Intrinsically Disordered Proteins

A. Keith Dunker
Department of Biochemistry and Molecular Biology
Indiana University School of Medicine
(kedunker@iupui.edu)

Tuesday, February 26, 2019

Order and Disorder in Biology and Human Disease:

Intrinsically Disordered Proteins

Institute for Quantiatative Biomedicine

Chemistry & Chemical Biology

Rutgers University

New Brunswick, NJ

Center For Computational Biology and Bioinformatics

School of Medicine INDIANA UNIVERSITY

Protein Structure/Function

Current Protein Structure/ Function Paradigm

Amino Acid Sequence

"Folding Problem"

↓

3-D Structure

Native = Ordered = Structured

↓

Protein Function

["Lock & Key"; "Induced Fit"]

Intrinsically Disordered Proteins (IDPs) and IDP Regions

- Some proteins & regions lack structure, yet carry out function.
- We call these intrinsically disordered proteins (IDPs) and IDP Regions.

Definition: Intrinsically Disordered Proteins (IDPs) and IDP Regions

Whole proteins and regions of proteins are intrinsically disordered if:

- they lack stable 3D structure under physiological conditions, and if:
- they are flexible molecules that form dynamic ensembles with inter-converting configurations and without particular equilibrium values for their coordinates.

What led me to become interested in Intrinsically Disordered Proteins (IDPs)?

1. An IDP region in TMV coat protein undergoes a disorder-to-order transition as it binds to TMV RNA during virus assembly.

   Holmes KC. Ciba Found Symp. 93:116-38 (1983)

2. Conversion of fd phage capsid from structure to molten globules enables the fd coat protein to insert into model membrane vesicles; fd coat protein loses structure but gains function.

   Dunker AK et al., FEBS Lett 292: 275-278 (1991)

Uversky's Rule of Three

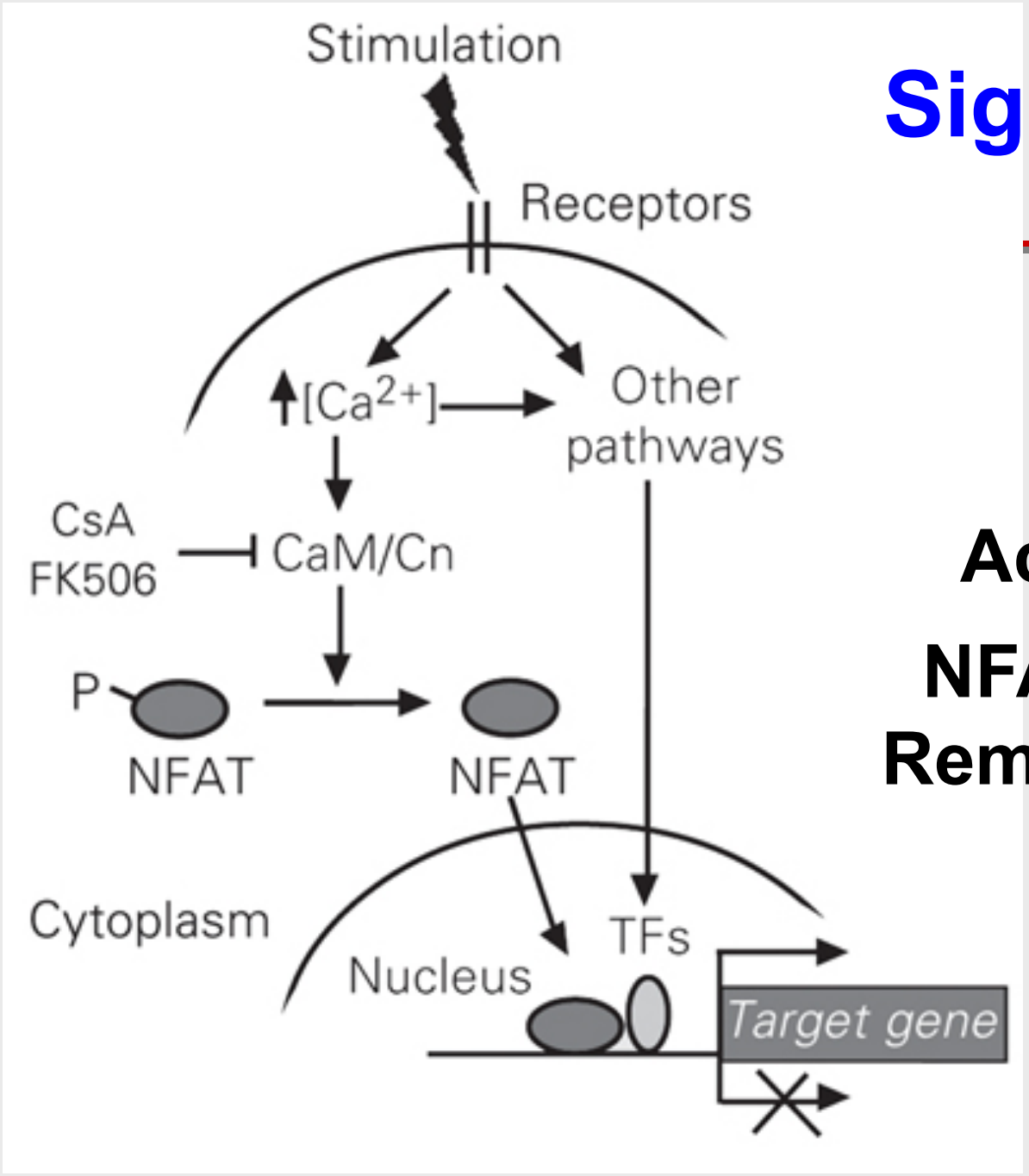
Vladimir Uversky

"Three encounters with IDPs are needed before a researcher takes them seriously."

Close IDP Encounter of the Third Kind, Trigger for my IDP Research

Seminar describing an important IDP
12 Noon to 1 PM, 15 November, 1995
Washington State University

Given By Chuck Kissinger
BS / MS Washington State University
PhD University of Washington
Johns Hopkins / MIT Post Doc
Aguoron Pharmaceuticals



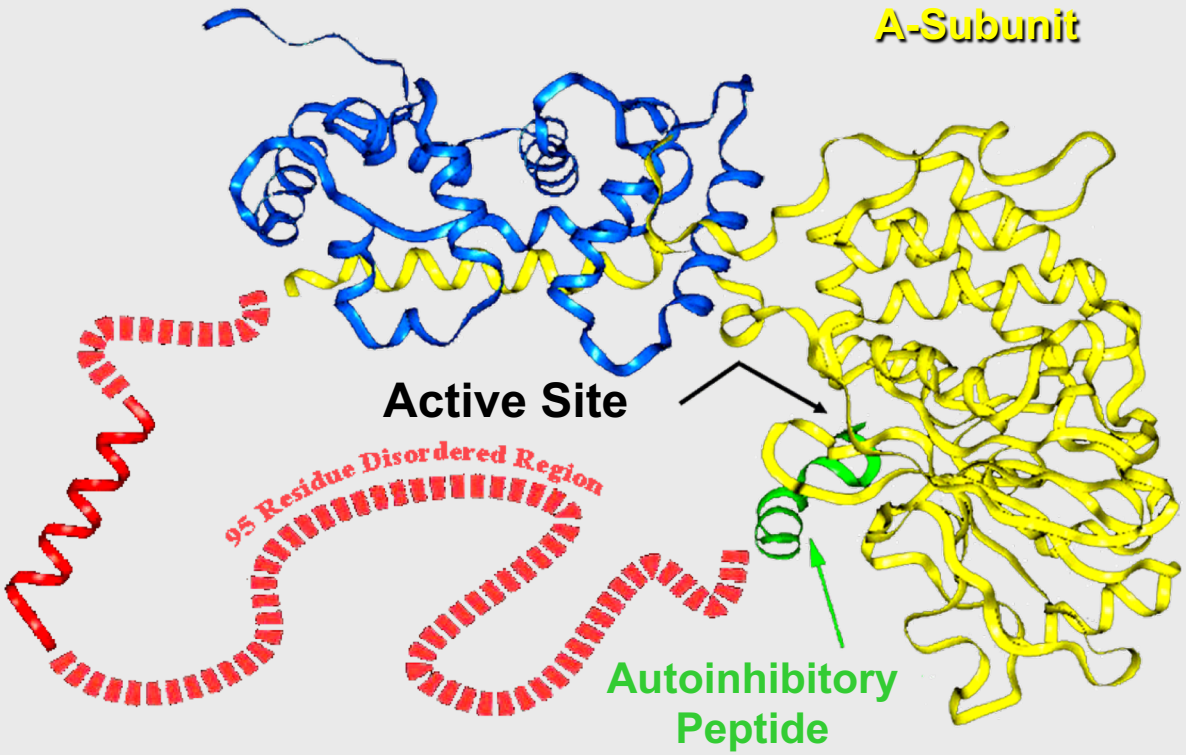
Signaling Pathway

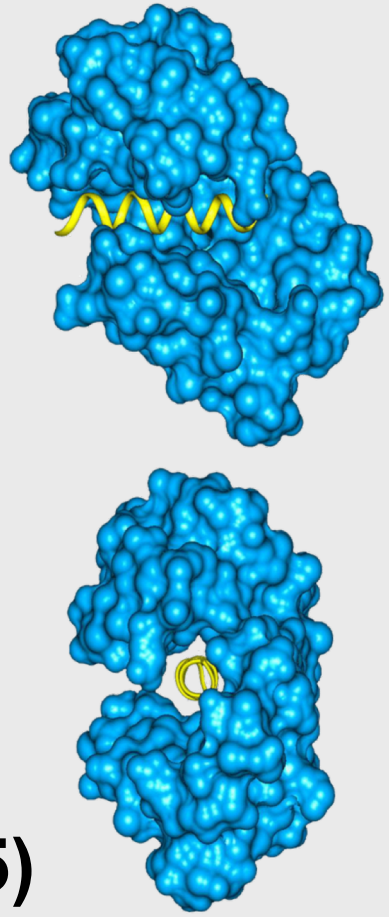
Calmodulin (CaM)
Calcineurin (Cn)
Nuclear Factor of Activated T- Cells (NFAT)

NFAT-poly-P in an IDP tail.
Remove Ps, activates NLS
→ NFAT → nucleus
→ turns on genes
→ T-cells activated
→ reject transplant

Calcineurin and Calmodulin

Meador W et al., *Science* 257: 1251-1255 (1992)





Kissinger C et al., *Nature* 378:641-644 (1995)

Key Points

- **IDP function: on-off switch for CaN;**
- **CaN activated by Ca^{2+}/ CaM – such activation is a well known, very important mechanism for regulating many enzymes and pathways;**
- **CaN is a phosphatase; phosphorylation / de-phosphorylation is a very important, frequently used mechanism for many signaling pathways;**
- **Overall, CaN's IDP region sits at the nexus of two extremely important signaling pathways!!**

After Seminar Questions: Nov 15, 1995

- Why don't IDPs and IDP regions fold into 3D structure?
- How common are IDPs and IDP regions?
- What are the functions of IDPs and IDP regions?

Why don't IDPs fold into 3D structure?

- *Amino acid composition* determines whether a protein will fold or remain unfolded.
- For compositions that favor structure, the sequence patterns of hydrophobic / hydrophilic groups determine which *3D structure* is formed.

Shakhnovich, E.I. and Gutin, A.M. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA* 90: 7195 – 7199 (1993). *Did not propose that IDPs exist in nature !*

Why don't IDPs fold into 3D structure?

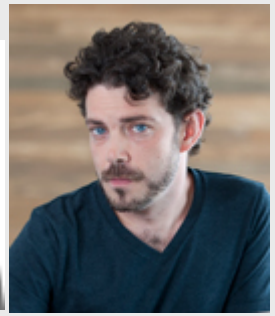
First step: collect structured proteins from PDB and also collect IDPs / IDP regions.

- X-ray Structures from PDB: structured regions and MED regions
- NMR Structures from PDB: invariant regions and highly variable regions
- Literature, one-by-one examples: whole protein disorder (IDPs) from CD or NMR spectra

Why don't IDPs fold into 3D structure?

Xie et al., *Genome Informatics* 9: 193-200 (1998)

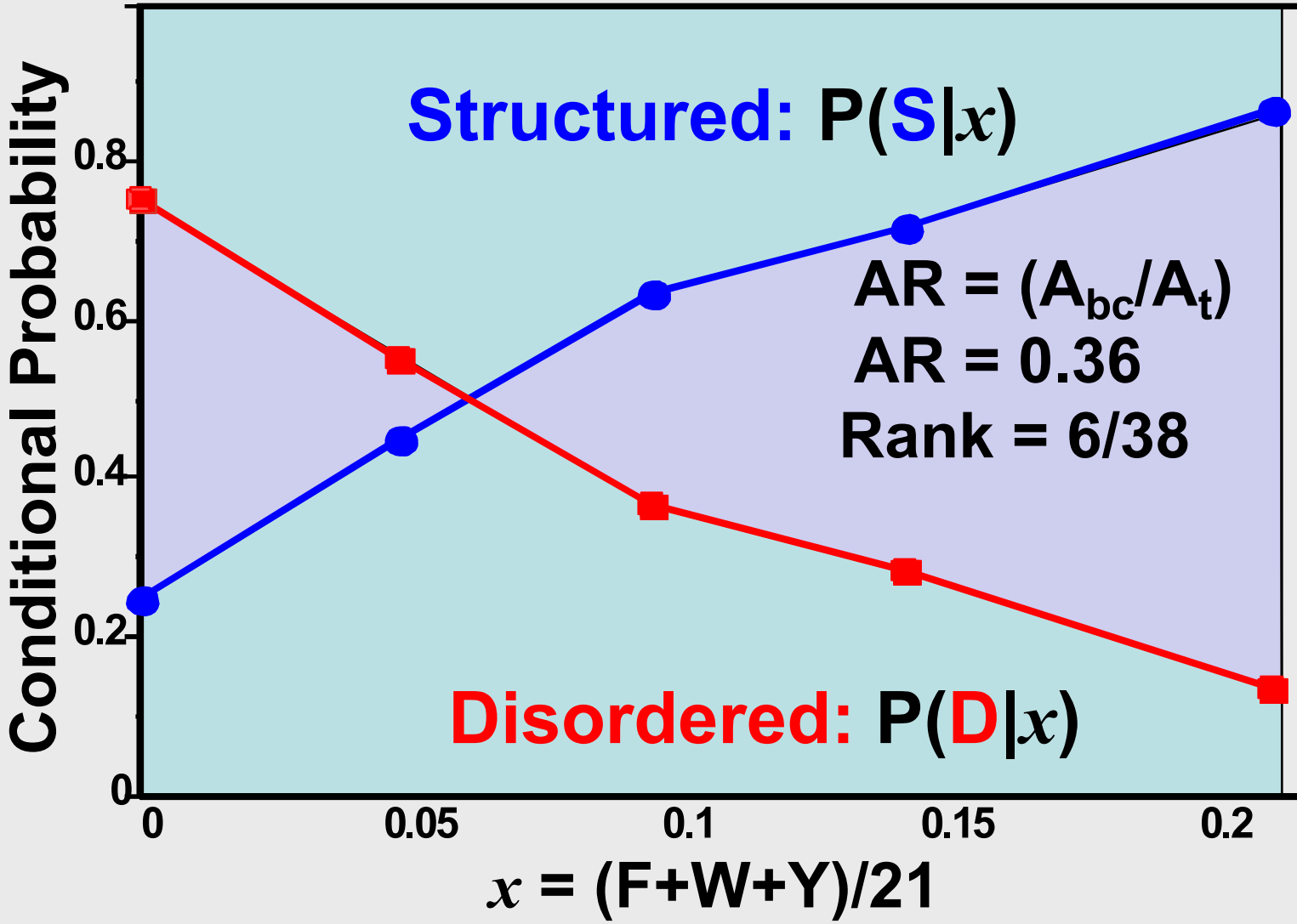
Qian Xie



Ethan Garner

Zoran Obradovic

Pedro Romero

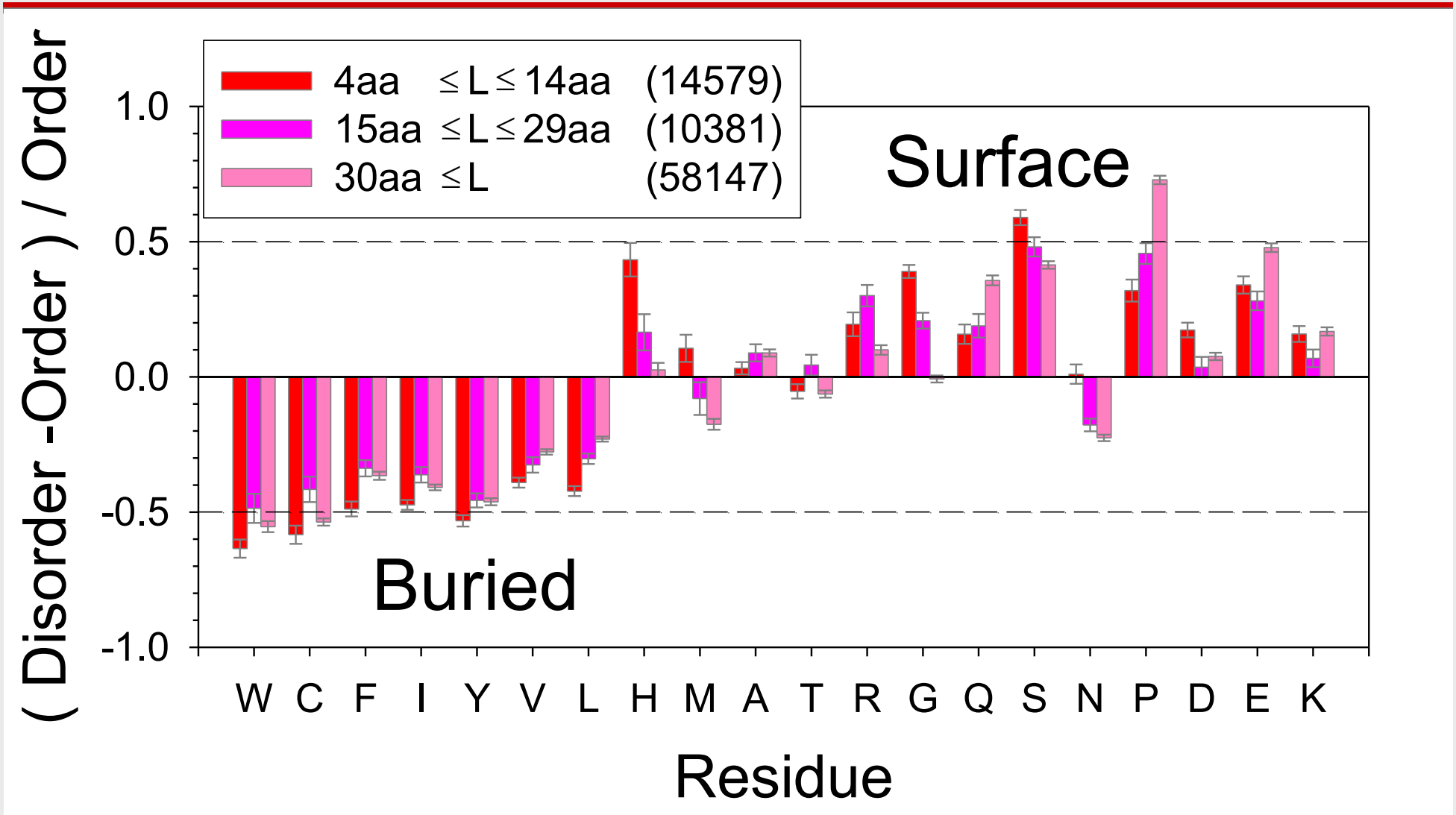


Why don't IDPs fold into 3D structure? Amino acid sequence favors nonfolding!

- IDPs have too few aromatics – aromatics are important for the stability of hydrophobic cores;
- IDP ratio of hydrophilic amino acids to hydrophobic amino acids is too high for folding;
- IDPs have too low of a sequence complexity
- IDPs have too large of a ***net*** charge – charge repulsion inhibits folding;
- IDPs have too many prolines – prolines cannot form backbone H–bond, so helices and sheets are destabilized by prolines.

Why don't IDPs fold into 3D structure?

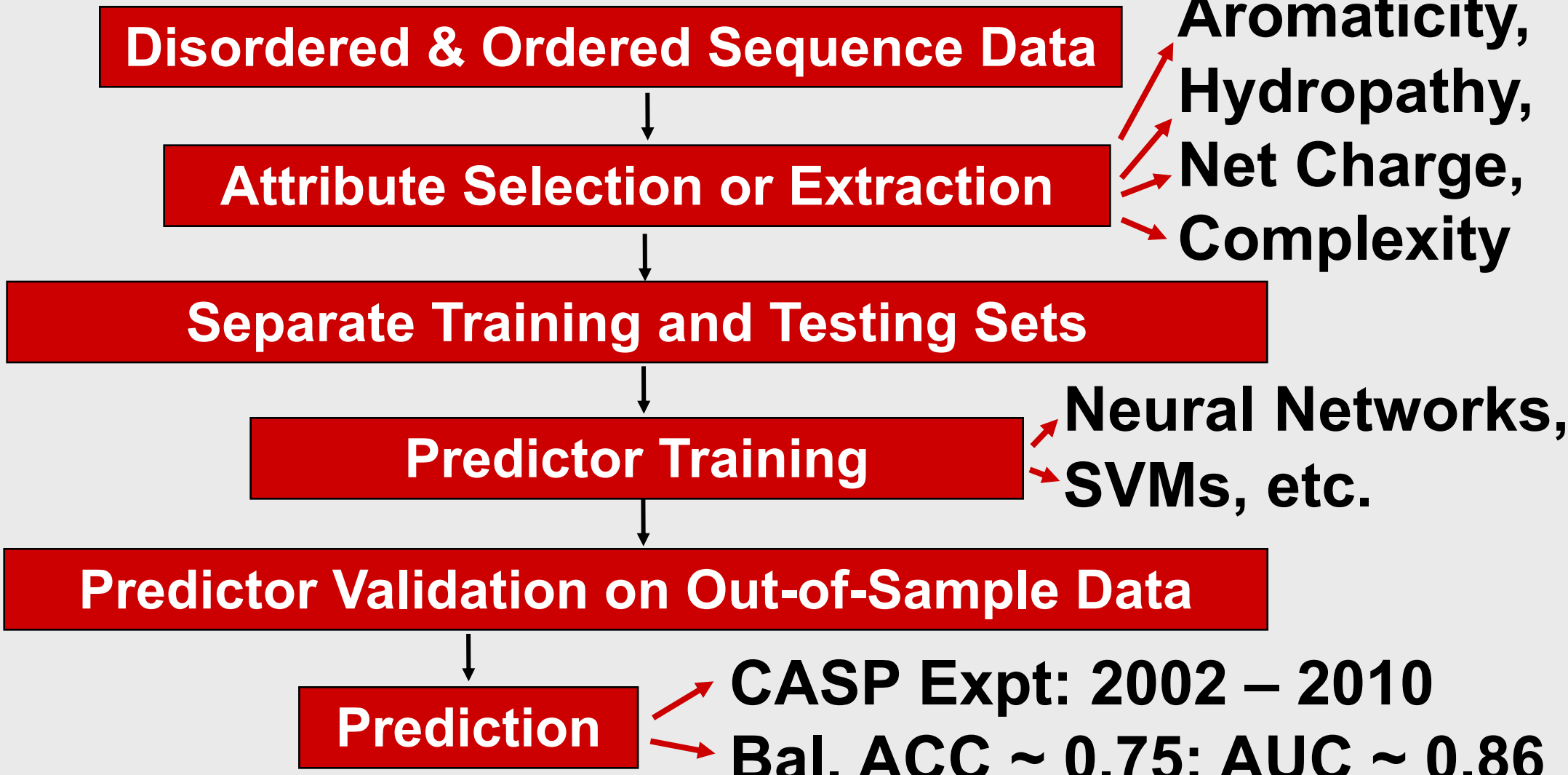
Dunker et al., *Adv. Prot. Chem.* 62: 25-37 (2002)



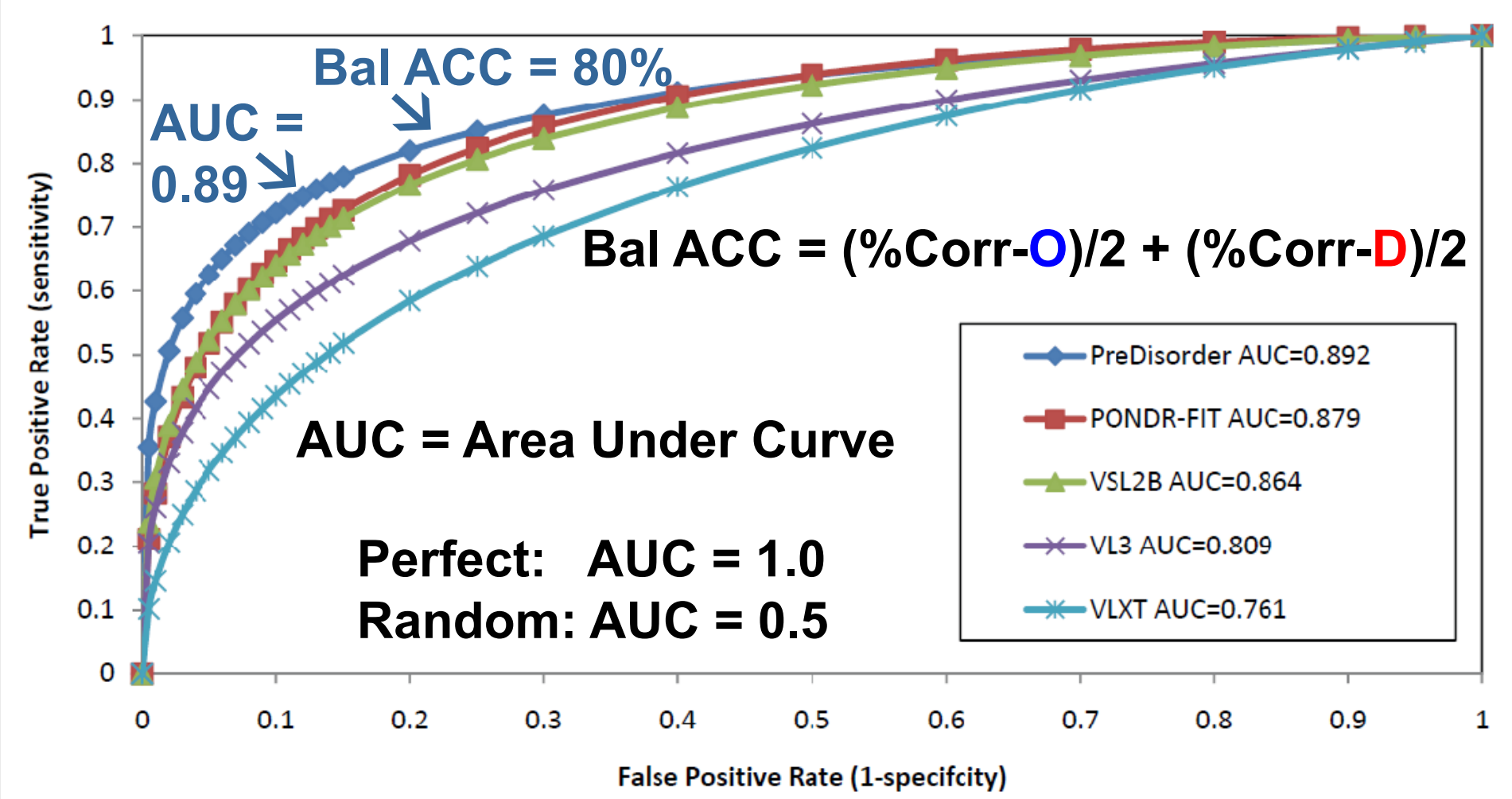
How common are IDPs?

- Using amino acid compositional differences between structured proteins and IDPs and IDP regions, develop order / disorder predictor;
- Validate predictor on “out-of-sample” data;
- Apply predictor to amino acid sequences of whole proteomes.

Prediction of Intrinsic Disorder

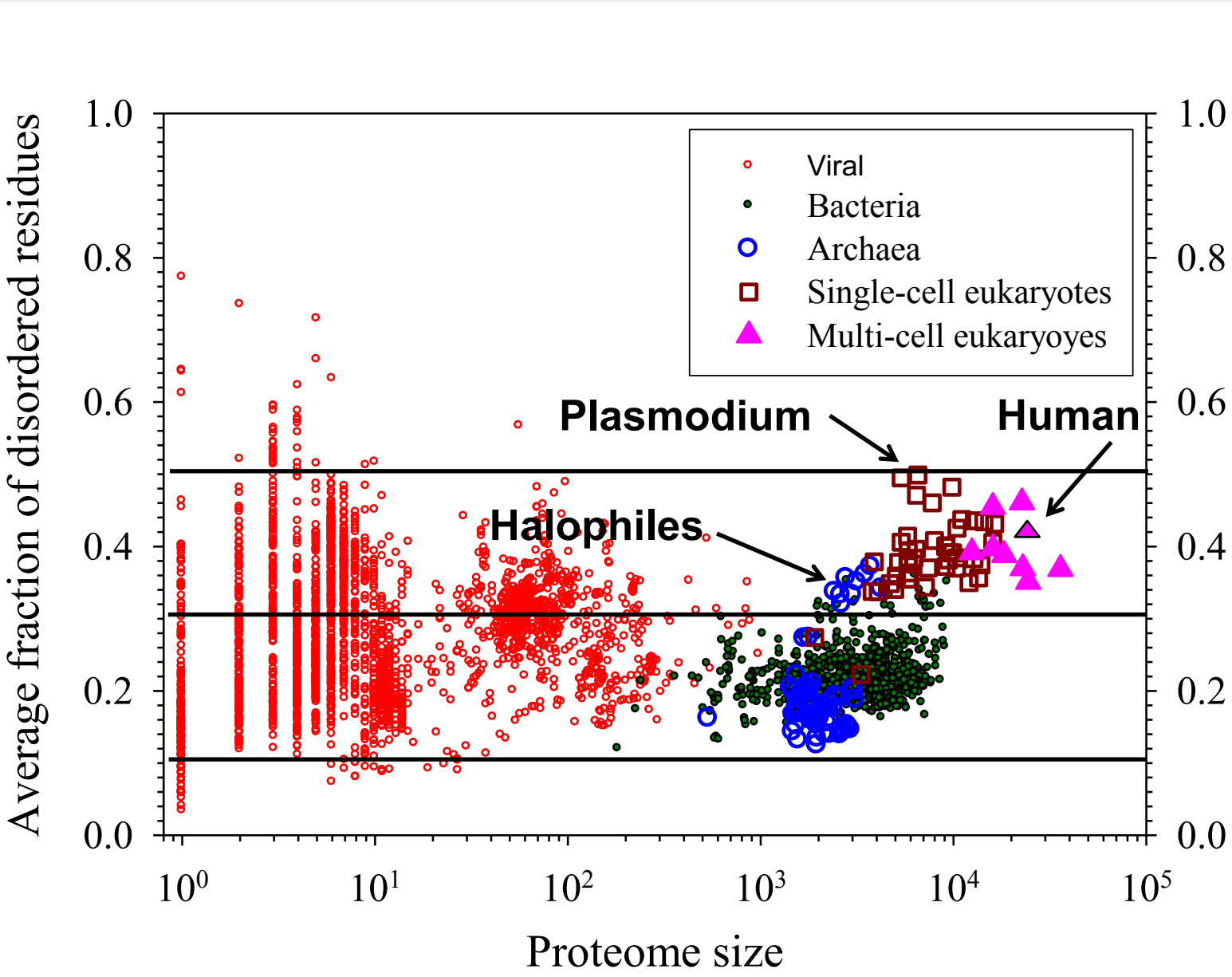


Comparison on CASP 8 Dataset



Zhang P, et.al. (unpublished results; not quite same as CASP evaluation)

How common are IDPs?



Bin Xue

Vladimir Uversky

Xue et al., *J Biomol Struct Dyn* 30: 137-149 (2012)

How common are IDPs?
More recent, improved approach

Combine structure / disorder prediction and structure prediction *by sequence similarity* to all currently known protein 3 D structures.

For the human proteome:

Fukuchi, S., et al., Binary classification of protein molecules into intrinsically disordered and ordered segments. BMC Struct Biol. 11:29 (2011); For Human: 35% residues are in IDPs or IDP regions. (Weakness → used Pfam for structured proteins)

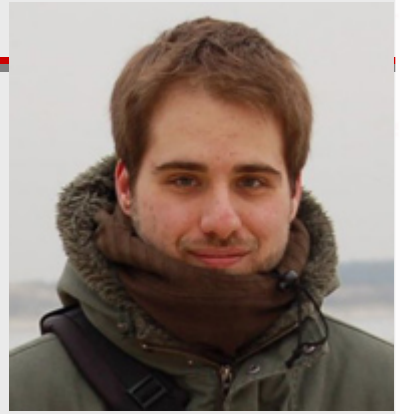
For 1,765 proteomes (8 different order / disorder predictors):

Oates, M.E. et al., D^2P^2: database of disordered protein predictions. Nucleic Acids Res. 41(Database issue):D508-516 (2013). For Human: 35% - 50% residues in IDPs or IDP regions.

(Strength → used SUPERFAMILY for structured proteins)

Human BIN1 from D²P²

Two transcripts from one gene;

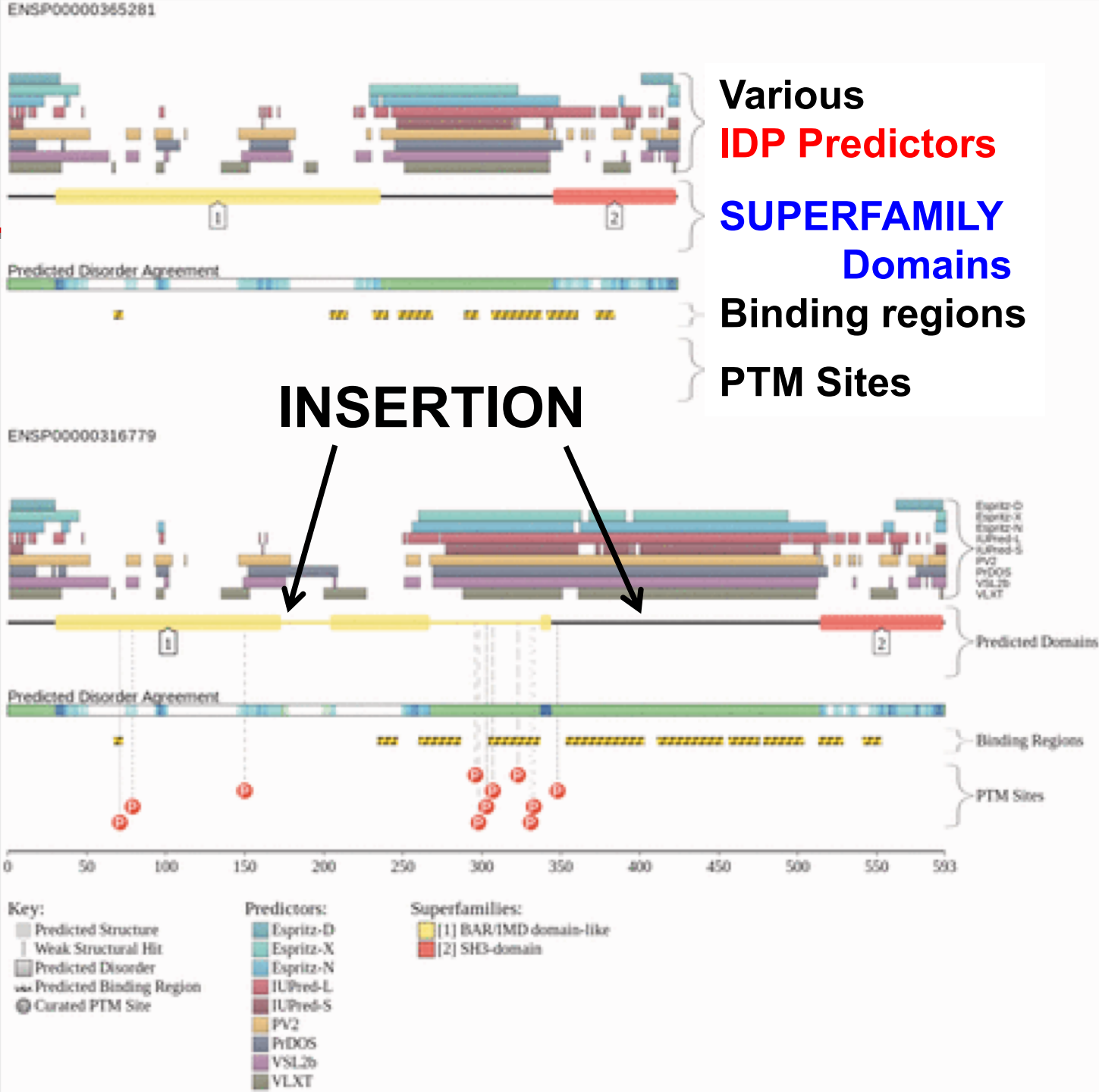


Matt Oates

Julian Gough

Insertion from alternative splicing.

Oates et al., NAR 41: D508-516 (2013)



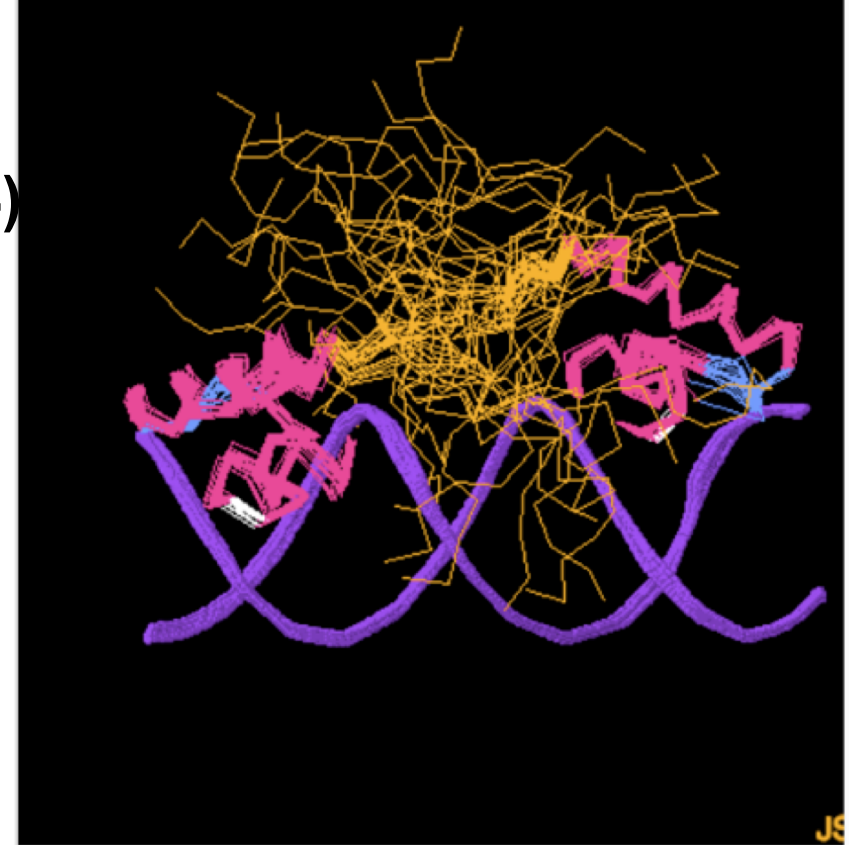
What are the functions of IDPs?

- Individual examples of IDPs and IDP regions and their functions: (calcineurin – CaN), lac repressor, signaling domain partners, p53, BRCA1; (p21/p27/p57)

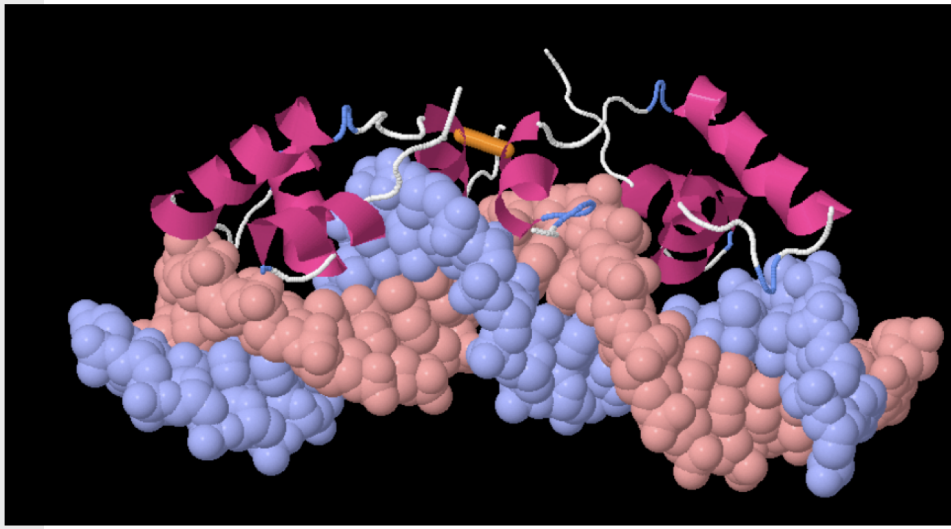
- Bioinformatics study to comprehensively determine functions of structured proteins and of IDPs and IDP regions.

The Lac Repressor

Kalodimos et al., Science 305:386-389 (2004)



- Upon binding to nonspecific DNA, a large segment of the Lac Repressor remains an IDP region that interacts transiently with DNA phosphates.
- Upon encountering its binding sequence, the IDP region → structure and is involved in recognizing the cognate DNA binding sequence and in increasing the binding affinity. Also, the DNA becomes bent.



Proteopedia, Life in 3D, the free, collaborative 3D Encyclopedia was used for these images – provided by:

Joel Sussman

IDPs & Function:
Signaling Domain Partners

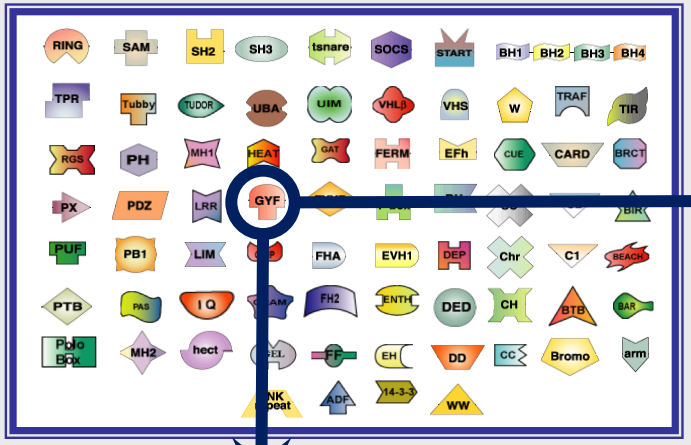
More than 100 signaling domains such as SH1, SH2, PDZ, GYF, etc. Most of these these domains bind to IDP regions. Discuss only GYF domain.

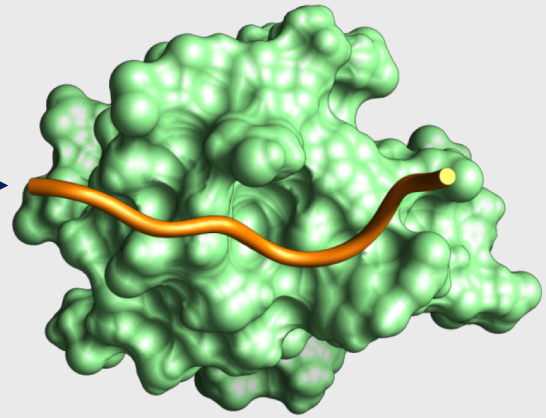
- **GYF domain: has GP[YF]xxxx[MV]xxx[GN]YF motif;**
- **GYF domain also known as CD2BP2 and other names;**
- **CD2: "cluster of differentiation"2 – on surface of T-cells;**
- **CD2 contains an IDP region that binds to the GYF domain.**

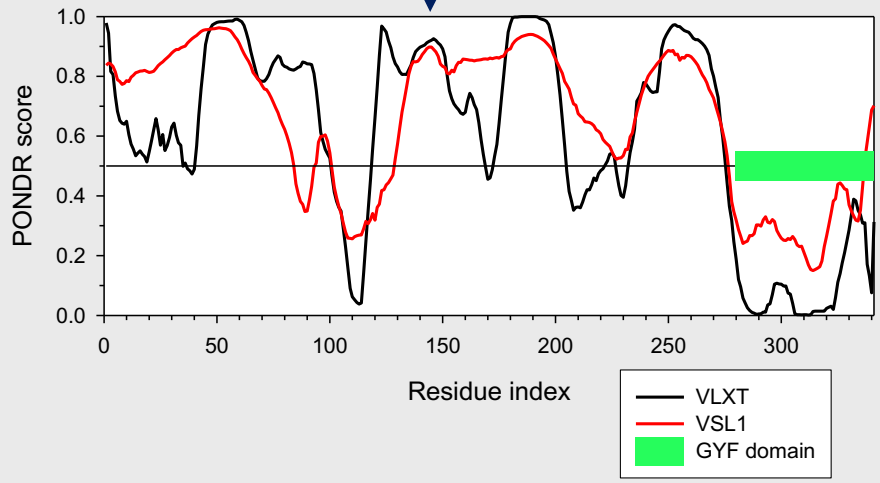
Signaling Domains (SH2, SH3) discovered by Tony Pawson

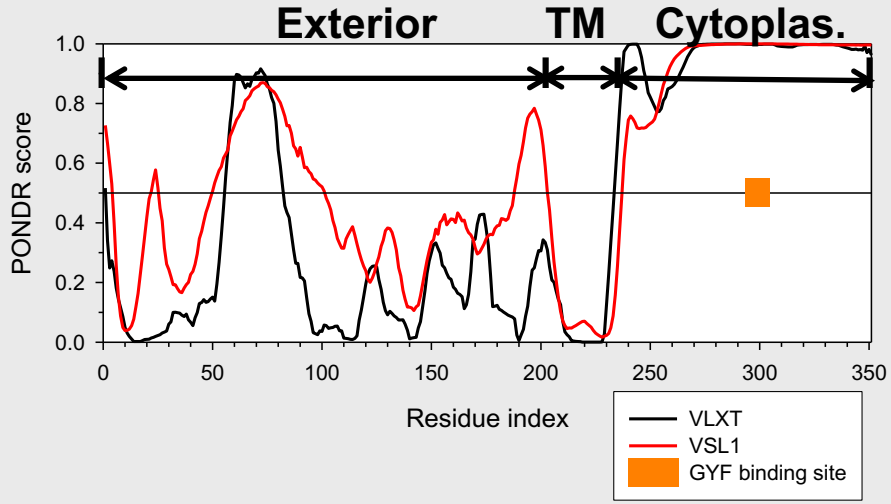
Protein Signaling Domain Example: GYF Domain Bound to CD2 IDP Region

Tony Pawson









See Also "Simple Modular Architecture Research Tool" (SMART)

IDPs & Function: p53

p53: main isoform ~ 400 AA residues

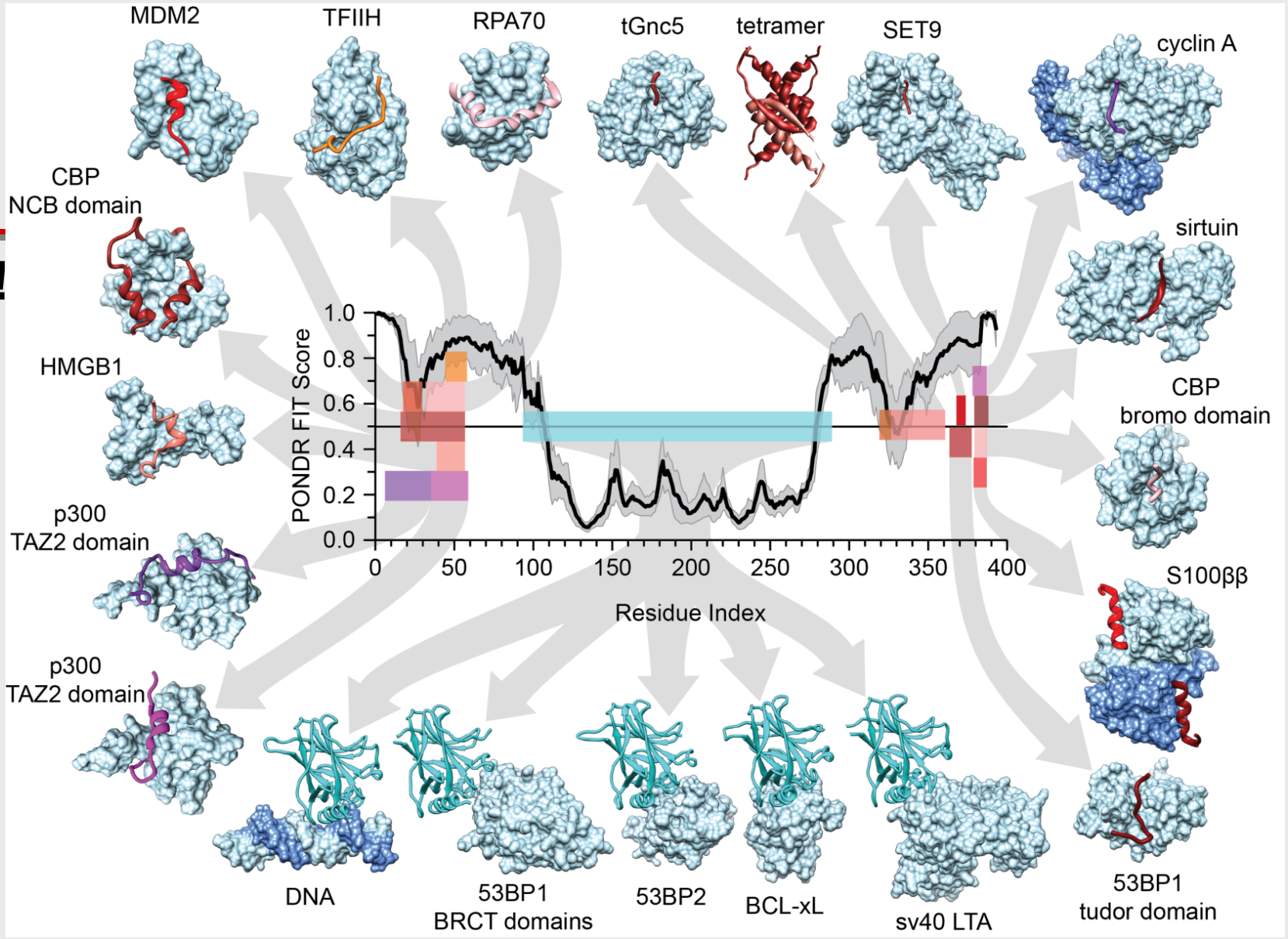
- **About 50% of this protein's residues are in two IDP regions, which are located at the two termini;**
- **This protein is a tumor suppressor, it initiates apoptosis, it arrests cell growth, it increases genome stability, it inhibits angiogenesis, and it activates the expression of hundreds of genes;**
- **This protein binds to DNA and to over 100 different protein partners; these many interactions enable the long list of functions given above.**

p53 binding

Note IDP tails!

Molecular Recognition Features (MoRFs)

Chris Oldfield



Modified from: Oldfield & Dunker, *Ann Rev Biochem* 83: 553 – 584 (2014)

IDPs & Function: BRCA1

BRCA1: main isoform ~ 1,860 AA residues

- About 83% of this protein is in one long, central IDP region of more than 1,500 residues;
- This protein is involved in DNA repair, in cell-cycle check point control, in transcription regulation, in apoptosis, in mRNA splicing, and in the activation of the expression of many genes;
- This protein binds to DNA and > 400 different protein partners; again these many interactions enable the long list of functions given above.

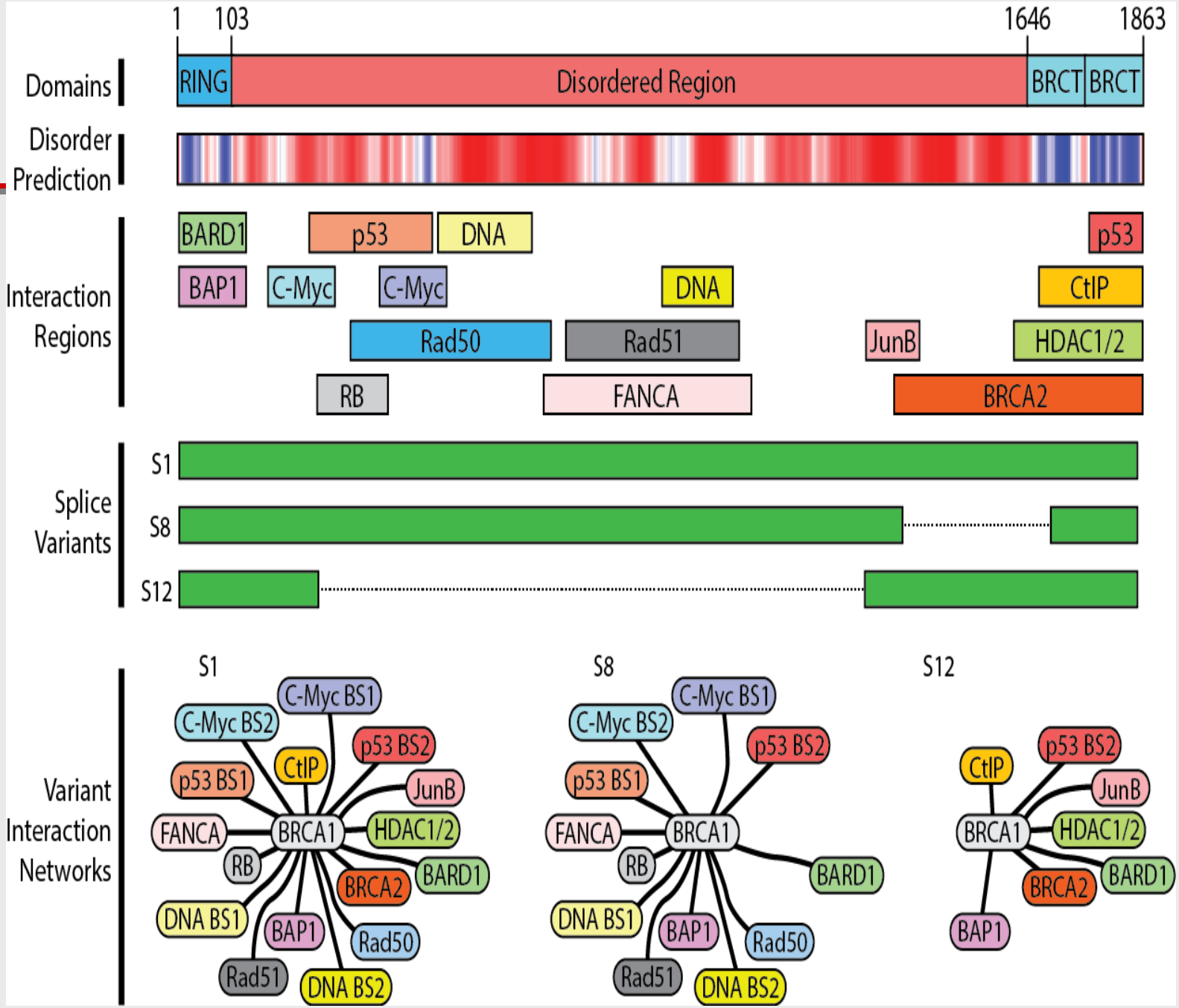
BRCA1

1863 residues;

103 ordered at the N-term;

217 ordered at the C-term;

1543 form one long IDP region in between.



Dunker AK et al. Semin Cell Devel Biol 37: 44-55 (2015)

IDPs & Function: p21/p27/p57

p21 / p27 / p57:

- **Each of these molecules is 100% IDP by both prediction and experiment;**
- **Each of these proteins is an inhibitor of the cyclin dependent kinase (CDK)-cyclin complex;**
- **Each of these proteins is involved in cell-cycle check point control;**
- **Removal of each of these proteins from the CDK-cyclin complex involves a multistep process that may act as a signal coordinator.**

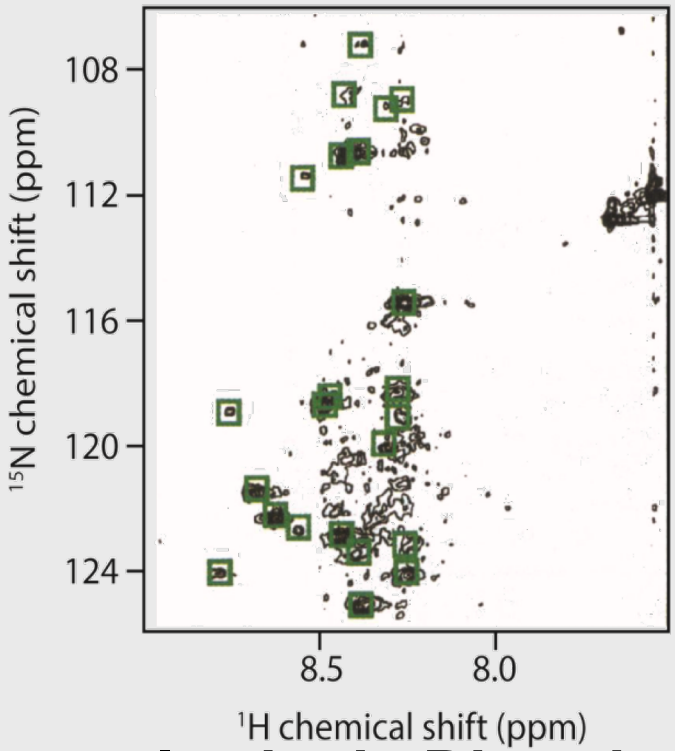
p21$^{Waf1/Cip1/Sdi1}$

p27^{Kip1}

p57^{Kip2}

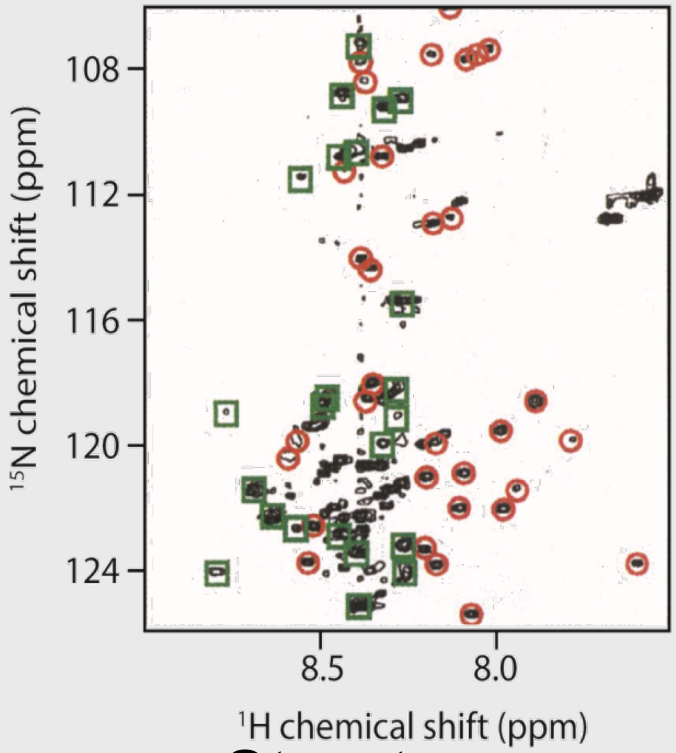
Reviewed in:

Dunker AK & Oldfield CJ IDPs Studied by NMR, Adv Expt Med & Biol; Felli & Pierattelli (eds), Springer International Publishing, Switzerland pp. 1-34 (2015)

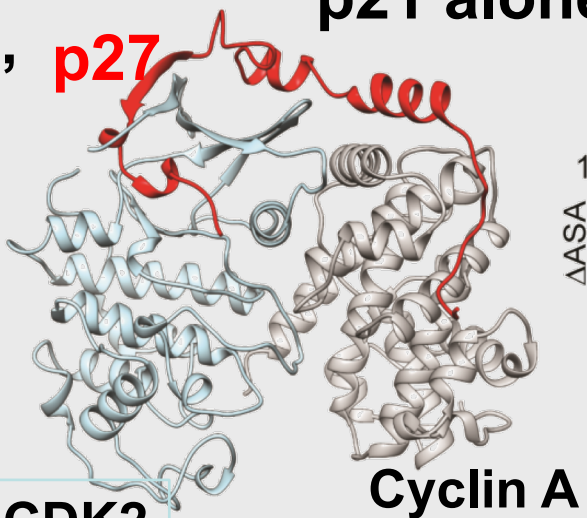


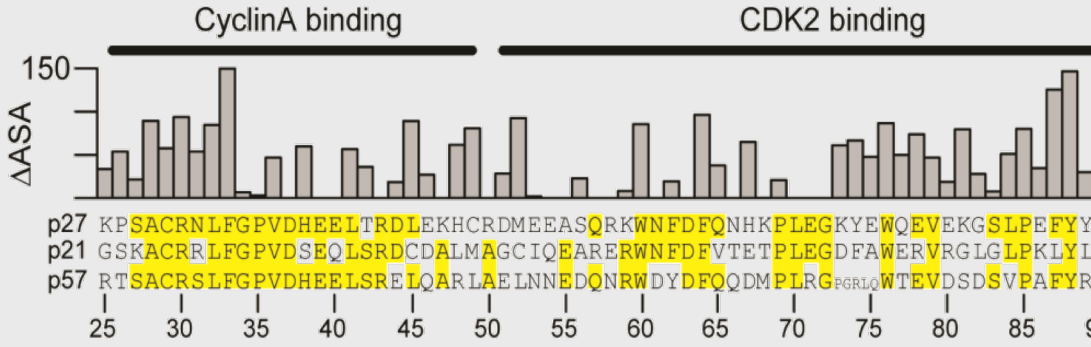
Intrinsic Disorder
p21 alone

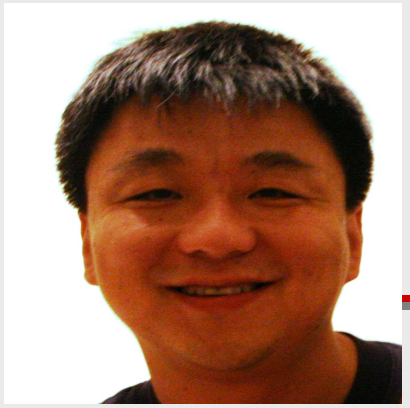
→



Structure
p21 + CDK







IDPs & Function Global Analysis

Hongbao Xie

Zoran Obradovic

- Collect SwissProt function-specific sequences;
- Collect matching random-function sequences; Repeat 1,000 times;
- Predict disorder for each function-specific & 1,000 random-function sets → all RFS ~ fit one Gaussian;
- Rank structure- and disorder-associated functions by Z-scores (Z-score = [x – <x>]/σ);

– values = more structure, + values = more disorder

IDPs & Function

Functional Key Word Categories	Number
High-prediction of disorder (> +1)	238
Intermediate (Z-score, –1 to +1)	170
Low-prediction of disorder (< –1)	302
TOTAL	710

Xie H et al. *J. Proteome Res.*. 6: 1882- 1898;
6:1899-1916; & 6:1917-1932 (2007)

Top 10 Biological Processes Most Strongly Associated with High-Prediction of Disorder

KEYWORDS	Proteins (number)	Families (number)	Length (Ave)	Z – Score
Differentiation	1406	422	439	18.8
Transcription	11223	1653	442	14.6
Transcription Regulation	9758	1554	413	14.3
Spermatogenesis	332	189	280	13.9
DNA Condensation	317	130	300	13.3
Cell Cycle	4278	612	494	12.2
mRNA Processing	1575	249	516	10.9
mRNA Splicing	716	180	459	10.1
Mitosis	718	215	620	9.4
Apoptosis	810	211	465	9.4

Xie H, et al., *J. Proteome Res* 6: 1882-1932 (2007)

Top 10 Biological Processes Most Strongly Associated with Low-prediction of Disorder (e.g. with Structure)

KEYWORDS	Proteins (number)	Families (number)	Length (Ave)	Z – Score
GMP Biosynthesis	225	3	473	–17.6
Amino-acid Biosynthesis	7098	212	361	–17.1
Transport	19888	2199	378	–14.9
Electron Transport	4633	346	272	–13.7
Lipid A Biosynthesis	533	13	291	–13.2
Aromatic Catabolism	320	105	300	–12.4
Glycolysis	2255	50	390	–12.1
Purine Biosynthesis	1208	28	445	–11.9
Pyrimidine Biosynthesis	1310	27	383	–11.7
Carbohydrate Metabolism	1797	180	404	–11.7

Xie H, et al., *J. Proteome Res* 6: 1882-1932 (2007)

What are the functions of IDPs?
IDPs Used for Signaling and Regulation!

- **Sequence → Structure → Function (Z < – 1)**
  - – Catalysis,
  - – Membrane transport,
  - – Binding to DNA, RNA, molecules or IDP regions.
- **Sequence → IDP Ensemble → Function (Z > + 1)**
  - – Signaling, Dunker AK, et al., *Biochemistry* 41: 6573-6582 (2002)
  - – Regulation, Dunker AK, et al., *Adv. Prot. Chem*. 62: 25-49 (2002)
  - – Recognition, Xie H, et al., *Proteome Res.* 6: 1882-1898 (2007)
  - – Control. Vucetic, S. et al., *Proteome Res* 6: 1899-1916 (2007)
    Xie H, et al., *Proteome Res* 6: 1917-1932 (2007)

Summary

Sequence → Structure → Function

- **Structured proteins are for catalysis, transport, and binding to molecules, to macromolecules, and to IDP regions;**

Sequence → IDP Ensembles → Function

- **IDPs are for signaling, regulation, recognition, and control.**

Intrinsically Disordered Proteins

THANK YOU!!!

(kedunker@iupui.edu)

Funding: NIH, NSF, INGEN
IUPUI Signature Centers Initiative