A glutamine-based single-helix scaffold to target globular proteins

Escobedo, Albert; Piccirillo, Jonathan; Aranda, Juan; Diercks, Tammo; Mateos, Borja; Garcia-Cabau, Carla; Sanchez-Navarro, Macarena; Topal, Busra; Biesaga, Mateusz; Staby, Lasse; Kragelund, Birthe B.; Garcia, Jesus; Millet, Oscar; Orozco, Modesto; Coles, Murray; Crehuet, Ramon; Salvatella, Xavier

Published in:
Nature Communications

DOI:
10.1038/s41467-022-34793-6

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Escobedo, A., Piccirillo, J., Aranda, J., Diercks, T., Mateos, B., Garcia-Cabau, C., Sanchez-Navarro, M., Topal, B., Biesaga, M., Staby, L., Kragelund, B. B., Garcia, J., Millet, O., Orozco, M., Coles, M., Crehuet, R., & Salvatella, X. (2022). A glutamine-based single-helix scaffold to target globular proteins. Nature Communications, 13, [7073]. https://doi.org/10.1038/s41467-022-34793-6
The binding of intrinsically disordered proteins to globular ones can require the folding of motifs into α-helices. These interactions offer opportunities for therapeutic intervention but their modulation with small molecules is challenging because they bury large surfaces. Linear peptides that display the residues that are key for binding can be targeted to globular proteins when they form stable helices, which in most cases requires their chemical modification. Here we present rules to design peptides that fold into single α-helices by instead concatenating glutamine side chain to main chain hydrogen bonds recently discovered in polyglutamine helices. The resulting peptides are uncharged, contain only natural amino acids, and their sequences can be optimized to interact with specific targets. Our results provide design rules to obtain single α-helices for a wide range of applications in protein engineering and drug design.
peptides have a low propensity to fold into stable α-helices, however, and the entropic cost of folding decreases both their affinity for their targets and their stability against proteolytic degradation, highlighting the need to develop new tools to stabilize their helical conformation.

Introducing non-natural amino acids that act as potent N-caps or substituting the \(i+i+4\) hydrogen bonds stabilizing this secondary structure by covalent, and therefore permanent, surrogates can be used to achieve this goal. Alternatively, specific amino acids can be introduced at positions \(i\) and \(i+3\), and \(i+4\) or \(i+7\), that are close in space in α-helices, and linked by different means\(^{20}\) such as by peptide stapling\(^{11,12}\). Peptide stapling is based on the use of pairs of synthetic α-methyl, α-alkenyl amino acids at relative positions \(i+3\), \(i+i+4\) or \(i+i+7\) where their side chains can react by ring-closing metathesis when the peptide folds into an α-helix, thus greatly stabilizing this conformation\(^{11,12}\). In some cases, such connections have been concatenated to obtain especially long and stable helices\(^{13}\). Although these approaches have shown applicability to inhibit protein-protein interactions, they have drawbacks that limit their range of applicability, such as their costly synthesis, limited solubility, and high rigidity.

We recently reported that Gln side chain to main chain hydrogen bonds stabilize the polyglutamine (polyQ) helix, a helical secondary structure based on the main chain NMR chemical shifts\(^{14}\). This approach offers several advantages over peptide stapling: it greatly reduces the entropic cost of folding, that Agadir\(^{17}\). This algorithm, similar to the Zimm-Bragg\(^{18}\) and Lifson-Jacobs\(^{19}\) helix-coil transition models, is based on statistical mechanics. The architectural weight of the helical state of any peptide segment depends on the free energy of it folding into an α-helix, that Agadir computes as the sum of different terms including one accounting for interactions between residues at positions \(i\) and \(i+4\), close in space in α-helices, that requires experimental parametrization. The current version of Agadir does not account for the Gln to Leu (\(\text{Gln}_{i+4} \rightarrow \text{Leu}_i\)) side chain to main chain interactions stabilizing polyQ helices and therefore underestimates the helicity of the polyQ tract in AR: peptide L4Q16, excised from AR and harboring four such interactions, has 38% helical propensity according to NMR experiments while Agadir predicts only 38%\(^{14}\). To address this we introduced an additional term to the free energy of folding into an α-helix accounting for this interaction (\(\Delta G_{\alpha\rightarrow\alpha}^{\text{GLQ}}\)) and by minimizing the prediction error with respect to the NMR-derived helicity (RMSD\(_{\text{Hel}}\)) see Supplementary Methods) obtained \(\Delta G_{\alpha\rightarrow\alpha}^{\text{GLQ}}\approx 0.6\) kcal mol\(^{-1}\) for L4Q16, in the range expected for one hydrogen bond in water\(^{20}\) (Supplementary Fig. 1a). To our surprise we found that the value of \(\Delta G_{\alpha\rightarrow\alpha}^{\text{GLQ}}\) that minimizes RMSD\(_{\text{Hel}}\) depends on polyQ tract length: it increases from −0.4 kcal mol\(^{-1}\) for L4Q8 to −0.7 kcal mol\(^{-1}\) for L4Q20 (Supplementary Fig. 1b). The effective strength of the Gln\(_{i+4} \rightarrow \text{Leu}_i\) interactions in AR depends thus on the number of equivalent interactions following them in the sequence, suggesting cooperativity.

To analyze the origin of this behavior and exploit it for peptide design we studied four peptides of identical amino acid composition but with two potential Gln\(_{i+4} \rightarrow \text{Leu}_i\) interactions (pink arrows in Fig. 1a) at different relative positions. The first such interaction is common to all peptides and the second one is shifted 1 (peptide P1-S), 2 (P2-6), 3 (P3-7), or 5 (P5-9) positions towards the C-terminus. After confirming that they were monomeric under our experimental conditions by size exclusion chromatography coupled to multiple angle light scattering (SEC-MALS) (Supplementary Fig. 1d), we used solution-state nuclear magnetic resonance (NMR) spectroscopy to probe the structural properties by exploiting the quantitative dependence of \(\delta_{\text{Cn}}\) and \(\delta_{\text{Hn}}\) NMR chemical shifts on residue-specific helical propensity, where larger \(\delta_{\text{Cn}}\) and lower \(\delta_{\text{Hn}}\) shifts indicate higher helicity\(^{21}\). An analysis of the NMR spectra indicated that in the sequence context of this family of peptides the strength of two Gln\(_{i+4} \rightarrow \text{Leu}_i\) interactions is maximal when the donor of the first and the acceptor of the second share a peptide bond such that the two interactions are concatenated, as in peptide P3-7; these results were confirmed by circular dichroism (CD) spectroscopy (Supplementary Fig. 1c).

We then used CD spectroscopy to study the secondary structure of peptides containing two and three pairs of concatenated Gln\(_{i+4} \rightarrow \text{Leu}_i\) interactions, (P3-7)\(_2\) and (P3-7)\(_3\), and obtained that they are highly helical and monomeric (\(\delta_{\text{O22}mm}/\delta_{\text{O20}mm} < 1\)) (Fig. 1c, d). An analysis of their secondary structure based on the main chain NMR chemical shifts indicated that the residue-specific helical propensity (p\(_{\text{Hel}}\)) is larger than 0.9 across 9 contiguous residues for (P3-7)\(_2\) and larger than 0.95 across 16 residues for (P3-7)\(_3\). We also characterized peptide (P3-7)\(_3\)\(_{C_{18p}}\) in which all Leu hydrogen bond acceptors were substituted by Ala, resulting, despite the higher helical propensity of Ala, in lower helicity (Fig. 1c, d) due to the relatively low stability of side chain to main chain hydrogen bonds accepted by Ala (see below)\(^{22}\). As it is not common for monomeric peptides to cooperatively fold into α-helices in the absence of tertiary interactions, we verified their monomeric state by SEC-MALS and native mass spectrometry (MS) (Supplementary Fig. 1d). In addition, we removed the N-terminal PGAS motif, which can facilitate helix nucleation\(^{1}\), from these peptides (Fig. 1c) and observed that the uncapped counterparts, u(P3-7)\(_2\) and u(P3-7)\(_3\), also have high helical propensity (Supplementary Fig. 1f).

Next, we investigated the thermal stability of the helices by CD spectroscopy at temperatures up to 368 K (Fig. 1e and Supplementary Fig. 1g). The spectra at 278 K were equivalent to those obtained upon cooling after thermal unfolding, indicating that the unfolded state of the peptide is soluble under our experimental conditions. We also characterized the structural properties of (P3-7)\(_3\) by NMR at physiological temperature, 310 K. A comparison of the \(^{13}\)C-detected 2D CACO spectra of (P3-7)\(_3\) at 278 and 310 K revealed only a small decrease in helical propensity (Fig. 1d, e) indicating that the peptide remains essentially fully folded at 310 K (p\(_{\text{Hel}}\)≈ 0.90 across 12 contiguous residues). Figure 1f shows the CD-monitored thermal denaturation of peptides (P3-7)\(_2\) and (P3-7)\(_3\)\(_{C_{18p}}\), which reports on the higher stability of the former. Finally, prompted by the observation that these peptides
remained highly helical at physiological temperature, we also investigated their stability in human serum as well as their internalization in HeLa cells. Both (P3-7)2 and (P3-7)3 control show high stability in human serum, with half-lives over 24 hours, well beyond that of Angiopep-222 that we included as a positive control (Supplementary Fig. 2a); in addition, both peptides were internalized by HeLa cells at 310 K (Supplementary Fig. 2b). In conclusion, concatenating Glni+4 \rightarrow \text{Leui} interactions allows obtaining stable \( \alpha \)-helices that contain only natural amino acids, remain folded under physiological conditions, are soluble upon thermal unfolding, are resistant to proteolytic degradation, and are readily internalized by living human cells.

**Structure of a Gln-based single \( \alpha \)-helix**

The high quality of the NMR spectra obtained for peptide (P3-7)2 allowed using this technique to study its structure at atomic resolution and further characterize the interactions stabilizing its helical conformation (Supplementary Fig. 3a). First, we measured \(^{15}N\) relaxation at 278 K (R1, R2, heteronuclear \(^{15}N{1H}\) NOE) at two magnetic field strengths (14.1 T and 18.8 T) for the main chain amide (NH) groups of residues Gly4 to Lys24 and for the side chain amide (N\(\varepsilon\)2H\(\varepsilon\)) group of all four Gln residues (Gln11, Gln14, Gln17, Gln20) (Fig. 2a and Supplementary Fig. 3b). We found that the main chain \(^{15}N\)R2/R1 ratios increase from the termini towards the center of the peptide until reaching plateau values (5.05 \(\pm\) 0.25 at 18.8 T, 3.20 \(\pm\) 0.15 at 14.1 T) between residues Leu10 and Gln17 (Fig. 2a). A similar trend is traced out by the \(^{15}N{1H}\) NOE, that reaches upper plateau values between 0.65 and 0.83 over a larger central segment, from Leu7 to Ala18. Although this result could be influenced by the well-characterized phenomenon of helix fraying the main chain amide \(^{15}N\) relaxation data localizes the region of highest structural rigidity on a 10 \(^{-4}\) to 10 ns timescale as the one defined by the pairs of Leu and Gln residues involved in concatenated Glni+4 \rightarrow \text{Leui} interactions, i.e., from Leu10 to Gln17; by contrast the Gln20 \rightarrow \text{Leu16} interaction, involving the last Gln residue, appears to be weaker.

The different properties of the first three Gln residues (11, 14, and 17) relative to the last Gln residue (20) are also evident in the side chain amide \(^{13}N{2H}\) relaxation data: whereas the positive (at 18.8 T) or close to zero (at 14.1 T) \(^{13}N{2H}{2H}\) NOE values suggest significantly less side chain mobility for the first three Gln residues, Gln20 shows clearly negative values, suggesting higher side chain mobility (Fig. 2a), also
reflected in the relaxation-derived spectral density map (Fig. 2b). These results agree with the notion that the first three Gln residues form stronger Glni+4 → Leui interactions than Gln20. The high NMR signal dispersion in both the 13C and 15N dimensions of the spectra of peptide (P3-7)2 also allowed studying the Gln side chains: the first three Gln residues show fully resolved Hβ and Hγ signals in the 15N-edited HSQC-TOCSY spectrum, whereas the Hβ signals overlap for Gln 20 (Fig. 2c). This indicates that the conformations of the side chains of the former are better defined than those of the latter, even more than in polyQ helices14,16, in agreement with the relaxation data.

To improve the description of the Gln side chains, we used CoMAND24 to infer rotamer populations from a diagonal-free 3D
Fig. 2 | Structure of a Gln-based single α-helix. a 15N NMR relaxation data for main chain (Nt) and Gln side chain (Ne) amide groups of peptide (P3-7)2, measured at 14.1 and 18.8 T at 278 K in 10% D2O (main chain) and 50% D2O (side chain). Left: Nt: R2/R1 ratios, error bars represent the error propagated SD of both exponential decay fits to R1 and R2 data and solid horizontal lines represent the theoretical R2/R1 ratio of a rigid body in isotropic motion with τc = 4.5 ns, whose field dependency agrees well with the experimental data. Right: [15N(1H)] heteronuclear NOE, where the error bars correspond to the NOE SD2 and the shade corresponds to the region with the highest degree of structuration. b Spectral densities derived by reduced spectral density mapping of all 15N relaxation data: the dashed line corresponds to the behavior expected for a polypeptide. c Strips from a 3D 15N-edited TOCSY-HSQC spectrum showing the H ε and H c resonances of all four Gln residues in (P3-7)2. d Residue-specific CoMAND fitting of NOE signal intensities from the 3D CNH-NOESY spectrum showing the H ε of Gln 4, 14 main chain N, a99sb-disp frame pool. Bottom: average R-factors obtained from 100 CoMAND iterations using either frame pool. e Density plots and mean values of the fraction of mt rotamer found in the 100 fitting iterations for all four Gln residues. The fraction of helical glutamates in mt configuration in the BBDep dataset31 is shown in blue. f Probability density for Gln4 Xε and Xc derived using the Gaussian mixture model (GMM). g Structural ensemble of (P3-7)2, with main chain conformations selected by CoMAND from both trajectory pools and side chain conformations generated by GMM sampling. Top: one of 20 global ensemble calculations aligned for the Leu10-Gln14 pair. Bottom: same iteration aligned for the Leu13-Gln17 pair. h Residual dipolar couplings (RDCs) for the main chain 1H-15N moieties and relaxation-derived main chain spectral density at zero frequency, J(0). i Correlation between experimental and back-calculated RDCs.

CNH-NOESY spectrum reporting on distances between protons bound to 15N and 13C. Briefly, CoMAND selects subsets of conformers from a pool, here obtained by molecular dynamics, that reproduce the NOE signal relaxation (Fig. 2d). The distributions were enriched in the mt Gln rotamer (Xε = −60° and Xc = 180°) that is required for the Glnn+1 → Leu interaction: the Gln residues involved in strong interactions (I1, I2, and I3) have a high mt population (0.80), whereas that obtained for Gln 20 was lower (0.58). Both values are higher than that obtained for Gln residues in α-helices of structures deposited in the PDB25, 0.44 (Fig. 2e); importantly, these results were robust to changes in the force field used to generate the pool26,27.

To generate a conformational ensemble, we used the residue-specific CoMAND ensembles to train a Gaussian mixture model (GMM) by inferring Xε and Xc probability densities for each residue (Fig. 2f and Supplementary Fig. 3c), modified the side chain conformations accordingly and, through R-factor minimization, obtained the set of representative conformers shown in Fig. 2g (Supplementary Table 1). To validate the ensemble we measured three sets of residual dipolar couplings (RDCs) under steric alignment. The main chain ≥15N values show a dipolar wave pattern, typical of α-helices, that matches well the period of the zero frequency spectral density, J(0) (Fig. 2h). Although the RDCs were not used as restraints they correlate well with the ensemble-averaged values (Q = 0.37, Fig. 2i), confirming that the ensemble is an accurate representation of peptide (P3-7)2 and that the design rules that we have put forward lead to single α-helices.

Ranking Glnn+1 → Xε interactions by strength

The stability of our Gln-based α-helices stems from concatenated Gln side chain to main chain hydrogen bonds accepted by Leu residues. This design decision was based on the fact that the polyQ tract found in AR, which is the most helical studied so far, is flanked by four Leu residues18. We sought to determine how other residues perform as acceptors to increase the versatility of the design rules and better understand the factors determining the strength of the interaction. For this, we used a host-guest approach in which we determined the secondary structure of LεXεGlnα peptides (Supplementary Table 2) by NMR. These peptides were obtained by substituting the fourth Leu residue of peptide LεQεGlnε, excised from AR, by 13 different representative amino acids (Fig. 3a and Supplementary Fig. 4a).

We measured the residue-specific helical propensities of the peptides by NMR by combining standard 1H-detected triple resonance with 13C-detected CACO and CON 2D NMR experiments: the high resolution in the CO dimension of the latter allowed the unambiguous assignment of all Gln residues, even in the variants with lowest signal dispersion (Supplementary Fig. 4b). Except L1TQGln and L1SQGln, all variants show a helicity profile approximately proportional to that of L1QGln (Fig. 3a): L1TQGln and L1SQGln instead show a different profile, likely because in these cases the substitution shifts the site of helix nucleation by introducing an S/T N-capping motif28. In these, Ser/Thr accept two concomitant hydrogen bonds donated by main chain amides of residues C-terminal to them: one by the main chain O and another one by the side chain O of the hydroxyl group.

To explain the range of helicities obtained we hypothesized that it is due to two main factors: the intrinsic helical propensity of residue Xε and, based on our previous work18, the ability of its side chain to shield the hydrogen bond. While the former has been extensively measured25,26,30, we quantified the latter by using molecular modeling, considering that the conformation of the residue accepting the hydrogen bond (Xε) can have an effect on the interaction of the Glnn+1-Hε21 donor with competing water molecules (Fig. 3b and Supplementary Fig. 7a). For this we computed 1 μs trajectories in different force fields28,31 for all 20 possible variants in which we constrained the secondary structure of the Leu2-Gln5 segment (Supplementary Figs. 5a, 6a) and increased the population of the Glnn+1 → Leu interaction with a soft restraint (Supplementary Figs. 5b, 6b) to facilitate sampling the relevant region of conformational space: we obtained, as expected, that the higher the frequency of the hydrogen bond, the lower the solvent accessible surface area (SASA) of the Hε21 atom of Gln 4, with high correlation (Fig. 3c).

Next, we quantified to what extent intrinsic helical propensity (x1) and Glnn+1-Hε21 SASA (x2) explain the experimental helical propensities (Fig. 3d) by multiple linear regression. Indeed we obtained that these two independent variables explain 73% of the variability (Fig. 3d). Remarkably, the Gln4-Hε21-SASA value (x2) is the most important factor in the correlation, as its weight in the fitted equation is 40% higher than that of intrinsic helicity (x1): the model allows the prediction of the average helicities of the LεXεGlnα variants not included in our experimental dataset (Fig. 3e). The correlation improves (from r2 = 0.73 to 0.85) when only the subset of residue types with apolar side chains is considered, suggesting that additional factors might play a role when charged or polar side chains are present (Fig. 3f), and the results are robust to changes in MD force field25,30 or intrinsic helical propensity scales (Fig. 3f). These data confirm that Leu is one of the best helicity-promoting acceptors, but that other residues such as Phe, Tyr, lle or Met are similarly good, and that Trp is a particularly good acceptor in spite of its low intrinsic helical propensity. Thus, residues other than Leu can be introduced as acceptors of Gln side chain to main chain interactions, increasing the versatility of our design rules.

These results prompted us to investigate the presence of (P3-7)n like motifs in nature. To do this we searched UniprotKB32, including the Swiss-Prot and TrEMBL databases, by using the motif search tool in ScanProsite33, which we queried with the motif Ω-X-X-(Ω-Q-X)21-X-X, with Ω denoting good acceptors of Gln side chain to main chain hydrogen bonds (namely W, L, F, Y, I, M). We found that 3451 proteins contain sequences matching the motif, belonging to organisms across the kingdoms of life with representatives of a wide variety of taxonomic lineages including archaea, bacteria, viruses, and a full range of eukaryotes, from unicellular organisms to metazoa including humans (Fig. 3g). There is experimental evidence for the existence of 94 of these proteins (UniprotKB annotation score >3), mostly belonging to extensively characterized metazoa (Supplementary Fig. 7b). Figure 3h
Gln side chain to main chain hydrogen bonds can be accepted by different residues. a Top: Sequence, numbering, and representation, as helical projection, of the L₃XQ₁₆ variants studied in this work. Bottom: residue-specific helical propensity of the L₃XQ₁₆ variants. The type of residue X (position 10) is indicated by the colored circles. Left: helical profile of the 7 most helical single variants. Center: helical profile of the least helical single variants and L₄Q₁₆. Right: Helical profile of the outlier variants (X = T, S) and L₄Q₁₆ (spectra in Supplementary Fig. 4b). b Effect of the rotameric state of the acceptor on the interaction of atom Hε₂¹ with H₂O. Left: two frames of the L₄Q₈ Charmm36m trajectory showing Gln4 involved in a bifurcated hydrogen bond with Leu 4 in either the mt (top) or tp (bottom) rotamer. Right: radial distribution function for Gln4 Hε₂¹ in the frames where Leu 4 populates the mt (red) or tp (gold) rotamer (shades show the 95% CI obtained from 10 block bootstrapping). c The frequency of the side chain to main chain hydrogen bond is strongly correlated with the solvent accessibility surface area (SASA) of Hε₂¹, which depends on the type of residue X. d Multiple regression correlating intrinsic helicity (x₁) (Pace and Scholtz scale) and SASA (x₂) with the average helicity (y): the data were standardized to estimate the relative weight of each variable in defining the model. e Measured versus predicted average L₃XQ₁₆ helicity. f Squared r correlation scores (r²) for the multiple regression shown in e, using different reported scales for intrinsic amino acid helicity and Hε₂¹ SASA values derived from two sets of 1μs MD trajectories independently generated with different force fields. The results are shown for all and apolar (I, L, V, F, Y, M, W, A) residues in position X. g Number of ScanProsite-identified protein sequences in UniprotKB (including the Swiss-Prot and TrEMBL databases) containing the (P₃₋₇)ₙ motif with an increasing number of Glnₙ₊₄→Ω pairs (Ω = W, Y, F, L; M; X = any amino acid). h Representative (P₃₋₇)ₙ motif with an increasing number of Glnₙ₊₄→Ω pairs (Ω = W, Y, F, L; M; X = any amino acid).
shows an alignment of some example sequences along with their UniprotKB accession code and the organism that they belong to.

**Design of a pH-sensitive conformational switch**

Gln to Glu substitutions in polyQ helices decrease helical character due to the inability of the Glu side chain to donate hydrogen bonds at physiological pH, where the carboxylate group is deprotonated (Fig. 4a). Re-protonation by decreasing the pH can restore the interaction, providing us with an opportunity to introduce a pH-sensitive conformational switch in our design rules. To explore this possibility we first sought to establish whether the loss of helicity upon substitution and its restoration upon pH decrease is strictly local or
Fig. 4 | Introduction of a pH-sensitive conformational switch. a, The side chains of Gln and Glu at pH 2.8, but not that of Glu at pH 7.4 can donate a hydrogen to the main chain CO. b, QEx variants of LQ$_n$, with pH-sensitive Glu$i$+4 $\rightarrow$ Leu$i$ interactions shown in red. c, Helical propensities at pH 7.4 (black) and 2.8 (dark red) compared to those of LQ$_n$, at pH 7.4 (orange). d, Residue-specific differences in helical propensity due to substitution of Gln by Glu at pH 7.4 (black) and pH 2.8 (dark red). e, Gln side chain N$_2$–H$_{4(i+2)}$ regions of the $^1$H$^1$N HSQC spectra of QEx variants at pH 7.4 (left) and 2.8 (right) with the spectrum of LQ$_n$ overlaid as an orange shape. The spectra of variants QIE and QFE with $^1$N labeling of only the Gln in position i+4 to the mutated residue are superimposed in green. b, QM/MM-derived hydrogen bond electron densities for the Glu$i$+4 $\rightarrow$ Leu$i$ interaction. Mean values for the Glu$i$+4 $\rightarrow$ Leu$i$ interaction are shown as an orange line. First and second panels: normalized histograms showing the distribution of the electron density $\rho(r)$ of the main chain to

whether its effects instead can propagate to other parts of the sequence due to cooperativity.

For this, we compared the residue-specific helicity of polyQ LQ$_n$, variants with Gln to Glu substitutions at positions 1 (Q1E), 4 (Q4E), and 5 (Q5E) in the polyQ tract (Fig. 4b and Supplementary Fig. 8a). In QIE the positions experiencing the strongest decrease in helicity are Leu 2 to Leu 4, which are tethered by the first bifurcated Glu$i$+4 $\rightarrow$ Leu$i$ hydrogen bond in LQ$_n$, as expected, but there is also a small decrease in helicity for Gln residues in positions 5 to 8 (Fig. 4d, left, black). Similar effects were observed in QIE and the loss of helical character in QSE was much smaller, likely because the interaction broken upon substitution, accepted by Glu, is weak even at physiological pH. This indicates that breaking the first Glui+4 $\rightarrow$ Leu$i$ interaction also weakens the interaction to which it is concatenated in the polyQ helix, and vice versa, again in line with the notion that these two interactions form cooperatively (Fig. 4c, d, black).

Experiments at pH 2.8, where the carboxylate group of the Glu side chain is protonated, showed that helicity was even higher than that of peptide LQ$_n$, at physiological pH (Fig. 4c, d, red); in addition the dispersion of Gln side chain N$_{H_2}$ and H$_{4(i+2)}$ resonances in the 2D $^1$H$^1$N HSQC spectrum, that is characteristic of polyQ helices, was restored (Fig. 4e–g). To investigate the physical basis of this, we simulated this interaction by using a hybrid QM/MM approach (Supplementary Fig. 5b, c). We found that the establishment of the side chain to main chain hydrogen bond weakened the main chain to main chain hydrogen bond: the associated average electron density decreased from 0.015 a.u. to 0.011 a.u. (Fig. 4h) and the main chain Glu$i+4$ (H) – Leu$i$ (O) hydrogen bond weakened the main chain to main chain hydrogen bond (0.038 a.u.) was 0.031 a.u., which is higher than that of Gln$\rightarrow$Leu$i$ interaction. Mean values for the Glu$i+4$ $\rightarrow$ Leu$i$ interaction are shown as an orange line. First and second panels: normalized histograms showing the distribution of the electron density $\rho(r)$ of the main chain to

Combining Glu$i+4$ $\rightarrow$ Leu$i$ and electrostatic interactions

The natural single $\alpha$-helices studied until now are stabilized by numerous electrostatic interactions between side chains of opposite charge at relative positions $i+3$ or $i+4$. We sought to investigate whether these electrostatic interactions can be combined with Glui+4 $\rightarrow$ Leu$i$ main chain to main chain hydrogen bonds to stabilize $\alpha$-helices. For this we studied the polyQ tract of the TATA-box binding protein (TBP), which has a primary structure that suggests the presence of an electrostatic interaction between either of two Glu residues immediately flanking the tract at the N-terminus (Glu9 and Glu10) and an Arg interrupting it (Arg13) (Fig. 5a). This interaction can occur concomitantly with two strong bifurcated hydrogen bonds accepted by Ile7 and Leu8, at position i+4 relative to the first two Gln residues of the tract. As observed for the polyQ tracts in AR$^4$ and, to a lesser extent, huntingtin$^{15}$ the CD spectrum of a monomeric (Supplementary Fig. 9) peptide spanning a tract of size 16 and its N-terminal flanking region, TBP-Q$_{10}$, showed it is strongly helical and that its expansion to 25 Gln residues increases its helicity (Fig. 5a, b).

We then used NMR to characterize TBP-Q$_{10}$ with residue resolution (Fig. 5d, e). At physiological pH, in agreement with the CD data, the peptide forms a fully folded helix between residues Glu9 and Glu14 and its helicity decreases progressively towards the C-terminus; at acidic pH, instead (Fig. 5c), at which Glu side chains are protonated, the helical propensity starts decreasing at position 12. In addition, both the spectral signature of concatenated Glui+4 $\rightarrow$ Leu$i$ interactions (Fig. 5f) and the rotamer selection associated with these interactions (Fig. 5g) are diminished for Glu14 (H) and, to some extent, Glu12 (H). These results are in agreement with the formation of a helix-stabilizing electrostatic interaction between Glu9 (or Glu10) and Arg13 that is lost upon protonation at low pH, confirming that electrostatic interactions can be combined with Gln side chain to main chain interactions. To further confirm that these interactions can co-exist we simulated a WTE-enhanced MD trajectory of TBP-Q$_{10}$ (253ön). Fig. 5h shows two frames of the trajectory in which both the Glu11 $\rightarrow$ Ile7 and the Glu12 $\rightarrow$ Leu8 bifurcated hydrogen bonds occur simultaneously with a salt bridge involving Arg13 and either Glu10 (left) or Glu9 (right).

Targeting the helices to globular domains

Gln-based $\alpha$-helical peptides can be modified to bind specific globular targets: the Ala residues in the (P3-7)$_n$ scaffolds (Fig. 1) can indeed be modified at will because they are not involved in the interactions that stabilize the helical structure. To prove this concept we modified the sequence of peptide (P3-7)$_n$ to interact with the C-terminal domain of RAP74 (RAP74-CTD), a small globular domain that binds to intrinsically disordered motifs that fold upon binding$^{27,28}$. We blended the sequence of (P3-7)$_n$ with that of two different motifs (centFCP1 and cterFCP1) derived from FCP1 that interact with this globular protein independently$^{29}$. This led to peptides $\delta$ and $\delta_{ctf}$; $\delta$ was designed to bind to RAP74-CTD whereas $\delta_{ctf}$ is a control sequence equivalent to $\delta$ where we replaced Leu by Ala that, despite having high intrinsic helical
propensity, are bad acceptors of the side chain to main interactions, thus decreasing helicity (Fig. 6a and Supplementary Fig. 10). To facilitate the comparison with established helix-stabilization methods, we also designed $\delta$Stpl, which features a chemical staple covalently linking the side chains of residues at positions 13 and 17 and where Leu hydrogen bond acceptors were substituted by Ala to weaken Gln side chain to main chain interactions.

An analysis of the structural properties of peptides $\delta$ and $\delta$ctrl by CD showed, as expected, that the former is more helical than the latter, especially at room temperature (298 K), further confirming
important role of Leu residues for the stability of the helical fold (Fig. 6a). At 278 K, δ and δStpl displayed equivalent helicity, although the later was slightly more helical at 298 K. We then analyzed the chemical shift perturbations (CSPs) in the $^1$H,$^1$N BEST-TROSY spectrum of RAP74-CTD induced by peptide binding. We obtained that δ induced perturbations in residues of the globular protein that define the binding site of FCP1 in this globular target, confirming a similar binding mode ($K_D = 131.01 \pm 19.96 \mu M$, Supplementary Fig. 11a); this interaction was also studied by isothermal titration calorimetry (ITC, Supplementary Fig. 11b). This was in contrast to the results obtained with δctrl and (P3-7)$_3$, which in both cases failed to interact (Figs. 6b, c). δStpl, instead, induced intermediate CSPs reporting on binding in the millimolar range (Supplementary Fig. 11a). The result obtained with δctrl indicates that the helical character of δ is key for its ability to interact with RAP74-CTD, whereas those obtained with (P3-7)$_3$ indicate high helical character does not suffice, and that the identity of the residues placed in the vacant positions is indeed key for binding, in agreement with our hypothesis.

To provide a second proof of concept we blended the sequence of (P3-7)$_2$ with that of a motif found in the activation domain of AR that also interacts with RAP74-CTD, to yield peptide γ (Fig. 6d and Supplementary Fig. 10).
interaction with small molecules or peptides is a potential avenue to treat castration-resistant prostate cancer. Linear peptides spanning the AR motif bind weakly to RAP74-CTD due, at least in part, to their low helical propensity, providing us with an additional opportunity to test the potential of our designs. In peptide y two GlnN→X hydrogen bond acceptor positions were modified to accommodate the binding motif following the rules learned previously (Fig. 3). We also designed peptides YSpl and YYcfl-analogous to their δ counterparts. CD spectra show higher helicity of γ when compared to either the WHTL motif or the Ycfl peptides. In contrast, Yspl is the most helical peptide of the series. In agreement with our hypothesis, we obtained that peptide y binds RAP74-CTD with a Kc in the mid-micromolar range, at least an order of magnitude more strongly than either the (P3-7)2, the WHTL or the Ycfl peptides (Fig. 6e, f and Supplementary Fig. 11c). Yspl also binds RAP74-CTD in the mid-micromolar range (Supplementary Fig. 11c), and likely compensates its rigidity with its higher helicity when compared to y. In summary, we have shown that the sequences of Gln-based single α-helices can be modified to interact with a specific globular protein.

Discussion

Our results show that Gln side chain to main chain hydrogen bonds can be used to design linear peptides that fold into α-helices (Fig. 2) with properties that make them attractive for various applications: they are highly soluble, even upon thermal denaturation, are not stabilized by electrostatic interactions, unlike the Gln and Lys/Arg-rich single α-helices reported until now; and display some degree of folding cooperativity due to the concatenation of side chain to main chain hydrogen bonds explicit in our design rules (Fig. 1). Our data sheds light on the potential basis of such cooperative effect, which occurs when the donor of a GlnN→Leu interaction and the acceptor of the next one share a peptide bond.

An important feature of our design rules is their versatility: the residue accepting the hydrogen bond donated by the Gln side chain can be any residue able to shield the interaction from the competition with water, such as Trp, Leu, Phe, Tyr, Met, and Ile (Fig. 3). Remarkably, we find natural sequences fulfilling our design in different kingdoms of life. We also found structural models for 2303 of these sequences in the AlphaFold Database (AFDB) in 61.4% of the cases, a DSSP analysis of the AlphaFold model shows that the (P3-7)3-like motif is helical, increasing to 79.8% when only motifs devoid of helix breaking residue types (P, O) in the central part of the sequence are considered (Supplementary Fig. 7c). This conclusion holds in a subset of 42 structural models calculated without multiple sequence alignments (MSAs), showing that the helicity of this motif is encoded in Alphafold learned structural preferences (Supplementary Fig. 7d). Thus, these motifs may represent a new class of uncharged single α-helices (SAHs) that had so far remained undetected. In addition, in suitable cases the design can be complemented by electrostatic interactions between side chains of opposite charge and Gln residues can be mutated to Glu to introduce a pH-responsive conformational switch that uses only natural amino acids and does not involve changes in oligomerization state (Figs. 4 and 5).

The key feature of our scaffold design is that it defines the identity of just a fraction of the peptide residues: the rest can be chosen or optimized for specific applications. To prove this concept, we designed two peptides to bind the globular target RAP74-CTD by using an approach analogous to previous motif-grafting attempts on folded scaffolds. In this specific case, naively blending these sequence features with the sequence of the designed (P3-7)3 scaffold proved sufficient for successful targeting. Of note, our strategy displayed similar or even superior binding compared to chemical stapling, known to impose backbone rigidity, which highlights the importance of geometric adaptability to the target surface conferred by the transient nature of the side chain to the main chain hydrogen bonds, and gives additional room for sequence optimization. In fact, even if the affinities that we have obtained in this initial exercise (Supplementary Fig. 11) can be sufficient for certain applications we anticipate that it will be possible to greatly improve them by systematically searching the sequence space available using techniques for affinity maturation based on high-throughput mutational scans, especially when taking advantage of the versatility of our design rules. We also showed that our scaffold peptides are highly resistant to proteolytic degradation in human plasma and readily internalized by living human cells (Supplementary Fig. 2), thus overcoming some of the most important hurdles for peptide therapeutics. Although this could be related to their amino acid composition and sequence and needs to be studied on a case-by-case basis for potential hits, our design constitutes a valuable platform from which to evolve peptides with favorable pharmacokinetic properties.

The link between polyQ tract expansion and disease onset in polyQ disorders has not been established. These tracts are found in intrinsically disordered regions, and much work has been devoted to investigating whether expansion changes their structural properties. The results obtained have been inconclusive: single-molecule Förster resonance energy transfer and NMR measurements showed little influence of tract length on the conformation of the tract found in huntingtin whereas recent studies from some of us have instead shown that the helical propensity of the tract found in the AR increases upon expansion, as does the tract found in TBP.

Our results help rationalize these observations by considering that polyQ tracts are in a polyQ helix-coil equilibrium. Its position can be influenced by the residues flanking the tract at its C-terminus, but, for a given set of solution conditions and tract length, it is mainly determined by the four residues flanking the tract at its N-terminus (Fig. 3). These are good acceptors of Gln side chain to main chain hydrogen bonds, as in AR (LLL), the polyQ helix is favored. Instead, when only two are good acceptors as in huntingtin (LKS), the coil is favored. The results obtained for TBP also fit this rationale: its flanking region (ILEE) contains two good acceptors (Ile, Leu) and the
two other residues (Glu) can establish, at physiological pH, an electrostatic interaction with the Arg residue three or four positions towards the C-terminus, favoring the polyQ helix.

The polyQ helix-coil equilibrium is sensitive to solution conditions: the entropic cost of folding results in higher stability of polyQ helices at relatively low temperatures whereas high temperatures favor the coil. This contributes in part to explain the discrepancy between the results obtained with huntingtin, where the experiments were carried out at room temperature, and those obtained for AR, where they were instead carried out at 278 K: indeed, CD studies of the structural properties of huntingtin showed an increase in helical propensity upon tract expansion at low temperature (263 K). Finally, our observation that pairs of the concatenated side chain to main chain interactions form cooperatively contributes to explaining how expansion shifts the equilibrium to the polyQ helix state both in AR and in huntingtin.

Our results thus suggest that polyQ tracts are in a helix-coil equilibrium that is governed by the N-terminal flanking region, by solution conditions and, due to cooperativity, by tract length. Given that interaction between low-populated helical conformations of huntingtin play a role in the early stages of its aggregation into amyloid fibrils, we propose that polyQ expansion leads to the onset of Huntington’s disease at least in part by stabilizing pre-nucleation oligomeric species, where the polyQ tract is partially helical, that are on-pathway to aggregation. Our proposal is in agreement with the very recently reported effects of amino acid substitutions in the N-terminal flanking region of the polyQ tract of exon 1 of huntingtin, where increases in helical propensity led to increases in aggregation propensity both in vitro and in cells.

In summary, we have shown how an appropriate concatenation of Gln side chain to main chain hydrogen bonds makes it possible, on the one hand, to design highly helical peptides that can be tailored to specific applications and, on the other hand, to rationalize the until now perplexing observations regarding the structural properties of polyQ tracts. We anticipate that the knowledge gained about this interaction will influence future developments in peptide design, particularly in the use of peptides as therapeutics, as well as contribute to better understanding the molecular basis of polyQ diseases.

Methods
Peptide sample preparation
Recombinant peptides with 15N or 13C,15N isotope enrichment were prepared as detailed elsewhere. Synthetic genes coding for the peptide of interest and codon-optimized for expression in Escherichia coli, with an N-terminal His6-Sumo tag fusion and cloned into the pDEST-17 vector, were directly obtained from GeneArt (Thermo Fisher Scientific, Waltham, MA, USA). Genes coding for the L12Q variants in the L12X and QXE series (see Supplementary Table 2) were obtained using the Q5 Site-Directed Mutagenesis Kit from New England Biolabs (Ipswich, MA, USA). Rosetta (DE3)pLysS competent cells (Novagen, Merck KGaA, Darmstadt, Germany) were grown in MOPS medium with 15NH4Cl as nitrogen source (310 K, induction at OD600 = 0.7, 1 mM IPTG, harvesting 3 h after induction). Soluble fractions of cell lysates were sonicated in lysis buffer (50 mM Tris-HCl, 1 M NaCl, 10 mM imidazole, pH 8.0) were purified by IMAC, and fractions containing the His6MBP-TEV-RAP-CTD fusion were dialyzed at 277 K overnight against cleavage buffer (50 mM Tris-HCl, 1 M NaCl, 0.5 mM EDTA, pH 8.0) in the presence of TEV protease (50 μg/mL). A second IMAC step was performed and the flow-through containing the RAP-CTD was loaded onto a HiTrap SP HP cation exchange column followed by a size exclusion chromatography step on a Superdex 75 GL 10/300 column equilibrated with NMR buffer (20 mM sodium phosphate, 0.1% trifluoroacetic acid, pH 7.4), both mounted in an Akta Purifier System (GE Healthcare, Chicago, IL, USA). The sample was concentrated to 100 μM using an Amicon Ultra 15 mL centrifugal filter (Merck KGaA, Darmstadt, Germany).

CD spectroscopy
Peptide samples for CD spectroscopy were diluted to a final concentration of 30 μM in a volume of 400 μL in either 20 mM phosphoric acid (pH 2.8) or sodium phosphate (pH 7.4) buffer. Spectra were obtained at 278 K (unless stated otherwise) in a Jasco 815 UV spectrophotometer with a 1 mm optical path cuvette using a data interval of 0.2 nm in the 190–260 nm range with a scanning speed of 50 nm/min and 20 accumulations. A blank spectrum acquired on the pertaining buffer under the same experimental conditions was subtracted from the sample spectrum. Thermal denaturation experiments were performed by acquiring a single accumulation spectrum with the same parameters at 10 K intervals, with a temperature ramp speed of 10 K/min and an equilibration time of 1 min.

NMR spectroscopy
Peptide samples for NMR spectroscopy were diluted to a final concentration of 100 μM in a volume of 400 μL in either 20 mM
phosphoric acid (pH 2.8) or sodium phosphate (pH 7.4) buffer with added 10% v/v D2O and 10 μM DSS for internal chemical shift referencing, then filled into Shigemi tubes (Shigemi Co. Ltd, Tokyo, Japan). All NMR experiments were recorded at 278 K (unless stated otherwise) on either a Bruker Avance III 600 MHz or a Bruker Avance NEO 800 MHz spectrometer, both equipped with TCI cryoprobes, using TopSpin 4.0.8 for data acquisition (Bruker, Billerica, MA). Unlabeled synthetic peptides P1-5, P2-6, P3-7, P5-9 were characterized by two-dimensional homonuclear (TOCSY and NOESY) and heteronuclear (1H-13C HSQC) experiments. The TOCSY and NOESY mixing times were set to 70 and 200 ms, respectively. Water suppression was achieved by excitation sculpting using a 2 ms long Squa100.10000 selective pulse. For peptide backbone resonance assignment, using uniformly 13N,15C labeled peptides, the following series of 3D triple resonance BEST-TROSY experiments were acquired with 25% non-uniform sampling (NUS) (HNCO, HN(CA)CO, HN(CO)CA, HNCA, and HN(CO)CACB). For some peptides, we resolved assignment ambiguities by also acquiring a 3D (1H)(15N)(CA)NH spectrum. Furthermore, 2D 13C-detected CACO and CON experiments were measured. Data processing was carried out with qMDD68 for non-uniform sampled data and with NMPIPE69 for all uniformly collected experiments. Data analysis was performed with CcpNmr Analysis version 2.4.0. Determination of the residue-specific helical propensity (ϕεψ) from backbone chemical shifts70,71,72,73,74 was performed using CheSPI68, which uses sequence and condition-corrected (temperature, pH) estimates for the reference random coil chemical shifts derived from backbone chemical shifts HN,N H,C.

CHESPI65, which uses sequence and condition-corrected (temperature, pH) estimates for the reference random coil chemical shifts derived from backbone chemical shifts HN,N H,C.

Furthermore, 2D 13C-detected CACO and CON experiments were measured. Data processing was carried out with qMDD68 for non-uniform sampled data and with NMPIPE69 for all uniformly collected experiments. Data analysis was performed with CcpNmr Analysis version 2.4.0. Determination of the residue-specific helical propensity (ϕεψ) from backbone chemical shifts70,71,72,73,74 was performed using CheSPI68, which uses sequence and condition-corrected (temperature, pH) estimates for the reference random coil chemical shifts derived from POTENC17. CHESPI was chosen over other algorithms as it intrinsically considers chemical shift changes derived from Glu side chain protonation, as the -0.6 ppm C chemical shift change reported before17 was observed in the TBF CACO spectra (Fig. 3d).

Side chain aliphatic 1H chemical shifts were obtained from 3D 15N-edited TOCSY-HSQC (75 ms mixing time) and NOESY-HSQC (200 ms mixing time) spectra. Glutamine side chain resonances were assigned using complementary 3D H(CC)(CO)NH and (H)(CC)(CO)NH spectra recorded with 25% NUS and 14 ms C,C-TOCSY mixing. To further confirm the side chain N resonance assignments for Gln5 in peptide Q4E we recorded 2D 1H-15N HSQC spectra of synthetic unlabeled peptides with specific 15N labeling only in the positions of interest.

For the detailed analysis of side chain rotamer distributions by the CoMAnd approach, a 3D CNH-NOE57,58 i.e., 3D [HJ,15N,H]NH-NOE-HSQC spectrum17 of [13C15N] labeled (P3-7), was recorded at 800 MHz, 278 K, with 400 ms NOE mixing time and 64(15N) × 86(13C) × 2048(1H) complex data points corresponding to 15.2 × 104.8 × 6.3 Hz FID resolution. The final 15N-HSQC module employed sensitivity-enhanced coherence selection by gradients and band-selective flip-back of H polarization to enable its faster re-equilibration during a shorter total interscan delay (experimentally optimized as 0.6 s). For the prior assignment of all aliphatic side chain 1H and 13C resonances and easy distinction between intra- and inter-residue NOE signals in the [HJ,15N,H]NH-NOE-HSQC spectrum, we furthermore recorded a set of 3D [HJ,15N)(CA)NH TOCSY (11.3 and 12.6 ms FLOPSY mixing) and [(H)(CC)(CO)]NH TOCSY spectra (9 and 18 ms FLOPSY mixing).

To study glutamine side chain 15N relaxation, a 15N labeled (P3-7) sample was prepared in NMR buffer (pH 7.4). To avoid bias due to dipole-dipole cross-correlated relaxation within the NHD3H4 moieties and thus allow a direct comparison with the main chain NH data, we sampled only their 50% semi-protonated NHD isotopomers in buffered 50% D2O and applied continuous deuterium decoupling during the 15N coherence evolution. Of note, the differential deuterium isotope shift of 15N2 in the NH3D2H2D2 vs NH2D2H2H2 species6 also allowed an unambiguous stereospecific signal assignment of the attached side chain carboxamide H6 (Supplementary Fig. 3b). To measure 15N R1 and R2 rates the conventional pulse sequences with sensitivity-enhanced coherence selection by gradients, water flip-back, and fully interleaved acquisition of relaxation delays were complemented with continuous deuterium decoupling during t2(15N) in order to suppress 1H(15N) line broadening from scalar relaxation (via J15N,1H coupling) for the glutamine side chain NHD isotopomers of interest. An exponential decay function was fitted to the data to obtain R1 and R2 values:

\[ I(t) = I_0 e^{-\frac{t}{\tau}} \]

with \( I_0 \) and \( I_1 \) corresponding to peak intensity at times 0 and \( t \), respectively. \( R_2/R_1 \) ratios were calculated as \( T_1/T_2 \), and errors were derived by propagating the SD of the fits. In contrast, the pulse sequence for measuring the 15N(1H) heteronuclear NOE (likewise with sensitivity-enhanced coherence selection by gradients and continuous deuterium decoupling during t2(15N)) required further critical adaptations to suppress detrimental antiphase signal components (in 1H(15N)) for the 50% glutamine NH2 isotopomers that impede a clean quantification of nearby NHD signals of interest. Thus, for the reference (non-saturated) spectrum, the first 90° 1H pulse in the sensitivity-enhanced relINEPT following t1(15N) had to be cycled (inverted) along with the receiver phase. For the 1H saturated spectrum, however, further antiphase contamination derives from some 4N1F1H multi-quantum coherence forming during the H saturation sequence that can be removed by its phase cycling and/or by appending a concatenated 1H spoil sequence (z-gradient 1 – 90°(1H) – z-gradient 2).

RDCs were determined from the difference between couplings observed for aligned versus unaligned (P3-7) samples (with U-15N,13C labeling) where alignment was achieved31 using a gel kit from New Era Enterprises, Inc. (Vineland, NJ, USA). For this, 7% acrylamide gels were dialyzed in ultrapure water for 3 h and NMR buffer overnight. The prepared gels were then soaked in the peptide sample (ca. 0.2 mM) overnight at 277 K and squeezed into open-ended 5 mm NMR tubes using a funnel and piston. The filled tube was closed with a bottom plug and a Shigemi top plunge. The sample alignment uniformity was assessed via the deuteron signal splitting. One bond 1H-15N RDCs were obtained from comparing aligned versus unaligned 2D BEST-TROSY spectra60 measured at 278 K and selecting either the HN TROSY or semi-TROSY signals in the direct dimension. One bond 13C-15N and two bond 13C-13H RDCs were derived by comparing the pertaining J3CJ15N splitting in the indirect (15N) and J3CH splitting in the direct (H) dimension, respectively, observed in the (not 13C decoupled) 2D 1H,15N HSQC spectra of aligned vs unaligned samples measured at 278 K. PALES3 was used to calculate the expected RDCs for each structure in the CoMAnd-derived ensemble, allowing for the calculation of an independent alignment tensor for each frame based only on its coordinates. Predicted RDCs were obtained as the average of the ca. 200 frames generated in the 20 iterations of converged, R-factor-minimizing CoMAnd global ensemble calculations, and scaled to minimize the RMSD against experimentally determined RDCs.

To study the interaction between the peptides and RAP74-CTD we measured the 2D 1H-15N BEST-TROSY spectrum of 15N-labeled RAP74-CTD (50 μM throughout, 298 K) in the presence of increasing concentrations of peptide. Before the experiment, both the protein and the peptide were dialyzed (277 K, two dialysis steps) in the same preparation of NMR buffer (20 mM sodium phosphate, 0.1% trifluoroacetic acid, pH 7.4), using a Pur-A-Lyzer (Sigma-Aldrich, Burlington, MA, USA) and a Micro Float-A-Lyzer (Spectrum Laboratories, San Francisco, CA, USA) respectively. The chemical shift assignment of RAP74-CTD was reported previously (BMRB code 27288). Averaged 1H and 15N chemical shift perturbations (CSPs) were calculated as:

\[ \Delta \delta_{AV} = \left( \frac{\Delta \delta_{1H}}{I} + \frac{1}{3} \Delta \delta_{15N} \right)^2 \]
Dissociation constants, $K_D$, and averaged CSP amplitude, $\Delta \delta_{\text{max}}$, for $\delta_\text{max}, \gamma$ and $\varphi_{\text{pep}}$ Peptide binding to RAP74-CTD were obtained by a global fitting (nonlinear regression) of the 10% peaks with the largest averaged CSPs to the following single-site binding model (1:1 stoichiometry):

$$
\Delta \delta = \Delta \delta_{\text{max}} \left[ \frac{[\text{Protein}] + [\text{Peptide}] + K_D}{[\text{Protein}] + [\text{Peptide}] + K_D + 4[\text{Protein}][\text{Peptide}]} \right]
$$

(3)

A Monte-Carlo simulation varying both protein and peptide concentrations within 20% experimental errors was used to derive error margins for the final $K_D$ values.

**Structural characterization using CoMANS**

To investigate the conformational tendencies of the (P3-7)$_2$ peptide we applied the CoMANS method (Conformational Mapping by Analytical NOESY Decomposition$^{[3]}$). This method analyzes a 3D CN$\text{H}$-NOESY spectrum (i.e. 3D [H]C$_{1\text{3}}$N-HSQC-NOESY-HSQC), which displays only NOE contacts between $^{1\text{3}}$N-bound and $^{1\text{3}}$C-bound protons and is therefore intrinsically diagonal-free$^{[4]}$. As a first step, one-dimensional $^{1\text{3}}$C sub-spectra (strips) were extracted from this spectrum ($\ell_{\text{m}}=400$ ns). Each strip is taken perpendicular to a specific $^{1\text{3}}$HN-HSQC position and represents contacts to a single $^{1\text{3}}$N-bound proton, edited by the $^{1\text{3}}$C shift of the attached carbon. For (P3-7)$_2$, we obtained strips for 18 main chain amide protons (residues L7 to K24) and all 4 glutamine side chain H$_{\text{glu}}$ protons (Q11, Q14, Q17, Q20). These strips are analyzed in terms of a quantitative R-factor expressing the agreement between experimental and back-calculated spectra. Global back-calculation parameters for CoMANS were optimized by grid searching, resulting in an overall correlation time of 2.0 ns and effective $^{1\text{3}}$C signal halfwidth of 14 Hz.

For a reconstruction of the experimental $^{1\text{3}}$C strips, CoMANS compiles a linear combination of strips back-calculated from a set of trial conformers that should reflect the conformational space of each residue. Here, we used the a99sb-disp and DES-amber MD trajectories and back-calculated 13002 frames from each trajectory for each of the 22 experimental strips. For each residue, the conformational ensemble producing the lowest R-factor was then compiled using the CoMANS stochastic optimization method$^{[4]}$. A starting conformer is randomly selected from the 20 conformers with lowest R-factors. All conformers are then tested in random order, with a new member added to the ensemble if it decreases the R-factor by more than a given threshold (0.0005). Convergence is achieved if no further conformer is found or if the ensemble reaches a maximum size, here set at 20 structures. Due to its stochastic nature, this selection procedure can be repeated to produce ensembles with similar R-factors, but sampling a wider range of conformers.

For (P3-7)$_2$, we applied a two-step protocol for each residue. In the first step we established the minimum R-factor by compiling 100 per-residue ensembles, optimizing over single experimental strips. These per-residue ensembles were also used to define a set of witness strips; i.e. those whose R-factors may be affected by conformational changes in the residue in question. In the second step, we obtained the conformational distribution by co-optimizing over these witness strips. For each set, 100 optimization trials were run for each MD trajectory frame pool, resulting in 100 ensembles per frame pool, each typically containing 5–15 members. After removing ensembles with R-factors significantly above average (90% confidence interval), the set of conformers used in co-optimization (typically over 1000) was pooled to represent the conformational diversity for each residue.

To quantify the conformational distributions, we clustered the data via GMM. A vector of $n$ features - here dihedral angles - was defined for each conformer which was then used to train a model describing the probability $p(x)$ that a data point $x$ is a member of cluster $k$. For each cluster, this probability is defined by an $n$-dimensional multivariate Gaussian distribution representing its center and shape, and by a prior probability, $p(k)$, corresponding to its relative population. These model parameters were fitted to the training data using the Expectation Maximization (EM) algorithm, modified to accommodate the periodic nature of dihedral angles. For (P3-7)$_2$, we applied the GMM method for the $\chi_1/\chi_2$ pairs of all leucine and glutamine residues. As the number of clusters that best describe the training data was not known a priori, we searched values from 1 to 9 systematically and assessed the fit via the Bayesian information criterion, a measure that includes a penalty for model complexity. EM initialization requires an arbitrary seed value for each cluster center. For $\chi_1/\chi_2$ pairs, it is convenient to select seeds at the center of a rotameric form, with seeds progressively added with increasing cluster number, according to their database frequency. The best scoring model was stored for each residue.

The GMM method provides a compact but detailed description of conformational landscapes for use in downstream calculations. Here we have applied it to Monte-Carlo conformational sampling as part of an extended “greedy” R-factor optimization protocol. The pooled a99sb-disp and DES-amber MD trajectories were systematically sampled and the conformer affording the greatest reduction in global R-factor was added in each iteration. Thus, 2–4 conformers were typically added to the ensemble, which was then further optimized by adjusting the side chain conformations for leucine and glutamine residues using $\chi_1/\chi_2$ combinations from the corresponding GMM model with a 0.05 probability cutoff. For each residue in the ensemble, up to 30 $\chi_1/\chi_2$ combinations were tested to find sterically acceptable conformations lowering the global R-factor. For glutamine, the $\chi_1$ angle was additionally sampled around population centers pertaining to each $\chi_1/\chi_2$ combination (five trials; standard deviation 8°). Note that enthalpic contributions from hydrogen bonding were not considered in testing conformers, such that their selection was primarily driven by the reduction in R-factors. This iterative process of ensemble selection and modification was repeated until no further conformers were added by the greedy step. This protocol was repeated 20 times to probe the consistency of results and an example was chosen as the final ensemble (Fig. 2g).

**Molecular dynamics simulations**

The trajectories for the (P3-7)$_2$ peptide used in the CoMANS analysis and for the TBP peptide shown in Fig. 5h were generated with the Well-Tempered Ensemble (WTE)$^{[45,46]}$ enhanced sampling algorithm starting from a fully helical conformation. We used 26 energy-biased replicas within the temperature range from 273 K to 500 K, and two unbiased replicas at 278 K and 298 K. The unbiased replica at 278 K (at which temperature the 3D [H]C$_{1\text{3}}$N-HSQC-NOESY-HSQC was measured) was then used to generate the conformations for our CoMANS analysis and Fig. 5h. For the biased replicas, the energy bias was increased during the first 500 ps and then kept constant. During the bias-deposition, a Gaussian with a height of 1.2 kJ mol$^{-1}$ and a width of 140 kJ mol$^{-1}$ was added every 0.5 ps. The bias factor was set to 16. All replicas were subsequently used in a production simulation for 200 ns, where conformations used by CoMANS were extracted from the last 150 ns. The exchange between replicas was monitored to ensure good replica diffusion in temperature space. The production simulation was run in the NPT ensemble. In order to generate frame pools for our CoMANS analysis that best reproduce the experimental results, we used two recent force fields for these simulations, DES-amber$^{[47]}$ and a99sb-disp$^{[48]}$, each with its pertaining TIP4P-D water model. These are two of the best force fields describing the helix-coil equilibrium, and we used both of them to test our model for robustness. Bonds with hydrogen atoms were constrained and a time-step of 2 fs was used. For our simulations, we used Gromacs 2019.4$^{[49]–[51]}$ patched with the PLUMED library$^{[52]}$ version 2.5.3$^{[53]}$ to enable the WTE sampling method.
Two sets of trajectories for segments of L₈X peptides (see Supplementary Table 2) with the sequence U₅L₇L₉K₆Q₄Q₇Q₄Q₉Q₈Q₉ (where X is any of the 20 natural proteinogenic amino acids) were calculated using either the Charmm36m force field with a TIP3P water model or the a99sb-disp force field with its TIP4P-D water model. In this case, our objective was to obtain accurate estimates of the donor Q₄₉ H₂₅₇ SASA depending on the nature and rotamer populations of the acceptor residue X₉ for this reason, we prioritized the choice of two force fields of different origins over an accurate description of the helix-coil equilibrium. Instead, to bias the simulations towards relevant conformations, calculations were started from a fully helical conformation, and the backbone φ and ψ dihedral angles of residues L₂ to Q₉ were restrained to −60° and −40°, respectively, with a spring constant k value optimized at 5 kJ mol⁻¹ degree⁻² (Supplementary Figs. 5a, 6a) that was maintained across the trajectory. Similarly, a bias to optimize the occurrence of Q₄ to X side chain to main chain hydrogen bonds was introduced by restraining the distance between the main chain O of residue X and the N₁₂ of residue Q₄ to 4 Å, with a spring constant k value optimized at 50 kJ mol⁻¹ nm⁻² for the Charmm36m simulations (Supplementary Fig. 5b) and 250 kJ mol⁻¹ nm⁻² for the a99sb simulations (Supplementary Fig. 6b). Structure minimization and thermalization at 278 K was performed in the NVT ensemble for 1 ns. For the 1 μs NPT production runs we used Gromacs 2019.475 and an equilibration period of 100 ns that was excluded from trajectory analysis.

A trajectory was calculated for peptide QIE using the CHARMM22 force field and TIP3P water model starting from a fully helical conformation to obtain different structures with Glu₁₋₈ → Leu₈ side chain to main chain hydrogen bonds involving the glutamic acid carboxyl group that served as seeds for the QM/MM simulations (see below). The dihedral angle χ₄ orienting the glutamic acid carboxyl group was restrained to 0° since its most stable conformation in CHARMM22, corresponding to χ₄ = −180°, is incompatible with the experimentally indicated side chain to main chain interaction. We previously reported an equivalent QM/MM study on the Glu₁₋₈ → Leu₈ interaction where we used the CHARMM22 force field to generate the seeds and treat the classical sub-system, so an equivalent configuration allowed us the direct comparison of the hydrogen bond electron densities and NPAs shown in Fig. 4h, i. A fully helical starting structure was thermalized (300 K) and equilibrated in the NVT ensemble for 1 ns. The 1 μs production run was obtained using ACEMD with a 100 ns equilibration period.

QM/MM simulations

Four starting structures were selected from the classical MD trajectory of peptide QIE conserving their box of water and ions. For the QM/MM simulations, we used AMBER 20 coupled to the QM Terachem 1.9 interface.58,59 The QM subsystem was described at the BLYP/6-31G* level of theory including dispersion corrections and comprised 66 atoms including linker atoms. The classical subsystem was treated with the CHARMM22 force field. The linker atom procedure was employed to saturate the valence of the frontier atoms and electrostatic embedding was used as implemented in AMBER. An electrostatic cutoff of 12 Å and periodic boundary conditions were employed throughout all QM/MM-MD simulations, using a time step of 1 fs. Structures were minimized, thermalized, and equilibrated for 10 ps at the QM/MM level prior to the 150 ps-long production runs. Finally, for each of the 150 ps QM/MM-MD runs, a Natural Bond Critical Point analysis was performed using NBO 7.0.

Database motif searches

UniProtKB (including both the Swiss-Prot and the TrEMBL databases) was queried for protein sequences containing motifs that fulfill our design rules using the ScanProsite motif search tool hosted at the Expasy website (https://prosite.expasy.org/scanprosite/). The query motif was introduced in Prosite format as: {LFYWIM}-X-{LFYWIM}-Q-X-{LFYWIM}-Q-X-{LFYWIM}-Q-X-{LFYWIM}-Q-X-Q-X-Q-X-Q-X to quest for sequences with n = 4 Q₋ₓ₄ - {LFYWIM}, pairs, with the number of central {LFYWIM}-Q-X- triplets increased stepwise for concomitant increases in the quested number of Q₋ₓ₄ - {LFYWIM}, pairs. UniProtKB annotation scores and taxonomic lineage information were obtained by programmatically accessing this information in UniProtKB using the accession codes from the ScanProsite searches.

Data analysis and plotting

Data were analyzed using Python 3.7.8 along with packages Pandas 1.3.5, Numpy 1.21.2, SciPy 1.7.3, Biopython 1.76, and MDtraj 1.9.3, whereas Matplotlib 3.5.2 and Seaborn 0.11.2 were used for data plotting and visualization. PyMOL 2.3.5 was used to generate the figures displaying macromolecular structures.

Hydrogen bond criteria

We considered two atoms to be hydrogen bonded if the distance between donor H and the acceptor O was shorter than 2.4 Å and their angle was larger than 120°.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The chemical shift assignments for peptides P1-5, P2-6, P3-7, P5-9, (P3-7)₂ (P3-7)₃, u(P3-7)₂, u(P3-7)₃, L₁Q₁₆ (pH 2.8), L₁WQ₁₆, L₁YQ₁₆, L₁FQ₁₆, L₁MQ₁₆, L₁JQ₁₆, L₁EQ₁₆, L₁RQ₁₆, L₁AQ₁₆, L₁KQ₁₆, L₁HQ₁₆, L₁VQ₁₆, L₁TQ₁₆, L₁SQ₁₆, Q₁E, Q₁F, Q₁L, Q₁N, Q₁P, Q₁R, Q₁S, Q₁T, Q₁V, Q₁W, Q₁Y, L₂Q₁₆, L₂WQ₁₆, L₂YQ₁₆, L₂FQ₁₆, L₂MQ₁₆, L₂JQ₁₆, L₂EQ₁₆, L₂RQ₁₆, L₂AQ₁₆, L₂KQ₁₆, L₂HQ₁₆, L₂VQ₁₆, L₂TQ₁₆, L₂SQ₁₆, Q₂E, Q₂F, Q₂L, Q₂N, Q₂P, Q₂R, Q₂S, Q₂T, Q₂V, Q₂W, Q₂Y, L₃Q₁₆, L₃WQ₁₆, L₃YQ₁₆, L₃FQ₁₆, L₃MQ₁₆, L₃JQ₁₆, L₃EQ₁₆, L₃RQ₁₆, L₃AQ₁₆, L₃KQ₁₆, L₃HQ₁₆, L₃VQ₁₆, L₃TQ₁₆, L₃SQ₁₆, Q₃E, Q₃F, Q₃L, Q₃N, Q₃P, Q₃R, Q₃S, Q₃T, Q₃V, Q₃W, Q₃Y, L₄Q₁₆, L₄WQ₁₆, L₄YQ₁₆, L₄FQ₁₆, L₄MQ₁₆, L₄JQ₁₆, L₄EQ₁₆, L₄RQ₁₆, L₄AQ₁₆, L₄KQ₁₆, L₄HQ₁₆, L₄VQ₁₆, L₄TQ₁₆, L₄SQ₁₆, Q₄E, Q₄F, Q₄L, Q₄N, Q₄P, Q₄R, Q₄S, Q₄T, Q₄V, Q₄W, Q₄Y, L₅Q₁₆, L₅WQ₁₆, L₅YQ₁₆, L₅FQ₁₆, L₅MQ₁₆, L₅JQ₁₆, L₅EQ₁₆, L₅RQ₁₆, L₅AQ₁₆, L₅KQ₁₆, L₅HQ₁₆, L₅VQ₁₆, L₅TQ₁₆, L₅SQ₁₆, Q₅E, Q₅F, Q₅L, Q₅N, Q₅P, Q₅R, Q₅S, Q₅T, Q₅V, Q₅W, Q₅Y, L₆Q₁₆, L₆WQ₁₆, L₆YQ₁₆, L₆FQ₁₆, L₆MQ₁₆, L₆JQ₁₆, L₆EQ₁₆, L₆RQ₁₆, L₆AQ₁₆, L₆KQ₁₆, L₆HQ₁₆, L₆VQ₁₆, L₆TQ₁₆, L₆SQ₁₆, Q₆E, Q₆F, Q₆L, Q₆N, Q₆P, Q₆R, Q₆S, Q₆T, Q₆V, Q₆W, Q₆Y, L₇Q₁₆, L₇WQ₁₆, L₇YQ₁₆, L₇FQ₁₆, L₇MQ₁₆, L₇JQ₁₆, L₇EQ₁₆, L₇RQ₁₆, L₇AQ₁₆, L₇KQ₁₆, L₇HQ₁₆, L₇VQ₁₆, L₇TQ₁₆, L₇SQ₁₆, Q₇E, Q₇F, Q₇L, Q₇N, Q₇P, Q₇R, Q₇S, Q₇T, Q₇V, Q₇W, Q₇Y, L₈Q₁₆, L₈WQ₁₆, L₈YQ₁₆, L₈FQ₁₆, L₈MQ₁₆, L₈JQ₁₆, L₈EQ₁₆, L₈RQ₁₆, L₈AQ₁₆, L₈KQ₁₆, L₈HQ₁₆, L₈VQ₁₆, L₈TQ₁₆, L₈SQ₁₆, Q₈E, Q₈F, Q₈L, Q₈N, Q₈P, Q₈R, Q₈S, Q₈T, Q₈V, Q₈W, Q₈Y. The sequences for which we calculated structural models using ColabFold with and without MSAs along with the results of the DSSP analysis. All structural calculation and visualization. PyMOL 2.3.5 was used to generate the figures displaying macromolecular structures.

References

1. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. Proc. Natl Acad. Sci. USA 93, 13–20 (1996).
Article

Nature Communications | (2022) 13:7073 | https://doi.org/10.1038/s41467-022-34793-6

2. Fuller, J. C., Burgoyne, N. J. & Jackson, R. M. Predicting druggable binding sites at the protein–protein interface. Drug Discov. Today 14, 155–161 (2009).

3. Tietz, O., Cortezon-Tamarit, F., Chalk, R., Able, S. & Vallis, K. A. Tricyclic cell-penetrating peptides for efficient delivery of functional antibodies into cancer cells. Nat. Chem. 14, 284–293 (2022).

4. Philippe, G. J. B., Craik, D. J. & Henriques, S. T. Converting peptides into drugs targeting intracellular protein–protein interactions. Drug Discov. Today 26, 1521–1531 (2021).

5. Ran, X. & Gestwicki, J. E. Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. Curr. Opin. Chem. Biol. 44, 75–86 (2018).

6. Rezaei Araghi, R. & Keating, A. E. Designing helical peptide inhibitors of protein–protein interactions. Curr. Opin. Struct. Biol. 39, 27–38 (2016).

7. Hack, V. et al. Efficient α-helix induction in a linear peptide chain by N-capping with a bridged-tricyclic diproline analogue. Angew. Chem. Int. Ed. Engl. 52, 9539–9543 (2013).

8. Patgiri, A., Yadav, K. K., Arora, P. S. & Bar-Sagi, D. An orthorhombic inhibitor of the Ras-Sos interaction. Nat. Chem. Biol. 7, 585–587 (2011).

9. de Araujo, A. D. et al. Comparative α-helicity of cyclic pentapeptides in water. Angew. Chem. Int. Ed. Engl. 53, 6965–6969 (2014).

10. Smith, S. J., Du, K., Radford, R. J. & Tezcan, F. A. Functional, metal-based crosslinkers for α-helix induction in short peptides. Chem. Sci. 4, 3740–3747 (2013).

11. Cromm, P. M., Spiegel, J. & Grossmann, T. N. Hydrocarbon stapled peptides as modulators of biological function. ACS Chem. Biol. 10, 1362–1375 (2015).

12. Pelay-Gimeno, M., Glas, A., Koch, O. & Grossmann, T. N. Structure-based design of inhibitors of protein–protein interactions: mimicking peptide binding epitopes. Angew. Chem. Int. Ed. Engl. 54, 8896–8927 (2015).

13. Hilinski, G. J. et al. Stitched α-helical peptides via bis ring-closing metathesis. J. Am. Chem. Soc. 136, 12314–12322 (2014).

14. Escobedo, A. et al. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat. Commun. 10, 2034 (2019).

15. EftekhariZadeh, B. et al. Sequence context influences the structure and aggregation behavior of a PolyQ tract. Biophys. J. 110, 2361–2366 (2016).

16. Urbanek, A. et al. Flanking regions determine the structure of the poly-glutamine in huntingtin through mechanisms common among glutamine-rich human proteins. Structure 28, 733–746 (2020).

17. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical proteins. Structure 26, 1521–1531 (2021).

18. Lifson, S. & Roig, A. On the theory of helix—coil transition in polypeptides. J. Chem. Phys. 34, 1963–1974 (1961).

19. Li, S. & Roig, A. The theory of the helix—coil transition in polypeptides. J. Chem. Phys. 34, 1963–1974 (1961).

20. Sheu, S.-Y., Yang, D.-Y., Selzle, H. L. & Schlag, E. W. Energetics of hydrogen bonds in peptides. Proc. Natl Acad. Sci. USA 100, 12683–12687 (2003).

21. Neal, S., Nip, A. M., Zhang, H. & Wishart, D. S. Rapid and accurate calculation of protein TH, 13C and 15N chemical shifts. J. Biomol. NMR 26, 215–240 (2003).

22. Wei, X. et al. Retro-inverso isomer of Angiopep-2: a stable d-peptide ligand inspires brain-targeted drug delivery. Mol. Pharm. 11, 3261–3268 (2014).

23. Presta, L. G. & Rose, G. D. Helix signals in proteins. Science 240, 1632–1641 (1988).

24. ElGamacy, M., Riss, M., Zhu, H., Truffault, V. & Coles, M. Mapping local conformational landscapes of proteins in solution. Structure 27, 853–865.e5 (2019).

25. Shapovalov, M. V. & Dunbrack, R. L. Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19, 844–858 (2011).

26. Piana, S., Robustelli, P., Tan, D., Chen, S. & Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein–protein complexes. J. Chem. Theory Comput. 16, 2494–2507 (2020).

27. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. Proc. Natl Acad. Sci. USA 115, E4758–E4766 (2018).

28. Wan, W. Y. & Milner-White, E. J. A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at alpha-helical N termini and in other situations. J. Mol. Biol. 286, 1651–1662 (1999).

29. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. Biophys. J. 75, 422–427 (1998).

30. Luque, I., Mayorga, O. L. & Freire, E. Structure-based thermodynamic scale of alpha-helix propensities in amino acids. Biochemistry 35, 13681–13688 (1996).

31. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods 14, 71–73 (2017).

32. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489 (2021).

33. de Castro, E. et al. ScanPosite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–W365 (2006).

34. Baker, E. G. et al. Local and macroscopic electrostatic interactions in single α-helices. Nat. Chem. Biol. 11, 221–228 (2015).

35. Bravo-Arrendondo, J. M. et al. The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. J. Biol. Chem. 293, 19613–19623 (2018).

36. Bonomi, M. & Parrinello, M. Enhanced sampling in the well-tempered ensemble. Phys. Rev. Lett. 104, 190601 (2010).

37. De Mol, E. et al. Regulation of androgen receptor activity by transient interactions of its transactivation domain with general transcription regulators. Structure 26, 145–152.e3 (2018).

38. Kamada, K., Roeder, R. G. & Burley, S. K. Molecular mechanism of recruitment of TFIIF-associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. Proc. Natl Acad. Sci. USA 100, 2296–2299 (2003).

39. Ung, Y. et al. NMR structure of a complex formed by the carboxyl-terminal domain of human RAP74 and a phosphorylated peptide from the central domain of the FCP1 phosphatase H. Biochemistry 48, 1964–1974 (2009).

40. Simm, D. & Kollmar, M. Waggawagga-CL: a command-line tool for predicting stable single α-helices (SAH-domains), and the SAH-domain distribution across eukaryotes. PLoS One 13, e0191924 (2018).

41. Swanson, C. J. & Sivaramakrishnan, S. Harnessing the unique structural properties of isolated α-helices. J. Biol. Chem. 289, 25460–25467 (2014).

42. Süveges, D., Gáspári, Z., Tóth, G. & Nyitray, L. Charged single alpha-helical peptides via bis ring-closing metathesis. J. Am. Chem. Soc. 134, 12322 (2012).

43. Tyniasavunkool, K. et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596 (2021).

44. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637 (1983).
45. Song, Z. et al. Modulation of polypeptide conformation through donor-acceptor transformation of side-chain hydrogen bonding ligands. Nat. Commun. 8, 92 (2017).

46. Harrington, L., Fletcher, J. M., Heermann, T., Woolfson, D. N. & Schwille, P. De novo design of a reversible phosphorylation-dependent switch for membrane targeting. Nat. Commun. 12, 1472 (2021).

47. Sia, S. K. & Kim, P. S. Protein grafting of an HIV-1-inhibiting epitope. Proc. Natl Acad. Sci. USA 100, 9756–9761 (2003).

48. Qu, J. et al. Specific knockdown of α-synuclein by peptide-directed proteasome degradation rescued its associated neurotoxicity. Cell Chem. Biol. 27, 751–762.e4 (2020).

49. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl Acad. Sci. USA 98, 3750–3755 (2001).

50. Ott, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. Annu. Rev. Neurosci. 30, 575–621 (2007).

51. Warner, J. B. 4th et al. Monomeric huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. J. Am. Chem. Soc. 139, 14456–14469 (2017).

52. Urbanek, A. et al. Evidence of the reduced abundance of proline cis conformation in protein poly proline tracts. J. Am. Chem. Soc. 142, 7976–7986 (2020).

53. Jayaraman, M. et al. A general strategy to access structural information at atomic resolution in polyglutamine homorepeats. Angew. Chem. Int. Ed. Engl. 130, 3660–3663 (2018).

54. Baias, M. et al. Structure and dynamics of the huntingtin exon-1 N-terminus: a solution NMR perspective. J. Am. Chem. Soc. 139, 1168–1176 (2017).

55. Katler, S. A. et al. Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR. Proc. Natl Acad. Sci. USA 116, 3562–3571 (2019).

56. Jayaraman, M. et al. Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. J. Mol. Biol. 415, 881–899 (2012).

57. Elena-Real, C. A. et al. The structure of pathogenic huntingtin exon-1 defines the bases of its aggregation propensity. bioRxiv https://doi.org/10.1101/2022.10.25.513681 (2022).

58. Escobedo, A., Chiesa, G. & Salvatella, X. Recombinant production of monomeric isoet-enriched aggregation-prone peptides: polyglutamine tracts and beyond. Methods Mol. Biol. 2141, 211–230 (2020).

59. Hwang, T. L. & Shaka, A. J. Water suppression that works. Excitation transfer. Nature Communications | 2022| 13:7073

60. Solyom, Z. et al. Backbone dynamics of a free and phosphopeptide-complexed Snc homology 2 domain studied by 15N NMR relaxation. Biochemistry 33, 5984–6003 (1994).

61. Chou, J. J., Gaemers, S., Howder, B., Louis, J. M. & Bax, A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. J. Biomol. NMR 21, 377–382 (2001).

62. Zwickler, M. NMR prediction of molecular alignment from structure using the PALES software. Nat. Protoc. 3, 679–690 (2008).

63. Deighan, M., Bonomi, M. & Pfendtner, J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. J. Chem. Theory Comput. 8, 2189–2192 (2012).

64. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56 (1995).

65. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718 (2005).

66. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2, 19–25 (2015).

67. MURAMDO consortium. Promoting transparency and reproducibility in enhanced molecular simulations. Nat. Methods 16, 670–673 (2019).

68. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. Comput. Phys. Commun. 185, 604–613 (2014).

69. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. Comput. Phys. Commun. 185, 604–613 (2014).

70. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Skell, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935 (1983).

71. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? Biophys. J. 100, L47–L49 (2011).

72. Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. J. Chem. Theory Comput. 5, 1632–1639 (2009).

73. Case, D. A. et al. AMBER 2020, University of California, San Francisco, (2020).

74. Uffmsvse, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. J. Chem. Theory Comput. 4, 222–231 (2008).

75. Uffmsvse, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units. 2. Direct semi-consistent-field (SCF) implementation. J. Chem. Theory Comput. 5, 3138–3138 (2009).

76. Uffmsvse, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. J. Chem. Theory Comput. 5, 2619–2628 (2009).

77. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. J. Comput. Chem. 32, 1456–1465 (2011).
Acknowledgements

We thank Luis Serrano for help with the Agadir predictions and helpful discussions, Ben Lehner and Ernest Giralt for helpful discussions and the ICTS NMR facility, managed by the scientific and technological centers of the University of Barcelona (CCiT UB), for their help in NMR. B.M. acknowledges funding from the Asociación Española contra el Cáncer (FCAECC project #POSTD211371MATE). C.G. acknowledges a graduate fellowship from MINECO (PRE2018-084684). M.S.-N. acknowledges funding from MINECO (PID2020-119810RB-I00). M.S.-N. holds a Ramón y Cajal contract (RYC2018-024759-I) from the Spanish Ministry of Science, Innovation, and Universities. X.S. acknowledges funding from AGAUR (2017 GR 324), MINECO (BIO2015-70092-R and PID2019-110198RB-I00), and the European Research Council (CONCERT, contract number 648201). B.B.K acknowledges funding from the Novo Nordisk Foundation (#NNF18OC0033926). M.O. acknowledges funding from the Instituto Nacional de Bioinformática, the EU BioExcel Centre of Excellence for HPC and the Spanish Ministry of Science (PID2021-122478NB-I00) and the Instituto de Salud Carlos III-Instituto Nacional de Bioinformáticas (ISCIII PT17/0009/0007 co-funded by the Fondo Europeo de Desarrollo Regional). M.O. is an ICREA Academy scholar and J.A. is a Juan de la Cierva fellow. M.C. was supported by institutional funds of the Max Planck Society. This project has been carried out using the resources of CSUC. IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain).

Author contributions

Conceptualization: A.E., J.G., R.C. and X.S. Experimental data acquisition, processing, and analysis: A.E., J.P., T.B., B.M., C.G., M.S.-N., M.B., L.S., J.G., and O.M. Simulations and data analysis: A.E., J.A., B.T. and R.C. Ensemble generation, and analysis: A.E., M.C. Supervision: A.E., B.B.K., O.M., R.C., and X.S. Writing original draft: A.E., J.G., R.C. and X.S. Writing final version: all authors. Funding acquisition: B.B.K., O.M., M.O., R.C., and X.S.

Competing interests

M.B. and X.S. are founders of Nuage Therapeutics. M.B. is an employee of Nuage Therapeutics. X.S. is a scientific advisor of Nuage Therapeutics. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34793-6.

Correspondence and requests for materials should be addressed to Albert Escobedo or Xavier Salvatella.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022