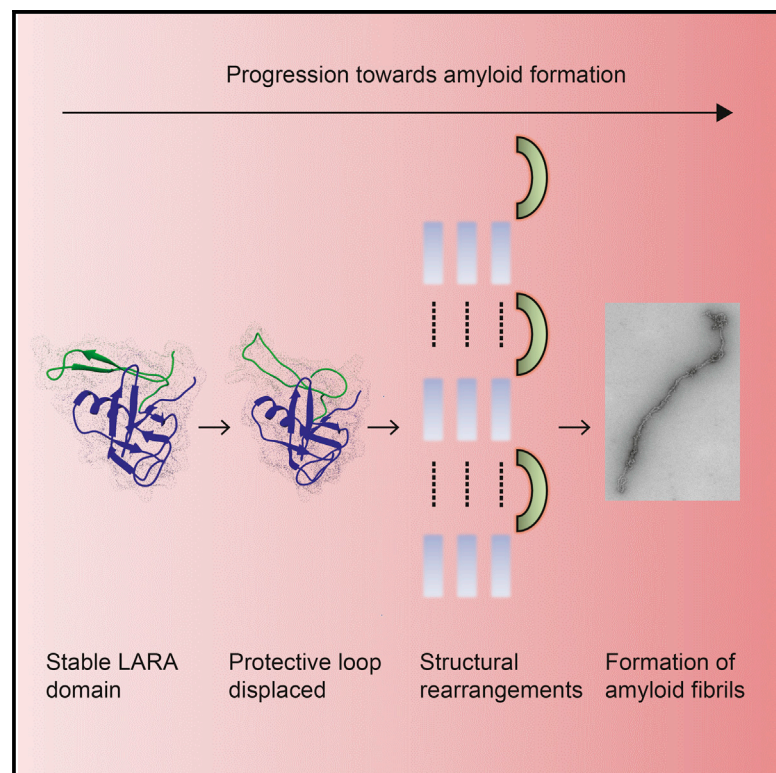


Structure

Mechanism of Amyloidogenesis of a Bacterial AAA+ Chaperone

Graphical Abstract



Authors

Sze Wah Samuel Chan, Jason Yau, Christopher Ing, ..., Régis Pomès, Simon Sharpe, Walid A. Houry

Correspondence

ssharpes@sickkids.ca (S.S.),
walid.houry@utoronto.ca (W.A.H.)

In Brief

Chan et al. make an unexpected observation that bacterial chaperone-like ATPase RavA and specifically the LARA domain within RavA readily form amyloids under acidic conditions at elevated temperatures. Experimental and theoretical studies revealed that the folded core of LARA is amyloidogenic and is protected by its N-terminal loop.

Highlights

- Bacterial chaperone ATPase RavA was unexpectedly observed to form an amyloid state
- LARA domain of RavA drives amyloidogenesis at low pH and high temperatures
- LARA stability and dynamics were studied in detail using NMR and MD simulations
- Destabilization of the N-terminal loop of LARA initiates amyloid formation



Mechanism of Amyloidogenesis of a Bacterial AAA+ Chaperone

Sze Wah Samuel Chan,^{1,5} Jason Yau,^{1,2,5} Christopher Ing,^{1,2} Kaiyin Liu,¹ Patrick Farber,² Amy Won,³ Vaibhav Bhandari,¹ Nareg Kara-Yacoubian,¹ Thiago V. Seraphim,¹ Nilmadhab Chakrabarti,² Lewis E. Kay,^{1,2,4} Christopher M. Yip,^{1,3} Régis Pomès,^{1,2} Simon Sharpe,^{1,2,*} and Walid A. Houry^{1,*}

¹Department of Biochemistry, University of Toronto, 1 King's College Circle, Medical Sciences Building, Toronto, ON M5S 1A8, Canada

²Molecular Structure and Function Program, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada

³Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3E1, Canada

⁴Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

⁵Co-first author

*Correspondence: ssharpe@sickkids.ca (S.S.), walid.houry@utoronto.ca (W.A.H.)

<http://dx.doi.org/10.1016/j.str.2016.05.002>

SUMMARY

Amyloids are fibrillar protein superstructures that are commonly associated with diseases in humans and with physiological functions in various organisms. The precise mechanisms of amyloid formation remain to be elucidated. Surprisingly, we discovered that a bacterial *Escherichia coli* chaperone-like ATPase, regulatory ATPase variant A (RavA), and specifically the LARA domain in RavA, forms amyloids under acidic conditions at elevated temperatures. RavA is involved in modulating the proper assembly of membrane respiratory complexes. LARA contains an N-terminal loop region followed by a β -sandwich-like folded core. Several approaches, including nuclear magnetic resonance spectroscopy and molecular dynamics simulations, were used to determine the mechanism by which LARA switches to an amyloid state. These studies revealed that the folded core of LARA is amyloidogenic and is protected by its N-terminal loop. At low pH and high temperatures, the interaction of the N-terminal loop with the folded core is disrupted, leading to amyloid formation.

INTRODUCTION

External influences such as thermal or acid stress can disrupt the native state of proteins, leading to exposure of aggregation-prone regions and giving rise to formation of non-native intermolecular β sheets. This can create a nucleus for protein aggregation, leading to the formation of amyloid fibrils (Eisenberg and Jucker, 2012). Amyloid fibrils are characterized by a cross β -sheet quaternary structure stacked perpendicular to the fibril axis (Moran and Zanni, 2014). This structure gives rise to protofilaments composed of core β -sheet strands strongly held together by a vast hydrogen-bonding network, which confers resistance to denaturation and proteolytic cleavage. Multiple protofilaments intertwine to form mature amyloid fibrils

containing a twisted spine arrangement having regular periodicity (Toyama and Weissman, 2011).

De novo kinetic formation of amyloids has typically been modeled by nucleation-dependent mechanisms comprising two distinct events: the formation of a thermodynamically disfavored nucleus followed by rapid elongation of thermodynamically favored fibrils (Knowles et al., 2014). However, key questions remain about the mechanisms that initiate the amyloid formation cascade. Considerable evidence in the field suggests that the process starts due to the exposure of amyloidogenic fragments that initiate assembly into large oligomers (Esteras-Chopo et al., 2005).

While many amyloids have been associated with diseases, bacterial amyloids have been shown to carry out diverse cellular roles. The most well-characterized functional bacterial amyloids are the extracellular curli fibers produced by many species of Enterobacteriaceae, and which are found to be involved in surface adhesion, cell aggregation, and biofilm formation (Barnhart and Chapman, 2006). The formation of amyloids has also been found to regulate protein toxicity. For example, microcin E492 from *Klebsiella pneumoniae* forms an amyloid state that abolishes its toxicity (Marcoleta et al., 2013). Hence, knowledge of the form and function of bacterial amyloids is important for understanding bacterial physiology and pathogenicity.

We present the first report of a novel intracellular bacterial amyloid formed by the chaperone termed regulatory ATPase variant A (RavA). RavA is a hexameric gammaproteobacterial ATPase that belongs to the MoxR family of the ATPases associated with diverse cellular activities (AAA+) superfamily. Based on the X-ray crystal structure of RavA (El Bakkouri et al., 2010), a RavA protomer can be divided into three domains (Figures 1A and S1): an N-terminal AAA+ domain (containing the Walker A and B conserved nucleotide-binding motifs), a discontinuous triple-helical bundle, and a LARA (LdcI-associating domain of RavA) domain. The LARA domain is a protein-protein interaction domain required for association of RavA with an acid stress enzyme, the inducible lysine decarboxylase (LdcI) (El Bakkouri et al., 2010; Kanjee et al., 2011; Snider et al., 2006). Structurally, the LARA domain contains an N-terminal loop region (Q329–S360) and a folded core (T361–E440) that forms a β -sandwich-like fold (Figure 1A). The binding of RavA to LdcI

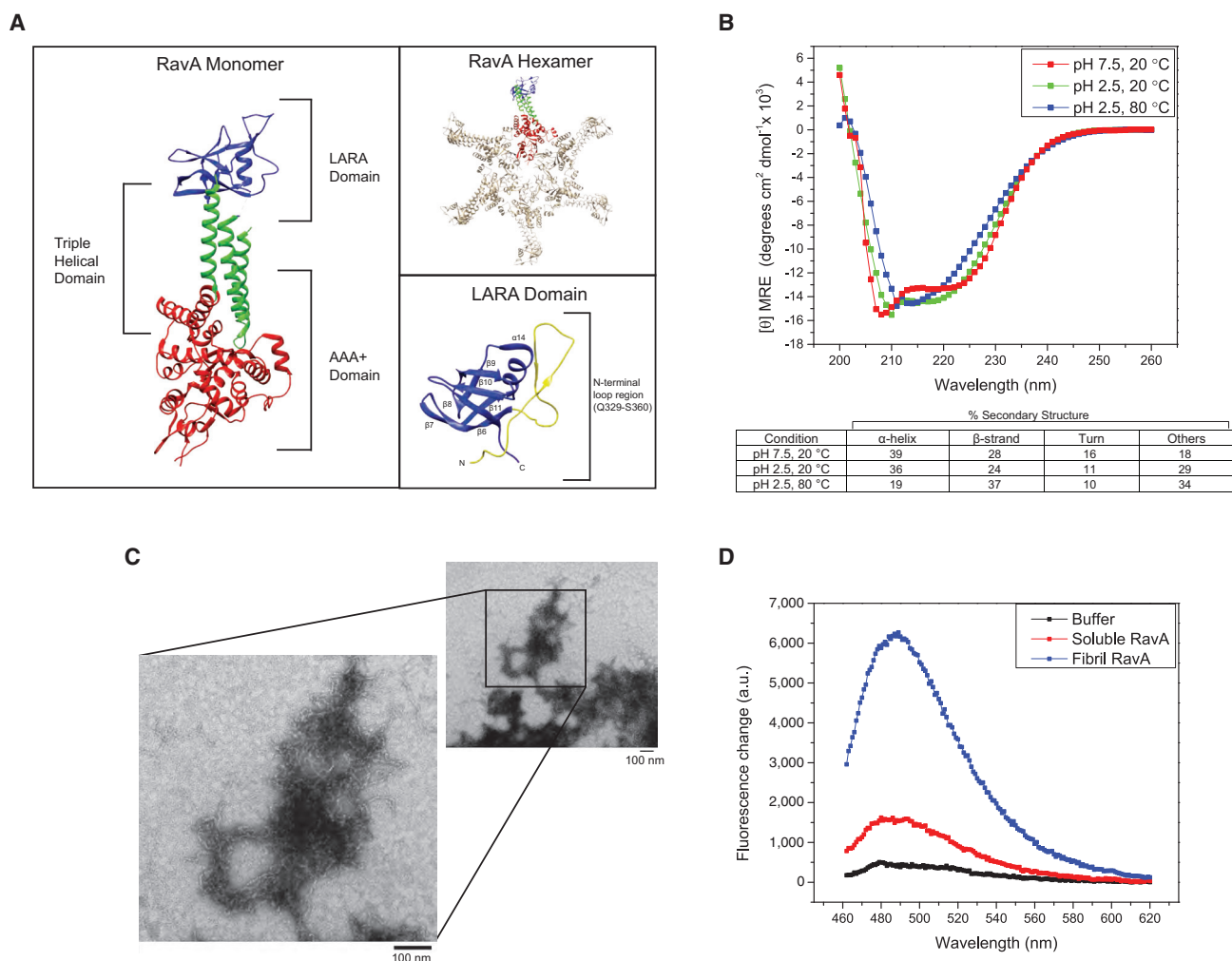


Figure 1. Formation of Amyloid Fibrils by RavA

(A) Structure of Regulatory ATPase Variant (RavA) as described in [El Bakkouri et al. \(2010\)](#). Highlighted on the left are the three main domains of RavA: the AAA+ domain (red), the triple-helical bundle (green), and the LARA domain (blue). A RavA hexamer is shown in the top right box. A close-up view of the LARA domain is shown in the bottom-right box with the N-terminal loop in yellow. N refers to the N terminus and C refers to the C terminus. Secondary structure elements are labeled. See also [Figure S1](#).

(B) CD spectra of RavA in pH 7.5 buffer at 20°C (red) and in pH 2.5 buffer at 20°C (green) and 80°C (blue). Predicted secondary structure content based on deconvolution using BeStSel ([Micsonai et al., 2015](#)) is shown underneath the figure.

(C) Negative-stain TEM images of RavA fibrils at 100,000× magnification.

(D) ThT fluorescence binding assay showing an emission wavelength scan from 462 to 620 nm for fibrillar RavA (blue), soluble RavA (red), and buffer (black).

results in the formation of a large cage-like structure that prevents the interaction of the alarmone, ppGpp, with LdcI ([El Bakkouri et al., 2010; Kanjee et al., 2011](#)). RavA has proposed chaperone activity involved in the maturation of the bacterial Nuo respiratory complex together with a cofactor protein termed ViaA ([Wong et al., 2014](#)). However, the functions and mechanisms of action of RavA, RavA-ViaA, and RavA-LdcI remain to be fully elucidated.

Full-length RavA was found to form an amyloid structure under conditions of low pH and elevated temperature, which is driven mainly by the LARA domain. Using various biochemical and biophysical approaches, we find that the N-terminal loop region of the LARA domain has to rearrange to allow for amyloid formation by the β-rich core of the domain.

RESULTS

RavA Forms an Amyloid State at Low pH and High Temperature

Since RavA interacts with the acid stress protein LdcI ([Snider et al., 2006](#)), we became interested in studying the structural properties of RavA under acidic conditions. At pH 7.5 and 20°C, the circular dichroism (CD) spectrum of RavA was consistent with a predominately α-helical structure, with characteristic minima at 208 and 222 nm ([Figure 1B](#)). At pH 2.5 and 20°C, RavA displayed a similar CD profile to that observed at pH 7.5, but with slightly lower β-strand and turn content. However, at pH 2.5 and 80°C, the proportion of β-strand content increased to become the dominant secondary structure, concurrent with a decrease

in the α -helical signal (Figure 1B). Under these conditions, RavA formed aggregates that were visualized using transmission electron microscopy (TEM), revealing distinct fibrillar morphology (Figure 1C). The fibrils were non-rod like, instead having a curly appearance. Most of the fibrils appeared to bundle or twist together to form worm-like structures. These fibrils gave rise to an increase in thioflavin-T (ThT) fluorescence emission at 480 nm compared with soluble protein (Figure 1D), indicating the presence of extensive cross- β organization in the fibrils (Biancalana and Koide, 2010). Overall, these experiments suggest that acidic conditions and high temperatures induce a conformational change in RavA, resulting in the formation of a predominantly β -sheet fibrillar assembly.

Formation of an Amyloid State by the LARA Domain

To determine which regions of RavA might be responsible for the formation of the amyloid state, we carried out *in silico* prediction of amyloidogenic sequences within RavA using MetAMYL (Emily et al., 2013), PASTA2.0 (Walsh et al., 2014), and AMYLPRED2 (Tsolis et al., 2013). Regions showing a consensus among two or three of the methods are indicated in Figure S1. Relatively few segments in the AAA+ domain and the triple-helical bundle had predicted amyloidogenic regions. Most of these predicted segments are in α helices, consistent with the suggestion that evolutionary pressure on protein sequence places amyloidogenic regions in α helices to prevent β -sheet aggregation (Mon-sellier and Chiti, 2007). On the other hand, the LARA domain exhibited many amyloidogenic regions that converged spatially around the β -sheet-rich core. Hence, we focused our subsequent studies on the isolated LARA domain.

As the LARA domain was heated at pH 6 from 20°C to 80°C, the protein structure was unfolded as shown by a loss of CD ellipticity at 217 nm (Figures 2A and S2A). Conversely, at pH 5 and lower, there was a gain in β -sheet structure as the temperature was increased (Figures 2A and S2A), with optimal conversion to a β -sheet-rich conformation at pH 4. Based on a sigmoidal fit of the melting curve (Figure 2A), the T_m for the structural conversion at pH 4 was 42°C \pm 0.1°C. We speculated that under these conditions, LARA was misfolding into an amyloid fibrillar state, which we name LARA^{fib} (as opposed to the native, soluble LARA^{sol}). Based on CD fits for experiments at pH 4 and 20°C, there was an average of 25% \pm 2% antiparallel β -sheet and 0% parallel β -sheet content predicted in the overall secondary structure for LARA^{sol} (Figure S2A). For LARA^{fib}, the predicted secondary structure was 34% \pm 5% antiparallel β -sheet and 11% \pm 1% parallel β -sheet content at pH 4 and 50°C (Figure S2A).

The amyloid nature of the LARA^{fib} state was confirmed by using ThT fluorescence. LARA^{fib} formed by heating at pH 4 exhibited increased ThT fluorescence compared with buffer and LARA^{sol} (Figure 2B). Based on Fourier transform infrared (FTIR) spectroscopy data, distinct structural differences between LARA^{sol} and LARA^{fib} were observed (Figure 2C). The amide I region (1,600–1,700 cm⁻¹) reflects the vibrational stretching of the amide backbone and is indicative of secondary structure (Sarroukh et al., 2013). Signal deconvolution using second derivative analysis was used to identify peaks that were subsequently fit to estimate secondary structure content. FTIR spectra of LARA^{sol} show a β -sheet-dominant structure,

while deconvolution of the LARA^{fib} FTIR spectra suggested a slight reduction in helix and coil content (Table S1). The shift in the β -sheet peaks to lower wavenumbers (for example, the dominant β -sheet peaks shifted from 1,640 cm⁻¹ in LARA^{sol} to 1,619 cm⁻¹ in LARA^{fib}) is indicative of rearrangement of globular β sheets to extended β sheets, which is consistent with previous FTIR reports of amyloid fibrils (Sarroukh et al., 2013).

The LARA^{sol} FTIR spectrum has a high-frequency band at 1,690 cm⁻¹ that indicates the presence of antiparallel β sheets. The FTIR spectrum of LARA^{fib} exhibited a distinct increase in a similar region at 1,693 cm⁻¹, suggesting that LARA^{fib} has higher antiparallel β -sheet content than LARA^{sol}. The β -sheet organizational index (intensity of the high-frequency component of antiparallel β sheets/intensity of the low frequency component of β sheets) is another measure used to compare antiparallel and parallel β -sheet content (Guo and Wang, 2012; Sarroukh et al., 2013; Zou et al., 2013). Antiparallel β -sheet content is correlated to the value of this index. For LARA^{sol}, the ratio is 0.22 compared with 0.39 for LARA^{fib}. The FTIR and CD data together suggest that LARA^{fib} has increased antiparallel β -sheet content compared with LARA^{sol}.

When imaged using TEM (Figures 2D and S2B), LARA domain fibrils were observed to have the same twisted, curly fibril morphology observed for RavA (Figure 1C) and exhibited a similar worm-like clumping behavior. The morphology of these fibrils may indicate similarity between the assembly of LARA^{fib} and fibrillar RavA. Preparation of LARA fibrils in different salt concentrations of 150 mM KCl, 50 mM KCl, and 10 mM KCl had no significant impact on the observed morphology of the fibrils (Figure S2B), indicating a lack of salt dependence within this salt concentration range.

The kinetics of LARA^{fib} formation were monitored by ThT fluorescence at 33°C and pH 4 (Figure 2E). The $t_{0.5}$ for the fibrillization reaction was 373 \pm 11 min obtained using a sigmoidal fit. Seeds created from sonicated fibrils decreased the $t_{0.5}$ to 277 \pm 12 min with a corresponding 30-min decrease in the lag phase (Figure 2E and Table S2). With decreasing protein concentration, the lag phase and $t_{0.5}$ increased (Table S2), showing the expected dependence of nucleation events on protein concentration.

Subsequently, to identify the core region of LARA^{fib}, the broad-spectrum endopeptidase proteinase K was used to digest the fibrils. As shown in Figure S2C, SDS-PAGE analysis demonstrated that the majority of LARA^{sol} was digested by 30–45 min. In contrast, several fragments of LARA^{fib} remained protease resistant after 60 min of digestion. Digestion with trypsin followed by mass spectrometry was used to identify the components of the A (full length), B, and C bands observed for the proteinase K-treated fibril samples. Based on this analysis (Figure S2D and Table S3), we find that peptides corresponding to the N- and C-terminal ends of the protein are removed by proteinase K, while a fibril core corresponding to residues Q349–R423 was most resistant to proteolysis.

Monitoring the Structural Changes Leading to LARA Amyloid Formation by NMR Spectroscopy

LARA^{sol} backbone ¹H, ¹⁵N, and ¹³C chemical shifts were assigned at pH 4 and 7°C, using a standard suite of triple-resonance experiments (Sattler et al., 1999) and were compared

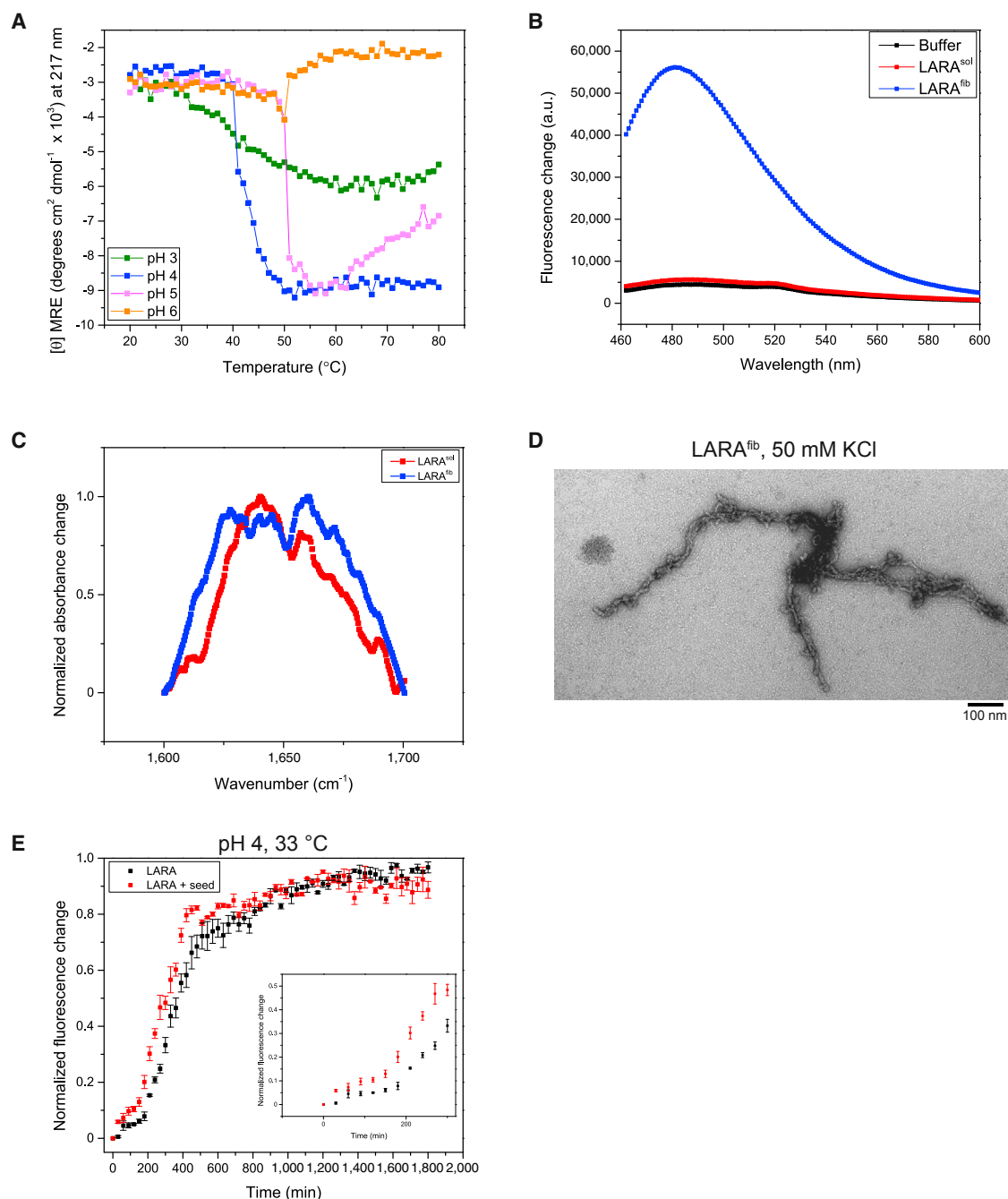


Figure 2. Amyloid Fibrils Formed by the LARA Domain of Rava

(A) The CD ellipticity of the LARA domain at 217 nm measured as a function of temperature (20°C–80°C) at pH 3 (green), pH 4 (blue), pH 5 (light magenta), and pH 6 (orange). See also [Figure S2A](#).

(B) ThT fluorescence emission spectra after addition of buffer (black), LARA^{sol} (red), and LARA^{fib} (blue).

(C) Amide I region from the FTIR spectra of LARA^{sol} (red) and LARA^{fib} (blue). See also [Table S1](#).

(D) TEM images of negatively stained LARA^{fib} prepared in 50 mM KCl at 120,000× magnification. See also [Figure S2B](#).

(E) ThT fluorescence emission at 480 nm monitored over time for LARA^{sol} alone (black) and LARA^{sol} + 2.5% (v/v) LARA^{fib} seeds (red). The inset shows a close-up of the lag phase from 0 to 300 min. Errors bars represent the SEM for each data point (n = 3). See also [Table S2](#).

with expected chemical shifts calculated from the crystal structure (El Bakkouri et al., 2010) (PDB: 3NBX) using SHIFTX2 (Han et al., 2011). A least-squared fit was performed comparing both C α (R² = 0.943) and C β (R² = 0.991) resonances (Fig-

ure S3A). In addition, the secondary structure of LARA^{sol} was predicted from the chemical-shift data using TALOS+ (Shen et al., 2009) and the Chemical Shift Index (Wishart and Sykes, 1994), the results of which were also compared with the

secondary structure elements identified in the crystal structure (Figure S3B). Both methods showed good agreement with the crystal structure (Figure S3B), indicating that the structure obtained at pH 6.5 using X-ray crystallography is similar to the LARA domain structure in solution at pH 4. The only structural differences suggested by our data were that the very N and C termini of the LARA domain were unstructured in solution, as expected for an isolated domain. Therefore, the X-ray structure was used as a model for the LARA domain when analyzing the experiments described below.

Heteronuclear single-quantum correlation (HSQC) spectra of LARA^{sol} were recorded from 10°C to 30°C in a 5°C step gradient (Figure 3A), and then from 30°C to 40°C in a 2°C step gradient (data not shown). Fibrillization occurred between 30°C and 40°C and resulted in the loss of most resonances at 40°C due to significant line broadening, in agreement with the formation of large protein assemblies (Figure 3B). In these experiments, only 19 residues could be identified at 40°C (Table S4) that are primarily located in the N and C termini of the LARA domain, and which likely remain highly disordered and mobile in the fibril state. During the pre-unfolding transition, 99 residues could be tracked across the protein structure, providing probes for the conversion process from LARA^{sol} to LARA^{fib}. During this transition the chemical shifts did not collapse into the random-coil region, suggesting that LARA^{sol} does not globally unfold prior to oligomer formation.

The changes in chemical shift as a function of temperature were analyzed as shown in Figure 3C (examples are shown for boxed areas of Figure 3A) and are summarized in Figure 3D. Of particular interest, residues localized to the N-terminal loop (for example, Q350 and G343 in Figure 3C) exhibited significantly larger changes in chemical shift as a function of temperature when compared with residues in the folded core (Q418 and A386 in Figure 3C). All residues exhibited a linear change in ¹H and ¹⁵N chemical shift with temperature, which has been shown to reflect the pre-unfolding transition (Farber et al., 2010). Overall trends from ¹⁵N Δδ showed trends similar to those of ¹H Δδ (data not shown). The ¹H temperature coefficients corresponding to chemical-shift change per °C were then calculated (Cierpicki and Otlewski, 2001; Cordier and Grzesiek, 2002; Hong et al., 2013). These ¹H temperature coefficients are sensitive probes of hydrogen-bond stretching as well as hydrogen-bond interactions (Hong et al., 2013). Coefficients that are more negative than −4.5 ppb/°C correspond to residues forming hydrogen bonds with H₂O, while coefficients that are less negative than −4.5 ppb/°C correspond to sites forming protein-protein hydrogen bonds (Baxter and Williamson, 1997). Three regions of high thermal sensitivity stand out in this analysis (Figure 3D and Table S5): the N-terminal loop region, backside region of the LARA domain β-rich core (G395–S416), and the C terminus. The backside region exhibited a high temperature sensitivity, which may be due to exposure of these non-planar β sheets to solvent as a function of temperature. The C-terminal region is more disordered, and contains fewer hydrogen bonds, resulting in increased thermal instability. The N-terminal loop was generally more sensitive to temperature, with the heart-shaped part of the loop (V338–Y351) exhibiting the highest local thermal sensitivity, suggesting that this loop region undergoes a conformational change to begin the misfolding process.

Information on fast timescale protein dynamics was obtained from ¹⁵N R₂, R₁ nuclear magnetic resonance (NMR) relaxation measurements and heteronuclear ¹H, ¹⁵N nuclear Overhauser effect measurements (pH 4, 24°C) carried out at 600 and 700 MHz. Lipari-Szabo model-free analysis (d'Auvergne and Gooley, 2008a, 2008b) was carried out on the LARA domain and the generalized order parameters (S²) were determined (Figure S4). The protein can be divided into several segments based on the dynamics data. We observe a highly dynamic heart-shaped N-terminal loop with S² values lower than 0.75 consisting of residues V338–Y351. In addition, there is a rigid protein core of residues T361–A433, which mostly have S² values greater than 0.75. This analysis shows that the N-terminal loop (Q329–S360) may be amenable to structural rearrangements necessary for amyloid fibril formation.

Probing Conformationally Excited States of the LARA Domain Induced by Low pH Using CPMG Relaxation Dispersion Experiments

To explore potential excursions from the ground state (LARA^{sol}) to sparsely populated, transiently formed conformers (referred to here as excited states), we performed ¹⁵N Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG-RD) experiments (Korzhnev and Kay, 2008; Tollinger et al., 2001) under mildly destabilizing conditions (pH 4 and 24°C). These experiments permit the atomic resolution study of folding/unfolding events occurring in proteins on the millisecond exchange timescale by exploiting a series of spin-echo refocusing pulses at various frequencies (νCPMG) (Palmer et al., 2001). For such processes, the effect of exchange on the NMR signal intensities can be modulated as a function of νCPMG that is subsequently reflected in changes in observed transverse relaxation rates (R_{2(eff)}). Analysis of R_{2(eff)} versus νCPMG provides quantitative information about the structural and kinetic properties of the excited state (Korzhnev and Kay, 2008).

Relaxation dispersion measurements of LARA^{sol} were obtained using three NMR field strengths (600, 700, and 800 MHz) (Figures 4A–4D), and the dispersion profiles were globally fit to a single exchange process (reduced χ² of 1.20, indicative of a reasonable fit), indicating local exchange of the native state with an excited state populated at 0.970% ± 0.003% and a k_{ex} of 630 ± 31 s^{−1}. The residues that exhibited significant dispersion (R_{ex} ≥ 3 Hz at 800 MHz) are labeled in Figure 4E, which shows that the majority of exchange with the excited state is localized primarily to the N-terminal loop region. The chemical-shift differences between the native and excited states (Δω) for residues in the N-terminal loop showed poor correlation with the chemical-shift differences between the native chemical shifts and those expected for random coil (R² of 0.14; Figure 4F). This indicates that the excited state of the N-terminal loop is not fully denatured but rather has a non-native but ordered structure.

Molecular Dynamics Simulations of LARA Domain

To gain further insight into the conformational changes that take place leading to amyloid formation by the LARA domain, we carried out molecular dynamics (MD) simulations at pH 4 and 7, and 300 K (26.85°C). The root-mean-square fluctuation (RMSF) for each residue was tracked and averaged across simulations (Figure 5A). Within each condition, higher RMSF values

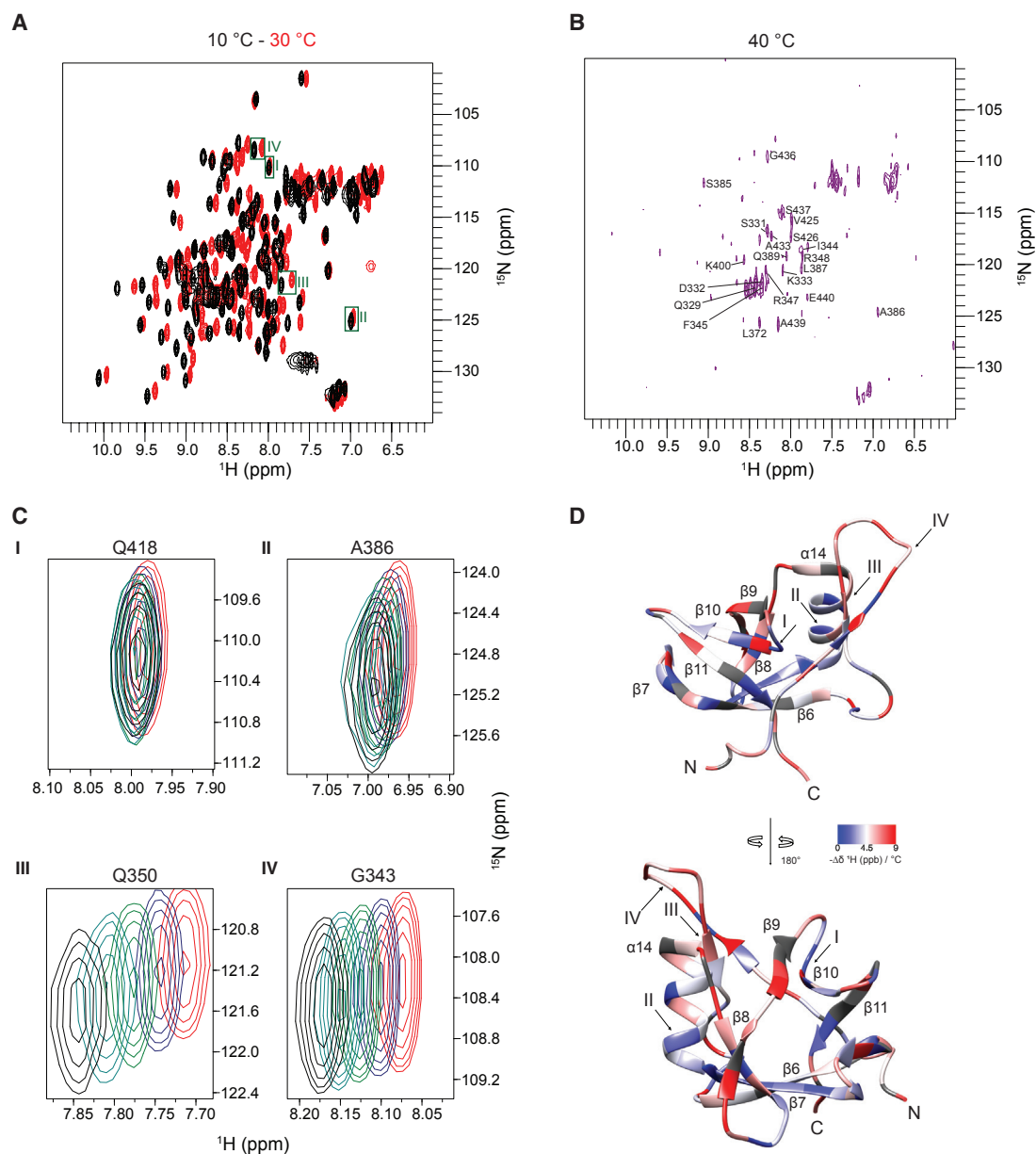


Figure 3. ^1H , ^{15}N HSQC Spectra of the Soluble LARA Domain at Different Temperatures

(A) ^1H , ^{15}N HSQC spectra of the LARA domain recorded at 10°C (black) and 30°C (red). Boxed peaks I and II show examples of residues with relatively low temperature coefficients, while III and IV are examples of residues having relatively high temperature coefficients.

(B) ^1H , ^{15}N HSQC spectrum of LARA^{flb} at 40°C. Backbone resonances that could be assigned at this temperature are labeled on the spectrum. See also Table S4.

(C) Examples of chemical-shift changes as a function of temperature for the residues identified in (A). Black is 10°C, turquoise is 15°C, green is 20°C, blue is 25°C, and red is 30°C.

(D) Temperature coefficients plotted on the structure of the protein with the locations of the residues highlighted in (A) and (C) indicated by roman numerals. Gray regions represent residues with unidentifiable coefficients due to inability to assign all the backbone resonances at all temperatures. The front (top) and back (bottom) faces of the LARA domain are shown, along with the corresponding color key (blue with low temperature coefficients and red with high temperature coefficients). See also Table S5.

were observed at the very N and C termini compared with the core, as expected. However, the heart-shaped region of the N-terminal loop (V338–Y351) exhibited significantly increased dynamics at pH 4 compared with pH 7 (Figure 5A). At pH 7, a major population was observed to form salt-bridge interactions (both direct hydrogen bonding and solvent mediated) between

the side chains of R340 and E413 (Figure 5B, top panel and Movie S1) or between R348 and E413 (Figure 5B, middle panel, and Movie S2). R340 and R348 were also observed to interchange salt-bridge interactions with E413 as shown in Movie S3, and were observed in some instances to interact simultaneously with E413 (Figure S5). Overall, these salt-bridge

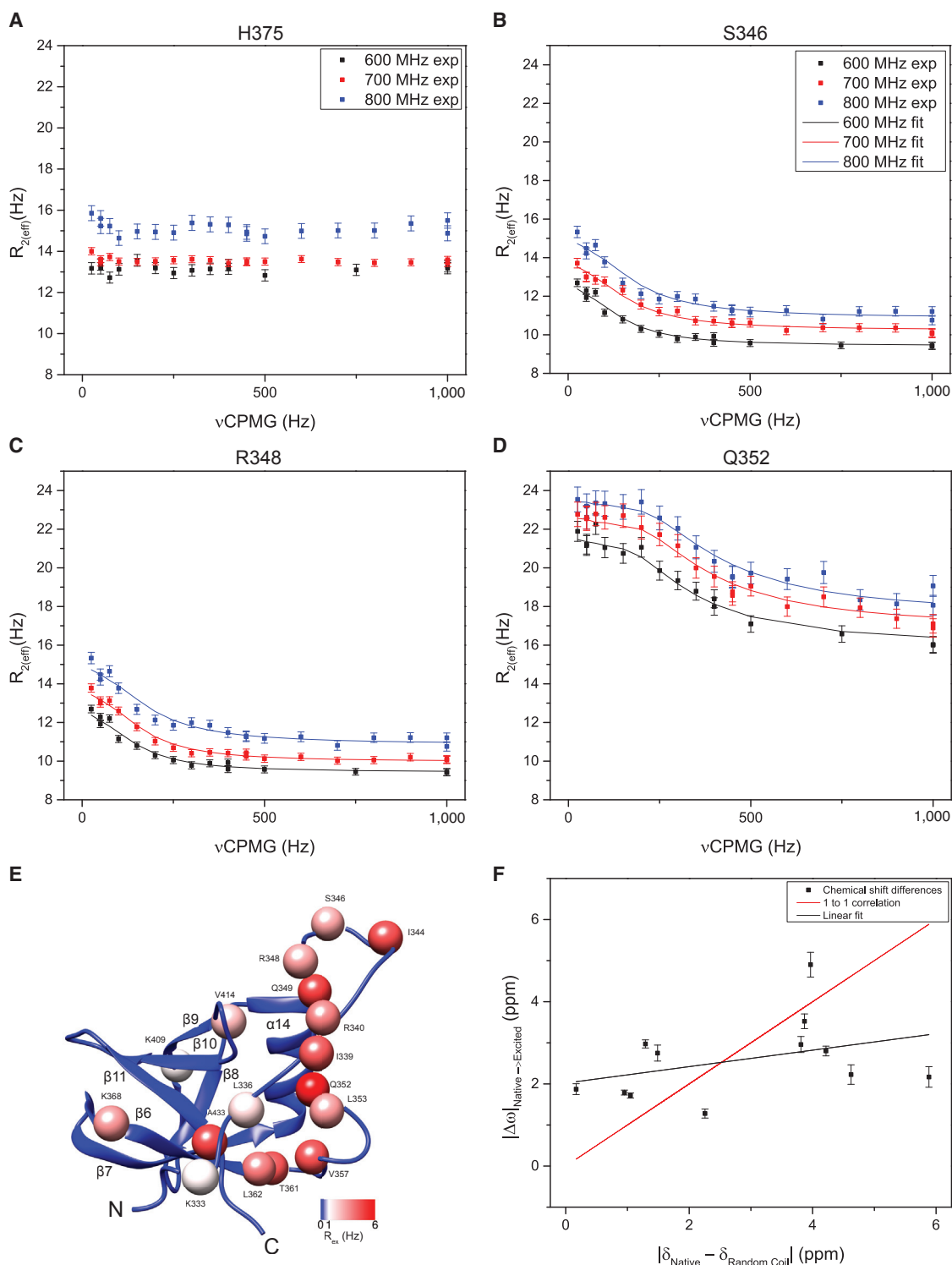


Figure 4. CPMG Relaxation Dispersion Analysis of LARA^{sol} at pH 4 and 24°C

(A–D) ^{15}N single-quantum $R_{2(eff)}$ as a function of CPMG frequency for several backbone amide resonances from the soluble LARA domain. CPMG-RD data were obtained at 600 MHz (black, lower), 700 MHz (red, middle), and 800 MHz (blue, top) 1H Larmor frequencies. (A) Shows a representative residue exhibiting no detectable relaxation dispersion, while (B–D) are examples of residues with significant relaxation dispersions due to chemical exchange. For each dataset showing significant relaxation dispersions due to chemical exchange (B–D), experimental data (exp) are shown along with global fitted values (fit) obtained using a two-state exchange between the ground and excited states. Errors bars shown are estimated errors in $R_{2(eff)}$ as described in the [Supplemental Experimental Procedures](#). Legend inset in (B) also applies to (C) and (D).

(legend continued on next page)

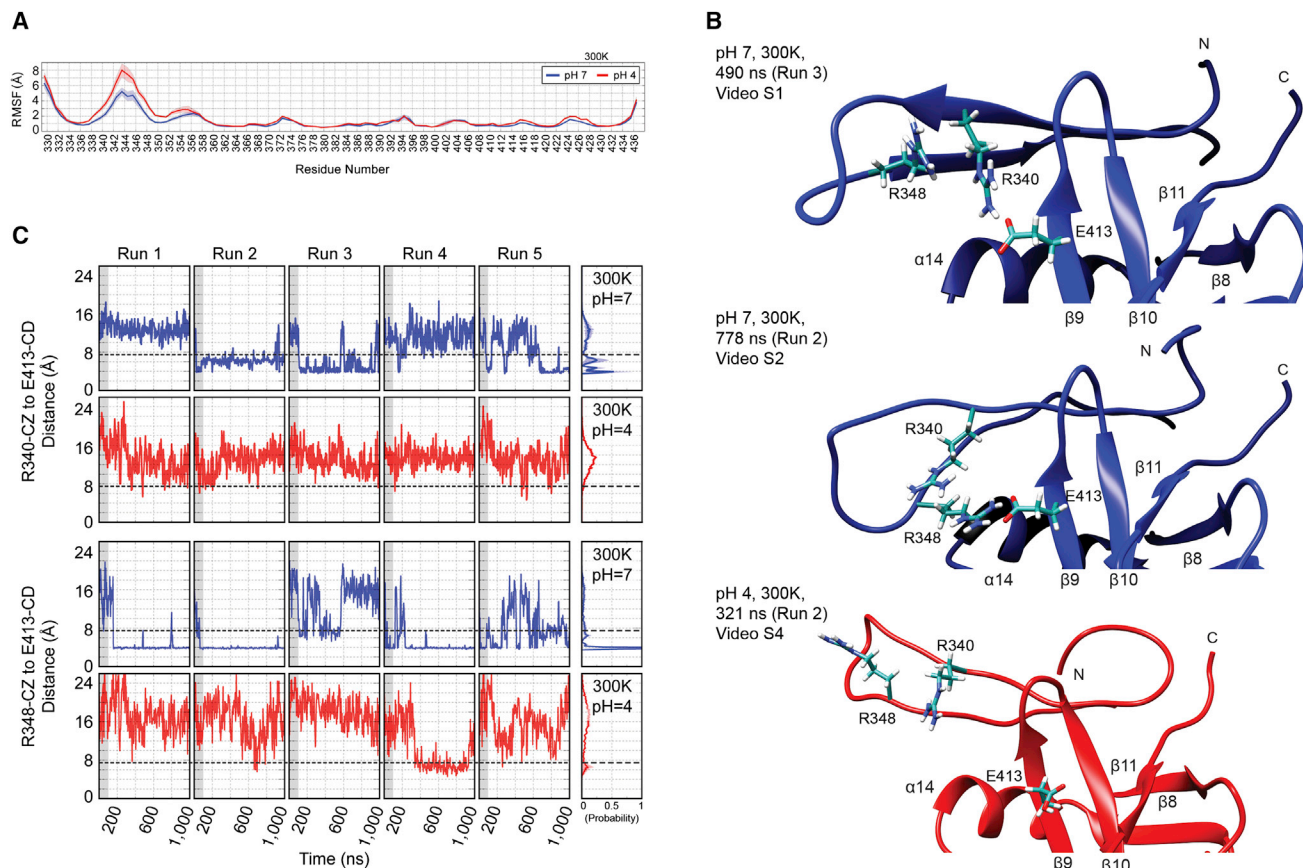


Figure 5. Molecular Dynamics Simulations of the LARA Domain

(A) Root-mean-square fluctuations (RMSF; Å) of residues of the LARA domain at pH 7 (blue) and pH 4 (red) at 300 K. Error bars are shaded and refer to the SEM ($n = 5$). See also Figure S5.

(B) The top panel shows an example of the salt bridge formed between R340 and E413 (at 490 ns in Movie S1, pH 7, run 3), while the middle panel shows the R348-E413 interaction (at 778 ns in Movie S2, pH 7, run 2). The bottom panel shows the loss of the salt bridges between R340-E413 and R348-E413 at pH 4 (at 321 ns in Movie S4, pH 4, run 2). See also Figure S5.

(C) Distances between the arginine C ζ and glutamic acid C δ for R340-E413 (top) and R348-E413 (bottom) through a series of 1,000-ns runs (runs 1–5) at pH 7 (blue) and pH 4 (red). The probability of distances across all runs is summed in the boxes to the right of run 5. The dashed line (7.5 Å) indicates the threshold of salt-bridge interactions, both solvent-mediated and direct contacts. The grayed-out regions (0–100 ns) were removed from the distance distribution calculation because the system may not be fully relaxed or equilibrated at these early time points.

interactions mediated loop closure over the β -rich core of the LARA domain. At pH 4, we observed disruption of these salt-bridge interactions due to protonation of E413 (Figure 5B, bottom panel, and Movie S4). Loss of these salt-bridge interactions at low pH likely gives rise to the increased dynamics of the heart-shaped region of the N-terminal loop in the simulations and as observed by NMR (Figures 3, 4, and S4). Note that the simulated protonation of the carboxylic acid moieties at low pH may lead to underestimation of the probability of salt-bridge formation compared with higher pH values.

If we define a Glu-Arg salt bridge as having a distance between the arginine ζ carbon and the glutamic acid δ carbon as <7.5 Å

(dotted line in Figure 5C), the probability of forming the salt bridge (both solvent-mediated and direct bonds) between R340 and E413 at pH 7 is $46\% \pm 19\%$ (Figure 5C, blue curves at top), while the probability of the R348-E413 salt bridge is $68\% \pm 15\%$ (Figure 5C, blue curves at bottom), giving a total salt-bridge formation probability of $86\% \pm 4\%$ (R340 and/or R348 bonded to E413). The probability of both R340 and R348 forming simultaneous salt bridges with E413 is $28\% \pm 17\%$. None of these salt-bridge interactions were observed for an entire 1- μ s trajectory; indeed, they were made and lost reversibly in four of the five simulations. In acidic conditions using the same criteria, the probability of forming similar salt bridges is $3\% \pm 8\%$

(E) R_{ex} values (600 MHz) derived from the global fit analysis are plotted onto the LARA domain structure with each sphere representing one residue.

(F) Shown is a comparison of the chemical-shift difference between the native and excited states ($|\Delta\omega|$) of the LARA domain N-terminal loop residues and the magnitude of the chemical-shift difference between native (experimental) and random-coil (predicted) states (Wishart et al., 1995). Data were fit to a straight line (black line), while the correlation line expected for a fully unfolded excited state is shown as red.

and $10\% \pm 9\%$ for E413-R340 and E413-R348, respectively (Figure 5C, red curves; total probability of $13\% \pm 9\%$). The formation of a combined salt bridge of R340 and R348 with E413 has a probability of $0.1\% \pm 0.1\%$ under these conditions. Note that a minor metastable interaction between R348 side chain and G394 backbone carbonyl oxygen was observed in run 4 at pH 4, which led to a locally trapped minimum (Figures 5C and S5C).

Mutations in the N-terminal Loop Region Enhance the Amyloid Formation Propensity of LARA

The HSQC and CPMG-RD experiments (Figures 3, 4, and S4) and MD results (Figure 5) strongly suggest that a conformational change in the N-terminal loop region is required for the formation of the amyloid state of the LARA domain. We speculated that the N-terminal loop must be displaced from the rest of the core LARA domain for amyloid formation to proceed. In this model, loop displacement exposes amyloidogenic regions, allowing for fibril formation through strand addition of LARA domain monomers to the exposed β -sheet edges in the core of the protein. To test this hypothesis, we generated a LARA domain construct, LARA(351–440), with a deletion of the N-terminal loop from residues 329–350, removing the Arg residues required for salt-bridge interactions with E413 (Figure 5).

CD spectra at 20°C and pH 4 of LARA(351–440) showed the presence of apparently increased β -sheet structure compared with the wild-type (WT) LARA domain at that same temperature and pH (Figure 6A). The altered shape of the CD spectrum might reflect a loss of random coil due to deletion of the N-terminal loop. Increasing the temperature at pH 4 resulted in even more β -sheet signal for LARA(351–440) (Figure 6A). LARA(351–440) also had a lower T_m compared with WT ($37.4^\circ\text{C} \pm 0.2^\circ\text{C}$ versus $42.3^\circ\text{C} \pm 0.1^\circ\text{C}$) as monitored by CD at pH 4, suggesting a lower stability of the truncated protein. To further confirm our results, we measured the aggregation propensity of WT LARA and LARA(351–440) using static light scattering (SLS) at 266 nm. As shown in Figure 6B, LARA(351–440) aggregated at a lower temperature than WT LARA with a T_{agg} (defined in Experimental Procedures) for LARA(351–440) of $30.5^\circ\text{C} \pm 0.5^\circ\text{C}$ relative to $40.4^\circ\text{C} \pm 0.7^\circ\text{C}$ for WT LARA. The higher starting SLS signal for LARA(351–440) may reflect a small population of pre-existing aggregates or fibrillar oligomers.

TEM analysis (Figure 6C) further confirmed the presence of worm-like fibrillar aggregates for LARA(351–440). Kinetic analysis of fibrillization showed that formation of LARA(351–440) amyloids was substantially faster than that of WT LARA, as monitored by ThT fluorescence at 33°C and pH 4 (Figure 6D and Table S2). This suggests that the barrier to amyloid formation is reduced for this mutant protein, consistent with our proposal that the N-terminal loop region plays a protective role in maintaining the native structure of the LARA domain.

FTIR analysis shows similarities between the estimated secondary structure of soluble LARA(351–440) and LARA^{fib} (Figure 6E). In addition, the peaks corresponding to β -sheet amyloid peaks ($1,620\text{ cm}^{-1}$ region) align well between LARA(351–440)^{sol} and LARA^{fib} (Figure 6E). Along with the estimated secondary structure content (Table S1), these data suggest that LARA(351–440) retains a folded structure and that soluble LARA(351–440) has a fold similar to that of LARA^{fib}. Furthermore, conversion of LARA(351–440)^{sol} to

the amyloid state resulted in loss of α -helical content (Table S1). Moreover, LARA(351–440)^{fib} has a smaller β -sheet peak at $1,634\text{ cm}^{-1}$ and an increased amyloid β -sheet signal at $1,618\text{ cm}^{-1}$ compared with LARA(351–440)^{sol}, suggesting extensive β -sheet rearrangements upon fibrillization by this truncation mutant.

Mutation of the positively charged R340, R347, and R348 to glutamate was carried out in the full-length LARA domain to further investigate the role of the salt-bridge interactions. The R347D mutation was made to prevent any compensatory salt-bridge interactions with E413. The triple mutant exhibited significantly faster amyloid formation kinetics relative to WT LARA (Figure 7A). The soluble and amyloid forms of the triple mutant and the WT LARA had generally similar FTIR spectra (Figure 7B, Table S1). Furthermore, TEM images of the LARA(R340D,R347D,R348D) amyloids showed fibrillar morphology similar to that of WT LARA^{fib} (Figure 7C).

Finally, a thermal melt was conducted using nanoDSF (see Experimental Procedures) to measure the thermal stability of the different LARA domain mutants (Figure 7D). The ratio of integrated fluorescence emission intensity at 350 and 330 nm (F_{350}/F_{330}) was monitored as a function of temperature. The T_m (defined in Experimental Procedures) was $42.6^\circ\text{C} \pm 0.04^\circ\text{C}$ for WT LARA, $32.4^\circ\text{C} \pm 0.2^\circ\text{C}$ for LARA(351–440), and $27.4^\circ\text{C} \pm 0.2^\circ\text{C}$ for LARA(R340D,R347D,R348D). These trends show that both LARA mutants have lower thermal stability than WT LARA. The results of these experiments further support our proposal that salt bridges formed between R340/R348 and E413 stabilize LARA^{sol} and reduce its rate of conversion to an amyloid state.

DISCUSSION

While many amyloids are associated with neurodegenerative diseases, the amyloid states of proteins have also been found to perform functional roles in many different organisms (Eichner and Radford, 2011). Current structural data suggest that both functional and pathological amyloid fibrils share a similar cross- β architecture (Greenwald and Riek, 2010), yet the precise mechanisms by which proteins assemble into these ordered aggregates remain largely unknown (Goldschmidt et al., 2010). This study provides the first report of the formation of an amyloid state by the cytoplasmic *E. coli* AAA+ protein, RavA, and its LARA domain. Several bacterial cytoplasmic proteins are now known to form amyloid states including the GroES cochaperone from *E. coli* (Higurashi et al., 2005), the Rho termination factor from *Clostridium botulinum* (Pallares et al., 2015), as well as an artificial RepA-WH1 fusion construct used to model proteinopathy in bacteria (Torreira et al., 2015).

Both RavA and the LARA domain formed misfolded aggregates with typical amyloid characteristics, including a β -sheet-dominant secondary structure, binding to ThT, sigmoidal aggregation kinetics, protease resistance, and distinct fibrillar morphology (Eichner and Radford, 2011; Knowles et al., 2014; Nilsson, 2004). Conversion into the amyloid state by RavA was accompanied by a decrease in α -helical content, indicating a loss of structure in either the triple-helical bundle and/or the AAA+ module (Figure 1B). For the LARA domain, conversion to the amyloid state was similar and accompanied by formation of extended β -sheet structures, as

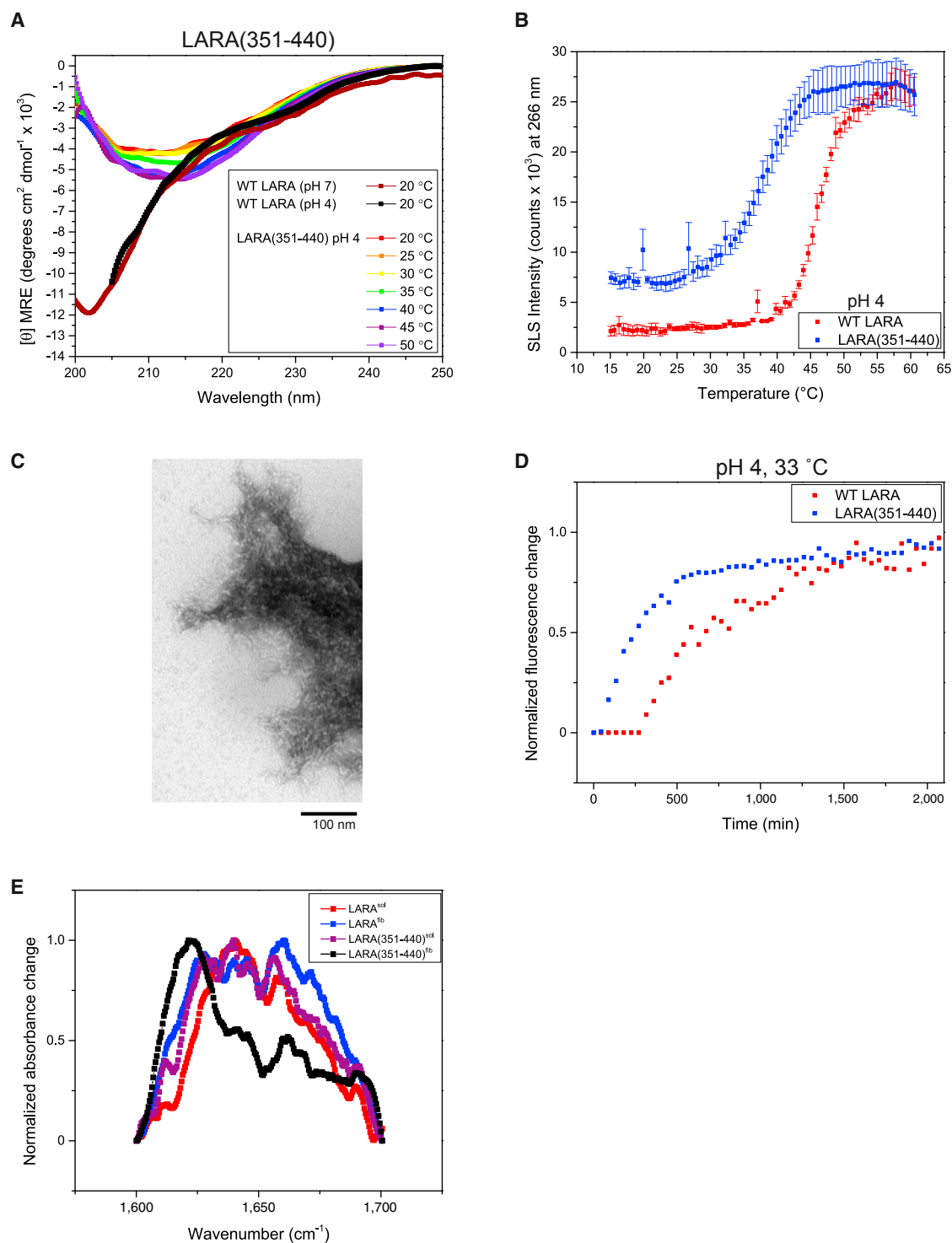


Figure 6. Amyloid Formation by LARA(351-440)

(A) CD spectra of LARA(351-440) recorded at pH 4 as a function of temperature. WT LARA^{sol} (20 °C) at pH 4 and pH 7 are shown as reference.

(B) Static light-scattering intensity at 266 nm for WT LARA (red) and the truncation mutant LARA(351-440) (blue) as a function of temperature. Error bars represent SEM (n = 3).

(C) TEM images of negatively stained LARA(351-440) fibrils viewed at 150,000 \times magnification.

(D) Representative curves of ThT fluorescence emission monitored over time for samples containing WT LARA (red) or LARA(351-440) (blue). Sample shown contained 15 μM protein at pH 4 and 33 °C. See also [Table S2](#).

(E) The Amide I region of FTIR spectra of LARA^{sol} (red), LARA^{fib} (blue), LARA(351-440)^{sol} (purple), and LARA(351-440)^{fib} (black). See also [Table S1](#).

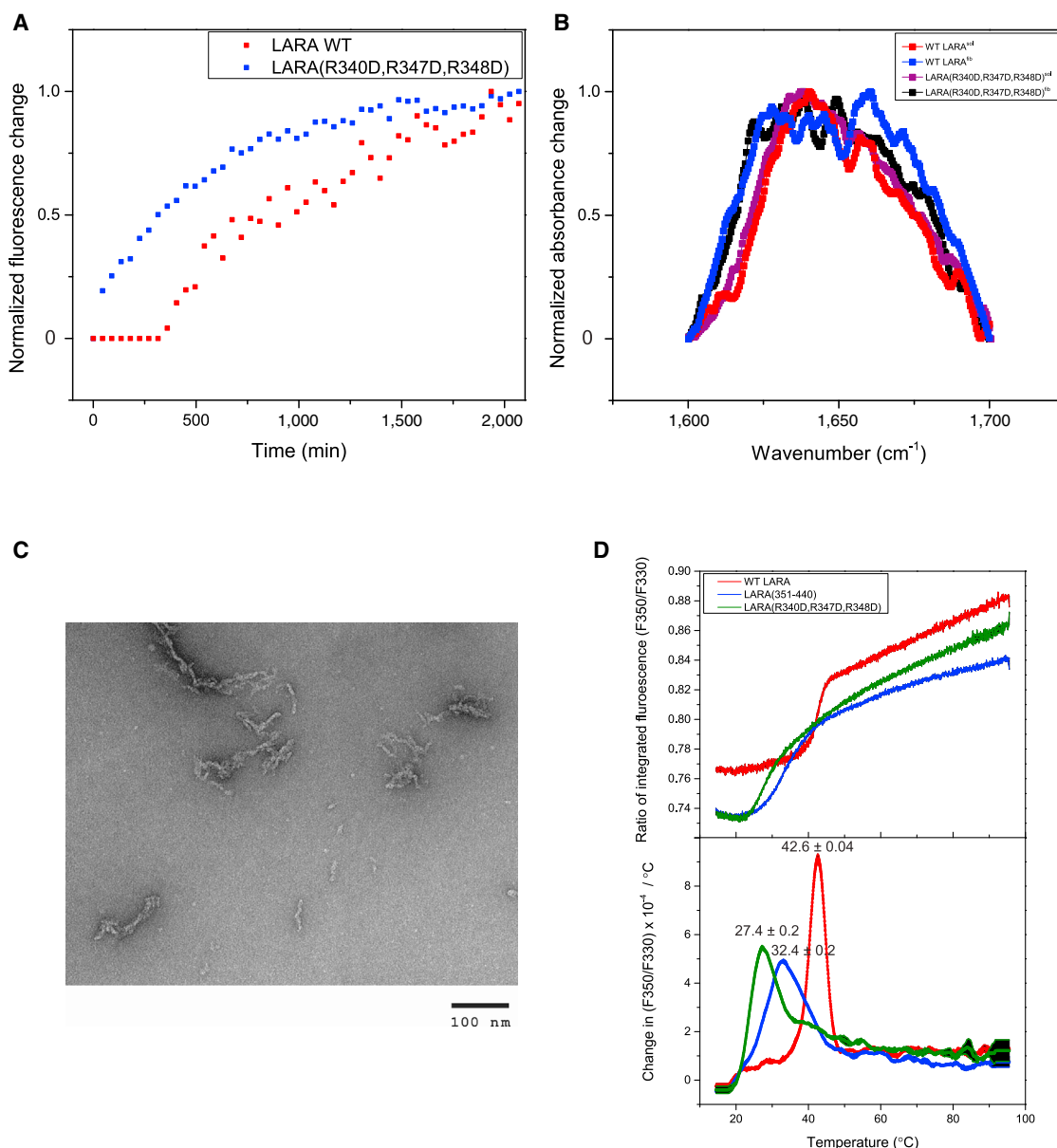


Figure 7. Amyloid Formation by LARA(R340D,R347D,R348D)

(A) Representative kinetic curves for amyloid formation monitored by ThT fluorescence emission are shown for LARA(R340D,R347D,R348D) in blue and for the WT LARA domain in red. Protein concentration was 10 μ M at pH 4 and 33°C. See also Table S2.

(B) Amide I region from the FTIR spectra of LARA^{sol} (red), LARA^{fib} (blue), LARA(R340D,R347D,R348D)^{sol} (purple), and LARA(R340D,R347D,R348D)^{fib} (black). See also Table S1.

(C) TEM images of negatively stained LARA(R340D,R347D,R348D) fibrils at 120,000 \times magnification.

(D) nanoDSF analysis of a temperature melt of WT LARA (red), LARA(351–440) (blue), and LARA(R340D,R347D,R348D) (green). The black shading shows SEM (n = 3). The top panel is the F_{350}/F_{330} ratio and the bottom panel is the first derivative of that curve with the T_m for each LARA domain construct indicated.

suggested by the shift of β -sheet FTIR peaks to lower wavenumbers (Figure 2C).

Differences in fibrillar morphology have been suggested to result from different assembly pathways (Gosal et al., 2005; Miti et al., 2015). Variations in monomer conformation or in the intermolecular interfaces adopted during fibrillization may lead to the formation of different superstructures in mature amyloid fibrils. Both RavA and LARA fibrillar morphologies were twisted

and worm-like with extensive self-association between fibrils, potentially resulting from similar mechanisms of assembly. The ubiquitous nature of these twists and turns may be due to structural defects in the fibril backbone that are compensated for by the curvature. Similar worm-like fibrils have been observed for the mouse prion protein under low pH conditions (Jain and Udgaonkar, 2011), and for β -lactoglobulin (vandenAkker et al., 2011). In the latter case, worm-like fibrils were found to have

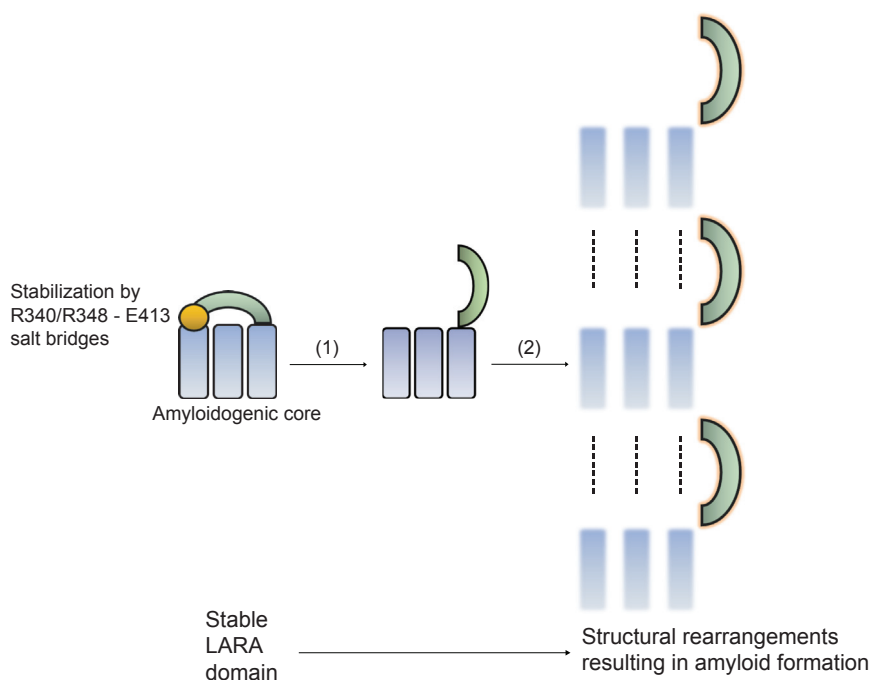


Figure 8. Model of the Mechanism of Amyloid Formation by the LARA Domain

On the left, the folded core of LARA (blue rectangles) is protected from aggregation by the N-terminal loop (green), which is held in position by the R340/R348-E413 salt-bridge interactions (orange circle). In step (1), acidic conditions disrupt the salt bridges, exposing the β -sheet edges. In step (2), partial destabilization of the core of the LARA domain and the N-terminal loop by increased temperature induces further structural rearrangements that promote non-native intermolecular contacts (dashed lines) and the formation of a nucleus that initiates the fibrillization process.

significantly lower β -sheet content relative to straight fibrils, suggesting differences in internal structure and assembly (van den Akker et al., 2011). A similar worm-like fibril morphology has been reported for amyloid fibrils formed by an SH3 domain mutant which mimics a folding intermediate (Neudecker et al., 2012). In this SH3 mutant, deletion of a protective β strand allowed exposure of aggregation-prone regions and promoted amyloid formation in a manner similar to protein misfolding. These observations suggest that exposure of the amyloidogenic-prone regions in the LARA domain could result in rapid nucleation events that parallel the aggregation process of a destabilized protein misfolding intermediate.

From the NMR thermal melts (Figure 3), CPMG-RD analysis (Figure 4), and MD data (Figure 5), the N-terminal loop region of the LARA domain is implicated in the amyloid formation process. MD showed a distinct mechanism for stabilization of the native fold through salt-bridge interactions of R340 and R348 with E413 that bring the heart-shaped loop region over the exposed β strands. This loop interaction is disrupted under conditions of low pH due to increased protonation of glutamic acid side chains. Deletion of the loop region, or mutagenesis of salt-bridge-forming residues, resulted in a significantly increased rate of in vitro amyloid fibril formation (Figures 6B, 6D, and 7A). Hence, we propose that the N-terminal loop inhibits amyloid formation by the LARA domain of RavA, and, therefore, plays a protective role in stabilizing the soluble native form of the protein. Since this region of the protein is relatively flexible under conditions of low pH and elevated temperature, the loop is able to access non-native conformations and lower the barrier to amyloid formation.

Another interesting observation is that residues of the $\beta 7$ strand (M375–E383) have the smallest, in absolute value, ^1H temperature coefficients ($\Delta\delta/^\circ\text{C}$ of -2.9 ppb, Table S5 and Figure 3D). This β strand is located within the core of the protein;

the surrounding core structures ($\alpha 14$, $\beta 6$, and $\beta 11$) are also relatively temperature insensitive (Figure 3 and Table S5). CPMG-RD analysis (Figure 4) showed that these core structural elements display almost undetectable relaxation dispersions under the mild conditions that destabilized the N-terminal loop. Based on multiple amyloid prediction algorithms (Figure S1), this β -sheet-rich core of the LARA domain, especially $\beta 6$ and $\beta 7$, contains highly amyloidogenic segments. Furthermore, the LARA domain construct lacking the N-terminal loop rapidly forms amyloid fibrils and displays a secondary structure similar to that of LARA^{fib} (Figure 6). Combined with the similarity in the proteolysis pattern between LARA^{sol} and LARA^{fib} (Figure S2C), the lack of a detectable global unfolding event from solution NMR (Figure 3A), and similarity in flexible residues after oligomer formation (Figure 3B), we propose that the core of the LARA amyloid fibrils retains a structure close to that of LARA^{sol}. The combination of low pH and high temperature is sufficient to destabilize the N-terminal loop, expose amyloidogenic regions, and slightly destabilize the packing of the β -sheet region to allow formation of a native-like nucleus that initiates amyloid formation (Figure 8).

The LARA domain is required for the association of RavA with Ldcl to form the RavA-Ldcl 3 MDa cage-like complex (Malet et al., 2014; Snider et al., 2006), in which the N-terminal loop of the LARA domain forms specific intermolecular contacts essential for RavA-Ldcl interactions (Malet et al., 2014) (Figure S6). It is interesting to note that Ldcl is induced at low pH and forms the RavA-Ldcl cage structure under these acidic conditions. While previous experiments suggest that the RavA, ViaA, and Ldcl might be required to modulate the assembly of respiratory complexes in *E. coli* (Erhardt et al., 2012; Wong et al., 2014), another function of the cage structure might be to stabilize the LARA domain of RavA and prevent the formation of RavA amyloids when the bacteria encounter low-pH and high-temperature conditions.

In conclusion, our study describes a new amyloid-forming protein and adds to the current list of amyloids formed by bacterial proteins. It provides critical insights into the mechanism of amyloidogenesis by the LARA domain of RavA. The physiological roles played by the RavA amyloids in the context of bacterial viability and pathogenesis remain to be elucidated.

However, the pH sensitivity of amyloid formation by a protein involved in the cellular acid stress response pathway raises intriguing functional possibilities for the role of such amyloids in pathogenesis.

EXPERIMENTAL PROCEDURES

Protein expression and purification and all biophysical and biochemical experiments are further described in detail in [Supplemental Experimental Procedures](#).

Protein Purification

Expression and purification of proteins were conducted according to standard protocols described previously ([El Bakkouri et al., 2010](#)). RavA and LARA were purified using nickel-nitrilotriacetic acid-agarose chromatography. Elutions were subsequently further purified using cation-exchange and flash-frozen with liquid nitrogen. For isotopic labeling, cells were induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside for 4 hr at 30°C in M9 Minimal Medium supplemented with 1 g/l $^{15}\text{NH}_4\text{Cl}$ and 2 g/l [^{13}C]glucose.

Amyloid Fibril Preparation

LARA was dialyzed into pH 4 buffer and RavA was buffer exchanged into pH 2.5 buffer. RavA and LARA were then converted into amyloids by heating from 20°C to 80°C at 1°C/min.

ThT Fluorescence Assay

ThT fluorescence change was measured with excitation set at 442 nm and emission set at 480 nm for kinetic experiments at a constant temperature of 33°C to initiate fibrillization. Fluorescence emission spectra were scanned from 462 to 600 nm to assess ThT binding. A concentration of 20 μM ThT was typically used in these experiments.

Transmission Electron Microscopy

RavA or LARA domain fibrils were prepared using general procedures outlined in [Walsh et al. \(2009\)](#). Samples were adsorbed onto continuous carbon films from copper rhodium grids (Electron Microscopy Sciences) and glow-discharged for 15 s at 30 mA negative discharge. TEM images were obtained using a JEOL 1011 microscope operating at 80 keV.

CD Spectroscopy

The secondary structure of full-length RavA and the isolated LARA domain were analyzed using a JASCO J-810 spectropolarimeter with a 1.0-mm path-length quartz cuvette. The temperature was raised at a rate of 1°C/min and spectra were typically collected at every 5°C (200–260 nm range) with a scanning speed of 50 nm/min.

Attenuated Total Reflectance (ATR)-FTIR Spectroscopy

A Nicolet Nexus 6700 FTIR spectrometer (Thermo Fisher Scientific) coupled with a Smart Orbit ATR unit was used to measure the infrared spectra. For each spectrum, 512 scans were collected with a resolution of 1 cm^{-1} over a scan range of 4,000–525 cm^{-1} using Omnic software. In the text, ATR-FTIR is referred to as FTIR.

NMR Spectroscopy

Bruker Avance III 600-MHz and 700-MHz spectrometers were used to collect data for all experiments using a 5-mm TXI triple-resonance probe equipped with z-gradient pulse-field gradient capabilities. Relaxation dispersion experiments were also recorded at 800 MHz on a Varian Unity Inova four-channel spectrometer using a 5-mm, triple-resonance, three-axis pulse-field gradient probe. Thermal melt of soluble ^{13}C , ^{15}N -labeled LARA was monitored using 2D ^1H , ^{15}N HSQC spectra. Data were acquired every 5°C for 7 min between 10°C and 30°C and every 2°C for 5 min between 30°C and 40°C. 4,4-dimethyl-4-silapentane-1-sulfonic acid was used as an external reference at each temperature.

^{15}N CPMG relaxation dispersion profiles of ^{15}N -labeled LARA domain were measured with a ^1H -decoupled ^{15}N CPMG-HSQC pulse sequence ([Tollinger et al., 2001](#)) at 24°C. CPMG refocusing pulse frequencies (vCPMG) between

0 and 1,000 Hz were employed over a constant CPMG time-relaxation interval of 40 ms. Relaxation dispersion profiles of LARA^{sol} were recorded at three fields (600, 700, and 800 MHz ^1H frequency).

MD Simulations

The LARA domain of RavA was used for MD simulations based on the PDB: 3NBX crystal structure. The protein was solvated in a rectangular cell for pH 7 simulations and in a rhombic dodecahedron cell for pH 4 simulations. Na^+ and Cl^- ions were added to each simulation box to obtain a net charge of zero and approximately 150 mM excess salt. All simulations were performed using GROMACS ([Pronk et al., 2013](#)). The optimized potentials for liquid simulations all-atom force field ([Jorgensen et al., 1996](#); [Kaminski et al., 2001](#)) were used to model the protein and ions. The transferable intermolecular potential 3-point model ([Jorgensen et al., 1983](#)) was used for water molecules.

ACCESSION NUMBERS

Resonance assignments have been deposited in the Biological Magnetic Resonance Bank database under accession number BMRB: 26735.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, five tables, and four movies and can be found with this article online at <http://dx.doi.org/10.1016/j.str.2016.05.002>.

AUTHOR CONTRIBUTIONS

S.W.S.C., K.L., S.S., and W.A.H. were involved in the study conception and experimental design. S.W.S.C., K.L., V.B., T.V.S., and N.K.Y. prepared samples and conducted the biophysical and biochemical experiments. J.Y. was responsible for the NMR spectroscopy data acquisition and along with P.F. and L.E.K. carried out the NMR data analysis. C.I., N.C., and R.P. were responsible for collecting and analyzing the MD data. A.W. was responsible for analyzing and collecting the ATR-FTIR data. C.M.Y., R.P., S.S., and W.A.H. helped provide experimental input. S.W.S.C., S.S., and W.A.H. were primarily responsible for drafting the manuscript.

ACKNOWLEDGEMENTS

We thank Mr. Wyatt Strutz, Senior Application Specialist for NanoTemper Technologies Inc., for his help with the nanoDSF experiments. S.W.S.C. was supported by the Undergraduate Research Opportunity Award from the University of Toronto. K.L. was supported by an Ontario Graduate Scholarship. T.V.S. was supported by CNPq-Brazil postdoctoral fellowship (202192/2015-6). S.S. holds a Canada Research Chair (Tier II) and L.E.K. holds a Canada Research Chair (Tier I). This work was supported by the Canadian Institutes of Health Research Operating Grant (MOP-130374) to W.A.H., and Natural Sciences and Engineering Research Council of Canada discovery grants to C.Y. (RGPIN-2015-043), S.S., and L.E.K. Supercomputer usage was performed on Mammouth parallèle 2 at Université de Sherbrooke, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), Ministère de l'Économie, de l'Innovation et des Exportations du Québec (MEIE), RMGA, and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT).

Received: February 13, 2016

Revised: April 27, 2016

Accepted: May 1, 2016

Published: June 2, 2016

REFERENCES

- Barnhart, M.M., and Chapman, M.R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.* 60, 131–147.
- Baxter, N.J., and Williamson, M.P. (1997). Temperature dependence of ^1H chemical shifts in proteins. *J. Biomol. NMR* 9, 359–369.

- Biancalana, M., and Koide, S. (2010). Molecular mechanism of thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta* 1804, 1405–1412.
- Cierpicki, T., and Otlewski, J. (2001). Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* 21, 249–261.
- Cordier, F., and Grzesiek, S. (2002). Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *J. Mol. Biol.* 317, 739–752.
- d'Auvergne, E.J., and Gooley, P.R. (2008a). Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* 40, 107–119.
- d'Auvergne, E.J., and Gooley, P.R. (2008b). Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR* 40, 121–133.
- Eichner, T., and Radford, S.E. (2011). A diversity of assembly mechanisms of a generic amyloid fold. *Mol. Cell* 43, 8–18.
- Eisenberg, D., and Jucker, M. (2012). The amyloid state of proteins in human diseases. *Cell* 148, 1188–1203.
- El Bakkouri, M., Gutsche, I., Kanjee, U., Zhao, B., Yu, M., Goret, G., Schoehn, G., Burmeister, W.P., and Houry, W.A. (2010). Structure of RuvA MoxR AAA+ protein reveals the design principles of a molecular cage modulating the inducible lysine decarboxylase activity. *Proc. Natl. Acad. Sci. USA* 107, 22499–22504.
- Emily, M., Talvas, A., and Delamarche, C. (2013). MetAmyl: a METa-predictor for AMYloid proteins. *PLoS One* 8, e79722.
- Erhardt, H., Steimle, S., Muters, V., Pohl, T., Walter, J., and Friedrich, T. (2012). Disruption of individual nuo-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in *Escherichia coli*. *Biochim. Biophys. Acta* 1817, 863–871.
- Esteras-Chopo, A., Serrano, L., and Lopez de la Paz, M. (2005). The amyloid stretch hypothesis: recruiting proteins toward the dark side. *Proc. Natl. Acad. Sci. USA* 102, 16672–16677.
- Farber, P., Darmawan, H., Sprules, T., and Mittermaier, A. (2010). Analyzing protein folding cooperativity by differential scanning calorimetry and NMR spectroscopy. *J. Am. Chem. Soc.* 132, 6214–6222.
- Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010). Identifying the amyloid, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* 107, 3487–3492.
- Gosal, W.S., Morten, I.J., Hewitt, E.W., Smith, D.A., Thomson, N.H., and Radford, S.E. (2005). Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid. *J. Mol. Biol.* 351, 850–864.
- Greenwald, J., and Riek, R. (2010). Biology of amyloid: structure, function, and regulation. *Structure* 18, 1244–1260.
- Guo, Y., and Wang, J. (2012). Spectroscopic evidence for polymorphic aggregates formed by amyloid-beta fragments. *Chemphyschem* 13, 3901–3908.
- Han, B., Liu, Y., Gininger, S.W., and Wishart, D.S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 50, 43–57.
- Higurashi, T., Yagi, H., Mizobata, T., and Kawata, Y. (2005). Amyloid-like fibril formation of co-chaperonin GroES: nucleation and extension prefer different degrees of molecular compactness. *J. Mol. Biol.* 351, 1057–1069.
- Hong, J., Jing, Q., and Yao, L. (2013). The protein amide (1)H(N) chemical shift temperature coefficient reflects thermal expansion of the N-H...O=C hydrogen bond. *J. Biomol. NMR* 55, 71–78.
- Jain, S., and Udgaonkar, J.B. (2011). Defining the pathway of worm-like amyloid fibril formation by the mouse prion protein by delineation of the productive and unproductive oligomerization reactions. *Biochemistry* 50, 1153–1161.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- Jorgensen, W.L., Maxwell, D.S., and TiradoRives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236.
- Kaminski, G.A., Friesner, R.A., Tirado-Rives, J., and Jorgensen, W.L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105, 6474–6487.
- Kanjee, U., Gutsche, I., Alexopoulos, E., Zhao, B., El Bakkouri, M., Thibault, G., Liu, K., Ramachandran, S., Snider, J., Pai, E.F., et al. (2011). Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *EMBO J.* 30, 931–944.
- Knowles, T.P., Vendruscolo, M., and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15, 384–396.
- Korzhnev, D.M., and Kay, L.E. (2008). Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* 41, 442–451.
- Malet, H., Liu, K., El Bakkouri, M., Chan, S.W., Effant, G., Bacia, M., Houry, W.A., and Gutsche, I. (2014). Assembly principles of a unique cage formed by hexameric and decameric *E. coli* proteins. *Elife* 3, e03653.
- Marcoleta, A., Marin, M., Mercado, G., Valpuesta, J.M., Monasterio, O., and Lagos, R. (2013). Microcin e492 amyloid formation is retarded by posttranslational modification. *J. Bacteriol.* 195, 3995–4004.
- Micsonai, A., Wien, F., Kernya, L., Lee, Y.H., Goto, Y., Refregiers, M., and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* 112, E3095–E3103.
- Miti, T., Mulaj, M., Schmit, J.D., and Muschol, M. (2015). Stable, metastable, and kinetically trapped amyloid aggregate phases. *Biomacromolecules* 16, 326–335.
- Monsellier, E., and Chiti, F. (2007). Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* 8, 737–742.
- Moran, S.D., and Zanni, M.T. (2014). How to get insight into amyloid structure and formation from infrared spectroscopy. *J. Phys. Chem. Lett.* 5, 1984–1993.
- Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundstrom, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., and Kay, L.E. (2012). Structure of an intermediate state in protein folding and aggregation. *Science* 336, 362–366.
- Nilsson, M.R. (2004). Techniques to study amyloid fibril formation in vitro. *Methods* 34, 151–160.
- Pallares, I., Iglesias, V., and Ventura, S. (2015). The Rho termination factor of *Clostridium botulinum* contains a prion-like domain with a highly amyloido-genic core. *Front. Microbiol.* 6, 1516.
- Palmer, A.G., 3rd, Kroenke, C.D., and Loria, J.P. (2001). Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339, 204–238.
- Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854.
- Sarroukh, R., Goormaghtigh, E., Ruyschaert, J.M., and Raussens, V. (2013). ATR-FTIR: a “rejuvenated” tool to investigate amyloid proteins. *Biochim. Biophys. Acta* 1828, 2328–2338.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.
- Snider, J., Gutsche, I., Lin, M., Baby, S., Cox, B., Butland, G., Greenblatt, J., Emili, A., and Houry, W.A. (2006). Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *J. Biol. Chem.* 281, 1532–1546.
- Tollinger, M., Skrynnikov, N.R., Mulder, F.A., Forman-Kay, J.D., and Kay, L.E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* 123, 11341–11352.

- Torreira, E., Moreno-Del Alamo, M., Fuentes-Perez, M.E., Fernandez, C., Martin-Benito, J., Moreno-Herrero, F., Giraldo, R., and Llorca, O. (2015). Amyloidogenesis of bacterial prionoid RepA-WH1 recapitulates dimer to monomer transitions of RepA in DNA replication initiation. *Structure* 23, 183–189.
- Toyama, B.H., and Weissman, J.S. (2011). Amyloid structure: conformational diversity and consequences. *Annu. Rev. Biochem.* 80, 557–585.
- Tsolis, A.C., Papandreou, N.C., Iconomidou, V.A., and Hamodrakas, S.J. (2013). A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One* 8, e54175.
- vandenAkker, C.C., Engel, M.F., Velikov, K.P., Bonn, M., and Koenderink, G.H. (2011). Morphology and persistence length of amyloid fibrils are correlated to peptide molecular structure. *J. Am. Chem. Soc.* 133, 18030–18033.
- Walsh, P., Simonetti, K., and Sharpe, S. (2009). Core structure of amyloid fibrils formed by residues 106–126 of the human prion protein. *Structure* 17, 417–426.
- Walsh, I., Seno, F., Tosatto, S.C., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* 42, W301–W307.
- Wishart, D., and Sykes, B. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* 4, 171–180.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995). ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 5, 67–81.
- Wong, K.S., Snider, J.D., Graham, C., Greenblatt, J.F., Emili, A., Babu, M., and Houry, W.A. (2014). The MoxR ATPase RavA and its cofactor ViaA interact with the NADH:ubiquinone oxidoreductase I in *Escherichia coli*. *PLoS One* 9, e85529.
- Zou, Y., Li, Y., Hao, W., Hu, X., and Ma, G. (2013). Parallel beta-sheet fibril and antiparallel beta-sheet oligomer: new insights into amyloid formation of hen egg white lysozyme under heat and acidic condition from FTIR spectroscopy. *J. Phys. Chem. B* 117, 4003–4013.

Structure, Volume 24

Supplemental Information

Mechanism of Amyloidogenesis

of a Bacterial AAA+ Chaperone

Sze Wah Samuel Chan, Jason Yau, Christopher Ing, Kaiyin Liu, Patrick Farber, Amy Won, Vaibhav Bhandari, Nareg Kara-Yacoubian, Thiago V. Seraphim, Nilmadhab Chakrabarti, Lewis E. Kay, Christopher M. Yip, Régis Pomès, Simon Sharpe, and Walid A. Houry

Figure S1. Related to Figure 1. Prediction of amyloidogenic regions in RavA.

(A) Shown is the amino acid sequence of RavA colored based on domains (same color scheme used in Figure 1). Highlighted in orange boxes are the regions predicted to be amyloidogenic by all three methods used: MetAMYL (Emily et al., 2013), PASTA2.0 (Walsh et al., 2014) and AMYLPRED2 (Tsolis et al., 2013). Purple boxes indicate regions predicted to be amyloidogenic by two of these methods. Secondary structure elements and residue numbers are placed above the sequence.

(B) The RavA protomer structure is shown with the different amyloidogenic regions indicated in (A) highlighted. Secondary structure elements are labeled for the LARA domain.

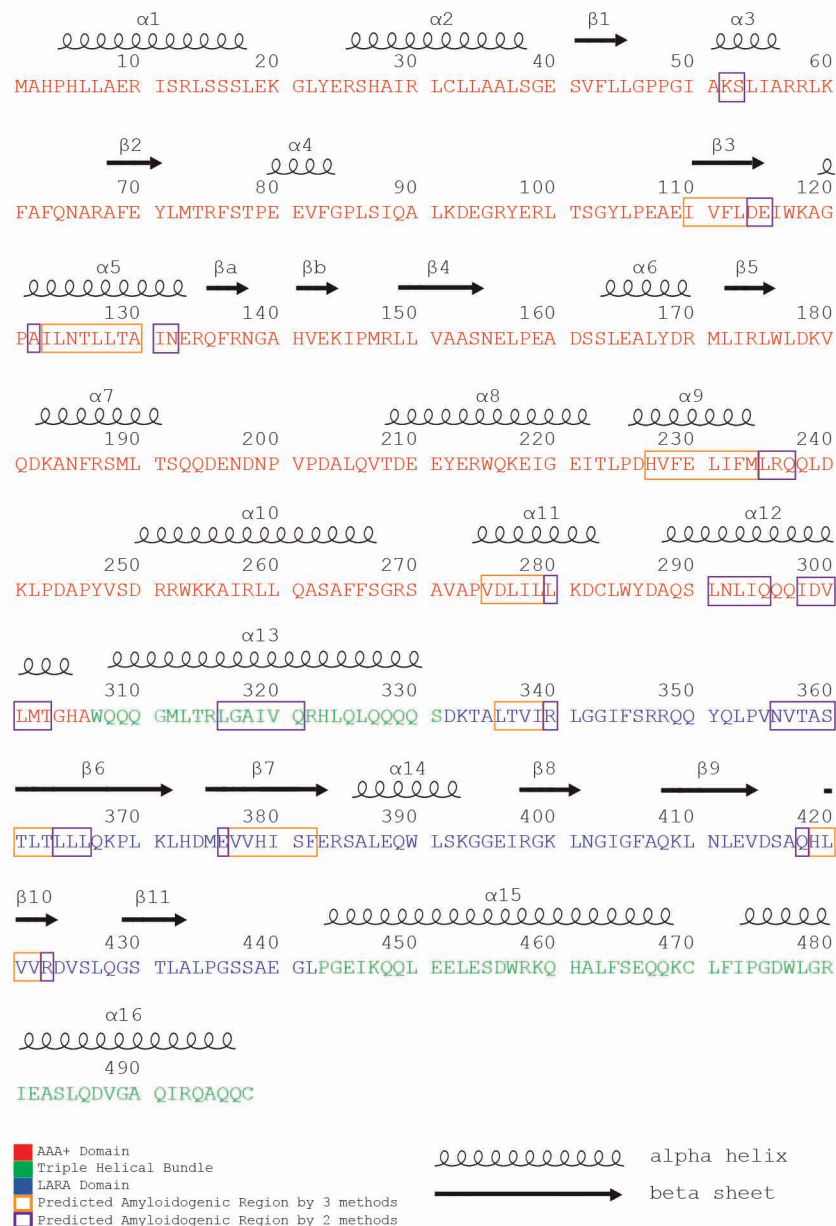
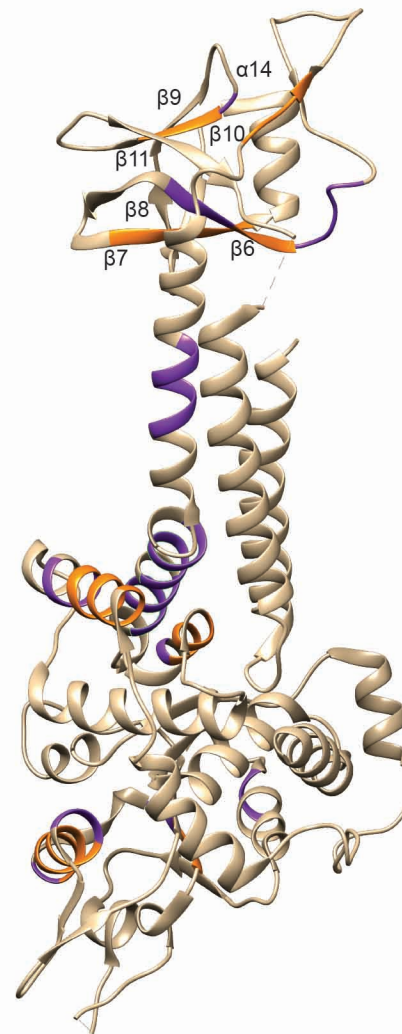
A**B**

Figure S2. Related to Figure 2. Characterization of the LARA domain amyloid state.

(A) CD spectra of the LARA domain at 20 °C, 50 °C, 80 °C in pH 3 (green), pH 4 (blue), pH 5 (light magenta), and pH 6 (orange) buffers.

(B) TEM images of negatively stained LARA^{fib} prepared in 150, 50 and 10 mM KCl viewed at 150,000x, 120,000x and 150,000x magnification, respectively.

(C) Time course of proteinase K digestion of LARA^{sol} at 0.5 mg/mL (left), LARA^{fib} at 0.5 mg/mL (middle) and LARA^{fib} at 1 mg/mL (right). Each lane corresponds to a different time point. Band A corresponds to the full length protein while bands B, C and D are the different protease-resistant fragments. Colors correspond to colors used in (D).

(D) The percent of total mass spectrometry signal intensity obtained for each peptide from each gel band in (C) are mapped onto the LARA domain amino acid sequence. Black bars are for band A, red bars are for band B, green bars are for band C. Peptides in gray were not detected.

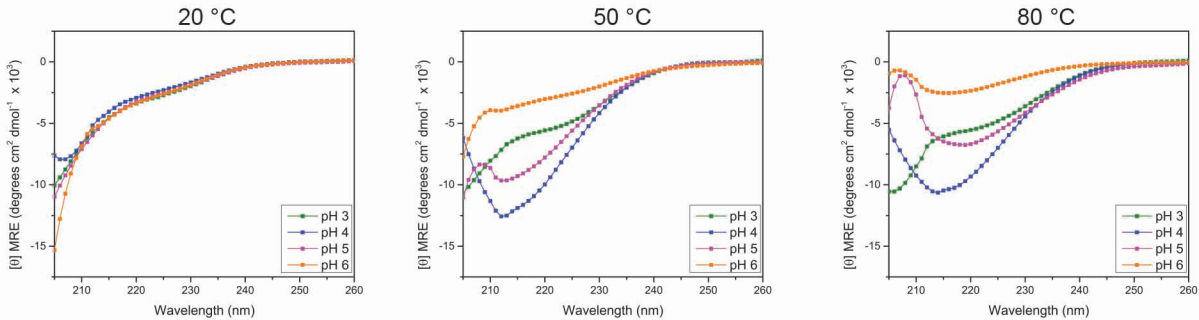
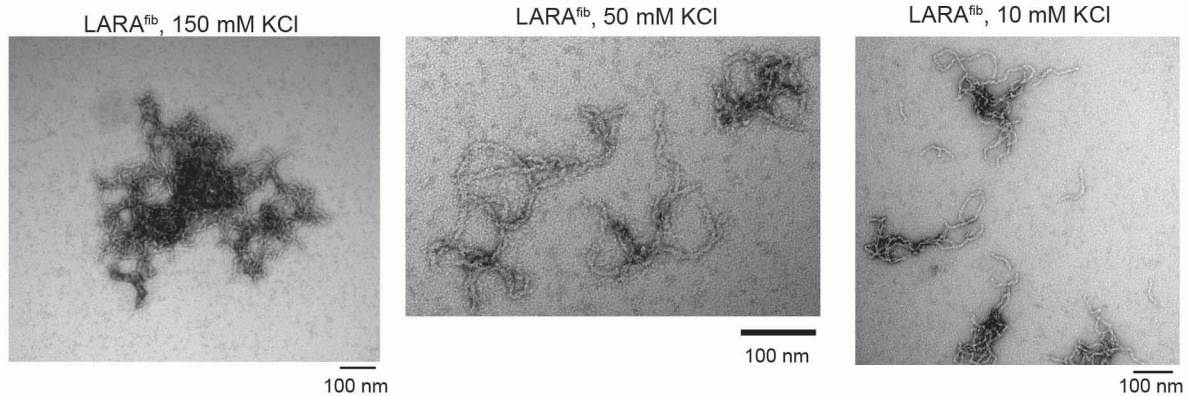
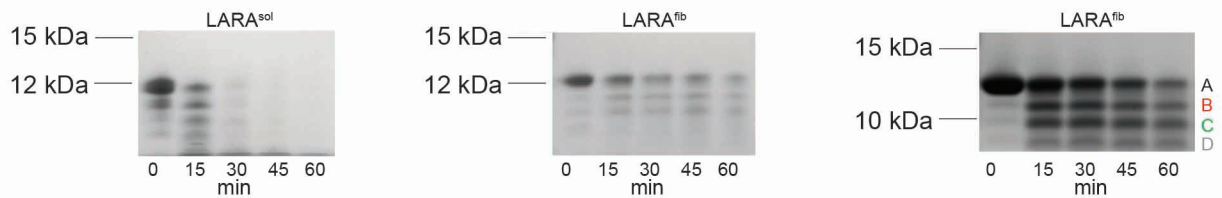
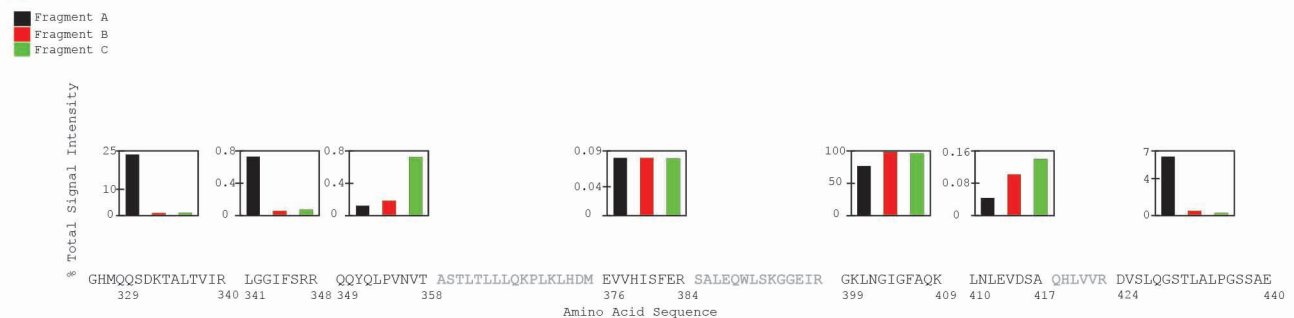
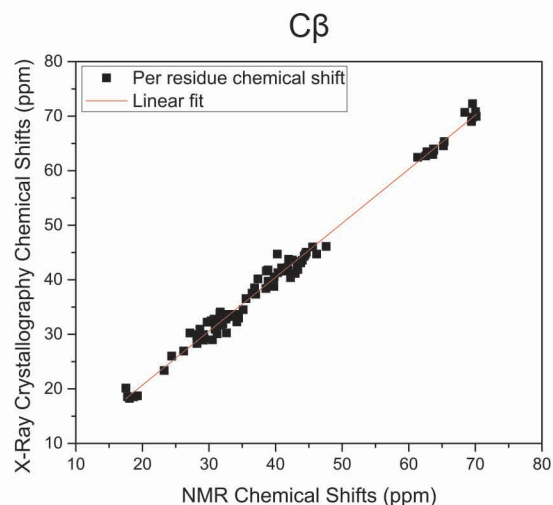
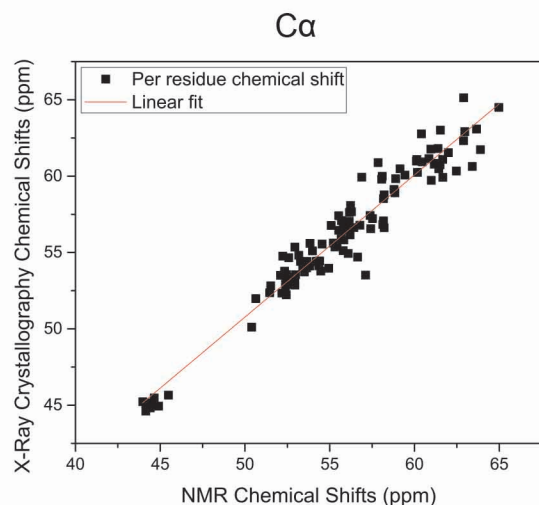
A**B****C****D**

Figure S3. Related to Figures 3 and 4. Comparison of NMR data with the X-ray crystal structure of LARA.

(A) Comparison of the chemical shifts predicted from the X-ray crystal structure of LARA calculated using SHIFTX2 (Han et al., 2011) and the assigned NMR chemical shifts. Shown are comparisons of C α (left, $R^2 = 0.943$) and C β (right, $R^2 = 0.991$) shifts.

(B) Comparison of the secondary structure calculated using TALOS+ (Shen et al., 2009), and Chemical Shift Index (CSI) (Wishart and Sykes, 1994) using observed NMR chemical shifts with the secondary structure defined by the X-ray crystal structure (PDB ID 3NBX) (El Bakkouri et al., 2010): α -helix (green), β -sheet (yellow) and random coil (no highlight).

A



B

```

          329
TALOS+ (NMR) : GHMQQSDKTALTVIRLGGIFSRQQYQLPVNVTASTLTLLQKPLKLHDMEVVHISFEES
CSI   (NMR) : GHMQQSDKTALTVIRLGGIFSRQQYQLPVNVTASTLTLLQKPLKLHDMEVVHISFEES
Crystal  : GHMQQSDKTALTVIRLGGIFSRQQYQLPVNVTASTLTLLQKPLKLHDMEVVHISFEES
          440

TALOS+ (NMR) : ALEQWLSKGGGEIRGKLNIGFAQKLNLEVDQAHLVVRDVSLQGSTLALPGSSAE
CSI   (NMR) : ALEQWLSKGGGEIRGKLNIGFAQKLNLEVDQAHLVVRDVSLQGSTLALPGSSAE
Crystal  : ALEQWLSKGGGEIRGKLNIGFAQKLNLEVDQAHLVVRDVSLQGSTLALPGSSAE
```


Figure S4. Related to Figures 3 and 4. Lipari-Szabo model-free analysis of the LARA domain.

(A-C) Shown are ^{15}N R_1 and R_2 NMR relaxation rates and ^1H - ^{15}N heteronuclear NOE values at pH 4 and 24 °C measured for LARA^{sol} at 700 MHz.

(D) Order parameter values (S^2) calculated from the relaxation data for each residue in LARA^{sol}. Secondary structure elements are plotted on top of each panel in (A-D).

(E) LARA^{sol} is shown with S^2 values mapped onto the structure, along with corresponding secondary structure labels. Residues for which S^2 could not be calculated are shown in black.

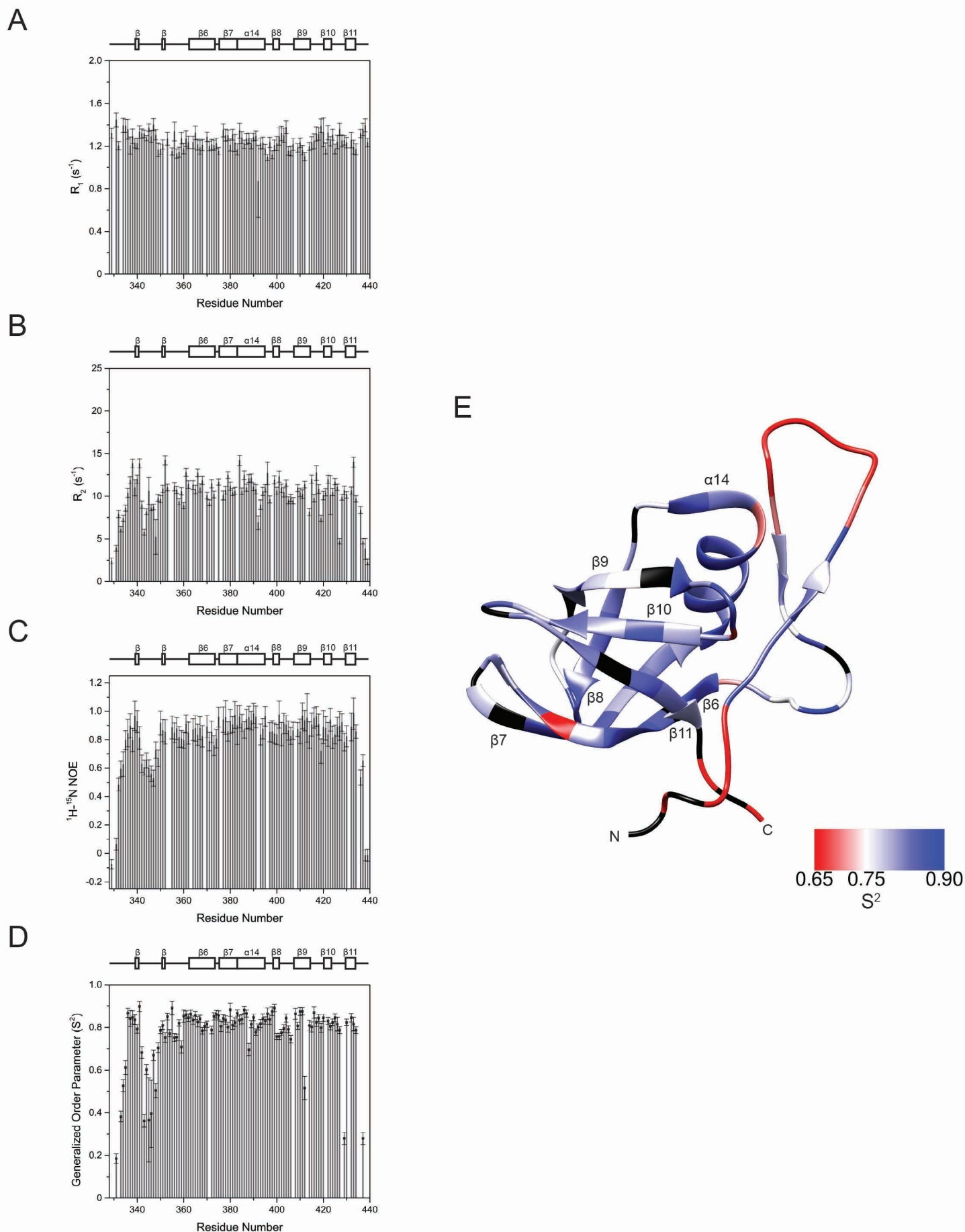


Figure S5. Related to Figure 5. Solvent-mediated salt bridge stabilization of the LARA domain.

(A) Example of a salt bridge interaction between R340 and E413 mediated by H₂O. A water molecule serves as an intermediary donor and acceptor that stabilizes the salt bridge. Snapshot from pH 7, Run 2, at 695 ns.

(B) Example of a salt bridge interaction between R348 and E413 mediated by H₂O. Snapshot from pH 7, Run 5, at 667 ns. In both A and B, direct hydrogen bonding between one arginine and E413 facilitates the other arginine to position the water molecule to form the salt bridge with E413.

(C) A metastable interaction was observed between the side chain of R348 and the backbone carbonyl oxygen of G394. Snapshot from pH 4, Run 4, at 630 ns - also refer to Figure 5C. The reason for the close distance between R348 and E413 was also a solvent mediated salt bridge. However, we suspect the nature of this salt bridge differs from those observed at pH 7 because the anchoring from G394 appears necessary to orient R348 to interact with E413.

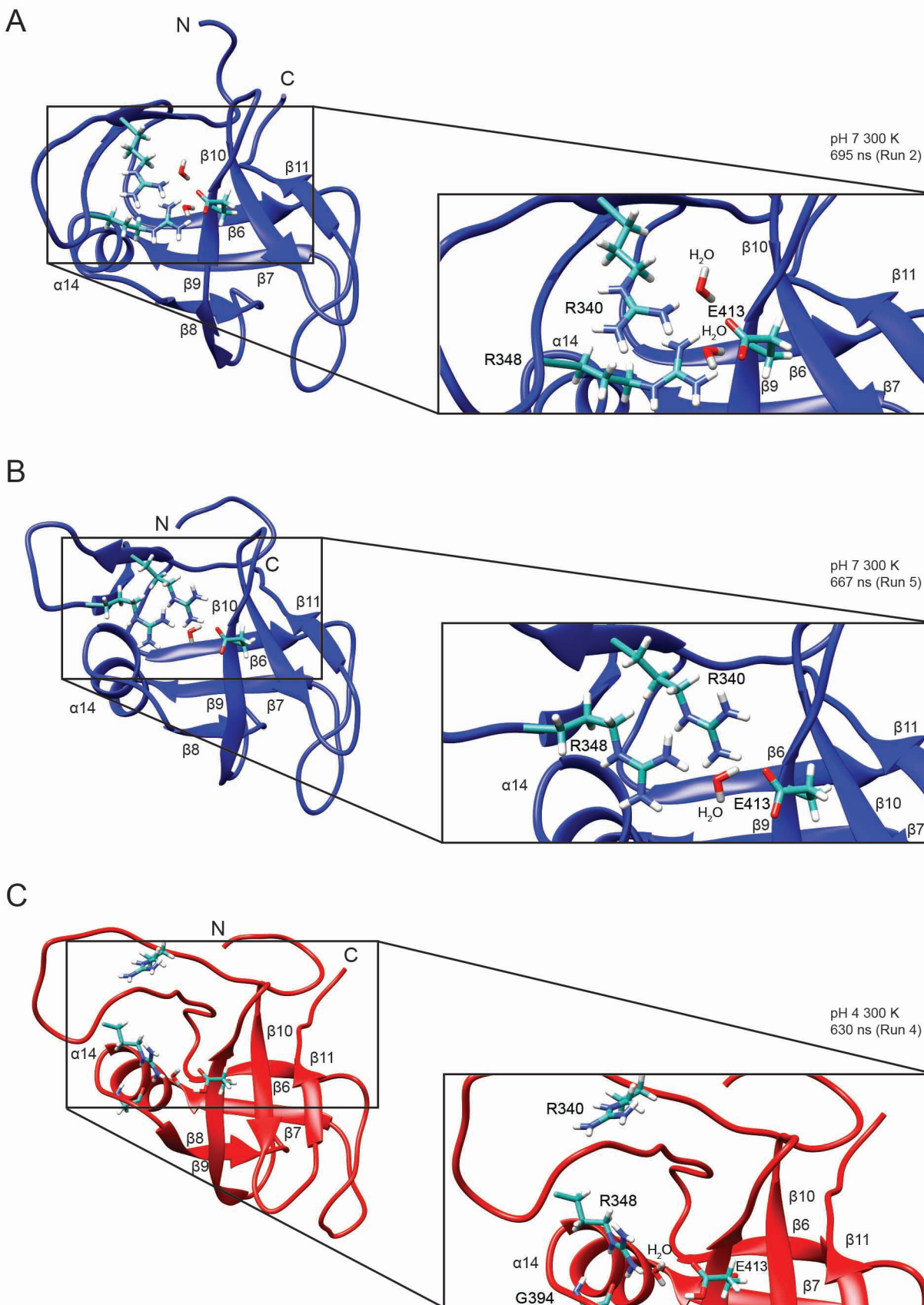
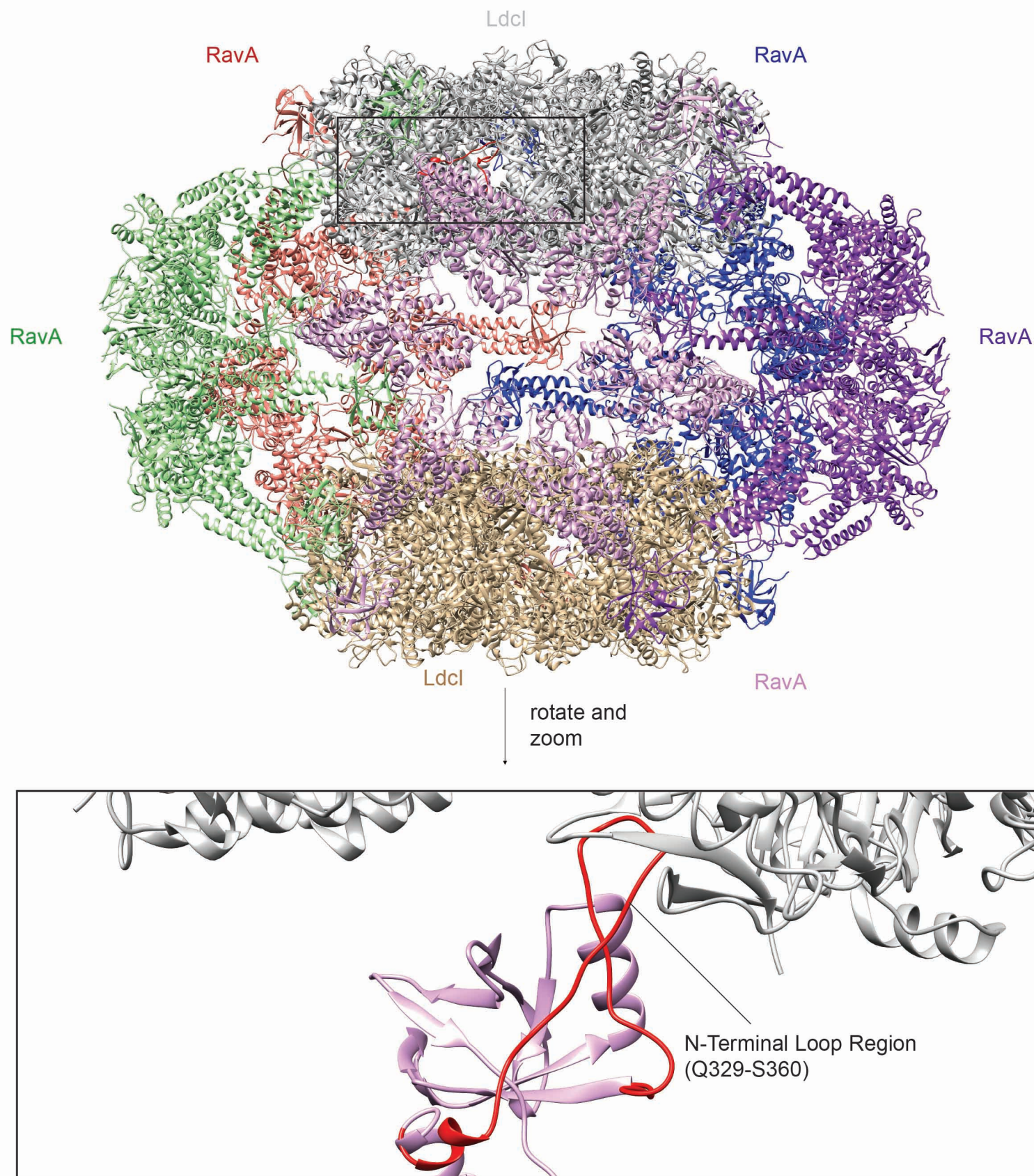


Figure S6. Related to Figure 8. RavA-LdcI cage structure.

The image is adapted from our cryo-EM structure of the RavA-LdcI cage (Malet et al., 2014) and our X-ray structures of RavA (El Bakkouri et al., 2010) and LdcI (Kanjee et al., 2011). LdcI decamers are at the top and bottom of the cage, while five RavA hexamers bridge the two LdcI decamers. The N-terminal loop region of LARA is shown in red in one of the protomers.



SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Related to Figure 1. Prediction of amyloidogenic regions in RavA.

(A) Shown is the amino acid sequence of RavA colored based on domains (same color scheme used in Figure 1). Highlighted in orange boxes are the regions predicted to be amyloidogenic by all three methods used: MetAMYL (Emily et al., 2013), PASTA2.0 (Walsh et al., 2014) and AMYLPRED2 (Tsolis et al., 2013). Purple boxes indicate regions predicted to be amyloidogenic by two of these methods. Secondary structure elements and residue numbers are placed above the sequence.

(B) The RavA protomer structure is shown with the different amyloidogenic regions indicated in (A) highlighted. Secondary structure elements are labeled for the LARA domain.

Figure S2. Related to Figure 2. Characterization of the LARA domain amyloid state.

(A) CD spectra of the LARA domain at 20 °C, 50 °C, 80 °C in pH 3 (green), pH 4 (blue), pH 5 (light magenta), and pH 6 (orange) buffers.

(B) TEM images of negatively stained LARA^{fib} prepared in 150, 50 and 10 mM KCl viewed at 150,000x, 120,000x and 150,000x magnification, respectively.

(C) Time course of proteinase K digestion of LARA^{sol} at 0.5 mg/mL (left), LARA^{fib} at 0.5 mg/mL (middle) and LARA^{fib} at 1 mg/mL (right). Each lane corresponds to a different time point. Band A corresponds to the full length protein while bands B, C and D are the different protease-resistant fragments. Colors correspond to colors used in (D).

(D) The percent of total mass spectrometry signal intensity obtained for each peptide from each gel band in (C) are mapped onto the LARA domain amino acid sequence. Black bars are for band A, red bars are for band B, green bars are for band C. Peptides in gray were not detected.

Figure S3. Related to Figures 3 and 4. Comparison of NMR data with the X-ray crystal structure of LARA.

(A) Comparison of the chemical shifts predicted from the X-ray crystal structure of LARA calculated using SHIFTX2 (Han et al., 2011) and the assigned NMR chemical shifts. Shown are comparisons of C α (left, $R^2 = 0.943$) and C β (right, $R^2 = 0.991$) shifts.

(B) Comparison of the secondary structure calculated using TALOS+ (Shen et al., 2009), and Chemical Shift Index (CSI) (Wishart and Sykes, 1994) using observed NMR chemical shifts with the secondary structure defined by the X-ray crystal structure (PDB ID 3NBX) (El Bakkouri et al., 2010): α -helix (green), β -sheet (yellow) and random coil (no highlight).

Figure S4. Related to Figures 3 and 4. Lipari-Szabo model-free analysis of the LARA domain.

(A-C) Shown are ^{15}N R_1 and R_2 NMR relaxation rates and ^1H - ^{15}N heteronuclear NOE values at pH 4 and 24 °C measured for LARA^{sol} at 700 MHz.

(D) Order parameter values (S^2) calculated from the relaxation data for each residue in LARA^{sol}. Secondary structure elements are plotted on top of each panel in (A-D).

(E) LARA^{sol} is shown with S^2 values mapped onto the structure, along with corresponding secondary structure labels. Residues for which S^2 could not be calculated are shown in black.

Figure S5. Related to Figure 5. Solvent-mediated salt bridge stabilization of the LARA domain.

(A) Example of a salt bridge interaction between R340 and E413 mediated by H₂O. A water molecule serves as an intermediary donor and acceptor that stabilizes the salt bridge. Snapshot from pH 7, Run 2, at 695 ns.

(B) Example of a salt bridge interaction between R348 and E413 mediated by H₂O. Snapshot from pH 7, Run 5, at 667 ns. In both A and B, direct hydrogen bonding between one arginine and E413 facilitates the other arginine to position the water molecule to form the salt bridge with E413.

(C) A metastable interaction was observed between the side chain of R348 and the backbone carbonyl oxygen of G394. Snapshot from pH 4, Run 4, at 630 ns - also refer to Figure 5C. The reason for the close distance between R348 and E413 was also a solvent mediated salt bridge. However, we suspect the nature of this salt bridge differs from those observed at pH 7 because the anchoring from G394 appears necessary to orient R348 to interact with E413.

Figure S6. Related to Figure 8. RavA-LdcI cage structure.

The image is adapted from our cryo-EM structure of the RavA-LdcI cage (Malet et al., 2014) and our X-ray structures of RavA (El Bakkouri et al., 2010) and LdcI (Kanjee et al., 2011). LdcI decamers are at the top and bottom of the cage, while five RavA hexamers bridge the two LdcI decamers. The N-terminal loop region of LARA is shown in red

in one of the protomers.

SUPPLEMENTAL VIDEOS

Video S1. Related to Figure 5. Molecular dynamics simulation of LARA at pH 7 showing the R340-E413 interaction.

Representative simulation of a sample at pH 7 (Run 3 from Figure 5C) highlighting interactions between R340 and E413. The peptide backbone is in blue, side chain carbons are in turquoise, nitrogen is blue and oxygen is red.

Video S2. Related to Figure 5. Molecular dynamics simulation of LARA at pH 7 showing the R348-E413 interaction.

Representative simulation of a sample at pH 7 (Run 2 from Figure 5C) highlighting interactions between R348 and E413. The peptide backbone is in blue, side chain carbons are in turquoise, nitrogen is blue and oxygen is red.

Video S3. Related to Figure 5. Molecular dynamics simulation of LARA at pH 7 showing switching between the R340-E413 and R348-E413 interactions.

Representative simulation of a sample at pH 7 (Run 5 from Figure 5C) highlighting the switch between R340-E413 and R348-E413 interactions. The peptide backbone is in blue, side chain carbons are in turquoise, nitrogen is blue and oxygen is red.

Video S4. Related to Figure 5. Molecular dynamics simulation of LARA at pH 4 showing no R340/R348-E413 interactions

Representative simulation of a sample at pH 4 (Run 2 from Figure 5C) highlighting the lack of interactions between R340/R348 and E413. The peptide backbone is in red, side chain carbons are in turquoise, nitrogen is blue and oxygen is red.

SUPPLEMENTAL TABLES

Table S1. Related to Figures 2, 6 and 7. Deconvolution of ATR-FTIR Amide I region spectra of various LARA constructs using second derivative analysis. * indicates no detectable level of secondary structure.

Construct	β -sheet (%)	α -helix (%)	β -turn (%)	Random coil (%)
LARA ^{sol}	46	22	20	12
LARA ^{fib}	52	16	27	4
LARA(351-440) ^{sol}	55	23	22	*
LARA(351-440) ^{fib}	68	11	17	4
LARA(R340D,R347D,R348D) ^{sol}	44	21	20	15
LARA(R340D,R347D,R348D) ^{fib}	48	12	24	16

Table S2. Related to Figure 2, 6 and 7. The time to reach half of the fibrillization reaction ($t_{0.5}$) was obtained from a sigmoidal fit to the ThT fluorescence emission kinetics data. The lag time (t_{lag}) was calculated from the x-intercept of a straight line having a slope of maximal velocity and passing through $t_{0.5}$. Reactions with no observable lag phase are indicated by an asterisk.

Protein	$t_{0.5}$ (min)	t_{lag} (min)
25 μ M WT LARA	373 \pm 11	127 \pm 28
25 μ M WT LARA + seed	277 \pm 12	99 \pm 22
20 μ M WT LARA	446 \pm 18	179 \pm 45
20 μ M LARA(351-440)	95 \pm 13	*
20 μ M LARA(R340D, R347D, R348D)	229 \pm 13	*
15 μ M WT LARA	552 \pm 30	212 \pm 90
15 μ M LARA(351-440)	306 \pm 16	*
15 μ M LARA(R340D, R347D, R348D)	356 \pm 16	*
10 μ M WT LARA	771 \pm 18	291 \pm 61
10 μ M LARA(351-440)	484 \pm 13	59 \pm 32
10 μ M LARA (R340D, R347D, R348D)	465 \pm 26	*

Table S3. Related to Figure 2. Peptide masses obtained from analysis of tryptic digests of the proteinase K resistant fragments of the LARA domain (refer to Figure S2D).

Peptide	Peptide Modified Sequence (number are additional Da)	Precursor m/z	Precursor Positive Charge	Precursor Neutral Mass (Da)
GHMQQSDKTALT VIR	GHM[+16]Q[+1]QSDKTALT VIR	567.629794	3	1699.868
GHMQQSDKTALT VIR	GHM[+16]Q[+1]QSDKTALT VIR	425.974164	4	1699.868
GHMQQSDKTALT VIR	GHM[+16]Q[+1]QSDKTALT VIR	567.957799	3	1700.852
GHMQQSDKTALT VIR	GHM[+16]Q[+1]QSDKTALT VIR	426.220168	4	1700.852
QSDKTALT VIR	Q[+1]SDKTALT VIR	1232.68449	1	1231.677
QSDKTALT VIR	Q[+1]SDKTALT VIR	616.845884	2	1231.677
QSDKTALT VIR	Q[+1]SDKTALT VIR	411.566348	3	1231.677
DKTALT VIR	DKTALT VIR	508.808573	2	1015.603
LGGIFSRR	LGGIFSRR	453.269418	2	904.5243
RQQYQLPVNVT	RQQYQLPVNVT	673.364775	2	1344.715
QQYQLPVNVT	QQYQLPVNVT	1189.62116	1	1188.614
QQYQLPVNVT	QQYQLPVNVT	595.31422	2	1188.614
EVVHISFER	EVVHISFER	558.29583	2	1114.577
GKLNIGGFAQKL	GKLN[+1]GIGFAQKL	623.861337	2	1245.708
GKLNIGGFAQKL	GKLN[+1]GIGFAQKL	416.243317	3	1245.708
GKLNIGGFAQKL	GKLNIGGFAQKL	415.915312	3	1244.724
GKLNIGGFAQK	GKLNIGGFAQK	566.827297	2	1131.64
GKLNIGGFAQK	GKLN[+1]GIGFAQK	567.319305	2	1132.624
GKLNIGGFAQK	GKLNIGGFAQK	566.827297	2	1131.64
GKLNIGGFAQ	GKLNIGGFAQ	1004.55236	1	1003.545
GKLNIGGFAQ	GKLNIGGFAQ	502.779816	2	1003.545
GKLNIGGFAQ	GKLN[+1]GIGFAQ	1005.53637	1	1004.529
GKLNIGGFAQ	GKLN[+1]GIGFAQ	503.271824	2	1004.529
GKLNIGGFAQ	GKLN[+1]GIGFAQ[+1]	1006.52039	1	1005.513
GKLNIGGFAQ	GKLN[+1]GIGFAQ[+1]	503.763832	2	1005.513
GKLNIGGFA	GKLN[+1]GIGFA	877.477794	1	876.4705
GKLNIGGFA	GKLN[+1]GIGFA	439.242535	2	876.4705
GKLNIGGFA	GKLNIGGFA	876.493778	1	875.4865
GKLNIGGFA	GKLNIGGFA	438.750527	2	875.4865
GKLNIGGF	GKLNIGGF	805.456664	1	804.4494
GKLNIGGF	GKLNIGGF	403.23197	2	804.4494
GKLNIGGF	GKLN[+1]GIGF	806.44068	1	805.4334
GKLNIGGF	GKLN[+1]GIGF	403.723978	2	805.4334
LNGIGFAQKL	LNGIGFAQKL	530.811116	2	1059.608
LNGIGFAQKL	LN[+1]GIGFAQKL	531.303124	2	1060.592
LNGIGFAQK	LNGIGFAQK	474.269084	2	946.5236
LNGIGFAQK	LNGIGFAQ[+1]K	948.514908	1	947.5076
LNGIGFAQK	LNGIGFAQ[+1]K	474.761092	2	947.5076

LNGIGFAQ	LNGIGFAQ	819.435929	1	818.4287
LNGIGFAQ	LNGIGFAQ	410.221602	2	818.4287
LNLEVDSAQ	LNLEVDSAQ	988.494566	1	987.4873
LNLEVDSAQ	LNLEVDSAQ	494.750921	2	987.4873
LNLEVDSA	LNLEVDSA	860.435989	1	859.4287
LNLEVDSA	LNLEVDSA	430.721632	2	859.4287
DVSLQGSTLALPGSSAE	DVSLQGSTLALPGSSAE	816.409774	2	1630.805

Table S4. Related to Figure 3. Residues in the LARA domain whose chemical shifts could be assigned at 40 °C, likely corresponding to sites that are highly mobile and disordered in the amyloid state.

N-Terminus	Core	C-Terminus
Q329, S331, D332, K333, I344, F345, R347, R348	L372, S385, A386, Q389, K400	V425, S426, G436, S437, A433, E440

Table S5. Related to Figure 3. Average ^1H temperature coefficients from defined structural elements within the LARA domain.

Region	Average ($\Delta\delta$ ^1H ppb/$^{\circ}\text{C}$)
Q329-S360 (N-terminal loop)	-5.2
V338-Y351 (heart shaped region)	-5.7
T361-L372 (β_6)	-4.1
H373-E383 (β_7)	-2.9
R384-G394 (α_{14})	-3.6
G395-S416 (β_8 , β_9 , β_{10} backside region)	-5.5
A417-Q428 (β_{11})	-3.4
G429-E440 (β_{12}) and C-terminus	-5.8

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Protein purification

The plasmid for expressing the LARA domain (Q329-E440), p11-LARA, was described previously (El Bakkouri et al., 2010). The domain is expressed with an N-terminal His₆-tag followed by a Tobacco Etch Virus (TEV) protease cleavage site that leaves residues GHM at the N-terminus after cleavage. RavA construct also has an N-terminal His₆-tag with a TEV cleavage site (El Bakkouri et al., 2010). Plasmids were transformed into a BL21(DE3) Gold pLysS strain of *Escherichia coli* (Stratagene) and expression was induced with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) for 4 hours at 30 °C as previously described (El Bakkouri et al., 2010). For isotopic labelling, cells were induced with 0.5 mM IPTG for 4 hours at 30 °C in M9 Minimal Media supplemented with 1 g/L ¹⁵NH₄Cl and 2 g/L ¹³C glucose. RavA and LARA cell lysates were purified as previously described (El Bakkouri et al., 2010) using nickel-nitrilotriacetic acid-agarose beads (Qiagen). Elutions were cleaved with TEV protease and further purified using an UnoS S1 column (Biorad) or a Mono S 5/50 cation-exchange column (GE Healthcare), concentrated, quantified using absorbance at 280 nm (same method for all experiments), flash frozen using liquid N₂ and stored at -80 °C.

Amyloid fibril preparation

Soluble LARA was dialyzed into pH 4, 25 mM citrate buffer with 50 mM KCl (LARA fibrillization buffer). RavA soluble protein was buffer exchanged into pH 2.5, 25 mM citrate buffer, 50 mM KCl and 1 mM DTT (RavA fibrillization buffer) using BioRad Micro Bio-Spin 6 columns as per manufacturer's protocol. RavA and LARA were converted into fibrils by heating the protein from 20 °C to 80 °C at a rate of 1 °C /min. Soluble LARA is typically referred to in the text as LARA^{sol} and fibrillized LARA as LARA^{fib}.

Circular dichroism (CD) spectroscopy

Secondary structures of full-length RavA and the isolated LARA domain were analyzed using a JASCO J-810 spectropolarimeter with a 1.0 mm path length quartz cuvette. For the pH range 3-6, a 25 mM citrate buffer was used with 50 mM KCl and titrated with KOH to achieve the desired pH. The LARA domain was diluted to 50 μM and RavA was diluted to 10 μM for measurement. For temperature melts, the temperature was raised at a rate of 1 °C/min. Spectra shown were the average of three wavelength scans from 200 – 260 nm at 50 nm/min measured every 5 °C. For other experiments, the 217 nm signal was monitored at 1 °C/min increments. Ellipticity was converted to mean residue ellipticity units and smoothed using a Savitzky-Golay smoothing algorithm with 30 points per window using a 3rd degree polynomial according to protocols described in (Greenfield, 2006). BeStSel (Micsonai et al., 2015) was used to fit spectra.

Thioflavin-T (ThT) fluorescence assay

ThT fluorescence change was measured on an EnSpire 2300 Multilabel Reader with excitation set at 442 nm and emission set at 480 nm for kinetic experiments. Spectra were scanned from 462 to 600 nm for assessment of ThT binding. All samples were measured in Greiner Bio-One black fluorescence 96-well plates in a final reaction volume of 200 μL. Soluble or fibril LARA domain was diluted to a final concentration of 50 μM in LARA fibrillization buffer supplemented with 20 μM of ThT for wavelength fluorescence scans. Soluble or fibril RavA was diluted to a final concentration of 25 μM in RavA fibrillization buffer and 20 μM ThT.

Analysis of fibrillization kinetics was performed with 25 μM LARA^{sol} in LARA fibrillization buffer set at a constant temperature of 33 °C to initiate fibrillization. The intensity of fluorescence emission at 480 nm (442 nm excitation) was recorded every 30 min. Seeds were prepared by bath sonication of 25 μM LARA^{fib} for 45 min and 2.5 % v/v of seeds were added to the final reaction. Kinetic experiments were carried out in triplicates using a volume of 200 μL in each well topped with 50 μL of mineral oil to prevent evaporation. Data were fitted to a Boltzmann sigmoidal function ($y = \frac{A_1 - A_2}{1 + e^{\frac{(x - x_0)}{dx}}} + A_2$) using OriginPro 9.0 (OriginLab). The midpoint of the graph was determined and the lag time was calculated based on the tangent at the midpoint of the curve with the time axis.

Transmission electron microscopy (TEM)

Samples of RavA or LARA domain fibrils were prepared for TEM following general procedures outlined in (Walsh et al., 2009). Continuous carbon films from copper rhodium grids (Electron Microscopy Sciences) were glow discharged for 15 s at 30 mA negative discharge. Protein samples were diluted 1 in 4 times in fibrillization buffer and 5 μL were adsorbed to grids for 2 min before rinsing with 10 μL of water for 10 seconds. Samples were blotted with

No. 2 Whatman filter paper and stained with 2% uranyl acetate for 15 seconds. TEM images were obtained using a Jeol 1011 microscope operating at 80 keV.

Attenuated total reflectance Fourier transform infrared radiation (ATR-FTIR) spectroscopy

Infrared spectra were taken with a Nicolet Nexus 6700 FTIR spectrometer (Thermo Fisher Scientific, Madison) coupled with a Smart Orbit ATR unit. Protein concentration of 5 mg/mL was used for ATR-FTIR analysis. 2 μ L of LARA was deposited on a clean diamond crystal. For each spectrum, 512 scans were collected with resolution of 1 cm^{-1} over a scan range of 4000 – 525 cm^{-1} using Origin software (version 8.3.103). Spectra were processed using OriginPro 9.0 (OriginLab) for the Amide I region (1600 – 1700 cm^{-1}) according to the methods described in (Yang et al., 2015). Values were normalized to the highest signal and a nine-point Savitzky–Golay second-derivative function was used for signal deconvolution analysis. Assignments were based on literature values described in (Kong and Yu, 2007) and (Sarroukh et al., 2013), while secondary structure proportions were determined by fitting assigned peaks using a Voigt function, which is a convolution of Gaussian and Lorentzian profiles (Manor et al., 2012).

Thermal melt monitored using static light scattering (SLS)

A temperature melt of the LARA domain was carried out using the Optim 2 (Avacta Analytical) according to the manufacturer's protocol. Triplicates were used (LARA concentration at 50 μ M in fibrillization conditions) for each melt with ramping at 1 $^{\circ}\text{C}/\text{min}$ and static light scattering was monitored at 266 nm. Data analysis was performed using the instrument software to analyze the SLS signal and to determine T_{agg} . The T_{agg} value was determined as follows: the curve was differentiated, and then the data was subjected to a Heaviside function to prevent complications from precipitation effects. A 10% change in signal was then defined as T_{agg} .

Thermal melt monitored by differential scanning fluorimetry (nanoDSF)

Temperature melts of LARA domain constructs were performed in triplicate, using the nanoDSF machine (Prometheus NT.48 from NanoTemper Technologies). Samples were aspirated into capillaries which were physically treated to have high quality surface properties. Each LARA domain construct was at 50 μ M in 25 mM citrate, pH 4, and 50 mM KCl. The UV detection system was used to monitor changes in the Trp and Tyr fluorescence emission using an excitation wavelength of 280 nm. The ratio of the emission intensities at 350 and 330 nm (F350/F330) was used to track structural changes in the protein as temperature was increased with a linear ramp of 1 $^{\circ}\text{C}/\text{min}$. Data analysis was conducted using the manufacturer's software where peaks in the first derivative were used to identify the T_m .

Protease digestion assay

LARA^{sol} or LARA^{fib} were mixed with proteinase K (Bioshop) using 0.5 mg/mL of LARA and 5 μ g/mL of proteinase K or 1.0 mg/mL of LARA and 10 μ g/mL of proteinase K. Proteins were incubated in 25 mM TrisHCl, pH 7.5 at 37 $^{\circ}\text{C}$ and digestion was stopped at different time points by boiling the samples in Laemmli buffer. Samples were then separated on SDS-PAGE gel, bands were excised, and in-gel trypsin digest was used to extract the peptides. Extracted peptides were sequenced using LC-MS/MS Q Exactive (Thermo Scientific) at the SPARC Biocentre at the Hospital for Sick Children (Toronto).

Prediction of amyloid propensity

The RavA amino acid sequence was analyzed for amyloidogenic propensity using MetAMYL (Emily et al., 2013), PASTA2.0 (Walsh et al., 2014) and AMYLPRED2 (Tsolis et al., 2013). MetAMYL was used with the highest global accuracy setting. For PASTA2.0, amyloidogenic regions were determined with top pairing of 22 and energy threshold of -2.8. For AMYLPRED2, residues were predicted to be amyloidogenic if at least half of the prediction parameters were in agreement.

Nuclear magnetic resonance (NMR) spectroscopy

All NMR experiments were typically recorded on Bruker Avance III 600 MHz and 700 MHz spectrometers using 5 mm TXI triple resonance probes equipped with z-gradient pulse-field gradient capabilities. The 800 MHz CPMG relaxation dispersion experiments were recorded on a Varian Unity Inova four channel spectrometer using a 5 mm, triple resonance, three axis pulse field gradient probe. Chemical shift assignments of ^{13}C , ^{15}N -labeled LARA domain were carried out at 7 $^{\circ}\text{C}$ using standard triple-resonance 3D backbone experiments HNCA, HNCACB, CBCA(CO)NH, HNCO, HN(CO)CA, HN(CA)CO, CC-TOCSY-(CO)HN, as provided from Bruker Biospin. NMR samples contained 25 mM citrate buffer pH 4.0, 50 mM KCl, 10% D_2O , and a final protein concentration of 450 μ M. All spectra were processed with NMRPipe processing scripts (Delaglio et al., 1995) and analyzed using CCPNMR

Analysis (Vranken et al., 2005). Secondary structure assignment was calculated from the chemical shifts using both Chemical Shift Index (CSI) (Wishart and Sykes, 1994) and TALOS+ (Shen et al., 2009). Predicted chemical shifts for the LARA domain were calculated using SHIFTX2 (Han et al., 2011) based on the crystal structure of RavA (PDB ID 3NBX). Resonance assignments have been deposited in the Biological Magnetic Resonance Bank (BMRB) database (accession number 26735).

NMR samples contained 25 mM citrate buffer pH 4.0, 50 mM KCl, 10% D₂O, and a final LARA concentration of 450 μ M. Thermal denaturation of ¹³C, ¹⁵N-labeled LARA was monitored using 2D ¹H-¹⁵N heteronuclear single quantum correlation (HSQC) spectra. Data were acquired every 5 °C for 7 min between 10 °C – 30 °C and every 2 °C for 5 min between 30 °C and 40 °C. Chemical shifts were externally referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) at each temperature. Amide ¹H chemical shifts were tracked and plotted as the change in chemical shift ($\Delta\delta$) as a function of temperature. The slope of the $\Delta\delta/^\circ\text{C}$ plot (temperature coefficient) was used to analyze trends in protein thermal sensitivity. Resonance assignments at each temperature were made based on the linear changes in ¹H-¹⁵N resonances relative to the assignment dataset at 7 °C.

¹⁵N R₂ (spin-spin) and R₁ (spin-lattice) relaxation times were derived from a pseudo-3D HSQC experiment recorded at 24 °C (Farrow et al., 1994; Kay et al., 1989) using a spin-echo sequence for R₂ and an inversion-recovery sequence for R₁ measurements. The 3rd dimension was set as a range of delay times (t) and NMR peak intensities were mathematically fitted to a simple exponential decay equation: $I(t) = I_0 e^{-t/T}$. R₁ delay times were incremented at 10 ms, 25 ms, 50 ms, 75 ms, 100 ms, 150 ms, 200 ms, 250 ms, 300 ms, 500 ms, 750 ms, 2000 ms and 2500 ms. R₂ experiments were incremented with a 16 ms delay block for 1, 2, 3, 4, 5, 6, 8, 10, or 12 blocks. Steady-state heteronuclear ¹H-¹⁵N NOE measurements were collected by measuring spectra with and without ¹H saturation and analyzed using CCPNMR Analysis (Vranken et al., 2005). Lipari-Szabo model-free analysis on the LARA domain was carried out using the relax GUI interface (Bieri et al., 2011; d'Auvergne and Gooley, 2008a, b).

Carr-Purcell-Meiboom-Gill (CPMG) R₂ relaxation dispersion profiles of soluble ¹⁵N-labeled LARA domain were measured with a ¹H-decoupled ¹⁵N CPMG-HSQC pulse sequence at 24 °C (Tollinger et al., 2001). CPMG refocusing pulses frequencies (νCPMG) between 0 - 1000 Hz were employed over a constant CPMG time relaxation interval of 40 ms. CPMG frequency is defined as $1/(4\delta)$ where 2δ is the time between consecutive refocusing pulses. Duplicate measurements at two or three νCPMG values were used to estimate errors in effective transverse relaxation rates (R_{2(eff)}) as described in (Korzhnev et al., 2004). Intensity line shape fitting was extracted using PINT by employing Gaussian shape fitting of peak volume intensities (Ahlner et al., 2013).

Relaxation dispersion profiles of LARA^{sol} were recorded at three fields (600, 700, 800 MHz ¹H frequency) and the CPMG profiles for each individual residue were fit across all datasets in order to determine which residues exhibited significant dispersion in the R₂ relaxation rate due to chemical exchange with a low-population excited state. After estimating an initial timescale for the exchange processes observed, all residues that exhibited dispersions were globally fit to Bloch-McConnell equations for two state exchange using an in-house program, ChemEx (Vallurupalli et al., 2012). Residues were then fit individually by fixing k_{ex} and pB (population of the minor state, B) to the globally determined values and comparing the ratio of the reduced χ^2 between the fixed fit and individual fits. This was to ensure that the residues being analyzed were exchanging in the same regime. Residues with a χ^2 ratio exceeding 2 were excluded. As a result, F345 was excluded from further analysis. The data were refit using the remaining data in a global fit to arrive at the final global exchange parameters. For structural analysis of the excited state, $\Delta\omega$ (chemical shift difference from the global analysis) was compared to the magnitude of the difference between native chemical shifts and predicted random coil chemical shifts (Wishart et al., 1995).

Molecular dynamics simulations

The LARA domain of RavA (residues 330-437) was extracted from PDB ID 3NBX and solvated in a rectangular cell for pH 7 simulations and a rhombic dodecahedron cell for pH 4 simulations. At low pH, all glutamic acid carboxylate groups (E376, E383, E388, E396, and E413) were singly protonated and all histidine imidazole groups were doubly protonated (H373, H379, H419). At neutral pH, all histidine imidazole groups were singly protonated at N-epsilon. Both N- and C-terminal ends of the protein were modeled as neutral moieties. Na⁺ and Cl⁻ ions were added to each simulation box as needed to obtain a net charge of zero and approximately 150 mM excess salt. The rectangular cell consisted of the LARA domain, 8706 water molecules, 24 Na⁺ atoms, and 27 Cl⁻ atoms. The dodecahedron cell consisted of the LARA domain, 5728 water molecules, 17 Na⁺ atoms, and 28 Cl⁻ atoms.

All simulations were performed with GROMACS (Pronk et al., 2013). The protein and ions were modeled with the optimized potentials for liquid simulations (OPLS) all-atom force field (Jorgensen et al., 1996; Kaminski et al., 2001), and the transferable intermolecular potential 3-point (TIP3P) model (Jorgensen et al., 1983) was used for water molecules. Lennard-Jones interactions were evaluated using a group-based cut off for separation distances less than 1.0 nm. Coulomb interactions were calculated using the smooth particle-mesh Ewald (PME) method (Darden et

al., 1993; Essmann et al., 1995) with a real-space cut off of 1.0 nm and a Fourier grid spacing of 0.12 nm. All simulations were done at 300 K. Simulation in the isothermal-isobaric (NPT) ensemble at pressure $P = 1$ atm was achieved with isotropic coupling to a Berendsen barostat (Berendsen et al., 1984) (1 bar) with a coupling constant of 4 ps and temperature coupling using velocity Langevin dynamics with coupling constants of 1 ps^{-1} (Van Gunsteren and Berendsen, 1988). Covalent bonds were constrained with SETTLE (Miyamoto and Kollman, 1992) and P-LINCS (Hess, 2008) for water and other molecules, respectively. The integration time step was 2 fs. The non-bonded pair-list was updated every 20 fs.

Before production simulations, both pH 4 and pH 7 systems were subjected to 1000 steps of steepest descent energy minimization and 10 ns of simulation with protein backbone atoms harmonically restrained to the initial conformation (with a force constant of $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$). Subsequent simulations were performed without positional restraints. Five 1000 ns repeats were performed for each model for a total of 10 microseconds of data. Based on the convergence of root-mean-square deviation with respect to the initial conformation, we removed the first 100 ns of each simulation from our analysis. Time trajectories were saved at 50 ps intervals. Distance calculations, root-mean-square calculations, and secondary structure were computed using MDTraj (McGibbon et al., 2015). Molecular visualizations of simulation trajectories were produced using VMD (Humphrey et al., 1996). Mean and standard error of mean values were computed over simulation repeats for each system.

Supplemental References

- Ahlner, A., Carlsson, M., Jonsson, B.H., and Lundstrom, P. (2013). PINT: a software for integration of peak volumes and extraction of relaxation rates. *J. Biomol. NMR* 56, 191-202.
- Berendsen, H.J.C., Postma, J.P.M., Vangunsteren, W.F., Dinola, A., and Haak, J.R. (1984). Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 81, 3684-3690.
- Bieri, M., d'Auvergne, E.J., and Gooley, P.R. (2011). relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μ s motion of proteins. *J. Biomol. NMR* 50, 147-155.
- Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98, 10089-10092.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 103, 8577-8593.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*. 33, 5984-6003.
- Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876-2890.
- Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* 4, 116-122.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33-38, 27-38.
- Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*. 28, 8972-8979.
- Kong, J., and Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim Biophys Sin (Shanghai)* 39, 549-559.
- Korzhev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430, 586-590.
- Manor, J., Feldblum, E.S., Zanni, M.T., and Arkin, I.T. (2012). Environment Polarity in Proteins Mapped Noninvasively by FTIR Spectroscopy. *The journal of physical chemistry letters* 3, 939-944.
- McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernandez, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J., and Pande, V.S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* 109, 1528-1532.
- Miyamoto, S., and Kollman, P.A. (1992). Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* 13, 952-962.
- Vallurupalli, P., Bouvignies, G., and Kay, L.E. (2012). Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* 134, 8148-8161.
- Van Gunsteren, W.F., and Berendsen, H.J.C. (1988). A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation* 1, 173-185.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687-696.

Yang, H., Yang, S., Kong, J., Dong, A., and Yu, S. (2015). Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* 10, 382-396.