Computational and Structural Biotechnology Reports 2 (2025) 100054

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Reports

journal homepage: www.sciencedirect.com/journal/csbr

Research article

Assessing the helical stability of polyXYs at the boundaries of intrinsically disordered regions with MD simulations

Mariane Gonçalves-Kulik [a], Luis A. Baptista [b], Friederike Schmid [c], Miguel A. Andrade-Navarro [a,*]

[a] Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University of Mainz, Hanns-Dieter-Hüsch-Weg 15, Mainz 55128, Germany
[b] Max Planck Institute for Polymer Research, MPI-P, Ackermannweg 10, Mainz 55128, Germany
[c] Faculty of Physics, Johannes Gutenberg University of Mainz, Staudingerweg 9, Mainz 55128, Germany

ARTICLE INFO

Keywords:
Intrinsically disordered regions
Low complexity regions
PolyXY
AlphaFold2
Molecular dynamics simulations

ABSTRACT

Intrinsically disordered regions (IDRs) of proteins lack a stable structure. Their high content of hydrophilic and charged residues prevents them from forming globular domains and contributes to their flexibility and accessibility. Naturally, regions with a reduced amino acid composition (low complexity regions; LCRs) are found within IDRs. Disorder and low complexity of protein sequences are linked to various biological functions, including phase separation, regulation, and molecular interactions, and mutations in these regions can contribute to several diseases, including cancer. Understanding these biological properties requires examining the structural properties of IDRs and the LCRs they contain, but their inherently dynamic nature requires specific approaches combining sequence analysis, structure predictions, and molecular dynamics (MD) simulations. Here, we leverage our previous work, where we identified that certain types of LCRs combining two residues (polyXY) are frequent within IDRs and confer them with a propensity to form helical conformations. We identified a significant accumulation of these polyXYs at the ends of IDRs, following alpha helices that begin outside the IDR and can extend through the polyXY into the IDR, particularly from the N-terminal end of the IDR. MD simulations support the dynamic nature of these helical conformations. Our results suggest a mechanism by which the evolutionary emergence of LCRs at IDR ends could provide proteins with flexible regions for fold-upon-binding.

1. Introduction

Intrinsically disordered regions (IDRs) are regions in protein sequences that do not adopt a stable tertiary structure and often not even secondary structures, remaining flexible and widely accessible by the solvent [1,2]. These regions are characterized by the accumulation of hydrophilic amino acid residues, making them unable to form the hydrophobic core necessary for the formation of globular domains. Another important attribute of IDRs is the accumulation of highly charged residues, which is believed to influence the compaction and expansion of these regions [3,4].

The natural reduction in amino acid diversity within these regions often leads to the accumulation of residues with similar physicochemical properties, resulting in low complexity regions (LCRs). While not limited to IDRs, these regions are common within them. They range from extremely low complexity, accumulating several copies of the same amino acid (homorepeats) to more complex patterns composed of different residues with similar characteristics [5–7].

Both disorder and low complexity in protein sequences have been extensively explored in recent years due to their biological relevance. Protein regions where these properties are associated perform various important biological processes in the cell, ranging from driving liquid-liquid phase separation and the formation of biomolecular condensates, to transcriptional and translational regulation, molecular recognition and binding, to mention a few [8–10]. Mutation in these regions could lead to diseases, including various neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), Alzheimer's, Parkinson's and Huntington's disease, neurodevelopmental disorders and numerous types of cancer [11,12].

Due to their plasticity, the experimental resolution of IDRs structures is challenging, requiring the application of specific techniques to ensure the correct capture of their multiple conformations. As a result, these

* Corresponding author.
E-mail addresses: magoncal@uni-mainz.de (M. Gonçalves-Kulik), baptistal@mpip-mainz.mpg.de (L.A. Baptista), friederike.schmid@uni-mainz.de (F. Schmid), andrade@uni-mainz.de (M.A. Andrade-Navarro).

https://doi.org/10.1016/j.csbr.2025.100054
Received 26 March 2025; Received in revised form 28 May 2025; Accepted 28 May 2025
Available online 31 May 2025

regions are underrepresented in structure repositories, and our knowledge of their structural dynamics and functional behavior is limited [13,14].

All-atom molecular dynamics (MD) simulations have been extensively used as a complementary methodology to experimental approaches to assess the dynamic properties of IDRs and expand our knowledge of the structural behavior of these regions. MD's ability to analyze local changes in structural attributes and various conformers over time constitutes a valuable source of information to capture meaningful features of IDRs. However, the application of traditional biomolecular force fields designed for globular regions has proven to be limited in obtaining accurate MD simulations of these regions. As a result, several strategies were developed in recent years to better address some of these issues, making MD more reliable for the analysis of IDRs [15–17].

Since the release of AlphaFold2 [18] and other deep learning based 3D structure predictors, we can finally glimpse a larger volume of structural conformations of IDRs. However, this expansion of available data brought with it new challenges [19]. AlphaFold2 and other IDR predictors do not always agree on the exact extent and position of IDRs, and according to a recent benchmark study, no IDR prediction tool presents an accuracy higher than 80 % on curated samples [20]. Low (between 50 and 70) or very low (below 50) values of AlphaFold's per-residue model confidence score (pLDDT) have been suggested to indicate disorder. Nonetheless, nearly 15 % of the predicted human IDR residues have a pLDDT greater than 70, and several of them exhibit some degree of helical structure. It is suggested that these helical formations indicate the ability of AlphaFold2 to predict conditionally folded regions, which are the conformations adopted by some IDRs upon contact with a ligand or binding partner. However, this hypothesis requires further investigation, since the number of curated regions that exhibit conditional folding is still very low, and the selection of negative cases that definitely do not fold upon binding is even harder [21].

All these uncertainties motivated our search for novel ways to evaluate the stability of these helical formations within IDRs. In particular, we focused on polyXYs, LCRs composed of a maximum of two different amino acids. In our previous research, we evaluated the general context of polyXYs present in human IDRs, first exploring their structural context using experimental structural data from the PDB [22] and subsequently expanding the available structural information using AlphaFold2 predictions [23]. We were able to determine that some polyXY types, such as EK, ER, DE and AE, show a higher helical content. Here, we demonstrate that IDRs near helical structures show accumulation of polyXYs at their helix-proximal end. To examine this effect, we focused on evaluating the specific subset of annotated polyXYs located at the ends of IDRs and found near helical formations outside the IDR. These polyXYs are particularly interesting because they often extend the helical conformation to the IDR. The composition of these polyXYs was analyzed, and finally, MD simulations were performed to evaluate their helical propensity and stability over time. Note that in this manuscript we will use the term "helical propensity" to describe, the likelihood of a given residue to adopt a helical conformation.

2. Methods

2.1. Selection and analysis of helical and non-helical neighboring regions

We followed the process of our previous work [23] for the annotation of human IDRs, polyXYs, and secondary structures using AlphaFold2 predicted regions and the DSSP database of secondary structure assignments. The following versions of the source databases were used: MobiDB version: 5.0 - Release: 2022_07 [24] and AlphaFold2 version 4 [25,26]. DSSP version 3.0.0, which includes a distinctive characterization of polyprolines (PPII helices), was also used [27]. The cutoff values applied in the previous work were maintained, with polyXYs composed of a minimum of 6 residues, overlapping with the containing IDR by at least 4 residues or 50 % of the polyXY region. IDRs have a minimum size of 20 residues, as set by the prediction methods.

Regular expressions were applied to both ends of the IDR to ascertain the presence of an adjacent helical formation. When located at the N-terminus of the IDR, these helices must contain at least four residues (1 turn) upstream of the IDR and end either within the IDR or at most 3 residues upstream of the IDR. Symmetric rules were applied to select helical formations at the C-terminus of the IDR (See Fig. 1-i to iii for a detailed schematic). The IDR ends were labeled as "Helical" and "Non - Helical" depending on whether they contained one of these adjacent helices (Table S1).

2.2. Subsetting polyXYs overlapping IDRs and helices at their termini

Since our analysis focused on the relationship between polyXYs within IDRs and neighboring helical structures, we separated IDRs containing a neighboring helix for the following analyses, respecting the IDR selection described in the previous section (a and b). Considering that an IDR can contain several polyXYs, c) we selected the polyXY closest to the target helix, d) independently of whether there is overlap between the helix and the polyXY (Fig. 1-iii, Tables S2-S3).

Based on the overlap of the polyXYs with the target helix, we classify them as "high helical overlap" (if they overlap by 6 or more residues), and "low/no helical overlap" (Fig. 1A-iii). Further analyses were performed considering the proximity between the center of the polyXY and the IDR or the helix ends.

2.3. Selecting and formatting targets for MD simulations

Despite recent advances in computational power and capacity, running all atom MD simulations remains computationally expensive. Therefore, the selection of targets for our analysis was unavoidable. First, we selected polyXYs at the helical N-termini of IDR ends and considered those with high helical coverage. Of these, we selected the 55.6 % located 20 residues or less from the helix end. We further filtered them by considering those with an average pLDDT value greater than 70.

From the resulting set, we selected two samples from each of the three polyXY types that most frequently overlap helices. Before proceeding, a visual inspection of the fragment structure was performed to confirm the helical structure of the region. A 100-residue segment surrounding the polyXY was excised from the AlphaFold2 prediction and submitted to simulation. We decided to use a fragment of the sequence to reduce runtimes, which increase with the size of the molecule under study, and to reduce the risk of artifacts generated by force fields designed for IDRs when applied to structures that also contain stable globular regions [28]. In the three selected polyXY types that commonly overlap helices, at least one of X or Y is a charged residue. To explore whether charge is enough to maintain this helical conformation, we experimented with mutations of the polyXYs to the counterpart charged residues. An AlphaFold2 prediction for the mutated protein was generated, which was then cut to 100 residues surrounding the polyXY and subjected to MD simulations under the same conditions.

2.4. Molecular dynamics simulations and analyses

Gromacs 2022.5 with the all-atoms force field CHARMM36IDPS [29] was used for all simulations. CHARMM36IDPS was optimized using the grid-based energy correction map (CMAP) method, using backbone dihedrals obtained from the PDB database, differentiating between globular and non-globular regions. For the setup of the system, cubic boxes with a solute-to-box wall distance of 2.0 nm, with periodic boundary conditions, were filled with TIP3P water [30], and sodium chloride concentrations were adjusted to human physiological conditions of 100 mM. The steepest-descent algorithm was used for the energy minimization step, with a maximum of 50,000 steps or an energy

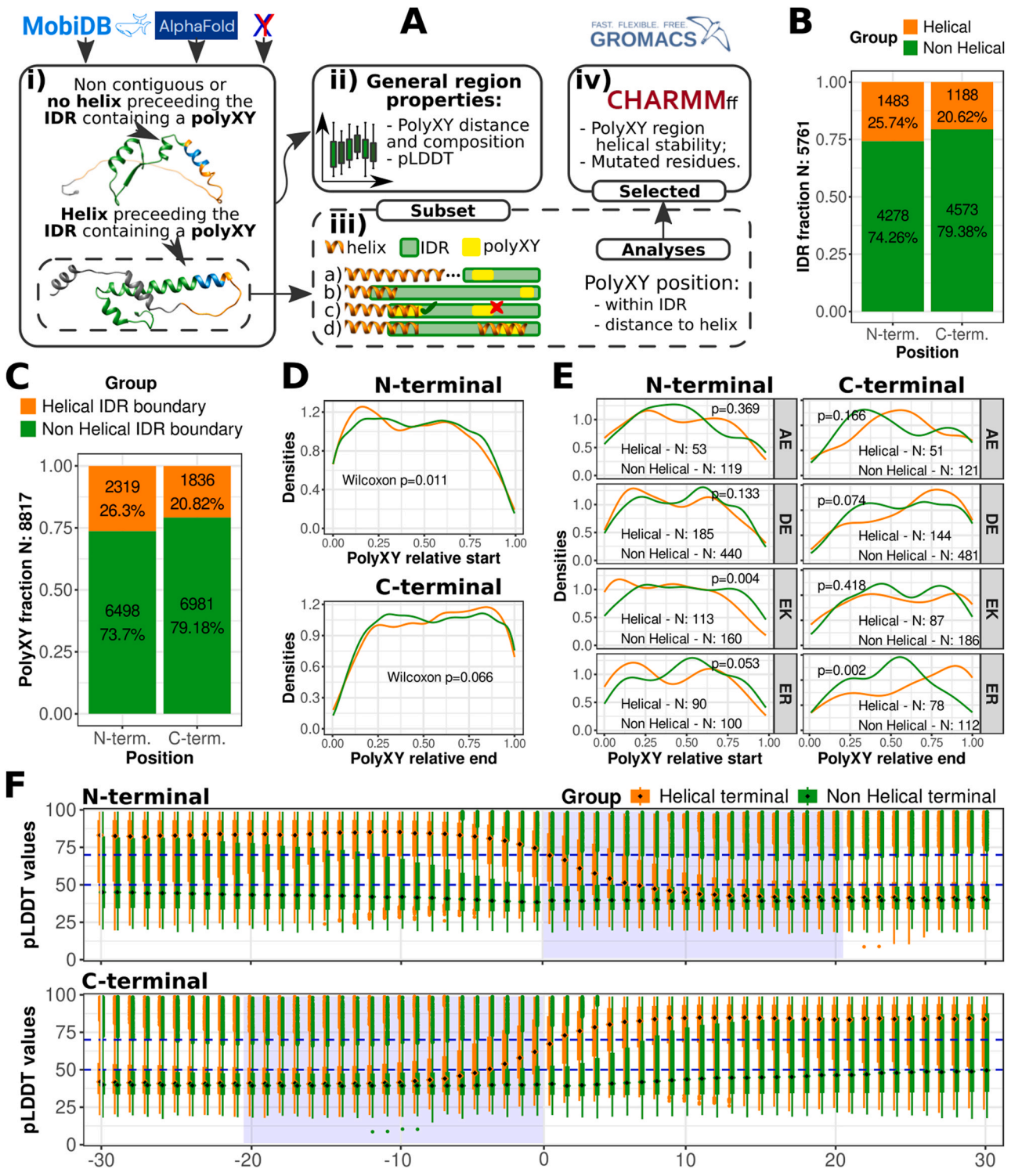

Fig. 1. Overview of the neighboring helices of IDRs containing polyXYs. (A) A schematic of the process followed to analyze and explore the properties of (i and ii) IDRs containing or not containing an adjacent helix; (iii) polyXYs within IDRs in the context of neighboring helices, considering different IDR, helix and polyXY positions along the region; and (iv) selection and presentation of helical polyXYs for MD simulations with the aim of analyzing helical stability over time. The subclassification in (iii) states for a) helices that end at most three residues before the IDR containing a polyXY begins; b) helices that partially or completely overlap with IDRs, regardless of the polyXY position; c) additional polyXYs within the same IDR were discarded from the second-level analysis, keeping only the first polyXY regardless of whether it overlaps a helix or not; d) even when multiple helices overlap with the same IDR, they are kept. (B) Counts and percentages of IDRs and (C) polyXYs within IDRs according to the presence or absence of a neighboring helix at each IDR termini. (D) Distribution of polyXYs globally and (E) among the top four polyXYs with highest helical content, considering the relative position within IDR coordinates and the presence or absence of a helix at each IDR termini. (F) Distribution of pLDDT values in the vicinity of IDRs, considering 30 residues before and 30 residues inside the IDRs. Average pLDDT is represented by diamonds, while upper and lower dots indicate outliers. The blue box indicates the minimum region covered by all IDRs (20 residues), while the dashed blue horizontal lines delimit regions of intermediate and high confidence, respectively.

difference less than 1000 kJ mol^{-1} nm^{-1} and an integration time step of 1 fs. The particle-mesh Ewald (PME) algorithm was used to calculate long-range electrostatics by cubic interpolation, with a cutoff radius of 1.0. The modified Berendsen thermostat algorithm was used for temperature coupling at 300 K, and the Parrinello-Rahman algorithm was used for pressure regulation; both steps were run every 100 ps. Production simulations were run for 1.2 μs, with the first 100 ns of each simulation discarded to reduce bias that AlphaFold2 predicted structures can cause in the initial states.

The simulation trajectories were subsequently analyzed using the Gromacs dssp function with the dssp mode version 3.0.0 for the creation of hydrogen pseudo atoms [27]. The canonical MD simulation metrics root mean squared deviation (RMSD) and surface area analysis (SASA) on the polyXY region, were also generated to support the findings observed in the secondary structure analyses of the first MD simulation and its mutations. These metrics present global average information about the distance between proposed structures along the trajectories and the proportion of these structures exposed to the surrounding solvent, respectively. To obtain local measurements, our analysis was focused on the target helical region and computed over a sliding window of six residues centered on the target residue.

2.5. Additional tools

All data were processed using in-house scripts developed in Python 3.8.10. The Biopython package was used to manipulate FASTA files and extract DSSP annotations [31]. All plots and tables were generated using scripts in R 4.4.0 and ggplot2 3.5.1, and p-values were calculated using Mann-Whitney U Test (Wilcoxon rank-sum test). Protein molecular structures were generated using Chimera 1.15 [32] and MD simulation analyses were performed using native Gromacs reports, with plots generated using Python scripts.

3. Results

3.1. IDRs neighboring helical structures show accumulation of polyXYs close to the helices

Our total set of human IDRs containing polyXYs (5761–23.31 % of all human IDRs) was divided into two groups, with and without helical structures adjacent to the IDR (see Methods for details). As proteins are synthesized from N- to the C-terminus, secondary structures N-terminal to the IDR may already be formed when the IDR starts to be synthesized, even within the ribosomal tunnel [33]. This could impose differences in the structural propensities of each IDR boundary, for example facilitating the formation of structures at the N-terminus. For this reason, we separately considered the N- and C-termini (Table S1). Consistent with this hypothesis, we observed that a slightly higher percentage of IDRs containing polyXYs present helices at their N-terminus (25.74 %) compared to their C-terminus (20.62 %) (Fig. 1B).

Next, we evaluated the polyXY content of these IDRs. Considering IDRs preceded by helices at their N-terminus, 2319 polyXYs were selected, with an average of 1.56 polyXYs per IDR, while for IDRs with helical formations at the C-terminus, a total of 1836 polyXYs were observed with a similar average of 1.55 polyXYs per IDR (Fig. 1C).

To further explore the relationship between polyXYs and helicity, we investigated whether IDRs preceded or followed by a helix presented an accumulation of polyXYs closer to the IDR end and the corresponding adjacent helix, compared to cases where no helix is present in the immediate vicinity of the disordered region.

We show that this is the case: an accumulation of polyXYs is observed within the group of IDRs neighboring helices, compared to the cases without helix, with a significant p-value of 0.011 for the N-terminus and a non-significant p-value of 0.066 for the C-terminus, according to the Wilcoxon rank test, with a threshold of $p < 0.05$ (Fig. 1D). The four polyXYs that presented the highest helical content among the polyXYs within the IDRs according to our previous work [23] were also evaluated. In cases with an IDR with a helical N-terminal, polyAE and polyDE show lower accumulation near the start of the IDR, whereas polyEK and polyER show higher accumulation, with the former presenting a significant p-value < 0.01 (Fig. 1E). This trend is attenuated for the C-terminal samples, except for polyER, with a significant p-value < 0.01, and polyDE with a non-significant accumulation to the end of the IDR, indicating a more scattered distribution of polyXY types at the C-termini.

Considering the presence of a helical boundary, one would expect AlphaFold confidence levels to be higher at the IDR termini, expanding at least marginally within the IDR, compared to samples without a boundary helix. The pLDDT distributions for both termini were analyzed (Fig. 1F), and these expectations were confirmed, with an average pLDDT above 75 for residues upstream of the IDR border and between 50 and 75 for the 6 first N-terminal and 4 last C-terminal residues of the IDRs.

3.2. Charged and helical prone polyXYs show a higher helical propensity within helical IDR ends

The following analyses were directed to polyXYs within IDRs with helical ends. Considering that an IDR can contain several polyXYs, only the first and last polyXY, closest to the IDR ends, were selected (Tables S2-S3). We analyzed their residue composition and their position relative to the helix leading into the IDR.

We evaluated 1435 polyXYs at N-termini, of which 225 (16 %) were covered by 6 or more helical residues (high helical coverage; Table S2). A similar distribution of polyXYs regarding their helical coverage could be observed at C-termini, with a total of 1153 polyXYs, of which 160 (14 %) had high helical coverage (Table S3). We observed that while polyXYs with helical coverage are significantly closer to the surrounding helix at N-termini, this is not the case at C-termini (Fig. 2). We interpret this result as indicating a stronger association of polyXYs with helical structures located at the N-termini of IDRs than those at the C-termini; such N-terminal helices could then be extended into the IDR with the help of polyXYs.

We next evaluated the residue composition of these polyXYs near helical IDR termini (Fig. 3). The most frequent polyXY types were similar at the N- and at C-termini. For low/no helical coverage, polyDE and polyAP were the most frequent types, followed by polyXY types containing serine, glycine, and proline. For high helical coverage, glutamic acid containing polyXY were the most frequent: polyER, polyEQ, polyEK, and polyAE.

In the next section, to complement our findings obtained from sequence analysis and structural data, we apply MD simulations to some examples selected among the polyXY types most prominently observed as associated with IDR helical ends, (i) to see if we can find dynamical information for some representative polyXY structures and (ii) to test the resilience of these structures to conservative mutations.

3.3. MD simulations of TERA polyER and variants

To analyze the dynamics and stability of an example of polyER situated at a helical N-terminal IDR, we focused on TERA, which, according to an electron microscopy solved structure, presents a transient helical coverage along the polyEK (PDB:7bpa; Fig. 4A) [34]. TERA stands for transitional endoplasmic reticulum ATPase (UniProt AC: P55072; also known as p97 and VCP). This protein is a hexameric AAA+ ATPase largely investigated due to its participation in several biological functions, including endoplasmic reticulum associated protein degradation and ubiquitin regulation. Due to its relevance to pathways related to cellular stress responses, some of its mutations are associated with diseases such as Amyotrophic lateral sclerosis and Paget's disease of bone; in addition, it is upregulated in certain cancer types [35,36]. A few residues downstream of the polyER are missing in

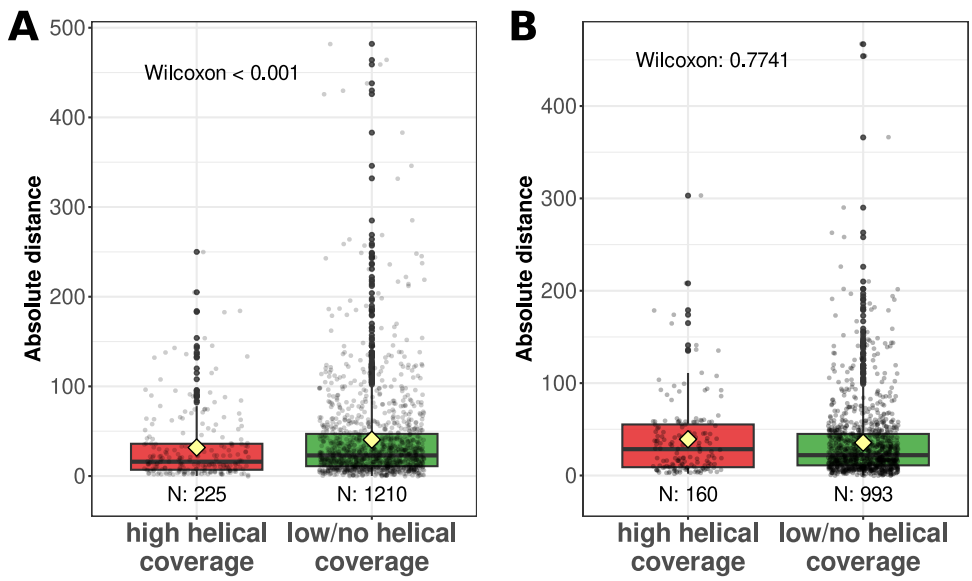

Fig. 2. Distances between polyXYs and helices surrounding IDRs. Distances from the center of the polyXY to (A) the end of the helix at the N-terminal of the IDR, and (B) to the start of the helix at the C-terminus of the IDR, considering just IDRs with adjacent helices. The total number of polyXYs in each group is shown. Black central dots indicate outliers, while gray dots show the distribution of distances in more detail. The Wilcoxon two-sided statistical test was applied, considering the null hypothesis as the existence of differences between the two groups of polyXYs with "high helical coverage" and "low/no helical coverage".

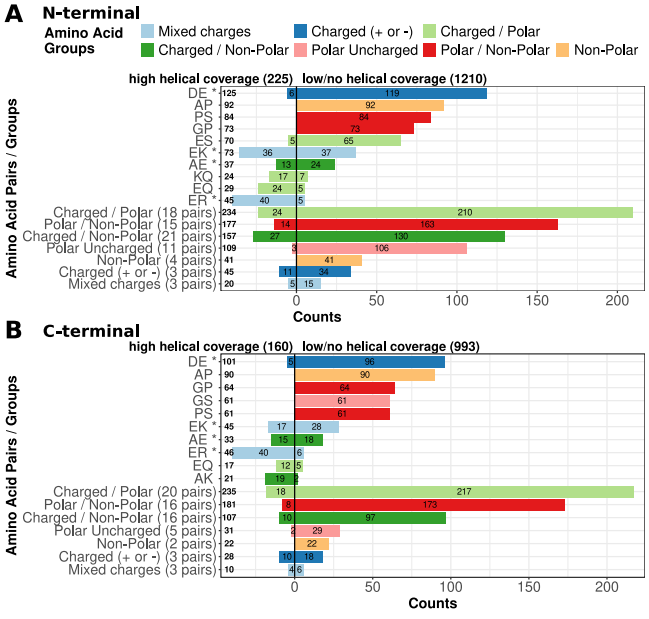

Fig. 3. Composition of polyXYs within helical IDR ends. Counts of polyXYs grouped as "high helical coverage" on the left and "low/no helical coverage" on the right. The five polyXYs with the highest counts in each group are detailed, while the remaining polyXY types were grouped by physicochemical composition of the residue pair, indicated by color. The four polyXYs with the highest overall helical coverage, according to the previous study [23], were highlighted with an asterisk. (A) Helical N-terminal ends, (B) helical C-terminal ends.

the solved structure (Fig. 4A, inset), suggesting that they adopt a flexible conformation. Examination of a total of 111 structures of TERA monomers deposited in the PDB shows 38 cases where the polyER is also completely helical, 13 cases where it is partially helical, and 30 cases where it is flexible, suggesting that the polyER could also be flexible.

To further explore the stability of the predicted helix and the conformational variability of the segment, we ran 1.2 μs MD simulations for a region of 100 residues surrounding the original polyER, RRERER, and for two mutated polyXYs: KKEKEK, replacing the Rs by similarly positively charged Ks; and KKDKDK, replacing both residue types with similarly charged residues. Like polyER, polyEK is a frequent polyXY with high helical coverage (Fig. 3), whereas polyDK is less common. The MD simulation structures indicate that the original polyER region, even when not exposed to the hexameric structure or ligands, still presents traces of a helical formation, while the two mutated structures completely lost the helical structure up to four residues upstream the polyXY (Fig. 4B).

Using the detailed propensities along the MD trajectory, we can analyze the behavior of specific residues in the helical segment surrounding the polyER (Figure S1). Generally, some instability can be seen in the helical region of six residues preceding the polyER starting at 350 ns, with most residues retaining some degree of helical propensity (Figure S1A). A similar pattern can be observed for the first four residues of the polyER, but the last two residues and the next two lose their helical propensity already at 50 ns (Figure S1B). The MS trajectories for the sequence mutated to polyEK present a different scenario (Figure S2A). The three residues preceding the polyEK lose their helical structure at 200 ns (Figure S2A), and the polyEK and following residues at 50 ns (Figure S2B). A similar scenario is observed in the MS trajectories for the polyER sequence mutated to polyDK, where the three residues preceding the polyDK, the LCR itself and the following two residues lose the helical structure at 150 ns (Figure S3A-B).

RMSD and SASA, two canonical MD simulation metrics (see Methods for details), were also generated to allow comparison between the polyXY regions and the adjacent helix, which presented differences in helical propensities between the three simulations. While there are significant differences in helical propensity between the original and mutated polyXY (Figure S4A), the RMSD and SASA values per residue are very similar. Across all three simulations, the RMSD and standard deviation values are found to be higher in the polyXY than in the preceding helix but lower than in the following IDR residues (Figure S4B), while SASA values are highest in the polyXY (Figure S4C). We interpret this result as indicating that changes in the sequence specifically affect the helical structural propensity of the sequence, without altering the primary properties that describe the dynamics of the sequence in the simulation.

These results suggest that the short polyER contributes to the

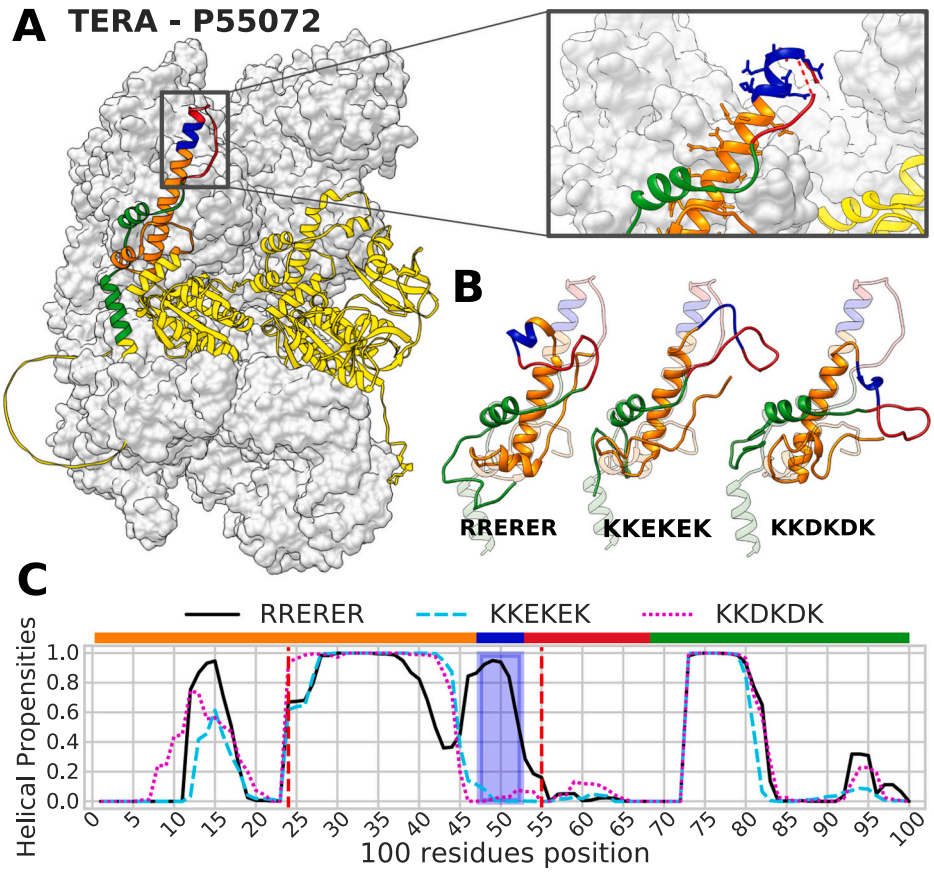

Fig. 4. TERA structure and MD simulation conformations after 1.2 μs for original and mutated polyER. (A) Electron microscopy of the solved structure of the TERA hexamer (PDB:7bpa) [34], with the AlphaFold2 predicted structure for TERA superimposed on chain D of the solved structure. Colors indicate: polyER (blue), IDR (red), and the 100-residue region included in the MD simulations (Orange and green for N- and C-terminus, respectively). The inset shows the original solved structure: the discontinuous line indicates missing residues, suggesting that some C-terminal residues of the polyER might adopt a flexible conformation. (B) The structure of the 100-residue region surrounding the polyER was predicted with AlphaFold2 for the original sequence and for two mutated polyXYs as indicated (faded colors) and MD simulations were applied for 1.2μs (final structures in stronger colors). (C) Helical propensities (likelihood of a given residue to adopt helical conformation) during MD simulations. The blue box delimits the polyER region and the red dashed vertical lines indicate the positions where the helix originally started and ended in the AlphaFold2 prediction. The first 100 ns were discarded to avoid a possible bias in the structure originally predicted by Alphafold2.

stability of the preceding helical residues and replacement by other residues, even with similar physicochemical properties, does not favor the same structure. The partial loss of structure upstream the polyER observed in the simulation of the original sequence may be the result of the absence of interacting partners, which are present in all solved structures of this hexameric protein. While the polyER does not directly interact with the neighboring chain, the two glutamic acid residues in the upstream helical region do, making the helical formation at this region essential for the stability of the protein [34]. Taken together, the stability observed in the original sequence and the AlphFold2 prediction reinforce the traits of a region with inherent flexibility and helical propensities.

3.4. MD simulations of other polyXYs with helical propensities

To extend the results observed with TERA, we selected examples of other frequent polyXYs with high helical coverage, situated at helical IDR N-termini, all with a median pLDDT within the polyXY region greater than 70 (see Methods for details): one additional polyER, two polyEK, and two polyEQ (Table S4), and subjected them to MD simulations. Unlike TERA, solved structures are not available for the helical region or the corresponding polyXYs, and we relied on AlphaFold2 for structural information.

The Heparan-sulfate 6-O-sulfotransferase 3 protein (HS6ST3; UniProt AC: Q8IZP7) is a 6-O-sulfation enzyme that catalyzes the transfer of a sulfate of heparan sulfate. The polyER (RREERR) targeted in our analysis belongs to the C-terminal IDR of the protein, which follows a sulfotransferase domain (Table S4). Due to its position within the sequence, the simulation was performed on an 85-residue fragment. While the AlphaFold2 predicted structure features a 46-residue long helix spanning almost half of the IDR and its N-terminal region, our resulting MD structure shows a shorter helical formation, with the C-terminal region of the polyER completely unfolded along with most of the N-terminal fragment of the IDR (Fig. 5A). When evaluating the helical propensities of the whole simulation, the same scenario can be observed, with a decrease of the helical structure at the end of the polyER, remaining helical around 60 % of the trajectories (Fig. 5B). By evaluating the helical stability of the polyER and the next two residues over time (Figure S5), we can observe that the C-terminal half of the polyER had a reduction in helical propensities for most of the simulation, with a complete loss between 400 and 550 ns, later recovering its helical structure around 800 ns and remaining helical until the end of the simulation.

Our first polyEK sample (KKEKKEKK) comes from the RNA polymerase-associated protein RTF1 homolog (UniProt AC: Q92541). The 100-residue region surrounding the polyEK features a long IDR, which spans 59 residues, including the polyEK and the whole C-terminus of the segment. The resulting MD simulation structure shows helical coverage only on the two last residues of the polyEK (Fig. 6A). The residues in the polyEK present values of helical propensity greater than

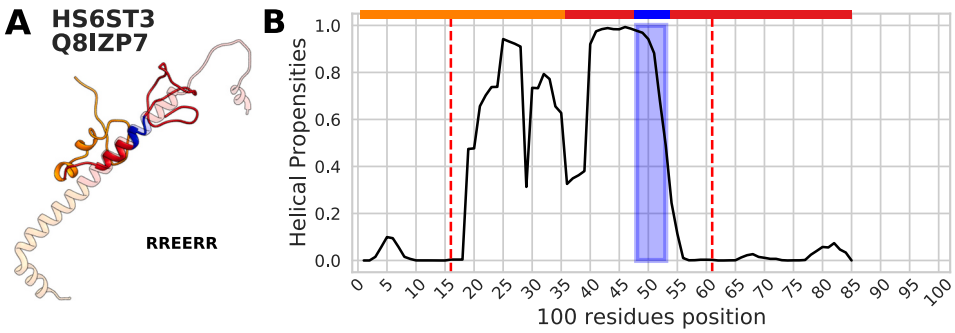

Fig. 5. MD simulation of a segment of HS6ST3 containing a polyER region. (A) Superimposition of the original AlphaFold2 prediction for a fragment of 85 residues surrounding the polyER region (faded colors) with the structure resulting from MD simulations after 1.2 μs (shown in stronger colors). Colors indicate the polyER region (blue), the remainder of the IDR (red), and the region reaching the 50 residues N-terminal to the polyER (orange). The segment includes the end of the protein sequence. (B) Helical propensities (likelihood of a given residue to adopt helical conformation) during MD simulation (first 100 ns were discarded). The blue box delimits the polyXY region and the red dashed vertical lines indicate the position where the helix originally started and ended in the AlphaFold2 prediction. The color bar above delimits the fragment regions, as described in (A).

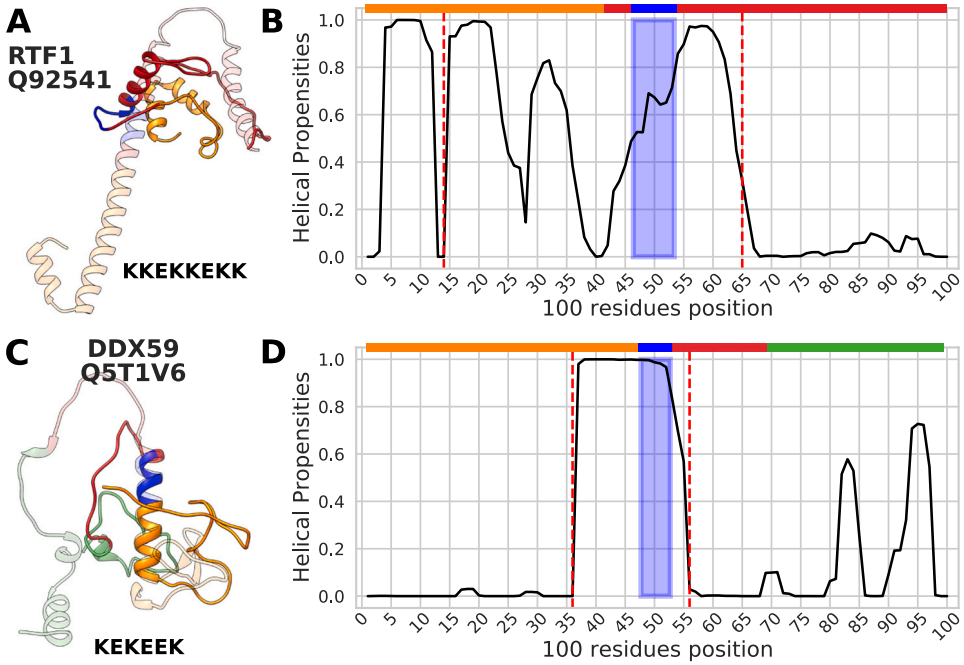

Fig. 6. MD simulations for two segments containing a polyEK region. (A) Superimposition of the original AlphaFold2 prediction for a fragment of 100 residues of RTF1 surrounding the polyEK region (faded colors) with the structure resulting from MD simulations after 1.2 μs (shown in stronger colors). Colors indicate the polyEK region (blue), the remainder of the IDR (red) and the region reaching the 50 residues N-terminal to the polyEK, which also contains the upstream helix (orange). (B) Helical propensities (likelihood of a given residue to adopt helical conformation) during MD simulation of RTF1 (first 100 ns were discarded). The blue box delimits the polyXY region and the red dashed vertical lines indicate the position where the helix originally started and ended in the AlphaFold2 prediction. The color bar above delimits the fragment regions as described in (A). (C) Superimposition of the original AlphaFold2 prediction for a fragment of 100 residues of DDX59 surrounding the polyEK region (faded colors) with the structure resulting from the MD simulations after 1.2 μs (shown in stronger colors). Colors indicate the polyEK region (blue), the remainder of the IDR (red), the region reaching to 50 residues to the N-terminal of the polyEK center, which also contains the preceding helix (orange) and the region reaching to the 50 residues to the C-terminal of the polyEK center (green). (D) Helical propensities (likelihood of a given residue to adopt helical conformation) during MD simulation of DDX59 (first 100 ns were discarded). The blue box delimits the polyXY region and the red dashed vertical lines indicate the position where the helix originally started and ended in the AlphaFold2 prediction. The color bar above delimits the fragment regions as described in (C).

50 % (Fig. 6B). However, MD trajectories show that at 1 μs only the last two residues remained helical (Figure S6B). All residues in the adjacent region at the N-terminus of the sequence also completely lost helical coverage at 1 μs (Figure S6A). An EK-rich region downstream of the polyEK exhibits a high helical propensity (KKQEEEQE) and remained mostly helical throughout the whole simulation (Figure S6C). This region could act as an extension of the polyEK region considered.

The second polyEK analyzed came from DDX59 (UniProt AC: Q5T1V6), a probable ATP-dependent RNA helicase. The center of the targeted polyXY region (KEKEEK) is located five residues from the end of the N-terminal helix, with a Zinc finger domain ending eight residues before the start of the polyEK (Table S4). The polyEK region presented a remarkably stable helical composition throughout the whole 1.2 μs, with some loss of structure at the last residue of the polyXY and the three following residues (Fig. 6C-D). This loss occurred as early as 100 ns into the simulation (Figure S6D).

The first selected polyEQ (QQEQEQ) was from PIEZO1 (UniProt AC: Q92508), a pore-forming subunit of the cation Piezo channel. After 1.2 μs of MD simulation, the helical segment which includes the polyEQ and some surrounding residues remains stable (Fig. 7A). The helical

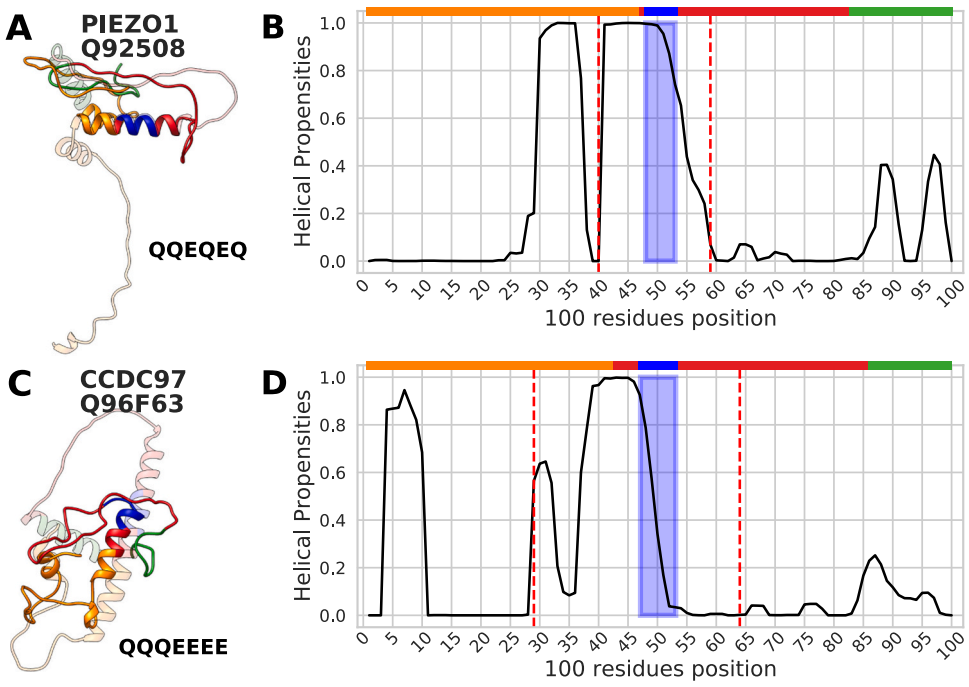

Fig. 7. MD simulations for two segments containing a polyEQ region. (A) Superimposition of the original AlphaFold2 prediction for a fragment of 100 residues of PIEZO1 surrounding the polyEQ region (faded colors) with the structure resulting from the MD simulations after 1.2 μs (shown in stronger colors). Colors indicate the polyEQ region (blue), the remainder of the IDR (red), the region reaching to 50 residues to the N-terminal of the polyEQ center, which also contains the upstream helix (orange) and the region reaching up to 50 residues to the C-terminus of the polyEQ center (green). (B) Helical propensities (likelihood of a given residue to adopt helical conformation) during MD simulation of PIEZO1 (first 100 ns were discarded). The blue box delimits the polyXY region and the red dashed vertical lines indicate the position where the helix originally started and ended in the AlphaFold2 prediction. The color bar above delimits the fragment regions as described in (A). (C) Superimposition of the original AlphaFold2 prediction for the 100 residues of CCDC97 surrounding the polyEQ region (faded colors) with the structure resulting from the MD simulations after 1.2 μs (shown in stronger colors). Colors indicate the polyEQ region (blue), the remainder of the IDR (red), the region reaching 50 residues to the N-terminal of the polyEQ center, which also contains the upstream helix (orange) and the region reaching 50 residues to the C-terminal of the polyXY center (green). (D) Helical propensities for the duration of the MD simulation of CCDC97 (first 100 ns were discarded). The blue box delimits the polyXY region and the red dashed vertical lines indicate the position where the helix originally started and ended in the AlphaFold2 prediction. The color bar above delimits the fragment regions as described in (C).

propensity is high in the first section of the polyEQ, with a 70 % reduction in the final section and lower after the polyEQ (Fig. 7B). We can observe the loss of helical structure in the last three residues of the polyEQ between 400 ns and 550 ns, with complete recovery at the end of the trajectory (Figure S7A). In the same time interval, the five residues after the polyEQ also lost their helical structure, gradually recovering thereafter (Figure S7B).

Our last sample, a second polyEQ (QQQEEEE), comes from CCDC97 (UniProt AC: Q96F63). In it, helical stability is observed in the glutamine portion of the polyEQ, which maintained a propensity greater than 40 % throughout the simulation as did the region upstream the polyEQ (Fig. 7C-D). Interestingly, among all the segments chosen for simulations, this is the only one that combines two homorepeats and also presented the highest rate of helical content loss across the entire polyXY region. Additional simulations would be required to investigate the relationship between the order of residues within a polyXY and its helical stability. The glutamic acid segment of the polyEQ lost most of its initial conformation during the simulation, as early as 200 ns, with a gradual recovery at the end (Figure S7C). The polyE following the polyEQ, predicted by AlphaFold2 to be helical, completely lost its conformation in the early stages of the simulation (Figure S7D).

In summary, most of the polyXYs studied in the selected fragments presented helical propensity values above 50 % at the end of the 1.2 μs simulation, with a tendency to lose structure at the C-terminus (away from the N-terminally situated adjacent helix). These findings support the AlphaFold2 predictions in these particular regions. Predictions for the helix preceding the polyXY were less stable, except for DDX59 (Fig. 6C-D) and PIEZO1 (Fig. 7A-B).

4. Discussion

In previous work, short LCRs composed of one or two amino acid types, denominated polyX and polyXYs, respectively, were characterized by exploring their position within protein sequences and IDRs, residue preferences, categories and protein biological functions [37,38]. Their structural properties when within IDRs were assessed for the human proteome, initially through homology to PDB structures, with some residues presenting helical propensities, including E and K, alone or associated, or helical avoidance, including polyEP and polyGX [22]. AlphaFold2 predictions increased the pool of structured candidates, not only confirming our previous findings for helical-prone structures, but also adding some new LCR types to this list, such as polyQ and polyER [23].

The importance of studying helical propensities, also called residual structures, within IDRs has been explored. The capacity to rapidly adopt a specific secondary structure upon-binding, a folding pathway called conformational selection, would increase the affinity of interactions between some disordered regions and their partners [39]. There are debates on how these regions benefit the formed complexes, either by enhancing the complex association rates and molecular recognition [40] or by reducing dissociation [41]. Nonetheless, these propensities are important to optimize IDR interactions. Short LCRs are not the only regions capable of exhibiting such propensities, but they may allow their occurrence. Shorter LCRs are often inserted in longer, less complex LCR tracts [42]. A better understanding of their local structural behavior can provide invaluable insights into how they contribute to the function of longer biased regions. Despite advances in the qualification and

quantification of these short LCRs within IDRs, the understanding of their specific local structural function still evades us.

AlphaFold2 represented a significant advance in current knowledge about the relationship between IDRs, LCRs and secondary structure. However, its predictions are known to sometimes be unrealistic and overrepresent helical formations. These characteristics can be especially noticed in IDR regions, where low pLDDT scores indicate a lack of confidence in the adopted structures [19] and in the large increase of helical structures within IDRs [43,44], explained as the result of training on a large set of protein-bound structures, thus not representing the unbound structure in solution [21,45,46].

These new challenges presented by AlphaFold2 provided an opportunity to evaluate the behavior of polyXYs within IDRs and to assess their local function. Here, we propose a specific function for a specific group of polyXYs: the extension and maintenance of helical structures formed outside the IDR region. Addressing this specific scenario would also bring some insight into the behavior of a transition region, often neglected by studies targeting only IDRs or globular regions separately. The amino acid composition of IDRs' neighboring regions has recently been explored, presenting characteristics that suggest a different nature compared to disordered and ordered regions, evolved and selected to support structural transitions allowing order to become disordered [47]. In this work, we explored the reverse perspective: how ordered regions can bring order to disordered regions and why this transition is meaningful.

4.1. The categorization of a subgroup of polyXYs within IDRs

To address our hypothesis, we first detected fitting candidates for our analysis, which were naturally found in a specific group of polyXYs in the vicinity of helices that span the ends of the IDR. This subset of polyXYs presents a similar composition at both IDR ends yet showing greater consistency in the N-terminal samples, an expected result of the natural direction of the synthesis of proteins at the ribosome [33]. AlphaFold2 pLDDT values within these regions also exhibit the expected behavior, considering that these are transition regions between order and disorder, showing higher values gradually fading when helices are present at the IDR ends, compared to cases where no helices are present.

By further analyzing the structural composition of the two groups in the targeted LCRs, with "high helical coverage" and "low/no helical coverage", we explored their composition, looking for distinctions between this subset of polyXY compared to the global scenario observed in our previous work. The same polyXY types were detected in this subgroup, however the preferences in helical composition differed: (i) polyDE mainly avoided helices, even when they were present at the IDR ends, (ii) polyEK presented mixed preferences, and (iii) polyER was overwhelmingly covered by helices when a neighboring helix was present, reaching more than 95 % coverage. Different polyXYs show a higher relevance in this subgroup, compared to the whole set of polyXYs covered by helices, with polyAE showing the same versatility as polyEK, and both polyKQ and polyEQ also showing a higher preference for helical coverage.

Furthermore, we demonstrate that some residues have different frequencies within polyXYs with helical coverage in the vicinity of neighboring IDRs; E, K, R, A and Q show preference; and D, P, S and G avoid these structures. This proposal is not new, as some of the involved amino acid residues are related to helical propensity in polyX [48,49]; however, here we show which types of polyXY are common or avoided in the specific scenario where helices are present in the proximity of IDRs according to AlphaFold2 predictions. This characterization may not only be important to define this target group, but may also support future sub-categorizations into other types of polyXYs within IDRs with helical propensities.

Interestingly, we found significantly smaller distances of polyXY regions covered by neighboring helices at helical N-terminal IDR ends but not at C-terminal IDR ends. Again, we take this as deriving from the direction of protein synthesis [33]: it might be easier to expand an already folded helix into an IDR during its synthesis, than propagating a helix into an already completed and flexible IDR.

4.2. Reinforcing the helical preferences predicted by AlphaFold2 and the role of polyXYs in local helical stability

The choice to run MD simulations to explore the proposed AlphaFold2 helical predictions in IDR regions came naturally, as an increasingly used technique to explore the conformational variety of the structural behavior of IDRs [50]. Our findings with the native TERA protein and its mutated polyXY versions suggest that the formation of a polyER at the end of the N-terminal helix acts as a stabilizer, ensuring that the upstream helix, essential for interaction, remains at least partially stable in our simulations. The mutated versions of TERA's polyER (polyEK and polyDK) indicated that the presence of a polyER in the region is absolutely essential for the adoption and maintenance of helical structure at the polyXY and at the residues immediately adjacent to the IDR, residues that present direct intermolecular interactions with the hexamer's partner protein.

To further explore the helical stability of not only highly helical polyER but also other polyXY types with helical propensities, we performed simulations for five additional polyXY containing segments: one polyER, two polyEK, and two polyEQ. Each MD simulation presents distinctive characteristics that require additional and focused exploration and understanding. Nonetheless, all of them showed a common aspect: they reinforced, at least partially, the helical preference of the region as predicted by AlphaFold2 and, more importantly for our analysis, they indicate that in all scenarios the helical conformation is sustained in the polyXY region.

Almost all samples showed a degradation of the helical structure towards the end of the polyXY, which could suggest that these short LCRs facilitate the extension of secondary structure into the IDR but not beyond the LCR. Only the simulation of RTF1 (polyEK) showed an increasing gradient of helical propensity towards the end of the polyXY region. The presence of a stable EK-rich region (Fig. 6A-B and Figure S6C) downstream the polyEK suggests that the definition of the polyEK region might have been too restrictive in this case.

4.3. Future perspectives and Conclusions

The functions of IDRs are largely studied in the context of how they contribute to protein interactions and secondary structure formation. A deeper knowledge of how different types of LCRs contribute to these structures is still lacking. In this work, we explored how specific types of polyXYs at the termini of IDRs may contribute to the stability of the helical secondary structure of regions preceding IDRs, especially when they are composed of positively charged residues paired with helical prone residues, such as alanine and glutamine. Our findings also support an explanation for why polyXYs accumulate at IDRs termini and why it is evolutionarily favorable to keep these short LCRs conserved in these regions.

Despite showing an interesting perspective into the local structural function of polyXYs within IDRs in the vicinity of helical structures, our work has some limitations.

First, the number of polyXYs with these characteristics was limited by the specificity of the scenario and the definitions chosen. This limitation eased the selection of suitable targets for MD simulation, but also limited our ability to further explore the amino acid composition of the surrounding region for additional evidence on why these regions may support helical structures and the interplay between the polyXY and its neighboring residues.

In addition, the number of suitable cases can greatly benefit from the addition of polyXYs with the same characteristics in proteins from other eukaryotic species without homology to human proteins. Increasing the number of cases and analyzing the surrounding amino acid composition

could also strengthen the findings derived from the MD simulations, which is not possible at this time with our results, due to the diverse scenarios presented by the targeted samples. With additional cases, we could then explore their domain context, binding regions, and charge composition, which could shed light on further categorization underlying the types of polyXYs analyzed. The search for longer LCRs in which these polyXYs are embedded, considering charge compositional bias and other different LCR compositions, could also further support this characterization.

Second, our approach was strongly affected by its scalability. MD simulations, despite being highly informative, remain extremely restrictive. Our choice to limit the simulation to a fraction of the protein reduced run times and ensured successful runs; however, it limited our understanding of the behavior of the entire protein. These fragments are impacted by long-range interactions [51], which cannot be accounted for in our analysis. For example, in Fig. 4C, a secondary structure element was observed at positions 5–20 with variable helical propensity; determining whether this variability is influenced by the nearby polyXY-associated helix would require modeling a larger fragment, even including interacting proteins.

Advances in computational power and the association of different techniques along with ways to reduce MD simulation runtimes [52,53] and the increasing development of force fields more suited for proteins with mixed ordered and disordered regions [28,54] may allow these limitations to be overcome in the future. The association of MD simulations with machine learning models could also improve the quality of the results, providing additional validation [55].

Additionally, new avenues can be explored to answer the questions raised by this work. Examples include further analyses of polyEK and polyAE's dual preferences for helical and flexible content (Fig. 3), and the analysis of flexible polyXYs in close vicinity to the external helix end, which could act as barriers to uncontrolled helical extension. Indeed, a systematic study of the helical stability of all polyXY cases with helical content predicted by AlphaFold2 would be desirable, but is currently beyond our computational capabilities.

We believe that our findings, which combine computational analyses and MD simulations pave the way for a better understanding of the relationship between short LCRs and structural propensity within IDRs. We have presented approaches that should translate future advances in protein structure resolution and prediction, as well as more powerful and specialized MD simulations, into a greater understanding of the structural dynamics of flexible regions in proteins.

Competing Interests Statement

The authors declare no competing interests.

Funding

This work received support from the following institutions: Mainz Institute of Multiscale Modeling (M3ODEL) for F.S. and M.A.A.N. for the position of M.G.K; L.A.B. gratefully acknowledges institutional funding by the Max Planck Society.

CRediT authorship contribution statement

Andrade Miguel: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Friederike Schmid:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition, Data curation, Conceptualization. **Baptista Luis Andre:** Writing – review & editing, Validation, Methodology. **Goncalves-Kulik Mariane:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgments

The authors gratefully acknowledge the computing time granted on the supercomputer MOGON 2 at Johannes Gutenberg University Mainz (hpc.uni-mainz.de) for the AlphaFold2 predictions and MD simulations.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbr.2025.100054.

Data availability

Data will be made available upon request.

References

[1] van der Lee R, et al. Classification of intrinsically disordered regions and proteins. Chem Rev Jul. 2014;114(13):6589–631. https://doi.org/10.1021/cr400525m.
[2] Uversky VN. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta May 2013;1834(5):932–51. https://doi.org/10.1016/j.bbapap.2012.12.008.
[3] Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol Jun. 2015;32:102–12. https://doi.org/10.1016/j.sbi.2015.03.008.
[4] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci USA Aug. 2013;110(33):13392–7. https://doi.org/10.1073/pnas.1304749110.
[5] Mier P, et al. Disentangling the complexity of low complexity proteins. Brief Bioinform Mar. 2020;21(2):458–72. https://doi.org/10.1093/bib/bbz007.
[6] Mier P, Elena-Real CA, Cortés J, Bernadó P, Andrade-Navarro MA. The sequence context in poly-alanine regions: structure, function and conservation. Bioinforma Oxf Engl Oct. 2022;38(21):4851–8. https://doi.org/10.1093/bioinformatics/btac610.
[7] Chavali S, Singh AK, Santhanam B, Babu MM. Amino acid homorepeats in proteins. Nat Rev Chem Aug. 2020;4(8):420–34. https://doi.org/10.1038/s41570-020-0204-1.
[8] Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans Oct. 2016;44(5):1185–200. https://doi.org/10.1042/BST20160172.
[9] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science Sep. 2017;357(6357):eaaf4382. https://doi.org/10.1126/science.aaf4382.
[10] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol Jan. 2015;16(1):18–29. https://doi.org/10.1038/nrm3920.
[11] Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet May 2018;19(5):286–98. https://doi.org/10.1038/nrg.2017.115.
[12] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008;37:215–46. https://doi.org/10.1146/annurev.biophys.37.032807.125924.
[13] Liu W, et al. Visualizing single-molecule conformational transition and binding dynamics of intrinsically disordered proteins. Nat Commun Aug. 2023;14(1):5203. https://doi.org/10.1038/s41467-023-41018-x.
[14] Nwanochie E, Uversky VN. Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. Int J Mol Sci Aug. 2019;20(17):4186. https://doi.org/10.3390/ijms20174186.
[15] Mu J, Liu H, Zhang J, Luo R, Chen H-F. Recent force field strategies for intrinsically disordered proteins. J Chem Inf Model Mar. 2021;61(3):1037–47. https://doi.org/10.1021/acs.jcim.0c01175.
[16] Rahman MU, Rehman AU, Liu H, Chen H-F. Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins. J Chem Inf Model Oct. 2020;60(10):4912–23. https://doi.org/10.1021/acs.jcim.0c00762.
[17] Liu H, Song D, Zhang Y, Yang S, Luo R, Chen H-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. Phys Chem Chem Phys PCCP Oct. 2019;21(39):21918–31. https://doi.org/10.1039/c9cp03434j.
[18] Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature Aug. 2021;596(7873):583–9. https://doi.org/10.1038/s41586-021-03819-2.
[19] Ruff KM, Pappu RV. AlphaFold and Implications for Intrinsically Disordered Proteins. J Mol Biol Oct. 2021;433(20):167208. https://doi.org/10.1016/j.jmb.2021.167208.
Necci M, Piovesan D, Predictors CAID, Curators DisProt, Tosatto SCE. Critical assessment of protein intrinsic disorder prediction. Nat. Methods 2021;18(5):472–81. https://doi.org/10.1038/s41592-021-01117-3.

Alderson TR, Pritišanac I, Kolarić Đ, Moses AM, Forman-Kay JD. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. Proc. Natl. Acad. Sci. U. S. A. 2023;120(44). https://doi.org/10.1073/pnas.2304302120. e2304302120.
[22] Gonçalves-Kulik M, et al. Low Complexity Induces Structure in Protein Regions Predicted as Intrinsically Disordered. Biomolecules Aug. 2022;12(8):1098. https://doi.org/10.3390/biom12081098.
[23] Gonçalves-Kulik M, Schmid F, Andrade-Navarro MA. One Step Closer to the Understanding of the Relationship IDR-LCR-Structure. Genes Aug. 2023;14(9):1711. https://doi.org/10.3390/genes14091711.
[24] Piovesan D, et al. MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res Jan. 2023;51(D1):D438–44. https://doi.org/10.1093/nar/gkac1065.
[25] Varadi M, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res Jan. 2022;50(D1):D439–44. https://doi.org/10.1093/nar/gkab1061.
[26] Varadi M, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res Jan. 2024;52(D1):D368–75. https://doi.org/10.1093/nar/gkad1011.
[27] Touw WG, et al. A series of PDB-related databanks for everyday needs (no. Database issue) Nucleic Acids Res Jan. 2015;43:D364–368. https://doi.org/10.1093/nar/gku1028.
[28] Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. pp. E4758–E4766 Proc Natl Acad Sci USA May 2018;115(21). https://doi.org/10.1073/pnas.1800690115.
[29] Liu H, Song D, Lu H, Luo R, Chen H-F. Intrinsically disordered protein-specific force field CHARMM36IDPSFF. Chem Biol Drug Des Oct. 2018;92(4):1722–35. https://doi.org/10.1111/cbdd.13342.
[30] MacKerell AD, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B Apr. 1998;102(18):3586–616. https://doi.org/10.1021/jp973084f.
[31] Cock PJA, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinforma Oxf Engl Jun. 2009;25(11):1422–3. https://doi.org/10.1093/bioinformatics/btp163.
[32] Pettersen EF, et al. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem Oct. 2004;25(13):1605–12. https://doi.org/10.1002/jcc.20084.
[33] Voorhees RM, Ramakrishnan V. Structural basis of the translational elongation cycle. Annu Rev Biochem 2013;82:203–36. https://doi.org/10.1146/annurev-biochem-113009-092313.
[34] Zhu K, et al. The phosphorylation and dephosphorylation switch of VCP/p97 regulates the architecture of centrosome and spindle. Cell Death Differ Oct. 2022;29(10):2070–88. https://doi.org/10.1038/s41418-022-01000-4.
[35] van den Boom J, Meyer H. VCP/p97-Mediated Unfolding as a Principle in Protein Homeostasis and Signaling. Mol Cell Jan. 2018;69(2):182–94. https://doi.org/10.1016/j.molcel.2017.10.028.
[36] Meyer H, Bug M, Bremer S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. Nat Cell Biol Feb. 2012;14(2):117–23. https://doi.org/10.1038/ncb2407.
[37] Mier P, Andrade-Navarro MA. PolyX2: Fast Detection of Homorepeats in Large Protein Datasets. Genes Apr. 2022;13(5):758. https://doi.org/10.3390/genes13050758.
[38] Mier P, Andrade-Navarro MA. Regions with two amino acids in protein sequences: A step forward from homorepeats into the low complexity landscape. Comput Struct Biotechnol J 2022;20:5516–23. https://doi.org/10.1016/j.csbj.2022.09.011.
[39] Yang J, Gao M, Xiong J, Su Z, Huang Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. Protein Sci Publ Protein Soc Nov. 2019;28(11):1952–65. https://doi.org/10.1002/pro.3718.
[40] Iešmantavičius V, Dogan J, Jemth P, Teilum K, Kjaergaard M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. Angew Chem Int Ed Engl Feb. 2014;53(6):1548–51. https://doi.org/10.1002/anie.201307712.
[41] Crabtree MD, Borcherds W, Poosapati A, Shammas SL, Daughdrill GW, Clarke J. Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein: Target Affinity by Altering the Lifetime of the Bound Complex. Biochemistry May 2017;56(18):2379–84. https://doi.org/10.1021/acs.biochem.7b00179.
[42] Kastano K, et al. Evolutionary Study of Disorder in Protein Sequences. Biomolecules Oct. 2020;10(10):1413. https://doi.org/10.3390/biom10101413.
[43] Triandafillou CG, Pan RW, Dinner AR, Drummond DA. Pervasive, conserved secondary structure in highly charged protein regions. PLoS Comput Biol Oct. 2023;19(10):e1011565. https://doi.org/10.1371/journal.pcbi.1011565.
[44] Bruley A, Bitard-Feildel T, Callebaut I, Duprat E. A sequence-based foldability score combined with AlphaFold2 predictions to disentangle the protein order/disorder continuum. Proteins Apr. 2023;91(4):466–84. https://doi.org/10.1002/prot.26441.
[45] Anbo H, Sakuma K, Fukuchi S, Ota M. How AlphaFold2 Predicts Conditionally Folding Regions Annotated in an Intrinsically Disordered Protein Database, IDEAL. Biology Jan. 2023;12(2):182. https://doi.org/10.3390/biology12020182.
[46] Versini R, et al. A perspective on the prospective use of AI in protein structure prediction. J Chem Inf Model Jan. 2024;64(1):26–41. https://doi.org/10.1021/acs.jcim.3c01361.
[47] Basu S, Bahadur RP. Do sequence neighbours of intrinsically disordered regions promote structural flexibility in intrinsically disordered proteins? J Struct Biol Feb. 2020;209(2):107428. https://doi.org/10.1016/j.jsb.2019.107428.
[48] Khaled M, Strodel B, Sayyed-Ahmad A. Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers. Front Mol Biosci 2023;10:1143353. https://doi.org/10.3389/fmolb.2023.1143353.
[49] Hong J-Y, et al. Structural and dynamic studies reveal that the Ala-rich region of ataxin-7 initiates α-helix formation of the polyQ tract but suppresses its aggregation. Sci Rep May 2019;9(1):7481. https://doi.org/10.1038/s41598-019-43926-9.
[50] Thomasen FE, Lindorff-Larsen K. Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins. Biochem Soc Trans Feb. 2022;50(1):541–54. https://doi.org/10.1042/BST20210499.
[51] Minor DL, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. Nature Apr. 1996;380(6576):730–4. https://doi.org/10.1038/380730a0.
[52] Yang YI, Shao Q, Zhang J, Yang L, Gao YQ. Enhanced sampling in molecular dynamics. J Chem Phys Aug. 2019;151(7):070902. https://doi.org/10.1063/1.5109531.
[53] Abriata LA, Dal Peraro M. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization. Comput Struct Biotechnol J 2021;19:2626–36. https://doi.org/10.1016/j.csbj.2021.04.050.
[54] Tian C, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J Chem Theory Comput Jan. 2020;16(1):528–52. https://doi.org/10.1021/acs.jctc.9b00591.
[55] Sisk TR, Robustelli P. Folding-upon-binding pathways of an intrinsically disordered protein from a deep Markov state model. Proc Natl Acad Sci USA Feb. 2024;121(6):e2313360121. https://doi.org/10.1073/pnas.2313360121.