



HAL
open science

Site-Specific Incorporation of Fluorinated Prolines into Proteins and their Impact on Neighbouring Residues

Carlos Elena-Real, Annika Urbanek, Amin Sagar, Priyesh Mohanty, Geraldine Levy, Anna Morató, Aurélie Fournet, Frédéric Allemand, Nathalie Sibille, Jeetain Mittal, et al.

► To cite this version:

Carlos Elena-Real, Annika Urbanek, Amin Sagar, Priyesh Mohanty, Geraldine Levy, et al.. Site-Specific Incorporation of Fluorinated Prolines into Proteins and their Impact on Neighbouring Residues. *Chemistry - A European Journal*, In press, 10.1002/chem.202403718 . hal-04865645

HAL Id: hal-04865645

<https://hal.science/hal-04865645v1>

Submitted on 6 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Site-Specific Incorporation of Fluorinated Prolines into Proteins and their Impact on Neighbouring Residues

Carlos A. Elena-Real,^{a,#} Annika Urbanek,^{a,#} Amin Sagar,^a Priyesh Mohanty,^b Geraldine Levy,^{c,d}
Anna Morató,^a Aurélie Fournet,^a Frédéric Allemand,^a Nathalie Sibille,^a Jeetain Mittal,^{b,e,f} Davy
Sinnaeve^{c,d,*} and Pau Bernadó^{a,*}

a- Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS
29 rue de Navacelles, 34090 Montpellier (France).

b- Artie McFerrin Department of Chemical Engineering, Texas A&M, College Station, TX
77843.

c- Univ. Lille, INSERM, Institut Pasteur de Lille, CHU Lille, U1167 - RID-AGE - Risk Factors
and Molecular Determinants of Aging-Related Diseases. 59000 Lille, France.

d- CNRS EMR9002 – Integrative Structural Biology, 59000 Lille, France

e- Department of Chemistry, Texas A&M University, College Station, TX 77843.

f- Interdisciplinary Graduate Program in Genetics and Genomics, Texas A&M University,
College Station, TX 77843.

These authors contributed equally to this work

* **Corresponding Author:** Davy Sinnaeve, Integrative Structural Biology
(davy.sinnaeve@univ-lille.fr); Pau Bernadó, Centre de Biologie Structurale
(pau.bernado@cbs.cnrs.fr)

Abstract

The incorporation of fluorinated amino acids into proteins provides new opportunities to study biomolecular structure-function relationships in an elegant manner. Several approaches have been developed for their incorporation, although they are not site-specific, limited to peptides or imply large, highly modified amino acids. Here, we present a chemical biology approach for the site-specific incorporation of three commercially available C γ -modified fluoroprolines that has been validated using a non-pathogenic version of huntingtin exon-1 (HttExon-1). ^{19}F , ^1H and ^{15}N NMR chemical shifts measured for multiple variants of HttExon-1 indicated that the *trans/cis* ratio was strongly dependent on the fluoroproline variant and the sequence context. By isotopically labelling the rest of the protein, we have shown that the extent of spectroscopic perturbations to the neighbouring residues depends on the number of fluorine atoms and the stereochemistry at C γ , as well as the isomeric form of the fluoroproline. We have rationalized these observations by means of extensive molecular dynamics simulations, indicating that the observed atomic chemical shift perturbations correlate with the distance to fluorine atoms and that the effect remains very local. These results validate the site-specific incorporation of fluoroprolines as an excellent strategy to monitor intra- and intermolecular interactions in disordered proline-rich proteins.

Introduction

The incorporation of non-canonical amino acids (ncAAs) into proteins is a powerful tool for studying protein structure, dynamics and function¹⁻³. Moreover, ncAAs expand the chemical diversity that is available in nature, incorporating novel and unique biophysical or functional properties in native proteins^{4,5}. For instance, ncAA can be used to (de)stabilize particular secondary structures or to disrupt specific interactions within proteins^{6,7}. The incorporation of chemical probes to monitor the local structure and dynamics of proteins remains one of the most common applications of ncAAs. Amino acids carrying stable radicals or cyano groups have been used to measure intramolecular distances by nuclear magnetic resonance (NMR) and electronic paramagnetic resonance and to monitor protein motions by infrared spectroscopy⁸⁻¹⁰. Among these probes, fluorine has become very popular due to its unique properties, especially for NMR experiments¹¹⁻¹³. ¹⁹F, which is the most abundant isotope, is almost non-existent in biological samples and provides very clean NMR spectra¹⁴. Moreover, its large gyromagnetic ratio and extremely broad chemical shift dispersion make ¹⁹F-NMR a highly sensitive tool to probe protein structure and dynamics at high resolution and for pharmacological applications^{12,15,16}. However, the quantitative interpretation of the resulting ¹⁹F chemical shifts is challenging as theoretical models, including quantum chemistry calculations, fail in providing accurate values^{17,18}.

Fluorine atoms can be incorporated into proteins using several strategies that can be global or site-specific^{16,19}. An example of a global strategy is the post-translational incorporation of fluorine. For this, highly reactive fluorinated moieties, such as $-CF_3$, are covalently anchored to solvent exposed cysteines or lysines²⁰⁻²². In order to achieve the systematic incorporation of a given fluorinated amino acid, some authors add the ¹⁹F-modified form of the amino acid or a suitable precursor into the cell culture²³⁻²⁵. Unfortunately, this approach severely reduces the expression yield and cannot be applied to all natural amino acids. The use of auxotrophic *Escherichia coli* strains can facilitate amino acid incorporation^{24,26}. For most of the cases, however, site-specific incorporation of fluorinated amino acids is desired. Solid-phase peptide synthesis is the most straightforward method to introduce fluorinated amino acids in a site-specific manner²⁷. The advantage of this approach is its flexibility in selecting the fluorinated amino acid, which can be chemically similar to the natural one. However, the size of the systems amenable to solid-phase peptide synthesis is severely limited, unless a subsequent chemical ligation is performed²⁸. A more elegant way to selectively introduce fluorinated amino acids into proteins is using an expanded genetic code via tRNA nonsense suppression¹⁻³. In this

strategy, the protein is overexpressed together with an engineered tRNA_{CUA} and its associated aminoacyl tRNA synthetase (aaRS) in the presence of the fluorinated amino acid. In the cell or in a cell-free (CF) reaction mixture, the aaRS aminoacylates the tRNA_{CUA} with the amino acid and is subsequently recognized by the ribosome and introduced in a position specific manner defined by the presence of an amber stop codon in the coding gene. However, in the vast majority of cases, the fluorinated amino acid must host important chemical differences to avoid recognition by the endogenous aaRS²⁹⁻³¹, precluding the usage of *quasi*-natural amino acids in which one or few hydrogen atoms are exchanged by fluorine.

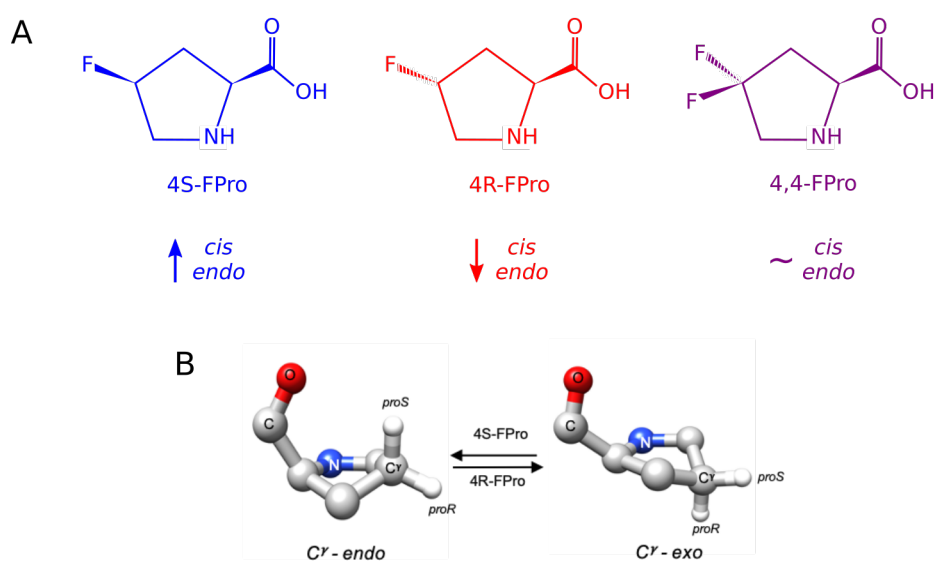


Figure 1. Conformational preferences of the studied fluoroprolines. (A) Chemical structure of the three fluoroprolines (4S-FPro, blue; 4R-FPro, red; and 4,4-FPro, purple) used in this study. The colour code is used throughout the study. Their conformational preferences trans/cis and endo/exo with respect to the canonical prolines are displayed below. (B) Cartoon of the endo/exo conformations found for the trans isomer of prolines where the position of the proS and proR H γ are indicated. The conformational equilibrium is biased for 4S-FPro (towards endo) and 4R-FPro (towards exo).

The availability of multiple commercial fluoroprolines, particularly those substituted at positions 3 (C β) and 4 (C γ), and their Fmoc/Boc derivatives has prompted intense research on the conformational effects of these ncAAs^{13,28,32-41}. The presence of an electron-withdrawing fluorine atom in prolines, especially in C γ , strongly influences the ring puckering, with the *S*- and *R*-stereochemistry favouring the C γ -endo and C γ -exo pucker, respectively (Fig. 1). As a consequence of the stabilizing $n \rightarrow \pi^*$ interactions between consecutive carbonyl groups, the C γ exo pucker, which is enriched in (2*S*,4*R*)-fluoroproline (4R-FPro), leads to an enhanced

preference for the *trans* rotamer^{36,42}. In contrast, the (2*S*,4*S*)-fluoroproline (4*S*-FPro), which is biased towards the *endo* pucker, presents an elevated *cis* population relative to regular proline. Indeed, the uniform substitution of proline by 4*R*-FPro results in a stabilization of the polyproline-II (PPII) conformation in oligoprolines, while 4*S*-FPro destabilizes this secondary structure⁴³. The impact of single substitutions of 4*R*-FPro and 4*S*-FPro residues in oligoprolines has been recently studied with NMR spectroscopy⁴⁴. Based on ¹³C chemical shifts, it was concluded that the impact on the PPII structure was minimal, likely only featuring some small perturbations on the neighboring residues due to preferred ψ dihedral angles of the FPro residue. Interestingly, when two fluorine atoms are introduced in C γ ((2*S*)-4,4-difluoroproline, 4,4-FPro), the above-mentioned structural effects are counteracted and its conformational behaviour is similar to that of the native proline^{33,39,45,46}. Indeed, a recent study suggested that the structural impact of a single 4,4-FPro substitution within oligoprolines was minimal⁴⁷.

Due to the tunability of the structural properties of fluoroprolines, these amino acids have become very popular tools in chemical biology to modulate protein and peptide stability and improve folding kinetics^{35,39–41,48}. In addition to solid-phase peptide synthesis, protein semi-synthesis and native chemical ligation²⁸, fluoroprolines have also been incorporated into recombinant proteins using auxotrophic *E. coli* strains, resulting in a global replacement of native proline by one of its fluorinated counterparts⁴⁹. These studies have shown that fluorinated prolines can alter the stability^{50,51}, the folding mechanism^{50,52,53}, the fluorescent properties⁵⁴ or the enzymatic activity^{55,56} of proteins. However, this approach does not allow the site-specific investigation of the perturbations exerted by these ncAAs.

We have recently developed a strategy that enables the incorporation of [¹⁵N,¹³C]-prolines in specific positions in proteins⁵⁷. This approach, which we named Site-Specific Isotopic Labelling (SSIL)⁵⁸, uses the previously mentioned genetic code expansion approach in a CF protein synthesis scheme. The ability to specifically incorporate proline isotopologues in a defined position relies on the external aminoacylation of the tRNA_{CUA} by an orthogonal proline tRNA synthetase (ProRS), which is subsequently added to the reaction mixture. In the present study, we show that the same ProRS derived from *Pyrococcus horikoshii* can be used to load the tRNA_{CUA} with three commercially available prolines fluorinated in C γ and selectively incorporate them into proteins to study their conformational specificities and evaluate the structural perturbation on neighbouring residues. Concretely, we have incorporated 4*R*-FPro, 4*S*-FPro and 4,4-FPro, which are the prototypical examples of commercially available fluorinated prolines presenting well defined conformational properties (Fig. 1A)^{35,39}. In order

to illustrate our methodology, we have used a non-pathogenic construct of huntingtin exon-1 (HttExon-1), the causing agent of Huntington's disease. The HttExon-1 construct used encompasses a 16 glutamine-long tract (poly-Q) immediately followed by a polyproline (poly-P) tract of 11 prolines that structurally influences poly-Q (Fig. 2A)^{59,60}. The incorporation of the three fluoroprolines at two positions of the poly-P has enabled the investigation of the perturbation of the *trans/cis* equilibrium exerted by these ncAAs in the context of a repetitive protein. Moreover, we have quantitatively explored the influence of these modified amino acids on the neighbouring glutamines and, using extensive molecular dynamics simulations, we have shown that the chemical shift changes exerted by fluoroprolines are distance dependent, but their effect is limited to very short distances. Overall, our study shows that, using our methodology, commercial fluoroprolines can be site-specifically incorporated in large protein constructs alongside with isotopic labelling of other amino acids, widening their already established scope as interesting probes for structural and functional studies.

Results

Efficient site-specific introduction of fluoroprolines in HttExon-1

In a previous study, we demonstrated the use of an orthogonal *P. horikoshii* ProRS/tRNA_{CUA} pair⁶¹ in CF to site-specifically incorporate isotopically labelled prolines into proteins in CF⁵⁷. In this procedure, the tRNA_{CUA} was aminoacylated *in vitro* by mixing it with the ProRS in the presence of proline and ATP at 37°C. In the present study, we first tested whether the same procedure could be applied to aminoacylate the tRNA_{CUA} with the three fluorinated prolines (Fig. 1A). The reaction conditions (concentrations, temperature and reaction time) for the three ncAA were explored, but no differences with respect to those previously established for canonical proline were found (see methods section). Urea-PAGE gels after aminoacylation with proline and the three fluoroprolines displayed two bands, corresponding to the loaded and empty tRNA_{CUA} (Fig. 2B). Although a precise quantification was not performed, the band corresponding to the aminoacylated tRNA_{CUA} was more intense than the empty one in all the cases, indicating that the presence of fluorine atoms in the amino acid does not preclude the enzymatic activity of the ProRS.

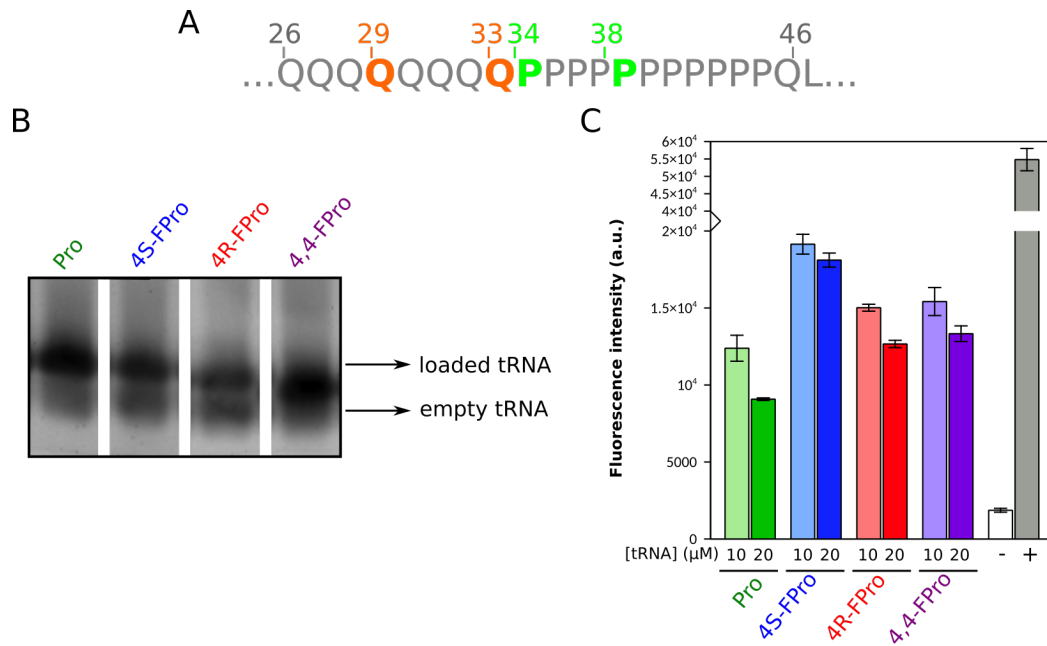


Figure 2. Incorporation of the three fluoroprolines into H16. (A) Fragment of H16 in which modifications have been performed. The three fluoroprolines have been incorporated in positions Pro34 or Pro38 (green). Gln29 or Gln33 (orange) have been isotopically labelled to monitor the structural and spectroscopic changes induced when incorporating fluoroprolines in position Pro34. (B) Urea-PAGE gels displaying the bands corresponding to the empty and loaded $tRNA_{CUA}$ after aminoacylation with *P. horikoshii* ProRS. (C) Endpoint fluorescence intensity plot for CF expression of H16-P34 with 10 and 20 μM of $tRNA_{CUA}$ loaded with proline (green), 4S-FPro (blue), 4R-FPro (red) and 4,4-FPro. + and - indicate the positive (wild-type H16) and negative (H16-P34 without $tRNA_{CUA}$) controls, respectively.

Then, we proceeded to evaluate the capacity of fluorinated prolines to be site-specifically incorporated into proteins using a CF approach. For this, we used a previously described HttExon-1 construct fused to the N-terminus of superfolder green fluorescent protein (sfGFP), hereafter called H16⁵⁸. In order to evaluate our procedure, the codon corresponding to Pro34, the first proline of the poly-P tract, was switched to an amber stop codon (TAG), yielding H16-P34. The addition of 10 and 20 μM of the $tRNA_{CUA}$ preparation to the CF reaction resulted in fluorescence signal for the native and the three fluorinated prolines, indicating the efficient synthesis of the protein with the desired amino acid (Fig. 2C). Similarly to previous studies of our group, a yield of 20-40% with respect to the plasmid in the absence of the stop codon was observed. Furthermore, a slight decrease in the H16-P34 production was noticed when increasing the loaded $tRNA_{CUA}$ concentration to 20 μM ^{57,58,62}. Interestingly, the production of the proteins in the presence of the fluorinated prolines was slightly more efficient than with the native one, although we did not further explore this observation.

A ^{19}F -NMR perspective of fluorinated prolines in HttExon-1

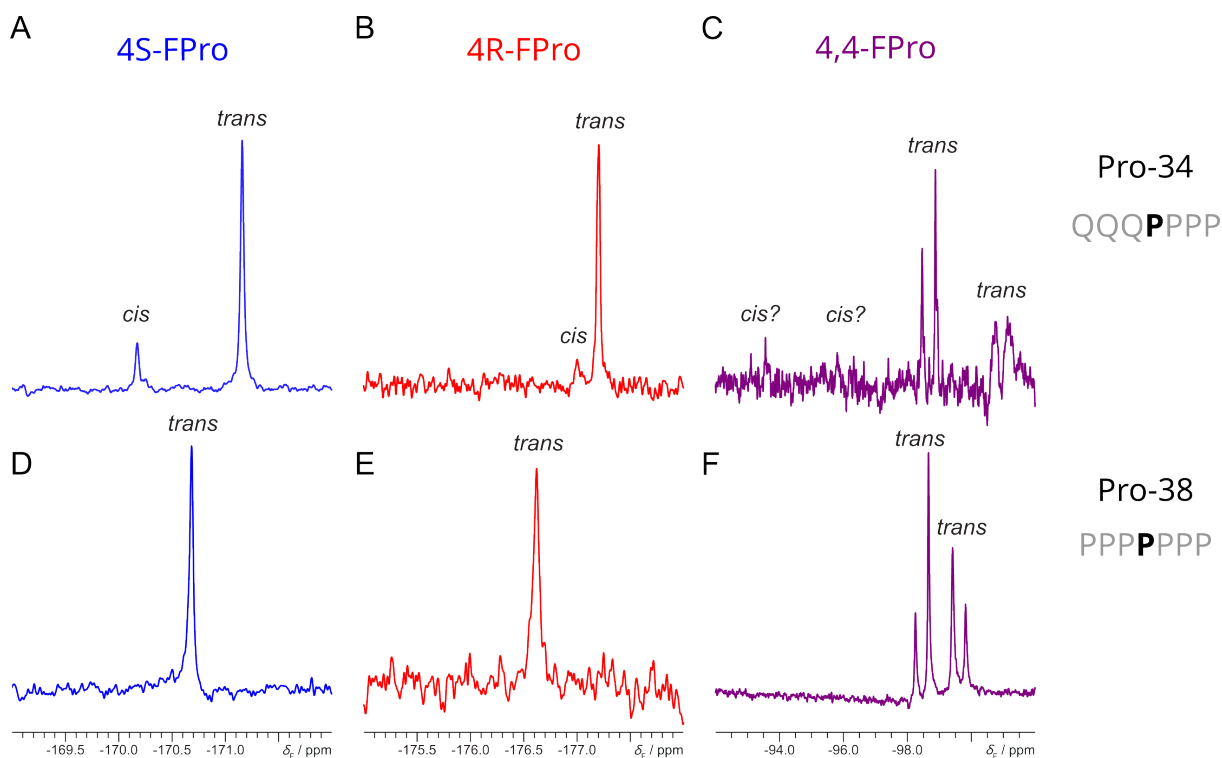


Figure 3. ^{19}F -NMR spectra of fluoroprolines incorporated in H16. Spectra of 4S-FPro (blue), 4R-FPro (red) and 4,4-FPro (purple) incorporated in positions Pro34 (A-C) and Pro38 (D-F). Signals assigned to *trans* and *cis* configurations are indicated.

The 1D ^{19}F -NMR spectra featuring 4R-FPro and 4S-FPro at position Pro34 revealed the presence of a major and a minor form in slow exchange, which we attributed to the presence of the *trans* and *cis* conformations of this residue, respectively (Fig. 3A,B). The introduction of 4,4-FPro in the same position clearly delivered the *trans*-form signals, but unfortunately the signal-to-noise ratio was insufficient to confidently identify the *cis*-form signals (Fig. 3C). The deconvolution of the ^{19}F -NMR signals of the fluorinated constructs indicated an important difference in the *cis* content for 4R-FPro and 4S-FPro with $9.1 \pm 3.7\%$ and $17.5\% \pm 1.3\%$, respectively. Note that the *cis* population for canonical proline in Pro34 was previously reported to be between 12% and 16% using SSIL samples^{57,60}. The ^{19}F -NMR experiments showed that 4R-FPro promotes the population of the *trans* form, while 4S-FPro increases the *cis* form. These observations are in line with previously reported *trans/cis* ratios for Ac-(F)Pro-NMe₂ model compounds⁶³.

Substitutions with all three fluoroprolines at position Pro38, in the core of the poly-P stretch, only yielded the *trans* signals, indicating that the *cis* content was very low (Fig. 3D-F). This observation is in agreement with previous observations that the *cis* conformer of proline within

the poly-P stretch was undetectably low⁵⁷. For 4,4-FPro, the chemical shift difference between both ¹⁹F doublets decreased when compared to the substitution at the Pro34 position, which suggested a more balanced (*endo/exo*) conformational equilibrium⁴⁷. Remarkably, for 4,4-FPro and independently of the position in the poly-P tract, the resonance intensities and line widths of both doublets were different, indicating distinct ¹⁹F relaxation properties for both fluorine atoms. This is more noticeable at position Pro34. Note that fluorine relaxation is largely governed by its chemical shift anisotropy (CSA)^{11,12}. Density Functional Theory (DFT) calculations for 4R-FPro and 4S-FPro have previously shown that the ¹⁹F CSA is significantly different for the *C^γ-exo* and *C^γ-endo* conformers, with the highest CSA values found when the fluorine is in the pseudo-axial position⁴⁴. The distinct relaxation properties of both fluorines in 4,4-FPro could thus be explained by a difference in the *C^γ-exo* and *C^γ-endo* populations, resulting in distinct average CSAs. The observation that the line widths are more different at position Pro34 than at Pro38 is thus consistent with the observed higher difference in chemical shift, implying a more biased ring conformation in this position due to the sequence context.

Fluoroprolines induce spectroscopic changes to the neighbouring glutamine

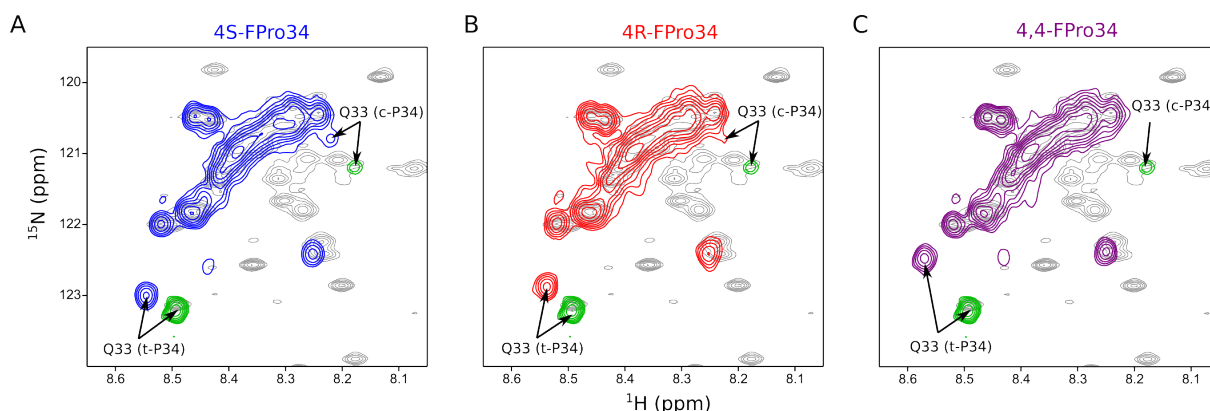


Figure 4. Probing the incorporation of fluoroprolines in H16. ¹⁵N-HSQC spectra of ¹⁵N-glutamine labelled H16-P34 after the incorporation of (A) 4S-FPro (blue), (B) 4R-FPro (red) and (C) 4,4-FPro (purple). Spectra are overlaid with the fully labelled H16 (grey) and the SSIL sample of Gln33⁶⁰ (green). Chemical shift perturbations on trans and cis peaks of Gln33 caused by the presence of fluoroprolines are indicated with arrows.

The efficient incorporation of fluoroprolines into H16 enabled the evaluation of the spectroscopic changes occurring in the neighbouring poly-Q tract (Table 1). For this, we separately incorporated the three fluoroprolines in position Pro34 of H16 in a [¹⁵N,¹³C]-Gln-containing CF reaction, and recorded the corresponding ¹⁵N-HSQC experiments (Fig. 4). Equivalently to the native protein, the three spectra displayed a large, unresolved density

corresponding to the glutamines present in H16^{60,64}, with some isolated signals corresponding to glutamines with specific chemical environments. Interestingly, the most downfield N^H-H^N correlation, which corresponds to Gln33 (see below), exhibits different chemical shifts depending on the specific fluoroproline introduced (Table 1). Concretely, the largest Gln33 N^H chemical shift change was observed for 4,4-FPro (-0.72 ppm), followed by 4R-FPro (-0.33 ppm), while 4S-FPro displayed the smallest one (-0.19 ppm). Although the ¹H chemical shift turned out to be less affected than that of ¹⁵N, changes induced by 4,4-FPro (0.074 ppm) were notably larger than those observed for 4R-FPro and 4S-FPro, 0.044 ppm and 0.048 ppm, respectively. These observations suggest that, not surprisingly, a larger effect is observed when two fluorine atoms are introduced in close proximity to the probed residue.

Table 1. sample characteristics and structural properties of the H16 constructs with the three fluoroprolines.

| Fluoroproline | Sample preparation | | <i>trans:cis</i> ^b | | <i>trans</i> | | <i>cis</i> | |
|---------------|----------------------------|-----------------------|-------------------------------|-------|-------------------------|------------------------|-------------------------|------------------------|
| | CF conc. (μM) ^a | NMR sample conc. (μM) | Pro34 | Pro38 | Δ ¹⁵ N (ppm) | Δ ¹ H (ppm) | Δ ¹⁵ N (ppm) | Δ ¹ H (ppm) |
| 4S-FPro | 2.6 | 21.0 | 82.5:17.5 | 100:0 | -0.19 | 0.048 | -0.42 | 0.039 |
| 4R-FPro | 1.6 | 8.7 | 90.9:9.1 | 100:0 | -0.33 | 0.044 | ND ^c | ND ^c |
| 4,4-FPro | 1.9 | 5.2 | ND | 100:0 | -0.72 | 0.074 | ND | ND |

ND: Non-determined frequencies due to overlap or low intensity.

a- H16 concentrations at the end of the CF reactions (10 mL).

b- Data obtained from ¹⁹F-NMR experiments.

c- Ill-defined peaks in the ¹⁵N-HSQC.

The above-described observations corresponded to the major (*trans*) conformation of the Gln33-Pro34 bond monitored on Gln33. Note that *cis* populations of $9.1 \pm 3.7\%$ and $17.5\% \pm 1.3\%$ were detected using ¹⁹F-NMR for this bond upon introducing 4R-FPro and 4S-FPro, respectively (Fig. 3). We also monitored the chemical shift perturbation of this minor conformation. By inspecting the unresolved glutamine signal density of H16-P34 containing 4S-FPro, we identified a small peak that was not present in the spectrum of the native protein (Fig. 4A). The identity of this signal as the *cis* form of the 4S-FPro manifested in Gln33 was subsequently confirmed (see below). Although less clear, a similar additional density in a similar position was observed for the 4R-FPro variant of H16-P34, which was also assigned to the *cis* form of this ncAA (Fig. 4B). Unfortunately, no signal was observed for the 4,4-FPro variant, which was probably overlapped in the unresolved glutamine density of the ¹⁵N-HSQC

spectrum. The chemical shift perturbation on Gln33 N^H induced by 4S-FPro was -0.42 ppm. Interestingly, this perturbation was notably larger than that observed for the *trans* isomer upon the incorporation of the same ncAA (-0.19 ppm). Although the perturbation induced by 4R-FPro on the *cis* conformation could not be accurately quantified due to its low intensity, our data suggest that the perturbation exerted is different than for 4S-FPro. While 4R-FPro has a stronger effect than 4S-FPro on the *trans* isomer, the *cis* isomer is more perturbed by 4S-FPro. The ensemble of our observations shows that the number of fluorine atoms, the stereochemistry in C^γ and the isomeric form of the fluoroproline play relevant roles in the extent of the spectroscopic perturbation exerted on neighbouring amino acids.

Two-site orthogonal nonsense suppression in cell-free

The severe frequency overlap of the fluorinated H16-P34 samples precluded the NMR analysis of the effect exerted by fluoroprolines on glutamines of the poly-Q tract other than Gln33. In order to overcome this limitation, we endeavoured to perform a two-site orthogonal nonsense suppression reaction in CF. Briefly, we aimed to simultaneously introduce a fluoroproline in position Pro34 and an isotopically labelled glutamine at specific positions of the poly-Q tract. For this, we modified the previously developed tRNA_{CUA}/aaRS pairs for the external tRNA_{CUA} aminoacylation of proline⁵⁷ and glutamine⁵⁸. In the absence of an engineered strain enabling the expansion of the genetic code^{65,66}, we simultaneously reassigned the amber (TAG) and opal (TGA) stop codons and the corresponding tRNA anticodons for the double orthogonal nonsense suppression. First, we tested the performance of the TGA nonsense codon with respect to the TAG one. For this, we added increasing amounts of proline-loaded tRNA_{CUA} and tRNA_{UCA} to a CF mixture containing the H16 plasmid with the amber and the opal stop codons in the Pro34 position, respectively (Fig. 5A). The end-point fluorescence measurement of the reaction showed an important decrease in the yield when the TGA nonsense was used. This decrease varied depending on the concentration of tRNA_{UCA} used, but it was more than 60% in all cases. This observation was expected as our lysate, which is depleted of the release factor (RF) 1,⁶⁷ still contains RF2 that competes for the opal nonsense codon, reducing the overall yield of the reaction.

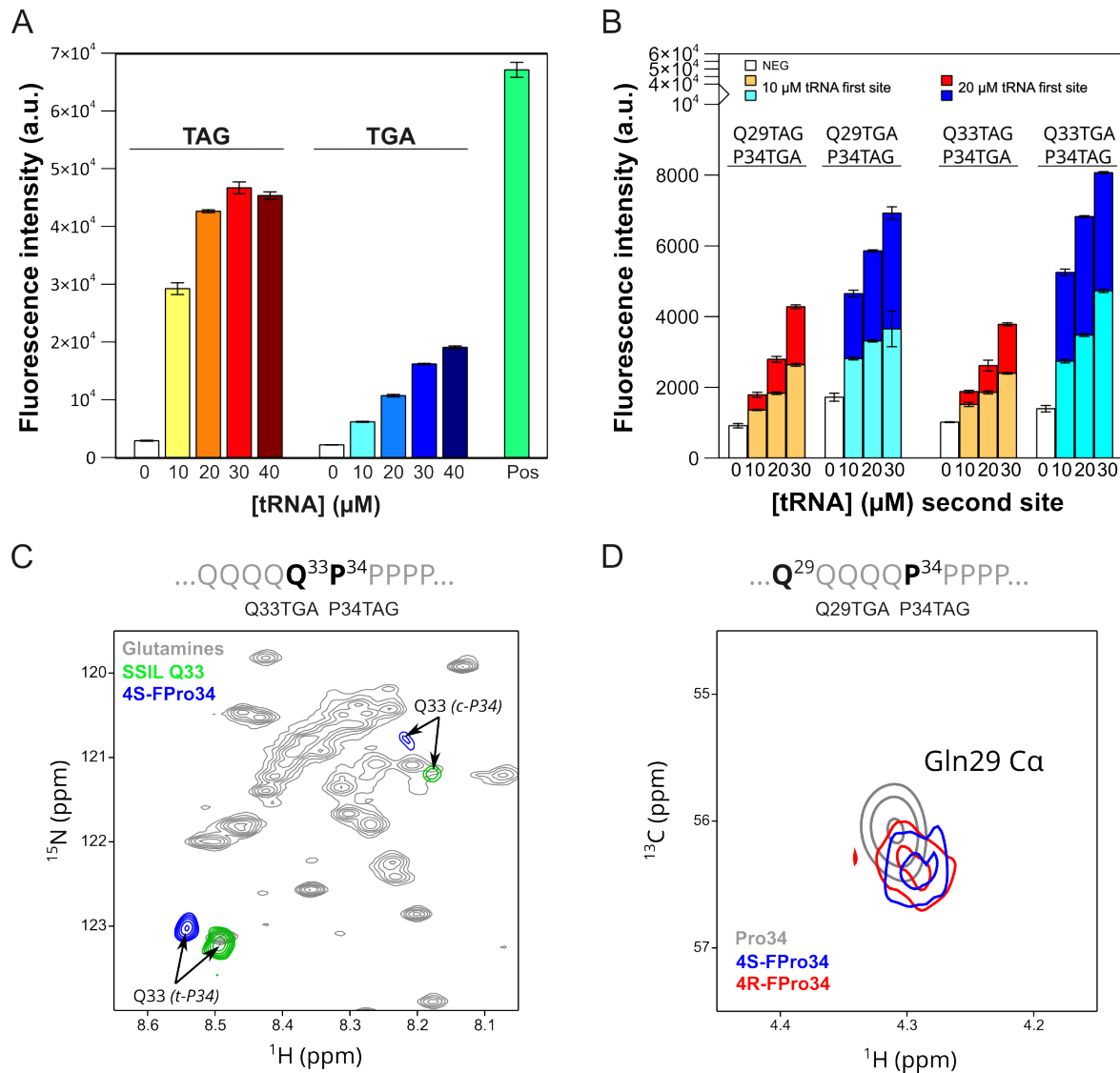


Figure 5. Validation and application of the two-site orthogonal nonsense suppression. (A) Fluorescence intensities measured upon the addition of increasing amounts of proline-loaded $t\text{RNA}_{\text{CUA}}$ (left) and $t\text{RNA}_{\text{UCA}}$ (right) to suppress the TAG and TGA stop codons, respectively, introduced in position Pro34 of H16. Yields are compared to the production of H16 without stop codon (positive control). **(B)** Combined titration of loaded $t\text{RNA}_{\text{CUA}}$ and $t\text{RNA}_{\text{UCA}}$ for the orthogonal suppression of Gln29/Pro34-H16 using either TAG/TGA (orange and red) or TGA/TAG (cyan and blue) stop codons. **(C)** ^{15}N -HSQC of an orthogonally suppressed H16 sample incorporating $^{15}\text{N},^{13}\text{C}$ -Gln in position Gln33 and 4S-FPro in position Pro34 (blue) overlaid with a SSIL H16-Q33 (green) and fully labelled H16 (grey) samples. **(D)** Zoom of the Gln29 $\text{C}\alpha$ - $\text{H}\alpha$ correlation of the ^{13}C -HSQCs of orthogonally suppressed H16 samples containing $^{15}\text{N},^{13}\text{C}$ -Gln in position Gln29 and 4S-FPro (blue) or 4R-FPro (red) in position Pro34. These two spectra are overlaid with the ^{13}C -HSQC of the SSIL H16-Q29 sample.

Second, we analysed whether the order of the nonsense codons in the plasmid affected the protein synthesis. In a first set of experiments, we positioned the TAG and the TGA codons in the position of Gln29 and Pro34, respectively. Increasing concentrations of glutamine-loaded

tRNA_{CUA} and proline-loaded tRNA_{UCA} were added to the CF reaction and the fluorescence endpoint was monitored (Fig. 5B). Not surprisingly, the higher the concentration of both tRNAs, the higher the yield. No saturation of the protein production was observed when large tRNA amounts (up to 50 μ M) were added to the CF mixture. This behaviour was similar to the single suppression experiments for TGA, but not for TAG (Fig. 5A). This observation suggested that suppressing the TGA codon was the bottleneck of the synthesis. An equivalent titration experiment was performed using an H16 plasmid where the two nonsense codons were swapped (TGA and TAG in positions of Gln29 and Pro34, respectively). For this codon arrangement, the positive correlation between tRNA concentration and protein production was also observed. However, this second set of experiments systematically yielded larger protein amounts than the first one, suggesting that the arrangement of nonsense codons played a relevant role during protein translation.

Finally, we tested whether the relative position of the two nonsense codons influenced the protein production. For this, we positioned them consecutively in the positions of Gln33 and Pro34 of H16, and we repeated the previously described titration experiments using the two nonsense codon arrangements (Fig. 5B). Interestingly, we observed that the presence of two consecutive stop codons did not impact the translation and yields were similar to those obtained with the construct with more spaced nonsense codons. Again, we observed that utilizing TGA to suppress glutamine and TAG for proline resulted in a higher protein production than the inverse.

In summary, these experiments validated the orthogonal site-specific incorporation of amino acids in proteins and defined the optimal arrangement of nonsense codons. Importantly, although the protein production yields were notably reduced when using this strategy, our results indicated that sufficient H16 could be produced for subsequent structural investigations.

Overcoming spectral crowding with two-site orthogonal nonsense suppression

In order to validate our double orthogonal nonsense suppression for structural applications, we produced an H16 sample in which the codons corresponding to Gln33 and Pro34 were changed to the TGA and TAG nonsense codons, respectively. We produced the sample from 10 mL of CF reaction, containing 20 μ M of tRNA_{UCA} loaded with [¹⁵N, ¹³C]-Gln and 10 μ M of tRNA_{CUA} loaded with 4S-FPro. The purified product provided a 7 μ M NMR sample that displayed two correlations with distinct intensity in the glutamine region of the ¹⁵N-HSQC spectrum (Fig. 5C). When overlaying this spectrum with the one previously obtained for H16-P34 suppressed with

4S-FPro (Fig. 4A), we observed a perfect overlap with the previously assigned *trans* and *cis* peaks probed by Gln33, validating our approach. Equivalent samples were produced by incorporating 4R-FPro and 4,4-FPro, also showing an excellent overlap with the previously identified major *trans* peaks of Gln33 (see Fig. S1 in SI). However, no signal corresponding to the *cis* conformation could be unambiguously identified in these spectra, most probably due to the low population of this conformer for the 4R-FPro and 4,4-FPro variants (see above) and the low concentration of the samples ($\sim 2 \mu\text{M}$).

Then, we used the double orthogonal nonsense suppression to investigate the extent of the effects exerted by fluoroproline. For this, we explored Gln29, a glutamine exhibiting a random coil conformational behaviour that follows the α -helical section of the poly-Q tract⁶⁰. We produced H16 variants with site-specific isotopic labelling for Gln29 and containing either 4S-FPro or 4R-FPro in position Pro34, as these two fluoroproline displayed the most distinct *trans/cis* according to previous investigations in peptide models^{38,44} and our ¹⁹F-NMR experiments on H16-P34 (see above). We monitored the C α -H α correlation in Gln29 in order to probe putative fluoroproline-induced secondary structure perturbations (Fig. 5D). Interestingly, the Gln29 signals from both the H16 variants overlapped, indicating the absence of a differential effect caused by the stereochemistry of the fluorine atom in Pro34. Importantly, these peaks presented chemical shifts very similar to the C α -H α correlation of the native H16, suggesting that the influence of fluoroproline positioned five residues apart was negligible.

Molecular dynamics simulations suggest that the fluorine effect is distance dependent

In order to rationalize the chemical shift perturbations observed in flanking glutamines when incorporating fluoroproline in position Pro34 of H16, we performed molecular dynamics (MD) simulations of a HttExon-1 fragment containing the N17, the poly-Q tract with 16 glutamines and five prolines. Twenty simulations of 1 μs each of this fragment with the Gln33-Pro34 peptide bond in *trans* and *cis* conformations were performed. The independent trajectories, with an aggregated time of $\sim 20 \mu\text{s}$ each, were generated at 293.15 K using the AMBER03ws force field⁶⁸ and frames at 200 ps intervals were saved for subsequent analyses.

First, we validated the trajectories by evaluating their capacity to reproduce experimental data measured on H16. The α -helical fractions computed from both trajectories were very similar, indicating that the H16 fragment presented an overall disorder while capturing transient populations of α -helices encompassing the N17 and different portions of the poly-Q tract (Fig. S2 in SI). Importantly, the α -helical fraction profiles displayed an excellent agreement with

those previously measured for H16⁶⁰, indicating that these trajectories are excellent representations of the conformational behaviour of the protein in solution.

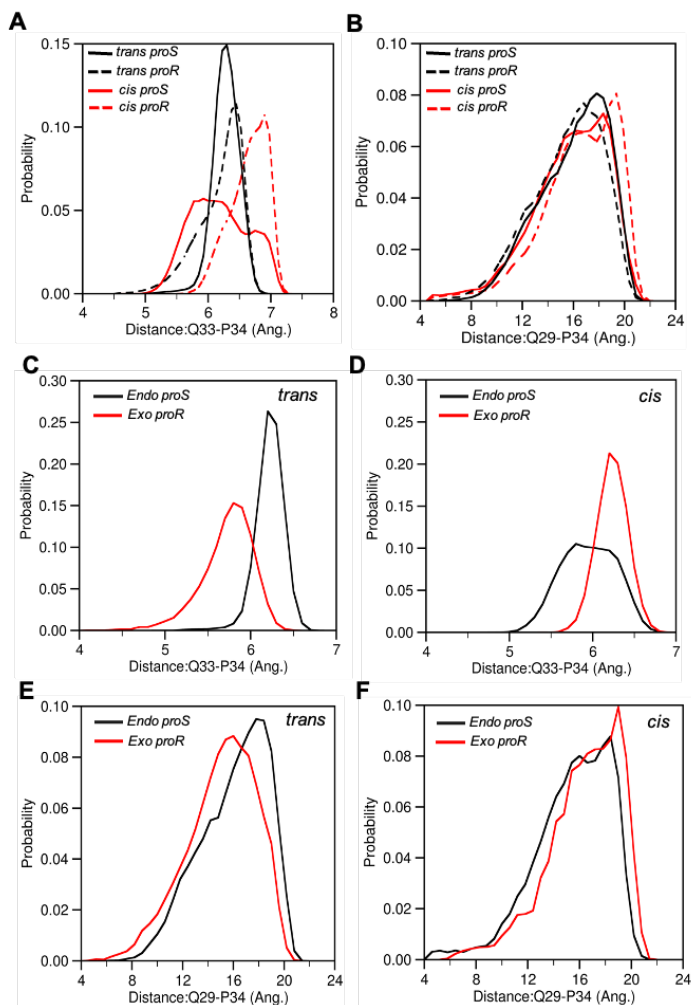


Figure 6. Fluorine distances to Gln33 and Gln29 computed from MD simulations. Computed distance distributions between the proS (solid line) and proR (dashed line) H^γ for Gln33 (A), and Gln29 N^H (B) for the MD trajectories in trans (black) and cis (red). (C,D) Distance distribution of the proS H^γ in endo (black) or the proR H^γ in exo to the Gln33 N^H corresponding to the trans (C) or the cis trajectory (D). (E,F) Distance distribution of the proS H^γ in endo (black) or the proR H^γ in exo to the Gln29 N^H corresponding to the trans (E) or the cis trajectory (F).

We aimed at evaluating whether the chemical shift perturbations that were experimentally probed can be explained by the distance to the fluorine atoms. From both trajectories, we calculated the distance distribution between the two Pro34 H^γ s (*proS* and *proR*) and the backbone N^H s for Gln33 and Gln29 (Fig. 1B), whose chemical shifts were monitored by NMR (see above). For Gln33, the *trans* trajectories showed that both H^γ s presented narrow distance

distributions centred around 6.5 Å, but the *proR* one displayed an additional shoulder corresponding to distances below 6.0 Å (Fig. 6A). The *proR* H γ thus on average comes closer to Gln33 N^H than the *proS* one. Very different distributions were observed when analysing the same distances for the trajectory with the Pro34 in a *cis* conformation. In this case, the *proS* H γ sampled a very broad range of distances to Gln33 N^H, including a large population of shorter distances below 6.0 Å (Fig. 6A). Interestingly, the *proR* H γ sampled a distribution that appeared shifted to higher distances when compared to the *trans* form. Therefore, for the *cis* form, on average, it is *proS* H γ that comes in close proximity to Gln33 N^H, and appears to do so than either H γ in *trans*. When doing the same analysis for Gln29 N^H, much broader distributions of distances, sampling from 5 to 20 Å, and very similar for both H γ s and for both Pro34 isomers were obtained (Fig. 6B). Overall, our observations suggest that the two H γ s of a proline do not sample the same space, presenting different distances to flanking residues, especially for the *cis* conformation, although this effect is rapidly lost when moving further away in the sequence.

Rationalizing the experimental chemical shift perturbations observed for the H16 constructs incorporating 4R-FPro and 4S-FPro, required adapting the trajectories to the specific conformational bias observed for both fluoroproline. Indeed, it is well established that the presence of a fluorine atom in C γ as well as its stereochemistry have a strong influence on the *exo/endo* conformational equilibrium in prolines^{36,38,39,42,44,46}. Concretely, the 4R-FPro has a strong preference for the C γ -*exo* conformation for both the *cis* and *trans* isomers (83% and 93%, respectively, within Ac-FPro-OMe model compounds), while the 4S-FPro shows an opposite behaviour with the C γ -*endo* conformation representing more than 99% for both isomers⁴⁶. Therefore, in order to evaluate the effect of the distance of the fluorine atom to the experimentally probed residues in H16, this conformational bias must be taken into account. For this, we first computed the distances within reduced conformational subsets of the HttExon-1 trajectories that selected for the Pro34 C γ -*exo* and C γ -*endo* conformations. For this, we calculated the Altona-Sundaralingam pseudorotation angle *P* of the five-membered ring to be used as a selection criterium (as detailed in the SI)^{69,70}. We found that $P = 18^\circ \pm 9^\circ$ captured the C γ -*exo* and $P = 180^\circ \pm 9^\circ$ the C γ -*endo* conformers^{67,68}, which are in agreement with the values expected for proline (Fig. S3 in SI). Within these conformationally-restricted subensembles, the distance distributions from *proR* and *proS* H γ s to Gln33 N^H were found to be highly influenced by the pucker of Pro34 (Fig. S4 in SI).

Figure 6C-F displays the distance distributions involving the particular H γ proton that is swapped for fluorine in fluoroproline, as derived from the conformationally biased subensemble that is relevant for that specific fluoroproline. Concretely, these are the distances from the *proS* H γ involving C γ -*endo* and the *proR* H γ involving C γ -*exo*, the major conformational states of 4S-FPro and 4R-FPro, respectively. When focusing on the distances involving Gln33 N^H computed from the *trans* trajectory (Fig. 6C), a clear difference between the *proR-exo* and *proS-endo* H's was observed, with the *proR-exo* H γ displaying shorter distances than the *proS-endo* one. Distances from *proR-exo* could be as short as 4.5 Å, although the maximum of the distribution was 5.7 Å, while distances from the *proS-endo* H γ displayed a very narrow peak centred in 6.4 Å. Note that these observations are in excellent agreement with our experimental observations, where the 4R-FPro exerted a stronger perturbation on Gln33 N^H chemical shift (-0.33 ppm) than the 4S-FPro (-0.19 ppm). Interestingly, these distance distributions were inverted when the *cis* trajectory was analysed, with the *proS-endo* H γ displaying shorter distances to the Gln33 N^H than the *proR-exo* H γ (Fig. 6D). Conversely, in the *cis* trajectory, the *proS-endo* H γ distance distribution sampled significantly shorter distances than in the *trans* one. This observation was in agreement with the stronger chemical shift perturbation observed for the *cis* isomer (-0.42 ppm) than for the *trans* one (-0.19 ppm) when the 4S-FPro was introduced in Pro34 (Table 1).

When performing an equivalent analysis for Gln29, much broader ranges of distances were obtained, with the vast majority being above 10 Å (Fig. 6E,F). Importantly, no systematic distance shifts were observed for *proR-exo/proS-endo* H's and *trans/cis* isomers. This is in agreement with the experimental observation that Gln29 chemical shifts were not perturbed in the presence of 4R-FPro or 4S-FPro (Fig. 5D), and suggests that atoms located at distances larger than 10 Å from fluorine are not spectroscopically perturbed.

Finally, for 4,4-FPro, both the *proR* and *proS* positions are fluorinated and a similar conformation is expected relative to proline, meaning both C γ -*exo* and C γ -*endo* conformers have to be considered. Since for *exo*, the *proR* position is on average found to be the closest to Gln33 N^H, while for *endo* this is the *proS* (Fig. S4), both fluorines are expected to significantly contribute to the chemical shift perturbation. This in agreement with the observation that 4,4-FPro in position Pro34 displays the largest chemical shift perturbation of the three fluoroprolines (Table 1).

Discussion

The incorporation of non-natural moieties, such as fluorine, into proteins enables to probe biomolecular structure and dynamics in a very elegant and clean way^{16,19}. In our study, we present tools for the incorporation of fluoroprolines in a site-specific manner by combining the tRNA suppressor strategy with CF protein synthesis. This technology expands to proteins what was previously done using solid-phase peptide synthesis^{28,44}. Although we have demonstrated our approaches for three commercial ncAAs, their application to less common C^β-modified or difluorinated prolines as well as post-translationally modified 4-hydroxyproline seems feasible^{35,46,71,72}. Crucially, this can be easily combined with additional isotopic labelling, either uniformly or for specific types of amino acids. This is an important advantage, as it paves the way for advanced NMR studies to assess the impact of the ncAA on the rest of the protein. In addition, by simultaneously using two stop codons in the construct, we have shown that the two-site orthogonal nonsense suppression is also achievable. Although the yields of such samples were substantially lower than for the single-site nonsense suppression, we have been able to produce enough labelled protein to explore the effects of the incorporation of the ncAAs to the inner part of the poly-Q tract of HttExon-1, which in the case of this homopolymer would have been impossible otherwise.

Assessing the nature of the perturbation exerted by fluoroprolines to the neighbouring residues is an important aspect in order to fully exploit ncAAs for structural purposes. Indeed, if protein structures are modified by the presence of non-natural moieties, the structural and dynamic information extracted for the protein from these probes can be biased. We have used a non-pathogenic HttExon-1 construct, which has been extensively characterized in the past^{57,60}, as a model system to record and structurally interpret the structural preferences of the three fluoroprolines and the chemical shift changes exerted on neighbouring glutamines. In line with previous measurements on model compounds and quantum chemistry calculations, we observed a stronger preference of the *cis* conformation for the 4S-FPro than for the 4R- and 4,4-FPro⁴⁶. Furthermore, our NMR data on isotopically enriched HttExon-1 forms unambiguously show that the number of fluorine atoms and the stereochemistry of the C^γ of prolines influence the chemical shift of the flanking glutamine. This observation holds for both *trans* and *cis* isomers, although the perturbation goes in opposite directions. The effect of fluorine vanishes for residues positioned five residues apart from the fluoroprolines, indicating that the extent of the perturbation is very local. We have performed extensive MD simulations using a state-of-the-art force field for disordered proteins to rationalize these observations and

to evaluate the origin of these changes. In all cases, experimentally observed strong chemical shift perturbations were associated with shorter distances between the fluorine and the atom probed by NMR. In very good agreement with our NMR analyses, our simulations show that distance distributions are strongly influenced by the isomeric form (*trans/cis*) and the pucker (*exo/endo*) of the proline. The comparison between the experimental data and the simulations strongly suggests that the fluoroprolines used in this study do not affect the structure of HttExon-1 and that the chemical shift changes observed are mainly due to the perturbation of the electronic environment induced by fluorine. Therefore, in the context of a disordered protein, the presence of fluorine in prolines does not strongly perturb the secondary structure of the region and chemical shift changes mainly originate from the proximity to fluorine. This observation underlines the use of fluorinated amino acids to monitor the intramolecular contacts in proteins or biomolecular interactions^{25,73}. A specially interesting application of the site-specific incorporation of fluoroprolines will be the study of the interactome of proline-rich proteins, for which high-resolution studies are extremely challenging. Note that poly-P tracts participate in protein-protein interaction networks, often through specific interactions with domains such as SRC homology 3 (SH3), the WW^{74,75} and the small actin-binding proteins profilins^{76,77}.

In summary, this study presents a chemical biology tool for the site-specific incorporation of commercially available fluoroprolines into proteins. The capacity to place one (or multiple) fluorine atoms in specific protein positions, in combination with tailored isotopic labelling schemes and without strongly modifying the structure, enables the measurement of precise NMR data reporting on intra- and intermolecular interactions.

Methods

Protein constructs

Plasmids were prepared as previously described⁵⁸. All synthetic genes of huntingtin exon1 (H16) were ordered from GeneArt (Thermo Fisher Scientific, Illkirch, France) and cloned into pIVEX 2.3d 3C-sfGFP-His₆. This includes the wild-type version and the genes encoding suppression stop codons. Amber stop codons (TAG) were used for single suppression samples at positions Gln33, Pro34 and Pro38. For double suppression samples, we tested all combinations of amber (TAG) and opal (TGA) codons in positions Gln29/Gln33 and Pro34 to evaluate the overall yield and the best codon arrangement.

Preparation of aminoacylated suppressor tRNA_{CUA}

Two pairs of tRNA_{CUA}/tRNA synthetase were prepared following previously described protocols to load amber stop codons with glutamine⁵⁸ and proline⁵⁷. The latter was also used to load 4S-FPro, 4R-FPro and 4,4-FPro. To use the opal stop codon, the glutamine and proline tRNA_{CUA} were modified to tRNA_{UCA} in order to recognise the TGA codon. Then, the same aminoacylation conditions as for tRNA_{CUA} were used. The successful loading of tRNA was confirmed by urea-PAGE (6.5% acrylamide 19:1, 8 M urea, 100 mM sodium acetate pH 5.2).

Standard cell-free expression

A lysate based on *Escherichia coli* strain BL21 Star (DE3)::RF1-CBD₃ was used for protein expression by CF reactions as previously described^{58,78}. The optimal tRNA concentrations for each fluorinated proline were determined by titrations of tRNA in 50 µL CF reactions incubated at 23 °C for 2-3 h. Then, the protein expression was monitored in a plate reader (Gen5 v3.03.14, BioTek Instruments, Colmar, France) by measuring sfGFP fluorescence (excitation at 485 nm, emission at 528 nm).

Preparation of NMR samples

Samples for NMR measurements were produced by 10 mL-CF reactions that were incubated at 23 °C and 450 rpm in a thermomixer for 3 h. Standard conditions were used as previously described⁶⁰. However, different labelling patterns required modifications in the amino acid mixtures used in each reaction. To uniformly label H16 samples, the standard amino acid mix was substituted by [¹⁵N, ¹³C]-labeled ISOGRO supplemented with [¹⁵N, ¹³C]-labeled Asn, Cys, and Trp (1 mM each) and 2 mM [¹⁵N, ¹³C]-Gln. For Gln33 single suppression and Gln33-Pro34 and Gln29-Pro34 double suppression samples, a non-labeled amino acid mixture was used. For

single Pro34 and Pro38 single suppression samples used for ^{19}F -NMR, a non-labelled amino acid mixture lacking glutamines was used and the CF reaction was supplemented with 2 mM [^{15}N , ^{13}C]-labeled Gln in order to monitor the stability of the sample via ^{15}N -HSQC spectra. In all suppression samples the corresponding suppressor tRNA was added at 10 μM or 20 μM for single or double suppression samples, respectively.

Prior to purification, samples were centrifuged and diluted 5-fold in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM imidazole. Proteins were then purified by affinity chromatography using a Ni gravity-flow column. After elution with 250 mM imidazole, samples were dialyzed overnight against NMR buffer (20 mM BisTris-HCl pH 6.5, 150 mM NaCl) and concentrated using 10 kDa MWCO Vivaspin centrifugal concentrators. Final protein concentrations were determined by fluorescence measurements in combination with a sfGFP calibration curve. Final NMR samples were supplemented with 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

NMR measurements

^{15}N and ^{13}C NMR experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe, operating at a ^1H frequency of 800 MHz. ^{15}N -HSQC and ^{13}C -HSQC were acquired for each sample in order to determine amide ($^1\text{H}_\text{N}$ and ^{15}N) and aliphatic ($^1\text{H}_\text{aliphatic}$ and $^{13}\text{C}_\text{aliphatic}$) chemical shifts, respectively. Spectra acquisition parameters were set up depending on the sample concentration. All spectra were processed with TopSpin v3.5 (Bruker Biospin) and analyzed using CCPN-Analysis software⁷⁹.

^{19}F -NMR experiments

All NMR samples were first concentrated up to a ca. 200 μL volume using Vivaspin centrifugal concentrators (Sartorius) with a 5 kDa MWCO at 4°C. 0.1 μL of a trimethylsilylpropanoic acid (TMSP) solution for chemical shift referencing and 10 μL of D_2O were added before the NMR measurement. All ^{19}F NMR experiments were performed on a Bruker Avance III HD spectrometer operating at a ^1H and ^{19}F frequencies of 600.13 MHz and 564.69 MHz, respectively, equipped with a CP-QCI-F cryoprobe with ^{19}F cryo-detection. All ^{19}F 1D experiments were performed at 293.0 K with ^1H decoupling during acquisition using waltz16 composite pulse decoupling. For all Pro34-modified samples, a ^{19}F spectral window of 20.1 ppm was used, with an acquisition time of 0.50 s and a relaxation delay of 1.0 s. For the Pro38-modified samples, a ^{19}F spectral window of 100.6 ppm was used, with an acquisition

time of 0.29 s and a relaxation delay of 0.5 s. For the Pro38-modified samples, concatenated 1D ^{19}F spectra of 128 transients each were acquired in order to monitor any spectral changes over time, which were not observed. These transients were then averaged to deliver the final spectrum. All ^{19}F spectra were referenced to the ^1H signal of TMSP using the unified chemical shift scale.

MD simulations and analyses.

Twenty structures of H16 (N17-polyQ₁₆-5P) were extracted from a ~ 20 μs long MD trajectory which was initiated from a coil conformation generated based on the Flexible-Meccano algorithm⁸⁰ using the ProtSA webserver⁸¹. To generate the corresponding *cis* conformers of Pro34, the Ω dihedral angle for the peptide bond connecting Gln33-Pro34 was set to 0° for each of the twenty structures in USCF Chimera⁸². Both *trans* and *cis* structures were solvated in octahedral boxes (edge length = 8.0 nm) with 150 mM NaCl, and additional counter ions added to neutralize the system. The system topology was modelled using the AMBER03ws force field⁶⁸ (https://bitbucket.org/jeetain/all-atom_ff_refinements/src/master/) and the TIP4P/2005 water model⁸³. Modified Lennard-Jones parameters proposed by Luo and Roux⁸⁴ were used for Na^+ and Cl^- ions to improve ion solubility.

The solvated systems were first subjected to energy minimization using the steepest descent algorithm in GROMACS-2020.4.⁸⁵ Following minimization, the systems were simulated for 100 ps using Nose-Hoover thermostat⁸⁶ ($\tau_c = 1$ ps) to attain a target temperature of 293.15 K. Following temperature equilibration, a 100 ps simulation was conducted using the Berendsen barostat⁸⁷ with isotropic coupling ($\tau_p = 5$ ps) for pressure control, to achieve a target pressure of 1 bar. Production simulations for each structure was performed for 1.0 μs in the NVT ensemble using the Langevin Middle Integrator⁸⁸ (friction coefficient = 1 ps^{-1}) within OpenMM-7.5.⁸⁹ Short-range nonbonded interactions were treated with a cut-off radius of 0.9 nm, while long-range electrostatics were computed using the Particle Mesh Ewald (PME) method.^{90,91} All bonds with hydrogen atoms were constrained using the SHAKE algorithm⁹² and hydrogen masses were increased by 1.5 times, allowing a simulation timestep of 4 fs.⁹³ α -helical fractions were computed based on the definition provided in the DSSP library⁹⁴ using the `gmx do_dssp` program. Gln29/Gln33-Pro34 distances and the Pro34 Altona-Sundaralingam P pseudorotation angle distributions for both *cis* and *trans* isomer trajectories were computed with custom python scripts using the MDanalysis library.⁹⁵

Acknowledgements

The authors thank Gottfried Otting (Australian National University, Canberra, Australia) for providing the BL21 (DE3) Star::RF1-CBD3 strain. This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020) / ERC Grant agreement n° [648030]. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively). J.M received funding from the National Institutes of Health (Grant No: NIGMS R35GM153388). D.S. acknowledges an « *accueil de talent* » grant from the Métropole Européenne de Lille (PUSHUP). The 600 MHz spectrometer for ¹⁹F NMR measurements is funded by the Nord Region Council, CNRS, Institut Pasteur de Lille, the European Community (ERDF), the French Ministry of Research and the Université de Lille and by the CTRL CPER co-funded by the European Union with the European Regional Development Fund (ERDF), by the Hauts-de-France Regional Council (contract n°17003781), Métropole Européenne de Lille (contract n°2016_ESR_05), and French State (contract n°2017-R3-CTRL-Phase1). Atomistic MD simulations were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

References

- 1 T. A. Cropp and P. G. Schultz, *Trends in Genetics*, 2004, 20, 625–630.
- 2 L. Wang, J. Xie and P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 225–249.
- 3 M. Kimoto and I. Hirao, *Front Mol Biosci*, 2022, 9.
- 4 L. Wang and P. G. Schultz, *Angewandte Chemie - International Edition*, 2004, 44, 34–66.
- 5 C. C. Liu and P. G. Schultz, *Annu Rev Biochem*, 2010, **79**, 413–44.
- 6 H. T. More, K. S. Zhang, N. Srivastava, J. A. Frezzo and J. K. Montclare, *Biomacromolecules*, 2015, **16**, 1210–1217.
- 7 T. Lugtenburg, A. Gran-Scheuch and I. Drienovská, *Protein Engineering, Design and Selection*, 2023, **36**, gzad003.
- 8 S. Peuker, H. Andersson, E. Gustavsson, K. S. Maiti, R. Kania, A. Karim, S. Niebling, A. Pedersen, M. Erdelyi and S. Westenhoff, *J Am Chem Soc*, 2016, **138**, 2312–2318.
- 9 A. J. Pell, G. Pintacuda and C. P. Grey, *Prog Nucl Magn Reson Spectrosc*, 2019, 111, 1–271.
- 10 Q. Miao, C. Nitsche, H. Orton, M. Overhand, G. Otting and M. Ubbink, *Chem Rev*, 2022, 122, 9571–9642.
- 11 A. Boeszoermyeni, B. Ogórek, A. Jain, H. Arthanari and G. Wagner, *J Biomol NMR*, 2020, **74**, 365–379.
- 12 J. L. Kitevski-LeBlanc and R. S. Prosser, *Prog Nucl Magn Reson Spectrosc*, 2012, 62, 1–33.
- 13 M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. Ye and B. Kokscha, *Chem Soc Rev*, 2012, **41**, 2135–2171.
- 14 B. D. Sykes and W. E. Hull, *Methods Enzymol*, 1978, **49**, 270–295.
- 15 E. N. G. Marsh and Y. Suzuki, *ACS Chem Biol*, 2014, 9, 1242–1250.
- 16 A. M. Gronenborn, *Structure*, 2022, 30, 6–14.
- 17 W. C. Isley, A. K. Urick, W. C. K. Pomerantz and C. J. Cramer, *Mol Pharm*, 2016, **13**, 2376–2386.
- 18 S. V. Fedorov and L. B. Krivdin, *J Fluor Chem*, 2020, **238**, 109625.
- 19 C. Di Pietrantonio, A. Pandey, J. Gould, A. Hasabnis and R. S. Prosser, in *Methods in Enzymology*, 2019, vol. 615, pp. 103–130.
- 20 J. Klein-Seetharaman, E. V. Getmanova, M. C. Loewen, P. J. Reeves and H. G. Khorana, *Proc Natl Acad Sci U S A*, 1999, **96**, 13744–13749.
- 21 P. A. Luchette, R. S. Prosser and C. R. Sanders, *J Am Chem Soc*, 2002, **124**, 1778–1781.
- 22 A. Manglik, T. H. Kim, M. Masureel, C. Altenbach, Z. Yang, D. Hilger, M. T. Lerch, T. S. Kobilka, F. S. Thian, W. L. Hubbell, R. S. Prosser and B. K. Kobilka, *Cell*, 2015, **161**, 1101–1111.
- 23 N. K. Mishra, A. K. Urick, S. W. J. Ember, E. Schonbrunn and W. C. Pomerantz, *ACS Chem Biol*, 2014, **9**, 2755–2760.
- 24 A. Boeszoermyeni, S. Chhabra, A. Dubey, D. L. Radeva, N. T. Burdzhiev, C. D. Chaney, O. I. Petrov, V. M. Gelev, M. Zhang, C. Anklin, H. Kovacs, G. Wagner, I. Kuprov, K. Takeuchi and H. Arthanari, *Nat Methods*, 2019, **16**, 333–340.
- 25 Y. J. Tan, E. H. Abdelkader, E. Tarcoveanu, A. Maleckis, C. Nitsche and G. Otting, *Biochemistry*, 2024, **63**, 1376–1387.
- 26 F. Agostini, L. Sinn, D. Petras, C. J. Schipp, V. Kubyshkin, A. A. Berger, P. C. Dorrestein, J. Rappsilber, N. Budisa and B. Kokscha, *ACS Cent Sci*, 2021, **7**, 81–92.
- 27 C. Jäckel and B. Kokscha, *European J Org Chem*, 2005, 4483–4503.
- 28 V. Y. Torbeev and D. Hilvert, *Proc Natl Acad Sci U S A*, 2013, **110**, 20051–20056.

- 29 S. E. Cellitti, D. H. Jones, L. Lagpacan, X. Hao, Q. Zhang, H. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz and B. H. Geierstanger, *J Am Chem Soc*, 2008, **130**, 9268–9281.
- 30 J. C. Jackson, J. T. Hammill and R. A. Mehl, *J Am Chem Soc*, 2007, **129**, 1160–1166.
- 31 F. Zhang, Q. Zhou, G. Yang, L. An, F. Li and J. Wang, *Chemical Communications*, 2018, **54**, 3879–3882.
- 32 V. J. Somovilla, I. A. Bermejo, I. S. Albuquerque, N. Martínez-Sáez, J. Castro-López, F. García-Martín, I. Compañón, H. Hinou, S.-I. Nishimura, J. Jiménez-Barbero, J. L. Asensio, A. Avenoza, J. H. Busto, R. Hurtado-Guerrero, J. M. Peregrina, G. J. L. Bernardes and F. Corzana, *J Am Chem Soc*, 2017, **139**, 18255–18261.
- 33 C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angewandte Chemie International Edition*, 2001, **40**, 923–925.
- 34 S. K. Holmgren, K. M. Taylor, L. E. Bretscher and R. T. Raines, *Nature*, 1998, **392**, 666–667.
- 35 S. J. M. Verhoork, P. M. Killoran and C. R. Coxon, *Biochemistry*, 2018, **57**, 6132–6143.
- 36 M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines and J. L. Markley, *J Am Chem Soc*, 2002, **124**, 2497–2505.
- 37 J. A. Hodges and R. T. Raines, *J Am Chem Soc*, 2005, **127**, 15923–15932.
- 38 A. K. Pandey, D. Naduthambi, K. M. Thomas and N. J. Zondlo, *J Am Chem Soc*, 2013, **135**, 4333–4363.
- 39 R. W. Newberry and R. T. Raines, in *Top. Heterocycl. Chem.*, 2016, pp. 1–25.
- 40 V. Kubyshkin, R. Davis and N. Budisa, *Beilstein Journal of Organic Chemistry*, 2021, **17**, 439–460.
- 41 V. Kubyshkin and M. Rubini, *Chem Rev*, 2024, **124**, 8130–8232.
- 42 R. W. Newberry and R. T. Raines, *Acc Chem Res*, 2017, **50**, 1838–1846.
- 43 J.-C. Horng and R. T. Raines, *Protein Science*, 2006, **15**, 74–83.
- 44 D. Sinnaeve, A. Ben Bouzayene, E. Ottoy, G. J. Hofman, E. Erdmann, B. Linclau, I. Kuprov, J. C. Martins, V. Torbeev and B. Kieffer, *Magnetic Resonance*, 2021, **2**, 795–813.
- 45 M. D. Shoulders, K. J. Kamer and R. T. Raines, *Bioorg Med Chem Lett*, 2009, **19**, 3859–3862.
- 46 G.-J. Hofman, E. Ottoy, M. E. Light, B. Kieffer, I. Kuprov, J. C. Martins, D. Sinnaeve and B. Linclau, *Chem. Commun.*, 2018, **54**, 5118–5121.
- 47 H. K. Ganguly, B. A. Ludwig, C. M. Tressler, M. R. Bhatt, A. K. Pandey, C. M. Quinn, S. Bai, G. P. A. Yap and N. J. Zondlo, *Biochemistry*, 2024, **63**, 1131–1146.
- 48 D. Chen, K. W. Drombosky, Z. Hou, L. Sari, O. M. Kashmer, B. D. Ryder, V. A. Perez, D. N. R. Woodard, M. M. Lin, M. I. Diamond and L. A. Joachimiak, *Nat Commun*, 2019, **10**, 2493.
- 49 S. L. Breunig and D. A. Tirrell, in *Methods in Enzymology*, 2021, vol. 656, pp. 545–571.
- 50 M. D. Crespo and M. Rubini, *PLoS One*, 2011, **6**, e19425.
- 51 C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angewandte Chemie International Edition*, 2001, **40**, 923–925.
- 52 J. O’ Loughlin, S. Napolitano and M. Rubini, *Chembiochem*, 2021, **22**, 3326–3332.
- 53 M. Rubini, M. A. Schärer, G. Capitani and R. Glockshuber, *Chembiochem*, 2013, **14**, 1053–7.
- 54 T. Steiner, P. Hess, J. H. Bae, B. Wiltschi, L. Moroder and N. Budisa, *PLoS One*, 2008, **3**, e1680.
- 55 B. Holzberger and A. Marx, *J Am Chem Soc*, 2010, **132**, 15708–15713.

- 56 M. S. Lukesch, T. Pavkov-Keller, K. Gruber, K. Zangger and B. Wiltschi, *Sci Rep*, 2019, **9**, 2697.
- 57 A. Urbanek, M. Popovic, C. A. Elena-Real, A. Morató, A. Estaña, A. Fournet, F. Allemand, A. M. Gil, C. Cativiela, J. Cortés, A. I. Jiménez, N. Sibille and P. Bernadó, *J Am Chem Soc*, 2020, **142**, 7976–7986.
- 58 A. Urbanek, A. Morató, F. Allemand, E. Delaforge, A. Fournet, M. Popovic, S. Delbecq, N. Sibille and P. Bernadó, *Angew Chem Int Ed Engl*, 2018, **57**, 3598–3601.
- 59 A. Bhattacharyya, A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer and R. Wetzel, *J Mol Biol*, 2006, **355**, 524–535.
- 60 A. Urbanek, M. Popovic, A. Morató, A. Estaña, C. A. Elena-Real, P. Mier, A. Fournet, F. Allemand, S. Delbecq, M. A. Andrade-Navarro, J. Cortés, N. Sibille and P. Bernadó, *Structure*, 2020, **28**, 733-746.e5.
- 61 A. Chatterjee, H. Xiao and P. G. Schultz, *Proc Natl Acad Sci U S A*, 2012, **109**, 14841–6.
- 62 C. A. Elena-Real, A. Urbanek, L. Imbert, A. Morató, A. Fournet, F. Allemand, N. Sibille, J. Boissouvier and P. Bernadó, *ACS Chem Biol*.
- 63 C. Siebler, B. Maryasin, M. Kuemin, R. S. Erdmann, C. Rigling, C. Grünenfelder, C. Ochsenfeld and H. Wennemers, *Chem. Sci.*, 2015, **6**, 6725–6730.
- 64 M. Baias, P. E. S. Smith, K. Shen, L. A. Joachimiak, S. Žerko, W. Koźmiński, J. Frydman and L. Frydman, *J Am Chem Soc*, 2017, **139**, 1168–1176.
- 65 D. L. Dunkelmann, J. C. W. Willis, A. T. Beattie and J. W. Chin, *Nat Chem*, 2020, **12**, 535–544.
- 66 C. Kofman, J. Lee and M. C. Jewett, *Cell Syst*, 2021, **12**, 593–607.
- 67 K. V Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa and G. Otting, *Angew Chem Int Ed Engl*, 2012, **51**, 2243–2246.
- 68 R. B. Best, W. Zheng and J. Mittal, *J Chem Theory Comput*, 2014, **10**, 5113–5124.
- 69 C. Altona and M. Sundaralingam, *J Am Chem Soc*, 1972, **94**, 8205–12.
- 70 F. A. A. M. de Leeuw, P. N. Van Kampen, C. Altona, E. Díez and A. L. Esteban, *J Mol Struct*, 1984, **125**, 67–88.
- 71 G.-J. Hofman, E. Ottoy, M. E. Light, B. Kieffer, J. C. Martins, I. Kuprov, D. Sinnaeve and B. Linclau, *J Org Chem*, 2019, **84**, 3100–3120.
- 72 A. Testa, X. Lucas, G. V Castro, K.-H. Chan, J. E. Wright, A. C. Runcie, M. S. Gadd, W. T. A. Harrison, E.-J. Ko, D. Fletcher and A. Ciulli, *J Am Chem Soc*, 2018, **140**, 9299–9313.
- 73 C. A. Elena-Real, A. Sagar, A. Urbanek, M. Popovic, A. Morató, A. Estaña, A. Fournet, C. Doucet, X. L. Lund, Z. D. Shi, L. Costa, A. Thureau, F. Allemand, R. E. Swenson, P. E. Milhiet, R. Crehuet, A. Barducci, J. Cortés, D. Sinnaeve, N. Sibille and P. Bernadó, *Nat Struct Mol Biol*, 2023, **30**, 309–320.
- 74 M. Mompeán, J. Oroz and D. V Laurents, *FEBS Open Bio*, 2021, **11**, 2390–2399.
- 75 A. A. Morgan and E. Rubenstein, *PLoS One*, 2013, **8**, e53785-.
- 76 A. E. Posey, K. M. Ruff, T. S. Harmon, S. L. Crick, A. Li, M. I. Diamond and R. V Pappu, *Journal of Biological Chemistry*, 2018, **293**, 3734–3746.
- 77 K. Krishnan and P. D. J. Moens, *Biophys Rev*, 2009, **1**, 71–81.
- 78 A. Morató, C. A. Elena-Real, M. Popovic, A. Fournet, K. Zhang, F. Allemand, N. Sibille, A. Urbanek and P. Bernadó, *Biomolecules*, , DOI:10.3390/biom10101458.
- 79 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687–696.
- 80 V. Ozenne, F. Bauer, L. Salmon, J. -r. Huang, M. R. Jensen, S. Segard, P. Bernado, C. Charavay and M. Blackledge, *Bioinformatics*, 2012, **28**, 1463–1470.

- 81 J. Estrada, P. Bernadó, M. Blackledge and J. Sancho, *BMC Bioinformatics*, 2009, **10**, 104.
- 82 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605–1612.
- 83 J. L. F. Abascal and C. Vega, *J Chem Phys*, 2005, **123**, 234505.
- 84 Y. Luo and B. Roux, *J Phys Chem Lett*, 2010, **1**, 183–189.
- 85 S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess and E. Lindahl, *J Chem Phys*, 2020, **153**, 134110.
- 86 D. J. Evans and B. L. Holian, *J Chem Phys*, 1985, **83**, 4069–4074.
- 87 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J Chem Phys*, 1984, **81**, 3684–3690.
- 88 Z. Zhang, X. Liu, K. Yan, M. E. Tuckerman and J. Liu, *J Phys Chem A*, 2019, **123**, 6056–6079.
- 89 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLoS Comput Biol*, 2017, **13**, e1005659.
- 90 T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089–10092.
- 91 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J Chem Phys*, 1995, **103**, 8577–8593.
- 92 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J Comput Phys*, 1977, **23**, 327–341.
- 93 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, *J Chem Theory Comput*, 2015, **11**, 1864–1874.
- 94 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- 95 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J Comput Chem*, 2011, **32**, 2319–2327.