Protein Characterization Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2022**, *61*, e202108361
International Edition: doi.org/10.1002/anie.202108361
German Edition: doi.org/10.1002/ange.202108361

Selective $^{1}H^{\alpha}$ NMR Methods Reveal Functionally Relevant Proline *cis/trans* Isomers in Intrinsically Disordered Proteins: Characterization of Minor Forms, Effects of Phosphorylation, and Occurrence in Proteome

*Fanni Sebák, Péter Ecsédi, Wolfgang Bermel, Burkhard Luy, László Nyitray, and Andrea Bodor**

Abstract: *It is important to identify proline cis/trans isomers that appear in several regulatory mechanisms of proteins, and to characterize minor species that are present due to the conformational heterogeneity in intrinsically disordered proteins (IDPs). To obtain residue level information on these mobile systems we introduce two $^{1}H^{\alpha}$-detected, proline selective, real-time homodecoupled NMR experiments and analyze the proline abundant transactivation domain of p53. The measurements are sensitive enough to identify minor conformers present in 4–15 % amounts; moreover, we show the consequences of CK2 phosphorylation on the cis/trans-proline equilibrium. Using our results and available literature data we perform a statistical analysis on how the amino acid type effects the cis/trans-proline distribution. The methods are applicable under physiological conditions, they can contribute to find key proline isomers in proteins, and statistical analysis results may help in amino acid sequence optimization for biotechnological purposes.*

Introduction

Intrinsically disordered proteins and protein regions (IDPs/IDRs) play vital roles in diverse biological processes.[1–2] These systems display an ensemble of conformations instead of a permanent secondary structure. The conformational flexibility enables them to bind to multiple targets and to become potential drug targets.[3–5] One factor contributing to the conformational heterogeneity is the *cis/trans* isomerization of proline (Pro) residues. The different *cis/trans*-Pro states are associated with tumorigenesis (in p53TAD[6] and ATR[7]), the formation of neurodegenerative diseases (in tau[8] and α-synuclein[9]) and in circadian rhythm regulation.[10] Previous studies suggested that the different isomeric forms can inhibit or support protein-protein interactions.[11–14] *Cis*-Pro residues are often localized in structurally or functionally important regions, and they regulate signaling pathways, drive membrane binding or even modulate protein folding and stability.[15] The Pro conformation gains further importance if the preceding Ser/Thr residue is phosphorylated by proline-directed kinases.[16,17] Several diseases—such as Alzheimer disease—were qualified as "cistauosis", as they would imply a *cis* conformation for the pThr/pSer-Pro prolyl bond.[8]

On the other hand, prolines are important in molecular recognition, mainly in intracellular signaling. Domains such as SH3 and WW bind to specific proline-containing interaction partners that are key parts of many signaling cascades.[18]

These results highlight that knowledge of the exact *trans* or *cis* proline conformation has high biological importance. IDPs are abundant in prolines[19] and for *N* prolines the number of possible conformations is 2^{N}, thus a multitude of isomers can co-exist in solution, and appropriate techniques have to be found for their detection.

The usual way to study *cis/trans*-Pro formation is by using proline analogs.[20] This means the system is shifted to extremes, and the targeted proline is fixed in either *cis* or *trans* form. A more useful way would be to determine the natural distribution of these isomers by an atomic level characterization. For these highly mobile systems X-ray crystallography and cryo-EM methods fail, leaving NMR spectroscopy to detect and discern the coexisting species.[21] Even in the NMR spectra only the most populated minor forms can be revealed, but each form should be analyzed accordingly to better understand their roles and functions.

[*] F. Sebák, Dr. A. Bodor
Eötvös Loránd University, Institute of Chemistry
Pázmány Péter s. 1/a, 1117 Budapest (Hungary)
E-mail: andrea.bodor@ttk.elte.hu
F. Sebák
Semmelweis University, Doctoral School of Pharmaceutical Sciences
Üllői út 26, 1085 Budapest (Hungary)
Dr. P. Ecsédi, Dr. L. Nyitray
Eötvös Loránd University, Department of Biochemistry
Pázmány Péter s. 1/c, 1117 Budapest (Hungary)
Dr. W. Bermel
Bruker BioSpin GmbH
Silberstreifen 4, 76287 Rheinstetten (Germany)
Dr. B. Luy
KIT-Institut für Organische Chemie, IBG4—Magnetische Resonanz
Fritz-Haber-Weg 6, 76131 Karlsruhe (Germany)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202108361.

The *cis/trans*-Pro isomerization occurs within a slow timescale (10^{-3}–10^{-2} s^{-1}),[22] and minor signals of varying intensities appear with peak multiplication for the Pro neighboring residues.[23–31] Detection and sequential connection of the minor X-Pro-Y and X-Pro-Pro regions represents a challenge. One possibility is to compare ^{1}H,^{15}N-HSQC spectra of a particular Pro containing sequence with the one corresponding to a mutant (Pro/Ala)[26] or to use site-specifically labeled proteins.[30] Using solely NMR techniques one can exploit the experiments based on the favorable signal dispersion in the ^{13}C-,[32–34] and ^{15}N-dimension.[35] Besides the many advantages of these techniques, sensitivity issues can be critical, especially if investigating minor species. As $^{1}H_{start}$ -$^{1}H_{detect}$ experiments are the most sensitive, this leaves an opportunity to exploit the ^{1}H-based detection schemes. $^{1}H^{N}$ detection suffers from signal loss issues at physiological conditions, and due to the lack of Pro amide proton a break occurs in the sequential connectivity. The 4D HNCOCANH[36] experiment enables connection of X-Pro-Y environments, but has low sensitivity, and it is inadequate for the study of minor species, and also not applicable for Pro-Pro motifs. Thus, $^{1}H^{\alpha}$ detected[37–39] approaches might offer the desired solution. To identify the particular Pro isomer the best indicator is the C^{β}-C^{γ} chemical shift difference (denoted $\Delta\beta\gamma$), where values ≈ 5 ppm, or ≈ 10 ppm reveal *trans*, or *cis* form, respectively.[40] To obtain the C^{β}, C^{γ} chemical shift values the classical $^{1}H^{N}$ detected 3D hCCCONH measurement can be used,[41] that has relatively low sensitivity and the sequential connectivity is sometimes hindered. Therefore, we aimed to develop proline-selective experiments based on the $^{1}H^{\alpha}$ detection approach with unambiguous determination of C^{β}, C^{γ} chemical shifts both for major and minor forms. This choice seems plausible, as we earlier showed that $^{1}H^{\alpha}$ detection combined with a real-time homodecoupling acquisition scheme has high sensitivity, and the proximity of the water signal does not disturb.[42,43]

To test the power of the new technique, we performed an in-depth characterization of the detectable Pro conformations in the transactivation domain of p53 ($p53TAD^{1-60}$). $p53TAD^{1-60}$ is a disordered, dynamic and multifunctional domain with several interaction sites,[44–47] with proline environments involved in *cis/trans*-Pro conformational regulation.[20] Numerous kinases recognize and phosphorylate pre-Pro sites[48] affecting the stability, activity and the interactions of p53.[44,49–51] These properties make $p53TAD^{1-60}$ a perfect subject for our work. Phosphorylation of Ser or Thr residues preceding a Pro are key events in signaling mechanisms, thus we investigated the modifications caused by the casein kinase 2 (CK2) enzyme and tested whether phosphorylation has an impact on the *cis/trans*-Pro equilibrium. Finally, the varying content of the minor conformers raises the question which parameters influence a higher/lower *cis*-Pro formation. Therefore, based on available experimental literature information, we aimed to find regularities about how the *cis*-Pro formation depends on the amino acid type.

Results and Discussion

Proline-selective $^{1}H^{\alpha}$-detected experiments for conformation determination. As minor proline peaks show an intensity of a few % compared to the major signals, their detection and conformation determination requires experimental setups of high sensitivity. In case of IDPs the condition of having minimal signal overlap should be also fulfilled. To unambiguously identify the proline *cis*- or *trans*- conformation, the C^{β},C^{γ} chemical shifts need to be determined (Figure 1 A). For this purpose, we designed a $^{1}H^{\alpha}$-detected, proline-selective triple-resonance experiment to accomplish a significant reduction in overlap by avoiding signals originating from other residues. As the chemically different ^{15}N spins of prolines resonate exclusively in the 130–145 ppm range, proline-selectivity can be best achieved in the ^{15}N-dimension. Thus, we introduce the Pro-(HCA)NCACBHA experiment (Figure 1 B), which is an extension of the HACAN sequence.[38,52] The transfer of magnetization has the following route:

$$\begin{aligned} &^{1}H^{\alpha}/^{1}H^{\delta} \xrightarrow{J_{HC}} {}^{13}C^{\alpha}/^{13}C^{\delta} \xrightarrow{J_{CN}} {}^{15}N_{Pro};\ t_1 \\ &\xrightarrow{J_{CC}J_{CN}} {}^{13}C^{\alpha}/^{13}C^{\delta},\ ^{13}C^{\beta}/^{13}C^{\gamma};\ t_2 \\ &\xrightarrow{J_{CC}} {}^{13}C^{\alpha}/^{13}C^{\delta} \xrightarrow{J_{H\alpha C\alpha}} {}^{1}H^{\alpha}/^{1}H^{\delta};\ t_3 \end{aligned} \quad (1)$$

where for each step the involved active couplings are indicated, and acquisition periods are denoted by t_i ($i = 1$–3). Starting from $^{1}H^{\alpha}$, transfer is achieved via band-selective pulses covering the $^{13}C^{\alpha}$ and the $^{15}N_{Pro}$ region, respectively. After t_1-evolution, magnetization is transferred back to $^{13}C^{\alpha}$ and further to $^{13}C^{\beta}$, where the t_2-evolution correlates the corresponding ^{13}C frequencies. Finally, back-transfer to $^{1}H^{\alpha}$ with homodecoupled acquisition using the BASEREX decoupling scheme[42–43] (Supporting Information, Figure S1) ensures sensitive detection with minimal signal widths. Since proline $^{1}H^{\delta}$, $^{13}C^{\delta}$, and $^{13}C^{\gamma}$ environments are also covered by the applied selective pulses, corresponding correlations of $^{15}N_{Pro}$ to $^{13}C^{\delta}/^{13}C^{\gamma}$ and $^{1}H^{\delta}$ are equally detected. The sequence allows selective detection of both proline C^{β} and C^{γ} chemical shifts, however, they will not be directly correlated. In order to achieve this direct correlation, a second optimized experiment has been developed. The Pro-(H)CBCGCAHA measurement focuses on high sensitivity achieved by maximizing transfer efficiencies and minimizing the number of transfer steps applied (Figure 1 C):

$$^{1}H^{\beta}/^{1}H^{\gamma} \xrightarrow{J_{HC}} {}^{13}C^{\beta}/^{13}C^{\gamma};\ t_1 \xrightarrow{J_{CC}} {}^{13}C^{\alpha};\ t_2 \xrightarrow{J_{H\alpha C\alpha}} {}^{1}H^{\alpha};\ t_3 \quad (2)$$

Starting from the ^{1}H side chain, selective transfer to $^{13}C^{\beta}$ and $13C^{\gamma}$ and a short constant time (CT) t_1-evolution ensures the possibility to detect the C^{β}, C^{γ} chemical shifts. Subsequently, highly optimized transfer from both carbons to C^{α} is applied, where the choice of Θ will result in different outcomes of correlations: choosing $\Theta = 45°$ pulse angle, the next transfer element equally transfers $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ magnetization to C^{α}. While setting $\Theta = 0°$ or $\Theta = 90°$ will result in 100 % transfer of only C^{β} or C^{γ} magnetization, respectively (Figure S2). Finally, highly resolved C^{α} evolution during t_2 and homodecoupled acquisition of H^{α}—similar to the SHACA-HSQC—will add up to a well-resolved 3D experiment,

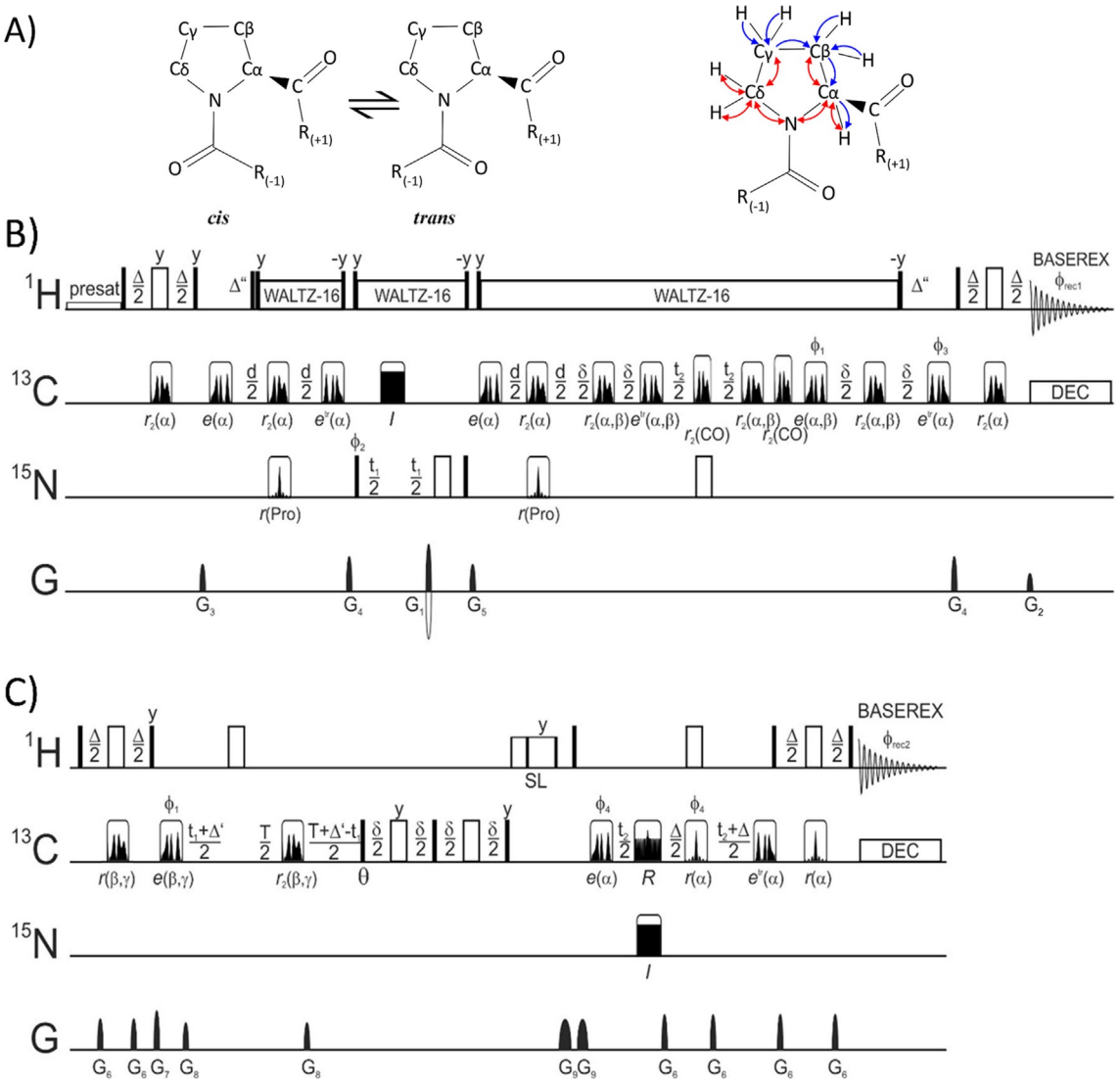


Figure 1. Selective experiments for optimal detection of the Pro conformation. A) Proline *cis* and *trans* conformers and the representation of the magnetization transfer pathway; B) Pro-(HCA)NCACBHA (red) and C) Pro-(H)CBCGCAHA (blue) pulse sequences. All pulses are applied along *x* unless indicated otherwise. Pulse phases are cycled according to $\varphi_1 = x, -x$; $\varphi_2 = x, x, -x-x$; $\varphi_3 = x, -x, x, -x, -x, x, -x, x$; $\varphi_4 = y, y, -y-y$; $\varphi_{rec1} = x, -x, -x, x, -x, x, x, -x$; $\varphi_{rec2} = x, -x, -x, x$. Filled thin bars represent 90° pulses, while open bars mark 180° hard pulses. Delays are $\Delta = 1/(2\,{}^1J_{CH})$, $\Delta' = 1/(4\,{}^1J_{CH})$, $\Delta'' = 1/(3\,{}^1J_{CH})$; $d = 1/(2\,{}^nJ_{CN})$; $\delta = 1/(2\,{}^1J_{CC})$. Shaped pulses are given with their amplitude patterns and a letter code. Uppercase letters indicate broadband pulses: (*I*) represents adiabatic Crp60(80) in (B), BIP[53] the inversion pulse in (C), (*R*) is the BURBOP refocusing pulse. Lowercase letters indicate selective pulses with selected nuclei in parentheses: (*e*) marks Q5-derived SEBOP excitation or 90° universal rotation pulses, (*r*) are the REBURP refocusing pulses, (r_2) are Q3-derived SURBOP refocusing pulses (for further details see the Supporting Information, Table S1).

suitable for studying larger IDPs with many prolines. In this experiment the selectivity of carbon resonances alone is not sufficient to fully exclude non-proline correlations, but proline signals are unambiguously identified, and the likelihood of overlap is dramatically reduced (Figure S3). On the other hand, long nitrogen transfer elements with corresponding signal losses are avoided (experimental details and pulse-programs are summarized in the Figure 1 legend and in the Supporting Information, Tables S1 and S2).

Both experiments result in a successful removal of the water peak, working with maximum 10% D_2O content. Regarding the choice of a measurement type, the sensitivity of our $^1H_{start}$,$^1H_{detect}$ experiment with fully decoupled signals will be 8 times higher than a $^1H_{start}$,$^{13}C_{detect}$ approach (based on $\gamma_{start}\cdot(\gamma_{detect})^{3/2}$), resulting in reduction of measurement time by a factor of approximately 64 for the same S/N. Considering the low intensity species, or having small amount of protein, clearly the $^1H^{\alpha}$-detected method is favored.

Proline isomers in p53TAD$^{1-60}$. We used the new pulse sequences to give a complete picture of the assignment, sequential connection, isomer-type determination and amount of minor form for all detectable signals in p53TAD$^{1-60}$ (Figure 2, top). Assignment of the major species was already available,[46] and for the characterization of minor resonances appearing in the ^1H,^{15}N-HSQC spectrum (Figure S4) other techniques were applied. Prolines were directly

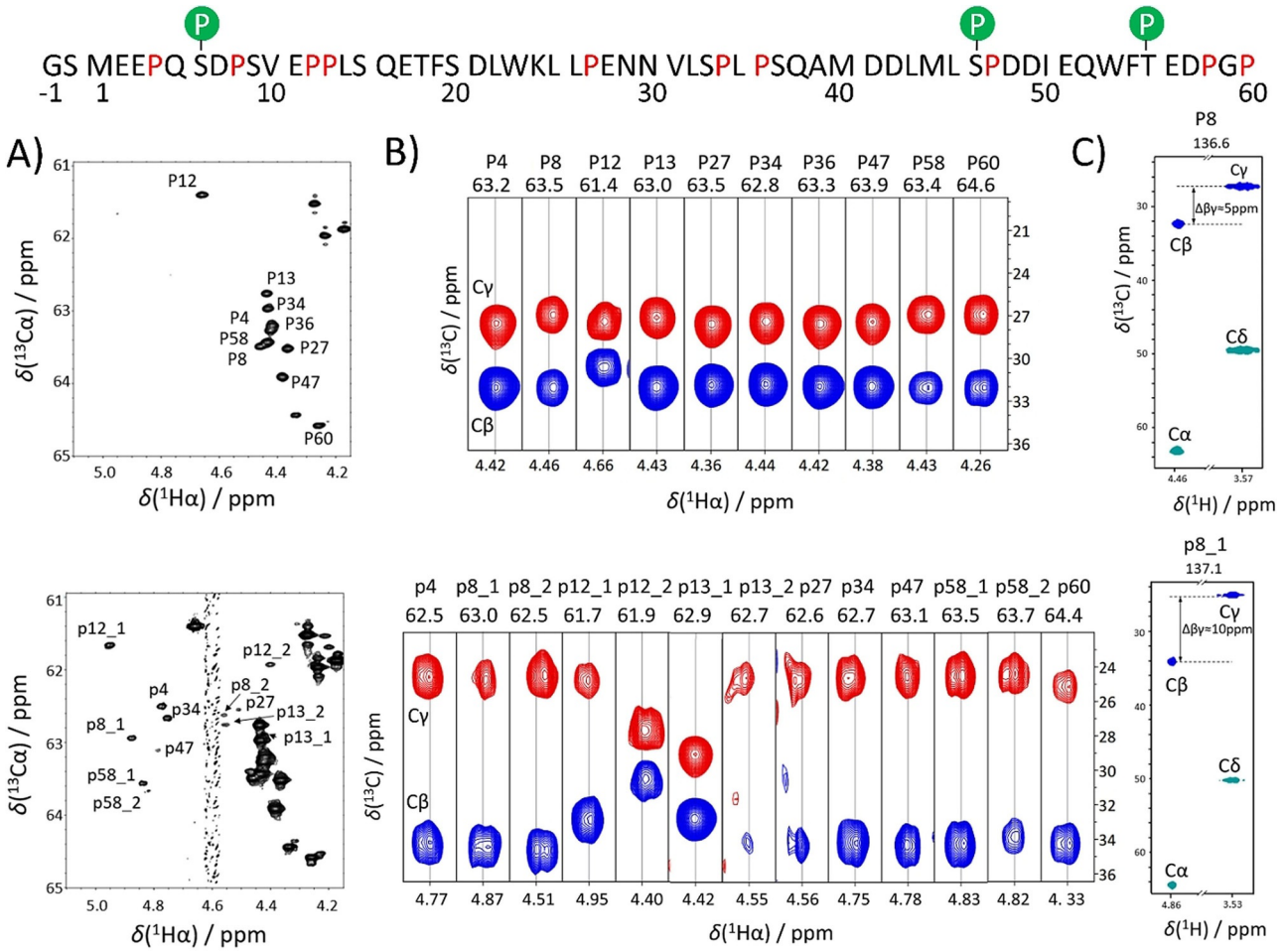


Figure 2. The amino acid sequence of p53TAD$^{1-60}$ (the first two residues are cloning tags; proline residues and the possible phosphorylation sites by CK2, predicted by Netphos 3.1[56] webserver are highlighted). Upper Figures show the results for the major prolines, while lower figures present the minor prolines. A) Assigned proline region of SHACA-HSQC (NS = 8, TD = 4096×1024 real points, d1 = 0.7 s). B) Resulting strips from the Pro-(H)CBCGCAHA measurement and C) a typical set of strips from the Pro-(HCA)NCACBHA experiment for selected environments. Minor peaks are labeled with lowercase letters and numbered 1,2 if more sets were detected.

detected via the previously introduced SHACA-HSQC measurement (Figure 2 A), and the X-Pro-Y, and X-Pro-Pro sequential connections were achieved from 3D HCAN, 3D HCACON and 3D HCANCO experiments (Figure S4). Assignment of minor resonances (Figure S4) proved that the effect of the conformation change vanishes at 2–4 residue distance from the proline, in accordance with previous literature findings.[23–27] Line widths on SHACA-HSQC for both major (*M*) and minor (*m*) forms are similar, meaning that there is no difference in the relaxation properties, also claimed in previous literature studies.[27] Thus, for quantitative evaluation of the minor content, peak integrals (Table S3) obtained from the corresponding extracted 1D ^{1}H rows of the SHACA-HSQC were used. The integrals of minor non-proline residues from the ^{1}H,^{15}N-HSQC were calculated using a similar approach. For a given proline-containing segment integrals of the consecutive residues do not show significant difference. The minor population was calculated using the $[m]/([m]+[M])$ equation (Table 1). Finally, the newly introduced pulse sequences allowed determination of the Δβγ chemical shift differences for all species and resulted in the unambiguous assessment of the proline isomer (Table 1). All prolines in the major conformer have C^{β} chemical shifts between 31–32 ppm (Pro12 lower as it is succeeded by Pro13), and C^{γ} values are around 27 ppm, indicating the expected *trans* isomer. In case of most minor peaks shifts in C^{β} to ≈ 34 ppm and C^{γ} to ≈ 24 ppm are observed, in agreement with a *cis* isomer (Figure 2). Regions containing multiple prolines show more than one set of minor peaks. Thus, 3 separate sets of minor peaks appear in the Pro8-Pro13 region: one belonging to Ser9-Val10-Glu11 and two sets of Glu11-Pro12-Pro13 fragments (Figure S4). For this the sequential connection is: p12 1^{cis}—p13 1^{trans} and p12 2^{trans}—p13 2^{cis}. Glu11 minors have similar signal intensity indicating a similar *cis* propensity for Pro8, Pro12 and Pro13 (Table S3). Two minor peaks were assigned for Gly59, due to Pro58 and Pro60 residues.

Secondary chemical shift values were calculated using C^{α}, C^{β} data[54–55] (Figure S4). Compared to the major, extended form, with all-*trans* prolines, in the *cis*-Pro containing minor forms the turn tendency is accentuated and a perturbance is

Table 1: Proline isomers in p53TAD$^{1-60}$ for major and minor conformers: Δβγ chemical shift differences and proline isomer form.

Pro	Major		Minor		
	Δβγ [ppm]	Pro conformation	Δβγ [ppm]	Pro conformation	Minor [%]
P4	4.7	*trans*	9.6	*cis*	11.5
P8	5.0	*trans*	9.4	*cis*	9.0
			9.7	*cis*	10.0
P12	3.1	*trans*	8.2	*cis*	13.0
			2.9	*trans*	8.0
P13	4.7	*trans*	4.6	*trans*	[a]
			9.7	*cis*	[a]
P27	4.6	*trans*	10	*cis*	5.5
P34	5	*trans*	9.5	*cis*	12.5
P36	4.7	*trans*		n.d.	
P47	4.7	*trans*	9.6	*cis*	4.5
P58	5.3	*trans*	9.6	*cis*	7.0
			9.5	*cis*	6.0
P60	5.1	*trans*	9.1	*cis*	34.5[b]

[a] Overlap with the water signal. [b] Pro60 is the last, C-terminal highly mobile residue. Not detected (n.d.).

observed in the nascent helical propensity of the Phe19-Leu25 region.

The performance of the introduced measurements is very good: using samples with 7%-10% D_2O, the residual water peak is minimal and an unambiguous determination of the Δβγ chemical shift difference is obtained. Good quality, high resolution Pro-(H)CBCGCAHA spectra ($4096 \times 512 \times 16$ real points) with minor peak detection is achieved within 15 h; while the corresponding 2D SHACA-HSQC takes ≈0.5 h (Table S2). Using non-uniform sampling, measurement time can be decreased by a factor 2, at the cost of slightly reduced sensitivity. Evaluation of the concentration shows—based on data in Table 1—that the amount of minor species varies between 4.5%-13.5%, meaning that conformers present in 50–150 μM concentration (protein concentration was ≈1 mM) are easily detectable. Figure S3 highlights that for other residue types which also have chemical shifts in this range, mostly one, typically low intensity signal—sometimes distorted or with reversed sign—is detected, indicating the ease with which proline residues can be selectively identified.

Phosphorylation by CK2 and its effect on *cis/trans*-Pro equilibrium. It is still little explored how post-translational modifications (PTMs), such as phosphorylation, modulate the existing structural ensembles of an IDP, including the major and minor forms, as well as the *cis/trans* proline isomers. To test if our method can analyze these changes, we studied the effect of CK2 phosphorylation on p53TAD$^{1-60}$. Ser/Thr sites are phosphorylated by CK2 if these residues are succeeded by one or more negative charge(s),[57] while amino acids with hydrophobic sidechains, and prolines can negatively affect the phosphorylation. It is known that CK2 can phosphorylate Ser386 at the p53 C-terminus,[58] but its effect on the N-terminal domain has not been studied comprehensively. Sun et al. reported that Ser46 and Thr55 in TAD can be phosphorylated *in vitro*.[59] Analyzing the CK2 consensus sequences and the residues around phosphorylation sites (Figure 2) Ser46 and Thr55 are succeeded by two negatively charged residues (Asp48-Asp49 and Glu56-Asp57, respectively), and they can be primary phosphorylation sites. Ser6 and Ser20 are followed by one negatively charged residue (Asp in both cases), and they can represent further possible phosphorylation sites. As Ser46 is positioned before a proline residue (Pro47), and Pro4, Pro8 and Pro58 are also situated close to potential phosphorylation sites, the multiple phosphorylation at Ser6, Ser46 and Thr55 would give us opportunity to investigate the effect on *cis/trans*-Pro equilibrium.

We tested this experimentally by monitoring the phosphorylation of p53TAD$^{1-60}$ with CK2 by ^{1}H,^{15}N-HSQC spectra at different time points (12 h and 36 h, respectively). After 36 hours incubation phosphorylation is complete for Thr55, Ser46 is 85% phosphorylated, while phosphorylation levels for Ser6 and Ser20 are ≈50% and <10%, respectively (Figure S5). These results are in accordance with the prediction. The lower phosphorylation level of Ser46 can be explained with the presence of Pro47 that limits the activity of CK2.[57] Analyzing the chemical shift changes, these occur only in the $i \pm 3$ region around the phosphorylated sites, detected both in ^{1}H,^{15}N-HSQC and SHACA-HSQC spectra. This indicates that the overall structural propensities are not affected (Figure S5).

Our *in vitro* results also show that Ser6 and Thr55 phosphorylation has no influence on the Pro environments in their vicinity (Pro4, Pro8 and Pro58), and phosphorylation levels are the same for both major and minor peaks. Instead, phosphorylation of Ser46 eliminates the *cis*-Pro47 isomer. The p47 minor peak is no longer detectable neither on the SHACA-HSQC (with a defined detection threshold below 0.9% as discussed in the Figure S4D), nor in the proline selective spectra; also the neighboring minor peaks are missing from the ^{1}H,^{15}N-HSQC spectrum. This proves that phosphorylation can change the *cis/trans* equilibrium of prolines and also that this effect might be present only in the neighboring prolines close to the modification (Figure 3; Figure S5).

Our results indicate that CK2 might be part of the kinases modifying p53 function and its interaction network. Thr55 is constitutively phosphorylated in unstressed cells,[60,61] while phosphorylation of Ser6, Ser20 and Ser46 is known to regulate p53TAD interactions.[60] Moreover, the observed effect of Ser46 phosphorylation on the *cis/trans* equilibrium of Pro47 can be connected to the proline isomer dependent p53-protein interactions. It is known that p53 binds to the apoptosis regulator Bcl-2-associated X (BAX) protein if Pro47 is in *cis* form. Consequently, in the bound state a *trans*-Pro47 rearrangement happens, that induces a conformational change in BAX, activating it.[20] This takes place either spontaneously by chance (not effective) or by contribution from peptidyl-prolyl *cis-trans* isomerase NIMA-interacting 1 (Pin-1).[20] Our *in vitro* experiments show that a ≈5% *cis*-Pro47 form exists under natural distribution, which can bind to BAX. However, if due to Ser46 phosphorylation the equilibrium is shifted towards *trans*-Pro47, then, in the absence of Pin-1, the free *cis*-Pro47 pool is decreased, and CK2 phosphorylation will keep the interaction level between p53 and BAX low. If phosphorylation happens after binding to BAX, then Pin-1, which requires pSer/pThr for recognition[62–63] can change *cis*-Pro47 to *trans*-Pro47 and activate

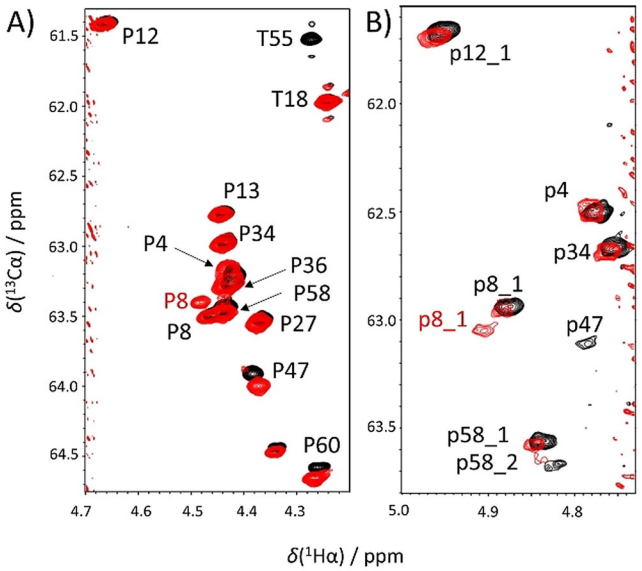


Figure 3. The effect of phosphorylation on p53TAD$^{1-60}$. Overlayed SHACA-HSQC spectra highlighting the proline region for the A) major and B) minor peaks, in native (black) and CK2 multiple phosphorylated form (red). The contour levels were lowered for the minor form (B).

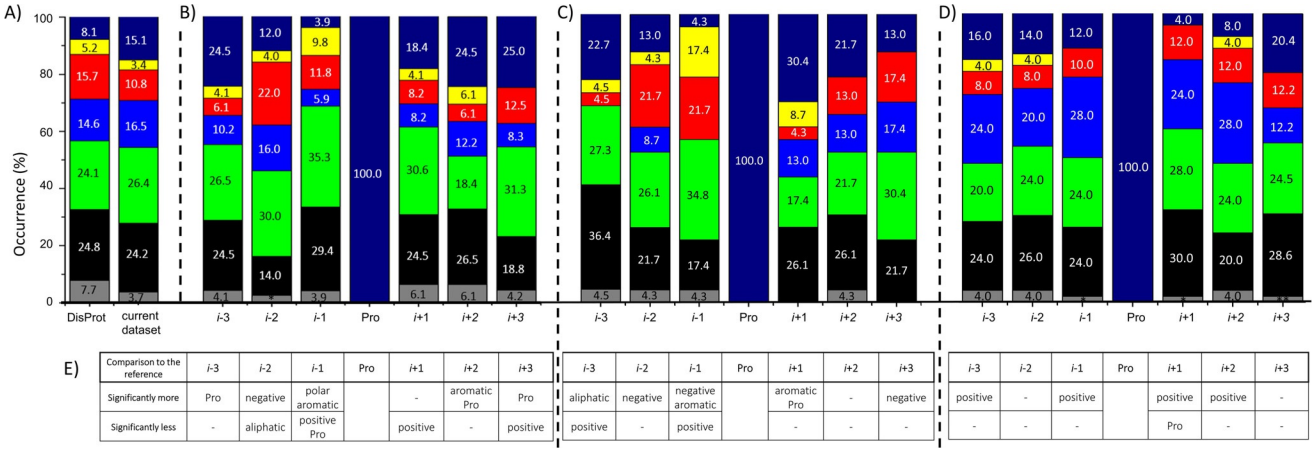
BAX.[20] Still, the *in vivo* effects of CK2 on p53 remain to be determined.

Sequence-dependent proline conformation distribution. Literature studies conducted on various proteins indicate that the amount of minor species varies typically between 5–15 %, and our data from p53TAD$^{1-60}$ also confirm these findings.[23–31] However, discrepancies may occur from the various experimental conditions and measurement approaches regarding how the minor proline contents were obtained (measurement type, that is, indirect evaluation from ^{1}H,^{15}N-HSQC spectra, direct evaluation using ^{13}C detection or in our case SHACA-HSQC; experimental parameters, that is, relaxation delay values; or temperature) (Table S4). As shown for p53TAD$^{1-60}$, the minor peak amount determined from the ^{1}H,^{15}N-HSQC and SHACA-HSQC led to similar results (Table S3). To test the influence of temperature, we monitored the minor peak content in the 278–323 K range by ^{1}H, ^{15}N-HSQC and SHACA-HSQC measurements. A variation of ~1 % could be detected, indicating that the influence of temperature is not significant, and this result is in agreement with the findings of Ahuja et al.[25] These observations suggest that reported data can be used for further analysis.

Cis-Pro content is influenced by the type of the surrounding amino acids. In order to find which residues have the most impact, we performed a statistical analysis taking into account the ±3 residues neighborhood, as appearance of the *cis*-Pro isomer causes peak multiplication for the neighbors at 2–4 residues distance. Thus, based on available literature information (Table S4), we collected 101 (N^{tot}) proline $i\pm3$ segments, where i is Pro, and this set contains 595 residues (n^{tot}). We determined the distribution of the residue types in this dataset—using the classification in 7 groups—and made a comparison to the distribution (occurrence %) found in the DisProt database (Figure 4 A).[19,64] Two-sided binomial tests at a significance level of 0.1 were conducted, considering the n^{tot} residues; and the expected occurrence range of values were calculated for each residue type (Table 2; Table S5).

Further on, we investigated the residue type distribution in each of the $i\pm3$ positions (Figure 4 B–D). Based on binomial tests, we determined which residue type occurrence shows a significant deviation from its expected value. In this respect, to obtain reliable information, we used the occurrence (%) from our dataset as a comparison, instead of DisProt. Based on the reported *cis*-Pro content we considered three cases: i) $N=51$ segments (from N^{tot}) fulfilling the >5 % condition, out of which ii) $N=23$ environments with >10 % and iii) $N=50$ regions with <5 % (Figure 4 B–D, columns i-3 to $i+3$). The statistical analysis enables us to formulate some general rules:

(i) An increased >10 % *cis*-Pro isomer may be found if positions i-1 or $i+1$ are aromatic residues. Indeed, it has been shown that aromatic residues in proline neighborhood can be involved in π·CH$_2$ stacking interactions,[10,27]



E)

Comparison to the reference	i-3	i-2	i-1	Pro	i+1	i+2	i+3
Significantly more	Pro	negative	polar aromatic		-	aromatic Pro	Pro
Significantly less	-	aliphatic	positive Pro		positive	-	positive

	i-3	i-2	i-1	Pro	i+1	i+2	i+3
	aliphatic	negative	negative aromatic		aromatic Pro	-	negative
	positive	-	positive		-	-	-

	i-3	i-2	i-1	Pro	i+1	i+2	i+3
	positive	-	positive		positive	positive	-
	-	-	-		Pro	-	-

Figure 4. A) Amino acid type occurrences (%) in DisProt and in our dataset. The determined occurrence (%) in positions $i\pm3$ for *cis*-Pro content in our dataset: B) >5 % ($N=50$), C) >10 % ($N=23$), and D) <5 % ($N=51$). Amino acids were classified in seven groups based on sidechain type: Gly (gray), aliphatic (black), polar (green), positively charged (blue), negatively charged (red), aromatic (yellow), and Pro (dark blue). E) for each of the cases (B–D) the result of the binomial analysis is shown at a 0.1 significance level. *3 %; **2 %.

Table 2: Results of our statistical analysis showing the differences in residue types between DisProt and our selected data set (Pro ± 3 residues). For each residue type the expected value range was determined. The occurrence (n) out of n^{tot} is compared to occurrence (%) in DisProt.

Residue type	Occurrence [%] in DisProt	Total selected residues ($n^{tot} = 595$)		
		Expected value range[a]	Occurrence [n]	Compared to DisProt
Gly	7.7	38–54	22	significantly less
Aliphatic	24.8	134–161	144	no difference
Polar	24.1	130–157	157	no difference
Positive	14.6	76–98	98	no difference
Negative	15.7	82–105	64	significantly less
Aromatic	5.2	24–38	20	significantly less
Pro	8.1	40–57	90	significantly more

[a] For each residue type, the occurrence (%) in DisProt was taken as reference.

causing rigidity to the segment. In all other positions aromatic residues have no significant effect. For a *cis*-Pro content $< 5\,\%$, no aromatic residues were found in positions $i \pm 1$.

(ii) For a $> 10\,\%$ *cis*-Pro content residues with negatively charged side chains are more frequent in positions *i*-2, *i*-1 and $i + 3$.

(iii) The *cis*-Pro content is increased if there is a Pro in $i + 1$ position. This means that *cis*-Pro-*trans*-Pro motifs may be more frequent in IDPs than *trans*-Pro-*cis*-Pro.

(iv) Residues possessing positively charged side chain in *i*-3 and *i*-1 positions can be an indicator of a decreased *cis*-Pro content. The electrostatic interaction between the positively charged side chain and the proline carbonyl group in *trans* isomer is energetically favored. Data also show that Arg and Lys residues are more common for $< 5\,\%$ *cis*-Pro content.

(v) Another observation that does not necessarily hold for each case is if position *i*-3 is occupied by an aliphatic residue with a large space filling sidechain (Val, Ile, Leu), then the likelihood to have a *cis*-Pro conformer is rather high.

To validate these criteria, we examined 9 out of 10 proline neighboring sequences of p53TAD$^{1-60}$. Since there are no aromatic residues in the analyzed segments, this is a good example to interpret the effect of other side chain types, which were not discussed before in the literature. The segments containing Pro4, Pro8, Pro12, Pro34, Pro58 belong to the $> 10\,\%$ class. Inspection of these segments shows that positions *i*-2 and *i*-1 are mostly occupied by residues bearing negative charge—in accordance with detected regularity (ii). Segments harboring Pro27, Pro47 show low, or no (Pro36) *cis*-Pro content (Table 1). We have to note, that besides the amino acid composition, the low *cis*-Pro content may also be the result of some secondary structural tendencies, as both Pro27 and Pro47 mark the end of regions with an α-helical propensity.

From a general point of view, phosphorylation means substitution of a polar residue to a negatively charged one. Similarly, phosphomimetic mutations of Ser to Glu/Asp also result in an increase of negative charge. According to our findings, such mutations can be advantageous for the formation of an increased *cis*-Pro amount. Our dataset analysis reveals that in the $> 5\,\%$ *cis*-Pro containing Ser/Thr-Pro sequences negatively charged residues are rare, therefore the mutation might alter the *cis/trans*-Pro ratio.[65] In p53TAD$^{1-60}$ the 2 Asp and 1 Leu neighbors of Pro47 can cause steric hindrance, leading to an unfavorable pSer46-*cis*-Pro47 arrangement.

Conclusion

In the present work we introduce a technique that enables the characterization of proline isomers present even in low concentration. The experiments allow unprecedented combined sensitivity and resolution in determining proline conformations at physiological pH and temperature, thereby passing a threshold that allows direct observation and quantification of minor species based on the detection of non-exchangeable $^{1}H^{\alpha}$ protons. Comparing to the less sensitive ^{13}C-detected measurements, our $^{1}H^{\alpha}$-detection based approach results in shorter experimental time, important especially if minor species are followed, or if for some proteins prolonged measurements are not possible. The method offers the possibility to follow the changes in the *cis/trans* proline composition via PTMs or protein interactions. There is no need for biological mutation/analogue incorporation, thus lowering the experimental costs and project elaboration time.

Based on all these aspects our new approach can help future research projects targeted to better understand the regulatory effects and mechanisms of *cis/trans* prolines in proteins, as we present for the case study of the p53TAD$^{1-60}$. For this protein we report and characterize 13 minor proline conformers of 10 proline residues. We highlight that phosphorylation by CK2 influences the *cis/trans*-Pro equilibrium, which might alter the function of p53.

The method is applicable for the study of larger IDPs/IDRs and continuous accumulation of new experimental data in this field will allow a more and more profound statistical analysis. This could at one point result in a reliable prediction of Pro *cis/trans* isomer distribution of any protein sequence and, based on the observed regularities, amino acid sequences with a certain *cis/trans*-Pro content can be designed.

Acknowledgements

The ELTE Thematic Excellence Programme 2020 supported by National Research, Development and Innovation Office (TKP2020-IKA-05 and Project no. 2018-1.2.1-NKP-2018-00005); NKFI Grants K124900, K137940 to A.B., K119359 to L.N.; Szint + Programme; EFOP-3.6.3-VEKOP-16–2017-00009 grant to F.S. and the Foundation for Hungarian Peptide and Protein Research are acknowledged. B.L. thanks the HGF program Information (43.35.02) and Virtmat for financial support.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: IDP datasets · NMR $^{1}H^{\alpha}$-detection · p53 · phosphorylation · proline *cis/trans* isomerization

[1] P. Tompa, *Trends Biochem. Sci.* **2002**, *27*, 527 – 533.
[2] V. N. Uversky, *Intrinsically Disordered Proteins*, 1 ed., Springer, **2014**.
[3] S. Ambadipudi, M. Zweckstetter, *Expert Opin. Drug Discovery* **2016**, *11*, 65 – 77.
[4] J. L. Neira, J. Bintz, M. Arruebo, B. Rizzuti, T. Bonacci, S. Vega, A. Lanas, A. Velázquez-Campoy, J. L. Iovanna, O. Abián, *Sci. Rep.* **2017**, *7*, 39732.
[5] L. I. Iconaru, D. Ban, K. Bharatham, A. Ramanathan, W. Zhang, A. A. Shelat, J. Zuo, R. W. Kriwacki, *Sci. Rep.* **2015**, *5*, 15686.
[6] X. X. Li, P. Dumont, A. Della Pietra, C. Shetler, M. E. Murphy, *J. Biol. Chem.* **2005**, *280*, 24245 – 24251.
[7] Y. Makinwa, P. R. Musich, Y. Zou, *Front. Cell. Dev. Biol.* **2020**, *8*, 281.
[8] K. Nakamura, A. Greenwood, L. Binder, E. H. Bigio, S. Denial, L. Nicholson, X. Z. Zhou, K. P. Lu, *Cell* **2012**, *149*, 232 – 244.
[9] J. Meuvis, M. Gerard, L. Desender, V. Baekelandt, Y. Engelborghs, *Biochemistry* **2010**, *49*, 9345 – 9352.
[10] C. L. Gustafson, N. C. Parsley, H. Asimgil, H. W. Lee, C. Ahlbach, A. K. Michael, H. Xu, O. L. Williams, T. L. Davis, A. C. Liu, C. L. Partch, *Mol. Cell* **2017**, *66*, 447 – 457.
[11] A. Brichkina, N. T. Nguyen, R. Baskar, S. Wee, J. Gunaratne, R. C. Robinson, D. V. Bulavin, *Cell Death Differ.* **2016**, *23*, 1592 – 1601.
[12] P. Sarkar, C. Reichman, T. Saleh, R. B. Birge, C. G. Kalodimos, *Mol. Cell* **2007**, *25*, 413 – 426.
[13] J. M. C. Teixeira, A. Guasch, A. Bicer, A. Aranguren-Ibanez, S. Chashmniam, J. C. Paniagua, M. Perez-Riba, I. Fita, M. Pons, *FEBS J.* **2019**, *286*, 1230 – 1239.
[14] K. P. Lu, G. Finn, T. H. Lee, L. K. Nicholson, *Nat. Chem. Biol.* **2007**, *3*, 619 – 629.
[15] P. Craveur, A. P. Joseph, P. Poulain, A. G. de Brevern, J. Rebehmed, *Amino Acids* **2013**, *45*, 279 – 289.
[16] M. Weiwad, A. Werner, P. Rucknagel, A. Schierhorn, G. Kullertz, G. Fischer, *J. Mol. Biol.* **2004**, *339*, 635 – 646.
[17] B. Chaves-Arquero, J. M. Pérez-Cañadillas, M. A. Jiménez, *Chemistry* **2020**, *26*, 5970 – 5981.
[18] M. J. Macias, S. Wiesner, M. Sudol, *FEBS Lett.* **2002**, *513*, 30 – 37.
[19] F. X. Theillet, L. Kalmar, P. Tompa, K. H. Han, P. Selenko, A. K. Dunker, G. W. Daughdrill, V. N. Uversky, *Intrinsically Disord Proteins* **2013**, *1*, e24360.
[20] A. V. Follis, F. Llambi, P. Merritt, J. E. Chipuk, D. R. Green, R. W. Kriwacki, *Mol. Cell* **2015**, *59*, 677 – 684.
[21] A. Bodor, *Amino Acids, Peptides and Proteins: 44*, The Royal Society of Chemistry, **2021**, 64 – 114.
[22] U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski, G. Fischer, *J. Mol. Biol.* **1998**, *279*, 449 – 460.
[23] T. R. Alderson, J. H. Lee, C. Charlier, J. F. Ying, A. Bax, *ChemBioChem* **2018**, *19*, 37 – 42.
[24] G. Gógl, B. Biri-Kovacs, A. L. Poti, H. Vadaszi, B. Szeder, A. Bodor, G. Schlosser, A. Acs, L. Turiak, L. Buday, A. Alexa, L. Nyitray, A. Remenyi, *FEBS J.* **2018**, *285*, 46 – 71.
[25] P. Ahuja, F. X. Cantrelle, I. Huvent, X. Hanoulle, J. Lopez, C. Smet, J. M. Wieruszeski, I. Landrieu, G. Lippens, *J. Mol. Biol.* **2016**, *428*, 79 – 91.
[26] M. Dujardin, V. Madan, N. S. Gandhi, F. X. Cantrelle, H. Launay, I. Huvent, R. Bartenschlager, G. Lippens, X. Hanoulle, *J. Biol. Chem.* **2019**, *294*, 13171 – 13185.
[27] B. Mateos, C. Conrad-Billroth, M. Schiavina, A. Beier, G. Kontaxis, R. Konrat, I. C. Felli, R. Pierattelli, *J. Mol. Biol.* **2020**, *432*, 3093 – 3111.
[28] B. Chaves-Arquero, D. Pantoja-Uceda, A. Roque, I. Ponte, P. Suau, M. A. Jiménez, *J. Biomol. NMR* **2018**, *72*, 139 – 148.
[29] O. Aitio, M. Hellman, B. Skehan, T. Kesti, J. M. Leong, K. Saksela, P. Permi, *Structure* **2012**, *20*, 1692 – 1703.
[30] A. Urbanek, M. Popovic, C. A. Elena-Real, A. Morato, A. Estana, A. Fournet, F. Allemand, A. M. Gil, C. Cativiela, J. Cortes, A. I. Jimenez, N. Sibille, P. Bernado, *J. Am. Chem. Soc.* **2020**, *142*, 7976 – 7986.
[31] P. Ludzia, B. Akiyoshi, C. Redfield, *Biomol. NMR Assignments* **2020**, *14*, 309 – 315.
[32] W. Bermel, I. Bertini, I. C. Felli, M. Piccioli, R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25 – 45.
[33] W. Bermel, I. Bertini, I. C. Felli, R. Kummerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56 – 64.
[34] I. C. Felli, W. Bermel, R. Pierattelli, *Magn. Reson.* **2021**, *2*, 511 – 522.
[35] S. Chhabra, P. Fischer, K. Takeuchi, A. Dubey, J. J. Ziarek, A. Boeszoermenyi, D. Mathieu, W. Bermel, N. E. Davey, G. Wagner, H. Arthanari, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1710 – E1719.
[36] L. E. Wong, J. Maier, J. Wienands, S. Becker, C. Griesinger, *J. Am. Chem. Soc.* **2018**, *140*, 3518 – 3522.
[37] M. J. Bottomley, M. J. Macias, Z. Liu, M. Sattler, *J. Biomol. NMR* **1999**, *13*, 381 – 385.
[38] V. Kanelis, L. Donaldson, D. R. Muhandiram, D. Rotin, J. D. Forman-Kay, L. E. Kay, *J. Biomol. NMR* **2000**, *16*, 253 – 259.
[39] S. Mantylahti, O. Aitio, M. Hellman, P. Permi, *J. Biomol. NMR* **2010**, *47*, 171 – 181.
[40] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, *J. Biomol. NMR* **2002**, *24*, 149 – 154.
[41] G. T. Montelione, B. A. Lyons, S. D. Emerson, M. Tashiro, *J. Am. Chem. Soc.* **1992**, *114*, 10974 – 10975.
[42] J. D. Haller, A. Bodor, B. Luy, *J. Magn. Reson.* **2019**, *302*, 64 – 71.
[43] A. Bodor, J. D. Haller, C. Bouguechtouli, F. X. Theillet, L. Nyitray, B. Luy, *Anal. Chem.* **2020**, *92*, 12423 – 12428.
[44] C. W. Lee, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, *Biochemistry* **2010**, *49*, 9964 – 9971.
[45] E. Bochkareva, L. Kaustov, A. Ayed, G. S. Yi, Y. Lu, A. Pineda-Lucena, J. C. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith, A. Bochkarev, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15412 – 15417.
[46] E. F. Dudás, G. Pálfy, D. K. Menyhárd, F. Sebák, P. Ecsédi, L. Nyitray, A. Bodor, *ChemBioChem* **2020**, *21*, 3087 – 3095.
[47] T. S. Wong, S. Rajagopalan, S. M. Freund, T. J. Rutherford, A. Andreeva, F. M. Townsley, M. Petrovich, A. R. Fersht, *Nucleic Acids Res.* **2009**, *37*, 6765 – 6783.
[48] L. M. M. Jenkins, S. R. Durell, S. J. Mazur, E. Appella, *Carcinogenesis* **2012**, *33*, 1441 – 1449.
[49] P. F. Lambert, F. Kashanchi, M. F. Radonovich, R. Shiekhattar, J. N. Brady, *J. Biol. Chem.* **1998**, *273*, 33048 – 33053.
[50] D. P. Teufel, M. Bycroft, A. R. Fersht, *Oncogene* **2009**, *28*, 2112 – 2118.
[51] P. Di Lello, L. M. M. Jenkins, T. N. Jones, B. D. Nguyen, T. Hara, H. Yamaguchi, J. D. Dikeakos, E. Appella, P. Legault, J. G. Omichinski, *Mol. Cell* **2006**, *22*, 731 – 740.
[52] A. C. Wang, S. Grzesiek, R. Tschudin, P. J. Lodi, A. Bax, *J. Biomol. NMR* **1995**, *5*, 376 – 382.
[53] M. A. Smith, H. Hu, A. J. Shaka, *J. Magn. Reson.* **2001**, *151*, 269 – 283.
[54] M. Kjaergaard, F. M. Poulsen, *J. Biomol. NMR* **2011**, *50*, 157 – 165.
[55] M. Kjaergaard, S. Brander, F. M. Poulsen, *J. Biomol. NMR* **2011**, *49*, 139 – 149.

[56] N. Blom, T. Sicheritz-Ponten, R. Gupta, S. Gammeltoft, S. Brunak, *Proteomics* **2004**, *4*, 1633 – 1649.
[57] N. St-Denis, M. Gabriel, J. P. Turowec, G. B. Gloor, S. S. C. Li, A. C. Gingras, D. W. Litchfield, *J. Proteomics* **2015**, *118*, 49 – 62.
[58] L. McKendrick, D. Milne, D. Meek, *Mol. Cell. Biochem.* **1999**, *191*, 187 – 199.
[59] X. Sun, H. J. Dyson, P. E. Wright, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021456118.
[60] C. Dai, W. Gu, *Trends Mol. Med.* **2010**, *16*, 528 – 536.
[61] A. M. Bode, Z. G. Dong, *Nat. Rev. Cancer* **2004**, *4*, 793 – 805.
[62] Y. M. Lee, Y. C. Liou, *Front. Oncol.* **2018**, *8*, 469.
[63] Y. Chen, Y. R. Wu, H. Y. Yang, X. Z. Li, M. M. Jie, C. J. Hu, Y. Y. Wu, S. M. Yang, Y. B. Yang, *Cell Death Dis.* **2018**, *9*, 17.
[64] A. Hatos, B. Hajdu-Soltesz, A. M. Monzon, N. Palopoli, L. Alvarez, B. Aykac-Fas, C. Bassot, G. I. Benitez, M. Bevilacqua, A. Chasapi, L. Chemes, N. E. Davey, R. Davidovic, A. K. Dunker, A. Elofsson, J. Gobeill, N. S. G. Foutel, G. Sudha, M. Guharoy, T. Horvath, V. Iglesias, A. V. Kajava, O. P. Kovacs, J. Lamb, M. Lambrughi, T. Lazar, J. Y. Leclercq, E. Leonardi, S. Macedo-Ribeiro, M. Macossay-Castillo, E. Maiani, J. A. Manso, C. Marino-Buslje, E. Martinez-Perez, B. Meszaros, I. Micetic, G. Minervini, N. Murvai, M. Necci, C. A. Ouzounis, M. Pajkos, L. Paladin, R. Pancsa, E. Papaleo, G. Parisi, E. Pasche, P. J. B. Pereira, V. J. Promponas, J. Pujols, F. Quaglia, P. Ruch, M. Salvatore, E. Schad, B. Szabo, T. Szaniszlo, S. Tamana, A. Tantos, N. Veljkovic, S. Ventura, W. Vranken, Z. Dosztanyi, P. Tompa, S. C. E. Tosatto, D. Piovesan, *Nucleic Acids Res.* **2020**, *48*, D269 – D276.
[65] 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, *Nat. Commun.* **2017**, *8*, 15233.

Manuscript received: June 23, 2021
Revised manuscript received: September 21, 2021
Accepted manuscript online: September 29, 2021
Version of record online: November 16, 2021