Supporting Information

Competing stress-dependent oligomerization pathways regulate self-assembly of the periplasmic protease-chaperone DegP

Materials and methods

Plasmid constructs and cloning

The DegP gene from *E. coli* (strain K12, UniProt ID P0C0V0) including the periplasmic signal sequence (corresponding to amino acids 1-474) was synthesized by GenScript (codon-optimized). An expression construct without the sequence for periplasmic localization (amino acids 27-474 hereafter renumbered as 1-448 and referred to as full-length) and with an N-terminal His_6-SUMO tag was subsequently obtained by subcloning into a *Champion* pET SUMO plasmid (kanamycin resistance) using Gibson Assembly® (New England BioLabs Inc.). Constructs for expression of S210A DegP, oligomerization mutants, and the isolated PDZ1 and PDZ2 domains were generated using *Phusion* DNA polymerase and the Quikchange site-directed mutagenesis method. V348W and V386W mutations were introduced into the PDZ1 and PDZ2 domain constructs, respectively, to increase the extinction coefficients of these two proteins so that accurate quantification of their concentrations could be made. The hTRF1 plasmid was obtained from previous studies (1).

Protein expression and purification

DegP constructs were expressed to the cytoplasm of *E. coli* and not the periplasm for increased protein yield (2). DegP contains an internal disulfide bond within its protease domain which normally forms through the assistance of the DsbA-DsbB machinery in the bacterial periplasm (2). DegP constructs containing the protease domain were therefore expressed in *SHuffle T7 Competent E. coli cells* which are engineered to promote disulfide formation in the cytoplasm. Isolated PDZ1 and PDZ2 domain constructs and the hTRF1 client protein were expressed in *Codon+ E. coli BL21*(*DE3*) cells.

Unlabeled DegP protein expression was achieved by growing cells in LB media. For samples with $^{13}CH_3$-labeling at Ileδ, Leuδ, Valγ, and Metε positions, where only one of the pair of isopropyl methyls of Leu and Val is $^{13}CH_3$ labeled (referred to as U-^{2}H, ILVM-$^{13}CH_3$ labeling), cells were grown in D_2O M9 minimal media supplemented with d_7-glucose as the sole carbon source along with the addition of precursors (60 mg L^{-1} 2-keto-3-d_2-4-^{13}C-butyrate for Ileδ, 100 mg L^{-1} 2-keto-3-methyl-d_3-3-d_1-4-^{13}C-butyrate for Leuδ,Valγ-$^{13}CH_3/^{12}CD_3$ (3), and 100 mg L^{-1} methyl-$^{13}CH_3$-methionine for Metε (4)) 1 hour prior to the induction of protein expression. The U-^{15}N,^{13}C labeled PDZ1 domain was produced in H_2O, M9 minimal media supplemented with uniformly-^{13}C labeled glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, while deuterated PDZ2 was obtained by overexpression in D_2O, M9 minimal media. For stereospecific assignments of the PDZ1 domain methyl groups, the PDZ1 domain was produced in H_2O, M9 minimal media supplemented with 10% ^{13}C-glucose:90% ^{12}C glucose as the sole carbon source (5). Proteins were expressed by incubating cells to an OD_{600} of ~0.6-1.0 at 37 °C at which point IPTG was added to 0.2 mM and the cells were allowed to grow for a further ~18 hours at 25 °C.

Protein purification of full length DegP was achieved by first lysing cells in lysis buffer (100 mM NaH_2PO_4, 500 mM NaCl, 6M Gdn-HCl, pH 8.0) followed by centrifugation of the lysate at 9000 rpm for 15 minutes. The supernatant was filtered and loaded onto a nickel affinity purification column equilibrated with lysis buffer. Contaminant proteins (including any DegP natively expressed from the *E. coli* genome) were removed by washing with three rounds of lysis buffer (50-100 mL each) containing increasing amounts of imidazole pH 8.0 (0, 10, and 20 mM). Elution of the protein of interest was achieved using ~25 mL of lysis buffer + 500 mM imidazole. The nickel column eluate, containing Gdn-denatured DegP, was concentrated to ~1-2

mL using a 3 kDa MWCO Amicon Ultra-15 concentrator. Natively refolded DegP was obtained by dilution into refolding buffer (25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA, pH 7.0) at 100X volume (i.e. ~100-200 mL) with rapid stirring. The N-terminal His_6-SUMO tag was cleaved by the addition of Ulp1 protease, followed by stirring the solution for ~1 hour. Refolded DegP was then further purified using hydrophobic interaction chromatography for removal of the free His_6-SUMO tag and of contaminant substrates or peptides which could influence the oligomerization of DegP. For this step, the refolded DegP solution was filtered and ammonium sulfate was added to ~1M. The solution was loaded onto a 5 mL Butyl-HiTrap column equilibrated with buffer A (25 mM NaH_2PO_4, 1M ammonium sulfate, 1 mM EDTA, pH 7.0). DegP was eluted using a flow rate of 3 mL min^{-1} and an initial step to 50% buffer B (25 mM NaH_2PO_4, 1 mM EDTA, pH 7.0), washing for 10 mL, a subsequent gradient to 100% buffer B over 25 mL and final washing for 25 mL. Fractions containing DegP were concentrated, with a final purification step using a Superdex 200 Increase10/300 size-exclusion column equilibrated with 25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA, pH 7.0. DegP's isolated PDZ1 and PDZ2 domains were purified using a similar procedure except that the cell lysis and nickel affinity purification steps were performed without 6M Gdn-HCl in buffers (*i.e.* no denaturation and subsequent refolding). After nickel affinity purification, N-terminal His_6-SUMO tags were cleaved by dialyzing with Ulp1 protease overnight and subsequently removed by an additional nickel affinity purification step. Final purification of the domains was achieved using a Superdex 75 16/600 size-exclusion column. The hTRF1 client protein for DegP was purified as previously described (1). In all cases, purified proteins were buffer exchanged into final buffers used for experiments (given in each section subsequently) via Amicon Ultra-15 concentrators. The concentrations of purified samples were measured using a NanoDrop spectrophotometer and

their extinction coefficients at 280 nm (7575 M^{-1} cm^{-1} for S210A and S210A/N45F DegP, 6085 M^{-1} cm^{-1} for S210A/Y444A DegP, 5500 M^{-1} cm^{-1} for PDZ1 V348W, 6990 M^{-1} cm^{-1} for PDZ2 V386W, and 24980 M^{-1} cm^{-1} for hTRF1). These were obtained from an on-line calculator (https://web.expasy.org/protparam/protparam-doc.html).

Measuring in-cell DegP concentrations

Cells for measuring *in-vivo* DegP concentrations were initially grown by streaking *E. coli DH5α*, *BL21(DE3) pLysS*, and *MG1655* stocks onto LB plates. Starter cultures for each strain were then grown by inoculation of LB media with a single colony and shaking overnight at 37 °C. Cultures were subsequently diluted to OD_{600} = 0.02 in LB and grown for ~3 hours at 37 °C to the exponential phase or overnight to the stationary phase. Cell counts were obtained with a 1 mL aliquot of cell culture that was pelleted through high-speed centrifugation for 1 minute and resuspended in 1 mL PBS. Exponential-phase samples were directly loaded into a Petroff-Hausser counting chamber on counting slides. Stationary-phase samples were diluted 3-fold (to OD_{600} ~1.5-2) prior to loading. Cells were allowed to settle on the counting slides for several minutes prior to initiating counting. A 40x objective on a Nikon optic microscope was used to count cells, following a protocol from (https://instr.bact.wisc.edu/book/displayarticle/108). Briefly, grids composed of 16 squares were selected for counting. Squares containing a maximum of three cells were included in the counting process. To ensure statistical significance, at least 100 cells were counted for each strain and growth condition. Average cell numbers per square were calculated as the number of cells counted/total number of squares used. Cell densities were then calculated as the average cell number per square/counting square volume

(0.05 mm × 0.05 mm × 0.02 mm for each square). These were used to obtain the in-cell concentrations of DegP as outlined below.

The moles of DegP in aliquots of each *E. coli* strain under a given growth condition were estimated by Western Blot analyses. A 1.5 mL aliquot of each cell culture was harvested by centrifugation at high speed for 1 minute. The cell pellets were resuspended in 1X SDS-PAGE loading buffer (100 and 300 μL for the exponential and stationary phase samples respectively). Resuspended cells were then boiled in a water bath for 10 minutes. These were finally centrifuged at high speed for 15 minutes at 4 °C. Western blots were performed by loading 10 μL aliquots of each supernatant onto a 10% SDS-PAGE gel, in addition to lanes containing known amounts of purified S210A DegP (~1-9 pmol) which were used to generate a standard curve for determining the moles of in-cell DegP. The SDS-PAGE gel was transferred onto a polyvinylidene difluoride (PVDF) membrane and detected by an anti-DegP/maltose-binding protein (MBP) antibody (1:10000) and an anti-rabbit-alkaline phosphatase (AP) conjugate antibody (1:10000). Nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) were added to the membrane for a colorimetric reaction catalyzed by AP which was scanned in an Azure illumination system and quantified with ImageJ to obtain DegP band intensities. The standard curve for S210A DegP was obtained by fitting the purified S210A DegP band intensities to the linear expression $I(\text{pmol DegP}) = m \times \text{pmol DegP} + I_0$ where $I(\text{pmol DegP})$ is the band intensity for a given amount of loaded S210A DegP, I_0 is the intensity at zero pmol DegP, and m is the slope of the standard curve. The unknown moles of cellular DegP were then determined using the corresponding DegP band intensities. The amount of DegP per cell was calculated as pmol DegP loaded/number of cells used per lane (determined from the cell densities and SDS-PAGE loading volumes above). The molar concentration of

DegP per cell was then calculated using the volume of the *E. coli* cell (6). The periplasmic concentration of DegP was obtained by scaling the whole-cell concentration by the periplasmic fraction of the *E. coli* cell (8%) (7), assuming that all of the DegP is localized to the periplasm. We note that higher periplasmic volume fractions of 20-40% have also been found for *E. coli* (8); therefore our calculated periplasmic concentrations of DegP represent an upper limit. These analyses were performed in triplicate and the periplasmic concentrations in Figure 1C are given as the mean ± SEM.

DLS measurements and autocorrelation analysis

DLS autocorrelation functions for generating 3D DLS datasets were recorded using a plate reader format Wyatt DynaPro DLS instrument with a 150° detector angle and 824 nm laser irradiation. For the DLS datasets shown in Figures 2A, E, F, Figure 5C, *SI Appendix* Figures S1-S6, and *SI Appendix* Figure S8, protein stocks were prepared in 25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA, pH 7.0 and diluted to obtain the desired total protein monomer concentration, M_T. Sample stocks used to collect the DLS datasets as a function of [NaCl] (Fig. 5A&5B) were prepared by buffer exchanging into 25 mM NaH_2PO_4, 1 mM EDTA, pH 7.0. The stocks were then diluted with an appropriate volume of 25 mM NaH_2PO_4, 5 M NaCl, 1 mM EDTA, pH 7.0 to obtain the desired [NaCl] and M_T. All buffers were filtered using a 0.2 μm syringe filter prior to use. Samples for DLS analysis (40 μL volumes) were prepared using Eppendorf tubes that were evacuated of dust with compressed air. All samples were centrifuged at 17500 rpm for 15 minutes prior to loading onto the DLS plate to pellet any large aggregates or residual dust that could obscure the light scattering measurement. Each well was evacuated of dust with compressed air and loaded with 20 μL from the top of the tube to avoid any pelleted

particles or dust. The loaded plate was subsequently centrifuged for 1 minute at 4680 rpm to settle the samples into the bottom of the wells. Infrared spectroscopy-grade paraffin oil (10 μL) was then placed on top of each well to prevent evaporation of the sample during the temperature ramp. The plate was centrifuged once more for 1 minute to settle the oil onto the samples before loading into the instrument. Autocorrelation functions for each well were measured over the temperature range 5-50 °C in discrete increments of 2.5 °C for a total sampling of 19 temperature points. In these discrete temperature experiments, the instrument was equilibrated at each temperature for approximately 10-15 minutes prior to measuring each well. Each well was measured 25 times per temperature with an acquisition time of 1 second per measurement. A final autocorrelation function for each well at a given temperature was obtained by averaging of the 25 replicates. In cases where an autocorrelation replicate was found to be artifactual due to dust or a bubble (characterized by a noisy baseline or an anomalously slow decay and small D_z value), the replicate was manually filtered from the dataset.

In order to obtain a crude estimate of the kinetics involved in DegP self-assembly, and to ensure that the DLS measurements described above are on a system in thermodynamic equilibrium, an additional experiment was performed in which the temperature was scanned continuously at an extremely slow rate of 0.03 °C min^{-1} (~33.3 min $°C^{-1}$). Changes were not observed in the resulting DLS profiles relative to an effective rate of approximately ~4-6 min $°C^{-1}$ that was used in all of the discrete sampling measurements (~10-15 min for each 2.5 °C increment) and similar fitted parameters were obtained (*SI Appendix* Fig. S4).

Translational diffusion constants (D_z) were extracted from the average autocorrelation functions using the cumulants method (9, 10). The autocorrelation functions were numerically fitted according to

$$g_2(\tau) = B + \beta e^{\left(-2\bar{\Gamma}\tau\right)}\left(1 + \frac{\mu_2}{2!}\tau^2\right)^2 \quad [1]$$

where τ is the autocorrelation delay time, B and β account for the baseline and amplitude of the autocorrelation function respectively, $\bar{\Gamma} = D_z q^2$ is the average decay constant (the first cumulant), and μ_2 (the second cumulant) is the variance of the Γ values about the average decay constant and is related to the polydispersity of the sample. In the expression for $\bar{\Gamma}$, $q = \frac{4\pi n}{\lambda_0}\sin\left(\frac{\theta}{2}\right)$ is the scattering wave vector where n is the solvent refractive index (a value of 1.3347 for water was used), λ_0 is the laser wavelength in a vacuum (824 nm), and θ is the scattering angle (150°×π/180° radians). Optimized values of B, β, D_z, and μ_2 were output from each fit.

AUC isotherm measurements

(i) Sample preparation

Concentration sets for AUC analysis of S210A DegP samples (2.1 μM – 170.9 μM) and S210A/Y444A DegP (2.1 – 170.9 μM) were prepared by diluting stock protein solutions with the corresponding buffer. AUC data for S210A DegP were collected at 8, 20, and 30 °C (25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA, 0.02% sodium azide, pH 7.0) to investigate the temperature-dependence of the path A equilibrium for comparison with the DLS results. AUC data for the S210A/Y444A trimer mutant were recorded at 20 °C, and with identical buffer conditions as for S210A DegP, as a reference point for the trimer particle, with each AUC experiment on S210A DegP. Experiments were also performed in low salt buffer (25 mM NaH_2PO_4, 25 mM NaCl, pH 7.0) at 20 °C since the path A hexamer dissociates more readily

with increasing temperature under high salt conditions, which could affect estimates of its sedimentation coefficient (however, differences were not observed).

(ii) *Sedimentation velocity experiments and analysis*

Sedimentation velocity (SV) AUC experiments were performed using a ProteomeLab XL-I ultracentrifuge (Beckman Coulter, Indianapolis, IN) following protocols described previously (11, 12) with minor modifications. Briefly, 400 µL samples were loaded into AUC cell assemblies with 3 mm or 12 mm charcoal-filled Epon double-sector centerpieces. After inserting cells into an An-50TI rotor and mounting in the centrifuge chamber, the temperature was equilibrated for ~3 hours (20 and 30 °C) or overnight (8 °C), followed by acceleration to 30,000 or 45,000 rpm. Sedimentation profiles for each sample were acquired using both interference and absorbance (230 nm) optical detection systems. In the study of S210A DegP at 20 °C, samples were remixed (verifying consistent signal levels) and re-run at 30 °C.

SV analysis was carried out with SEDFIT (V16p34b) using the standard c(s) model which does not account for the non-ideality of sedimentation and diffusion (13) as these effects are small, below ~200 µM M_T (~9.4 mg mL^{-1}) for the particles studied here. Sedimentation coefficient distributions were normalized to water at 20 °C using SEDNTERP (14), and plotted and integrated using GUSSI (15). An isotherm of the signal-weighted average sedimentation coefficient, $s_{20,w}$, was generated for the concentration series at each temperature (8, 20, and 30 °C) for subsequent modeling.

NMR measurements

NMR experiments were recorded on a Bruker Avance III HD 18.8T spectrometer equipped with a cryogenically cooled, pulsed-field X,Y,Z-gradient, triple-resonance probe. All

NMR spectra were processed using NMRPipe (16) and visualized using NMRFAM-SPARKY (17) and nmrglue (18).

Sequential backbone resonance assignments of DegP's isolated PDZ1 domain were obtained using standard triple-resonance experiments (19) including HNCO, HNCACO, HNCACB, HBCBCACONNH, and HNN recorded on a U-^{15}N,^{13}C-labeled PDZ1 domain sample in 25 mM HEPES free acid, 100 mM NaCl, 1 mM EDTA, 2% D_2O, pH 7.0 at 25 °C. Assignments of methyl sidechains of Ile, Leu, Val, and Met residues were obtained through (H)C(CO)NH-TOCSY, H(CCO)NH-TOCSY, and long-range ^{13}C-^{13}C J-correlation experiments (19, 20). Stereospecific assignments of the PDZ1 methyl groups were obtained from a constant-time ^{13}C-^{1}H HSQC experiment (21, 22) measured on a sample of the PDZ1 domain isotopically-labeled according to (5). Triple resonance experiments were acquired using non-uniform sampling (NUS) with Poisson gap schemes (23) and sampling densities of 20%; other experiments were obtained using uniform sampling in the indirect dimensions. NUS-acquired datasets were processed using SMILE reconstruction (24) in NMRPipe (16). Methyl side chain assignments at 25 °C were mapped to 50 °C using a constant-time ^{13}C-^{1}H HSQC temperature series and transferred to S210A DegP.

^{15}N-^{1}H and ^{13}C-^{1}H (constant-time) HSQC experiments where 100 μM U-^{15}N,^{13}C-labeled PDZ1 was titrated with unlabeled PDZ2 (0-5 mM, Fig. 3) were recorded at 25 and 50 °C in 25 mM HEPES free acid, 225 mM NaCl, 1 mM EDTA, 2% D_2O, pH 7.0 (pD 7.4). Spectra in Figure 4A were obtained as HMQC datasets that exploit the methyl-TROSY effect (25, 26). These were recorded as a concentration series of U-^{2}H, ILVM-$^{13}CH_3$-labelled S210A DegP, from 25 μM–2 mM, or as a titration of U-^{2}H, ILVM-$^{13}CH_3$-labelled PDZ1 (100 μM) with 0-~4.8 mM U-^{2}H

PDZ2. These two titrations were measured at 50 °C in 25 mM HEPES free acid, 200 mM NaCl, 1 mM EDTA, 100% D_2O, pH 7.0 (pD 7.4). All NMR titrations made use of 3 mm NMR tubes.

Single particle electron cryo-microscopy

(i) Preparation of electron cryo-microscopy samples

All samples were vitrified on a FEI Vitrobot mark III at 4 °C and ~100% humidity in 25 mM NaH_2PO_4, 200 mM NaCl, 1 mM EDTA, pH 7.0. 2.5 µL of sample mixtures were applied to nanofabricated holey gold grids with a hole size ~2 µm, before plunge freezing into a 60:40 propane:ethane mixture held at liquid nitrogen temperature. Grids were glow discharged in air for 15 seconds and blotted for 18 seconds. hTRF1-bound DegP samples (Fig. 5B) were prepared by incubating DegP at 100 µM with 150 µM hTRF1 before freezing. The *E. coli* substrate-bound DegP (Fig. 5C) was frozen at a final DegP concentration of 100 µM.

(ii) Electron microscopy

Datasets for both the hTRF1-bound DegP and *E. coli* substrate-bound DegP (Fig. 5B&5C) were collected on an FEI Tecnai F20 electron microscope operating at 200 kV and equipped with a Gatan K2 summit direct detector device used in electron counting mode at 400 frames/sec. Movies were recorded as 30 fractions over a 15 second exposure. Defocuses ranged from 0.7 to 3.2 µm. Movies were recorded at a nominal magnification of 25000× corresponding to a calibrated pixel size of 1.45 Å and with an exposure rate of 5 electrons/pixel/s, and a total exposure of 35 electrons/$Å^2$. For the hTRF1 and *E. coli* substrate datasets 170 and 201 movies were collected, respectively, utilizing the Digital Micrograph software package.

(iii) Image analysis

All image analysis was carried out in *cryoSPARC* v3 (27). For both datasets, patch-based alignment and exposure weighting was done with a 10×10 grid and the resulting averages of frames were used for patch-based contrast transfer function (CTF) determination and particle picking. Templates for particle selection were generated by 2D classification of manually selected particles. For the DegP:hTRF1 dataset (Fig. 5B) 79,281 particle images were extracted in 200×200-pixel boxes and subjected to 2D classification. Classes corresponding to 12-mer particles were selected yielding 62,156 particle images that were then used in refinement with tetrahedral symmetry to yield a final reconstruction with a global resolution of $FSC_{0.143}$=4.4 Å. The tetrahedral nature of these particles was also apparent in reconstructions with C1 symmetry. Notably, a small fraction of particle images from this 2D classification resembled hexamers (2,287 particle images) as seen in *SI Appendix* Figure S9. For the DegP:*E. coli* substrate dataset (Fig. 5C), 55,598 particle images were initially picked, extracted in 256×256-pixel boxes, and Fourier cropped to 128×128-pixel boxes for further analysis. Particle images were subject to 2D classification to filter the dataset (for example, to remove images from damaged particles), followed by *Ab initio* reconstruction and classification. After selecting the best *ab initio* class, corresponding to an 18-mer, 26,720 particle images were refined using C1 symmetry to a global resolution of $FSC_{0.143}$=10.1 Å. 3,828 particle images selected from 2D class averages corresponding to 24-mer particles were refined with octahedral symmetry to a final resolution of $FSC_{0.143}$=9.9 Å. Figures of maps and models were generated in *UCSF ChimeraX* (28).

Global fitting of DLS datasets for DegP self-assembly

(i) The two-pathways model

The DLS data, reporting on DegP oligomerization, were described in terms of a model containing two self-assembly pathways (Fig. 2B) (29, 30). As each of these pathways originates from the trimer state of DegP (M_3), the model can be built by initially considering each path individually and combining the resultant expressions. In path A, the association of DegP trimers to form canonical hexameric particles ($M_{6,A}$) is described by the macroscopic association constant,

$$K_A = \frac{[M_{6,A}]}{[M_3]^2} \ . \qquad [2]$$

We have described path B in terms of a phenomenological polymerization reaction where trimers associate into higher-order oligomers ($M_{3i,B}$) in a step-wise manner, with a single macroscopic association constant for each of the steps,

$$K_B = \frac{[M_{6,B}]}{[M_3]^2} = \frac{[M_{9,B}]}{[M_3][M_{6,B}]} = \frac{[M_{3i,B}]}{[M_3][M_{3(i-1),B}]} \ . \qquad [3]$$

As all steps within path B are assumed equivalent, it can be shown that

$$[M_{3i,B}] = K_B^{i-1}[M_3]^i \ . \qquad [4]$$

Note that the concentrations of each DegP oligomer, K_A, and K_B are functions of temperature but this notation is omitted above and in what follows for simplicity. We further assume that the temperature dependencies of K_A and K_B are given by the Van 't Hoff equation

$$K_i = K_{i,0} e^{\left[\frac{\Delta H_{i,0}}{R}\left(\frac{1}{T_0}-\frac{1}{T}\right)+\frac{\Delta C_{p,i}}{R}\left(\ln\left(\frac{T}{T_0}\right)+\frac{T_0}{T}-1\right)\right]} \qquad [5]$$

where $K_{i,0}$ and $\Delta H_{i,0}$ ($i \in (A,B)$) are the association constants and enthalpy changes at the reference temperature T_0 (298.15 K in our fits), $\Delta C_{p,i}$ is the heat capacity change (assumed to be

independent of temperature), and R is the ideal gas constant (8.3145 × 10^{-3} kJ mol^{-1} K^{-1}). The total concentration of DegP monomers (M_T) in the two-pathways model is given by

$$M_T = 3[M_3] + 6[M_{6,A}] + \sum_{i=2}^{\infty} 3i[M_{3i,B}] \quad [6]$$

where the first, second, and third terms account for the protomers within trimers, path A hexamers, and path B oligomers (hexamers and larger particles), respectively. The stoichiometric factors in each term (*3*, *6*, *3i*) reflect the total number of protomers in a given oligomeric state of DegP. Here it is assumed that trimers (i=1) can polymerize into infinitely large oligomers (i=∞) in path B since this leads to a useful approximation, as shown subsequently. When combined with Eqs. [2] and [4], Eq. [6] becomes

$$M_T = 6K_A[M_3]^2 + \sum_{i=1}^{\infty} (3i)K_B^{i-1}[M_3]^i . \quad [7]$$

Eq. [7] can be recast using dimensionless total monomer (X_T) and trimer (X_3) concentrations, as has been done previously (29, 30), along with the ratio of path A and B association constants, α, defined as

$$X_T = \frac{M_T K_B}{3}, \quad [8]$$

$$X_3 = K_B[M_3], \quad [9]$$

and

$$\alpha = \frac{K_A}{K_B}. \quad [10]$$

Substitution of Eqs. [8]-[10] into Eq. [7] yields the dimensionless expression

$$X_T = 2\alpha X_3^2 + \sum_{i=1}^{\infty} iX_3^i \quad [11]$$

and since $\sum_{i=1}^{\infty} i x^i$ converges to $\frac{x}{(x-1)^2}$ when $|x|<1$ (29, 30) it follows that

$$X_T = 2\alpha X_3^2 + \frac{X_3}{(X_3-1)^2}. \quad [12]$$

Notably, the condition $|X_3|<1$ is satisfied in all fits of the DLS data. The use of dimensionless concentration terms in Eq. [12] provides numerical stability during parameter optimization (in this case $K_{i,0}$, $\Delta H_{i,0}$, and $\Delta C_{p,i}$ ($i \in (A,B)$)), as described previously (29, 30).

The D_z values reporting on DegP oligomerization are related to the molar concentration and molecular weight of each species within the two-pathways model according to

$$D_z = \frac{m_3^2[M_3]D_3 + m_6^2[M_{6,A}]D_{6,A} + \sum_{i=2}^{N} m_{3i,B}^2[M_{3i,B}]D_{3i,B}}{m_3^2[M_3] + m_6^2[M_{6,A}] + \sum_{i=2}^{N} m_{3i,B}^2[M_{3i,B}]} \quad [13]$$

where m_{3i} is the molecular weight of the i^{th} diffusing species, and D_{3i} (path A and B notation omitted here) is the corresponding diffusion constant under ideal conditions (*i.e.* infinitely dilute concentrations). We note that D_z values can be influenced by non-ideality effects at high concentrations (13). We have not explicitly accounted for these in our analyses as we restricted our global fits to the biologically relevant range of M_T = ~10-200 μM where these effects are small. In order to calculate the D_z values for comparison with the experimental data in the global fitting routine, the temperature dependencies of D_{3i} must first be established. We obtained an estimate of $D_{3,0}$ and $D_{6,A}$ at low temperature (5 to 10 °C where path B oligomerization is minimal) from the S210A/Y444A DegP and S210A DegP experimental data, respectively, by linear extrapolation of the measured D_z values to zero M_T. These were found to be in good

agreement with the corresponding D_0 values calculated using HYDROPRO (31) where the hexamer crystal structure (PDB 3MH4) and one of its constituent trimers were supplied as inputs (see below). We note that the protease domain loops are missing in the hexamer crystal structure and in order to obtain accurate estimates of D_0 values using HYDROPRO, these loops were modeled using SWISSMODEL (32). The hydrodynamic radii for spherical particles can be calculated according to the Stokes-Einstein relationship

$$r_{3i} = \frac{k_B T}{6\pi\eta D_{3i}} \quad [14]$$

where k_B is the Boltzmann constant, T is the absolute temperature, η is the solution viscosity (assumed to be a function of temperature), and r_{3i} is the hydrodynamic radius. The η values for each DLS buffer used in this study were calculated as a function of temperature using SEDNTERP (http://www.rasmb.bbri.org/) and the resulting viscosities fit to a 3rd order polynomial to obtain empirical parameters that permit the calculation of a given buffer viscosity at any temperature within the experimental range of 5-50 °C. If the r_{3i} values (which are assumed to be temperature independent) are known, D_{3i} can be calculated over the entire temperature range. As the number of species in path B is large, it is not possible to obtain unique solutions for each of the r_{3i} values if they are treated as optimization parameters. Therefore, in order to constrain our global fits, we employed an empirical scaling law that has previously been applied in calculating diffusion constants of higher-order oligomers for self-assembling proteins (33, 34). In this approach D_{3i} values for i>1 are obtained by assuming that the ideal diffusion constants for DegP oligomers scale according to

$$D_{3i} = D_3 i^n \quad [15]$$

where n is an empirically determined scaling constant, which can range from -1/3 to -1/2 for globular proteins (34). We calculated the scaling constant for the path A hexamer using the experimentally determined D_3 and $D_{6,A}$ values at low temperature as

$$n = \frac{\ln\left(\frac{D_{6,A}}{D_3}\right)}{\ln(2)} \qquad [16]$$

which we found to range between -0.19 to -0.21 for temperatures between 5-10 °C, while from computations using HYDROPRO n = -0.227 was obtained. Fits performed using any value in this range gave nearly identical results and n = -0.227 was used in Figure 2. Since we could not resolve the path A and B hexamers in the DLS data, n = -0.227 was applied in calculating both of their ideal diffusion constants over the complete experimental temperature range. Furthermore, we could not establish the precise structural properties of the larger path B oligomers through DLS measurements alone. Given that higher-order substrate-bound DegP particles have been shown to be roughly spherical, and in the absence of more detailed information on the apo path B DegP particles, we used an n value of -1/3 corresponding to spherical oligomers (33) for all assemblies in path B (i>2).

Having established an approach to calculate the $[M_{3i}]$ and D_{3i} values that are required to evaluate D_z via Eq. [13], the fitting parameters $\zeta = \left[K_{A,0}, \Delta H_{A,0}, \Delta C_{p,A}, K_{B,0}, \Delta H_{B,0}, \Delta C_{p,B}\right]$ were obtained for the two-pathways model as summarized in the following computer pseudocode:

Initialize global fit with guesses for ζ.
    For T in experimental temperatures:
        For M_T *in experimental concentrations:*
            Solve $X_3(M_T, T, \text{parameters})$ *according to Eq* [12].
            Calculate $[M_{3i}](M_T, T, \text{parameters})$ *according to Eqs.* [4] *and* [9].
            Calculate $D_z(M_T, T, \text{parameters})$ *according to Eqs.* [13]-[16] *and save these values.*
    Calculate the fit quality statistic by comparing the calculated D_z *to the experimental* D_z.

If the fit quality is high and unchanging (optimized):
End the minimization and return the optimal parameters.
Otherwise:
Increment the parameters and repeat calculations.

The above minimization procedure was achieved using a program developed in-house, written in Python 3.7 and the lmfit fitting package (https://lmfit.github.io/lmfit-py/index.html). To solve for X_3 an additional minimization was performed whereby the real, positive root, ρ, of the polynomial expression given by Eq. [12], $0 \le \rho \le \frac{M_T K_B}{3}$, was obtained numerically using the *optimize.root* solver in Python's *SciPy* library by passing the values of X_T and α calculated from the current set of thermodynamic parameters, ζ, in addition to an initial guess for X_3. The optimal value of X_3 was then used to calculate [M_3] and all other [M_{3i}] up to i=N. The oligomer concentrations were combined with the D_{3i} to finally calculate the D_z values for comparison with the experimental data. The global fit proceeded by minimizing the weighted residual sum-of-squares (RSS_w)

$$RSS_w = \sum_{j=1}^{Z}\sum_{i=1}^{Y}\left(\frac{D_{z,sim.}\left(M_{T,i},T_j,\zeta\right) - D_{z,\exp t.}\left(M_{T,i},T_j\right)}{D_{z,\exp t.}\left(M_{T,i},T_j\right)}\right)^2 \qquad [17]$$

where the inner and outer sums run over the total number of sampled M_T and temperature points, $D_{z,sim.}\left(M_{T,i},T_j,\zeta\right)$ and $D_{z,\exp t.}\left(M_{T,i},T_j\right)$ are the simulated and experimental D_z values, respectively, and ζ is defined as above. In general, a total of 19 temperatures × 10 M_T values/temperature = 190 points were fit to extract the optimal thermodynamic parameters. From these the fractional population of subunits within each oligomeric species of the two-pathways model (Fig. 2C, D) was obtained as

$$P_{3i} = \frac{3i[M_{3i}]}{M_T}. \qquad [18]$$

It is worth noting that identical global fit results are produced with N = 50 and 100; increasing N beyond 50 only increases computation time (3N is the maximum oligomer size). This is due to the fact that oligomers consisting of 150 or more protomers are not appreciably populated under the experimental conditions we have considered (5-50 °C and ~10-200 μM M_T). In the limit that N is large (*e.g.* ≥50), the resultant oligomer concentrations approximate those in the case where trimers polymerize indefinitely with the K_B values measured here.

Estimates of the errors in the fitted thermodynamic parameters and confidence intervals for the Van 't Hoff plot of Figure 2B and were obtained through Monte-Carlo analyses (35). This first involved generating a synthetic "error-free" D_z dataset using the optimal thermodynamic parameters, comprising Z temperature points and Y M_T values. Subsequently, $Z \times Y$ total values (the size of the DLS dataset) were randomly drawn from a normal distribution with a standard deviation equal to the optimal fit RMSD and a mean of zero. These points were then added to the synthetic data and the perturbed dataset was fit using the same global fitting routine as for the experimental data. The fitted parameters were stored and this procedure was repeated 1000 times, with the errors taken as the standard deviation of the 1000 values for each thermodynamic parameter (*SI Appendix* Table S1). Confidence intervals in ln(K) values from the Van 't Hoff plot of Figure 2B were generated by simulating 1000 ln(K_A) and ln(K_B) curves using the set of generated Monte-Carlo parameters, with the plotted 95% confidence intervals shown (±1.96 standard deviation; light curves).

(ii) The canonical DegP particles model

We additionally fit the S210A DegP DLS data to a simpler model (in terms of the number of states) which assumes that DegP hexamers can form canonical 12-mers and 24-mers in a sequential association mechanism that involves only a single association constant. The association constant for this model is related to concentrations of oligomers as

$$K_1 = \frac{[M_{12}]}{[M_6]^2} = \frac{[M_{24}]}{[M_{12}]^2} \quad [19]$$

and the expression for M_T is

$$M_T = 6[M_6] + 12[M_{12}] + 24[M_{24}] \quad [20]$$

which can be recast using Eq. [19] as

$$M_T = 6[M_6] + 12K_1[M_6]^2 + 24K_1^3[M_6]^4 . \quad [21]$$

The global fit of this model to the experimental data was performed in a similar manner to that described for the two-pathways model above. Briefly, the concentrations of each DegP oligomer as a function of M_T, temperature, and input thermodynamic parameters were solved using a root finding algorithm and, along with D_{3i} values, used to generate ensemble averaged diffusion constants (D_z) for comparison with the experimental data. The RSS_w value was then calculated and the parameters were incremented until the parameter set corresponding to the minimum RSS_w value was obtained. This model fit the data poorly, as shown in *SI Appendix* Figure S5A.

We have also considered a slightly more complex model that includes trimer particles

$$M_3 \overset{K_1}{\longleftrightarrow} M_6 \overset{K_1}{\longleftrightarrow} M_{12} \overset{K_1}{\longleftrightarrow} M_{24} \quad [22]$$

from which it follows that

$$M_T = 3[M_3] + 6K_1[M_3]^2 + 12K_1^3[M_3]^4 + 24K_1^7[M_3]^8 \quad [23]$$

Fits of this model to the DLS data were also not satisfactory (*SI Appendix* Fig. S5B).

Global fitting of AUC datasets for DegP self-assembly

The AUC $s_{20,w}$ isotherms (corrected to 20 °C in water) for S210A DegP were globally fit in an analogous manner to the DLS datasets. Since the AUC data were primarily sensitive to the trimer and path A hexamer species, our analysis was simplified by considering only the path A arm of the two-pathways model. Thus, only two parameters are included in the fit, $\zeta_A = \left[K_{A,0}, \Delta H_{A,0} \right]$, with $\Delta C_{p,A}$ constrained to the optimal value obtained from DLS fitting (-6.9 kJ mol^{-1} K^{-1}), as the temperature range and sampling of the AUC isotherms (8, 20, and 30 °C) was not sufficient to obtain a unique solution for this parameter. Focusing on path A only,

$$\begin{aligned} M_T &= 3[M_3] + 6[M_{6,A}] \\ M_T &= 3[M_3] + 6K_A[M_3]^2 \end{aligned} \qquad [24]$$

from which we obtain

$$[M_3] = \frac{-3 + \sqrt{9 + 24K_A M_T}}{12K_A} \quad . \qquad [25]$$

and $[M_{6,A}]$ from Eq. [24]. The measured $s_{20,w}$ values are related to the concentrations of trimer and hexamer particles according to (36)

$$\begin{aligned} s_{20,w} &= \frac{3[M_3]}{M_T} s_3 + \frac{6[M_{6,A}]}{M_T} s_{6,A} \\ s_{20,w} &= P_3 s_3 + P_{6,A} s_{6,A} \end{aligned} \qquad [26]$$

where s_3 and $s_{6,A}$ are the sedimentation coefficients for the trimer and path A hexamer, respectively. We obtained experimental estimates for s_3 of ~6.1-6.2 S from independent AUC measurements of the S210A/Y444A DegP trimer mutant. A value of ~10.3 for s_6 was measured using S210A DegP at 8 °C and 21 μM M_T in high salt buffer (25 mM NaH_2PO_4, 200 mM NaCl,

1 mM EDTA, 0.02% sodium azide, pH 7.0), while a value of ~10.4 for s_6 was measured in low salt buffer (25 mM NaH_2PO_4, 25 mM NaCl, pH 7.0) at 20 °C and 21 μM M_T. These are consistent with the previously measured value of 10.4 (37). For consistency with the scaling law of Eq. [15] used in the analysis of the DLS data, we calculated s_6 in AUC global fits as

$$s_6 = s_3 2^{(1+n)} \quad [27]$$

where n = -0.227, yielding s_6 =10.48 S with s_3 set to 6.13 S. Global fits were carried out in a similar manner as for the DLS fitting to obtain estimates of $K_{A,0}$ and $\Delta H_{A,0}$ (*SI Appendix* Table S1). We note that estimates of $\Delta H_{A,0}$ from AUC analyses are somewhat more exothermic than from the DLS fits (-212 ± 15 versus -112 ± 27 kJ mol^{-1} respectively), although the AUC measurements were performed over a narrower temperature range (8-30 °C) and were sampled at only three temperatures (8, 20, and 30 °C). Additionally, the fits of the AUC isotherms do not take into account any of the oligomers in path B, albeit these are expected to be present in only very small concentrations under the conditions of our experiments. The AUC-based $\Delta H_{A,0}$ value must be considered an estimate, therefore. Nevertheless, the two orthogonal methodologies indicate that the formation of path A hexamers is associated with a large exothermic enthalpy change. Errors in the extracted thermodynamic parameters were taken from standard deviations obtained from 1000 Monte-Carlo global fitting iterations. The 95% confidence intervals for the $s_{20,w}$ isotherms, illustrated by the coloured bands in Figure 2G, were calculated from the 1000 fitted Monte-Carlo isotherms at each temperature (±1.96 standard deviation).

Estimating ΔC_p values for oligomerization via changes in solvent-accessible surface areas

ΔC_p values for (i) formation of the canonical closed hexamer from a pair of trimers (path A) and for (ii) the path B oligomerization reactions were estimated from changes in solvent-

accessible surface areas (ΔASA) calculated from the crystal structures of DegP (PDBs 3MH4 and 3CS0). These estimates take into account the changes in heat capacity that arise through hydration effects which are the dominant contribution to ΔC_p in protein folding (38). Importantly, they provided an orthogonal measure of ΔC_p which could be used to compare with values obtained from DLS global fitting analyses. In order to calculate accurate ΔASA values for the formation of the canonical hexamer (path A), it was first necessary to model the protease domain loops using SWISS-MODEL (32), as these are absent from the crystal structure due to their high mobility. The relevant ΔC_p value for path B oligomerization was estimated from the association of individual PDZ1:PDZ2' domains. This estimation was based on our NMR results (Fig. 4C&4D) showing that PDZ1:PDZ2' interactions drive apo-DegP assembly, as observed previously for oligomerization of substrate-bound DegP (39). As described in the text, a full complement of PDZ1:PDZ2' interactions are not present in the substrate-free state at physiological protein concentrations, but likely approximately two interactions per pair of trimers are established. The structure of the PDZ1:PDZ2' dimer in the apo-DegP state was modeled using the 24-mer cage crystal structure (PDB 3CS0) (39) that was solved in the presence of substrate. *ASA* values were obtained using the program *Naccess* (http://www.bioinf.manchester.ac.uk/naccess/) which individually calculates the polar and non-polar contributions to the *ASA* using an input structure. The polar and non-polar ΔASA values that are relevant for path A ($\Delta ASA_{j,A}$) and path B ($\Delta ASA_{j,PDZ1:PDZ2'}$) are given by

$$\begin{aligned} \Delta ASA_{j,A} &= ASA_{j,M_{6,A}} - 2\,ASA_{j,M_3} \\ \Delta ASA_{j,PDZ1:PDZ2'} &= ASA_{j,PDZ1:PDZ2'} - ASA_{j,PDZ1} - ASA_{j,PDZ2'} \quad j \in \{polar, nonpolar\} \end{aligned} \qquad [28]$$

ΔC_p values were subsequently calculated as

$$\Delta C_p = W_{polar}\Delta ASA_{polar} + W_{non-polar}\Delta ASA_{non-polar} \qquad [29]$$

where W_{polar} and $W_{non\text{-}polar}$ are empirical scaling factors from studies of the polar and non-polar contributions to ΔC_p in protein unfolding and upon transferring model compounds such as amino acids into water (40). We note that there is some variation in the published W_{polar} and $W_{non\text{-}polar}$ values (38, 40), though, in general, W_{polar} assumes negative values ranging from -0.38 to -1.09 J mol^{-1} K^{-1} $Å^2$ and $W_{non\text{-}polar}$ takes positive values from 1.17 to 2.14 J mol^{-1} K^{-1} $Å^2$. We therefore used the average of the values for W_{polar} and $W_{non\text{-}polar}$ (40) in calculating estimates of ΔC_p, obtaining values of -13.1 and -0.9 kJ mol^{-1} K^{-1} for the formation of the canonical hexamer and a single PDZ1:PDZ2' interaction respectively. The signs of these values are consistent with the predominant burial of hydrophobic surface area that commonly occurs in protein folding and binding reactions (41). Their relative magnitudes are also in agreement with the extent of buried surface area in the formation of each interaction, *i.e.* forming a hexamer buries more surface area than forming two PDZ1:PDZ2' heterodimers. Notably, these values are in reasonable agreement with those obtained for path A and B via global fits of the DLS data (*SI Appendix*, Table S1).

Analysis of NMR data

(i) Calculation and fitting of chemical shift perturbation (CSP) profiles (*Figs. 3C&3D, 4C*)

Amide and methyl correlations were used to monitor the interaction of isolated PDZ1 and PDZ2 domains in a series of correlation spectra in which PDZ1 peak positions were monitored as a function of the addition of unlabeled PDZ2 (see *NMR measurements* for the isotope labeling schemes used). The majority of the correlations titrated linearly between their respective free PDZ1 and PDZ1:PDZ2-bound endpoints with little changes to lineshapes. We subsequently fit the obtained chemical shift changes to a simple two-state binding model assuming fast exchange

on the NMR chemical shift timescale so as to obtain estimates of the PDZ1:PDZ2 binding affinity. CSP profiles as a function of PDZ2 concentration were calculated according to

$$\Delta\nu_{obs.} = \sqrt{\Delta\nu_i^{\,2} + \Delta\nu_j^{\,2}} \qquad [30]$$

where $\Delta\nu_{obs.}$ is the observed CSP for a given correlation (Hz) and $\Delta\nu_i$ is the frequency change in dimension *i*, at a given PDZ2 concentration relative to the frequency prior to addition of PDZ2. These profiles (Fig. 3C&3D and Fig. 4C) were subsequently fit using the following expressions

$$\begin{aligned} &P + L \rightleftarrows PL \\ &K_A = \frac{[PL]}{[P][L]} \\ &P_T = [P] + [PL] \\ &L_T = [L] + [PL] \end{aligned} \qquad [31]$$

where *P*, *L*, and *PL* denote PDZ1, PDZ2, and PDZ1:PDZ2, respectively, K_A is the association constant, and P_T and L_T are the total PDZ1 and PDZ2 concentrations respectively. These expressions can be rearranged to yield

$$[L] = \frac{(K_A L_T - K_A P_T - 1) + \sqrt{(K_A P_T - K_A L_T + 1)^2 + 4K_A L_T}}{2K_A} \qquad [32]$$

from which the concentrations of all other species are calculated (Eq. [31]). The fraction of PDZ1:PDZ2 complex, $F_{PDZ1:PDZ2}$, is defined as

$$F_{PDZ1:PDZ2} = \frac{[PL]}{P_T} \qquad [33]$$

and is related to $\Delta\nu_{obs.}$ according to

$$\Delta\nu_{obs.} = \Delta\nu_{PDZ1:PDZ2} F_{PDZ1:PDZ2} \qquad [34]$$

where $\Delta\nu_{PDZ1:PDZ2}$ is the maximum CSP corresponding to the completely PDZ2-bound endpoint. The fitting procedure was performed using an in-house Python script and the lmfit minimization

package (https://lmfit.github.io/lmfit-py/index.html). The fit output consisted of K_A and $\Delta\nu_{PDZ1:PDZ2}$ values for each CSP profile. The K_As given in Figure 3 are the mean ± SD of the set of individually fitted K_A*s*.

(ii) Lineshape analyses (*Fig. 3E&3F*)

During the course of the PDZ1-PDZ2 titrations a number of cross-peaks broadened (for example, correlations derived from the amides of A284 and N273). Titration data for these peaks were globally fit using the 2D lineshape fitting program TITAN (42) in order to extract estimates of the association (k_{on}) and dissociation (k_{off}) rates for the PDZ1:PDZ2 interaction according to a two-state $PDZ1 + PDZ2 \rightleftarrows PDZ1{:}PDZ2$ binding model. The fitting process was performed by simulating HSQC spectra (using the default pulse scheme supplied by TITAN) for the selected correlations as a function of PDZ2 concentration and comparing these with the experimental data in a minimization routine. Experimental spectra supplied to TITAN were processed using an exponential window function with 5 Hz line broadening in each dimension, with the intrinsic transverse relaxation rates of the PDZ1 domain amide spins in each of the two dimensions given as $R_{2,0} = R_{2,apparent} - R_{2,window}$ for the fits (for both bound and free states). A set of optimized parameters including PDZ1 peak centers as a function of PDZ2 concentration, "free" and "bound" $R_{2,0}$ values for both ^{1}H and ^{15}N/^{13}C spins, K_D (dissociation), and k_{off} were obtained, from which $k_{on} = k_{off}/K_D$ was calculated. Errors in the optimal parameters reported in the main text were obtained through a boot-strap analysis built into TITAN. The fitted lineshapes in Figure 3E&3F were visualized using nmrglue (18).

(iii) Calculating fraction bound profiles for S210A DegP (*Fig. 4C*)

As described in the text, insight into the mechanism of path B oligomer formation can obtained by comparing the fraction of PDZ1:PDZ2' interactions that form as a function of protein concentration in the S210A DegP titration, $F_{PDZ1:PDZ2'}$, with predictions based on simple models of trimer assembly (see next section). $F_{PDZ1:PDZ2'}$ is given by

$$F_{PDZ1:PDZ2'} = \frac{\Delta\nu_{obs.}}{\Delta\nu_{PDZ1:PDZ2'}} \quad [35]$$

where $\Delta\nu_{obs.}$ is defined as the difference in peak position at a given DegP protein concentration from the corresponding correlation in a spectrum of the S210A/Y444A trimer mutant at 100 μM M_T (which approximates the infinite dilution case), and $\Delta\nu_{PDZ1:PDZ2'}$ is given by the chemical shift differences between fully bound and completely unbound PDZ1 correlations obtained from fits of CSP profiles in a titration of the U-^{2}H, ILVM-$^{13}CH_3$ PDZ1 domain with U-^{2}H PDZ2.

(iv) Estimating the influence of DegP oligomerization on the apparent NMR peak position

The calculation of the fraction bound profiles for the S210A DegP titration relies on the assumption that the apparent frequencies of the methyl correlations accurately reflect the thermodynamic fraction of PDZ1:PDZ2' interactions throughout the titration series (*i.e.* infinitely fast exchange limit). Peak positions used for analysis (Fig. 4B) titrated linearly between their respective trimer and PDZ1:PDZ2' bound endpoints, and the similar fraction bound profiles obtained for methyl groups with different CSPs between free and bound states (Fig. 4C, circles) suggests that the DegP trimers and higher-order oligomers are in relatively rapid exchange (k_{ex} >~1800 s^{-1}, based on maximum CSPs). We have examined the influence of exchange rates and differences in transverse relaxation rates of interconverting states on fraction bound profiles through a series of simulations using the Bloch-McConnell equations (43), where

for simplicity we have assumed a $M_3 \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} M_6$ model to mimic the exchange beween trimers and path B hexamers (*SI Appendix* Fig. S7A). The magnetization exchange matrix used to perform these simulations has been presented in detail elsewhere (44–46). In all simulations values of M_T were chosen to coincide with DegP concentrations used experimentally and a $K_A = k_1/k_{-1}$ value of 1.3×10^3 M^{-1} was selected, as obtained from the titration of the PDZ domains at 50 °C (Fig. 3F). Values of k_1 and k_{-1} were increased, from 3.9×10^5 M^{-1} s^{-1} and 300 s^{-1} in (*SI Appendix* Fig. 7B) to 3.9×10^7 M^{-1} s^{-1} and 3×10^4 s^{-1} (*SI Appendix*, Fig. 7D; the upper limit values for the isolated PDZ1:PDZ2 domain interaction) in one order of magnitude steps. Simulations were performed assuming $R_2(M_3) = 23$ s^{-1}, the average ^{13}C R_2 estimated from the L272, L276, and M280 methyl resonances in the U-^{2}H,^{13}C-ILVM S210A/Y444A DegP trimer mutant ^{13}C-^{1}H HMQC spectrum, and $R_2(M_6) = 2R_2(M_3)$. The resultant time-domain datasets were Fourier-transformed, and the frequency domain spectra shown in *SI Appendix* Figure S7B-D as a function of M_T. Notably, the titration data for the S210A DegP (*SI Appendix* Fig. S7F) are most consistent with the simulations of fast-intermediate to fast exchange kinetics (*SI Appendix* Fig. S7C&7D) as described above, where peaks titrate linearly between free and PDZ1:PDZ2' bound endpoints, without the appearance of additional correlations that are characteristic of slow chemical exchange. A plot of the apparent fraction of hexamer ($P_{6,app.}$) calculated from the titration series with a minimum k_{ex} of ~3000 s^{-1} (SI Appendix Fig. S7C) *vs* the P_6 values used in the simulations is shown in SI Appendix Figure S7E. Importantly, the deviations between the apparent and simulated P_6 values are at most only ~2-3% and occur at the smallest M_T values, with even better agreement at higher M_T where the exchange time-scale increases. Additional simulations performed with a larger $R_2(M_6)$ (=$4R_2(M_3)$) produced similarly small deviations between the apparent and simulated P_6 values. Although the DegP oligomerization pathway is

more complex than the simple trimer-hexamer equilibrium considered here, the present analysis nevertheless provides a measure of confidence that the fraction bound values obtained experimentally are only weakly dependent on differential relaxation of the exchanging particles.

(v) Modeling of fraction bound profiles for S210A DegP self-assembly (*Fig. 4C&4D*)

Profiles of the fraction of PDZ1:PDZ2' interactions formed as a function of DegP concentration have been assessed using very simple models of DegP path B oligomerization (Fig. 4D). Although the true assembly behavior of DegP is most certainly more complicated than the schemes considered here, these models, nevertheless can provide mechanistic insights, as described in the main text. In what follows, we establish simple relationships for the fraction of bound PDZ1 domains in the different hypothetical DegP oligomers that are illustrated in Figure 4D.

a. Figure 4D blue scheme

The formation of the i^{th} (linear) oligomeric species occurs through the edge-wise association of a trimer (Fig. 4D blue triangle) and an i-1 (linear) oligomer via inter-trimer PDZ1:PDZ2' domain interactions (Fig. 4D black circles). Only one PDZ1:PDZ2' interaction connects each pair of trimers (black circle connecting triangles; note that all PDZ domains form interactions in the substrate-bound cages (39)). It is further assumed, for simplicity, that trimers associate with other oligomers in a bi-directional, linear manner. We selected a maximum size of 8 associated trimers for these schemes (corresponding to a 24-mer DegP particle of ~1.1 MDa molecular mass) as our NMR measurements become progressively less sensitive to particles as they grow beyond this size.

We note that

$$M_T = \sum_{i=1}^{8} 3i[M_{3i}]$$
$$9K_{PDZ1:PDZ2'} = \frac{[M_6]}{[M_3]^2} \qquad [36]$$
$$3K_{PDZ1:PDZ2'} = \frac{[M_9]}{[M_6][M_3]} = \frac{[M_{12}]}{[M_9][M_3]} = \ldots = \frac{[M_{24}]}{[M_{21}][M_3]}$$

where $K_{PDZ1:PDZ2'}$ is the microscopic association constant for the formation of a single PDZ1:PDZ2' interaction, which we treated as a fitting parameter in the analysis of the NMR titration data. The multiplicative factor of 9 for $K_{PDZ1:PDZ2'}$ in the expression for the association of two trimers (where $9K_{PDZ1:PDZ2'}$ is a macroscopic equilibrium constant) takes into account the 3-fold rotational symmetry of each trimer (3 × 3 = 9), *i.e.* a trimer can associate with any other oligomer through a single PDZ1:PDZ2' interaction formed at any of its three edges (indicated by white stars in Fig. 4D). The multiplicative factor of 3 for the association constants governing the formation of all other oligomers accounts for the rotational symmetry of the trimer and the fact that trimers can associate (dissociate) at either of the two ends of higher-order oligomers (3 × (2 ways to associate/2 ways to dissociate) = 3). The set of expressions in Eq. [36] can be solved numerically to yield the concentrations of each oligomeric state, from which the fractional populations of subunits within each oligomer are given by Eq. [18].

The fraction of PDZ1:PDZ2' interactions *vs.* M_T can then be calculated as (47)

$$F_{PDZ1:PDZ2'} = \sum_{i=2}^{N} w_{B,3i} P_{3i} \qquad [37]$$

where $w_{B,3i}$ is fraction of bound PDZ1:PDZ2' interactions in the *3i*-mer particle and P_{3i} is the fraction of protomers that are contained within the *3i*-mer. In Eq [37] the sum starts from *i=2* since trimers do not contribute to the PDZ1:PDZ2'-bound signal, and *N* is the number of trimers in the largest species considered (*N*=8 for the 24-mer DegP). Values of $w_{B,3i}$ are determined by

counting the number of PDZ1:PDZ2' interactions formed in each oligomer and dividing by the total number of possible interactions (equal to the total number of PDZ1 domains in the oligomer; *3i* is the maximum number of PDZ1 interactions in particle M_{3i}). For example, a trimer has 0 bound PDZ1 domains and the potential for three bound PDZ1 interactions (where the PDZ1 domains derive from the trimer in question), thus $w_{B,3} = 0/3 = 0$. For the hexamer, a single PDZ1:PDZ2' interaction is formed out of a total of 6 possible interactions, so that $w_{B,6} = 1/6$. Each subsequent addition of a trimer in this scheme leads to one further PDZ1:PDZ2' interaction and an additional two non-interacting PDZ1 domains. Thus, $w_{B,3i} = \frac{i-1}{3i}$, and $w_{B,\infty} = 1/3$, as expected, since each trimer:trimer interface supports only a single PDZ1:PDZ2' interaction in this scheme. Eq. [37] becomes

$$F_{PDZ1:PDZ2'} = \sum_{i=2}^{8} \frac{i-1}{3i} P_{3i} \qquad [38]$$

where the $w_{B,3i}$ coefficients do not reach the theoretical maximum value of 1/3 as *N* is small. In this case, the maximum value of $F_{PDZ1:PDZ2'} = \sim 0.29$ is achieved if M_{24} becomes fully populated. We note that this model cannot capture the experimental data for S210A DegP as $F_{PDZ1:PDZ2'}$ rapidly increases above 0.3 at low M_T. Eqs. [37]-[38] lead to the important conclusion that $F_{PDZ1:PDZ2'}$ will never reach 1 for models of DegP self-assembly where the full complement of PDZ1:PDZ2' interactions are not formed. This is in contrast to the formation of canonical DegP cages where all PDZ1:PDZ2' interactions are satisfied (green scheme).

b. Fig. 4D pink scheme

This model is identical to the one discussed immediately above except that it is assumed that a pair of interacting trimers are connected by two PDZ1:PDZ2' interactions, rather than one, so that a total of two PDZ1:PDZ2' contacts are made per trimer interface. We can write,

$$18K_{PDZ1:PDZ2'} = \frac{[M_6]}{[M_3]^2}$$
$$6K_{PDZ1:PDZ2'} = \frac{[M_9]}{[M_6][M_3]} = \frac{[M_{12}]}{[M_9][M_3]} = \ldots = \frac{[M_{24}]}{[M_{21}][M_3]} \quad [39]$$

where the scaling factors for $K_{PDZ1:PDZ2'}$ are multiplied by factors of two relative to the corresponding equations given for the previous model to take into account that there are two PDZ1:PDZ2' interactions formed at each trimer:trimer interface. For example, the scaling factor in the expression for formation of a hexamer contains the rotational symmetry factor of 9 (for the association of two trimers) × 2 PDZ1:PDZ2' interactions = 18. The binding affinity for the first and second PDZ1:PDZ2' interactions are assumed to be the same. In this model $w_{B,3i} = \frac{2(i-1)}{3i}$ ($w_{B,\infty} = 2/3$) so that

$$F_{PDZ1:PDZ2'} = \sum_{i=2}^{8} \frac{2(i-1)}{3i} P_{3i} \quad , \quad [40]$$

where the maximum fraction of PDZ1:PDZ2' interactions formed is ~0.58 in the event that M_{24} is completely populated. This model provides a reasonable fit to the experimental fraction bound profiles for S210A DegP in Figure 4C over the physiological protein concentration range and extending until approximately 1.2 mM. The fitted $K_{PDZ1:PDZ2'}$ value of (4.0 ± 0.3) × 10^3 M^{-1} is approximately a factor of four larger than the association constant that has been measured for the interaction between isolated PDZ1 and PDZ2 domains, likely reflecting avidity effects in the context of PDZ interactions in DegP. Notably, the macroscopic constant $6K_{PDZ1:PDZ2'}$ = 24 × 10^3 M^{-1}, 50 °C, (Eq. [39]) for formation of M_9 and larger oligomers that is obtained from fits of the fraction bound NMR data (U-^{2}H,^{13}C-ILVM labelled S210A DegP in D_2O buffer, 200 mM added NaCl) to the simple linear chain model with two interactions for each pair of trimers (Fig. 4D,

pink) is in reasonable agreement with the phenomenological macroscopic association constant derived for path B, K_B = 34 × 10^3 M^{-1}, 50 °C, (Fig. 2B, 6A), from analysis of the DLS data (unlabelled S210A DegP in H_2O buffer, 200 mM added NaCl) using the two-pathways model.

c. Fig. 4d green scheme

Here it is assumed that trimers spontaneously assemble into canonical 12- and 24-mer cages where all PDZ1:PDZ2' contacts are satisfied and no other species is populated, so that $w_{B,3i}$ = 1 and $F_{PDZ1:PDZ2'}$ = 1.

Supplementary Figures

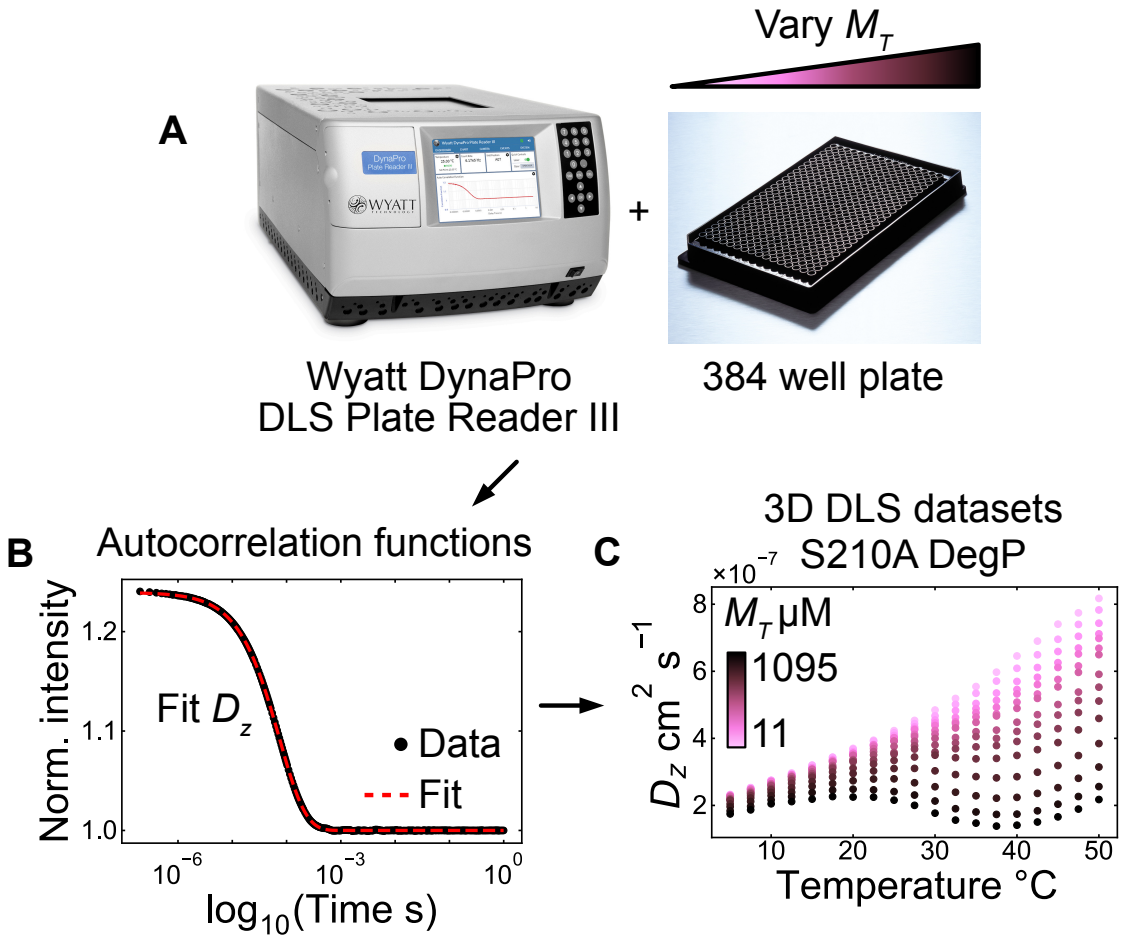
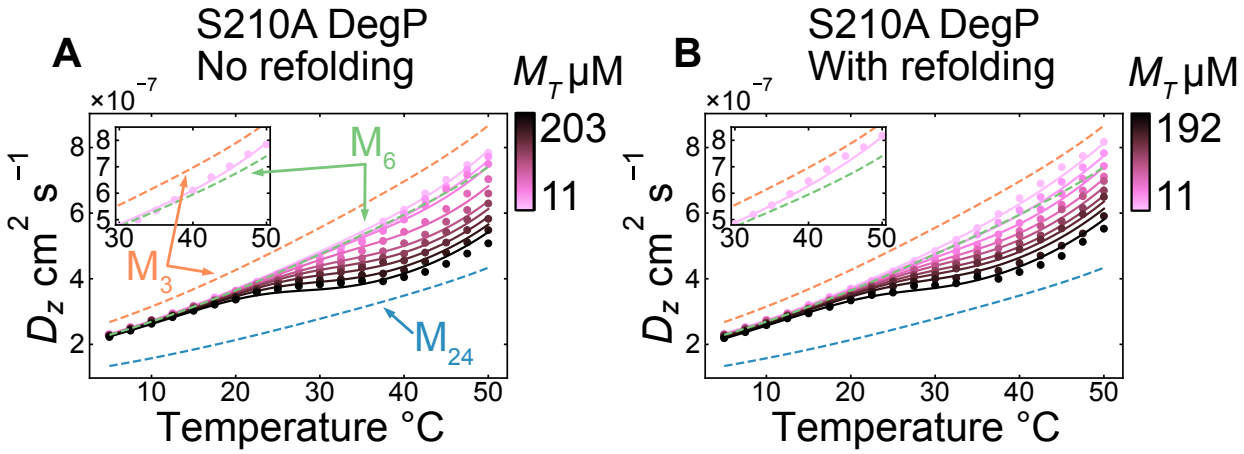


Figure S1. Measuring 3D DLS datasets. (A) Samples are prepared for high-throughput DLS analysis in a well plate format by varying an experimental condition such as M_T or ionic strength over the wells in each plate. Light scattering data for each well are measured using a Wyatt DynaPro DLS plate reader III as a function of temperature. The images of the DLS instrument and the 384 well plate were obtained from the Wyatt Technologies website (https://www.wyatt.com/products/instruments/dynapro-plate-reader-iii-summary-landing-page.html) and the Corning products e-catalog (https://ecatalog.corning.com/life-sciences/b2c/CA/en/Microplates/Assay-Microplates/384-Well-Microplates/Corning%C2%AE-384-well-Black-Clear-and-White-Clear-Bottom-Polystyrene-Microplates/p/3540), respectively.

(B) The resultant autocorrelation functions are numerically fitted to extract D_z values. (C) 3D DLS datasets are obtained as plots of D_z versus temperature for each well on the DLS plate. The resultant D_z profiles are color-coded according to M_T, varying from 11 μM (pink) to 1095 μM (black) and can be fit to a model of DegP oligomer assembly.

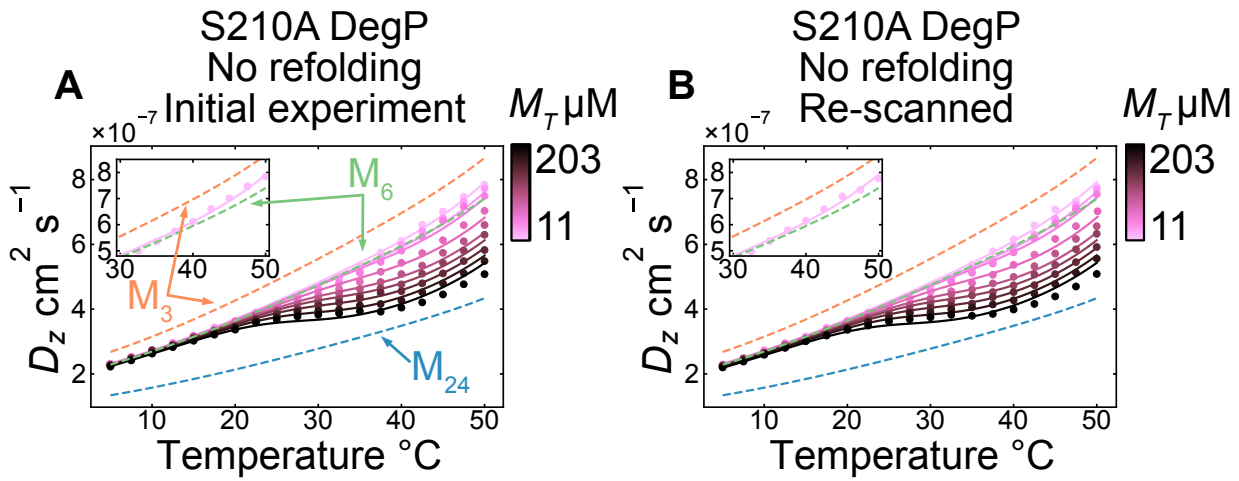


Two pathways global fit parameters

$K_{A,0}$ = (2.5 ± 1.5) × 10^6	$K_{A,0}$ = (0.9 ± 0.4) × 10^6
$\Delta H_{A,0}$ = -107.9 ± 38.5	$\Delta H_{A,0}$ = -112.0 ± 27.2
$\Delta C_{p,A}$ = -5.4 ± 1.9	$\Delta C_{p,A}$ = -6.9 ± 1.5
$K_{B,0}$ = (2.0 ± 0.6) × 10^5	$K_{B,0}$ = (1.2 ± 0.3) × 10^5
$\Delta H_{B,0}$ = -26.3 ± 21.5	$\Delta H_{B,0}$ = -24.6 ± 14.9
$\Delta C_{p,B}$ = -1.3 ± 1.3	$\Delta C_{p,B}$ = -1.3 ± 0.9

Figure S2. Refolding S210A DegP preserves its oligomerization properties. 3D DLS data for S210A DegP (circles) purified (A) without and (B) with a Gdn-HCl unfolding/refolding step, shown with overlaid curves globally fitted according to the two-pathways model (lines). Note that while the data in panels (A) and (B) are sampled over a similar concentration range (~10-

200 µM M_T), sampling has occurred at different concentrations intervals. Parameters extracted from global fits to each dataset are shown below each panel, given as the mean ± SD from 1000 Monte Carlo fitting iterations. $K_{i,0}$, $\Delta H_{i,0}$, and $\Delta C_{p,i}$ ($i \in (A,B)$) are given in M^{-1}, kJ mol^{-1}, and kJ mol^{-1} K^{-1} respectively.

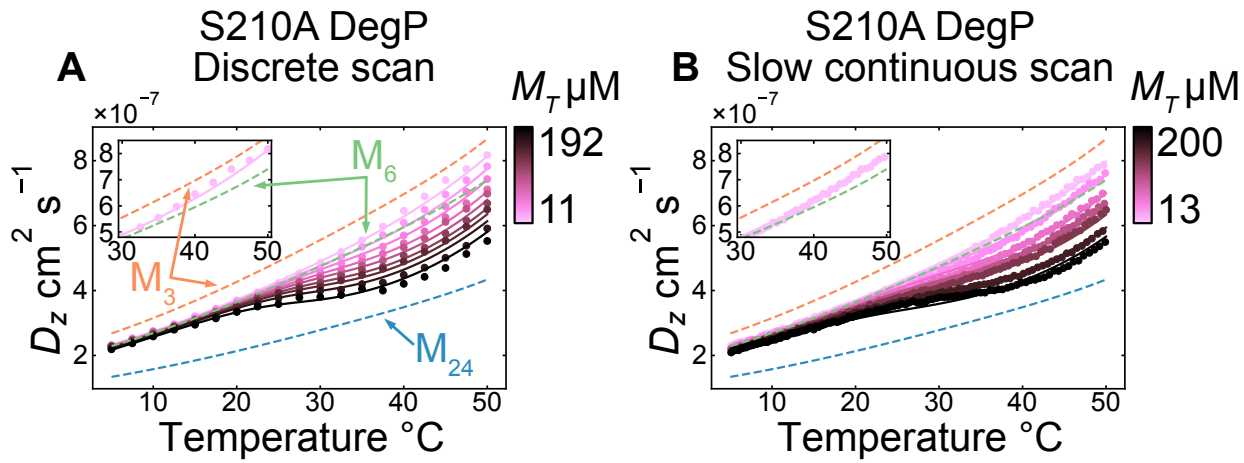


Two pathways global fit parameters

$K_{A,0}$ = (2.5 ± 1.5) × 10^6	$K_{A,0}$ = (2.5 ± 1.3) × 10^6
$\Delta H_{A,0}$ = -107.9 ± 38.5	$\Delta H_{A,0}$ = -107.5 ± 32.9
$\Delta C_{p,A}$ = -5.4 ± 1.9	$\Delta C_{p,A}$ = -6.3 ± 1.7
$K_{B,0}$ = (2.0 ± 0.6) × 10^5	$K_{B,0}$ = (2.0 ± 0.5) × 10^5
$\Delta H_{B,0}$ = -26.3 ± 21.5	$\Delta H_{B,0}$ = -26.7 ± 18.1
$\Delta C_{p,B}$ = -1.3 ± 1.3	$\Delta C_{p,B}$ = -1.5 ± 1.1

Figure S3. S210A DegP oligomerization is reversible. (A,B) 3D DLS datasets for apo-DegP (not refolded, circles) where the samples in the wells for the initial experiment (A) were re-scanned from 5-50 °C after cooling (B). Each dataset is shown with overlaid curves globally fitted according to the two-pathways model (lines). Parameters extracted from global fits to each

dataset are shown below each panel, given as the mean ± SD from 1000 Monte Carlo fitting iterations. $K_{i,0}$, $\Delta H_{i,0}$, and $\Delta C_{p,i}$ ($i \in (A,B)$) are given in M^{-1}, kJ mol^{-1}, and kJ mol^{-1} K^{-1} respectively.



Two pathways global fit parameters

$K_{A,0}$ = (0.9 ± 0.4) × 10^6	$K_{A,0}$ = (0.9 ± 0.3) × 10^6
$\Delta H_{A,0}$ = -112.0 ± 27.2	$\Delta H_{A,0}$ = -100.4 ± 21.0
$\Delta C_{p,A}$ = -6.9 ± 1.5	$\Delta C_{p,A}$ = -5.7 ± 1.2
$K_{B,0}$ = (1.2 ± 0.3) × 10^5	$K_{B,0}$ = (1.4 ± 0.2) × 10^5
$\Delta H_{B,0}$ = -24.6 ± 14.9	$\Delta H_{B,0}$ = -25.6 ± 10.5
$\Delta C_{p,B}$ = -1.3 ± 0.9	$\Delta C_{p,B}$ = -1.2 ± 0.8

Figure S4. Assessing the influence of the temperature scanning rate on apo-DegP oligomerization. (A,B) 3D DLS data for S210A DegP (circles) collected using discrete temperature sampling (with ~10-15 minutes temperature equilibration prior to measurement, corresponding to an effective temperature ramp of ~0.25-0.17 °C min^{-1} (~4-6 min $°C^{-1}$), (A)) or

with a slow continuous temperature ramp rate of 0.03 °C min^{-1} (~33.3 min $°C^{-1}$, (B)). Note that the slow continuous ramping rate in (B) allows for the collection of many more data points than in the discrete case in (A). Only every 3rd data point in (B) is shown. In addition, while the data in panels (A) and (B) are sampled over a similar concentration range (~10-200 μM M_T), sampling has occurred at different intervals. In each panel, curves globally fitted according to the two pathways model (solid lines) are shown. Parameters extracted from global fits to each dataset are shown below each panel, given as the mean ± SD from 1000 Monte Carlo fitting iterations. $K_{i,0}$, $\Delta H_{i,0}$, and $\Delta C_{p,i}$ ($i \in (A,B)$) are given in M^{-1}, kJ mol^{-1}, and kJ mol^{-1} K^{-1} respectively.

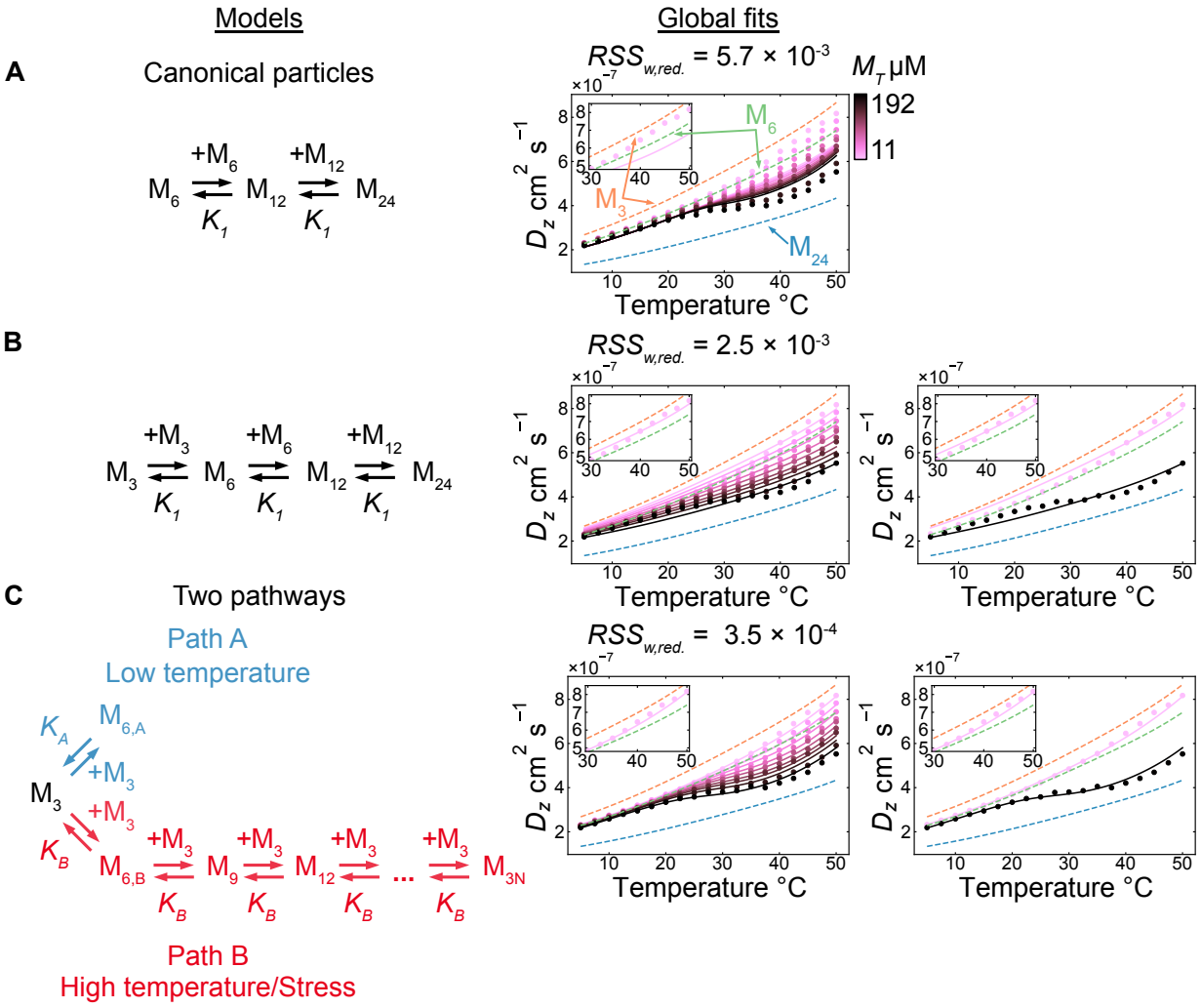


Figure S5. Comparison of self-assembly models in global fits of 3D DLS datasets for S210A DegP. (A) A model based on the formation of canonical DegP particles, whose structures have been previously solved in the apo- and substrate-engaged states (37, 39, 48). (B) The model in (A) modified to account for trimer particles. (C) The two pathways model. In each case, the corresponding global fit of the model (lines) to the S210A DegP DLS data (circles) is shown to the right, with the fit quality indicated in the form of the reduced RSS_w ($RSS_{w,red.}$) calculated as $RSS_{w,red} = RSS_w$ (Eq. [17])/(# points - # fitted parameters). In (B) and (C), additional panels to the right are shown for the D_z profiles and their fitted curves at the lowest and highest M_T values (11

and 192 μM respectively) in order to highlight differences in the fits of these two models. The two pathways model provides the best fit to the experimental data (smallest $RSS_{w,red.}$).

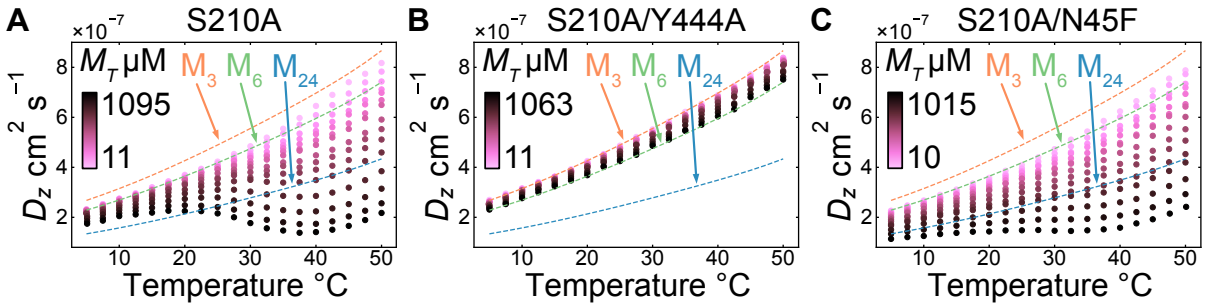


Figure S6. Validating the two-pathways model using mutations. (A) 3D DLS dataset for S210A DegP. (B) 3D DLS dataset for S210A/Y444A DegP (PDZ1:PDZ2' domain interaction mutant). (C) 3D DLS dataset for S210A/N45F DegP (protease:protease' domain interaction mutant). Note that DLS data for these DegP constructs are shown up to ~1 mM in M_T (a non-biological concentration, see Figure 1C) to emphasize the differences in their oligomeric properties.

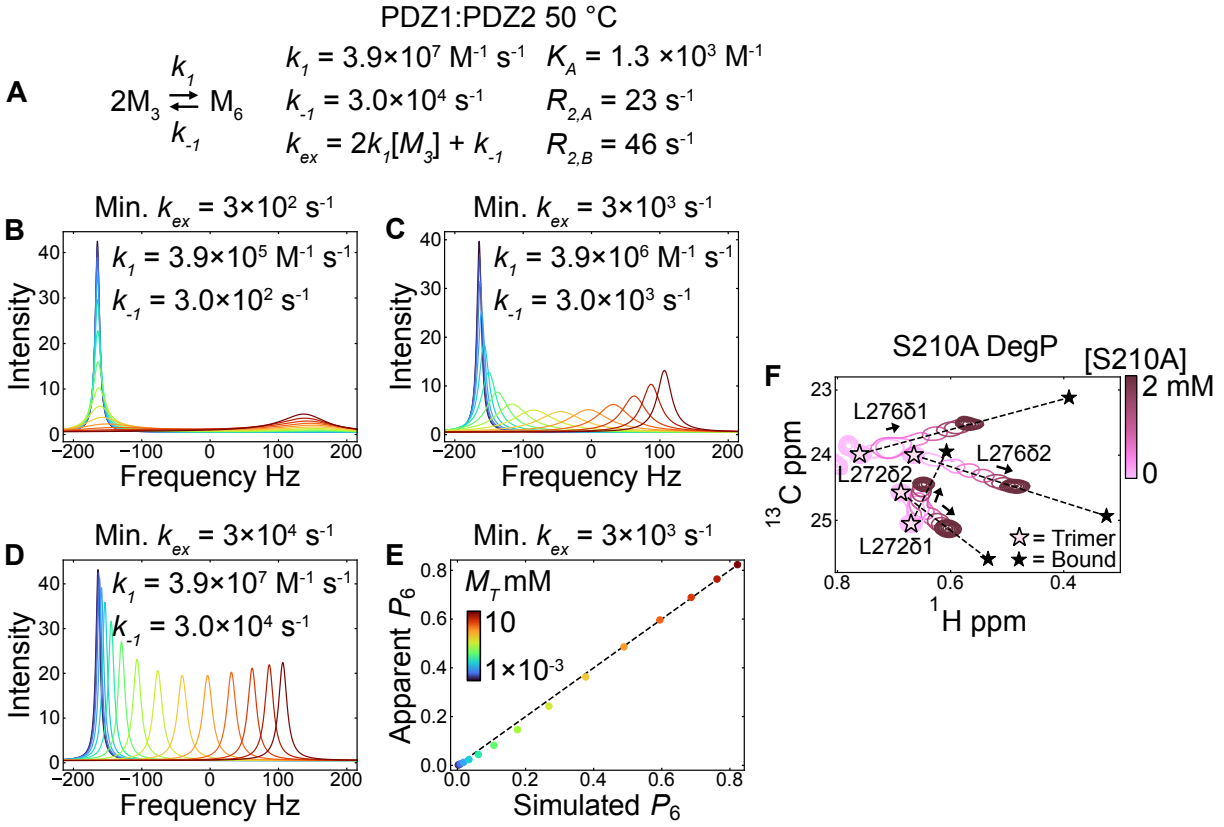


Figure S7. Bloch-McConnell simulations of exchange between trimers and path B hexamers and the influence of exchange rates on the apparent fast exchange peak position. (A) Thermodynamic model and simulation affinity and rate constant parameters (based on fits of experimental data from the titration of individual PDZ domains, 50 °C). (B) Simulations with a minimum k_{ex} of ~300 s^{-1} produce NMR data in the slow exchange regime (free and bound peaks are observed), a result inconsistent with our observed S210A DegP titration profile (see panel (F)). (C, D) Simulations with k_{ex} 3000 s^{-1} (C) and 30,000 s^{-1} (D) produce NMR data in the fast-intermediate to fast exchange regime, which are consistent with the S210A DegP titration (panel (F)). (E) Apparent fraction of the hexamer state ($P_{6,app.}$) calculated from the simulations in (C) where the kinetics are in the fast-intermediate exchange regime *vs* P_6 values used in the simulation

(Simulated P_6). The apparent P_6 was calculated as $P_{6,app.} = \frac{\Delta\nu_{i,app.}}{\Delta\nu_{max}}$, where $\Delta\nu_{i,app.}$ is the displacement of the peak after the i^{th} titration point from its initial position, and $\Delta\nu_{max}$ is the chemical shift difference between a reporter peak in M_3 and M_6. Deviations of only a few percent are observed and these occur at the lowest M_T values (~1-100 μM) where k_{ex} is smallest. In all simulations, ν_A (M_3) = -165 Hz, ν_B (M_6) = 165 Hz, $R_{2,A}$ = 23 s^{-1} (the average ^{13}C R_2 estimated from the L272, L276, and M280 methyl resonances in the U-^{2}H,^{13}C-ILVM S210A/Y444A DegP trimer mutant ^{13}C-^{1}H HMQC spectrum), and $R_{2,B}$ = 46 s^{-1}. The maximum frequency difference $\Delta\nu_{max} = \nu_B - \nu_A$ was selected based on the observed CSPs for the L276 methyls in the U-^{2}H,^{13}C-ILVM S210A DegP and U-^{2}H,^{13}C-ILVM PDZ1:U-^{2}H PDZ2 domain titrations from Figure 4A. The titration for S210 A DegP is shown in (F) for reference where pink stars indicate the trimer peak positions, and black stars show the fully PDZ1:PDZ2 bound endpoints.

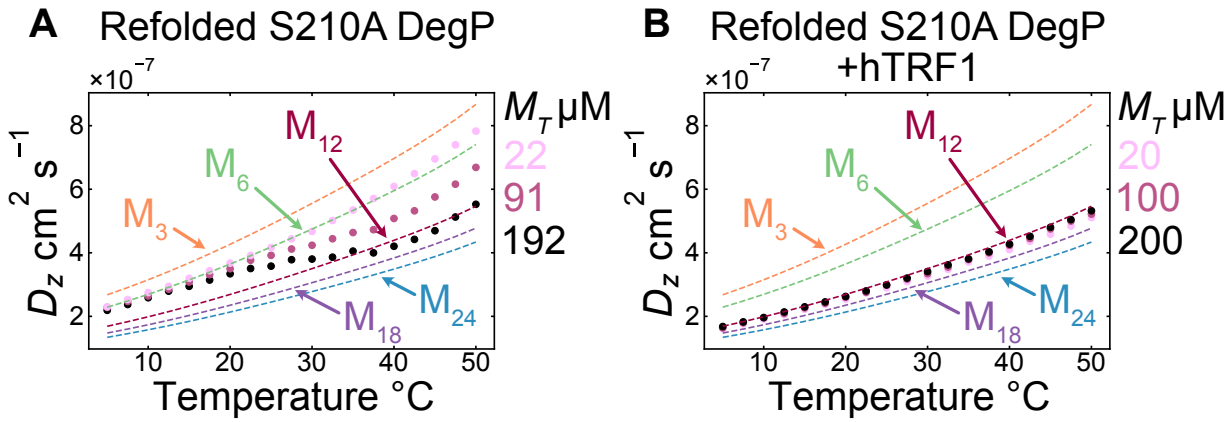


Figure S8. Substrate binding remodels the DegP free energy landscape. DLS profiles for apo S210A DegP are sensitive to M_T (A, from Fig. 2A), as a distribution of oligomers is produced that varies with concentration, while DLS profiles collected for refolded S210A DegP in the presence of 400 μM hTRF1 (B) are insensitive to M_T over the biological concentration range of

DegP. The data at each M_T in (B) have been corrected to a first approximation for the free concentration of hTRF1 assuming that DegP assembles exclusively into 12-mer cages and that all DegP protomers are saturated with hTRF1 (1:1 binding), according to $D_z = D_{z,measured} - D_{z,hTRF1}$, where $D_{z,hTRF1} = \frac{m_{hTRF1}[hTRF1]^2 D_{hTRF1}}{m_{hTRF1}[hTRF1]^2 + m_{12}[M_{12}]^2}$ and m_{hTRF1} and m_{12} are the molecular weights of hTRF1 and DegP 12-mers respectively, [*hTRF1*] and [M_{12}] are their molar concentrations, and D_{hTRF1} is the ideal diffusion constant for hTRF1 calculated from the Stokes-Einstein equation assuming a constant hydrodynamic radius of 1.97 nm computed using the solution NMR structure in HYDROPRO (31, 49). The slight overcorrection of the pink points (20 μM M_T) at higher temperatures (~30-50 °C) reflects the unfolding of an increasing fraction of hTRF1 with temperature, leading to a slight increase in hTRF1 hydrodynamic radius with increasing temperature (1) that is not taken into account in the correction.

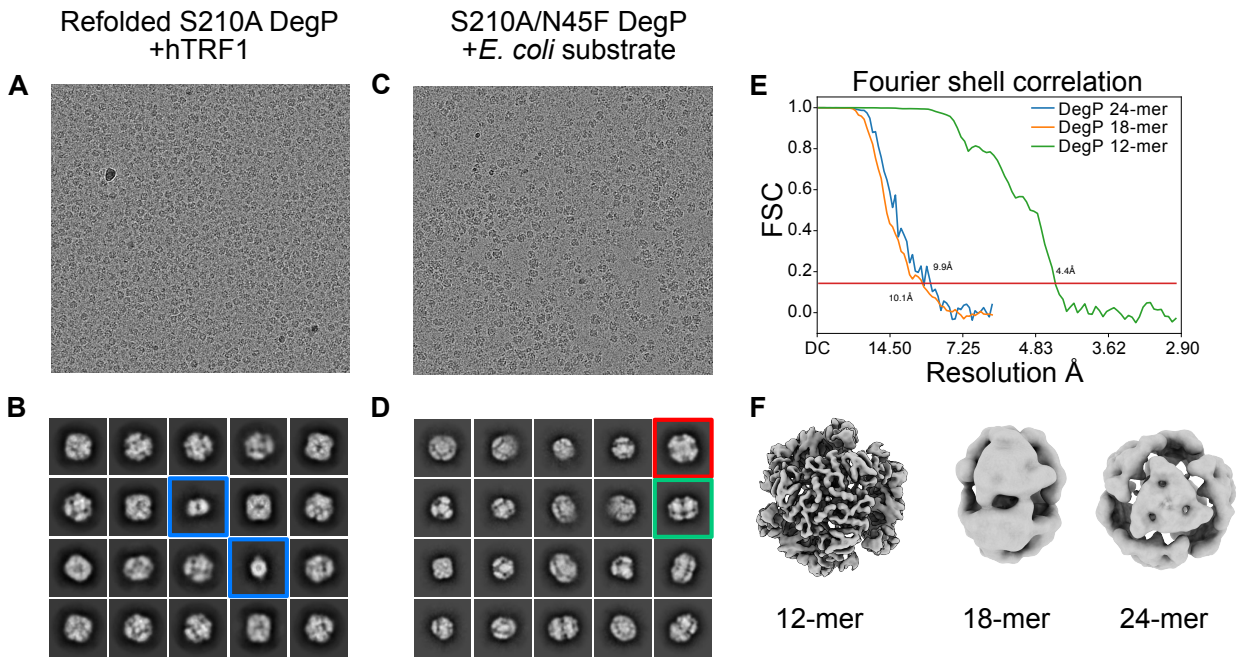


Figure S9. Electron cryo-microscopy of DegP cages. (A) Representative micrograph from the refolded S210A DegP + hTRF1 dataset, and (B) 2D class averages for picked particle images. A small fraction appeared to be hexamers outlined in blue. (C) Representative micrograph from the S210A/N45F DegP + *E. coli* substrate dataset, and (D) 2D class averages for picked particle images; a 24-mer class is outlined in red, and an 18-mer class is outlined in green. (E) Fourier shell correlation for the 24-mer, 18-mer, and 12-mer cage reconstructions, with the 24-mer and 18-mer images Fourier cropped to a Nyquist of 5.8 Å. (F) Density map for the 12-mer cage that has not been low-pass filtered, along with the 18-mer and 24-mer cages (not to scale).

Supplementary Table

Table S1. Thermodynamic parameters for the DegP two-pathway self-assembly model.

Parameter	DLS	AUC	NMR PDZ1:PDZ2 titration[d]
$K_{A,0}$[a]	$(9.1 \pm 4.3) \times 10^5$	$(8.1 \pm 1.0) \times 10^5$	-
$\Delta H_{A,0}$[a]	-112.0 ± 27.2	-212.3 ± 14.7	-
$\Delta C_{p,A}$[b]	-6.9 ± 1.5 (-13.1)	-6.9[c]	-
$K_{B,0}$[a]	$(1.2 \pm 0.3) \times 10^5$	-	-
$\Delta H_{B,0}$[a]	-24.6 ± 14.9	-	-15.3 ± 5.5[d]
$\Delta C_{p,B}$[b]	-1.3 ± 0.9 (-0.9)	-	-

[a]Reported at the reference temperature of 25 °C; "*A*" and "*B*" refer to pathways A and B, respectively. [b]ΔC_p values in parentheses were calculated using solvent-accessible surface areas and empirical estimates of non-polar and polar contributions to protein folding, see *SI Appendix Estimating* ΔC_p *values for oligomerization via changes in solvent-accessible surface areas* for details. [c]Constrained to the optimal value from DLS global fitting. [d]Parameters for the isolated PDZ1:PDZ2 interaction for comparison with $\Delta H_{B,0}$. All values are defined in the association direction and given as the mean ± SD based on 1000 Monte-Carlo global fitting iterations. K_0, ΔH_0, and ΔC_p values are reported in units of M^{-1}, kJ mol^{-1}, and kJ mol^{-1} K^{-1} respectively.

Supplementary References

1. A. Sekhar, R. Rosenzweig, G. Bouvignies, L. E. Kay, Mapping the conformation of a client protein through the Hsp70 functional cycle. *Proc. Natl. Acad. Sci.* **112**, 10395–10400 (2015).
2. S. Schlegel, *et al.*, Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb. Cell Fact.* **12**, 1–12 (2013).
3. V. Tugarinov, L. E. Kay, An isotope labeling strategy for methyl TROSY spectroscopy. *J.*

Biomol. NMR **28**, 165–172 (2004).

4. I. Gelis, *et al.*, Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **131**, 756–769 (2007).
5. D. Neri, T. Szyperski, G. Otting, H. Senn, K. Wüthrich, Stereospecific Nuclear Magnetic Resonance Assignments of the Methyl Groups of Valine and Leucine in the DNA-Binding Domain of the 434 Repressor by Biosynthetically Directed Fractional ^{13}C Labeling. *Biochemistry* **28**, 7510–7516 (1989).
6. B. Volkmer, M. Heinemann, Condition-Dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* **6**, 1–6 (2011).
7. L. L. Graham, T. J. Beveridge, N. Nanninga, Periplasmic space and the concept of the periplasm. *Trends Biochem. Sci.* **16**, 328–329 (1991).
8. J. B. Stock, B. Rauch, S. Roseman, Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.* **252**, 7850–7861 (1977).
9. D. E. Koppel, Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants. *J. Chem. Phys.* **57**, 4814–4820 (1972).
10. B. J. Frisken, Revisiting the method of cumulants for the analysis of dynamic light-scattering data. *Appl. Opt.* **40**, 4087 (2007).
11. H. Zhao, C. A. Brautigam, R. Ghirlando, P. Schuck, Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. *Curr. Protoc. Protein Sci.* **7**, 20.12.1 (2013).
12. H. Zhao, *et al.*, Quantitative Analysis of Protein Self-Association by Sedimentation Velocity. *Curr. Protoc. Protein Sci.* **101**, 1–15 (2020).

13. S. E. Harding, P. Johnson, The concentration-dependence of macromolecular parameters. *Biochem. J.* **231**, 543–547 (2015).

14. T. M. Laue, B. D. Shah, T. M. Ridgeway, S. L. Pelletier, "Computer-aided interpretation of analytical sedimentation data for proteins" in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S. E. Harding, A. J. Rowe, J. C. Horton, Eds. (The Royal Society of Chemistry, 1992), pp. 90–125.

15. C. A. Brautigam, Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).

16. F. Delaglio, *et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

17. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

18. J. J. Helmus, C. P. Jaroniec, Nmrglue: An Open Source Python Package for the Analysis of Multidimensional NMR Data. *J. Biomol. NMR* **55**, 355–367 (2013).

19. M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).

20. A. Bax, D. Max, D. Zax, Measurement of Long-Range ^{13}C-^{13}C J Couplings in a 20-kDa Protein-Peptide Complex. *J. Am. Chem. Soc.* **114**, 6923–6925 (1992).

21. W. Vuister, A. D. Bax, Resolution Enhancement and Spectral Editing of Uniformly ^{13}C-Enriched Proteins by Homonuclear Broadband ^{13}C Decoupling. *J. Magn. Reson.* **98**, 428–435 (1992).

22. J. Santoro, G. C. King, A constant-time 2D overbodenhausen experiment for inverse

correlation of isotopically enriched species. *J. Magn. Reson.* **97**, 202–207 (1992).

23. S. G. Hyberts, K. Takeuchi, G. Wagner, Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **132**, 2145–2147 (2010).

24. J. Ying, F. Delaglio, D. A. Torchia, A. Bax, Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **68**, 101–118 (2017).

25. J. E. Ollerenshaw, V. Tugarinov, L. E. Kay, Methyl TROSY: Explanation and experimental verification. *Magn. Reson. Chem.* **41**, 843–852 (2003).

26. V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).

27. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290-+ (2017).

28. T. D. Goddard, *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

29. P. A. Korevaar, *et al.*, Pathway complexity in supramolecular polymerization. *Nature* **481**, 492–496 (2012).

30. D. Van Der Zwaag, *et al.*, Kinetic Analysis as a Tool to Distinguish Pathway Complexity in Molecular Assembly: An Unexpected Outcome of Structures in Competition. *J. Am. Chem. Soc.* **137**, 12677–12688 (2015).

31. A. Ortega, D. Amorós, J. García De La Torre, Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.*

101, 892–898 (2011).

32. A. Waterhouse, *et al.*, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

33. A. K. Attri, C. Fernández, A. P. Minton, Self-association of Zn-insulin at neutral pH: Investigation by concentration gradient-static and dynamic light scattering. *Biophys. Chem.* **148**, 23–27 (2010).

34. A. D. Hanlon, M. I. Larkin, R. M. Reddick, Free-solution, label-free protein-protein interactions characterized by dynamic light scattering. *Biophys. J.* **98**, 297–304 (2010).

35. E. Koehler, E. Brown, S. J. P. A. Haneuse, On the assessment of Monte Carlo error in simulation-based Statistical analyses. *Am. Stat.* **63**, 155–162 (2009).

36. S. K. Chaturvedi, V. Sagar, H. Zhao, G. Wistow, P. Schuck, Measuring Ultra-Weak Protein Self-Association by Non-ideal Sedimentation Velocity. *J. Am. Chem. Soc.* **141**, 2990–2996 (2019).

37. S. Kim, R. A. Grant, R. T. Sauer, Covalent linkage of distinct substrate degrons controls assembly and disassembly of DegP proteolytic cages. *Cell* **145**, 67–78 (2011).

38. N. V. Prabhu, K. A. Sharp, Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **56**, 521–548 (2005).

39. T. Krojer, *et al.*, Structural basis for the regulated protease and chaperone function of DegP. *Nature* **453**, 885–890 (2008).

40. V. V. Loladze, D. N. Ermolenko, G. I. Makhatadze, Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Sci.* **10**, 1343–1352 (2008).

41. P. R. Connelly, J. A. Thomson, Heat capacity changes and hydrophobic interactions in the binding of FK506 and rapamycin to the FK506 binding protein. *Proc. Natl. Acad. Sci. U.*

S. A. **89**, 4781–4785 (1992).

42. C. A. Waudby, A. Ramos, L. D. Cabrita, J. Christodoulou, Two-Dimensional NMR Lineshape Analysis. *Sci. Rep.* **6**, 1–8 (2016).

43. H. M. McConnell, Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28**, 430–431 (1958).

44. E. Rennella, A. Sekhar, L. E. Kay, Self-Assembly of Human Profilin-1 Detected by Carr-Purcell-Meiboom-Gill Nuclear Magnetic Resonance (CPMG NMR) Spectroscopy. *Biochemistry* **56**, 692–703 (2017).

45. A. Sekhar, A. D. Bain, J. A. O. Rumfeldt, E. M. Meiering, L. E. Kay, Evolution of magnetization due to asymmetric dimerization: Theoretical considerations and application to aberrant oligomers formed by apoSOD1^{2SH}. *Phys. Chem. Chem. Phys.* **18**, 5720–5728 (2016).

46. A. G. Palmer, C. D. Kroenke, J. P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238 (2001).

47. R. W. Harkness, Y. Toyama, L. E. Kay, Analyzing multi-step ligand binding reactions for oligomeric proteins by NMR: Theoretical and computational considerations. *J. Magn. Reson.* **318**, 106802 (2020).

48. T. Krojer, M. Ehrmann, T. Clausen, M. Garrido-Franco, R. Huber, Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455–459 (2002).

49. T. Nishikawa, A. Nagadoi, S. Yoshimura, S. Aimoto, Y. Nishimura, Solution structure of the DNA-binding domain of human telomeric protein, hTRF1. *Structure* **6**, 1057–1065 (1998).