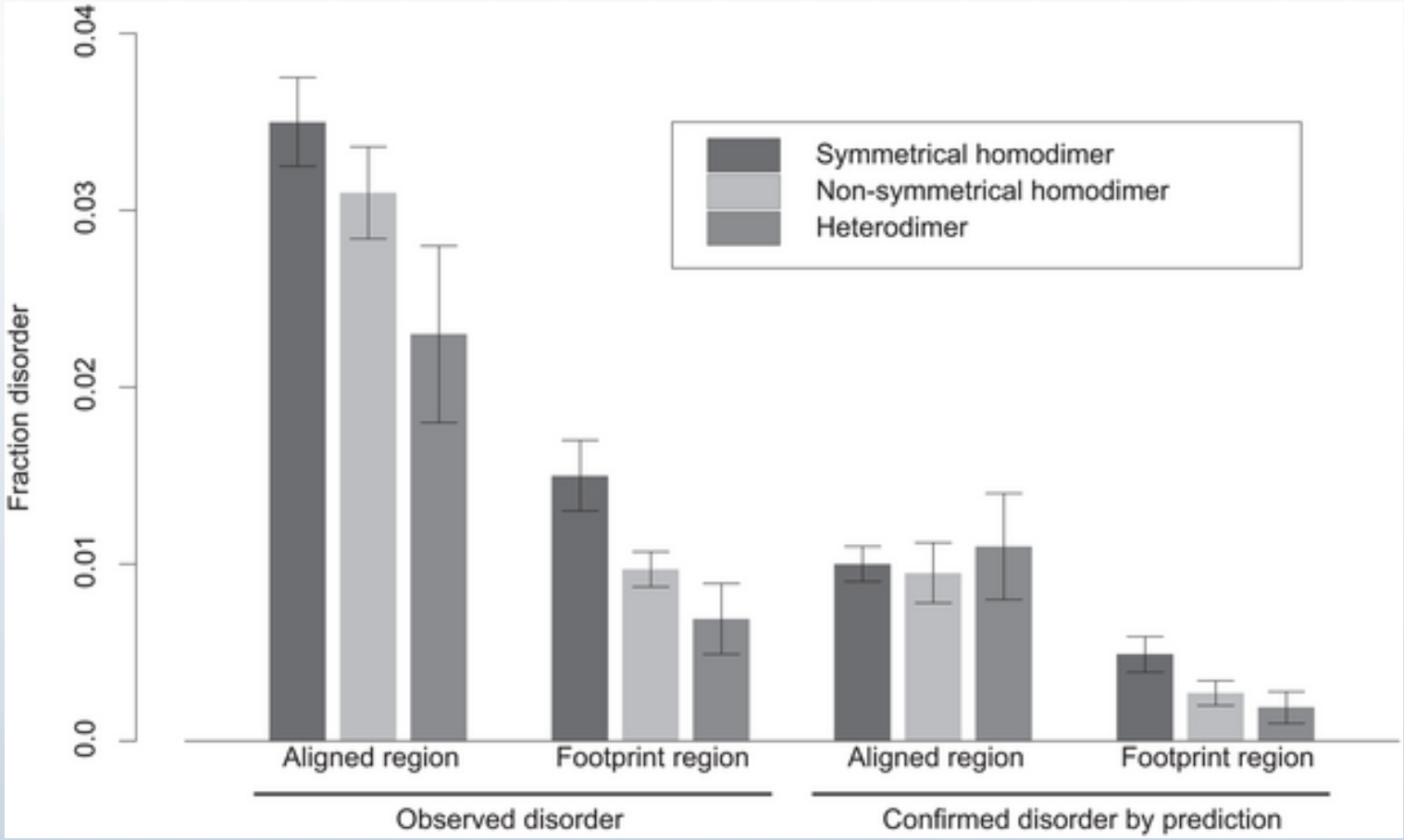
Protein Complexes Results



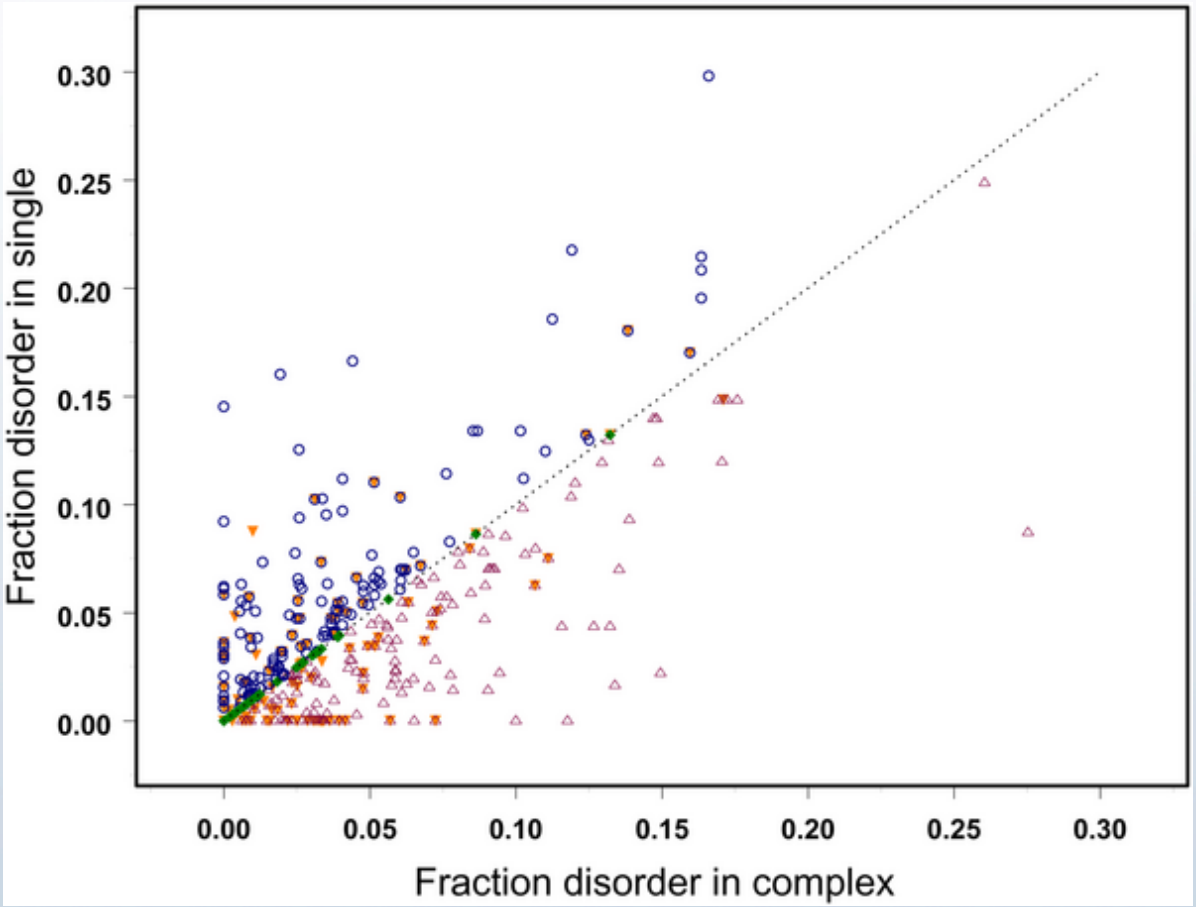
Examples

Family name, interacting partner, CBM	PDB code	DO, %	Function of disordered region
Chaperone hchA PRK04155 - PRK04155, CBM#83,80	1PV2	8.8 (8.4)	Disorder of loops D2 and D3 leads to the exposure of a hydrophobic patch of dimer interface that helps in binding to client proteins.
Holliday junction resolvases cd00523 – cd00523, CBM#5	1OB8	6.3 (8.8)	Catalytic Ser is located on disordered loop on the junction-binding surface. Accounts for specific binding of four-way DNA junctions.
Pyridoxamine 5′-phosphate oxidase PRK05679-PRK05679, CBM#26	1WV4	25 (28)	Disordered domain can rotate to allow passage of pyridoxal 5′-phosphate.
2-dehydro-3-deoxyphosphooctonate aldolase PRK05198-PRK05198, CBM#202,203,206	1D9E	7.0 (6.5)	Possible role of disorder in homotetrameric enzyme to be involved in synthase kinetic mechanism.
Thymidylate kinase PRK07933 – PRK07933, CBM#1192	1N5K	5.9 (7.2)	Disordered LID region closes on the phosphoryl donor when it binds. It anchors Mg ion, which establishes a link, through Glu166 and Asp9, between the P-loop and the LID region.
Lysin cd00243 - cd00243, CBM#3	2LYN	0.0 (6.9)	Disordered N- and C-termini are involved in the cleft formation which in turn is involved in an initial species-specific binding of the lysin dimer to VERL.
2-methylisocitrate lyase PRK11320 – PRK11320, CBM#72, 77, 74	1S2V	3.6 (2.3)	Disordered loop which is located near dimerization interface serves to gate the PEP mutase active site, converting between an open conformation that allows substrate binding and product release and a closed conformation that separates the reaction site from the solvent during catalysis.
HPr Serine kinase C-terminus, PTS HPr pfam07475 – pfam00381, CBM 9	1KKL	16.3 (4.8)	In complex with serine-phosphorylated Hpr, the disordered loop is a part of interaction interface. The phosphoserine forms an additional residue contact that helps to stabilize the loop.

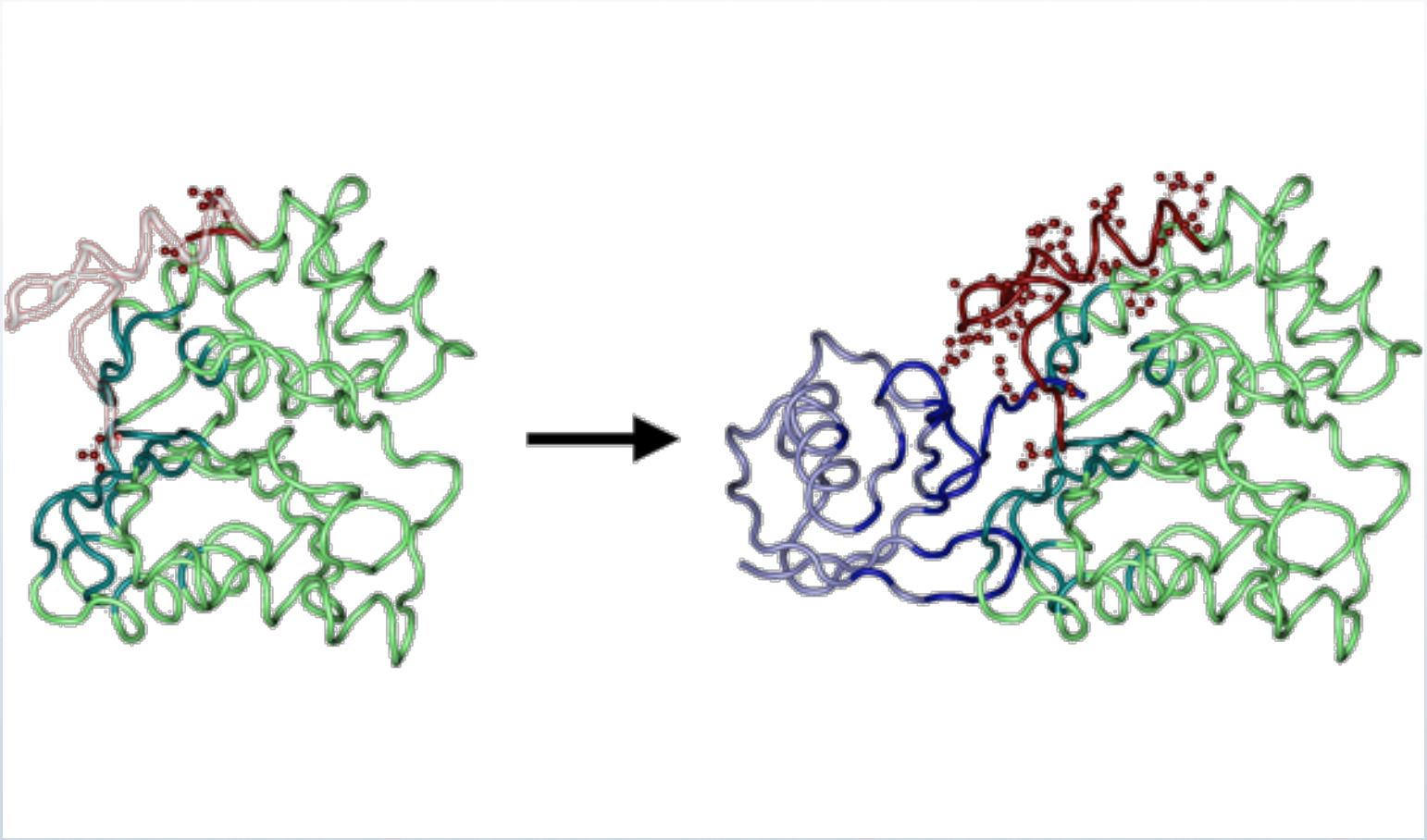
Homodimer vs heterodimer



Binding: Order or Disorder?



Order Upon Binding



Remarks and Observations

- Intrinsic disorder is functionally important in:
  - regulating the specificity of interactions between the dimer complexes and their interacting partners,
  - establishing the links between different residues upon allosteric regulation,
  - influencing the kinetics.
- The disorder content in homodimers, especially in symmetrical homodimers, is significantly higher than in heterodimers.

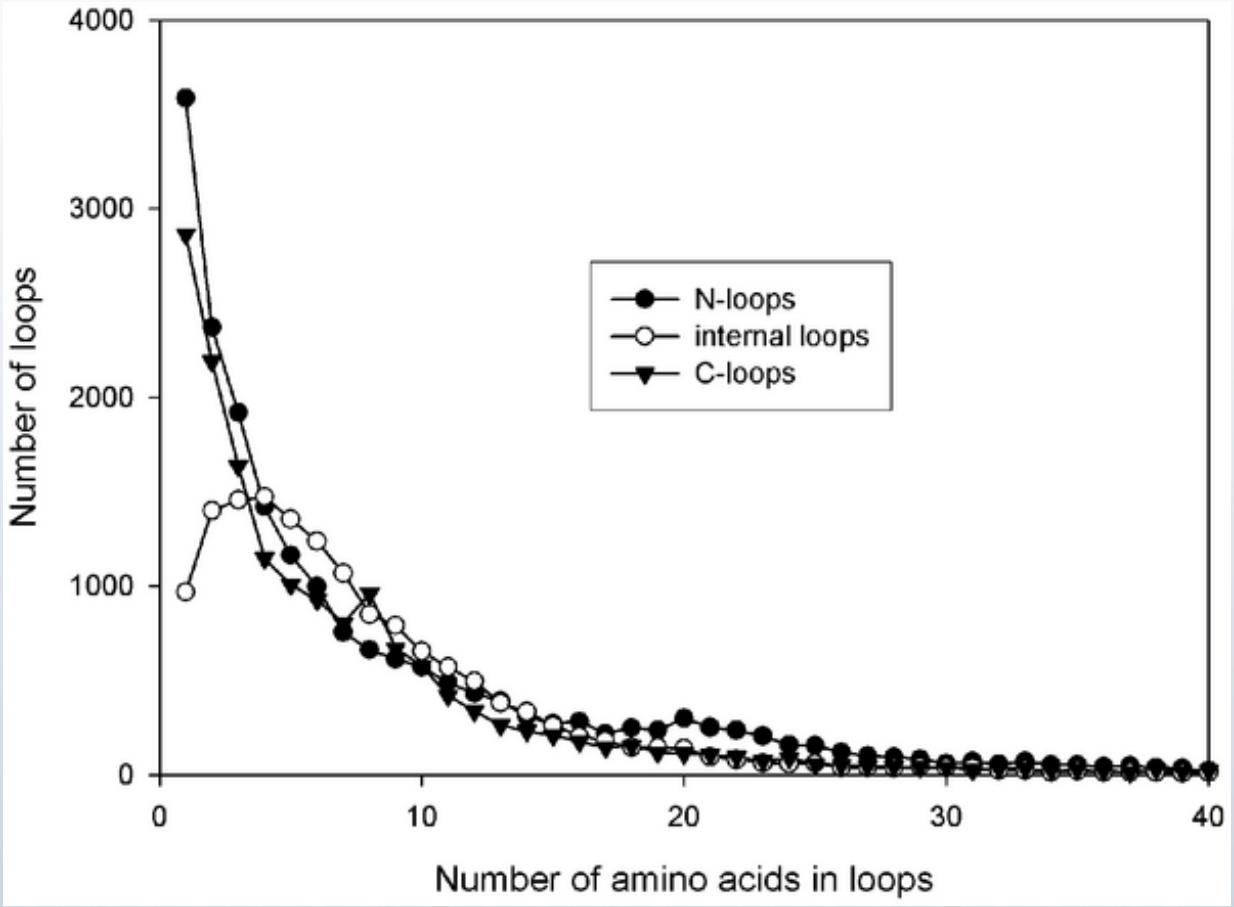
Resources on Disordered Proteins

- DISPROT – database of disordered proteins,
  – 500 structures.
- Predictions methods on disordered regions:
  - AAs analysis, physico-chemical properties,
  - Evolutionary conservation.
- Motif discovery:
  - PROSITE, InterPRo, Pfam, Casp, PEST.
- A library of disordered patterns.

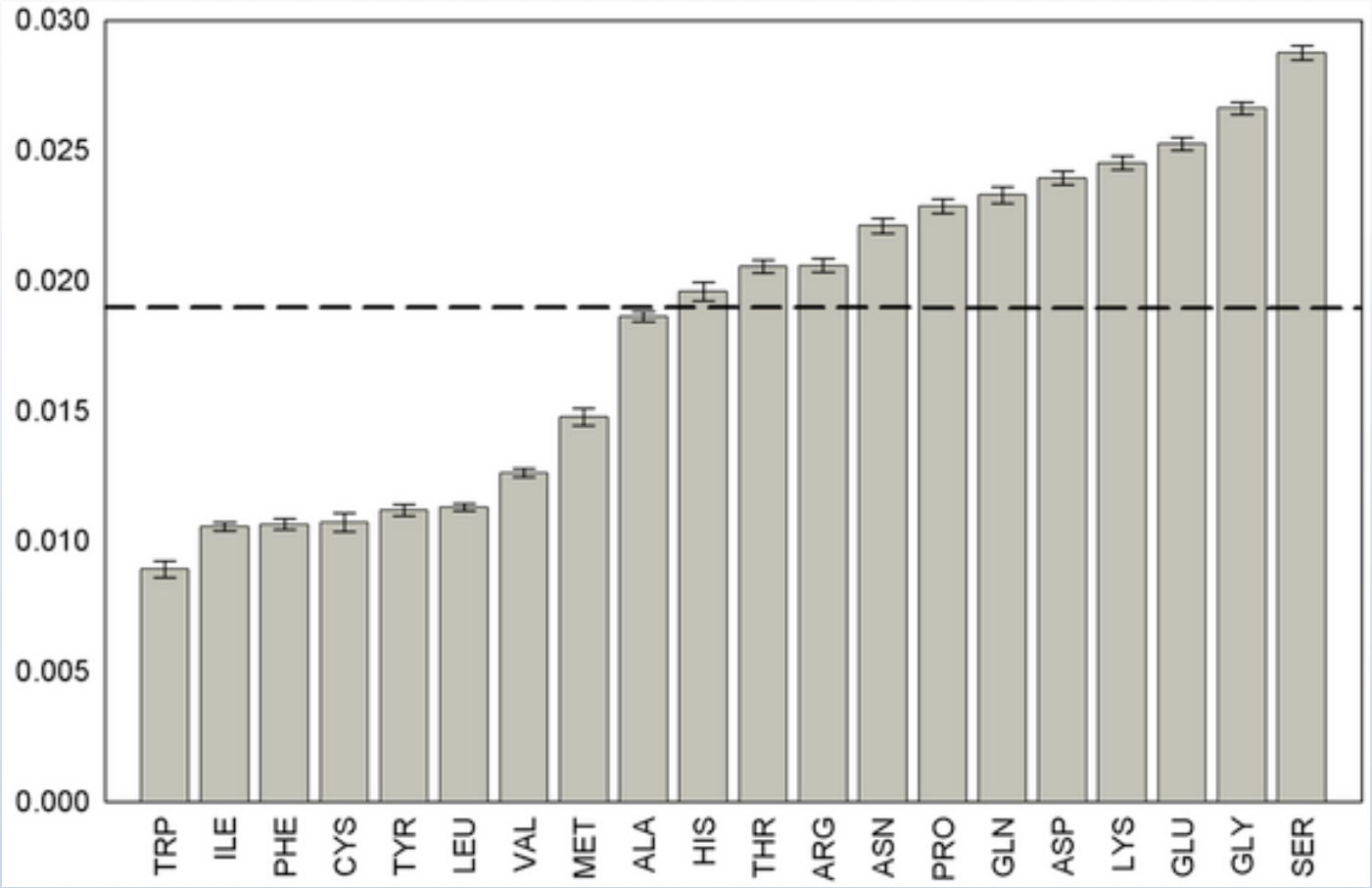
Methods

- Calculate degree of disorder for each residues.
- Sequence selection from PDB -> Disordered Residue Data Base (DRDB).
- Statistical quality assessment.
- Selection of patterns:
  - all residues of the fragment are disordered,
  - the length of a fragment is at least 6 Aas,
  - the fragment has occurrences in at least 5 other unique chains from DRDB.

Location, Location, Location



Frequency of Disordered Residues in the Middle Part of Protein Chain



AAs Distribution in Disordered Regions

a.a.	TRP	ILE	PHE	CYS	TYR	LEU	VAL	MET	ALA	HIS
N-40	0.032	0.054	0.061	0.044	0.055	0.077	0.069	0.351	0.134	0.427
C-40	0.029	0.046	0.045	0.047	0.038	0.063	0.054	0.065	0.090	0.376
middle	0.009	0.011	0.011	0.011	0.011	0.011	0.013	0.015	0.019	0.020
whole	0.015	0.022	0.022	0.022	0.021	0.028	0.027	0.093	0.046	0.166
a.a.	**THR**	**ARG**	**ASN**	**PRO**	**GLN**	**ASP**	**LYS**	**GLU**	**GLY**	**SER**
N-40	0.110	0.108	0.115	0.143	0.121	0.108	0.105	0.112	0.167	0.219
C-40	0.079	0.087	0.092	0.107	0.100	0.097	0.104	0.117	0.114	0.123
middle	0.021	0.021	0.022	0.023	0.023	0.024	0.025	0.025	0.027	0.029
whole	0.043	0.044	0.046	0.053	0.051	0.046	0.050	0.054	0.060	0.075

doi:10.1371/journal.pcbi.1000958.t002

Occurrence of Patterns in Eukaryotic Proteomes

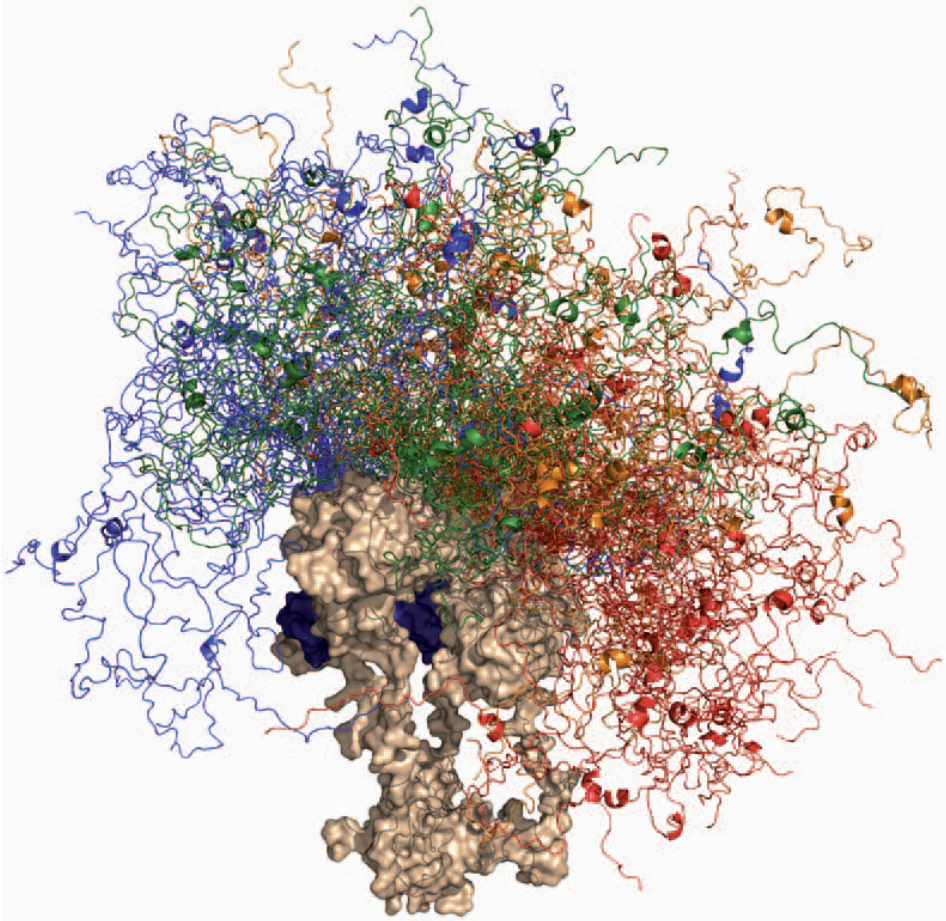
Pattern	Number of groups, identity inside group >20%	Fraction of disordered residues in the patterns from the DRDB	Probability of occurrence of the patterns in protein	Occurrence in the human proteome/in the DRDB	Occurrence in the fruit fly proteome/in the DRDB	Occurrence in the nematode worm proteome/ in the DRDB
PPPPPP	15	0.70	0.00017	703/32	304/32	247/32
QQQQQQ	11	0.66	0.00004	331/17	869/17	249/17
EEEDEE	55	0.65	0.00015	242/55	42/55	54/55
QPPPPP	9	0.74	0.00013	163/16	66/16	32/16
APAPAP	17	0.51	0.00067	121/30	44/30	34/30
HHHHHH	1227	0.93	0.00002	99/5423	133/5423	57/5423
EDEDEE	23	0.64	0.00014	97/29	27/29	42/29
DEEEED	12	0.68	0.00014	83/16	26/16	39/16
GGGGGSG	17	0.65	0.00028	78/29	80/29	8/29
GSSGSS	66	0.68	0.00120	67/93	35/93	19/93
PPPPPK	18	0.81	0.00027	62/31	24/31	32/31
DDEDED	14	0.64	0.00013	53/16	31/16	26/16
SGGGGSG	10	0.82	0.00022	31/29	19/29	2/29
KKKGKK	26	0.55	0.00181	27/56	8/56	13/56
EEEEAP	12	0.66	0.00028	26/21	6/21	9/21
KKRKRK	12	0.54	0.00067	25/19	6/19	7/19
SGGGSGG	12	0.68	0.00024	20/17	17/17	5/17
SHHHHH	558	0.98	0.00005	19/1566	27/1566	12/1566
GGSGSGG	17	0.77	0.00027	14/50	23/50	6/50
NHHHHH	19	0.83	0.00003	10/25	14/25	8/25

doi:10.1371/journal.pcbi.1000958.t003

Conclusions & Observations

- Library of disordered patterns: 109 patterns.
- Functional categories in GO for 3 patterns: PPPPPP, QQQQQQ and HHHHHH.
- Molecular functions for the proteins including disordered patterns:
  - actin binding, calcium ion binding, DNA binding, nucleic acid binding, protein binding, protein serine/threonine kinase activity, receptor activity, Rho GTPase binding, RNA binding, SH_3 domain binding, signal transducer activity, transcription coactivator activity, transcription factor activity, tropomyosin binding, voltage-gated potassium channel activity, and zinc ion binding.

Protein Mess vs Eaton Mess



ThankYOU!