Intrinsically disordered proteins and their interactions

Zsuzsanna Dosztányi

MTA–ELTE Momentum Bioinformatics Group
Department of Biochemistry
Eötvös Loránd University, Budapest

4 December 2019, Bangalore


Central and Eastern Europe
Stockholm
SWEDEN
Goteborg
Gotland
Tallinn
ESTONIA
RUSSIA
Moscow
Riga
LATVIA
Baltic Sea
DENMARK
North Sea
Bornholm
Kaliningrad
RUSSIA
LITHUANIA
Vilnius
Minsk
Smolensk
Hamburg
Gdansk
Amsterdam
Bremen
Berlin
Hrodna
BELARUS
NETHERLANDS
Warsaw
Brussels
Cologne
Leipzig
POLAND
Brest
BELGIUM
GERMANY
Wroclaw
Rivne
Kiev
Luxembourg
Prague
CZECH REPUBLIC
Krakow
UKRAINE
Strasbourg
Stuttgart
Brno
SLOVAKIA
FRANCE
Munich
Bratislava
Vienna
Budapest
Chisinau
Bern
Geneva
Vaduz
AUSTRIA
Cluj-Napoca
MOLDOVA
Odesa
HUNGARY
Lyon
Milan
Ljubljana
ROMANIA
Zagreb
Genoa
CROATIA
Bucharest
Constanta
Marseille
Belgrade
SERBIA
Varna
Black Sea
Corsica
ITALY
Sarajevo
Pristina
BULGARIA
Sofia
Rome
Podgorica
Skopje
Istanbul
Sardinia
Naples
Tirana
F.Y.R.O.M.
ALBANIA
Tyrrhenian Sea
Cagliari
GREECE
Aegean Sea
TURKEY
Mediterranean Sea
Palermo
Ionian Sea
Sicily
Tunis
Rhodes
TUNISIA
Valletta
MALTA
300 Kilometers
300 Miles

IDPs/IDRs

- Intrinsically disordered proteins/regions (IDPs/IDRs)
- Do not adopt a well-defined structure in isolation under native-like conditions
- Ensemble of very different conformations
- Functional proteins
- Often involved in PPIs

Article No. jmbi.1999.3110 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (1999) **293**, 321–331

JMB

Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm

Peter E. Wright* and H. Jane Dyson*

Experimental detection of disorder

From the literature

- Failed attempts to crystallize
- Lack of NMR signals
- Heat stability
- Protease sensitivity
- Increased molecular volume

NMR

- HSQC
- chemical shifts (CS)
- residual dipolar couplings (RDCs)
- paramagnetic relaxation enhancement (PRE)

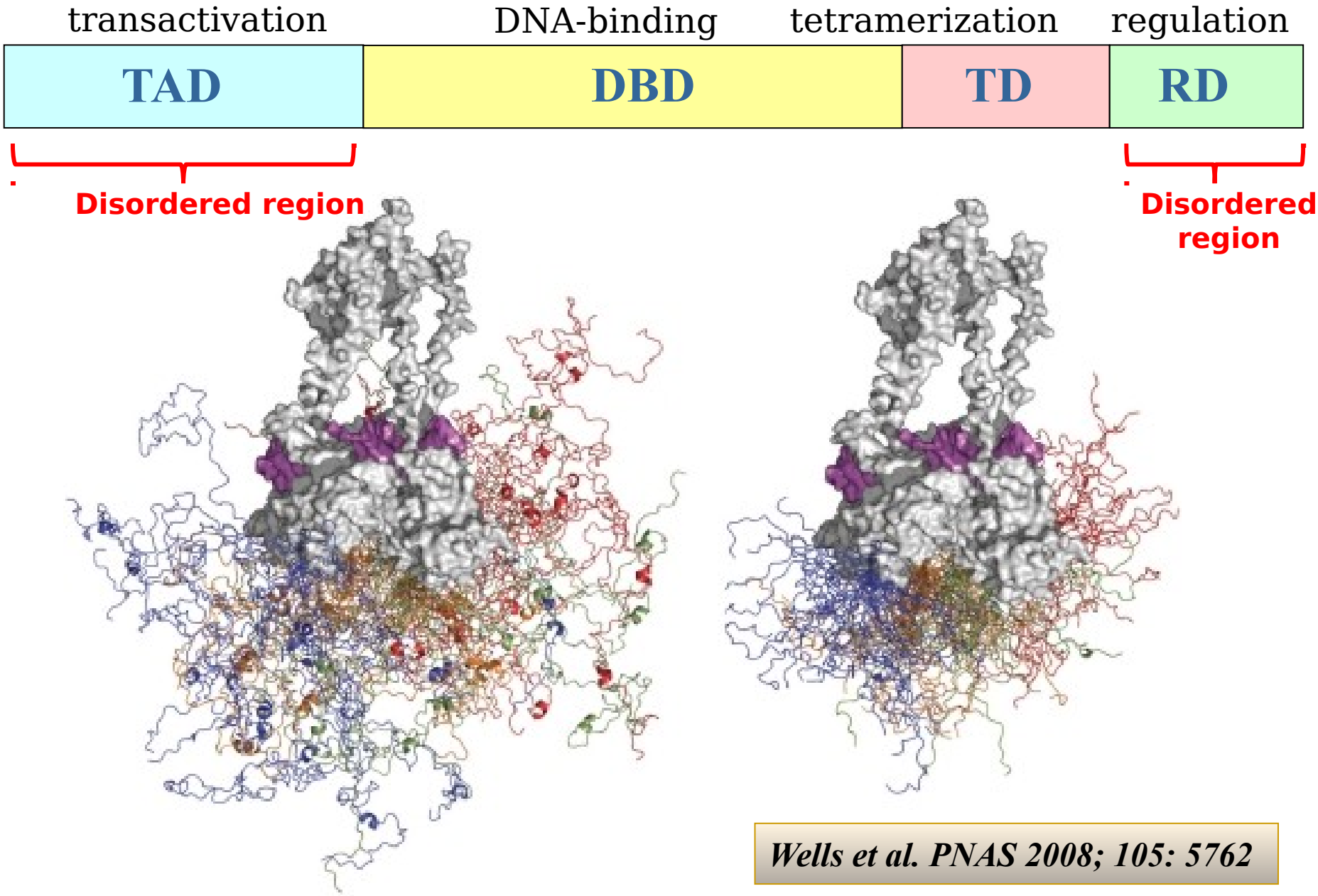
Disordered proteins

In the PDB

Missing electron density regions from the PDB

NMR structures with large structural variations

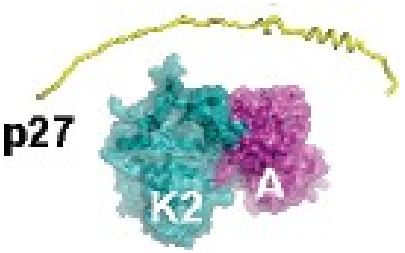
p53 tumor suppressor



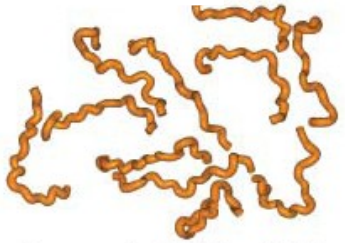
Wells et al. PNAS 2008; 105: 5762

Heterogeneity in protein disorder

Transient structures

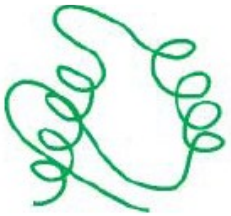


RC-like



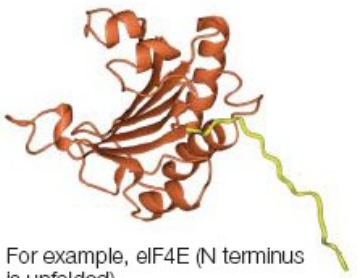
For example, ACTR (no NCBD)

Compact



For example, NCBD (no ACTR)

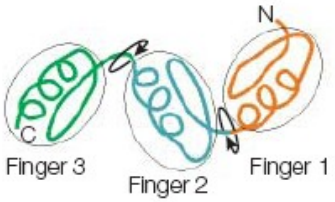
Mostly folded, local disorder



For example, eIF4E (N terminus is unfolded)

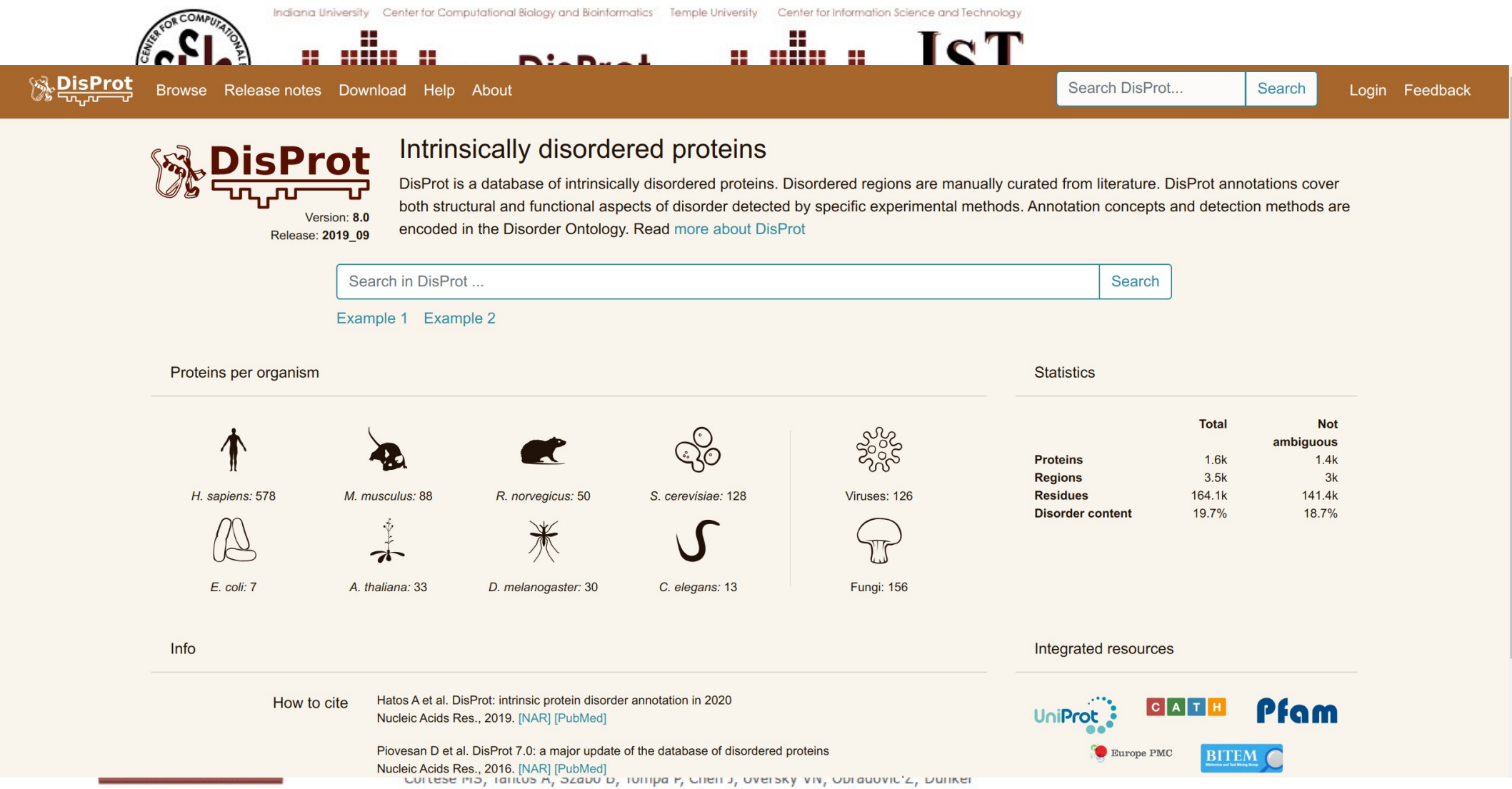
Flexible loop

Linked folded domains
(beads on a string)



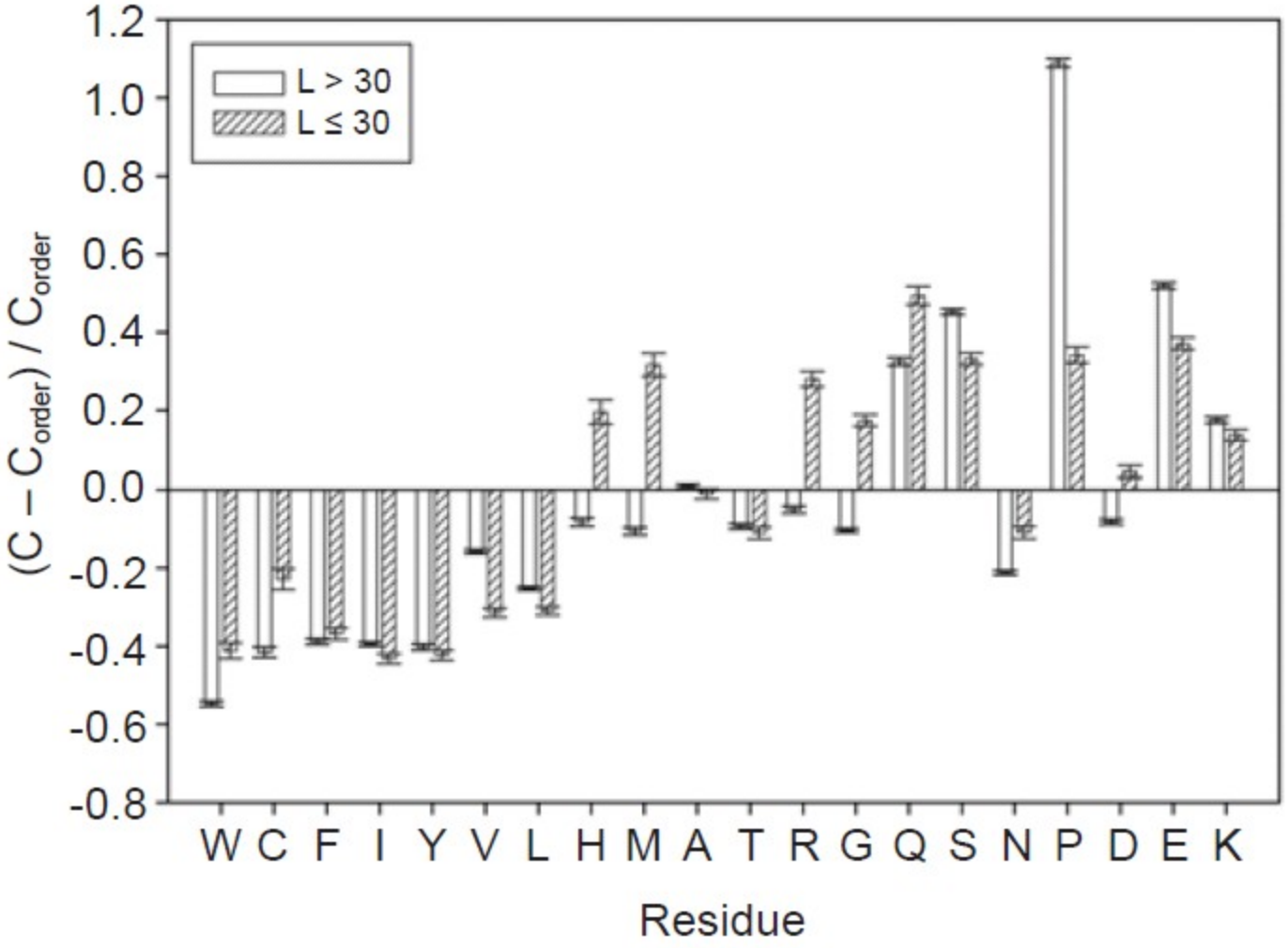
For example, zinc fingers (no DNA)

DisProt



Disprot – old version

Amino Acid Compositions



Protein disorder is encoded in the amino acid sequence

Prediction methods for protein disorder

Over 50 methods …

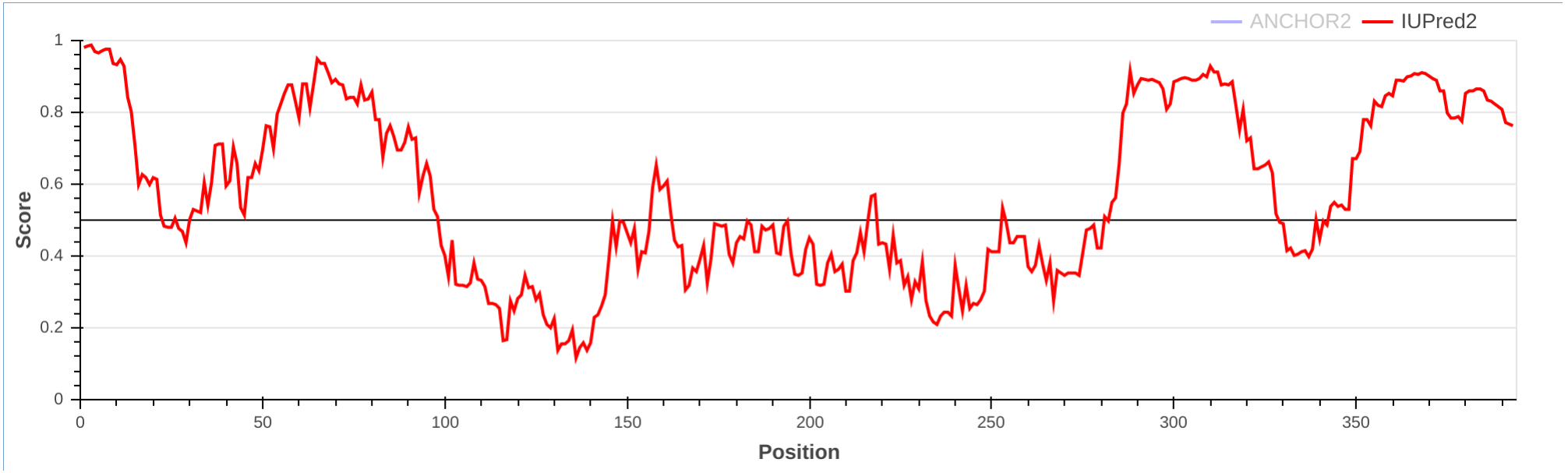
- Based on amino acid propensity scales
  - GlobPlot (secondary structure propensities)
- Simplified biophysical models
  - **IUPred**
- Machine learning approaches
  - Disopred, Espritz
- Meta servers
- Deep learning

IUPred

If a residue cannot form enough favorable interactions within its sequential environment, it will not adopt a well defined structure → it will be disordered

- Based on an energy estimation method
- Parameters calculated from statistics of globular proteins
- No training on disordered proteins

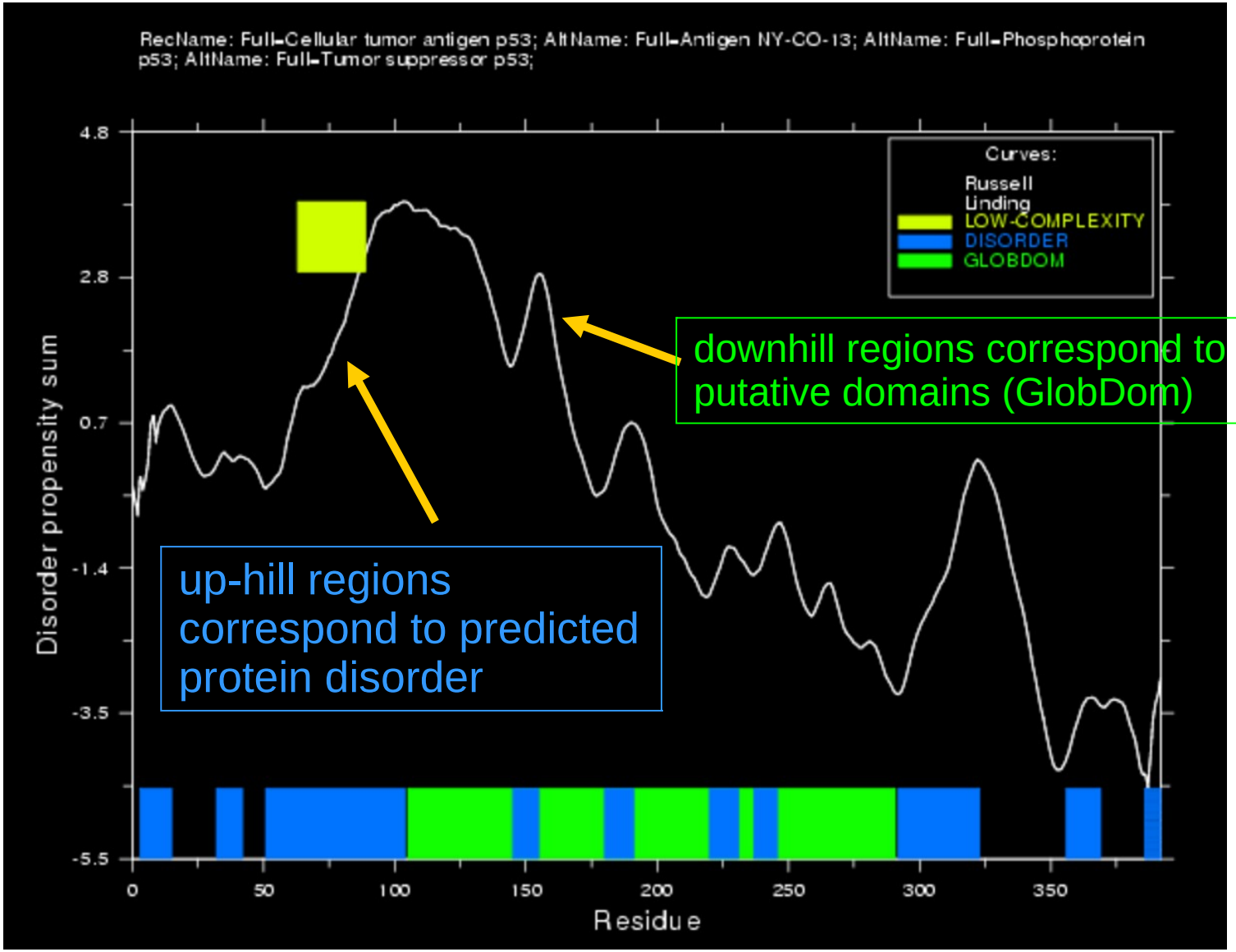
A typical output (IUPred2A)



Predictions are on a per residue basis

Dosztányi et. a. (2005) JMB 347, 827
Erdős et al. (2018) ***NAR 46(W1):W329***

GlobPlot: http://globplot.embl.de/

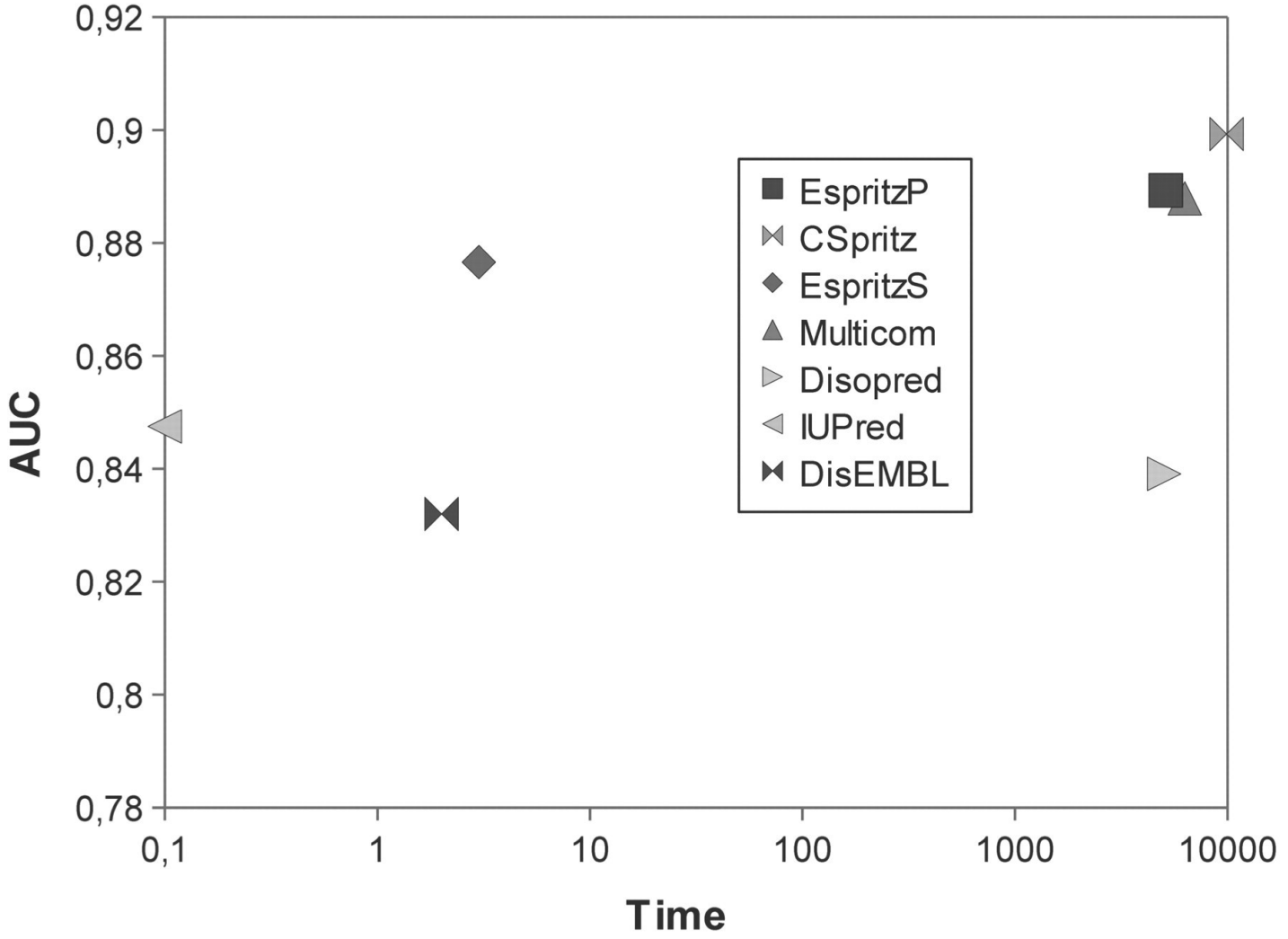


Where are the ordered domains, disordered regions?

Prediction of protein disorder

- Disordered is encoded in the amino acid sequence
- Can be predicted from the sequence
- ~80% accuracy (>0.8 AUC)
  - *Necci et al. 2018. Bioinformatics 34, 445*
  - *Neilsen&Mulder 2019. Sci Rep, 9, 5137*
  - *CAID*
- Challenges
  - Small, noisy datasets
  - Disorder is heterogeneous
  - “Flavors” of disorder

Which is the best method? Speed

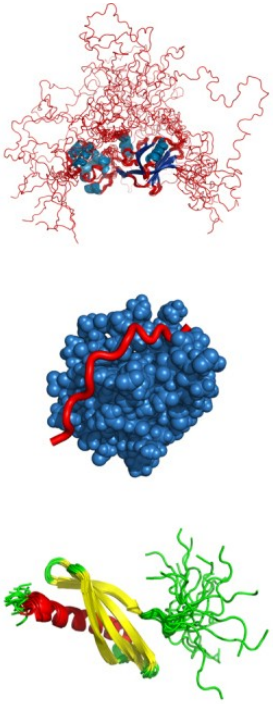


Time versus performance plot for different predictors.

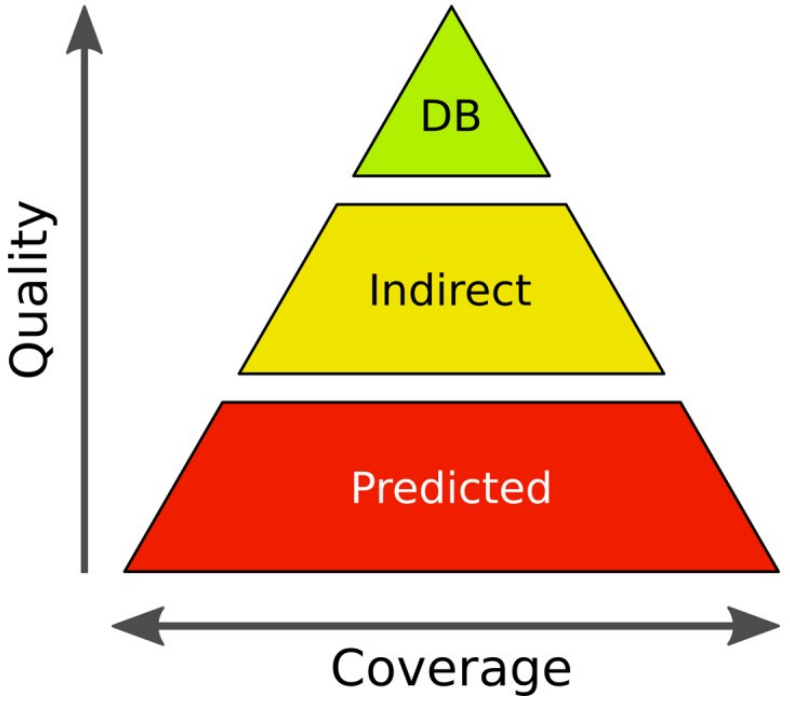
Ian Walsh et al. Bioinformatics 2012;28:503

MobiDB

A

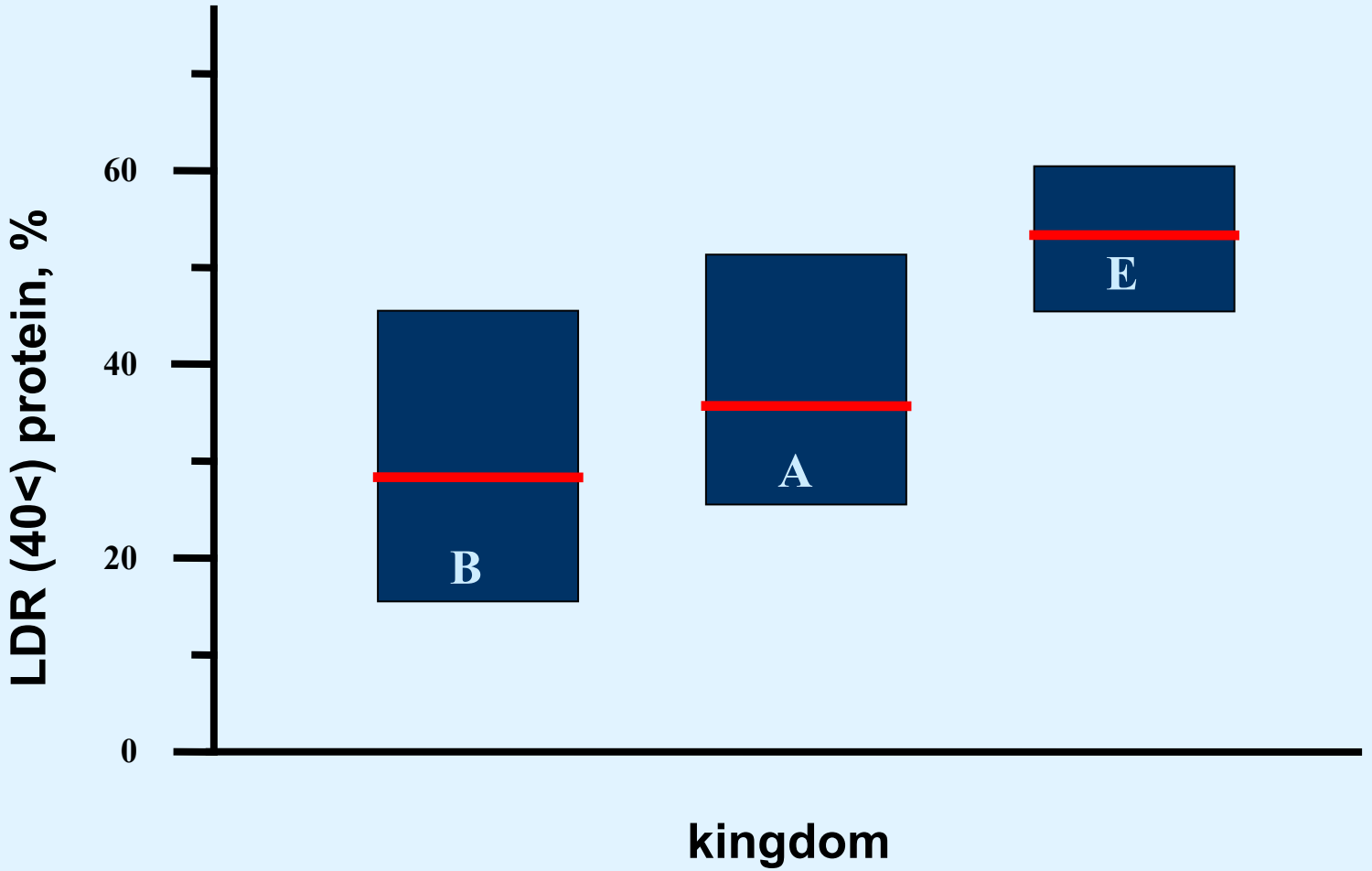


B



Piovesan D. et al. NAR. 2018;46(D1):D471–D476.

Protein disorder is prevalent



Protein disorder complements the functional repertoire of globular proteins

Table 2. Correlation and anticorrelation of structural disorder with Swiss-Prot functional categories

Top functions that correlate with long disorder[a]	Top functions that anticorrelate with long disorder
Differentiation	GMP biosynthesis
Transcription	Amino acid biosynthesis
Transcription regulation	Transport
Spermatogenesis	Electron transport
DNA condensation	Lipid A biosynthesis
Cell cycle	Aromatic hydrocarbons catabolism
mRNA processing	Glycolysis
mRNA splicing	Purine biosynthesis
Mitosis	Pyrimidine biosynthesis
Apoptosis	Carbohydrate metabolism
Protein transport	Branched-chain amino acid biosynthesis
Meiosis	Lipopolysaccharide biosynthesis

Xie H et al. J Proteome Res. 2007, 6, 1882

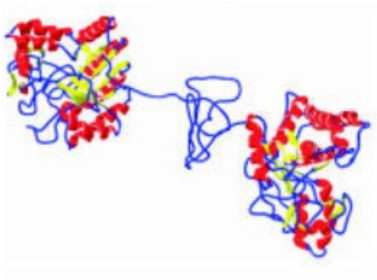
Protein disorder is important

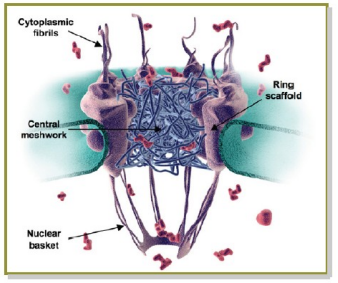
- Prion protein → Prion disease
- CFTR → Cystic fibrosis
- τ → Alzheimer's
- α-synuclein → Parkinson's
- p53, BRCA1 → cancer

How IDPs carry out their functions?

- Entropic chains

  Function directly results from disordered state





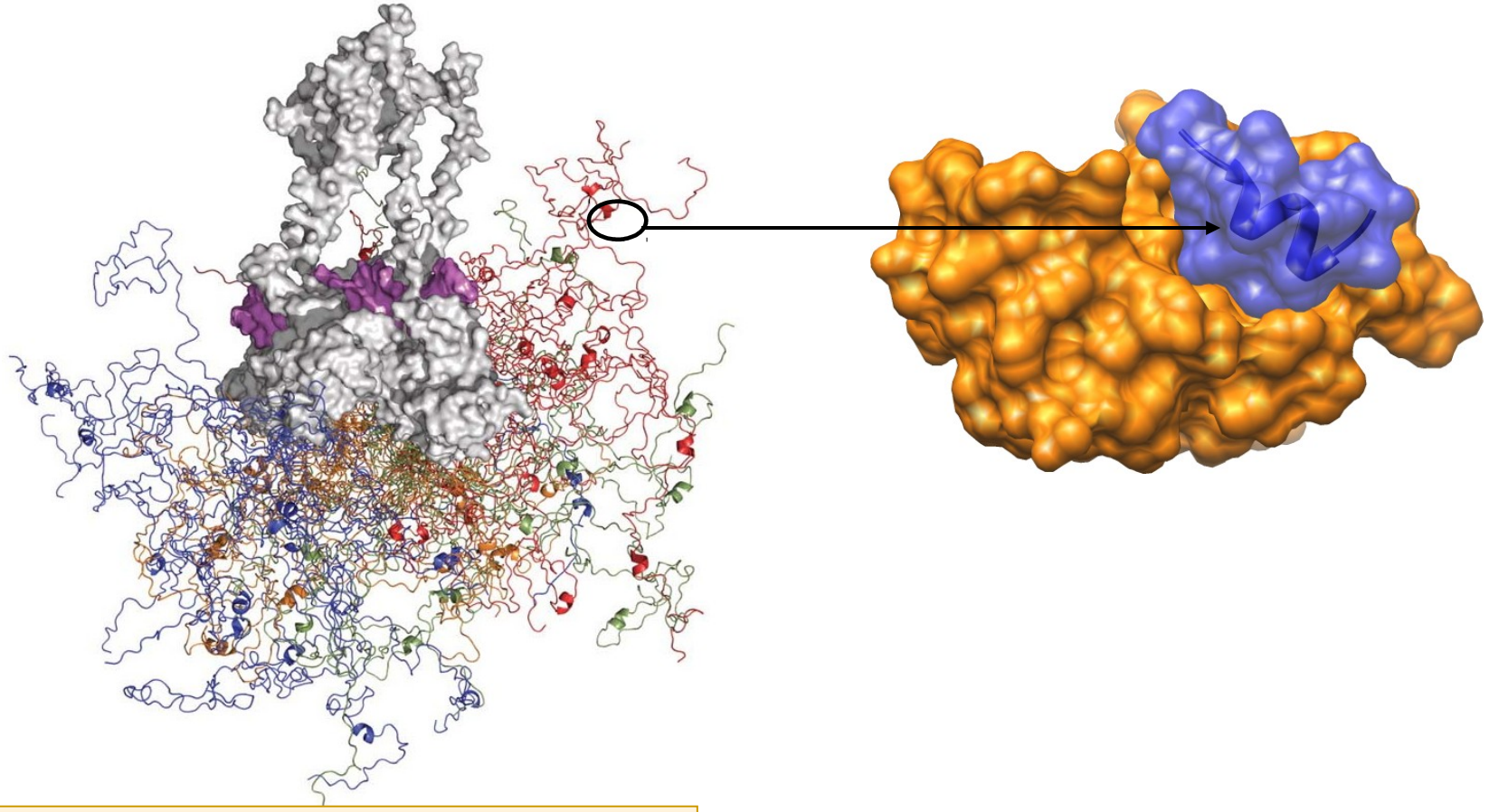
- Molecular recognition

  Coupled folding and binding

- “Assemblages”

  Functional sites formed by phase separation

Interaction of IDPs



Complex between p53 és MDM2

Coupled folding and binding

Binding regions within IDPs

- **SLIMs: Short linear motifs**

  3-11 residues long, average size 6-7 residues

  enriched in IDRs

- **Disordered binding sites, Morfs, LIPs**

  undergo disorder to order transition upon binding

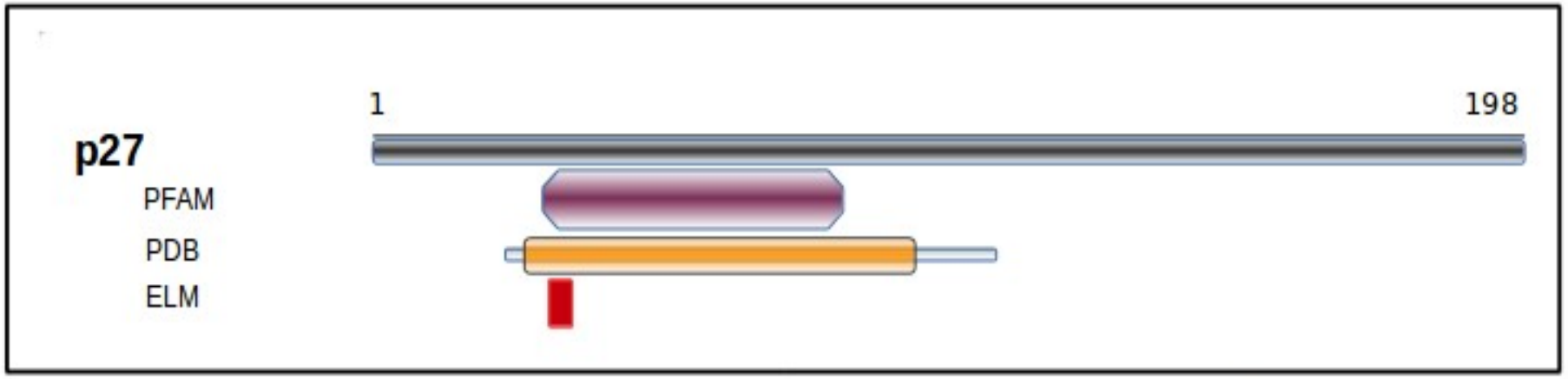
  usually less then 30 residues, can be up to 70

- **Intrinsically disordered domains**

  evolutionary conserved disordered segments

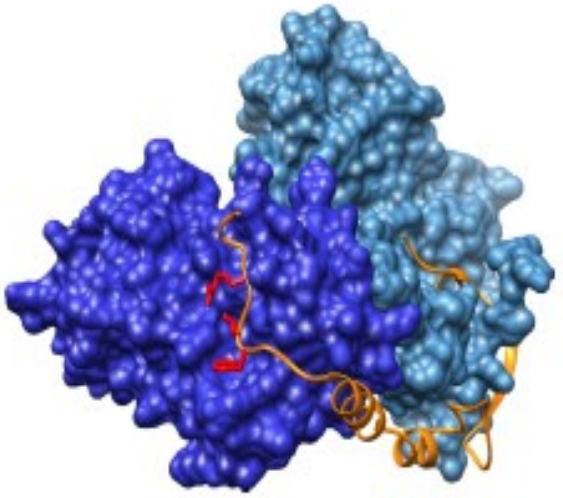
p27

Inhibitor of CDK2-CyclinA complex.



[RK].L.{0,1}[FYLIVMP]

CDN1B_HUMAN	30-33	HPKPSACRNLFGPVDHEEL
MPIP1_HUMAN	11-15	PEPPHRRRLLFACSPPPAS
CDC6_HUMAN	94-98	HSHTLKGRRLVFDNQLTIK
RB_HUMAN	873-877	SNPPKPLKKLRFDIEGSDE
P53_HUMAN	381-385	GQSTSRHKKLMFKTEGPDS
VE1_HPV18	127-130	SGQKKAKRRLFTISDSGYG

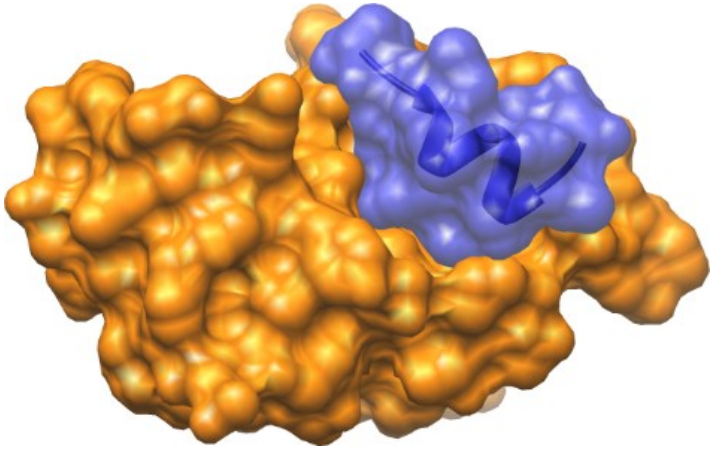


Database of Disordered Binding Sites

DIBS

- Experimentally verified to be disordered
  - Manual, Disprot, ELM, IDEAL, DisBind
- Forms a complex with ordered partners
- **Kd values**

773 proteins
(1577 structures)



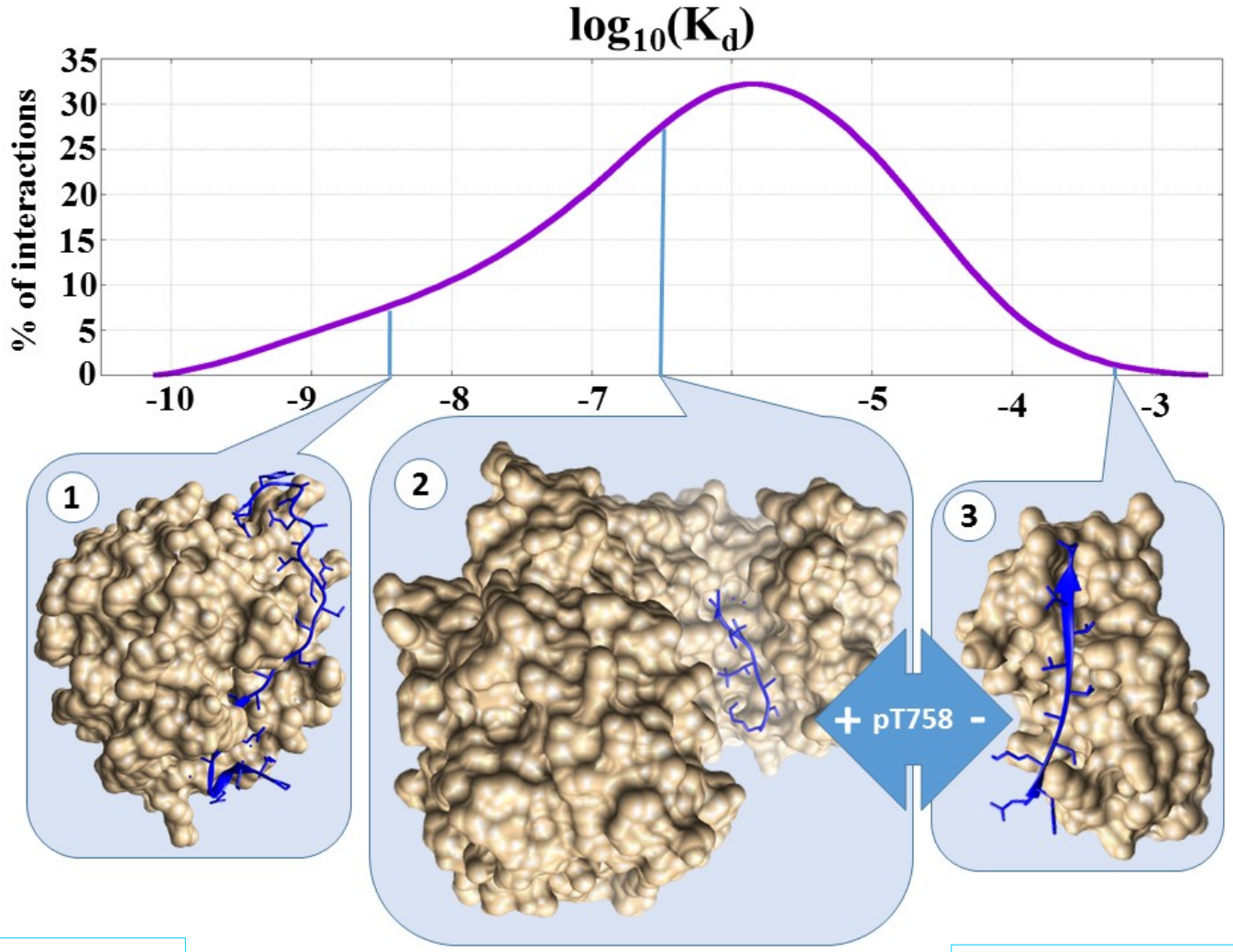
Databases and ontologies

DIBS: a repository of disordered binding sites mediating interactions with ordered proteins

Eva Schad[1], Erzsébet Fichó[1], Rita Pancsa[2], István Simon[1], Zsuzsanna Dosztányi[3,*] and Bálint Mészáros[1,3,*]

Bioinformatics (2017)
doi: 10.1093/bioinformatics/btx640

Kd values



Thrombin-anophelin

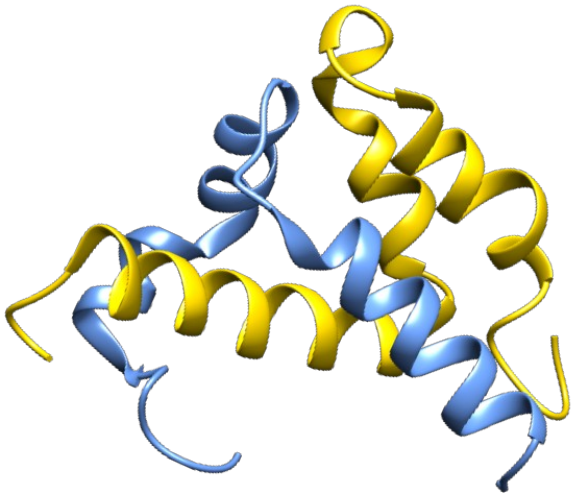
14-3-3ζ-integrin β2

filamin-integrin β2

Mutual Folding Induced by Binding

MFIB

- Both partners are experimentally verified to be disordered
  - Manual, Disprot



205 proteins
(1406 structures)

Databases and ontologies

MFIB: a repository of protein complexes with mutual folding induced by binding

Erzsébet Fichó[1], István Reményi[2], István Simon[1,*] and Bálint Mészáros[1,*]

Bioinformatics (2017)
doi: 10.1093/bioinformatics/btx486

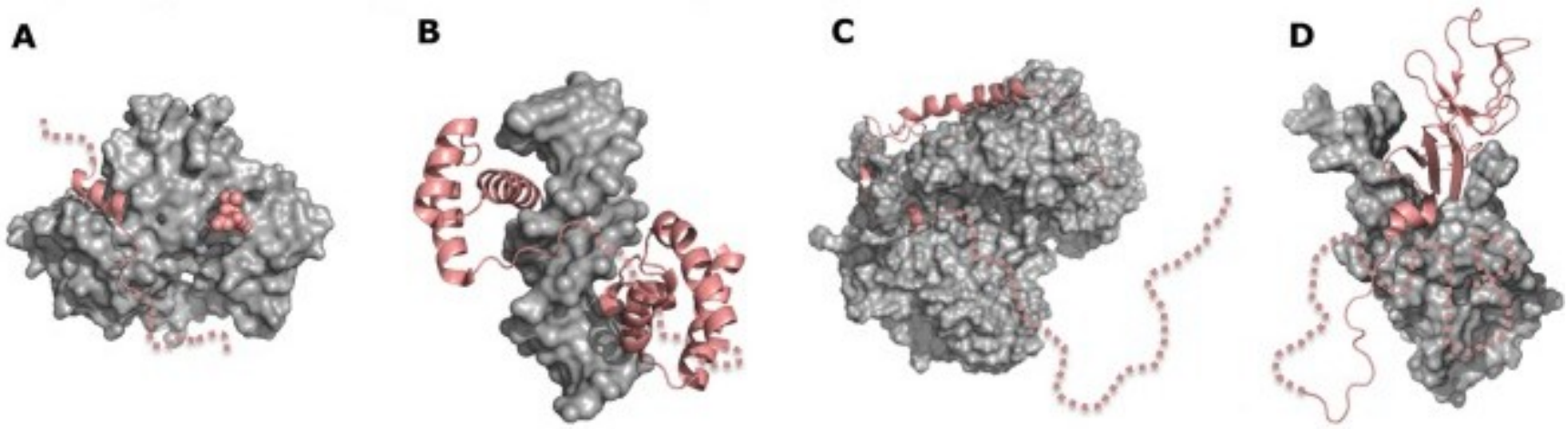
FuzDB

FUZZY COMPLEXES DATABASE

Home | Browse | Search | Methods | Classes | Upload | Analysis | Help | FAQs

Database of fuzzy protein complexes
FuzDB v3.3

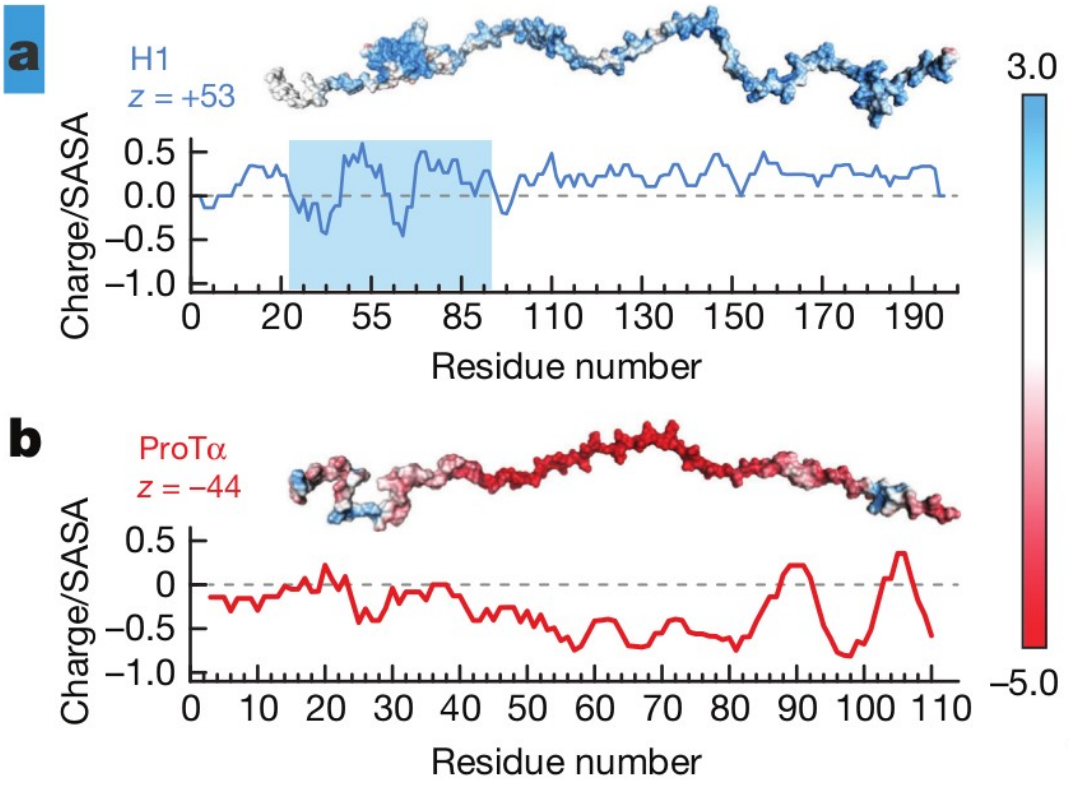
Citation: Nucleic Acids Research (2017) Jan 4;45(D1):D228-D235. abstract full [PDF]

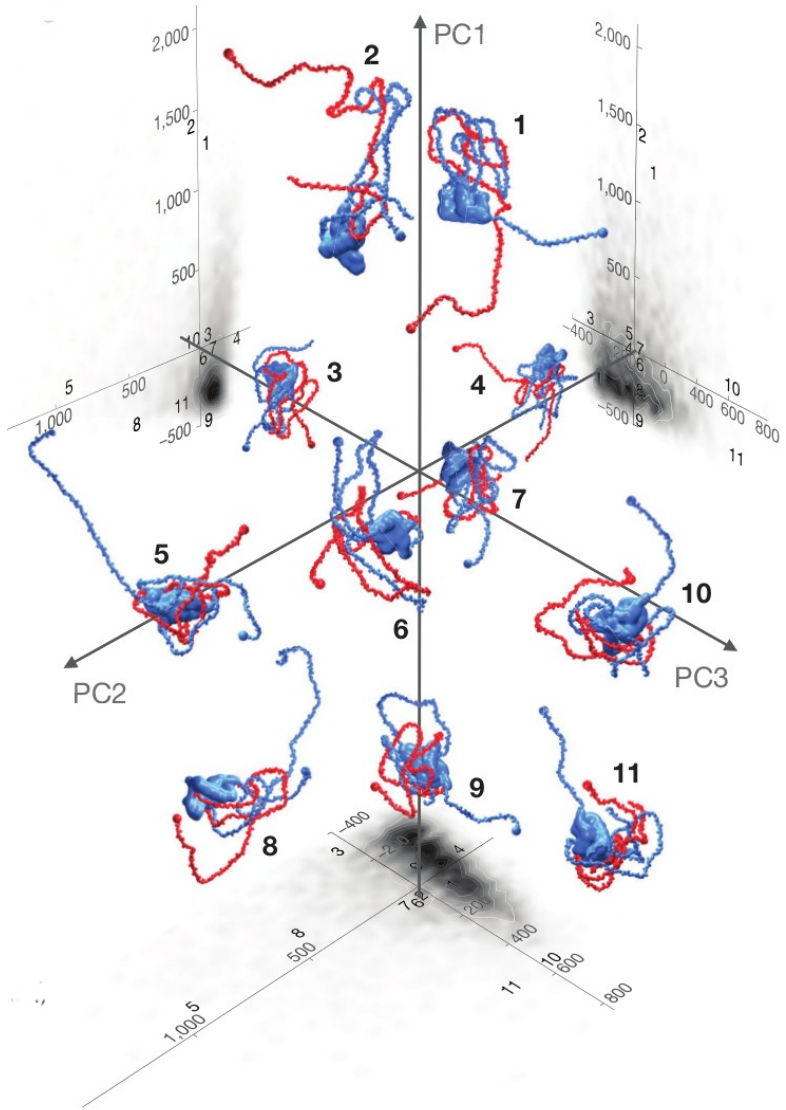


polymorphic clamp flanking random

Fuzzy complexes

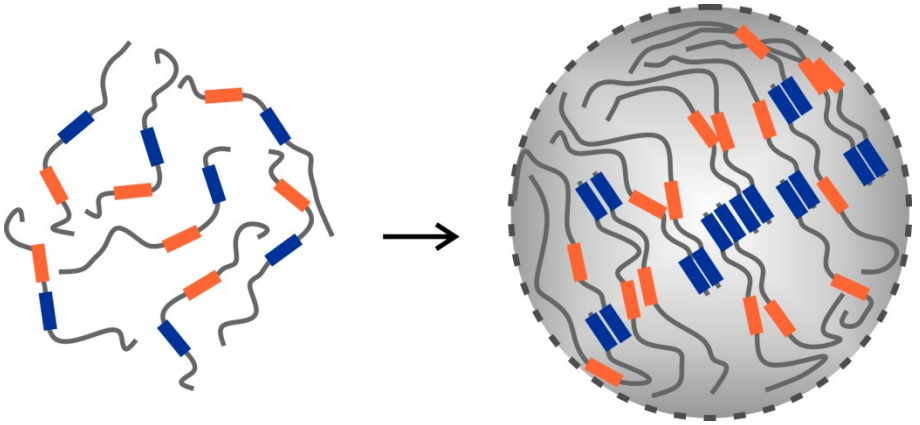
Extreme disorder in complex





Borgia et al. Nature 2018, 555, 61.

Liquid-liquid phase separation



formation of liquid/gel-like condensates of protein, RNA, and other biomolecules

driven by multivalent weak interactions

delimited by a phase boundary (no membrane)

Nucleolus, stress granules, P-bodies, Germ granules, heterochromatin, post-synaptic densities,

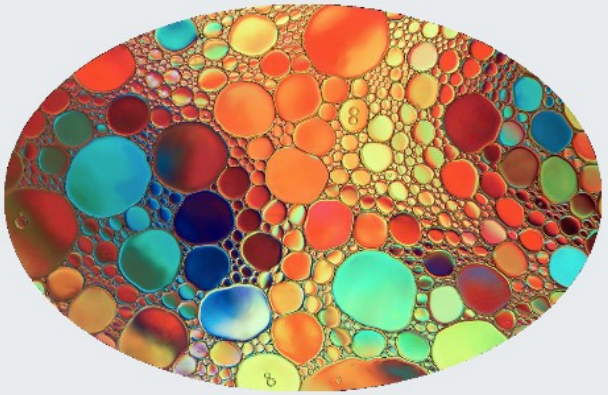
PhaSePro: https://phasepro.elte.hu/

Home | Browse/Search | About/Help | Candidates | Statistics | Download | Annotate

Welcome to PhaSePro!

PhaSePro is the comprehensive database of proteins driving liquid-liquid phase separation (LLPS) in living cells. LLPS is a molecular process employed by all living organisms to form membraneless organelles, mediating crucial cellular functions. PhaSePro is manually curated, it is solely based on experimentally verified cases of LLPS, integrating a wide range of information on the biophysical driving forces, biological function and regulation of these molecular systems.

Learn more »



Search for gene names, full or partial common/UniProt protein names, or UniProt accessions. Example 1 Example 2

Getting started

To get an introduction to the structure and use of PhaSePro, you can visit the selected examples (FUS and TDP-43) by clicking the buttons above, or read the About/Help pages by clicking below.

View details »

Explore

You can start searching the database by entering keywords in the above field, or by browsing the available entries by clicking below.

Browse/Search entries »

Annotate

Help us expand the knowledge about proteins involved in liquid-liquid phase separation by submitting new entries into PhaSePro.

View details »

Partners

TYPE OF PARTNER
Protein-protein binding
Protein-DNA binding
Protein-RNA binding
Protein-lipid binding
Protein-metal binding
Protein-inorganic salt binding

IDR – functional ontology (CV)

I) molecular function of disorder (MFUN)
- type of functional readout of function (such as MF in GO)

II) type of molecular transition (TRAN)
- necessary for function (such as disorder-to-order transition)

III) molecular partner (PART)
- type of partner recognized (protein, RNA, DNA, metabolite)

Predicting function for IDRs

Prediction of binding sites

- Interaction sites are usually linear (consist of only 1 part)
- Enrichment of interaction prone amino acids
- Can be predicted from sequence without predicting the structure
- ~ 10 Methods
  - Biophysical methods (*ANCHOR*)
  - Machine learning (*MorfPred, Morf$_{chibi}$, DISOPRED3)*
  - Evolutionary approaches (*SlimPrints, PhyloHMM*)

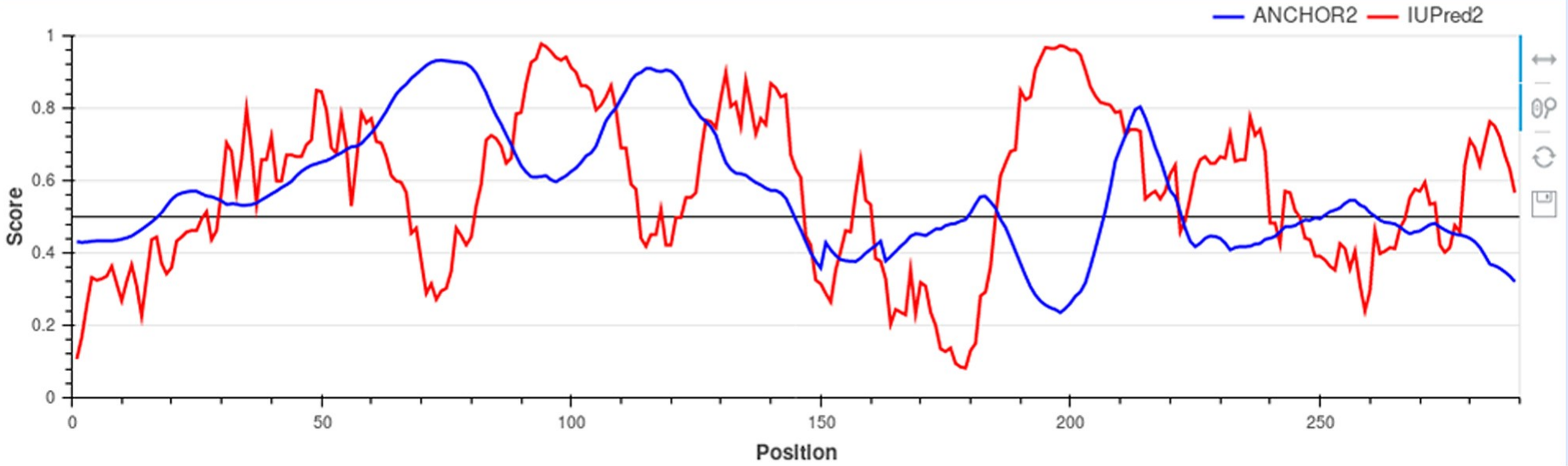
Prediction of disordered binding sites – ANCHOR

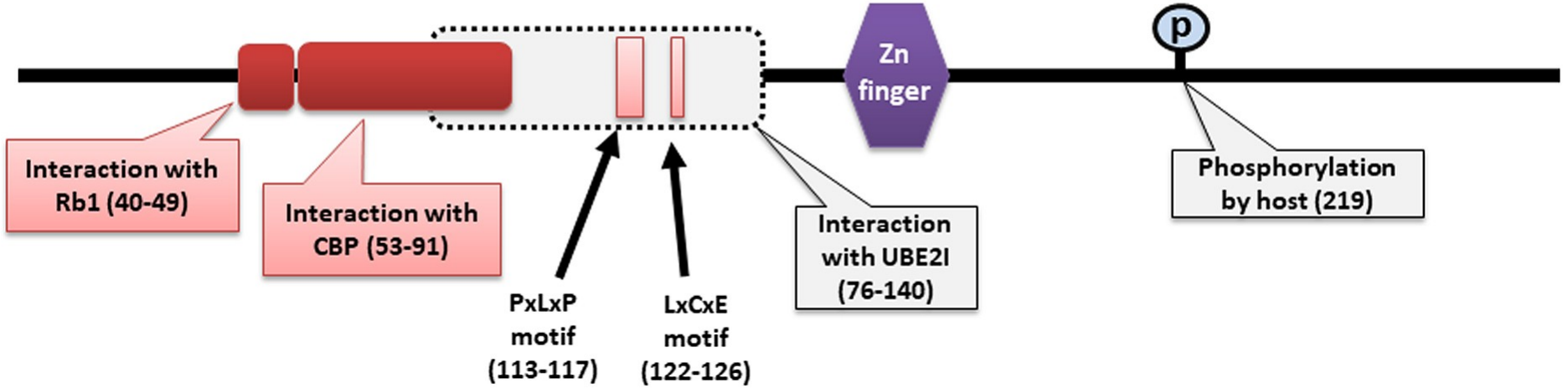
- What discriminates disordered binding regions?
  - A cannot form enough favorable interactions with their sequential environment
  - It is favorable for them to interact with a globular protein

- Based on simplified physical model
  - Based on an energy estimation method using statistical potentials

ANCHOR2

>sp|P03255|E1A_ADE05 Early E1A protein OS=Human adenovirus C serotype 5 PE=1 SV=1





Human adenovirus C early E1A protein