
PART II.

Prediction of functional regions within disordered proteins

Zsuzsanna Dosztányi

MTA-ELTE Momentum Bioinformatics Group

Department of Biochemistry Eotvos Lorand University,

Budapest, Hungary

dosztanyi@caeser.elte.hu

Large-scale analysis of IDPs

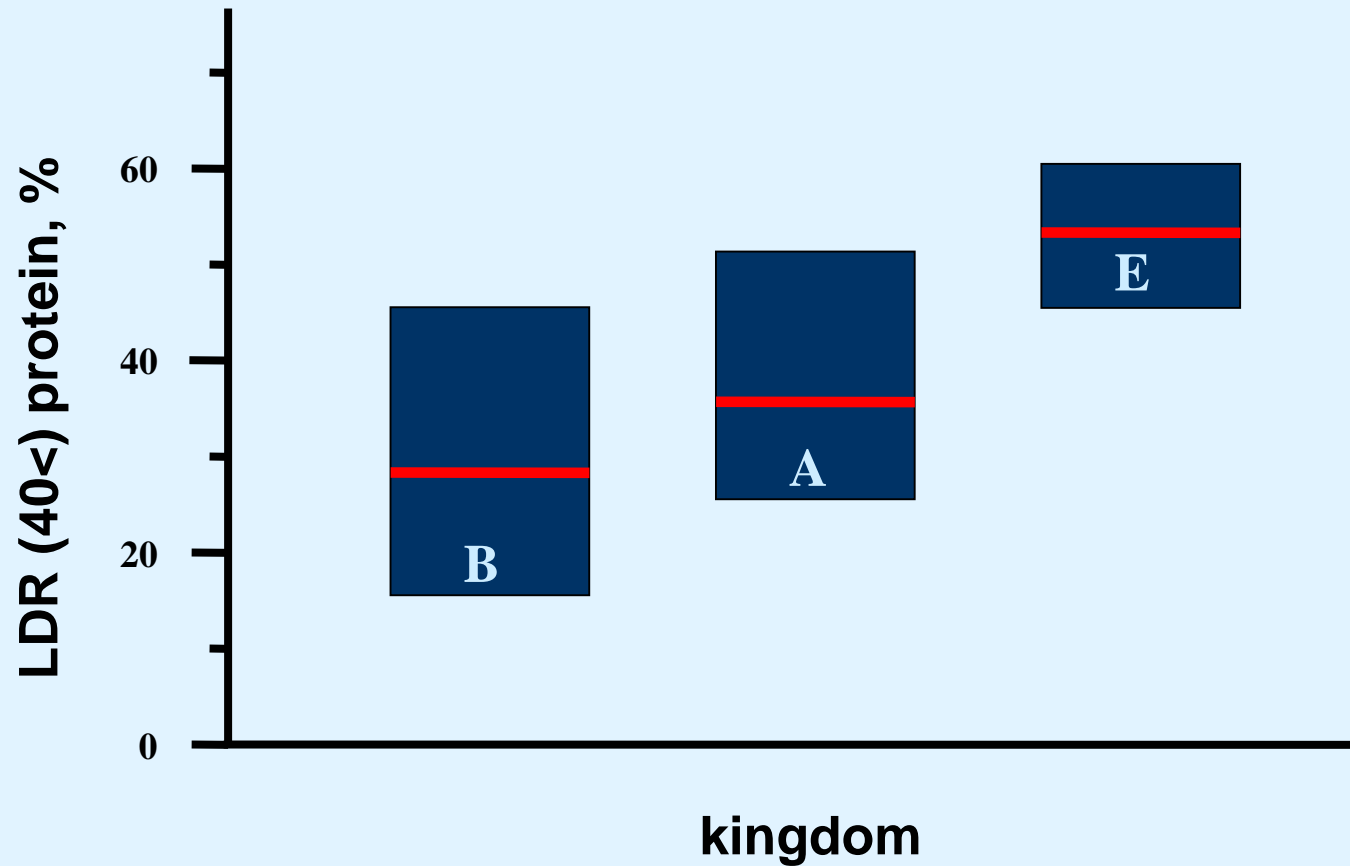
Possible through prediction methods

- Functional properties
 - Evolutionary properties
 - Disorder characterization
 - Percentage of properties with long (>30 or >40 aa)
 - Percentage of disordered residues
-

How common is protein disorder?

- Around 50% of human proteins have long disordered regions
 - Around 30% of residues in the human proteome are predicted as disordered
 - Disorder content increases with evolutionary complexity
-

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

Functions of intrinsically disordered proteins

I Entropic chains

II Linkers

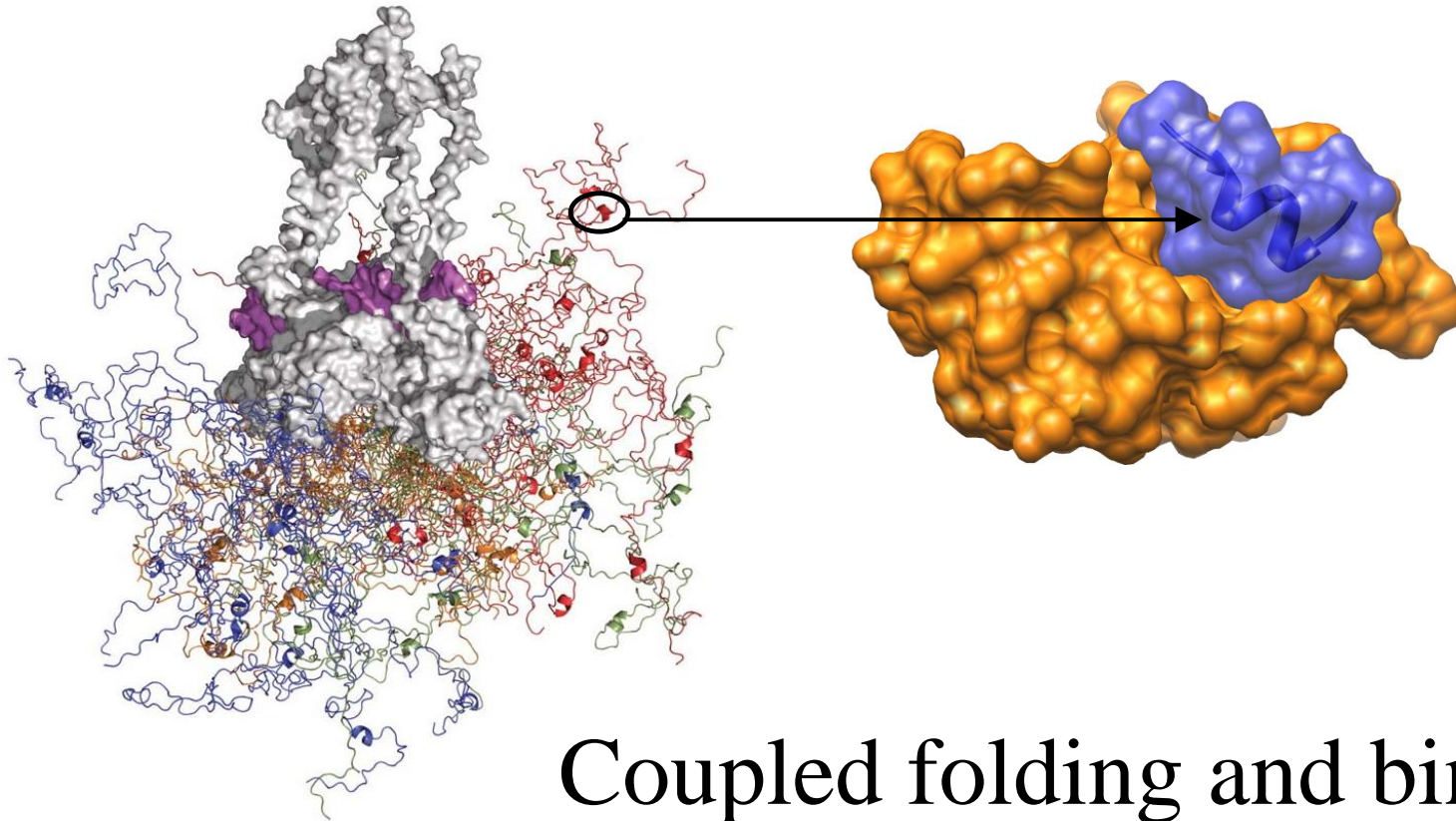
III Molecular recognition

IV Protein modifications (e.g. phosphorylation)

V Assembly of large multiprotein complexes

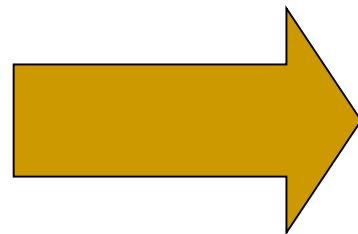
Interaction of IDPs

Complex between p53 and MDM2



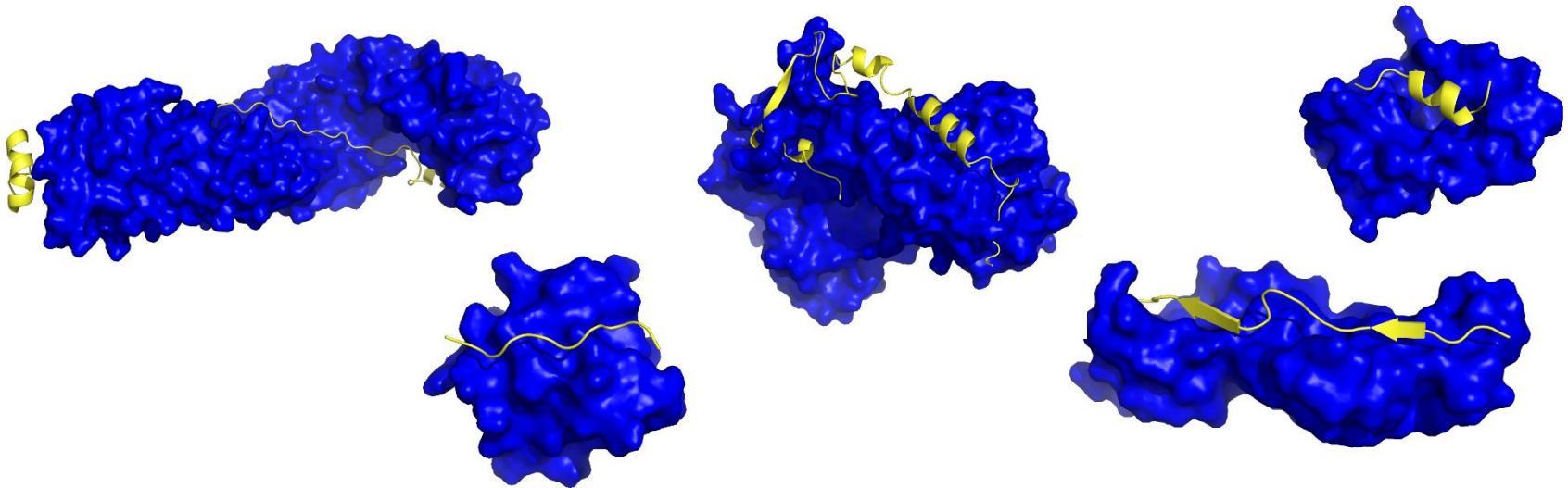
Coupled folding and binding

- Functional advantages
 - Weak transient, yet specific interactions
 - Post-translational modifications
 - Flexible binding regions that can overlap
 - Evolutionary plasticity



*Signaling
Regulation*

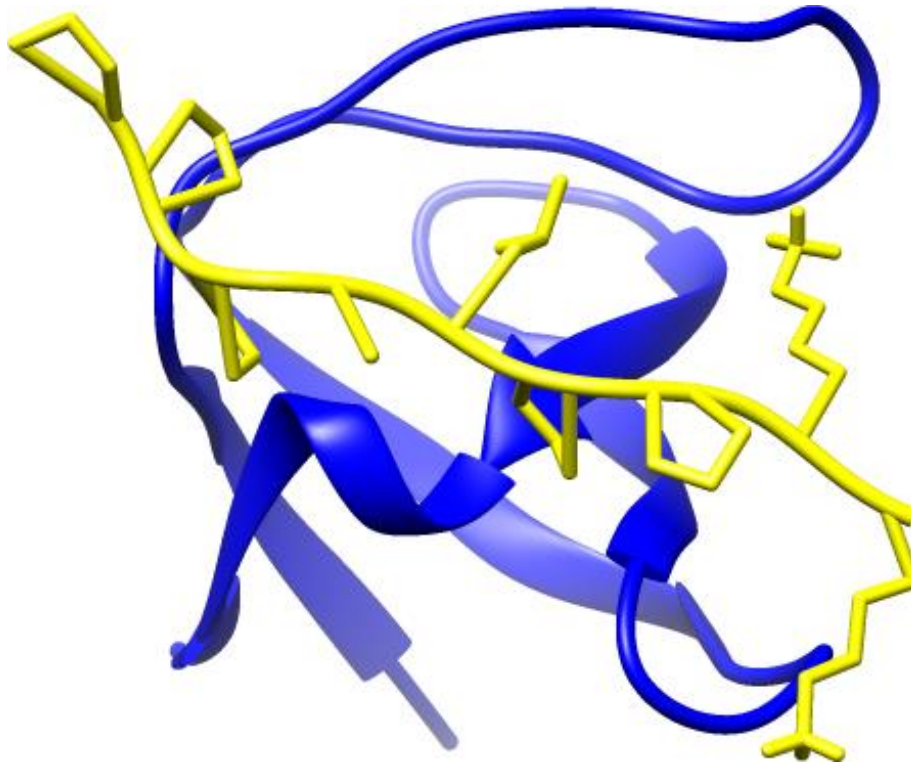
Various complexes of IDPs



- Can be grouped according the adopted secondary structure elements
 - ❑ alpha helical
 - ❑ beta strand
 - ❑ polyproline
 - ❑ irregular

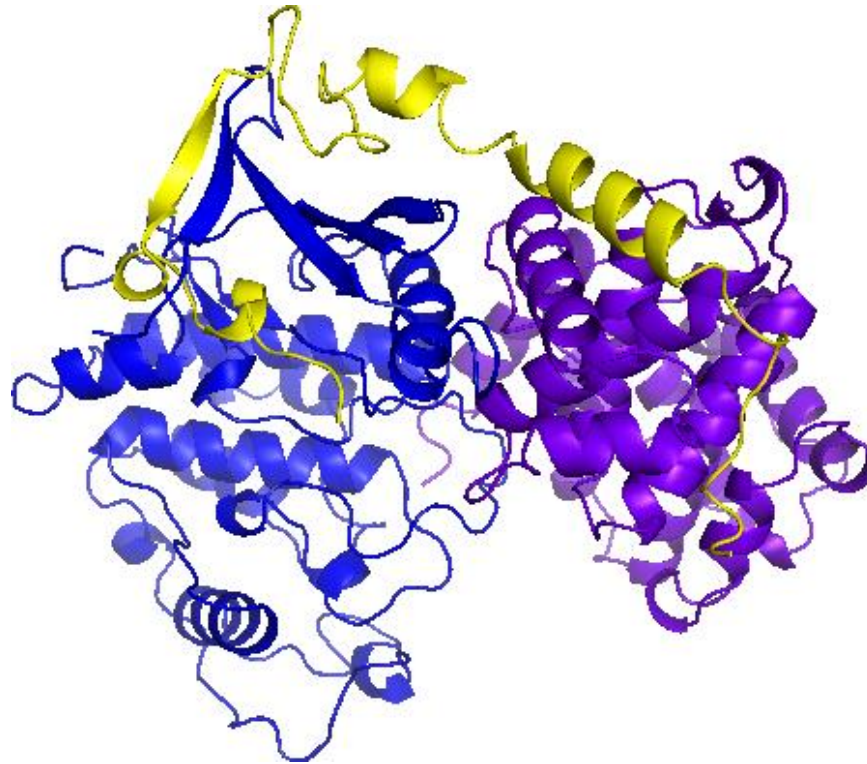
Small interfaces

SH3 domain

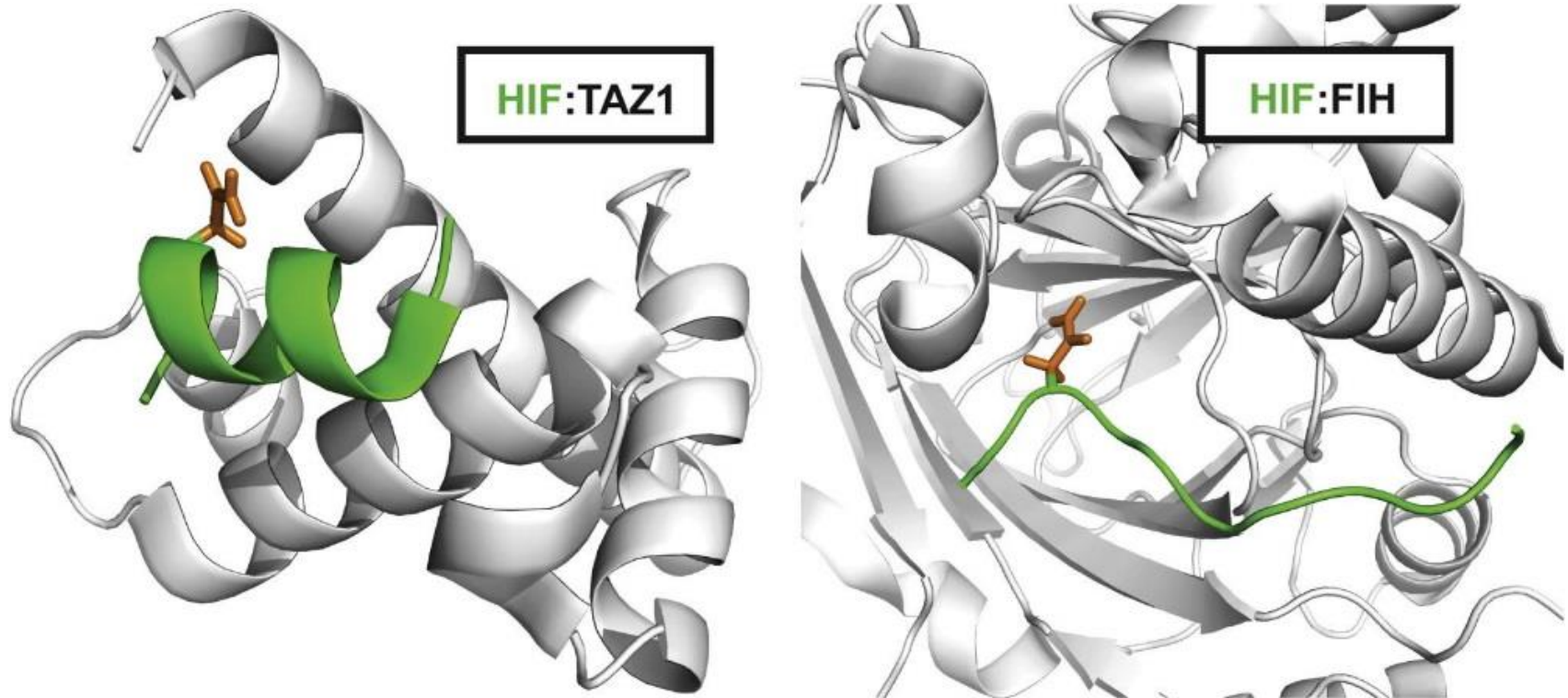


Large interfaces

- Cyclin-dependent kinase (Cdk) inhibitor, p27Kip1 (p27)
- Binds to cdk-cyclin complex and inhibits their activity
- Fully disordered protein

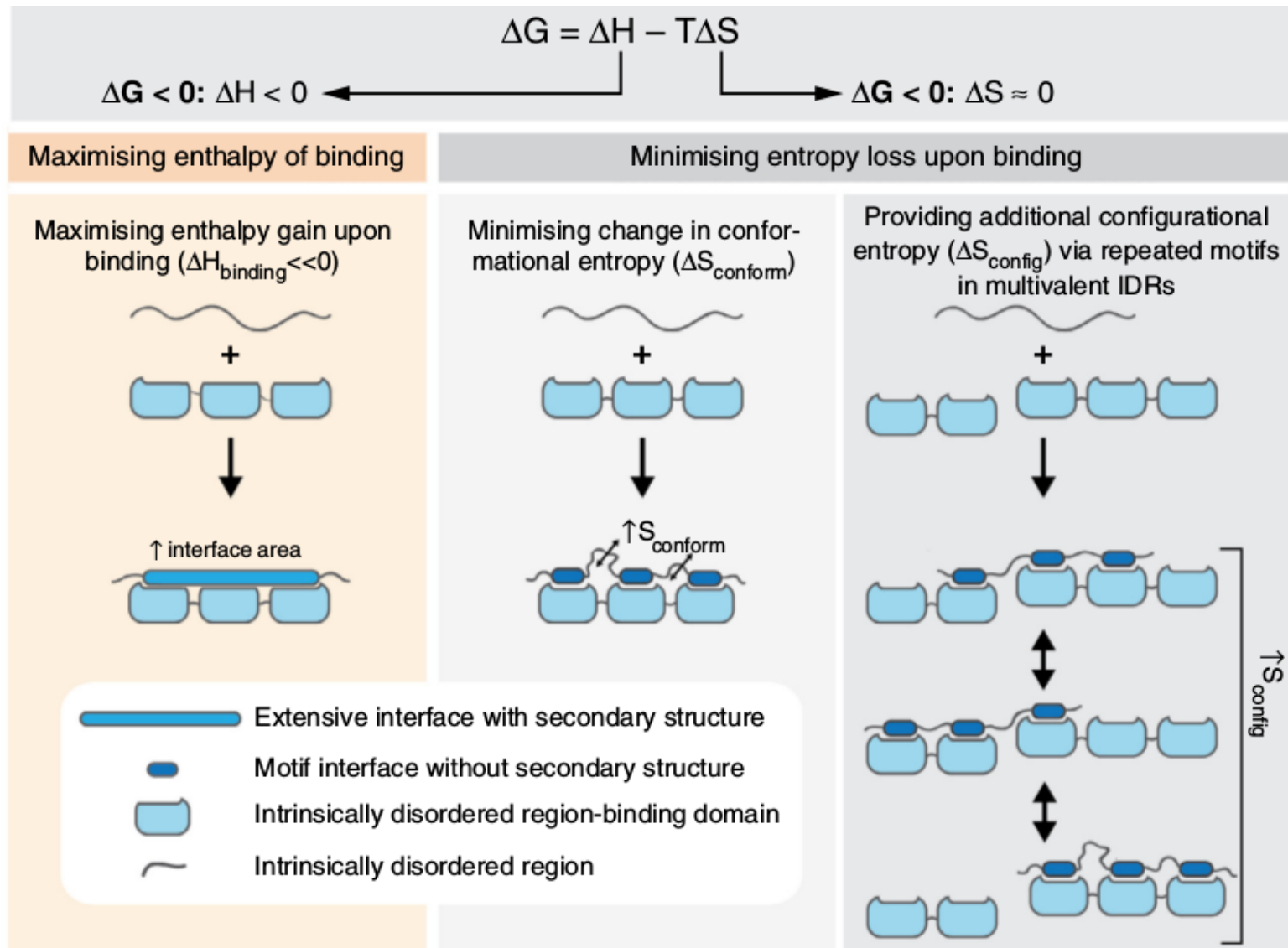


Conformational plasticity

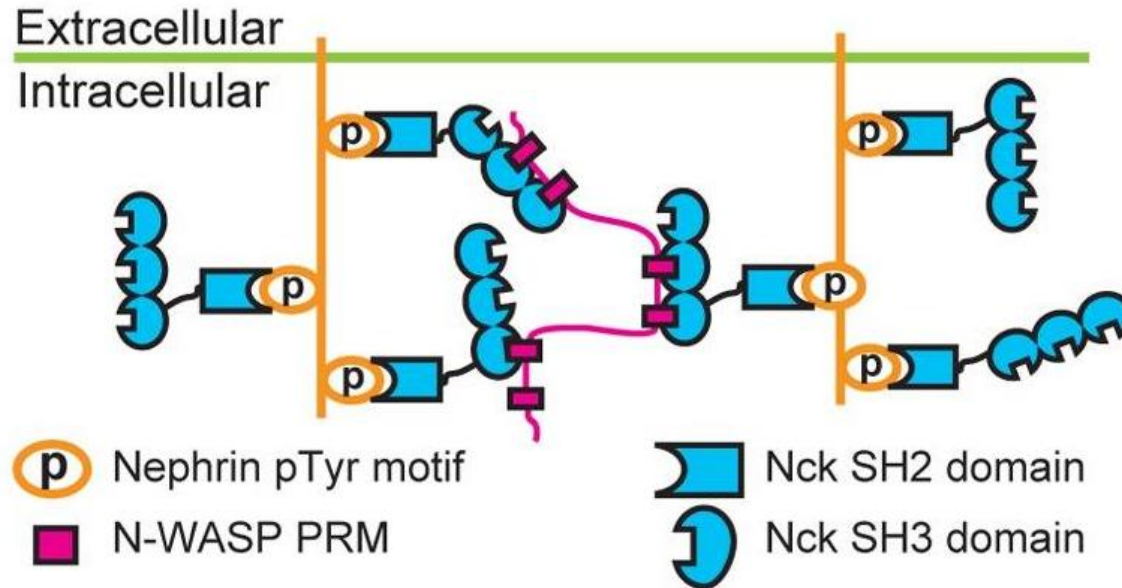


C-terminal transactivation domain (CTAD) of the hypoxia inducible factor-1 α

Fine tuning the entropic component



Phase transitions



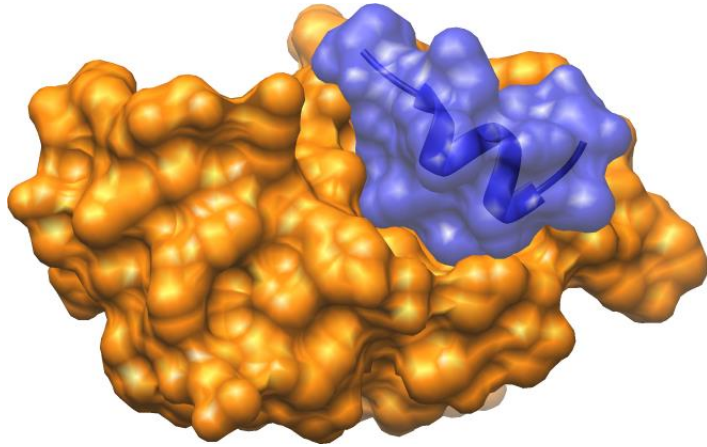
- ❑ Multivalency and weak interactions
- ❑ Regulated by phosphorylation
- ❑ Transition from small complexes and large, dynamic supramolecular polymers.

Disordered binding regions

- Complexes of IDPs in the PDB: ~ 200
- Known instances: ~ 2 000
- Estimated number of such interactions in the human proteome: ~ 1 000 000

- Experimental characterization is very difficult
- Computational methods

Disordered protein complexes



Complex between p53 and MDM2

- Interaction sites are usually *linear* (consist of only 1 part)
- enrichment of interaction prone amino acids

Sequence



No need for structure,
binding sites can be
predicted from
sequence alone

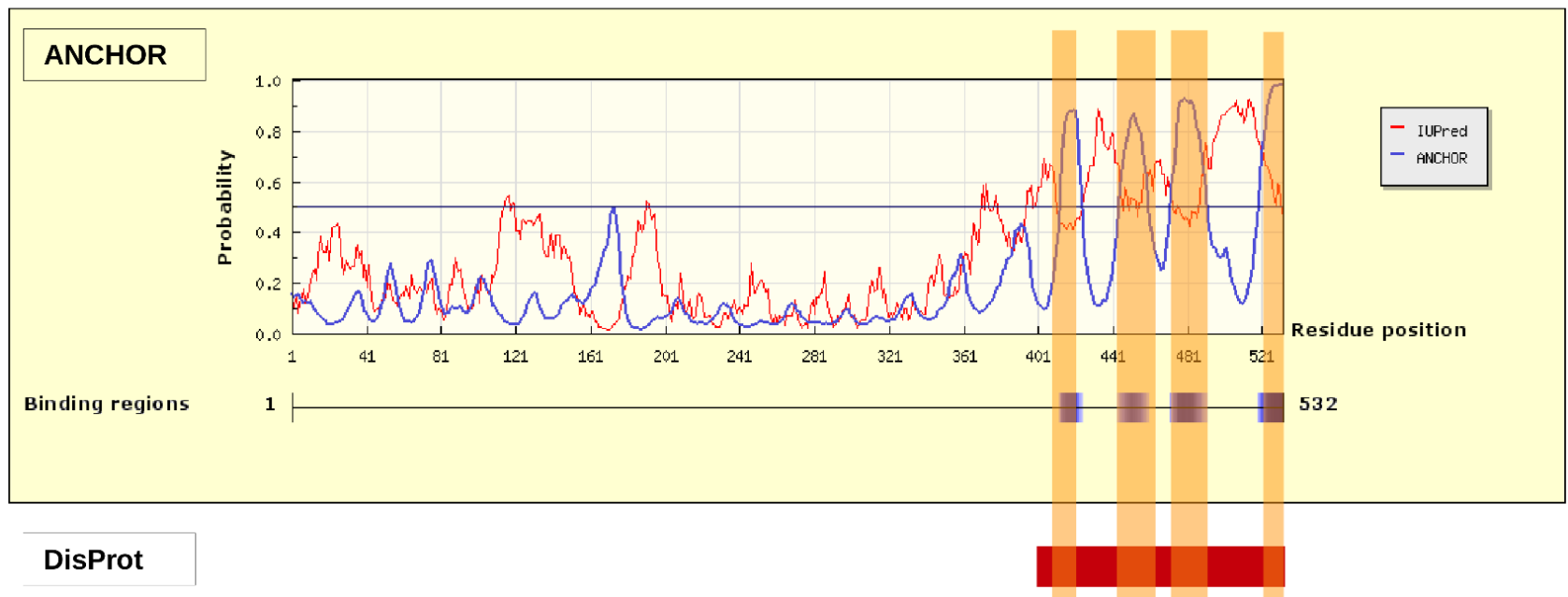
Binding sites

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

nucleoprotein from Nipah virus (DP00697)



Machine learning approaches

MORFchibi

- ❑ Uses two SVMs
 - One recognizes the different amino acid composition of flanking regions compared to the binding region
 - One recognizes the similarity to known binding regions
- ❑ trained on short chains in complex

Machine learning approaches

DISOPRED3

- ❑ Uses three SVMs
 - Simple sequence profile
 - PSI-Blast profiles (very slow)
 - PSI-Blast profiles with global features
 - ❑ trained on short chains in complex
-

Amount of disordered binding regions

- What is the amount of disordered protein regions in the human proteome?
 - ❑ ANCHOR: 93429
 - ❑ MORFchibi: 275013
 - ❑ DISOPRED3: 63848
 - We cannot tell what is the *false positive rate* of these methods
-

Conservation

The functionality of a protein segment is often approached by investigating the evolutionary history of its primary sequence

Can this approach used for disordered proteins?

Sometimes ...

Constrained and flexible disorder

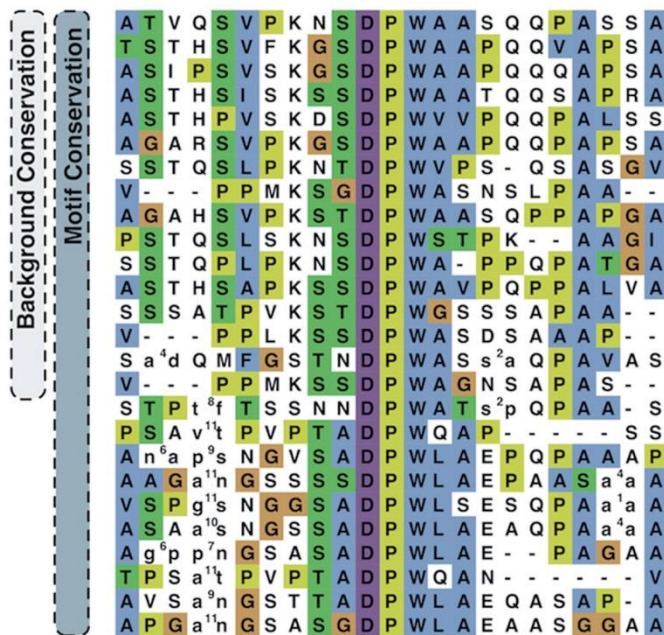
- ‘constrained’,
 - if both features (amino acid sequence and the property of disorder) are conserved
- ‘flexible’,
 - if only disorder is conserved
- ‘non-conserved’ positions
 - where disorder is not conserved

DISCONS

novel server (preliminary)

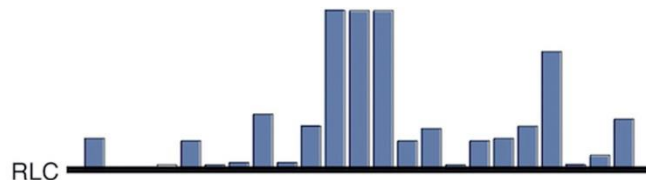
Conservation patterns of linear motifs

- No evolutionary constraints to keep the structure
- Strong constraints on functional site



C

Island-like conservation



SlimPrints

- Generates sequence alignments of orthologous sequences
 - Relative conservation score per position
 - Filters out less reliable regions
 - Fails if sequences are too divergent, or too similar
 - <http://bioware.ucd.ie/slimprints.html>.
-

The next challenge:

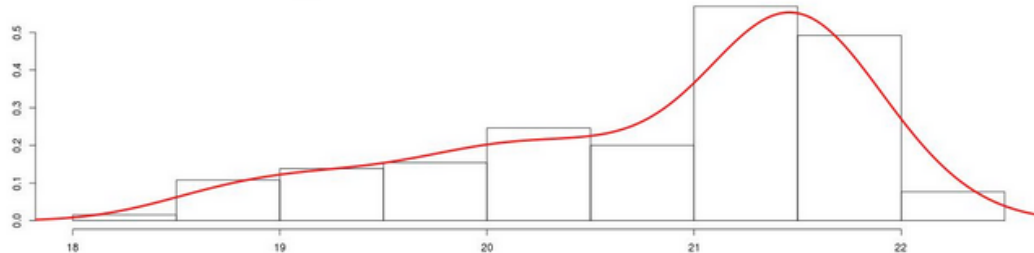
- Characterizing the ensemble of conformations of IDPs
 - And their relationship with function
-

Ensemble characterization for IDPs

- Experimental methods cannot detect a single conformation, only time or ensemble averages
 - Combination of methods are needed (NMR, SAXS)
 - Methods are used to characterize
 - Radius of gyration
 - Transient secondary structure elements
 - Transient long range contacts
-

PED database

Rg distribution of Ensemble #1



Ensemble index	# of conformers	Rg mean	Dmax mean	Show/Hide
1	130 conformers	20.81+-0.94 A	70.89+-4.49 A	Show List

Select a conformer from the list below to display it

Conformer	Rg	Dmax	Display
Conformation 1	Rg: 20.19	Dmax: 65.17	View conformer - Up
Conformation 2	Rg: 18.99	Dmax: 60.44	View conformer - Up
Conformation 3	Rg: 20.16	Dmax: 67.9	View conformer - Up

