JOHANNES GUTENBERG UNIVERSITÄT MAINZ

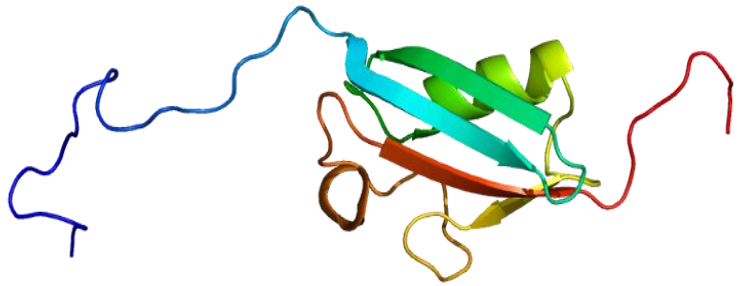
Protein disorder: Rethinking our image of proteins

Eric Schumbera

Miguel Andrade
CBDM Group
Biocenter I
e.schumbera@uni-mainz.de

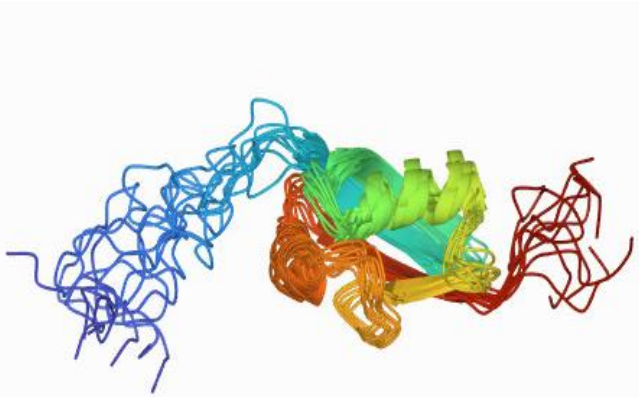
What is protein disorder?

Intrinsically disordered proteins (IDPs): proteins with regions that lack a single well-defined 3D structure in native conditions.

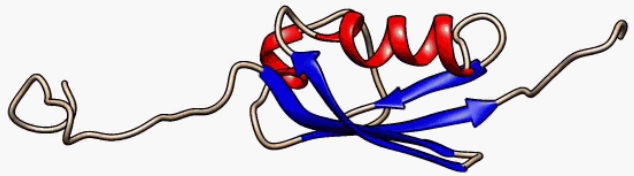


(PDB:1a5r)

experimentally determined NMR structure



10 individual NMR structures layered on top of each other



consensus structure

Intrinsically disordered regions (IDRs): regions within a protein that lacks a well-defined 3D structure (in native conditions).

The “discovery” of disorder

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

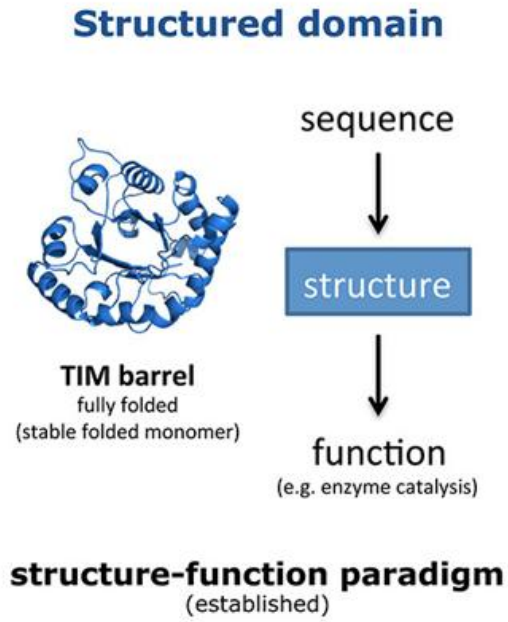
JMB

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

Peter E. Wright* and H. Jane Dyson*

Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA 92037, USA

A major challenge in the post-genome era will be determination of the functions of the encoded protein sequences. Since it is generally assumed that the function of a protein is closely linked to its three-dimensional structure, prediction or experimental determination of the library of protein structures is a matter of high priority. However, a large proportion of gene sequences appear to code not for folded, globular proteins, but for long stretches of amino acids that are likely to be either unfolded in solution or adopt non-globular structures of unknown conformation. Characterization of the conformational propensities and function of the non-globular protein sequences represents a major challenge. The high proportion of these sequences in the genomes of all organisms studied to date argues for important, as yet unknown functions, since there could be no other reason for their persistence throughout evolution. Clearly the assumption that a folded three-dimensional structure is necessary for function needs to be re-examined. Although the functions of many pro-

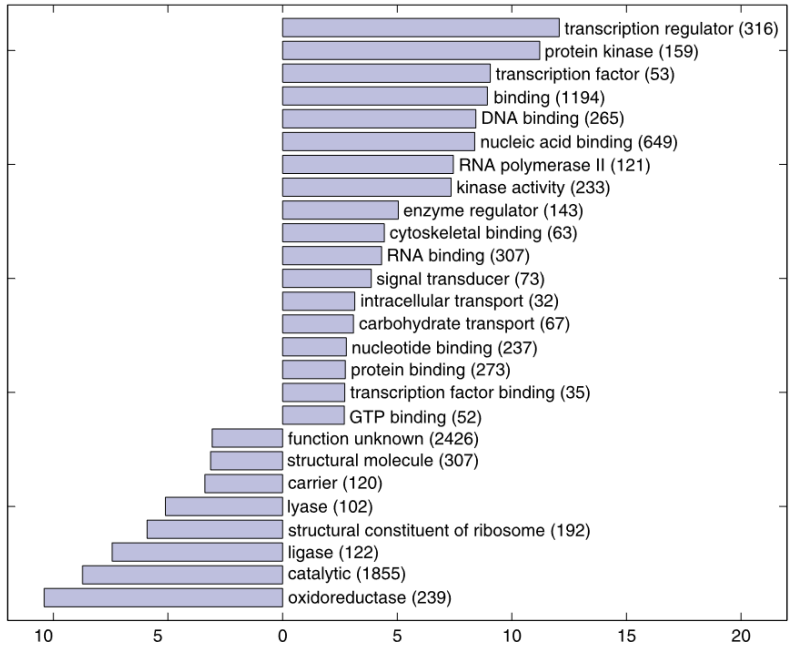


→ one of the first major works on IDR (in 1999!)
→ claims the old function paradigm needs top be re-examined
→ unfolded proteins have functions!

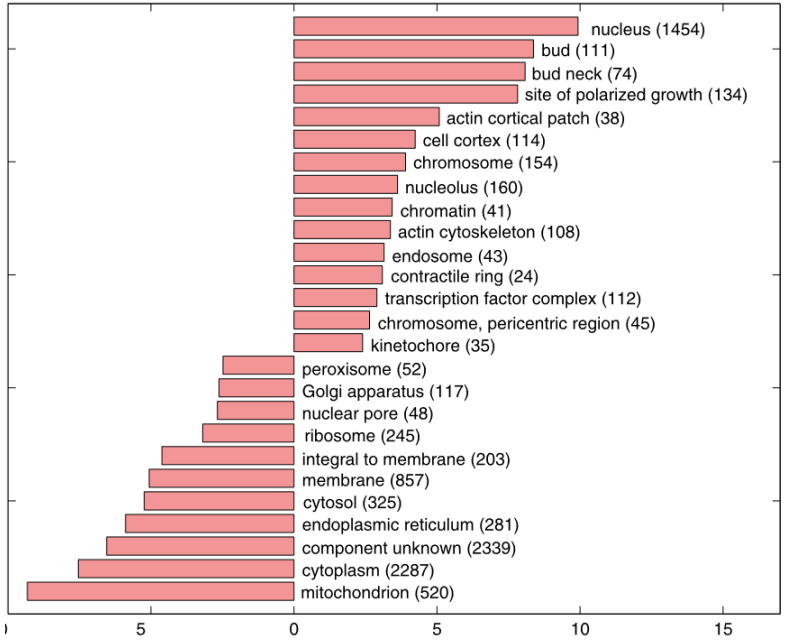
How often does disorder appear in biology?

"Putative, long (>30 residue) disordered segments are found to occur in **2.0% of archaean**, **4.2% of eubacterial** and **33.0% of eukaryotic proteins**."

GO Terms molecular functions



GO Terms - cellular component



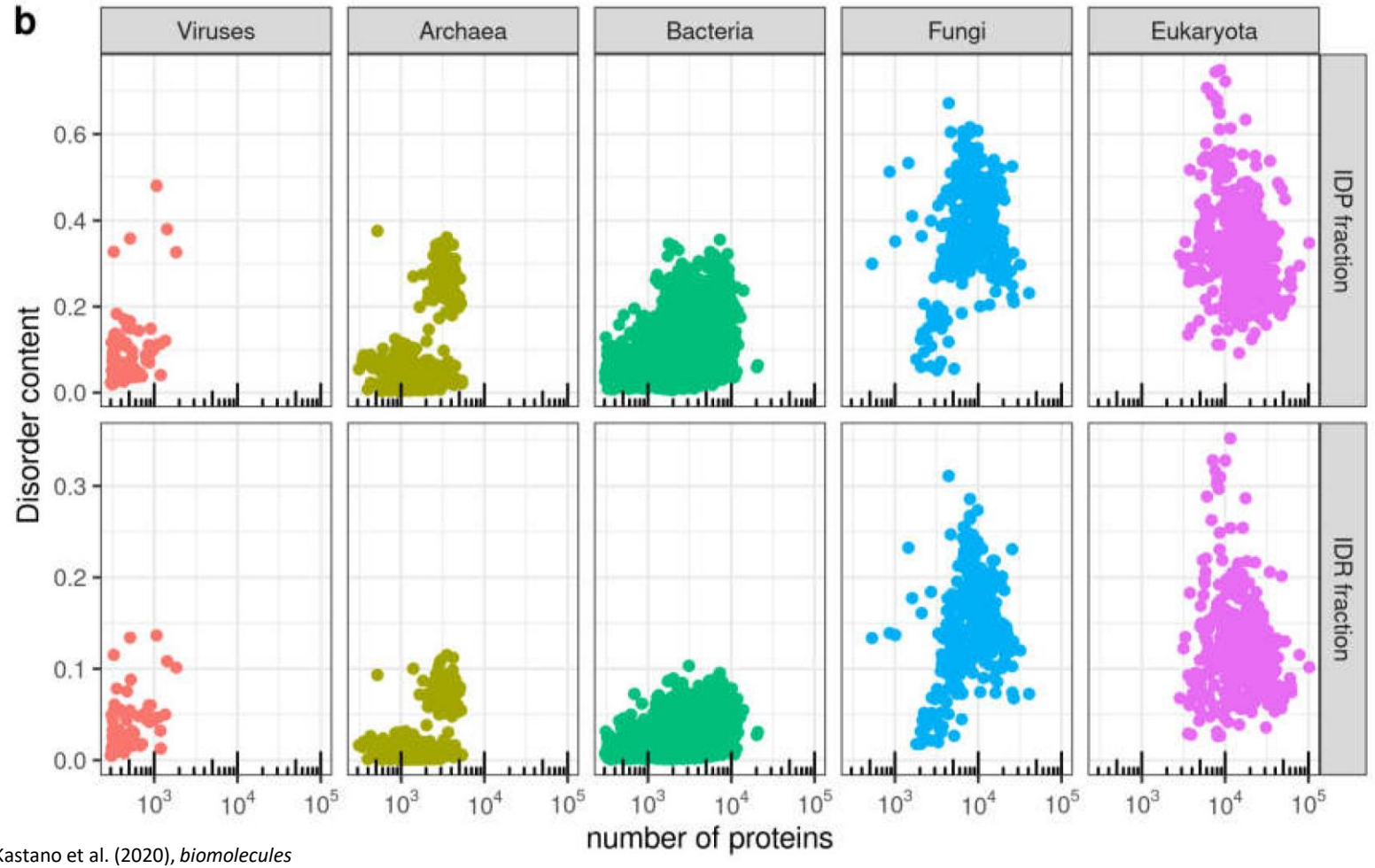
over- and underrepresented molecular functions and cellular locations in set of (predicted) disordered proteins

→ **functional meaning?**

Ward et al. (2004), *Journal of Molecular Biology*

IDR: intrinsically disordered region
IDP: intrinsically disordered protein

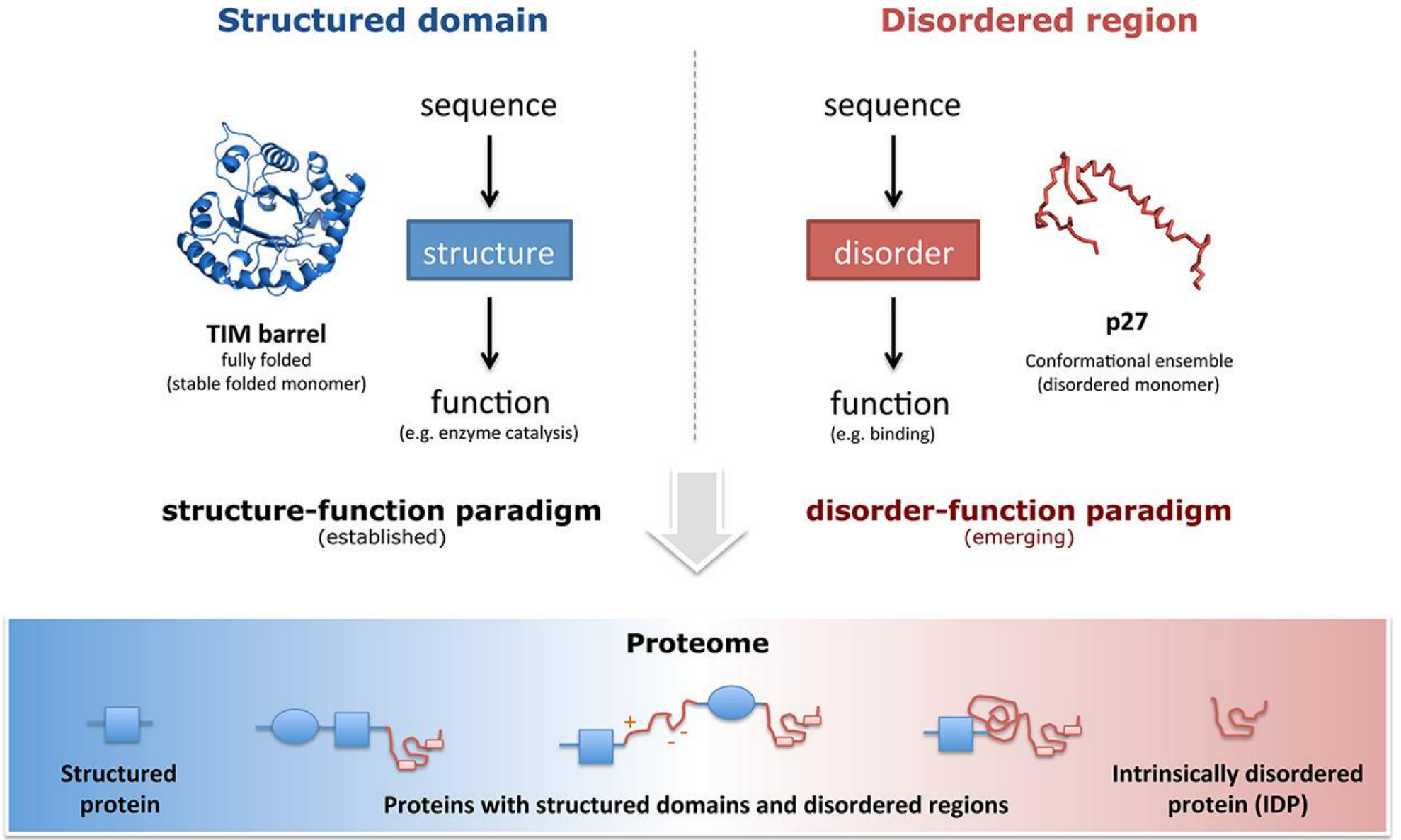
How often does disorder appear in biology?



Kastano et al. (2020), *biomolecules*

Relationship between disorder content and number of proteins in a proteome

→ IDRs and IDPs can be found in all taxonomic groups of life!
→ they must be of functional importance!

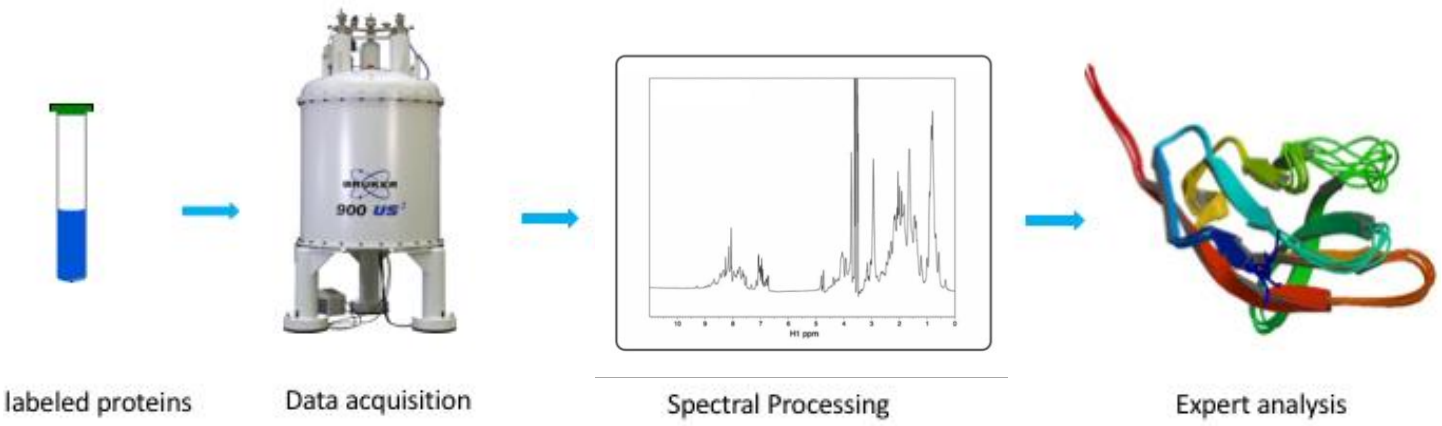


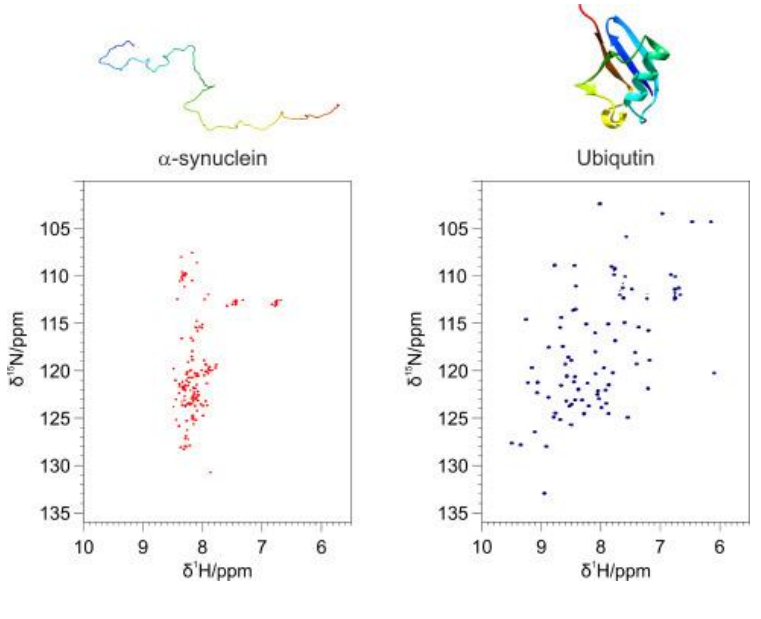
van der Lee (2014), *Chem. Rev.*

Experimental methods to detect IDRs:

1. NMR (Nuclear Magnetic Resonance)

most common quantitative technique used for studying IDPs

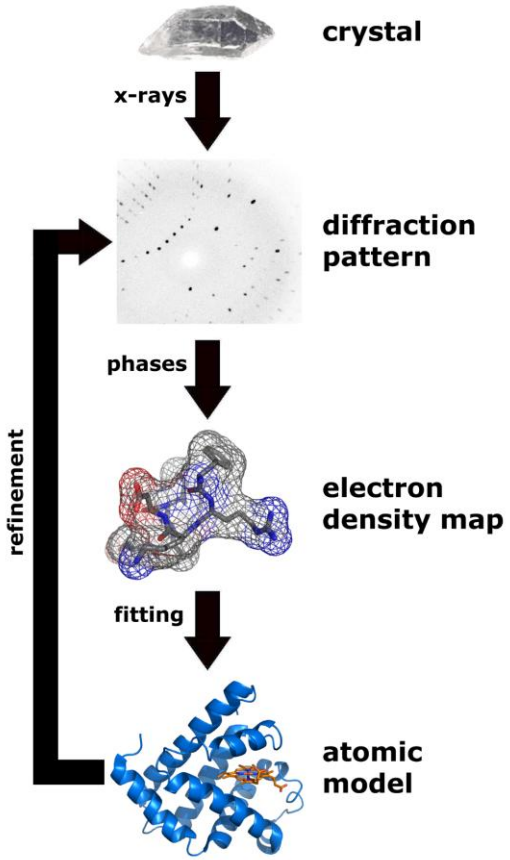


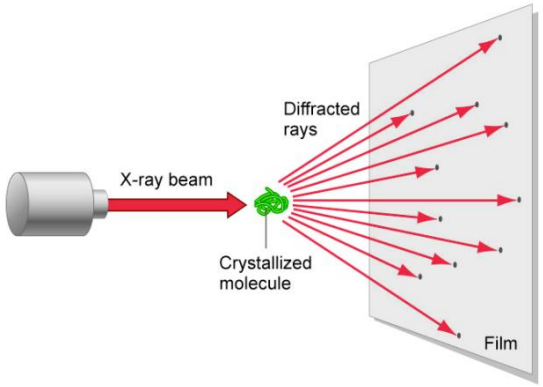


→ **high-resolution, residue-specific information** about their conformational ensembles and dynamics in solution
→ large NMR disorder databases (BMRB) → after „discovery“ of IDRs very helpful resource!

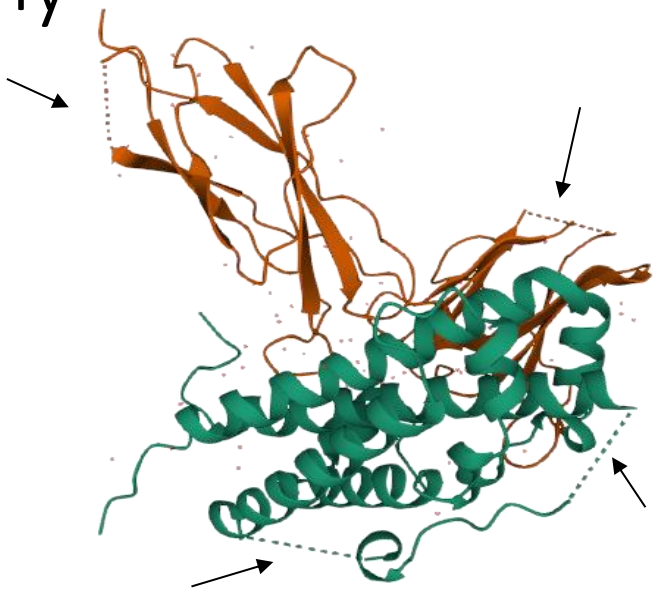
Experimental methods to detect IDRs:

2. X-ray crystallography





→ diffraction of rays can be calculated into an electron density map!
→ disordered regions have **no clear electron density map**



Missing electron densities in X-ray crystallography from PDB

- indirect method, but high resolution (1–2 Å)
- difficult crystallization necessary

Experimental methods to detect IDRs:

3. Other methods

- **Circular Dichroism (CD):**
  - α-helix, β-sheet, turn, PPII helix, and coil conformations are determined via far-UV spectroscopy (190–230 nm)
    → distinct peaks!
  - disordered proteins/regions lack these peaks! → indirect discovery
- **Small-Angle X-ray Scattering (SAXS):**
  - can measure molecular mass, volume, radius of gyration, folding state (even disorder-to-order transitions!)
  - but low resolution (10-30 Å)

  → highly complementary to NMR and X-ray crystallography
- **Cryo-Electron Microscopy (Cryo-EM):**
  - allow the structural characterization of proteins near-native state with high resolution (< 4 Å)
  - no crystallization!

  → highly complementary with NMR

Sequence Characteristics of Intrinsically Disordered Proteins

order-promoting residues

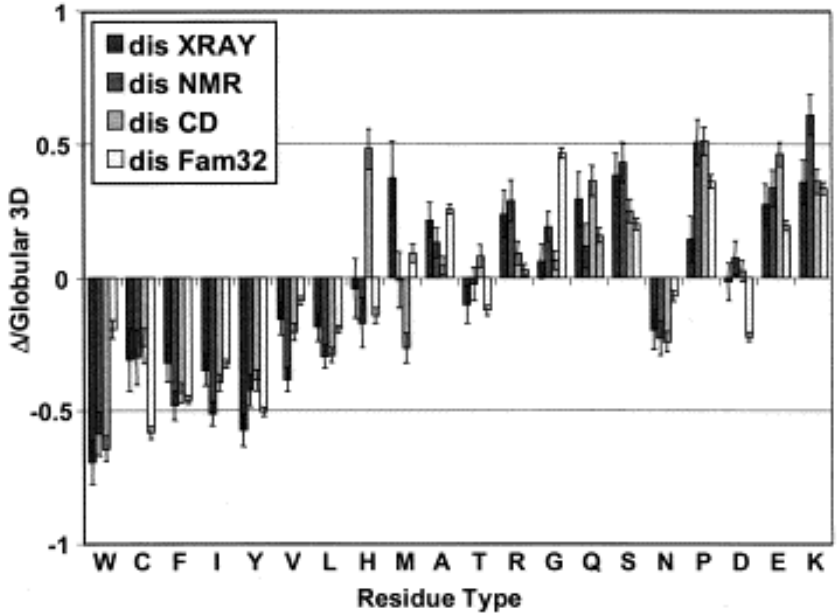
hydrophobic

Asparagine (N)
Cysteine (C)
Isoleucine (I)
Leucine (L)
Valine (V)
Phenylalanine (F)
Tryptophan (W)
Tyrosine (Y)

aromatic

→ more common in the ordered/structured segments of the proteins

amino-acid compositions of disordered datasets



Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

disorder-promoting residues

Arginine (R)
Lysine (L)
Glutamic Acid (E)
Alanine (A)
Glycine (G)
Glutamine (Q)
Proline (P)
Serine (S)

pos. charged

neg. charged

hydrophobic

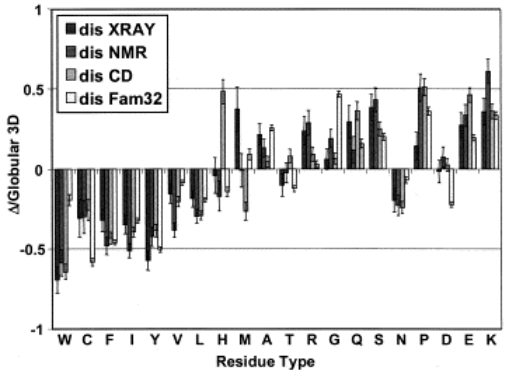
→ more common in IDPs/IDRs

Sequence Characteristics of Intrinsically Disordered Proteins

order-promoting residues

Asparagine (N)
Cysteine (C)
Isoleucine (I)
Leucine (L)
Valine (V)
Phenylalanine (F)
Tryptophan (W)
Tyrosine (Y)

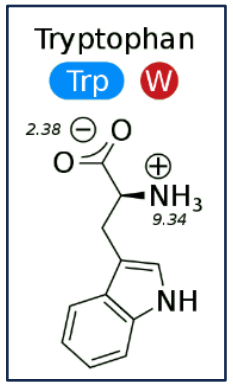
amino-acid compositions of disordered dataset

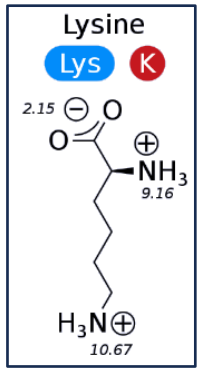


Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

disorder-promoting residues

Arginine (R)
Lysine (K)
Glutamic Acid (E)
Alanine (A)
Glycine (G)
Glutamine (Q)
Proline (P)
Serine (S)





W C F I Y V L H M A T R G Q S N P D E K
Less Flexible ⟷ More Flexible

Order Promoting
Disorder Promoting
Inconsistent

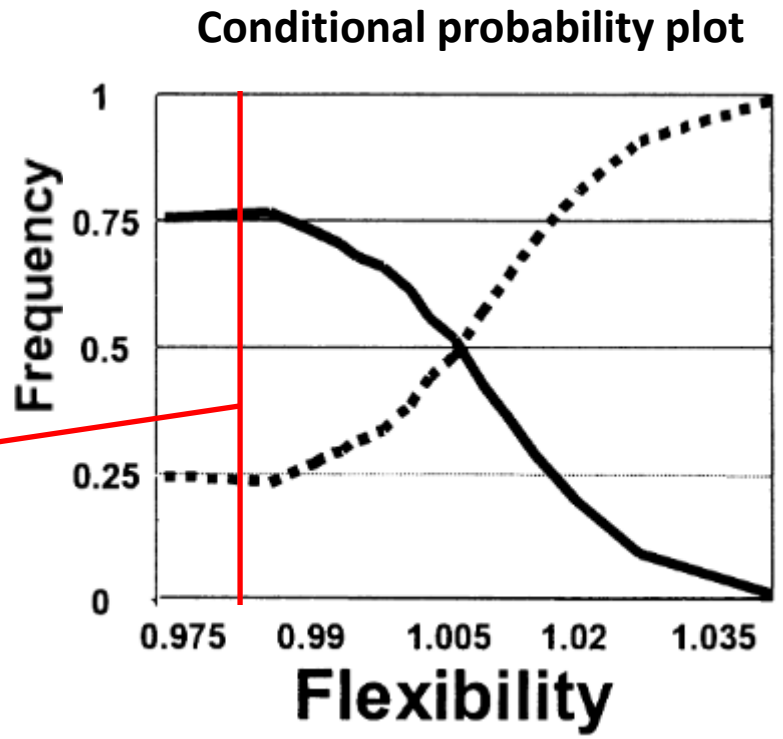
DeForte et al. (2016), *MDPI molecules*

Sequence Characteristics of Intrinsically Disordered Proteins

ratio of structured proteins

ratio of disordered proteins

75% of proteins are structured
25% of proteins are disordered



Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

→ indicate the ability of a given attribute **to discriminate between order** (solid line) and **disorder** (dashed line)

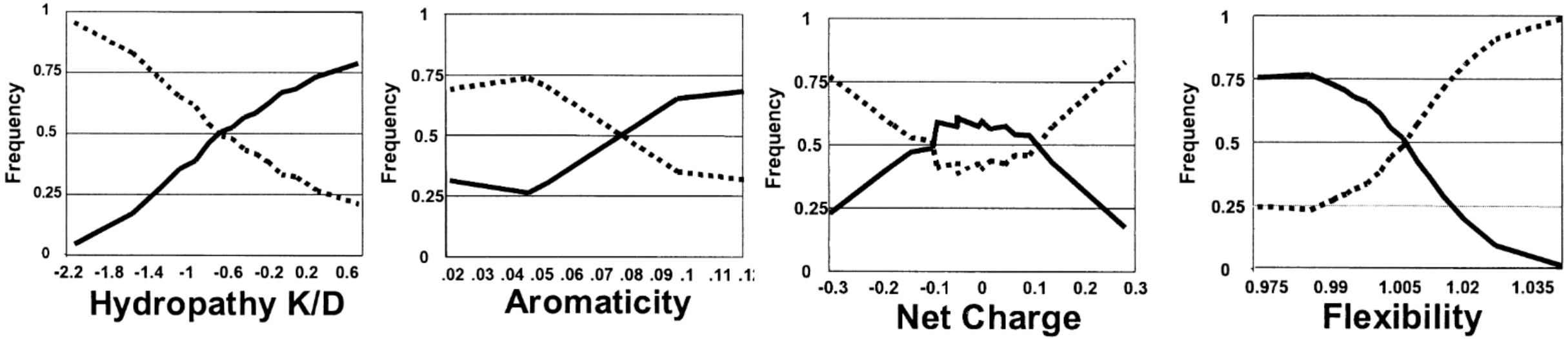
Sequence Characteristics of Intrinsically Disordered Proteins

ratio of structured proteins

ratio of disordered proteins

Conditional probability plots

→ indicate the ability of a given attribute **to discriminate between order** (solid line) and **disorder** (dashed line)



Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

Disorder prediction tools

SEQUENCE-BASED

FoldIndex

→ based on: charge/hydropathy of sequence

Prilsuky et al. (2005), *Bioinformatics*

IUPred

→ based on: pairwise interaction energies (physical principle)

Walsh et al. (2012), *Bioinformatics*

MACHINE LEARNING

ESpritz

→ based on: neural networks (BRNN) trained on curated exp. data (NMR, Xray,…)

Walsh et al. (2012), *Bioinformatics*

DISOPRED2

→ based on: support vector machines (SVM) trained on X-ray datasets

Ward et al. (2004), *Bioinformatics*

CONSENSUS-BASED

MobiDB-lite

→ based on: consensus of 8 different disorder predictors

Necci et al. (2020), *Bioinformatics*

"INDIRECT" METHOD

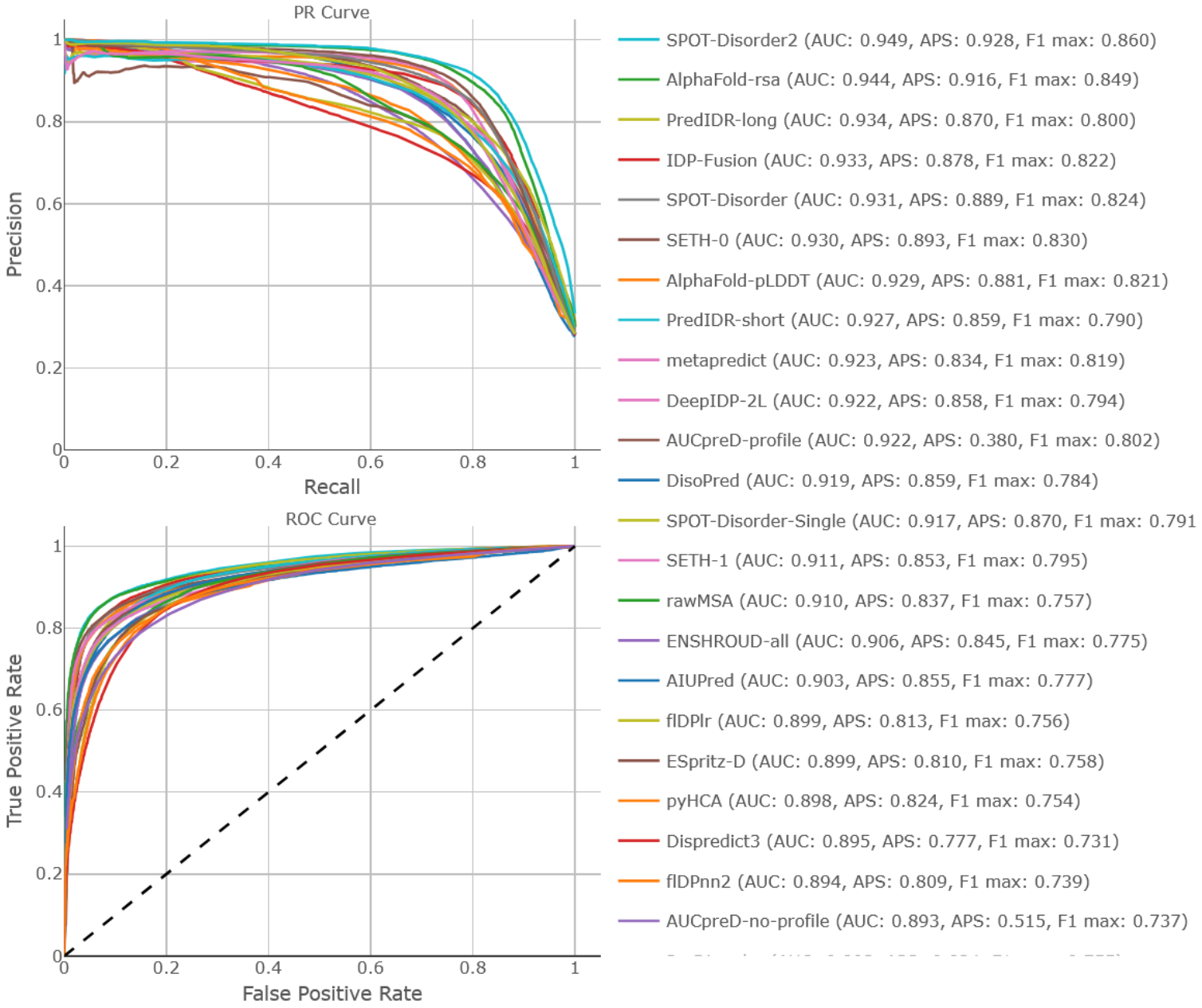
Alphafold

→ based on: indirect method by low confidence scoring

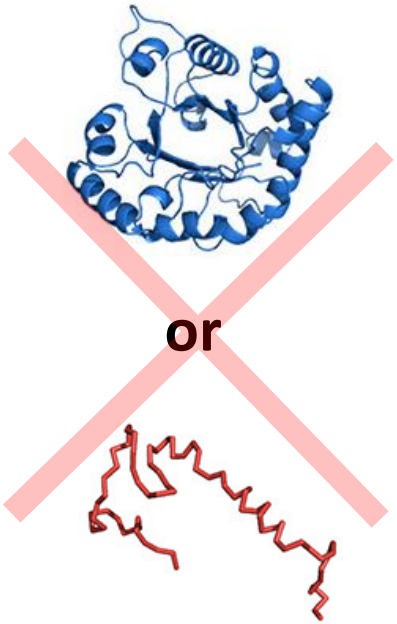
Jumper et al. (2021), *Nature*

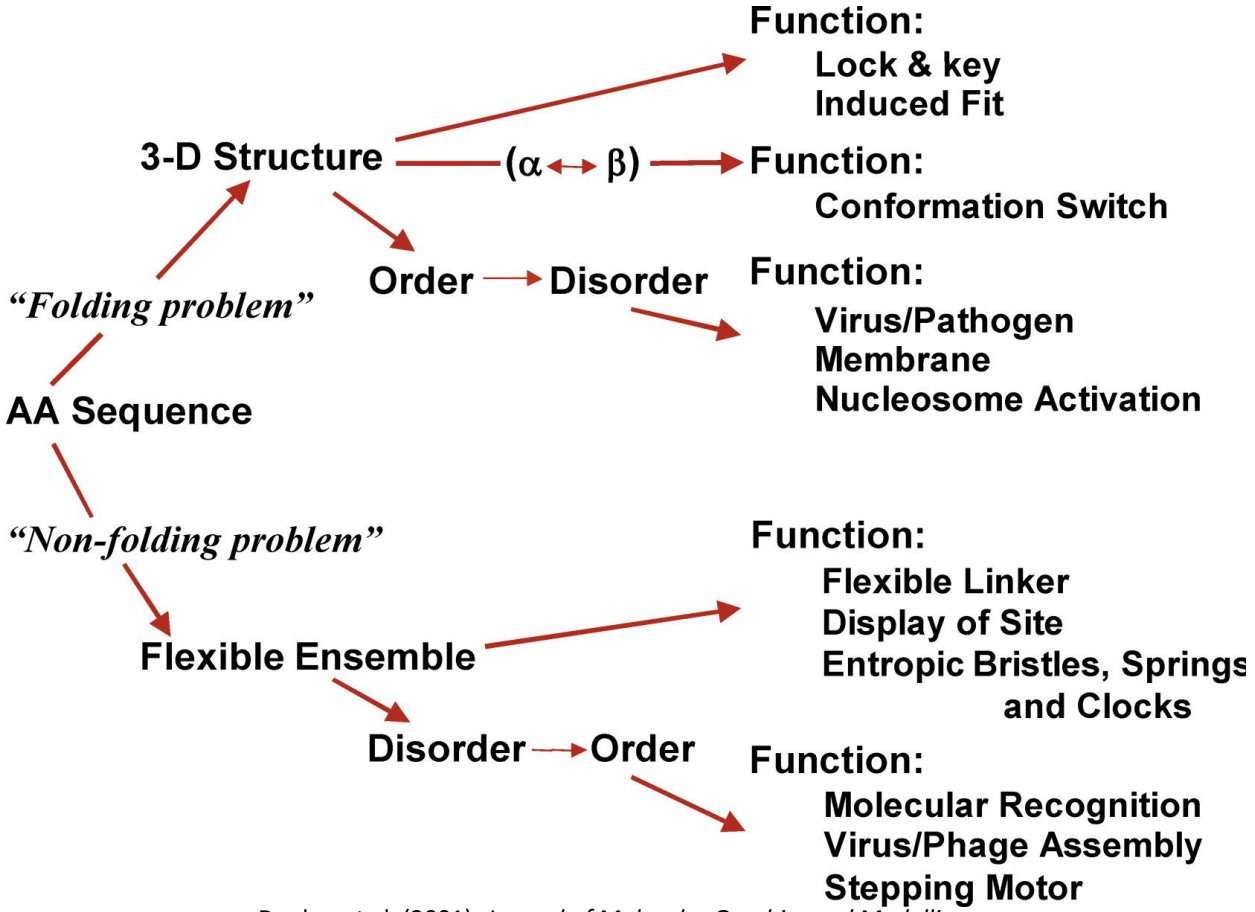
CAID 2

- CAID *(Critical Assessment of Intrinsic Disorder Prediction)* is the equivalent of CASP *(Critical Assessment of Structure Prediction)*



Proteins in Motion: Bridging Order and Disorder

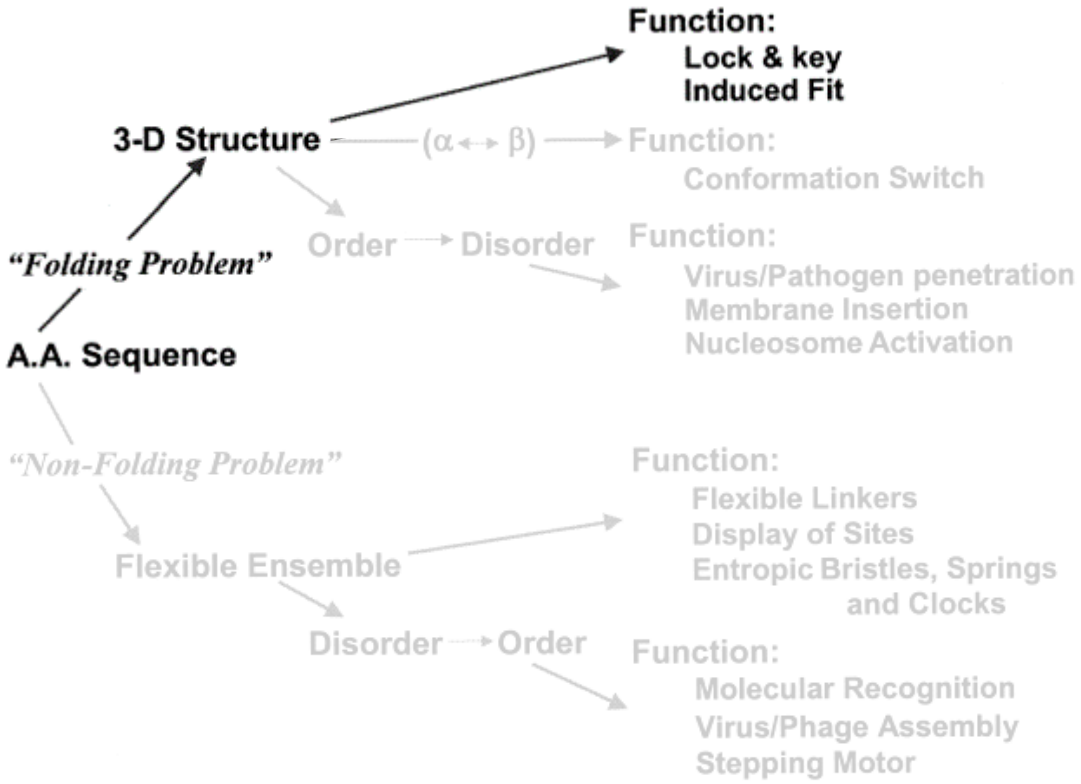




Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

Structure and disorder are not exclusive!

Proteins in Motion: Bridging Order and Disorder

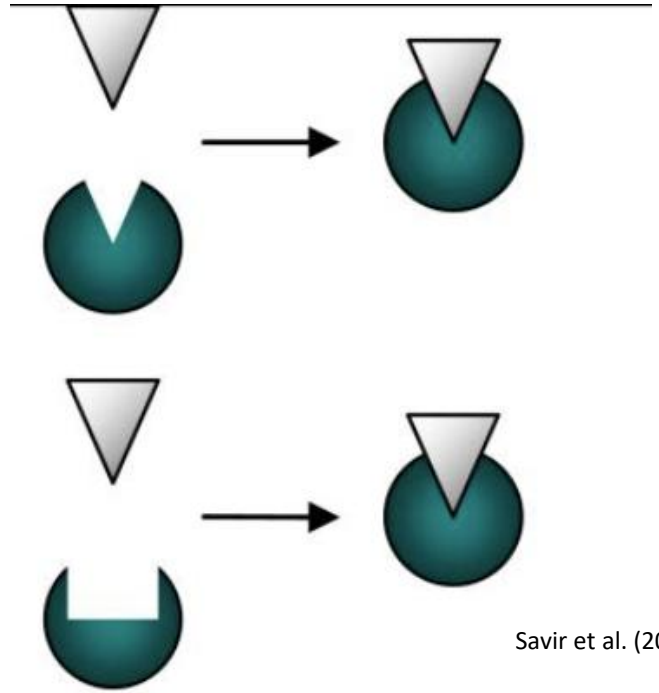


Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

1. Function through rigid shape, no (considerable) transition

„Lock & key model“

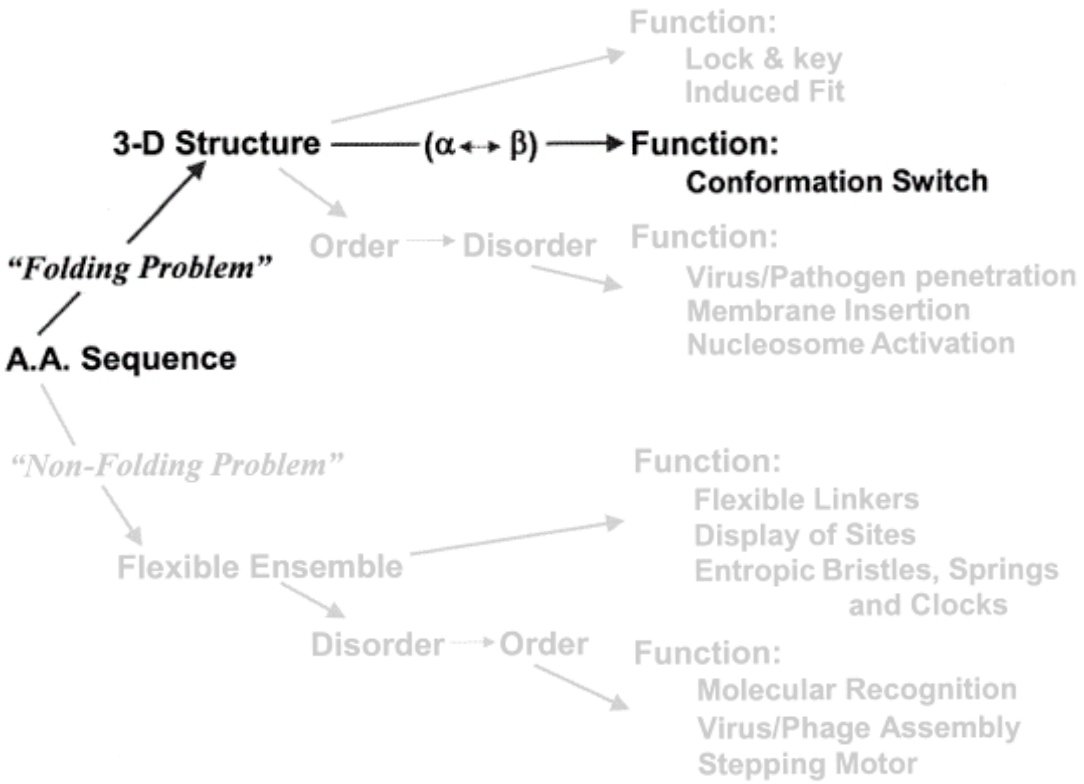
„Induced Fit“



Savir et al. (2007), *PLOS ONE*

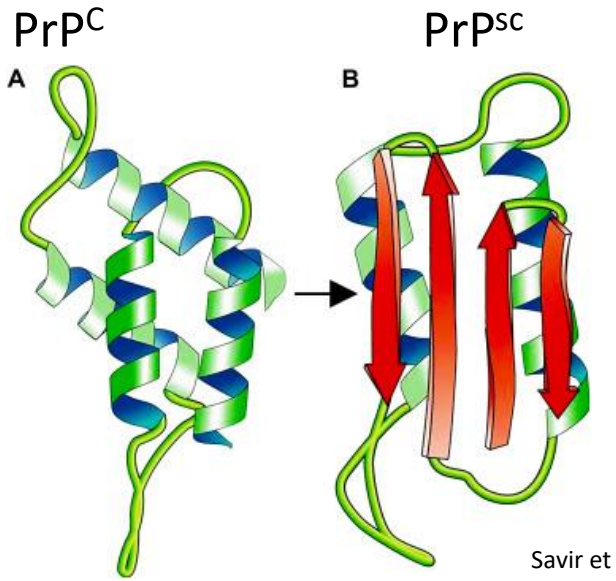
e.g. enzymatic reactions like lactate dehydrogenase and pyruvate

Proteins in Motion: Bridging Order and Disorder



Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

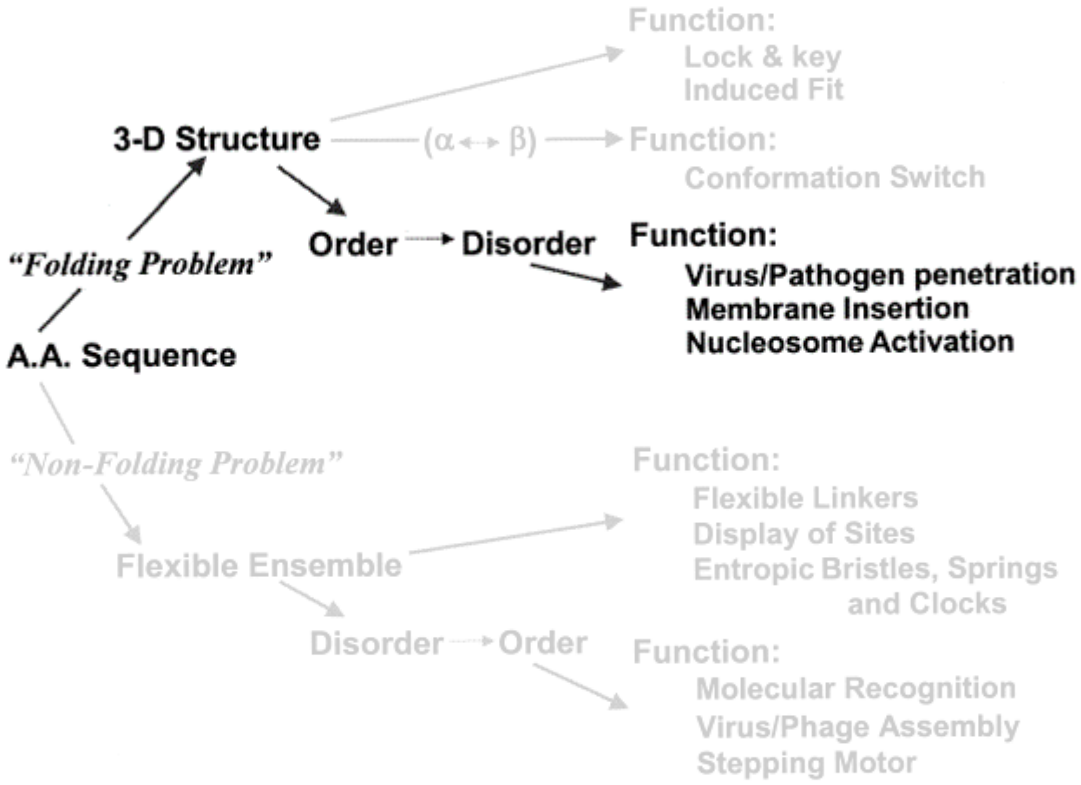
2. function changes through conformation switch (alpha helix <-> beta sheet)



Savir et al. (2007), *PLOS ONE*

➔ loss of function → pathogenic

Proteins in Motion: Bridging Order and Disorder



Dunker et al. (2001), *Journal of Molecular Graphics and Modelling*

3. Order-to-Disorder transition

➔ the disordered state is responsible for function
➔ rare but has been observed!

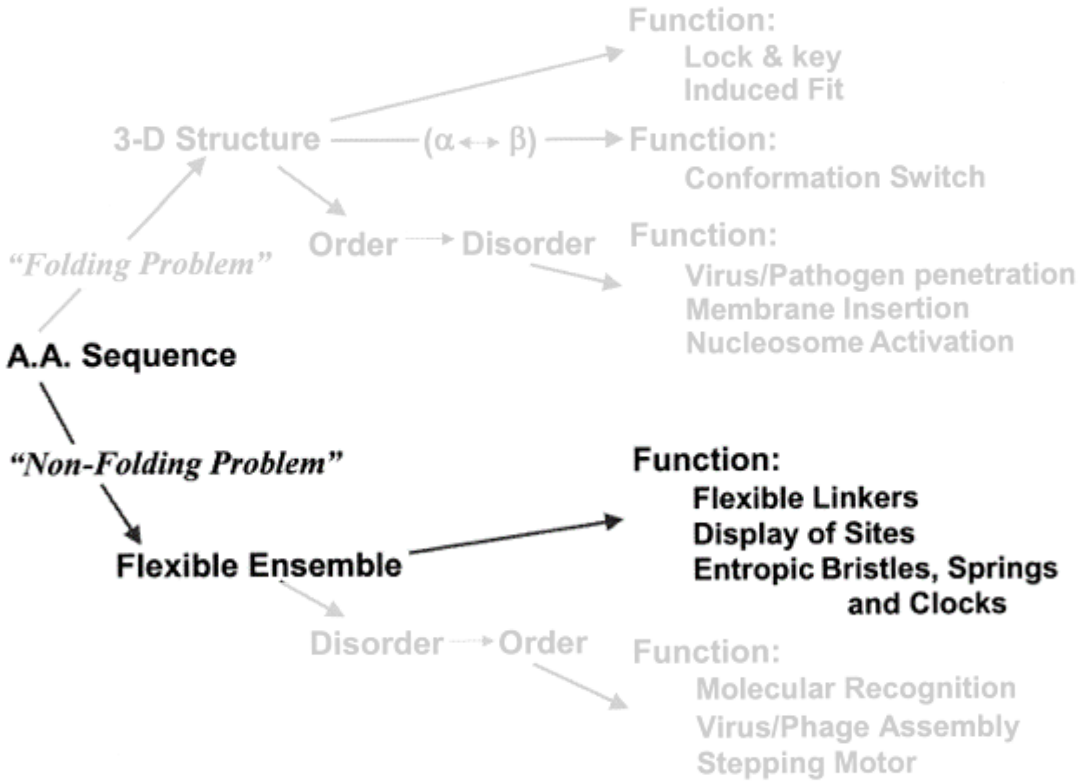
e.g. hyperacetylated nucleosomes

"[…] hyperacetylation makes nucleosome core particles **less rigid** […]"
→ higher levels of gene trancription

e.g. membrane penetration of fd phage

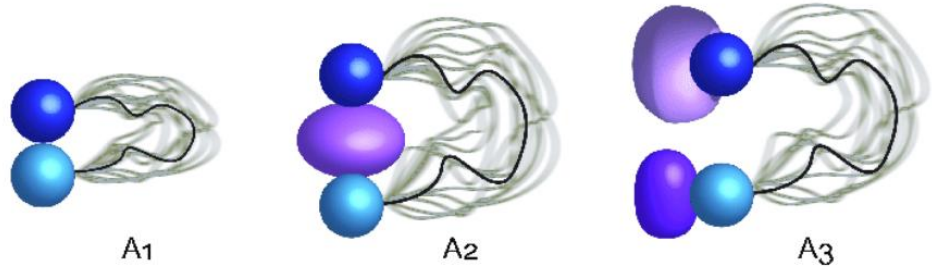
"In a process that likely mimics infection**, fd phage** converts from the **ordered into the disordered** molten globular state."
→ enables/facilitates entry into host

Proteins in Motion: Bridging Order and Disorder

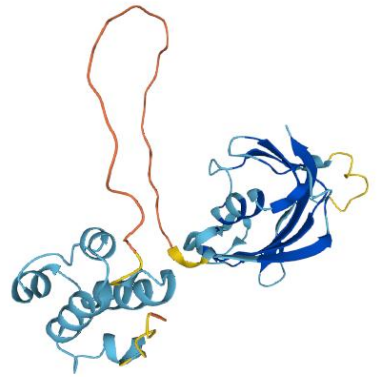


https://www.sciencedirect.com/science/article/pii/S1093326300001388#FIG9

4. Function through disordered shape, no (considerable) transition

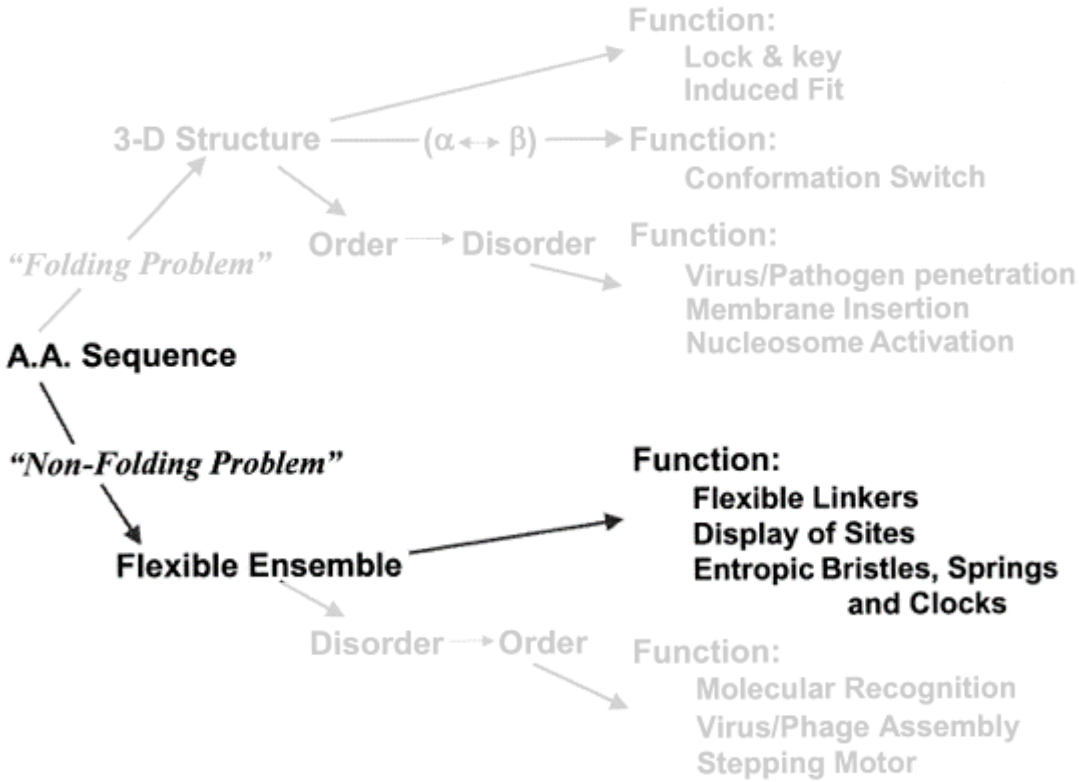


Clerc et al. (2021), *Computational and Structural Biotechnology Journal*

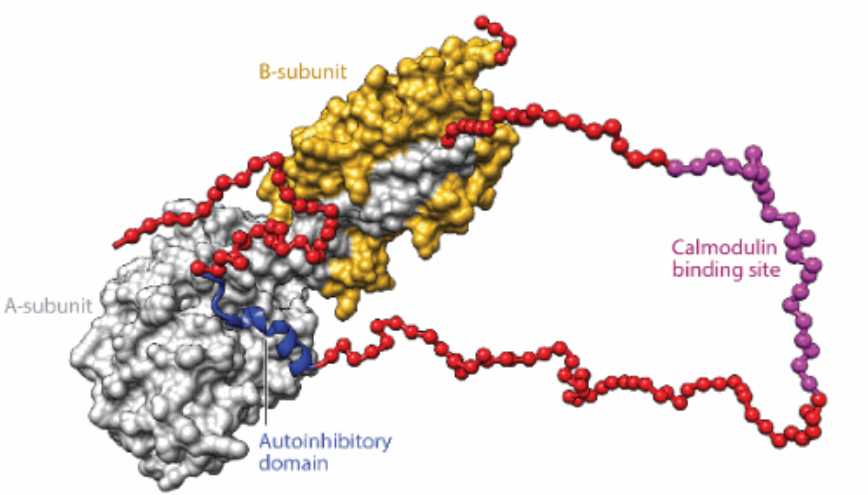


AlphaFold structure prediction of Peptidyl-prolyl cis-trans isomerase **FKBP3**

Proteins in Motion: Bridging Order and Disorder



4. Function through disordered shape, no (considerable) transition



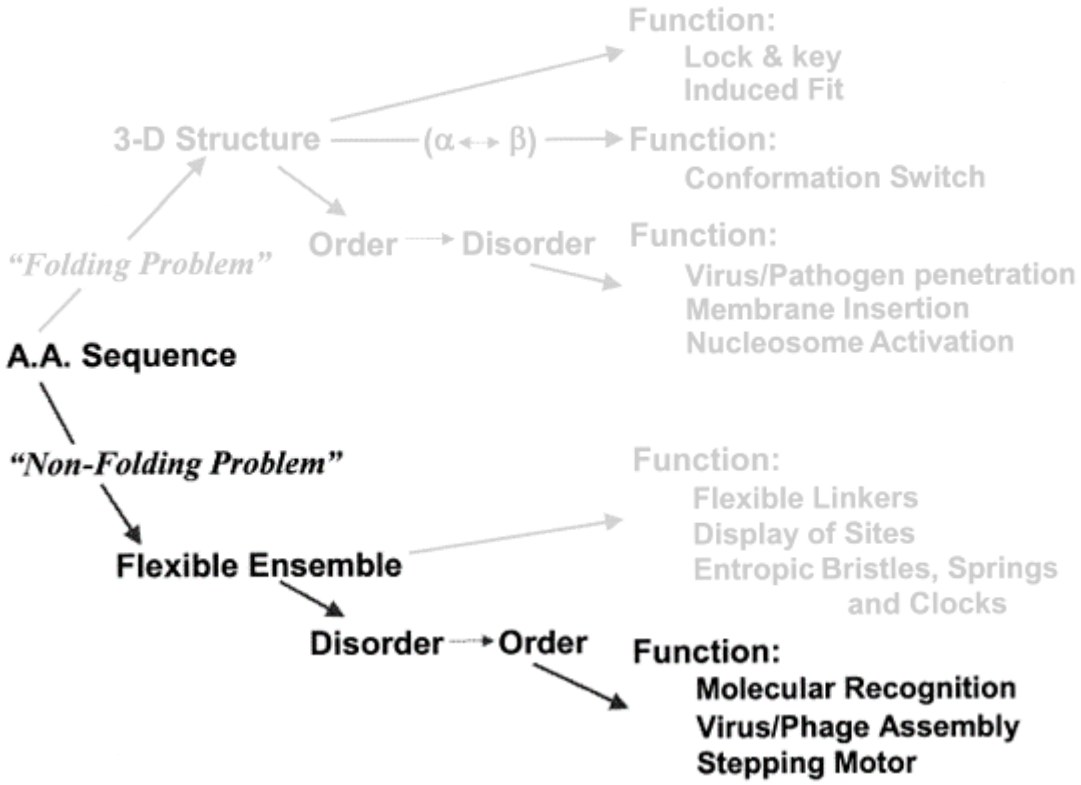
Oldfield CJ, Dunker AK. 2014. Annu. Rev. Biochem. 83:553–84

remains disordered even after binding!

→ disorder is essential to function (flexibility)

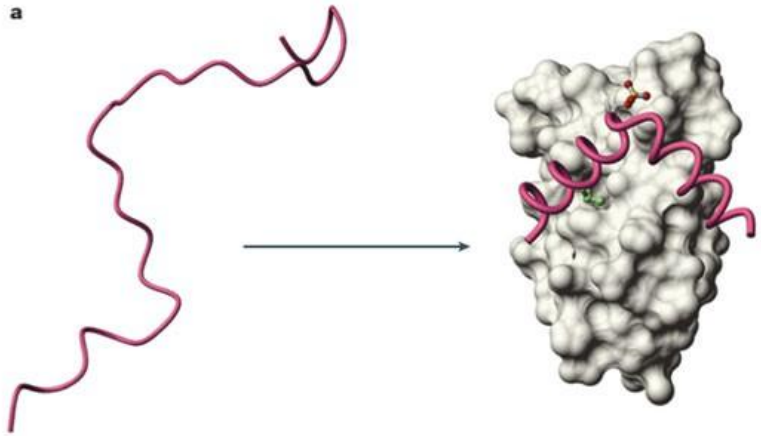
https://www.sciencedirect.com/science/article/pii/S1093326300001388#FIG9

Proteins in Motion: Bridging Order and Disorder



5. Function through disorder-to-order transition

→ in most cases of disorder, proteins undergo a disorder-to-order conformational transition (termed as 'induced folding')

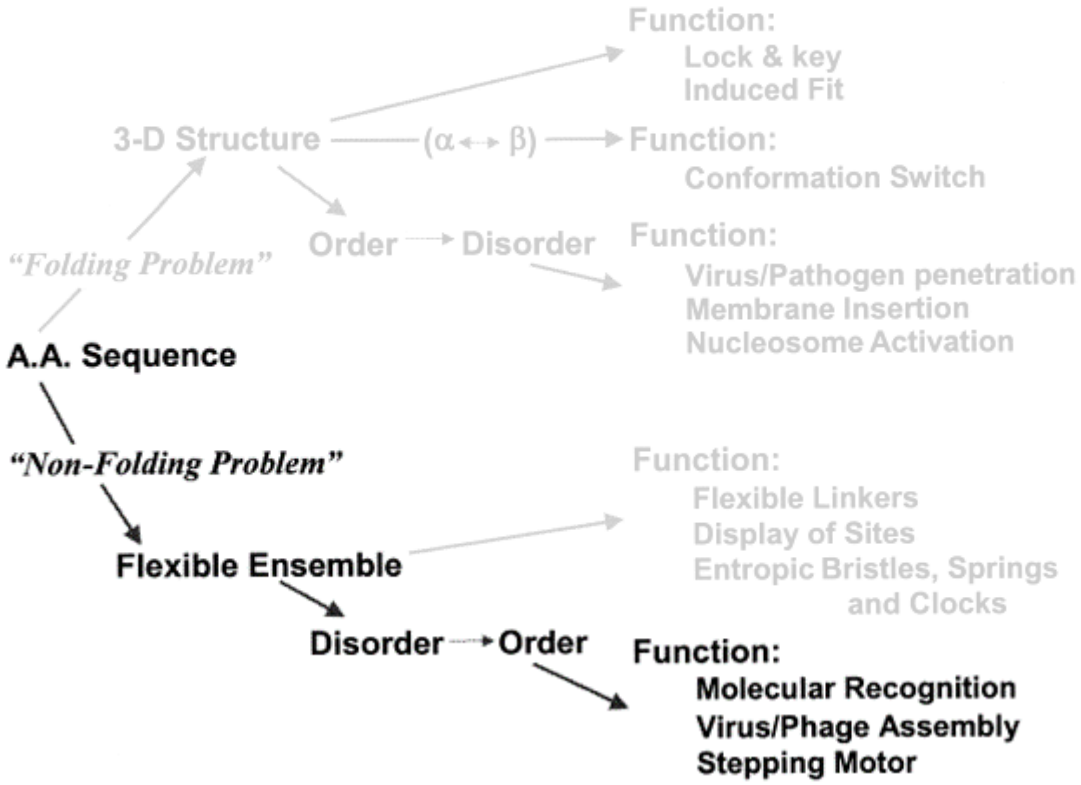


disordered pKID domain

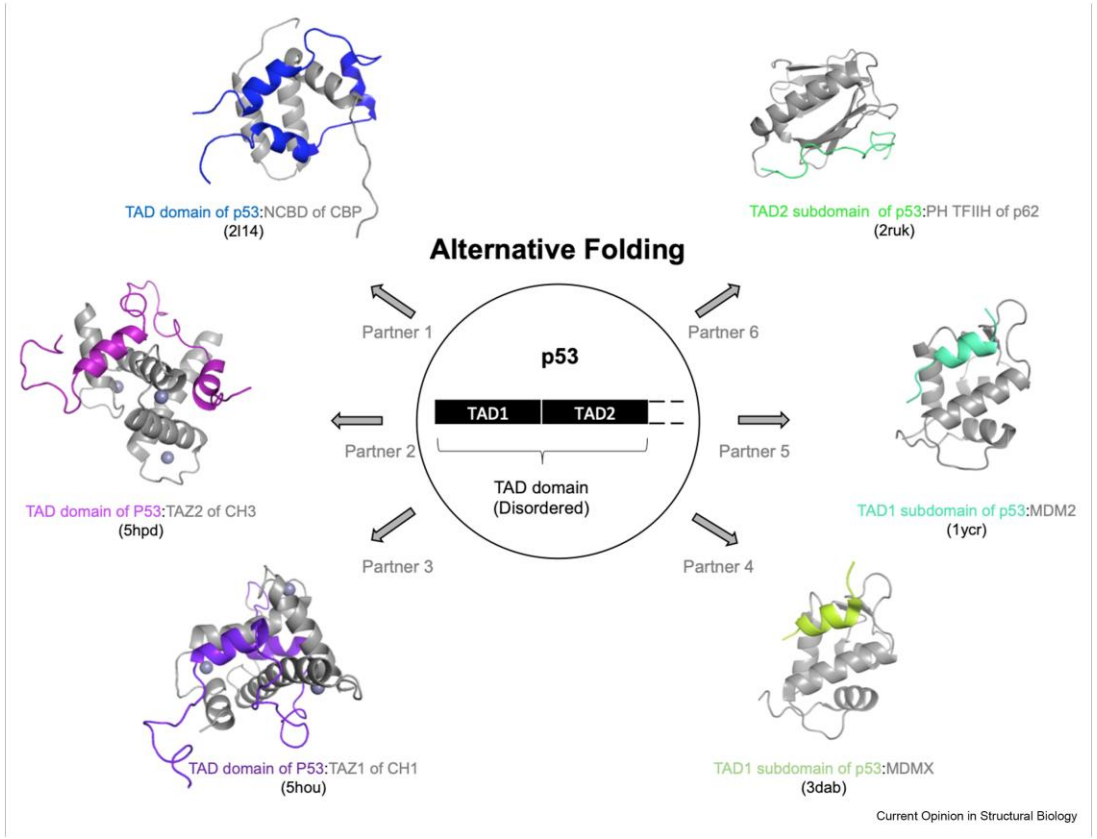
pKID domain interacting with KIX domain

https://www.sciencedirect.com/science/article/pii/S1093326300001388#FIG9

Proteins in Motion: Bridging Order and Disorder

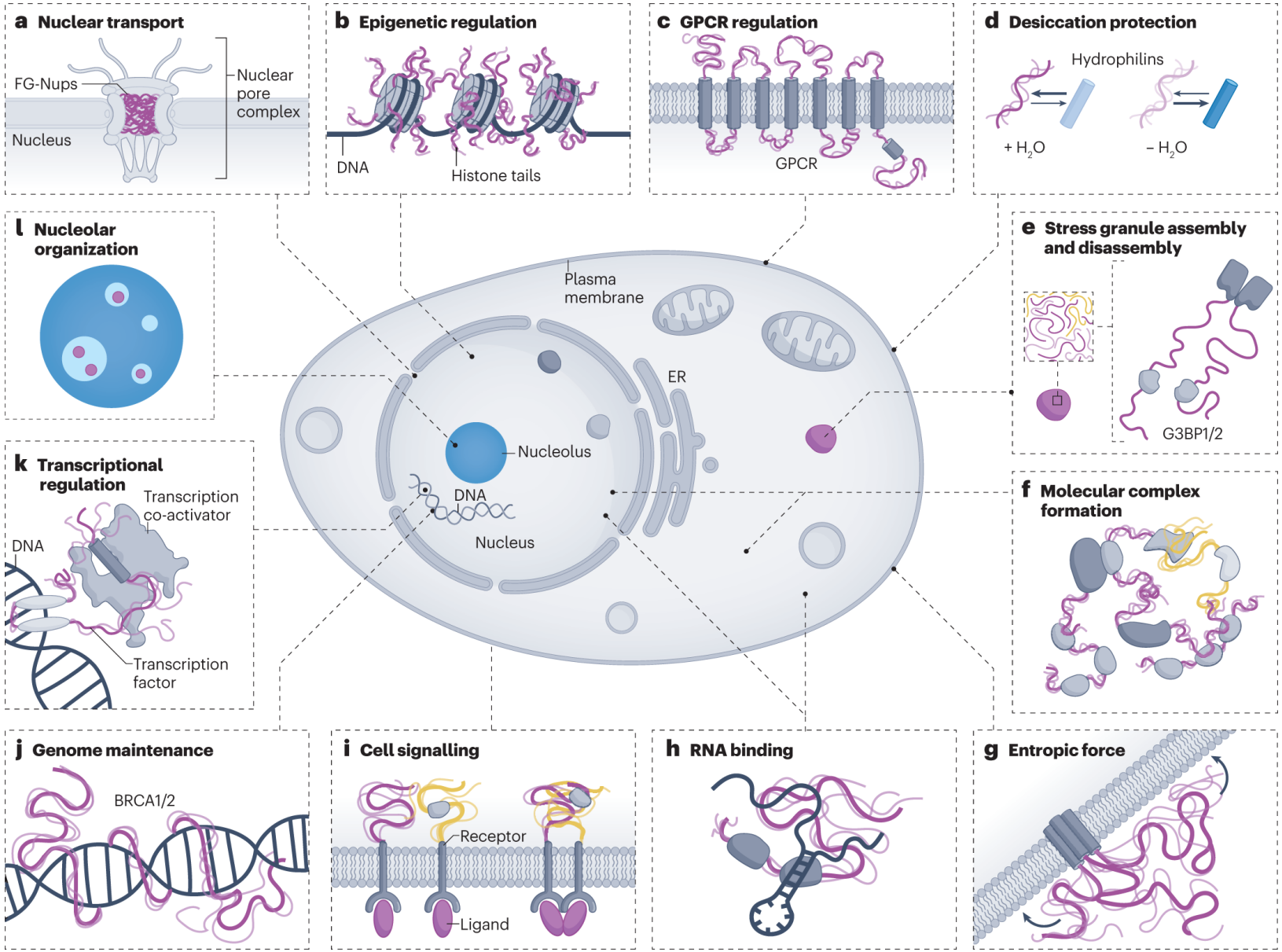


5. Function through disorder-to-order transition



https://www.sciencedirect.com/science/article/pii/S1093326300001388#FIG9

Intrinsically disordered regions are central to cellular function



Holehouse et al. (2023),*nature reviews molecular cell biology*

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

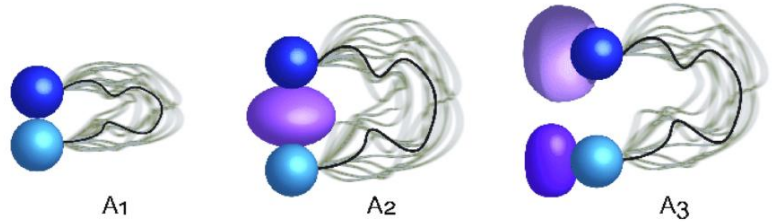
Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

1.Entropic chains

➔ conformational flexibility of IDR **regulate movement** of domains and **regulate inter-domain distances**
➔ **linkers and spacers!**



Clerc et al. (2021), *Computational and Structural Biotechnology Journal*

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

Molecular Scavengers
store and/or neutralize small molecules

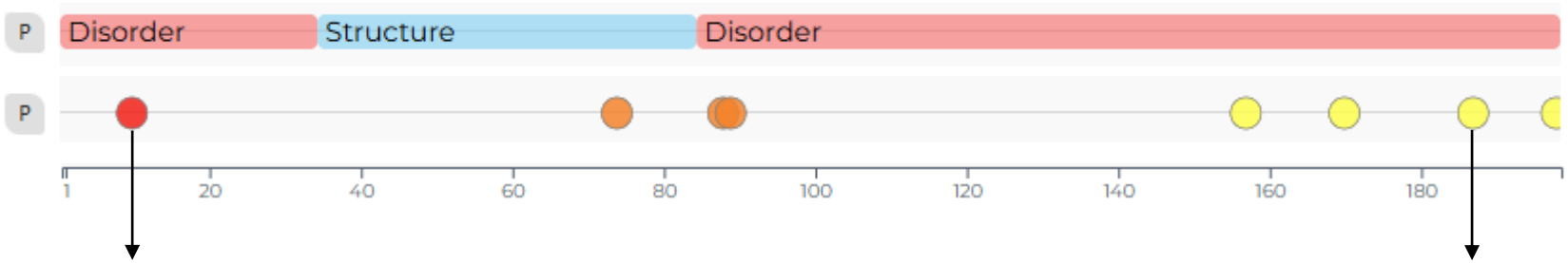
Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

2.Modification sites

➔ Most of the regulatory and signaling proteins possess IDRs
- ➔ **Flexibility of IDRs affords them an advantage over structured domains!**
- ➔ sites within their polypeptide chains are highly accessible!

e.g.: P46527 - Cyclin-dependent kinase inhibitor 1B (p27Kip1)



phosphorylation of Ser10:
promotes its interaction with the shuttling protein, CRM1, leading to export from the nucleus
Ishida et al. (2002), *J. Biol. Chem.*

phosphorylation of Thr187:
promotes recognition by ubiquitination machinery and degradation
Tsvetkov et al. (1999), *Curr. Biol*

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

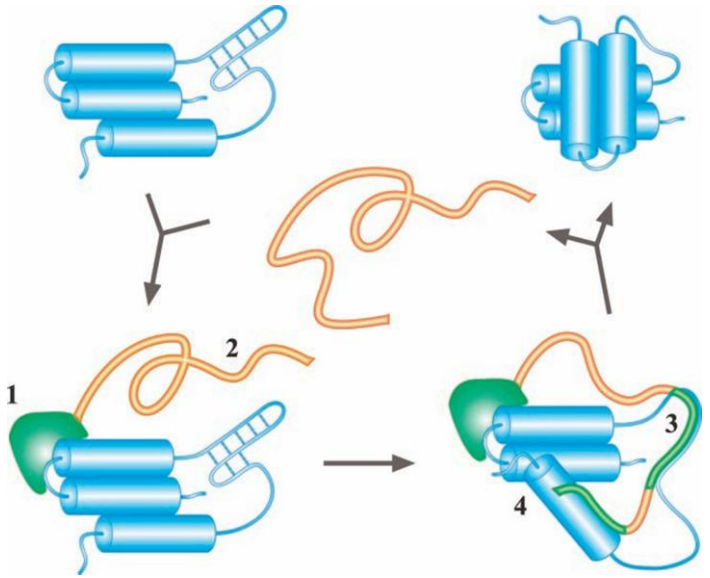
3.Disordered chaperones

chaperones: *proteins that assist RNAs and other proteins in their conformational folding or unfolding*

➔ **50% of RNA chaperone** sequences and **33% of the protein chaperones** are disordered in nature

➔ advantageous due to structure "adaptation" and increasing lifetime of the complex

partially misfolded protein



correctly folded protein

Tompa et al. (2004) *FASEB J.*

<u>some well-studied examples:</u>
hnRNP A1, GroEL, α-crystallin, Hsp33

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

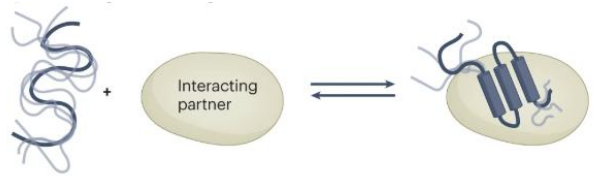
Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

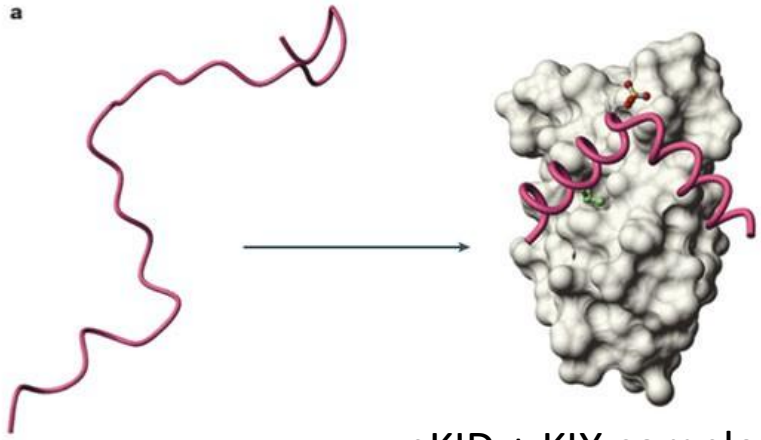
Functional classes of IDRs

4.Molecular effectors

➔ Modifying other proteins through permanent binding
➔ often in combination with ‘disorder to order’ transition (coupled folded and bonding)



Drerer et al. (2022), *Cellular and Molecular Life Science*



pKID domain (part of CREB)

pKID + KIX complex (part of CBP)

➔ allosteric effect on CBP
➔ affects the binding affinity of CBP to multiple transcriptional regulators

➔ **notice the phosphorylation!!!**

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

Molecular Scavengers
store and/or neutralize small molecules

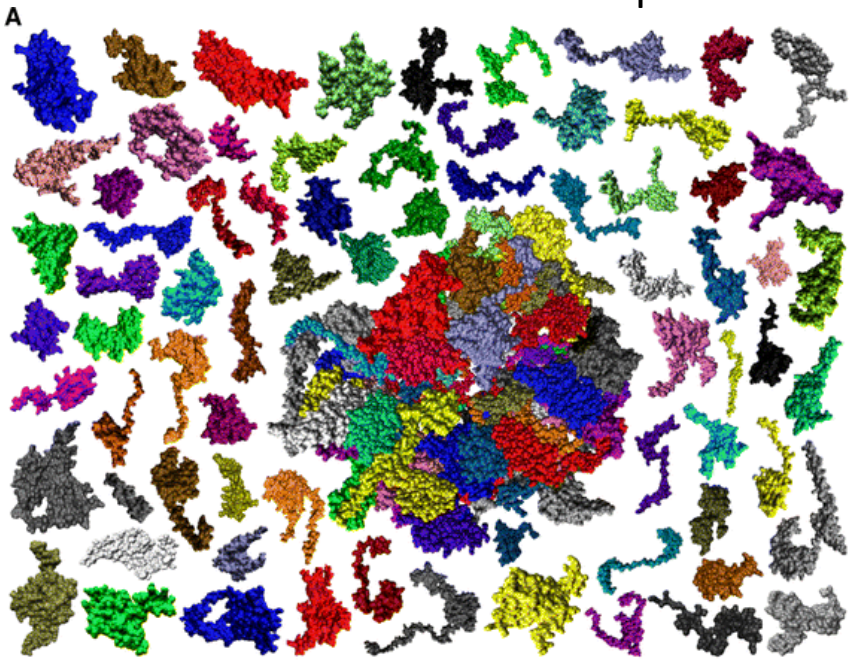
Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

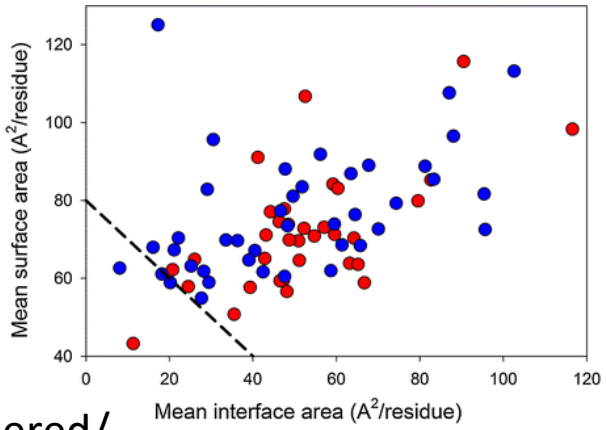
5.Molecular assemblers

multivalency allows an "assembler" function
→ multiple proteins and/or RNA forming higher-order complexes with IDRs

ribosome complex



"more disordered"



"more ordered/ structured"

Peng et al. (2013), *Cellular and Molecular Life Sciences*

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

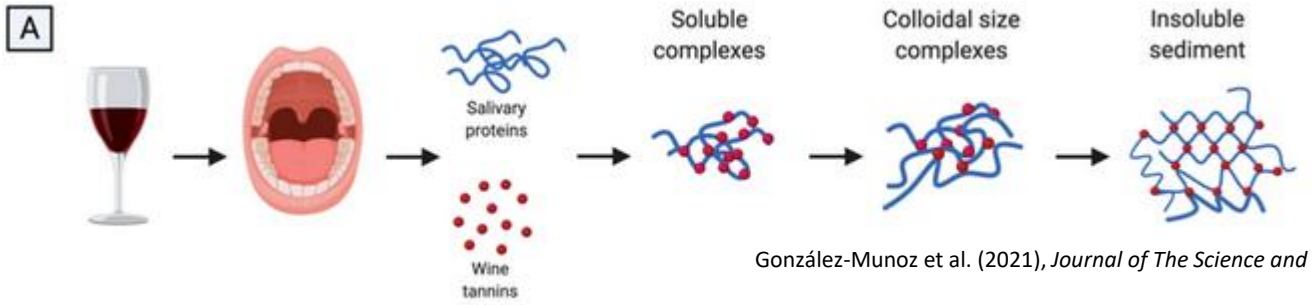
Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

6. Molecular scavengers

➔ stores and neutralizes small ligands



González-Munoz et al. (2021), *Journal of The Science and Food Agriculture*

➔ salivary proteins "catch" the tannins and therefore create certain effects on our taste!

<u>other examples:</u>

→ the VviDHN4 isoform of dehydrin acts as a scavenger by **removing reactive oxygen species** from the cellular environment
(Vazquez-Hernandez et al. (2021) *Plant Physiol. Biochem.*)

→ SmbP protein of *N. Europaea*, **binds to** divalent cations, especially **copper** → prevents cellular toxicity
(Barney et al. (2004), *Biochemistry*)

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

Molecular Effectors
modulating partner molecules through permanent binding

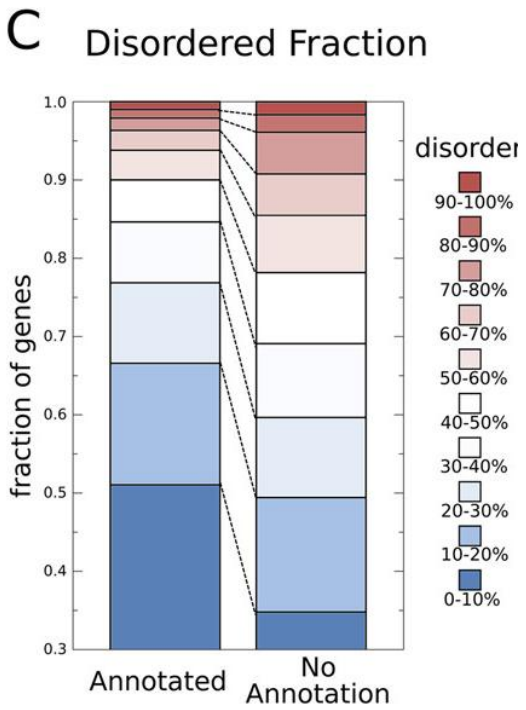
Molecular Assemblers
assemble complexes through permanent binding

Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs

7.Unknown



van der Lee (2014), *Chem. Rev.*

→ large unknown fraction of IDRs/IDPs where functional annotation is still necessary!

Entropic chains
no binding, but flexibility added to protein

Modification sites
sites for post-translational modifications

Disordered chaperones
assists in folding (RNA and proteins) by transient binding

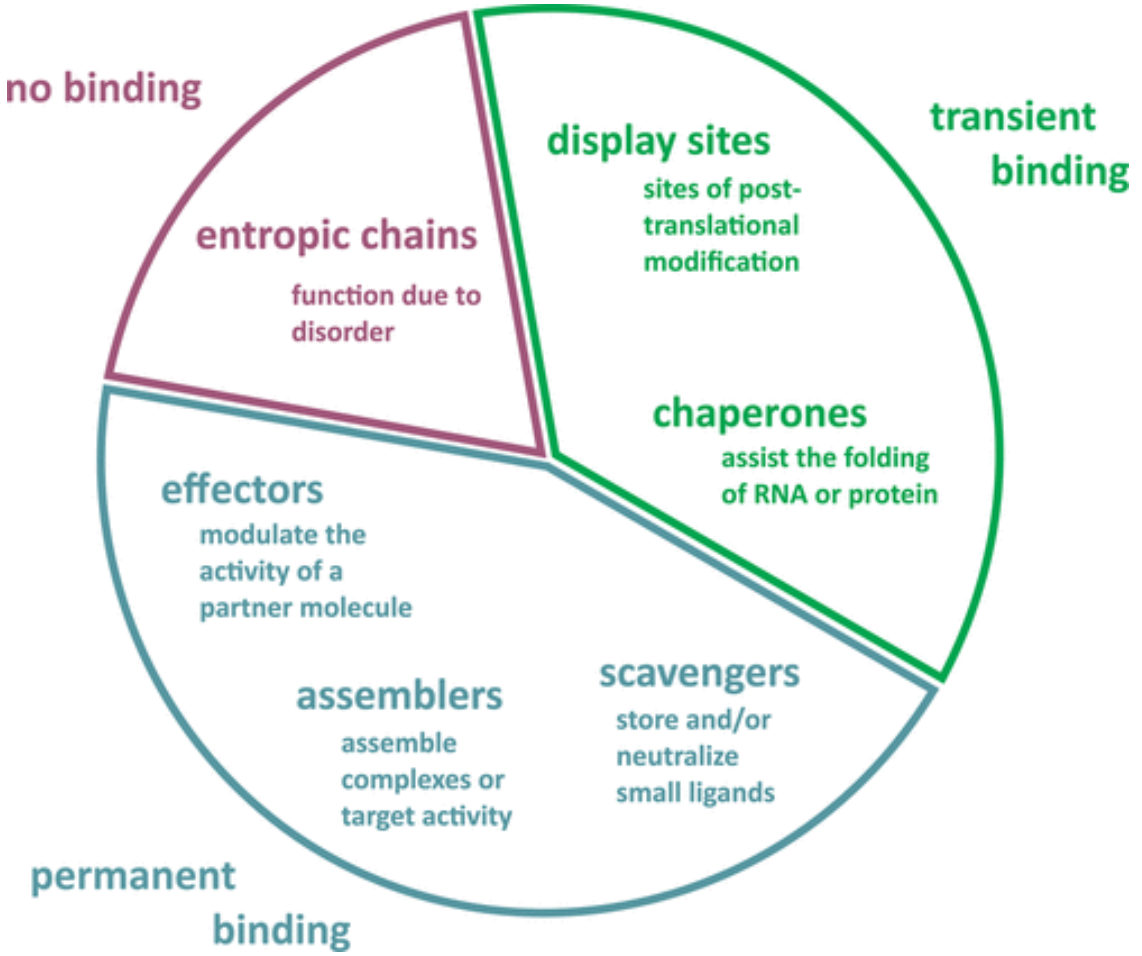
Molecular Effectors
modulating partner molecules through permanent binding

Molecular Assemblers
assemble complexes through permanent binding

Molecular Scavengers
store and/or neutralize small molecules

Unknown
no evidence supporting OR rejecting any function for IDRs

Functional classes of IDRs



van der Lee (2014), *Chem. Rev.*

Functional regions/features in IDRs

Short Linear Motifs (SLiMs)

- 3-10 residues long
- modification sites
- docking motif
  - increase the specificity and efficiency of modification events
- post translational processing

Molecular Recognition Features (MoRFs)

- 10-70 residues long
- undergo disorder-to-order transitions upon binding
- can bind multiple partners to perform multiple functions

Intrinsically Disordered Domains (IDDs)

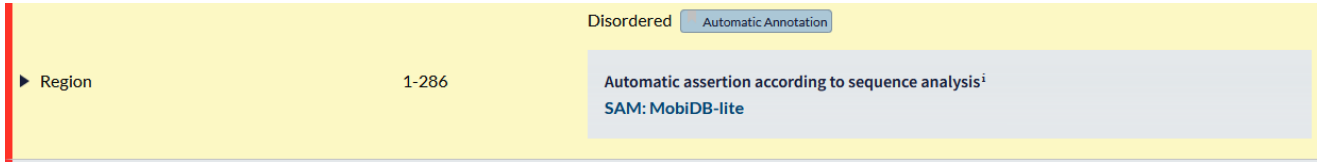
- > 70 residues long
- largely fully disordered
- mostly involved in DNA, RNA and protein binding
- conserved function, conserved sequence, and conserved disorder

length of region

large overlaps between these different features → no clear distinction

Important resources for studying IDRs

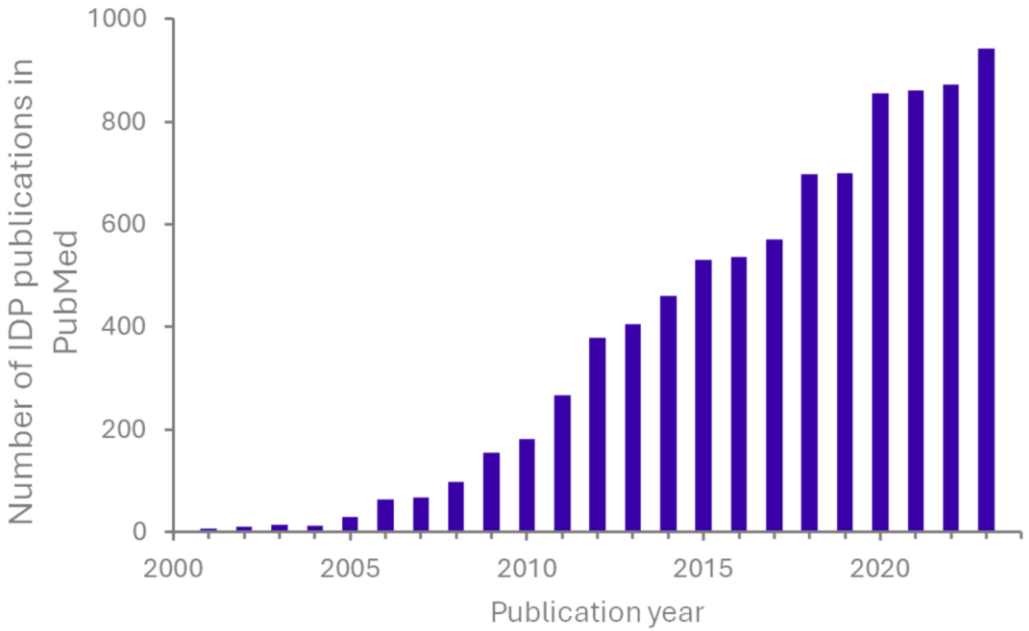
- **Uniprot:**
  - shows structure information (experimental (coming from PDB) and/or Alphafold)
  - also shows it as automatic annotations from other IDR tools



- **Disprot:**
  - manually curated repository of Intrinsically Disordered Proteins

- **MobiDB:**
  - comprehensive database for **protein disorder**, **flexibility**, and **mobility**, including experimental annotations, computational predictions, and curated data for intrinsically disordered proteins and regions

Conclusions and outlook

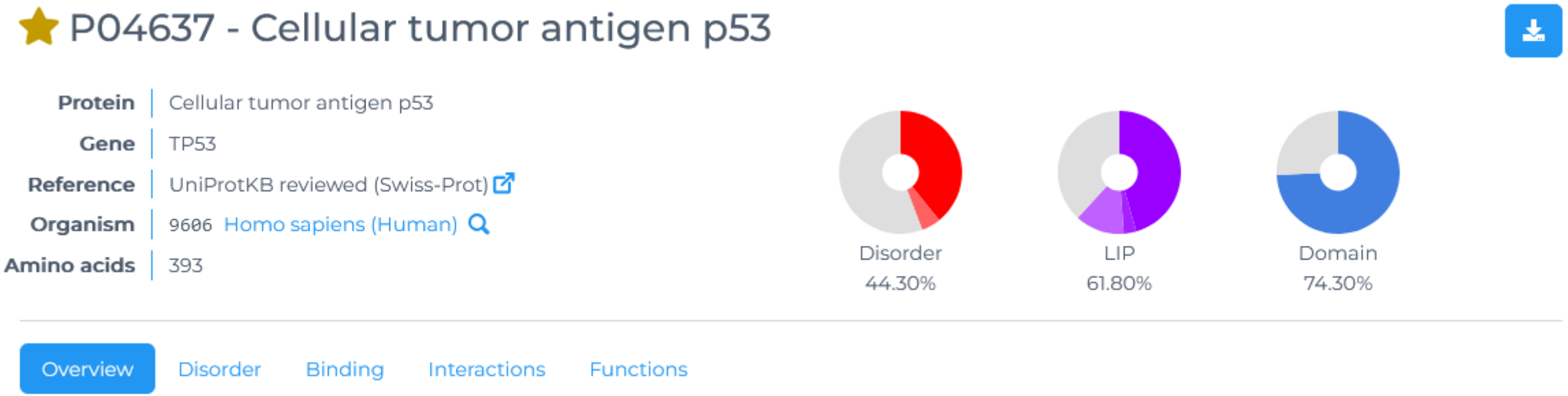


➔ order and disorder are everywhere in biology and very co-existing!

- still emerging field
  - esp. beyond eukaryotes and the evolution behind them!
- difficulties in detection and study mostly overcome
- still large challenges as to functionality!
  - projects like **DisProt** and **MobiDB** are major steps in this area

Exercise

- Go to MobiDB (https://mobidb.org/) and search for the protein *p53* (P04637)

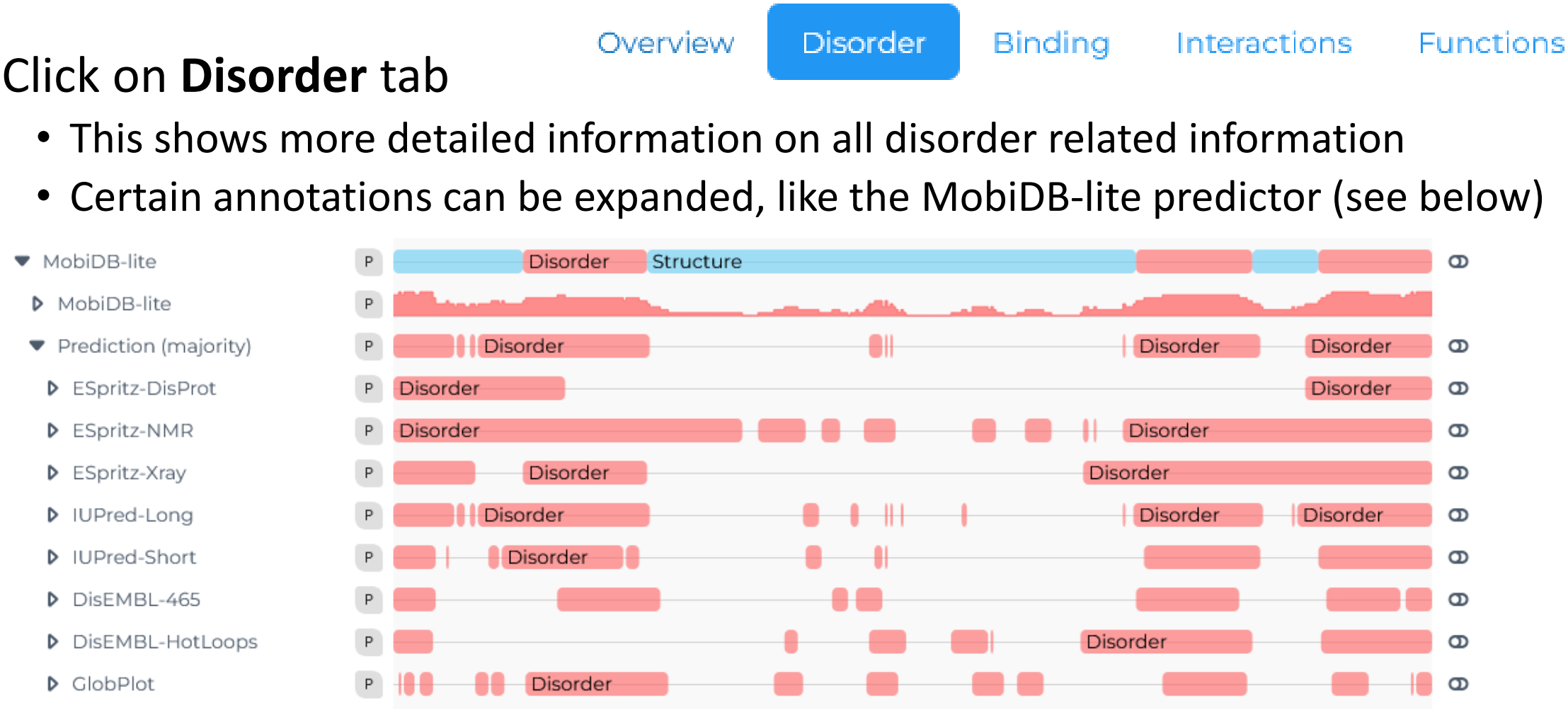


Exercise

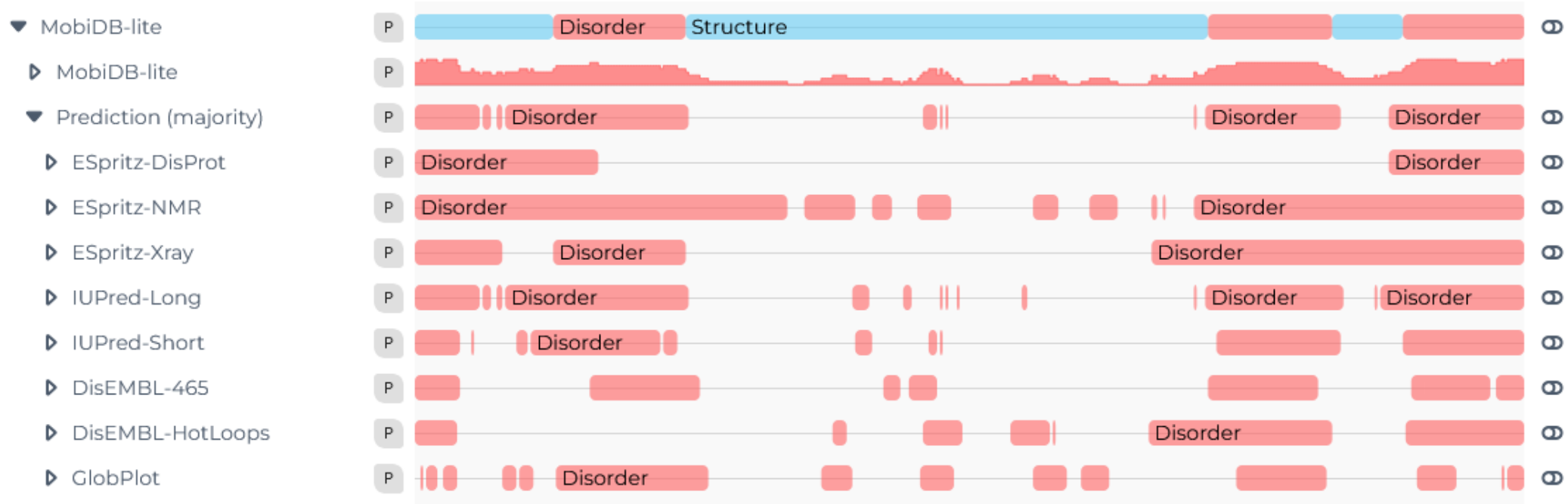
- Go to MobiDB (https://mobidb.org/) and search for the protein *p53* (UniprotID: P04637)
  - Answer these questions:
  1. Which regions of *p53* sequence are predicted to be disordered?
     1. Why are there differences between the predictors?(esp. between Espritz-NMR and Espritz-Xray?
     2. Which one would you trust and why?
     3. How would you make a decision on which regions are disordered? (without looking for exp. annotation)
  2. Check out the curated info on the disorder of *p53*:
     1. Can you find all 12 studies that show curated experimental proof of disorder in the first disordered region of that protein (amino acids 1-93)
     2. Can you name some functions of that region (and their respective studies)?

Exercise 1

- Click on **Disorder** tab
  - This shows more detailed information on all disorder related information
  - Certain annotations can be expanded, like the MobiDB-lite predictor (see below)



Exercise 1

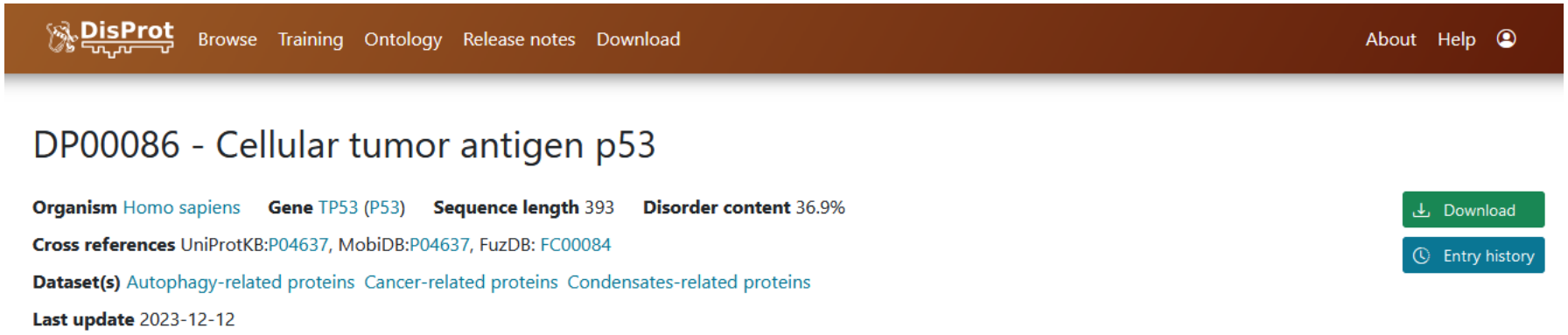


1. Which regions of *p53* sequence are predicted to be disordered?
    1. Why are there differences between the predictors?(esp. between Espritz-NMR and Espritz-Xray?
    2. Which one would you trust and why?
    3. How would make decision on which regions are disordered? (without looking for exp. annotation)

Different predictors show different results of disorder, we can generally see that there seems to be a large disordered region in the beginning (ca. 100 residues) and one or two regions in the end of the sequence.
The difference in these predictors are how they were built. In the case of the Espritz versions (DisProt, NMR, or Xray), this ML-based predictor was trained with different datasets, with different biases. These versions then also create different predictors.
It is always important to not blindly trust a prediction, but to realize that this prediction could be wrong. A majority-based predictor like MobiDB-lite is better, because it will only predict a disordered region, when 6/8 predictors agree. However, even MobiDB-lite can be wrong.

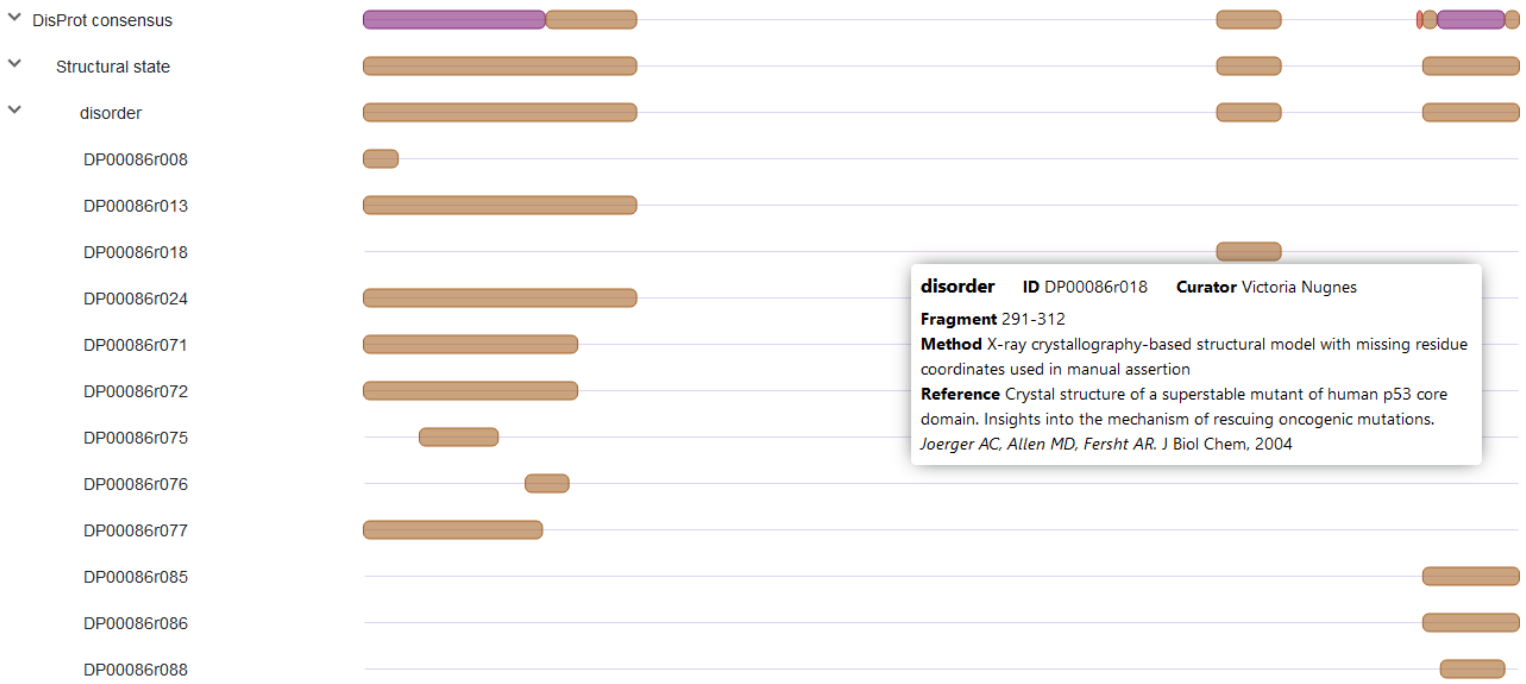
Exercise 2

- Scroll to the bottom of the page and go to the cross references:
  - Here you can see all the sources that MobiDB combines and curates together
  - In this case we want to learn more about the disorder of the protein, esp. the functionality, so we will click on the DisProt Identifier (DP00086)
  - However, check out the other annotation and see what else you can learn about this protein!



Exercise 2

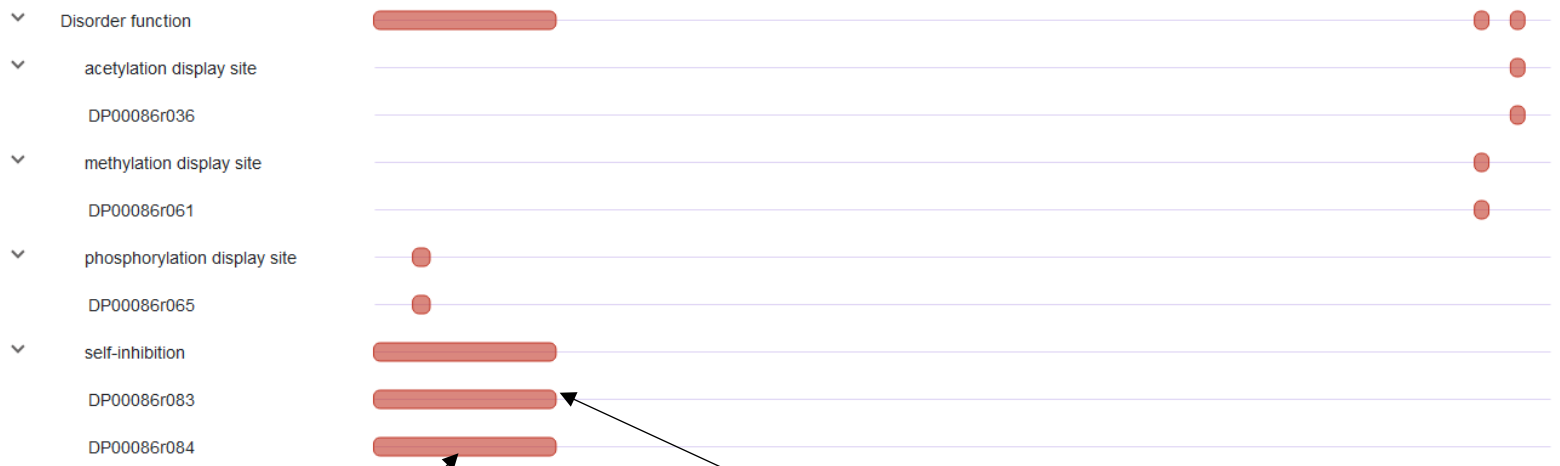
1. Check out the curated info on the disorder of *p53*:
    1. Can you find all 12 studies that show curated experimental proof of disorder in the first disordered region of that protein (amino acids 1-93)



When we open the information about the disorder in this protein we see a total of 12 manually curated sources. Each of these blocks depicts the region of disorder that this publication found and the method that it was found with. At the bottom of the page, you can directly learn more about this paper and the source!

Exercise 2

1. Check out the curated info on the disorder of *p53*:
2. Can you name some functions of that region (and their respective studies)?



self-inhibition **ID** DP00086r084 **Curator** Edoardo Salladini
Fragment 1-61
Method nuclear magnetic resonance spectroscopy evidence used in manual assertion
Reference Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Krois AS, Dyson HJ, Wright PE.* Proc Natl Acad Sci U S A. 2018

self-inhibition **ID** DP00086r083 **Curator** Edoardo Salladini
Fragment 1-61
Method fluorescence anisotropy evidence used in manual assertion
Reference Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Krois AS, Dyson HJ, Wright PE.* Proc Natl Acad Sci U S A. 2018

When you scroll down and open the disorder function, you can see which regions have even been annotated with a function. In the example on the left, we can see that there was a study in 2018, that detemerings the region of 1-61 residues has a self-inhibition function. It actually shows up twice, because they discovered it with two different methods!
You can read more about it in the mentioned paper and the methodology used!

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

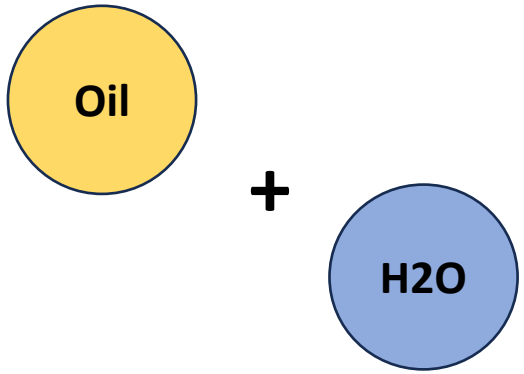
Phase-separation in biology

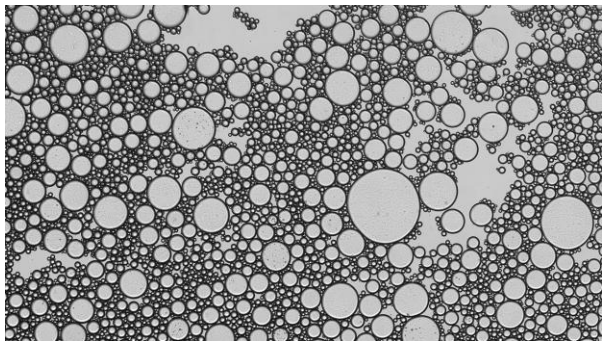
Eric Schumbera

Miguel Andrade
CBDM Group
Biocenter I
e.schumbera@uni-mainz.de

(Liquid-liquid) Phase separation(LLPS) as a concept

In daily life

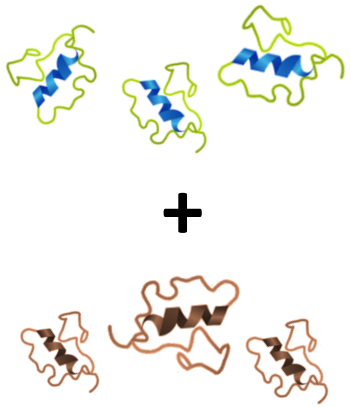




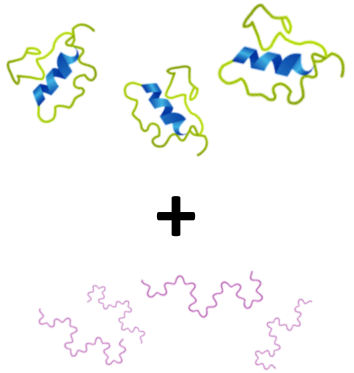
Water/oil emulsion

In biology

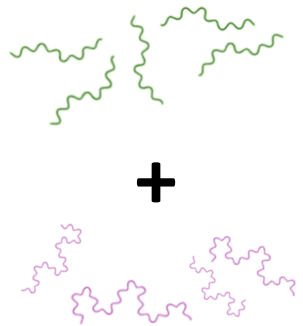
Protein-Protein (I)

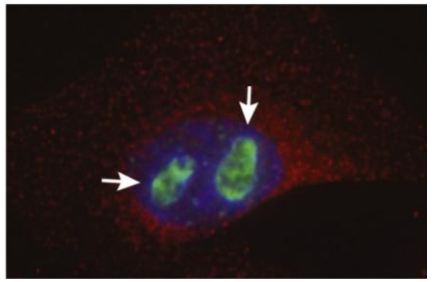


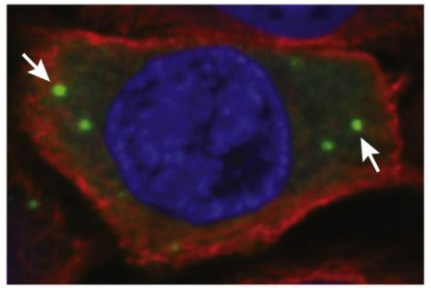
Protein-RNA (II)



RNA-RNA (III)

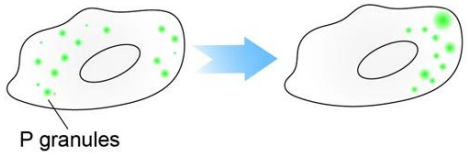






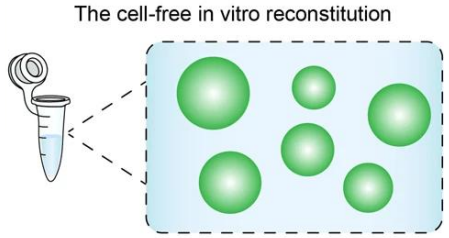
granule
(bio-)condensate
RNP droplet
RNP body
paraspeckle
assembly
…

The young history of LLPS in biology



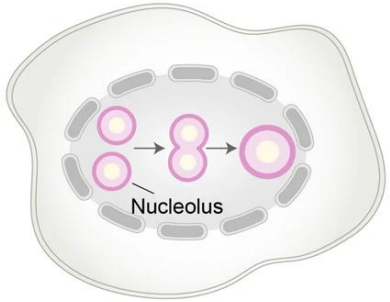
Hyman & Brangwyne:
P-granules exhibits liquid droplet like behaviors and rapidly dissolve and condense!

Steven McKnight:
First in-vitro model for the architecture and formation of RNA granules



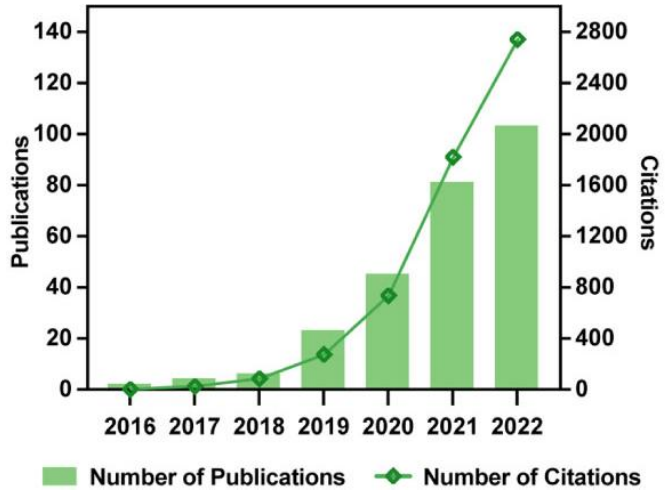
2009 | 2011 | 2012 | 2013 - present

Hyman & Brangwyne:
Nucleolus exhibits liquid droplet like behaviours!

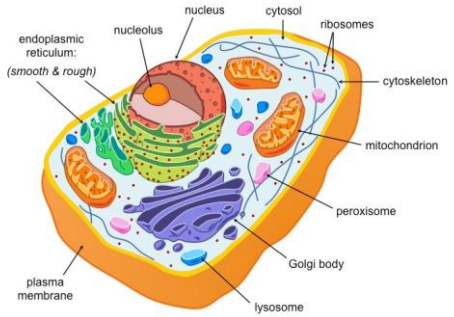


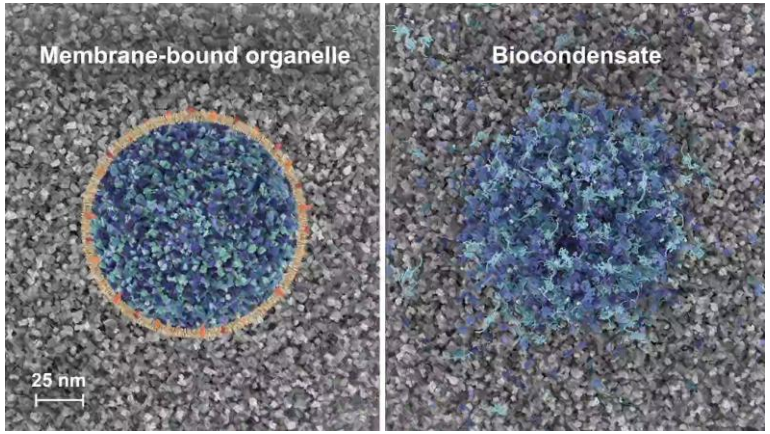
LLPS shown to be involved in many aspects:

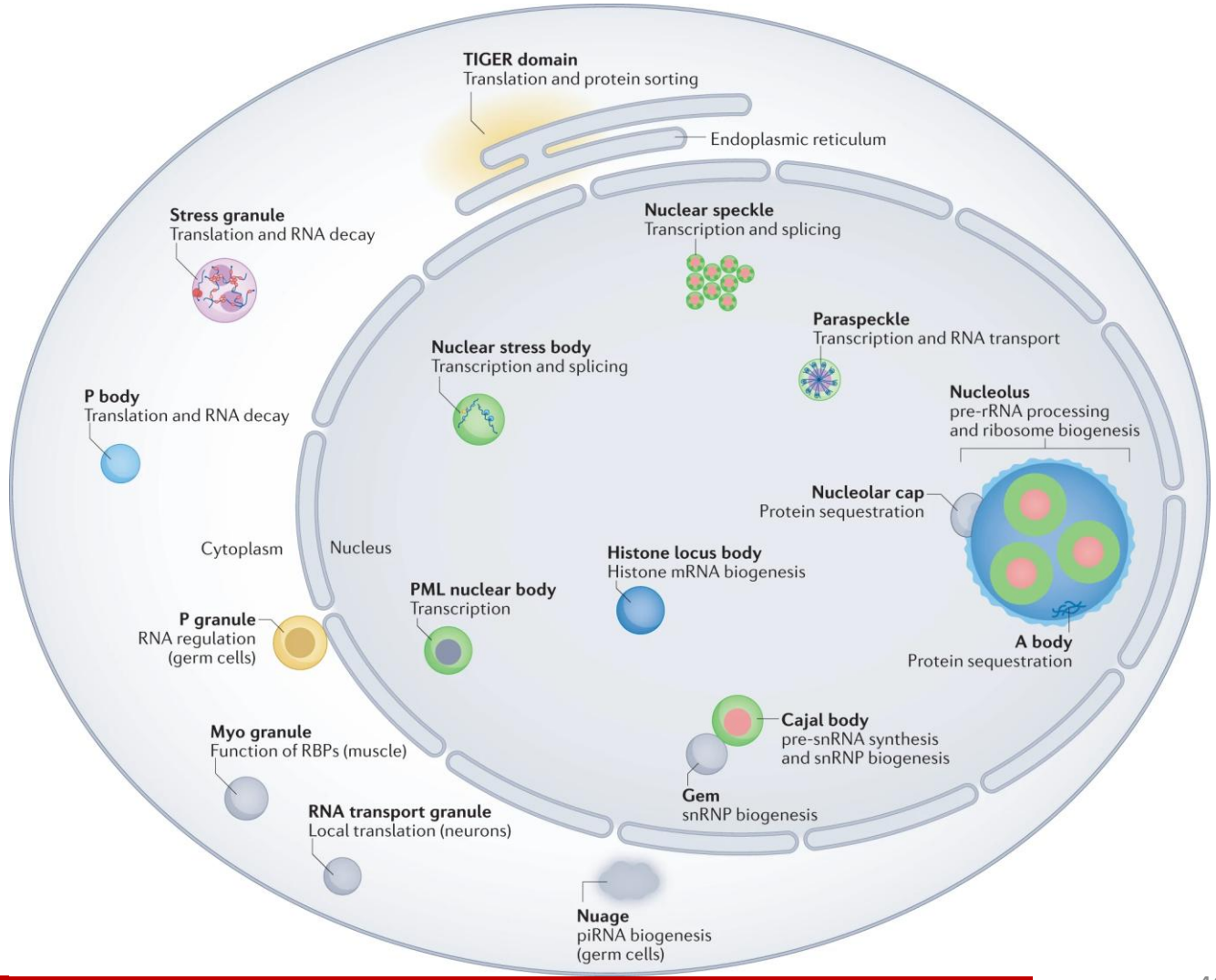
- mediation of the assembly of adhesion complexes
- condensation of cGAS activates innates immune signaling
- LLPS promotes T cell receptor signal transduction
- promotes miRISC assembly
- …



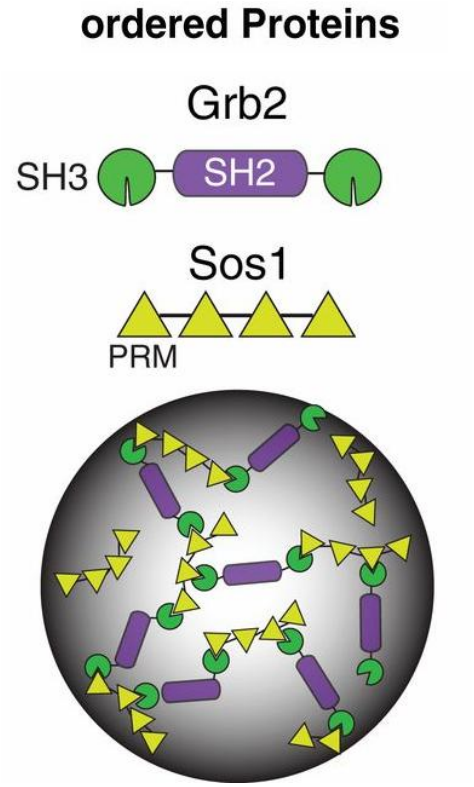
Membrane-less organelles (MLOs)

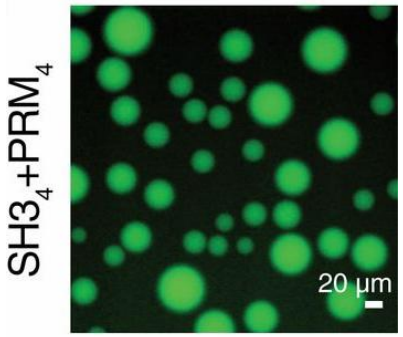




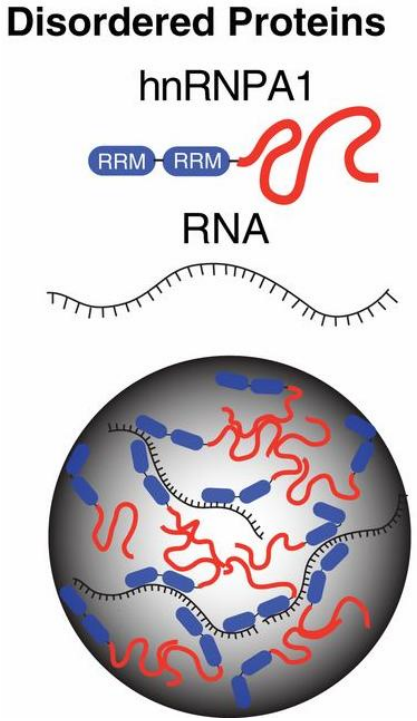


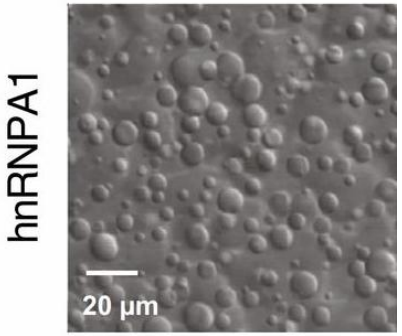
The mechanisms behind LLPS


ordered Proteins
Grb2
SH3
SH2
Sos1
PRM



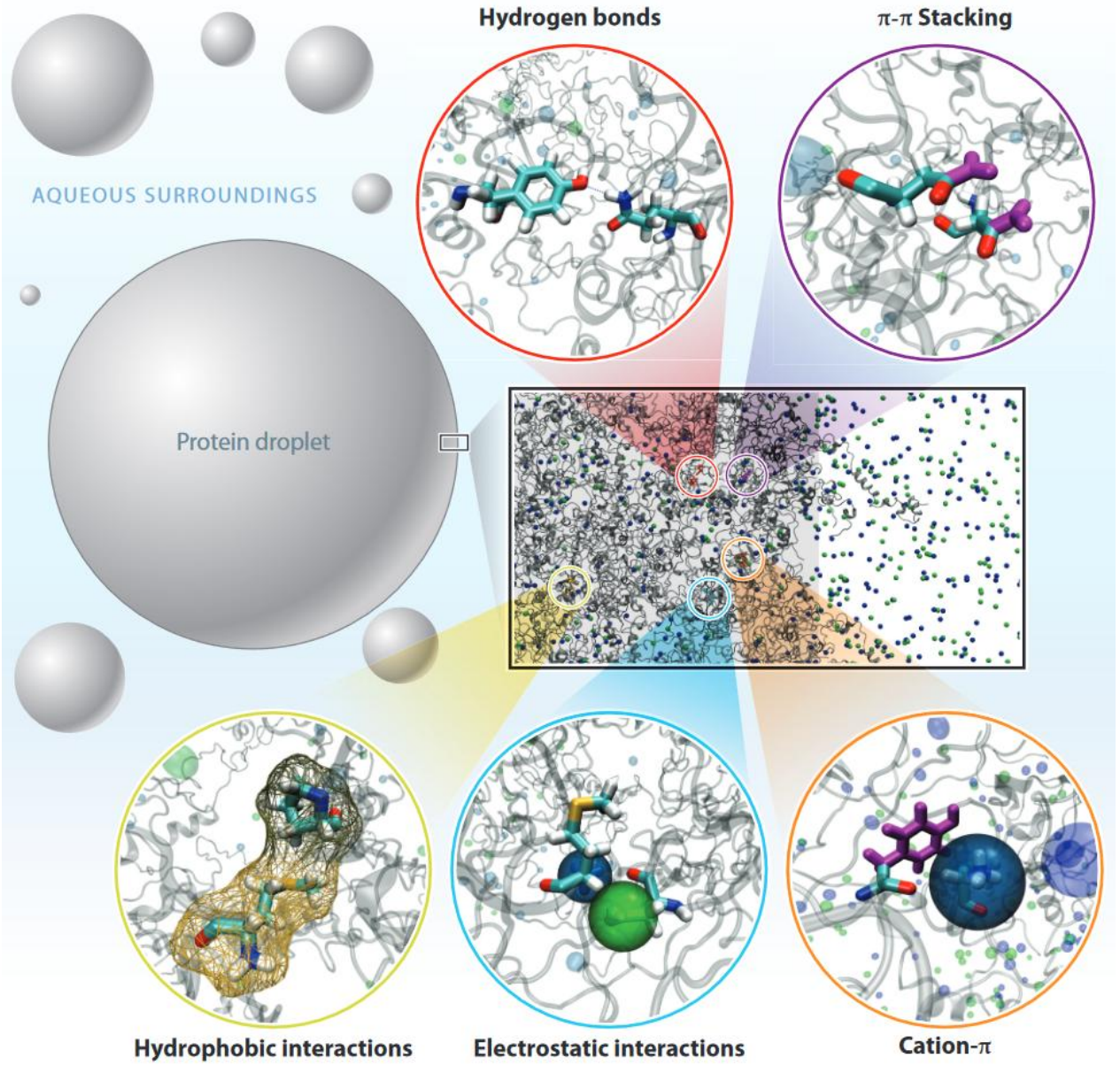
<u>driven by:</u>
- “strong” protein interactions between folded domains





<u>driven by:</u>
- weak, transient **multivalent** interactions between IDRs and/or RNAs:

1. Hydrogen bonds
2. π - π stacking
3. Hydrophobic interactions
4. Electrostatic interactions
5. Cation-π stacking



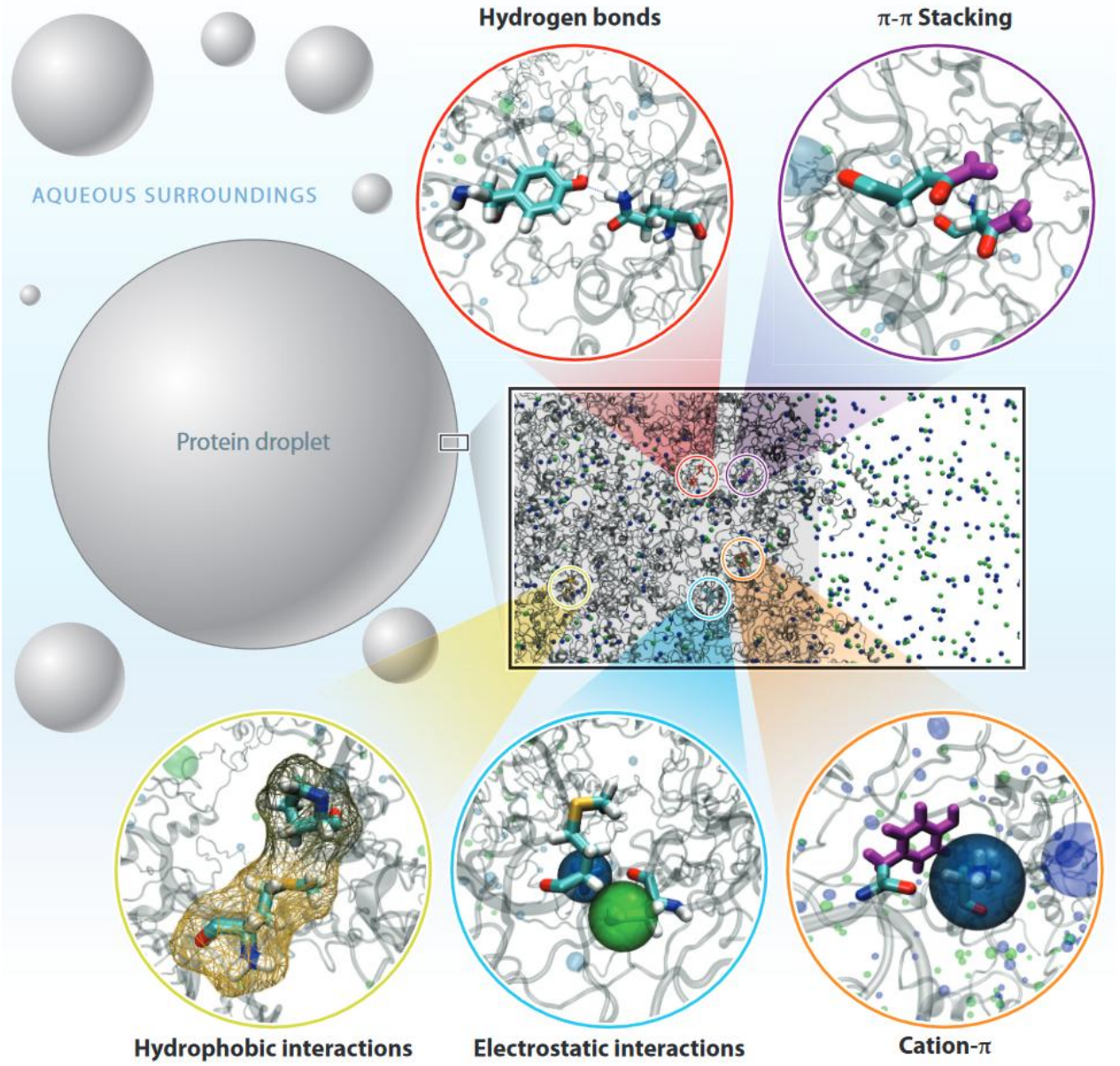
Hydrogen bonds

- shown to stabilize phase separation
- all amino acids able to participate, but mostly from polar ones
- Likely very important for incorporating RNA/DNA in condensates

Arginine Arg (R)
Lysine Lys (K)
Histidin His (H)
pos. charged

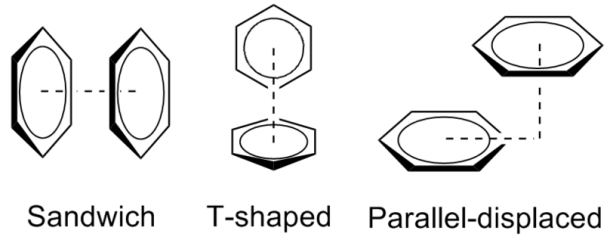
Aspartate Asp (D)
Glutamate Glu (E)
neg. charged

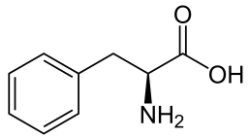
Asparagine Asn (N)
Glutamine Gln (Q)
not charged but strong polar



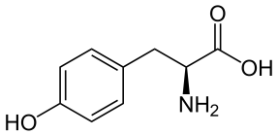
π – π stacking

- interactions between sp^2-hybridized groups
  - mostly **aromatic rings**
  - possibly the strongest of the 5 interaction types

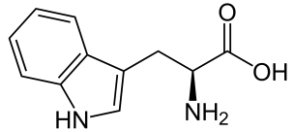




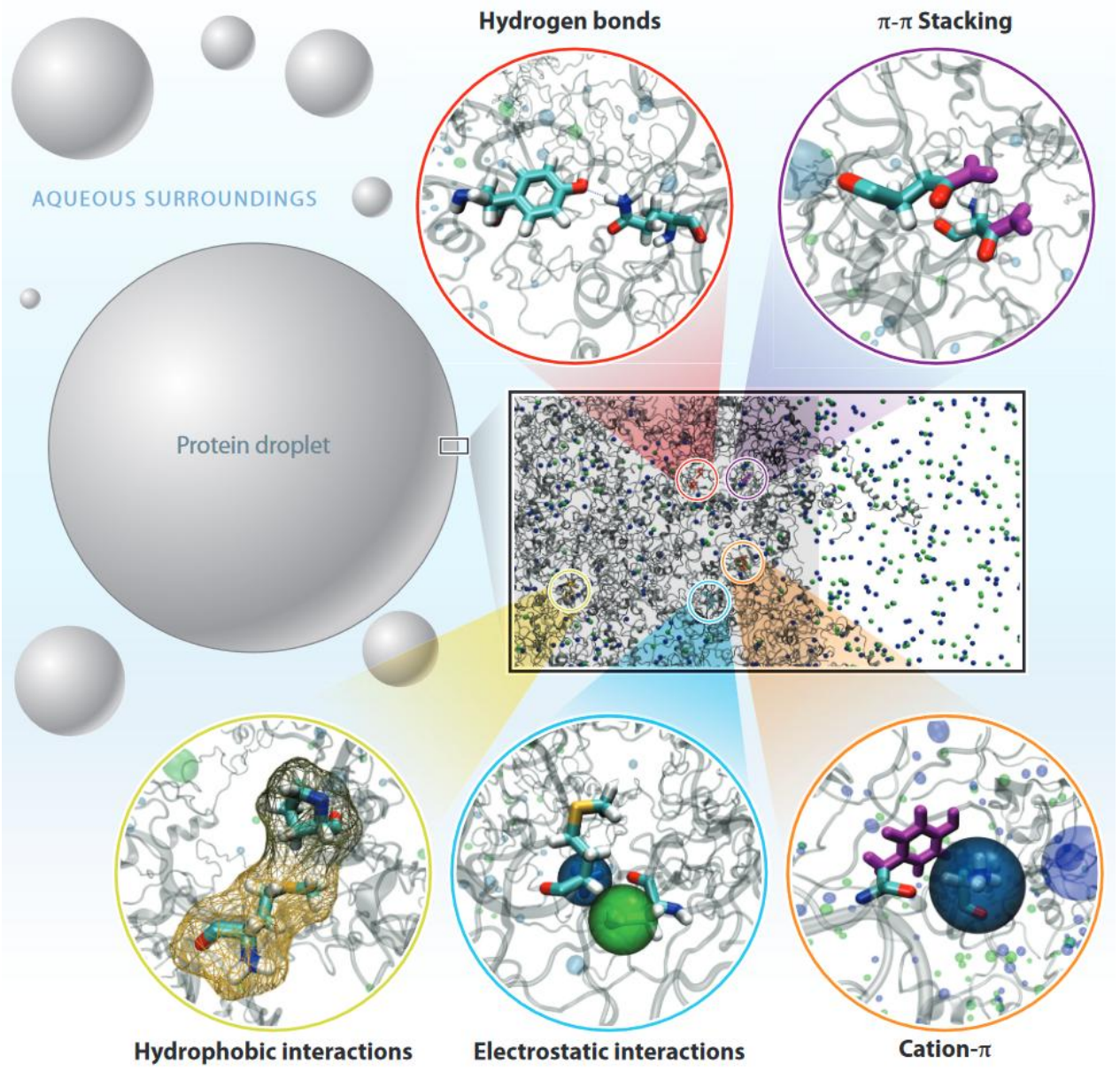
Phenylalanine
Phe (F)



Tyrosine
Tyr (Y)



Tryptophan
Trp (W)

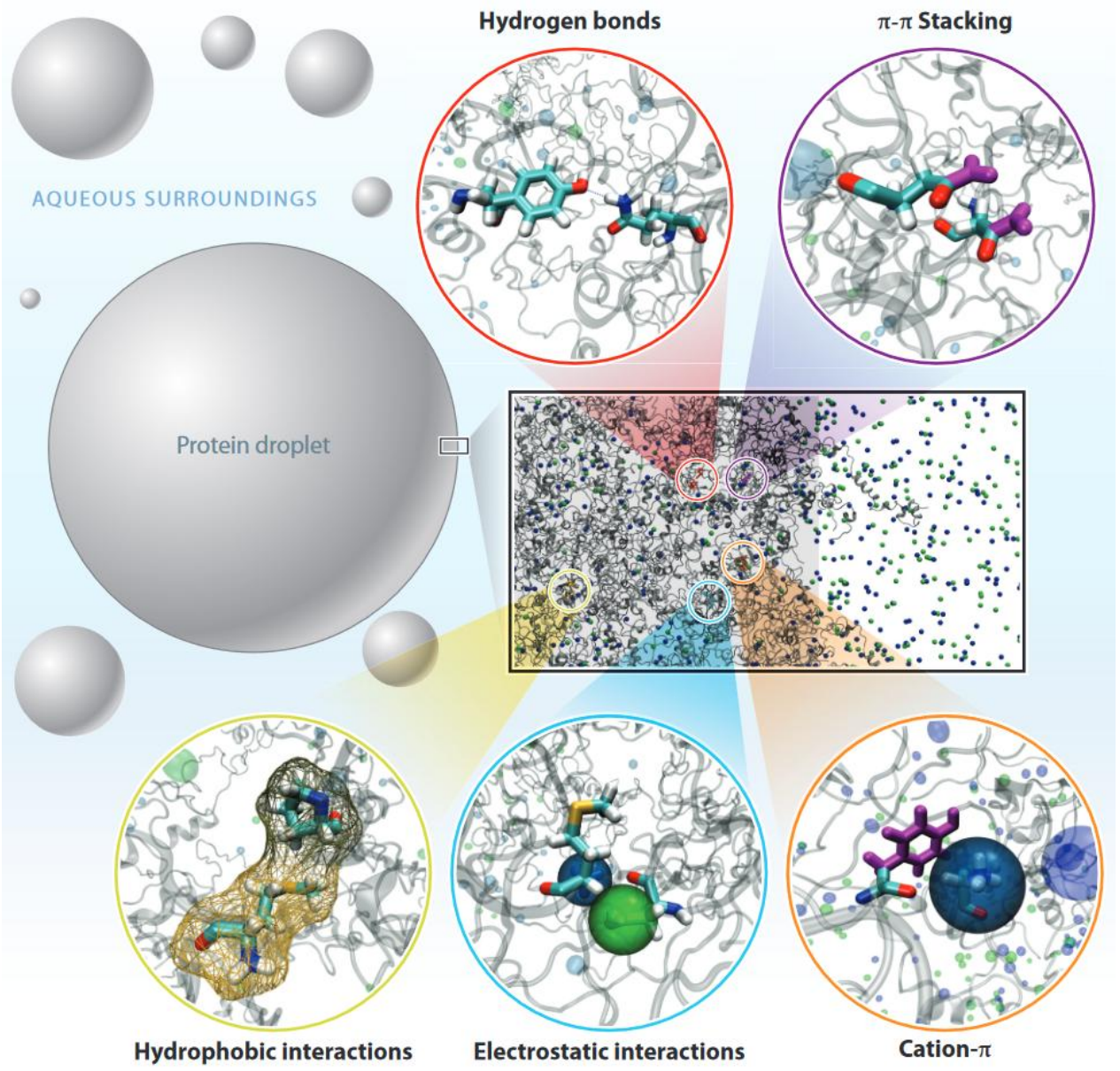


Hydrophobic interactions

- generally a major force of **protein folding**
  - **BUT:**
    - hydrophobic leucine-rich helices in the processing body (MLO) **stabilizes** the formation as condensate
    - important role in recruiting specific ligands into the condensed phase

hydrophobic (aliphatic)

Isoleucine	**Valine**	**Leucine**	**Alanine**	**Glycine**
Ile (I)	Val (V)	Leu (L)	Leu (L)	Leu (L)



Electrostatic interactions

- intrinsically disordered proteins (IDPs) are usually **enriched in charged amino acids**
- strongly affected by pH, salt concentration

Arginine
Arg (R)

Lysine
Lys (K)

Histidin
His (H)

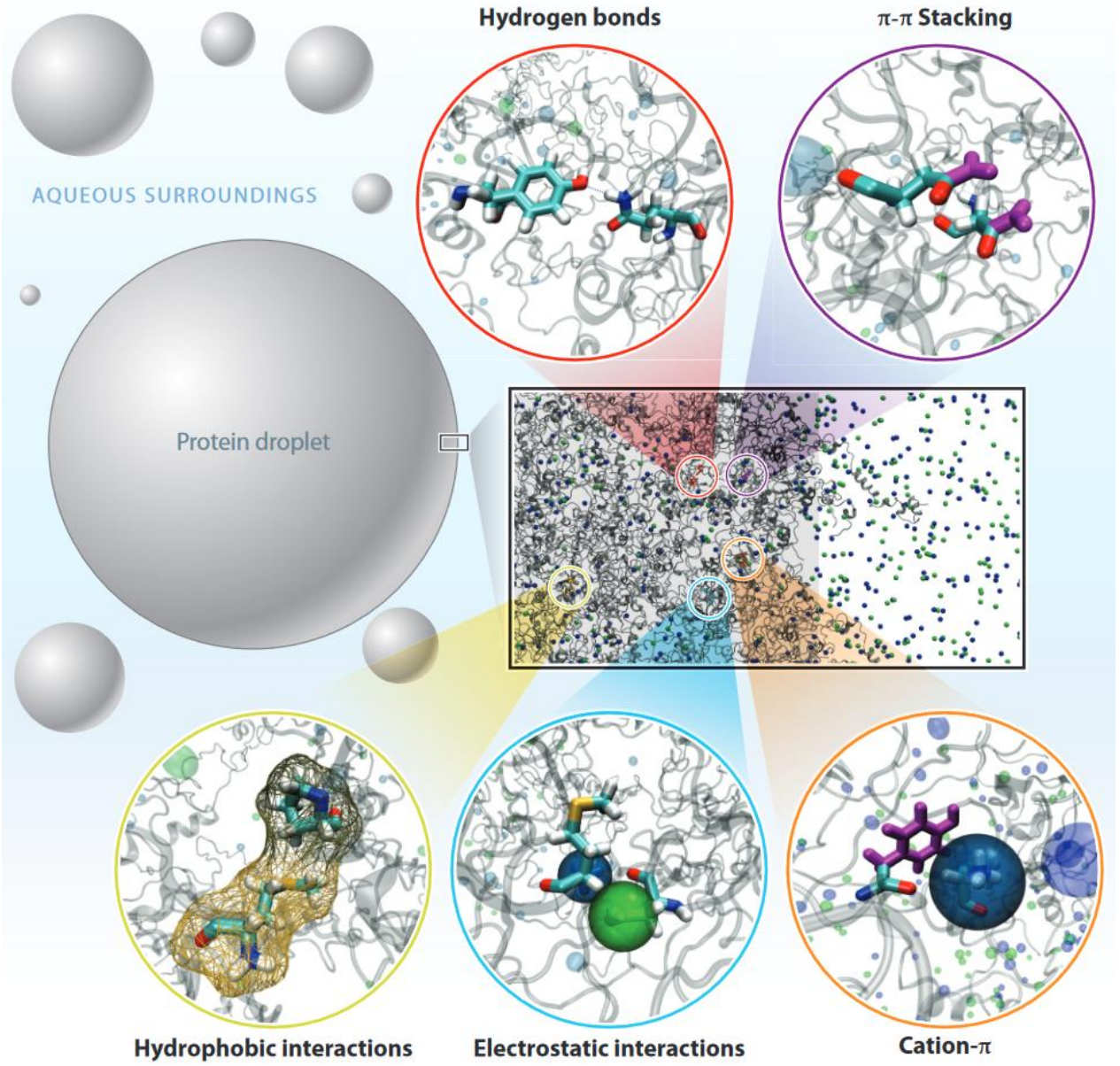
pos. charged

Aspartate
Asp (D)

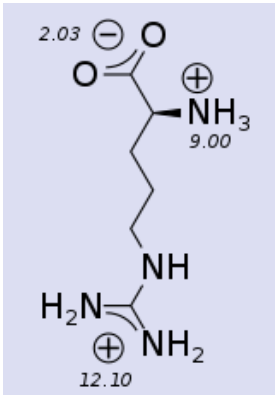
Glutamate
Glu (E)

neg. charged

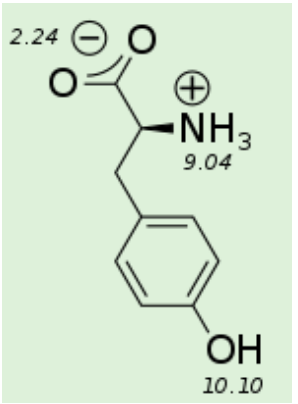
Cation-π interaction



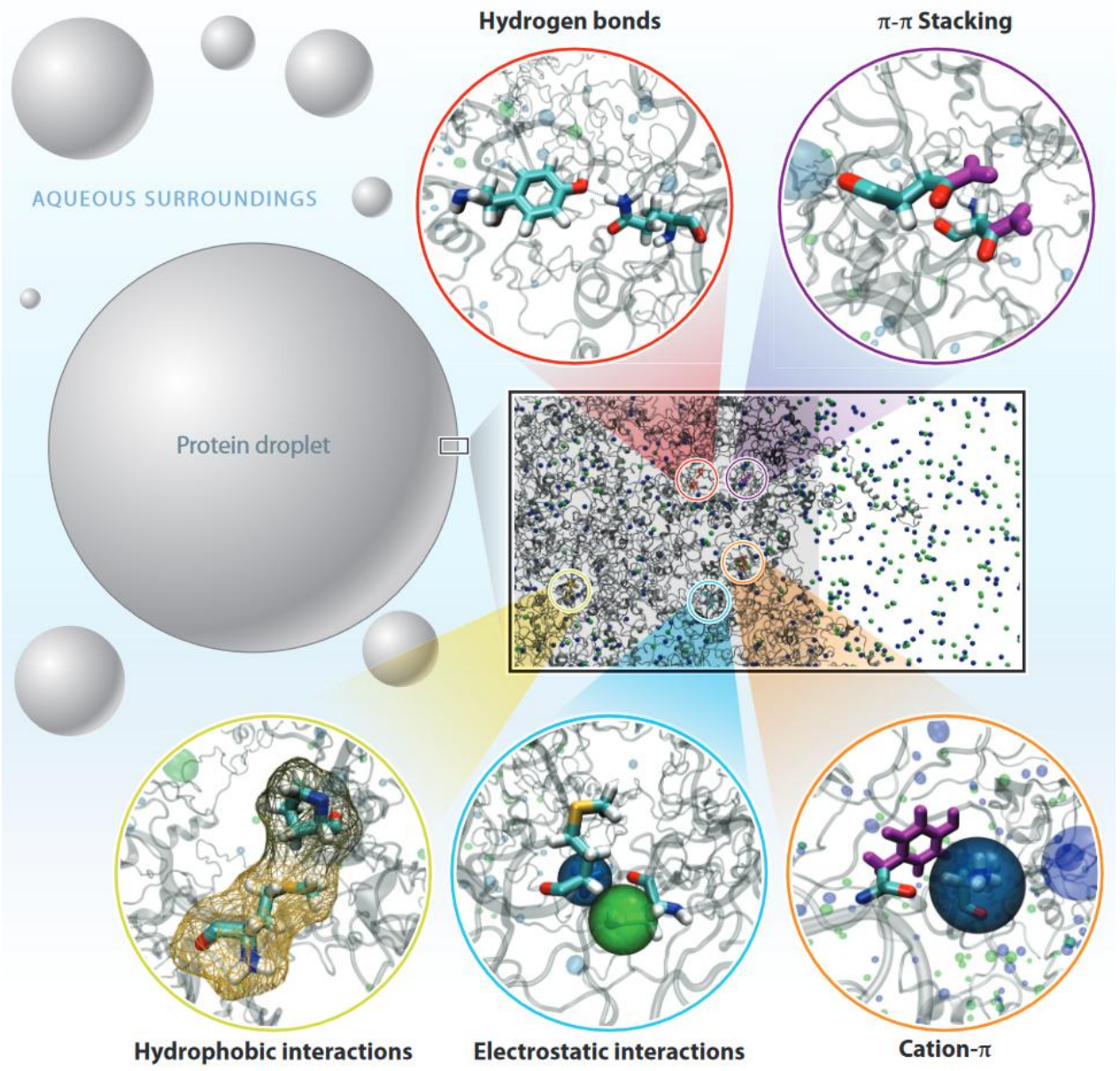
- aromatic rings and cationic residues, particularly between **arginine** and **tyrosine** residues



Arginine
Arg (R)

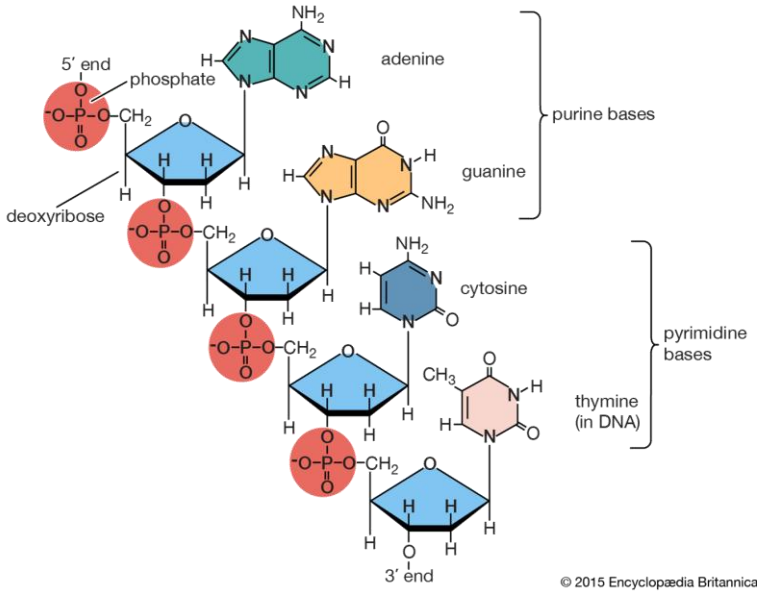


Tyrosine
Tyr (Y)



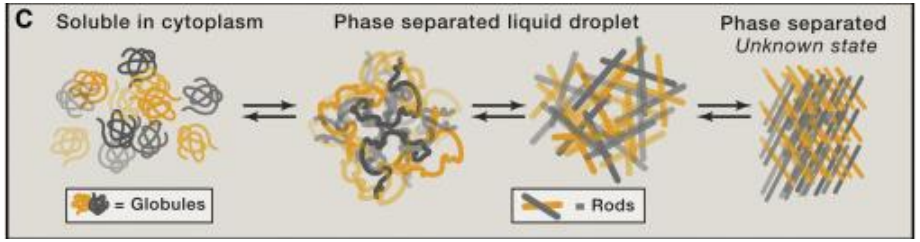
Cation-π interaction

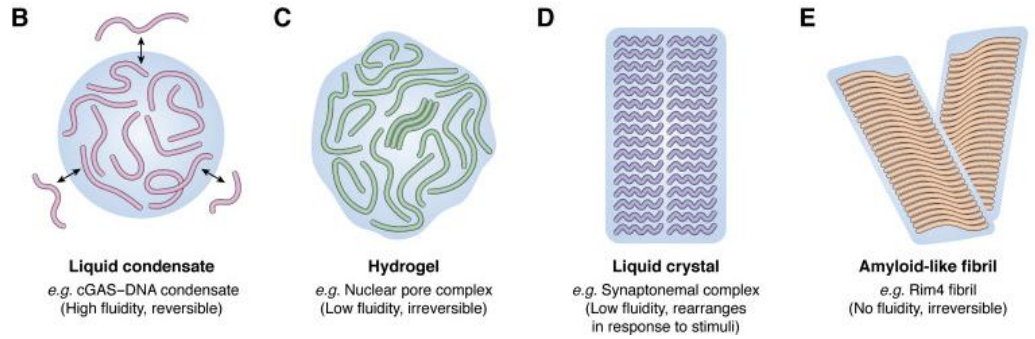
- this and π - π stacking apply also to the aromatic rings of **RNA** and **ssDNA**
  - Why not dsDNA?

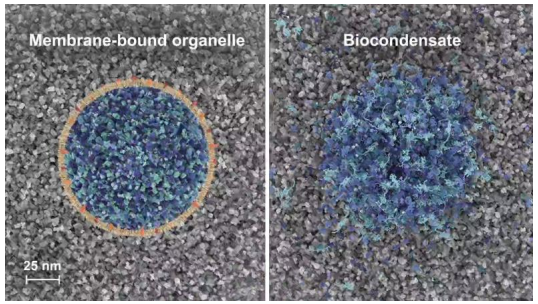


The dynamic nature of phase separation

- organelles with membranes are not dynamic
- MLOs are highly dynamic!
  - quick assembly and disassembly
  - change of state
  - interaction with surrounding phase





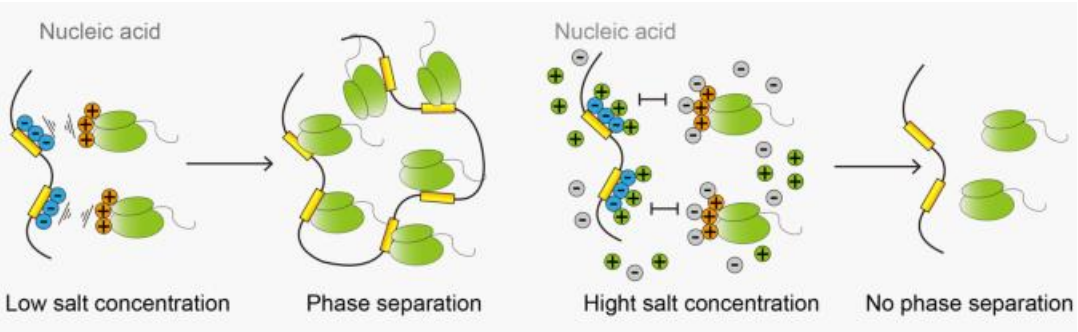


How does the organism "control" LLPS behavior?

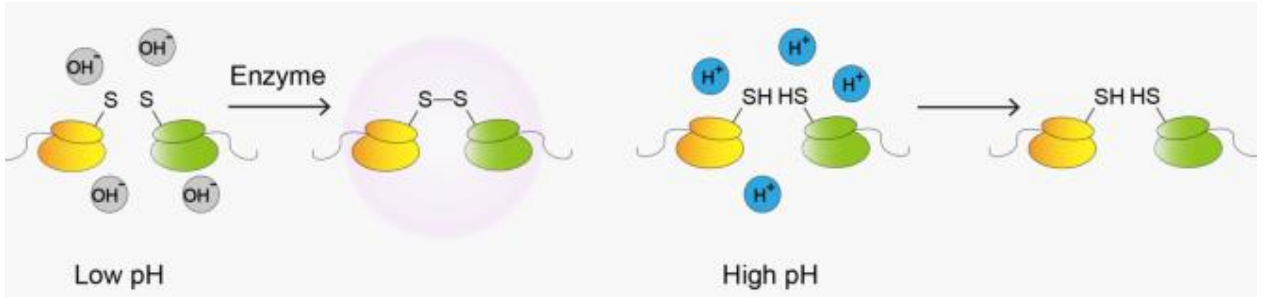
The role of temperature, pH and salt concentration in LLPS

- biocondensates usually have a lower critical solution temperature (LCST) and an upper critical solution temperature (UCST)
  - not that applicable **in-vivo** but important for **in-vitro** studies of LLPS

- salt concentration may induce or prevent phase separation
  - not only explained by charge changes



- pH
  - reduced pH causes stress granules (SG) formation
  - but most likely affects LLPS in both direction



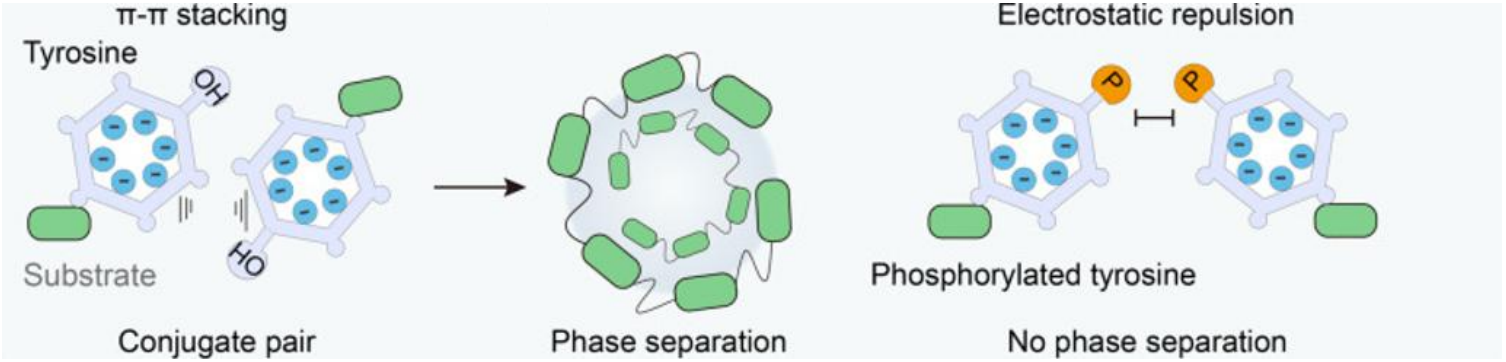
CONTEXT DEPENDENCY

Post-translational modifications (PTMs)

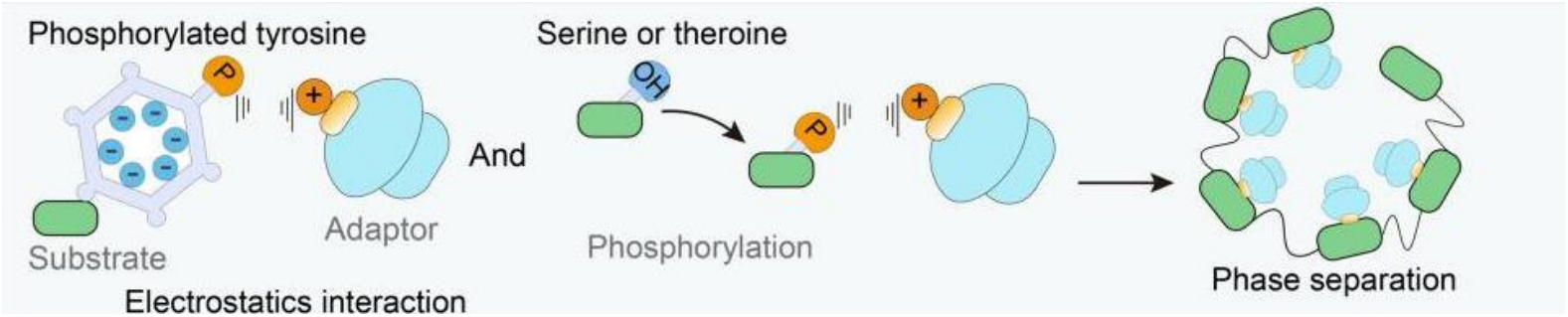
IDRs are more prone to PTMs than folded domains!

PTMs change the chemical properties of the residues → multivalency directly affected

Phosphorylation **hinders** phase separation

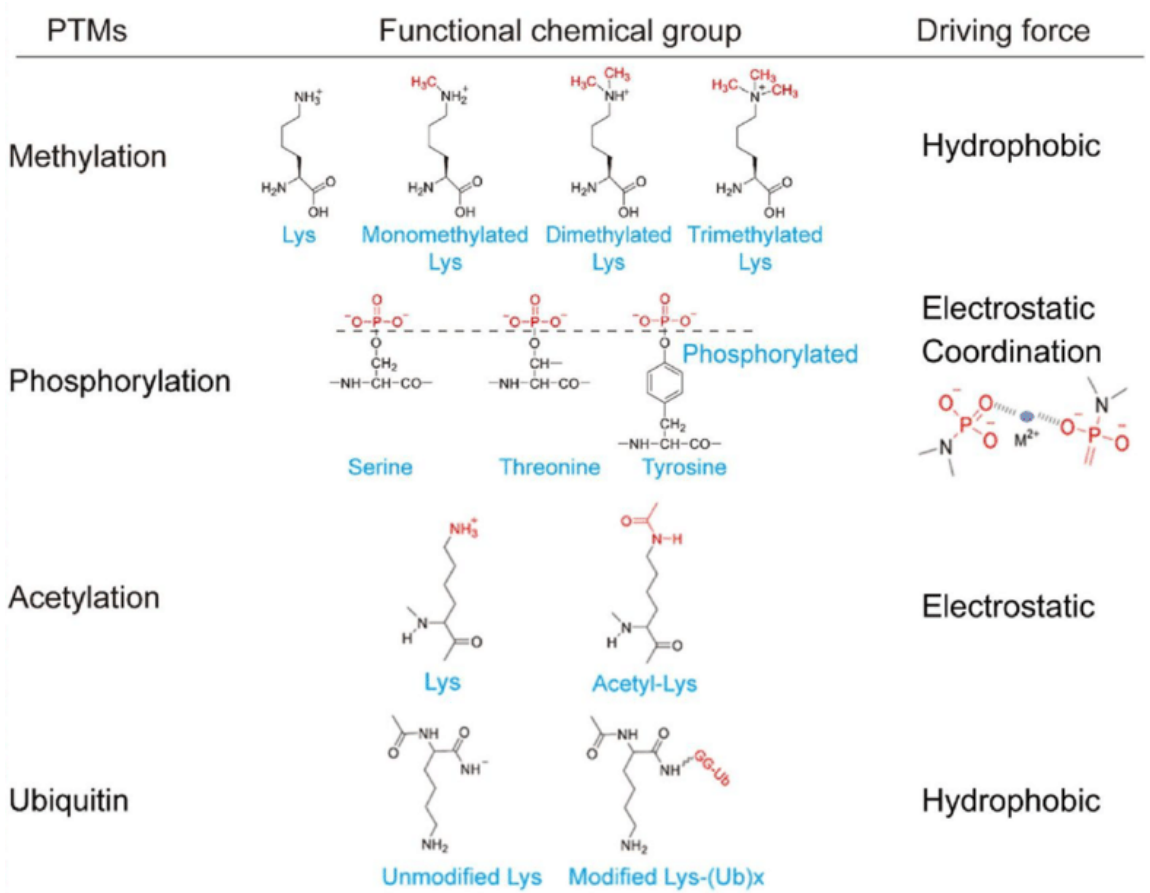


Phosphorylation **facilitates** phase separation



CONTEXT DEPENDENCY

Post-translational modifications (PTMs)

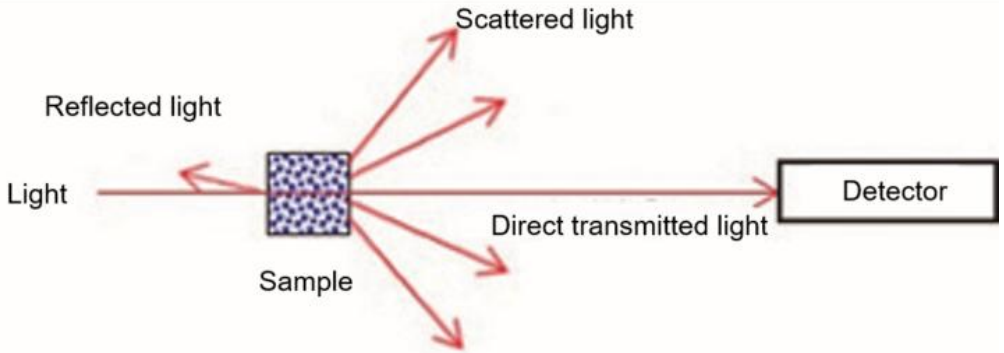


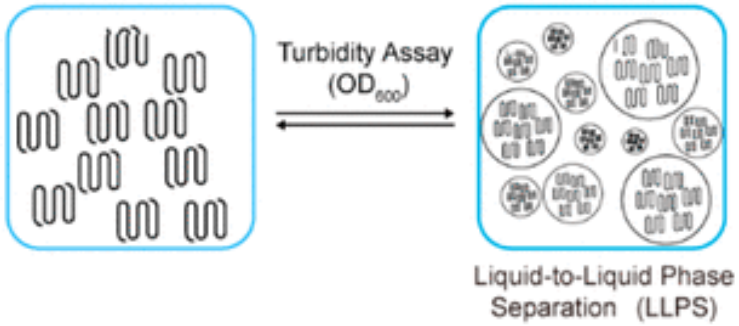
more PTMs suspected to play large roles in LLPS formation!

Methods to study LLPS – 1. Turbidity

no LLPS induced

LLPS induced

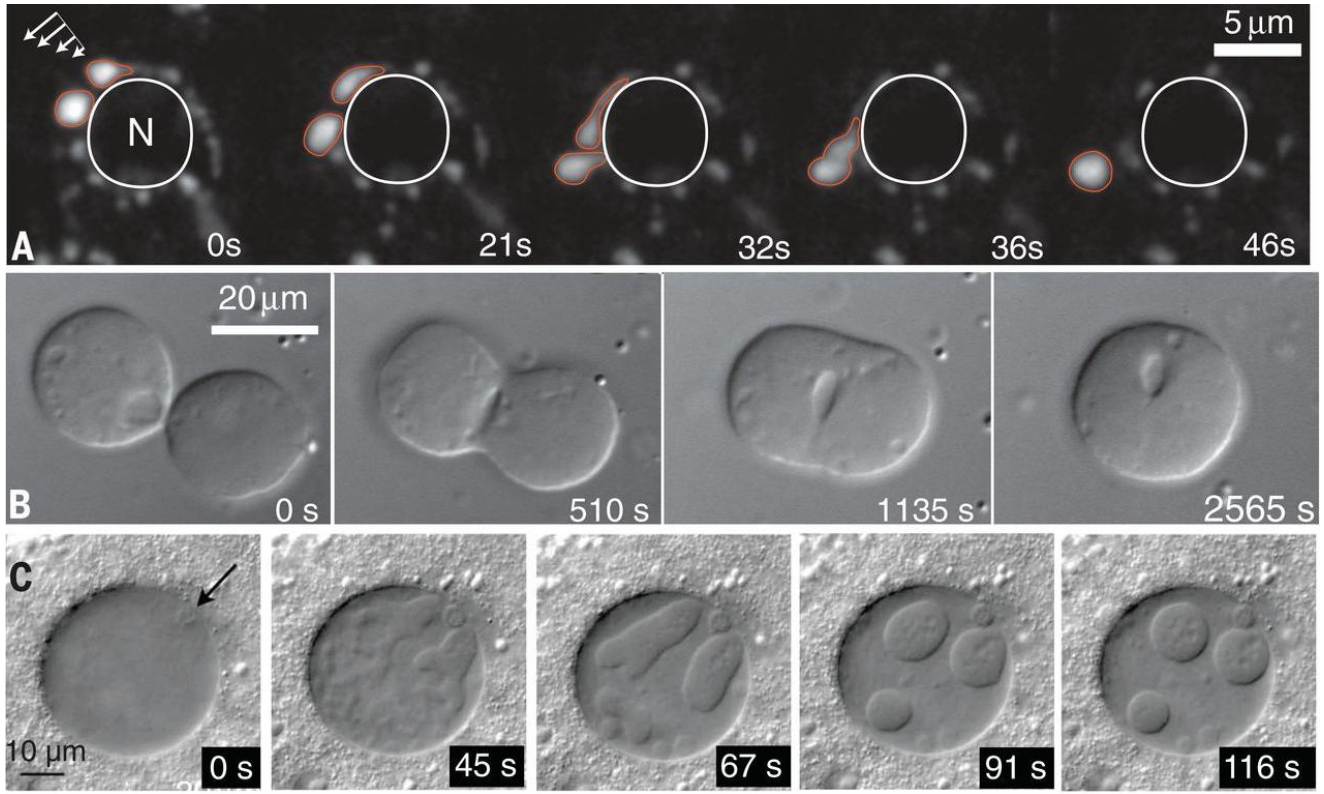




- only in-vitro applicable
- simple, cheap, useful for preliminary experiment

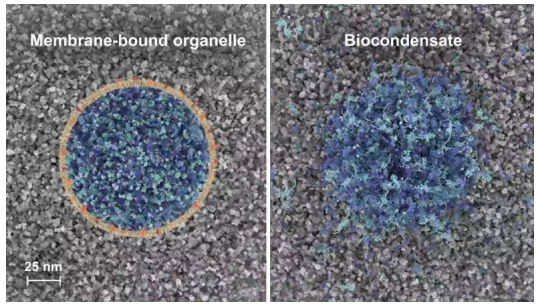
Methods to study LLPS – 2. (Fluorescence) microscopy

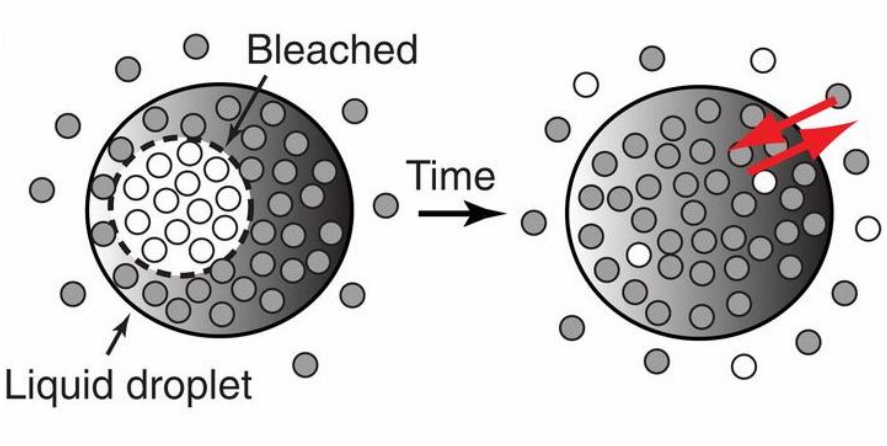
- in-vitro **AND** in-vivo
- common and very flexible
- antibody-staining or fluorescent labeling necessary

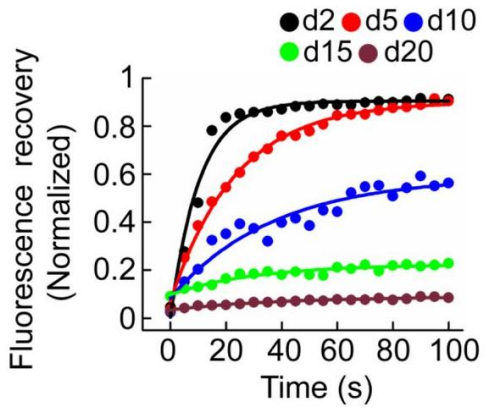


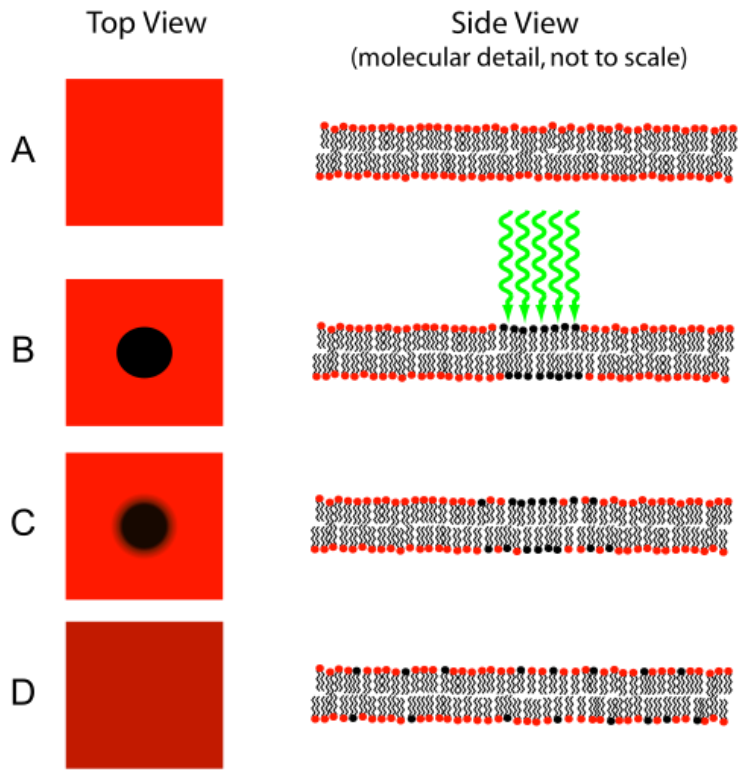
Methods to study LLPS – 3.Fluorescence recovery after photobleaching (FRAP)

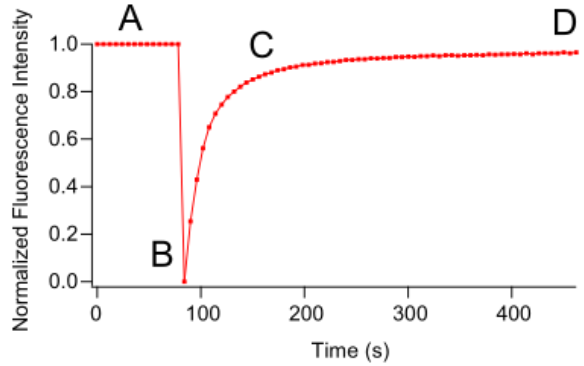
can quantify the "fluidity" of a biocondensate!



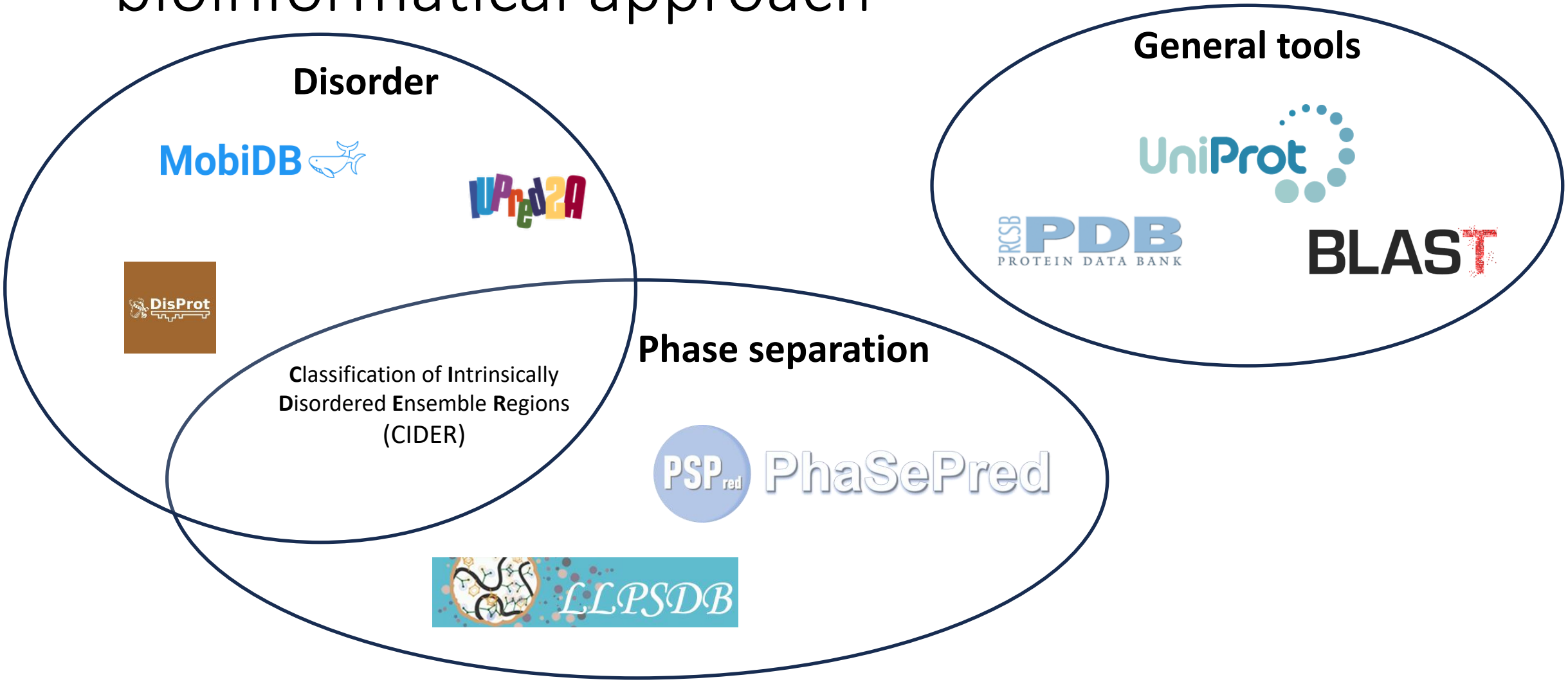








Database/resources and predictors for a bioinformatical approach



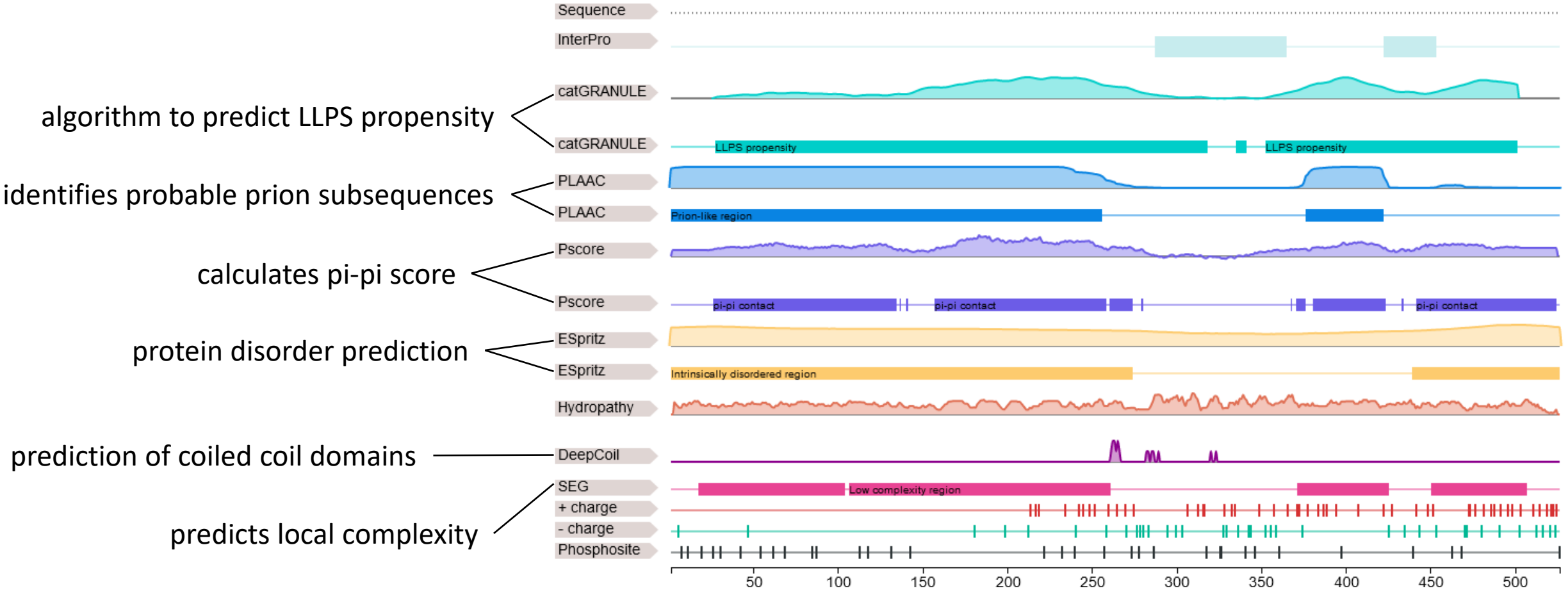
PhaSePred: A Meta-predictor For Phase-Separating Proteins

Protein Information					
Uniprot Entry	P35637	**Entry name**	FUS_HUMAN	**Gene name**	FUS TLS
Length	526	**Status**	reviewed	**Organism**	Homo sapiens (Human)

PhaSePred Scores				
	Score (8-feature)	**Rank (8-feature)**	**Score (10-feature)**	**Rank (10-feature)**
PS-Self score (Proteins that can self-assemble to form condensates)	0.928	0.993	0.930	0.994
PS-Part score (Proteins whose phase separation behaviors are regulated by protein or nucleic acid partner components)	0.817	0.967	0.911	0.981

*** The 8-feature model incorporates Hydropathy, FCR, IDR, LCR, PScore, PLAAC, catGRANULE, and DeepCoil.**
The 10-feature model incorporates the 8 features described above plus Phos frequency and DeepPhase. This model is only available for human proteins.

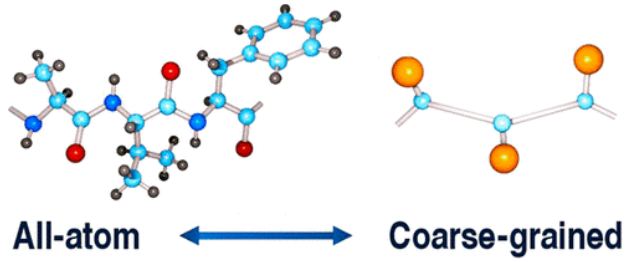
PhaSePred: A Meta-predictor For Phase-Separating Proteins



Molecular dynamics (MD) for LLPS simulation

computer simulation method for analyzing the physical movements of atoms and molecules

- for LLPS used: **coarse-grained** models
  - e.g. MARTINI model
  - each amino acid is one „element“

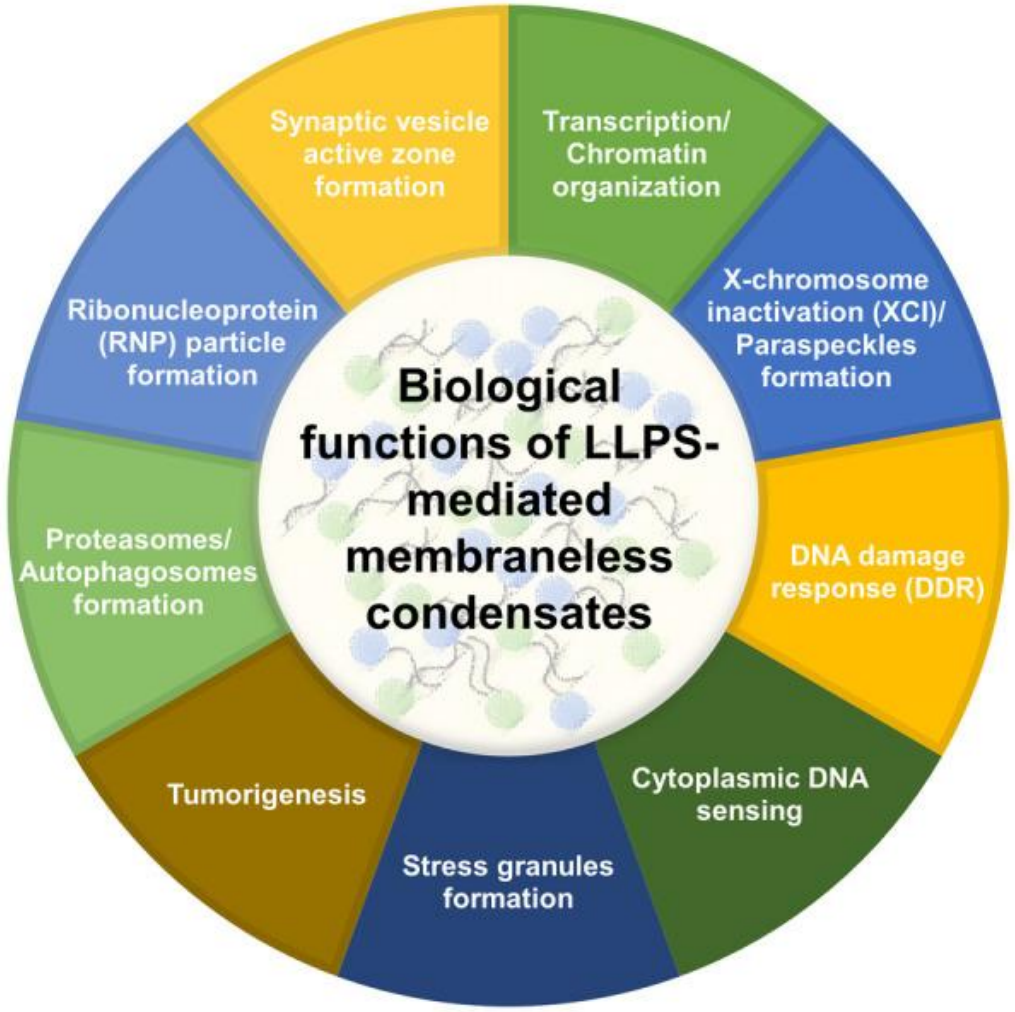


MD simulation promise easy, cheap and fast research on LLPS!

BUT

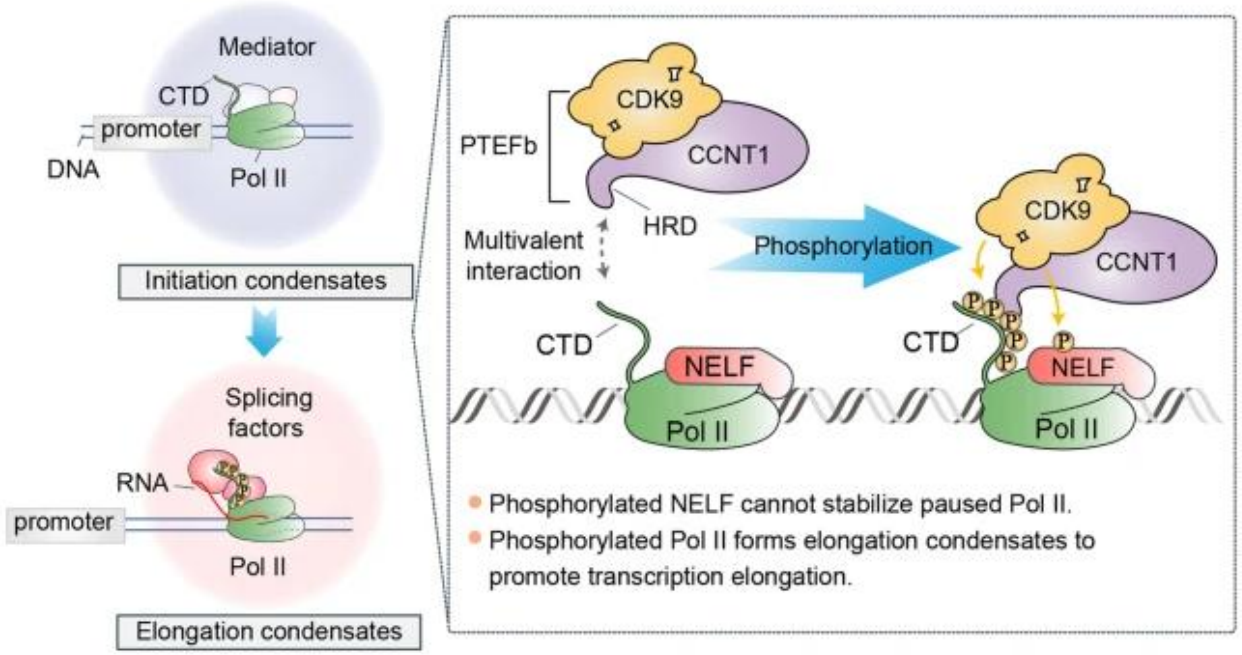
- simplified model loses information (folded domains not considered)
- parametrization of model complicated
- simulation of an artificial subset of the cell
- …

Functions of Phase separation in biology



Understanding function enables us to manipulate LLPS for therapeutic application!

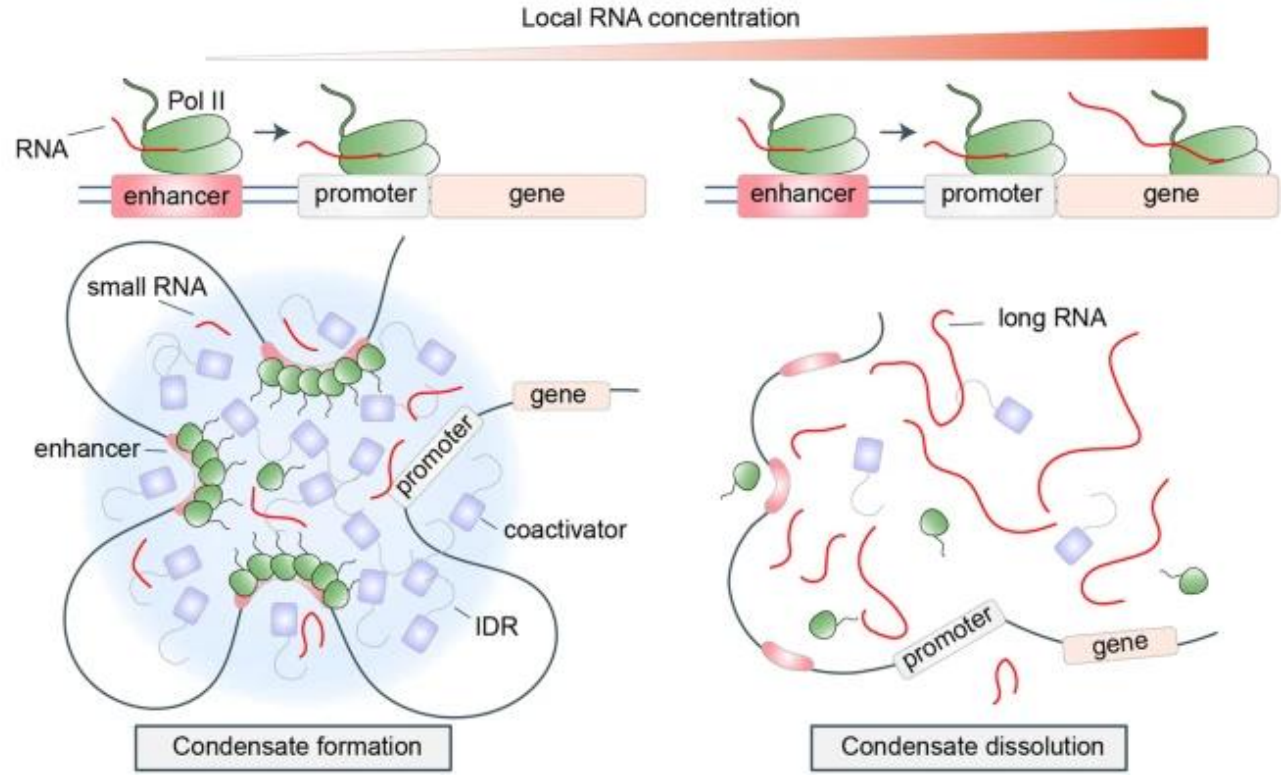
How LLPS regulates transcription



- transcription consists of initiation, elongation and termination
- the initiation complex forms by **condensate** of RNA polymerase II (Pol II), transcription factors and coactivators
- PTEFb recruited through multivalent LLPS mechanics!
- Phosphorylated Pol II (with different multivalency) forms elongation condensates with very different factors (splicing factors)

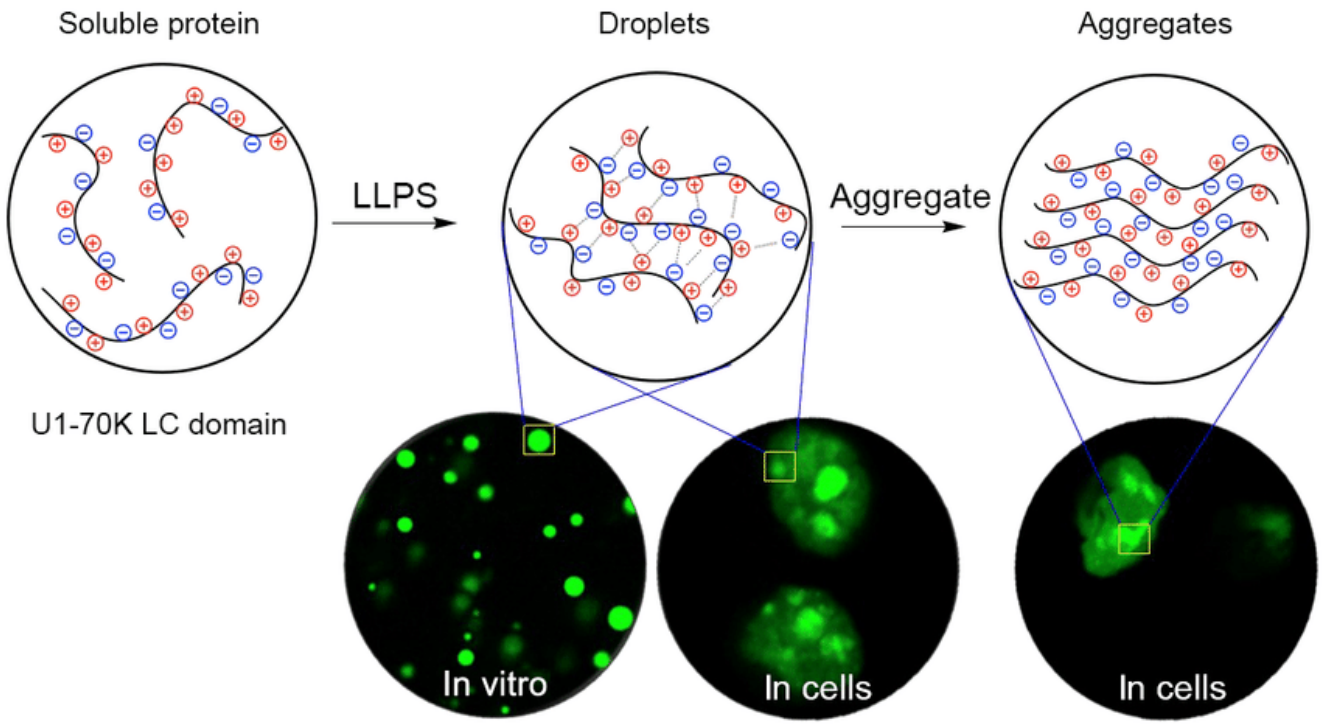
How LLPS regulates transcription

- feedback mechanism in (super)enhancers is controlled by phase separation!
- low levels of RNA **promote** the formation of transcriptional condensates (based around Pol II)
- high levels of RNA can **dissolve** the transcription condensates

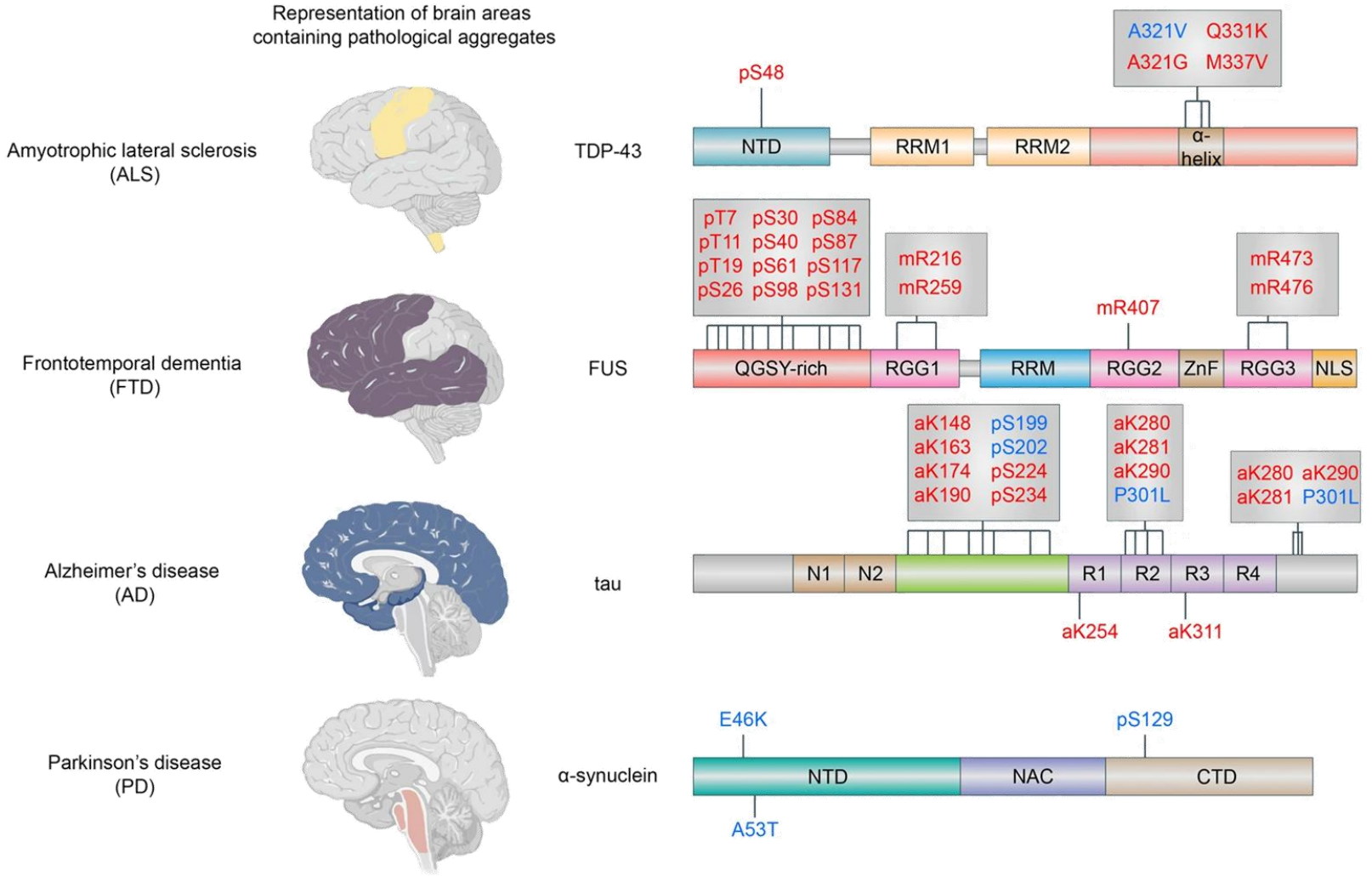


LLPS in neurodegenerative diseases

protein aggregation is largely aberrant phase separation!

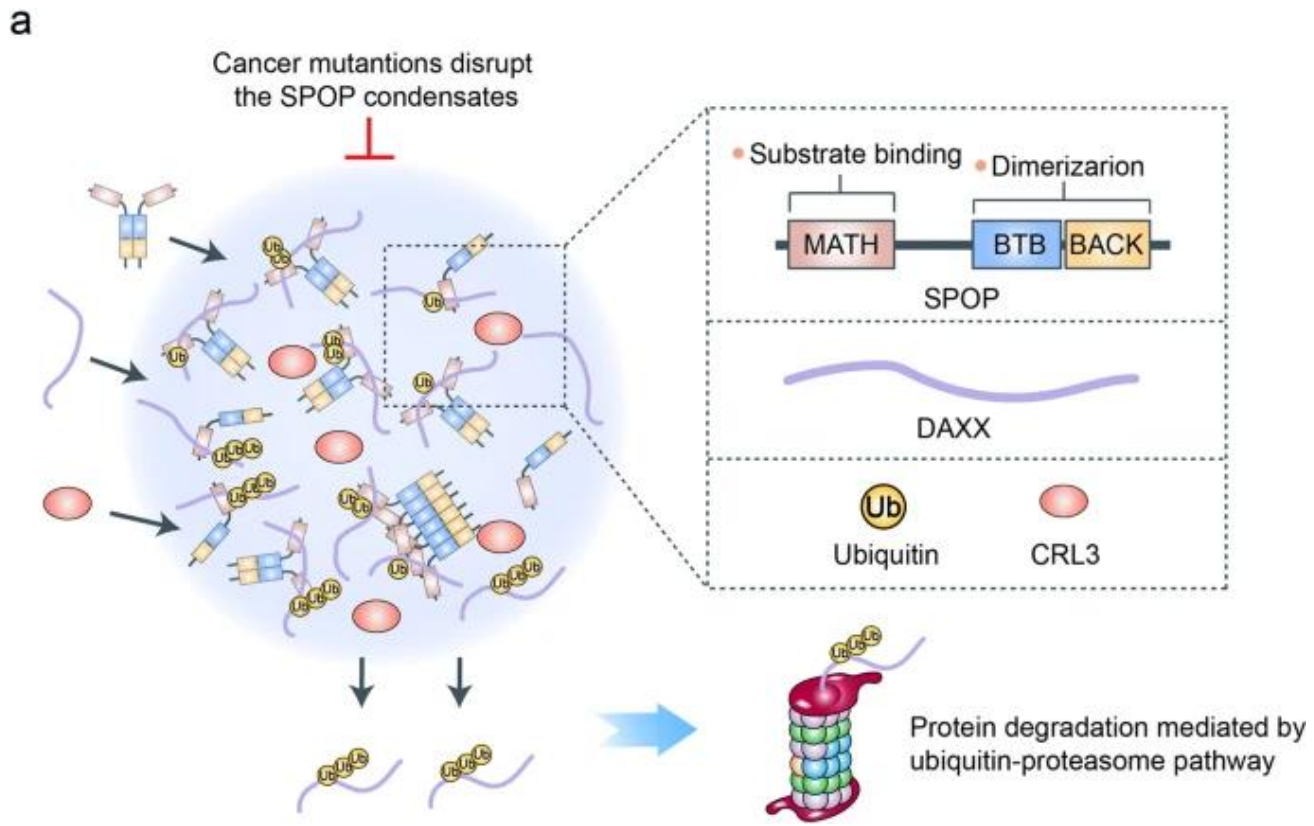


LLPS in neurodegenerative diseases



LLPS in cancer

lots of identifications of specific mutations in proto-oncogenes and tumor-suppressor genes that could cause cancer, but we don't know why!

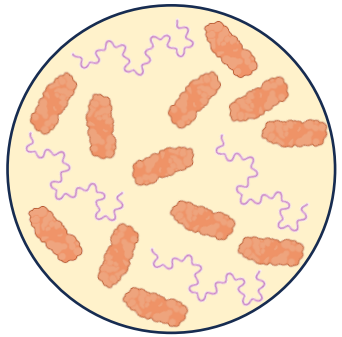


current ongoing RNA therapeutics clinical research to stabilize the condensate even with the mutation!

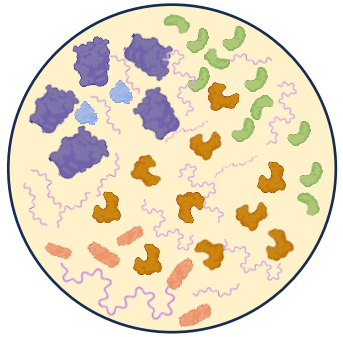
Future challenges and key questions

Complexity of natural condensates

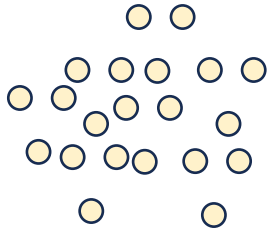
Expectation:
homogenous 1 or 2 phasesystem



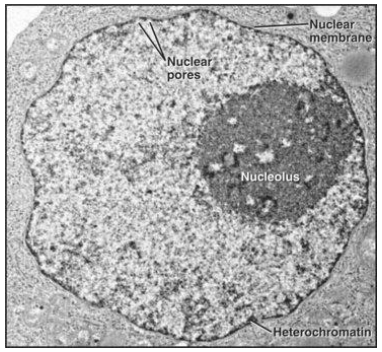
Reality:
up tp 100 different molecules & complex suborganization



The question of size



< 100 nm in size



up to 3,5µm in size!

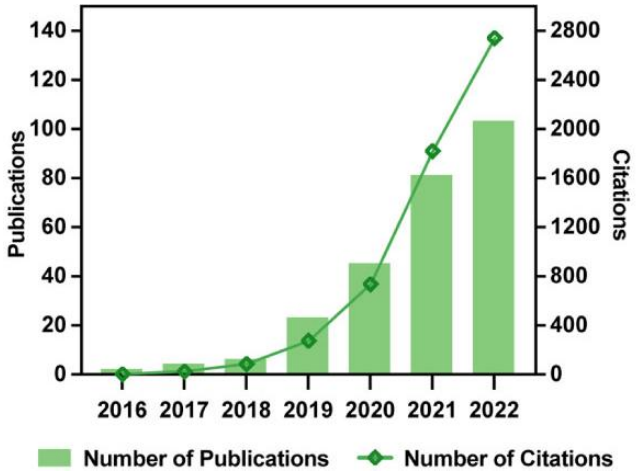
Classifying LLPS sequence features

Arginine & **Glycine**-rich regions

- certain LLPS properties
- RNA binding properties
- more biological functions?

Evolutionary considerations

Conclusions



1. LLPS exists in many biological areas and functions
2. LLPS is extremely complex in its formation and dynamic character and heavily context dependent
3. The importance of LLPS for therapeutic targets