

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/260377318>

N-terminal Protein Tails act as Aggregation Protective Entropic Bristles: The SUMO case.

ARTICLE *in* BIOMACROMOLECULES · FEBRUARY 2014

Impact Factor: 5.75 · DOI: 10.1021/bm401776z · Source: PubMed

CITATIONS

5

READS

41

4 AUTHORS, INCLUDING:



[Patrizia Marinelli](#)

Autonomous University of Barcelona

3 PUBLICATIONS 11 CITATIONS

[SEE PROFILE](#)



[David Reverter](#)

Autonomous University of Barcelona

41 PUBLICATIONS 1,480 CITATIONS

[SEE PROFILE](#)



[Salvador Ventura](#)

Autonomous University of Barcelona

163 PUBLICATIONS 3,615 CITATIONS

[SEE PROFILE](#)

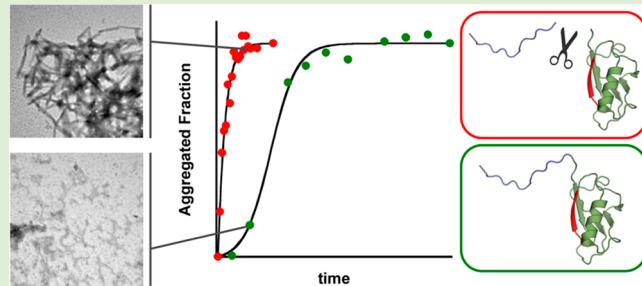
N-Terminal Protein Tails Act as Aggregation Protective Entropic Bristles: The SUMO Case

Ricardo Graña-Montes,[†] Patrizia Marinelli,[†] David Reverter, and Salvador Ventura*

Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

 Supporting Information

ABSTRACT: The formation of β -sheet enriched amyloid fibrils constitutes the hallmark of many diseases but is also an intrinsic property of polypeptide chains in general, because the formation of compact globular proteins comes at the expense of an inherent sequential aggregation propensity. In this context, identification of strategies that enable proteins to remain functional and soluble in the cell has become a central issue in chemical biology. We show here, using human SUMO proteins as a model system, that the recurrent presence of disordered tails flanking globular domains might constitute yet another of these protective strategies. These short, disordered, and highly soluble protein segments would act as intramolecular entropic bristles, reducing the overall protein intrinsic aggregation propensity and favoring thus the attainment and maintenance of functional conformations.



INTRODUCTION

Protein misfolding and aggregation into β -sheet enriched amyloid-like structures are associated with a large set of human disorders, including Alzheimer's disease, diabetes, and some types of cancer.^{1–4} However, the adoption of cytotoxic amyloid-like conformations is not restricted to disease-linked proteins and seems to constitute a generic property of polypeptide chains,^{5,6} likely because the noncovalent contacts that stabilize native structures resemble those leading to the formation of amyloids.⁷ Indeed, a majority of proteins contain at least one and often several aggregation-promoting sequences,^{8,9} in many cases buried in the hydrophobic core of the native structure. Therefore, productive protein folding and deleterious aggregation are continuously competing in the cell. Because the formation of compact globular proteins comes at the expense of an inherent sequential aggregation propensity, identification of the strategies that enable proteins to remain functional and soluble in the cell is a central issue in biology.¹⁰

Organisms have evolved different mechanisms to survey and minimize side aggregation reactions,^{11–15} including sophisticated and highly conserved protein quality control machineries.^{16,17} In addition, during the course of evolution, proteins have adopted negative design strategies to prevent or diminish their intrinsic propensity to aggregate, by incorporating β -sheet breakers at structurally critical positions,^{18,19} avoiding the presence of β -strands on the edge of protein structures²⁰ or placing gatekeeper residues at the flanks of aggregation-prone segments.⁸ It has been recently suggested that the recurrent presence of disordered segments adjacent to folded domains in proteomes might be yet another strategy evolved to overcome aggregation.²¹ Random movements of these tails around the

point of attachment to the folded domain would sweep out a large area in space, acting thus as entropic bristles (EB).^{22,23} Recently, the ability of long and highly disordered tails, either natural or artificial, to act as EB was tested experimentally by fusing them to different target proteins and expressing the fusions recombinantly in bacteria. Proteins fused to these EB were significantly more soluble than their natural counterparts.²² Several evidences indicate that disordered terminal tails might also play an antiaggregational effect in the context of natural proteins. In this way, it has been shown that the disordered C-terminal region of NEIL1, a human homologue of *Escherichia coli* DNA glycosylase endonuclease VIII, is necessary for its soluble recombinant expression.²⁴ A similar role has been proposed for the highly charged and disordered C-terminal tail of α -synuclein in the context of α -synuclein-GST fusions.²⁵

SUMO belongs to the ubiquitin-like (Ubl) protein family. The members of the Ubl family are small size (<10 kDa) post-translational modifiers, which are attached to protein substrates via an isopeptide bond between a C-terminal glycine and an acceptor lysine residue of the substrate.²⁶ Despite the low degree of sequence homology displayed between the members of the Ubl family, they all share a common protein fold and a similar mechanism of conjugation.^{27,28} However, in contrast to ubiquitin, SUMO proteins contain an N-terminal extension of 15–20 amino acid residues, which constitutes a flexible tail that protrudes from the core protein.²⁹ Using N-terminal deletion mutants that include only the globular functional domain of

Received: December 4, 2013

Revised: February 6, 2014

Published: February 24, 2014



SUMO, we have previously shown that when the conformational stability of the SUMO1, SUMO2, and SUMO3 human isoforms is compromised, these small globular proteins aggregate into amyloid-like structures.³⁰ SUMO aggregation might have important physiological implications because the SUMO pathway is essential in mammals and in budding yeast. Therefore, it is likely that these proteins might have evolved strategies to minimize their aggregation propensity. By comparing the conformational, thermodynamic and aggregational properties of SUMO variants with and without the N-terminal extension, we show here that this tail acts as an EB, slowing down the aggregation kinetics of the globular domain. Moreover, by fusing the SUMO N-terminal tail to the highly amyloidogenic A β 42 peptide and monitoring its impact on intracellular aggregation in bacteria,³¹ we demonstrate that the solubilizing effect of this N-terminal extension can act in trans. We provide here experimental evidence for the antiaggregational effect of EB in a natural protein context. Computational analysis indicates that in fact this might be a generic function of disordered N- and C-terminal extensions in globular proteins.

MATERIALS AND METHODS

SUMO Domains Expression and Purification. Human SUMO2 full-length and a 14-residue Nter deletion variant (residues 15–95), referred here as SUMO2 and Δ Nt-SUMO2, and SUMO1 full-length and a 17-residue Nter deletion variant (residues 18–101), denoted SUMO1 and Δ Nt-SUMO1, respectively, were cloned into a pET-28b vector to encode either a Cter SENP2-cleavable hexahistidine fusion protein for full-length domains, or a Nter thrombin-cleavable hexahistidine fusion for Nter deletion variants. Cultures of *E. coli* BL21(DE3) cells transformed with these plasmids were grown in lysogeny broth (LB) medium with 50 μ g·mL⁻¹ kanamycin at 37 °C and 250 rpm to an OD at 600 nm of 0.5–0.6 before induction with 1 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) for 4 h at 30 °C. Next, the cultures were centrifuged and the cell pellets were frozen at –20 °C. After cell lysis, SUMO proteins were purified under native conditions by affinity chromatography on a FF-Histrap histidine-tag resin (General Electric). To cleave the histidine-tag, SUMO domains were incubated for 16 h at 4 °C with either SENP2 protease or thrombin, accordingly. The cleaved tags were removed by gel filtration on a HiLoad Superdex 75 prep grade column (General Electric). Protein buffer was further exchanged with the appropriate assay buffer on a Sephadex G-25 (General Electric) column prior to storage of SUMO samples at –80 °C. Unless otherwise stated, assays on SUMO domains were performed on 50 mM phosphate buffer at pH 7.

Intrinsic Fluorescence. SUMO2 and Δ Nt-SUMO2 intrinsic fluorescence was monitored by recording Tyr emission spectra between 280 and 400 nm upon excitation at 268 nm. Spectra were registered, after equilibration at 298 K of a 50 μ M protein sample, as the accumulation of three consecutive scans in a Jasco FP-8200 spectrophotometer (Jasco).

Intrinsic Fluorescence Quenching Assays. Quenching of SUMO2 and Δ Nt-SUMO2 intrinsic fluorescence was analyzed by monitoring Tyr emission in the presence of acrylamide. Tyr fluorescent emission was recorded between 280 and 400 nm upon excitation at 268 nm, and after equilibration of 50 μ M protein samples with final quencher concentrations ranging from 0 to 0.21 M at 298 K. Spectra were registered as the accumulation of three consecutive scans in a Jasco FP-8200 spectrophotometer (Jasco).

Bis-ANS Binding. Binding of 4,4'-bis(1-anilinonaphthalene 8-sulfonate) dye to soluble and aggregated forms of SUMO2 and Δ Nt-SUMO2 was followed by recording the fluorescence spectra of protein–dye mixtures between 400 and 600 nm after excitation at 370 nm. Spectra were registered, after equilibration of the sample at 298 K, as the accumulation of three consecutive scans in a Jasco FP-8200 spectrophotometer (Jasco). Final protein and bis-ANS concentrations

in the mixtures were 50 and 25 μ M, respectively. Aggregated SUMO domains correspond to samples incubated at 343 K for 400 min.

Circular Dichroism (CD). SUMO2 and Δ Nt-SUMO2 far-UV CD spectra were recorded for their soluble and aggregated forms between 195 and 270 nm at 298 K with a spectral resolution of 1 nm, using a Jasco 810 spectropolarimeter (Jasco). For spectra acquisition, protein samples were placed in a 0.1 cm path-length quartz cell at a final concentration of 20 μ M and 20 scans were averaged for each spectrum.

Near-UV CD spectra were obtained for SUMO2 and Δ Nt-SUMO2 soluble forms between 250 and 320 nm at 298 K with a spectral resolution of 1 nm. Spectra were obtained by accumulating 50 scans of 100 μ M protein samples in a 0.2 cm cell, employing a Jasco 810 spectropolarimeter (Jasco).

Thermal Denaturation. SUMO2 and Δ Nt-SUMO2 thermal denaturation was monitored by following the change of its CD signal at 222 nm and of its intrinsic fluorescence at 305 nm upon excitation at 268 nm. Signal change was recorded between 288 and 368 K using a 1 K·min⁻¹ gradient at a final protein concentration of 20 μ M. Experimental data were fitted to a two-state transition model, whose signals for the folded and unfolded states are dependent on the temperature, using the nonlinear least-squares algorithm provided with KaleidaGraph (Synergy Software).

Gel Filtration Chromatography. Gel filtration chromatography was employed routinely to remove purification tag peptides cleaved from SUMO domains, but it also allowed to infer differences between the effective hydrodynamic volumes of SUMO2 and Δ Nt-SUMO2. For this analysis, ≈0.5 μ moles of previously purified SUMO2 or Δ Nt-SUMO2 were injected in a HiLoad Superdex 75 prep grade column (General Electric) and eluted with 20 mM Tris, 250 mM NaCl, and 1 mM β -mercaptoethanol at pH 8.

Thioflavin-T (Th-T) Binding. Binding of the Th-T amyloid dye to SUMO2 and Δ Nt-SUMO2 samples aggregated by incubating them for 400 min at 343 K was evaluated by recording the mixtures fluorescence spectra. Aggregated samples were diluted to 50 μ M in Th-T, resulting in a final dye concentration of 25 μ M. Spectra were recorded between 460 and 600 nm, after sample equilibration at 298 K and upon excitation at 440 nm, as the accumulation of three consecutive scans in a Jasco FP-8200 spectrophotometer (Jasco).

Transmission Electron Microscopy (TEM). Protein morphology was evaluated under TEM for SUMO2 and Δ Nt-SUMO2 samples incubated for 400 min at 343 K. These samples were diluted 10-fold in distilled water and negatively stained by placing 10 μ L of the dilutions on carbon-coated copper grids and allowing its deposition for 5 min, after sample removal the grids were stained with 2% uranyl acetate which was further removed after 1 min deposition. The grids were imaged in a Hitachi H-7000 transmission electron microscope operating at a 75 kV accelerating voltage.

Aggregation Kinetics. To follow its aggregation kinetics, soluble SUMO domains were prepared at a final protein concentration of 200 μ M. In order to avoid the presence of preformed oligomeric species, protein samples were previously filtered through a 0.22 μ m polyvinylidene fluoride filter. SUMO aggregation was triggered by incubating the proteins under 500 rpm agitation at temperatures close to their T_m : 343 K for SUMO2 and Δ Nt-SUMO2, or 333 K for SUMO1 and Δ Nt-SUMO1. Aggregation kinetics was monitored by evaluating the binding to the amyloid dye Thioflavin-T (Th-T) of samples from each aggregation reaction taken at different time intervals. Samples corresponding to each time point were diluted in Th-T and the fluorescence signal at 475 nm of the mixtures was recorded at 298K upon excitation at 440 nm in a Jasco FP-8200 spectrophotometer (Jasco). Final protein and dye concentrations in the mixtures were 100 and 25 μ M, respectively.

The aggregated fraction of the protein (f) along the kinetics was calculated by normalizing Th-T fluorescence relative to the initial and end point intensities. Kinetic data from the full-length domains could be accurately fitted to an autocatalytic reaction model³² according to the following equation:

Table 1. Physical, Intrinsic Disorder, and Solubility Properties Predicted for the N-ter Fragments of SUMO Family Members

protein	UniProt accession No.	net charge	pI	length (aa)	VLXT ^a (%)	VSL2 ^a (%)	mean hydrophathy ^b	WH solubility ^c (%)		
								full protein	ΔNter protein	Nter tail
Hs SUMO1	P63165	-3.1	4.42	18	67	100	-1.8	73.1	58.1	97.0
Hs SUMO2	P61956	-1.1	4.89	14	79	100	-1.7	64.6	61.4	80.3
Ce SUMO	P55853	-3.1	3.44	11	82	82	-0.8	86.8	73.1	97.0
At SUMO2	Q9FLP6	-2.1	4.42	14	93	100	-1.9	57.6	49.8	93.1
Sc Smt3	Q12306	-3.1	4.42	21	100	100	-1.4	81.1	72.5	97.7

^aPercentage of amino acids found disordered using the PONDR predictors VLXT and VSL2. ^bA measure that distinguishes ordered from disordered proteins. ^cSolubility prediction according to the Wilkinson–Harrison method for the full-length protein, the protein without the Nter tail and the Nter fragment alone. The figures indicate the percentage of soluble protein.

$$f = \frac{\rho(e^{[(1+\rho)kt]} - 1)}{1 + \rho e^{[(1+\rho)kt]}} \quad (1)$$

where k is the product of the elongation constant of the aggregation reaction (k_e) times the protein concentration and ρ represents the dimensionless ratio of the nucleation constant (k_n) to k . k and ρ were derived by regression of f against time (t , in minutes) using the nonlinear least-squares algorithm provided with KaleidaGraph (Synergy Software). For aggregation kinetics with sigmoidal behavior the lag time was obtained by extrapolating the growth phase of the aggregation curve to $f = 0$, the half aggregation time and the time to aggregation completeness were derived using eq 1 and considering $f = 0.50$ or 0.99, respectively.

Analysis of Protein Solubility and Disorder in the SUMO Family. Theoretical analysis of the solubility and intrinsic disorder properties in the SUMO family was performed employing SUMO sequences from different species (Table 1). Protein solubility was estimated using the revised Wilkinson–Harrison solubility predictor,³³ and protein disorder was predicted with the PONDR-VLXT³⁴ and PONDR-VSL2³⁵ algorithms. All predictions were performed using the methods default settings.

Cloning and Expression of SUMO2-A β 42-GFP Fusion. The insert encoding for human SUMO2 full-length was subcloned into a pET-28a vector already containing a fusion of Amyloid β 42 with the Green Fluorescent Protein (A β 42-GFP), and was introduced upstream to the A β 42 coding sequence. Next, a PCR was performed to amplify the whole plasmid but the region encoding for residues 15–95 of SUMO2 and the Cter hexahistidine tail, so the resulting product encodes for a ternary fusion of SUMO2 Nter tail (residues 1–14) with A β 42-GFP (S2Nt-A β 42-GFP). *E. coli* BL21(DE3) competent cells were transformed with the plasmid encoding for S2Nt-A β 42-GFP or A β 42-GFP without the SUMO2 Nter fragment. Bacterial cultures were grown at 37 °C and 250 rpm in LB medium containing 50 μ g·mL⁻¹ kanamycin and 34 μ g·mL⁻¹ chloramphenicol. When cultures reached an OD at 600 nm of 0.5, protein expression was induced with IPTG for 4 h. Cultures were further incubated for 16 h at 4 °C, then cells were pelleted and washed with phosphate-buffered saline (PBS) for three times.

GFP Fusions Fluorescence Determinations. GFP fluorescence in intact cells expressing the A β 42-GFP or S2Nt-A β 42-GFP fusions was measured at a cell density with an OD at 600 nm of 0.1, using a Jasco FP-8200 spectrofluorimeter (Jasco). GFP emission spectra were recorded between 500 and 600 nm at 298 K, after excitation at 470 nm.

Intact cells expressing GFP fusions were also imaged by phase-contrast microscopy and fluorescence microscopy under UV light using a Leica Q500 MC fluorescence DMBR microscope (Leica Microsystems) employing a 718 ms exposure. For microscopic analysis, 5 μ L of washed cells were deposited on top of glass slides.

Aggregation Properties of Disordered Patterns at the Termini of Globular Proteins. Aggregation properties were analyzed for a set of protein fragments derived from the library of disordered patterns built by Galzitskaya and co-workers.³⁶ The sequences belonging to the library were filtered to exclude sequences with three or more consecutive His residues in order to avoid the

presence of fragments which likely belong to affinity purification tags, thus having an artificial nature. The aggregation properties were analyzed for the resulting set, consisting of 71 sequences, using the AGGRESCAN³⁷ and Waltz³⁸ algorithms employing default settings. The AGGRESCAN aggregation parameters computed for the disordered patterns were compared with those calculated for a reference set consisting of 71 sequences from the subset of sequences with less than 40% identity of the ASTRAL Compendium (Astral40), retrieved using a randomizing function.

RESULTS AND DISCUSSION

Conformational properties of SUMO2 and ΔNt-SUMO2. Full-length human SUMO2 is a 95 residues protein (Figure 1) in which the first 14 N-terminal residues correspond

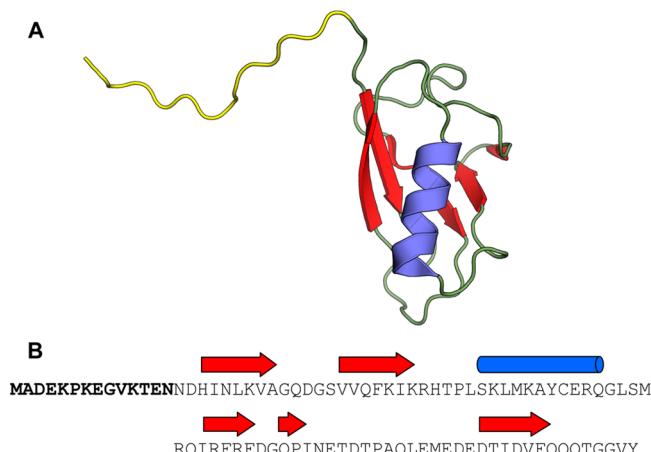


Figure 1. Structure and sequence of SUMO2. (A) Cartoon representation of the full-length SUMO2 solution structure (PDB 2AWT). The N-terminal tail is shown in yellow. (B) Sequence and regular secondary structure elements of SUMO2. The N-terminal tail sequence is highlighted in bold.

to a disordered tail according to its NMR 3D solution structure (PDB: 2AWT). This N-terminal extension is not involved in SUMO function, at least in vitro, and after its deletion, the remaining protein remains competent for the conjugation machinery.^{39,40}

We expressed and purified SUMO2 and a variant in which the first 14 N-terminal residues were deleted, thus, containing only the SUMO globular domain (ΔNt-SUMO2). We compared the secondary structure content of both protein variants using circular dichroism (CD) in the far-UV region (Figure 2A). Both spectra are essentially identical, deconvolution of the spectra using the K2D3 software⁴¹ suggests that the β -sheet signal (34–45%) is the major contributor to the

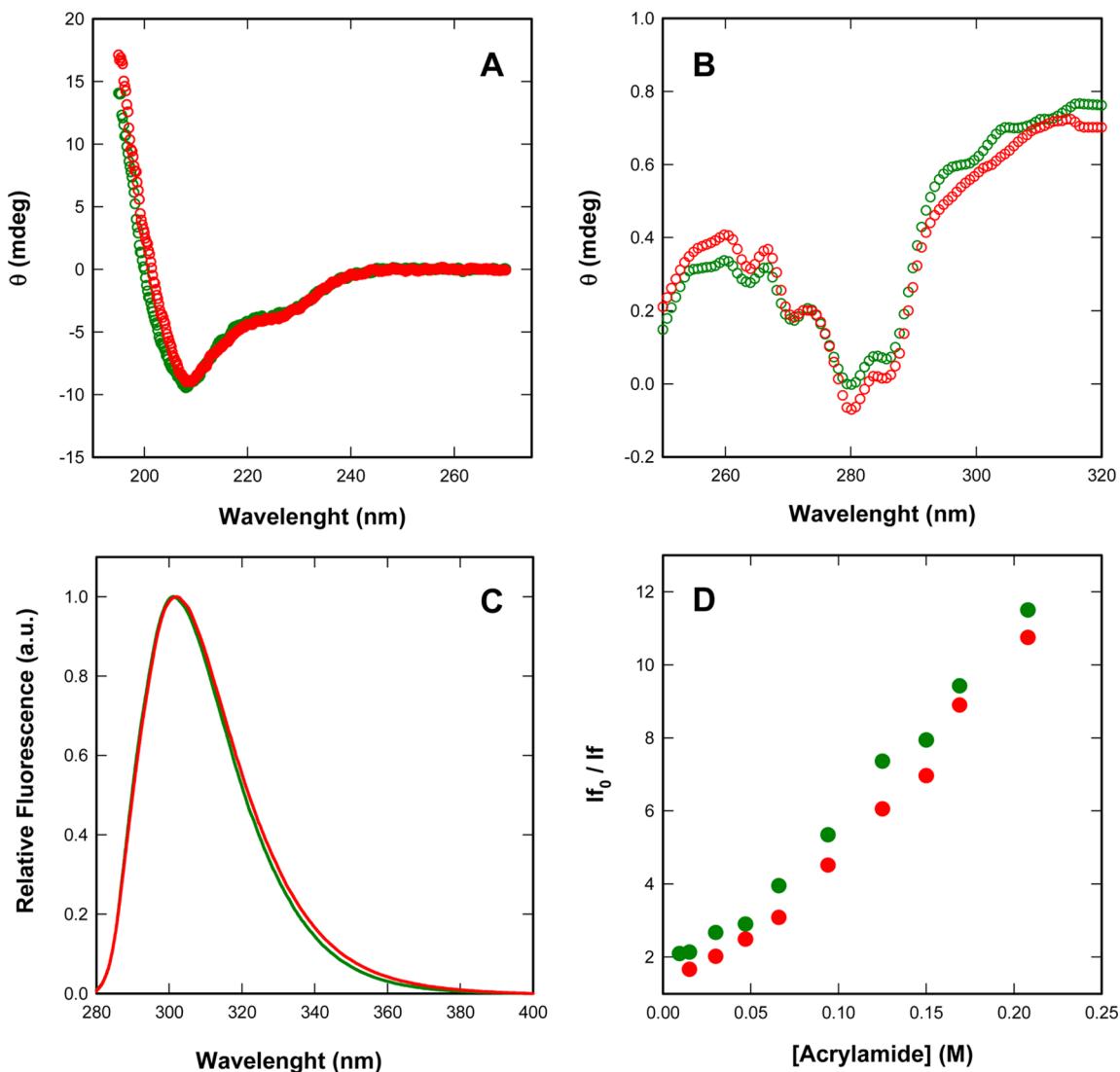


Figure 2. Conformational Properties of SUMO2 and ΔNt -SUMO2. (A) CD spectra in the far-UV region. (B) CD spectra in the near-UV region. (C) Intrinsic tyrosine fluorescence. (D) Stern–Volmer plot for the acrylamide quenching of tyrosine fluorescence. Data for SUMO2 and ΔNt -SUMO2 are shown in green and red, respectively.

spectra. We analyzed the CD spectra in the near-UV region to monitor differences in tertiary structure between the two variants. Also, in this region, the shapes of SUMO and ΔNt -SUMO2 spectra were almost identical (Figure 2B). Since SUMO2 lacks Trp residues we recorded the intrinsic fluorescence of Tyr residues in the two SUMO2 variants to monitor if they exhibit spectral differences. The proteins were excited at 268 nm and fluorescence recorded between 280 and 400 nm. The fluorescence spectra overlap, sharing the characteristic Tyr emission maximum at 305 nm (Figure 2C). Tyr fluorescence quenching by acrylamide was used to get more specific information on the location of these residues in both SUMO variants. As expected, in both cases, emission at 305 decreased with increasing acrylamide concentration. Stern–Volmer plots indicated that Tyr residues in the two proteins were in similar environments (Figure 2D). Overall, these data are in agreement with previous structural studies,³⁹ indicating that deletion of the N-terminal tail does not affect significantly the secondary and tertiary structure of the globular domain.

Thermal Stability of SUMO2 and ΔNt -SUMO2. The thermal stabilities of the two SUMO2 variants were analyzed by monitoring Tyr intrinsic fluorescence and far-UV CD changes at 305 and 222 nm, respectively, in the 293–368 K range. The transition curves from heat induced fluorescence emission changes in the SUMO2 variants are shown in Figure 3A. The thermal denaturation curves, followed by far-UV CD, are shown in Figure 3B. In both cases, a single cooperative transition was observed, and the data could be fitted to a two-state temperature-induced unfolding model ($R \geq 0.99$). The two probes reported essentially identical thermal transitions indicating that the secondary and tertiary structures are lost simultaneously upon heating, thus, supporting a two-state thermal unfolding mechanism for the two proteins.

The calculated melting temperatures (T_m) were 346.24 ± 0.35 and 345.06 ± 1.19 K, for SUMO2 and 345.65 ± 0.47 K and 344.08 ± 0.87 K for ΔNt -SUMO2, by intrinsic fluorescence and far-UV CD, respectively. These data indicate that deletion of the N-terminal tail does not affect the thermodynamic stability of SUMO2, which can be thus

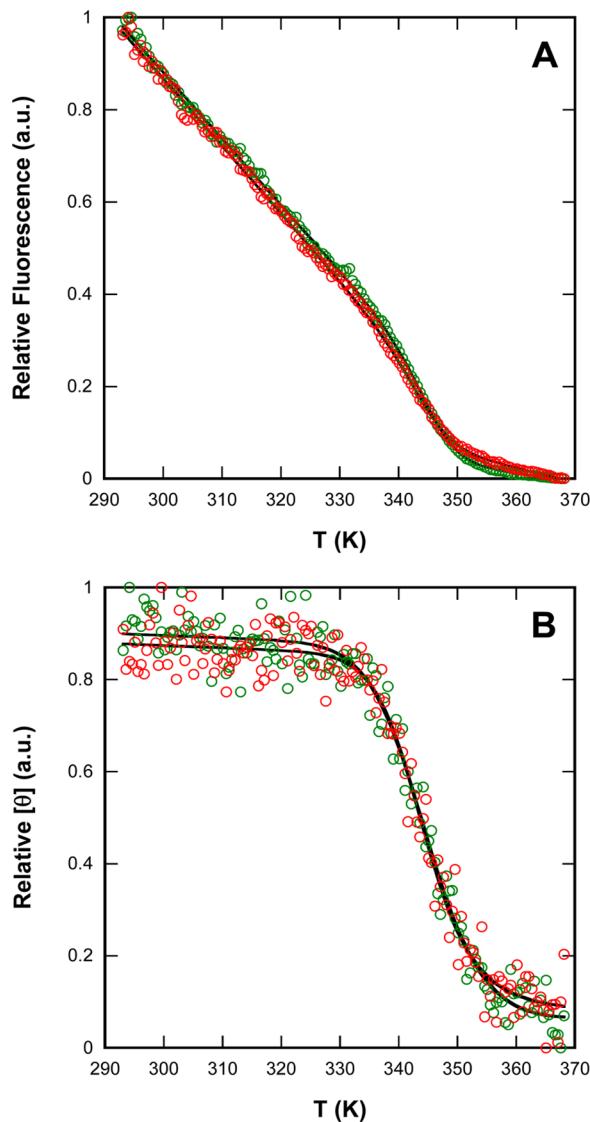


Figure 3. Thermal Denaturation of SUMO2 and Δ Nt-SUMO2. Thermal denaturation was monitored by following changes in (A) intrinsic fluorescence and (B) CD signal at 222 nm of SUMO2 (green) and Δ Nt-SUMO2 (red).

unequivocally attributed to the interactions sustaining the globular domain of the protein.

ΔNt-SUMO2 Exposes Hydrophobic Residues to Solvent. The N-terminal tail of SUMO2 might, in principle, affect the hydrodynamic properties of the full-length protein, when compared with that of Δ Nt-SUMO2. We used gel-filtration chromatography to investigate if this is the case. As expected, SUMO2 elutes first than Δ Nt-SUMO2 (Figure 4). However, the calculated difference in apparent molecular weight between the two proteins is \sim 10 kDa, which is much higher than the one expected for a tail of only 14 residues (\sim 1.5 kDa). This suggests that the large difference in hydrodynamic volume between the two proteins can be attributed to the disordered nature of SUMO2 N-terminal extension.

Despite all previous data indicated that the globular domains in the two proteins were conformationally identical, the presence of a fluctuating tail might affect the domain surface properties. We monitored the presence of exposed hydrophobic clusters in the native structure of SUMO2 and Δ Nt-SUMO2 by

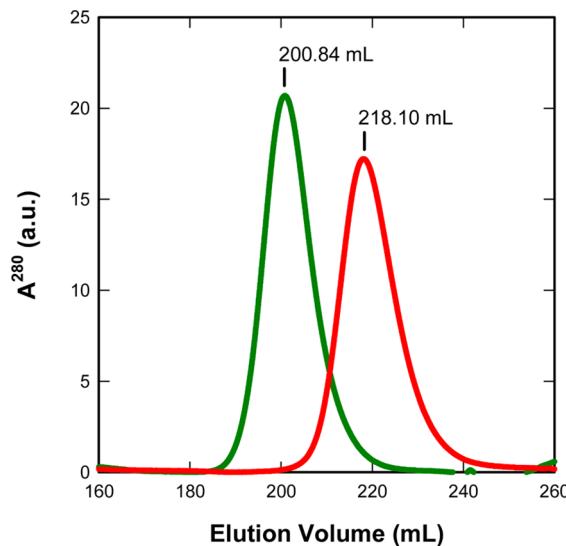


Figure 4. Hydrodynamic properties of SUMO2 and Δ Nt-SUMO2. Gel filtration of SUMO2 (green) and Δ Nt-SUMO2 (red) onto a Superdex 75 column. The elution volumes of the two proteins are indicated.

measuring their binding to 4,4'-bis(1-anilinonaphthalene 8-sulfonate) bis-ANS, a dye that increases its fluorescence emission upon interaction with these nonpolar regions.⁴² We found that they display differential binding to this dye, Δ Nt-SUMO2 exhibiting <2 times higher fluorescence emission than SUMO2 (Figure 5A). This implies that certain residues totally or partially protected from the solvent in SUMO2 increase their exposure after deletion of the N-terminal tail. According to the conformational and stability data this cannot be attributed to global or partial unfolding of the globular domain in the absence of the tail. Therefore, we speculated that the tail might shield, at least partially, exposed hydrophobic regions in the globular domain. An inspection of the 20 energy-minimized conformers in the NMR solution structure of the full-length protein indicates that this is likely the case (Figure 5B). Despite being devoid of any ordered motif, the N-terminus seems to dock preferentially on top of the globular domain, partially shielding β -strand 2, in which Val 29, Val30, and Phe32 are exposed to solvent.

N-Terminal Tail Protects SUMO Proteins from Aggregation into Amyloid-Like Structures. We have shown in a recent study that β -strand 2 is the most amyloidogenic region in the sequence of SUMO2 and it is fully protected in the amyloid-fibrils formed by the globular domain under destabilizing conditions.³⁰ The observation that the N-terminal tail could partially shield this strand from the solvent might imply a protective role of the tail against the aggregation of the globular domain. To test this possibility we incubated SUMO2 and Δ Nt-SUMO2 at 343 K for 400 min and afterward we monitored their binding to the amyloid detecting dye Thioflavin-T (Th-T; Figure 6A). Despite both samples promoted an increase in Th-T fluorescence emission, the presence of the N-terminal tail reduced Th-T fluorescence emission by 5-fold in comparison to the globular domain alone. Analysis of the two incubated protein solutions by far-UV CD after cooling at 293 K (Figure 6B), indicates that SUMO2 maintains/recovers essentially its native spectrum in terms of shape and intensity. In contrast, in the case of Δ Nt-SUMO2, despite the spectrum is essentially native, the intensity is 3-fold

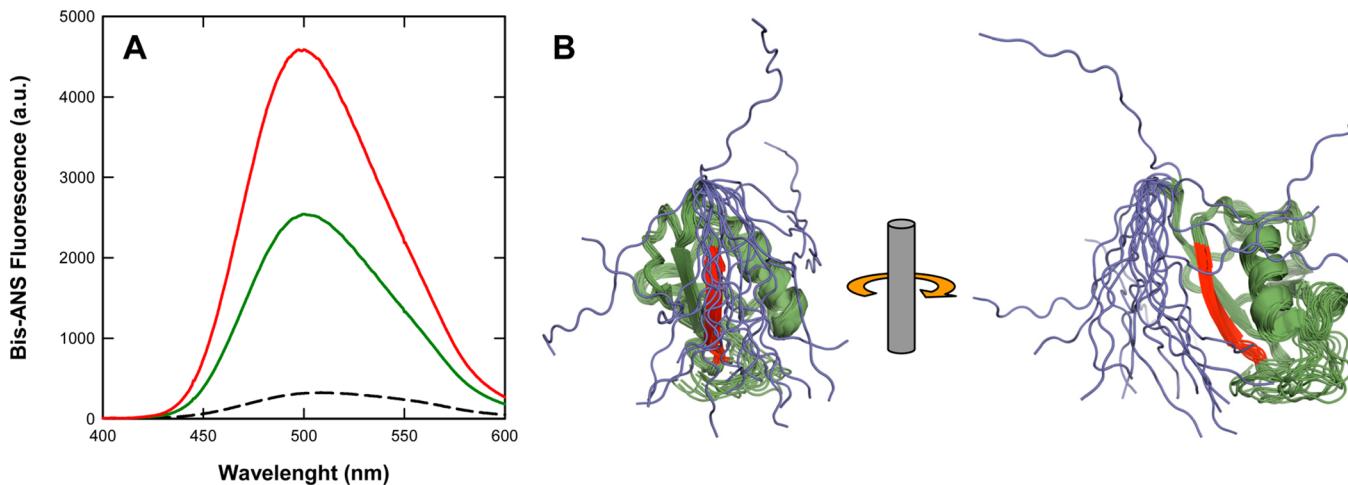


Figure 5. Exposure of Hydrophobic Clusters in SUMO2 and Δ Nt-SUMO2. (A) bis-ANS binding to SUMO2 (green) and Δ Nt-SUMO2 (red). The slashed line represents free bis-ANS emission spectrum. (B) Cartoon representations of the 20 conformers with lowest energy derived from the full-length (residues 1–95) SUMO2 solution structure determination (PDB: 2AWT). The conformers are structurally aligned considering solely the structured globular ubiquitin-like domain (residues 15–93). The Nter unstructured tail is shown in blue, and a previously identified aggregation-prone region [27] in red. The representations are rotated 90° over the z-axis one respect the other.

lower than in the case of SUMO2, suggesting that, in agreement with the Th-T binding data, a significant fraction of the protein has precipitated out of the solution. We used transmission electron microscopy to confirm these data. As it can be seen in Figure 6C, the Δ Nt-SUMO2 solution exhibits abundant amyloid fibrils, whereas the presence of the N-terminal tail completely abrogated the presence of long fibrils and only small aggregates were detected in SUMO2 solutions.

We addressed if the observed differences in aggregation properties might have a kinetic origin by monitoring the time course of SUMO2 and Δ Nt-SUMO2 aggregation at 343 K using Th-T (Figure 7A). The kinetics of Δ Nt-SUMO2 aggregation is fast, following an hyperbolic curve, reaching a plateau after 300 min and lacking any detectable lag phase. In contrast, SUMO2 exhibits a canonical sigmoidal aggregation curve that reflects a nucleation-dependent growth mechanism, with a lag time of 385 min, half aggregation time ($t_{1/2}$) of 824 min and reaching the final aggregated state after 1747 min, being thus exceedingly slow when compared with the Δ Nt-SUMO2 aggregation process.

As in the case of SUMO2, human SUMO1 possesses a N-terminal disordered tail consisting in this case of 18 residues (Figure 8). To assess if SUMO1 N-terminal tail plays also a protective role against aggregation, we expressed and purified SUMO1 and a N-terminal deletion mutant consisting only of the globular domain (Δ Nt-SUMO1), we incubated these two proteins at their melting temperature (333 K) and followed their aggregation kinetics as explained above. As it can be observed in Figure 7B, also in the case of SUMO1, the presence of the N-terminal tail results in a significantly slower aggregation reaction.

SUMO N-Terminal Tails Resemble Entropic Bristles. The presence of N-terminal extensions is a recurrent feature of SUMO proteins in eukaryotic organisms, from yeast to humans (Figure 8).

Table 1 shows the characteristics of disorder and solubility predictions for the N-terminal tails of human SUMO1 and SUMO2 and the SUMO proteins from *C. elegans*, *S. cerevisiae*, and *A. thaliana*. PONDR predictors VLXT and VSL2^{34,35} confirm a high disorder propensity in all N-terminal extensions.

In fact, despite the absence of sequence homology between SUMO N-terminal tails, they all share a net negative charge and a low mean hydrophobicity, a characteristic of disordered proteins. In terms of composition they only exhibit 7% of the so-called order-promoting amino acids (W, Y, F, I, L, V, C, and N), whereas 67% of the residues correspond to disorder-promoting amino acids (A, R, G, Q, S, E, K, and P). Using the Wilkinson–Harrison solubility model³³ we also show that all N-terminal tails are predicted to be highly soluble peptides and that their presence in full length SUMO proteins is predicted to increase the solubility of the corresponding globular domain in all cases. Overall, despite being shorter, N-terminal SUMO tails share many features with canonical EB, defined as long heterologous intrinsically disordered sequences able to enhance the solubility of aggregation-prone proteins when fused in trans at their N-terminus.²²

SUMO2 N-Terminal Tail Reduces the Intracellular Aggregation Propensity of the $\text{A}\beta42$ Peptide. We investigated if as it happens with EB, the SUMO2 N-terminal tail can also act in trans by reducing the aggregation propensity of an insoluble polypeptide when fused at its N-terminal side. We used a fusion to the highly amyloidogenic $\text{A}\beta42$ peptide as a proof of principle. We have previously shown that, when fused to the green fluorescent protein (GFP) and expressed in *E. coli*, the $\text{A}\beta42$ -GFP fusion forms inclusion bodies (IB) displaying amyloid-like properties.⁴³ Using a large set of mutants, we have demonstrated that, in this system, the presence of active GFP in the aggregates depends on the aggregation propensity of the $\text{A}\beta42$ variant.⁴⁴ The faster the fusion protein aggregates, the lower the IB fluorescence emission is and vice versa, in such a way that the fluorescence of IBs reports on the *in vivo* protein aggregation.⁴⁵

We fused the N-terminal tail of SUMO2 at the N-terminus of $\text{A}\beta42$ -GFP and expressed the new fusion (NterS2- $\text{A}\beta42$ -GFP) in bacteria. The original $\text{A}\beta42$ -GFP fusion was expressed as a control. Next, we measured the GFP fluorescence of intact cells expressing the two variants using spectrofluorimetry. As shown in Figure 9A, the two fusions differ in their fluorescence emission, cells expressing NterS2- $\text{A}\beta42$ -GFP variant being three times more fluorescent than control cells. We used

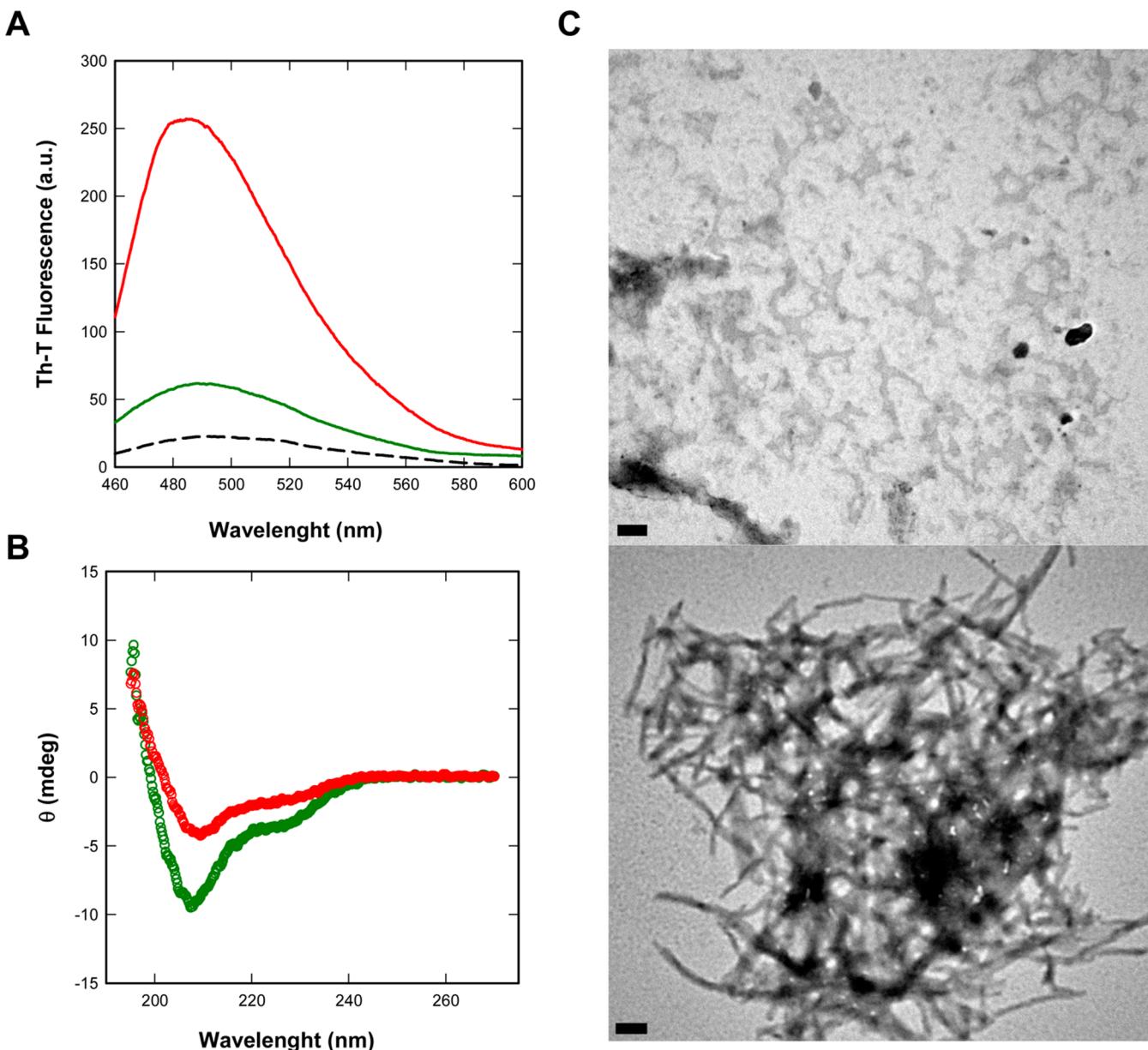


Figure 6. Morphological, Dye Binding, and Structural Properties of SUMO2 and Δ Nt-SUMO2 Aggregates. (A) Th-T emission spectra and (B) CD spectra of SUMO2 (green) and Δ Nt-SUMO2 (red) aggregates after 400 min incubation. The slashed line represents the emission spectrum of free Th-T. (C) Representative TEM micrographs showing aggregate morphology after 400 min incubation of SUMO2 (top) and Δ Nt-SUMO2 (bottom). The scale bar represents 1 μ m.

fluorescence microscopy to identify the cellular location of the detected GFP emission (Figure 9B). In both cases, the fluorescence is confined mainly in the IBs at the poles of the cell. However, the IBs formed by NterS2-A β 42-GFP are clearly more fluorescent than those formed by A β 42-GFP alone, indicating that the fusion in trans of the N-terminal tail of SUMO2 significantly reduces the intracellular aggregation propensity of the amyloidogenic A β 42 peptide. This solubilizing effect resembles the one exerted by an artificial segment comprising 19 repeats of the tetrapeptide sequence NANP, when fused at the N-terminus of A β 42 peptide.⁴⁶

Disordered Patterns at the N- and C-Termini of Globular Domains Display Low Aggregation Propensity. We wondered whether the antiaggregational properties of the N-terminal SUMO tails might be in fact a generic property of short sequences flanking globular domains. To study this

possibility we exploited a library of disordered patterns in 3D structures developed by Galzitskaya and co-workers.³⁶ They identified 109 disordered patterns of different lengths in disordered regions in the PDB. We examined the properties of the patterns identified at the first (N-tail) and last (C-tail) 40 residues of proteins. We analyzed their overall aggregation propensity and the presence of aggregation-prone regions using AGGRESCAN and their amyloidogenicity as predicted by WALTZ. We excluded from the analysis all the patterns containing three or more consecutive His residues, since they might correspond to artificial His-tags, reducing the data set to 71 different patterns (Table S1). AGGRESCAN indicates that the aggregation tendency of these sequences is extremely low, with an average value of -59.06, which contrasts with an average aggregation propensity of 0.023 for the ensemble of protein sequences in Swiss-Prot (the more negative the value

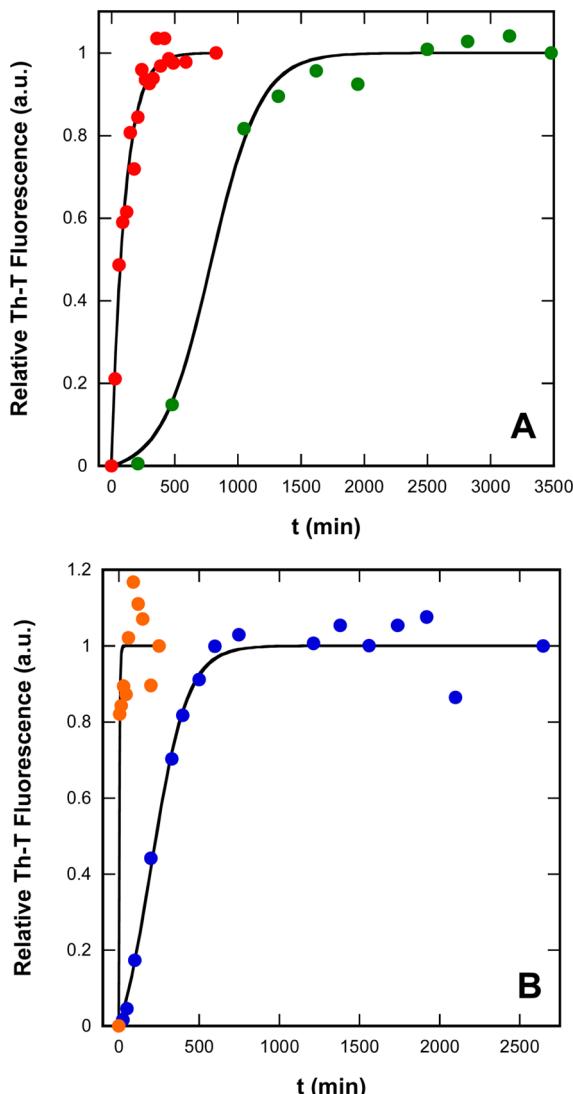


Figure 7. Aggregation Kinetics of SUMO Domains. The aggregation kinetics was monitored by following the change in relative Th-T fluorescence at different time points for (A) of SUMO2 (green) and Δ Nt-SUMO2 (red), and (B) SUMO1 (blue) and Δ Nt-SUMO1 (orange).

Hs SUMO1	MSDQEAKPSTEGLDKKEGEYIKLKVIQGQDSSEIHFKVK	39
Hs SUMO2	MADEKPKEGVKTENNDHINLKVGAGQDGSSVQFKIK	35
Ce SUMO	MADDAAQAGDNAEYIKIKVVGQDSNEVHFRVK	32
At SUMO2	MSATPEEDKKPDQGAHINLKVKGQDGNEVFFRIK	34
Sc Smt3	MSDSEVNQEAKPEVKPEVKPETHINKVS-DGSSEIFFKIK	40
	: * : * * : . . : * : * :	

Figure 8. Alignment of SUMO Sequences from Different Species. SUMO Nter region sequences from *Homo sapiens* (Hs), *Caenorhabditis elegans* (Ce), *Arabidopsis thaliana* (At), and *Saccharomyces cerevisiae* (Sc). The highlighted segment denotes the fragment which does not belong to the ubiquitin-like fold and that is unstructured in Hs SUMO1 and SUMO2, according to their solution structures (PDBs 1A5R and 2AWT, respectively).

the more soluble the protein). Comparison of the distribution of aggregation propensities of these sequential patterns with those of the sequences of 71 proteins randomly selected from the SCOP-derived ASTRAL40 data set⁴⁷ illustrates the comparatively low aggregation tendency of protein termini (Figure 10). In fact, 97% of the patterns are predicted to be

devoid of aggregation-prone or amyloidogenic sequences, thus, being potential EB.

CONCLUSIONS

It is now widely accepted that protein misfolding and aggregation impact cell physiology, reducing organisms cell fitness. Thus, selection against protein aggregation is expected to act as an important constraint in the evolution of protein sequences. Accordingly, proteins have adopted different structural and sequential strategies to prevent or reduce their aggregation propensity.

Disordered regions have been found in a large number of eukaryotic proteins.⁴⁸ These stretches are not equally distributed within protein sequences, with residues at the protein termini displaying on the average higher disorder propensity.³⁶ Disordered tails are not just flexible protrusions, instead they are being found associated with an increasing range of protein functions.²³ Computational simulations have suggested that one of the functions of disordered segments in proteins might be reduction of the aggregation propensity of the flanked domains.²¹ In particular, multiscale simulations have shown that the aggregation of fungal hydrophobins at air–water interfaces relies on the reduction of the flexibility of an inner loop, which is disordered in aqueous solution, where these proteins remain soluble.⁴⁹ The characterization of the conformational and aggregative properties of SUMO and Δ Nt-SUMO variants in the present study provides yet another experimental evidence for the antiaggregational role of the usually highly flexible termini in the context of a natural protein.

In vitro, the SUMO2 N-terminal tail does not affect significantly the secondary and tertiary structure of the globular domain, neither its conformational stability nor its function. However, it might transiently shield exposed hydrophobic residues in the amyloidogenic β -strand 2. This β -strand forms a functional intermolecular β -sheet with SUMO partners containing SUMO binding domains, being a major contributor to complex stability. Therefore, its intrinsic aggregation propensity responds to functional constraints and cannot be evolitively suppressed. The flexible N-terminal tail would contribute to decrease this aggregation tendency without compromising function.

The sequences of SUMOs N-tails in different organisms resemble the so-called EB. The solubilizing effects of disordered EB on globular proteins respond to two major effects: (1) a relatively larger surface able to interact favorably with water and (2) a large excluded volume that would restrict intermolecular contacts between aggregation-prone regions in proteins. These two factors are highly inter-related and would act increasing the entropic cost for protein aggregation resulting in a higher free energy barrier. In fact, the behavior of EB resembles that of polymers like polyethylene glycol,²² which chemical attachment to globular proteins significantly increases their solubility.⁵⁰ This generic ability to increase solubility explains why these disordered segments, as we show here for SUMO2 N-terminal tail, can act in trans.²² The computational analysis of the most frequent sequential patterns in disordered N- and C- termini shows that they are highly soluble and skip the presence of dangerous aggregation/amyloidogenic sequences, arguing that their recurrent presence in proteins might constitute a negative design strategy to maintain solubility. Our results suggest that dynamic aspects should be taken into account when addressing the aggregation/solubility properties of proteins.

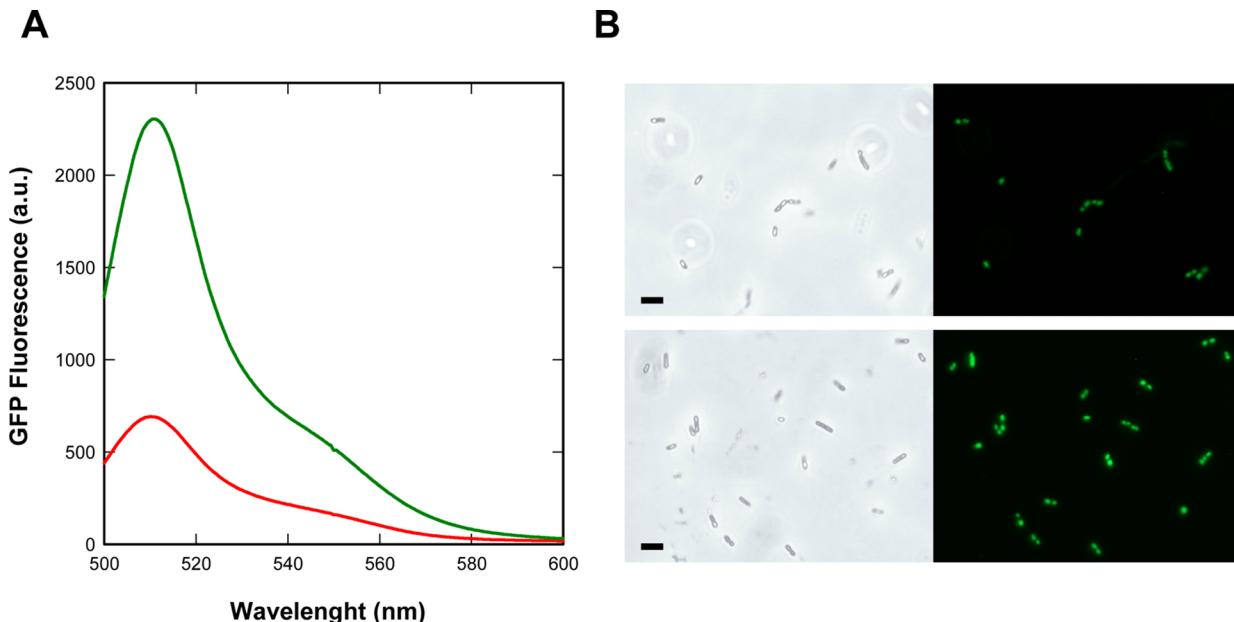


Figure 9. In vivo analysis of GFP fusions solubility. (A) Emission spectra of intact cells suspensions expressing the A β 42-GFP fusion (red) or the S2Nt-A β 42-GFP fusion (green). (B) Intact cells imaged using phase-contrast microscopy (left) and fluorescence microscopy (right). Upper panels show bacteria expressing the A β 42-GFP fusion and lower panels cells expressing the S2Nt-A β 42-GFP fusion. The scale bar represents 5 μ m.

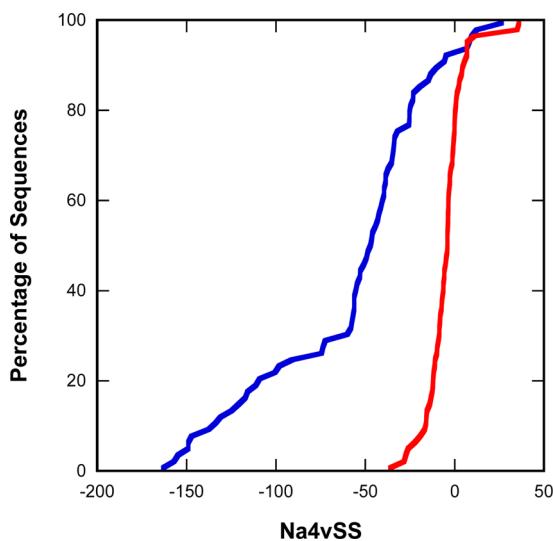


Figure 10. Aggregation properties of disordered patterns and globular proteins. Distribution of the average aggregation propensity (Na4vSS) computed by the AGGRESCAN algorithm for 71 disordered patterns (blue) retrieved from Galzitskaya and co-workers library³⁶ and 71 globular proteins (red) randomly selected from the Astral40 data set.

ASSOCIATED CONTENT

Supporting Information

Table S1. List of the 71 patterns retrieved from the disordered patterns library constructed by Galzitskaya and co-workers.³⁶ This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 34-93-5868956. Fax: 34-93-5811264. E-mail: salvador.ventura@uab.es.

Author Contributions

[†]These authors contributed equally (R.G.-M. and P.M.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Work in our lab is supported by Grants BFU2010-14901 from Ministerio de Ciencia e Innovacion (Spain) and 2009-SGR-760 from AGAUR (Generalitat de Catalunya). S.V. has been granted an ICREA Academia award (ICREA). R.G.-M. is the beneficiary of the FPU AP2009-0948 fellowship from the Ministerio de Educaci \tilde{n} n (Spain) and P.M. is the beneficiary of a FI-DGR 2012 fellowship from the AGAUR (Generalitat de Catalunya).

REFERENCES

- (1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333.
- (2) Fernandez-Busquets, X.; de Groot, N. S.; Fernandez, D.; Ventura, S. *Curr. Med. Chem.* **2008**, *15*, 1336.
- (3) Silva, J. L.; Rangel, L. P.; Costa, D. C.; Cordeiro, Y.; De Moura Gallo, C. V. *Biosci. Rep.* **2013**, *33*, e00054.
- (4) Xu, J.; Reumers, J.; Couceiro, J. R.; De Smet, F.; Gallardo, R.; Rudyak, S.; Cornelis, A.; Rozenski, J.; Zwolinska, A.; Marine, J. C.; Lambrechts, D.; Suh, Y. A.; Rousseau, F.; Schymkowitz, J. *Nat. Chem. Biol.* **2011**, *7*, 285.
- (5) Jahn, T. R.; Radford, S. E. *FEBS J.* **2005**, *272*, 5962.
- (6) Dobson, C. M. *Semin. Cell Dev. Biol.* **2004**, *15*, 3.
- (7) Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L. *J. Mol. Biol.* **2004**, *342*, 345.
- (8) Rousseau, F.; Serrano, L.; Schymkowitz, J. W. *J. Mol. Biol.* **2006**, *355*, 1037.
- (9) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 3487.
- (10) Chiti, F.; Dobson, C. M. *Nat. Chem. Biol.* **2009**, *5*, 15.
- (11) Gsponer, J.; Babu, M. M. *Cell Rep.* **2012**, *2*, 1425.
- (12) Sanchez de Groot, N.; Torrent, M.; Villar-Pique, A.; Lang, B.; Ventura, S.; Gsponer, J.; Babu, M. M. *Biochem. Soc. Trans.* **2012**, *40*, 1032.
- (13) Monsellier, E.; Chiti, F. *EMBO Rep.* **2007**, *8*, 737.

- (14) Tartaglia, G. G.; Pechmann, S.; Dobson, C. M.; Vendruscolo, M. *Trends Biochem. Sci.* **2007**, *32*, 204.
- (15) Reumers, J.; Maurer-Stroh, S.; Schymkowitz, J.; Rousseau, F. *Hum. Mutat.* **2009**, *30*, 431.
- (16) Tyedmers, J.; Mogk, A.; Bukau, B. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 777.
- (17) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. *Nature* **2011**, *475*, 324.
- (18) Steward, A.; Adhya, S.; Clarke, J. *J. Mol. Biol.* **2002**, *318*, 935.
- (19) Parrini, C.; Taddei, N.; Ramazzotti, M.; Degl'Innocenti, D.; Ramponi, G.; Dobson, C. M.; Chiti, F. *Structure (Oxford, U. K.)* **2005**, *13*, 1143.
- (20) Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2754.
- (21) Abeln, S.; Frenkel, D. *PLoS Comput. Biol.* **2008**, *4*, e1000241.
- (22) Santner, A. A.; Croy, C. H.; Vasanwala, F. H.; Uversky, V. N.; Van, Y. Y.; Dunker, A. K. *Biochemistry* **2012**, *51*, 7250.
- (23) Uversky, V. N. *FEBS Lett.* **2013**, *587*, 1891.
- (24) Bandaru, V.; Cooper, W.; Wallace, S. S.; Doublie, S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1142.
- (25) Park, S. M.; Jung, H. Y.; Chung, K. C.; Rhim, H.; Park, J. H.; Kim, J. *Biochemistry* **2002**, *41*, 4137.
- (26) Johnson, E. S. *Annu. Rev. Biochem.* **2004**, *73*, 355.
- (27) Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425.
- (28) Hay, R. T. *Mol. Cell* **2005**, *18*, 1.
- (29) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenike, R.; Becker, J. *J. Mol. Biol.* **1998**, *280*, 275.
- (30) Sabate, R.; Espargaro, A.; Grana-Montes, R.; Reverter, D.; Ventura, S. *Biomacromolecules* **2012**, *13*, 1916.
- (31) Villar-Pique, A.; de Groot, N. S.; Sabate, R.; Acebron, S. P.; Celaya, G.; Fernandez-Busquets, X.; Muga, A.; Ventura, S. *J. Mol. Biol.* **2012**, *421*, 270.
- (32) Sabate, R.; Villar-Pique, A.; Espargaro, A.; Ventura, S. *Biomacromolecules* **2012**, *13*, 474.
- (33) Wilkinson, D. L.; Harrison, R. G. *Bio/Technology* **1991**, *9*, 443.
- (34) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38.
- (35) Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z. *BMC Bioinf.* **2006**, *7*, 208.
- (36) Lobanov, M. Y.; Furletova, E. I.; Bogatyreva, N. S.; Roytberg, M. A.; Galzitskaya, O. V. *PLoS Comput. Biol.* **2010**, *6*, e1000958.
- (37) Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S. *BMC Bioinf.* **2007**, *8*, 65.
- (38) Maurer-Stroh, S.; Debulpaep, M.; Kuemmerer, N.; Lopez de la Paz, M.; Martins, I. C.; Reumers, J.; Morris, K. L.; Copland, A.; Serpell, L.; Serrano, L.; Schymkowitz, J. W.; Rousseau, F. *Nat. Methods* **2010**, *7*, 237.
- (39) Reverter, D.; Lima, C. D. *Nature* **2005**, *435*, 687.
- (40) Pichler, A.; Knipscheer, P.; Oberhofer, E.; van Dijk, W. J.; Korner, R.; Olsen, J. V.; Jentsch, S.; Melchior, F.; Sixma, T. K. *Nat. Struct. Mol. Biol.* **2005**, *12*, 264.
- (41) Louis-Jeune, C.; Andrade-Navarro, M. A.; Perez-Iratxeta, C. *Proteins* **2011**, DOI: 10.1002/prot.23188.
- (42) de Groot, N. S.; Parella, T.; Aviles, F. X.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732.
- (43) Morell, M.; Bravo, R.; Espargaro, A.; Sisquella, X.; Aviles, F. X.; Fernandez-Busquets, X.; Ventura, S. *Biochim. Biophys. Acta* **2008**, *1783*, 1815.
- (44) de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Ventura, S. *FEBS J.* **2006**, *273*, 658.
- (45) Villar-Pique, A.; de Groot, N. S.; Sabate, R.; Acebron, S. P.; Celaya, G.; Fernandez-Busquets, X.; Muga, A.; Ventura, S. *J. Mol. Biol.* **2012**, *421*, 270.
- (46) Finder, V. H.; Vodopivec, I.; Nitsch, R. M.; Glockshuber, R. J. *Mol. Biol.* **2010**, *396*, 9.
- (47) Chandonia, J. M.; Hon, G.; Walker, N. S.; Lo Conte, L.; Koehl, P.; Levitt, M.; Brenner, S. E. *Nucleic Acids Res.* **2004**, *32*, D189.
- (48) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635.
- (49) De Simone, A.; Kitchen, C.; Kwan, A. H.; Sunde, M.; Dobson, C. M.; Frenkel, D. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 6951.
- (50) Milla, P.; Dosio, F.; Cattel, L. *Curr. Drug Metab.* **2012**, *13*, 105.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article posted ASAP on March 6, 2014. Numerous changes have been made throughout the paper. The correct version posted March 10, 2014.