

Intrinsically disordered proteins and their interactions

Zsuzsanna Dosztányi

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

5-10 November 2018, Rome

IDPs

- 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

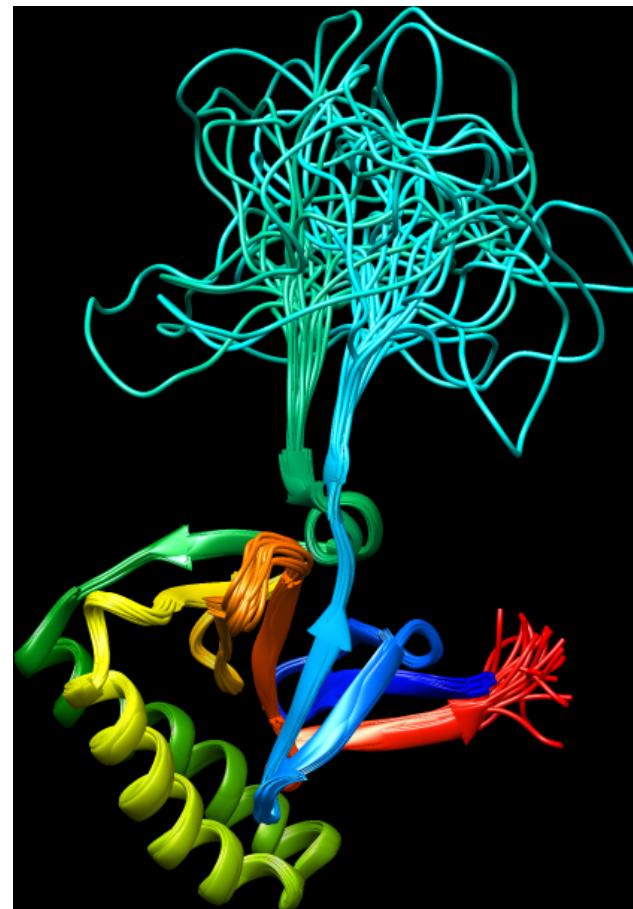
- HSQC
- chemical shifts (CS)
- residual dipolar couplings (RDCs)
- paramagnetic relaxation enhancement (PRE)

Disordered proteins

In the PDB



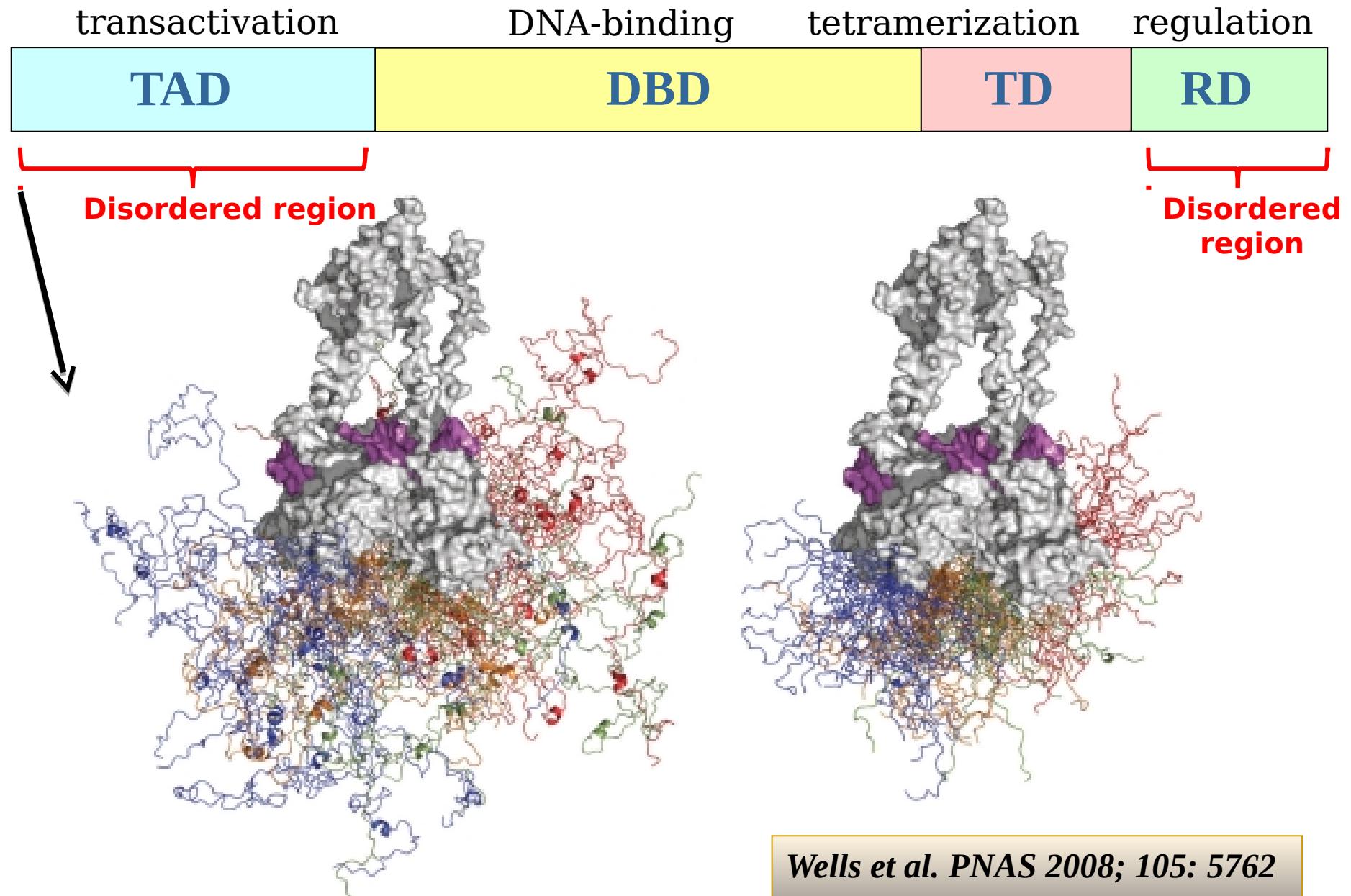
Missing electron density regions from
the PDB



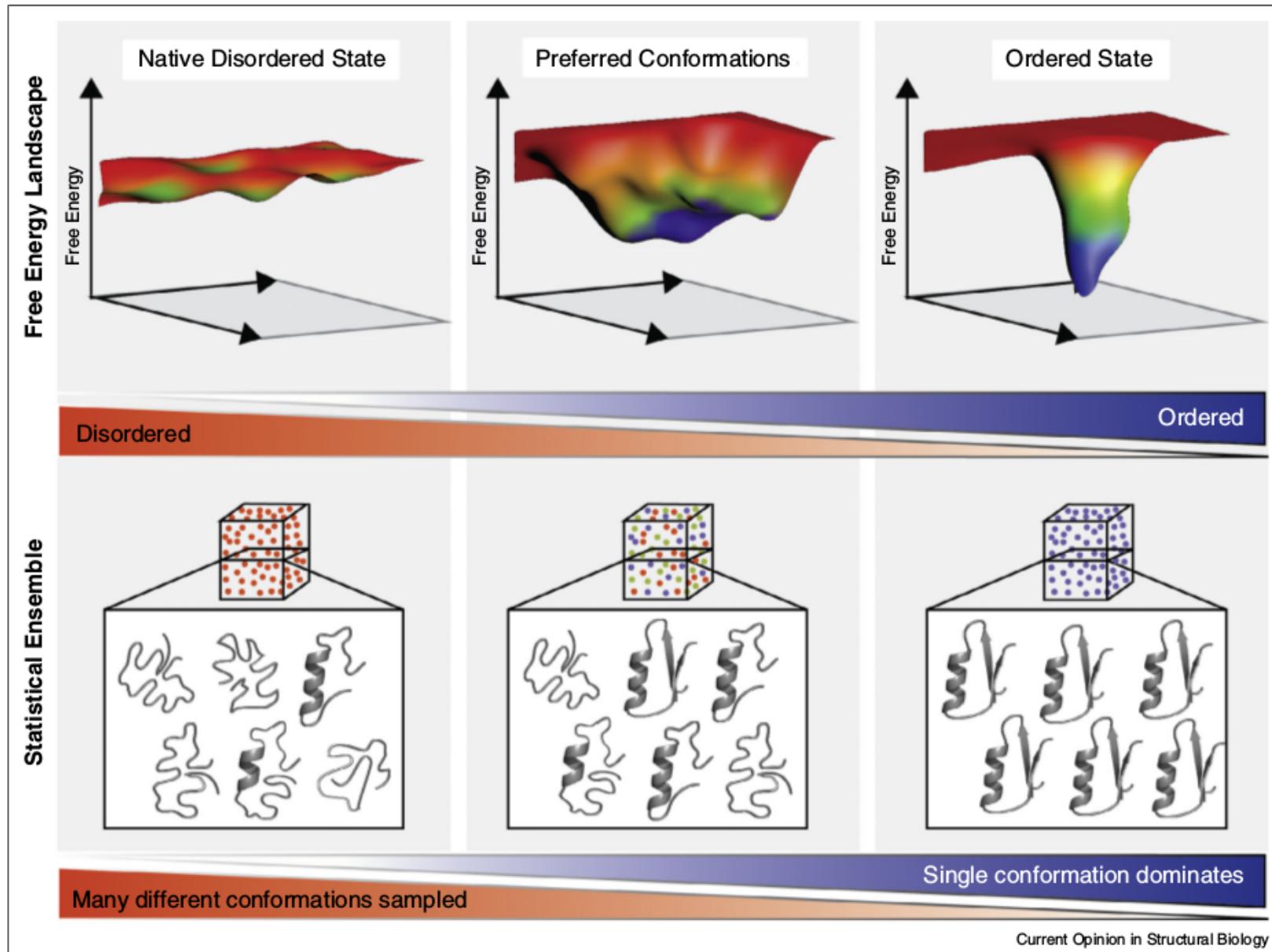
NMR structures with large structural
variations

Primary methods

p53 tumor suppressor



Flexibility vs Disorder



DisProt 7



Indiana University Center for Computational Biology and Bioinformatics Temple University Center for Information Science and Technology

Home Search Browse

DisProt News

Disprot is now running on the new 48 core server. If you encounter difficulties, please let us know at disprot@disorder.compbio.iupui.edu.

Alpha source for the Intrinsically Disordered Protein Ontology ([Idpo.obo](#)) is here.
The [IDP_Ontology](#) interest group listserver is now up and running.

Current Disprot release: **6.02**
Release date: **05/24/2013**
Number of proteins: **694**
Number of disordered regions:
1539

Release notes

Latest additions:

- Natriuretic peptides A
- Peptidyl-prolyl cis-trans isomerase
- Atrial natriuretic factor
- PfEMP1 variant 1 of strain MC
- Adapter molecule crk [Isoform 2]
- more...

Download DisProt

Download DisProt in FASTA or XML format.

Disorder Predictors

Predict disorder, and browse links to other predictors.

Disprot – old version

Database of I

The Database of P information about | states, either in the Center for Compute Medicine and Cen

New Version

DisProt 7 v0.5 11-05-2017

The [old DisProt](#) is still available!

Statistics

Proteins 803
Regions 2167

Start

You can do a [complex search](#) from the Browse page or use [Blast](#) from the Search page.

Citing DisProt

Piovesan D et al. [DisProt 7.0: a major update of the database of disordered proteins](#) Nucleic Acids Res., 2016.

[Go to PubMed](#)

[Go to NAR](#)

Welcome to DisProt

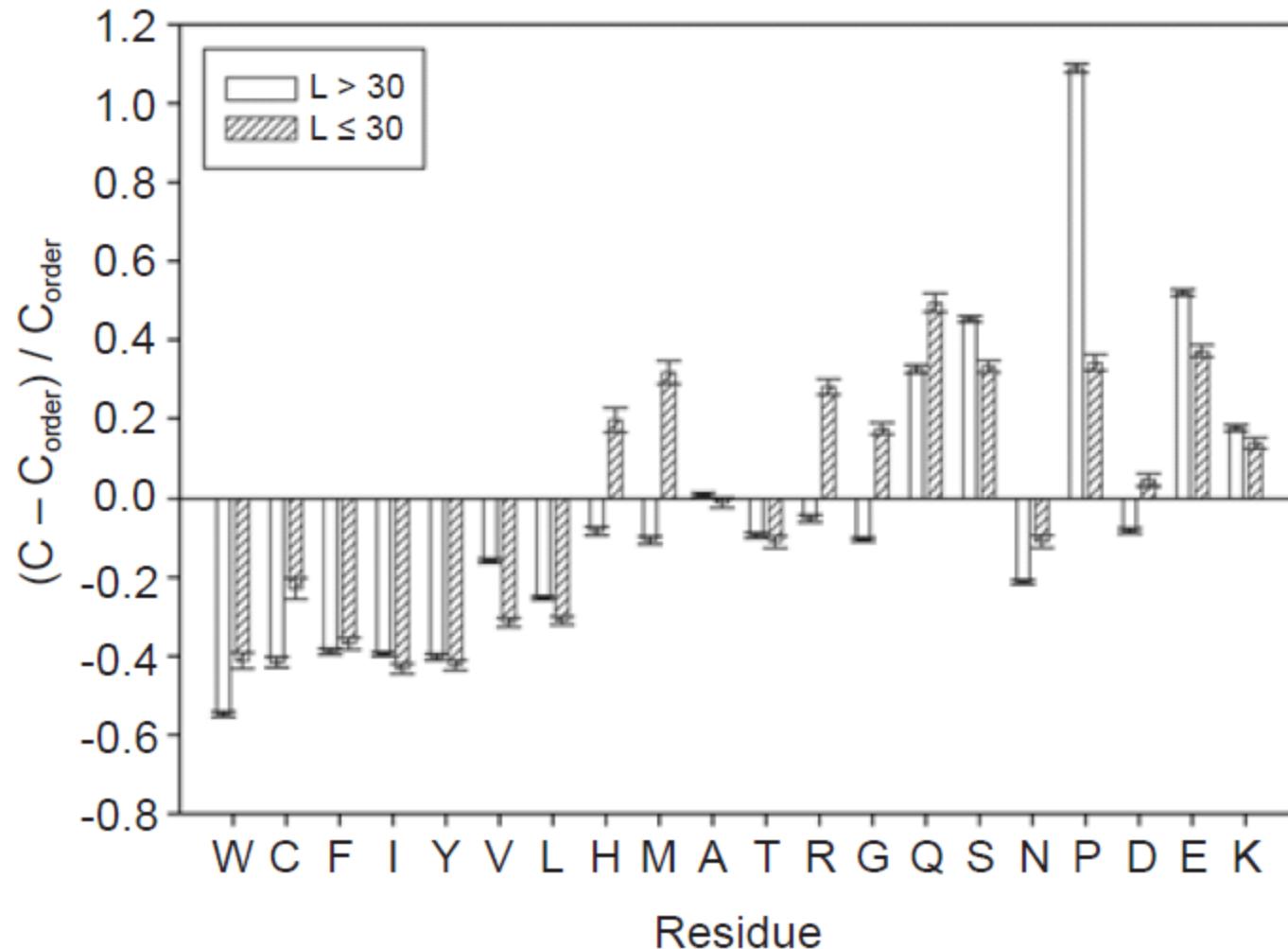


DisProt is a [community resource](#) annotating protein sequences for intrinsically [disorder regions](#) from the literature.

It classifies intrinsic disorder based on [experimental methods](#) and three ontologies for [molecular function](#), [transition](#) and [binding partner](#).

See [About](#) page for more information.

Amino Acid Compositions



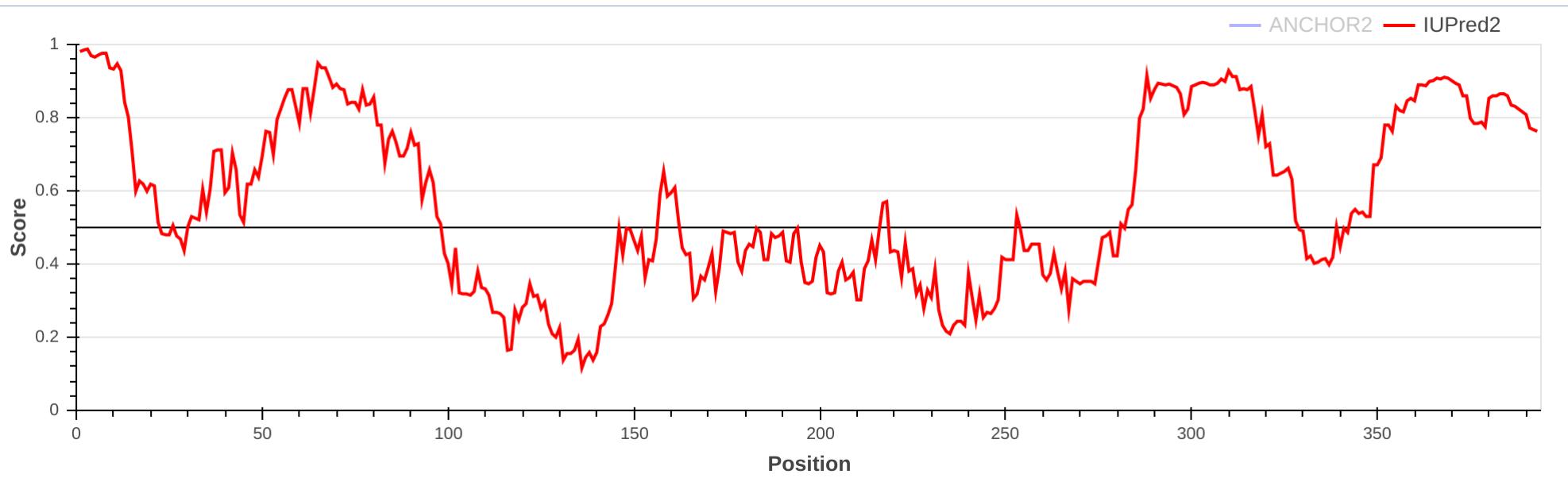
Protein disorder is encoded in the amino acid sequence

Prediction methods for protein disorder

Over 50 methods ...

- Based on amino acid propensity scales
 - GlobPlot (secondary structure propensities)
- Simplified biophysical models
 - IUPred
Disordered protein cannot form enough favorable interactions to ensure the stability of a folded structure
- Machine learning approaches
 - Disopred, Espritz
- Meta servers

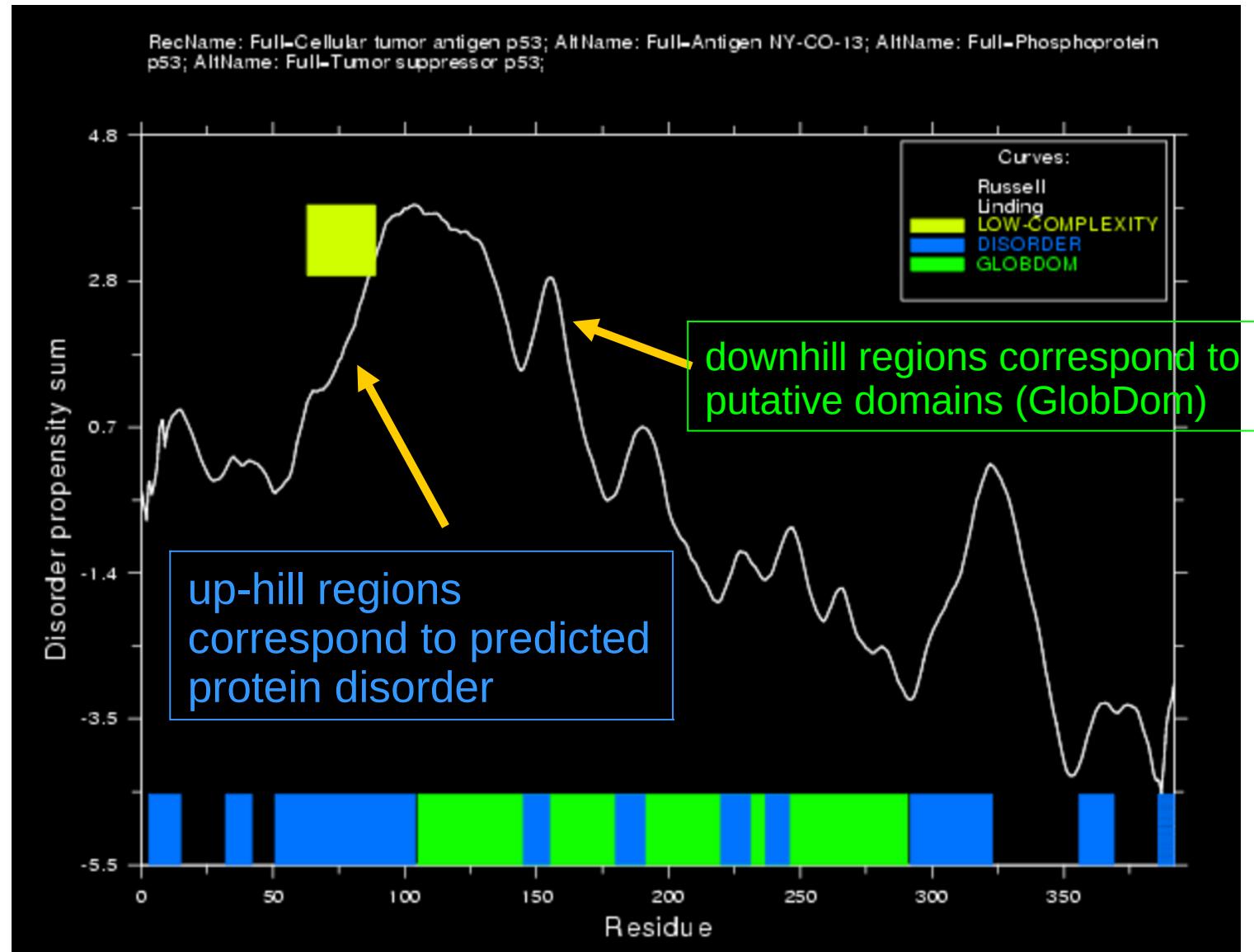
A typical output (IUPred2A)



Predictions are on a per residue basis

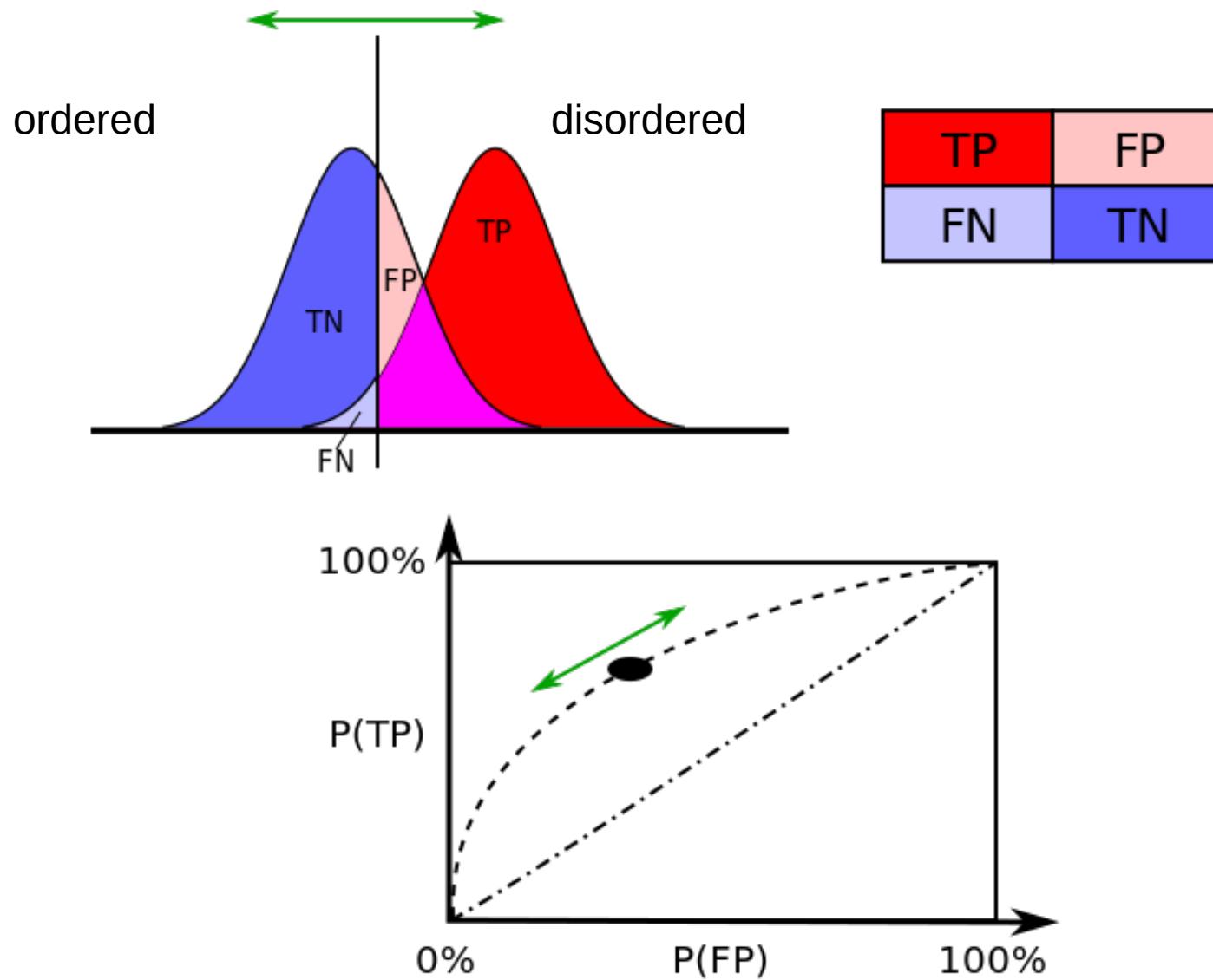
Dosztányi et. al. (2005) JMB 347, 827
Erdős et al. (2018) NAR 46(W1):W329

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



Where are the ordered domains, disordered regions?

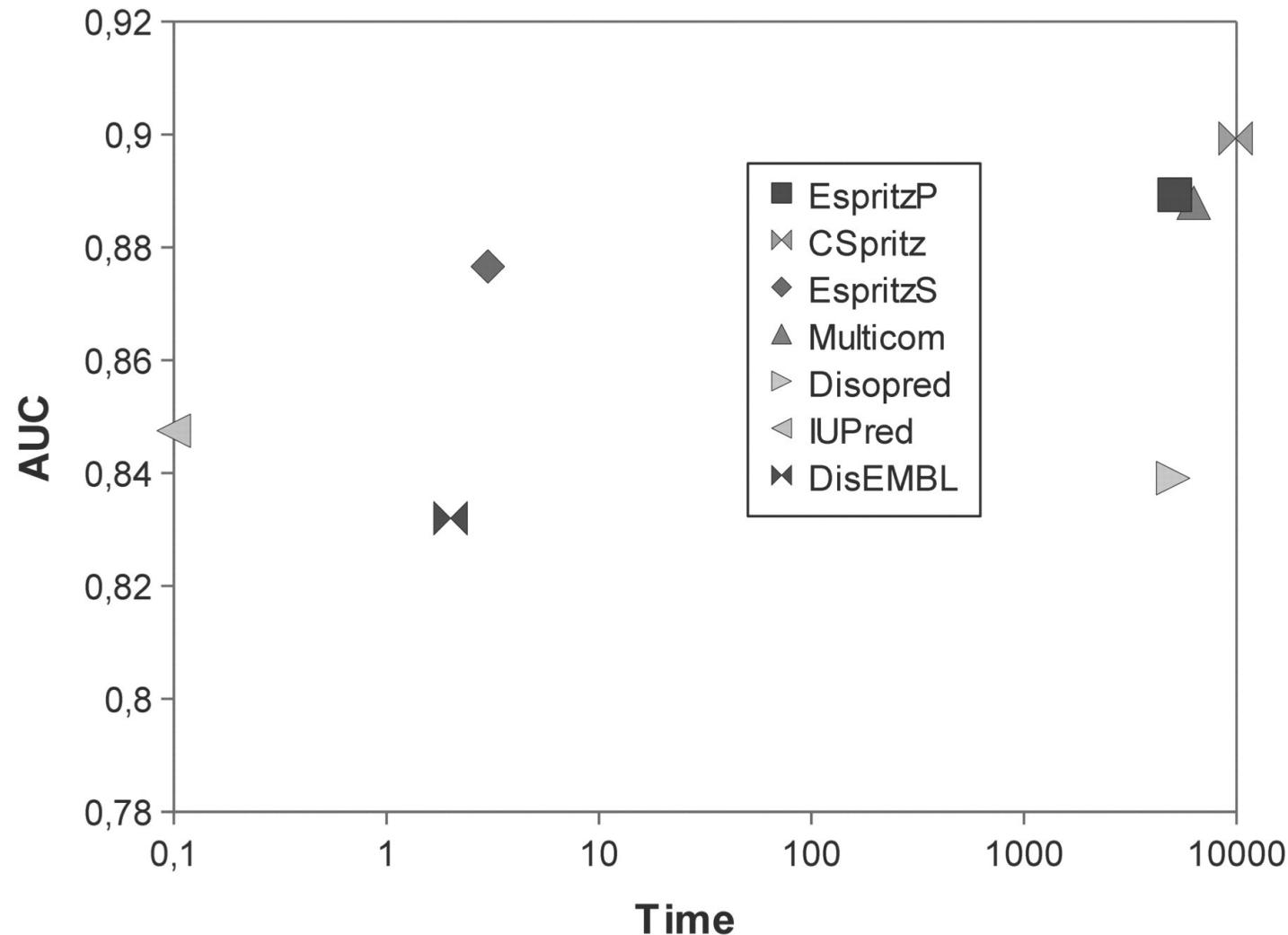
Evaluation of prediction methods



Prediction of protein disorder

- Disordered is encoded in the amino acid sequence
- Can be predicted from the sequence
- ~80% accuracy (>0.8 AUC)
- Challenges
 - Small, noisy datasets
 - Disorder is heterogeneous

Which is the best method? Speed

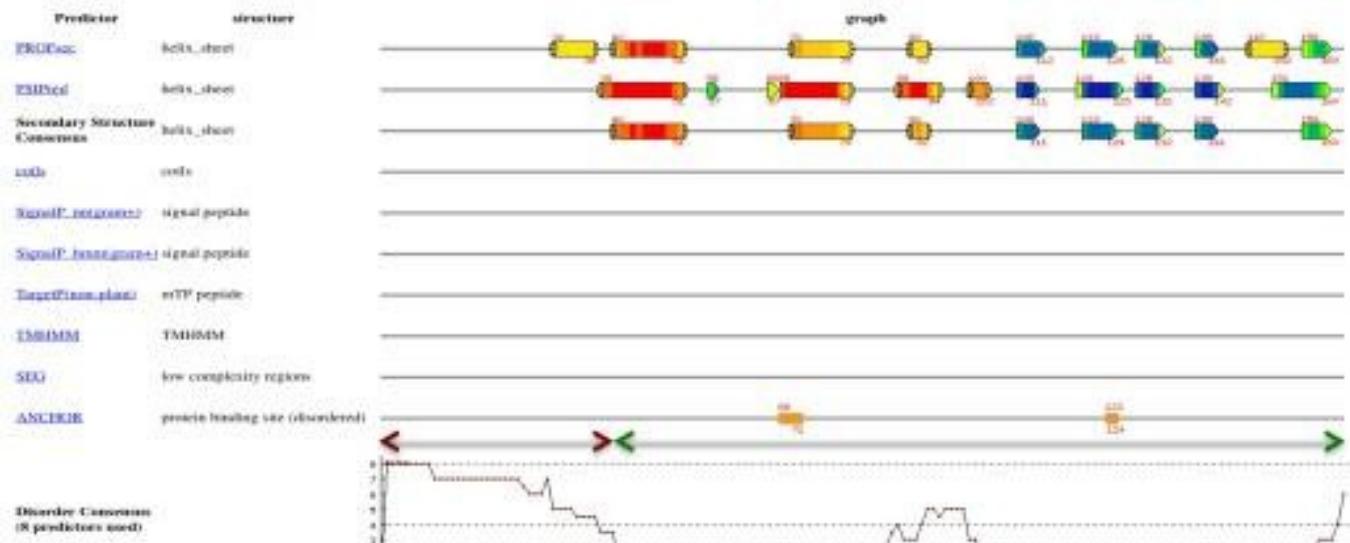


Time versus performance plot for different predictors.

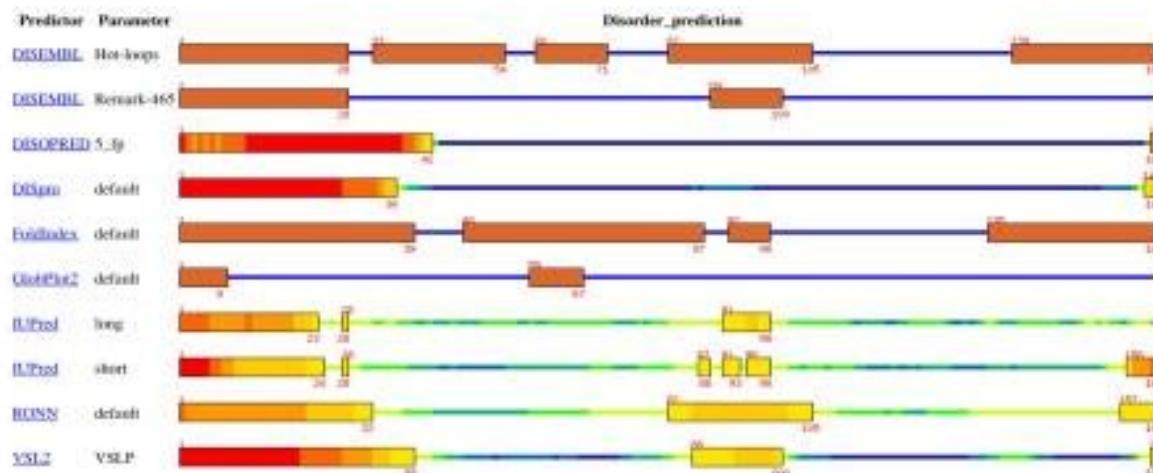
Which is the best method?

Specific application – Construct optimization

A. Disorder Prediction



B. Disorder Prediction



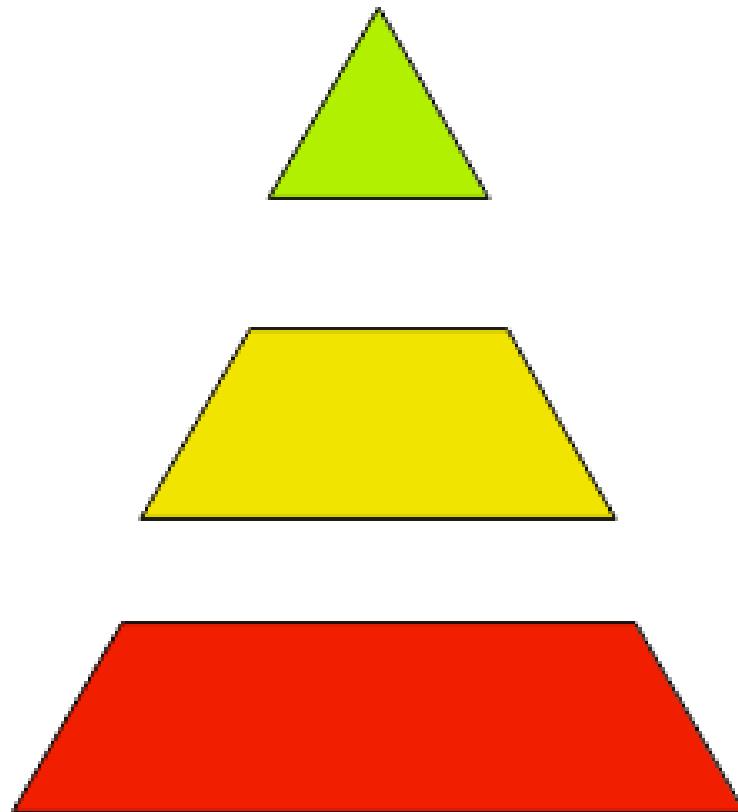
How to predict disorder?

- Use multiple disorder prediction methods
- Incorporate information about structure, domains
- Evolutionary information (families)
- Use additional prediction tools
 - Topology, coiled coil, repeats, low complexity, secondary structure

MobiDB



MobiDB



Database

Manually curated annotations from external databases

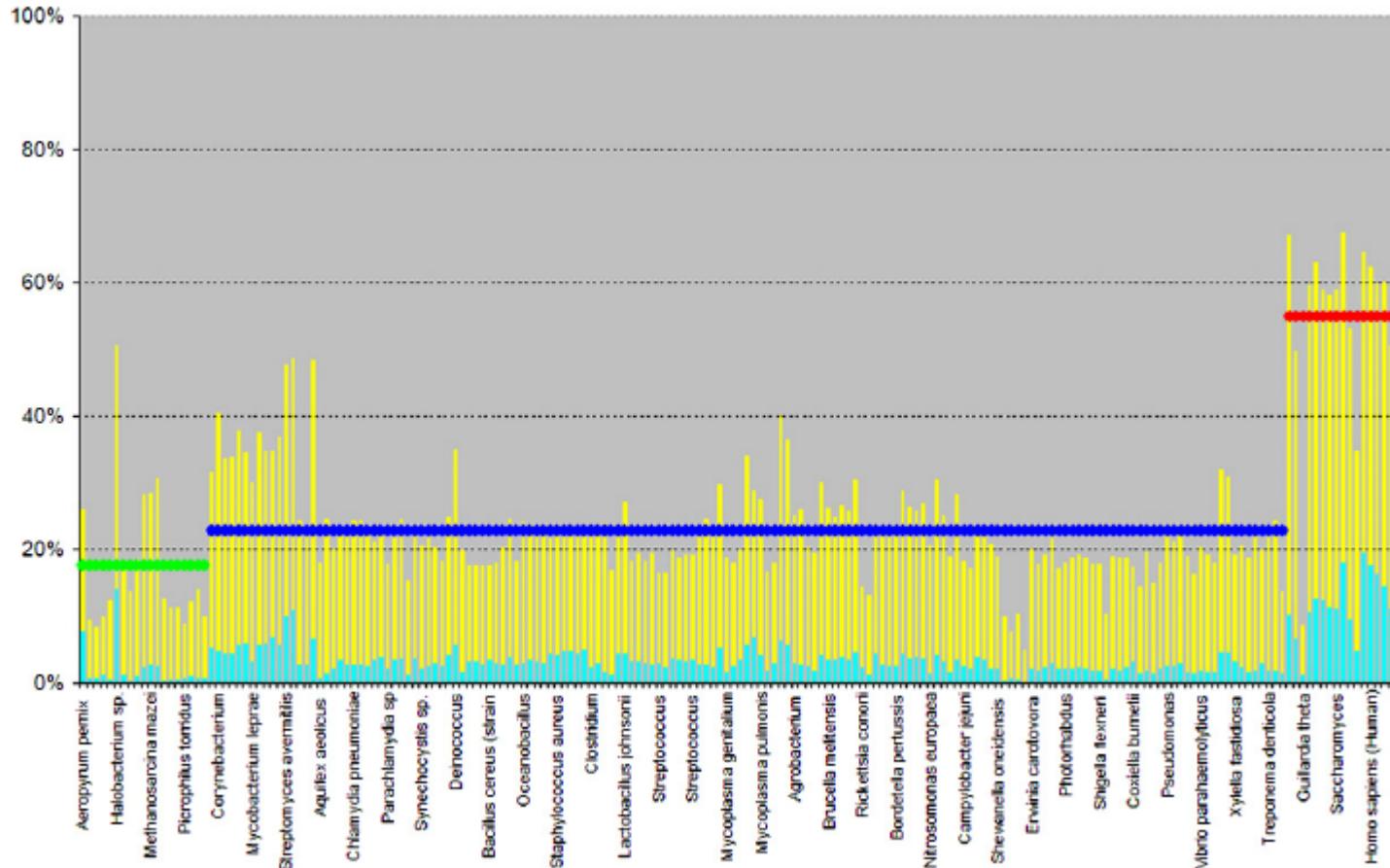
Indirect

Derived/calculated information from experimental data, i.e. PDB structures and/or chemical shifts

Predictions

Predicted annotations

Protein disorder is prevalent



Lobanov MY, Galzitskaya OV.
Mol Biosyst. 2012;8:327-337.

Protein disorder complements the functional repertoire of globular proteins

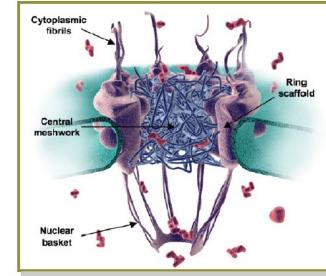
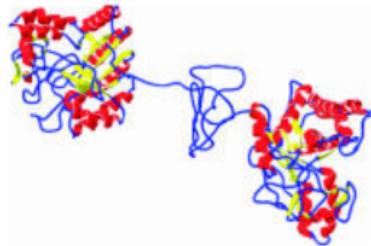
Table 2. Correlation and anticorrelation of structural disorder with Swiss-Prot functional categories

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

How IDPs carry out their functions?

- Entropic chains

Function directly results from disordered state



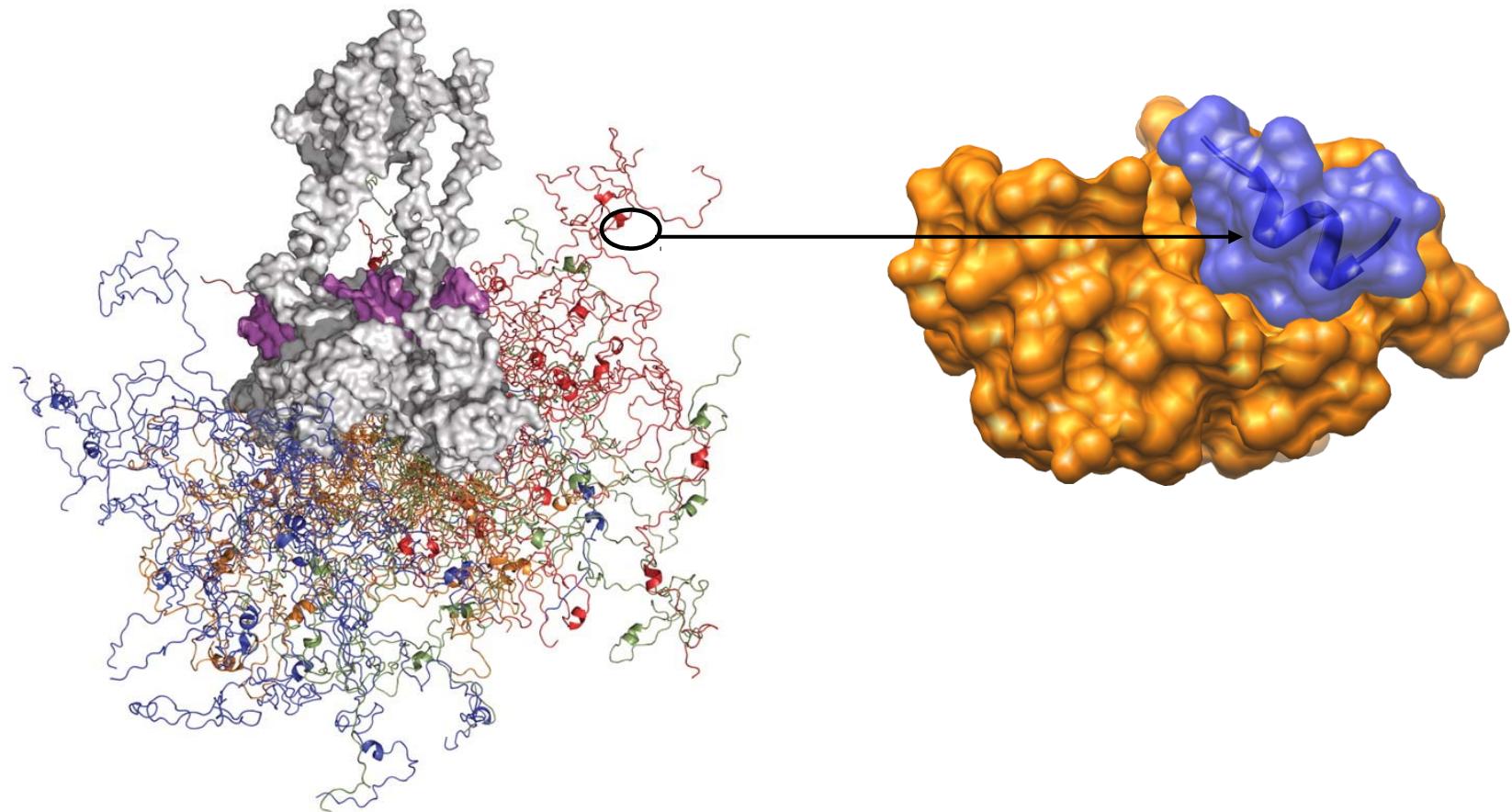
- Molecular recognition

Coupled folding and binding

- “Assemblages”

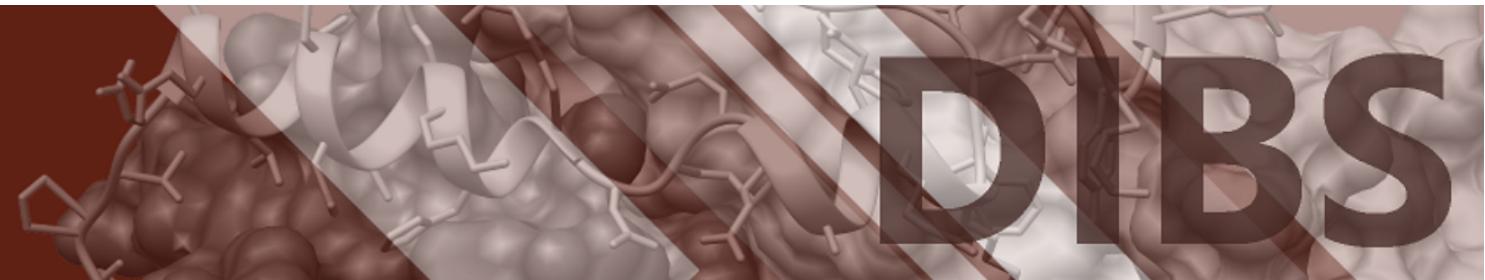
Functional sites formed by phase separation

Interaction of IDPs

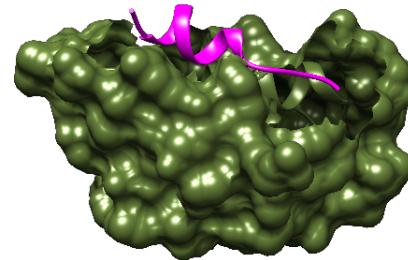


Complex between p53 és MDM2

Coupled folding and binding



- 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

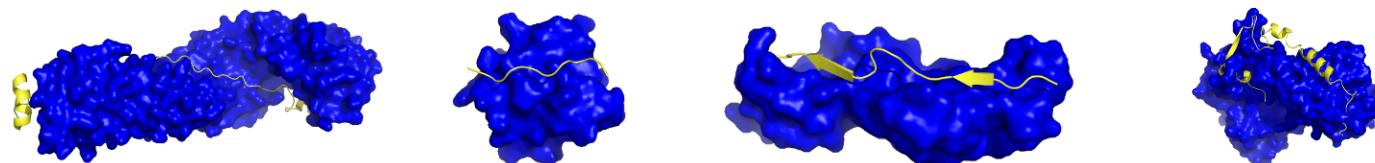
DIBS: a repository of disordered binding sites mediating interactions with ordered proteins

Eva Schad¹, Erzsébet Fichó¹, Rita Pancsa², István Simon¹,
Zsuzsanna Dosztányi^{3,*} and Bálint Mészáros^{1,3,*}

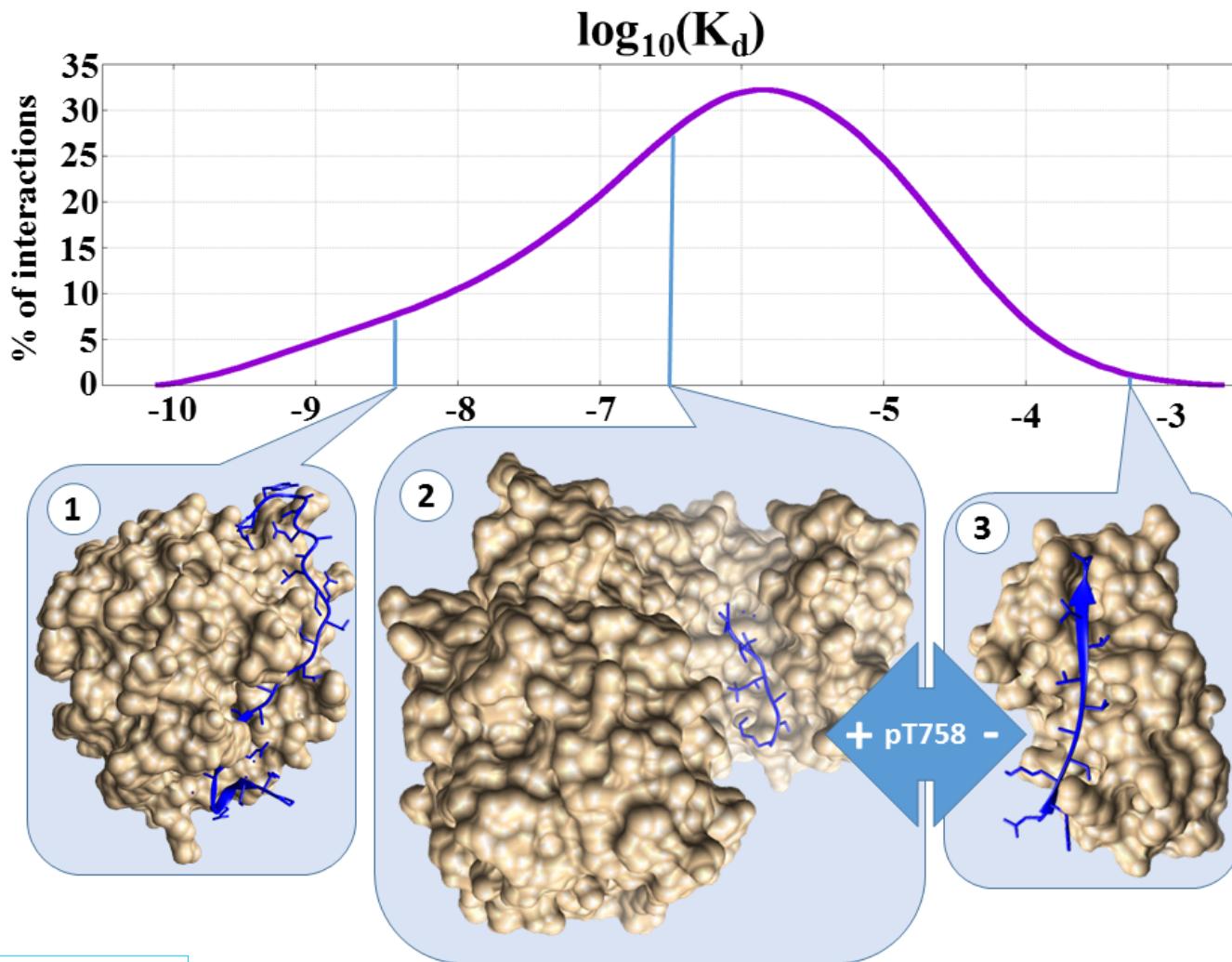
Bioinformatics (2017)
doi: 10.1093/bioinformatics/bt
x640

Various types of IDP complexes

- Adopted secondary structure
- Size of interaction surface
- PTMs
- Disorder in the bound states (fuzziness)
- Preformed structural elements
 - Transient secondary structure elements



Kd values



Thrombin-anophelin

filamin-integrin $\beta 2$

14-3-3 ζ -integrin $\beta 2$

Predicting functional regions within IDPs

(~10, as opposed to the more than 50 disorder prediction methods)

Morfs, LIPs, PresMos, ...

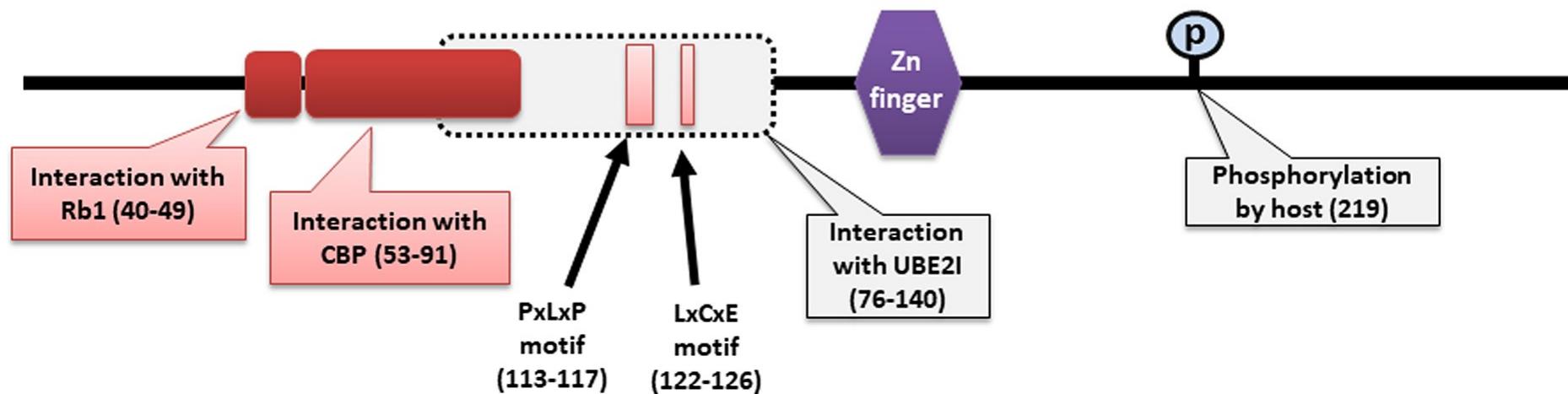
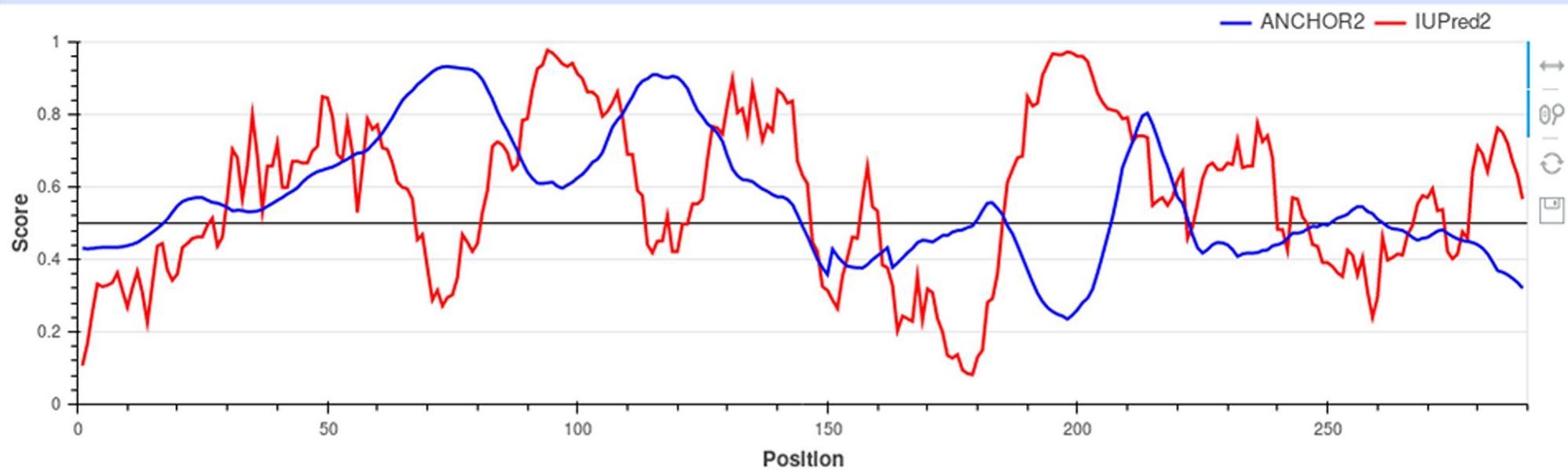
- ❑ Biophysical properties (**ANCHOR**)
- ❑ Machine Learning methods
(MorfPred, Morf_{chibi}, DISOPRED3)
- ❑ Linear motifs
(Regular Expression, PSSMs)
- ❑ Conservation patterns (**SlimPrints, PhyloHMM**)
- ❑ Non-binding regions (e.g. linkers)
(DFLpred)

Prediction of disordered binding regions – 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
 - Captures sequential context

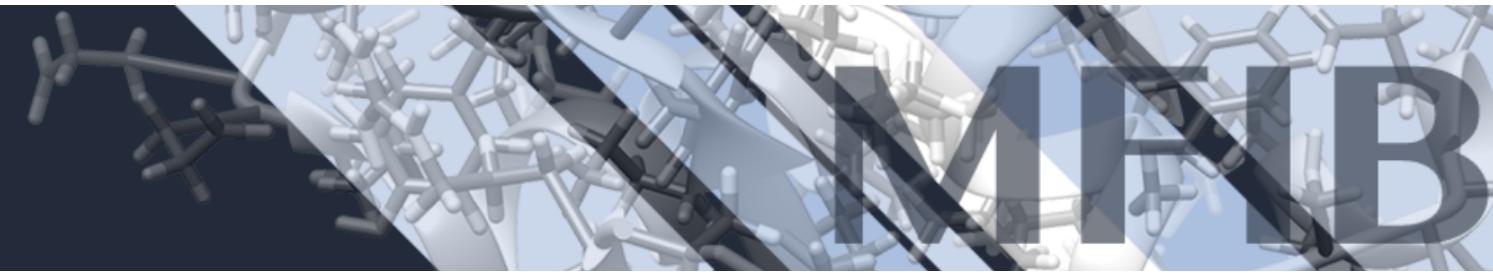
ANCHOR

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



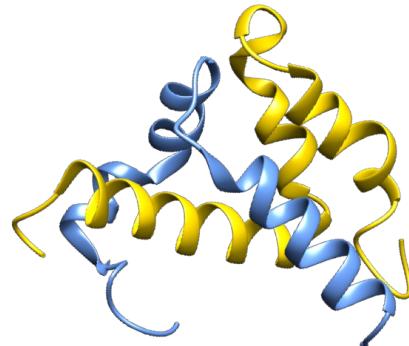
Human adenovirus C early E1A protein

Mutual Folding Induced by Binding



- 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

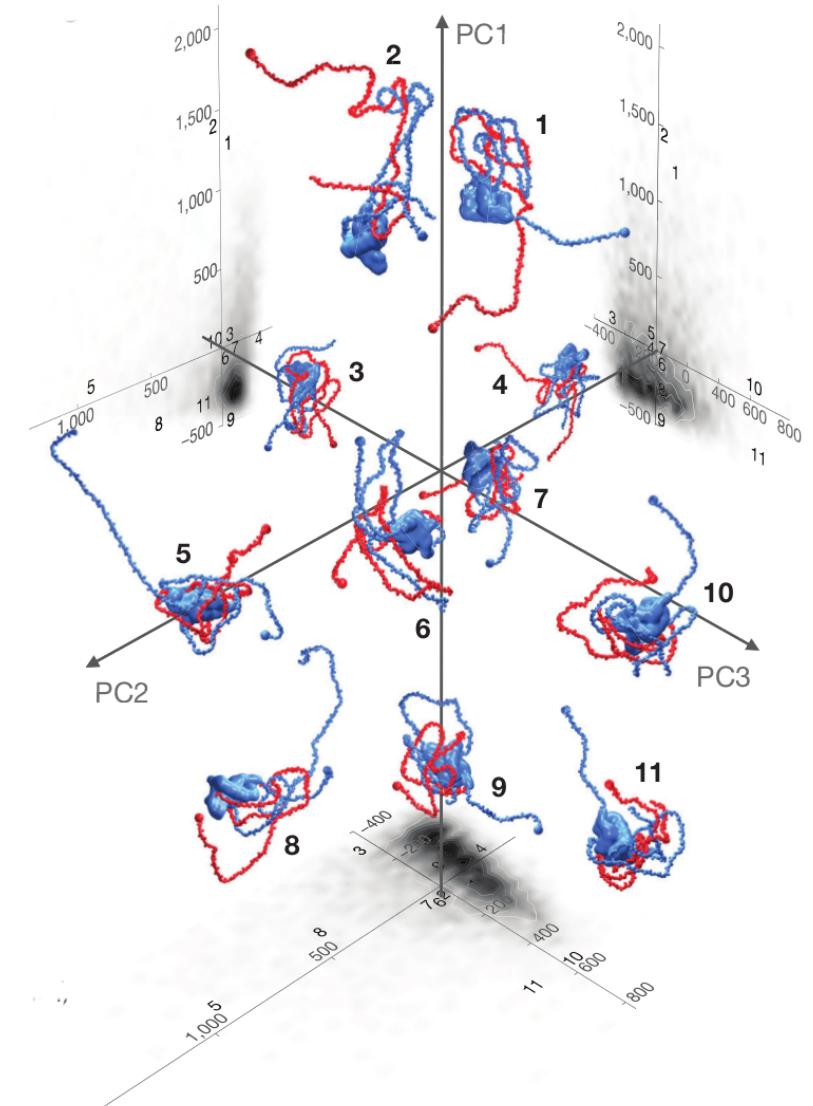
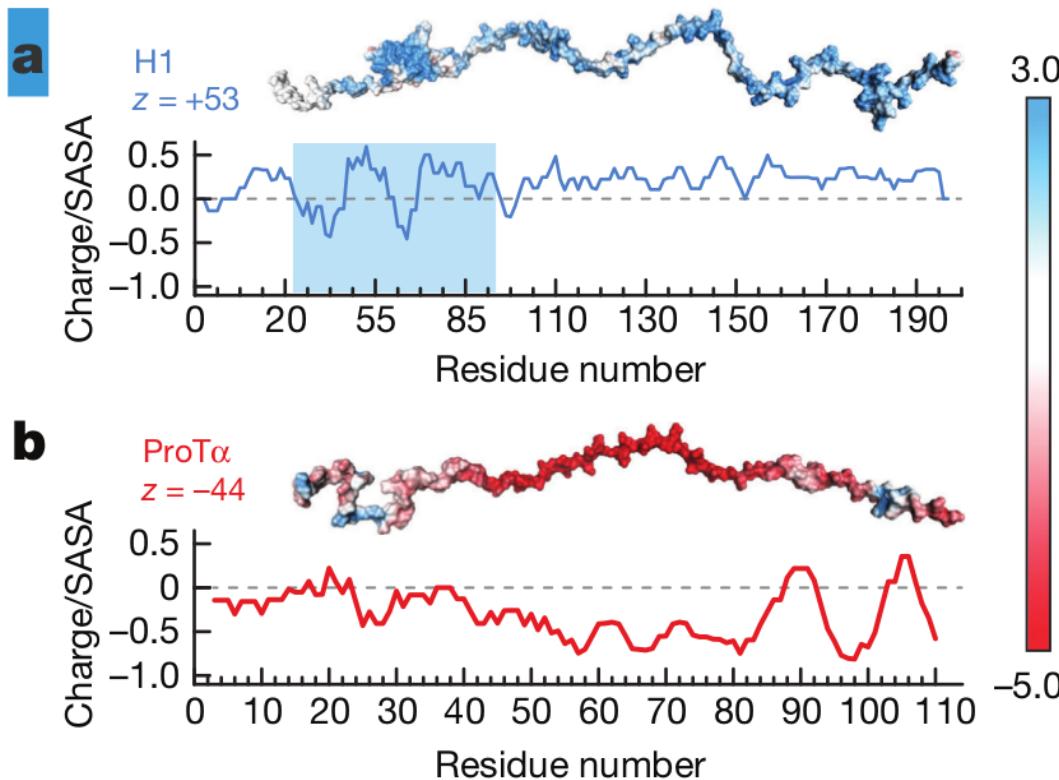
Erzsébet Fichó¹, István Reményi², István Simon^{1,*}
and Bálint Mészáros^{1,*}

Bioinformatics (2017)

PMID: 29036655

doi: 10.1093/bioinformatics/btx486

Extreme disorder in complex



Fuzzy complex

Borgia et al. Nature 2018, 555, 61.

Disprot – functional annotation

- 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)

Six basic functions

