Site-specific introduction of alanines for the nuclear magnetic resonance investigation of low-complexity regions and large biomolecular assemblies

Carlos A Elena-Real, Annika Urbanek, Lionel Imbert, Anna Morató, Aurélie Fournet, Frédéric Allemand, Nathalie Sibille, Jérôme Boisbouvier, Pau Bernadó

pubs.acs.org/acschemicalbiology

Articles

Site-Specific Introduction of Alanines for the Nuclear Magnetic Resonance Investigation of Low-Complexity Regions and Large Biomolecular Assemblies

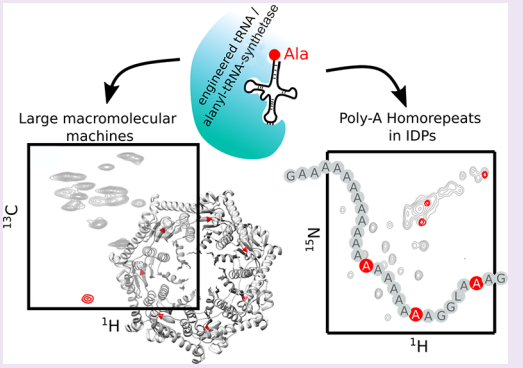
Carlos A. Elena-Real, Annika Urbanek, Lionel Imbert, Anna Morató, Aurélie Fournet, Frédéric Allemand, Nathalie Sibille, Jérôme Boisbouvier,* and Pau Bernadó*

Cite This: https://doi.org/10.1021/acschembio.3c00288

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nuclear magnetic resonance (NMR) studies of large biomolecular machines and highly repetitive proteins remain challenging due to the difficulty of assigning frequencies to individual nuclei. Here, we present an efficient strategy to address this challenge by engineering a *Pyrococcus horikoshii* tRNA/alanyl-tRNA synthetase pair that enables the incorporation of up to three isotopically labeled alanine residues in a site-specific manner using in vitro protein expression. The general applicability of this approach for NMR assignment has been demonstrated by introducing isotopically labeled alanines into four distinct proteins: huntingtin exon-1, HMA8 ATPase, the 300 kDa molecular chaperone ClpP, and the alanine-rich Phox2B transcription factor. For large protein assemblies, our labeling approach enabled unambiguous assignments while avoiding potential artifacts induced by site-specific mutations. When applied to Phox2B, which contains two poly-alanine tracts of nine and twenty alanines, we observed that the helical stability is strongly dependent on the homorepeat length. The capacity to selectively introduce alanines with distinct labeling patterns is a powerful tool to probe structure and dynamics of challenging biomolecular systems.



INTRODUCTION

Nuclear magnetic resonance (NMR) has become a fundamental tool for the high-resolution structural, dynamic, and functional characterization of biomolecules in solution. Indeed, NMR is nowadays routinely applied for the investigation of all types of biomolecules from very large biomolecular assemblies (up to 1 MDa) to intrinsically disordered proteins (IDPs) and nucleic acids.[1−4] However, the power of NMR to reveal this detailed information relies on the capacity to identify resonance frequencies of individual nuclei. In this context, the availability of high magnetic fields in combination with multidimensional pulse sequences and isotopic labeling strategies has reduced the signal overlap of the resulting spectra, facilitating the subsequent frequency assignment.[5−7] Well-established procedures have been designed for the unambiguous frequency assignment of backbone nuclei of small- and medium-size proteins (up to ≈80 kDa). All these strategies are based on the specific spectroscopic features of natural amino acids and their connectivity, which emerge through couples of multidimensional heteronuclear experiments.[8] Therefore, the absence of signals for certain nuclei, due, for instance, to chemical exchange processes, or the degeneracy of two or more frequencies can hamper the assignment process. A clear example of severe signal overlap is low-complexity regions (LCRs) in proteins, which present a strong compositional bias that reduces the diversity of chemical environments and, as a consequence, the chemical shift dispersion.[9] The site-specific NMR frequency assignment of large biomolecular assemblies is also extremely challenging. Due to their slow tumbling in solution, most of the backbone signals broaden beyond detection, even when using high temperatures and/or perdeuteration. The structural and dynamic investigation of these macromolecular machines has been accomplished by using methyl-NMR.[10−14] The fast rotation of the methyl group and the presence of three equivalent hydrogen atoms make methyl-NMR experiments especially suited to study large biomolecular assemblies. In order to obtain appropriate samples for these experiments, one or a few methyl-containing amino acids (leucine, isoleucine, valine, alanine, threonine, and methionine) are incorporated in their $^{13}CH_3$-labeled form, while the rest of the protein is deuterated.[15−17] This labeling scheme can be achieved by

Received: May 16, 2023
Accepted: August 1, 2023

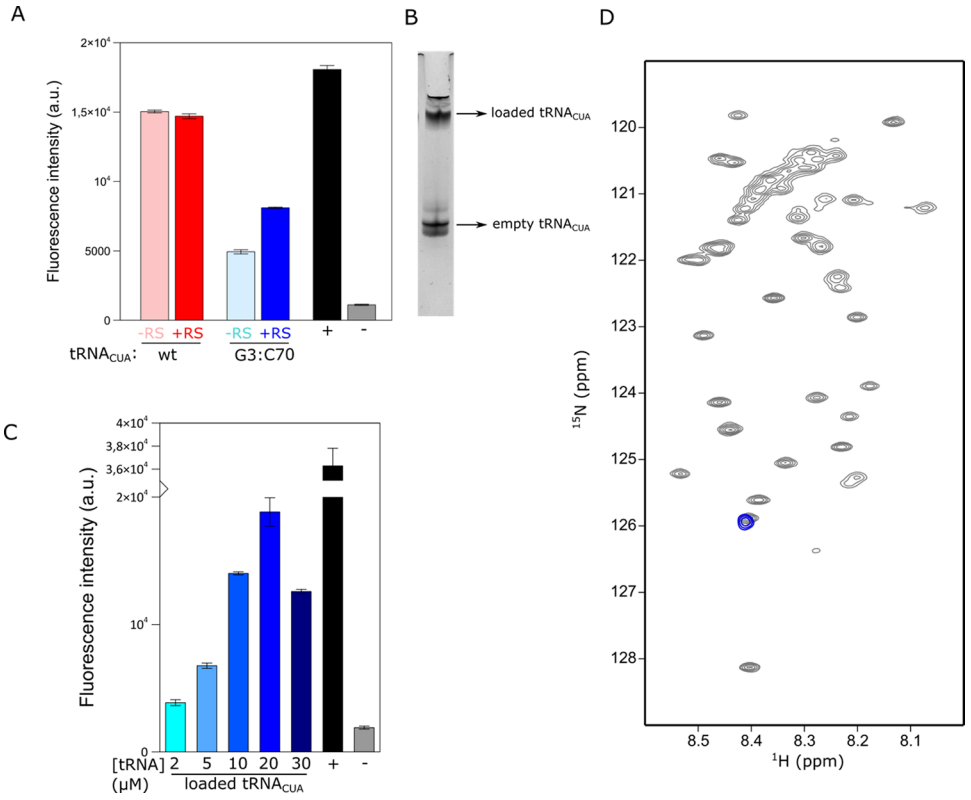


Figure 1. (A) Endpoint fluorescence intensity plot for CF expression of HttExon1-A53 using wt or G3:C70 $tRNA_{CUA}$ and in the absence or presence of AlaRS. + and − indicate the positive (wild-type HttExon1) and negative (HttExon1-A53 without $tRNA_{CUA}$) controls. (B) Urea-PAGE monitoring the efficiency (∼50%) of the suppressor $tRNA_{CUA}$ loading. (C) Titration of CF reactions with increasing concentrations of alanine-loaded $tRNA_{CUA}$. (D) ^{15}N-HSQC spectrum of HttExon1 fully labeled (gray) and HttExon1-A53 SSIL sample (blue) where a single peak, corresponding to A53, was observed.

supplying the bacterial culture with the appropriate isotopologues or with molecules that are precursors in the enzymatic chain of the amino acid synthesis.[18−21] However, in methyl-labeled samples, traditional sequential assignment strategies are not possible. For these large machines, mutagenesis-based[22,23] or "divide-and-conquer"[14,24] approaches are normally applied. These are laborious approaches that are prone to errors caused by local structural perturbations induced by mutations or changes in protein constructs.[22,23]

In order to overcome the above-mentioned limitations, we have developed the site-specific incorporation of up to three isotopically labeled alanines into proteins to be applied in large macromolecular machines and LCRs. To achieve this aim, we combined tRNA nonsense suppression with cell-free (CF) protein synthesis.[25] We show that the application of this strategy enormously reduces the complexity of the recorded spectra, enabling the study of biomolecular systems not amenable to traditional NMR approaches. Concretely, we have demonstrated our approach for the tetradecameric ClpP protease and the disordered alanine-rich C-terminal region of the Phox2B transcription factor. These proteins represent prototypical examples of very large macromolecular machines and highly degenerated sequences, respectively. On the one hand, ClpP protomers self-assemble into two stacked heptameric rings to form a 300 kDa barrel-shaped serine protease containing a single degradation chamber involving all 14 copies of the catalytic triad. These proteases play an important role in protein homeostasis and are conserved in bacteria, chloroplasts, and mitochondria.[26,27] On the other hand, the C-terminal region of Phox2B is an example of a LCR rich in alanines, glycines, serines, and prolines. It contains two poly-alanine (poly-A) tracts of 9 and 20 alanines, which is the longest one found in the human proteome.[28,29] Interestingly, the expansion of the long poly-A tract with +5, +7, or +11 alanines causes a severe developmental pathology, the congenital central hypoventilation syndrome (CCHS).[30,31] The structural changes occurring in Phox2B upon the alanine expansion that trigger the pathology remain unknown, equivalently to eight other rare diseases linked to the abnormal expansion of poly-A tracts.[29,32]

RESULTS AND DISCUSSION

Development of an Efficient Nonsense Suppressor for Alanine Site-Specific Labeling. The site-specific isotopic labeling (SSIL) approach uses a synthetic nonsense suppressor tRNA ($tRNA_{CUA}$) and aminoacyl-tRNA synthetase (aaRS) pair with a maximum orthogonality with respect to the *Escherichia coli* translation system.[25] $tRNA_{CUA}$ enables the insertion of a labeled amino acid specifically at the position encoded by an amber stop codon, while its orthogonality to endogenous tRNA/aaRS pairs prevents its reloading with the non-labeled amino acid present in the CF reaction. To incorporate alanines, a previously characterized tRNA/alanyl-RS (AlaRS) pair from *Pyrococcus horikoshii* was chosen[33,34] with the anticodon mutated to obtain the $tRNA_{CUA}$. First, the capacity of the recombinant AlaRS to load the synthetic $tRNA_{CUA}$ with alanine was evaluated using urea-PAGE (see Methods section). When optimized conditions were used, an aminoacylation level of ∼50% was achieved (see Figure S1 in

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

Supporting Information). The orthogonality of this pair was subsequently tested by monitoring the CF expression of huntingtin exon-1 protein (UniProt ID P42858) containing 16 consecutive glutamines N-terminally fused to a superfolder green fluorescent protein and a His-TAG (HttExon1), a protein construct that we have extensively used to test the SSIL strategy.[25,35] For this specific application, the A53 codon (GCA) was exchanged by the amber stop codon (TAG) (HttExon1-A53). When the $tRNA_{CUA}$ was used in the absence of the AlaRS, the protein was synthesized at similar levels as in the positive control, without the amber stop codon (Figure 1A). This observation indicated that the $tRNA_{CUA}$ could be loaded by the endogenous *E. coli* AlaRS, showing that the *P. horikoshii* $tRNA_{CUA}$ was not orthogonal to the *E. coli* translational machinery.

In order to improve the required orthogonality, we engineered the *P. horikoshii* pair. Previous studies highlighted the relevance of the highly conserved G3:U70 base pair located in the acceptor stem of the $tRNA^{Ala}$ for the selective aminoacylation with alanine.[36,37] In a recent study, it was shown that the substitution of this base pair by G3:C70 impaired alanine loading by *E. coli* AlaRS.[38] Interestingly, the efficient aminoacylation of the G3:C70 $tRNA^{Ala}$ was recovered when highly conserved Asn303 and Asp400 residues in *E. coli* AlaRS were mutated to alanine. Based on these observations and the sequence alignment of several AlaRS (Figure S2), we designed a new *P. horikoshii* $tRNA_{CUA}$/AlaRS pair containing both modifications. On the one hand, the G3:C70 mutation in the $tRNA_{CUA}$ would hamper the recognition by *E. coli* AlaRS. On the other hand, the introduction of two alanines at positions corresponding to Asn360 and Asp459 in the *P. horikoshii* AlaRS would enable an efficient external loading of the $tRNA_{CUA}$. Using similar aminoacylation conditions as for the wild-type pair (see Methods section), ∼50% of loaded $tRNA_{CUA}$ was obtained, suggesting a minimal impact of the designed mutations in the *P. horikoshii* pair (Figure 1B). Importantly, in the absence of the mutant AlaRS, a substantial reduction of HttExon1-A53 CF expression was observed (Figure 1A), indicating an improvement in the $tRNA_{CUA}$ orthogonality upon introducing the G3:C70 base pair. Addition of the mutant AlaRS to the CF reaction increased the protein expression, evidencing the recognition of suppressor $tRNA_{CUA}$ by the enzyme under the experimental conditions, which notably differ from those found for the optimal external loading. When increasing amounts of aminoacylated $tRNA_{CUA}$ were added to the CF reaction, protein synthesis improved, reaching its maximum at 20 μM loaded $tRNA_{CUA}$ (Figure 1C). As previously observed,[25] further increasing the $tRNA_{CUA}$ concentration compromised the protein synthesis. In order to validate our approach for NMR studies, a 5 mL CF reaction was prepared using 10 μM [^{13}C,^{15}N]-alanine-loaded $tRNA_{CUA}$ to express HttExon1-A53 (Table S1). After purification, the ^{15}N-HSQC spectrum showed a single correlation that overlapped with that previously assigned for A53 using multidimensional NMR (Figure 1D).[39] Given the requirement of the systematic application of SSIL for the NMR characterization of proteins, it is important to evaluate the cost of the CF production. We have estimated the material cost of a SSIL sample for NMR obtained from 5 mL of CF using our own lysate to be ≈40 Euros. Note that two thirds of this price (25 Euros) correspond to the production of the aminoacylated $tRNA_{CUA}$ and the rest to the batch CF mixture. The inclusion of isotopically labeled amino acids at 2 mM in the CF mixture slightly increases the price by around 3 Euros per labeled amino acid.

Quantitative Orthogonality Estimation of the Mutated *P. horikoshii* $tRNA_{CUA}$/AlaRS Pair. The ability to produce isotopically labeled samples in a site-specific manner allowed the quantification of the orthogonality improvement when using the mutated *P. horikoshii* $tRNA_{CUA}$/AlaRS couple. For this, we used the N-terminally His-tagged nucleotide-binding domain (NBD) of the HMA8 ATPase[40] (UniProt ID B9DFX7, fragment 557-689), which we have used for CF methodological developments in the past.[41] Concretely, by introducing the amber stop codon in position A619 (originally GCA) of the coding gene, two SSIL samples of this protein were prepared using either the wild-type or the G3:C70 mutated $tRNA_{CUA}$ aminoacylated with $^{13}CH_3$-labeled alanine (Table S1). A third sample with all (12) alanine methyl sites labeled was produced using standard CF incorporating $^{13}CH_3$-labeled alanine and used as a reference for complete isotopic incorporation. To quantify the level of incorporation of $^{13}CH_3$-labeled alanine at the amber codon-encrypted site, the three samples were produced in the presence of $^{13}CH_3$-labeled methionine, and the average intensity measured for the two non-assigned methionine peaks (I_{Met}) was used as an internal reference (Figure 2A). An enormous decrease in the peak intensity ratio I_{A619}/I_{Met} of the SSIL sample using the wild-type $tRNA_{CUA}$ (Figure 2B) was observed when compared with the fully labeled sample (Figure 2A). This observation indicated that the majority of the protein produced in these conditions contained non-isotopically enriched alanine, underlining the lack of orthogonality of wild-type $tRNA_{CUA}$. Interestingly, for the SSIL sample produced with the aminoacylated G3:C70-mutated $tRNA_{CUA}$ (Figure 2C), the I_{A619}/I_{Met} ratio was increased by a factor of 3.2 with respect to the wild-type (Figure 2B). This observation demonstrated the reduced $tRNA_{CUA}$ aminoacylation with non-labeled alanine from the CF mixture by endogenous *E. coli* AlaRS. Comparison with the spectra acquired for the fully labeled sample (Figure 2A) enabled us to quantify the level of incorporation to be 60.2% using G3:C70-mutated $tRNA_{CUA}$. Although the mutated $tRNA_{CUA}$ was not fully orthogonal, it enabled the systematic application of our strategy for the NMR assignment of challenging biomolecular systems.

Application of the Alanine SSIL Approach to ClpP, a 300 kDa Chaperone. The above-described results validated the use of the engineered *P. horikoshii* $tRNA_{CUA}$/AlaRS pair as a tool to isotopically label single alanines in a site-specific manner. Next, we applied this technology to study the *Thermus thermophilus* ClpP (UniProt ID Q5SKM8). First, we analyzed the signal dispersion of alanines in this macromolecular complex. For this, ClpP was produced *in vitro* in a D_2O-based buffer and in the presence of a mixture of perdeuterated amino acids, supplemented with an excess of [2-^{2}H,3-^{13}C]-L-alanine. This procedure yielded an optimally labeled sample for the acquisition of high-resolution 2D-methyl-TROSY experiments (Figure 3A).[42,43] Although most of the 24 alanine methyl signals were well resolved, the characteristically slow overall tumbling of this 300 kDa particle precluded the acquisition of triple-resonance experiments necessary for sequence-specific assignment. To test the SSIL strategy in the context of a large protein assembly, we targeted two alanines with expectedly different features. On the one hand, A96 is located in a globular section of the complex, in close

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

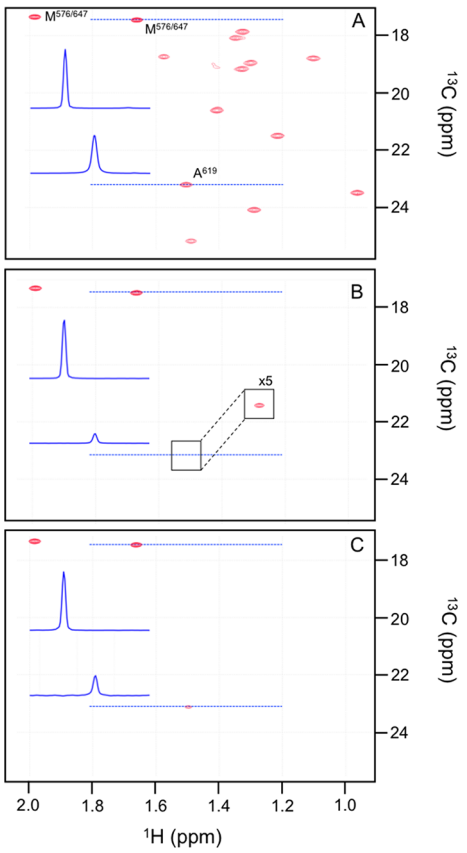


Figure 2. 2D methyl-TROSY spectra of the NBD domain of the HMA8 ATPase with (A) non-assigned methionine (2) and alanine (12) residues $^{13}CH_3$-labeled. For spectra presented in (B) and (C), samples were $^{13}CH_3$-labeled on methionines and site-specifically labeled on A619 using wild-type $tRNA_{CUA}$ (B) or G3:C70-mutated $tRNA_{CUA}$ (C). The first contours of all spectra are displayed at a level corresponding to 10% of the most intense methionine signal, except for the insert in (B), which displays spectra at a 5-fold lower contour level. The 1D traces for a reference methionine signal and A619 are displayed in blue in each panel.

proximity of the active site. On the other hand, a more flexible behavior was expected for A11, which is involved in ClpP gating and target recognition.[10] To identify these two residues, we used the above-described nonsense suppressor $tRNA_{CUA}$ loaded with [2-^{2}H,3-^{13}C]-L-alanine to prepare two ClpP samples. A well-resolved alanine signal was observed in both cases (Figure 3B,C), allowing direct assignment of the two methyl probes. The low intensity observed for the A11 methyl signal is presumably due to an exchange between different conformations of the N-terminal gate segment of ClpP.[44] Note that this approach does not suffer from the secondary chemical shift (SCS) perturbation as observed using mutagenesis-based approaches.[22]

Resolving the Poly-A Tracts in Phox2B. In order to validate our strategy for alanine-rich LCRs, we applied the alanine SSIL strategy to investigate the two poly-A tracts present at the C-terminal domain of Phox2B (UniProt ID Q99453, residues 136−314). This 179-residue long fragment is a LCR highly enriched in alanines (45), glycines (33), prolines (19), and serines (17). Despite the fact that the resulting ^{15}N-HSQC spectrum of this fragment N-terminally fused to sfGFP exhibited a large region with overlapped frequencies, several resolved peaks (98) could be identified (Figure S4). Note that the ^{15}N-HSQC spectrum was recorded at 800 MHz, suggesting that the use of higher magnetic fields would not solve this massive overlap. When the protein was produced in CF using [^{13}C,^{15}N]-L-alanine as the only isotopically enriched amino acid, the ^{15}N-HSQC exhibited multiple alanine signals, but many of them clustered in a narrow region of the spectrum (Figure S4). Similarly, the Cα-Hα and Cβ-Hβ regions of the ^{13}C-HSQC displayed few broad connected signals, but no isolated peaks could be identified (Figure S4). The ensemble of these observations indicated that traditional frequency assignment with multidimensional NMR spectra was not applicable to characterize this protein.

In order to overcome the severe overlap, we applied our strategy to study two alanines located at the center of the short (A163) and long (A250) poly-A tracts of Phox2B (Figure 4A). Two NMR samples were prepared by introducing a

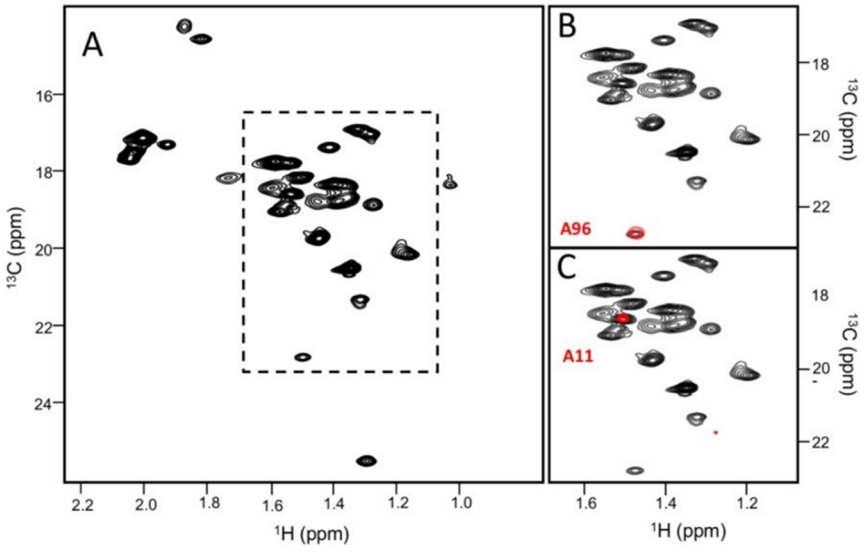


Figure 3. 2D methyl-TROSY of the 300 kDa ClpP. (A) Reference spectrum of perdeuterated ClpP where all alanines and methionines were $^{13}CH_3$-labeled. Zoom corresponding to the dashed delimited reference spectrum overlaid with SSIL spectra (red) where a single alanine methyl signal was observed for A96 (B) and A11 (C).

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

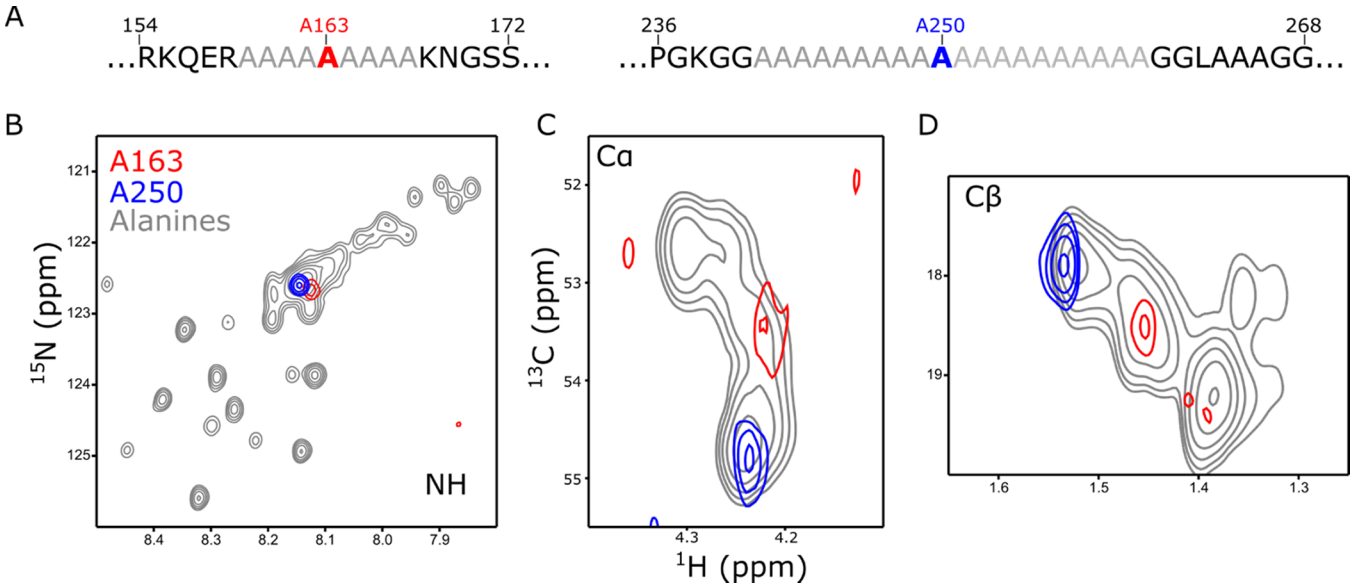


Figure 4. (A) Sequences of short (left) and long (right) poly-A stretches present in the Phox2B C-terminal domain. Alanines probed by SSIL are highlighted. Overlays of the single alanine ^{15}N-HSQC and ^{13}C-HSQC spectra of the Phox2B C-terminal domain measured for samples with all alanines ^{15}N,^{13}C-labeled (gray), showing NH (B), Cα-Hα (C), and Cβ-Hβ (D) correlations. Red (A163) and blue (A250) colors are consistent in all panels.

[^{13}C, ^{15}N]-L-alanine in these two positions while keeping all the other amino acids unlabeled. Note that deuterated alanine was used in the CF reaction in order to avoid spurious signals arising from natural abundance.[35] Well-resolved signals were obtained for N-H, Cα-Hα, and Cβ-Hβ correlations (Figure 4B−D), demonstrating the power of SSIL to resolve degenerated frequencies in homorepeats. However, despite the concentration of both NMR samples being similar (see Table S1), the NMR intensities for A163 were only slightly above the noise level, while A250 exhibited more intense peaks. We attributed this difference to the presence of intermediate conformational exchange processes in the short poly-A tract that increased the transversal relaxation rates. Interestingly, despite the fact that both alanines, A163 and A250, experience the same sequence context, they displayed different chemical shifts for the three studied correlations. Indeed, the largest differences between both alanines were observed for the Cα and Cβ frequencies, which are the most sensitive atoms to secondary structure propensities. To further demonstrate this point, SCS analysis was performed by comparing the experimental Cα and Cβ chemical shifts with these derived with POTENCI, a neighbor-corrected random-coil library (Table S2 in Supporting Information).[45] The SCS analysis revealed that both A163 and A250 adopt a helical conformation, although with a different propensity, with A250 exhibiting a higher structuration level (3.3 ppm) than A163 (1.5 ppm).

Multisite-Specific Isotopic Labeling in Phox2B. Next, we aimed at validating our recently reported multisite-specific isotopic labeling (multi-SSIL) approach[46] for the systematic assignment of alanines, which only exhibits partial orthogonality (see above). In multi-SSIL, three sites are simultaneously labeled. Note that the incorporation of three isotopically labeled amino acids using tRNA suppression in concatenated samples reduces to $N/2$ the number of samples required to assign N residues.[46] However, the increase in the number of suppressed sites systematically reduces the yield of the CF reaction[46,47] (see Table S1) and increases the possibility of signal overlap. In this context, the incorporation of three sites is a good compromise. Importantly, the decrease in the CF protein yield of multi-SSIL samples can be partially modulated by adding a larger amount of aminoacylated tRNA$_{CUA}$ to the reaction. To validate this approach, three additional alanines were targeted: A253 and A259, which are located in the long poly-A tract, and A265, which is placed in the center of a short three-alanine segment next to the long tract (Figure 5A). The chemical shifts of these three additional alanines enabled a more detailed exploration of the effect of the poly-A length on the helical stability.

In order to increase the protein yield of the multi-SSIL sample (A253, A259, and A265) and obtain a 7 μM NMR sample of 200 μL, 20 μM loaded tRNA$_{CUA}$ was added to 10 mL of CF reaction. Three isolated frequencies could be identified in the NH, Cα-Hα, and Cβ-Hβ regions of the spectra, overcoming the severe overlap observed when all alanines of Phox2B were isotopically labeled (Figure S4). Interestingly, despite the tRNA$_{CUA}$/AlaRS pair being only partially orthogonal, we had a good enough signal-to-noise ratio to perform multi-SSIL. In order to assign these correlations, we produced two other multi-SSIL samples in which only one of the three labeled alanines coincided with the original one. Concretely, A242, A253, and A264 were isotopically labeled in the first sample, while A159, A259, and A266 were labeled in the second one (Figure 5A). These two additional samples would, in principle, provide the concatenated chemical shifts that will enable the unambiguous assignment of the three initially targeted alanines (A235, A259, and A265). Note that the selection of the other four sites (A242 and A264 for sample 1 and A159 and A266 for sample 2) was arbitrary and they were selected according to their position in the poly-A tracts, which we expected would have distinct chemical shifts and facilitate the subsequent assignment of A235, A259, and A265. The validation of multi-SSIL for alanines and the presented design would also enable future experiments to obtain a systematic assignment of both poly-A tracts in Phox2B.

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

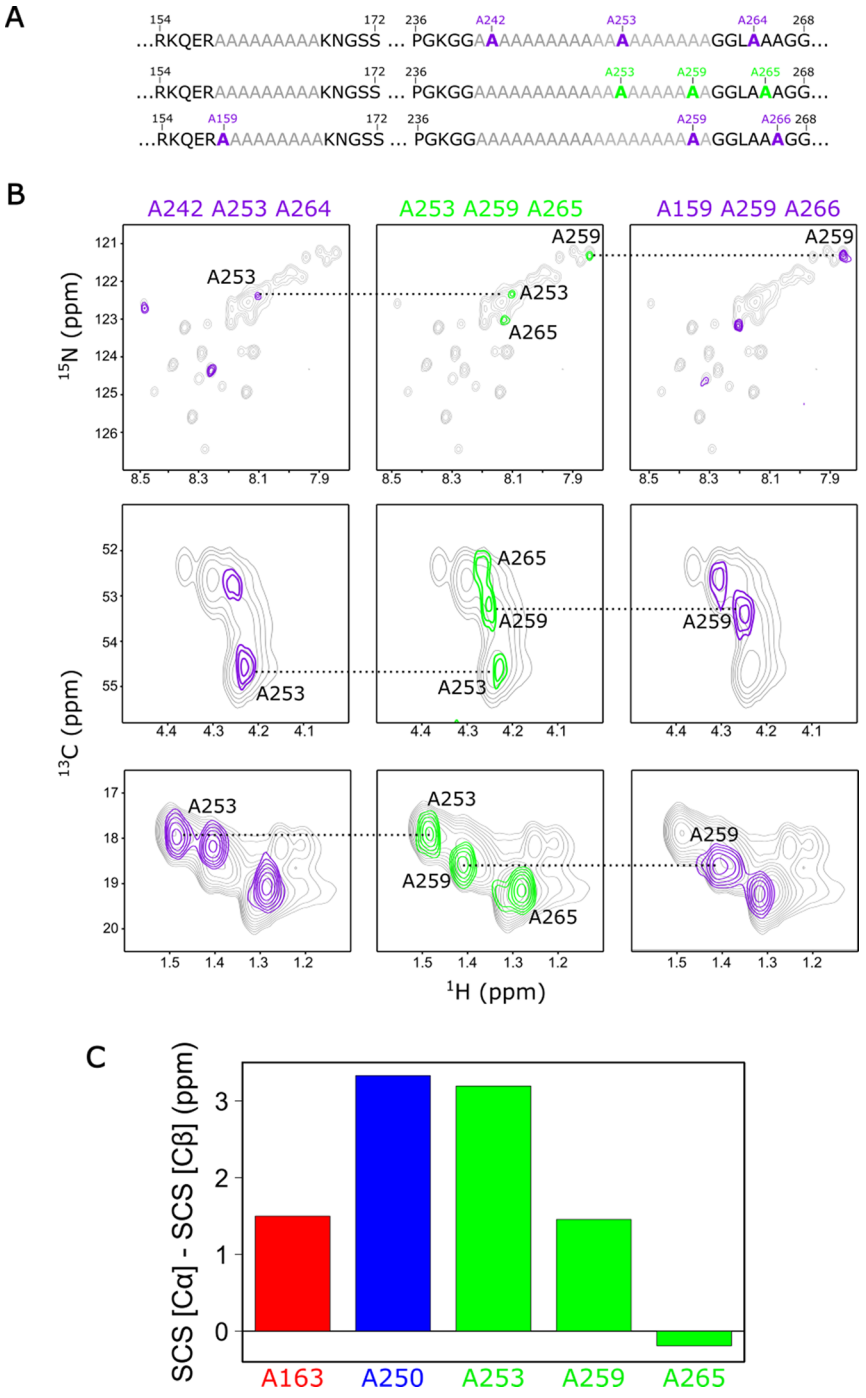


Figure 5. (A) Labeling scheme for the three multi-SSIL samples of this study. Central sequence with the three residues targeted (A253, A259, and A265—green) and the two concatenated sequences enabling its unambiguous assignment (purple). (B) Zooms of the NMR spectra displaying the NH, Cα-Hα, and Cβ-Hβ correlations (from top to down) for the A253-A259-A265 specifically labeled sample (in green) and its two concatenated samples (A242-A253-A264, purple, left) and A159-A259-A266 (purple, right) that enable the unambiguous assignment. Horizontal dashed lines indicate the common peaks in two concatenated spectra. (C) SCS analysis using experimental Cα and Cβ chemical shifts and a random-coil library. A163 (red) and A250 (blue) measured using SSIL and A253, A259, and A265 (green) measured in a single sample using multi-SSIL.

This strategy enabled us to unambiguously assign the three targeted alanines. Each of the ^{15}N-HSQC spectra of these two additional samples yielded three isolated NH frequencies and only one of them overlapped with the original spectrum, enabling the direct assignment of A253 and A259, while the only unconnected frequency was identified as A265 (Figure 5B). Due to the limited carbon frequency dispersion in Phox2B, a slightly different scenario occurred for Cα-Hα and Cβ-Hβ for which only two isolated peaks were observed in some of the spectra. Despite the reduced number of isolated frequencies, the overlap of concatenated spectra also yielded the unambiguous assignment of Cα-Hα and Cβ-Hβ correlations for A253, A259, and A265. Note, however, that further concatenation could produce some assignment ambiguities, which could be eventually overcome with strategically designed singly labeled alanine samples. Therefore,

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

despite the very limited frequency dispersion of alanine carbon correlations in Phox2B, these experiments demonstrate that the concatenation of multi-SSIL samples is feasible with the engineered *P. horikoshii* $tRNA_{CUA}$/AlaRS pair.

We then performed an SCS analysis including the three additional alanines, which displayed very different helical propensities (Table S2). Similar to A250, A253, which is also located in the central region of the long poly-A tract, displayed a strong helical propensity (Figure 5C). Conversely, A259, which is positioned toward the C-terminus of the same tract, exhibited a reduced helical behavior. Indeed, this terminal alanine had a structuration level very similar to that observed for A163, located in the center of the short poly-A tract. Interestingly, A265, which probed the helical propensity of the three-alanine stretch, displayed a random-coil conformational behavior.

Next, we investigated the helical fraction (HF) of the poly-A tracts of Phox2B. The precise calculation of the HF from sparse chemical shift data is difficult. Cα chemical shifts for several poly-A tracts in a forced α-helical conformation thanks to designed N- and C-capping residues have been experimentally determined, indicating a deviation of 2.3 ppm from random-coil estimated values.[48] Assuming a linear relationship between chemical shift deviation and helicity,[49] our Cα data indicate that the long poly-A tract of Phox2B has an HF of ∼90% at the central region that decreases to ∼25% at its C-terminus (A259). Interestingly, the short poly-A tract displays a maximum HF of ∼25%.

This structural propensity found in our NMR investigation of Phox2B is in line with the pioneering study of Gratzer and Doty[50] and recent protein NMR studies that unambiguously identified poly-A stretches containing up to 25 consecutive alanines as partially α-helical.[51−53] Note that these observations are in contradiction with other studies on synthetic poly-A peptides in which the residues were identified as disordered with some prevalence for extended poly-proline-II conformations.[54,55] Our data, as well as other structural information collected from poly-A tracts in their protein context, indicate a positive correlation between homorepeat length and helical propensity. This increased helical length and stability could be key in enhancing protein self-interaction through coiled-coils[56] and trigger aggregation[57] or phase separation[58] of pathogenic forms of poly-A-hosting proteins. In addition to the inherent α-helical propensity of poly-A tracts, it has been recently suggested that, in a protein context, this conformational behavior could be tuned by the nature of the flanking residues, where the presence of N- or C-capping residues could further stabilize the inherent helical propensity of the poly-A tracts.[59]

CONCLUDING REMARKS

In summary, despite the technical advances in multiple fronts, the high-resolution investigation of large biomolecular machines and LCRs by NMR remains a challenge. Here, we have presented an engineered $tRNA_{CUA}$/AlaRS pair enabling the site-specific incorporation of up to three alanines in four distinct proteins in terms of size, structure, and sequence composition, demonstrating the generality of the approach. With this methodology, the assignment of biological machines is facilitated and rendered independent of the specific mutation introduced in mutagenesis-based approaches. Furthermore, it enables the structural and dynamic characterization of alanine-rich proteins, including transcription factors causing poly-A-expansion developmental and neurodegenerative diseases. The extension of the SSIL strategy to other amino acids will permit structural and functional studies of other challenging relevant biomolecular systems, including the kinetic investigation of large catalytic machines or the structural bases of the emerging phenomenon of LCR-driven liquid−liquid phase separation.[60,61]

METHODS

Unless specified otherwise, all chemicals were obtained from Sigma-Aldrich (St. Quentin Fallavier, France).

PROTEIN CONSTRUCTS FOR SUPPRESSION SAMPLES

All plasmids were prepared as previously described.[25] Synthetic genes of wild-type huntingtin exon1 with 16 consecutive glutamines (HttExon1Q16), Phox2B C-terminal domain (residues from 136 to 314), and carrying amber codons (TAG) instead of alanine codons, e.g., A53 (HttExon1Q16-A53), were ordered from GeneArt (Thermo Fisher Scientific, Illkirch, France). All genes were cloned into pIVEX 2.3d 3C-sfGFP-His_6. Synthetic genes of ClpP from *Thermus thermophilus*, wild-type and mutants carrying amber codons (TAG) instead of alanine codons, e.g., A96-ClpP or A11-ClpP, were ordered from Genecust (Boynes, France). The sequence of all plasmids was confirmed by sequencing by GENEWIZ (Leipzig, Germany).

STANDARD CF EXPRESSION CONDITIONS

Lysate was prepared as previously described.[25] It is based on the *Escherichia coli* strain BL21 Star (DE3)::RF1-CBD_3, generously provided by Prof. Gottfried Otting (Australian National University, Canberra, Australia).[62] CF protein expression was performed as described in.[46,63] Nevertheless, the magnesium acetate (5−20 mM) and potassium glutamate (60−200 mM) concentrations were optimized for each new batch of S30 extract. A titration of both compounds was performed to obtain the maximum yield.

PREPARATION OF AMINOACYLATED SUPPRESSOR $TRNA_{CUA}$

A $tRNA_{CUA}$/tRNA synthetase pair from *Pyrococcus horikoshii*[33,34] was prepared in house. Substitution of the G3:U70 by a G3:C70 base pair in the $tRNA_{CUA}$, as well as N360A and E459A mutations on the alanine tRNA synthetase (AlaRS), was designed to improve orthogonality.[38] The optimized genes encoding the wild-type and double mutation (N306A and E459A) of truncated N752 *P. horikoshii* alanyl-tRNA synthetase (*Ph* AlaRS) were ordered from IDT DNA Technologies. These genes were cloned in pET26b vector between *NdeI* and *XhoI* sites in frame with the $(HIS)_6$-tag at the C-terminal end. Plasmids were transformed into the *E. coli* BL21(DE3) strain. Wild-type and mutated proteins were expressed in *E. coli* BL21(DE3) cells in LB medium at 37 °C supplemented with 50 μg/mL kanamycin. Protein expression was induced by IPTG (0.5 mM), and cells were grown for 3 h before harvesting them by centrifugation for 20 min at 6000 × *g* at 4 °C. The pellet was resuspended in 20 mM Tris−HCl, pH 7.5, 300 mM NaCl, and 2 mM DTT (buffer A) and stored at −80 °C. Cells were supplemented with protease inhibitors (cOmplete EDTA-free protease inhibitor cocktail), lysed by sonication, and insoluble proteins and cell debris were sedimented by centrifugation at 40000 × *g* at 4 °C for 30

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

min. The soluble fraction was aliquoted in 2 mL tubes, heated at 70 °C for 20 min, and centrifuged at 15000 × *g* at RT for 10 min. The supernatant was pooled and supplemented with imidazole to a final concentration of 10 mM, filtered through 0.45 μm filters, and loaded onto an affinity column (5 mL of HisTrap Excel, Cytiva), equilibrated with buffer B (buffer A containing 10 mM imidazole). The column was washed with buffer B, and proteins were eluted with a linear 0−100% gradient of buffer C (buffer A containing 0.5 M imidazole). The peak fractions were analyzed by SDS-PAGE. Fractions containing tagged *Ph* AlaRS were pooled and dialyzed overnight at 4 °C against 25 mM Tris−HCl, pH 7.5, 5 mM DTT, and 50 μM zinc acetate and concentrated to 10 mg/mL.

The suppressor $tRNA_{CUA}$ and the G3:C70 mutant (3′-GGGCCGGUAGCUCAGCCUGGUAUGAGCGCCGCC-C U c u a A A G G C G G A G G C C C C G G G U U -CAAAUCCCGGCCGGCCCACCA-5′) were transcribed *in vitro* and purified by phenol-chloroform extraction. Prior to use, suppressor $tRNA_{CUA}$ was refolded in 100 mM HEPES-KOH, pH 7.5, and 10 mM KCl at 70 °C for 5 min, and a final concentration of 5 mM $MgCl_2$ was added just before the reaction was placed on ice. The refolded $tRNA_{CUA}$ was then aminoacylated with [^{15}N,^{13}C]-L-alanine (CortecNet, Les Ulis, France) or [2-^{2}H,3-^{13}C]-L-alanine (NMR-Bio, France) in a standard aminoacylation reaction: 20 μM $tRNA_{CUA}$, 4 μM AlaRS (wild-type or mutant), 0.4 mM of the specific alanine in 100 mM sodium acetate, pH 5.0, 10 mM KCl, 20 mM $MgCl_2$, 0.5 mM TCEP, and 5 mM ATP. After incubation at 45 °C for 1 h, loaded suppressor $tRNA_{CUA}$ was precipitated with 300 mM sodium acetate, pH 5.2, and 2.5 volumes of 96% EtOH at −80 °C and stored as dry pellets at −20 °C. Successful loading was confirmed by polyacrylamide gels with 10% acrylamide (19:1), 6 M urea, and 37 mM PIPES, pH 6.0. To enable the analysis of empty and loaded tRNA, the free amine of the loaded amino acid was modified with sNHS-biotin and conjugated to streptavidin as reported by Pütz et al.[64] and the Suga lab.[65,66]

OPTIMIZATION OF CF SUPPRESSION CONDITIONS

In order to optimize the loaded $tRNA_{CUA}$ concentrations for the CF reaction, a titration was performed with final $tRNA_{CUA}$ concentrations between 0 and 30 μM. Protein expression was monitored by measuring sfGFP fluorescence in a plate reader/incubator (Gen5 v3.03.14, BioTek Instruments, Colmar, France) at 485 nm (excitation) and 528 nm (emission). Assays were carried out in a reaction volume of 50 μL dispensed in 96-well plates and incubated at 23 °C for 5 h. A similar test was performed for ClpP.

PREPARATION OF NMR SAMPLES

Samples for NMR studies were produced by CF at a 5−20 mL scale and incubated at 23 °C and 450 rpm in a thermomixer for 4 h or at 27 °C at 20 rpm in a hybridization oven (Techne) for 3 h for ClpP. Uniformly labeled NMR samples of HttExon1 and Phox2B were obtained by substituting the standard amino acid mix with 3 mg/mL [^{15}N,^{13}C]-labeled ISOGRO[67] (an algal extract lacking four amino acids: Asn, Cys, Gln, and Trp) and additionally supplying [^{15}N, ^{13}C]-labeled Asn, Cys, and Trp (1 mM each) and 2 mM Gln (CortecNet, Les Ulis, France). The perdeuterated ClpP sample, specifically $^{13}CH_3$-labeled on all the alanine and methionine residues, was obtained using a mix of 20 [^{2}H,^{15}N]-labeled amino acids (3 mg/mL; CIL, USA), supplemented with an excess of [2-^{2}H,3-^{13}C]-L-alanine (0.5 mg/mL; NMR-Bio, FR) and L-methionine (2,3,3,4,4-$^{2}H_5$, methyl-$^{13}CH_3$) (0.8 mg/mL CIL, USA). 15 μM [^{15}N,^{13}C]-alanine suppressor $tRNA_{CUA}$ was added to perform single site-specific suppression of HttExon1-A53 and Phox2B samples, while 20 μM [^{15}N,^{13}C]-L-alanine-loaded $tRNA_{CUA}$ was added for multiple site-specific suppression. For Phox2B suppression samples [^{2}H]-L-alanine and serine residues (2 mM) were used to eliminate the presence of natural abundance signals. The site-specifically labeled $^{13}CH_3$-A11-ClpP and $^{13}CH_3$-A96-ClpP samples were produced *in vitro* using a mix of 20 [^{2}H,^{15}N]-labeled amino acids (3 mg/mL; CIL, USA), supplemented with 20 μM suppressor $tRNA_{CUA}$ loaded with [2-^{2}H,3-^{13}C]-L-alanine. The NBD of the P1B-type ATPase HMA8 from *Arabidopsis thaliana* (16.7 kDa, 155 amino acids) sample specifically $^{13}CH_3$-labeled on all the alanine and methionine residues was produced in CF using a mixture of the 20 amino acids at a concentration of 1 mM each. [2-^{2}H,3-^{13}C]-L-alanine (NMR-Bio, FR) and L-methionine (2,3,3,4,4-$^{2}H_5$, methyl-$^{13}CH_3$, CIL, USA) were mixed with 18 other amino acids in unlabeled form. The site-specifically labeled $^{13}CH_3$-A619-HMA8-NBD samples were produced *in vitro* using a mixture containing L-methionine (2,3,3,4,4-$^{2}H_5$, methyl-$^{13}CH_3$, 1 mM), 19 unlabeled amino acids (1 mM each), supplemented with 15 μM suppressor $tRNA_{CUA}$ (wild-type or mutated) loaded with [2-^{2}H,3-^{13}C]-L-alanine.

PROTEIN SAMPLE PURIFICATION

The purification of HttExon1 and Phox2B C-terminal domain and their suppression mutants was performed at room temperature. The CF reaction was diluted 10-fold with buffer A (50 mM Tris−HCl, pH 7.5, 1 M NaCl) before incubating it for 1 h with 1.5 mL of Ni-resin (cOmplete His-TAG Purification Resin). The matrix was packed by gravity-flow and washed with increasing concentrations of imidazole (10, 15, 25, 50, 100, and 200 mM). Elution fractions were checked under UV light and fluorescent fractions were pooled, protease inhibitors were added (cOmplete EDTA-free protease inhibitor cocktail); 1 mM DTT was added to Phox2B C-terminal domain samples. Then, proteins were dialyzed against NMR buffer (20 mM BisTris−HCl, pH 6.5, 150 mM NaCl) at 4 °C using SpectraPor 4 MWCO 12−14 kDa dialysis tubing (Fisher Scientific, Illkirch, France). Dialyzed proteins were then concentrated with 10 kDa MWCO Vivaspin centrifugal concentrators (3500 × *g*, 4 °C) (Sartorius, Göttingen, Germany). Protein concentrations were determined by means of fluorescence using a sfGFP calibration curve. Final NMR sample concentrations ranged from 8 to 40 μM. Protein integrity was analyzed by SDS-PAGE.

T. thermophilus ClpP samples produced by CF were purified using a heat shock (60 °C for 20 min) followed by chromatography using a Resource Q resin (GE, USA). Elution fractions containing ClpP were concentrated with 50 kDa MWCO Amicon centrifugal concentrators (3500 × *g*) (Millipore, FR) in the final NMR buffer (99,8% D_2O, 20 mM Tris−HCl, pH 7.5, 100 mM NaCl, $MgCl_2$ 5 mM). Final NMR sample concentrations ranged from 20 to 400 μM (ClpP monomer concentration). The ATPase HMA8 NBD samples were purified in a single step using Ni-NTA affinity chromatography and concentrated (20−100 μM) in the NMR buffer (99,8% D_2O, 20 mM, pH 8, 100 mM NaCl).

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

NMR EXPERIMENTS AND DATA ANALYSIS

All NMR HttExon1 and Phox2B samples contained final concentrations of 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). NMR experiments for HttExon1 and Phox2B were performed at 293 K on a Bruker Avance III spectrometer (Bruker Biospin, Wissembourg, France) operating at a ^1H frequency of 800 MHz. ^{15}N-HSQC and ^{13}C-HSQC were acquired in order to determine amide (^1H_N and ^{15}N) and aliphatic ($^1H_{aliphatic}$ and $^{13}C_{aliphatic}$) chemical shifts, respectively. Spectrum acquisition parameters were set up depending on the sample concentration. For ClpP samples, 2D ^1H-^{13}C SOFAST methyl-TROSY experiments[43] were performed at 333 K on a Bruker Avance III HD spectrometer equipped with a 5 mm cryogenically cooled, pulsed-field-gradient triple-resonance probe operating at a ^1H frequency of 600 MHz. Extra signals corresponding to protein impurities were greatly reduced by subtracting NMR spectra acquired on a control sample produced in vitro using the same protocol but in the absence of the ClpP plasmid. 2D ^1H-^{13}C SOFAST methyl-TROSY experiments[43] for the NBD of ATPase HMA8 were recorded at 298 K on a Bruker Avance III HD spectrometer equipped with a 5 mm cryogenically cooled, pulsed-field-gradient triple-resonance probe operating at a ^1H frequency of 850 MHz.

All spectra were processed with TopSpin v3.5 (Bruker Biospin, Wissembourg, France) and analyzed using CCPN-Analysis software v2.4.[68] Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the $^1H/X$ frequency ratio of the zero point according to Markley *et al.*[69]

Random-coil chemical shifts for Phox2B were predicted using POTENCI, a pH, temperature, and neighbor-corrected IDP library (https://st-protein02.chem.au.dk/potenci/).[45] SCSs were obtained by subtracting the predicted value from the experimental one (SCS = $\delta_{exp} - \delta_{pred}$).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.3c00288.

CF reaction conditions and yields; structural parameters derived from the chemical shifts for phox2B; gels for the $tRNA_{CUA}$ loading and the purification of the proteins; sequence alignment of the alanine tRNA synthetases; and additional NMR spectra (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jérôme Boisbouvier − *Univ. Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale (IBS), F-38044 Grenoble, France*; orcid.org/0000-0003-3278-3639; Email: jerome.boisbouvier@ibs.fr

Pau Bernadó − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*; orcid.org/0000-0001-7395-5922; Email: pau.bernado@cbs.cnrs.fr

Authors

Carlos A. Elena-Real − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Annika Urbanek − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Lionel Imbert − *Univ. Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale (IBS), F-38044 Grenoble, France*

Anna Morató − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Aurélie Fournet − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Frédéric Allemand − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Nathalie Sibille − *Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.3c00288

Author Contributions

P.B. and J.B. conceived the project. C.A.E.-R., A.U., L.I., J.B., and P.B. designed experiments. C.A.E.-R, A.U., L.I., A.M., A.F., and F.A. performed experiments. N.S., J.B., and P.B. supervised experiments. C.A.E.-R., J.B., and P.B. wrote the manuscript with the help of all the co-authors. C.A.E.-R., A.U., and L.I. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020)/ERC Grant agreement *n*° [648030] and MUSE-App 2021 Ondine ANR-16-IDEX-0006 awarded to P.B. and ProteaseInAction ANR-19-CE11-0022 awarded to J.B. The CBS is a member of France-BioImaging (FBI), and both CBS and IBS are members of the French Infrastructure for Integrated Structural Biology (FRISBI), supported by ANR-10-INBS-04-01 and ANR-10-INBS-05 grants, respectively. This work used facilities at the Grenoble Instruct-ERIC Center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), with the support of GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

REFERENCES

(1) Camacho-Zarco, A. R.; Schnapka, V.; Guseva, S.; Abyzov, A.; Adamski, W.; Milles, S.; Jensen, M. R.; Zidek, L.; Salvi, N.; Blackledge, M. NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem. Rev.* **2022**, *122*, 9331−9356.

(2) Liu, B.; Shi, H.; Al-Hashimi, H. M. Developments in Solution-State NMR Yield Broader and Deeper Views of the Dynamic Ensembles of Nucleic Acids. *Curr. Opin. Struct. Biol.* **2021**, *70*, 16−25.

(3) Schütz, S.; Sprangers, R. Methyl TROSY Spectroscopy: A Versatile NMR Approach to Study Challenging Biological Systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **2020**, *116*, 56−84.

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

(4) Rosenzweig, R.; Kay, L. E. Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR. *Annu. Rev. Biochem.* **2014**, *83*, 291−315.
(5) Ohki, S.; Kainosho, M. Stable Isotope Labeling Methods for Protein NMR Spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 208−226.
(6) Narayanan, R. L.; Dürr, U. H. N.; Bibow, S.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. *J. Am. Chem. Soc.* **2010**, *132*, 11906−11907.
(7) Pustovalova, Y.; Mayzel, M.; Orekhov, V. Y. XLSY: Extra-Large NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2018**, *57*, 14043−14045.
(8) Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear Multidimensional NMR Experiments for the Structure Determination of Proteins in Solution Employing Pulsed Field Gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93−158.
(9) Urbanek, A.; Elena-Real, C. A.; Popovic, M.; Morató, A.; Fournet, A.; Allemand, F.; Delbecq, S.; Sibille, N.; Bernadó, P. Site-Specific Isotopic Labeling (SSIL): Access to High-Resolution Structural and Dynamic Information in Low-Complexity Proteins. *ChemBioChem* **2020**, *21*, 769.
(10) Mas, G.; Guan, J.-Y.; Crublet, E.; Debled, E. C.; Moriscot, C.; Gans, P.; Schoehn, G.; Macek, P.; Schanda, P.; Boisbouvier, J. Structural Investigation of a Chaperonin in Action Reveals How Nucleotide Binding Regulates the Functional Cycle. *Sci. Adv.* **2018**, *4*, No. eaau4196.
(11) Gauto, D. F.; Macek, P.; Barducci, A.; Fraga, H.; Hessel, A.; Terauchi, T.; Gajan, D.; Miyanoiri, Y.; Boisbouvier, J.; Lichtenecker, R.; Kainosho, M.; Schanda, P. Aromatic Ring Dynamics, Thermal Activation, and Transient Conformations of a 468 KDa Enzyme by Specific (1)H-(13)C Labeling and Fast Magic-Angle Spinning NMR. *J. Am. Chem. Soc.* **2019**, *141*, 11183−11195.
(12) Lapinaite, A.; Simon, B.; Skjaerven, L.; Rakwalska-Bange, M.; Gabel, F.; Carlomagno, T. The Structure of the Box C/D Enzyme Reveals Regulation of RNA Methylation. *Nature* **2013**, *502*, 519−523.
(13) Rosenzweig, R.; Moradi, S.; Zarrine-Afsar, A.; Glover, J. R.; Kay, L. E. Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction. *Science* **2013**, *339*, 1080−1083.
(14) Sprangers, R.; Kay, L. E. Quantitative Dynamics and Binding Studies of the 20S Proteasome by NMR. *Nature* **2007**, *445*, 618−622.
(15) Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. Methyl-Specific Isotopic Labeling: A Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 113−122.
(16) Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, *1*, 749−754.
(17) McShan, A. C. Utility of Methyl Side Chain Probes for Solution NMR Studies of Large Proteins. *J. Magn. Reson. Open* **2023**, *14-15*, No. 100087.
(18) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* **1996**, *263*, 627−636.
(19) Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. A Robust and Cost-Effective Method for the Production of Val, Leu, Ile (Delta 1) Methyl-Protonated 15N-, 13C-, 2H-Labeled Proteins. *J. Biomol. NMR* **1999**, *13*, 369−374.
(20) Hoogstraten, C. G.; Johnson, J. E., Jr. Metabolic Labeling: Taking Advantage of Bacterial Pathways to Prepare Spectroscopically Useful Isotope Patterns in Proteins and Nucleic Acids. *Concepts Magn. Reson. A* **2008**, *32A*, 34−55.
(21) Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late Metabolic Precursors for Selective Aromatic Residue Labeling. *J. Biomol. NMR* **2018**, *71*, 129−140.
(22) Amero, C.; Asunción Durá, M.; Noirclerc-Savoye, M.; Perollier, A.; Gallet, B.; Plevin, M. J.; Vernet, T.; Franzetti, B.; Boisbouvier, J. A Systematic Mutagenesis-Driven Strategy for Site-Resolved NMR Studies of Supramolecular Assemblies. *J. Biomol. NMR* **2011**, *50*, 229−236.
(23) Crublet, E.; Kerfah, R.; Mas, G.; Noirclerc-Savoye, M.; Lantez, V.; Vernet, T.; Boisbouvier, J. *A Cost-Effective Protocol for the Parallel Production of Libraries of 13CH3-Specifically Labeled Mutants for NMR Studies of High Molecular Weight Proteins BT - Structural Genomics: General Applications*; Chen, Y. W., Ed.; Humana Press: Totowa, NJ, 2014; pp. 229−244.
(24) Gelis, I.; Bonvin, A. M. J. J.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **2007**, *131*, 756−769.
(25) Urbanek, A.; Morató, A.; Allemand, F.; Delaforge, E.; Fournet, A.; Popovic, M.; Delbecq, S.; Sibille, N.; Bernadó, P. A General Strategy to Access Structural Information at Atomic Resolution in Polyglutamine Homorepeats. *Angew. Chem. Int. Ed.* **2018**, *57*, 3598−3601.
(26) Baker, T. A.; Sauer, R. T. ClpXP, an ATP-Powered Unfolding and Protein-Degradation Machine. *Biochim. Biophys. Acta Mol. Cell Res.* **2012**, *1823*, 15−28.
(27) Olivares, A. O.; Baker, T. A.; Sauer, R. T. Mechanistic Insights into Bacterial AAA+ Proteases and Protein-Remodelling Machines. *Nat. Rev. Microbiol.* **2016**, *14*, 33−44.
(28) Lavoie, H.; Debeane, F.; Trinh, Q.-D.; Turcotte, J.-F.; Corbeil-Girard, L.-P.; Dicaire, M.-J.; Saint-Denis, A.; Pagé, M.; Rouleau, G. A.; Brais, B. Polymorphism, Shared Functions and Convergent Evolution of Genes with Sequences Coding for Polyalanine Domains. *Hum. Mol. Genet.* **2003**, *12*, 2967−2979.
(29) Amiel, J.; Trochet, D.; Clément-Ziza, M.; Munnich, A.; Lyonnet, S. Polyalanine Expansions in Human. *Hum. Mol. Genet.* **2004**, *13*, R235−R243.
(30) Matera, I.; Bachetti, T.; Puppo, F.; Di Duca, M.; Morandi, F.; Casiraghi, G. M.; Cilio, M. R.; Hennekam, R.; Hofstra, R.; Schöber, J. G.; Ravazzolo, R.; Ottonello, G.; Ceccherini, I. Mutations and Polyalanine Expansions Correlate with the Severity of the Respiratory Phenotype and Associated Symptoms in Both Congenital and Late Onset Central Hypoventilation Syndrome. *J. Med. Genet.* **2004**, *41*, 373.
(31) Ramanantsoa, N.; Gallego, J. Congenital Central Hypoventilation Syndrome. *Respir. Physiol. Neurobiol.* **2013**, *189*, 272−279.
(32) Darling, L. A.; Uversky, N. V. Intrinsic Disorder in Proteins with Pathogenic Repeat Expansions. *Molecules* **2017**, *22*, 2027.
(33) Sokabe, M.; Okada, A.; Yao, M.; Nakashima, T.; Tanaka, I. Molecular Basis of Alanine Discrimination in Editing Site. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11669−11674.
(34) Sokabe, M.; Ose, T.; Nakamura, A.; Tokunaga, K.; Nureki, O.; Yao, M.; Tanaka, I. The Structure of Alanyl-TRNA Synthetase with Editing Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11028−11033.
(35) Morató, A.; Elena-Real, C. A.; Popovic, M.; Fournet, A.; Zhang, K.; Allemand, F.; Sibille, N.; Urbanek, A.; Bernadó, P. Robust Cell-Free Expression of Sub-Pathological and Pathological Huntingtin Exon-1 for Nmr Studies. General Approaches for the Isotopic Labeling of Low-Complexity Proteins. *Biomolecules* **2020**, *10*, 1458.
(36) Hou, Y.-M.; Schimmel, P. A Simple Structural Feature Is a Major Determinant of the Identity of a Transfer RNA. *Nature* **1988**, *333*, 140−145.
(37) McClain, W. H.; Foss, K. Changing the Identity of a TRNA by Introducing a G-U Wobble Pair Near the 3′ Acceptor End. *Science* **1988**, *240*, 793−796.
(38) Chong, Y. E.; Guo, M.; Yang, X.-L.; Kuhle, B.; Naganuma, M.; Sekine, S.-I.; Yokoyama, S.; Schimmel, P. Distinct Ways of G:U Recognition by Conserved TRNA Binding Motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 7527−7532.
(39) Urbanek, A.; Popovic, M.; Morató, A.; Estaña, A.; Elena-Real, C. A.; Mier, P.; Fournet, A.; Allemand, F.; Delbecq, S.; Andrade-Navarro, M. A.; Cortés, J.; Sibille, N.; Bernadó, P. Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX

Mechanisms Common among Glutamine-Rich Human Proteins. *Structure* **2020**, *28*, 733−746.e5.
(40) Mayerhofer, H.; Sautron, E.; Rolland, N.; Catty, P.; Seigneurin-Berny, D.; Pebay-Peyroula, E.; Ravaud, S. Structural Insights into the Nucleotide-Binding Domains of the P1B-Type ATPases HMA6 and HMA8 from Arabidopsis Thaliana. *PLoS One* **2016**, *11*, No. e0165666.
(41) Imbert, L.; Lenoir-Capello, R.; Crublet, E.; Vallet, A.; Awad, R.; Ayala, I.; Juillan-Binard, C.; Mayerhofer, H.; Kerfah, R.; Gans, P.; Miclet, E.; Boisbouvier, J. In Vitro Production of Perdeuterated Proteins in H2O for Biomolecular NMR Studies. In *Structural Genomics: General Applications*; Chen, Y. W., Yiu, C.-P. B., Eds.; Springer US: New York, NY, 2021; pp. 127−149.
(42) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. Cross-Correlated Relaxation Enhanced 1H−13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* **2003**, *125*, 10420−10428.
(43) Amero, C.; Schanda, P.; Durá, M. A.; Ayala, I.; Marion, D.; Franzetti, B.; Brutscher, B.; Boisbouvier, J. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *J. Am. Chem. Soc.* **2009**, *131*, 3448−3449.
(44) Vahidi, S.; Ripstein, Z. A.; Bonomi, M.; Yuwen, T.; Mabanglo, M. F.; Juravsky, J. B.; Rizzolo, K.; Velyvis, A.; Houry, W. A.; Vendruscolo, M.; Rubinstein, J. L.; Kay, L. E. Reversible Inhibition of the ClpP Protease via an N-Terminal Conformational Switch. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E6447−E6456.
(45) Nielsen, J. T.; Mulder, F. A. A. POTENCI: Prediction of Temperature, Neighbor and PH-Corrected Chemical Shifts for Intrinsically Disordered Proteins. *J. Biomol. NMR* **2018**, *70*, 141−165.
(46) Elena-Real, C. A.; Urbanek, A.; Lund, X. L.; Morató, A.; Sagar, A.; Fournet, A.; Estaña, A.; Bellande, T.; Allemand, F.; Cortés, J.; Sibille, N.; Melki, R.; Bernadó, P. Multi-Site-Specific Isotopic Labeling Accelerates Highresolution Structural Investigations of Pathogenic Huntingtin Exon-1. *Structure* **2023**, *31*, 644−650.e5.
(47) Martin, R. W.; Des Soye, B. J.; Kwon, Y.-C.; Kay, J.; Davis, R. G.; Thomas, P. M.; Majewska, N. I.; Chen, C. X.; Marcum, R. D.; Weiss, M. G.; Stoddart, A. E.; Amiram, M.; Ranji Charna, A. K.; Patel, J. R.; Isaacs, F. J.; Kelleher, N. L.; Hong, S. H.; Jewett, M. C. Cell-Free Protein Synthesis from Genomically Recoded Bacteria Enables Multisite Incorporation of Noncanonical Amino Acids. *Nat. Commun.* **2018**, *9*, 1203.
(48) Heitmann, B.; Job, G. E.; Kennedy, R. J.; Walker, S. M.; Kemp, D. S. Water-Solubilized, Cap-Stabilized, Helical Polyalanines: Calibration Standards for NMR and CD Analyses. *J. Am. Chem. Soc.* **2005**, *127*, 1690−1704.
(49) Weinstock, D. S.; Narayanan, C.; Baum, J.; Levy, R. M. Correlation between 13Calpha Chemical Shifts and Helix Content of Peptide Ensembles. *Protein Sci.* **2008**, *17*, 950−954.
(50) Gratzer, W. B.; Doty, P. A Conformation Examination of Poly-L-Alanine and Poly-D,L-Alanine in Aqueous Solution. *J. Am. Chem. Soc.* **1963**, *85*, 1193−1197.
(51) Hong, J.-Y.; Wang, D.-D.; Xue, W.; Yue, H.-W.; Yang, H.; Jiang, L.-L.; Wang, W.-N.; Hu, H.-Y. Structural and Dynamic Studies Reveal That the Ala-Rich Region of Ataxin-7 Initiates α-Helix Formation of the PolyQ Tract but Suppresses Its Aggregation. *Sci. Rep.* **2019**, *9*, 7481.
(52) Chen, T.-C.; Huang, J. Musashi-1: An Example of How Polyalanine Tracts Contribute to Self-Association in the Intrinsically Disordered Regions of RNA-Binding Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 2289.
(53) Roberts, S.; Harmon, T. S.; Schaal, J. L.; Miao, V.; Li, K.; Hunt, A.; Wen, Y.; Oas, T. G.; Collier, J. H.; Pappu, R. V.; Chilkoti, A. Injectable Tissue Integrating Networks from Recombinant Polypeptides with Tunable Order. *Nat. Mater.* **2018**, *17*, 1154−1163.
(54) Chen, K.; Liu, Z.; Kallenbach, N. R. The Polyproline II Conformation in Short Alanine Peptides Is Noncooperative. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15352−15357.
(55) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9190−9195.
(56) Pelassa, I.; Corà, D.; Cesano, F.; Monje, F. J.; Montarolo, P. G.; Fiumara, F. Association of Polyalanine and Polyglutamine Coiled Coils Mediates Expansion Disease-Related Protein Aggregation and Dysfunction. *Hum. Mol. Genet.* **2014**, *23*, 3402−3420.
(57) Polling, S.; Ormsby, A. R.; Wood, R. J.; Lee, K.; Shoubridge, C.; Hughes, J. N.; Thomas, P. Q.; Griffin, M. D. W.; Hill, A. F.; Bowden, Q.; Böcking, T.; Hatters, D. M. Polyalanine Expansions Drive a Shift into α-Helical Clusters without Amyloid-Fibril Formation. *Nat. Struct. Mol. Biol.* **2015**, *22*, 1008−1015.
(58) Basu, S.; Mackowiak, S. D.; Niskanen, H.; Knezevic, D.; Asimi, V.; Grosswendt, S.; Geertsema, H.; Ali, S.; Jerković, I.; Ewers, H.; Mundlos, S.; Meissner, A.; Ibrahim, D. M.; Hnisz, D. Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **2020**, *181*, 1062−1079.e30.
(59) Mier, P.; Elena-Real, C. A.; Cortés, J.; Bernadó, P.; Andrade-Navarro, M. A. The Sequence Context in Poly-Alanine Regions: Structure, Function and Conservation. *Bioinformatics* **2022**, *38*, 4851−4858.
(60) Franzmann, T. M.; Alberti, S. Prion-like Low-Complexity Sequences: Key Regulators of Protein Solubility and Phase Behavior. *J. Biol. Chem.* **2019**, *294*, 7128−7136.
(61) Bremer, A.; Farag, M.; Borcherds, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering How Naturally Occurring Sequence Features Impact the Phase Behaviours of Disordered Prion-like Domains. *Nat. Chem.* **2022**, *14*, 196−207.
(62) Loscha, K. V.; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angew. Chem. Int. Ed.* **2012**, *51*, 2243−2246.
(63) Apponyi, M. A.; Ozawa, K.; Dixon, N. E.; Otting, G. Cell-Free Protein Synthesis for Analysis by NMR Spectroscopy. In *Structural Proteomics: High-Throughput Methods*; Kobe, B., Guss, M., Huber, T., Eds.; Humana Press: Totowa, NJ, 2008; pp. 257−268.
(64) Pütz, J.; Wientges, J.; Schwienhorst, A.; Sissler, M.; Giegé, R.; Florentz, C. Rapid Selection of Aminoacyl-TRNAs Based on Biotinylation of α-NH2 Group of Charged Amino Acids. *Nucleic Acids Res.* **1997**, *25*, 1862−1863.
(65) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A Highly Flexible TRNA Acylation Method for Non-Natural Polypeptide Synthesis. *Nat. Methods* **2006**, *3*, 357−359.
(66) Goto, Y.; Katoh, T.; Suga, H. Flexizymes for Genetic Code Reprogramming. *Nat. Protoc.* **2011**, *6*, 779−790.
(67) Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. Cell-Free Production and Stable-Isotope Labeling of Milligram Quantities of Proteins. *FEBS Lett.* **1999**, *442*, 15−19.
(68) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **2005**, *59*, 687−696.
(69) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids. *J. Mol. Biol.* **1998**, *280*, 933−952.

https://doi.org/10.1021/acschembio.3c00288
ACS Chem. Biol. XXXX, XXX, XXX−XXX