

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*

Abstract: Homorepeat (HR) proteins are involved in key biological processes and multiple pathologies, however their high-resolution characterization has been impaired due to their homotypic nature. To overcome this problem, we have developed a strategy to isotopically label individual glutamines within HRs by combining nonsense suppression and cell-free expression. Our method has enabled the NMR investigation of huntingtin exon1 with a 16-residue polyglutamine (poly-Q) tract, and the results indicate the presence of an N-terminal α -helix at near neutral pH that vanishes towards the end of the HR. The generality of the strategy was demonstrated by introducing a labeled glutamine into a pathological version of huntingtin with 46 glutamines. This methodology paves the way to decipher the structural and dynamic perturbations induced by HR extensions in poly-Q-related diseases. Our approach can be extended to other amino acids to investigate biological processes involving proteins containing low-complexity regions (LCRs).

While the majority of protein sequences are aperiodic and feature most of the 20 proteinogenic amino acids, many protein sequences harbor low-complexity regions (LCRs). LCRs are unusually simple sequences with a strong bias in their amino acid composition.^[1] A subfamily of LCRs termed homorepeats (HRs) presents sequences with repetitions of just one amino acid. HRs are abundant in eukaryotes, in which around 15% of the proteome contains at least one repeat, and this represents an important source of genetic variability. Furthermore, HRs are conserved in specific functional classes of proteins, such as transcription factors and developmental proteins, where they increase functional

versatility by mediating biomolecular interactions and facilitate spatial organization.^[2,3] The abundance of HR-hosting proteins is tightly regulated due to their potential detrimental effects, which explains why more than 40% of proteins with multiple HRs are involved in human diseases.^[4] Although HRs from all amino acids have been reported, the most abundant are hydrophilic, and are characterized by a high level of disorder, which in many cases precludes crystallization.^[2,5,6] As a consequence, they are underrepresented in the Protein Data Bank (PDB), thus highlighting the need for novel approaches for their structural and dynamic characterization.

The most common type of HRs in eukaryotes are polyglutamine (poly-Q) repeats.^[2] Moreover, poly-Q expansions form the basis of at least ten neurodegenerative disorders, including Huntington's Disease (HD), Kennedy's disease, and several ataxias.^[6,7] HD is the most abundant and most well studied poly-Q-related disorder. The causative agent of this deadly neurodegenerative pathology is the huntingtin protein (htt).^[4,7] Although full-length htt has more than 3000 residues, the first exon (exon1; Figure 1a), which contains the poly-Q tract, is sufficient to replicate much of the pathology in cell and animal models.^[8] Importantly, only individuals with more than 35 consecutive glutamines (pathological threshold) develop HD symptoms, and disease onset and severity are correlated to the length of the poly-Q expansion.^[7] All poly-Q-related diseases present similar pathological thresholds, thus suggesting a common amyloidogenic mechanism.^[9]

A hallmark of HD is the presence of neuronal inclusions containing htt exon1 aggregates that have been linked to the aggregation behavior of poly-Q constructs in vitro.^[7,9] So far, structural studies on htt exon1 constructs in solution have used computational^[10] or low-resolution approaches,^[10–13] and focused mainly on the overall properties of the poly-Q tract. Unfortunately, these studies have reported contradictory observations regarding the conformational preferences of the HR. Similar observations have been obtained from crystallographic studies of non-aggregated poly-Q constructs: while a peptide of 10 consecutive glutamines in complex with an antibody showed an extended structure,^[14] a construct with 17 Qs in the context of htt exon1 showed multiple conformations (helix, random coil, and extended loop), and only the initial residues were visible on the map.^[15] More recently, Baias et al. assigned the backbone resonances and studied the secondary-structure propensities of a shortened exon1 construct using pH titrations and high-dimensional (5D) NMR experiments.^[16] While this study managed to assign 17 glutamines, it is questionable whether this approach could

[*] Dr. A. Urbanek, Dr. A. Morató, Dr. F. Allemand, Dr. E. Delaforge, A. Fournet, Dr. M. Popovic, Dr. N. Sibille, Dr. P. Bernadó
Centre de Biochimie Structurale (CBS), INSERM, CNRS
Université de Montpellier
29 rue de Navacelles, 34090 Montpellier (France)
E-mail: pau.bernado@cbs.cnrs.fr

Dr. S. Delbecq

Laboratoire de Biologie Cellulaire et Moléculaire
(LBCM-EA4558 Vaccination Antiparasitaire), UFR Pharmacie
Université de Montpellier, Montpellier (France)

Supporting information for this article can be found under:
<https://doi.org/10.1002/anie.201711530>.

© 2018 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial, and no modifications or adaptations are made.

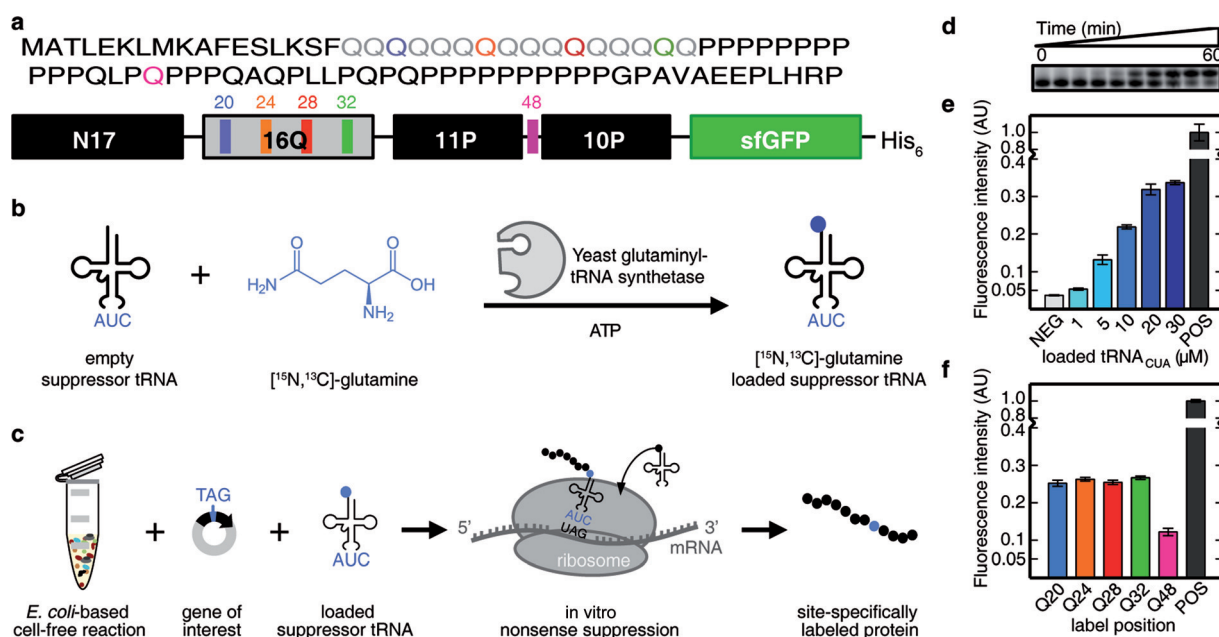


Figure 1. General Scheme, validation, and optimization of the method. a) Amino acid sequence of huntingtin exon1 containing 16 consecutive glutamines and schematic representation of the sfGFP fusion construct used in this study. N17 = 17 N-terminal residues; 16Q = polyglutamine tract with 16Q; 11P and 10P = polyproline tracts; sfGFP = superfolder GFP. The glutamines studied by NMR are highlighted in different colors (Q20 blue, Q24 orange, Q28 red, Q32 green, and Q48 magenta). The same color scheme is used throughout the manuscript. b) Enzymatic loading of suppressor tRNA_{CUA}. In vitro transcribed suppressor tRNA_{CUA} is incubated with the yeast glutaminyl-tRNA synthetase, $[^{15}\text{N},^{13}\text{C}]$ -glutamine, and ATP to yield loaded suppressor tRNA_{CUA}. c) Site-specific labeling is achieved by adding the loaded suppressor tRNA_{CUA} to the cell-free reaction in the presence of a plasmid coding for htt exon1 with an amber stop codon replacing the codon for the residue to be studied by NMR. d) Time course following the enzymatic loading of suppressor tRNA_{CUA}. Upper and lower bands correspond to loaded and unloaded suppressor tRNA_{CUA}, respectively. Samples were taken at $t = 0, 0.16, 0.5, 1, 2, 5, 10, 15, 30,$ and 60 min. e) Titration of cell-free reactions with increasing concentrations of loaded tRNA_{CUA}. The plasmid with an amber stop codon replacing the codon for Q20 (H16Q20) was used. NEG = negative control without suppressor tRNA_{CUA}. POS = positive control using a plasmid without the amber stop codon. f) Screening of several constructs for which the amber stop codon was placed at the position of Q20, Q24, Q28, Q32, or Q48 using 10 μM of loaded suppressor tRNA_{CUA}.

be applied to constructs with longer poly-Q tracts where the similarity of the chemical environments within the poly-Q would yield poorly resolved NMR spectra. The lack of general approaches to perform NMR studies on poly-Q-hosting proteins precludes the understanding of structural features such as residue-specific conformational preferences, structural cooperativity within HRs, and the presence of intramolecular interactions. Consequently, the structural basis of the pathological threshold, and the putative presence of a subpopulation of toxic conformers beyond this threshold remain speculative.

Herein, we present a strategy that overcomes the aforementioned limitations and enables the detailed characterization of poly-Q HRs independent of their lengths. We site-specifically introduce a unique $[^{15}\text{N},^{13}\text{C}]$ -labeled glutamine into an otherwise unlabeled htt exon1 poly-Q tract with 16 glutamines (H16) by combining nonsense suppression^[17] and cell-free (CF) protein synthesis^[18] (Figure 1). In contrast to earlier studies,^[17,19,20] our method makes use of a synthetic orthogonal and in vitro transcribed suppressor tRNA/glutaminyl-tRNA synthetase (GlnRS) pair based on the tRNA^{Gln} from *Saccharomyces cerevisiae* (Figure 1b). The GlnRS allows rapid (<60 minutes) charging of up to 90% of the nonsense suppressor tRNA (tRNA_{CUA}) with commercially

available $[^{15}\text{N},^{13}\text{C}]$ -glutamine (Figure 1d). The loaded tRNA_{CUA} is then added to a CF reaction that uses release factor-1 (RF1)-depleted *Escherichia coli* extracts originating from BL21 Star (DE3)::RF1-CBD₃ cells, which has been shown to increase the yield of CF suppression reactions.^[18] In order to determine the optimal suppression conditions, we optimized the CF reaction conditions on a small scale (50 μL) by adding increasing amounts of loaded tRNA_{CUA} (Figure 1e). To monitor the titration, we used a plasmid coding for the htt exon1 construct fused to superfolder GFP (sfGFP), in which the Q20 codon was substituted with the amber stop codon (see Figure 1a and the Supporting Information).

The titration showed that increasing concentrations of loaded tRNA_{CUA} resulted in better suppression yields (ca. 37% of the positive control at 30 μM of loaded tRNA_{CUA}). However, the yield increase was not linear above concentrations of 10 μM . Thus, we used 10 μM of loaded tRNA_{CUA} for large-scale protein production. We subsequently investigated the position effect on the suppression efficiency by placing the amber stop codon in a set of non-consecutive glutamine positions spanning the whole htt exon1, including the HR tract (Q20, Q24, Q28, Q32) and the proline-rich region (Q48; Figure 1a). While the suppression efficiency within the poly-Q tract remained stable (ca. 25%), the yield of H16Q48 was

approximately half (Figure 1f). We attribute this lower efficiency to the intercalated position of the glutamine between two prolines (Figure 1a), which present the slowest translational velocity among the 20 natural amino acids.^[21]

Using optimized CF conditions, we prepared residue-specifically [¹⁵N,¹³C]-labeled htt exon1 samples of the aforementioned glutamine positions. A 5 mL reaction was sufficient to conduct ¹⁵N and ¹³C heteronuclear NMR experiments and structurally probe these residues (Figure 2). The simplicity of the spectra allowed the precise measurement of backbone and side-chain chemical shifts, thus avoiding the inherent overlap present in the fully labeled version of the protein (Figure 2b,d and Table S1). Inspection of the spectra regarding the different glutamine residues indicates a reasonable dispersion of the ¹H-¹⁵N correlations for residues closer to the N17 region, however Q32 (at the end of the HR) and Q48 (in the proline-rich region) overlap (Figure 2d). Interestingly, the Q48 ¹H-¹⁵N HSQC presents a second less-intense peak (ca. 1:0.2) that corresponds to the population of this residue when preceding a *cis*-proline (Figure 2b and Figure S3). The dispersion of the ¹H-¹³C correlations diminishes along the side chain (from H α -C α to H γ -C γ), thus indicating

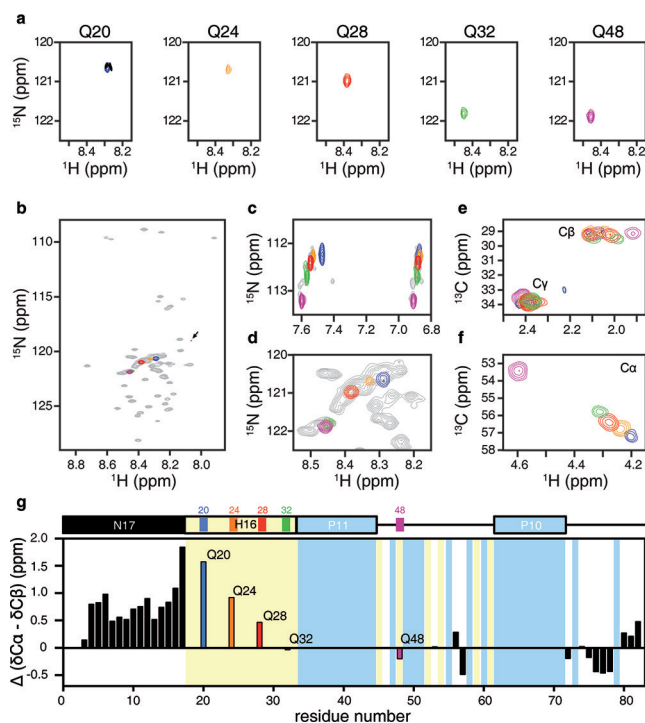


Figure 2. NMR analyses of H16. a) Individual backbone resonances of site-specifically labeled glutamines H16Q20, H16Q24, H16Q28, H16Q32, and H16Q48 in context of htt exon1. b) Overlay of the single-residue ¹H-¹⁵N HSQC spectra of H16 with that of the fully labeled H16 htt exon1. The arrow indicates the population of Q48 preceding a *cis*-proline. c) Enlarged view of the H16 ¹H-¹⁵N HSQC spectra showing the glutamine side chains. d) Zoom on the poly-Q region of the H16 ¹H-¹⁵N HSQC spectra showing individual Qs overlaid on fully labeled htt exon1. e) ¹H-¹³C HSQC NMR spectra showing the C β and C γ (e) and the C α (f) regions. g) Secondary chemical-shift analysis on H16 using experimental C α and C β chemical shifts and a random-coil library.^[22] 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.

an increase in mobility and a reduced influence of the backbone conformation.

We then proceeded to the assignment of H16 by applying 3D NMR experiments to two samples produced under CF conditions using [¹⁵N,¹³C]-labeled algal extract with or without added [¹⁵N,¹³C]-glutamine (Figure S1 in the Supporting Information). Following this strategy, 28 out of the 30 non-proline, non-glutamine backbone resonances of htt exon1 were assigned (Figure S2). Access to backbone chemical shifts enabled the derivation of secondary-structure elements of the protein. Secondary chemical shift (SCS) values for H16 at pH 6.5 were computed as the difference between the measured chemical shifts and their amino acid specific random-coil values using neighbor amino acid sequence corrections employing a well-established random-coil chemical-shift database.^[22] Figure 2g, which displays the SCS difference between C α and C β for htt exon1, indicates that the 17 N-terminal residues (N17) have a strong propensity for α -helical structure that increases when approaching the HR, which is in agreement with previous NMR studies.^[16,23] Along the poly-Q tract, a systematic decrease in helical propensity is observed, with Q32 presenting random-coil chemical shifts. The helical tendency within the poly-Q region causes the up-field shift of the glutamine peak cluster in the ¹H-¹⁵N-HSQC (Figure 2d), and increases the dispersion of the H α -C α correlations (Figure 2f). Finally, at the end of the C-terminal proline-rich region, deviations from the random-coil chemical shifts become less profound, with a greater tendency towards extended structure, which is most likely due to an enrichment in polyproline-II conformations.

We explored the potential for applying our strategy to longer poly-Q tracts by isotopically labeling Q20 in htt exon1 with 46 consecutive glutamines (H46Q20). An NMR sample was obtained from a 10 mL CF production upon addition of 10 μ M [¹⁵N,¹³C]-glutamine-loaded tRNA_{CUA} (Figure S6). Interestingly, the ¹H-¹⁵N HSQC displays two peaks for Q20 in the context of a longer poly-Q tract (Figure S6). This observation suggests the coexistence of at least two conformations in slow exchange that are not present in the non-pathological construct. Further experiments will be necessary to fully characterize the origin of this difference.

In summary, we present a novel strategy that combines CF expression and nonsense suppression to gain high-resolution structural and dynamic information on inaccessible poly-Q regions. We demonstrated the robustness of our approach using huntingtin, the causative agent of HD, as a model protein, and the results reveal the presence of helical regions at the N-terminus, encompassing N17 and the first portion of the poly-Q tract. With this method, NMR observables such as residual dipolar couplings and paramagnetic relaxation enhancement can be precisely measured to decipher local conformational propensities, structural cooperativity, and intramolecular interactions in HR regions. The experiments performed on a pathological htt (H46) demonstrate that, in contrast to other strategies based on the use of high-dimensional NMR spectra, our method is independent of the length of the poly-Q expansion.^[16,24] Consequently, this approach paves the way for investigation of the structural basis of the pathological threshold found in at least ten different poly-Q

diseases. The development of other orthogonal tRNA/synthetase pairs or the use of established synthetic tools, such as flexizyme^[25] or chemical aminoacylation,^[17] to load nonsense tRNA with natural amino acids could extend the described strategy for the exploration of other LCRs. In this way, this important but structurally elusive family of proteins can be investigated at high resolution to decipher the structural basis of crucial biological processes and to guide drug design.

Acknowledgements

The authors thank Gottfried Otting 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] and Labex EpiGenMed, an « Investissements d'avenir » program (ANR-10-LABX-12-01) awarded to P.B.. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), two national infrastructures supported by the French National Research Agency (ANR-10-INSB-04-01 and ANR-10-INSB-05, respectively). A.U. is supported by a grant from the *Fondation pour la Recherche Médicale* (SPF20150934061). The authors thank Lionel Imbert, IBS cell-free facility, for his technical help and valuable advice. This work used the Cell-Free facility at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from Instruct (PID: 1552) within the Grenoble Partnership for Structural Biology (PSB).

Conflict of interest

The authors declare no conflict of interest.

Keywords: genetic code expansion · homorepeat proteins · huntingtin · NMR spectroscopy · protein structures

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 3598–3601
Angew. Chem. **2018**, *130*, 3660–3663

- [1] J. C. Wootton, *Curr. Opin. Struct. Biol.* **1994**, *4*, 413–421.
- [2] J. Jorda, A. V. Kajava, *Adv. Protein Chem. Struct. Biol.* **2010**, *79*, 59–88.
- [3] S. Chavali, P. L. Chavali, G. Chalancon, N. S. de Groot, R. Gemayel, N. S. Latysheva, E. Ing-Simmons, K. J. Verstrepen, S. Balaji, M. M. Babu, *Nat. Struct. Mol. Biol.* **2017**, *24*, 765–777.
- [4] S. Karlin, L. Brocchieri, A. Bergman, J. Mrazek, A. J. Gentles, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 333–338.

- [5] J. Jorda, B. Xue, V. N. Uversky, A. V. Kajava, *FEBS J.* **2010**, *277*, 2673–2682.
- [6] A. L. Darling, V. N. Uversky, *Molecules* **2017**, *22*, 2027.
- [7] A. Adegbiyuro, F. Sedighi, A. W. Pilkington, S. Groover, J. Legleiter, *Biochemistry* **2017**, *56*, 1199–1217.
- [8] L. Mangiarini, K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trotter, H. Lehrach, S. W. Davies, et al., *Cell* **1996**, *87*, 493–506.
- [9] H. Y. Zoghbi, H. T. Orr, *Annu. Rev. Neurosci.* **2000**, *23*, 217–247.
- [10] J. B. Warner, K. M. Ruff, P. S. Tan, E. A. Lemke, R. V. Pappu, H. A. Lashuel, *J. Am. Chem. Soc.* **2017**, *139*, 14456–14469.
- [11] A. Bhattacharyya, A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer, R. Wetzel, *J. Mol. Biol.* **2006**, *355*, 524–535.
- [12] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R. V. Pappu, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16764–16769.
- [13] C. Peters-Libeu, J. Miller, E. Rutenber, Y. Newhouse, P. Krishnan, K. Cheung, D. Hatters, E. Brooks, K. Widjaja, T. Tran, et al., *J. Mol. Biol.* **2012**, *421*, 587–600.
- [14] P. Li, K. E. Huey-Tubman, T. Gao, X. Li, A. P. West, M. J. Bennett, P. J. Bjorkman, *Nat. Struct. Mol. Biol.* **2007**, *14*, 381–387.
- [15] M. W. Kim, Y. Chelliah, S. W. Kim, Z. Otwinowski, I. Bezprozvanny, *Structure* **2009**, *17*, 1205–1212.
- [16] M. Baias, P. E. S. Smith, K. Shen, L. A. Joachimiak, S. Žerko, W. Koźmiński, J. Frydman, L. Frydman, *J. Am. Chem. Soc.* **2017**, *139*, 1168–1176.
- [17] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.
- [18] K. V. Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chem. Int. Ed.* **2012**, *51*, 2243–2246; *Angew. Chem.* **2012**, *124*, 2286–2289.
- [19] S. Peuker, H. Andersson, E. Gustavsson, K. S. Maiti, R. Kania, A. Karim, S. Niebling, A. Pedersen, M. Erdelyi, S. Westenhoff, *J. Am. Chem. Soc.* **2016**, *138*, 2312–2318.
- [20] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schulz, D. E. Wemmer, *J. Am. Chem. Soc.* **1992**, *114*, 7959–7961.
- [21] M. Y. Pavlov, R. E. Watts, Z. Tan, V. W. Cornish, M. Ehrenberg, A. C. Forster, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 50–54.
- [22] M. Kjaergaard, F. M. Poulsen, *J. Biomol. NMR* **2011**, *50*, 157–165.
- [23] A. K. Thakur, M. Jayaraman, R. Mishra, M. Thakur, V. M. Chellgren, I.-J. L. Byeon, D. H. Anjum, R. Kodali, T. P. Creamer, J. F. Conway, et al., *Nat. Struct. Mol. Biol.* **2009**, *16*, 380–389.
- [24] B. Eftekharzadeh, A. Piai, G. Chiesa, D. Mungianu, J. García, R. Pierattelli, I. C. Felli, X. Salvatella, *Biophys. J.* **2016**, *110*, 2361–2366.
- [25] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357–359.

Manuscript received: November 9, 2017

Revised manuscript received: December 28, 2017

Accepted manuscript online: January 23, 2018

Version of record online: March 7, 2018