Intrinsically Disordered Tardigrade Proteins Self-Assemble into Fibrous Gels in Response to Environmental Stress

Anas Malki, Jean-Marie Teulon, Aldo R. Camacho-Zarco, Shu-wen W. Chen, Wiktor Adamski, Damien Maurin, Nicola Salvi, Jean-Luc Pellequer, and Martin Blackledge*

anie_202109961_sm_miscellaneous_information.pdf
Material and methods

Protein expression, purification, and mutagenesis

CAHS-8 was cloned into pET-28b(+) between Ncol and Xhol cleavage sites with a C-terminal 8His-tag preceded by a TEV cleavage site (ENLYFQG). They were expressed in E. coli BL21(DE3) cells for 5 hours after induction at an optical density of 0.6 with 1 mM of isopropyl-β-d-thiogalactopyranoside. The unlabeled protein expression was performed in LB medium whilst isotopically labelled protein expression for NMR (\(^{15}\)N and \(^{13}\)C) was performed in M9 minimal media prepared in H\(_2\)O or in D\(_2\)O to produce \(^2\)H-labelled proteins.

The cysteine mutations (A134C and A185C) were made by site directed mutagenesis using the quick change method [1]. The proteins were sonicated in 50 mM Tris, pH = 8.0 and 150 mM NaCl (lysis buffer) and then boiled at 95°C for 20 min for standard nickel beads purification. The elution was performed in the same buffer with 300 mM imidazole, then the protein was incubated overnight with TEV protease while dialyzing against the lysis buffer, subjected to a second nickel beads affinity column and the flow-through was loaded into a Superdex 75/200 column for size exclusion chromatography (SEC) in 30 mM NH4HCO3 and lyophilized. The obtained proteins were weighed and resuspended in 50 mM phosphate buffer at pH = 6.5 and 100 mM NaCl (AFM buffer) or 500 mM (NMR buffer) at the desired concentration. Paramagnetic labelling was done using maleimide chemistry. In short, the protein was incubated overnight with 10 mM DTT, then dialyzed against 50 mM NaP (pH = 7.0) and 150 mM NaCl. A 5 time molar excess of maleimide reactant (4-maleimido-TEMPO dissolved in DMSO) was added for overnight incubation. The excess of label was then removed through in the NMR buffer.

ANP32a IDR was expressed and purified as described [2]. Concentrations were verified using a Thermo Fischer NanoDrop 2000.

Sequence of the IDR of ANP32a: DRDDKEAPDSDAEGYVEGLDDEEDEDEDEDEDEDEDEDEDEDEDEEDEEEDVSEEGEEEEDEEGYNDGEVDDEEDEDEEDEDEDEDEDEEDD

N\(_{\text{T}A\text{IL}}\) IDR was expressed and purified as described [3].

Sequence of the IDR of N\(_{\text{T}A\text{IL}}\): MSGGDGAYHKPTGGGAIENVLDNADIDLTEAHADQDARGWGGESGERWARQVSGHFVTLHGAERLEEETNDEDVSDIERIRRAMLAERRQEDSATHDEGNVGDHDEDDDAAAVAGIGGI

SH3A from CD2AP was expressed and purified as described [4].

NMR Spectroscopy

NMR experiments were measured on 600, 700, 850 and 950 MHz Bruker spectrometers in the NMR buffer using 3 mm NMR tubes with 10% D\(_2\)O. Backbone resonance assignment was performed using triple resonance experiments for Ca, CB and CO chemical shift correlation. Secondary Structure Propensity (SSP) and secondary chemical shift were done using the software SSP [5].

The relaxation rate experiments on the CAHS-8 free form were measured at 350 \(\mu\)M protein concentration whilst the experiments with the gel were measured at a total CAHS-8 concentration of 1.12 mM.

\(^{15}\)N R1\(_{p}\) rates were measured as described using a spin lock of 1.5 kHz [6] with a set of delays of 1, 20, 50, 70, 90, 120, 160 and 200 ms with the repetition of two delays. A recycle delay of 1.5s was used.

\(^{15}\)N R1 and \(^{1}H\)-\(^{15}\)N heteronuclear NOE rates were measured as described [7] with a set of delays for R1 of 0, 0.1552, 0.2716, 0.388, 0.5044, 0.6208, 0.7372, 0.9312 and 1.164 s with the repetition of one delay. A recycle delay of 1.5s was used.

\(^{15}\)N-\(^{1}H\) CSA/DD transverse cross correlated rates (\(\eta_{xy}\)) were measured as described [8,9] with a cross-relaxation delay of 50 ms.

The rates were extracted using an in-house software with Monte Carlo simulation-based noise-estimation. R2 rates were determined from R1 and R1\(_{p}\) as described [10].

NMRPipe [11] was used to process all spectra and CCPNMR Analysis [12] to analyze them. PRE effects were determined from the peak intensity ratios in a 15N HSQC between the paramagnetic state and the diamagnetic state obtained by addition of 5 mM ascorbic acid. A recycle delay of 1.575s was used.

For residual dipolar couplings (RDCs) measurement, the protein was aligned using ASLA Biotech pf1 phages at 15 mg/mL concentration. The scalar couplings were measured in CCPN and the RDCs...
values were obtained by subtracting the couplings in the isotropic and anisotropic (phage aligned) measurements.

DOSY were measured with gradient strength of 1.016, 4.269, 7.522, 10.775, 17.280, 20.533, 23.786, 27.039, 30.292, 33.544, 36.797, 40.050, 43.303, 46.556, 49.808 G/cm. ANP32a total concentration was 150 µM, NTAIL 300 µM and SH3A 150 µM.

SAXS
SAXS experiments were performed at BM29 BioSAXS in the ESRF Grenoble in the NMR buffer. The measurement used an incident beam wavelength of 0.99 Å and with a distance between the sample and the Pilatus3 2M detector of 2.867 m. This gives a scattering momentum transfer comprised between 0.0039 Å⁻¹ and 0.49 Å⁻¹. Every measurement comprised 10 frames which were analyzed and rejected in case of radiation damage. The solvent was measured before and after each sample and its scattering was subtracted from the sample’s. The data were processed and analyzed using ScAtter.

10000 conformers were calculated using flexible-meccano[13] assuming the helical domains α₁ (E95-F150) and α₂ (V167-L194) were either populated or not. Small angle scattering curves were calculated for each conformer using the program Crysol.[14] The ensemble selection algorithm ASTEROIDS[15] was used to calculate combinations of equally populated conformers that reproduced experimental data. Ensembles were calculated with 2 to 20 members. The optimal number was estimated to be 7. Radii of gyration were calculated for each conformer and the average over the best fitting ensemble is quoted in the main manuscript.

Dynamic Light Scattering (DLS)
DLS measurement were performed on a Malvern Zetasizer Nano S at an angle of 173° in a Malvern ZEN0040 cuvette. A temperature trend from 318K to 278K was used with 3 measurements (comprising 10 runs of 60s) performed every 5K after 300s of equilibration. The protein phase was determined from the correlation function.[16]

AFM
Images were recorded on a Bruker multimode 8 microscope with a Nanoscope V controller using the PeakForce Tapping mode in liquid and soft tapping in air. For liquid imaging, a silicon nitride cantilevers (ScanAsyst Fluid, Bruker AFM probes, Camarillo, CA, USA) with the following nominal characteristics: k = 0.7 N/m, Fq = 150 kHz, r = 20 nm were used. A semi-automated ScanAsyst mode was used by switching off the autogain. A scan rate of 1 Hz was commonly used with a setpoint of about 0.1 V. For imaging in air, etched silicon cantilevers (RFESP-75, Bruker AFM probes) with the following nominal characteristics: k = 3 N/m, Fq = 75 kHz, r = 8 nm were used. A target amplitude of 300 mV (~13 nm) with a setpoint of about 185 mV was commonly used with a scan rate of 1Hz. For air imaging, 2 µL of CAHS8 at 30 mg/mL, kept at 42°C, were deposited on mica or glass and dried for 5 min in air. For liquid imaging, 2 µL of CAHS8 at 30 mg/mL, kept at 42°C, was deposited on mica with 50 µL of AFM buffer, also kept at 42°C. Images were processed using Gwyddion[17]: raw images were leveled by mean-plane subtraction, rows were then aligned by the median method, strokes were corrected and the minimal intensity was set to 0. For further analysis, images were denoised using DeStripe[16] and improved visibility was performed using the Laplacian weight method[19,20].

REFERENCES
[1] J. Braman, C. Papworth, A. Greener, in In Vitro Mutagenesis Protocols (Ed.: M.K. Trower), Humana Press, Totowa, NJ, 1996, pp. 31–44.
[2] A. R. Camacho-Zarco, S. Kalayil, D. Maurin, N. Salvi, E. Delaforge, S. Milles, M. R. Jensen, D. J. Hart, S. Cusack, M. Blackledge, Nat Commun 2020, 11, 3656.
[3] K. Houben, D. Marion, N. Tarbouriech, R. W. H. Ruigrok, L. Blanchard, J. Virol. 2007, 81, 6807–6816.
[4] J. L. O. Roldan, M. Blackledge, N. A. J. van Nuland, A. I. Azuaga, J. Biomol. NMR 2011, 50, 103–117.
[5] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, Protein Sci. 2006, 15, 2795–2804.
[6] N.-A. Lakomek, J. Ying, A. Bax, Journal of Biomolecular NMR 2012, 53, 209–221.
[7] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, J. Am. Chem. Soc. 2016, 138, 6240–6251.
[8] P. Pelupessy, G. M. Espallargas, G. Bodenhausen, *Journal of Magnetic Resonance* **2003**, *161*, 258–264.

[9] P. Pelupessy, F. Ferrage, G. Bodenhausen, *J. Chem. Phys.* **2007**, *126*, 134508.

[10] M. Akke, A. G. Palmer, *J. Am. Chem. Soc.* **1996**, *118*, 911–912.

[11] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *Journal of Biomolecular NMR* **1995**, *6*, 277–293.

[12] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687–696.

[13] V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinformatics* **2012**, *28*, 1463–1470.

[14] D. Svergun, C. Barberato, M. Koch, *J. Appl. Crystallogr.* **1995**, *28*, 768–773.

[15] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *Journal of the American Chemical Society* **2010**, *132*, 8407–8418.

[16] M. Shibayama, T. Norisuye, *BCSJ* **2002**, *75*, 641–659.

[17] D. Nečas, P. Klapeřek, *Open Physics* **2012**, *10*, 181–188.

[18] S. W. Chen, J.-L. Pellequer, *BMC Struct Biol* **2011**, *11*, 7.

[19] S. W. Chen, J.-M. Teulon, C. Godon, J.-L. Pellequer, *J Mol Recognit* **2016**, *29*, 51–55.

[20] S.-W. W. Chen, A.-S. Banneville, J.-M. Teulon, J. Timmins, J.-L. Pellequer, *Nanoscale* **2020**, *12*, 22628–22638.

[21] M. R. Jensen, M. Blackledge, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1557–1558.

[22] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc* **2008**, *130*, 8055–8061.

[23] J. H. Brown, C. Cohen, D. A. Parry, *Proteins* **1996**, *26*, 134–145.

[24] A. N. Lupas, M. Gruber, in *Advances in Protein Chemistry*, Elsevier, **2005**, pp. 37–38.
Figure S1. Paramagnetic Relaxation Enhancement (PRE) experiment with the spin label on positions 185 (top) 134 (bottom). The absence of significant enhancement beyond the immediate neighboring residue indicates that the protein is extended and is lacking a tertiary structure. Values that are greater than 1 are non-physical, and are due to experimental noise, also in part due to potential differences in pH between oxidized and reduced samples.

Figure S2. Representations of the dimensions of CAHS-8 in the presence of the helical domains (A), and assuming that these regions sample random coil statistics (B). The two ensembles have radii of gyration of 59±11 and 47±10 Å respectively.
Figure S3. Temperature dependent conformational sampling. Temperature compensated secondary chemical shift for each backbone $^{13}\text{C}^a$ are shown for 278K (blue), 293K (red) and 303K (orange). A clear increase in helical propensity in the $\alpha_2$ helix is visible at 278K. Many peaks from the $\alpha_1$ helix are not visible at this temperature.
Figure S4. A Transverse $^{15}$N relaxation ($R_2$) measured at 303K (blue) and 278K (red) at 850 MHz, showing very large increase in molecular tumbling at the lower temperature. B Transverse $^{15}$N relaxation ($R_2$) of Measles virus nucleoprotein C-terminal disordered domain measured at 298K (blue) and 278K (red) at 850 MHz. The helical element (positions 490-500) is populated to approximately 90% in the central region.[21] C. Transverse $^{15}$N relaxation ($R_2$) of Sendai virus nucleoprotein C-terminal disordered domain measured at 298K (blue) and 278K (red) at 850 MHz.[7] The helical element (positions 478-491) is populated to approximately 80% in the central region.[22] The increase in relaxation rates (and disappearance of signal) in the helical region of CAHS-8 is significantly greater than for the two other viral proteins that do not show signs of oligomerization at 278K.
Figure S5. (A-Q) Liquid mode AFM images of the kinetic of fibrils formation. The 17 frames of the movies are separated by 8 min. The images show the fibrillation starting from a few already fully formed fibres to a mesh of longer and smaller fibrils. The blue arrow denotes oligomers thought to be either aborted fibrils (possibly “off pathway oligomers”) or prefibrillar oligomers. The green circle highlights small nascent circular fibrils that later strengthen to an early fibril. (R). Air mode AFM image of the gel obtained by letting the sample imaged in (A-Q) dry displaying a network of interwined fibres.
Image processing of Fig. 3A using the triple L-weight filter. The L-weight filter suggests a twisted pair of proteins stacking onto each other with their local long axis perpendicular to the fiber extension. The measured width of a single protein stacking is about 9 nm; which could correspond to the full length of the central folded α-helix. The original image has a physical size of 2 x 2 µm² with 512 x 512 pixel². The inset is made using a computer magnification (X4) using the Schaum interpolation method with Gwyddion. The size of the white square is 64 x 64 pixel² representing a physical size of 250 x 250 nm². Because the intensity values of processed images using the L-weight filter do not represent height values, there is no corresponding colour scale bar present on this filtered image. The colour ramp of pixel intensities has been manually adjusted to improve visibility.

Fibres are continuously growing and shrinking during AFM imaging and consequently, the thickness of fibers changes over time. In addition, some fibres lie on top of others, not directly in contact with the mica surface, and thus the tip-effect in such fibers is expected to be larger than for fibres directly adsorbed on the mica surface. The apparent twisting, evidenced by the change in height along a single fibre, is not expected to be an artefact of the tip-effect.

We have tested the processing on several images and obtained a similar result, suggesting a fibre made of pairs of proteins.
Left, the same L-weight filter processed image as in Fig. 6a and on the right the corresponding original height image. The original image has been cropped from Fig. 3a with a physical size of 250 x 250 nm² with 64 x 64 pixel², then the imaged was made using a computer magnification (X4) using the Schaum interpolation method with Gwyddion. Because the intensity values of processed images using the L-weight filter do not represent height values, there is no scale bar present on the left image whereas the color bar on the right represents true height values. It can be seen that the middle of the cropped fibre is flatter (about 3 nm) than the two twisted extremities (about 6 nm).
Comparison of the properties of folded and unfolded proteins in the isotropic phase and in the gel.

A – SH3A domain from CD2AP. Red - $^{15}$N-$^1$H HSQC spectrum in free solution, blue - spectrum in CAHS-8 gel. The intensity of the peaks is severely diminished in the gel, suggesting that a significant fraction of the protein becomes invisible. The visible fraction has the same fold, and (B) has a diffusion constant that is significantly slower than in free solution (Free : $2.18808203 \times 10^{-10}$ m$^2$s$^{-1}$ compared to $1.31157081 \times 10^{-10}$ m$^2$s$^{-1}$ in the gel).

The conditions are 150 µM concentration of protein, 298K, 600 MHz in NaP 50 mM, NaCl 500 mM, pH 6.5. The gel is at 20 mg/mL.

C – The disordered domain of Sendai virus nucleoprotein (N$^\text{Tails}$) (red - $^{15}$N-$^1$H HSQC spectrum in free solution, blue - spectrum in CAHS-8 gel). D - The protein has a diffusion constant that is again significantly slower than in free solution (Free : $1.83816502 \times 10^{-10}$ m$^2$s$^{-1}$ compared to $1.16630247 \times 10^{-10}$ m$^2$s$^{-1}$ in the gel).

Experimental conditions were 150 µM concentration of protein, 298K, 600 MHz in NaP 50 mM, NaCl 500 mM, pH 6.5. The gel is at 30 mg/mL.
Figure S8. Sequence alignment of CAHS-8 with other CAHS from *Hypsibius exemplaris* (HYPDU), the closely related tardigrade *Ramazzottius varieornatus* (RAMVA) and the more distant *Paramacrobiotus richtersi* (PARRC). A, B and C correspond respectively to CAHS-8 disordered N-terminus (M1–T94), its helical domain (E95–F194) and its disordered C-terminus (A195–N227). Blue arrows indicate hidden residues from the aligned proteins in order to visualize alignment on CAHS-8 only. The sequence alignment shows a significant conservation of T39 onward.

(B) The helical domain displays the highest amount of conservation and the analysis of its sequence reveals an irregular coiled coil [23,24] comprising 9 a-b-c-d-e-f-g canonicals heptads, 3 d-e-f-g stutters.
(M119-E120-K121-K122, A130-E131-A132-E133 and A173-K174-R175-E176) and one $d-f-f'$-g five residues insertion (L142-E143-K144-Q145-H146). The position $a$ and $d$ being preferentially occupied by hydrophobic amino acids while the $e$ and $g$ by charged residues of opposite charges. The irregularity of this coiled coil would make it unstable whose affinity may be modulated by conditions. In agreement with NMR measurements on the gel (fig.4), the helical domain of CAHS-8 would then be the fibrillation core with the coiled coil formation being its first step toward the formation of paired helical filament as already seen. The high conservation of the sequence between these CAHS indicate that they most probably have a similar structure (alpha helix that would assemble into coiled-coil), fibrillation and gelation properties.