Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

A General Strategy to Access Structural Information at Atomic Resolution in Polyglutamine Homorepeats

*Annika Urbanek, Anna Morató, Frédéric Allemand, Elise Delaforge, Aurélie Fournet, Matija Popovic, Stephane Delbecq, Nathalie Sibille, and Pau Bernadó**

anie_201711530_sm_miscellaneous_information.pdf

Table of Contents

Experimental Procedures

DNA constructs

H16 and H46

The H16 (htt exon1 with 16 consecutive glutamines) and superfolder GFP (sfGFP)[1] synthetic genes were ordered from Integrated DNA Technologies (IDT). The sequences carrying H16 and sfGFP were cleaved by NcoI and XhoI endonucleases and cloned into pIVEX 2.3d by an In-Fusion® (Clontech) reaction, giving rise to pIVEX-H16-sfGFP-His_6. The sfGFP fusion protein can be removed by HRV 3C protease. Five H16 genes carrying the amber codon (TAG) instead of the initial glutamine codon at positions Q20, Q24, Q28, Q32, or Q48, respectively, were ordered from IDT. These suppressor mutants were cloned into the pIVEX-H16-sfGFP plasmid by In-Fusion® between NcoI and KpnI sites. A synthetic gene coding for H46 (htt exon1 with 46 consecutive glutamines) carrying the amber codon at position Q20 was ordered from GeneArt® and subcloned into the pIVEX plasmid. The sequence of all plasmids was confirmed by sequencing by GENEWIZ® and chromatograms for H16Q24 and H46Q20 are shown in Figure S5 as example.

GLN4

The cDNA sequence coding for GLN4 (glutaminyl-tRNA transferase, UniProt ID P13188) was ordered from IDT and subcloned into pET22 between the NdeI and XhoI restriction sites to yield the pET22-Gln4 vector. The final GLN4 construct carried a HRV 3C protease recognition site followed by GST-His_6 at its C-terminus.

Artificial suppressor $tRNA_{CUA}$

A DNA duplex containing the sequence of the T7 RNA polymerase promoter followed by the gene coding for Gln2 tRNA from *Saccharomyces cerevisiae*[2] with a G36→A anticodon mutation was ordered from IDT. The primers (5' GGCGTAATACGACTCACTATAG 3' and 5' mUmGGAGGTCCCACCCGGAT 3') to amplify the template by the means of PCR and insert two O-methylated bases at the 3' end of the final template[3] were also ordered from IDT.

Protein purification

H16 and H46

5 mL of frozen CF reaction were thawed on ice and diluted with 10 mL H16 buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM imidazole) before loading onto a Ni gravity-flow column of 1 mL bed volume (cOmplete™ His-Tag Purification Resin, Sigma Aldrich). The column was washed with H16 buffer B (50 mM Tris-HCl pH 7.5, 1000 mM NaCl, 5 mM imidazole) and the target protein was eluted with H16 buffer C (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 250 mM imidazole). Elution fractions were checked under UV light and fluorescent fractions were pooled and dialyzed against H16 NMR buffer (20 mM BisTris-HCl pH 6.5, 150 mM NaCl) at 4°C using SpectraPor 1 MWCO 6-8 kDa dialysis tubing (Spectrum Labs). Dialyzed protein was then concentrated with a 10 kDa MWCO Vivaspin centrifugal concentrator (3,500 xg, 4°C) (Sartorius). Protein concentrations were determined by means of fluorescence using an sfGFP calibration curve. Protein integrity was analyzed by SDS-PAGE.

GLN4

E. coli BL21 (DE3) cells were transformed with pET22-Gln4 and grown overnight at 25°C in ZYM 5052 auto-inducing medium supplemented with 50 µg/mL kanamycin[4]. Cells were harvested by centrifugation (6,000 xg, 20 min, 4°C), the pellet was resuspended in 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 2 mM β-mercaptoethanol (GLN4 buffer A) and stored at -80°C. Cells were

supplemented with a cOmplete™ EDTA free protease inhibitor tablet (Roche) and lysed by sonication. Cell debris and insoluble proteins were sedimented by centrifugation (40,000 xg, 30 min, 4°C). The supernatant was supplemented with imidazole to a final concentration of 10 mM and loaded onto a gravity affinity column (Ni sepharose 6 FF 5 mL, GE Life Sciences), equilibrated with GLN4 buffer B (GLN4 buffer A containing 10 mM imidazole). The column was washed with 50 mL GLN4 buffer B and the target protein was eluted with GLN4 buffer C (GLN4 buffer A containing 250 mM imidazole). Fractions were analyzed by SDS-PAGE and fractions containing GLN4 were pooled and dialyzed against GLN4 buffer A overnight at 4°C. The dialyzed protein was then loaded on a 5 mL gravity GST column (glutathione sepharose 4B, GE Life Sciences) equilibrated in GLN4 buffer A. The resin was washed with 60 mL of GLN4 buffer A and proteins were eluted with GLN4 buffer D (GLN4 buffer A containing 10 mM glutathione). Protein fractions were analyzed by SDS-PAGE and fractions containing GLN4 were pooled and dialyzed against GLN4 buffer E (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT) and concentrated to 6 mg/mL with Vivaspin centrifugal concentrators (Sartorius Stedim Biotech).

***In vitro* transcription and aminoacylation of $tRNA_{CUA}$**

Artificial yeast Gln suppressor $tRNA_{CUA}$ was produced by *in vitro* run-off transcription directly from the PCR product using the HiScribe™ T7 High Yield RNA Synthesis Kit (New England Biolabs) following the manufacturer's instructions. The resulting transcript was purified by phenol-chloroform extraction using a mixture of phenol-chloroform-isoamylalcohol (25:24:1) (Applichem), precipitated in 2 M NH_4-acetate and 2.5 volumes of 96% EtOH at -80°C and stored as dry pellets at -20°C. Prior to use the pellets were dissolved in nuclease-free water and refolded. To this end, $tRNA_{CUA}$ in HK buffer (100 mM HEPES-KOH pH 7.5, 10 mM KCl) was heated to 70°C for 5 min and a final concentration of 5 mM $MgCl_2$ was added just before placing the reaction on ice. The refolded $tRNA_{CUA}$ was then aminoacylated with [^{15}N-^{13}C]-glutamine (CortecNet) at 37°C for 1 hour. A standard aminoacylation reaction contained 20 µM $tRNA_{CUA}$, 0.5 µM GLN4, 0.1 mM [^{15}N-^{13}C]-Gln in 100 mM HEPES-KOH pH 7.5, 10 mM KCl, 20 mM $MgCl_2$, 1 mM DTT and 10 mM ATP[5]. GLN4 was removed by addition of glutathione beads (glutathione sepharose 4B, GE Life Sciences) to the aminoacylation reaction. The reaction was incubated for another 30 min before the glutathione beads were removed. Successful loading was confirmed by urea-PAGE (6.5% acrylamide 19:1, 8 M urea, 100 mM Na-acetate pH 5.2)[5].The loaded suppressor $tRNA_{CUA}$ was precipitated in 300 mM Na-acetate pH 5.2 and 2.5 volumes of 96% EtOH at -80°C and stored as dry pellets at -20°C.

Lysate preparation

E. coli lysates were prepared as described by Apponyi *et al*. and Loscha *et al*. using BL21 Star (DE3)::RF1-CBD_3 cells[6,7], but with slight modifications. The cells were grown at 37°C in a fermenter with 3 L of Z-medium (41,2 mM potassium dihydrogen phosphate, 166 mM potassium phosphate dibasic, 10 g/L yeast extract, 110 mM glucose, 10 mg/L thiamine, 1 mM $MgSO_4$ and 50 µg/mL Kanamycin). When the cell density (OD_{600}) reached 1, isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) was added to the cell culture medium to induce T7 RNA polymerase synthesis. The cells were harvested in the mid-log phase (OD_{600} ~3-4) and washed with S30 α buffer (10 mM Tris-acetate, pH 8.2, 16 mM potassium acetate, 14 mM magnesium acetate, 0.5 mM PMSF, 1 mM DTT and 7.2 mM β-mercaptoethanol) before storing the pellets at -80°C until use. The thawed cells were suspended in buffer S30 α (1.3 mL of buffer per gram of cells) and disrupted in a French press cell (Emulsiflex C-3, Avestin) at a constant pressure of 20,000 psi. The lysate was cleared by centrifugation twice (30 min; 30,000 xg; 4°C), discarding the pellet after each centrifugation. To remove release factor 1 (RF1), the lysate was passed over chitin resin (New England Biolabs) as described by Loscha *et al*.[7]. All dialysis steps were performed at 4°C. The final supernatant was dialysed against buffer S30 β (buffer S30 α without PMSF and β-mercaptoethanol) using SpectraPor#4 dialysis tubing (12-14 kDa MWCO) (Spectrum Laboratories Inc.) for 3x 1 hour, followed by a dialysis against 50% PEG 8000 in S30 buffer until the volume of the extract was reduced to half volume. Residual traces of PEG were removed by a short dialysis against S30 β (~15 min) before changing to buffer S30 γ (S30 β with 400 mM NaCl) overnight. Subsequently the dialysis tubes were placed into a 250 mL Pyrex glass bottle, filled with prewarmed buffer S30 γ (42°C), and incubated for 45 minutes at 42°C in a water-bath with gentle shaking. The lysate was then dialyzed against buffer S30 β for 4 h. Finally, the extract was cleared by centrifugation (10 min; 30,000 xg; 4°C) and the supernatant was aliquoted, flash frozen and stored at -80°C until use.

Cell-free expression

Cell-free protein expression was performed in batch mode as described by Apponyi *et al*.[6]. The time-course of protein synthesis was monitored using a fluorescence read-out (sfGFP) in combination with a plate reader/incubator (Gen5, BioTek Instruments, 485 nm (excitation), 528 nm (emission)). Most assays were carried out as triplicates in a reaction volume of 50 µL dispensed in 96-well plates. The reactions were incubated at 23-30°C for 5 hours. The standard reaction mixture consisted of the following components: 55 mM HEPES-KOH (pH 7.5), 1.2 mM ATP, 0.8 mM each of CTP, GTP and UTP, 1.7 mM DTT, 0.175 mg/mL *E. coli* total tRNA mixture (from strain MRE600), 0.64 mM cAMP, 27.5 mM ammonium acetate, 68 µM 1-5-formyl-5,6,7,8-tetrahydrofolic acid (folinic acid), 1 mM of each of the 20 amino acids, 80 mM creatine phosphate (CP), 250 µg/mL creatine kinase (CK), plasmid (16 µg/mL) and 22.5% (v/v) S30 extract. The concentrations of magnesium acetate (5 - 20 mM) and potassium glutamate (60 - 200 mM) were adjusted for each plasmid and for each new batch of S30 extract. A titration of both compounds was performed to obtain the optimum concentrations.

Samples for NMR studies were produced at large scale (5 x 1 mL for H16 and 10 x 1 mL for H46) and incubated at 23°C for 5 hours at 750 rpm. To produce site-specifically labeled samples 10 µM of Gln suppressor $tRNA_{CUA}$ were added in addition to the standard reaction mixture (see above). Yields between 12-17% and 8% were obtained for the H16 samples and H46Q20, respectively. Uniformly labeled NMR samples were prepared by substituting the standard amino acid mix with 3 mg/mL ISOGRO®[8] (an algal extract lacking four amino acids: Asn, Cys, Gln and Trp) and additionally supplying Asn, Cys, Gln and Trp (1 mM each). To simplify the assignment, two different samples were used: ISOGRO® -Gln -Glu (unlabeled Gln and the standard 80 mM potassium glutamate were used) and ISOGRO® with added 1 mM [^{15}N, ^{13}C] labeled Gln and 80 mM ammonium acetate instead of 80 mM potassium glutamate.

NMR

All NMR samples contained 20 mM BisTris-HCl pH 6.5, 150 mM NaCl, 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were recorded at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a ^{1}H frequency of 700 MHz for H16Q24, H16Q28, H16Q32, H16Q48, H46Q20 , or 800 MHz for H16Q20. A ^{1}H-^{15}N-HSQC was used to determine ^1H_N and ^{15}N amide chemical shifts[9]. For the determination of $^1H_{aliphatic}$ and $^{13}C_{aliphatic}$ chemical shifts a ^{1}H-^{13}C HSQC was used[10]. For H16Q20, H16Q24, H16Q28 and H16Q32 ^{1}H-^{15}N HSQC spectra were acquired in ~21 hours using 512 scans, 128 increments and a spectral width of 21 ppm in the indirect dimension. ^{1}H-^{13}C-HSQC spectra were acquired in 28 hours using 512 scans, 128 increments and spectral width of 60 ppm in the indirect dimension. In order to observe the *cis/trans* isomerization of H16Q48 ^{1}H-^{15}N-HSQC spectra were acquired in 42 hours using 1024 scans, 128 increments and a spectral width of 21 ppm in the indirect dimension. The ^{1}H-^{13}C-HSQC spectra were acquired with the same settings as those for the other Qs. For H46Q20 the ^{1}H-^{15}N HSQC spectrum was acquired over 23 hours using 736 scans, 96 increments and a spectral width of 16 ppm in the indirect dimension. Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the ^{1}H/X frequency ratio of the zero point according to Markley *et al*.[11]. For the sequential assignments of H16 ISOGRO® -Gln -Glu (39 µM), four triple resonance (3D) experiments were used: HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB.

Random coil chemical shifts were predicted according to the Poulsen IDP/IUP random coil chemical shift predictor (https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/)[12–14] or to Tamiola's neighbor corrected IDP library (http://nmr.chem.rug.nl/ncIDP/)[15]. Secondary chemical shifts (SCS) were obtained by subtracting the predicted value from the experimental one ($SCS=\delta_{exp}-\delta_{pred}$). For a better reliability of the results regarding possible referencing errors, we used the combined Cα and Cβ secondary chemical shifts (SCS(Cα)-SCS(Cβ)).

Results and Discussion

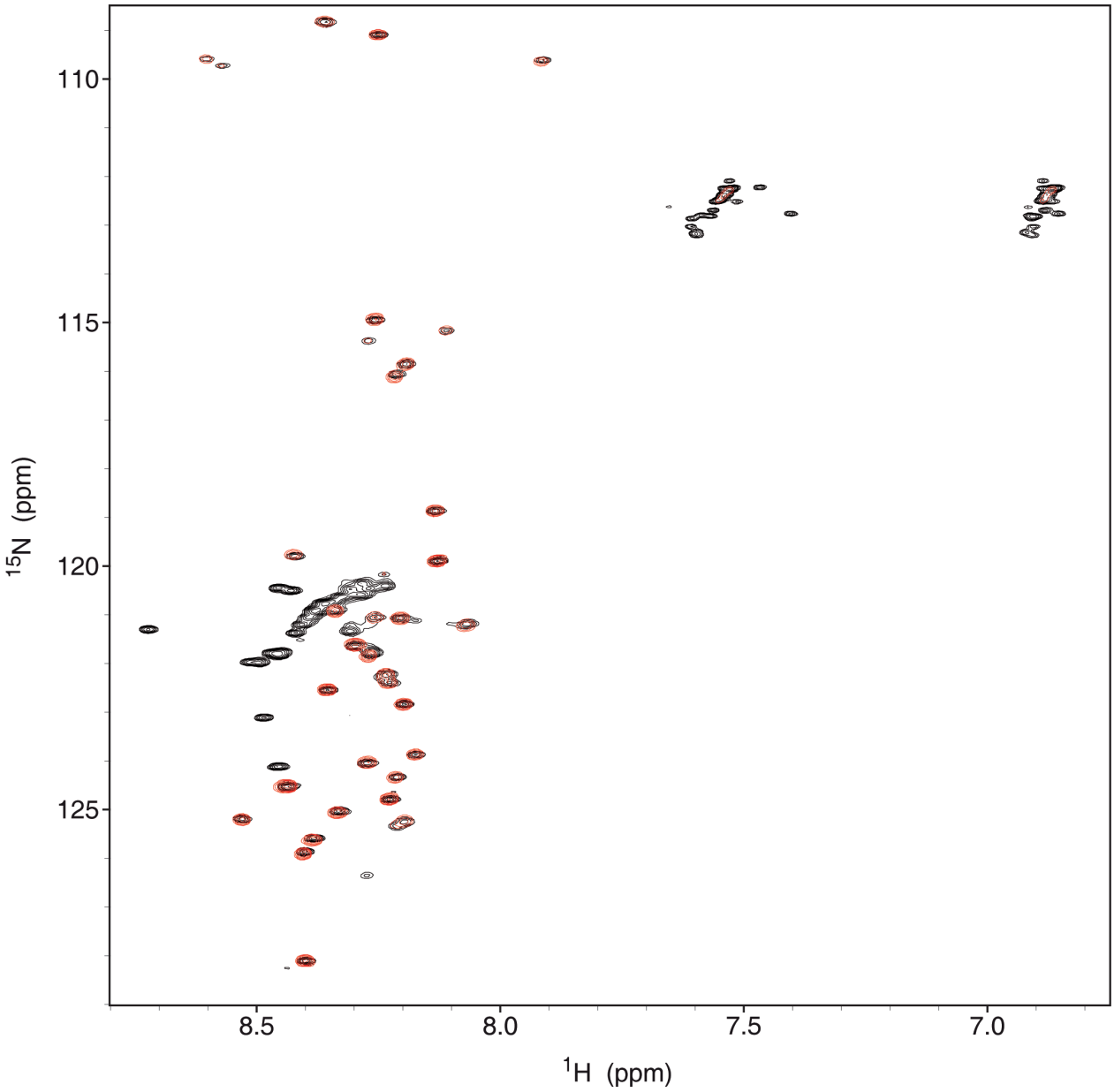

Figure S1: Overlay of ISOGRO® +Q/+E and -Q/-E ^{1}H-^{15}N HSQC NMR spectra. The spectrum of fully labeled htt exon1 is shown in black. The spectrum of partially labeled htt exon1 (glutamines and glutamates not labeled) is shown in red.

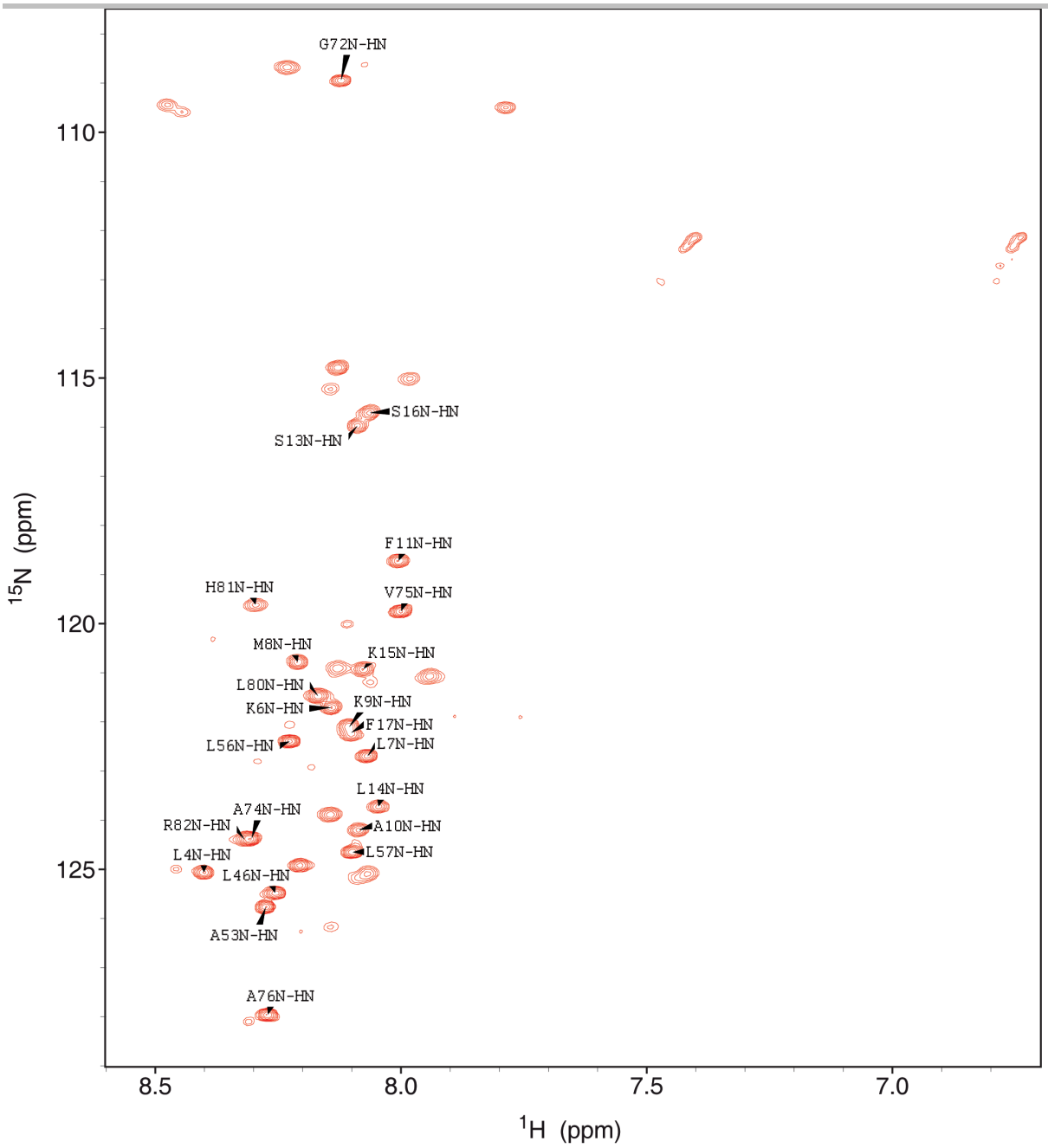

Figure S2: Assigned ISOGRO® -Q/-E ^{1}H-^{15}N HSQC NMR spectrum. The spectrum of partially labeled htt exon1 (glutamines and glutamates not labeled) is shown in red. Additional peaks, corresponding to sfGFP or the linker region, are not assigned.

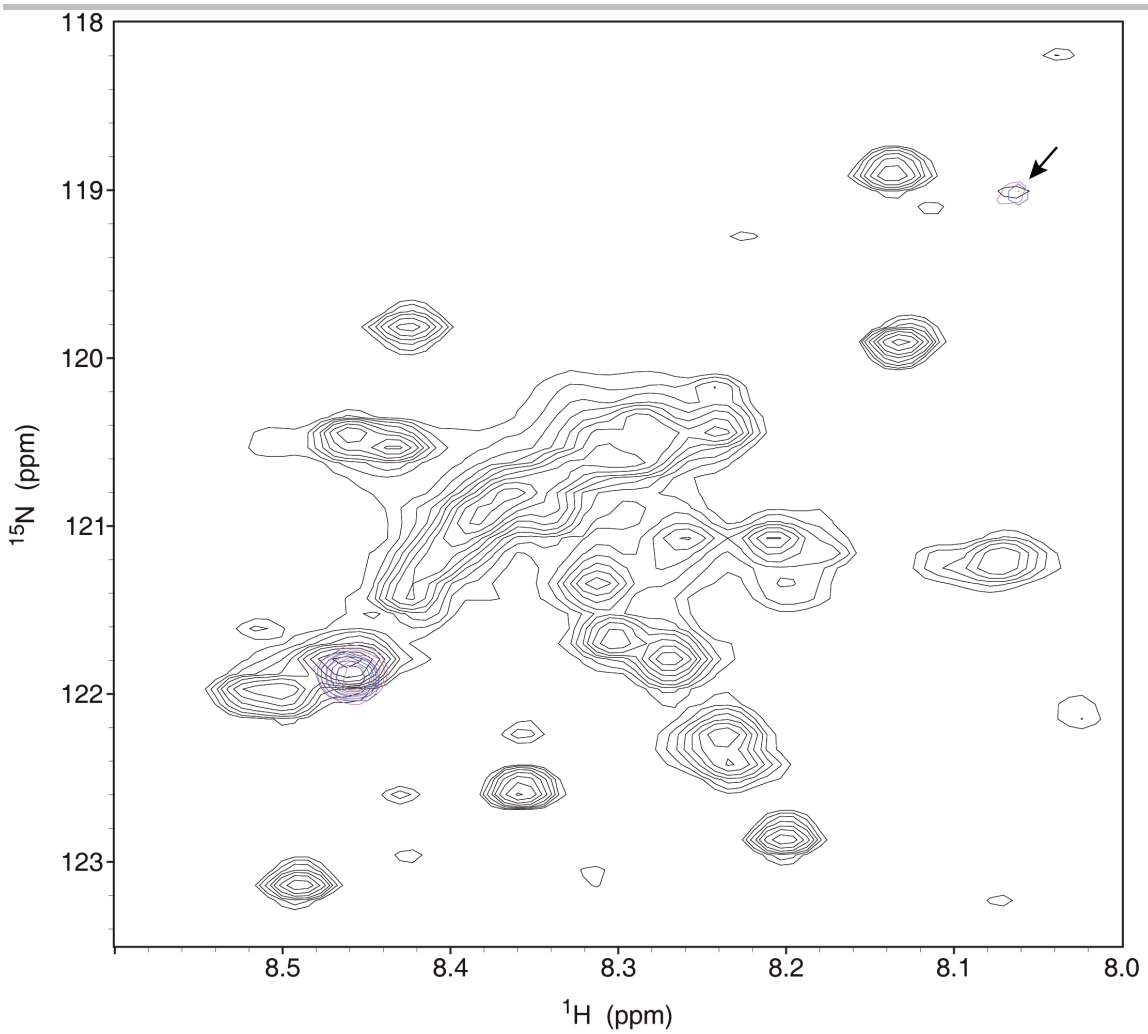

Figure S3: Spectra overlay of ISOGRO + Q (grey) and site-specifically labeled Q48 (recorded on two different samples; magenta and blue). Our labeling strategy allows us to identify peaks resulting from the proline *cis-trans* isomerization (see black arrow). The second, less intense peak, corresponds to Q48 preceding a *cis*-proline, and has a population of ≈1:0.2.

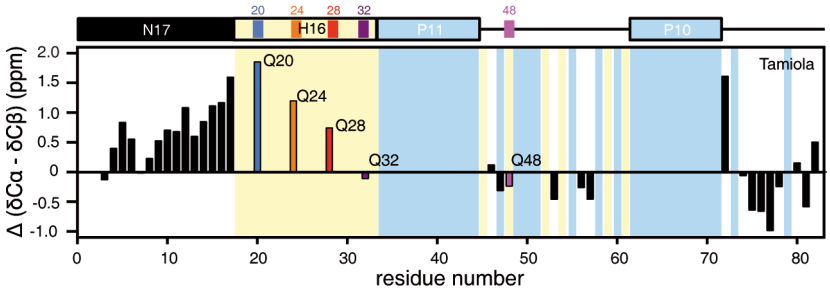

Figure S4: Secondary chemical shift analysis using Cα and Cβ chemical shifts using Tamiola *et al.*'s random-coil library[15]. A schematic representation of htt exon1 is shown to indicate the different domains and the positions of Q and P residues are highlighted in yellow and blue, respectively. The results obtained with this random coil library are in excellent agreement with the ones obtained using the Poulsen IDP/IUP random coil chemical shift predictor[13,14] shown in Figure 2g in the main text.

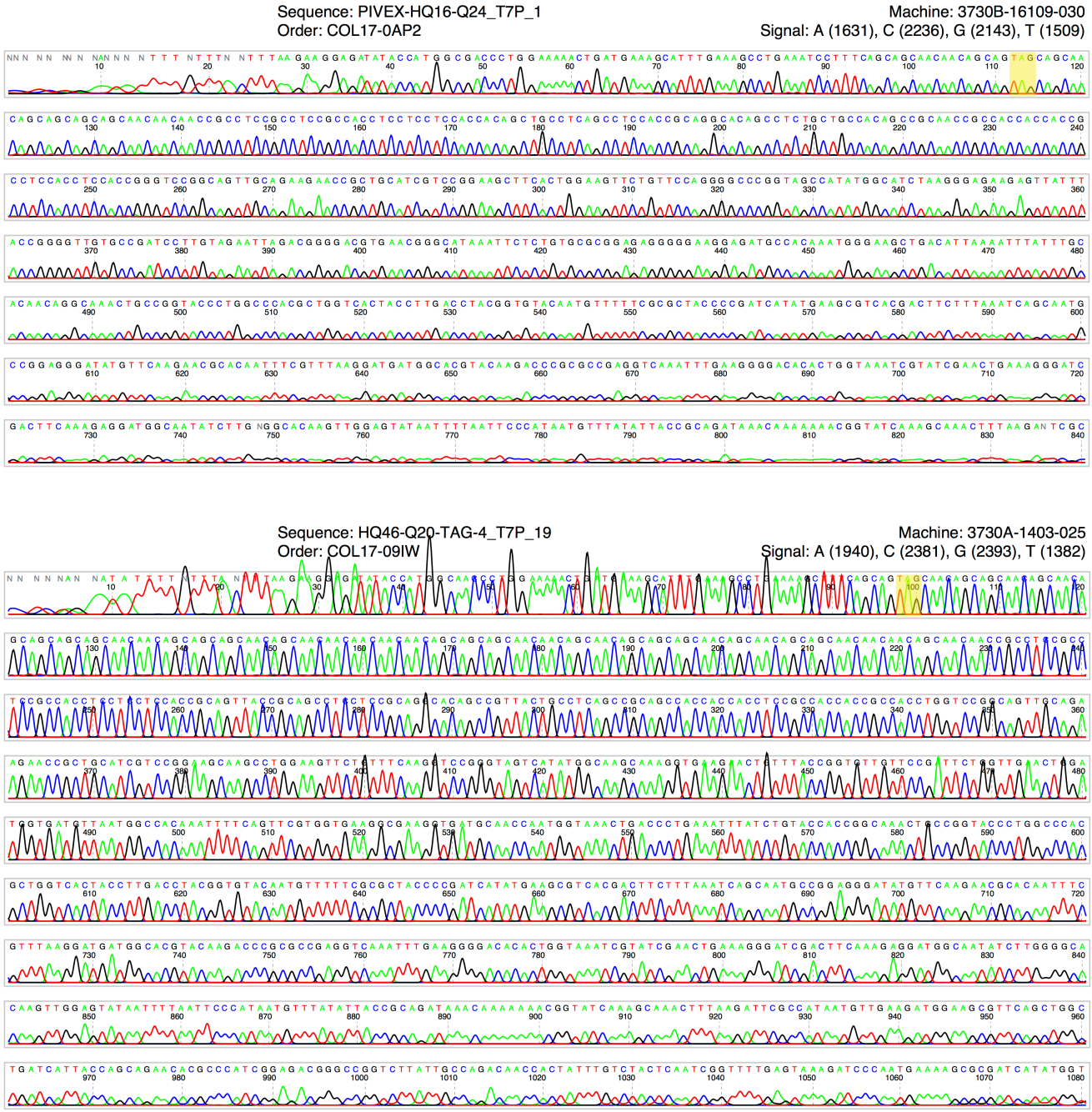

Figure S5: Sequencing results for H16Q24 (top) and H46Q20 (bottom). The chromatograms of the forward sequencing reactions are shown, with A in green, C in blue, G in black and T in red. The amber codon mutation is highlighted in yellow. Poly-Q and Poly-P tracts are well resolved. The integrity of the C-terminal sfGFP fusion was confirmed by reverse sequencing (data not shown).

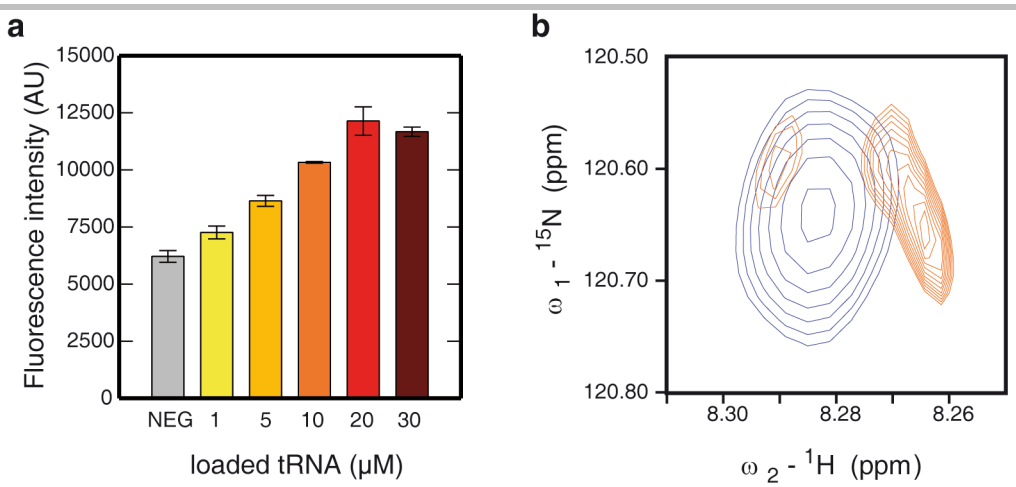

Figure S6: H46Q20 experiments. a) Titration of cell-free reactions containing a plasmid for H46 with an amber codon at position Q20 with increasing concentrations of loaded $tRNA_{CUA}$. Increasing concentrations of loaded $tRNA_{CUA}$ yielded increasing amounts of H46Q20. NEG - negative control without loaded $tRNA_{CUA}$. b) ^{1}H-^{15}N HSQC NMR spectrum showing two backbone peaks for Q20 in the context of H46 (orange) overlaid with the single peak found for that residue in the context of H16 (blue).

Table S1: Extracted N, HN, Cα, Cβ, Cγ, Nε2, Hα, Hβ, Hγ and Hε2 chemical shifts, Cα-Cβ secondary chemical shifts and sample concentrations for H16Q20, H16Q24, H16Q28, H16Q32, H16Q48 and H46Q20. Random coil chemical shifts of htt exon1 were predicted according to Poulsen[13,14] or Tamiola[15].

	Residue #					
	H16Q20	**H16Q24**	**H16Q28**	**H16Q32**	**H16Q48**	**H46Q20**
N	120.68	120.69	120.99	121.81	121.89	120.599 (minor) 120.655 (major)
HN	8.289	8.328	8.382	8.448	8.456	8.219 (minor) 8.265 (major)
Cα	57.24	56.79	56.4	55.8	53.43	-
Cβ	29.04	29.13	29.29	29.48	29.14	-
Cγ	33.93	33.87	33.86	33.79	33.518	-
Nε2	112.24	112.27	112.42	112.71	113.21	-
Hα	4.204	4.239	4.278	4.312	4.59	-
$H\beta^1$ / $H\beta^2$	2.056 / 2.115	2.041 / 2.125	2.028 / 2.120	1.982 / 2.092	1.916 / 2.082	-
$H\gamma^1$ / $H\gamma^2$	2.372 / 2.4234	2.395	2.392	2.367	2.4182	-
$H\varepsilon2^1$ / $H\varepsilon2^2$	6.866 / 7.469	6.868 / 7.529	6.8826 / 7.546	6.8852 / 7.566	6.909 / 7.599	-
Cα – Cβ (Poulsen)	1.501	0.961	0.411	-0.024	-0.182	-
Cα – Cβ (Tamiola)	1.78	1.24	0.69	-0.1	-0.22	-
[protein]	6.3 µM	2.4 µM	8.4 µM	6.1 µM	5.4 µM	2.3 µM
Number of scans						
^{1}H-^{15}N-HSQC	512	512	512	512	1024	736
^{1}H-^{13}C-HSQC	512	512	512	512	512	-
Time (h)						
^{1}H-^{15}N-HSQC	20	21	21	21	42	23
^{1}H-^{13}C-HSQC	31	28	28	28	28	-

References

[1] J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2006**, *24*, 79–88.

[2] E. F. Whelihan, P. Schimmel, *EMBO J.* **1997**, *16*, 2968–2974.

[3] C. Kao, M. Zheng, S. Rüdisser, *RNA* **1999**, *5*, 1268–1272.

[4] F. W. Studier, *Protein Expr. Purif.* **2005**, *41*, 207–234.

[5] S. E. Walker, K. Fredrick, *Methods* **2008**, *44*, 81–86.

[6] M. A. Apponyi, K. Ozawa, N. E. Dixon, G. Otting, *Methods Mol. Biol.* **2008**, *426*, 257–268.

[7] K. V Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chemie - Int. Ed.* **2012**, *51*, 2243–2246.

[8] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, S. Yokoyama, *FEBS Lett.* **1999**, *442*, 15–19.

[9] S. Mori, C. Abeygunawardana, M. O. Johnson, P. C. van Zijl, *J. Magn. Reson. B* **1995**, *108*, 94–98.

[10] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. J. Glaser, O. W. Sørensen, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 301–306.

[11] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J. Mol. Biol.* **1998**, *280*, 933–952.

[12] S. Schwarzinger, G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright, H. J. Dyson, *J. Am. Chem. Soc.* **2001**, *123*, 2970–2978.

[13] M. Kjaergaard, F. M. Poulsen, *J. Biomol. NMR* **2011**, *50*, 157–165.

[14] M. Kjaergaard, S. Brander, F. M. Poulsen, *J. Biomol. NMR* **2011**, *49*, 139–149.

[15] K. Tamiola, B. Acar, F. A. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003.

Author Contributions

P.B. and S.D. conceived the project. F.A. and A.F. designed and cloned the protein constructs. A.F. and F.A. produced GlnRS. P.B., A.U. and A.M. designed and analyzed the suppression experiments. A.U. and A.M. prepared the protein NMR samples. N.S. designed, and E.D., A.U., M.P. and N.S. performed and analyzed the NMR experiments. The manuscript was prepared by A.U. and P.B. with input from all coauthors.