Electronic Supplementary Information for

Histidine-based hydrogels via singlet-oxygen photooxidation

Michelle S. Liberato, Nayara G. S. Cavalcante, P. Abinaya Sindu, Mônica J. Rodrigues-Jesus, Pavel Zelenovskii, Ana C. O. Carreira, Maurício S. Baptista, Mari C. Sogayar, Luís C. S. Ferreira, Luiz H. Catalani

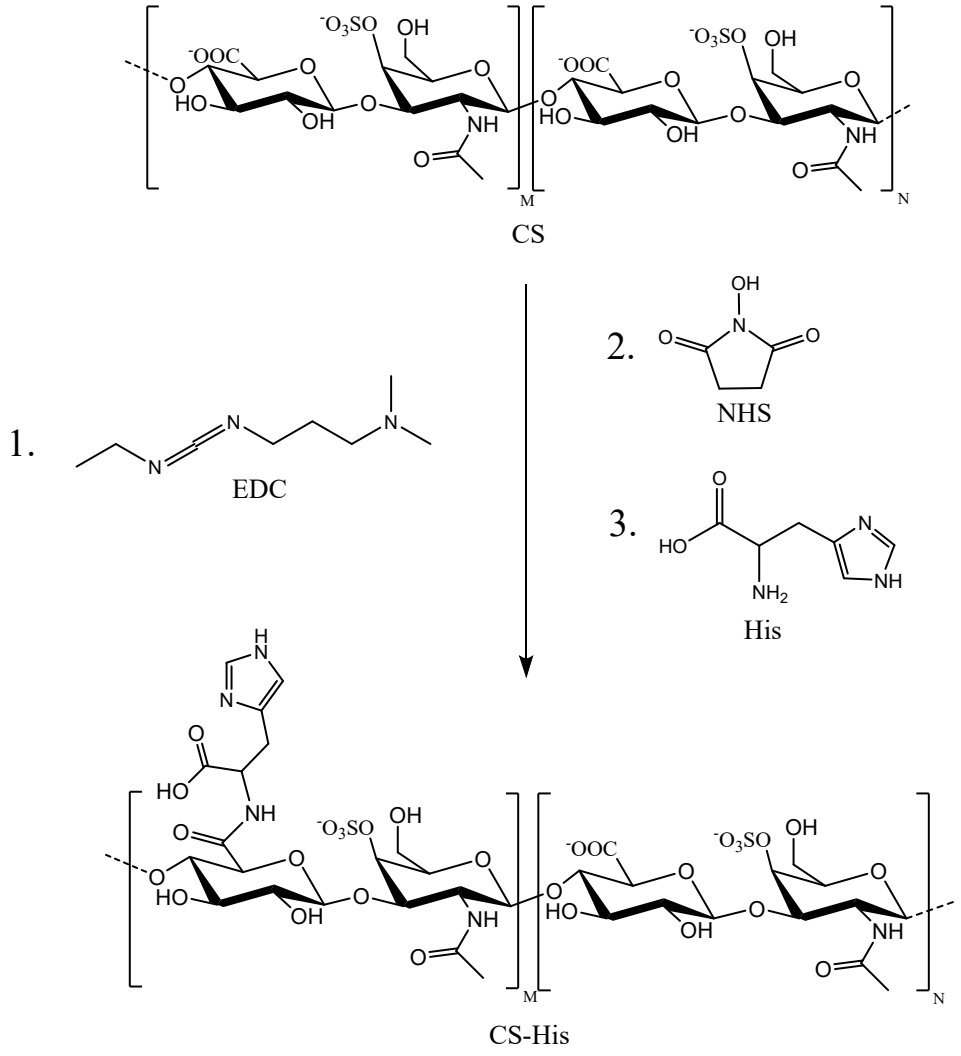

Figure S1. Scheme of CS functionalization by histidine.

NMR analysis of CS-His

The ^{1}H NMR spectrum of the reaction product was recorded in order to verify the functionalization of the CS molecules by His (Figure S2a). The measurements were done in the Analytical Center of the Chemistry Institute (University of Sao Paulo, Sao Paulo, Brazil) using Bruker AIII500 (500 MHz) spectrometer. The sample was diluted in the D_2O with a concentration of 12% (w/v). The obtained spectrum consists of lines attributed to both His groups and CS glycosidic chains. For example, the signals referring to the glucuronic acid protons of C2 and C3 carbons are located at 3.3 ppm and 3.5 ppm. The resonance at 3.9 ppm is attributed to the C2 proton of N-acetylgalactosamine. The signal at 3.7 ppm from the protons linked to C3 and C5 carbons of N-acetylgalactosamine is superimposed on the C5 proton of glucuronic acid. The protons attached to the C1 of the two monosaccharides overlap at 4.4 ppm [S1].

The chemical shifts of CS-His were predicted by the ChemNMR module built into the ChemDraw Professional software (v. 16.0.1.4). The predicted chemical shifts presented in the bottom part of Figure S2a correlate with the experimental spectrum though some signal mismatch occurs. This confirms the functionalization of CS by histidine.

Different protons in the His residue shown in the inset in Figure S2 with blue letters give rise to the chemical shifts are marked with blue lines. It is worth noting that because of the limited choice of the solvents in the ChemNMR module, the prediction was made for the DMSO. However, the solvent typically influences the chemical shift of mobile protons in amine or hydroxyl groups, whereas other protons are less sensitive to the solvent. This is well seen for NH group in the His residue (labeled *e* in the inset in Figure S2). The predicted chemical shift for this proton in DMSO is 8.32 ppm, whereas this signal is completely absent in the experimental spectrum due to the proton-deuterium substitution.

The ^{13}C NMR spectrum (not presented here) corroborates the results shown above for the ^{1}H NMR spectrum. The imidazole carbon atoms in His residue located at 116.7, 129.6, and 133.2

ppm correspond to δ, γ, ε carbons, respectively (see inset in Figure S2) [S2]. The carbon atoms C1 to C6 of the glucuronic acid are located at 100.8, 54.0, 80.4, 75.6, 74.7, and 61.1 ppm, respectively. The carbon atoms C1 to C5 of N-acetylgalactosamine are 103.7, 72.3, 73.6, and 76.4 ppm, respectively. The amide carbon (CH_3) is observed at 22.5 ppm, corroborating what was exposed [S3].

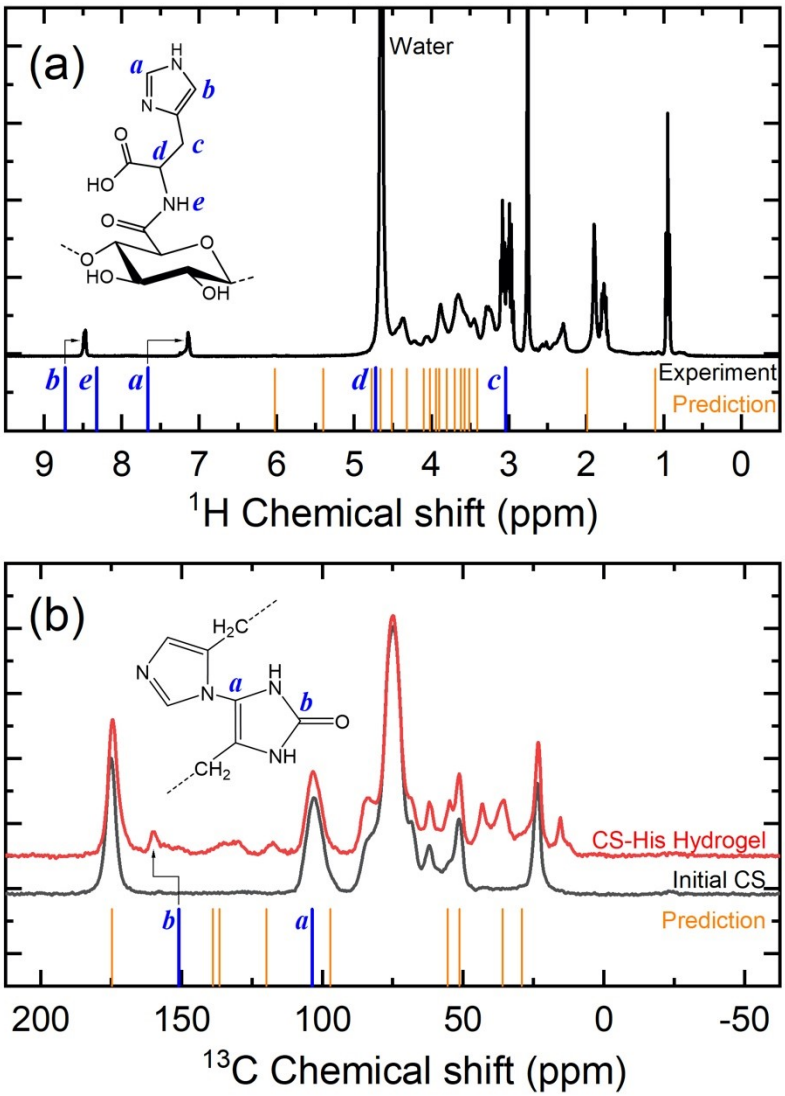

Figure S2. NMR spectra of different CS molecules. (a) Experimental ^{1}H NMR spectrum and predicted chemical shifts for CS-His molecule. The inset shows the fragment of the CS-His molecule with His residue. Blue lines correspond to the protons in His labeled with blue letters. (b) Experimental ^{13}C NMR spectra for initial CS molecules and CS-His hydrogel. Predicted chemical shifts were calculated for His-His cross-link. Blue lines correspond to carbon atoms essentially changed during the cross-link.

^{13}C solid-state NMR spectra of initial CS and lyophilized CS-His hydrogel were recorded to verify the formation of His-His cross-link between CS-His molecules (Figure S2b). The measurements were done in the Department of Chemistry of University of Aveiro (Aveiro, Portugal) using Bruker Avance III WB 400 MHz spectrometer (9.4 T) at a magic angle spinning (MAS) frequency 12 kHz. 4 mm double-resonance MAS probe (Bruker) was used. Chemical shifts were calibrated using glycine as a secondary chemical shift standard with C=O carbon at 176.03 ppm. The obtained spectrum for the CS-His hydrogel contains lines corresponding to initial CS molecules (black spectrum in Figure S2b) as well as additional lines corresponding to the His functionalization. The chemical shifts of the cross-linked His molecules were predicted by the ChemNMR module built into the ChemDraw Professional software (v. 16.0.1.4) to predict the signature of the cross-linking in CS-His hydrogel. Despite the fact the ChemNMR module predicts chemical shifts for molecules in solution, this does not influence the ^{13}C chemical shifts. Therefore, the obtained shifts' values correlate with the experimental solid-state spectrum (Figure S2b). According to the scheme presented in Figure 1a (see main text), the His-His cross-linking occurs by forming a new C-N bond between imidazole rings. The predicted chemical shift for the involved carbon atom (labeled *a* in the inset in Figure S2b), 103.6 ppm, overlaps with the lines of initial CS molecules, thus hampering the proof of the cross-linking. However, the crosslinking also involves the attachment of an oxygen atom to another carbon atom of the imidazole ring (labeled *b* in the inset in Figure S2b). The predicted chemical shift for this atom, 151.0 ppm, well corresponds to the experimental line located at 161.1 ppm (Figure S2b). The presence of this line confirms the His-His cross-linking in CS-His hydrogel.

ELP expression and characterization

ELP expression test in strains of E.coli Arctic-ELP and BL21DEIII-ELP

The ELP protein has an amino acid sequence based on the 60 repetitions of pentapeptide VPGHG (Scheme S1). The fragment of the encoding gene to ELP (951 bp) was cloned into the expression vector *pET28a* (5349 bp) using the restriction enzymes ECORI and XhoI, obtained commercially and called *pET28-elp* (6305 bp) (GENE SCRIPT). The *Escherichia coli DH5α* (cloning cell), *Arctic Cell,* and *BL21-DEIII (expression cells)* strains chemo-competent [S4] were transformed with 100 ng of the plasmid *pET28-elp* by thermal shock. Initially, the *pET28*-elp was amplificated in DH5α, (Figure S3), and *Arctic*-ELP and *BL21DEIII*-ELP strains were submitted to an initial expression test. Three isolated clones of both lineages were grown in Terrific Broth (TB) medium plus 50 μg / mL of the kanamycin antibiotic, with expression induction after addition isopropyl-beta-D-thiogalactopyranoside 0.5 mM (IPTG), for 4 h at 37 °C in an orbital shaker at 220 rpm. Aliquots before and after induction were removed for analysis of expression in SDS-PAGE (12.5%). The culture was centrifuged for 5 min at 10,000 rpm (Eppendorf). The cell mass was resuspended in TRIS buffer (Tris 0.1M; NaCl 0.5M; Glycerol 20%; 1 mM of the protease inhibitor phenylmethylsulfonyl fluoride - PMSF; pH 8.5), and cells were lysed in 2 cycles of 5 min "ON" and 2 min "OFF" under pressure (600-800 mBar) in a homogenizer (ARTEPEÇAS APLAB-10). The cell lysate was centrifuged at 10,000 rpm for 15 min, and the soluble and insoluble fractions were separated and analyzed on SDS-PAGE. Aliquots of before (T0) and after (T4) induction were evaluated by electrophoresis on gel 12.5% polyacrylamide under denaturing conditions (Figure S4a).

VPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGV

PGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVP

GHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGAVTGRGDSPASSVPGHGVP

GHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPG

HGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGHGVPGH

GVPGHGVPGHGVPGHGVPGHGVPGHG

Scheme S1. The amino acid sequence of ELP-His studied in this work.

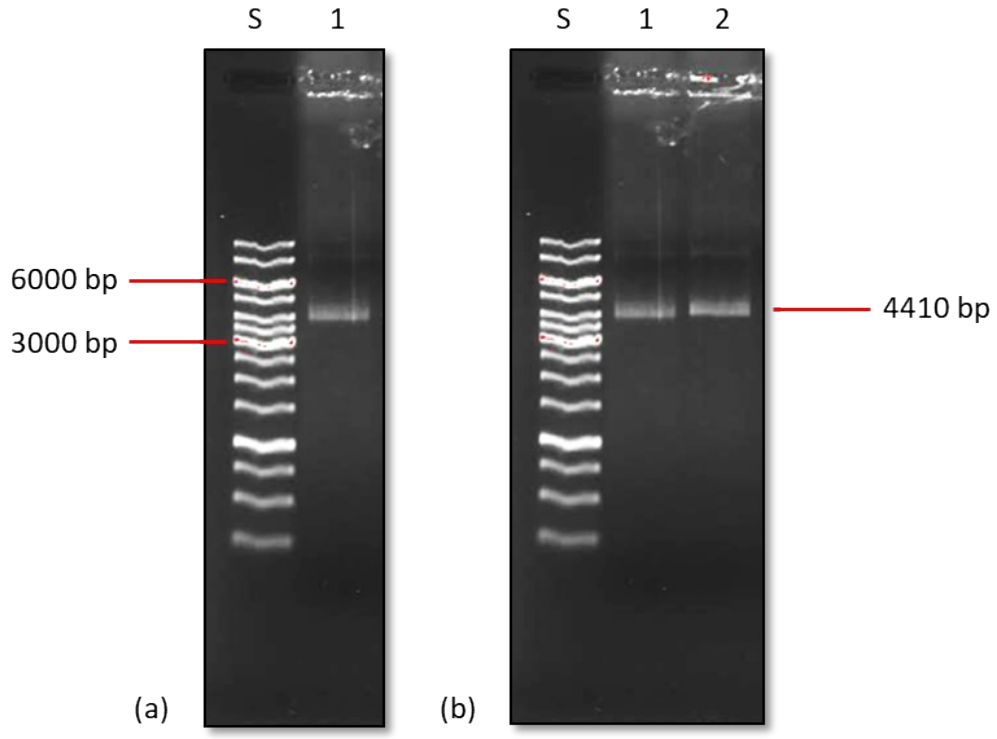

Figure S3. Confirmation of the *pET28-elp* vector amplification and the presence of the ELP gene. The *pET28-elp* expression vector obtained commercially was amplified in *E.coli DH5α* strain, and the resulting plasmid DNA was analyzed on 0.8% agarose gel. The vector was digested with enzymes ECORI and XhoI for its release and observation on the gel to confirm the presence of the fragment encoding the ELP gene. (a) Samples: Standard (Gene Ruler 1Kb # SM1163); 1: *pET28-elp*. (b) Samples: Standard (Gene Ruler 1Kb # SM1163); 1: *pET28-elp*; 2: fragment of ELP gene.

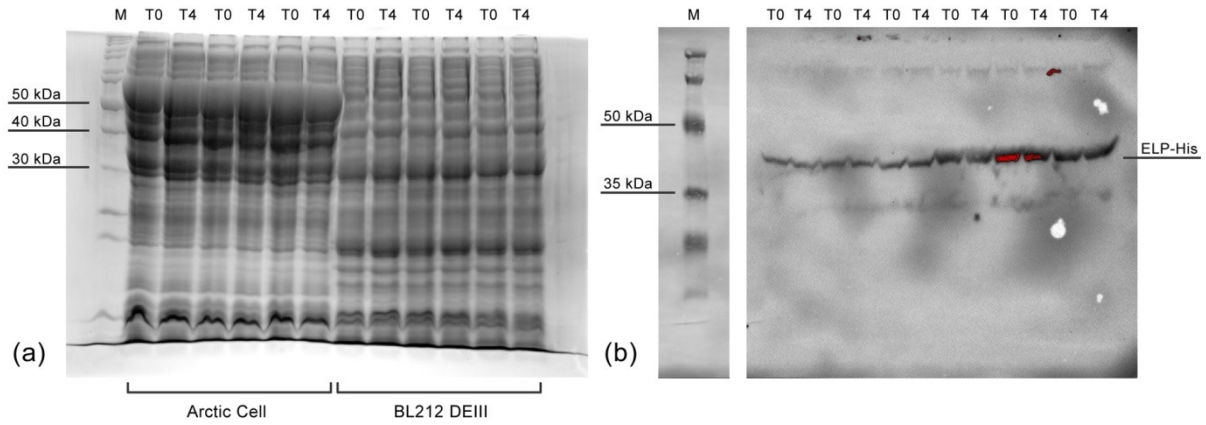

Figure S4. ELP-His was best expressed in the *E.coli BL21DEIII*-ELP lineage. The expression of the recombinant ELP-His protein was tested in the *E.coli Arctic*-ELP and *BL21DEIII*-ELP strains. Three clones of *E.coli* strains transformed with *pET28-elp* were selected for expression testing. After cultivation, aliquots before and after induction had their electrophoretic profile analyzed in 12.5% polyacrylamide gel under denaturing conditions. The expression validation was performed in an immunodetection assay with α-HisTag polyclonal antibodies. ELP-His had a molecular weight of approximately 40 kDa, different from its predicted molecular weight (26.9 kDa); however, this is expected due to its excess positive charge that interferes with its migration in the gel. (a) SDS-PAGE of the ELP-His protein expressed in the Arctic Cells and BL21 DEIII strains before (T0) and after (T4) the induction at 37 °C for 4 h (b) Immunodetection assay confirms the ELP-His expression with anti-HisTag antibody.

Immunodetection assay for ELP in induced E.coli (Western Blotting)

Immunodetection was performed to confirm the expression level of the clones selected in *Arctic*-ELP and *BL21DEIII*-ELP and the molecular weight of the expressed ELP. Aliquots of 10 µl before and after induction of the three isolated colonies each strain were submitted to electrophoresis on 12.5% polyacrylamide gel and transferred to a nitrocellulose membrane (GE Healthcare Life Sciences). The membrane was blocked using a 0.05% PBS-Tween solution (PBS-T) plus 5% skimmed milk for 18h at 4 °C. Primary labeling was performed at 37 °C (1.5h) with anti-HIS monoclonal antibody (SIGMA - MA121315) at a 1:10000 dilution. After washing

the membrane 3x with PBS-T, the secondary staining was performed with anti-mouse IgG antibody (SIGMA - A2304) conjugated to the peroxidase enzyme 1:3000 dilution in blocking buffer for 1 h. The resulting reactive bands were identified by exposing the membranes to the Luminol-Hydrogen Peroxide (H_2O_2) solution, according to the manufacturer's instructions (Figure S4b).

Expression of the recombinant in BL21DEIII-ELP and purification by ITC (Inverse Transition Cycle)

Clone 2 of the *BL21DEIII*-ELP strain was selected to continue the expression experiments. The pre-inoculum was cultivated in 50 ml of LB medium plus 50 µg/ml kanamycin and 20 µg/ml gentamicin using an orbital agitator at 37 °C and 220 rpm to 24h. The inoculums were cultivated in TB medium plus 50 µg/ml kanamycin and 1:100 of pre-inoculum on orbital agitator at 37 °C and 200 rpm to reach the absorbance values in 600 nm of 2. The expression of recombinant protein was induced by using 0.5 mM IPTG at 37 °C to 24h. The protein was expressed in the soluble fraction. The purification of the expressed ELP from the soluble fraction of bacterial lysate to remove contaminants was done using the inverse transition cycle (Figure S5) [S5]. The ELP-rich lysate is heated above the lower critical solution temperature (LCST), forcing ELP to aggregate into micron-scale co-acervates. Then, the ELP is separated by centrifugation (hot centrifugation stage, 37°C). The ELP rich sediment is then resuspended in a cold buffer, where it becomes soluble and centrifuged again at 4°C (cold centrifugation stage). The sediment now containing insoluble contaminants is discarded, while the supernatant containing soluble ELP is retained. The alternating cycles of hot and cold rotations are repeated until the desired purity is achieved, as determined with SDS-PAGE, molecular weight predicted of 26.9 kDa.

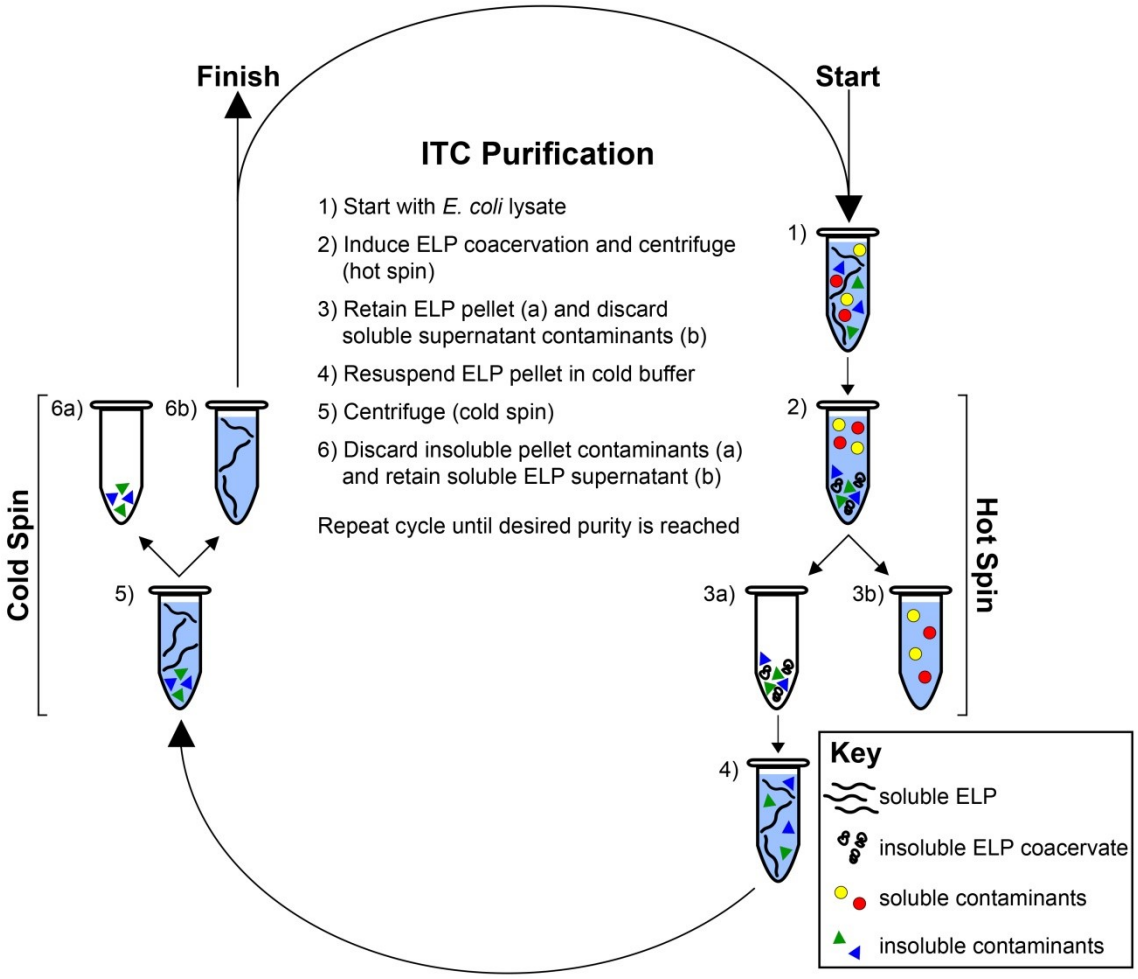

Figure S5. Reverse transition cycle for ELP purification. [S5]

Figure S6. Organization of the thiol groups in the helical structure of ELP-Cys.

Yellow spheres represent sulfur atoms.

Determination of the ELP-His protein structure from SAXS measurements

The suspension of the ELP-His protein in pD carbonate buffer (pH = 9.4) with a concentration 5 mg/mL was studied by small-angle X-ray scattering (SAXS) using Xenocs XEUSS™ setup equipped with a GENIX™ source of the X-ray radiation (Cu Kα edge, λ = 1.54 Å). The beam was focused at the sample into a 1 × 1 mm spot by FOX2D™ optics. A Dectris Pilatus™ 300K detector collected the 2D scattering data. The sample to detector distance was 0.86 m, giving a q range from 0.012 to 0.424 Å^{-1}, where q is the reciprocal space momentum transfer modulus, defined as $q = (4\pi \sin\theta)/\lambda$, where 2θ is the scattering angle, and λ is the radiation wavelength. The acquisition time was 1800 s, and the temperature was 25 °C. The sample was placed in a reusable quartz capillary of 1.5 mm in diameter, mounted on stainless steel cases, allowing the sample and buffer measurements in precisely the same conditions.

The obtained scattering curve $I(q)$ is presented in Figure 4a (see main text). The scattering intensity rapidly decreases with q increasing and remains almost constant at q > 0.07 Å^{-1}. Therefore the further analysis of the curve was performed in the range from 0.012 to 0.07 Å^{-1}. This range determines the resolved particle sizes' limits ($a \approx \pi/q$) from 45 to 260 Å.

First of all, the shape of the protein particles in the solution was determined. Computer modeling predicted the shape of the individual molecule close to the ellipsoid with the axes a, b, and c around 45, 70, and 150 Å, respectively. However, the aggregation of molecules usually occurs in the solution. Tthe middle part of the SAXS curve was analyzed (Figure 4c, main text) to determine the shape of the aggregates. It is known that the scattering intensity in this region is proportional to q^{-D}, where D is a fractal dimension of the scattering particles [S6]. In the case of the studied protein D = 2.6±0.1 thus demonstrating the formation of branched 2D aggregates, which can be roughly considered in the further analysis as **globular** particles due to their developed surface.

Analysis of the Guinier region ($0.012 < q < 0.022$ Å^{-1}) allowed determining the radius of gyration, R_g, of the aggregates (Figure 4d, main text), which was found to be 278 Å, and the forward intensity, $I(0)$ = 0.43, which was used to estimate the molecular weight, M, of the aggregates [S7]:

$$M = \frac{I(0)N_A}{c(\Delta\rho_M)^2},$$

where c is the concentration of the protein (5 mg/mL), $\Delta\rho_M$ is the excess scattering length density per unit mass (for proteins $\Delta\rho_M = 2\times10^{10}$ cm/g [S7]), N_A is the Avogadro's number, $I(0)$ is the forward intensity obtained from Guinier plot (Figure 4d, main text). The value of the molecular mass obtained by this method is around 128 kDa. This is 4.7 times bigger than the estimated molecular mass of the protein molecule is 27 kDa. Therefore, it is reasonable to suppose that the aggregates consist, on average, of 5 individual protein molecules.

To better understand the internal structure of the aggregates, an indirect Fourier transformation (IFT) was performed using the GNOM program [S8] built-in the ATSAS package [S9], and a pair distance distribution function, $P(r)$, was evaluated. Figure 4a (see main text) shows quite well-fitting of the experimental curve (the estimated total quality is 0.9052), which allowed getting a reasonable $P(r)$ function (see inset in Figure 4a in the main text) and determined the maximum intermolecular distance D_{max} = 736 Å. The shape of the $P(r)$ function is close to that usually obtained for globular particles, however, consists of two components with peak distances at D_1 = 227 Å and D_2 = 400 Å. These distances can be represented as a sum of axes a, b, and c of an effective ellipsoid expressing the protein molecule: $D_1 \approx 2\cdot a+c$ = 241 Å and $D_2 \approx 2\cdot c+b$ = 370 Å. Such decomposition is in line with estimating several individual protein molecules made based on the molecular weight of the aggregate.

Thus, SAXS measurements demonstrated that ELP-His protein forms in the solution aggregates with the shape and size, correlating with the computer simulation results. The radius of gyration

of the aggregates is 278 Å, with a molar mass around 128 kDa, slightly elongated shape with characteristic sizes 227×400 Å, and thus consist of 5 ELP-His molecules on average.

Supporting references:

S1 Pomin, V. H. Holothurian Fucosylated Chondroitin Sulfate. *Mar Drugs.* **2014**. *12*, 232–254.

S2 Frei, M. H.; Opella, S. J. The Effect of pH on Solid-State ^{13}C NMR Spectra of Histidine. *J. Magn. Reson.* **1986**, *66*, 144- 147.

S3 Hamer, G. K.; Perlin, A. S. A 13c-n.m.r. spectral study of chondroitin sulfates A, B, and C: evidence of heterogeneity. *Carbohydr. Res.* **1976**, *49*, 37-48.

S4 Molecular Cloning: A Laboratory Manual, 3rd ed.; Sambrook J. F., Russell, D. W., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

S5 MacEwan, S. R.; Hassouneh, W.; Chilkoti, A. Non-chromatographic Purification of Recombinant Elastin-like Polypeptides and their Fusions with Peptides and Proteins from *Escherichia coli*. *J. Vis. Exp.* **2014**, (88), e51583.

S6 Anitas, E. M. Small-Angle Scattering (Neutrons, X-Rays, Light) from Complex Systems. Fractal and Multifractal Models for Interpretation of Experimental Data, SpringerBriefs in Physics, Springer International Publishing, 2019.

S7 Oliveira, C. L. P. Investigating Macromolecular Complexes in Solution by Small Angle X-Ray Scattering, in *Current Trends in X-Ray Crystallography*; Chandrasekaran A., Ed.; IntechOpen: London, 2011; p 2442

S8 Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495-503.

S9 Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Cryst.* **2017**, *50*, 1212-1225.