Charles University

Faculty of Science

**Study programme:** Special Chemical and Biological Programmes

**Field of study:** Molecular Biology and Biochemistry of Organisms


Michal Hub

# Recognition by intrinsically disordered proteins

# Rozponávání pomocí neuspořádaných oblastí proteinů

**Supervisor:** Mgr. Marian Novotný Ph.D.

Bachelor's thesis

Prague 2023

I would like to thank my thesis supervisor Mgr. Marian Novotný Ph.D. for his invaluable time and expert advice, who has been very supportive throughout this project. It was a valuable experience, and I am truly grateful for having the chance to work on such an interesting topic in my thesis.

I declare that this bachelor's thesis titled "Recognition by Intrinsically Disordered Proteins" is solely my own work, and I have carried it out independently, using only the cited literature and consultations with my supervisor.

In Prague 2023

…….

Author's signature

## Abstrakt

Neuspořádanost je jedna z vlastností mnoha přirozeně se vyskytujících proteinů, která ovlivňuje jejich funkčnost v biologických systémech. Tato práce popisuje poznatky o mechanismech, kterými mohou neuspořádané proteiny nebo neuspořádané proteinové domény přispívat ke specifickému rozpoznávání na molekulární úrovni. Jsou zde shrnuty obecné charakteristiky neuspořádaných proteinů, míra jejich zastoupení napříč různými organismy a na konkrétních příkladech jsou zde představeny možné způsoby molekulárního rozpoznávání. Dále se tato práce zaměřuje na přechody mezi uspořádaným a neuspořádaným stavem vyvolané interakcí s vazebným partnerem. V posledních dvou kapitolách se věnuje charakteristickým rysům neuspořádaných proteinů ovlivňujících buněčnou signalizaci, kterými jsou vazebná promiskuita v podobě signaling hubs, alternativní splicing nebo post-translační modifikace.



**Klíčová slova:** molekulární rozpoznávání, vnitřně neuspořádané proteiny, neuspořádané oblasti, změna konformace



## Abstract

Intrinsic disorder is one of the many traits that can affect the functionality of multiple naturally occurring proteins in biological systems. This thesis reports on the latest findings on mechanisms that intrinsically disordered proteins or intrinsically disordered regions utilize in specific recognition at the molecular level. Here, the general characteristics of intrinsically disordered proteins are summarized, along with the extent of their abundance throughout different lifeforms and the variety of their molecular recognition mechanisms depicted on specific examples. Furthermore, this thesis focuses on protein transitions between ordered and disordered states induced by interaction with its' binding partner. In the last two chapters, characteristic features of intrinsically disordered proteins are described, and attention is paid to the way these features influence cellular signaling pathways such as interactional promiscuity, the role of signaling hubs, alternative splicing, and post-translational modification.



**Key words:** molecular recognition, intrinsically disordered proteins, disordered regions, conformational change

## List of abbreviations

| | |
|---|---|
| AS | alternative splicing |
| IDAA | intrinsically disordered amino acid |
| IDD | intrinsically disordered domain |
| IDP | intrinsically disordered protein |
| IDR | intrinsically disordered region |
| MoRF/MoRE | molecular recognition feature/element |
| PreSMo | pre-structured motif |
| PTM | post-translational modification |
| SLiMs | short linear motifs |

# Contents

# 1. Introduction

Protein molecular recognition is a fundamental biological process that involves the specific binding and interaction of proteins with other molecules, such as nucleic acids, proteins, and small molecules. This recognition is critical for a multitude of biological processes, including gene expression, enzyme catalysis, and signal transduction, and as such, it demonstrates its crucial importance in the functioning of living organisms.

One of the key elements in molecular recognition that enables proteins to recognize and interact with other molecules is their structure. Specific folds and surface features allow selective recognition, enabling proteins to perform their biological functions (Voet, 2016).

An example of a structured protein domain involved in specific binding interactions is the SH2 (Src homology 2) domain. This domain is approximately 100 amino acids in length and mainly consists of an antiparallel β-sheet with one α-helix on each side. One pocket of the peptide's binding surface is located near the helices and the edge of the sheet, utilizing two of the protein's arginine and one of its' lysine residues in the binding of the phosphate group of phosphotyrosine. The other binding pocket of the domain binds three amino acids adjacent to the phosphotyrosine (Figure 1). This highly specific binding of phosphotyrosine is largely determined by the specific amino acid composition of these pockets, which can vary between different SH2 domains and therefore contribute to their varying specificity for particular phosphorylated tyrosine-containing peptides, which is essential for the proper functioning of many cellular signaling pathways (Liljas *et al.,* 2017).

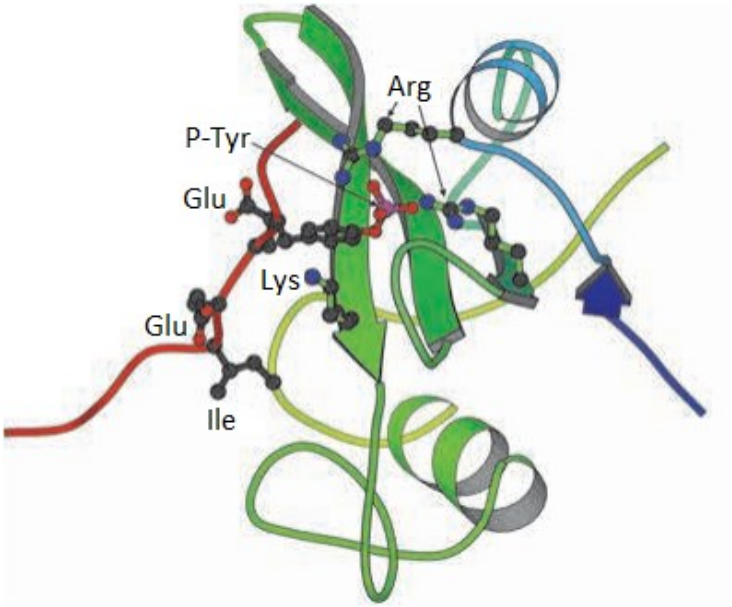


*Figure 1: The SH2 domain of the src protein with the peptide YEEI (red) in complex. The binding of the phosphate group of phosphotyrosine (P-Tyr) is mediated by two arginine residues and one lysine residue. Adopted from "Textbook of structural biology, 2nd edition" (Liljas et al., 2017).*

While structured proteins and protein domains are crucial for molecular recognition, another class of proteins or protein domains known as intrinsically disordered proteins (IDPs) or intrinsically disordered protein regions (IDRs) / protein domains (IDDs) have also gained significant attention in the last 30 years. Unlike structured proteins, IDPs lack a fixed three-dimensional structure and can adopt a range of conformations while maintaining their physiological activity. The intrinsic flexibility of many IDPs enables them to interact with a diverse set of binding partners, often with weaker affinity compared to their globular counterparts. The diversity of their interactional partners may be further modulated by alternative splicing, creating protein isoforms with different representations of residues, or via post-transcriptional modifications, such as phosphorylation of serine or acetylation of lysine residues, which can affect the charge or hydrophobicity of the residue (Dunker *et al.*, 2015).

SUMO-1, a small protein with about 100 amino acids in length that plays a role in determining protein localization inside the cell, is an example of a protein that contains disordered regions. Although it adopts a structured ubiquitin fold consisting of a four-stranded β-sheet and a α-helix, the protein also contains N- and C-terminal regions that lack a stable three-dimensional structure when unbound, as shown in *Figure 2*. The function of the C-terminus is, like in ubiquitin, to attach its' carboxyl group to the target protein's lysine. The variable length of the flexible N-terminal region in the SUMO protein family, which is not present in ubiquitin, suggests possible involvement in interactions with other proteins (Bayer *et al.*, 1998; Liebelt *et al.*, 2019).

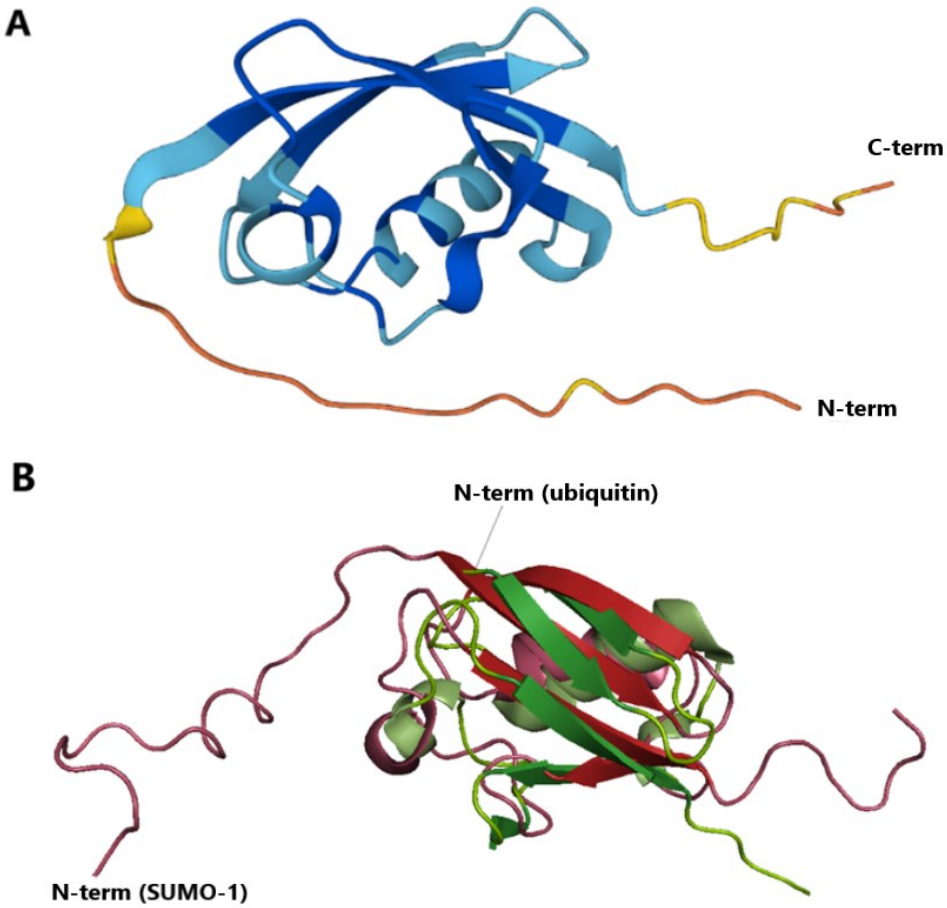


*Fig. 2: A*) Depiction of protein SUMO-1 in AlphaFold Protein Structure Database under UniProt entry number "P63165," where red- and yellow-colored residues have a lower confidence score and are more likely disordered in their native state, and blue residues have a higher confidence score and possess native structure as depicted. B) Superposition of human SUMO-1 protein (colored in shades of red) and ubiquitin (colored in shades of green). The picture was created in PyMOL software while using NMR structures of two proteins stored in the PDB database under the accession IDs "1A5R" and "ID3Z", respectively.

Despite their lack of a fixed structure, IDPs possess distinct structural features that enable their function, such as short linear motifs and regions of intrinsic disorder that undergo disorder-to-order transitions upon binding to their targets. IDPs also tend to interact with their binding partner in a highly dynamic and flexible manner without forming a stable, ordered structure. These complexes where one of the partners retains some flexibility are called "fuzzy complexes" (Kulkarni and Uversky, 2019).

Due to their irreplaceable role in cellular processes, IDPs, or proteins with IDRs, significantly contribute to many human pathologies. Neurodegenerative diseases such as Alzheimer's and Parkinson's disease are caused by the misfolding and aggregation of IDPs such as tau and α-synuclein and mutations or aberrant expression of IDPs such as p53 and BRCA1 lead to cancer development (Uversky *et al.,* 2008).

Although the unique properties of IDPs have made them challenging to study, IDPs have gained significant attention in recent years due to their critical roles in many cellular processes and their involvement in human pathologies. And even though advances in experimental and computational techniques have enabled researchers to better understand their structure or intramolecular interactions, many recognition mechanisms involving disordered proteins are still to be clarified.

# 2. Protein structure

In a cellular environment, most biological processes are driven by proteins, large biomolecules often designed to bind other proteins, nucleic acids, or other bioactive molecules. Proteins comprise one or several chains of different amino acid residues, which affect protein biological, chemical, and physical properties and determine protein primary structure (amino acid sequence) (Liljas *et al.*, 2017).

## 2.1. Secondary structure

Protein secondary structure is an arrangement of polypeptide backbone atoms in space held together by hydrogen bonds between backbone amide and carbonyl. Due to the partial double-bond nature of the C-N bond, it cannot rotate and is destined only to possess the form of a cis or trans isomer, of which trans is more stable. $C_\alpha$-N and $C_\alpha$-C bonds are however freely rotable, and their torsion angles φ ($C_\alpha$-N bond) and ψ ($C_\alpha$-C bond) are illustrated in *Figure 3.*

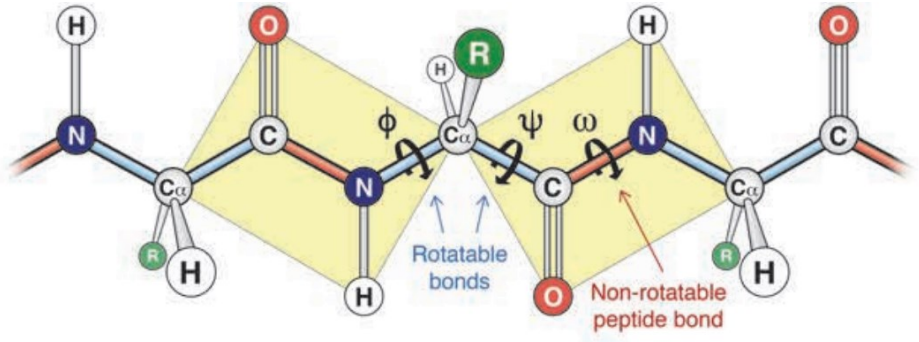


*Figure 3: Torsion angles of the polypeptide backbone. The backbone consists of $C_\alpha$, C and N atoms. Rotations of the protein backbone are permitted only about $C_\alpha$-N (φ angle) and $C_\alpha$-C (ψ angle) bonds, which are colored blue. N-C (peptide) bonds depicted in orange are non-rotable due to their partial character of double bond. Adopted from "Textbook of structural biology, 2nd edition" (Liljas et al., 2017).*

In reality, some of these torsion angles do not naturally occur due to the sterical barriers between atoms of the peptide backbone and its sidechains. This fact is depicted in Ramachandran's diagram (*Figure 4*), indicating allowed combinations of φ and ψ angles (Liljas *et al.*, 2017).

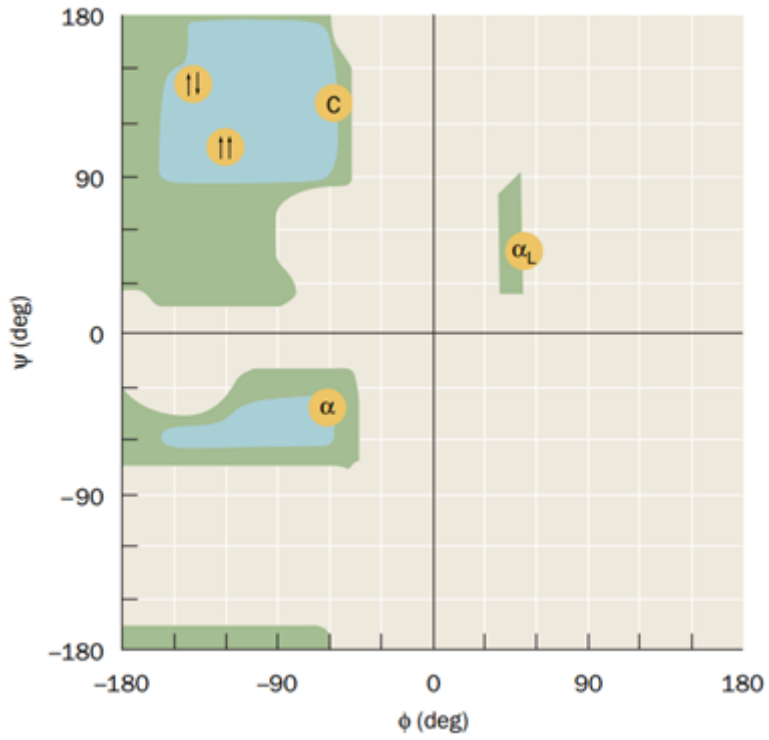


*Figure 4: The Ramachandran diagram. Blue-colored regions represent allowed regions of φ and ψ angle combinations, and green regions represent the more crowded areas of φ and ψ angles. Distinct secondary structures are present in areas of yellow circles with corresponding symbols: right-handed α-helix (α), left-handed α-helix ($\alpha_L$), collagen helix (C), parallel β-sheet (↑↑) and antiparallel β-sheet (↑↓). Adopted from "Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition" (Voet, 2016).*

Most of the space in Ramachandran's diagram is occupied by forbidden conformations of the polypeptide chain. The torsional angles of most residues dwell in just three small areas of the plot. The existence of polypeptide chains in these areas favors the formation of regular secondary structures such as α-helixes, β-sheets and β-turns. Common secondary structures also tend to prefer some amino acid residues over others. Larger, hydrophobic amino acids are typical for β-sheets, where prolines may be found too for their strand-bending nature. α-helices typically feature longer amino acid stretches that are capable of forming hydrogen bonds. However, proline and glycine residues are generally less abundant in these structures. Protein parts without any fixed secondary structure are called random coils. These coils are typically functionally versatile and are present in natively unstructured proteins (Voet, 2016). This type of disorder is found in natively unstructured regions of proteins and is commonly referred to as intrinsic disorder.

## 3. Characterization of intrinsically disordered proteins

According to the traditional understanding of molecular biology, proteins fold to acquire a predefined three-dimensional structure, allowing them to perform their biological activity. Thus, their function can be predicted by their structure. The main driving force behind the folding of typical soluble proteins is the maximalization of polar interactions, such as salt bridges and hydrogen bonds, while hiding hydrophobic residues in the protein core. This concept proved insufficient since many highly functional proteins without a solid 3D structure have been discovered over the history of research (Lermyte, 2020).

The proteins without regular 3D structure are called intrinsically disordered proteins (IDPs), referring to their lack of rigid structure in physiological conditions (disordered) and the relation of this disorder to the sequence (intrinsically) (Wright and Dyson, 1999).

IDPs can assume a wide range of transitory conformations with similar free energy and without any significant thermodynamical barrier between them. This characteristic of IDPs is apparent in an example of the pKID domain of the CREB protein. When unbound, this domain is highly disordered. Relatively shallow funnels (with a slope of -100 kJ/mol) in its' free energy landscape prevent pKID from forming a stable structure, making it fluctuate between many partially folded states. However, if bound to the KIX domain of the CREB binding protein, it assumes the conformation of an α-helix with a free energy of -226 kJ/mol (Chong and Ham, 2019).

The degree of disorder in IDPs ranges from partially unstructured polypeptide segments of varying length (IDRs) to fully unstructured proteins. This absence of 3D structure does not, however, exclude them from engaging in a variety of important biological activities. In addition to protein and DNA/RNA recognition, IDPs are also targets of several posttranslational modifications (PTMs) and alternative splicing (AS) of pre-mRNA that codes for IDP regions (Dunker *et al.*, 2001; Oldfield and Dunker, 2015).

Their mobility and structural instability are caused mainly by their amino acid composition. The main cause of their inability to fold is a lack of hydrophobic residues and an abundance of polar residues and proline (Uversky and Gillespie, 2000). In general, Pro, Glu, Ser, Gln and Lys can be considered disorder-promoting amino acids. Cys, Trp, Ile, Tyr, Phe, Leu, Met, His and Asn, on the other hand, are perceived to be structure-promoting (Figure 5). Ala, Gly, Thr, Arg and Asp are indifferent regarding structural order in proteins (Vacic *et al.*, 2007).

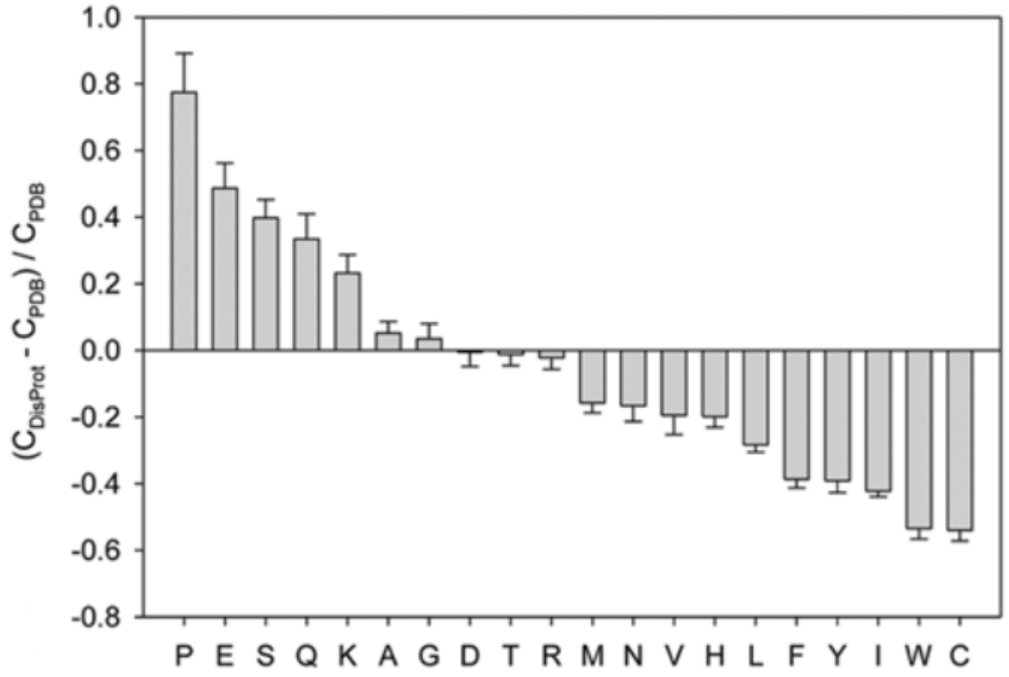


*Figure 5: Difference in the amino acid composition of IDPs in the DisProt database. The fractional difference is computed as ($C_D$-$C_P$)/$C_P$, where $C_D$ stands for the content of a given residue in DisProt and $C_P$ is the content of the same amino acid in PDB. Bars above the 0 threshold indicate the presence of the particular amino acid in IDPs, and bars below the threshold imply a lack of it in IDPs. Adopted from "The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins" (Theillet et al., 2013).*

## 3.1. Distribution of IDPs through different life forms

In comparison to prokaryotes and archaea, the proteomes of eukaryotes contain substantially more predicted intrinsically disordered proteins, of which multicellular eucaryotes are richer in IDPs than unicellular eucaryotes. Viruses show the largest variability in the fraction of IDRs in their genome, where small viruses and viruses expressing one polyprotein precursor have the highest IDR fraction - up to 77% in the case of the Avian Carcinoma virus, in contrast to the human coronavirus NL63, which has an estimated IDR fraction of only 7.3% as an example of larger viruses. The content of proteins with IDRs in proteomes ranges in common bacteria from ~15% to~30%, in eukaryotes from~30% to~50% and in archaea from~12% to~24% as depicted in *Figure 6* (Xue *et al.*, 2012). Structural prediction tools are often insufficient when assessing the scale of disorder in the proteins of extremophiles, as these organisms thrive in conditions that differ greatly from those of mesophiles. The physiological environment in which extremophilic proteins operate can be vastly different, such as in halophiles, which live in highly saline conditions. This environment results in proteins having a greater number of polar and hydrophilic amino acids on the surface, which can impact their structural disorder and complicate accurate predictions using traditional methods (Pancsa *et al.*, 2019).

Listing these statistics, it is important to mention that many of these proteins only contain locally disordered regions or possess disordered linkers connecting folded domains and are otherwise structured (Lermyte, 2020). The abundance of longer disordered residues (with a length of 30 amino acids or more) is higher in eucaryotes, taking up to one third of all eucaryotic proteins on average (Ward *et al.*, 2004).

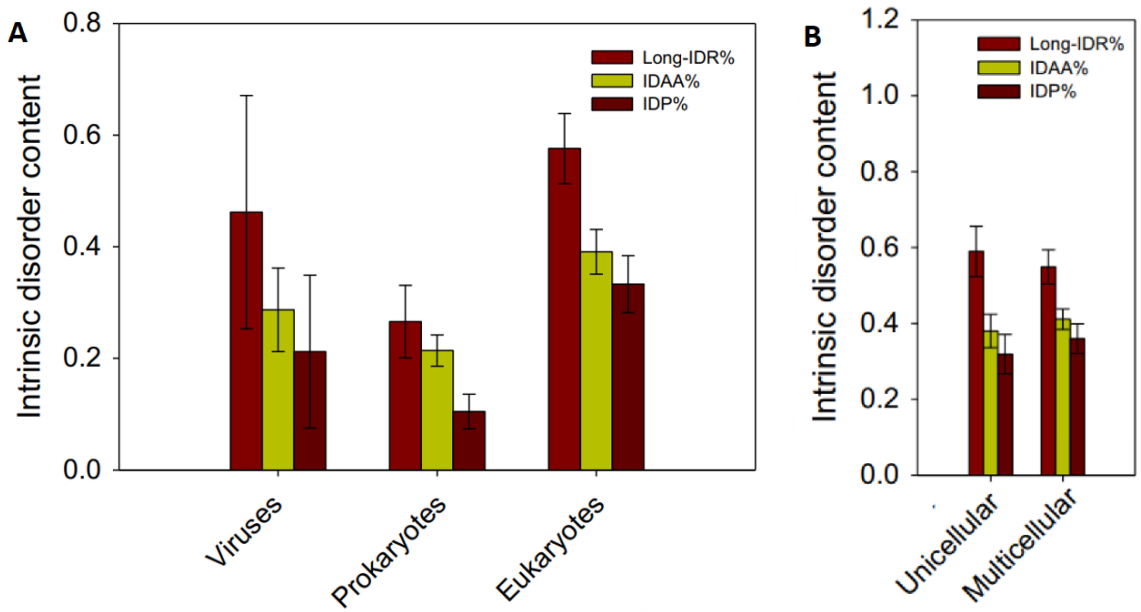


*Figure 6: Correlation between the intrinsic disorder content and complexity of organism: Average content of fully disorder proteins (IDP%), disordered residues (IDAA%) and long disordered regions (Long-IDR%) is shown in comparison of eucaryotes to viruses and bacteria (A) and in comparison, of multicellular eucaryotes to unicellular (B). Adopted from "Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life" (Xue, Dunker and Uversky, 2012).*

Depending on which predictor is used, the fraction of disordered proteins in the human proteome ranges between 30 and 50%. According to Ruff and Pappu in 2021, the human proteome is expected to be composed of 30% proteins with IDRs 30 or more residues long, of which 15–25% fold conditionally (Deiana *et al.*, 2019). suggests that 51% of human proteins contain at least one disordered region, and 32% of these proteins are disordered in at least 30% of their length. The distribution of IDPs also varies among the different types of proteins. IDR are prevalent in proteins that undergo alternative splicing and are frequent protein targets of post-translational modifications. Evolutionary highly conserved enzymes, however, tend to be extremely poor in disordered regions (Niklas *et al.*, 2018).

# 4. Protein recognition

The structure-function relationship is often described in terms of the "lock-and-key" model, where the protein is the lock, and the ligand or substrate is the key. The binding site on the protein is complementary in shape and chemical properties to the ligand or substrate, allowing for specific recognition and binding. The function of proteins is often performed by well-structured regions called functional domains. Particular functional domains can provide a variety of activities ranging from enzymatic function to molecular recognition, which can be represented by the kinase and DNA-binding domains, respectively. Domains can exist and function independently of the rest of the protein. When laying in proximity, some secondary structures may arrange into distinct three-dimensional structures upon interaction with the ligand. These structures are called structural motifs and often function as binding sites for proteins, nucleic acids and small molecules (Voet, 2016).

One of the structurally well-defined protein-binding motifs is a β-sandwich, which can be found in the Ig (Immunoglobulin) fold present in antibodies or in the PTB (phosphotyrosine binding) domain found in tensin and which binds to the cytoplasmic tails of β integrin. Another structured binding domain is the SH3 domain. The SH3 domain is about 60 amino-acid-long protein domain that is commonly found in eukaryotic proteins involved in intracellular signaling. This domain is composed of five-stranded antiparallel β-barrels and typically binds to proline-rich sequences found in other proteins, that, while binding, are present in the conformation of the polyproline-helix. This binding to the proline-rich peptide takes place in a groove between two loops of SH3. This binding interaction is typically weak and transient, allowing for dynamic protein-protein interactions (Liljas *et al.*, 2017).

Among the most important structures in DNA recognition are coiled coil motifs. These motifs consist of dimers or trimers of α-helices tightly bound together, where each helix interacts with another via hydrophobic interactions mediated by aliphatic amino acid residues. These helices have an amphipathic nature, meaning their multimerization sites consist of small hydrophobic residues, whereas exposed parts contain polar residues. The sequence of coiled coils is quite repetitive, and due to the frequent presence of leucine in the fourth position, it is also called the leucine zipper *(Figure 7)*. Among other well-known structures that play a role in nucleic acid recognition are the $Ca^{2+}$-binding EF-hand with a helix-loop-helix motif or the $Zn^{2+}$-binding zinc-finger motif. Structural motifs, however, are not considered domains due to their structural instability when unbound (Lodish, 2017).

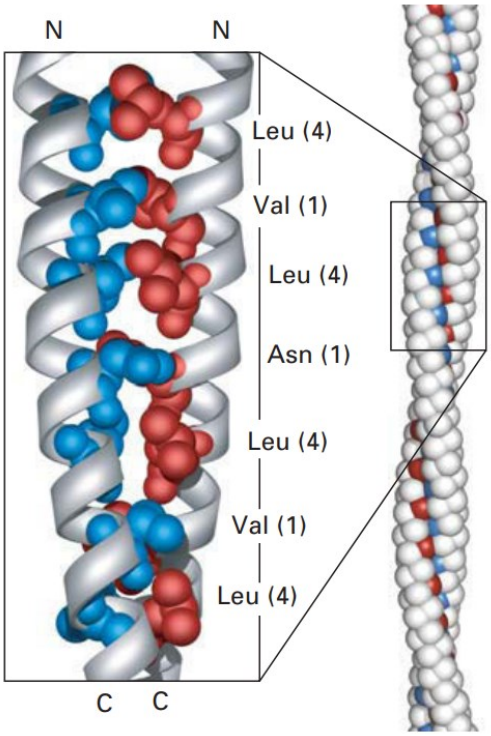


*Figure 7: Example of a protein motif (coiled coil) Two α-helices twisted together are apparent, hence the name. In the detailed picture of the interaction between two helices, amino acid residues are shown with their positions in parenthesis. Hydrophobic residues are often found in the 1 and 4 positions, with leucin present mainly on the 4 position. Adopted from "Molecular Cell Biology, 8th edition" (Lodish, 2017).*

## 5. IDPs in molecular recognition

IDPs can bind to a protein, nucleic acid, or metal ion such as zinc, and when they do, they often form structured complexes by undergoing a disorder-to-order transition. This phenomenon is called coupled folding and occurs, for instance, upon binding of the leucine zipper motif to the major groove of DNA or the zinc-finger motif to the zinc cation, which causes approximately 30 residues to adopt a rigid structure (Sugase *et al.*, 2007; Dunker *et al.*, 2015).

Complexes consisting of at least one IDP might assume multiple transient structures or can stay unstructured while still being partially bound, however, under different conditions. Such protein complexes capable of possessing a wide range of binding modes are called "fuzzy complexes" (Sharma *et al.*, 2015). The size of the IDP interactional interface does not significantly differ from that of a structured protein, however notable difference can be seen in the ratio of the length of the interactional site compared to the length of a whole protein, where chains of interactional sites of IDPs are much smaller than those of structured globular proteins relative to their size. These sites tend to utilize more of their residue area,

resulting in a larger total interactional area per residue. Intrinsically disordered proteins also use a larger portion of their surface for interaction due to the existence of multiple binding sites on one IDP, which implies the existence of multiple binding partners, as depicted in *Figure 8(a, b).* As for amino acid composition, IDPs are generally richer in polar and charged residues, however, the complete opposite applies to interactional sites of IDPs. These sites are noticeably richer in exposed hydrophobic residues as opposed to their counterparts in structured proteins (*Figure 8(c))*. The hydrophobic residues may participate in various interactions, while little or none of them take part in the formation of the hydrophobic center as observed in structured proteins (Mészáros *et al.*, 2007).

While it is commonly observed that protein complexes with extremely weak interactions typically involve an IDP and those with exceptionally strong binding are found only in globular complexes, IDPs are not limited only to weak interactions, as suggested by Lazar *et al.* in 2022. There is also a linear relationship between the size of the interface and the strength of the binding in globular proteins, with an increase of about 0.02 kcal/mol/Å2, suggesting an increase in the binding strength due to cooperative interactions between different binding parts. In the case of IDP interfaces, increasing the interface size has a much smaller effect on the strength of the binding. In *Figure 8(d),* the threshold of 9 kcal/mol and 1500 Å2 divides IDPs into two categories. The first group of IDPs below this threshold is able to achieve the same binding free energy while retaining a smaller interactional interface than the interfaces of the globular proteins. However, IDP interfaces above this threshold must be larger to reach the same level of binding strength.

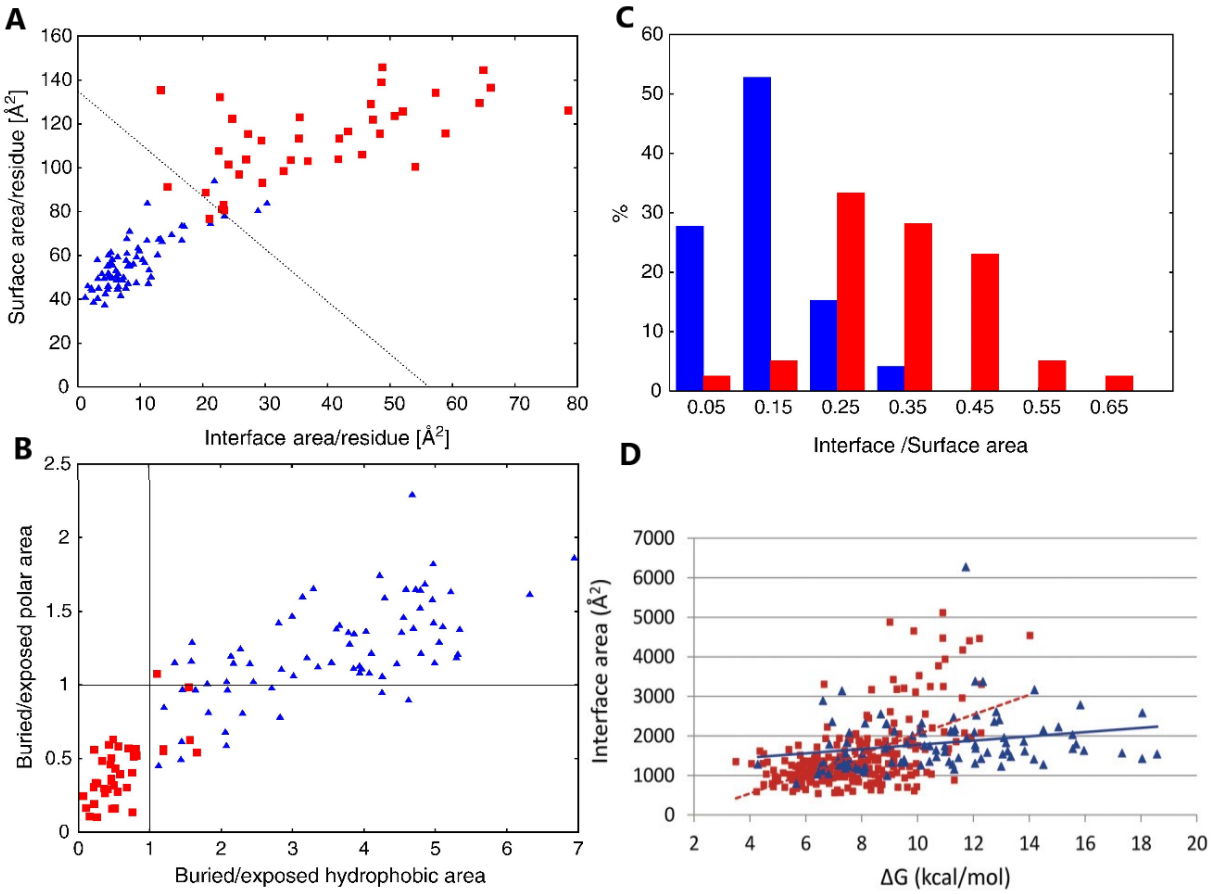


*Figure 8: Comparison of differences in characteristics in the interactional sites of ordered proteins (blue) and IDPs (red). A) Surface area per residue graphed to interface area per residue, where the blue triangles represent smaller chains of ordered complexes and the red squares represent disordered proteins complexed with an ordered protein. B) Comparison of the area of the interactional interface to the total surface area of the protein. C) The ratio of buried and exposed residues in the polar areas to those in the hydrophobic areas. Adopted from "Molecular Principles of the Interactions of Disordered Proteins" (Mészáros et al., 2007). D) The connection between the free energy of binding and the binding interface size compared for disordered (red squares) and globular (blue triangles). Adopted from "Intrinsic protein disorder uncouples affinity from binding specificity" (Lazar et al., 2022).*

Formation of fuzzy complexes is often strongly conditioned by interaction with other proteins or post-translational modification, as suggested in the case of 4E-BP binding to eIF4E, where phosphorylation of 4E-BP hinders its' ability to form a complex with eIF4E (Gosselin *et al.*, 2011). Interaction interfaces in fuzzy complexes are generally more hydrophilic compared to those in structured complexes. However, the disorder-to-order capabilities of one region may be dependent on the binding context. This conditioned folding upon forming complexes is exemplified by the ordering of the 134–161 residues in β-catenin when complexed with E-cadherin, while the same region remains unstructured in complex with the ICAT protein (Fuxreiter, 2020).

Due to the amino acid composition of IDPs, some of these proteins are highly charged, possessing a great excess of positively or negatively charged residues. Such oppositely charged proteins may then form complexes based on ionic interaction while remaining in their native disorder. The interaction between histone H1 and its chaperone ProTα is an example of this strong and structureless interaction (Borgia *et al.*, 2018). IDPs with negative charges have the ability to interact with positively charged metal ions such as calcium, zinc, or copper ions, leading to the formation of coordinate covalent bonds. This bond is formed by oxygen or nitrogen atoms, which serve as electron pair donors, usually from glutamate, aspartate, histidine, or the peptide backbone. The calcium binding sites of IDRs resemble those present in EF-hand or Excalibur motifs, thus being rich in closely spaced negatively charged residues. MIIA disordered domain of bacterium *V. coralliilyticus* calcium binding protein follows trends of previously mentioned motifs by coordinating calcium with the aspartate triad (Hoyer *et al.*, 2019).

One of the many distinctive characteristics of IDPs is their ability to bind multiple proteins due to their inherited structural flexibility. However, this should not be mistaken for a lack of specificity in their binding, as ID regions can still perform a number of site-specific interactions. An example of such specificity are natively unstructured TND-interacting motifs (TIMs), which play a role in RNA transcription and show high sequential conservation among eucaryotes. TIMs can be found in several elongation factors interacting with TFIIS N-terminal domains (TNDs), highly structurally conserved domains that consist of five grouped α-helices. Hydrophobic residues of helices in TNDs show strong conservation, although overall sequential conservation between some TNDs is lower than 10%. IWS1 transcription elongation factor possesses a total of three TIMs, where each TIM preferentially binds only one TND located on TFIIS, LEDGF, and PPP1R10 proteins, respectively (Cermakova *et al.*, 2021).

As mentioned above, IDPs are characterized by the ability to bind to multiple different proteins and often recognize common binding site(s) shared by several IDRs of diverse proteins. (Dunker *et al.*, 2005). This feature is enabled by the flexibility of IDPs, which allows them to fit onto a variety of surfaces and, in many cases, makes such proteins function as cellular interactional hubs. An example of such promiscuous binding is the oncogenic adenoviral E1A protein, which is encoded by the HAdV-5 *E1A* gene and, during early infection, is expressed in the form of a 243- or 289-residue-long product. Both proteins are natively disordered, but highly sequentially conserved regions on the N-terminus play an important role in the disruption of the cell cycle by binding several cellular proteins and forcing cells into S-phase. These interactions of this hub protein are allowed by unstructured, short linear motifs (SLiMs), one of which is the "EVIDLT" sequence, located on 118–123

residues of the 289-residue long isoform (*Figure 9*), which binds to the N-terminus SUMO conjugase UBC9. This sequence mimics a similar motif that is also present in SUMO-1 protein, and upon binding to UBC9, E1A protein effectively inhibits progression of poly-SUMOylation via UBC9 protein, possibly leading to uncontrolled cell proliferation (Yousef *et al.*, 2010; Pelka *et al.*, 2008).

Another example of a natively unstructured but conserved region in the E1A protein is the TAD. This region, which is positioned between residues 53 and 91, binds to the TAZ2 domain of the CREB-binding protein, as shown in Figure 9. Upon interacting with this domain, the TAD region undergoes a disorder-to-order transition and adopts a well-structured α-helical conformation. These types of unstructured motifs that form structured elements upon binding to their targets are known as MoRFs (molecular recognition features), as described by Sharma et al. in 2022.

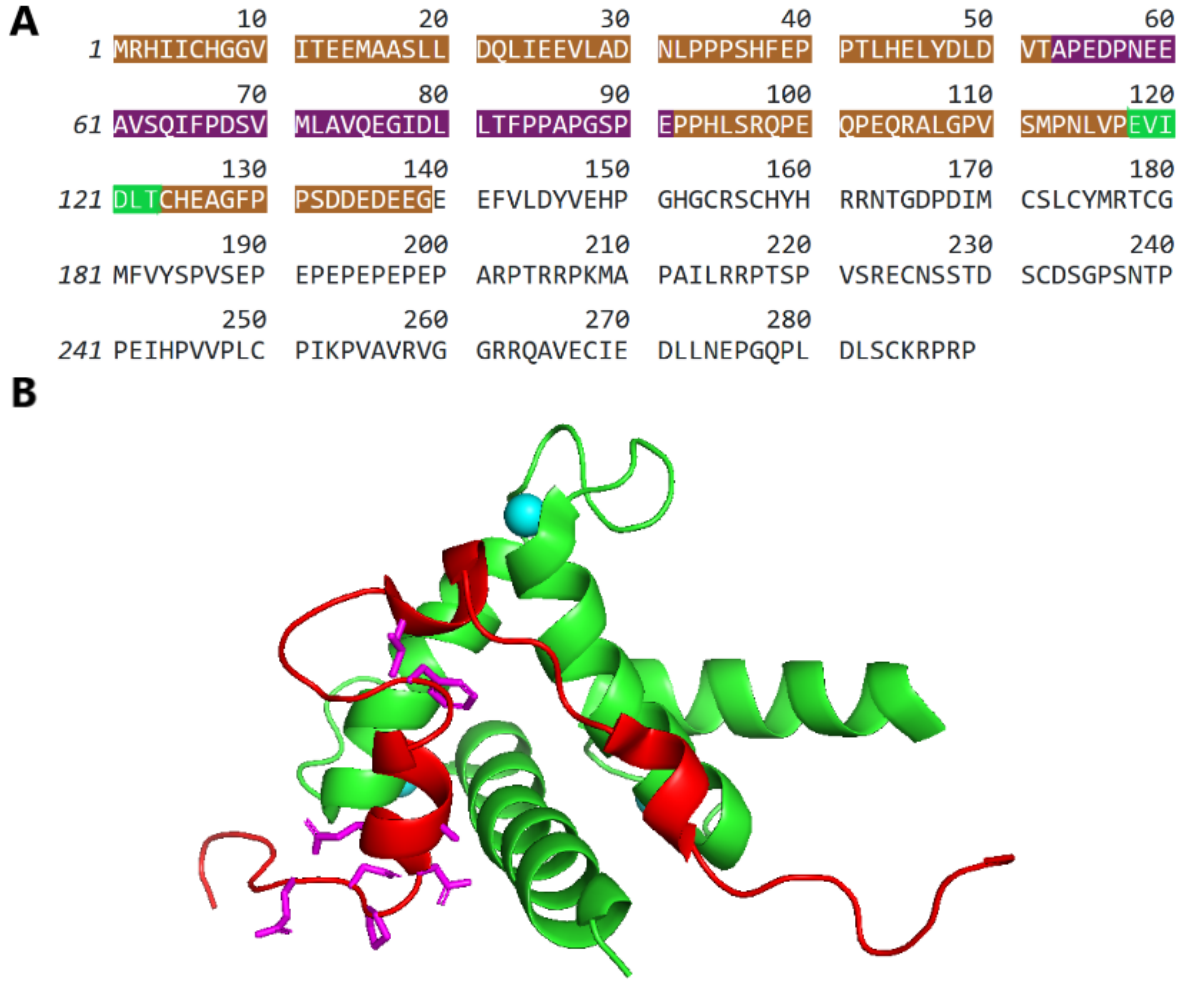


*Figure 9: A) Sequence of E1A protein with 48.1% disorder content adopted from the DisProt database depicting binding MoRF (green) of UBC9 and TAD binding MoRF (purple) within the natively disordered region (orange). B) Structure of TAD MoRF (red) with highly conserved residues (purple) complexed with the TAZ2 domain of the CREB binding protein (green) adopted from the PDB database under the accession ID "2KJE" and adjusted in PyMOL software.*

## 5.1. MoRFs

Molecular recognition features/elements (MoRFs or MoREs) are short regions of a protein that shift from a disordered to an ordered state upon binding to their interaction partner(s) and represent a class of intrinsically disordered regions with molecular recognition and binding capabilities (Mohan *et al.*, 2006).

There are three basic types of MoRFs regarding their structure formed after binding to their partner: α-MoRFs forming α-helixes, β-MoRFs adapting the structure of β-sheet and ι-MoRFs with irregular 3D structure when bound. The amino acid composition of MoRFs interfaces and their binding partners varies, as depicted in *Figure 10.* α-MoRFs and their binding partners lack tyrosine and histidine in their interfaces, and unlike β- and ι-MoRFs also lack proline due to its helix-breaking nature. Beside proline, β- and ι-MoRFs are also rich in aromatic amino acids, whereas α-MoRFs lack these residues. β-MoRFs are depleted in tryptophan and cysteine, but their binding partners are remarkably rich in cysteine. All the above-mentioned types of MoRFs have fewer residues involved in the binding interface compared to the common protein-protein interface (Mohan *et al*., 2006; Yan *et al.,* 2016).

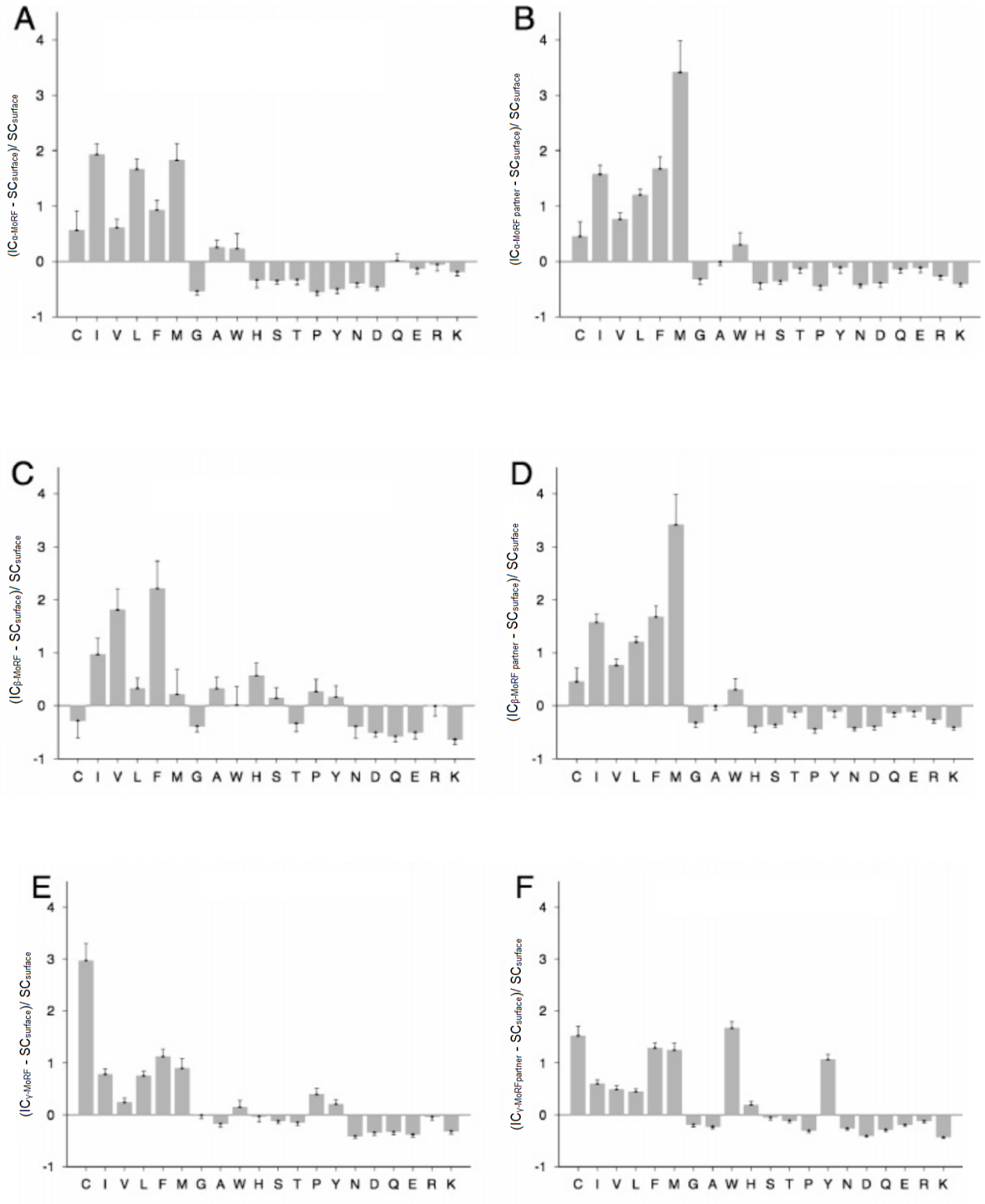


*Figure 10: Composition profiles of the interface residues of MoRFs. Interface compositions (IC) are shown relative to all surface residues (SC) of given structure on examples of α-MoRFs (A), α-MoRFs partner (B), β-MoRFs (C), β-MoRFs partner (D), and ι-MoRFs (E) and their partner (F). Adopted from "Characterization of molecular recognition features, MoRFs, and their binding partners" (Vacic, Oldfield, et al., 2007).*

Different MoRFs can bind to the same or overlapping binding site of the target protein or can bind to an entirely distinct binding site. Alternatively, IDP can use numerous MoRFs to bind several target proteins or just different sites of the same target protein (Vacic, Oldfield, *et al.*, 2007).

According to Yan *et al.* (2016), even though IDP abundance is significantly increased in eucaryotes, the amino acid composition of MoRFs is similar across the three domains of life. MoRFs are also commonly found in ribosomal proteins and proteins interacting with nucleic acids in all organisms.

The disorder-to-order transition accompanying the formation of secondary structure by MoRF is hypothesized to be associated with a considerable penalty in free energy. In the case of α-helical MoRFs, average value of this folding penalty is around 2 kcal/mol. Hadži and Lah in 2022 suggested that one of the ways to lower this energetic threshold would be through the formation of the fuzzy complex. Retaining some disorder in the bound state reduces the folding penalty of the binding motif by roughly 30%. The same drop in the free energy penalty (e.g., 30%) can be observed in disordered motifs forming pre-folding structures where the ID motif is already partially helical.

## 5.2. PreSMos:

Even though IDPs lack stable 3D structure while unbound, some disordered regions tend to form transient secondary structures before binding to their partners. These regions are commonly agreed to be called pre-structured motifs (PreSMos) (Kim and Han, 2018).

By definition, PreSMos are target-binding regions of unbound IDPs/IDRs with a length of at least 40 residues that are already present in a conformation similar to the conformation occurring in target-bound complexes. In contrast to MoRFs - disordered regions, whose structure can be detected only in a complex upon binding to target protein, structures of PreSMos are experimentally detectable by NMR analysis in unbound IDPs (*Figure 11*). Thus, unlike in the case of MoRFs, the function of a given IDR can be predicted by the PreSMos structure. Even though NMR is well suited to study IDPs, it is basically restricted to structured complexes and, to some degree, to PreSMo. More populated PreSMos also often contain longer IDRs (Fuxreiter *et al.*, 2004). The transcriptional activation domain (TAD) domain of the p53 protein, consisting of 73 residues, can serve as an example. An NMR study proved it to possess a form of amphipathic helix, where certain residues in the unbound form of p53 TAD pre-existed in a helical conformation (Lee *et al.*, 2000).

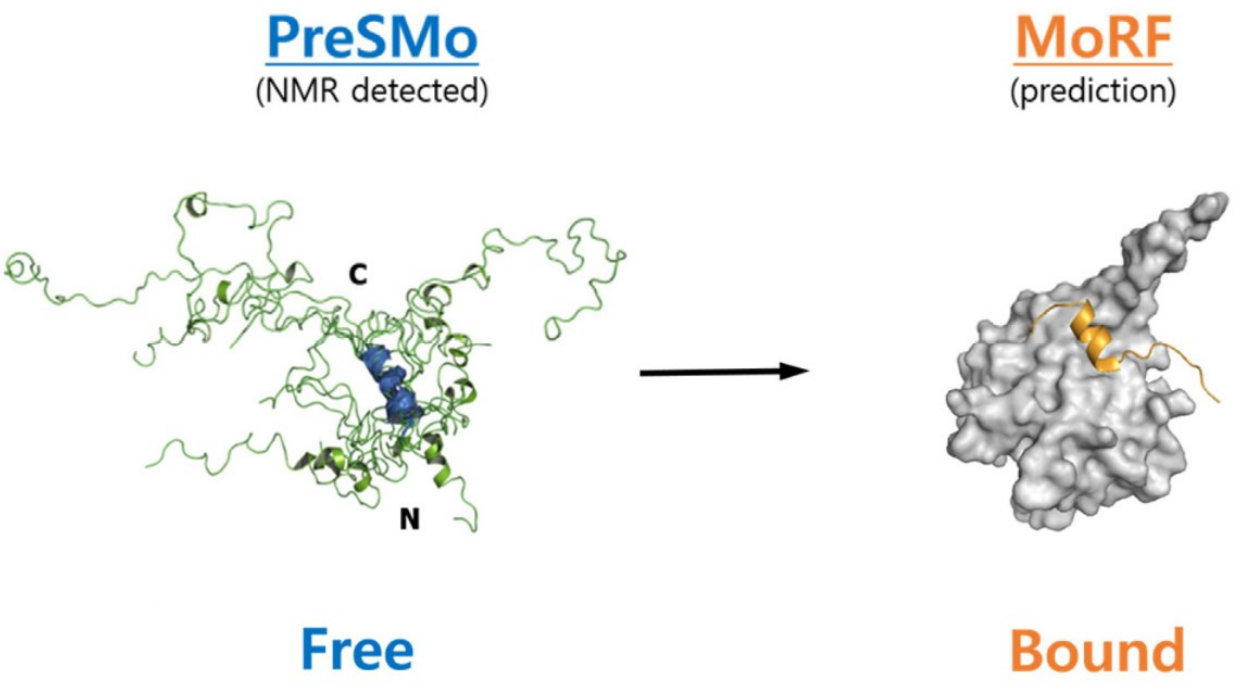


*Figure 11: Diagram of differences between a PreSMo and a MoRF illustrated on the 4EBP1 protein. Unlike MoRFs in target-free states, PreSMos can be observed by NMR experiments and can be superposed to the structure bound to the target molecule. After binding, PreSMo might become a MoRF, α-MoRF in the case of 4EBP1 bound to eIF4E. Adopted from "PreSMo Target-Binding Signatures in Intrinsically Disordered Proteins" (Kim and Han, 2018).*

## 6. Signaling hubs

IDPs significantly contribute to the complexity of cellular signaling in higher eucaryotes. They stand out as functional signaling molecules due to their fast association rates, allowing them to rapidly switch signals. IDPs that contain multiple binding motifs can utilize these features in allosteric regulation or can function as signaling hubs, binding several different macromolecules (Wright and Dyson, 2015).

Signaling hubs can bind multiple disordered proteins or might be disordered themselves. Such a disordered hub may be represented by HMGA (high mobility group protein A), a fully disordered nuclear protein binding to DNA and several other proteins, most of which are transcription factors. HMGA possess three DNA-binding motifs called AT-hooks with consensus sequence PRGRG surrounded by a number of positively charged residues on each side, that bind to the minor groove of AT-rich DNA sequence and form a partially structured (fuzzy) HMGA-DNA complex (Reeves, 2001). The C-tail of this protein is responsible for protein-protein interactions and engages in gene transcription regulation.

A well-known example of an interactional hub is the p53 protein. Most of the binding sites on p53 are located in the disordered regions, and while some of these sites fall under the general definition of a MoRF, other sites tend to form a more fuzzy structure when bound. As a signaling hub, p53 can bind to several structurally different partners. This is depicted in *Figure 12*, where p53 interacts with a variety of binding partners via its' highly structured DNA-binding domain, but it also makes use of its' natively unstructured parts, which are the flexible C- and N-terminal domains in DNA binding and the N-terminal domain in SV40 binding. An extreme example of binding plasticity is MoRF, located on the C-terminus of the p53 protein. Not only is this region (from 374 to 388 residues) part of several interactional sites that bind different binding partners, but it also forms different secondary structures upon binding. That are α-helix in S100β binding, β-sheet with sirtuin, and a coil when bound to CBP and cyclin A (Oldfield *et al.*, 2008).

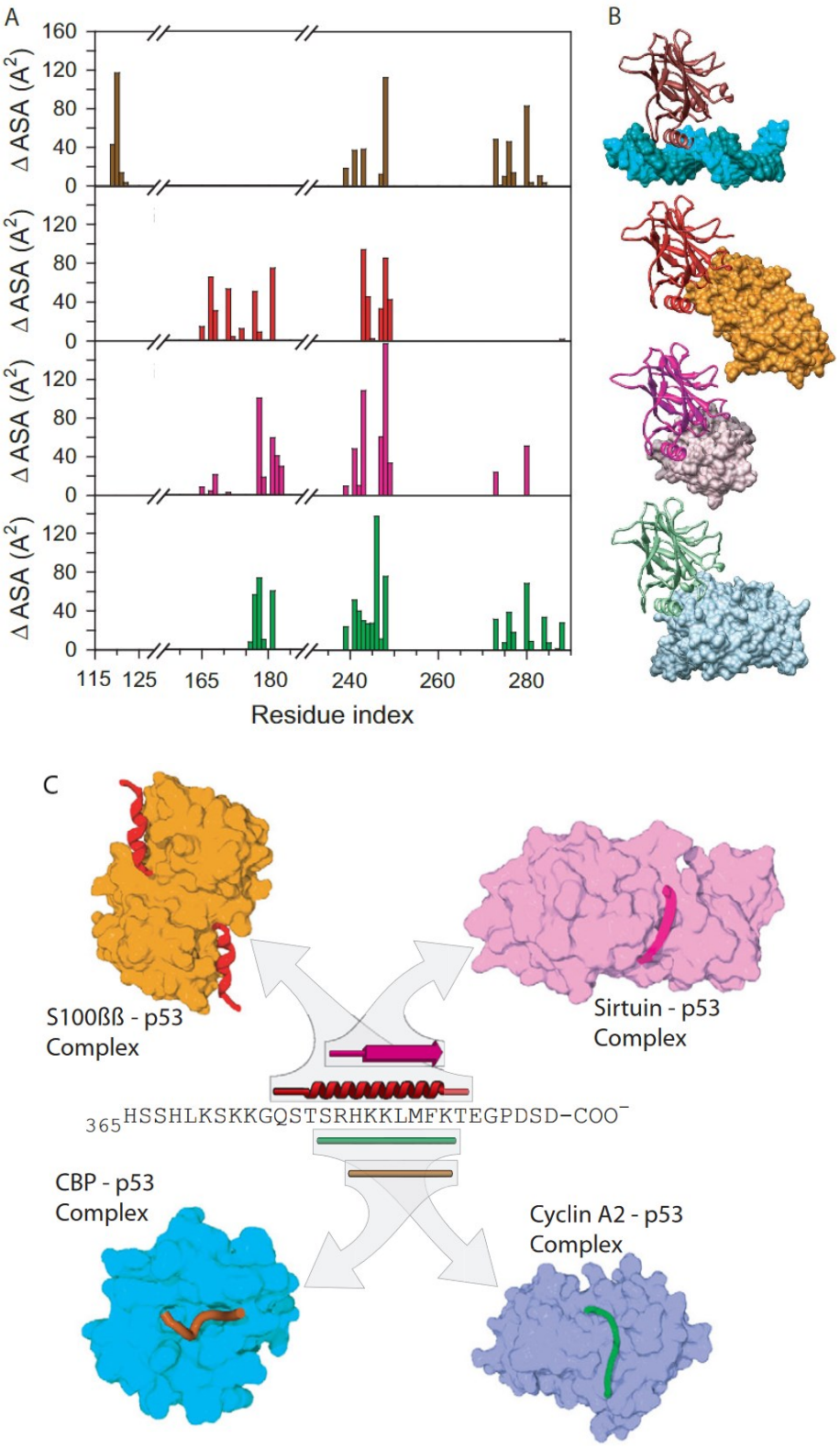


*Figure 12: DNA-binding domain of p53 in interaction with different binding partners. Interaction profiles, assigning accessible surface area values (in A2) to certain residues (A), and pictures of complexes of p53 with DNA, 53BP1, 53BP2 and sv-40 in that order (B). In subfigure (C), the binding sites of four different interactional partners share one common binding sequence in p53. Adopted from "Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners" (Oldfield et al., 2008).*

In addition to their physiological function, IDPs are also notable for their pathological roles in degenerative amyloid diseases, such as Parkinson's disease, where the accumulation of Lewy Bodies, consisting mainly of Alpha synuclein, results in the loss of dopaminergic neurons. Alpha synuclein is a natively disordered protein that functions as a signaling hub.

Alpha synuclein has intrinsically disordered N- and C-terminal domains, which play crucial roles in its' aggregation. Half (or more) of the N-terminal domain is structured in the core structure of all fibrils. The C-terminal domain retains its intrinsically disordered character in fibrils. The charge content of the C-terminal domain is sensitive to environmental changes, e.g., pH. Charge-charge interactions mediated by the C-terminal domain are critical regulators of the balance between the early stages of aggregation. The disordered C-terminal domain in the fibril acts to recruit monomers to the fibril *via* interactions with the monomer N-terminal domain. When bound to a synaptic vesicle, it changes its' conformation into an amphipathic α-helix, where it multimerizes and assembles the SNARE complex. Under pathological conditions, however, Alpha-synuclein precipitates from its' soluble state and forms amyloid fibrils, β-sheet-rich oligomers, which can spread throughout the neuronal network, causing Parkinson's disease and Lewy body dementia (Burré *et al.*, 2018). Although exact mechanism of propagation of Alpha-synuclein fibrils is still unclear, the importance of disordered N- and C-terminal domains in its aggregation process is undeniable. More than half of the N-terminal domain forms the structured core of the fibrils, whereas the C-terminal domain remains disordered and recruits monomers by interacting with their N-terminus. The C-terminus of alpha-synuclein also acts as a pH dependent regulator of fibril aggregation. When put at a physiological (7.4) or mildly acidic pH, the overall negative charge of the C-terminus does not allow aggregation, however in lower pH (5.5–6), the repulsion of negatively charged residues is lowered, which results in the formation of fibrils. Apart from pH change, acetylation of its' N-terminal domain has also been proven to negatively affect fibril growth rate (Khare *et al.*, 2023).

Interactional promiscuity of signaling hubs can be increased by alternative splicing (AS) and post-translational modifications (PTMs). Alternative splicing, a mechanism of diversifying transcripts of one gene by altering its' pre-mRNA, commonly occurs only in multicellular organisms as a supposed cause of its' increased complexity and effectiveness in reducing selective pressure on a given gene. Splicing has generally been shown to take place in intrinsically disordered regions, which leads to the formation of various protein products of the same gene (proteoforms) (Romero *et al.*, 2006).

## 7. Post-translational modifications

Another means of expanding IDPs interactional diversity is post-translational modification. From the numerous PTMs commonly occurring in living cells, some of them have proven to be more frequently present in IDPs than others. Those being phosphorylation on residues of Ser, Thr and Tyr and N-methylation (single or multiple) of lysine and arginine. Phosphorylation is one of the most common PTMs, as more than a third of all eucaryotic proteins are estimated to be phosphorylated. Most of these phosphorylated residues take place in disordered regions, and in mammals, about 90% of them are serines, with the rest being threonine and tyrosine. NMR analysis of proteins undergoing modifications by these IDP-preferring PTMs showed evidence of a transition from disordered to ordered states as the hydrodynamic properties, compactness and charge (if phosphorylated or acetylated) of proteins alter after being post-translationally modified. IDPs can be further modified by cell-type-specific PTMs that remodel generic protein pathways into cell-type/tissue-specific ones, adding even more interactional potential to a single gene (Gao *et al.*, 2016).

Disordered-to-order transition can be exemplified by the phosphorylation induced folding of the 4E-BP2 protein, which plays a role in the control of mRNA translation initiation and binds to eukaryotic initiation factor E4, part of the eIF4E complex. In its unphosphorylated, unbound state, 4E-BP2 lacks any solid structure, however, upon phosphorylation of Thr37/Thr46 followed by another tyrosine and two serines, 4E-BP2 adopts the conformation of β-sheet, weakening affinity to 4E, which enables eIF4E to recruit a small ribosomal subunit to mRNA. When unphosphorylated, 4E-BP2 binds to the eIF4E complex, undergoing a partial disorder-to-helix transition, which prevents eIF4E from proceeding with mRNA translation (Lukhele *et al.*, 2013).

Both mechanisms (PTMs and AS) take place in the creation of the final form of the protein p53. Due to AS, p53 commonly occurs in nine isoforms, of which each can be further post-translationally modified on at least 60 amino acid residues, resulting in hundreds of functionally different proteoforms (Uversky, 2016). PTMs can exist in succession within a relatively short sequence. An example of such a region would be the vastly acetylated C-tail of p53 (Figure 13). The N-terminal tail of p53 is negatively charged and does not interact with DNA. This follows the trend of disordered tails being more abundant in DNA-binding proteins. These disordered tails also tend to be longer and richer in positively charged, closely clustered residues in comparison to non-DNA-binding protein tails (Vuzman and Levy, 2012). Histone proteins are another well-studied example of post-translational modification of disordered tails affecting DNA binding. Core histones (H1, H2A, H2B, H3 and H4) participate in packaging DNA into nucleosomes and high-level chromatin fibers. The N-

tails of core histone proteins are highly conserved in their amino acid sequence, with a high frequency of positively charged lysine and arginine (Zheng and Hayes, 2003). Successive acetylation can be again noticed on H3's N-terminal tail. Acetylation of lysine residues makes nucleosomal DNA more accessible by weakening histone-DNA ionic interactions, whereas deacetylation accomplishes the exact opposite (Eberharter and Becker, 2002).

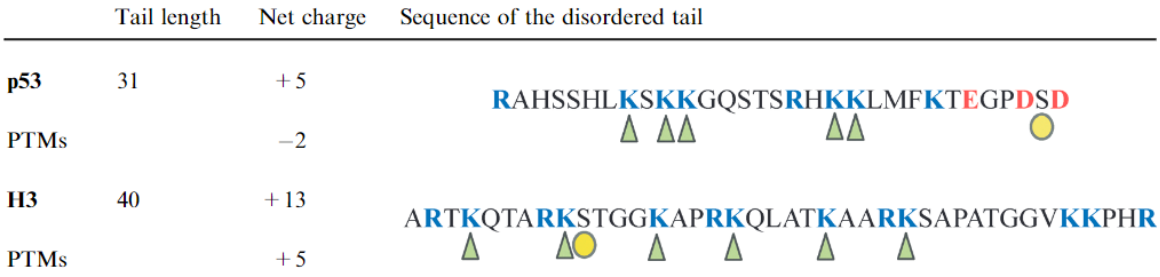


*Figure 13: Sequences and post-translational modification of p53s' C-tail and H3s' N-tail. Acetylation of certain residue is represented by green triangle and phosphorylation by yellow circle. Adopted from "Intrinsically disordered regions as affinity tuners in protein-DNA interactions" (Vuzman and Levy, 2012).*

Phosphorylation of IDPs can function as an allosteric regulator, response from their binding partners may differ depending on their number and which combination of them is phosphorylated. Some phosphorylations of serine or threonine show a strong preference for residues in IDP areas (Gao *et al.*, 2010). Beside simple phosphorylation, where addition or removal of phosphate simply modulates, as in the cAMP-regulated transcription factor CREB, there are many cases of complex response modulating mechanisms via multiple phosphorylation. First of them being threshold switches, where multiple residues are phosphorylated in close proximity and function in a cooperative manner to otherwise low-affinity receptors—for example, signaling in the process of degradation of the Cdk1 inhibitor Sic1 in transition between G1 and S phases in yeast. Sic1 gets phosphorylated up to six times, which increases its' affinity for Cdc4, a subunit of ubiquitin ligase, destined for degradation. Other signaling mechanisms, including multiple phosphorylation, involve a gradual increase in affinity for the receptor with each additional phosphorylation, which is the case with phosphorylation and acetylation of both p53 terminal tails during genotoxic stress and eventually leads to apoptosis (Wright and Dyson, 2015). Another example of PTMs in signaling is the IDR of p27kip1, which regulates the G1/S transition by binding to the cyclinA-Cdk2 complex. Removal of this protein is achieved by ubiquitination and both T and Y phosphorylation (Galea *et al.*, 2008).

# 8. Conclusions

Recognition by intrinsically disordered proteins was proven to participate in many cellular processes such as signaling, cell cycle regulation, transcription, and translation, thus being crucial for the proper functioning of the cell. This thesis summarizes recent works done on the recognition of IDPs, highlighting their recognition features and pointing out examples of IDP recognition in eucaryotes.

Intrinsically disordered proteins are present in all realms of life, with eucaryotes being most rich in them, as proteins with long IDRs take up to one third of the human proteome. Some proteins can be disordered just partially, therefore being said to contain intrinsically disordered regions. Beside their lack of rigid structure and thus large variety of backbone torsion angles, IDPs can be characterized by their amino acid composition. Hydrophilic residues are generally more frequent due to their structure-destabilizing nature. The high frequency of serine and threonine makes them ideal targets for phosphorylation, which is highly utilized in cellular signaling. Post-translational modifications of intrinsically disordered regions can happen in succession, resulting in short sequences with multiple modified residues clustered near each other.

Many biological processes include the interaction of IDPs with another biomolecule, such as protein or DNA. Upon binding to another molecule, IDPs may perform a disorder-to-order transition, resulting in the formation of structured regions that were disordered in the unbound state. These motifs responsible for recognition in disordered regions are agreed upon to be called MoRFs. MoRFs may or may not turn ordered and form secondary structure-like formations upon binding to the partner proteins. This structured complex might then be detectable by conventional methods. Depending on the structure recognized by them and the structure forming after binding, the amino acid composition of MoRF might vary. In some cases, recognition sites of IDPs can form transitory secondary structures while unbound, these transient structures are called PreSMos, and unlike other parts of IDPs, they can be detected by NMR. However, this term is not so deep-seated yet.

IDPs often function as signaling hubs—proteins with a wide range of interactional partners that participate in cellular signaling. This could be due to their ability to share the same binding site with many different proteins, the effect of alternative splicing, which presents binding diversity by creating different protein isoforms and therefore new binding sites, or post-translational modifications, which affect charge or polarity of the interactional regions. This thesis presented the p53 protein as an example of such a signaling hub for its'

ability to employ all of these three features characteristic of IDPs mentioned above, together with the disorder-to-order transition of its' MoRF. Multiple partner binding, however, should not be mistaken for a lack of binding specificity, as exemplified in the case of TIM motifs, where vital function depends on the interactional specificity of an IDR.

The binding sites of IDPs are characterized by a significantly larger area of interaction compared to globular proteins. Additionally, IDRs tend to have a higher content of hydrophobic amino acids compared to the rest of the protein or structured counterparts. However, the increase in strength of IDR binding in correlation with residue count per one binding site does not show a clear trend as in globular proteins.

As advancements in computational and experimental techniques regarding prediction and observation of the structure of intrinsically disordered proteins and their interactions have grown rapidly in the last few years, IDPs are becoming increasingly recognized as important players in various biological processes. Future research on IDPs has the potential to clarify many unanswered questions regarding intracellular processes, IDP-caused diseases, or evolutional conservation of the protein structure.

# 9. References


Bah A. and Forman-Kay J.D. (2016). Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications, The Journal of Biological Chemistry, 291(13):6696-6705, doi: 10.1074/jbc.R115.695056

Bayer P., Arndt A., Metzger S., Mahajan R., Melchior F., Jaenicke R., Becker J. (1998). Structure determination of the small ubiquitin-related modifier SUMO-1, Journal of Molecular Biology, 280(2), 275-286, doi: 10.1006/jmbi.1998.1839

Borgia A., Borgia, M., Bugge, K. et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex, Nature, 555(7694):61-66, doi: 10.1038/NATURE25762

Burré J., Sharma M. and Südhof T.C. (2018). Cell Biology and Pathophysiology of α-Synuclein, Cold Spring Harbor Perspectives in Medicine, 8(3), doi: 10.1101/CSHPERSPECT.A024091

Cermakova, K. *et al.* (2021). A ubiquitous disordered protein interaction module orchestrates transcription elongation, *Science*, 374(6571):1113-1121, doi: 10.1126/SCIENCE.ABE2913

Chong, S.H. and Ham, S. (2019). Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins, Scientific Reports, 9(1):1-9, doi: 10.1038/s41598-019-50825-6

Deiana A., Forcelloni S., Porrello A., Giansanti A. (2019). Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell, PLOS ONE, 14(8), doi: 10.1371/journal.pone.0217889

Dunker A.K. et al. (2001). Intrinsically disordered protei', Journal of Molecular Graphics and Modelling, 19(1):26-59, doi: 10.1016/S1093-3263(00)00138-8

Dunker A.K., Cortese M.S., Romero P., Iakoucheva L.M., Uversky V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks, The FEBS Journal, 272(20):5129–5148, doi: 10.1111/j.1742-4658.2005.04948.x

Dunker A.K., Bondos S.E., Huang F., Oldfield C.J. (2015). Intrinsically disordered proteins and multicellular organisms, Seminars in Cell and Developmental Biology, 37:44-55, doi: 10.1016/j.semcdb.2014.09.025

Eberharter, A. and Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics, EMBO Reports, 3(3):224-229, doi: 10.1093/embo-reports/kvf053

Fuxreiter M., Simon I., Friedrich P., Tompa P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, Journal of Molecular Biology, 338(5):1015-1026, doi: 10.1016/j.jmb.2004.03.017

Fuxreiter M. (2020). 'Fuzzy protein theory for disordered proteins', Biochemical Society Transactions, 48(6):2557-2564, doi: 10.1042/BST20200239

Galea C.A., Nourse A., Wang Y., Sivakolundu S.G., Heller W.T., Kriwacki R.W. (2008). Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1, Journal of Molecular Biology, 376(3):827-838, doi: 10.1016/j.jmb.2007.12.016

Gao J., Thelen J.J., Dunker A.K., Xu D. (2010). Musite, a tool for global prediction of general and kinase-specific phosphorylation sites, Molecular and Cellular Proteomics, 9(12):2586-2600, doi: 10.1074/mcp.M110.001388

Gao J. and Xu D. (2012). Correlation between posttranslational modification and intrinsic disorder in Protein, Pacific Symposium on Biocomputing 2012, 94-103,doi: 10.1142/9789814366496_0010.

Gosselin P., Oulhen N., Jam M., Ronzca J., Cormier P., Czjzek M., Cosson B. (2011). The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS, Nucleic Acids Research, 39(8):3496-3503, doi: 10.1093/nar/gkq1306

Hadži S. and Lah, J. (2022). The free energy folding penalty accompanying binding of intrinsically disordered α-helical motifs, Protein Science, 31(7), doi: 10.1002/pro.4370

Hoyer E. et al. (2019). Calcium binding to a disordered domain of a type III-secreted protein from a coral pathogen promotes secondary structure formation and catalytic activity, Scientific Reports, 9(1):1-13, doi: 10.1038/s41598-019-42898-0

Khare S.D., Chinchilla P. and Baum J. (2023). Multifaceted interactions mediated by intrinsically disordered regions play key roles in alpha synuclein aggregation, Current Opinion in Structural Biology, 80, doi: 10.1016/j.sbi.2023.102579

Kim D.H. and Han K.H. (2018). PreSMo Target-Binding Signatures in Intrinsically Disordered Proteins, Molecules and Cells, 41(10):889-899, doi: 10.14348/molcells.2018.0192

Kulkarni P. and Uversky V.N. (2019). Intrinsically Disordered Proteins in Chronic Diseases', Biomolecules, 9(4):147, doi: 10.3390/biom9040147

Lazar T., Tantos A., Tompa P. and Schad E. (2022). Intrinsic protein disorder uncouples affinity from binding specificity, Protein Science, 31(11), doi: 10.1002/pro.4455

Lee H., Mok K.H., Muhandiram R., Park K.H., Suk J.E., Kim D.H., Chang J., Sung Y.C., Choi K.Y., Han K.H. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53, The Journal of Biological Chemistry, 275(38):29426-29432, doi: 10.1074/jbc.M003107200

Lermyte, F. (2020) 'Roles, Characteristics, and Analysis of Intrinsically Disordered Proteins: A Minireview', Life (Basel, Switzerland), 10(12):1-19, doi: 10.3390/life10120320

Liebelt F., Jansen N.S., Kumar S., Gracheva E., Claessens L.A., Verlaan-de Vries M., Willemstein E., Vertegaal A.C.O. (2019). The poly-SUMO2/3 protease SENP6 enables assembly of the constitutive centromere-associated network by group deSUMOylation, Nature Communications, 10(1):1-18, doi: 10.1038/s41467-019-11773-xLukhele

Liljas A., Liljas L., Ash M.R. (2017). Textbook of Structural Biology, World Scientific, 2nd Edition.

Lodish H.F. (2008). Molecular Cell Biology, W.H. Freeman, 8th Edition.

Lukhele S., Bah A., Lin H., Sonenberg N., Forman-Kay J.D. (2013). Interaction of the eukaryotic initiation factor 4E with 4E-BP2 at a dynamic bipartite interface', Structure, 21(12):2186-2196, doi: 10.1016/j.str.2013.08.030

Mészáros B., Tompa P., Simon I. and Dosztányi Z. (2007). Molecular Principles of the Interactions of Disordered Proteins, Journal of Molecular Biology, 372(2):549-561, doi: 10.1016/j.jmb.2007.07.004

Mohan A., Oldfield C.J., Radivojac P., Vacic V., Cortese M.S., Dunker A.K. and Uversky V.N. (2006). Analysis of molecular recognition features (MoRFs), Journal of Molecular Biology, 362(5):1043-1059, doi: 10.1016/j.jmb.2006.07.087

Niklas K.J., Dunker A.K. and Yruela, I. (2018). The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins, Journal of Experimental Botany, 69(7):1437-1446, doi: 10.1093/jxb/erx493

Oldfield C.J., Meng J., Yang J.Y., Yang M.Q., Uversky V.N. and Dunker A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, BMC Genomics, 9, doi: 10.1186/1471-2164-9-S1-S1

Pancsa R., Kovacs D. and Tompa P. (2019). Misprediction of Structural Disorder in Halophiles, Molecules, 24(3):479, doi: 10.3390/molecules24030479

Pelka P., Ablack J.N., Fonseca G.J., Yousef A.F., Mymryk J.S. (2008). Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes, Journal of Virology, 82(15):7252, doi: 10.1128/JVI.00104-08.

Reeves R. (2001). Molecular biology of HMGA proteins: hubs of nuclear function, Gene, 277(1–2):63-81, doi: 10.1016/S0378-1119(01)00689-8.

Romero P.R., Zaidi S., Fang Y.Y., Uversky V.N., Radivojac P., Oldfield C.J., Cortese M.S., Sickmeier M., LeGall T., Obradovic Z., Dunker A.K. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, Proceedings of the National Academy of Sciences of the United States of America, 103(22):8390-8395, doi: 10.1073/pnas.0507916103

Ruff K.M. and Pappu R.V. (2021). AlphaFold and Implications for Intrinsically Disordered Proteins, Journal of Molecular Biology, 433(20):167208, doi: 10.1016/j.jmb.2021.167208

Sharma N., Gadhave K., Kumar P. and Giri R. (2022). Transactivation domain of Adenovirus Early Region 1A (E1A): Investigating folding dynamics and aggregation, Current Research in Structural Biology, 4:29, doi: 10.1016/j.crstbi.2022.01.001

Sharma R., Raduly Z., Miskei M. and Fuxreiter M. (2015). Fuzzy complexes: Specific binding without complete folding, FEBS Letters, 589(19):2533-2542, doi: 10.1016/j.febslet.2015.07.022

Sugase K., Dyson H.J. and Wright P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein, Nature, 447(7147):1021-1025, doi: 10.1038/nature05858

Theillet F.X., Kalmar L., Tompa P., Han K.H., Selenko P., Dunker A.K., Daughdrill G.W. and Uversky V.N. (2013). The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins, Intrinsically Disordered Proteins, 1(1), doi: 10.4161/idp.24360

Uversky V.N. (2016). p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept, International Journal of Molecular Sciences, 17(11):1874, doi: 10.3390/ijms17111874

Uversky V.N., Oldfield C.J. and Dunker A.K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annual Review of Biophysics, 37:215-246, doi: 10.1146/annurev.biophys.37.032807.125924

Vacic V., Oldfield C.J., Mohan A., Radivojac P., Cortese M.S., Uversky V.N. and Dunker A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners, Journal of Proteome Research, 6(6):2351, doi: 10.1021/pr0701411

Vacic V., Uversky V.N., Dunker A.K. and Lonardi S. (2007). Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinformatics, 8: 211, doi: 10.1186/1471-2105-8-211

Voet, Voet and Pratt (2016). Fundamentals of Biochemistry: Life at the Molecular Level, John Wiley & Sons, 5th Edition.

Vuzman D. and Levy Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions, Molecular Biosystems, 8(1):47-57, doi: 10.1039/c1mb05273j

Ward J.J., Sodhi J.S., McGuffin L.J., Buxton B.F., Jones D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, Journal of Molecular Biology, 337(3):635-645, doi: 10.1016/j.jmb.2004.02.002

Uversky V.N.,Gillespie J.R., Fink A.L. (2000). Why are 'natively unfolded' proteins unstructured under physiologic conditions?, Proteins, 41(3):415-427, doi: 10.1002/1097-0134(20001115)41:3<415::aid-prot130>3.0.co;2-7

Wright P.E. and Dyson H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, Journal of Molecular Biology, 293(2):321-331, doi: 10.1006/jmbi.1999.3110

Wright P.E. and Dyson H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation, Nature reviews. Molecular cell biology, 16(1)-18-29, doi: 10.1038/nrm3920

Xue B., Dunker A.K. and Uversky V.N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life, Journal of Biomolecular Structure and Dynamics, 30(2):137-149, doi: 10.1080/07391102.2012.675145

Yan J., Dunker A.K., Uversky V.N., Kurgan L. (2016). Molecular recognition features (MoRFs) in three domains of life, Molecular Biosystems, 12(3):697-710, doi: 10.1039/c5mb00640f

Yousef A.F., Fonseca G.J., Pelka P., Ablack J.N., Walsh C., Dick F.A., Bazett-Jones D.P., Shaw G.S., Mymryk J.S. (2010). Identification of a molecular recognition feature in the E1A oncoprotein that binds the SUMO conjugase UBC9 and likely interferes with polySUMOylation, Oncogene, 29(33):4693-4704, doi: 10.1038/onc.2010.226

Zheng C. and Hayes J.J. (2003). Structures and interactions of the core histone tail domains, Biopolymers, 68(4):539-546, doi: 10.1002/bip.10303