JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Intrinsically Disordered Proteins

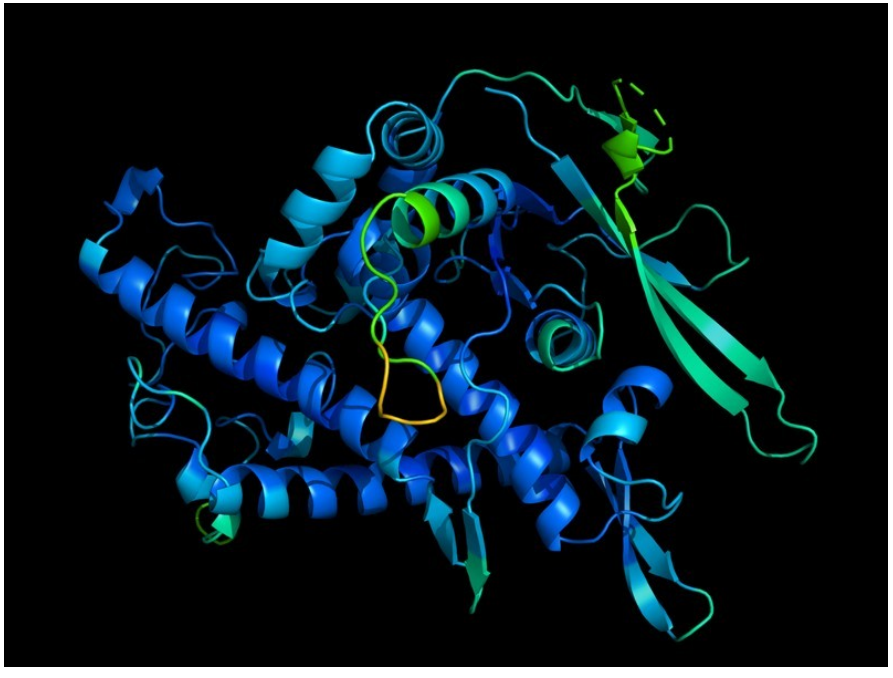
Miguel Andrade
Kristina Kastano
Mariane Gonçalves-Kulik

Faculty of Biology,
Johannes Gutenberg University
Mainz, Germany
magoncal@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.

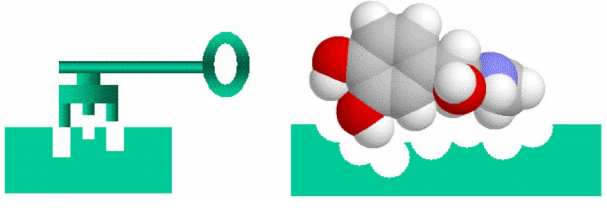
Representation of the 3D structure of a protein



Structure-function paradigm:

Structure → function

Emil Fisher's key-lock model



RCSB PDB | Deposit | Search | Visualize | Analyze | Download | Learn | More | Documentation | Careers | MyPDB

RCSB PDB PROTEIN DATA BANK

188085 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB Archive

Advanced Search | Browse Annotations | Help

PDB-101 | Worldwide PDB | EMDataResource | NDB Nucleic Acid Database | Worldwide Protein Data Bank Foundation

Developers: Join the RCSB PDB Team — Explore Open Positions

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

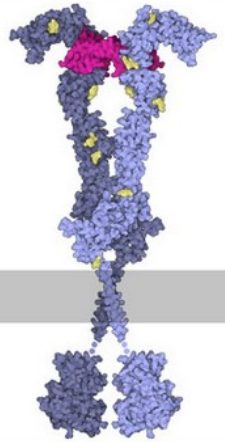
A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

March **Molecule of the Month**



Vascular Endothelial Growth Factor (VegF) and Angiogenesis

Latest Entries

As of Tue Mar 08 2022

7MM8

Features & Highlights

Streamline Search Results

Use the Group option to simplify results based on sequence identity, UniProt ID, and group depositions.

Search for Structures or Feature Help, News, and PDB-101

Use the pulldown menu to target queries of the *PDB archive* or *Documentation*

Deposition of Half-maps for Certain EM Entries Now Mandatory

Deposition of half-maps for single-

News | Publications

Register for Exploring Structural Database Use in Crystallography

The U.S. National Committee for Crystallography of the National Academies of Sciences, Engineering, and Medicine is providing an online workshop series starting March 21 for students and researchers

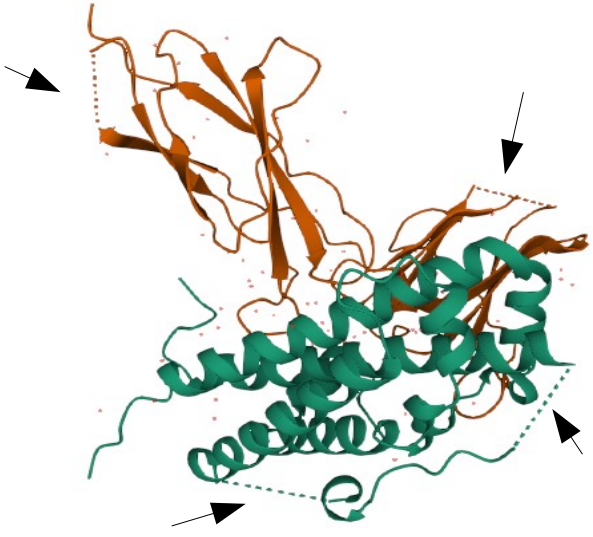
» *03/08/2022*

Molecular Landscapes in The New Yorker

JCVI-syn3A Minimal Cell is highlighted at

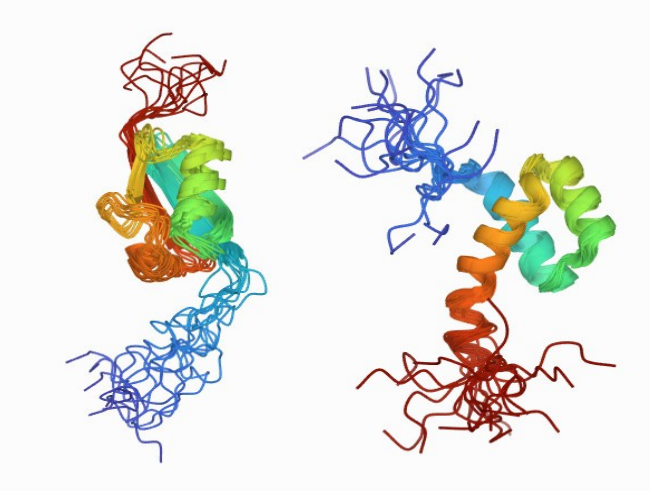
What does it look like?

Missing electron densities in X-ray crystallography from PDB

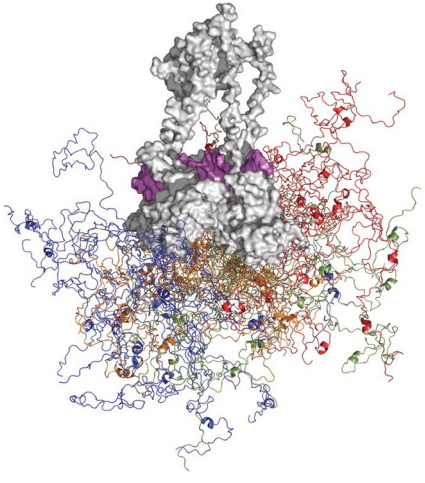


Human Growth hormone bound to receptor

NMR (nuclear magnetic resonance) ensembles of SUMO-1 and antennapedia from PDB



Model of tumor suppressor p53 using X-ray, NMR and SAXS



Disordered in 40% of its length!

Wells et al 2008

~30% of PDB structures have such regions!

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-

→ Argues that disordered regions in themselves have important functions!

Disordered region functions

- Flexible linkers/spacers between domains
- Entropic chains (contribute to the structure energy)
- Molecular recognition:
  - binding to proteins, nucleic acid polymers, membrane, metal ions
  - As enzymes that undergo disorder-to-order transitions
  - Formation of multiprotein complexes
- Protein modifications and regulation (e.g. phosphorylation)

See Dunker et al 2002 for more details

Why study IDRs?

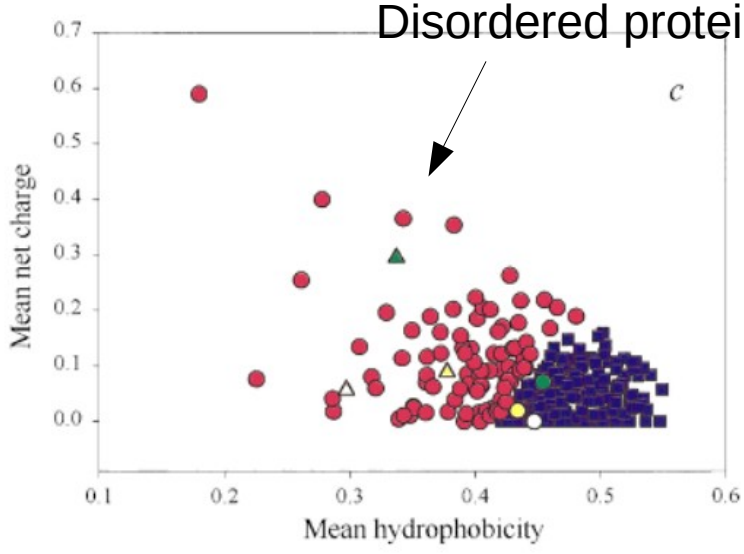
- Involved in diseases:
  - Cancer
  - Neurodegenerative diseases (Parkinson’s, Dementia, Alzheimer’s)
  - Diabetes
  - Cardiovascular diseases
- Can be used in drug delivery (synthetic IDRs);
- Understanding of protein complexes interactions.

Revised lock-and-key model for IDPs

Disordered proteins can bind to many structured partners!

Disordered sequences

- Low content of bulky hydrophobic amino acids (Val, Leu, Ile, Phe, Trp, Tyr, Met)
- High content of charged and polar Ser, Pro, Glu, Lys)



Used in disorder predictor FoldIndex

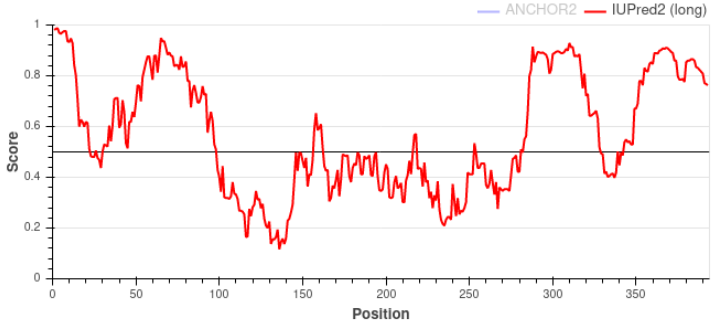
Uversky et al 2000

Disorder prediction

Prediction methods based on

- Physical/chemical features (FoldIndex)
- Machine learning algorithms (DISOPRED2, Spritz, PONDR)
- Energy estimation (IUPred2):

  Globular proteins form many favorable interactions to ensure the stability of the structure.

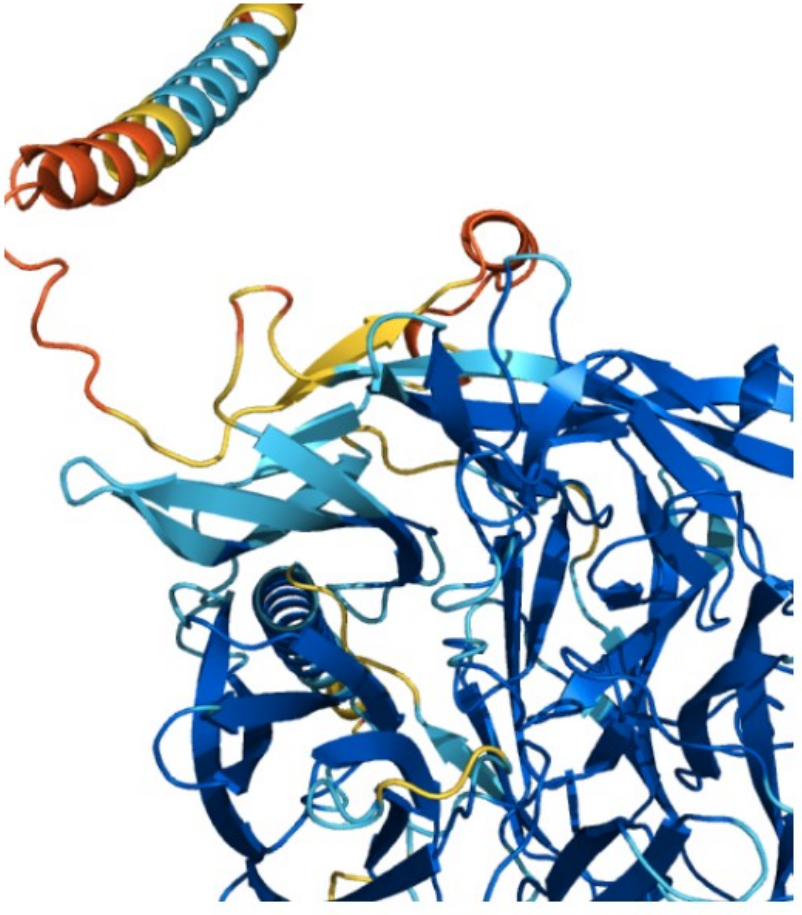


IUPred output for p53

What about order prediction?

AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.

DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) have partnered to create AlphaFold DB to make these predictions freely available to the scientific community. The latest database release contains over 200 million entries, providing broad coverage of UniProt (the standard repository of protein sequences and annotations). We provide individual downloads for the human proteome and for the proteomes of 47 other key organisms important in research and global health. We also provide a download for the manually curated subset of UniProt (Swiss-Prot).



Q8I3H7: May protect the malaria parasite against attack by the immune system.
Mean pLDDT 85.57.

View protein

Disorder databases

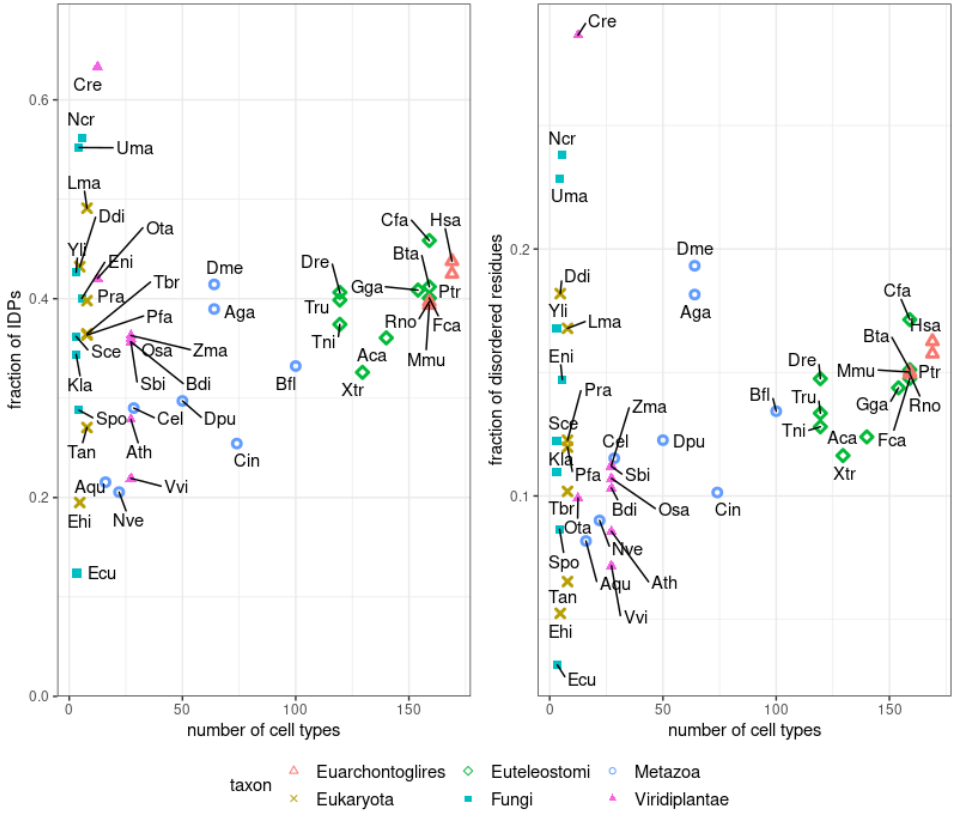
- Disprot: 

  database of almost 3.4k experimentally verified IDPs
- IDEAL: database of experimentally verified 995 IDPs
- MobiDB: 

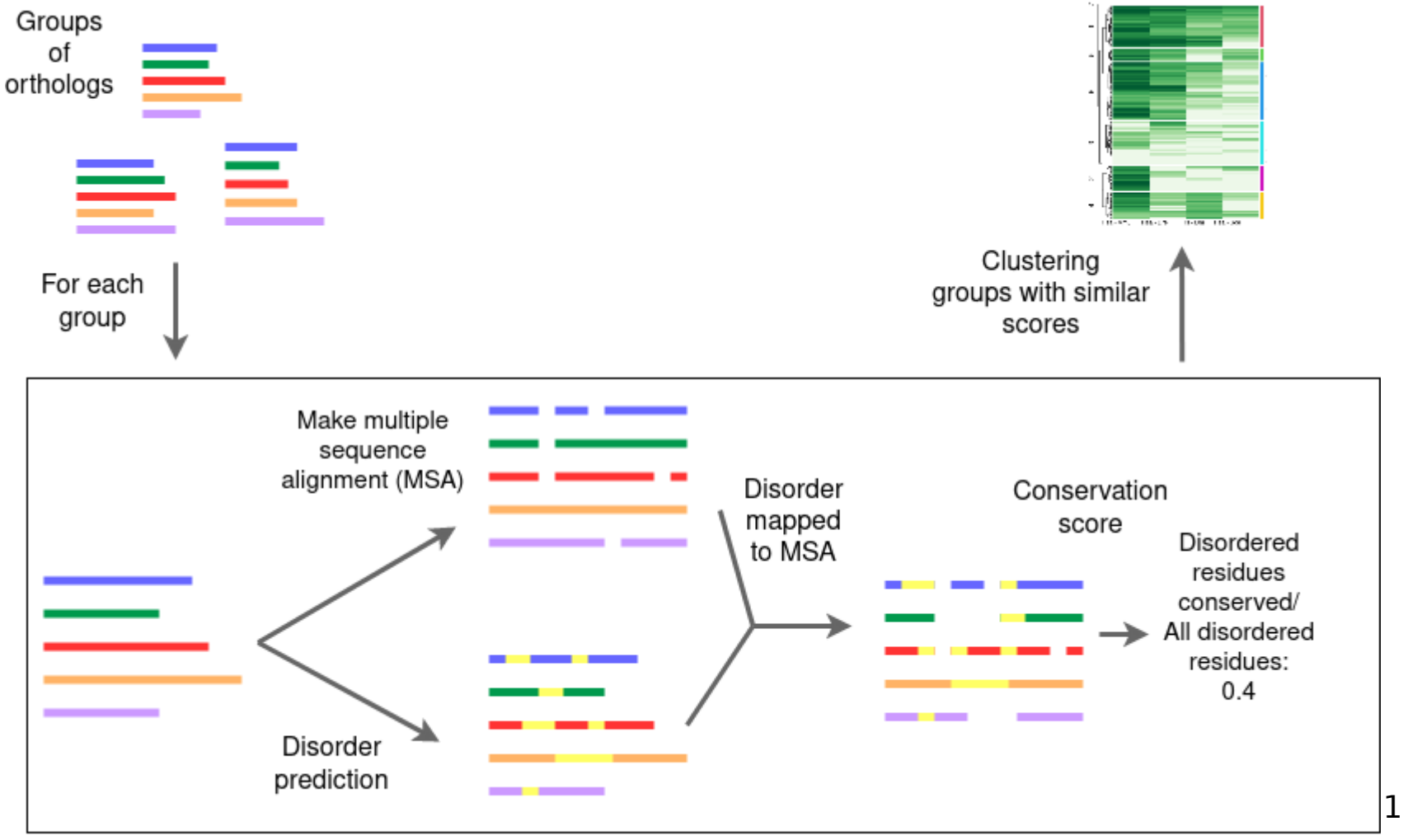
  centralized resource that combines experimental and predicted data into a consensus annotation
- DIBS: Disordered Binding Sites (DIBS) with 1,576 complexes with curated interactions on the IDR region.

Natural abundance and phylogenetic distribution

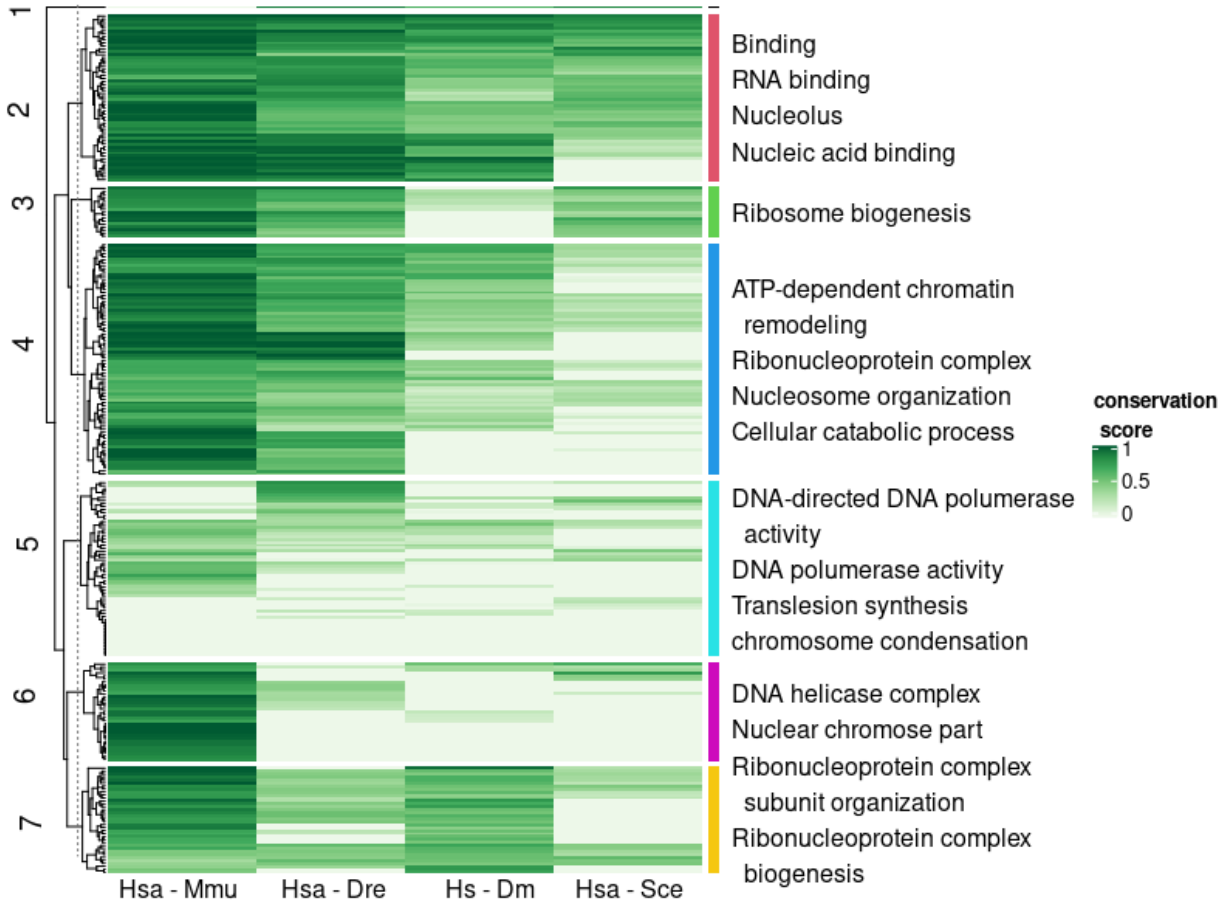
- ~40% of human proteins predicted to be IDPs
- ~30% of Eukaryotic proteins predicted to be IDPs



Evolutionary study of disorder



Evolutionary study of disorder

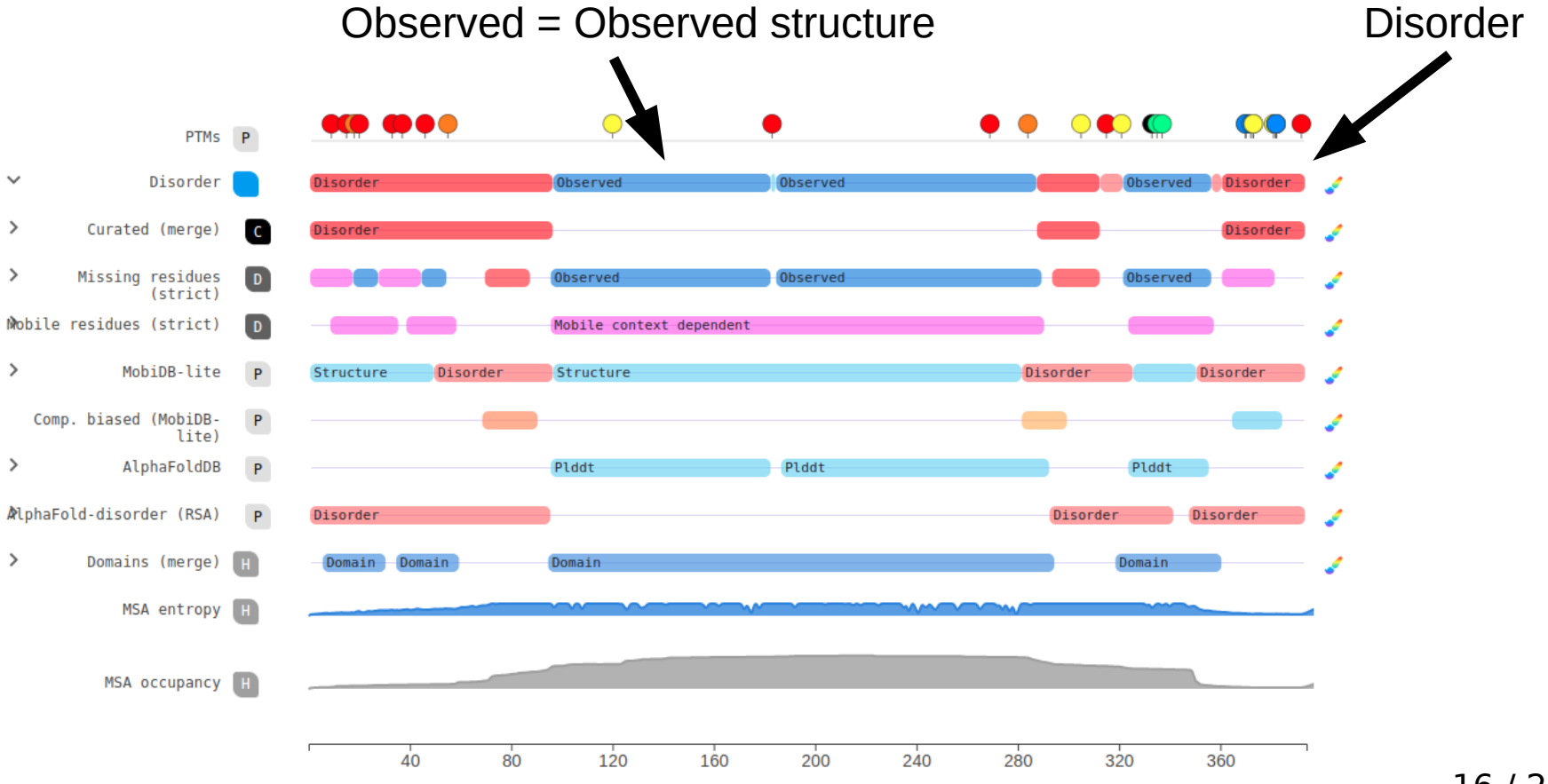


Correlating
similar disorder
conservation patterns
with protein functions

Kastano et al 2020

MobiDB https://mobidb.bio.unipd.it/

Database that provides information about intrinsically disordered regions (IDRs) and related features from various sources and prediction tools.



P04637 - Cellular tumor antigen p53

MobiDB advanced view

MobiDB new Alphafold disorder

Exercises

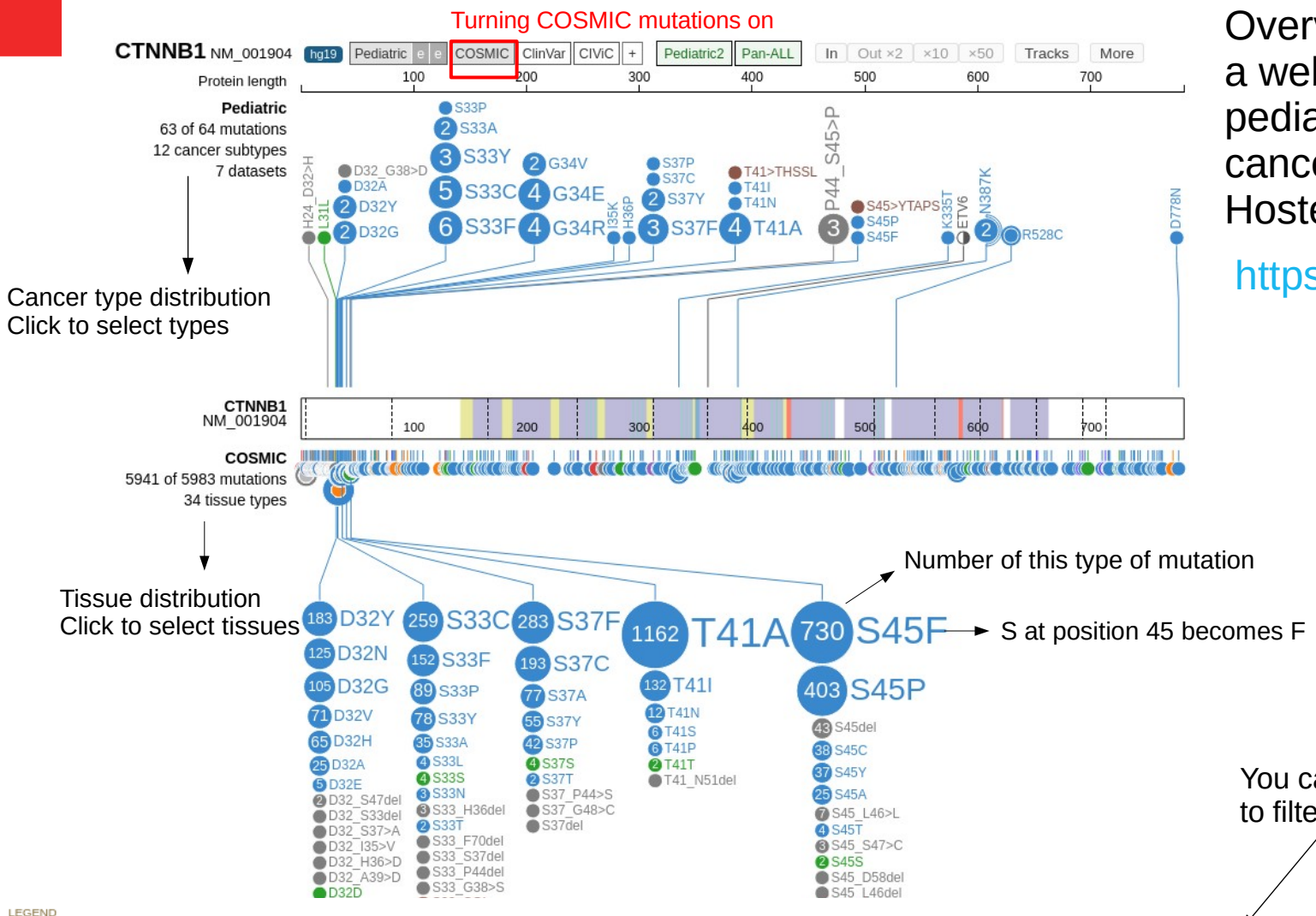
In the next 4 slides there are some exercises using some databases and webtools. Answer each of them and submit your answers at:

https://take.quiz-maker.com/QJRLDKS59

Exercise: MobiDB and DisProt

- Search beta-catenin-1 (P35222) on MobiDB. Go to the advanced view.
  - What kind of disorder annotations are there for this protein?
  - The different annotations don’t give exactly the same results. Which annotation gives the longest IDRs (Intrinsically Disordered Regions)?
- Go to the Disprot entry linked on the bottom.
  - How many types of methodologies (indicated as “Evidence”) were used to identify the disorder?

Exercise: IDPs and cancer mutations



Overview of proteinpaint, a webtool to explore pediatric and adult cancer mutations. Hosted at:

https://pecan.stjude.cloud

Exercise: IDPs and cancer mutations

DIBs: Database of Disordered Binding Sites

- General Information
- Function and Biology
- Structure Summary
  - Chain A
  - Chain B
- Evidence
- Related Structure(s)

Click to go to evidence for structural status

Domain Type:
Bromodomain

Structure Summary

Entry contents: 2 distinct polypeptide molecules

Chains: A, B

Notes: *No modifications of the original PDB file.*

Chain A

Structural status

Name: Cellular tumor antigen p53 **Disordered** **Confirmed**

Source organism: *Homo sapiens*

Length: 20 residues

Sequence: SHLKSKKGQSTSRHKKLMFK

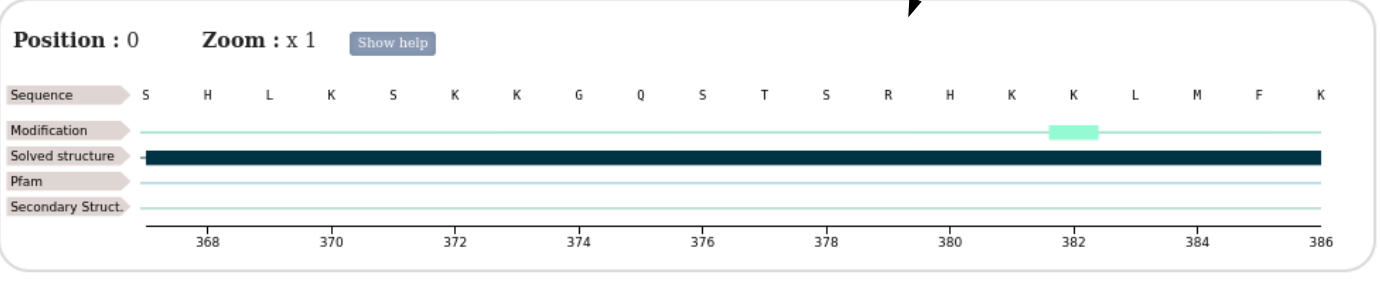
The sequence contains the following modified/non-standard residues:

- N(6)-acetyllysine (K) at position 382 (PDB position: 382)

UniProtKB AC: P04637 (positions: 367-386) **UniProt** Coverage: 5.1%

UniRef90 AC: UniRef90_P04637 (positions: 367-386) **UniRef90**

Residues of this chain involved in binding



Chain B

Name: CREB-binding protein **Ordered**

Source organism: *Homo sapiens*

Length: 121 residues

Exercise: IDPs and cancer mutations

- Go on PECAN at https://pecan.stjude.cloud and search for beta-catenin (CTNNB1).Where in the sequence (residue number) are most of the mutations located in pediatric cancers?
- In what type of cancer are the mutations common?
- Turn on COSMIC (Catalogue Of Somatic Mutations In Cancer) mutations. In what location are mutations most common?
- Go on DIBS (Database of Disordered Binding Sites) at http://dibs.enzim.ttk.mta.hu/search.php and search for beta-catenin (P35222). Which entry contains the mutated region?
- Highlight the most mutated positions in the structure. Are they structured? What evidence is there for their status (ordered/disordered)?
- What is the binding partner?

References

Wright, Peter E, and H.Jane Dyson. “Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm.” Journal of Molecular Biology 293, no. 2 (October 1999): 321–31. https://doi.org/10.1006/jmbi.1999.3110.

Uversky VN, Gillespie JR, Fink AL. “Why are "natively unfolded" proteins unstructured under physiologic conditions?” Proteins. 2000 Nov 15;41(3):415-27. PMID: 11025552.

Uversky, Vladimir N. “A Decade and a Half of Protein Intrinsic Disorder: Biology Still Waits for Physics.” Protein Science: A Publication of the Protein Society 22, no. 6 (June 2013): 693–724. https://doi.org/10.1002/pro.2261.

Mark Wells, Henning Tidow, Trevor J. Rutherford, Phineus Markwick, Malene Ringkjobing Jensen, Efstratios Mylonas, Dmitri I. Svergun, Martin Blackledge, Alan R. Fersht. “Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain” Proceedings of the National Academy of Sciences Apr 2008, 105 (15) 5762-5767; https://doi.org/10.1073/pnas.0801353105

Dunker, A. Keith, Celeste J. Brown, J. David Lawson, Lilia M. Iakoucheva, and Zoran Obradović. “Intrinsic Disorder and Protein Function.” Biochemistry 41, no. 21 (May 28, 2002): 6573–82. https://doi.org/10.1021/bi012159+.

Kastano, Kristina, Gábor Erdős, Pablo Mier, Gregorio Alanis-Lobato, Vasilis J. Promponas, Zsuzsanna Dosztányi, and Miguel A. Andrade-Navarro. “Evolutionary Study of Disorder in Protein Sequences.” Biomolecules 10, no. 10 (October 6, 2020): 1413. https://doi.org/10.3390/biom10101413.

Uversky, Vladimir N. Intrinsically Disordered Proteins. SpringerBriefs in Molecular Science. Cham: Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-08921-8.