800 ANNI 1222·2022

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
MATEMATICA

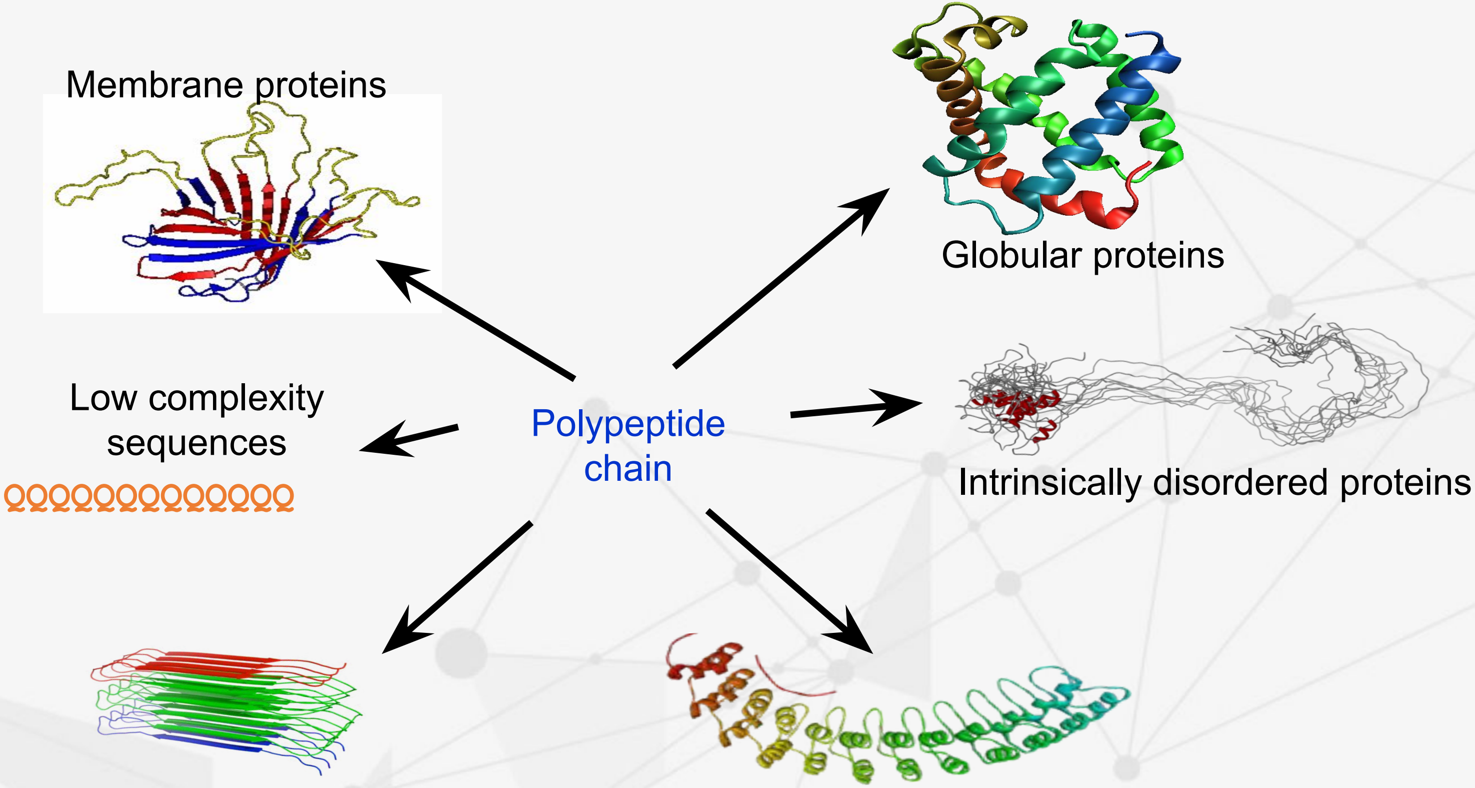
DATA SCIENCE
UNIVERSITY OF PADOVA

NON-GLOBULAR PROTEINS

Master of Science in Data Science

Damiano Piovesan

Non-globular proteins



Polypeptide chain

Membrane proteins

Globular proteins

Low complexity sequences

QQQQQQQQQQQQQ

Intrinsically disordered proteins

Aggregates, amyloids

Proteins with tandem repeats

1222·2022
800 ANNI

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
MATEMATICA

DATA SCIENCE
UNIVERSITY OF PADOVA

INTRINSICALLY DISORDERED PROTEINS

Master of Science in Data Science

Damiano Piovesan

Intrinsically disordered proteins?

BREAKING THE PROTEIN RULES

If dogma dictates that proteins need a structure to function, then why do so many of them live in a state of disorder?

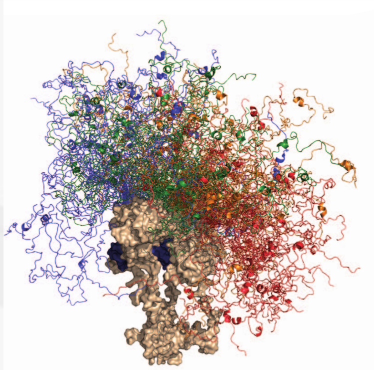
BY TANGUY CHOUARD

Keith Dunker's life is a mess. His desk is so swamped with books, old chocolate bars, half-reviewed manuscripts, pens, diet coke bottles and — somewhere — a stray sock, that he ends up printing papers again rather than wading in to find the original. "I'm so disorganized," he crows, "some people have called me Dr Disorder." But he remembers with great precision the moment that disorder invaded his scientific life. It was 15 November 1995, at 12:40 p.m., halfway through a seminar by crystal-

disordered from beginning to end[1]. This part of the protein universe had largely been ignored because disordered protein segments impede crystal formation — a prerequisite for X-ray diffraction, the predominant way structures are deduced — and structural biologists clip them out whenever they can.

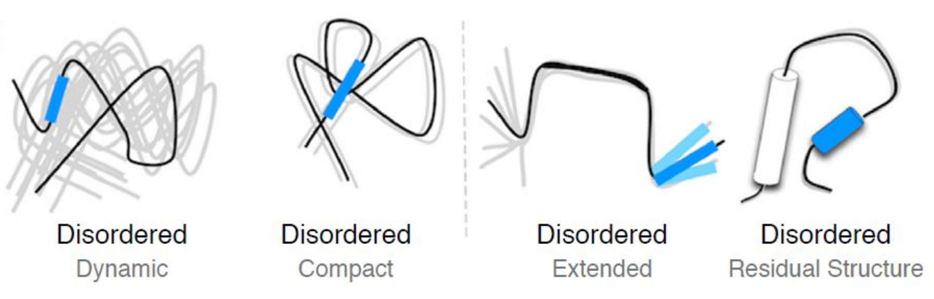
Today, though, "the recognition of disorder has grown dramatically", Peter Wright, a protein biophysicist at the Scripps Research Institute in La Jolla, California, told the American Association for the Advance-

10 MARCH 2011 | VOL 471 | NATURE | 151

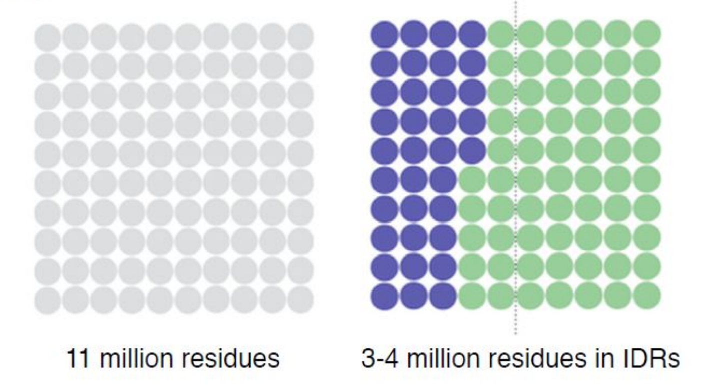


natively denatured	*Schweers et al., 1994*
natively unfolded	*Weinreb et al., 1996*
intrinsically unfolded	*Baskakov et al., 1999*
intrinsically unstructured	*Wright and Dyson, 1999*
intrinsically disordered	*Dunker et al., 2000*
exceptionally flexible	*Ahmed et al., 2007*
natively unstructured	*Schlessinger et al., 2007*
naturally flexible	*Uversky et al., 2009*

Intrinsically Disordered Proteins (IDPs) and Regions (IDRs)



Human proteome



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
MATEMATICA

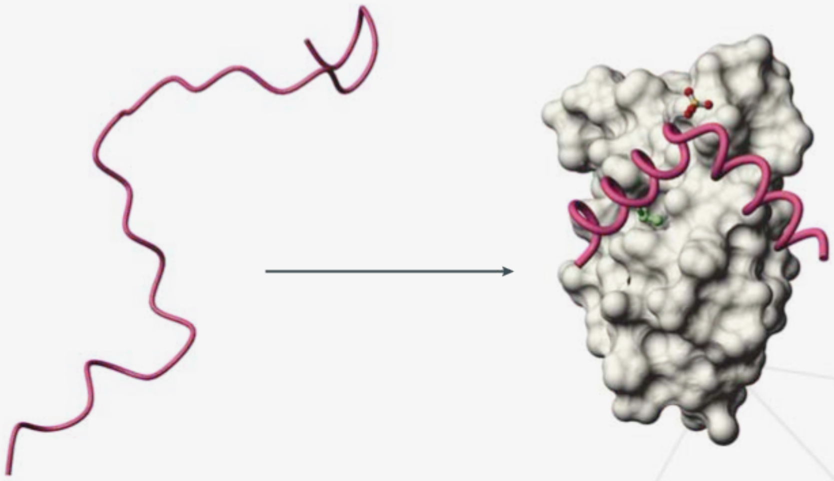
DATA SCIENCE
UNIVERSITY OF PADOVA

IDP FUNCTIONS

Master of Science in Data Science

Damiano Piovesan

Coupled folding and binding



pKID domain of CREB
Phosphorylated kinase-inducible domain of the *transcription factor cyclic-AMP-response-element-binding protein*

KIX domain of CBP
KID-binding domain of the CREB-binding protein

- The unstructured region is **induced** to form a stable structure, but only in the presence of the interacting structural domain
- Such interactions tend to be **transient** and have **low binding affinities**
- Well suited for mediating functions that require a **fast response** to changing stimuli

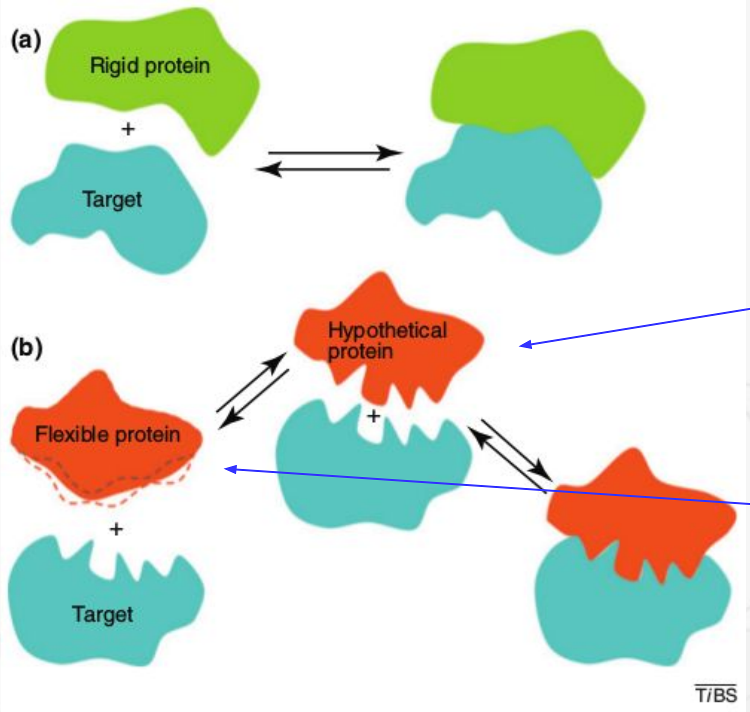
(Dyson & Wright, Nat Rev Mol Cell Biol, 2005)

Induced fit

- Important when the **cellular concentrations** of a regulatory protein and its target are **low**, as is the case for many signaling and transcriptional processes
- Less prone to **environmental stress**, preserve function in unstable conditions, e.g. high temperature
- Example
    - the dynamic interaction between a **kinase** and its **phosphorylation site**
    - the kinase domain needs to interact with other substrates

Miller M, Curr Protein Pept Sci, 2009, 10(3):244-69
Shoemaker et al. 2000

Induced fit energy



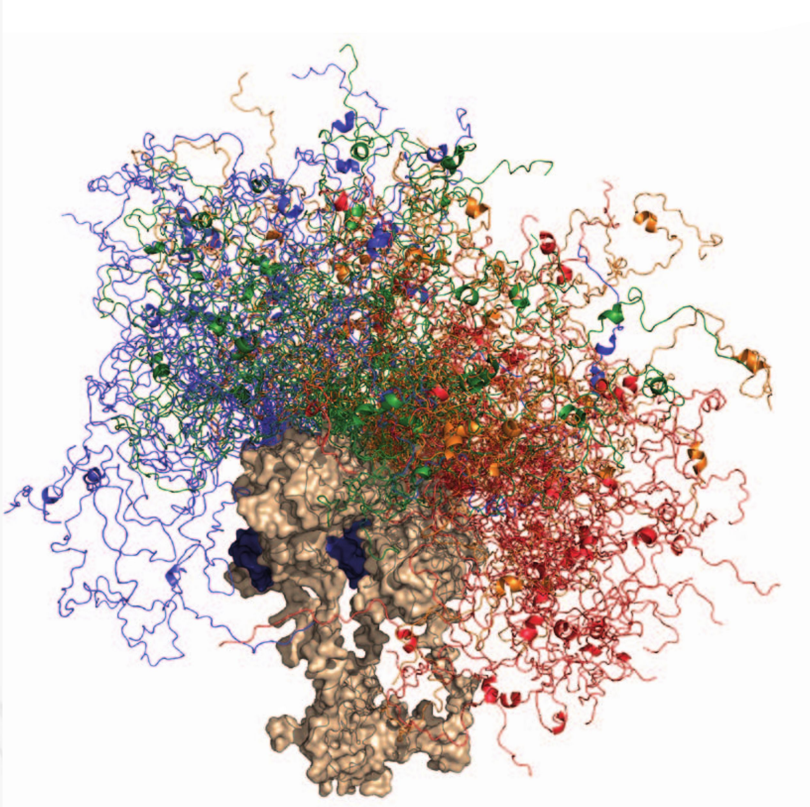
A. Rigid chain. Simple, **smooth** interaction surface
Low specificity, high affinity

B. Disordered **dynamic** chain
High specificity, low affinity (versatility)

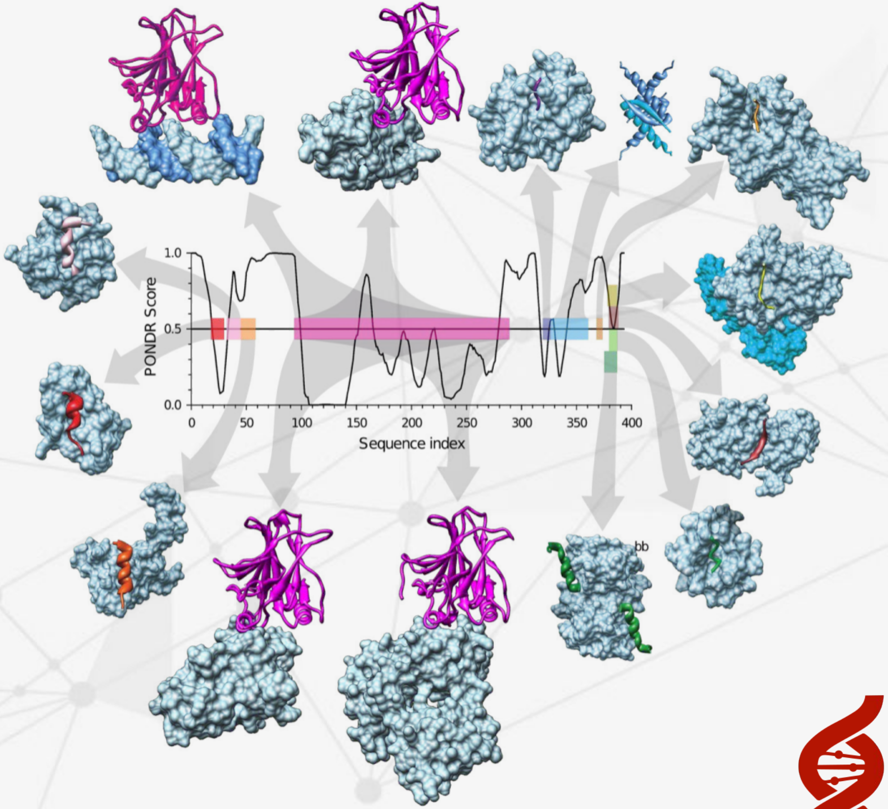
- A hypothetical disordered protein with the bound conformation even when unbound
  - Unlikely →High free energy to preserve the folding
- The flexible protein adapt to protrusions and indents
  - **Large surface** area of interaction
  - **Entropic cost** of folding
  - **Versatility**
  - *High specificity, low affinity*

(Zhou HX, Trends Biochem Sci. 2012)

Versatility

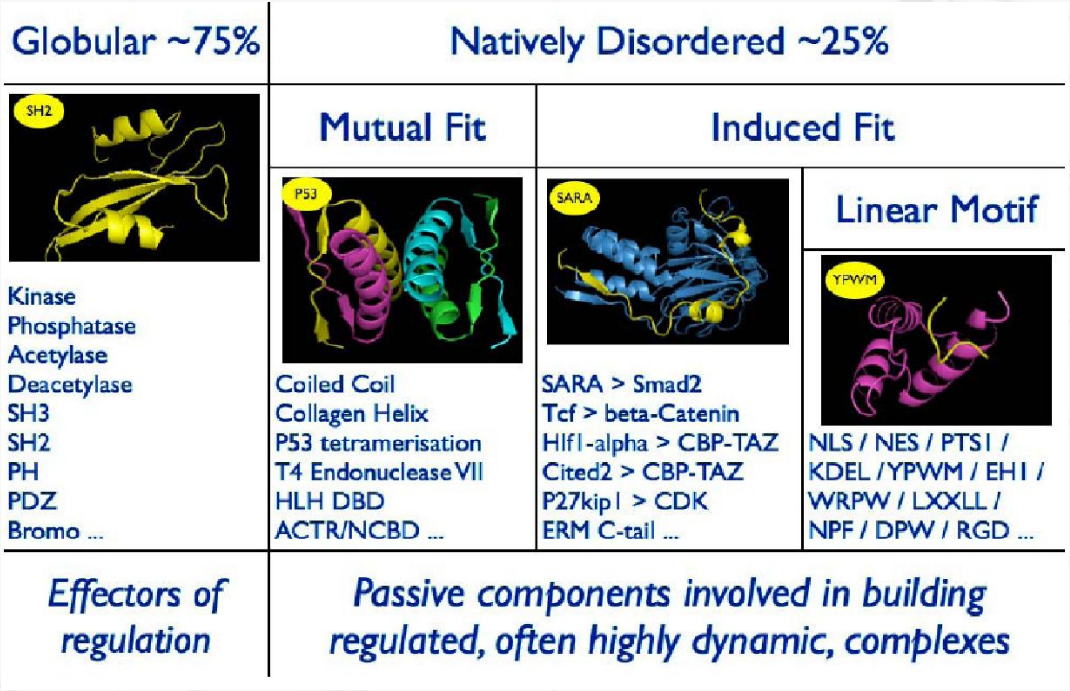


p53



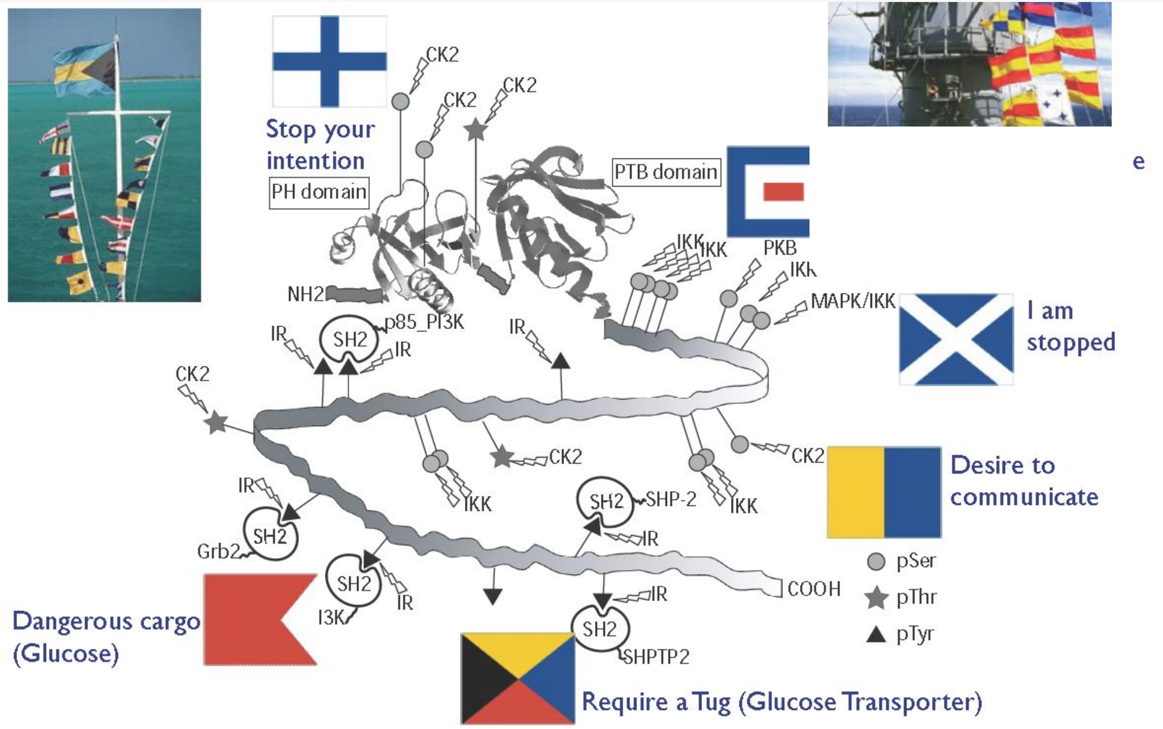
Uversky & Dunker, Biochim Biophys Acta, 2010

Protein Architecture Modules in Cell Regulation



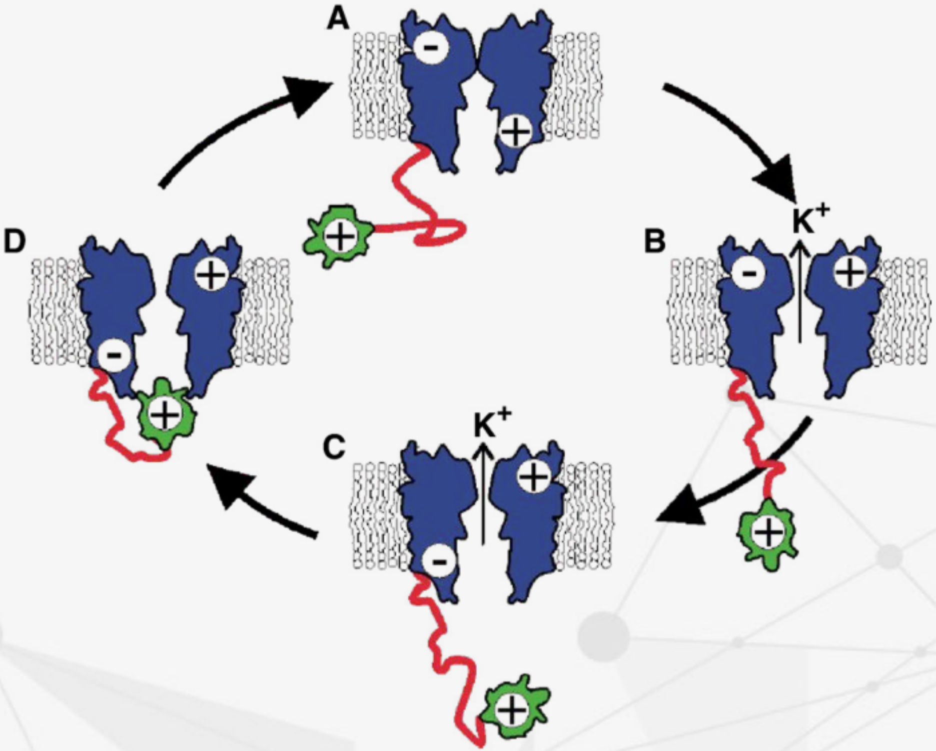
Globular ~75%	Natively Disordered ~25%		
	Mutual Fit	Induced Fit	
			Linear Motif
Kinase Phosphatase Acetylase Deacetylase SH3 SH2 PH PDZ Bromo ...	Coiled Coil Collagen Helix P53 tetramerisation T4 Endonuclease VII HLH DBD ACTR/NCBD ...	SARA > Smad2 Tcf > beta-Catenin HIf1-alpha > CBP-TAZ Cited2 > CBP-TAZ P27kip1 > CDK ERM C-tail ...	NLS / NES / PTS1 / KDEL / YPWM / EH1 / WRPW / LXXLL / NPF / DPW / RGD ...
Effectors of regulation	*Passive components involved in building regulated, often highly dynamic, complexes*		

Molecular flagpoles

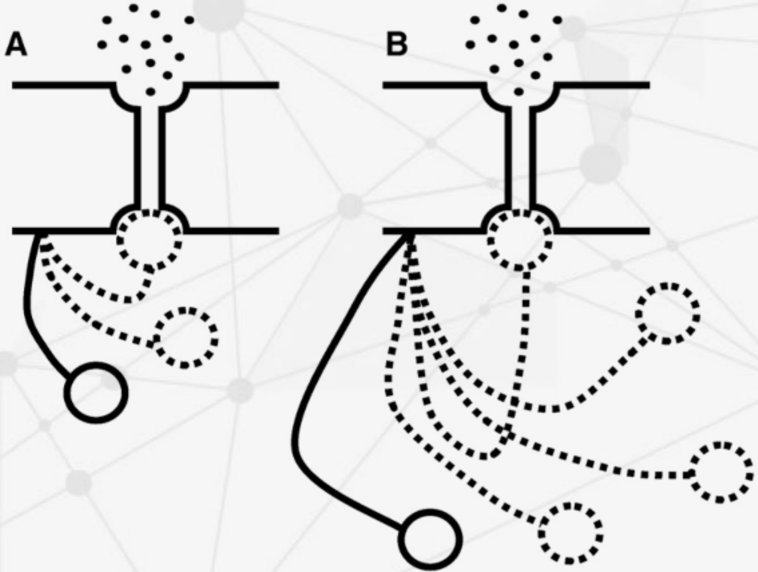


The Insulin Receptor Substrate (IRS) protein

“Ball and chain” timing mechanism

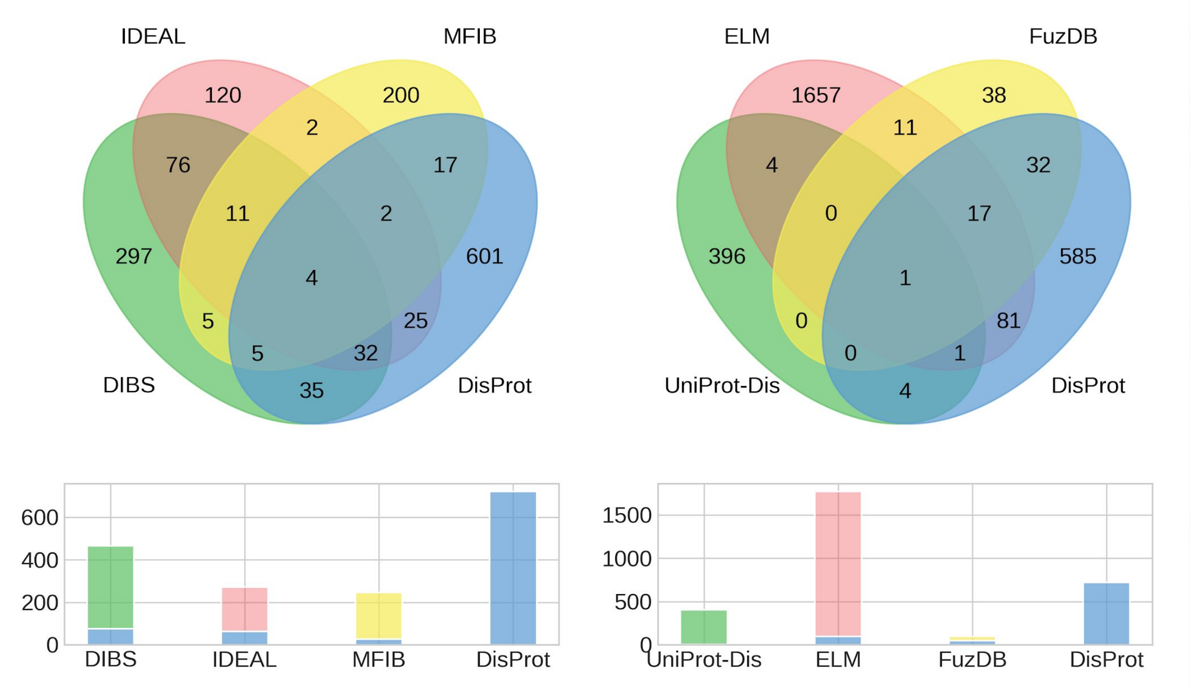


Shaker-type voltage-gated K+ ion channel

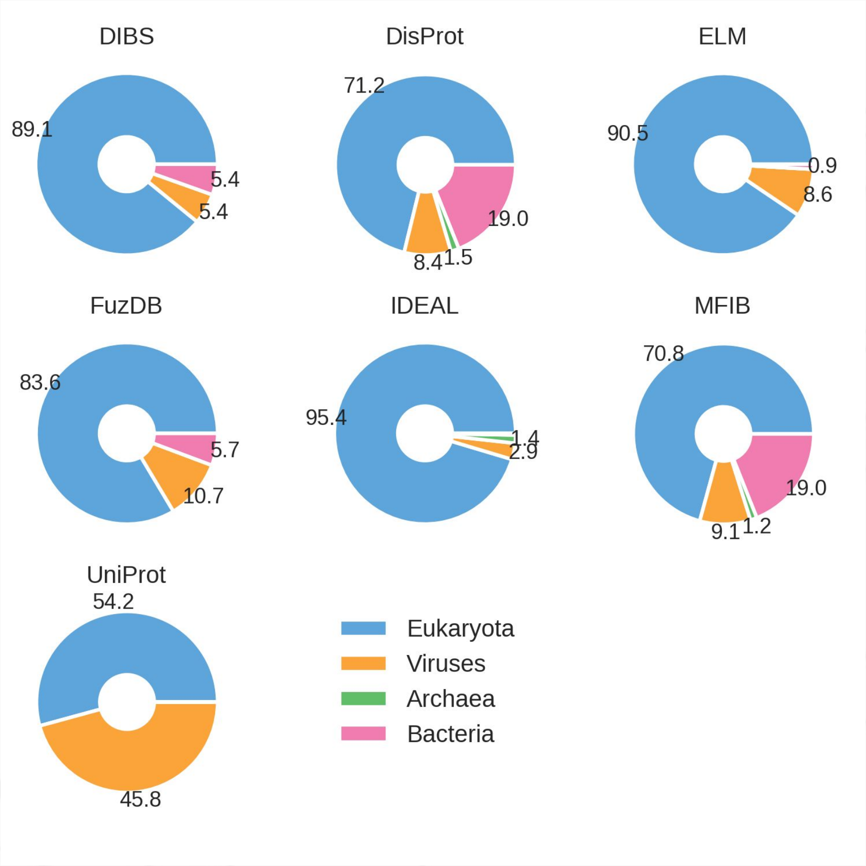


(Uversky & Dunker, Biochim Biophys Acta, 2010)

Manually curated disorder databases



Taxonomy


DIBS
89.1
5.4
5.4
DisProt
71.2
19.0
8.4
1.5
ELM
90.5
0.9
8.6
FuzDB
83.6
5.7
10.7
IDEAL
95.4
1.4
2.9
MFIB
70.8
19.0
9.1
1.2
UniProt
54.2
45.8
Eukaryota
Viruses
Archaea
Bacteria

Nuclear magnetic resonance (NMR)

Master of Science in Data Science

Damiano Piovesan

Crystallization (X-ray) bias

- Isolation and **homogenization** of the cells
- Protein **activity** testing
- **Fractionation**, chromatography and/or gel filtration
- Fractions tests and **purification** of the active protein
- Active protein is **sequenced**
- The **three-dimensional structure** is determined

Biased towards folded proteins

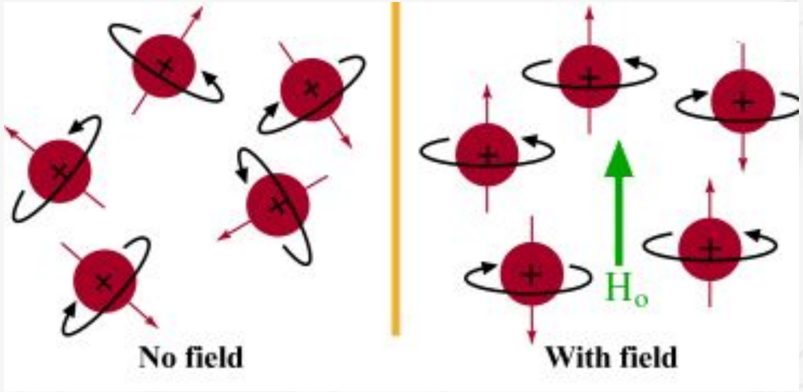
- Released **proteases** digest unfolded regions
- **Regulatory proteins** are present in few copies and their activity is difficult to assay

NMR spectroscopy

- **Proton** Nuclear Magnetic Resonance (paramagnetic isotopes: ^{1}H, ^{15}N, ^{19}F, ^{13}C, ^{31}P)
- Based on **absorption/relay** of **electromagnetic radiation by nuclei**
- Radio frequency 4 – 900 MHz
- Gives information about the **number and relations of nuclei** in a compound

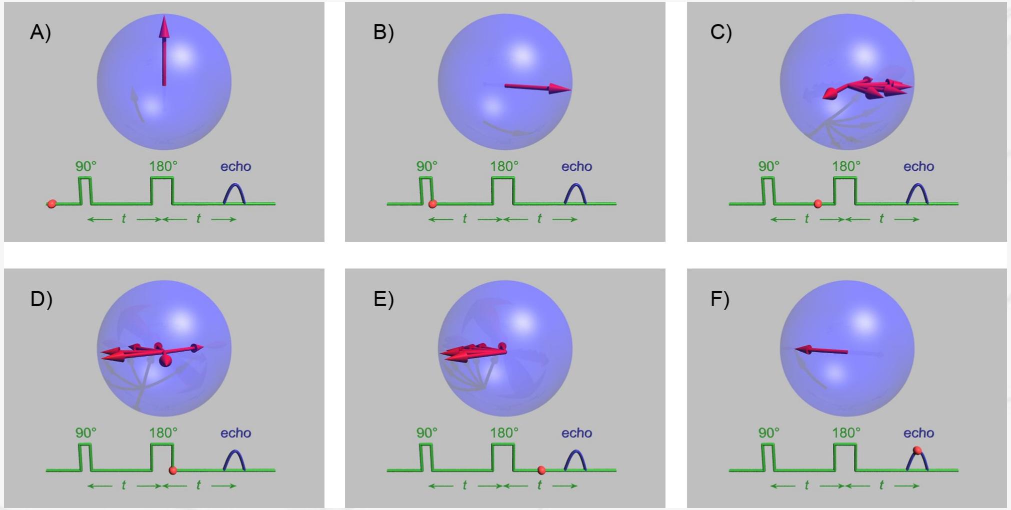
Principles

- **Without** an external magnetic field, the **nuclei spins are random** in direction
- **With** an external magnetic field, the **nuclei align** themselves (with or against the field)



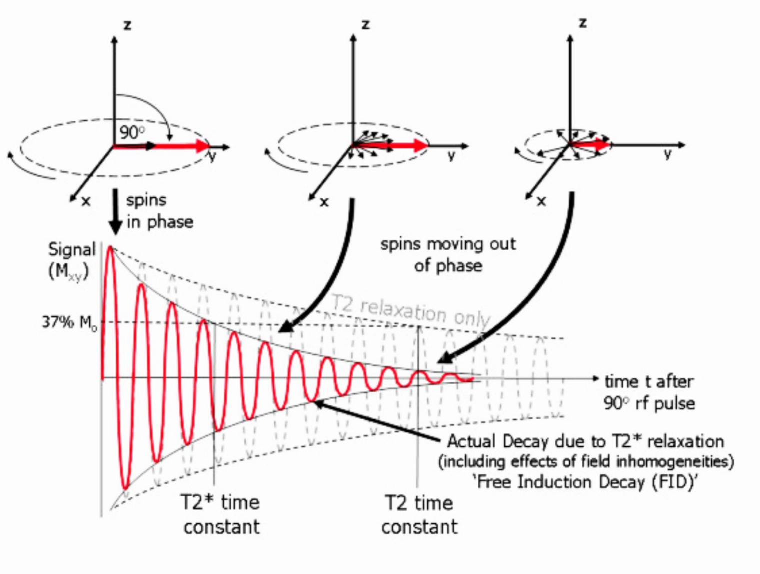
Principles

- From excited to ground state **nuclei emit energy** (**relaxation**)
- High electron density around a nucleus **shields** the nucleus and affect relaxation (field inhomogeneities)



Principles

Free Induction Decay (**FID**) - Relaxation

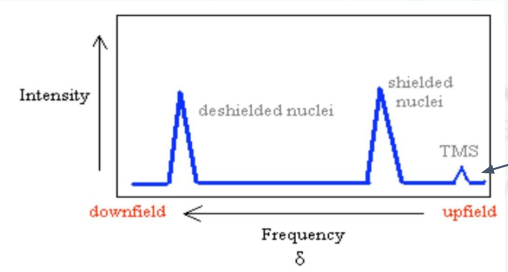


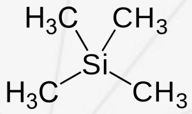
Principles

The **relaxation frequency** is transformed to provide a **NMR spectrum** (**peaks**)

Feature in Time Domain (FID)	Feature in Frequency Domain (Spectrum)
Oscillation Frequency	Peak Position (Chemical Shift)
Slow Decay (Long T_2)	Narrow, Sharp Peak
Fast Decay (Short T_2)	Broad, Flat Peak

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}} ,$$





TMS - tetramethylsilane

Reference compound

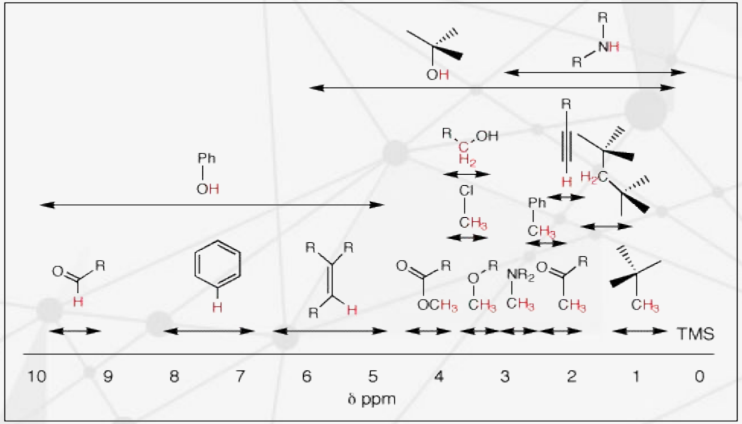
Chemical shift

Depending on the **environment** signals are **upfield** (or **downfield**) in the spectrum

Position of intensity peaks are compared with a **reference** (**δ, chemical shift**)

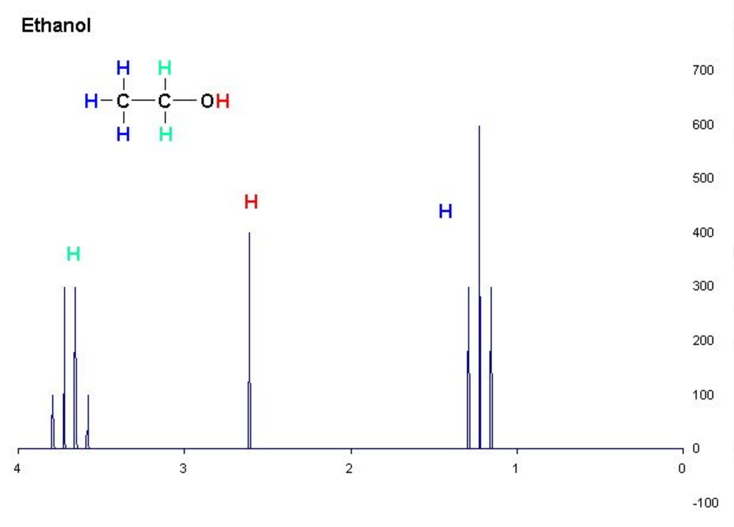
Delta is affected by

- Electronegative groups (deshielding)
- Hydrogen bonding (deshielding)
- Magnetic anisotropy of π-systems (different values along different directions)



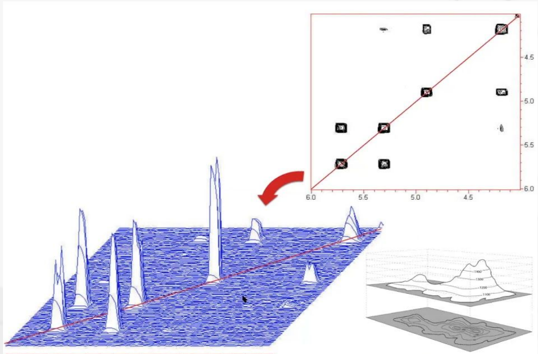
Chemical shift

- **Number of blocks** → different kinds of protons
- **Position** → magnetic environment (shielding)
- **Intensity** → number of protons
- **Splitting** (spin-spin coupling) → number of nearby nuclei (only non equivalent protons couple)

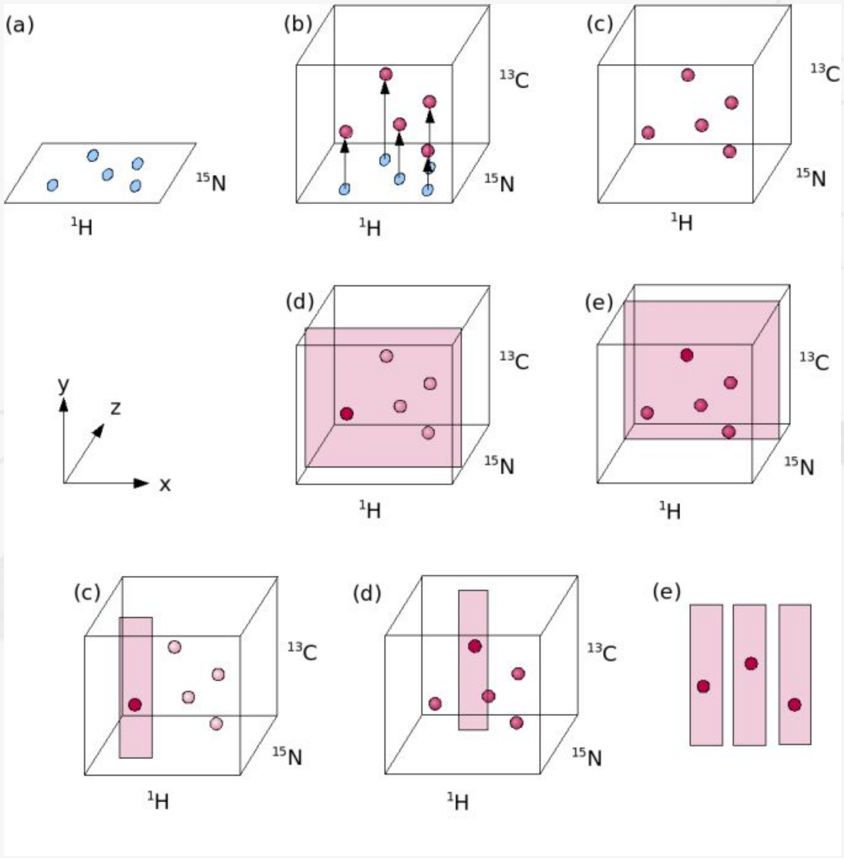


Multi-dimensional NMR

- **Cross-peaks** correspond to **nuclear coupling** (connected protons of the same type)
- Combine multiple different experiments (add 3rd dimension using different isotopes, ex. ^{13}C)
- **Strips** identify similar (or same) amino acids

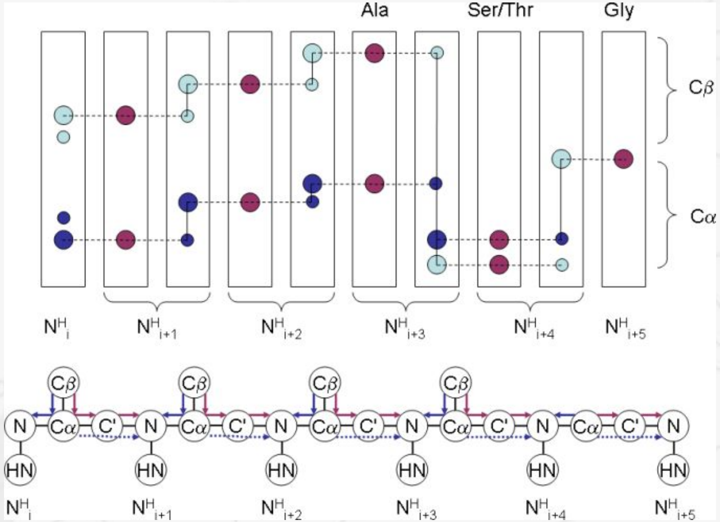


https://www.acornnmr.com/codeine/cosy/
https://www.greenbelly.co/pages/contour-lines



NMR assignment

- Combining additional experiments it is possible to **assign backbone positions**
- One experiment to identify spectra of the different amino acids in the sequence (previous slide)
- Another experiment(s) to track *i* → *i - 1* relationships



BMRB database

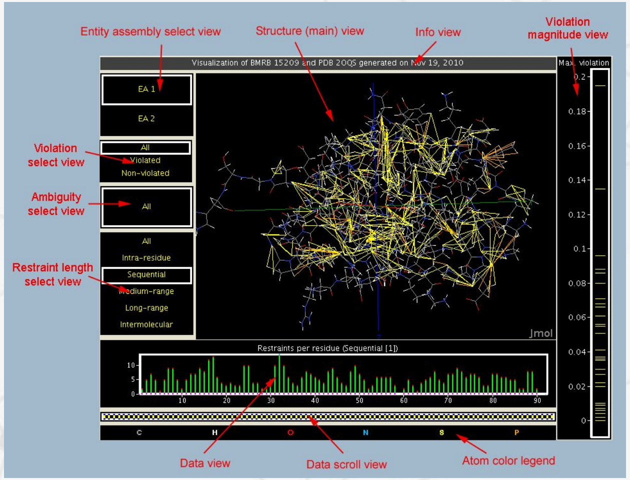
BMRB makes bio-NMR data FAIR.
Findable, Accessible, Interoperable, Re-usable.

BMRB collects, annotates, archives, and disseminates spectral and quantitative data derived from NMR spectroscopic investigations of biological macromolecules and metabolites.

Recently released at BMRB:

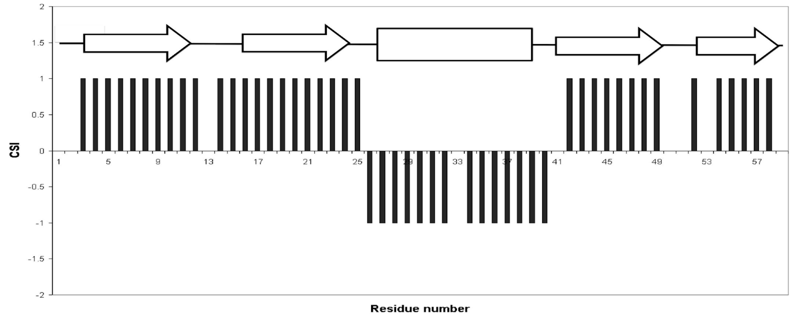
Entry 50987: NMR solution structures of d(GGCCTG) n repeats associated with spinocerebellar ataxia type 36

https://bmrb.io



Chemical Shift Index (CSI) predictor

- Chemical shifts ($^1H_\alpha$) in **random coil** for all **amino acid types** as reference
- **Helices** shifted upfield (+1)
- **Beta strands** shifted downfield (-1)

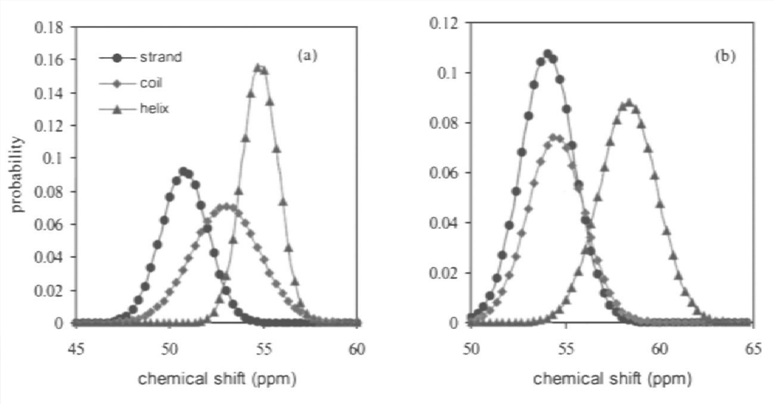


Wishart D. S., 1992

Secondary Structure Probability (SSP) predictor

Delta distribution for all 20 amino acids for different types of NMR

- (a) Strand Vs Coil
- (b) Helix Vs Coil

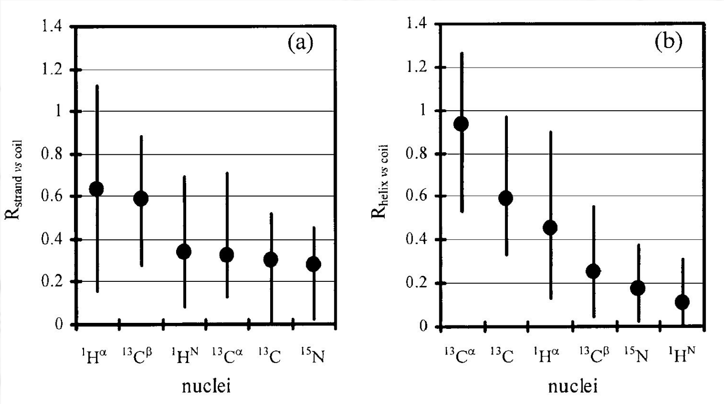


ALA

MET

Strand vs coil

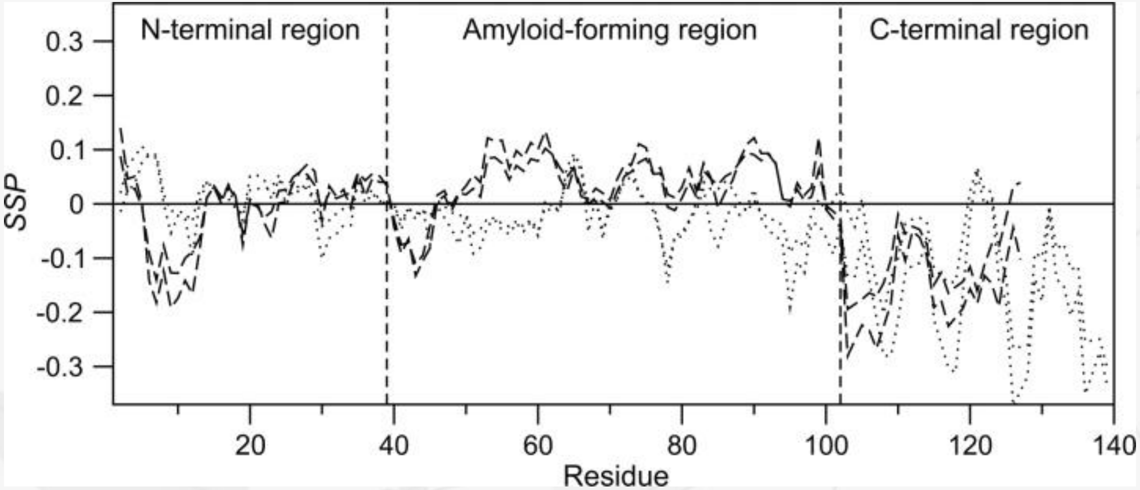
Helix vs coil



WANG Y. AND JARDETZKY O. Protein Science (2001)

Secondary structure propensity (SSP)

- **α-synuclein** (dotted lines), **γ-synuclein** (dashed lines)
- **Positive** values represent **α-structure**, **negative** values represent **β-structure**
- Different lines are $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts



Marsh, Singh, Jia, Forman-Kay. Protein Sci. 2006