



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators / Mateos, Borja; Conrad-Billroth, Clara; Schiavina, Marco; Beier, Andreas; Kontaxis, Georg; Konrat, Robert; Felli, Isabella C.; Pierattelli, Roberta. - In: JOURNAL OF MOLECULAR BIOLOGY. - ISSN 0022-2836. - STAMPA. - 432:(2020), pp. 3093-3111. [10.1016/j.jmb.2019.11.015]

Availability:

The webpage <https://hdl.handle.net/2158/1179700> of the repository was last updated on 2020-06-05T21:47:38Z

Published version:

DOI: 10.1016/j.jmb.2019.11.015

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

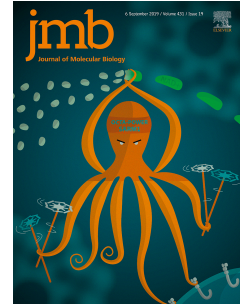
La data sopra indicata si riferisce all'ultimo aggiornamento della scheda del Repository FloRe - The above-mentioned date refers to the last update of the record in the Institutional Repository FloRe

(Article begins on next page)

Journal Pre-proof

The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators

Borja Mateos, Clara Conrad-Billroth, Marco Schiavina, Andreas Beier, Georg Kontaxis, Robert Konrat, Isabella C. Felli, Roberta Pierattelli



PII: S0022-2836(19)30680-1

DOI: <https://doi.org/10.1016/j.jmb.2019.11.015>

Reference: YJMBI 66333

To appear in: *Journal of Molecular Biology*

Received Date: 20 August 2019

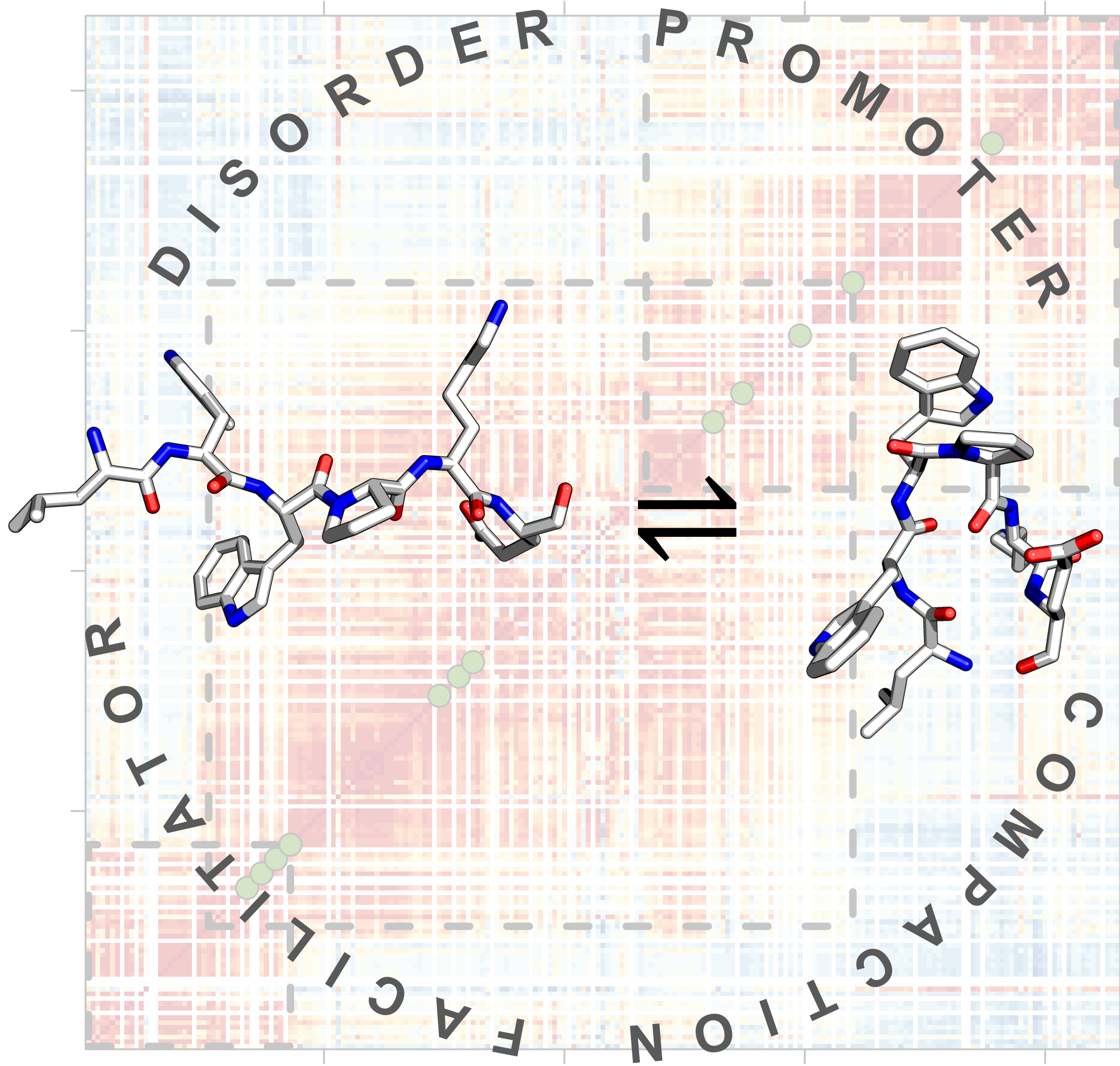
Revised Date: 23 October 2019

Accepted Date: 14 November 2019

Please cite this article as: B. Mateos, C. Conrad-Billroth, M. Schiavina, A. Beier, G. Kontaxis, R. Konrat, I.C. Felli, R. Pierattelli, The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators, *Journal of Molecular Biology*, <https://doi.org/10.1016/j.jmb.2019.11.015>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Ltd.



Title:

The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators

Author names and affiliations:

Borja Mateos^a, Clara Conrad-Billroth^a, Marco Schiavina^b, Andreas Beier^a, Georg Kontaxis^a, Robert Konrat^{a,*}, Isabella C. Felli^{b,*} and Roberta Pierattelli^{b,*}

^a Department of Structural and Computational Biology, University of Vienna, Max Perutz Labs, Vienna Biocenter Campus 5, 1030 Vienna, Austria.

^b CERM and Department of Chemistry “Ugo Schiff”, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy.

Corresponding authors:

Robert Konrat: robert.konrat@univie.ac.at; Phone +43-1-4277-52202; Fax +43-1-4277-852202

Isabella C. Felli: felli@cerm.unifi.it; Phone +39 055 4574242; Fax +39 055 4574923

Roberta Pierattelli: roberta.pierattelli@unifi.it; Phone +39 055 4574265; Fax +39 055 4574923

Abstract

Intrinsically disordered proteins (IDPs) carry out many biological functions. They lack a stable three-dimensional structure, but rather adopt many different conformations in dynamic equilibrium. The interplay between local dynamics and global rearrangements is key for their function. In IDPs, proline residues are significantly enriched. Given their unique physicochemical and structural properties, a more detailed understanding of their potential role in stabilizing partially folded states in IDPs is highly desirable. Nuclear magnetic resonance (NMR) spectroscopy, and in particular ^{13}C -detected NMR, is especially suitable to address these questions. We applied a ^{13}C -detected strategy to study Osteopontin, a largely disordered IDP with a central compact region. By employing the exquisite sensitivity and spectral resolution of these novel techniques we gained unprecedented insight into *cis*-Pro populations, their local structural dynamics and their role in mediating long-range contacts. Our findings clearly call for a reassessment of the structural and functional role of proline residues in IDPs. The emerging picture shows that proline residues have ambivalent structural roles. They are not simply *disorder promoters* but rather can, depending on the primary sequence context, act as nucleation sites for structural compaction in IDPs. These unexpected features provide a versatile mechanistic toolbox to enrich the conformational ensembles of IDPs with specific features for adapting to changing molecular and cellular environments.

Keywords:

Proline, intrinsically disordered proteins, Osteopontin, ^{13}C -detection, NMR spectroscopy.

Intrinsically disordered proteins (IDPs) exist as ensembles of rapidly interconverting conformers, therefore cannot be described by a single three-dimensional structure [1–3]. Their unique physicochemical properties derive from the polymer-like properties [4] of polypeptide chains in combination with the specific characteristics of the interaction pattern resulting from the primary sequence. Among the naturally occurring amino acids, proline has peculiar properties due to its cyclic structure, which links the backbone nitrogen with its sidechain generating a pyrrolidine ring that constrains its available conformational space. These steric restraints reduce the energy difference between *cis/trans* conformers involving peptide bonds between proline residues and the amino acids preceding them (Xaa-Pro peptide bonds). While for non-cyclic amino acids the occurrence of the *cis* conformation is less than a 0.5%, the cyclic nature of prolines slightly disfavors the *trans* conformation, leading to increased *cis* populations, typically ranging from 5 to 10% [5,6]. The energetic barrier in *cis* and *trans* interconversion is estimated to be around 84 kJ/mol¹, high enough to have distinct separate species with a significant lifetime (i.e. interconversion rates of 10⁻³ s⁻¹), existing in ‘slow’ exchange. Peptide bonds involving prolines differ substantially from those involving other amino acids and among other factors [7,8] play an important role in protein folding processes [9].

The role of proline residues in highly flexible disordered polypeptide chains has been a long discussed topic, and has involved many hypotheses [9–13]. The peculiarities of this amino acid have been a focus of study due to its pivotal role in protein-protein interactions [14], backbone rearrangements [8,15], post-translational modifications [16,17] and secondary structure formation [18–21]. Experimental data, however, were sparse mainly due to methodological limitations.

Nuclear magnetic resonance (NMR) is a unique tool to characterize the structure and dynamics of proteins at atomic resolution. Most methods used in the study of biomolecules

rely on the detection of ^1H , which offers better inherent sensitivity due to its higher gyromagnetic ratio compared to ^{13}C and ^{15}N . The general approach for a well-structured protein consists in exploiting the amide proton ($^1\text{H}^{\text{N}}$) in combination with the backbone nitrogen (^{15}N) as readout due to its ease of detection and the resulting favorable overall cross-peak dispersion that allows to observe the 'fingerprint' of a protein of interest by simple 2D spectra. However, this method has limitations when dealing with IDPs, in particular when approaching physiological conditions. On one side, $^1\text{H}^{\text{N}}$ is highly sensitive to its environment, in particular to pH and temperature, which may compromise its use for studies under such conditions due to rapid exchange with the solvent. In addition, it does not provide information about proline residues, which lack the amide proton. This also complicates their sequence specific assignment, although approaches have been developed to partially 'by-pass' this limitation [22,23]. Alternative detection schemes for prolines involve aliphatic proton-carbon ($^1\text{H}-^{13}\text{C}^{\text{ali}}$) correlation spectroscopy [12,16,24]. Despite many observable correlations ($^1\text{H}^{\alpha}-^{13}\text{C}^{\alpha}$ and $^1\text{H}^{\text{ali}}-^{13}\text{C}^{\text{ali}}$), these experiments have very poor chemical shift dispersion characteristics and are largely insensitive to their backbone environment. Moreover, the large number of proline residues in IDPs leads to extensive cross-peak overlap that hampers their investigation at atomic resolution.

An attractive option to obtain NMR spectroscopic information on proline residues is the use of direct ^{13}C observation [25]. To our knowledge, the first reference making use of direct ^{13}C -detection for the study of prolines dates back to 1973 and was proposed by the Bovey group [26] and few years later, in 1976, by the Wüthrich group [27]. However, this approach was abandoned due to the low sensitivity of these experiments. The advances in obtaining $^{13}\text{C}/^{15}\text{N}$ -enriched recombinant proteins and the development of cryogenically-cooled NMR probeheads for the direct detection of heteronuclei [28], which have boosted sensitivity, allowed the viability of direct ^{13}C and ^{15}N detection to study macromolecules [29–33]. More interestingly, as the chemical shifts of an NMR cross-peak depend on the

two nuclei involved in the peptide bond, the 2D correlation of carbonyl carbon and amide nitrogen nuclear spins across the peptide bond ($^{13}\text{C}'_i\text{-}^{15}\text{N}_{i+1}$) results in exquisite signal dispersion in the so-called CON spectrum (**Fig. 1A**) [34,35]. This is largely independent of tertiary structure and provides well-resolved information about all amino acids including proline, even in IDPs. In other words, CON spectra provide the most complete two-dimensional fingerprint of an IDP and they are a key tool for their complete characterization [36,37].

Here we make use of direct ^{13}C -detection tools [38–42] to fully characterize the *cis/trans* equilibrium of proline residues in *Coturnix japonica* Osteopontin (OPN), an extracellular IDP that exerts its function by binding to numerous components of the extracellular matrix (ECM), including two types of membrane receptors: integrins and CD44. Its activity is highly regulated by post-translational modifications (phosphorylation, glycosylation and protein cleavage), splicing of OPN and CD44 and several binding events with different extracellular ligands. The correct interpretation of the extracellular stimuli is key for cell behavior and a deregulation of OPN has been related to metastatic and abnormal cell function [43–47]. OPN contains regions with significant compaction that enable the correct exposure of many of the interacting motifs, like the RGD motif (**Fig. 1C**) [48], indicating a potential role of these compact states for its function. Thus, OPN is an ideal case to study the role of proline residues in IDPs and the relevance of *cis/trans* conformers for stabilizing local conformation and long-range contacts.

RESULTS**Observation and signal assignment of *cis* isomers of proline residues by ^{13}C -detection.**

$^{13}\text{C}'_{i-1}$ - $^{15}\text{N}_i$ cross-peaks, where i is a proline, due to their peculiar nitrogen chemical shifts, appear in a well-separated region of CON spectra (**Fig. 1A**). A careful inspection of the spectra reveals the existence of peaks due to minor species (**Fig. 1B**).

Given the low population of the minor species special emphasis was put into optimizing the sensitivity of the standard assignment experiments exploiting ^{13}C detection (3D (H)CBCACON and 3D (H)CBCANCO) [29,35,49] (see Method details). Signal assignment cannot be achieved simply by comparison with the major form chemical shifts because cross-peaks displacements of the minor form relative to the main peaks are substantial and do not follow a simple regular pattern. Standard assignment strategies based on the correlation of C^α , C^β and C' of consecutive residues (**Fig. 2A**) allow the assignment of most of the residues when spectral overlap is not pronounced. The 3D (H)CCCON [34] provides additional information about side chain shifts of aliphatic carbons (C^{ali}), a key aspect to distinguish residues that have similar C^α and C^β chemical shifts. However, this strategy fails when dealing with similar sequence repeats. In particular, proline sidechain carbons share very similar chemical shifts and several ambiguities might remain during the sequence-specific assignment procedure. The 3D COCON experiment [38,39] is an excellent complementary experiment because it correlates carbonyl signals of three or more consecutive residues by transferring the magnetization across the $^3\text{J}_{\text{C}'\text{C}}$ couplings (0.5-3.0Hz) via isotropic mixing. In order to detect cross peaks involving the minor species, characterized by low signal-to-noise ratio, a sensitivity improvement was introduced into the original 3D COCON experiment. By taking advantage of the higher ^1H polarization, initial magnetization was transferred from $^1\text{H}^\alpha$ to $^{13}\text{C}^\alpha$ and then to $^{13}\text{C}'$ prior to the beginning

of the COCON experiment. This resulted in an approximate increase of the S/N ratio per unit time by a factor of 1.65 (**Fig. S1A-B**). The detection of sequential connectivities in the 3D (HCA)COCON (**Fig. 2B**) in combination with those observed in the 3D (H)CBCACON, 3D (H)CBCANCO and 3D (H)CCCON allowed the assignment of both major and minor forms in OPN. Under the present experimental conditions, the (HCA)COCON spectra gave the most complete information by correlating at least three (from $\delta(^{15}\text{N})_i$ to $i+2$) residues in all cases and, in some cases even five residues (from $\delta(^{15}\text{N})_{i-1}$ to $i+3$). The signal intensity of the off-diagonal signals is directly dependent on the value(s) of the intervening $^3J_{\text{C}'\text{C}}$ coupling constants. In the subspectra associated with the minor forms some of the cross-peaks connecting proline residues to their *following* residues show low to vanishing intensity (**Fig. S1C**). As the connections to the preceding residues are usually still present, this is not simply an issue of reduced spectral sensitivity reflecting the low effective concentration of the minor species, but rather suggests a consistently small value for the $^3J_{\text{C}'\text{C}_{i+1}}$ indicative of a specific backbone geometry. This is consistent with the cyclic nature of the Pro sidechain limiting the $|\varphi|$ angles of the Pro residue (in position i) to values around $[-60^\circ, -90^\circ]$, thus corresponding to low values of the $^3J_{\text{C}'\text{C}}$ coupling based on the Karplus equation [50,51] (**Fig. 2B** and **Fig. S1D-F**).

The chemical shifts of side chain carbon atoms of proline residues obtained through the analysis of the 3D (H)CCCON were very useful to confirm that the minor form observed derives from *cis/trans* isomers of Xaa-Pro peptide bonds. Indeed, as initially proposed by Schubert and others [52,53] *cis* conformations can be identified by the difference between C^β and C^γ chemical shifts, which is larger in the case of *cis*-Pro relative to *trans*-Pro. As shown in Figure 3A, the minor forms observed in the proline region follow this pattern and are mostly in the *cis* conformation (**Fig. S2**).

Cis/trans isomerization at positions of proline residue also affects the neighboring residues, similarly to $^1\text{H}^{\text{N}}$ shift differences observed between *cis/trans*-induced forms in α -

synuclein [11]. Specifically, we observed that (i) the percentage of *cis*-Pro conformers is conserved along the signals of minor forms of consecutive residues surrounding the proline (**Fig. S3A**) and (ii) the difference in chemical shifts between major and minor forms is large for residues $i\pm 1$ and gradually decreases until position $i\pm 3$, where it coincides with the position of the major form (**Fig. S3B-C**), indicating that proline isomerization has a drastic effect at the local level in a range spanning residues $i\pm 3$.

Another observation that can be easily depicted is the position of C^α and C^β shifts of residues preceding a proline (BMRB ID: 27443 and **Table S4**), which reflects the fact that residues preceding a Pro are biased towards an extended conformation. A typical indication of a proline in the following position is the up-field shift of C^α . This is easily recognized for all residues (see examples of V81, V90, F121, A125, A128, Q178, L196, R207 present in OPN). More interestingly, for some of these residues (V81, V90, F121 and R207) the C^β carbons experience a comparable downfield shift, thus indicating specific conformational changes at these sites.

To conclude, given the substantially improved sensitivity of the new pulse sequences, the CON cross peaks of both *trans* and *cis* forms, including their sidechains, of the 12 proline residues present in OPN could be assigned. Additionally, the minor forms of several proline neighbors were assigned (11 out of 12 in the $i-1$ positions; 6 out of 12 in the $i+1$ positions; and some in $i\pm 2$ and $i\pm 3$ positions). The assignment of the minor forms can be found in the **Table S4**. This also further extends the assignment already achieved for OPN (deposited with BMRB as ID 27443) comprising 99.1% of N, C' , C^α and C^β (major form).

Quantification of the *cis* populations.

Relative populations of *trans* and *cis* species were determined by taking peak volumes from the ^{13}C -detected 2D CON spectra (**Fig. 1B**). It is well-known that prolines display high *cis* population, ranging from 5 to 10% in proteins, while even higher percentages have

been observed for small peptides [54–56]. In addition, this equilibrium can be modulated by the chemical environment (for example, by the occurrence of phosphorylation [17] and/or by secondary structure propensities). In the case of OPN, experimental data clearly indicate that all proline residues sample both the *trans* and the *cis* conformation, with populations for *cis* conformers that significantly deviate from percentages reported for other IDPs (around 5-10%) like α -synuclein [11] or tau [12]. P88, P91, P122 and P126 also show an additional minor form with a *trans* signature (**Fig. 1B** and **Fig. 3B**) likely coupled with the *cis* conformation of neighboring proline residues.

It is worth noting that for P88, P122, P129 and P185, the population of the *cis* forms is greater than 10% (**Fig. 3B**) and this higher *cis* population correlate with the presence of aromatic residues (Phe or Trp) in either $i\pm 1$ position (**Fig. 1C** and **Fig. S4A**). Indeed, as an example, the mutation to alanine of phenylalanine 121 (preceding proline 122), induces a lower *cis* population of the peptide bond linking the two aminoacids compared to the wild-type protein (from 13% to 7%, **Fig. S5**) confirming the importance of the direct interaction between aromatic and proline residues to stabilize the *cis* conformer.

The structural dynamics of *cis* forms.

To further characterize the role of proline *cis/trans* isomerism and its impact on the structural dynamics of OPN, we measured ^{15}N relaxation parameters [57] (**Fig. 4A** and **Fig. 5**). A comparison of ^{15}N - R_2 relaxation parameters in the major and minor forms for residues other than proline showed comparable local dynamics in both states, although some specific changes occur (**Fig. 4B**). Interestingly, increased ^{15}N transverse relaxation rates were consistently found for residues within the region of residues 140-175. This region is known to be the most compact segment of OPN comprising most of the binding regions of this protein and housing both hydrophobic and charged residues to exploit for its stabilization [48]. Additionally, residues in the regions 120-132 and 175-190 and flanking the core region of OPN also show reduced conformational dynamics resulting in higher ^{15}N

transverse relaxation rates. Notably, the mutation of F121A reduces the $^{15}\text{N-R}_2$ rates of these flanking regions accompanied by a moderate decompaction of the region 140-175 (**Figure S5B**). This dynamic pattern is largely conserved in the minor forms and thus suggests that there are no substantial global differences in dynamics between signals stemming from proline *cis/trans* isomerization. However, $^{15}\text{N-R}_2$ and heteronuclear $^{15}\text{N}\{-^1\text{H}\}$ NOEs (**Fig. 4B**) provide some evidence for reduced local mobility in the *cis*-forms presumably due to the tilt of the polypeptide chain. This is consistent with the observed increased $^{15}\text{N-R}_2$ rates for the ^{15}N of *cis*-Pro (**Fig. 5**). Due to the pronounced flexibility in IDPs ^{15}N spin relaxation is typically dominated by local fluctuations on different timescales rather than the global tumbling of the protein [58–62]. In a fully extended conformation ^{15}N transverse relaxation is dominated by fast time scale motions around the C^α - C^α direction [58,59,63–65], as described by the (GAF) Gaussian Axial Fluctuation model [66]. Bending of the polypeptide chain due to formation of the *cis*-form may restrict this dynamic mode and increase the effective local correlation time which would be accompanied by faster $^{15}\text{N-R}_2$ relaxation. Several mechanisms could explain these differences in relaxation, i.e. interactions that generate local compaction. A plausible explanation could be the increase of conformations resembling those found in PPI helices in contrast to the conventional PPII form often sampled by proline-rich segments [67]. Although the changes in backbone dihedral angles are rather small the switch from PPII to PPI results in a significant compaction of the polypeptide chain. While PPII is an elongated left-handed helix PPI is a compacted right-handed helix. The average helical pitch per residue in PPI is around 5.6 Å and thus very similar to the value found in a typical α -helix. In contrast, the more extended PPII motif has a helical pitch of 9.4 Å. In order to test the hypothesis of a potential PPII-to-PPI switch we performed a quantitative analysis of secondary structure propensities by the program $\delta 2\text{D}$ [68] separately using the shifts of major and minor forms, which shows that the *cis*-Pro leads to lower PPII populations compensated by either coil or β -strand

conformations (**Fig. 6**). Interestingly, the largest changes in ^{15}N - R_2 rates were observed for P82, P85, P88, P122, P126 and P236. Strikingly, most of them (all except P236) belong to stretches containing consecutive P-X-X-P motifs and is thus more likely to form a continuous PPI helix. However, PPI-PPII transitions are challenging to pick up solely from chemical shift data and a cause-effect direct connection could not be obtained from these results alone.

On the role of proline isomerization in modulating global rearrangements.

In order to probe putative changes in long-range contacts resulting from *cis/trans* isomerization of proline residues, we carried out a series of paramagnetic relaxation enhancement (PRE) experiments. We exploit ^{13}C -detected PRE by observing intensities of the paramagnetic and diamagnetic forms [69]. $^1\text{H}^\alpha$ -start experiments are more sensitive to paramagnetic effects because protons, with their high gyromagnetic ratio compared to carbon, are reintroduced in the coherence transfer pathway to transfer coherence from protons to heteronuclei (^{13}C or ^{15}N). Thus, we used $^1\text{H}^\alpha$ -start experiments to measure ^{13}C -detected PRE to observe prolines and get higher dataset completeness. Overall, the PRE profiles of both major and minor forms obtained for the various spin label positions showed a very similar behavior (**Fig. 7**). Notable exceptions only occurred for residues located in the region 120-132 (i.e. D120, F121, T123 and E124) which showed slightly larger PRE effects when the spin label is located in position 188, as well as the region 80-90 (paralleled by an increased ^{15}N R_2 rate) with the spin label at position 108.

Additionally, putative changes in long-range correlations were also probed by measuring $^1\text{H}^\text{N}$ - R_2 rates using standard $^1\text{H}^\text{N}$ -detected experiments. In the interpretation of these rates care has to be taken as $^1\text{H}^\text{N}$ - R_2 rates may also contain slight contributions from intermolecular exchange with bulk water. We find the largest differences between major (stemming from *trans*-Pro) and minor (stemming from *cis*-Pro) forms in the proline-rich regions located at the termini (120-130 and 180-190) of the central region of OPN

characterized by largest compaction (100-190) (**Fig. S6A-B**). Summing up, PRE rates were largely comparable in both Pro forms suggesting that the overall (tertiary) compaction is maintained and only local dynamics are different between *trans*-Pro and *cis*-Pro forms.

Finally, to obtain more specific information about differential side-chain structural dynamics we performed $^{15}\text{N}\{-^1\text{H}\}$ heteronuclear NOE and ^{15}N 3D NOESY-HSQC experiments. We use the side-chain signals of tryptophan residues in the stretch (W¹⁸³-W¹⁸⁴-P¹⁸⁵) as proxies for the dynamic measurements (**Fig. S9**). We were particularly interested in the impact of proline *cis/trans* isomerization on the side-chain dynamics of W183 and W184 as this would provide prototypical information about the consequences of Pro-aromatic side-chain packing on local conformational dynamics. It can be clearly seen that minor (stemming from *cis*-Pro) and major (stemming from *trans*-Pro) forms of W183 and W184, which are directly adjacent to P185, display distinctly different conformational dynamics (**Fig. S9**). While the two forms of W183 show nearly identical $^{15}\text{N}\{-^1\text{H}\}$ NOEs, the minor form of W184 is evidently different and exhibits significantly restricted motions in the (fast) ps timescale (**Fig. 8A-B**) and increased water exchange properties (**Fig. S9C**) while keeping similar $^{15}\text{N}\text{-T}_1$ and $^{15}\text{N}\text{-T}_2$ relaxation (**Fig. S9A-B**). Additionally, different indole $^1\text{H}^{\text{N}}$ NOEs to protons of the side-chains further corroborate the finding of changes in side-chain orientations in the minor form of W184 (**Fig. 8C**). It can thus be concluded that steric repulsion due to distinct $\pi\text{-CH}$ -aromatic interactions lead to restricted local mobility in the *cis*-Pro form, as illustrated in **Fig. 8D**.

DISCUSSION

Proline residues are highly abundant in IDPs, although their role is still elusive. A unique property of prolines in proteins is their significantly higher than other residues' tendency to form *cis* peptide bond isomers. The percentage of *cis*-Pro population is known to vary between 5-10% and has been reported for previously studied IDPs like α -synuclein [11] and tau [12]. In addition, this population can be drastically affected by phosphorylation [17] and specific protein-protein interactions [70]. In the present manuscript, we used ^{13}C -detection NMR experiments to unambiguously access information about Xaa-Pro isomerization in OPN and to characterize the local and global effects of this conformational equilibrium.

The tunable stabilization of *cis* conformers by π -CH interaction between the aromatic sidechain and the proline ring has been described mainly for small molecules and peptides [55,56,71] and recently for the TAD domain of BMAL1 [72]. This stabilization is limited to prolines (at position *i*) and aromatics in positions $i\pm 1$ and $i-2$. Our study provides, to our knowledge, the first experimental evidence that these weak interactions remain relevant in the context of a large IDP. We found that OPN shows particularly high *cis*-Pro populations, reaching *cis*-Pro percentages around 20% in some cases. Most importantly, the highest populations of *cis*-Pro conformers were found for prolines in close proximity to aromatic residues. The mutation of an aromatic residue to alanine leads to a drastic reduction in *cis*-Pro population. We thus conclude that π -CH interactions between aromatic sidechains and proline rings are highly relevant for the stabilization of local structural segments in IDPs. Aromatic amino acids, leucines and proline residues are well-conserved in IDPs [5], but hydrophobic and proline residues have an opposite behavior with respect to their relative abundance when comparing ordered and disordered proteins [73]. Prolines are highly abundant in IDPs and are usually considered as disorder-promoting residues [5,74]. We thus performed a statistical analysis of the discussed motifs in globular proteins and IDPs.

In our data sets (PDB for globular proteins (<https://www.rcsb.org/>) and DISPROT for disordered segments (<http://www.disprot.org/>)) (**Fig. S4B-C**), we found Pro residues to be more abundant in IDPs (7.3%) than in globular proteins (4.4%), whereas for aromatics the opposite was found (IDPs: 5.1 %, globular proteins: 9.5 %), in line with values previously reported [75]. The comparison between globular proteins vs IDPs with respect to the prevalence of aromatic-proline pairs revealed that in IDPs 14% of prolines have aromatic residues in positions $i\pm 1$ and $i-2$ and thus in positions where the π -CH stacking interaction can take place, whereas, in globular proteins this number is 27%. Thus, our experimental findings that these sequence motifs are found in regions of significant structural compaction and that their mutations reduce, to a certain degree, this state are suggesting that these motifs might serve as nucleation sites for facilitating the formation of compact states in IDPs. In addition, it points out to the relevance of aromatic-proline interactions for the stabilization of underlying structure in IDPs. Although systematic studies on the biological significance of aromatic-proline amino acids pairs in disordered segments are currently lacking, a first hint was provided with the Pro-Trp pair in BMAL1 [72] which is strongly conserved and key for the regulation of circadian rhythms. Another example is given by the unique domain of Src-family kinases which contains an unusual number of hydrophobic residues for an IDR and which are intercalated with prolines. Arbesú et al. [76] have shown that phenylalanine residues in this IDR promote the compaction of this protein, they hypothesize that proline residues may compensate the entropic cost of this effect by retaining correlated local flexibility, as proposed elsewhere [77]. Further evidence for the relevance of this sequence motif is provided by the fact that aromatic residues are found in pre-structured molecular recognition motifs (also called MoRFs and PreSMOs [78]). Our findings that π -CH interactions significantly reduce the conformational dynamics of the minor *cis*-Pro form in W184 can thus be regarded as a prototypical case for a general mechanism to modulate backbone as well as side-chain dynamics in IDPs. The

compaction modulation provided by π -CH stacking interactions could reshape the conformational space sampled by OPN and in consequence, affects the binding events, such as the one with heparin. Previous work [48] showed an unfolding-upon-binding mechanism upon heparin interaction. This binding mechanism is possible if a certain degree of residual structure is retained in the apo-form to gain flexibility in other regions upon binding [79].

More in general, there is growing evidence that IDPs are far from being natively unfolded or disordered but rather exist as heterogeneous conformational ensembles. The distinct heterogeneities of these conformational ensembles and correlated conformational fluctuations within them can be quantitatively visualized by the Pearson correlation maps derived from PRE (and PRI) rates [80]. This is illustrated with two divergent OPNs (*Homo sapiens* vs. *Coturnix japonica* PRE rates in **Fig. S6A and S6C**, respectively) (**Fig. 9**). *Coturnix japonica* OPN contains three regions of correlated fluctuations indicating considerable structural compaction and the existence of –at least– one compact state in OPN (**Fig. 9 upper**). Most importantly, extremely high concentrations of urea and high-salt concentrations are needed to completely denature (unfold) the protein [48], thus indicating that both backbone structural elements and electrostatic interactions are involved in generating these compact states. Interestingly, human OPN showed a different pattern of correlated segments, where oppositely charged regions are less correlated than its evolutionary far homolog (**Fig. 9 lower**). In order to assess the potential influence of different charge distributions we followed a strategy proposed by Pappu and co-workers [81], who have demonstrated the intimate relation between sequence distributions and ensemble conformations of polyampholites with simulations and introduced the parameter κ , which takes into account not only the overall charge of the polymer but additionally also the distribution of these charges along the primary sequence. The κ value for *Coturnix japonica* OPN (or quail OPN) is 0.262, corresponding to a typical value of a pre-molten

globule in the interface between a weak and strong polyampholyte (79). Interestingly, despite the low interspecies protein sequence similarity (**Fig. S7**), the parameter κ (**Table S5**) is largely identical for the individual OPN family members (**Fig. S7**). The observed differences in correlated fluctuations (**Fig. 9**) can therefore not be explained by charge pattern but result from context-dependent interactions along OPN primary sequence. Regarding the number of prolines and their location in OPN (**Fig. S8 coloured circles**), human OPN is significantly divergent from the quail form. In particular, the human N-terminal segment is enriched in prolines and the quail form is depleted. On the other side, aro-Pro motifs are conserved next to the positive charged patch at the central region of the protein (**Fig. S8 red circles**), suggesting a relevant role of those in the central compact state.

Based on our findings we thus conclude that in addition to the specific distribution of charges along the primary sequence (quantified by the parameter κ) the specific location of proline-aromatic sequence motifs (relative to electrostatic patches) plays a decisive role in determining the conformational space sampled by IDPs.

It is also instructive to compare our findings with a recent sequence analysis deciphering the evolutionary constraints acting on IDPs to disable a negative function (aggregation) [82]. In this study it was shown that representative ensembles of folded proteins and IDPs respond differently to randomization (while maintaining the overall amino acid distribution) of primary sequences. Most importantly, this study unveiled that in contrast to the widespread belief that protein disorder triggers aggregation-induced diseases, IDP sequences evolved during evolution to exactly avoid amyloidogenesis and increase solubility. In other words, the finding that amyloids are sometimes associated to IDPs obscures the fact that the vast majority of IDPs emerged under the evolutionary constraint to avoid aggregation. Thus, the significance of this protein property (aggregation propensity) and its relevance for disease results from its scarcity in IDPs. Based on our

finding that aro-Pro sequence motifs are significantly less abundant in IDPs we postulate an analogous negative evolutionary constraint for IDP sequences. Since aro-Pro sequence motifs are depleted in IDPs their occurrence points to their significance. Our study of OPN showed that the aro-Pro sequence motifs exhibit significant local rigidification of the backbone chain and is thus in stark contrast to the conventional notion that proline residues are the origin of flexibility in IDPs. These structurally constrained sites can on the contrary act as nucleation sites for the formation of compact states in IDPs as their restricted motion reduces the entropic penalty for the formation of collapsed states. Most interestingly, the compact state(s) of OPN is (are) characterized by substantial long-range order [83]. Since there is typically a low sequence homology in related IDPs, we postulate that the conservation of (aro-Pro) sequence patterns indicates the existence of relevant and specific aro-Pro interactions and therefore suggests the existence of compact segments. It will be interesting to see whether this information can be used as a primary-sequence search criterion to identify IDPs with distinct structural properties, as proposed elsewhere [84].

We therefore suggest reassessing the structural and functional role of prolines in IDPs. It is generally accepted that prolines are highly abundant in IDPs and on this basis are described as disorder promoting amino acids [73,74]. Based on our findings we suggest a more subtle and ambivalent role for prolines in IDPs. Proline residues are often found in specific local secondary structural elements called PPII helices. The unique structural arrangement of residues in PPII helices does not allow for tight packing interactions, and residues in PPII helices have therefore a smaller number of direct neighbours and consequently lower structural stabilization (or disorder for that matter). However, the data obtained in our study clearly show that the structural dynamics of proline residues depends significantly on the amino acid sequence context (i.e. neighbouring amino acids next to prolines in the primary sequence). Sequence patches including proline and aromatic

amino acids display significant reduction of conformational flexibility, higher populations of *cis*-Pro configurations, and different π -CH stacking interactions between the proline and the aromatic ring systems.

To conclude, the observed novel proline features provide a versatile mechanism to modulate and tune the conformational ensembles of IDPs. For example, *cis-trans* isomerization will modulate the relative orientation of locally folded protein segments (i.e. regions displaying correlated structural fluctuations). The example of OPN provides a first glimpse on how the modulation of conformational ensembles in IDPs can be realized. Aro-Pro sequence motifs act as nucleation sites for the structural compaction reminiscent of diffusion-collision type of processes. The motional restriction (as evidenced by ^{15}N relaxation data) of these segments reduce the entropic penalty inherent to structure formation in IDPs. Moreover, our finding that these Pro-aromatic patches are located in hinge regions linking the three observed structural segments in OPN may point towards a general building principle. Since *cis-trans* isomerization is not only abundant in Pro-rich segments but can also be modulated by external factors (i.e. presence of prolyl-peptidyl isomerases, PPI, post-translational modifications such as phosphorylation of neighbouring Tyr or Ser/Thr) it can be anticipated that these modulation possibilities are realized by nature in order to endow IDPs with the stunning structural plasticity and malleability necessary to adapt to the continuously changing molecular environment, cellular context, and thereby fulfilling their diverse biological functions.

MATERIALS AND METHODS**Expression and purification of the *Coturnix japonica* and *Homo Sapiens* OPN forms.**

Coturnix japonica OPN sequence was cloned in pET11d plasmid. Several cysteine mutants and a phenylalanine-to-alanine mutation (F121A) were introduced by site-directed mutagenesis (Thermo Fisher[®]) and produced as the wild-type form. A total of 10 cysteine-mutants of the *Coturnix japonica* OPN (S54C, E77C, S108C, R132C, A161C, D174C, S188C, S206C, S228C and S247C) were used. The plasmid was transformed into the *E. coli* strain BL21(DE3) phage resistant via heat-shock method. Protein expression was induced at OD₆₀₀ of ca. 0.8 by adding 0.4 mM IPTG (Isopropyl-β-D-Thiogalactopyranoside). For expression of isotopically labelled protein for NMR studies, before induction the cells were harvested and pellets from 4 L LB resuspended in 1L M9-Minimal Media (containing 1g/L ¹⁵NH₄Cl and 3g/L ¹³C-Glucose for ¹⁵N and ¹³C labelling respectively). The expression was carried out at 27°C and 140rpm over-night. The cells were harvested and resuspended in 40 mL of cold PBS (2 mM KH₂PO₄, 8 mM Na₂HPO₄, 2.5 mM KCl, 140 mM NaCl, 5 mM EDTA, pH 7.3); freshly prepared 1mM DTT solution was added in the case of cysteine-mutated proteins preparations. The suspension was sonicated and warmed at 95°C for 10min. The lysate was centrifuged at 18000 rpm and ammonium sulfate precipitation (saturation of 50%) carried out with the supernatant. The pellet was resuspended in PBS, diluted 1:2 with water to lower the salt concentration and an anion-exchange chromatography (HiTrap Q, GE healthcare) performed. The column was equilibrated with PBS and a gradient of 30% High-Salt Buffer (PBS containing 1 M NaCl) run in 20 minutes at 2 mL/min.

A pETM11 plasmid encoding for the *Homo sapiens* OPN with a histidine tag was transformed in *E. coli* BL21(DE3) Rosetta phage resistant strain. The steps were the same as its homolog described above. Nine cysteine-mutants of the *Homo sapiens* OPN (G25C, Q50C, D90C, D130C, T185C, D210, S239C, S267C and S311C) were produced.

Regarding the purification, the harvested cell suspension was sonicated and cooked at 95°C for 10min. The lysate was centrifuged at 18000 rpm and the supernatant loaded onto a HisTrap® (GE Healthcare) affinity column. The protein was eluted with high-imidazole Buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 500 mM Imidazole, pH 8). The protein was concentrated in TEV-Cleavage Buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 1 mM DTT, 1 mM EDTA, pH 8.0), TEV protease was added in a ratio of 1:50 and incubated over-night at 4°C under agitation to get rid of the affinity tag. Further purification was carried out with ion-exchange chromatography (HiTrapQ® - GE Healthcare). The column was equilibrated with PBS and a gradient of 50% High-Salt Buffer (PBS containing 1 M NaCl) run in 30 minutes at 1 mL/min. Protein purity was confirmed via SDS-PAGE.

The protein was concentrated using a centrifugal filter and the final concentration measured by absorbance at 280 nm. The sample concentration for the assignment by ¹³C-detection and ¹⁵N relaxation rates was 2 mM. ¹³C-detected PRE experiments were measured with protein samples about 1 mM while ¹H^N-detected PRE experiments were measured with protein samples about 0.5 mM (*C. japonica*) and 0.3 mM (*H. sapiens*) in PBS buffer at pH 6.5.

MTSL-tagged cysteine mutants for PRE. For the MTSL-tagging, the protein was incubated with an excess of DTT (10 mM) for 15min at room temperature. A PD-10 Desalting-Column was equilibrated with 100 mM sodium phosphate, 1 mM EDTA and pH 8.0 buffer, the protein loaded to a total volume of 2.5 mL and eluted with 3.4 mL of buffer. The free thiol concentration was measured using a 300 μM DTNB (5,5'- dithiobis-(2-nitrobenzoic acid))-solution at 412 nm. A 3-fold excess of MTSL was added and the sample incubated for 3 hours at 37°C during agitation. The free thiol concentration was measured again to confirm complete tagging of the protein. Protein purity was confirmed via SDS-PAGE electrophoresis.

¹³C-detection assignment and ¹H^α-start CON PRE. NMR experiments were recorded at 310 K with a 16.4 T Bruker Avance NEO spectrometer operating at 700 MHz ¹H frequency and equipped with a cryogenically cooled probehead optimized for ¹³C direct detection. It has to be noted that the presence of D₂O (lock solvent) in the sample also gives rise to cross peaks resulting from isotopic shifts in the CON spectra due to H/D exchange. Although these signals are easy to identify due to their regular pattern of displacement, they can lead to unwanted signal overlap and complicate spectral analysis. Therefore, the lowest possible concentration of D₂O (typically 1-2%) has been used during the measurements to reduce the intensity of this set of additional cross peaks below detection level.

In the ¹³C detected experiments for sequence specific assignment of N, C', C^α, C^β resonances at 310 K (2D CON, 3D (H)CBCACON, 3D (H)CBCANCO, 3D (HCA)COCON and 3D (H)CCCON)[34,38,85] the ¹³C carrier was placed at 174.5 ppm, the ¹⁵N carrier at 123 ppm and the ¹H carrier at 4.7 ppm (H₂O). Band selective ¹³C pulses were applied at 174.5 and 39 ppm to excite or invert ¹³C' and ¹³C^{ali} spins respectively. The following band-selective pulses were used: 300 μs with Q5 and time reversed Q5 shapes for C' and C^{ali} excitation, 220 μs Q3 shapes for C'/C^{ali} inversion/refocusing. Homonuclear virtual decoupling in the direct dimension was achieved through the IPAP approach. Detailed information about the acquisition parameters is provided (**Table S1 and S2**). NMR data were processed using Bruker TopSpin 4.0. The programs CARA[86] and Sparky[87] were used to analyze and annotate the spectra. All the statistical analysis and fittings were carried out with RStudio.

¹H^N-T₂ PRE acquisition parameters. NMR experiments were recorded in on a 18.8 T Bruker Avance III HD+ spectrometer operating at 800 MHz. For the ¹H^N-T₂ PRE measurements, the measurements were carried out at 298 K and paramagnetic samples were reduced using a 3-fold excess of ascorbic acid for the diamagnetic control. The ¹H^N-

T_2 relaxation delays were 0.001, 0.005, 0.01, 0.02, 0.05 and 0.1 s. The paramagnetic effects were quantified as

$$\Delta\Gamma_2 = R_{2,\text{para}} - R_{2,\text{dia}} \quad (1)$$

For $^{15}\text{N}-T_1$, $^{15}\text{N}-T_2$, CLEANEX and $^{15}\text{N}-\{^1\text{H}\}$ heteronuclear NOE, the measurements were carried out at 310 K and 10% D_2O was added as the lock-solvent. See **Table S3** for further details. Analysis and plotting was carried out with RStudio. For the difference plots showed in Fig. 4, Fig. 7, Fig. S5 and Fig. S6 the error associated (δQ) was calculated by propagating the fitting errors of the experimental data (δa and δb) as:

$$\delta Q = (\delta a^2 + \delta b^2)^{1/2} \quad (2)$$

^{15}N -relaxation acquisition parameters. NMR experiments were recorded in on a 18.8 T Bruker Avance III HD+ spectrometer operating at 800 MHz. The $^{15}\text{N}-T_1$ relaxation delays were 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64 and 1.28 s. The $^{15}\text{N}-T_2$ relaxation delays were 0, 0.017, 0.034, 0.068, 0.136, 0.271, 0.407 and 0.543 s. $^{15}\text{N}-\{^1\text{H}\}$ heteronuclear NOE was measured with an interscan delay of 10 s to avoid systematic errors in the measurement of steady-state NOE as a consequence of incomplete equilibration of the magnetization in the reference spectrum [88].

For the 2D ($\text{H}^\alpha\text{C}^\alpha$)CON^{Pro} experiments to determine the ^{15}N R_2 the delay used were 0.080, 0.160, 0.240, 0.320, 0.400, 0.480, 0.560, 0.640 and 0.800 s. The RF field strength used for the CPMG block was 3.1 kHz.

Trp sidechain dynamics. We measured relaxation rates of the $^1\text{H}_{\epsilon 1}-^{15}\text{N}_{\epsilon 1}$ major and minor forms (**Fig. S9D**). Minor forms were also observed for $^1\text{H}_{\epsilon 3}-^{13}\text{C}_{\epsilon 3}$ and $^1\text{H}_\eta-^{13}\text{C}_\eta$ pairs selectively-labelled (**Fig. S9E**).

AUTHOR CONTRIBUTIONS

BM, RK, ICF, and RP conceived the project; BM, RK, ICF, and RP designed the experiments; BM and CCB produced samples; BM, MS and CCB performed NMR experiments with RP and ICF; BM and MS processed and analyzed the data together with RK, ICF, and RP; AB and GK contributed with new analytic tools; BM, RK, ICF, and RP wrote the manuscript, with contributions from all the other authors.

ACKNOWLEDGEMENTS

The support and the use of resources of the CERM/CIRMMP center of Instruct-ERIC, a Landmark ESFRI project, is gratefully acknowledged. This work has been supported in part by iNEXT, Grant number 653706, funded by the Horizon 2020 program of the European Commission, PID: 3972 and 2676. Grant of the Fondazione CR Firenze/AIRC to RP and by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development to RK. BM is grateful to the NGP-net COST action (ID: 38248) for funding a short-term mission to the Magnetic Resonance Center of the University of Florence. BM also acknowledges the DK Integrative Structural Biology funded by the Austrian Science Fund (FWF, project number AW0125821). AB is funded by the Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology, Austria.

- [1] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradović, Intrinsic disorder and protein function, *Biochemistry*. 41 (2002) 6573–82.
- [2] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* 6 (2005) 197–208. doi:10.1038/nrm1589.
- [3] J. Habchi, P. Tompa, S. Longhi, V.N. Uversky, Introducing protein intrinsic disorder, *Chem. Rev.* 114 (2014) 6561–88. doi:10.1021/cr400514h.
- [4] H. Hofmann, a. Soranno, a. Borgia, K. Gast, D. Nettels, B. Schuler, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy, *Proc. Natl. Acad. Sci.* 109 (2012) 16155–16160. doi:10.1073/pnas.1207719109.
- [5] F. Theillet, L. Kalmar, P. Tompa, K. Han, P. Selenko, A.K. Dunker, G.W. Daughdrill, V.N. Uversky, The alphabet of intrinsic disorder, *Intrinsically Disord. Proteins.* 1 (2013) e24360. doi:10.4161/idp.24360.
- [6] C.M. Deber, B. Brodsky, A. Rath, Proline residues in proteins, *Encycl. Life Sci.* (2001) 1–6. doi:10.1002/9780470015902.a0003014.pub2.
- [7] C. Dugave, L. Demange, Cis-trans isomerization of organic molecules and biomolecules: Implications and applications, *Chem. Rev.* 103 (2003) 2475–2532. doi:10.1021/cr0104375.
- [8] U. Reimer, G. Fischer, Local structural changes caused by peptidyl-prolyl cis/trans isomerization in the native state of proteins, *Biophys. Chem.* 96 (2002) 203–212. doi:10.1016/S0301-4622(02)00013-3.
- [9] S. Osváth, M. Gruebele, Proline can have opposite effects on fast and slow protein folding phases, *Biophys. J.* 85 (2003) 1215–1222. doi:10.1016/S0006-3495(03)74557-3.
- [10] R.B. Perez, A. Tischer, M. Auton, S.T. Whitten, Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins., *Proteins.* (2014) 3373–3384. doi:10.1002/prot.24692.
- [11] T.R. Alderson, J.H. Lee, C. Charlier, J. Ying, A. Bax, Propensity for cis-Proline formation in unfolded proteins, *ChemBioChem.* 19 (2018) 37–42. doi:10.1002/cbic.201700548.
- [12] P. Ahuja, F.X. Cantrelle, I. Huvent, X. Hanouille, J. Lopez, C. Smet, J.M. Wieruszeski, I. Landrieu, G.

- Lippens, Proline conformation in a functional Tau fragment, *J. Mol. Biol.* 428 (2016) 79–91. doi:10.1016/j.jmb.2015.11.023.
- [13] R. Glaves, M. Baer, E. Schreiner, R. Stoll, D. Marx, Conformational dynamics of minimal elastin-like polypeptides: The role of proline revealed by molecular dynamics and nuclear magnetic resonance, *ChemPhysChem*. 9 (2008) 2759–2765. doi:10.1002/cphc.200800474.
- [14] G. Fernandez-Ballester, C. Blanes-Mira, L. Serrano, The Tryptophan switch: changing ligand-binding specificity from type I to type II in SH3 domains, *J. Mol. Biol.* 335 (2004) 619–629. doi:10.1016/j.jmb.2003.10.060.
- [15] M.W. MacArthur, J.M. Thornton, Influence of proline residues on protein conformation, *J. Mol. Biol.* 218 (1991) 397–412. doi:10.1016/0022-2836(91)90721-H.
- [16] E.W. Martin, A.S. Holehouse, C.R. Grace, A. Hughes, R. V. Pappu, T. Mittag, Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation, *J. Am. Chem. Soc.* 138 (2016) 15323–15335. doi:10.1021/jacs.6b10272.
- [17] E.B. Gibbs, F. Lu, B. Portz, M.J. Fisher, B.P. Medellin, T.N. Laremore, Y.J. Zhang, D.S. Gilmour, S.A. Showalter, Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain, *Nat. Commun.* 8 (2017) 1–11. doi:10.1038/ncomms15233.
- [18] R. V. Pappu, G.D. Rose, A simple model for polyproline II structure in unfolded states of alanine-based peptides, *Protein Sci.* 11 (2002) 2437–2455. doi:10.1110/Ps.0217402.
- [19] J.A. Vila, H.A. Baldoni, D.R. Ripoll, A. Ghosh, H.A. Scheraga, Polyproline II helix conformation in a proline-rich environment: a theoretical study, *Biophys. J.* 86 (2004) 731–742. doi:10.1016/S0006-3495(04)74151-X.
- [20] W.A. Elam, T.P. Schrank, A.J. Campagnolo, V.J. Hilser, Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites, *Protein Sci.* 22 (2013) 405–17. doi:10.1002/pro.2217.
- [21] M.D. Crabtree, W. Borchers, A. Poosapati, S.L. Shamma, G.W. Daughdrill, J. Clarke, Conserved helix-flanking prolines modulate IDP: target affinity by altering the lifetime of the bound complex, *Biochemistry*. (2017) acs.biochem.7b00179. doi:10.1021/acs.biochem.7b00179.
- [22] L.E. Wong, J. Maier, J. Wienands, S. Becker, C. Griesinger, Sensitivity-enhanced four-dimensional amide–amide correlation NMR experiments for sequential assignment of proline-rich disordered

proteins, *J. Am. Chem. Soc.* 140 (2018) 3518–3522. doi:10.1021/jacs.8b00215.

- [23] Y. Yoshimura, N. V. Kulminskaya, F.A.A. Mulder, Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone ^{15}N or ^{13}C chemical shifts of multiple contiguous residues in highly resolved 3D spectra, *J. Biomol. NMR.* 61 (2015) 109–121. doi:10.1007/s10858-014-9890-7.
- [24] V. Kanelis, L. Donaldson, D.R. Muhandiram, D. Rotin, J.D. Forman-Kay, L.E. Kay, Sequential assignment of proline-rich regions in proteins: Application to modular binding domain complexes, *J. Biomol. NMR.* 16 (2000) 253–259. doi:10.1023/A:1008355012528.
- [25] M.G. Murrall, A. Piai, W. Bermel, I.C. Felli, R. Pierattelli, Proline fingerprint in intrinsically disordered proteins, *ChemBioChem.* 19 (2018) 1625–1629. doi:10.1002/cbic.201800172.
- [26] D.E. Dorman, F.A. Bovey, Carbon-13 magnetic resonance spectroscopy. Spectrum of proline in oligopeptides, *J. Org. Chem.* 38 (1973) 2379–2383. doi:10.1021/jo00953a021.
- [27] C. Grathwohl, K. Wüthrich, The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides, *Biopolymers.* 15 (1976) 2025–2041. doi:10.1002/bip.1976.360151012.
- [28] H. Kovacs, D. Moskau, M. Spraul, Cryogenically cooled probes - A leap in NMR technology, *Prog. Nucl. Magn. Reson. Spectrosc.* 46 (2005) 131–155. doi:10.1016/j.pnmrs.2005.03.001.
- [29] W. Bermel, I. Bertini, I.C. Felli, M. Piccioli, R. Pierattelli, ^{13}C -detected protonless NMR spectroscopy of proteins in solution, *Prog. Nucl. Magn. Reson. Spectrosc.* 48 (2006) 25–45. doi:10.1016/j.pnmrs.2005.09.002.
- [30] K. Takeuchi, G. Heffron, Z.Y.J. Sun, D.P. Frueh, G. Wagner, Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments, *J. Biomol. NMR.* 47 (2010) 271–282. doi:10.1007/s10858-010-9430-z.
- [31] S. Chhabra, P. Fischer, K. Takeuchi, A. Dubey, J.J. Ziarek, A. Boeszoermyeni, D. Mathieu, W. Bermel, N.E. Davey, G. Wagner, H. Arthanari, ^{15}N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins, *Proc. Natl. Acad. Sci.* 115 (2018) E1710–E1719. doi:10.1073/pnas.1717560115.
- [32] R.B. Pritchard, D.F. Hansen, Characterising side chains in large proteins by protonless ^{13}C -detected NMR spectroscopy, *Nat. Commun.* 10 (2019) 1747. doi:10.1038/s41467-019-09743-4.
- [33] A. Escobedo, B. Topal, M.B.A. Kunze, J. Aranda, G. Chiesa, D. Mungianu, G. Bernardo-Seisdedos,

- B. Eftekharzadeh, M. Gairí, R. Pierattelli, I.C. Felli, T. Diercks, O. Millet, J. García, M. Orozco, R. Crehuet, K. Lindorff-Larsen, X. Salvatella, Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor, *Nat. Commun.* 10 (2019) 2034. doi:10.1038/s41467-019-09923-2.
- [34] W. Bermel, I. Bertini, I.C. Felli, R. Kümmerle, R. Pierattelli, Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins, *J. Magn. Reson.* 178 (2006) 56–64. doi:10.1016/j.jmr.2005.08.011.
- [35] W. Bermel, I. Bertini, I.C. Felli, R. Peruzzini, R. Pierattelli, Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins, *ChemPhysChem.* 11 (2010) 689–695. doi:10.1002/cphc.200900772.
- [36] S. Gil, T. Hošek, Z. Solyom, R. Kümmerle, B. Brutscher, R. Pierattelli, I.C. Felli, NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions, *Angew. Chemie - Int. Ed.* 52 (2013) 11808–11812. doi:10.1002/anie.201304272.
- [37] E.B. Gibbs, R.W. Kriwacki, Direct detection of carbon and nitrogen nuclei for high-resolution analysis of intrinsically disordered proteins using NMR spectroscopy, *Methods.* 138–139 (2018) 39–46. doi:10.1016/j.ymeth.2018.01.004.
- [38] W. Bermel, I. Bertini, I.C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins, *J. Am. Chem. Soc.* 128 (2006) 3918–3919. doi:10.1021/ja0582206.
- [39] I.C. Felli, R. Pierattelli, S.J. Glaser, B. Luy, Relaxation-optimised Hartmann–Hahn transfer using a specifically Tailored MOCCA-XY16 mixing sequence for carbonyl–carbonyl correlation spectroscopy in ^{13}C direct detection NMR experiments, *J. Biomol. NMR.* 43 (2009) 187–196. doi:10.1007/s10858-009-9302-6.
- [40] S. Balayssac, B. Jiménez, M. Piccioli, ^{13}C direct detected COCO-TOCSY: A tool for sequence specific assignment and structure determination in protonless NMR experiments, *J. Magn. Reson.* 182 (2006) 325–329. doi:10.1016/j.jmr.2006.06.021.
- [41] W. Bermel, I. Bertini, J. Chill, I.C. Felli, N. Haba, V. Kumar M V, R. Pierattelli, Exclusively heteronuclear ^{13}C -detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs), *ChemBioChem.* 13 (2012) 2425–32. doi:10.1002/cbic.201200447.

- [42] I.C. Felli, L. Gonnelli, R. Pierattelli, In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins, *Nat. Protoc.* 9 (2014) 2005–2016. doi:10.1038/nprot.2014.124.
- [43] A. Pietras, A.M. Katz, E.J. Ekström, B. Wee, J.J. Halliday, K.L. Pitter, J.L. Werbeck, N.M. Amankulor, J.T. Huse, E.C. Holland, Osteopontin-CD44 signaling in the glioma perivascular niche enhances cancer stem cell phenotypes and promotes aggressive tumor growth., *Cell Stem Cell.* 14 (2014) 357–69. doi:10.1016/j.stem.2014.01.005.
- [44] L.A. Shevde, R.S. Samant, Role of osteopontin in the pathophysiology of cancer, *Matrix Biol.* 37 (2014) 131–141. doi:10.1016/j.matbio.2014.03.001.
- [45] C.C. Kazanecki, D.J. Uzwiak, D.T. Denhardt, Control of osteopontin signaling and function by post-translational phosphorylation and protein folding, *J. Cell. Biochem.* 102 (2007) 912–924. doi:10.1002/jcb.21558.
- [46] P.H. Anborgh, J.C. Mutrie, A.B. Tuck, A.F. Chambers, Role of the metastasis-promoting protein osteopontin in the tumour microenvironment, *J. Cell. Mol. Med.* 14 (2010) 2037–2044. doi:10.1111/j.1582-4934.2010.01115.x.
- [47] Y.U. Katagiri, J. Sleeman, H. Fujii, P. Herrlich, H. Hotta, K. Tanaka, S. Chikuma, H. Yagita, K. Okumura, M. Murakami, I. Saiki, A.F. Chambers, T. Uede, CD44 variants but not CD44s cooperate with beta1-containing integrins to permit cells to bind to osteopontin independently of arginine-glycine-aspartic acid, thereby stimulating cell motility and chemotaxis., *Cancer Res.* 59 (1999) 219–26.
- [48] D. Kurzbach, T.C. Schwarz, G. Platzner, S. Höfler, D. Hinderberger, R. Konrat, Compensatory adaptations of structural dynamics in an intrinsically disordered protein complex, *Angew. Chemie - Int. Ed.* 53 (2014) 3840–3843. doi:10.1002/anie.201308389.
- [49] W. Bermel, I.C. Felli, L. Gonnelli, W. Kołmiński, A. Piai, R. Pierattelli, A. Zawadzka-Kazimierczuk, High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins, *J. Biomol. NMR.* 57 (2013) 353–361. doi:10.1007/s10858-013-9793-z.
- [50] J.-S. Hu, A. Bax, Measurement of three-bond ^{13}C – ^{13}C J-couplings between carbonyl and carbonyl/carboxyl Carbons in isotopically enriched proteins, *J. Am. Chem. Soc.* 118 (1996) 8170–8171. doi:10.1021/ja9616239.

- [51] F. Li, J.H. Lee, A. Grishaev, J. Ying, A. Bax, High accuracy of Karplus equations for relating three-bond J Couplings to protein backbone torsion angles, *ChemPhysChem*. 16 (2015) 572–578. doi:10.1002/cphc.201402704.
- [52] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics, *J. Biomol. NMR*. 24 (2002) 149–154. doi:10.1023/A:1020997118364.
- [53] Y. Shen, A. Bax, Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts, *J. Biomol. NMR*. 46 (2010) 199–204. doi:10.1007/s10858-009-9395-y.
- [54] H.J. Dyson, M. Rance, R.A. Houghten, R.A. Lerner, P.E. Wright, Folding of immunogenic peptide fragments of proteins in water solution, *J. Mol. Biol.* 201 (1988) 161–200. doi:10.1016/0022-2836(88)90446-9.
- [55] K.M. Thomas, D. Naduthambi, N.J. Zondlo, Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction, *J. Am. Chem. Soc.* 128 (2006) 2216–2217. doi:10.1021/ja057901y.
- [56] F. Nardi, J. Kemmink, M. Sattler, R.C. Wade, The cisproline(i-1)-aromatic(i) interaction: Folding of the Ala-cisPro-Tyr peptide characterized by NMR and theoretical approaches, *J. Biomol. NMR*. 17 (2000) 63–77.
- [57] A.G. Palmer, NMR characterization of the dynamics of biomacromolecules, *Chem. Rev.* 104 (2004) 3623–3640. doi:10.1021/cr030413t.
- [58] G. Parigi, N. Rezaei-Ghaleh, A. Giachetti, S. Becker, C. Fernandez, M. Blackledge, C. Griesinger, M. Zweckstetter, C. Luchinat, Long-Range Correlated Dynamics in Intrinsically Disordered Proteins, *J. Am. Chem. Soc.* 136 (2014) 16201–16209. doi:10.1021/ja506820r.
- [59] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R.W.H. Ruigrok, M.R. Jensen, M. Blackledge, Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR Relaxation, *J. Am. Chem. Soc.* (2016) jacs.6b02424. doi:10.1021/jacs.6b02424.
- [60] S.N. Khan, C. Charlier, R. Augustyniak, N. Salvi, V. Déjean, G. Bodenhausen, O. Lequin, P. Pelupessy, F. Ferrage, Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation, *Biophys. J.* 109 (2015) 988–999. doi:10.1016/j.bpj.2015.06.069.
- [61] N. Salvi, A. Abyzov, M. Blackledge, Solvent-dependent segmental dynamics in intrinsically disordered proteins, *Sci. Adv.* 5 (2019). doi:10.1126/sciadv.aax2348.

- Journal Pre-proof
- [62] N. Rezaei-Ghaleh, G. Parigi, A. Soranno, A. Holla, S. Becker, B. Schuler, C. Luchinat, M. Zweckstetter, Local and global dynamics in intrinsically disordered synuclein, *Angew. Chemie - Int. Ed.* 57 (2018) 15262–15266. doi:10.1002/anie.201808172.
- [63] G. Wagner, NMR relaxation and protein mobility, *Curr. Opin. Struct. Biol.* 3 (1993) 748–754. doi:10.1016/0959-440X(93)90059-T.
- [64] J. Wirmer, C. Schlörb, J. Klein-Seetharaman, R. Hirano, T. Ueda, T. Imoto, H. Schwalbe, Modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations, *Angew. Chemie - Int. Ed.* 43 (2004) 5780–5785. doi:10.1002/anie.200460907.
- [65] M. Xie, L. Yu, L. Bruschweiler-Li, X. Xiang, A.L. Hansen, R. Brüschweiler, Functional protein dynamics on uncharted time scales detected by nanoparticle-assisted NMR spin relaxation, *Sci. Adv.* 5 (2019) eaax5560. doi:10.1126/sciadv.aax5560.
- [66] R. Brueschweiler, P.E. Wright, NMR order parameters of biomolecules: A new analytical representation and application to the Gaussian Axial fluctuation model, *J. Am. Chem. Soc.* 116 (1994) 8426–8427. doi:10.1021/ja00097a084.
- [67] L. Shi, A.E. Holliday, H. Shi, F. Zhu, M.A. Ewing, D.H. Russell, D.E. Clemmer, Characterizing intermediates along the transition from polyproline I to polyproline II using ion mobility spectrometry-mass spectrometry, *J. Am. Chem. Soc.* 136 (2014) 12702–12711. doi:10.1021/ja505899g.
- [68] C. Camilloni, A. De Simone, W.F. Vranken, M. Vendruscolo, Determination of secondary structure populations in disordered states of proteins using Nuclear Magnetic Resonance chemical shifts, (2012).
- [69] B. Mateos, R. Konrat, R. Pierattelli, I.C. Felli, NMR characterization of long-range contacts in intrinsically disordered proteins from paramagnetic relaxation enhancement in ^{13}C direct-detection experiments, *ChemBioChem.* 20 (2019) 335–339. doi:10.1002/cbic.201800539.
- [70] T.R. Alderson, J.L.P. Benesch, A.J. Baldwin, Proline isomerization in the C-terminal region of HSP27, *Cell Stress Chaperones.* 22 (2017) 639–651. doi:10.1007/s12192-017-0791-z.
- [71] N.J. Zondlo, Aromatic–Proline interactions: electronically tunable CH/ π interactions, *Acc. Chem. Res.* 46 (2013) 1039–1049. doi:10.1021/ar300087y.
- [72] C.L. Gustafson, N.C. Parsley, H. Asimgil, H.W. Lee, C. Ahlback, A.K. Michael, H. Xu, O.L. Williams, T.L. Davis, A.C. Liu, C.L. Partch, A slow conformational switch in the BMAL1 transactivation domain

- [73] A. Campen, R. Williams, C. Brown, J. Meng, V. Uversky, A. Dunker, TOP-IDP scale: A new amino acid scale measuring propensity for intrinsic disorder, *Protein Pept. Lett.* 15 (2008) 956–963. doi:10.2174/092986608785849164.
- [74] J. A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, *Biophys. J.* 98 (2010) 2383–90. doi:10.1016/j.bpj.2010.02.006.
- [75] R.M. Williams, Z. Obradovi, V. Mathura, W. Braun, E.C. Garner, J. Young, S. Takayama, C.J. Brown, A.K. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, *Pac. Symp. Biocomput.* 100 (2001) 89–100.
- [76] M. Arbesú, M. Maffei, T.N. Cordeiro, J.M.C. Teixeira, Y. Pérez, P. Bernadó, S. Roche, M. Pons, The Unique Domain forms a fuzzy intramolecular complex in Src Family Kinases, *Structure.* 25 (2017) 630–640.e4. doi:10.1016/j.str.2017.02.011.
- [77] M.C. Baxa, E.J. Haddadian, J.M. Jumper, K.F. Freed, T.R. Sosnick, Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations, *Proc. Natl. Acad. Sci.* 111 (2014) 15396–15401. doi:10.1073/pnas.1407768111.
- [78] C.J. Oldfield, Y. Cheng, M.S. Cortese, P. Romero, V.N. Uversky, A.K. Dunker, Coupled folding and binding with α -helix-forming molecular recognition elements, *Biochemistry.* 44 (2005) 12454–12470. doi:10.1021/bi050736e.
- [79] D. Kurzbach, G. Platzer, T.C. Schwarz, M. a Henen, R. Konrat, D. Hinderberger, Cooperative unfolding of compact conformations of the intrinsically disordered protein osteopontin, *Biochemistry.* 52 (2013) 5167–75. doi:10.1021/bi400502c.
- [80] D. Kurzbach, A. Vanas, A.G. Flamm, N. Tarnoczi, G. Kontaxis, N. Maltar-Strmečki, K. Widder, D. Hinderberger, R. Konrat, Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference, *Phys. Chem. Chem. Phys.* 18 (2016) 5753–5758. doi:10.1039/C5CP04858C.
- [81] R.K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proc. Natl. Acad. Sci.* 110 (2013) 13392–13397. doi:10.1073/pnas.1304749110.
- [82] Y. Naranjo, M. Pons, R. Konrat, Meta-structure correlation in protein space unveils different selection

rules for folded and intrinsically disordered proteins., *Mol. Biosyst.* 8 (2012) 411–6. doi:10.1039/c1mb05367a.

- [83] D. Kurzbach, A. Beier, A. Vanas, A.G. Flamm, G. Platzner, T.C. Schwarz, R. Konrat, NMR probing and visualization of correlated structural fluctuations in intrinsically disordered proteins, *Phys. Chem. Chem. Phys.* 19 (2017) 10651–10656. doi:10.1039/C7CP00430C.
- [84] I. Pritišanac, R.M. Vernon, A.M. Moses, J.D. Forman Kay, Entropy and information within intrinsically disordered protein regions, *Entropy*. 21 (2019) 662. doi:10.3390/e21070662.
- [85] W. Bermel, I. Bertini, V. Csizmok, I.C. Felli, R. Pierattelli, P. Tompa, H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins, *J. Magn. Reson.* 198 (2009) 275–281. doi:10.1016/j.jmr.2009.02.012.
- [86] R.L.J. Keller, *Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment*, 2005. doi:10.3929/ETHZ-A-005068942.
- [87] W. Lee, M. Tonelli, J.L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy, *Bioinformatics*. 31 (2015) 1325–1327. doi:10.1093/bioinformatics/btu830.
- [88] F. Ferrage, A. Piserchio, D. Cowburn, R. Ghose, On the measurement of ^{15}N - ^1H nuclear Overhauser effects, *J. Magn. Reson.* 192 (2008) 302–313. doi:10.1016/j.jmr.2008.03.011.

FIGURE LEGENDS

Fig. 1. Observation of prolines by ^{13}C -detection NMR. (A) ^{13}C -detected 2D CON spectrum recorded on *Coturnix japonica* Osteopontin. Note the good dispersion of cross-peaks and the presence of prolines signals around 134-140 ppm $\delta(^{15}\text{N})$. (B) Proline region in the 2D CON spectrum. The main forms of prolines are observed (left) and minor forms appear (right) when the contour levels are lowered. In the inset, the level of the 2D slice and the relative intensity of major and minor forms are reported. Tentative assignments lacking further connectivities are indicated with an asterisk. (C) Primary sequence of the *Coturnix japonica* Osteopontin construct used in this work. Proline residues are indicated in red, the RGD motif (integrin binding site) is in blue. Underlined residues indicate amino acid spacing of ten residues (note that the sequence starts at residue M45). Residues located in the green box are part of the compact region in OPN. This compact state has extensively been characterized by paramagnetic relaxation enhancement (PRE) and interference (PRI) data, displaying cooperative unfolding under denaturing conditions [48].

Fig. 2. Assignment of *cis* forms in prolines. Assignment strips for both *trans* and *cis* forms of *Coturnix japonica* Osteopontin OPN. (A) Strip plots of the region 233-238 of the 3D (H)CBCACON allowing the connectivity of dipeptides by C^β and C^α carbons. The corresponding patterns of major and minor forms are very similar because they involve the same amino acid type in both *trans*-Pro and *cis*-Pro forms. (B) 3D (HCA)COCON strip plots of region 233-238 showing the connectivity between at least three consecutive C' signals for both *trans* and *cis* forms. As the amplitude of $^3\text{J}_{\text{C}'\text{C}'}$ determines the intensity of the off-diagonal signals, the connectivity between a *cis*-proline (i-1) and its C-terminal neighbor (i) can have vanishingly low sensitivity due to a preferred ϕ angle of $[-60^\circ, -90^\circ]$ (see pair P236-A237).

Fig. 3. Characterization of *cis*-Pro in Osteopontin. (A) Sidechain C γ shifts of valines (upper) and C α , C β , C δ and C γ shifts of prolines (lower) of Val-Pro pairs in OPN. Characteristic C α , C β and C δ of *cis*/*trans* isoforms are indicated with dashed lines. (B) Percentage of *cis* population obtained by measuring the ratio of peak volumes between *cis* and *trans* forms for proline signals in *Coturnix japonica* Osteopontin. Orange bars correspond to proline residues (i) with an aromatic in position $i\pm 1$.

Fig. 4. ^{15}N relaxation of both major and minor forms. (A) 800 MHz ^{15}N -R $_1$, ^{15}N -R $_2$ and ^{15}N - $\{^1\text{H}\}$ NOE relaxation parameters of Osteopontin measured at 310K for both major (red) and minor (black) forms of non-Pro residues. Error bars indicate the fitting errors (^{15}N -R $_1$ and ^{15}N -R $_2$) and the error propagation of intensity ratios based to the noise level (hetNOE). (B) Differences in ^{15}N relaxation parameters between minor and major forms. Statistically meaningless values are depicted in grey. Proline positions are indicated as yellow circles.

Fig. 5. ^{15}N -R $_2$ relaxation rates for *trans* prolines (orange) and *cis* prolines (cyan) in *Coturnix japonica* OPN measured with a ^{13}C -detection scheme. Note the increased R $_2$ values in *cis* prolines, especially for the residues in regions with an accentuated difference in PPII content between *cis* and *trans* conformers, as shown in Figure S5.

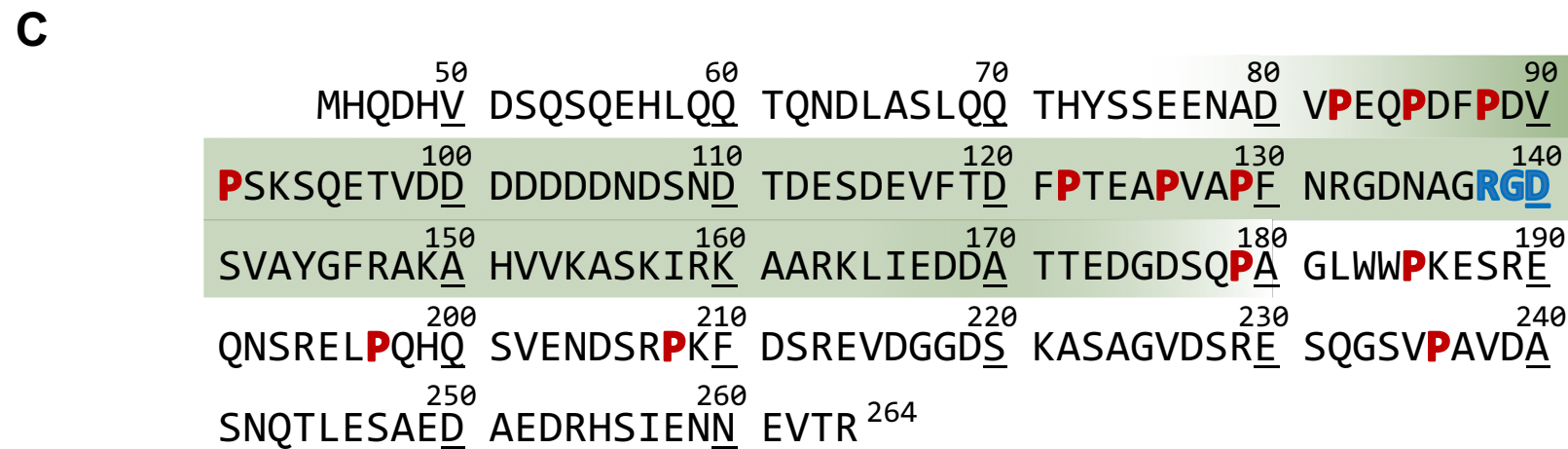
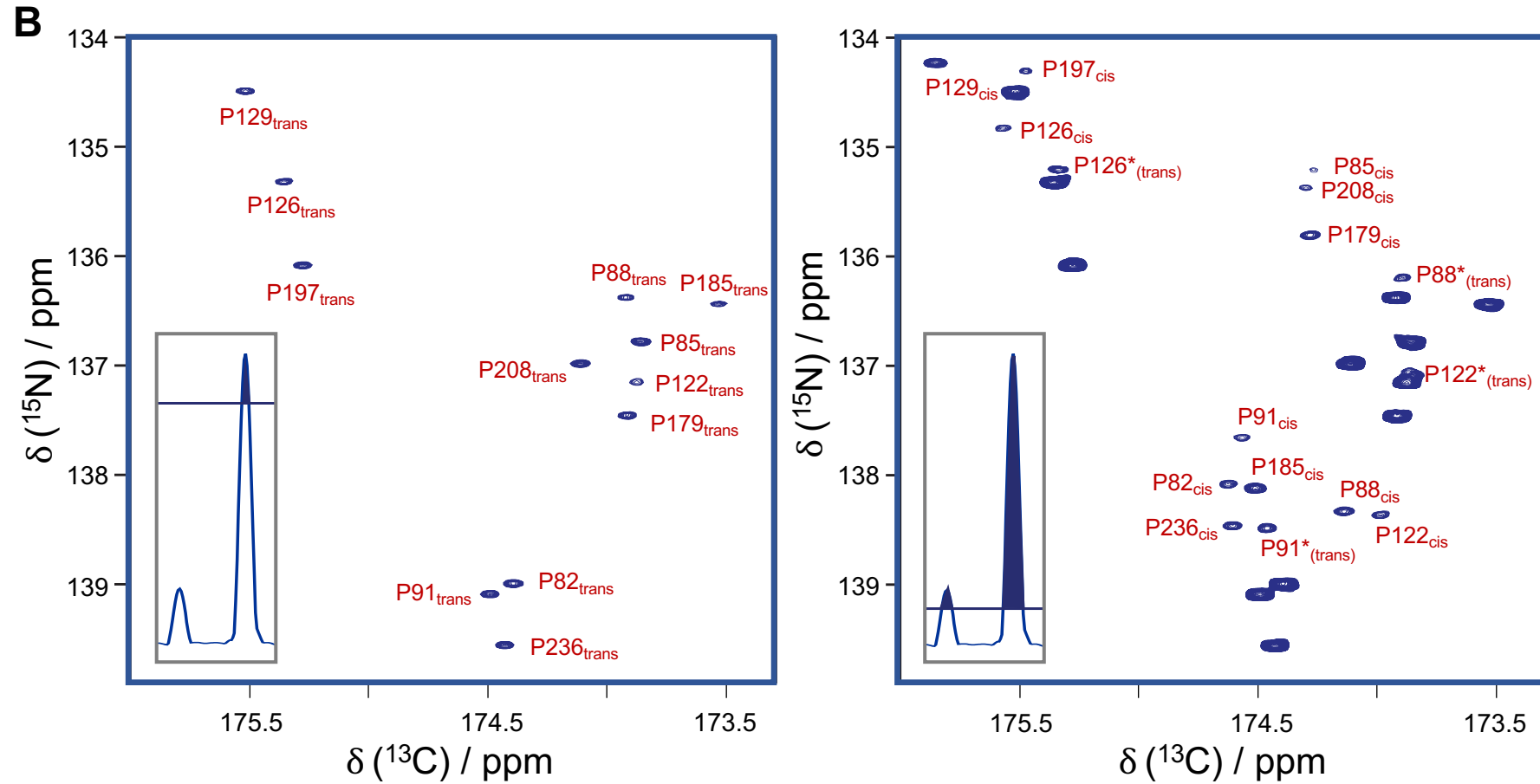
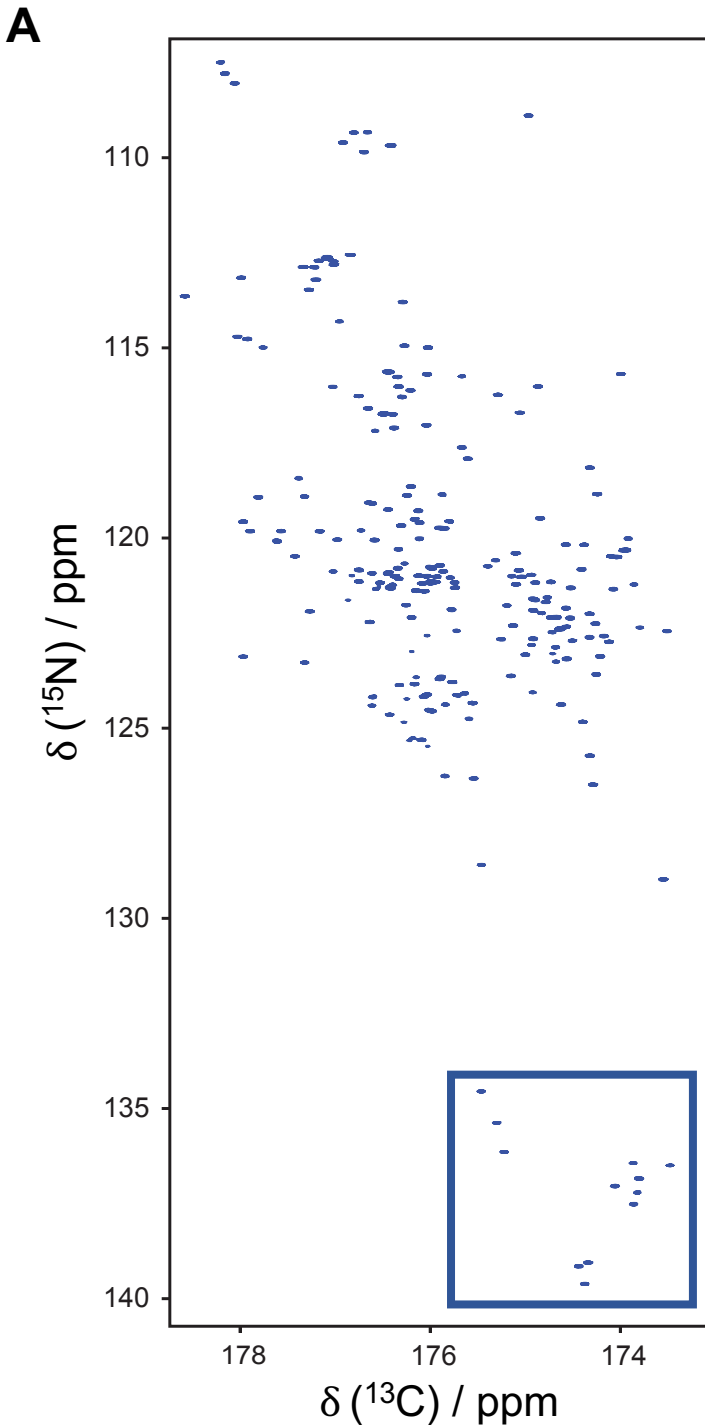
Fig. 6. Secondary structure populations from NMR shifts. (A) Secondary structure populations extracted using the $\delta 2\text{D}$ software [68] of the major forms arising from *trans*-Pro (left) or the minor forms arising from *cis*-Pro (right). (B) Secondary structure populations differences. The secondary structure populations calculated with the minor forms shifts, arising from nearby *cis* prolines, were subtracted from the corresponding major forms arising from *trans* prolines. The minor forms are characterized by lower PPII populations compensated by either coil or β -strand conformations.

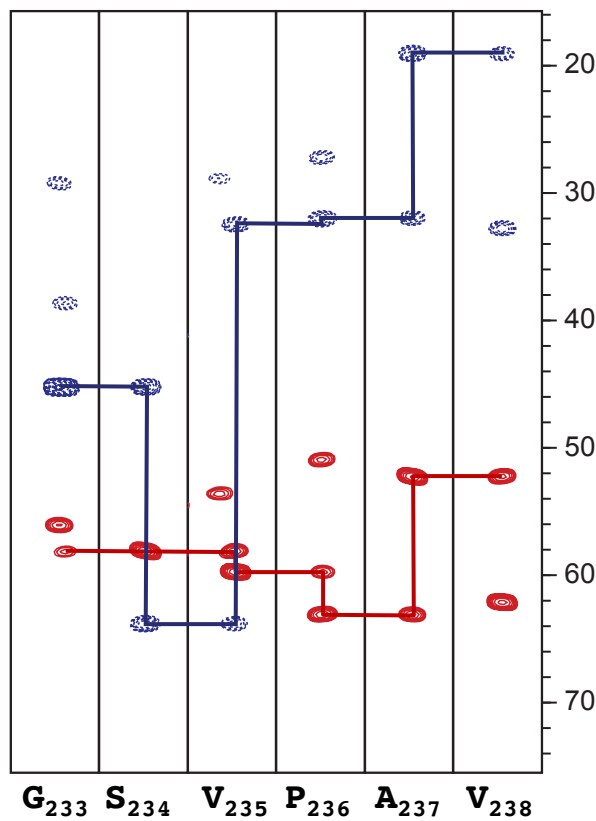
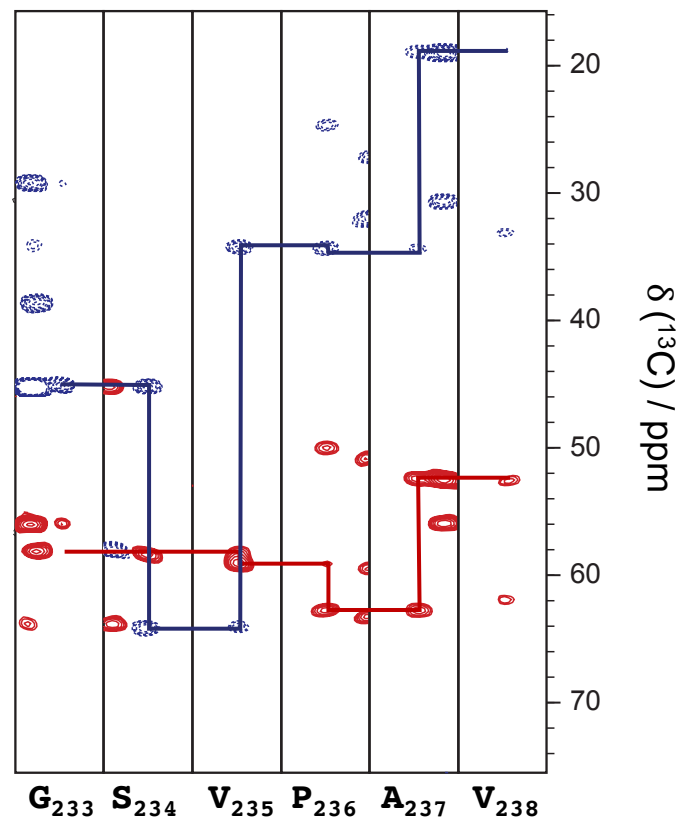
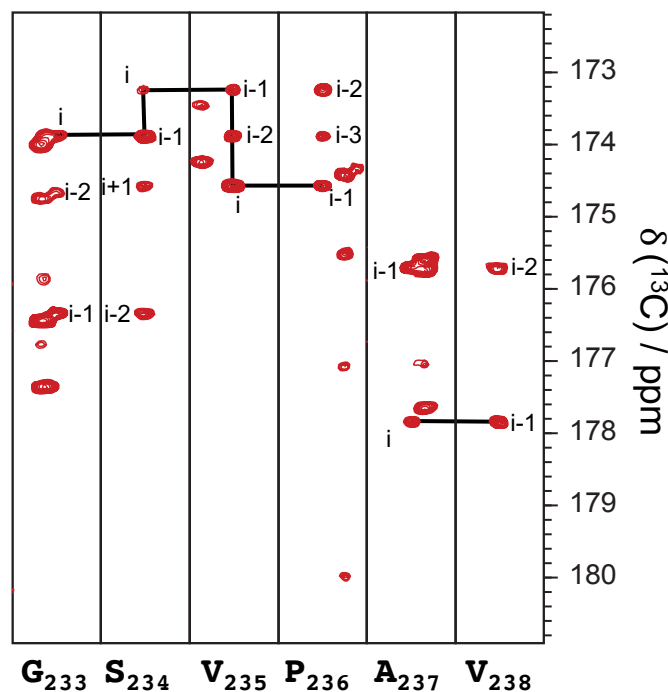
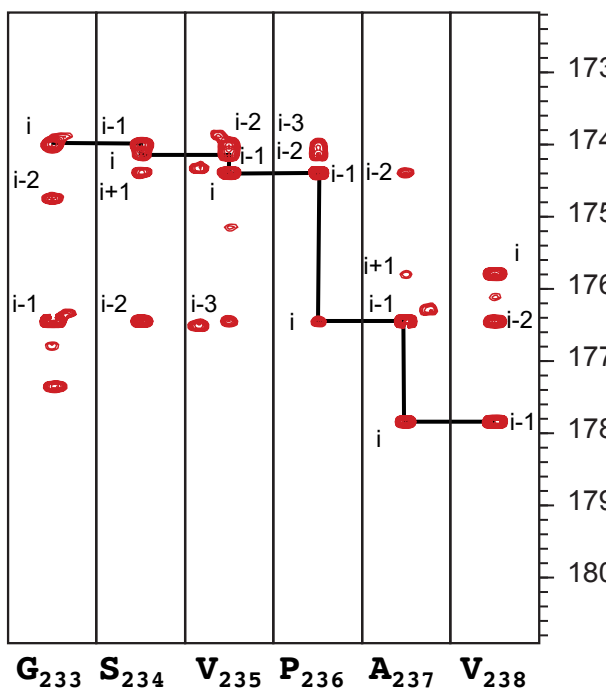
Fig. 7. ^{13}C -detected PRE experiments. (A) ^{13}C '-detected PRE profiles (H $^\alpha$ -CON) of Cys mutants 54, 108, 188 and 247 (orange circles) at 310K. Grey bars correspond to major

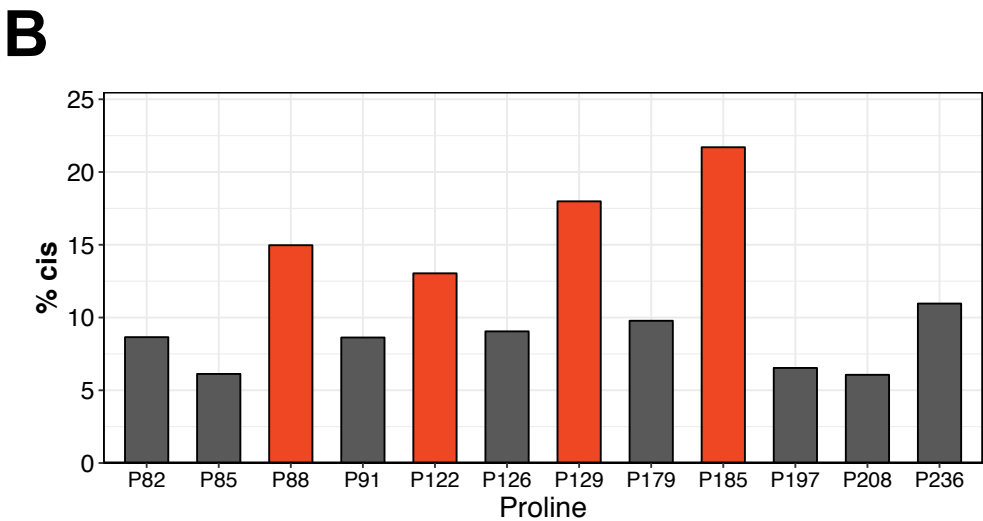
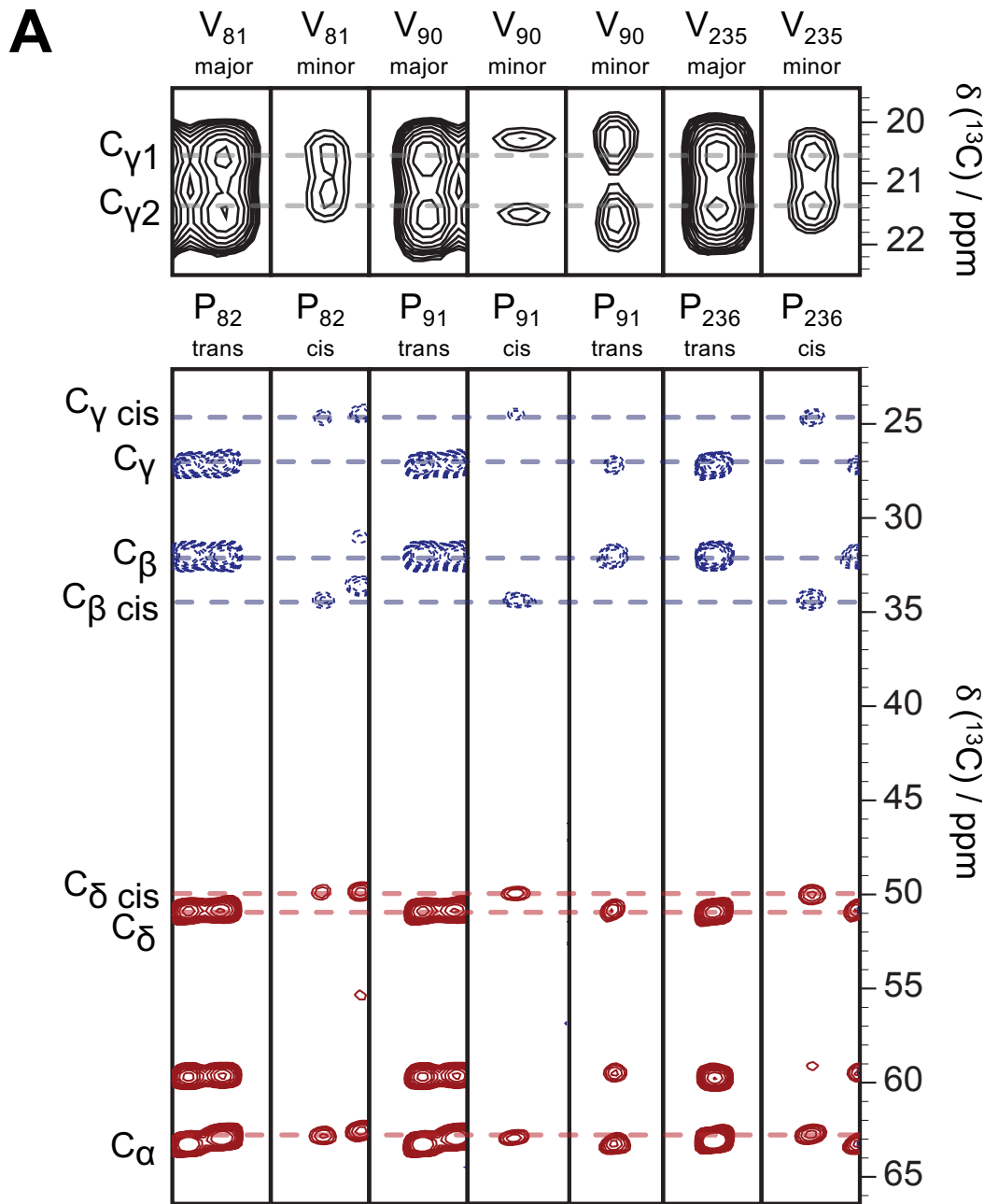
forms and green bars to the minor ones. Error bars represent the quantification uncertainty based on the noise level. (B) Differences in $^{13}\text{C}'$ -detected PREs between minor (stemming from *cis*-Pro) and major (stemming from *trans*-Pro) forms are plotted as a function of residue position. Negative values (red) indicate more effective PRE effects in the minor form (lower signal intensity due to larger paramagnetic effect), while positive differences are indicative of smaller paramagnetic relaxation. Statistically insignificant changes are depicted in grey. Proline positions are indicated as yellow circles. The position of the spin label is indicated by large orange circles.

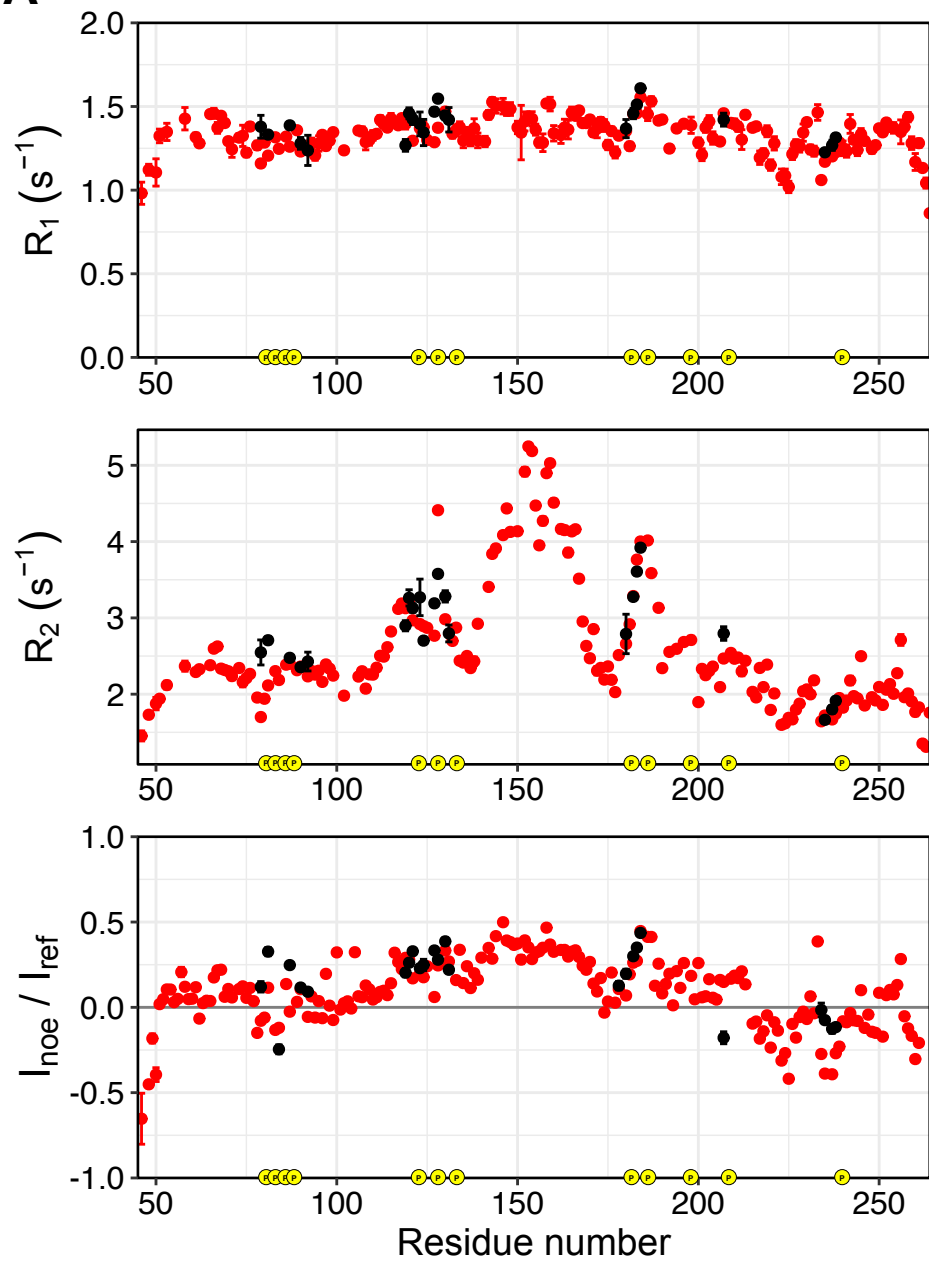
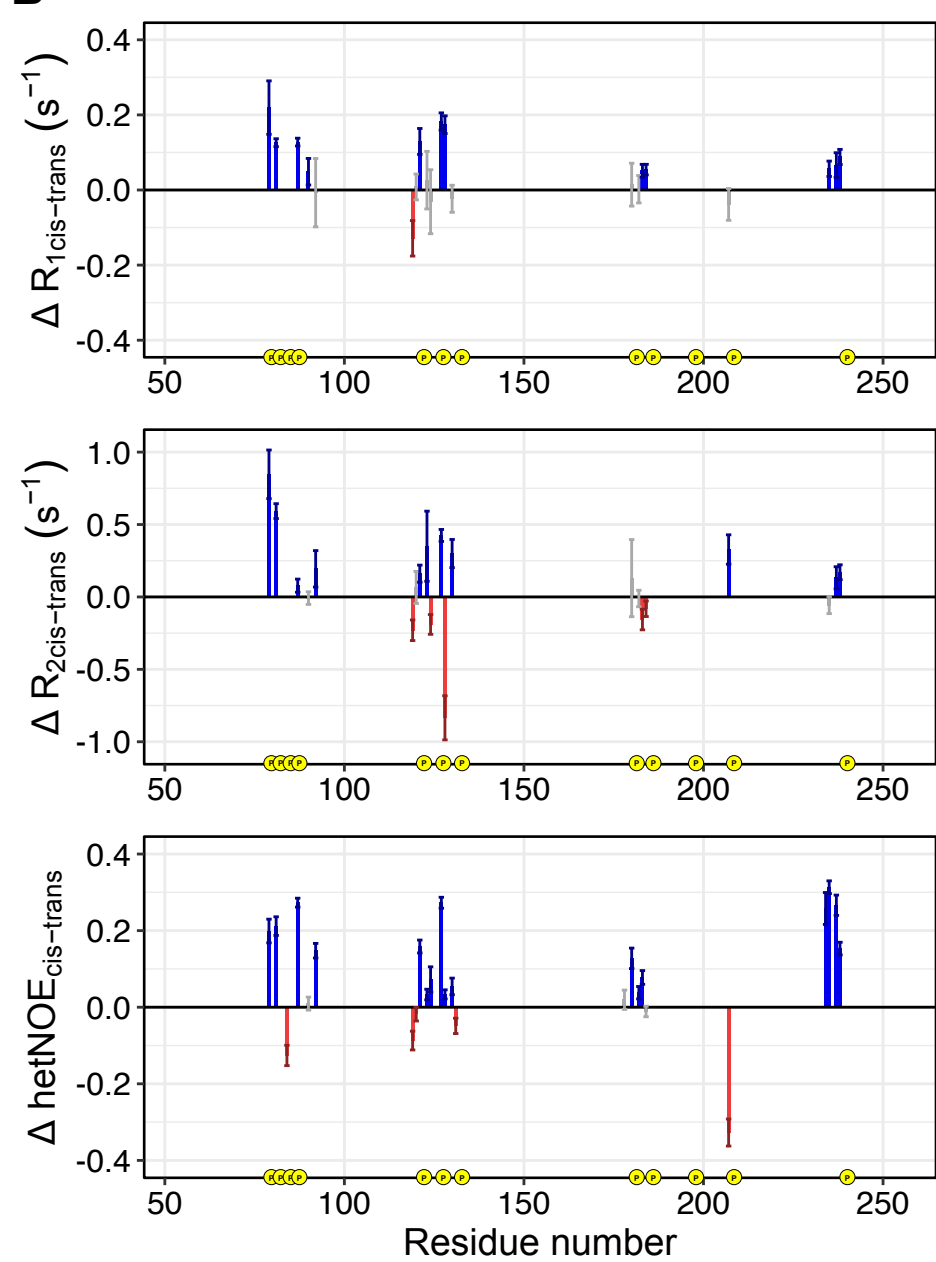
Fig. 8. Differential Trp side-chain structural dynamics of minor (stemming from *cis*-Pro) and major (stemming from *trans*-Pro) forms. (A) ^{15}N - $\{^1\text{H}\}$ NOE experiments (left: reference; right: ^1H -saturation), (B) ^{15}N - $\{^1\text{H}\}$ intensity ratios ($I_{\text{sat}}/I_{\text{ref}}$) for minor and major forms of W183 and W184. (C) Amide $^1\text{H}^{\text{N}}$ to side-chain ^1H NOEs for minor and major forms of W183 and W184, obtained in a ^{15}N 3D NOESY-HSQC experiment. Differences in NOESY patterns are indicated by circles, and shifted resonances are marked with arrows. (D) Schematic 3D structural model of the sequence motif –WWP– in *cis*-Pro and *trans*-Pro forms. In the *cis*-Pro distinct CH-aromatic interactions lead to restricted local mobility.

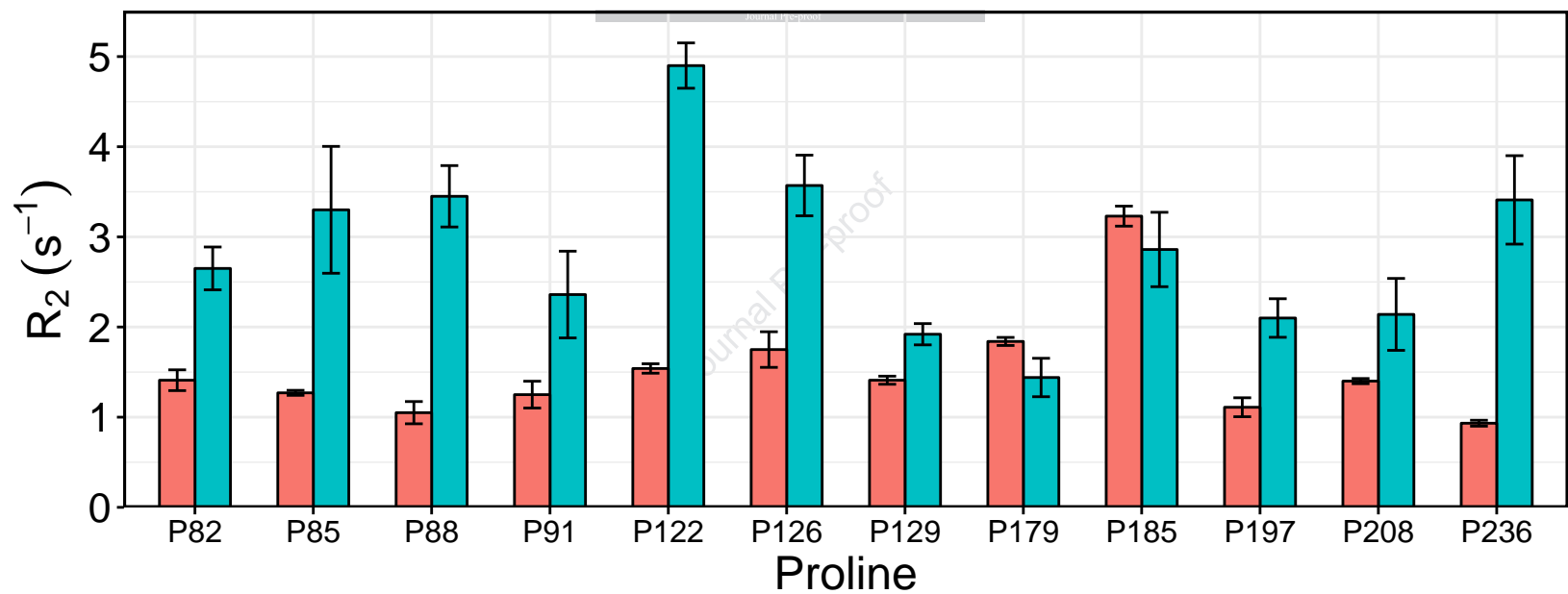
Fig. 9. Pearson correlation map calculated from the $^1\text{H}^{\text{N}}$ - T_2 PRE rates of 10 cysteine mutants (*Coturnix japonica*) or 9 cysteine mutants (*Homo sapiens*). All measurements were carried out at 800MHz and 298K. Plots on top of the Pearson correlation map represent the charge and proline distributions. Red and purple dots correspond to context-dependent peculiar proline residues (i.e. aro-Pro and pSer-Pro sites, respectively). Green circles within the correlation map indicate the positions of proline residues along the protein sequence. Dashed squares represent regions of distinct structural compaction. Residues within one of these regions undergo correlated structural fluctuations. Coupling to other regions can be correlated (orange to red), anti-correlated (light-blue to dark-blue) or uncorrelated (light-yellow to light-blue).

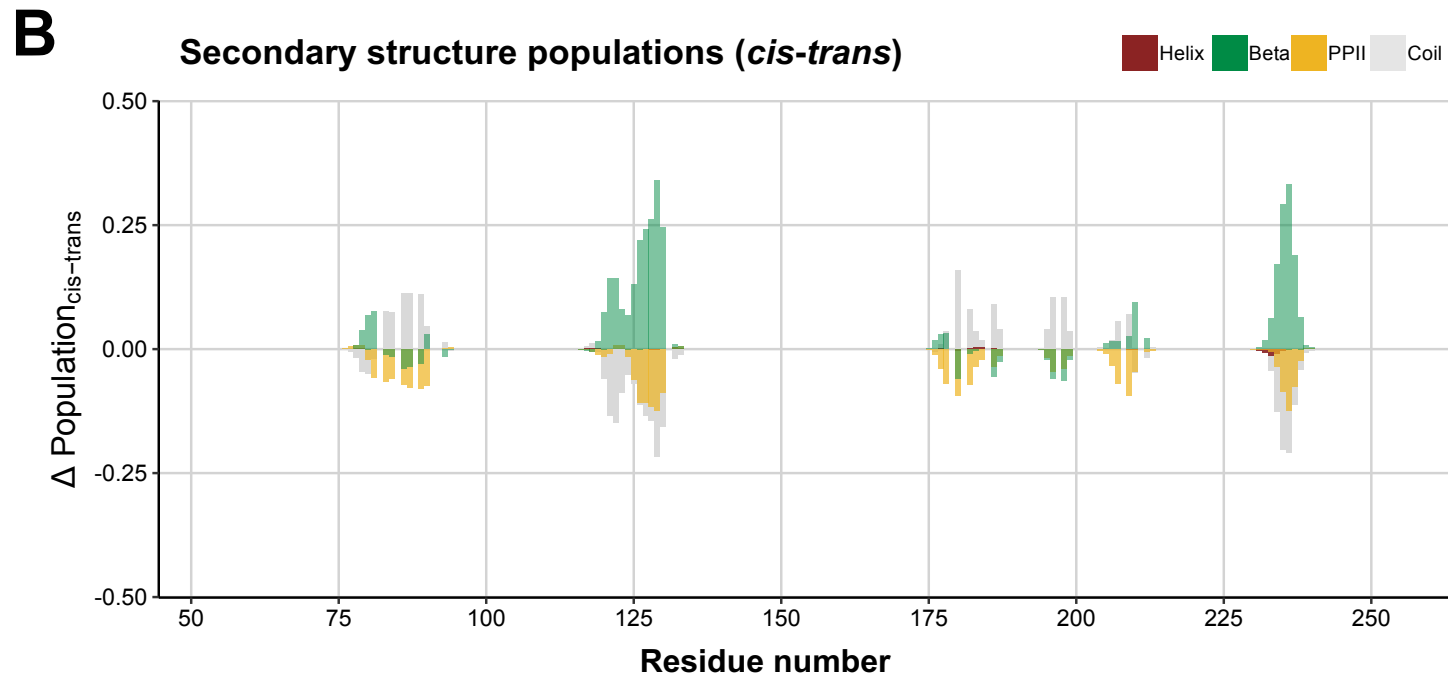
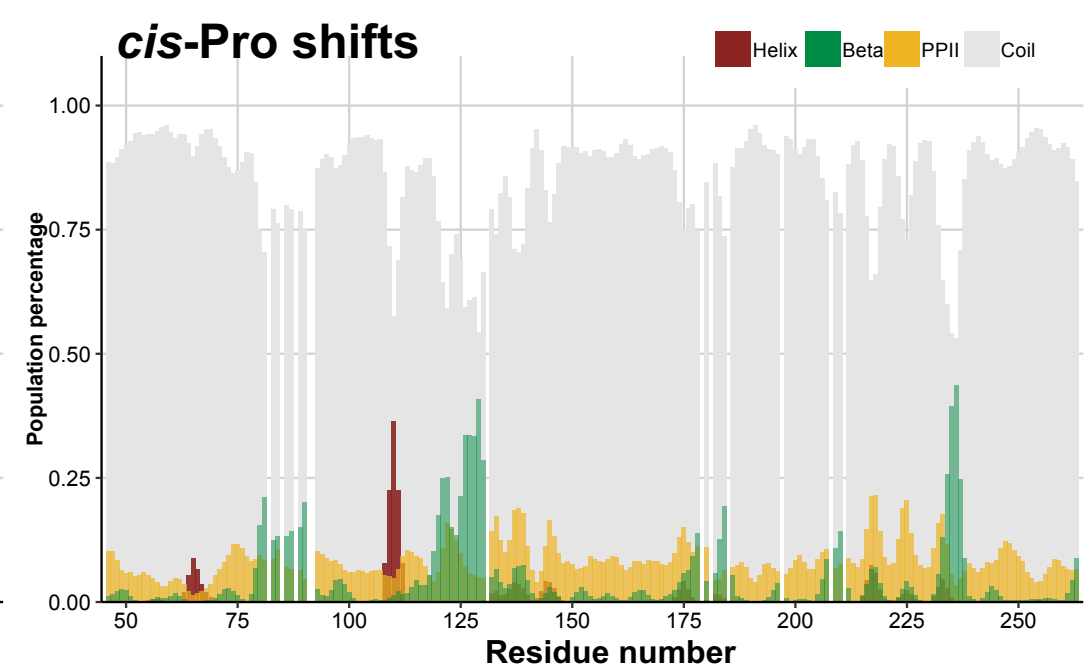
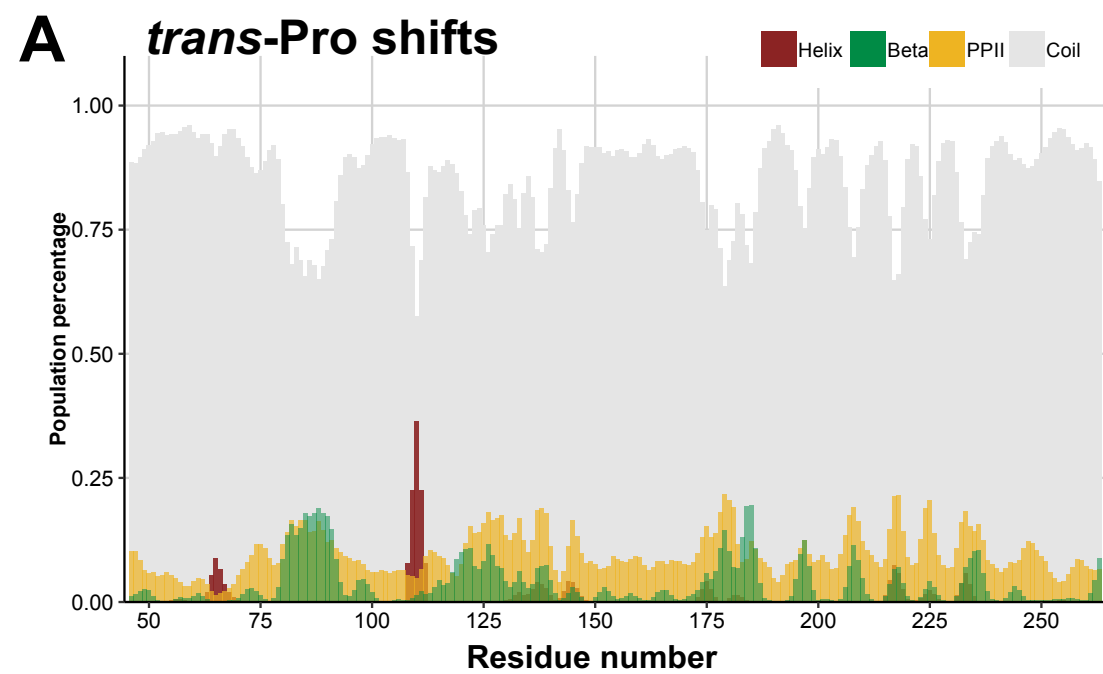


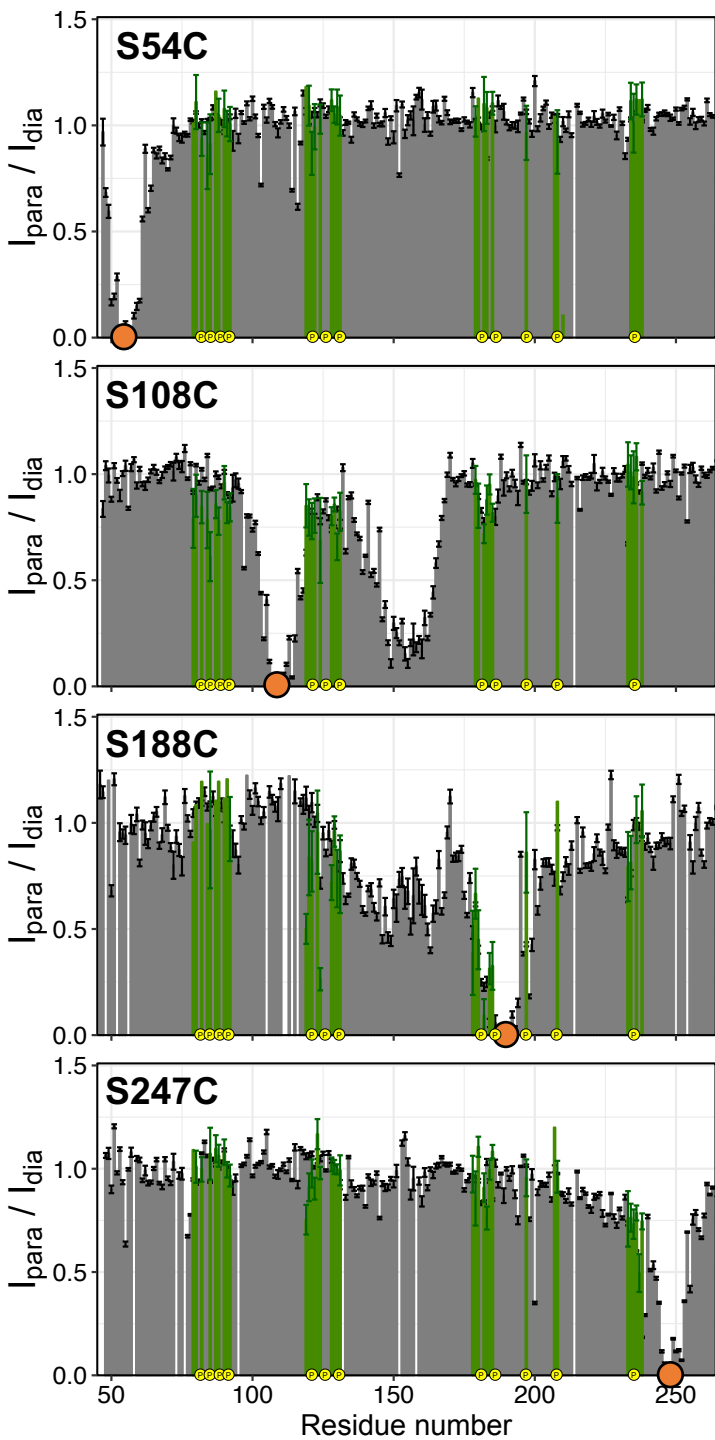
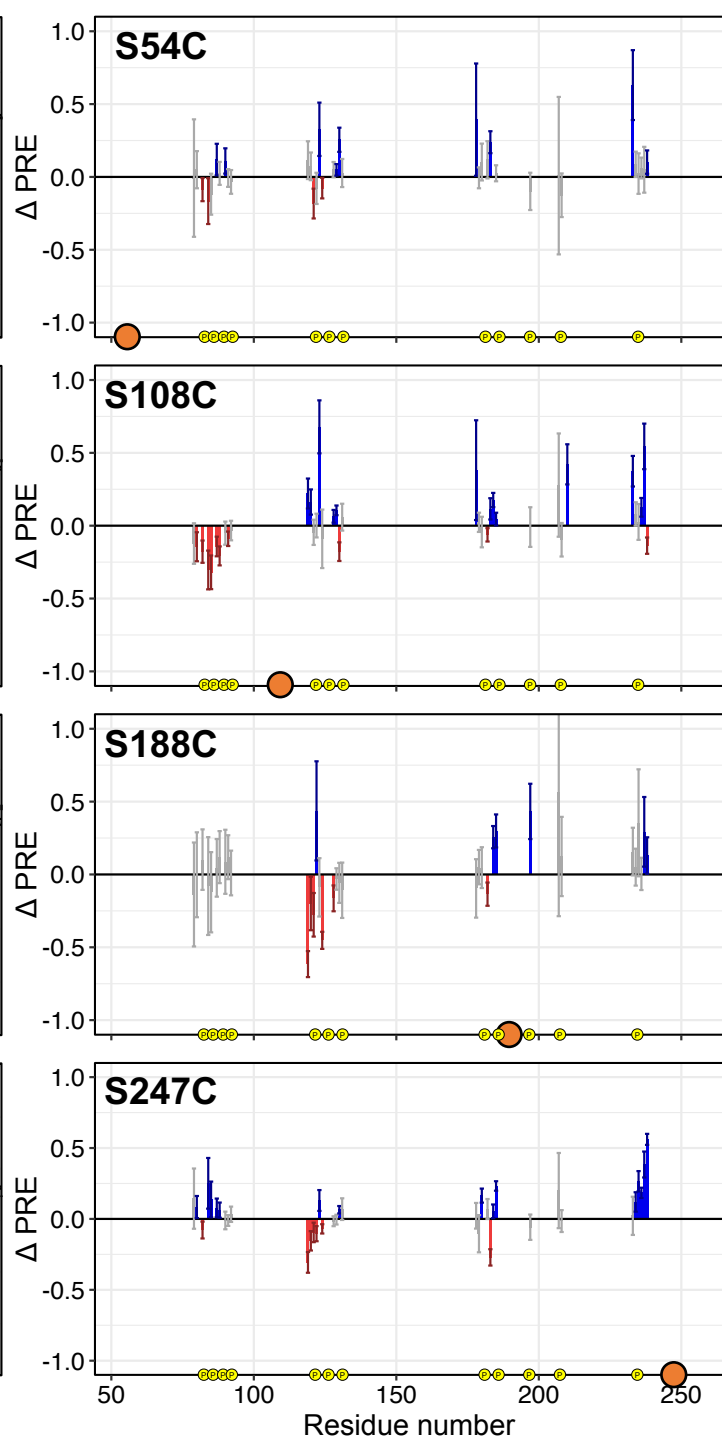
A**Major forms****Minor forms****B**

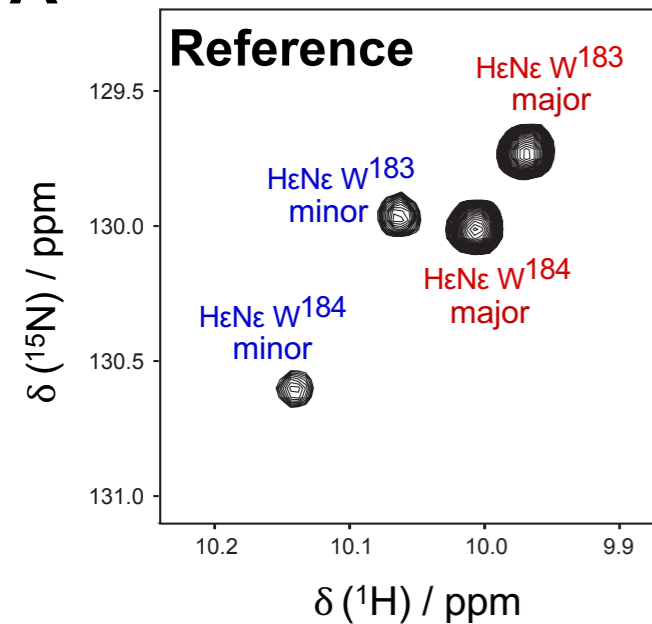
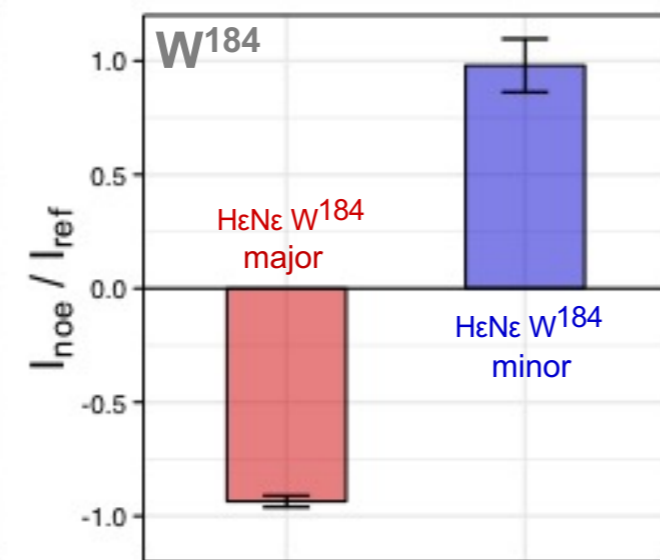
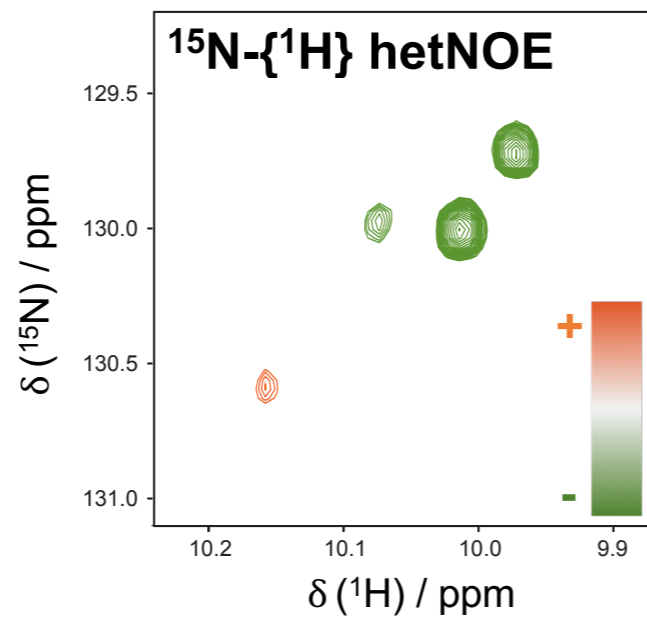
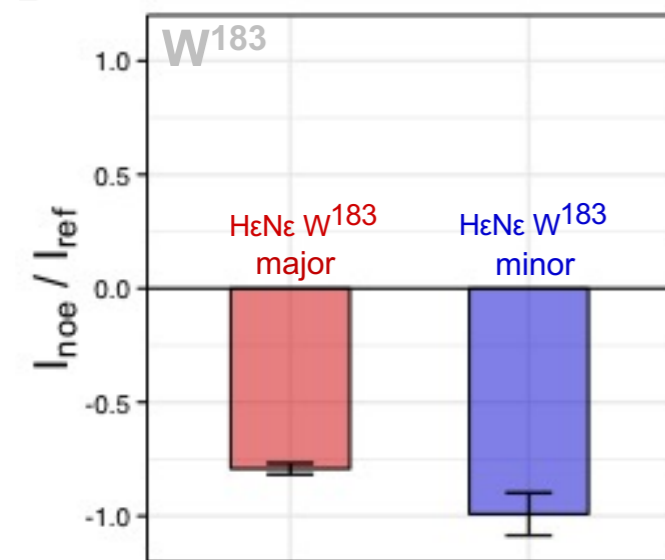
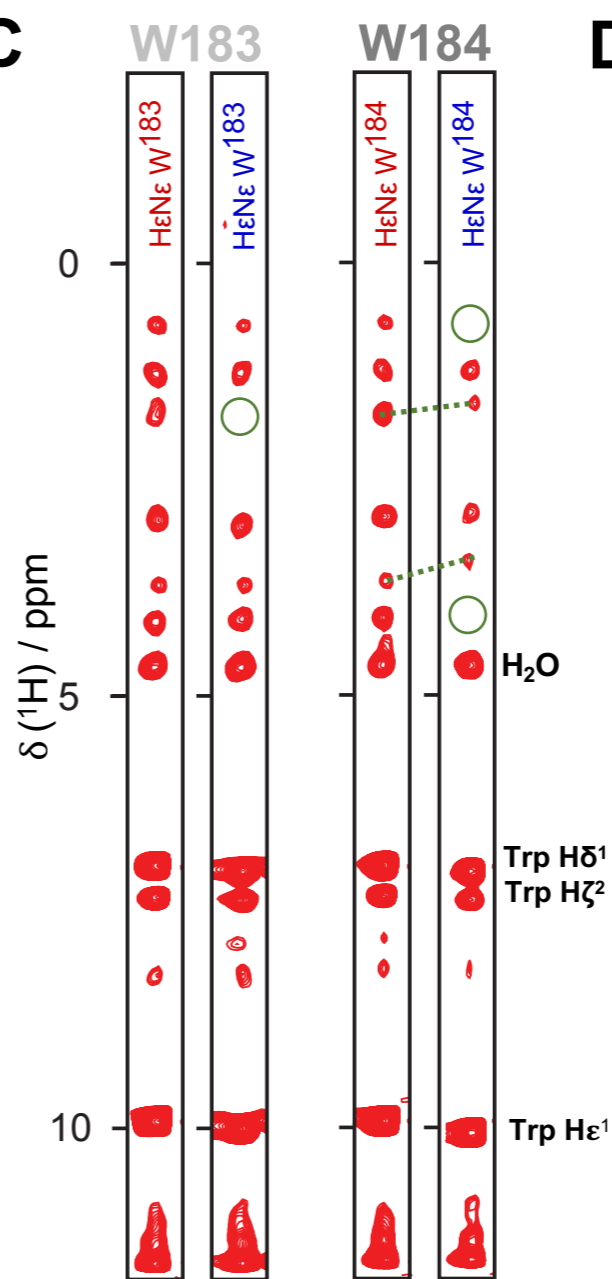
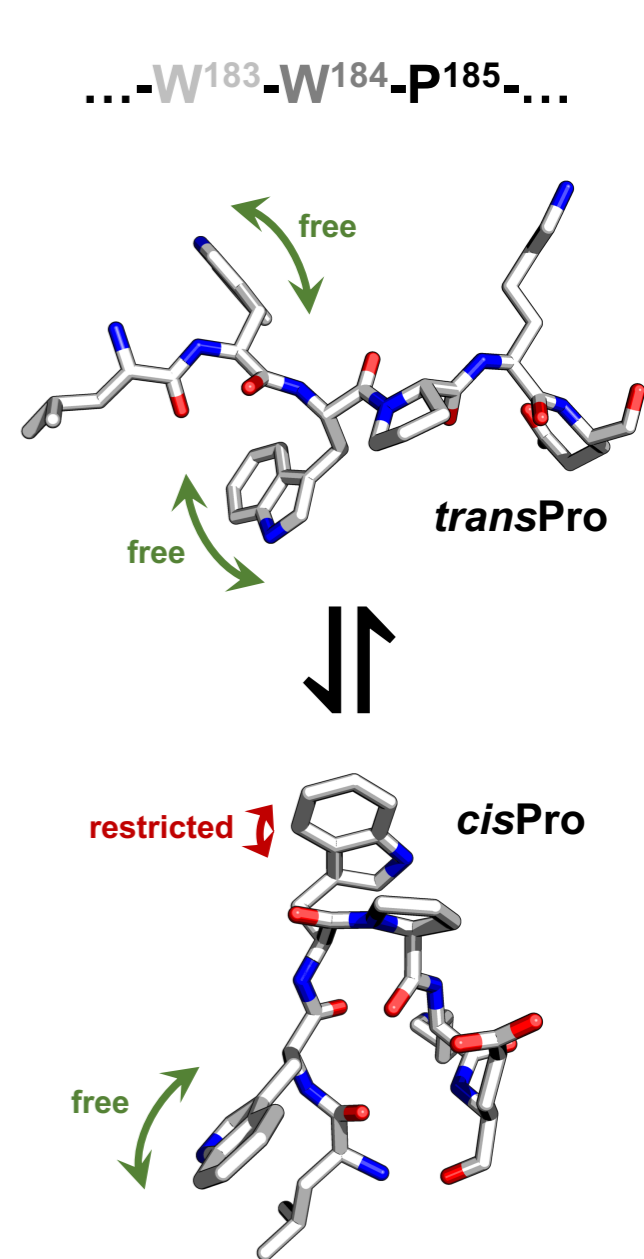


A**B**

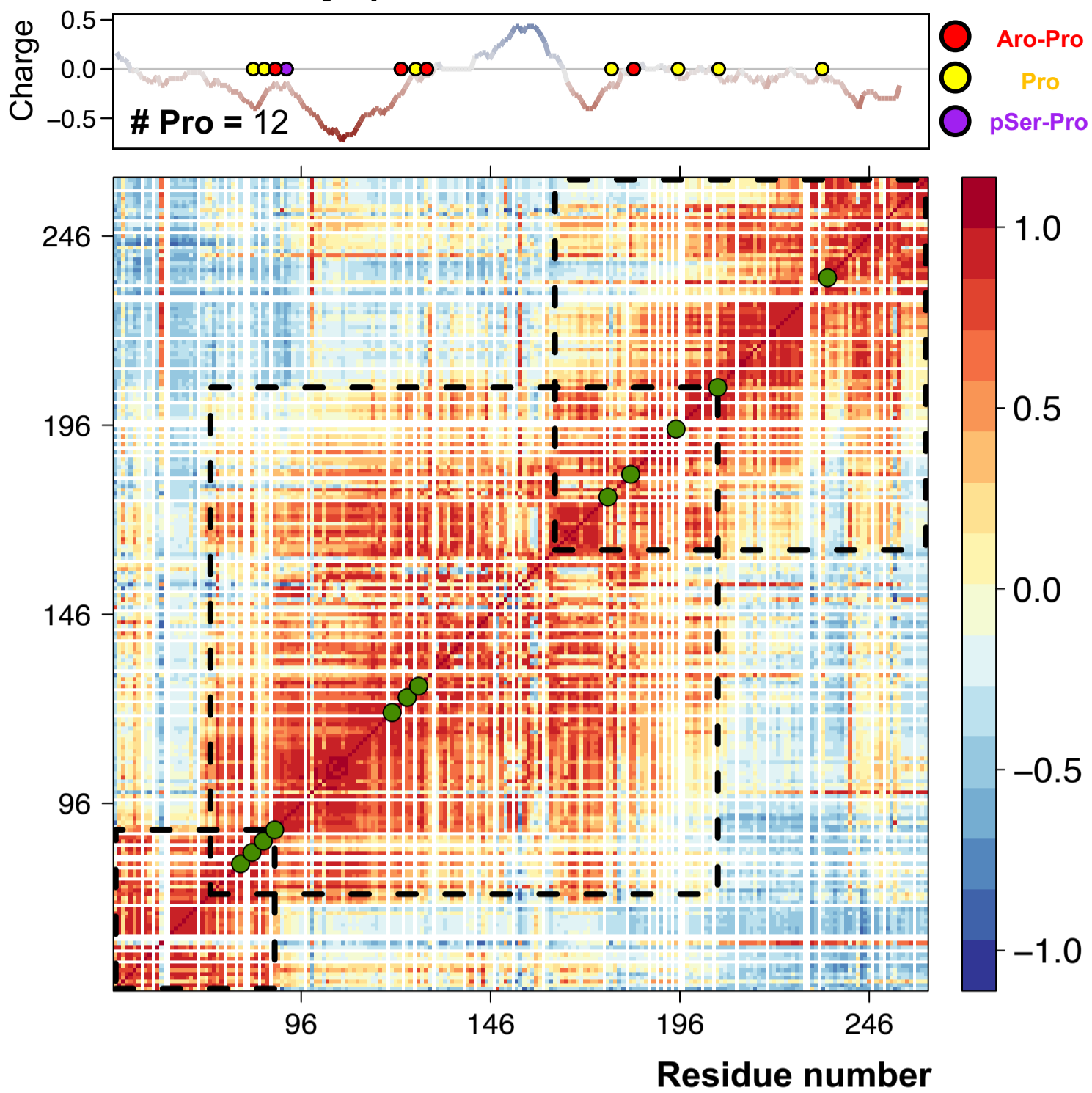




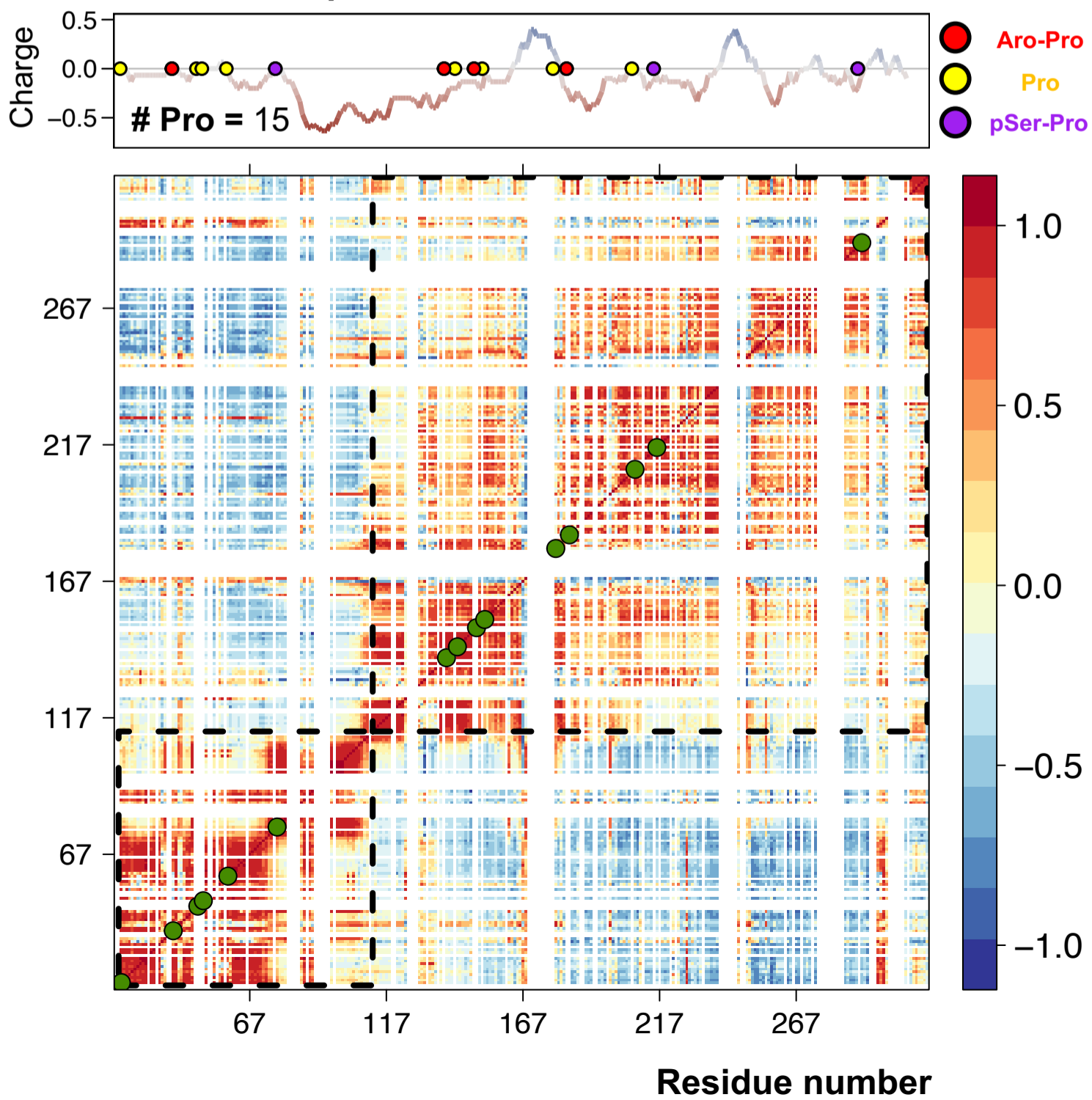
A**B**

A**B****C****D**

Coturnix japonica OPN



Homo sapiens OPN



Research highlights:

- ^{13}C -detection NMR enables the characterization of Pro residues in Osteopontin
- High *cis* populations of Xaa-Pro bonds are observed in proximity of aromatic residues
- ^{15}N relaxation and PRE enable the study of structural and dynamic properties of *cis/trans* isomers
- We discover that Pro, typically considered just a *disorder promoter*, has subtler roles in IDPs