Observing protein degradation by the PAN-20S proteasome by time-resolved neutron scattering

Emilie Mahieu, Jacques Covès, Georg Krüger, Anne Martel, Martine Moulin, Nico Carl, Michael Härtlein, Teresa Carlomagno, Bruno Franzetti, Frank Gabel

Observing protein degradation by the PAN-20S proteasome by time-resolved neutron scattering

Emilie Mahieu[1], Jacques Covès[1], Georg Krüger[2], Anne Martel[3], Martine Moulin[3], Nico Carl[3], Michael Härtlein[3], Teresa Carlomagno[2,4], Bruno Franzetti[1], Frank Gabel[1,3,*]

[1]Univ. Grenoble Alpes, CEA, CNRS, IBS, 71 avenue des Martyrs, F-38044 Grenoble, France

[2]Leibniz University Hannover, Centre for Biomolecular Drug Research, Schneiderberg 38, D-30167 Hannover, Germany

[3]Institut Laue-Langevin, 71 avenue des Martyrs, F-38042 Grenoble, France

[4]Helmholtz Centre for Infection Research, Group of Structural Chemistry, Inhoffenstrasse 7, D-38124 Braunschweig, Germany

[*]Corresponding author: frank.gabel@ibs.fr

Running title: Observing protein degradation by SANS

Abstract

The proteasome is a key player of regulated protein degradation in all kingdoms of life. While recent atomic structures have provided snapshots on a number of conformations, data on substrate states and populations during the active degradation process in solution remain scarce. Here, we use time-resolved small angle neutron scattering of a deuterium-labelled GFPssrA substrate and an unlabeled archaeal PAN-20S system to obtain direct structural information on substrate states during ATP-driven unfolding and subsequent proteolysis in solution. We find that native GFPssrA structures are degraded in a bi-exponential process, which correlates strongly with ATP hydrolysis, the loss of fluorescence and the build-up of small oligopeptide products. Our solution structural data support a model where the substrate is directly translocated from PAN into the 20S proteolytic chamber, after a first successful unfolding process which represents a point of no return, and thus prevents dissociation of the complex and the release of harmful, aggregation-prone products.

Statement of significance

Protein degradation by cells is an essential process for maintaining proteome homeostasis and for eliminating superfluous, dysfunctional or otherwise potentially dangerous proteins. While static structures of the large macromolecular complexes responsible for specific protein degradation have become increasingly available over the past years, the time-course and kinetics of protein substrates during the process remains largely unknown. The present study, combining protein substrate deuteration, solvent contrast variation, and time-resolved neutron scattering with online fluorescence provides new insights into substrate states and populations as a function of time, inaccessible to static techniques such as crystallography and cryo-EM. The results allow to propose a dynamic, regulated substrate processing mechanism.

Introduction

The cellular proteome determines the specific function and behavior of a cell at any given time. Protein levels need to be tightly regulated throughout the cellular life cycle, during cell differentiation in multicellular organisms and as a response to internal or external stimuli. While the regulation of gene expression represents an early and efficient mechanism to adjust protein levels, controlled protein degradation allows the cell to remodel the existing proteome at later stages (1-4).

Controlled protein degradation inside cells is a complex process, involving networks of multiple partners in all three domains of life (5, 6). Dysfunctional or superfluous proteins need to be identified, targeted and destroyed with exquisite specificity by the cellular proteolytic machinery (3, 7, 8). In general, specific structural surface features, so-called degrons (tags, hydrophobic patches and others), mark proteins for degradation; these marks are recognized by energy-dependent regulatory particles (RP), associated with core proteolytic particles (CP) (9). The proteasome represents the main intracellular proteolytic system in archaeal and eukaryotic cells; the endo-proteolytic sites responsible for the trimming of target proteins are sequestered within the proteasomes and thus inaccessible to natively folded proteins (9-11). The ATP-driven RPs act as "unfoldases" and are required to unfold and translocate substrates into the CP (12). A key constituent of regulatory particles are AAA+ ATPases (13), a large family of proteins, including DNA helicases and motor proteins (14), and which are generally involved in macromolecular remodeling processes .

Over the past decades, our understanding of the molecular mechanisms of protein degradation has improved significantly (9, 12). In particular, single-molecule experiments have shed light on the relationship~~s~~ between substrate unfolding and translocation as well as their correlation with ATP hydrolysis (15); in parallel, structural studies of substrate-free RPs have revealed ATP-binding patterns as well as the associated conformational changes of the oligomeric complexes (16). Recently, cryo-electron microscopy (cryo-EM) structures of substrate-engaged human (17) and yeast (18) 26S proteasome have revealed the correlations between ATP-hydrolysis, coordinated conformational changes of the 19S RP, and substrate interaction and translocation from the RP into the CP. However, despite these tremendous advances, we are still lacking structural data from the unfolding machinery during the catalytic cycle in solution. In particular, structural information on native substrates, putative unfolded intermediate states and proteolytic products remain scarce (19, 20), as well as information on the respective populations during the catalytic process.

Here, we present a structural kinetics study by time-resolved small angle neutron scattering (TR-SANS) which monitors the degradation of a protein substrate by the proteasome system from the extremophilic archaeon *Methanocaldococcus jannaschii*, during the catalytic process in solution. The archaeal proteasome consists of a AAA+ ATPase RP called PAN and a 20S protease core. We were able to exploit temperature activation at 55 °C of the hyperthermophilic system to trigger and fine-tune the rate of the degradation process and make it amenable to an *in vitro* structural study on the sub-minute time scale. As a substrate, we used ssrA-tagged GFP (21, 22), which allowed obtaining complementary information on the unfolding process *via* online fluorescence measurements, conducted in parallel to the TR-SANS experiments (20).

SANS allows monitoring structures and conformational changes of solubilized biomacromolecules or complexes (23, 24). In contrast to its more common and widespread sister technique small angle X-ray scattering (SAXS), SANS is able to distinguish between different (isolated or interacting) proteins present in solution by means of selective macromolecular deuteration (25) and solvent contrast variation (H_2O/D_2O exchange) (26-28). In particular, at about 42% D_2O, natural (i.e. hydrogenated) proteins are made "invisible", and the SANS signal can be attributed almost exclusively to the

deuterated proteins (29). This process is known as “contrast matching” and has been used in this work to mask the strong signal of the very large hydrogenated (h) particles (PAN, 20S) and focus exclusively on the signal of the much smaller, deuterated (d) GFP substrate during the degradation process in solution.

With this experimental setup, we monitored the structures of the native GFP substrate and its degradation products throughout the unfolding and degradation process in the presence of the PAN-20S complex, and quantified the respective populations at a time resolution of 30 seconds. We found that after a rapid initial exponential decay (characteristic time ≈ 1.5 min) of natively folded GFP and the concomitant appearance of oligopeptide products, the degradation process continued at a slower rate (characteristic time ≈ 10 min). The disappearance of native GFP structures, observed by SANS, correlated strongly with ATP hydrolysis, which displayed the same two distinct rates. GFP molecules aggregated strongly and instantly in the presence of PAN and ATP, while the addition of 20S prevented aggregation and triggered the emergence of small oligopeptide products. Importantly, our data strongly support a model where the substrate, once successfully unfolded, is directly translocated from PAN into the tightly bound 20S core particle, without intermediate release of unfolded, aggregation-prone substrates. Finally, our data are compatible with the hypothesis of a proteasome-substrate complex, where an elongated, unfolded intermediate substrate spans the PAN central channel and extends into the 20S catalytic chamber, potentially serving as a “tether” for the two enzymatic components until proteolysis is successfully completed. However, our SANS data also indicate that this intermediate state, if it exists, must be very weakly populated throughout the degradation process.

Material and methods

Expression and purification of perdeuterated GFPssrA

Plasmid construct: GFPssrA is a variant (https://www.fpbase.org/protein/alphagfp/) of the green fluorescent protein from *Aequorea victoria* containing an A206K mutation to minimize dimerization and with its C-terminus fused to the 11-residue peptide ssrA (AANDENYALAA). For the sake of simplicity we mostly use the notation “GFP” for “GFPssrA” throughout the manuscript. The GFP gene was codon-optimized for expression in *Escherichia coli*, synthesized and cloned by *GeneCust Europe* into a pET30a vector using NdeI/NotI cloning sites.

GFPssrA amino acid sequence:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDF
FKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFK
IRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELAANDENYALAA

Expression and purification: Deuterated GFP (d-GFP) was expressed at the Deuteration Laboratory of the Institut Laue-Langevin (ILL D-Lab) (Grenoble, France) in *E. coli* BL21 (DE3) cells. High-cell-density fermentation process in D_2O Enfors minimal medium, with d_8-glycerol (fully deuterated glycerol) as a carbon source and D_2O as solvent (25, 30), was used to grow bacteria at 30 °C until the culture OD_{600nm} reached 11. Protein expression was induced by using 1 mM IPTG until the OD_{600nm} reached 16.6. Bacteria were harvested by centrifugation at 3,000 x g for 20 min at 4 °C. Cells from 1 L of culture were re-suspended in 50 mL of 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.1% Triton X-100, and supplemented with 0.25 mg/mL lysozyme (Euromedex), 0.05 mg/mL DNase I grade II (Roche), 0.2 mg/mL RNase (Roche), 1 mg/mL Pefabloc SC (Roche) and 0.01 mg/mL $MgSO_4$.

Cells were lysed using a cell disruptor with 4 cycles at 18 kpsi at 7 °C (Constant System L_{TD}), and the lysate was clarified by centrifugation at 10,000 x g for 1 h at 4 °C. GFP was purified by organic extraction followed by Phenyl Sepharose 6 Fast Flow hydrophobic interaction chromatography as previously described (31). The protein was loaded on a Superose 12 10/300 GL size exclusion column (GE Healthcare), equilibrated in a 42 % D_2O buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl and 10 mM $MgCl_2$. The peak fractions were combined and concentrated at 50 mg/mL using Amicon Ultra-15 filtration units (3 kDa cutoff). Protein absorbance at 280 nm was used to determine the concentration of the proteins, and purified proteins were stored at -80 °C.

Expression and purification of hydrogenated PAN and 20S proteasome

Plasmid construct: The PAN gene from *Methanocaldococcus jannaschii* (PAN) was optimized for expression in *Escherichia coli*, synthesized and cloned by *GeneCust Europe* into a pET30a vector using NdeI/XhoI cloning sites. The expressed protein contains a poly-histidine tag at its C-terminal end. The genes of the α and βΔ6 subunits of the 20S proteasome from *M. jannaschii* (20S) were also optimized for expression in *E. coli*, synthesized and cloned by *GeneCust Europe* into a pETDuet-1 vector using, respectively, NcoI/AflII and NdeI/XhoI cloning sites. The expressed protein contains a poly-histidine tag on the N-terminal end of the α subunits.

PAN amino acid sequence:

MVFEEFISTELKKEKKAFTEEFKEEKEINDNSNLKNDLLKEELQEKARIAELESRILKLELEKKELERENLQLMKENEILRRE
LDRMRVPPLIVGTVVDKVGERKVVVKSSTGPSFLVNVSHFVNPDDLAPGKRVCLNQQTLTVVDVLPENKDYRAKAMEVDERPN
VRYEDIGGLEKQMQEIREVVELPLKHPELFEKVGIEPPKGILLYGPPGTGKTLLAKAVATETNATFIRVVGSELVKKFIGEGA
SLVKDIFKLAKEKAPSIIFIDEIDAIAAKRTDALTGGDREVQRTLMQLLAEMDGFDARGDVKIIGATNRPDILDPAILRPGRF
DRIIEVPAPDEKGRLEILKIHTRKMNLAEDVNLEEIAKMTEGCVGAELKAICTEAGMNAIRELRDYVTMDDFRKAVEKIMEKK
KVKVKEPAHLDVLYRLEHHHHHH

20S amino acid sequence:

α subunit

MGHHHHHHVPPSAYDRAITVFSPEGRLYQVEYAREAVRRGTTAIGIACKDGVVLAVDRRITSKLVKIRSIEKIFQIDDHVAAA
TSGLVADARVLIDRARLEAQIYRLTYGEEISIEMLAKKICDIKQAYTQHGGVRPFGVSLLIAGIDKNEARLFETDPSGALIEY
KATAIGSGRPVVMELLEKEYRDDITLDEGLELAITALTKANEDIKPENVDVCIITVKDAQFKKIPVEEIKKLIEKVKKKLNEE
NKKEEENREETKEKQEE

βΔ6 subunit

MTTTVGLICDDAVILATDKRASLGNLVADKEAKKLYKIDDYIAMTIAGSVGDAQAIVRLLIAEAKLYKMRTGRNIPPLACATL
LSNILHSSRMFPFLTQIIIGGYDLLEGAKLFSLDPLGGMNEEKTFTATGSGSPIAYGVLEAGYDRDMSVEEGIKLALNALKSA
MERDTFSGNGISLAVITKDGVKIFEDEEIEKILDSMKAKPKKKTTKRSRRKSK

Expression and purification: PAN and 20S were over-expressed in *E. coli* strain BL21 (DE3). Bacteria were grown in LB medium at 37 ˚C with constant shaking (160 rpm) until the OD_{600nm} of the cultures reached 0.6. Protein expression was induced by using 1 mM isopropyl-β-D-thio-galactoside (IPTG) for 4 h at 37 °C, or overnight at 20 °C for PAN and 20S, respectively. The cells were harvested by centrifugation at 3,000 x g for 20 min at 4 °C. Cells from 1L of culture were re-suspended in 50 mL of 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton X-100 and 5 mM imidazole, and supplemented with 0.25 mg/mL lysozyme (Euromedex), 0.05 mg/mL DNase I grade II (Roche), 0.2 mg/mL RNase (Roche), 1 mg/mL Pefabloc SC (Roche) and 0.01 mg/mL $MgSO_4$.

Cells were lysed using a cell disruptor with 4 cycles at 18 kpsi at 7 °C (Constant System L_{TD}). The lysates were incubated 20 min at 70 °C (PAN) or 85 °C (20S), and clarified by centrifugation at 10,000 x g for 1 h at 4 °C. The clear lysates were filtered (0.2 µM) and loaded on a Nickel column (GE Healthcare),

equilibrated in buffer composed of 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl, supplemented with 5 mM imidazole. After washing with equilibration buffer containing first 20, and subsequently 50 mM imidazole for 5 CV (column volumes), bound proteins were eluted with a 10 CV linear imidazole gradient (50-400 mM imidazole in the equilibration buffer). PAN and 20S were eluted at about 150 and 200 mM imidazole, respectively, and dialyzed over night against the buffer.
PAN was further purified by a Resource Q anion exchange column (GE Healthcare), equilibrated in the same buffer as before (20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl). After washing for 4 CV, bound proteins were eluted with a 15 CV linear NaCl gradient (100-300 mM NaCl). PAN was eluted at about 180 mM NaCl.
PAN and 20S were loaded on a Superose 6 Increase 10/300 GL size exclusion column (GE Healthcare), equilibrated in a 42 % D_2O buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl and 10 mM $MgCl_2$. The peak fractions were combined and concentrated to 20 mg/mL (PAN) or 45 mg/mL (20S) by using Amicon Ultra-15 filtration units (50 kDa cutoff for PAN and 100 kDa cutoff for 20S). Absorbance at 280 nm was used to determine protein concentration. The purified proteins were stored at -80 °C.

Offline optimization of SANS experimental conditions

Due to the limited neutron flux and weak contrast, SANS experiments from dilute biomacromolecular solutions require sample volumes of several hundred µL and protein concentrations of several mg/mL in order to yield good signal (29). As a consequence, the SANS sample composition (Table S1) differed significantly from those typically used in biochemical assays (29, 32, 33). In particular, protein and ATP (100 mM) concentrations were much more elevated. In order to match the elevated ATP concentrations in the SANS samples, the $MgCl_2$ concentrations in the buffers were also increased from 10 to 200 mM. We therefore investigated an array of temperatures and ATP concentrations by offline fluorescence spectroscopy (Fig. S1) in order to optimize the SANS experimental conditions, under the requirement that each SANS curve (frame) would be collected in 30 seconds and that a complete experimental series would last 45 minutes.

ATP and $MgCl_2$ concentration assays
Reaction mixtures were prepared with hydrogenated proteins (70 µM GFPssrA, 20 µM PAN and 20 µM 20S) in an H_2O buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl and 10 mM $MgCl_2$. The following conditions were measured: isolated GFP, GFP and PAN, or GFP, PAN and 20S with 0 mM ATP & 10 mM $MgCl_2$, 10 mM ATP & 20 mM $MgCl_2$, 50 mM ATP & 100 mM $MgCl_2$, and 100 mM ATP & 200 mM $MgCl_2$. The denaturation of GFPssrA was monitored (λ_{ex} = 400 nm and λ_{em} = 509 nm) by using a plate reader spectrophotometer (BioTek Synergy® H4 Hybrid Multi-Mode Microplate Reader). Fluorescence was measured over 1 hour at 55 °C with a time resolution of 40 seconds. Fluorescence curves as a function of time were generated by normalization to the first measurement point, which was defined as 100%.

Temperature assay
Reaction mixtures were prepared with hydrogenated proteins (70 µM GFPssrA, 20 µM PAN and 20 µM 20S) in a H_2O working buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 100 mM ATP, and 200 mM $MgCl_2$. The following conditions were measured: isolated GFP, GFP and PAN, and GFP, PAN and 20S at different temperatures (30, 40, 45, 50, 55, 60 and 65 °C). The denaturation of the GFPssrA was monitored (λ_{ex} = 400 nm and λ_{em} = 509 nm) using a plate reader spectrophotometer (BioTek Synergy® H4 Hybrid Multi-Mode Microplate Reader). Fluorescence was measured over 1 hour at each

temperature with a time resolution of 40 seconds. Fluorescence curves as a function of time were generated by normalization to the first measurement point, which was defined as 100%.

ATPase activity assays

Offline ATP hydrolysis under SANS conditions

The rate of ATP hydrolysis by PAN was measured offline in the same conditions as the SANS experiments. For this purpose, the amount of inorganic phosphate was recorded as a function of time. Reaction mixtures were prepared with hydrogenated proteins (70 µM GFPssrA, 20 µM PAN and 20 µM 20S) in an H_2O buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 100 mM ATP and 200 mM $MgCl_2$. The following conditions were measured in triplicate: isolated PAN, PAN and GFP, PAN, 20S and GFP, isolated GFP, isolated 20S, and buffer. The samples were incubated at 55 °C and monitored over 1 hour with a 20 µL reaction mixture, measured at each time point. The aliquots were incubated on ice to stop the reaction and diluted 150 times in the buffer. 4 µL were then supplemented with 36 µL of assay buffer and 200 µL of malachite green reagent (Sigma MAK113). After 30 minutes of incubation at room temperature, absorbance at 630 nm was analyzed by using a plate reader spectrophotometer (BioTek Synergy® H4 Hybrid Multi-Mode Microplate Reader). A standard phosphate curve was used to estimate the amount of phosphate released during the hydrolysis of ATP. Phosphate concentrations were fitted with a bi-exponential process (Eq. 1), where intensities were replaced by concentrations.

ATP/ADP mixture

To evaluate the inhibitory effect of ADP on the degradation reaction, we measured the rate of ATP hydrolysis by PAN in an ATP/ADP mixture in the presence of GFPssrA. Reaction mixtures were prepared with hydrogenated proteins (70 µM GFPssrA and 20 µM PAN) in an H_2O buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, and 200 mM $MgCl_2$. The following conditions were measured in triplicate: 20 mM ATP, 80 mM ADP, 20 mM ATP + 80 mM ADP. The samples were incubated at 55 °C and monitored over 1 hour with a 20 µL reaction mixture measured at each time point. The assay proceeded as described above.

SANS data collection, reduction and analysis

SANS data sets (https://doi.ill.fr/10.5291/ILL-DATA.8-03-935) were recorded on the D22 diffractometer (http://www.ill.eu/instruments-support/instruments-groups/instruments/d22) at the European Neutron Source Institut Laue Langevin (ILL) in Grenoble, France.

Reaction mixtures containing 42% D_2O (Table S1) were measured at a collimator/detector configuration of 2.0 m/2.8 m and allowed the matching of all hydrogenated (h) proteins (h-PAN and h-20S). For the reaction mixture of d-GFP and h-PAN with 100 mM ATP, an additional independent measurement was performed at a collimator/detector configuration of 8.0m/8.0 m in order to access lower q-values for a better characterization of the aggregates. 42% D_2O was chosen since it corresponds to the experimental contrast match point (CMP) of h-PAN, determined in a previous study (20). It is in good agreement with its theoretical match point (40.9% D_2O), calculated from its amino acid sequence (http://psldc.isis.rl.ac.uk/Psldc/). The experimental CMP of h-20S was not

determined in the present study explicitly, since its amino acid sequence predicts a theoretical match point of 41.2% D_2O, virtually identical to the one of h-PAN.

For each sample (Table S2), 310 µl of the reaction mixture were prepared and pipetted into 1 mm path length quartz cells (Hellma 110-QS), just before transferring them to the thermo-stated sample rack, already equilibrated at 55 °C in order to start the reaction. 55 °C was chosen as a compromise due to practical limitations of the SANS experiments: at higher temperatures, PAN is more active (Fig. S1, and (34)), but the reaction would have been too fast for the 30 s frames, in particular in the first, rapid regime, and at lower temperatures the reaction would have been too slow for the limited measurement time of the individual SANS experiments (45 min), in particular regarding the second, slower regime. The scattering curves were recorded with an individual exposure time of 30 s. An initial dead time of about 20 s was necessary to evacuate the experimental hutch and switch on the neutron beam. A neutron wavelength of 6 Å was used for data collection. A total of 90 individual scattering curves were recorded during the 45 min exposure times for each condition (apart from the control experiments, Fig. S2, were only 45 were collected due to limited beam time).

The corresponding buffers, empty cell, empty beam and boron/cadmium were measured in the same collimator/detector setups to perform data reduction (correction for detector efficiency, electronic noise, and sample holder scattering). Data reduction and radial integration over a two-dimensional image were performed using standard ILL software LAMP (35), and the one-dimensional sample and buffer curves were presented in absolute units (cm^{-1}), calibrated against the direct beam. The scattering curves of the buffers were subtracted from the respective sample curves with the software PRIMUS from the ATSAS suite (36). The radii of gyration R_G (which report on the extension of a particle) and forward scattered intensities $I(0)$ (which report on its molecular weight), as well as the respective errors, were extracted by using the Guinier approximation (37), i.e. a linear fit in an $\ln[I(q)]$ vs. q^2 plot, with the same program. The validity of the Guinier approximation ($R_G \cdot q_{max} < 1.3$) was fulfilled in all cases apart from the d-GFP + h-PAN samples where aggregates occurred. In this latter case, the lowermost three data points were used to extract a lower estimation of R_G and $I(0)$.

The program OLIGOMER (36) was used to fit the experimental scattering curves of GFP in the presence of PAN and 20S by using either two (natively folded GFP and decapeptide product) or three components (natively folded GFP, decapeptide product and partially unfolded GFP intermediate). All fits were done in the same q-range (0.03 to 0.35 $Å^{-1}$). The form factor inputs (i.e. theoretical SANS curves) of each component in 42% D_2O were calculated by the program CRYSON (38) and loaded into OLIOGMER. As a representative model of native GFP, we scored an ensemble of atomic models, including a flexible C-terminus, from previous work (39), and selected the model that fitted the d-GFP SANS data (Fig. 1A) best at t = 0 min. As a representative decapeptide product, the first ten residues (N-terminus) of the same model were used. The partially unfolded intermediate was created from the native GFP model by deleting the flexible C-terminus and the first, connected beta-barrel strand (up to P213) using PyMOL. Subsequently, unfolded peptide residues (1 to 28) and the first six connected ubiquitin residues from a blocked substrate solved by CryoEM (18) were fused to P213 to yield a 247 residue GFP construct (Fig. 3, red model). The completely unfolded GFP (Fig. S10) was generated from native GFP by Coot (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/).

Complementary to OLIGOMER we estimated the number of independent GFP-derived species in solution (in the presence of PAN and 20S) by a Singular Value Decomposition (SVD) analysis using the

program BioXTAS RAW (40). Intensities were integrated over the whole q-range and for all 30 s frames during the 45 min experimental series (Fig. 13S).

The time course of $I(0)$ intensities, $I(0,t)$, were fitted by using a bi-exponential function with the Igor Pro Software (https://www.wavemetrics.com/):

$$I(0,t) = I_0 + I_1[1 - \exp(-t/t_1)] + I_2[1 - \exp(-t/t_2)] \quad \text{(Eq. 1)}$$

I_0 is an initial intensity at t = 0 min, I_1 and I_2 are the weight fractions of two different exponential decay processes with characteristic time rates t_1 and t_2. The same function was also used to fit the evolution of the radii of gyration (R_G) and the fluorescence intensities (Fig. 2), as well as the ATP hydrolysis (Fig. S3). Eq. 1 was also used to describe the time behavior of the natively folded GFP population from the OLIGOMER data (Fig. 3, Fig. S4). In this case a bi-exponential fit did not converge, possibly due to the limited signal/noise of the original SANS data that limits the convergence of the OLIGOMER fits and/or due to fact that the first SANS curve did not yield a 100% population of natively folded GFP. Indeed, due to the short experimental delay (20 s, see above) until neutron acquisition is started and averaged over 30 seconds, the reaction is already active. Therefore, a single exponential decay function (t_2 = 0) was used.

Online fluorescence measurement

In parallel to all TR-SANS measurements, we carried out online fluorescence measurements. To this end, a special setup with an *in situ* spectrofluorimeter (Ocean Optics, 65000 pro) was implemented on the D22 beamline at the ILL as reported previously (20). Fiber optics (Ingénerie Développement Instrument Laser) were connected to the SANS quartz cell containing 310 µl of the reaction volume. The excitation was ensured by a laser diode at 405 nm and the emission wavelengths were filtered with a passband filter centered on 500 nm with a full width at half maximum (FWHM) of 40 nm (Thorlab), respectively. The fluorescence emission spectra (1 s) were recorded every 30 seconds and the fluorescence data were registered as NEXUS file format, in the same file as the SANS data. All spectra were subsequently extracted with the python hd5py package (http://www.h5py.org) and fitted with a Gaussian function to determine the fluorescence intensity at 509 nm. The error of the fluorescence intensity was estimated based on the standard deviation of the fitting parameters. The decreases in percentage of fluorescence (Table 1) were calculated with respect to time t = 0 min for each sample. The time course of the online fluorescence intensities was fitted by Eq. 1.

Mass spectrometry measurements

LC/electrospray ionization-TOF-MS analysis was carried out on d-GFP (6210, Agilent Technologies). The mass for d-GFP (28956.80 Da) gave an estimate of 69% deuteration (without the exchangeable H) due to the H_2O-based MS buffers and ~92% deuteration, combined with the calculated mass (with deuteration of exchangeable H) in 42% D_2O buffer. The Biological Scattering Tools website (http://psldc.isis.rl.ac.uk/) was used for the calculation of the number of exchangeable H in each protein.

MALDI-TOF analysis was carried out on the reaction h-GFP, h-PAN, and h-20S with or without ATP (Autoflex, Bruker Daltonics). The reaction mixture was prepared as described in Table S1 with hydrogenated proteins in a final volume of 50 µL and incubated for 1 h at 55 °C in a H_2O buffer containing 20 mM Tris-HCl pH7.5, 100 mM NaCl. h-GFP was purified in the same conditions as d-GFP.

Western blot analysis

Western-blot anti-GFP analysis was performed on the reaction products of the following samples: isolated h-GFP, h-GFP and h-PAN, and h-GFP, h-PAN, and h-20S with ATP. The reaction mixture was prepared in the presence of ATP as described in Table S1 with hydrogenated proteins in a final volume of 50 µL and incubated for 1 h at 55 °C in an H_2O buffer containing 20 mM Tris-HCl pH7.5, 100 mM NaCl. The samples were diluted in the same buffer to deposit 7.5 ng of GFP for each sample on an SDS-PAGE 15% acrylamide gel. The gel was then incubated for 10 min in the transfer buffer (3.03 g/L Trizma-base, 14.4 g/L glycine, 20% ethanol). The samples were transferred to the nitrocellulose membrane for 1 h at 100 V in transfer buffer. The membrane was then rinsed in TBS-Tween buffer pH7.6 (2.4 g/L Trizma base, 8 g/L sodium chloride, 0.1% Tween^{TM}20), before being incubated for 1 h at room temperature in TBS-Tween containing 5% (w/v) skimmed milk powder (Regilait). After washing in TBS-Tween buffer pH7.6, the membrane was incubated for 1 h at room temperature in TBS-Tween buffer pH7.6 containing the primary anti-GFP antibody (1/5000 - abcam®), then rinsed again. The membrane was finally incubated for 1 h at room temperature in TBS-Tween buffer pH7.6, the secondary rabbit antibody conjugated to horseradish peroxidase (HRP) (1/2000 - abcam®), and washed in TBS-Tween buffer pH7.6 before revealing the signal by dipping the membrane in detection solutions A (luminol) and B (peroxide) in a 1:1 ratio (Amersham® ECL®, GE Healthcare). The membrane was exposed for about 30 s and the signal was visualized using a CCD camera imager (Kodak). The exposure time was adjusted according to the signal/noise contrast.

Results

The PAN-20S complex degrades GFP at a bi-exponential rate that follows ATP consumption

In the absence of h-PAN and h-20S (h: hydrogenated), ssrA-tagged d-GFP (d: deuterated) conserved its native structure for 45 minutes at 55 °C: all SANS curves, acquired at 42% D_2O (where hydrogenated proteins are invisible), were identical and corresponded to the native GFP structure (Fig. 1A). Both the radius of gyration R_G and the forward scattered intensity $I(0)$ remained constant over time (Fig. 2A, Table 1) demonstrating that neither the compactness of the protein (R_G) nor its molecular weight ($I(0)$) were altered (23). This demonstrates that GFP did not unfold or form aggregates during the experiment (Fig. S5, left). Likewise, d-GFP conserved its native structure in the presence h-20S and ATP (but in the absence of h-PAN), and in the presence of h-PAN (or h-PAN and h-20S) but in the absence of ATP (Fig. S2). The moderate decrease of fluorescence at 509 nm (Fig. 2A, Table 1) may be attributed to a diminished quantum efficiency of fluorescence at higher temperatures (33), to minor conformational changes (34), or to structural degradation caused by the strong UV illumination in the exposed sample area (Fig. S5, left).

The addition of (contrast-matched) h-PAN and ATP in a freshly prepared d-GFP sample induced significant changes in the SANS curves (Fig. 1B): the radius of gyration R_G and the $I(0)$ intensity

increased by more than 10-fold and 1,000-fold, with respect to isolated d-GFP, while the fluorescence signal showed a considerable decrease (Fig. 2B, Table 1). The evolution of all molecular parameters was compatible with the rapid appearance of GFP aggregates (Fig. S5, middle), with average molecular weight and linear dimensions a thousand and ten times larger, respectively, than native GFP. The decrease of the fluorescence intensity could be described by a bi-exponential process (Fig. 2B, Eq. 1) with a fast (t_1) and a slow (t_2) characteristic time constant (Table 1). The increase of $I(0)$ displayed a more complex behavior (e.g. initial sigmoidal lag time), probably due to the process of growing aggregates, and was not fitted by a bi-exponential function.

Addition of (contrast matched) h-20S (and ATP) to a freshly prepared h-PAN/d-GFP sample yielded SANS curves that could be fitted by the structure of natively folded d-GFP at low angles, while at high angles the curves increasingly deviated from the native structure over time (Fig. 1C). Both the $I(0)$ intensity and the fluorescence decreased rapidly over time in a bi-exponential process at identical fast ($t_1 \approx 1.5$ min) and slow ($t_2 \approx 10$ min) rates (Fig. 2C, Table 1, Eq. 1). A single exponential process could not describe the decay in a satisfactory manner (Fig. S6). Thus, the SANS data indicate that GFP is efficiently degraded by 20S, leading to a progressive decrease of the population of native GFP (Fig. S5, right). These findings were supported by a Western blot analysis (as well as by mass spectrometry results, see next section) using an anti-GFP primary antibody on the end product of the reaction, where all native GFP disappeared in the presence of PAN, 20S and ATP, in conditions similar to SANS (Fig. S7).

Interestingly, ATP hydrolysis by the PAN-20S complex in the presence of GFP, measured off-line in conditions identical to the SANS experiment (Fig. S3), followed the same bi-exponential decay process as the $I(0)$ and fluorescence intensities, with identical fast and slow time constants ($t_1 \approx 1.5$ min and $t_2 \approx 10$ min, Table 1). This correlation indicates a direct relationship between ATP hydrolysis and the change in structural and functional parameters of the GFP substrate.

The GFP substrate is degraded into small oligopeptide products by the PAN-20S complex

The addition of 20S to the reaction volume efficiently prevented the appearance of GFP aggregates (Fig. 1C, Fig. S5, right), which were observed in the presence of PAN alone (Fig. 1B, Fig. S5, middle). As mentioned above, the SANS curve of d-GFP, in the presence of h-PAN, h-20S and ATP, was in good agreement with the native GFP structure only at t = 0 min (Fig. 1C, topmost red data points and fit), while the curves at subsequent times could no longer be fitted by native GFP alone over the entire q-range (Fig. 1C, black data, and Fig. S8). The scattering intensity of these curves decreased continuously over the reaction time at lower q-values, while it progressively increased at higher angles ($q > 0.15$ Å^{-1}) (Fig. 1C, right black arrow).

The increase of $I(q)$ at higher angles indicated the appearance of smaller reaction products that scatter homogeneously over a wider q-range (37, 41). Fittingly, a mass spectrometry analysis of the reaction products revealed a distribution of oligopeptides in the range from 4 to 14 amino acids (Fig. S9). In order to verify whether the SANS data could be fitted with a combination of natively folded GFP and oligopeptide products, we generated an average structure of a decapeptide to represent the low-molecular-weight products. This crude approximation is justified by the limited statistics of the 30 s SANS data at higher angles, which makes the curves insensitive to different conformations and moderate size variations of the oligopeptides. Furthermore, a Singular Value Decomposition (SVD) analysis with the program BioXTAS RAW (40) suggested the presence of a maximum of two significant populations of different species (Fig. S13).

The SANS curves could be fitted well ($0.8 < \chi < 1.1$) by a mixture of native GFP and the representative decapeptide by using the program OLIGOMER (36) (Fig. S8A to C); correspondingly, the time dependence of the populations of both species could be determined with good statistical precision (Table S3, left). The GFP population disappeared in a single exponential decay process (Fig. S4) to a final level of 13% (volume fraction), with a characteristic time of 5 min; concomitantly, oligopeptide products built up at the same rate to reach a final level of 87% (volume fraction). The rate of GFP proteolysis (5 min) lay between the slow and fast rates of the bi-exponential processes of ATP hydrolysis and fluorescence decay (Table 1).

Does a putative, elongated intermediate substrate state exist?

Recently, the structures of two eukaryotic homologs of the archaeal PAN-20S complex were solved by cryo-EM, in the presence of blocked substrates (17, 18). We combined information from the substrate-engaged structure of the yeast cryo-EM complex (18) and from single molecule spectroscopy (42) to generate a *bona fide* structural model of a putative, partially unfolded GFP intermediate (see Material and Methods). This model was added to the natively folded GFP structure and the decapeptide product to fit the TR-SANS data with the program OLIGOMER.

As in the case of the two-species fits, the three-species fits showed a rapid exponential decline of native GFP over the first 10 minutes, accompanied by an increase of the decapeptide population occurring at the same rate (Fig. S8D-F, Fig. 3). The population of the putative intermediate, (partially) unfolded state could not be determined with high precision and remained below 5-10% throughout the experiment (Fig. 3, Table S1, right). In addition, the quality of the three-species fits was not better than that of the two-species fits, as judged by the χ values (Fig. S8) and the populations of native GFP and decapeptide products did not change significantly between the fits. A significant population of a GFP intermediate with a high degree of unfolding is incompatible with the SANS curves at low q-values (Fig. S10). Together, our SANS data are compatible with the presence of a putative elongated and unfolded intermediate state as suggested by recent, substrate-engaged cryo-EM models, but indicate also that such a state, if it exists in solution, must be very transient and weakly populated during the degradation reaction.

Discussion

The time rates of the diminution of native GFP structures correlate with ATP hydrolysis

The experimental setup of the GFP-PAN-20S reaction allowed monitoring the disappearance of native GFP structures, the appearance of degradation products, and the decrease of fluorescence by TR-SANS and online fluorescence (Fig. 1C, 2C and 3). The results could be compared with ATP hydrolysis assays, measured off-line under identical conditions (Fig. S3), and revealed a strong correlation between the three experimental datasets (Table 1, Fig. 4). Moreover, the loss of natively folded GFP structure, determined from the SANS data by OLIGOMER (-87%), was identical to the loss of fluorescence (-86%), indicating a direct relationship between the structural and functional parameters.

At the beginning of the reaction, ATP was rapidly hydrolyzed by PAN and the disappearance of native GFP molecules and the fluorescence decay followed strictly the rate of ATP hydrolysis. After the first 10 minutes, ATP hydrolysis continued at a slower pace (Fig. S3), likely because of the inhibitory effect of accumulating ADP, which is known to compete with ATP for the PAN nucleotide-binding sites (16).

In an offline control experiment, we observed that ATP hydrolysis was significantly slowed down in an ADP-ATP mixture with respect to ATP only, while at the same time GFP unfolding was inhibited (Fig. S11). Indeed, the slow (t_2) rates measured for the disappearance of native GFP and for the fluorescence decay (Fig. 2C) correlated strongly with the slow rate measured for ATP consumption (Table 1, Fig. 4, light grey bars). Thus, the slow-down of ATP hydrolysis by the inhibitory effect of ADP resulted in a similar slow-down of GFP degradation at longer times.

Previous biochemical assays (33) have demonstrated the dependence of ATP hydrolysis on the presence of substrate or the 20S particle, which were both shown to stimulate ATP consumption. These assays were conducted at ATP concentrations of ≈1 mM, a value close to the K_m of ATP hydrolysis by PAN (0.5 mM) (43). Under these conditions, ATP hydrolysis is expected to be sensitive to modulations of the substrate or downstream effector concentrations. In our conditions, which were optimized for SANS experiments, the concentration of ATP was much higher (100 mM) in order to ensure the presence of sufficient ATP during the long reaction times, as well as to counteract the inhibitory effect of accumulating ADP. Thus, we expect our assays to be much less sensitive to the modulation of the K_m or k_{cat} of ATP hydrolysis. Consequently, we did not observe any dependence of the characteristic time constants on the presence of substrate (Fig. S3).

GFP is translocated directly from PAN to 20S in a processive degradation mechanism

The comparison of the SANS data in the presence of PAN and PAN+20S (Table 1, Fig. S5) shows that the GFP substrate is directly translocated from the unfoldase to the protease, without release of (partially) unfolded, aggregation-prone intermediate GFP states: GFP oligomers and/or aggregates were observed in the presence of PAN (Fig. 1B) but were completely absent in the reaction mixture containing both PAN and 20S (Fig. 1C). Since SANS intensities in the forward scattering direction, $I(0)$, are proportional to the squared molecular weight of a given molecular species (23), they are exquisitely sensitive to the presence of oligomers and/or aggregates: a fraction as low as 0.1% of GFP released in the form of an aggregate 100 times its size (estimated from the "GFP + PAN" data, Table 1) would have resulted in a readily detectable 10% increase of $I(0)$. We thus conclude that, even though the PAN-20S complex is known to be short-lived in the presence of ATP in solution (44, 45), a stoichiometric ratio as low as 1:1 (Table S1) is sufficient to keep both partners tightly associated, cause complete digestion of GFP, and prevent the release of aggregation-prone products.

Other studies have suggested that a successful degradation event is preceded by numerous unsuccessful unfolding attempts, followed by substrate release (12). Our SANS data indicate that, if substrate is indeed released after such attempts, it would have to be in a form ("destabilized GFP") that, while not necessarily identical with native GFP at an atomic level, would have to be structurally very similar to it. In particular, it must not display any altered features (loose C-terminal, exposed hydrophobic patches…) that would increase its propensity to aggregate (Fig. 5), and would therefore also have to differ from the (partially) unfolded intermediate state.

In conclusion, our SANS data strongly support a model where the translocation and proteolysis processes are spatially and temporarily coupled, with the unfolding intermediate substrate potentially serving as a "tether" for the PAN-20S complex during translocation, in agreement with recent cryo-EM structures of substrate-engaged eukaryotic proteasomes (17, 18) and previous enzymatic studies using inhibitors and mutational analyses (34).

A putative, partially unfolded intermediate substrate state may exist in solution, but would be weakly populated and transient

The TR-SANS data, recorded during the degradation reaction in solution, could be fitted satisfactorily using a distribution of only two species: native GFP and a representative decapeptide product (Fig. S8, A-C), and the evolution of their respective populations during time could be determined with good precision (Table S3, Fig. S4). Inspired by recent cryo-EM studies of substrate-engaged eukaryotic 26S proteasomes using ubiquitinylated peptide substrates (17, 18), we tested the existence of a third species representing a partially unfolded intermediate state (Fig. S8, D-F). The addition of this third species did not improve, but rather worsened, the quality of the SANS data fits and resulted in increased errors of the populations of native GFP. Minor populations (< 5-10%) of the putative intermediate state were compatible with the SANS curves at individual time points throughout the reaction (Fig. 3). These time points were distributed stochastically with no specific pattern or correlation (such as, e.g., a peak distribution during a specific time period). Another possibility would be a more fully unfolded intermediate state, displaying probably an ensemble of flexible conformations. While such ensembles can in principle be generated and modeled to small-angle data (46, 47), we believe that the statistics of our data, especially at higher angles ($q > 0.2$ Å^{-1}, e.g. Fig. S8) do not warrant such a sophisticated approach. We therefore estimated the impact of an ensemble by fitting a completely unfolded intermediate state as an upper case, and found a strong disagreement with the experimental SANS data at low angles (Fig. S10). The direct fits of the SANS curves agree well with a model-free SVD analysis (Fig. S13) which suggested a maximum of two non-negligible populations of independent species (i.e. native GFP and a representative oligopeptide population). In conclusion, while the SANS data do not strongly indicate the presence of a GFP state of different shape than the native protein, they would be compatible with a very transient and weakly populated form of a partially unfolded and elongated intermediate substrate, if it is structurally not too different from native GFP.

Results from force-induced, single-molecule unfolding of single and double (two fused) GFP molecules (15) or multi-domain filamin A substrates (48) indicate that the bacterial ClpXP AAA+ / protease complex unfolds individual protein domains in a cooperative manner within milliseconds, before translocating them during several seconds. Both studies report ratios of about 1:3 between translocation and dwell times (periods of up to 100s during which ATP is being consumed by ClpXP in futile unfolding attempts). The low population of GFP substrates in the process of unfolding and their homogeneous spread over time, without a specific pattern (Fig. 3, red data points), are compatible with these current models which describe protein unfolding by AAA+ proteases as a stochastic process (12). Finally and importantly, the fast time constant t_1 of 1.5 min observed by TR-SANS and online fluorescence (Table 1) agrees very well with the dwell times of about 100 s from single-molecule experiments, and would correspond to the half-life of a substrate unfolding process in solution.

Conclusions

A multitude of recent studies, including single molecule, cryo-EM and NMR, have impressively increased our knowledge of the structure and molecular mechanisms of the proteasome complex. However, structural data on the substrate states during the degradation process remain scarce. The TR-SANS data of the GFPssrA-PAN-20S system presented here provide direct and selective structural information on the ensemble of GFP substrate states during the unfolding and proteolysis processes in solution.

The proteolysis of native GFP follows a bi-exponential process consisting of a first, fast reaction on the time scale of approximately one minute and a second, slower reaction with time constant of the order of ten minutes. Both steps of the process correlate very strongly with the loss of fluorescence, measured online in parallel to SANS, and the consumption of ATP, as well as with the generation of small peptides (Fig. 4). The slowdown of the reaction rate after the first minutes is likely due to an inhibitory effect of accumulating ADP on the PAN unfoldase activity.

The comparison of SANS data in the presence of PAN alone and of PAN and 20S (Fig. 1B and 1C, both in the presence of ATP) strongly suggest that the unfolding GFP substrate is directly and quickly translocated from PAN into the 20S catalytic chamber, without intermediate release in solution, and might serve as a "tether" for both partners, assuring that they stay associated during unfolding, translocation and proteolysis (Fig. 5). The translocating substrate could function as a "safety mechanism" against the release of partially unfolded, and potentially harmful, aggregation-prone products into the cell.

Our results highlight the functional synergy between the regulatory protein AAA+ PAN and the catalytic particle 20S, which together build a high-processivity degradation machinery. While several attempts of unfolding are conceivable, the first successful unfolding attempt marks a "point of no return" (Fig. 5) beyond which the system cannot release misfolded proteins in the cytosol, as they would be toxic to the cell. These findings imply that the PAN-20S assembly must be closely regulated *in vivo* to avoid accumulation of denatured protein products. It is also conceivable that PAN is inactive in isolation *in vivo.*

Temperature activation of a hyperthermophilic archaeal system, in combination with deuterium labelling and SANS contrast variation represents a promising route to gain insight into a number of other substrate-processing events. It may be used to study functional time-resolved processes of a variety of large macromolecular assemblies such as cytoskeletal molecular motors (myosin, kinesin, dynein), RNA polymerase, DNA polymerase, chaperons, DNA repair machinery, or RNA degradation machinery. In this context, hyperthermophilic proteins remain interesting targets, given their higher stability compared to the mesophilic homologues and the possibility they offer to tune the reaction rates to the acquisition times required by TR-SANS.

Author contributions

E.M., A.M. and F.G. performed SANS experiments and analyzed data; E.M. performed wet lab experiments and analyzed data; J.C. and B.F. supervised wet lab experiments; G.K. and T.C. designed and discussed wet lab experiments; M.M. and M.H. designed and performed deuteration experiments; N.C. performed SANS data reduction; F.G. designed the project, analyzed data and wrote the manuscript together with E.M., B.F. and T.C. All authors reviewed, approved and contributed to the final version of the manuscript.

Acknowledgements

We acknowledge financial support from the French Agence Nationale de la Recherche (grant "PROTStretch" ANR-15-CE11-0026-01). This work used the platforms of the Grenoble Instruct-ERIC Centre (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INBS-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The authors

would like to thank Francesca Rossano for assistance with offline tests (Western blots, fluorescence), Luca Signor for mass spectrometry experiments and data reduction, Ziad Ibrahim and Gaëlle Hogrel for discussions on SANS sample experimental conditions, Eric Girard the generation of a strongly unfolded GFP molecule and Cristina Cocho Martinez for IT support during the TR-SANS-fluorescence experiments. The open access fee was covered by FILL2030, a European Union project within the European Commission's Horizon 2020 Research and Innovation program under grant agreement N°731096.

Supporting citations

References 36, 38 and 40 appear in the Supporting Material.

References

1. Wolf, D. H., and R. Menssen. 2018. Mechanisms of cell regulation – proteolysis, the big surprise. FEBS Letters 592(15):2515-2524.
2. Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2008. Molecular biology of the cell. Garland Science, New York, Adingdon.
3. Goldberg, A. L. 2003. Protein degradation and protection against misfolded or damaged proteins. Nature 426(6968):895-899.
4. Schoenheimer, R. 1946. The dynamic state of body constituents. Harvard Univ. Press, Cambridge, Mass.
5. Guerrero, C., T. Milenkovic, N. Przulj, P. Kaiser, and L. Huang. 2008. Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. Proceedings of the National Academy of Sciences of the United States of America 105(36):13333-13338.
6. Ciechanover, A. 2005. Proteolysis: from the lysosome to ubiquitin and the proteasome. Nature Reviews Molecular Cell Biology 6(1):79-87.
7. Gandhi, J., C. Antonelli Anthony, A. Afridi, S. Vatsia, G. Joshi, V. Romanov, V. J. Murray Ian, and A. Khan Sardar. 2019. Protein misfolding and aggregation in neurodegenerative diseases: a review of pathogeneses, novel detection strategies, and potential therapeutics. In Reviews in the Neurosciences. 339.
8. Brehm, A., and E. Krüger. 2015. Dysfunction in protein clearance by the proteasome: impact on autoinflammatory diseases. Seminars in Immunopathology 37(4):323-333.
9. Sauer, R., and T. Baker. 2011. AAA+ proteases: ATP-fueled machines of protein destruction. Annu Rev Biochem 80:587-612.
10. Majumder, P., and W. Baumeister. 2019. Proteasomes: unfoldase-assisted protein degradation machines. In Biological Chemistry.
11. Maupin-Furlow, J. 2012. Proteasomes and protein conjugation across domains of life. Nature Reviews Microbiology 10(2):100-111.
12. Olivares, A. O., T. A. Baker, and R. T. Sauer. 2015. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. Nature Reviews Microbiology 14:33. Review Article.
13. Yedidi, R. S., P. Wendler, and C. Enenkel. 2017. AAA-ATPases in Protein Degradation. Frontiers in Molecular Biosciences 4(42). Review.
14. Miller, J. M., and E. J. Enemark. 2016. Fundamental Characteristics of AAA+ Protein Family Structure and Function. Archaea (Vancouver, B.C.) 2016:9294307-9294307.

15. Maillard, R. A., G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante. 2011. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. Cell 145(3):459-469.
16. Kim, Y.-C., A. Snoberger, J. Schupp, and D. M. Smith. 2015. ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. Nature Communications 6:8520. Article.
17. Dong, Y., S. Zhang, Z. Wu, X. Li, W. L. Wang, Y. Zhu, S. Stoilova-McPhie, Y. Lu, D. Finley, and Y. Mao. 2019. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. Nature 565(7737):49-55.
18. de la Peña, A. H., E. A. Goodall, S. N. Gates, G. C. Lander, and A. Martin. 2018. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. Science 362(6418):eaav0725.
19. Augustyniak, R., and L. E. Kay. 2018. Cotranslocational processing of the protein substrate calmodulin by an AAA+ unfoldase occurs via unfolding and refolding intermediates. Proceedings of the National Academy of Sciences of the United States of America 115(21):E4786-E4795.
20. Ibrahim, Z., A. Martel, M. Moulin, H. S. Kim, M. Härtlein, B. Franzetti, and F. Gabel. 2017. Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. Scientific Reports 7:40948.
21. Gottesman, S., E. Roche, Y. Zhou, and R. T. Sauer. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes & development 12(9):1338-1347.
22. Weber-Ban, E. U., B. G. Reid, A. D. Miranker, and A. L. Horwich. 1999. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. Nature 401(6748):90-93.
23. Svergun, D. I., M. H. J. Koch, P. A. Timmins, and R. P. May. 2013. Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules. Oxford University Press, Oxford.
24. Putnam, C. D., M. Hammel, G. L. Hura, and J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly Reviews of Biophysics 40(3):191-285.
25. Haertlein, M., M. Moulin, J. M. Devos, V. Laux, O. Dunne, and V. Trevor Forsyth. 2016. Chapter Five - Biomolecular Deuteration for Neutron Structural Biology and Dynamics. Methods in Enzymology. Z. Kelman, editor. Academic Press, pp. 113-157.
26. Feigin, L. A., and D. I. Svergun. 1987. Structure Analysis by Small-Angle X-Ray and Neutron Scattering. Plenum Press, New York and London.
27. Jacrot, B. 1976. The study of biological structures by neutron scattering from solution. Reports on Progress in Physics 39(10):911-953.
28. Stuhrmann, H. B. 1974. Neutron Small-Angle Scattering of Biological Macromolecules in Solution. J. Appl. Cryst. 7:173-178.
29. Mahieu, E., and F. Gabel. 2018. Biological small-angle neutron scattering: recent results and development. Acta Crystallographica Section D 74(8):715-726.
30. Artero, J. B., M. Haertlein, S. McSweeney, and P. Timmins. 2005. A comparison of refined X-ray structures of hydrogenated and perdeuterated rat gammaE-crystallin in H2O and D2O. . Acta Crystallographica Section D 61:1541-1549.
31. Yakhnin, A. V., L. M. Vinokurov, A. K. Surin, and Y. B. Alakhov. 1998. Green Fluorescent Protein Purification by Organic Extraction. Protein Expression and Purification 14(3):382-386.
32. Kenniston, J. A., T. A. Baker, J. M. Fernandez, and R. T. Sauer. 2003. Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine. Cell 114(4):511-520.

33. Benaroudj, N., P. Zwickl, E. Seemüller, W. Baumeister, and A. L. Goldberg. 2003. ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. Molecular Cell 11(1):69-78.
34. Zhang, F., Z. Wu, P. Zhang, G. Tian, D. Finley, and Y. Shi. 2009. Mechanism of substrate unfolding and translocation by regulatory particle of the proteasome from Methanocaldococcus jannaschii. Molecular Cell 34:485-496.
35. Richard, D., M. Ferrand, and G. J. Kearley. 1996. Analysis and visualisation of neutron-scattering data. Journal of Neutron Research 4(1-4):33-39.
36. Franke, D., M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. Journal of Applied Crystallography 50(Pt 4):1212-1225.
37. Guinier, A., and G. Fournet. 1955. Small Angle Scattering of X-rays. John Wiley & Sons, New York.
38. Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proceedings of the National Acadamy of Sciences USA 95(5):2267-2272.
39. Kim, Henry S., A. Martel, E. Girard, M. Moulin, M. Härtlein, D. Madern, M. Blackledge, B. Franzetti, and F. Gabel. 2016. SAXS/SANS on Supercharged Proteins Reveals Residue-Specific Modifications of the Hydration Shell. Biophysical Journal 110(10):2185-2194.
40. Hopkins, J. P., R. E. Gillian, and S. Skou. 2017. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J. Appl. Crystallogr. 50:1545-1553.
41. Glatter, O., and O. Kratky. 1982. Small Angle X-ray Scattering. Academic Press, London, New York, Paris, San Diego, San Francisco, Sao Paolo, Sydney, Tokyo, Toronto.
42. Nager, A. R., T. A. Baker, and R. T. Sauer. 2011. Stepwise unfolding of a β barrel protein by the AAA+ ClpXP protease. Journal of molecular biology 413(1):4-16.
43. Wilson, H. L., M. S. Ou, H. C. Aldrich, and J. Maupin-Furlow. 2000. Biochemical and Physical Properties of the Methanococcus jannaschii 20S Proteasome and PAN, a Homolog of the ATPase (Rpt) Subunits of the Eucaryal 26S Proteasome. Journal of Bacteriology 182(6):1680.
44. Smith, D. M., N. Benaroudj, and A. Goldberg. 2006. Proteasomes and their associated ATPases: A destructive combination. Journal of Structural Biology 156(1):72-83.
45. Smith, D. M., G. Kafri, Y. Cheng, D. Ng, T. Walz, and A. L. Goldberg. 2005. ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. Molecular Cell 20(5):687-698.
46. Perkins, S. J., D. W. Wright, H. Zhang, E. H. Brookes, J. Chen, T. C. Irving, S. Krueger, D. J. Barlow, K. J. Edler, D. J. Scott, N. J. Terrill, S. M. King, P. D. Butler, and J. E. Curtis. 2016. Atomistic modelling of scattering data in the Collaborative Computational Project for Small Angle Scattering (CCP-SAS)This article will form part of a virtual special issue of the journal, presenting some highlights of the 16th International Conference on Small-Angle Scattering (SAS2015). Journal of Applied Crystallography 49(6):1861-1875.
47. Bernado, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society 129(17):5656-5664.
48. Aubin-Tam, M.-E., Adrian O. Olivares, Robert T. Sauer, Tania A. Baker, and Matthew J. Lang. 2011. Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. Cell 145(2):257-267.

Figure legends

Figure 1: Experimental SANS curves as a function of time. d-GFP alone (A), d-GFP in the presence of h-PAN (B), d-GFP in the presence of h-PAN and h-20S (C) on a double logarithmic plot (h: hydrogenated, d: deuterated). Continuous lines display fits to the atomic PDB model with CRYSON (38). Insets in (A) and (C) show the respective Guinier fits used to extract radii of gyration (R_G). GFP was per-deuterated in all cases and had a strong contrast in the 42% D_2O buffers. PAN and 20S were in their natural, hydrogenated state and effectively invisible in the 42% D_2O (and are therefore depicted as transparent ribbon). All samples were measured at 55 °C and in the presence of ATP. The arrows indicate the direction of the change of I(q) over time. In (C) the experimental points of the last curve (45 min) are connected by a line to facilitate their visualization at high angles. A total of 90 SANS frames of 30 seconds were recorded for all samples. For reasons of clarity, only a selection of frames is shown. Likewise, error bars are shown only for the first and last dataset in case (B). Typical error bars of the datasets in cases (A) and (C) are show in Figure S12.

Figure 2: Time-dependence of R_G, $I(0)$ and fluorescence. GFP alone (A), GFP in the presence of PAN (B) and GFP in the presence of PAN and 20S (C). All samples were measured at 55 °C and in the presence of ATP. Due to the contrast-matching of hydrogenated PAN and 20S, the SANS curves report exclusively on (deuterated) GFP and derived species and states (native, unfolded, hydrolyzed products and aggregates) in solution. R_G and $I(0)$ were extracted from the data in Fig. 1 with the Guinier approximation (see Material and Methods), the online fluorescence intensities are provided in arbitrary units. Continuous lines represent linear or bi-exponential (Eq. 1) fits to the data. A total of 90 SANS and fluorescence frames were recorded for all samples. For reasons of clarity, all frames were used for t < 8 min (to fit the fast processes in B and C) but only 1 out of 10 was used for t > 8 min.

Figure 3: Time-dependent populations of native GFP, a representative decapeptide product, and the putative intermediate state. The respective populations (volume fractions) were fitted with the program OLIGOMER (36), based on form factors (theoretical curves) of native GFP, the representative decapeptide product, and the model of the putative intermediate state. The structural model of the intermediate was generated based on cryo-EM and single-molecule data (18, 42) (see Material and Methods). Likewise, the PAN and 20S structures shown in this figure were exceptionally taken from the cryo-EM study of their eukaryotic homologs and are depicted as transparent ribbon to indicate that their SANS signal was contrast-matched by the solvent at 42% D_2O. The continuous lines represent the fits with a single exponential decay function (Eq. 1 with $t_2 = 0$).

Figure 4: Comparison of characteristic time constants from different experimental data. Characteristic time constants t_1 and t_2 (Eq. 1) from fits of $I(0)$ SANS intensities (indicative of molecular weight), fluorescence at 509 nm (indicative of the percentage of folded state), ATP consumption and OLIGOMER populations (Table 1). All

data refer to GFP in the presence of both PAN and 20S. In the case of OLIGOMER, a single exponential decay was sufficient to fit the data (Fig. 3); the corresponding characteristic time can be considered as an average of the t_1 and t_2 measured in the other experiments.

Figure 5: Schematic overview of GFP degradation process by PAN and 20S. Proposed mechanism of GFP degradation and interaction with the PAN-20S complex, based on TR-SANS data, coupled to online fluorescence. Our data are compatible with a trial-and-error model of unfruitful unfolding attempts and substrate release in a form (denominated "destabilized GFP", and distinct from the bound, intermediate form) that has no propensity to aggregate. At this stage, dissociation of the PAN-20S complex is still possible. Once unfolding is successfully engaged, the process has reached a "point of no return" and substrate is no longer released in solution, neither on the N-terminal nor on the C-terminal side of PAN. This implies that the substrate is efficiently transferred through the central PAN channel into the 20S protease catalytic chamber and hydrolyzed subsequently. The elongated intermediate substrate, in the process of translocation, might indeed act as a "tether" and keep PAN and 20S tightly associated and thus prevent the release of harmful, aggregation-prone substrate states.

Tables

	GFP	**GFP + PAN**	**GFP + PAN + 20S**
SANS			
$R_{G,start}$ [Å]	19.4 ± 1.1	20.8 ± 0.7	19.7 ± 1.2
$R_{G,final}$ [Å]	20.7 ± 1.1	195.9 ± 4.7 (*)	17.3 ± 4.8
ΔR_G [%]	+ 6.7%	+ 842% (*)	- 12.2%
$I(0)_{start}$ [cm^{-1}]	0.132 ± 0.005	0.130 ± 0.003	0.127 ± 0.005
$I(0)_{final}$ [cm^{-1}]	0.136 ± 0.005	20.27 ± 1.21 (*)	0.033 ± 0.004
ΔI(0) [%]	+ 3.0%	+ 15592% (*)	**- 74%**
t_1 [min]	N.D.	N.D. (**)	**1min 17s ± 27s**
t_2 [min]	N.D.	N.D. (**)	**11min 29s ± 4min 43s**
ATP hydrolysis			
ΔATP [%]	N.D.	- 83%	**- 84%**
t_1 [min]	N.D.	1min 55s ± 1min 8s	**1min 22s ± 41s**
t_2 [min]	N.D.	8min 10s ± 2min 45s	**10min 25s ± 5min 31s**
OLIGOMER			
ΔGFP_{native} [%]	N.D.	N.D.	**- 87%**
t_1 [min]	N.D.	N.D.	**4min 55s ± 41s**
Fluorescence			
$Fluo_{start}$ [a.u.]	57.0 ± 0.9	50.7 ± 0.9	50.9 ± 0.6
$Fluo_{final}$ [a.u.]	35.6 ± 1.0	5.5 ± 0.4	7.3 ± 0.4
ΔFluo [%]	- 38%	- 89%	**- 86%**
t_1 [min]	141min ± 136min (***)	40s ± 5s	**1min 23s ± 8s**
t_2 [min]	N.D. (***)	10min 14s ± 23s	**9min 10s ± 52s**

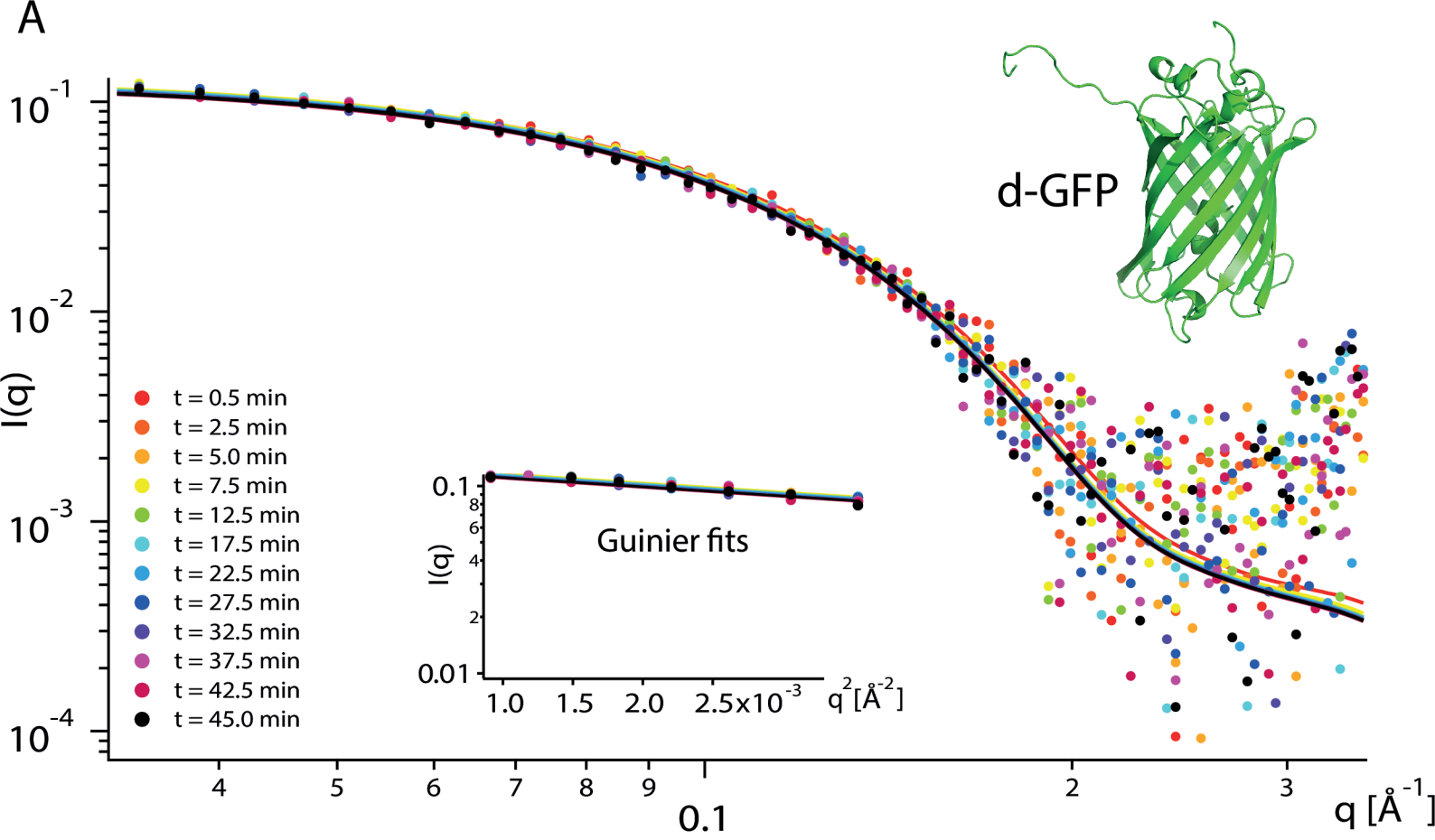
Table 1: Overview of parameters extracted from the experimental data. Relative changes of R_G, $I(0)$, fluorescence and populations of GFP, and ATP consumption. R_G, $I(0)$ and fluorescence at 509 nm (Fig. 2) were extracted from the SANS curves (Fig. 1) and from the online fluorescence measurements. Populations were determined by fitting the SANS curves with OLIGOMER (Fig. S8). ATP hydrolysis was measured using a standard assay (Fig. S3). Characteristic time constants t_1 and t_2 were fitted in all cases using Eq. 1, either with a bi-exponential function, or a single exponential function (all OLIGOMER populations, and fluorescence of isolated GFP). Errors represent standard deviations from the fits. "N.D.": not determined. "a.u.": arbitrary units.

(*): R_G and $I(0)$ intensities of the aggregated sample ("GFP + PAN") are "apparent" values since the experimental q-range did not cover the Guinier range correctly ($R_G \cdot q_{min} > 1$). As a consequence, both values are underestimated and represent lower limits of the real values.

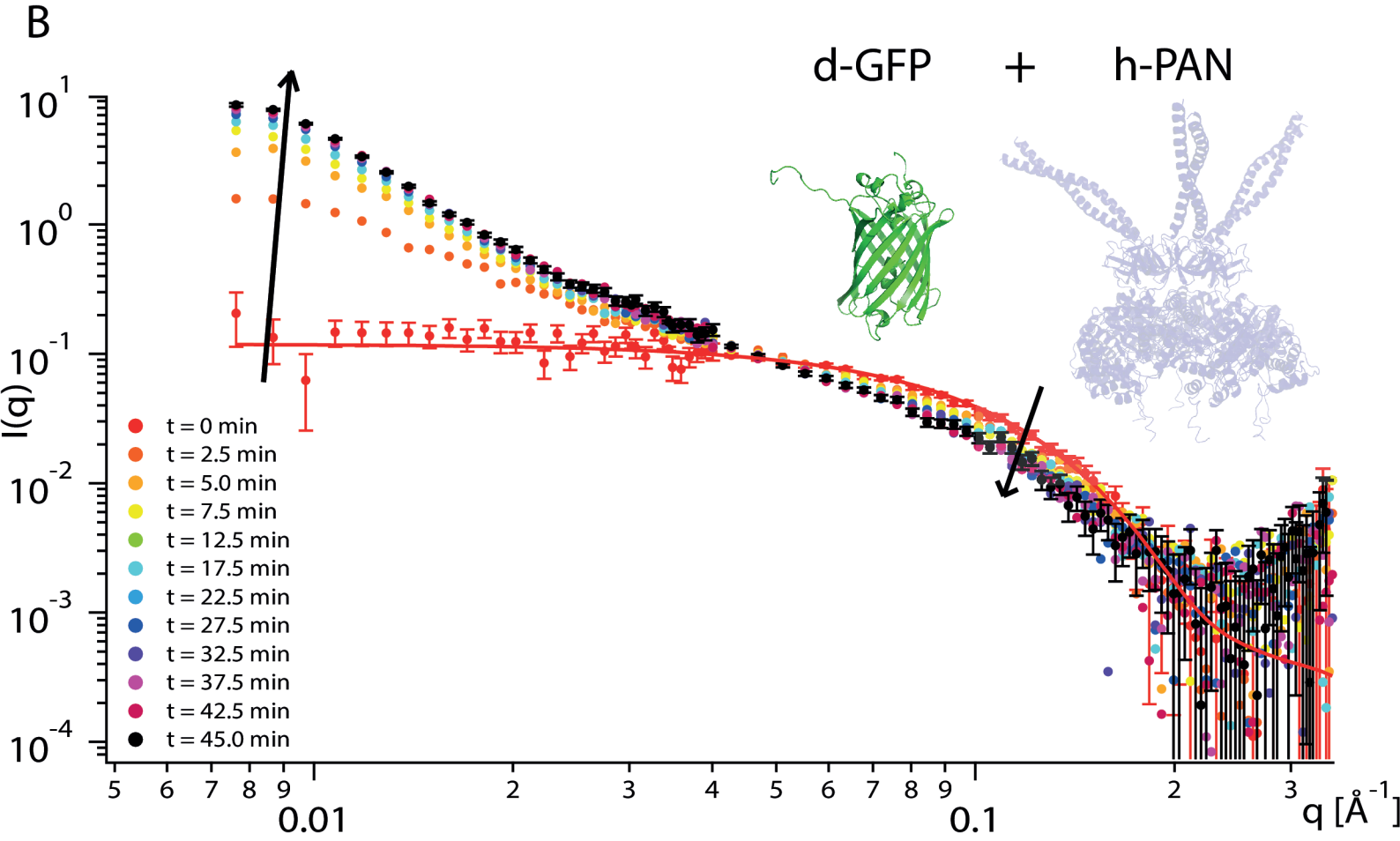
(**) $I(0)$ intensities of the aggregated samples display a sigmoidal behavior at short times. They can be fitted by a bi-exponential function only for t > 2min (not shown). Since the characteristic times include the growth process of the aggregates and can therefore not be compared directly to the characteristic times of the fluorescence decay or ATP hydrolysis, we did not attempt a quantitative analysis of the data.

(***) The fluorescence decay of isolated GFP can be fitted by a linear function or a single exponential function (as done here). The addition of a second exponential function does not improve the fit significantly.

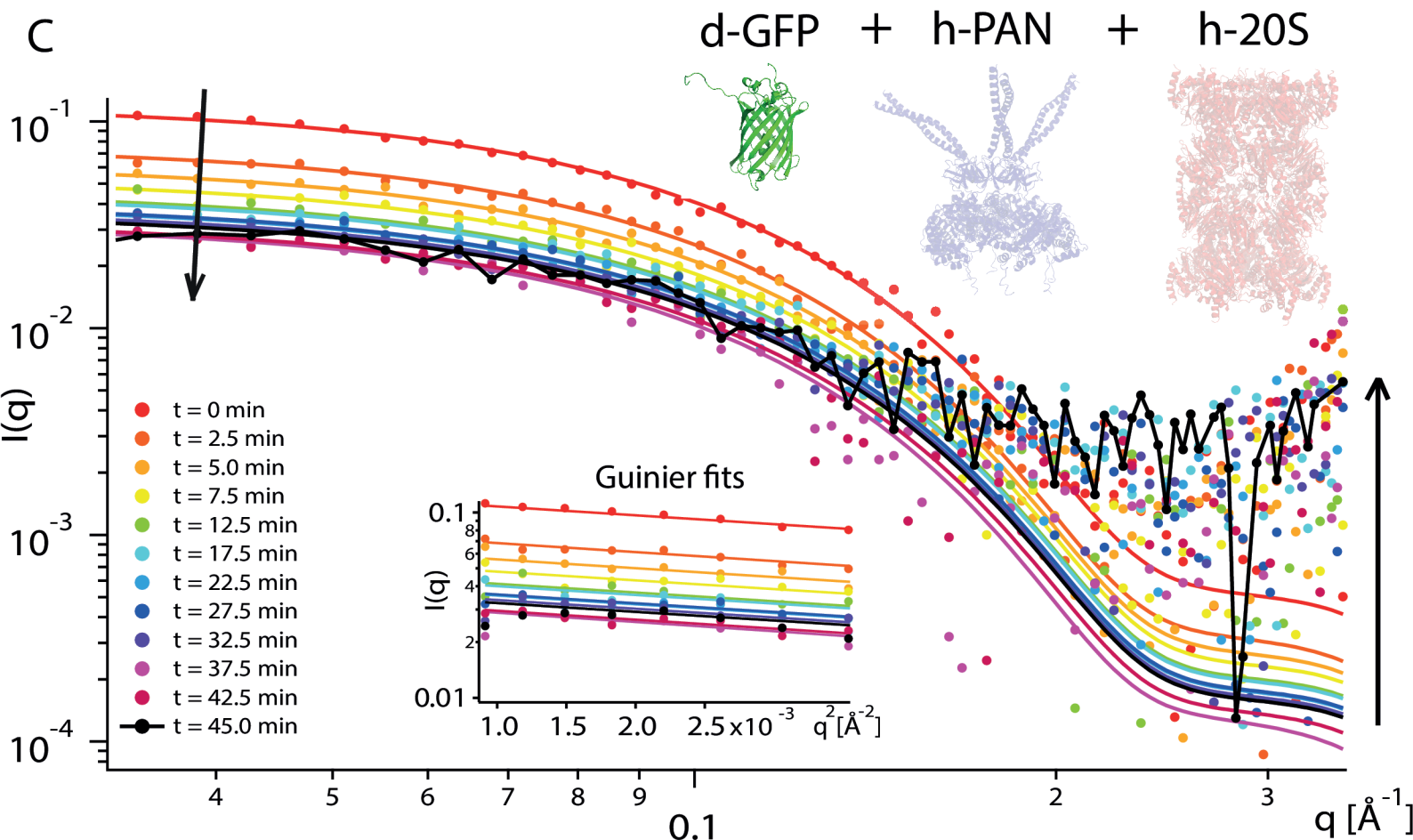
A


d-GFP
t = 0.5 min
t = 2.5 min
t = 5.0 min
t = 7.5 min
t = 12.5 min
t = 17.5 min
t = 22.5 min
t = 27.5 min
t = 32.5 min
t = 37.5 min
t = 42.5 min
t = 45.0 min
Guinier fits
I(q)
q²[Å⁻²]
q [Å⁻¹]

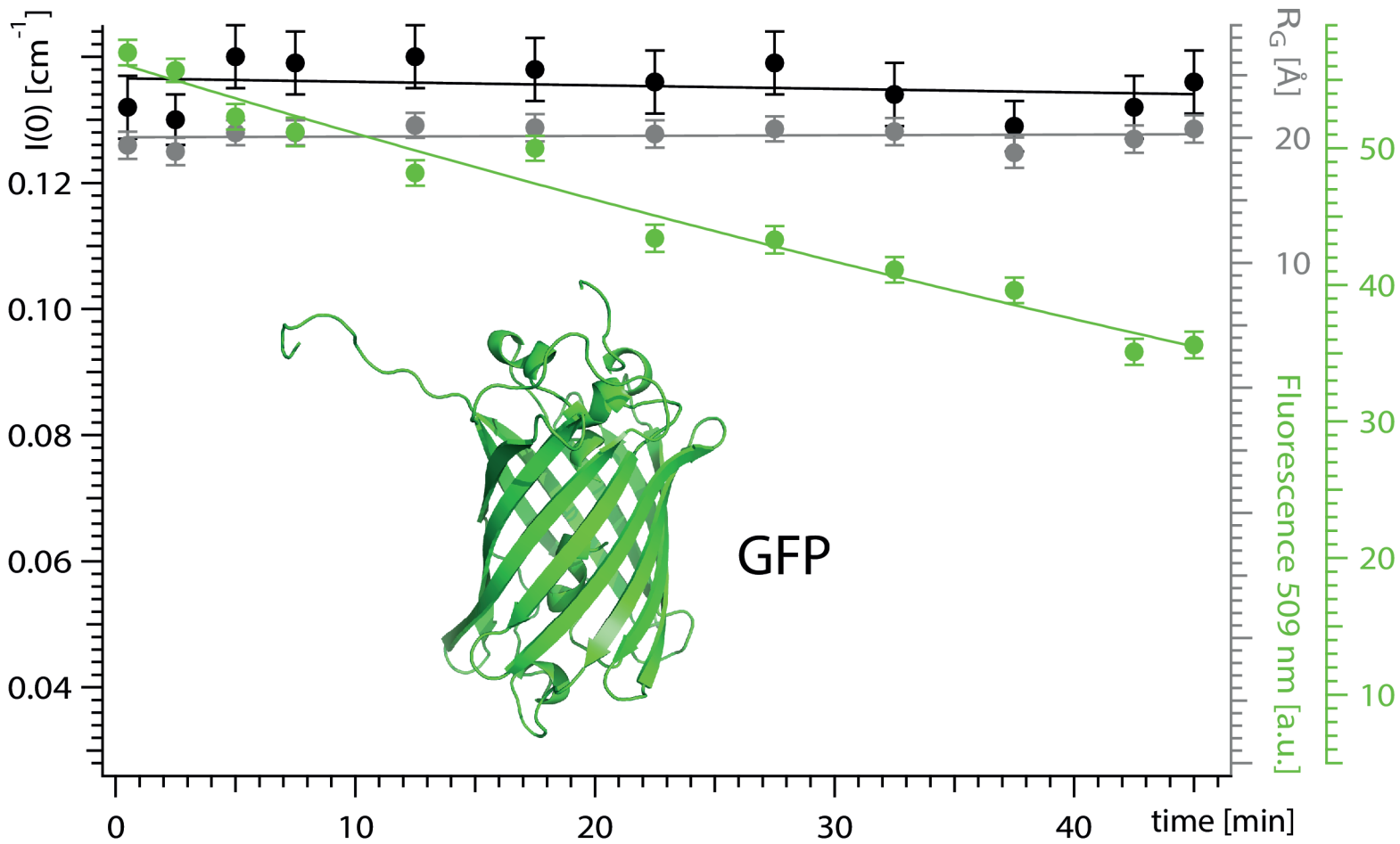
B


d-GFP + h-PAN
t = 0 min
t = 2.5 min
t = 5.0 min
t = 7.5 min
t = 12.5 min
t = 17.5 min
t = 22.5 min
t = 27.5 min
t = 32.5 min
t = 37.5 min
t = 42.5 min
t = 45.0 min
I(q)
q [Å⁻¹]

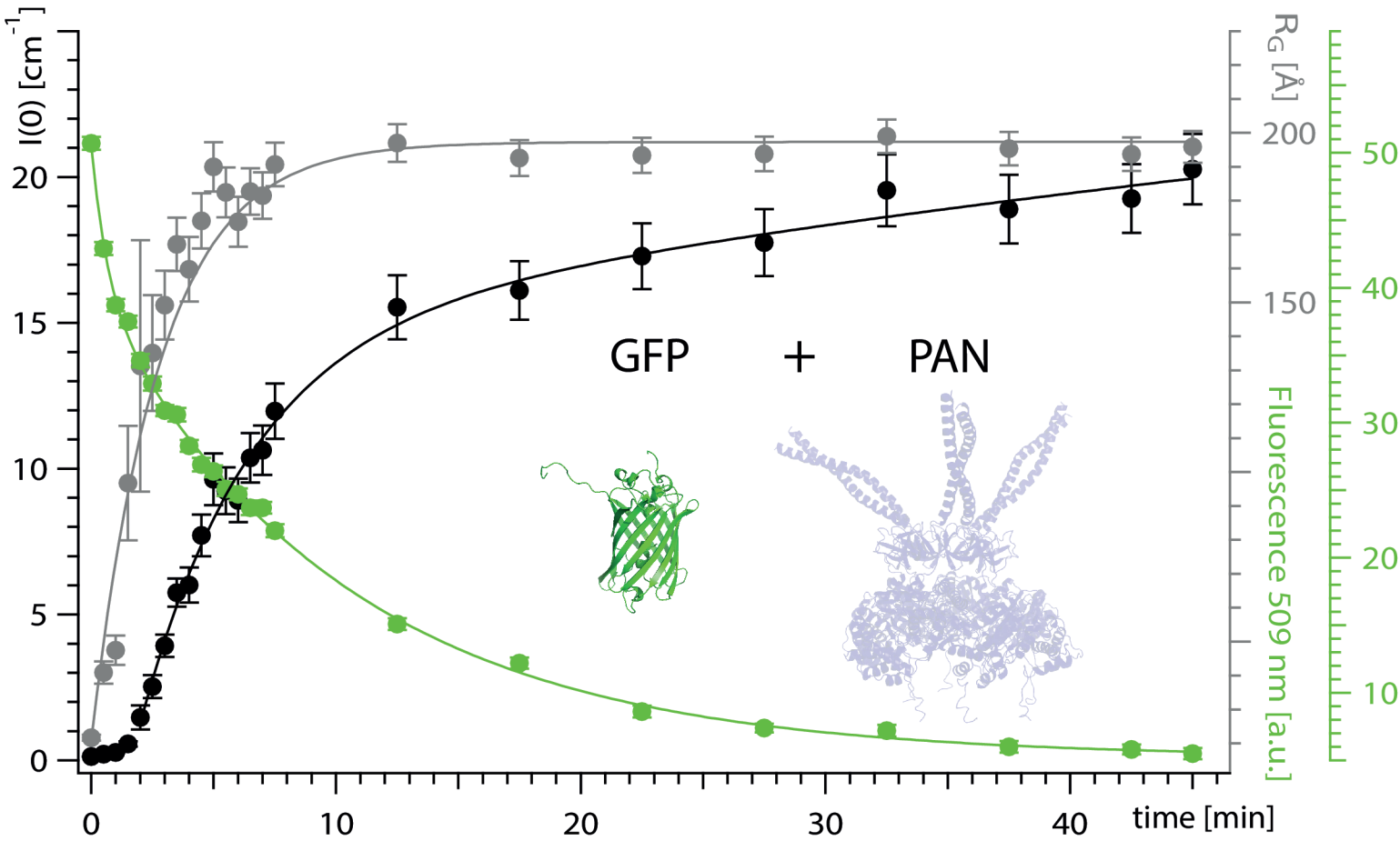
C


d-GFP + h-PAN + h-20S
t = 0 min
t = 2.5 min
t = 5.0 min
t = 7.5 min
t = 12.5 min
t = 17.5 min
t = 22.5 min
t = 27.5 min
t = 32.5 min
t = 37.5 min
t = 42.5 min
t = 45.0 min
Guinier fits
I(q)
q²[Å⁻²]
q [Å⁻¹]

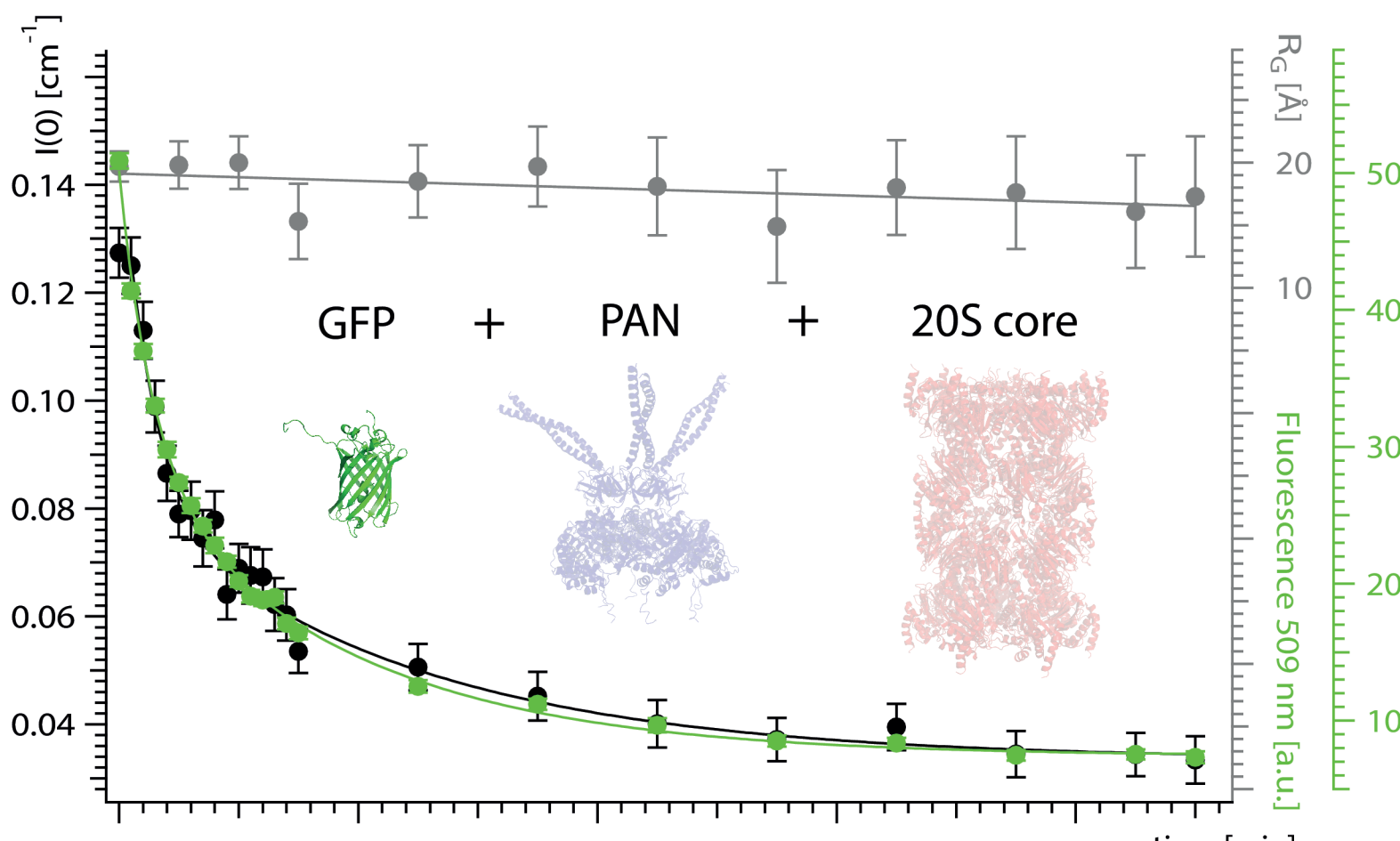
A

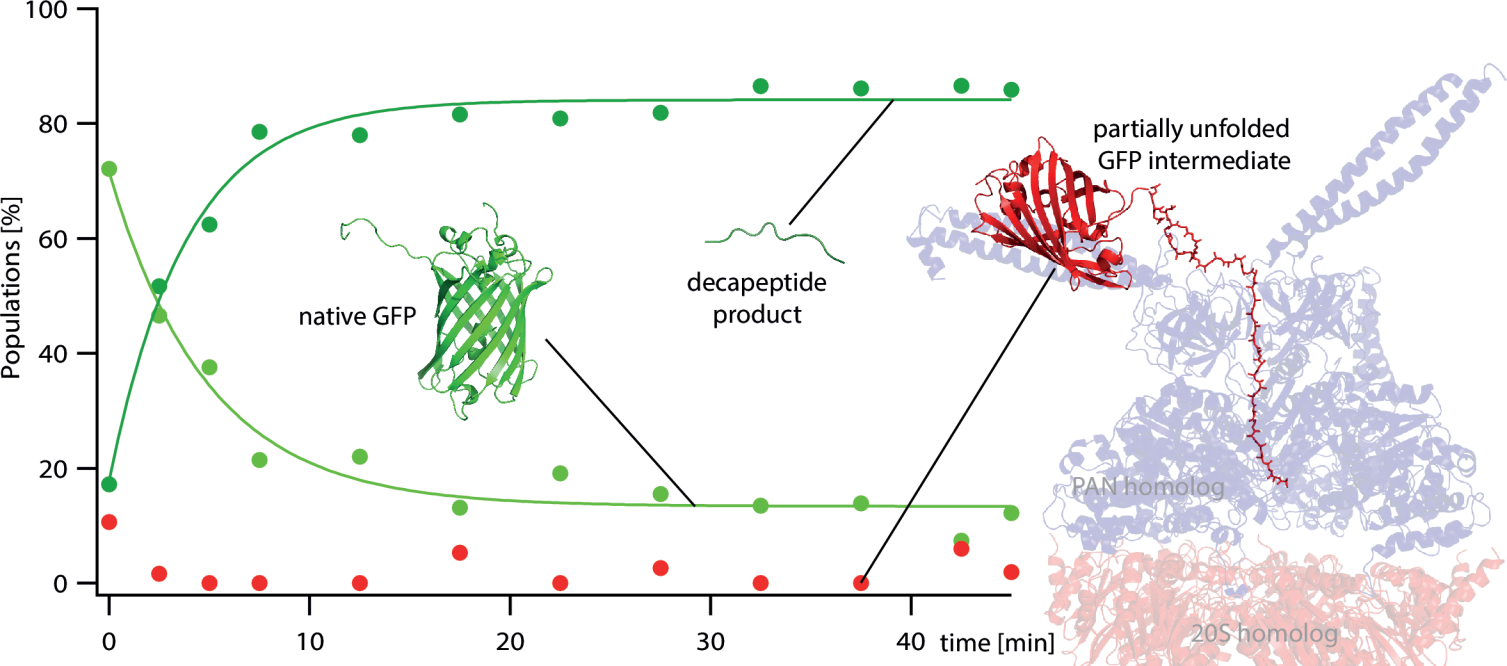

I(0) [cm^{-1}]
0.12
0.10
0.08
0.06
0.04
R_G [Å]
20
10
Fluorescence 509 nm [a.u.]
50
40
30
20
10
GFP
0
10
20
30
40
time [min]

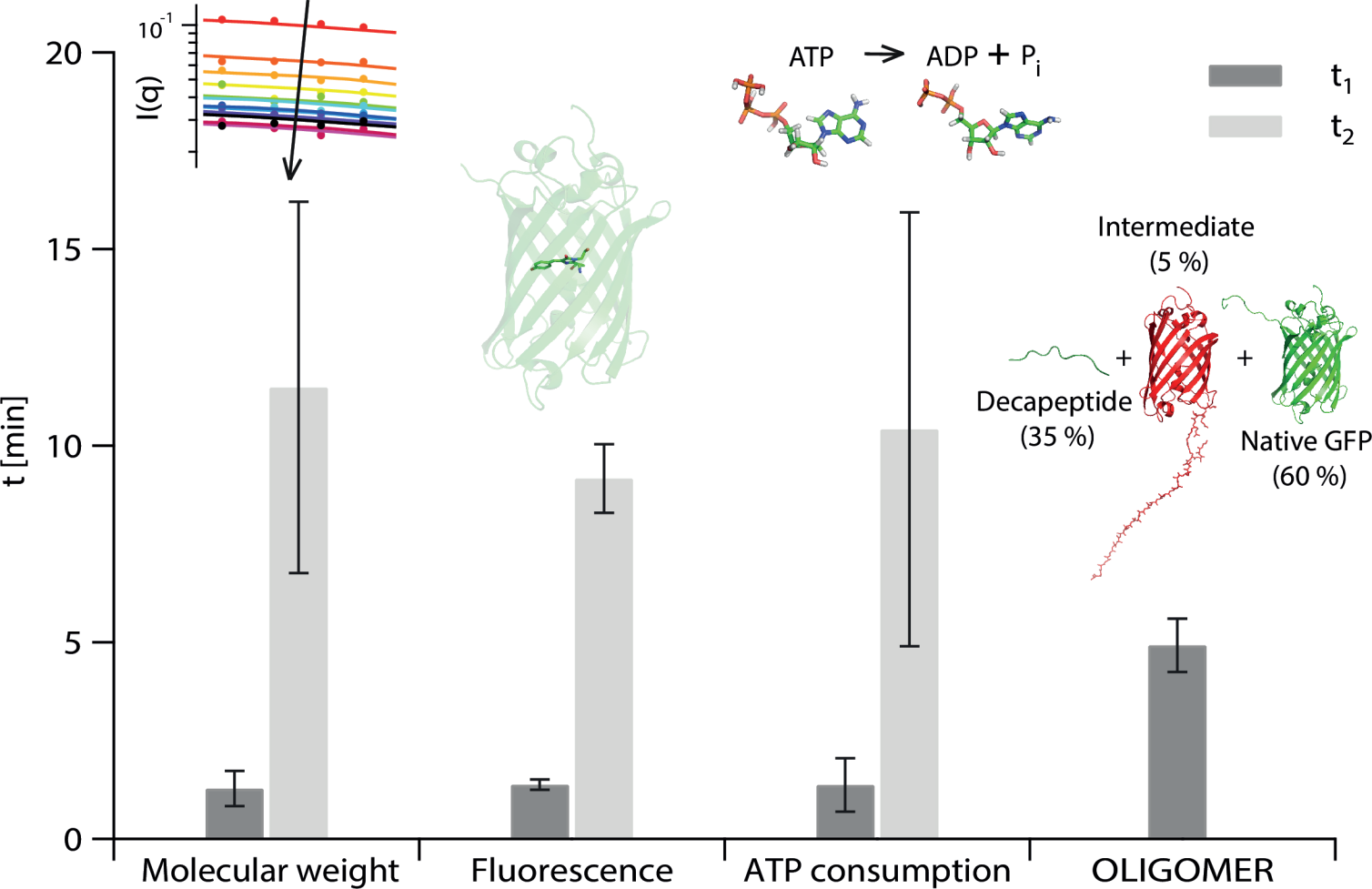
B

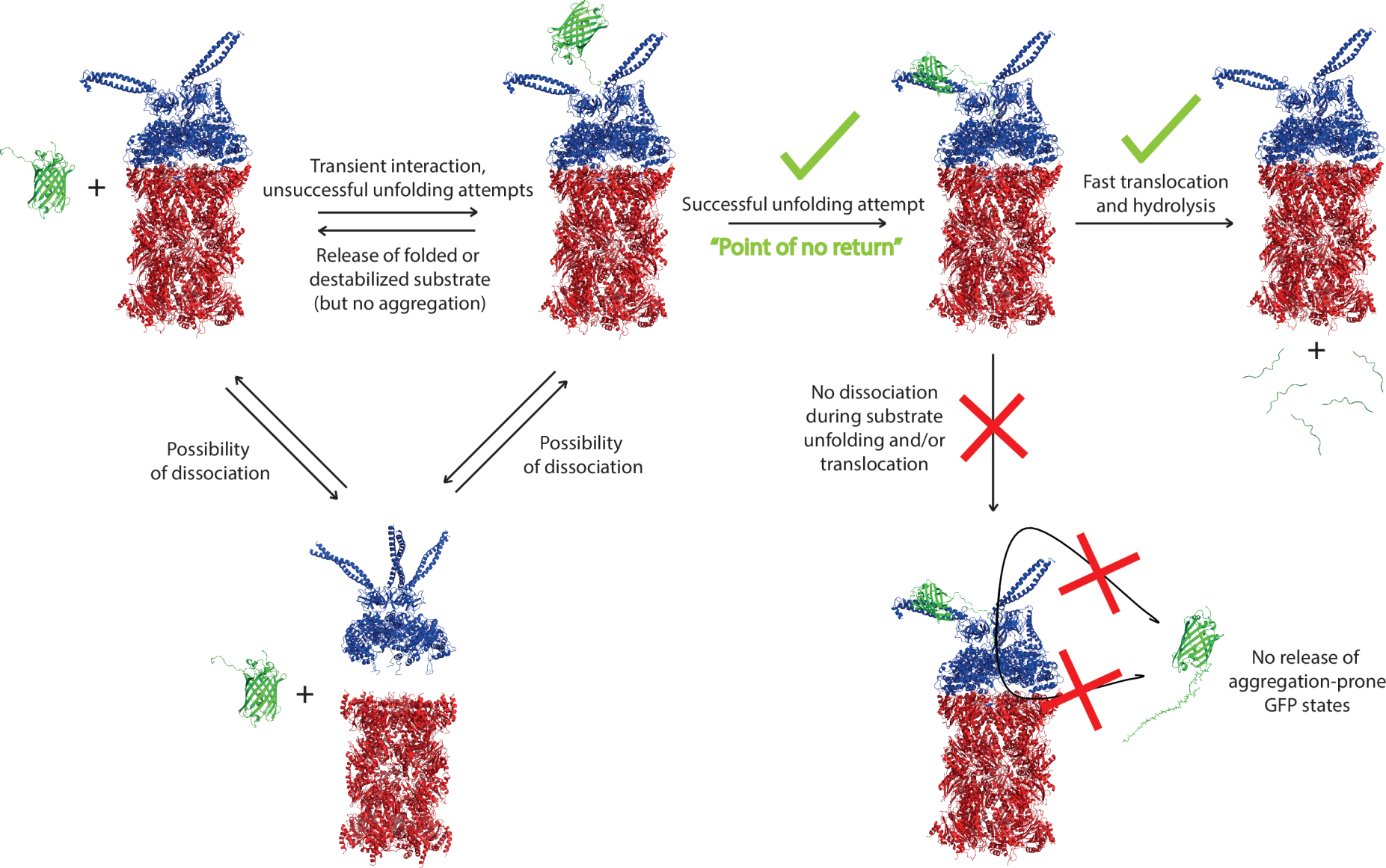

I(0) [cm^{-1}]
20
15
10
5
0
R_G [Å]
200
150
Fluorescence 509 nm [a.u.]
50
40
30
20
10
GFP + PAN
0
10
20
30
40
time [min]

C


I(0) [cm^{-1}]
0.14
0.12
0.10
0.08
0.06
0.04
R_G [Å]
20
10
Fluorescence 509 nm [a.u.]
50
40
30
20
10
GFP + PAN + 20S core
0
10
20
30
40
time [min]


100
80
60
40
20
0
Populations [%]
0
10
20
30
40
time [min]
native GFP
decapeptide
product
partially unfolded
GFP intermediate
PAN homolog
20S homolog


20
15
10
5
0
t [min]
10^{-1}
I(q)
ATP → ADP + P_i
t_1
t_2
Intermediate
(5 %)
Decapeptide
(35 %)
Native GFP
(60 %)
Molecular weight
Fluorescence
ATP consumption
OLIGOMER


Transient interaction,
unsuccessful unfolding attempts
Release of folded or
destabilized substrate
(but no aggregation)
Successful unfolding attempt
"Point of no return"
Fast translocation
and hydrolysis
Possibility
of dissociation
Possibility
of dissociation
No dissociation
during substrate
unfolding and/or
translocation
No release of
aggregation-prone
GFP states