

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.

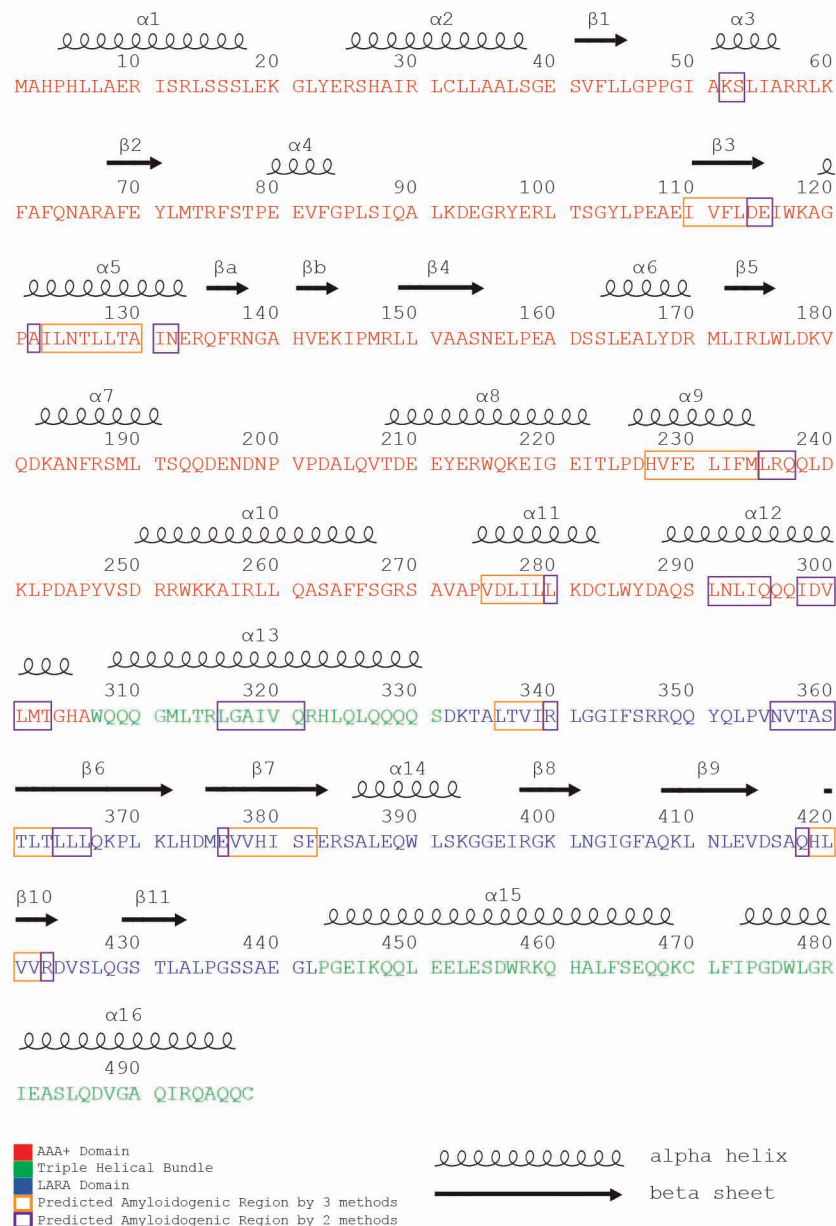
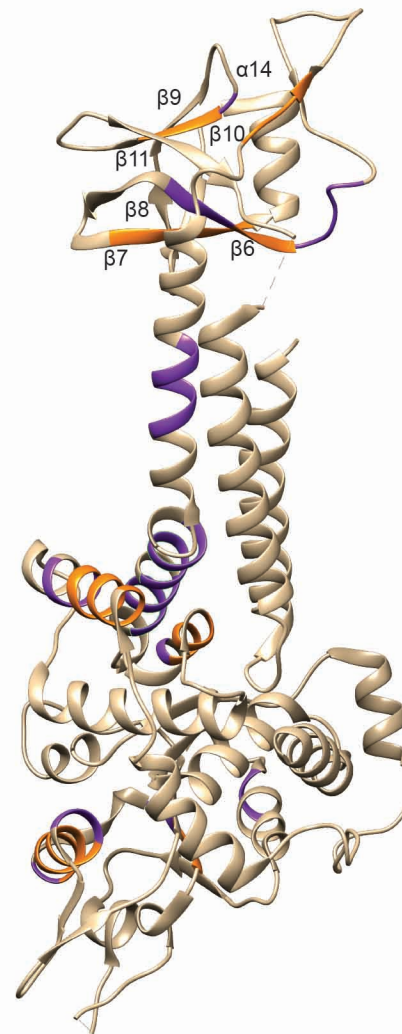
A**B**

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.

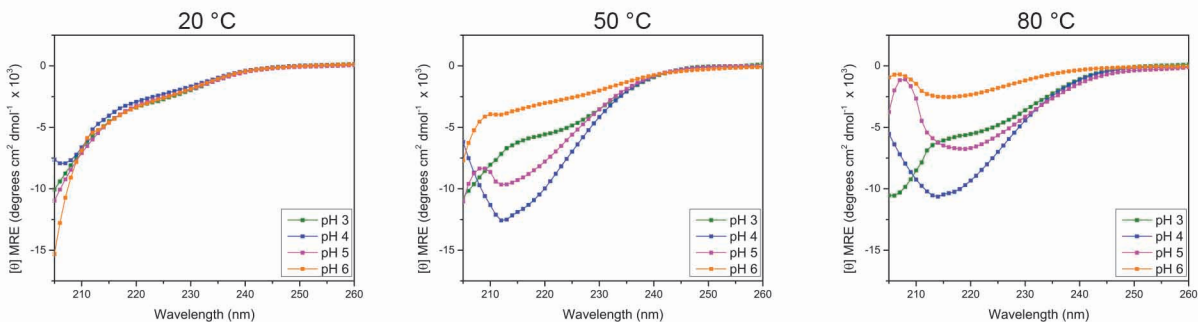
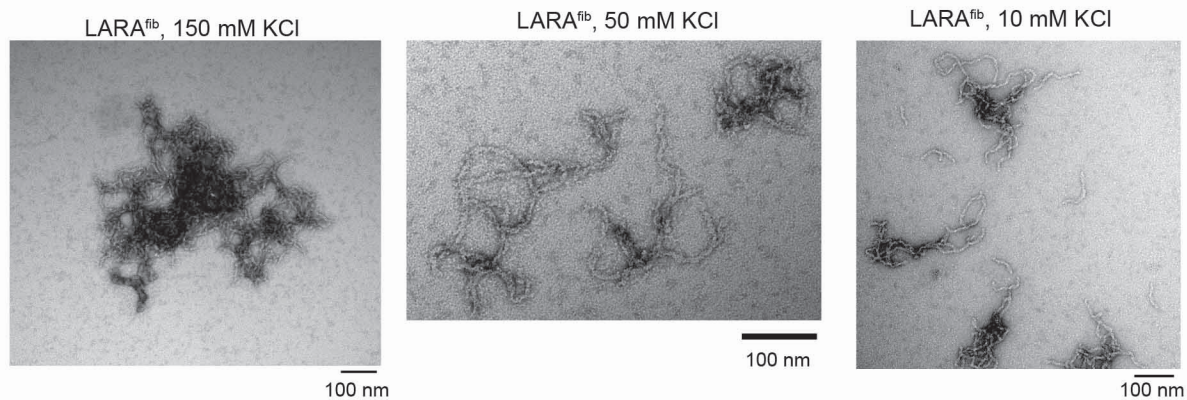
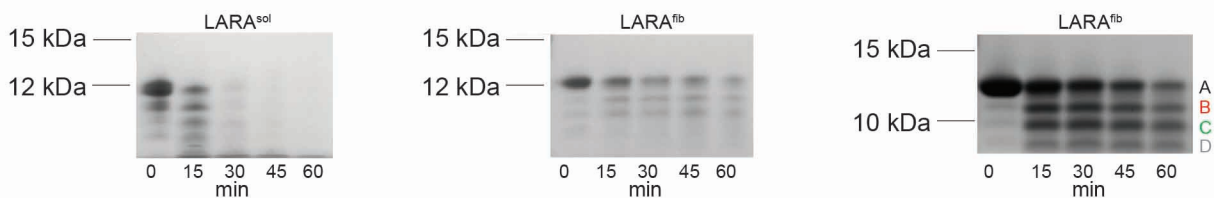
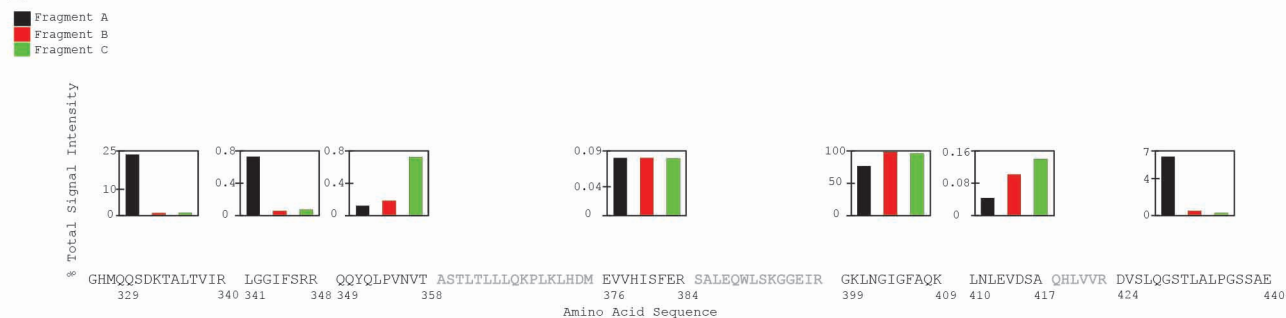
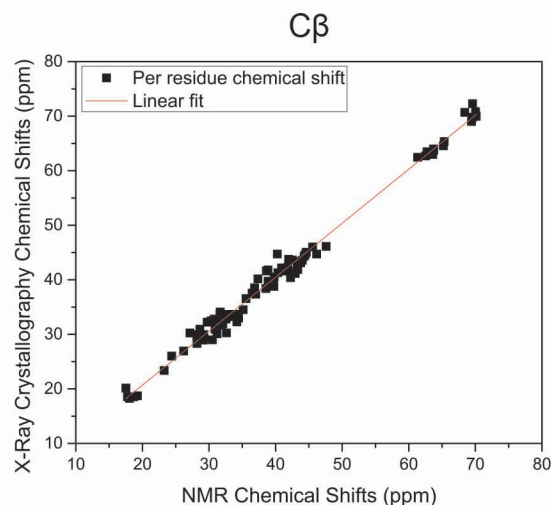
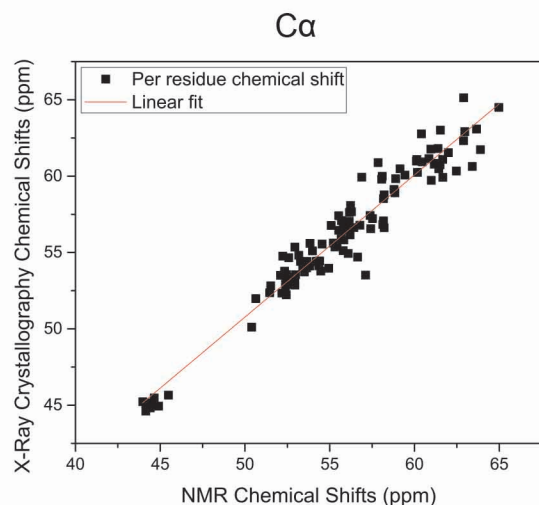
A**B****C****D**

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, R² = 0.943) and C β (right, R² = 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.

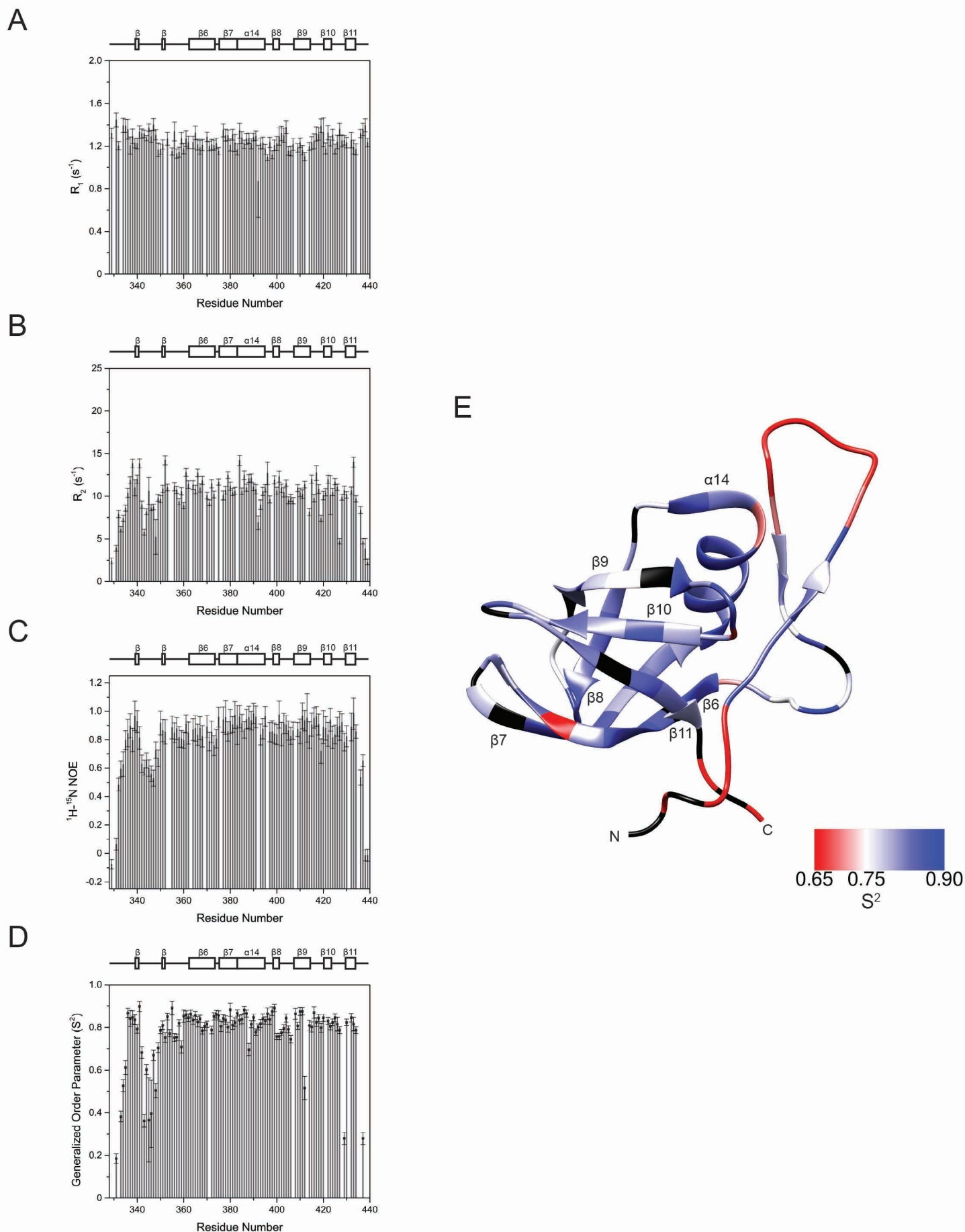


Figure S5. Related to Figure 5. Solvent-mediated salt bridge stabilization of the LARA domain.

(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.

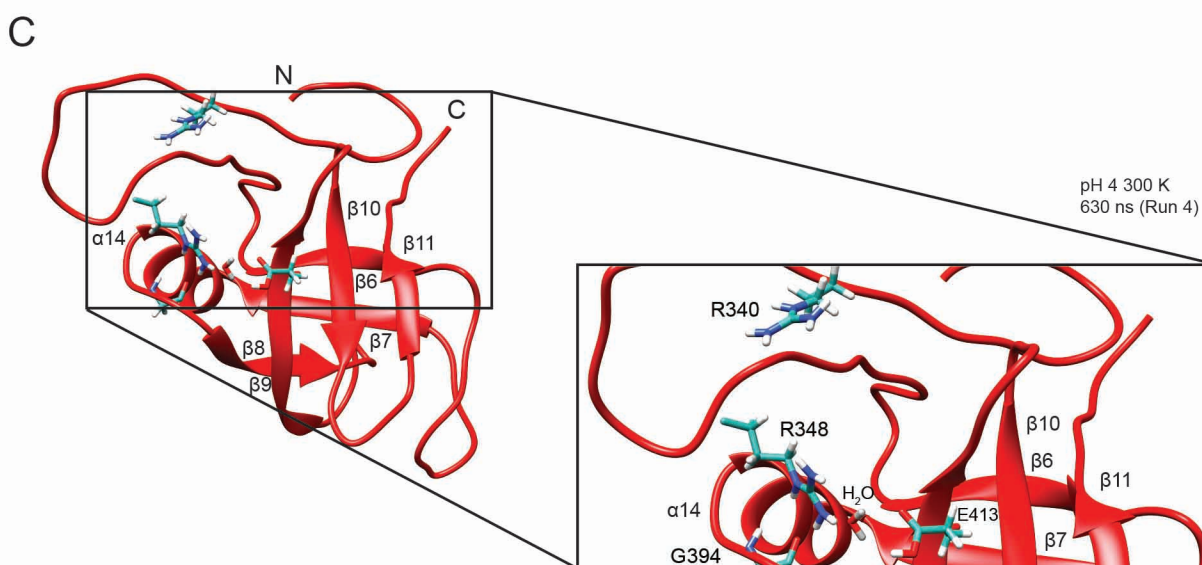
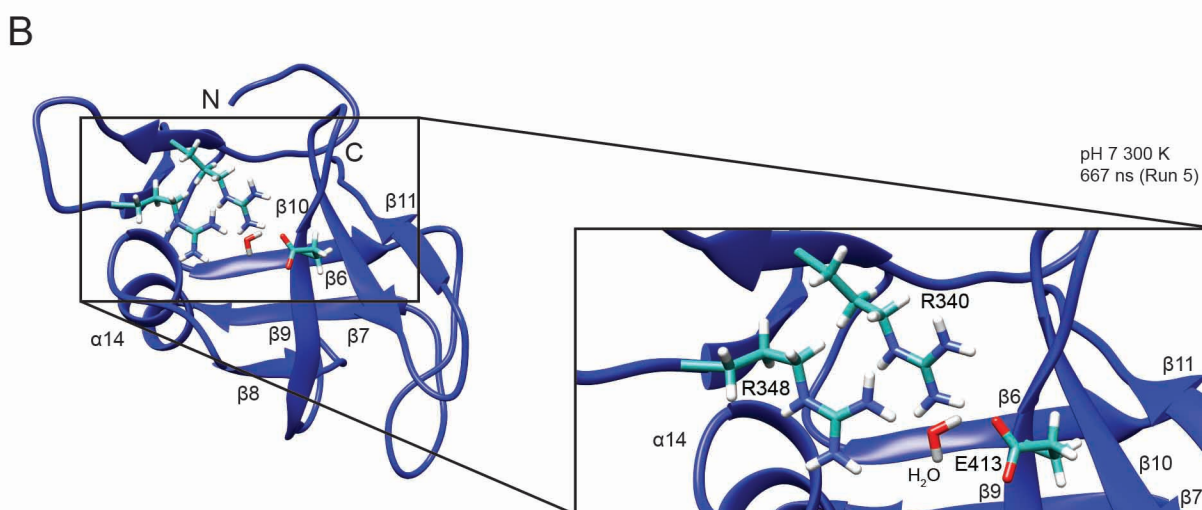
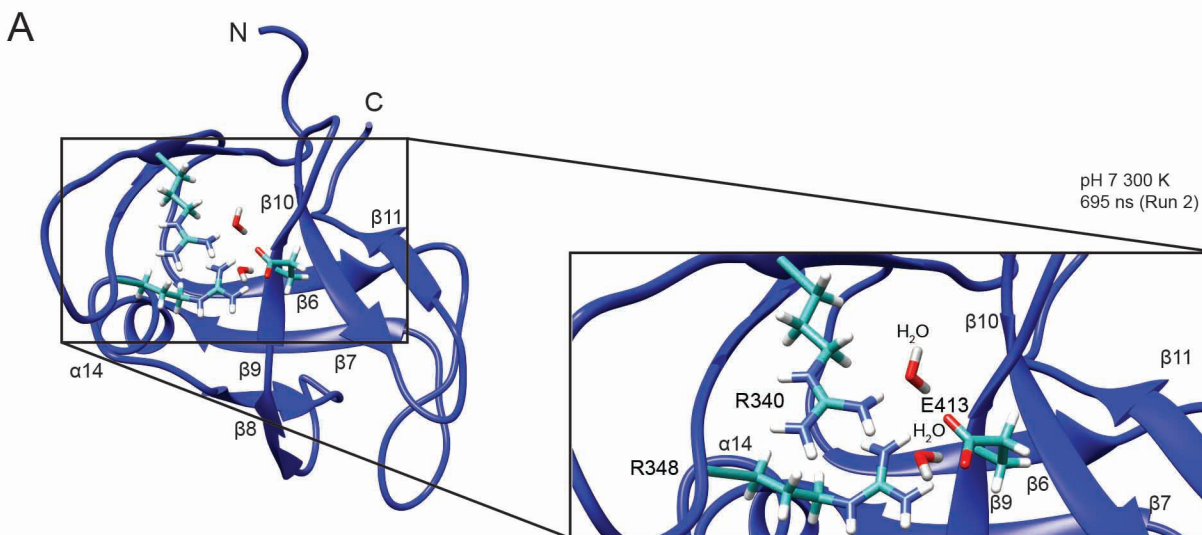
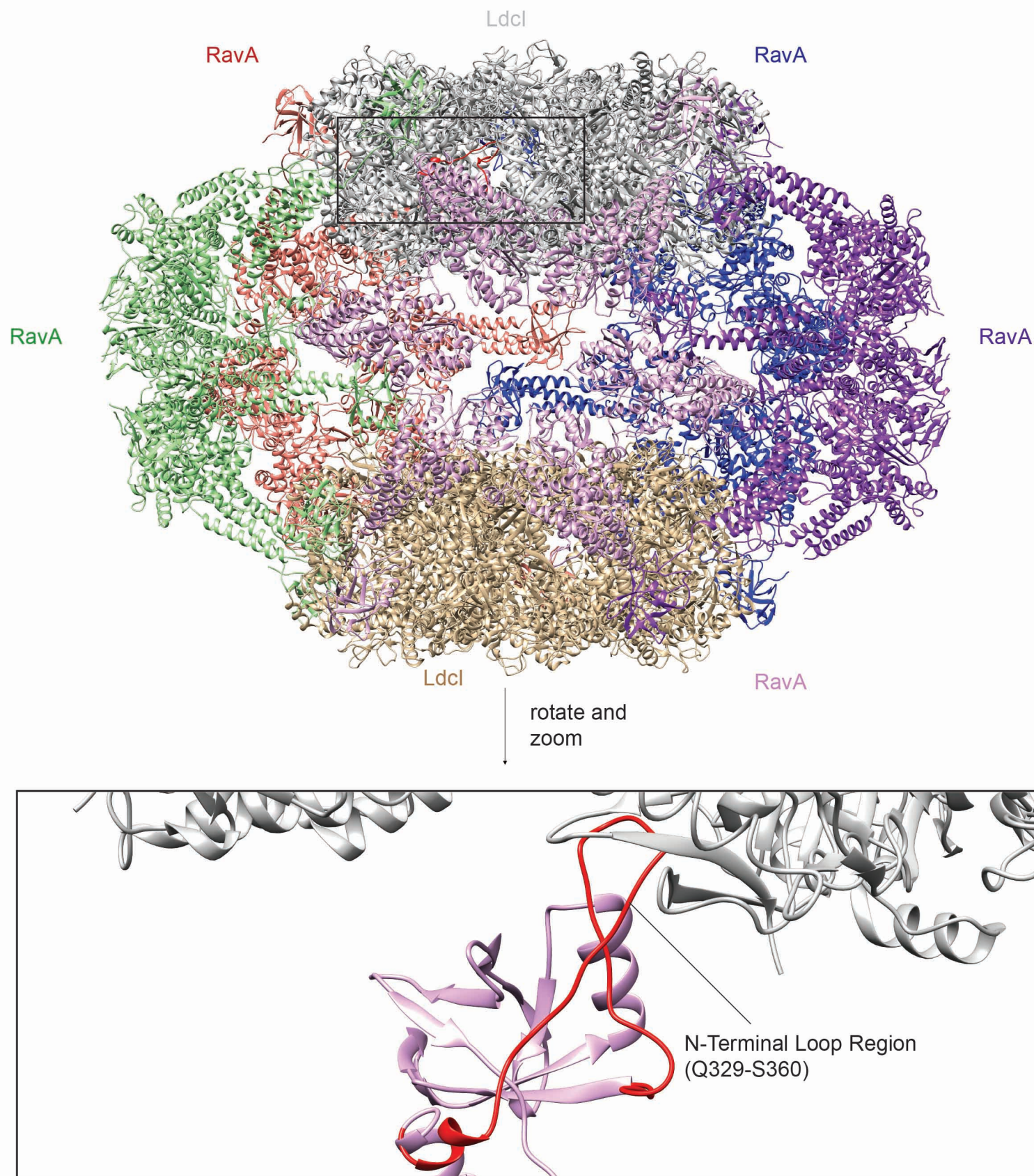


Figure S6. Related to Figure 8. RavA-LdcI cage structure.

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 Omnic 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.