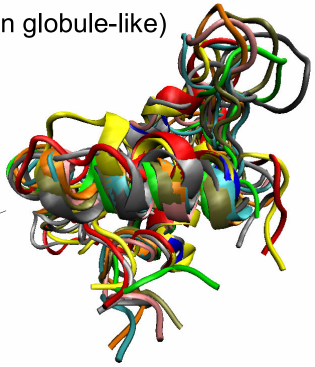
Disorder in Proteins: Current Characterization Efforts

Lukasz Kurgan

GLBIO 2012

Disorder in proteins

- Intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs) exist as dynamic conformational ensembles which can be
  - collapsed-disordered (molten globule-like)
  - partially collapsed-disordered (pre-molten globule-like)
  - extended disordered (coil-like)



NMR solution structures (10 conformations) of the globular domain (residues 41–113) of the *G. gallus* (chicken) histon H1

Peng Z, Mizianty M, Xue B, Kurgan L, Uversky V. *Molecular BioSystems* 2012, DOI: 10.1039/c2mb25102g

Disorder in proteins

- DisProt
  - curated database
  - 653 proteins and 1421 disordered regions
- Disorder is abundant and (relatively) little explored
  - functions of disorder were studied in human and yeast proteomes
    - involved in regulation of transcription, cell signalling, kinase activity, and nucleic acid binding

Kingdom	**# sequences**	**% disorder**
Archaea (6 species)	11,742	3.8
Bacteria (13 species)	35,389	5.7
Eukaryota (5 species)	88,531	18.9
PDB (non-redundant at 95% sequence identity)	7,169	3.2

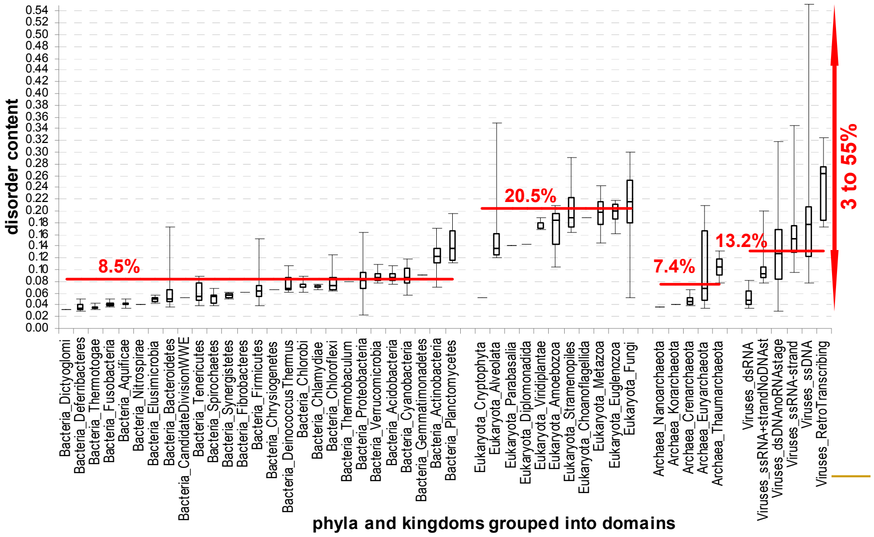
Sickmeier M, et al. *Nucleic Acids Res.* 2007; 35:D786-93
Ward JJ, et al. *J Mol Biol.* 2004; 337(3):635-45
Lobley A, et al., *PLoS Comput Biol.* 2007; 3(8):e162
Pentony MM and Jones DT, *Proteins* 2010; 78(1):212-21

Large scale characterization of disorder

- The setup
  - 965 complete proteomes and 6.4+ million proteins
    - 3.6% proteins from 59 species in archaea
    - 66.6% proteins from 471 species in bacteria
    - 29.5% proteins from 110 species in eukaryota
    - 0.3% proteins from 325 viral proteomes
  - consensus of 5 disorder predictors
    - compared to one predictor used in prior studies
  - analysis of
    - generic disorder characteristics (abundance and distribution)
    - disorder and post-translational modifications (PTMs) based on UniProt
    - disorder and protein functions based on Gene Ontology (GO)
    - disorder in cellular components based on GO

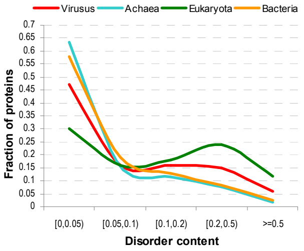
Large scale characterization of disorder

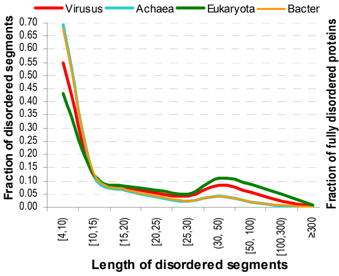
- Abundance of disorder

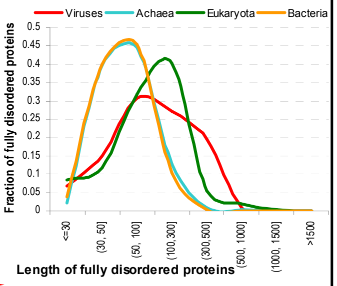


Large scale characterization of disorder

- Generic characteristics of disorder







79% archaea
77% bacteria
63% viruses
46% eukaryota

36% eukaryota
21% viruses

69% archaea
67% bacteria
55% viruses
43% eukaryota

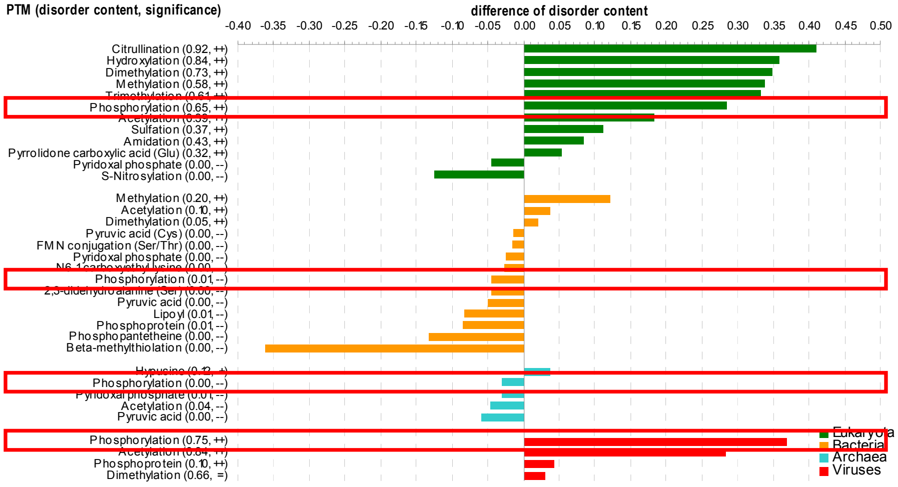
25% eukaryota
16% viruses

89% bacteria
86% archaea
53% viruses
52% eukaryota

20% viruses
8% eukaryota

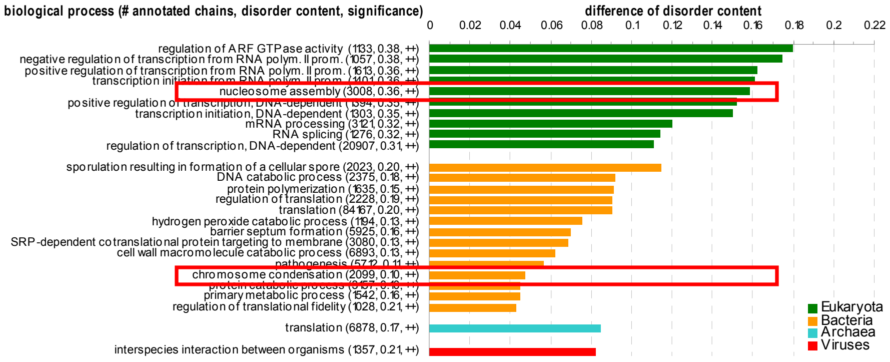
Large scale characterization of disorder

- Disorder in PTM sites



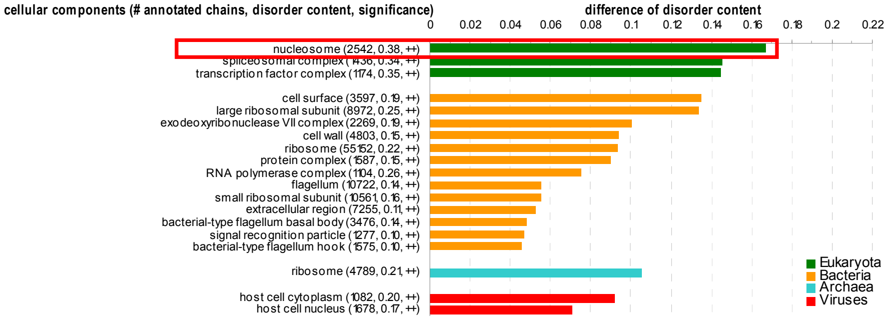
Large scale characterization of disorder

		Archaea	Bacteria	Eukaryota	Viruses
biological processes	total # of processes	12	318	104	2
	# of processes enriched in disorder	1	14	10	1



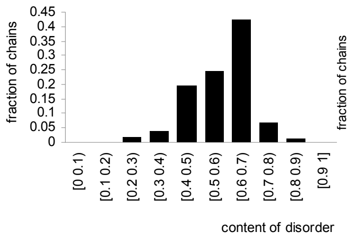
Large scale characterization of disorder

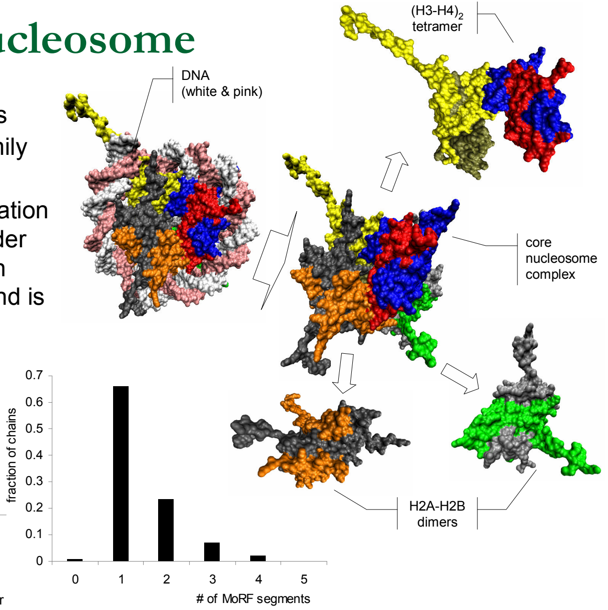
		Archaea	Bacteria	Eukaryota	Viruses
cellular components	total # of components	6	61	50	5
	# of components enriched in disorder	1	13	3	2



Disorder in nucleosome

- 2007 histones; 746 species
- all members of histone family are intrinsically disordered
- plays role in heterodimerization and formation of higher order oligomers, interactions with DNA and other proteins, and is enriched in PTM sites
- is highly conserved





Peng Z, Mizianty M, Xue B, Kurgan L, Uversky V. *Molecular BioSystems* 2012, DOI: 10.1039/c2mb25102g

Acknowledgements

People

Zhenling Peng	Ph.D. student
Marcin Mizianty	Ph.D. student
Jing Yan	Ph.D. student
Dr. Keith A. Dunker	Indiana University
Dr. Vladimir Uversky	University of South Florida

Funding

NSERC
CRSNG
Natural Sciences and Engineering Research Council of Canada

Alberta Innovates
Technology Futures