Thermodynamic stability of the transcription regulator PaaR2 from *Escherichia coli* O157:H7

Running title: Thermodynamic stability of PaaR2

Pieter De Bruyn[1,2]*, San Hadži[1,2,3]*, Alexandra Vandervelde[1,2]*+, Albert Konijnenberg[1,2,4], Maruša Prolič-Kalinšek[1,2], Yann G.-J. Sterckx[1,2,5], Frank Sobott[4,6,7], Jurij Lah[3], Laurence Van Melderen[8], Remy Loris[1,2]

[1] Structural Biology Brussels, Department of Biotechnology, Vrije Universiteit Brussel, B-1050 Brussel, Belgium

[2] Center for Structural Biology, Vlaams Instituut voor Biotechnologie, B-1050 Brussel, Belgium

[3] Department of Physical Chemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, 1000 Ljubljana, Slovenia

[4] Biomolecular and Analytical Mass Spectrometry group, Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

[5] Laboratory of Medical Biochemistry, University of Antwerp (UA), Campus Drie Eiken, Universiteitsplein 1, 2610 Wilrijk, Belgium

[6] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

[7] School of Molecular and Cellular Biology, University of Leeds, Leeds, UK

[8] Cellular and Molecular Microbiology, Faculté des Sciences, Université Libre de Bruxelles (ULB), rue des Professeurs Jeener et Brachet 12, B-6041 Gosselies, Belgium

Correspondence to: Remy Loris (reloris@vub.ac.be; Remy.Loris@VIB-VUB.be)

Tel. 0032 2 6291989

Fax 0032 2 6291963

* These authors contributed equally to this work and should be considered joint first author.

+ Present address: KULeuven Laboratory of Dynamics in Biological Systems, O&N I Herestraat 49 - box 802, 3000 Leuven

Abstract

PaaR2 is a putative transcription regulator encoded by a three-component *parDE*-like toxin-antitoxin module from *E. coli* O157:H7. While this module's toxin, antitoxin and toxin-antitoxin complex have been more thoroughly investigated, little remains known on its transcription regulator PaaR2. Using a wide range of biophysical techniques (CD spectroscopy, SEC-MALLS, DLS, SAXS and native mass spectrometry), we demonstrate that PaaR2 mainly consists of α-helices, displays a concentration-dependent octameric build-up in solution and that this octamer contains a global shape that is significantly non-spherical. Thermal unfolding of PaaR2 is reversible and displays several transitions suggesting a complex unfolding mechanism. The unfolding data obtained from spectroscopic and calorimetric methods were combined into a unifying thermodynamic model, which suggests a five-state unfolding trajectory. Furthermore, the model allows the calculation of a stability phase diagram, which shows that, under physiological conditions, PaaR2 mainly exists as a dimer that can swiftly oligomerize into an octamer depending on local protein concentrations. These findings, based on a thorough biophysical and thermodynamic analysis of PaaR2, may provide important insights into biological function such as DNA binding and transcriptional regulation.

Introduction

Toxin-antitoxin (TA) modules are small operons abundant in chromosomes and mobile genetic elements of bacteria and archaea (for reviews see refs. (1–3)). Most prevalent are the type II modules, where both toxin and antitoxin genes encode proteins. TA modules were initially discovered on low copy number plasmids where they contribute to plasmid stabilization via a mechanism called "post-segregational killing" (4). This mechanism depends on the differential proteolytic degradation of toxin and antitoxin as initially observed for F-plasmid antitoxin CcdA (5) and later on observed for many other TA antitoxins.

Progress to understanding the biological function(s) of TA modules has been cumbersome and hampered by the lack of a clear phenotype in strains in which individual TA modules have been knocked-out (6). TA modules have been shown to be activated via a variety of stresses (for a review see ref. (7)). In addition, ectopic overexpression of many toxins leads to a reversible stasis, at least within a certain time window (8–10). Their involvement in the generation or maintenance of the persister state remains unclear as the cumulative effects of deleting multiple TA modules in *E. coli* were recently attributed to the inadvertent infection

of mutant strains with bacteriophage φ80 and technical flaws (11, 12). Several other studies that claim the involvement of individual TA modules in persistence looked at dormancy rather than true persistence (13, 14). Nevertheless, evidence for the involvement of TA modules in the survival of *Salmonella* within macrophages remains unchallenged (15).

Most type II toxin-antitoxin modules encode only two proteins. In such classic "two-component" modules, the antitoxin contains a globular DNA binding domain linked to an intrinsically disordered toxin-neutralizing domain (For a review, see ref. (16)). Interestingly, when the antitoxin neutralizes the toxin, the DNA-binding ability is retained but the affinity of the antitoxin for its operator DNA can be strongly influenced by binding of the toxin. This mechanism of transcriptional regulation, where the ratio of toxin to antitoxin ensures repression or derepression, is called conditional cooperativity and is best understood for *relBE*, *ccdAB* and *phd/doc* (17–19). Other TA modules employ a less complex mechanism of transcription regulation where the toxin only acts to weaken the affinity of the antitoxin for its operator. Examples of the latter are found in *higBA* type of TA modules (20, 21). For several families of TA modules, members exist where the antitoxin is split into two independent proteins: a DNA-binding regulator and a toxin-neutralizing antitoxin. The best-known example of such three-component systems is the ω-ε-ζ module from *Streptococcus pyogenes* plasmid pSM19035 (22), which is related to the classic two-component *pezAT* modules (23). Here, regulation is independent of the formation of a TA complex and depends only on the regulator ω (24, 25).

The chromosome of *E. coli* O157:H7 contains two three-component TA modules related to the *parDE* family, which were termed *paaR1-paaA1-parE1* and *paaR2-paaA2-parE2* (26). Here, the operons encode a toxin (ParE1 or ParE2), an antitoxin (PaaA1 or PaaA2) and an additional regulator (PaaR1 or PaaR2). The antitoxin and toxin components of one of these modules, *paaR2-paaA2-parE2*, have been studied in more detail. The antitoxin PaaA2 is intrinsically disordered, lacks a DNA binding domain and forms two transient α-helices (27). Upon toxin binding, PaaA2 wraps around ParE2 in a mostly α-helical conformation and the resulting antitoxin-toxin complex assembles into an octamer of PaaA2-ParE2 heterodimers (i.e., a heterohexadecamer) (28). ParE2 adopts a fold very similar to ParE from *Caulobacter crescentus* (29), although no GyrA-binding activity was detected for ParE2 (28). The functional relevance of this PaaA2-ParE2 heterohexadecamer remains unclear (28). In contrast to PaaA2

3

and ParE2, very little is known about the regulator PaaR2. Previously, Hallez and colleagues have identified PaaR2 as a putative transcriptional regulator containing features typical of the DicA transcriptional repressor family and a potential role in transcription regulation of the three-component *paaR2-paaA2-parE2* TA module was proposed (26). In this paper, we demonstrate that, in solution, PaaR2 assembles into an octamer with a non-spherical shape in a concentration-dependent manner. Based on a unifying thermodynamic model (which was constructed from spectroscopic and calorimetric thermal unfolding data), the octamer assembly seems to occur via a hierarchical pathway consisting of 5 species. Importantly, the proposed model suggests that, under physiological conditions, the predominant PaaR2 species in solution is a dimer that may relatively quickly oligomerize into an octamer depending on local conditions. Finally, we discuss that these findings may be relevant to understand the function of PaaR2 as a transcriptional regulator.

Materials and methods

Cloning, mutagenesis and transformation

The genes coding for PaaR2, PaaA2 and ParE2 were amplified from a purified colony of *E. coli* MC1061 carrying a pET15b expression plasmid with the *paaR2-paaA2-parE2* operon. The polyhistidine-tag was placed C-terminally on PaaR2 using primers PaaR2_1 and PaaR2_2. Primers PaaR2_3 and PaaR2_4 were used for *paaA2-parE2*. The sequences of the primers are shown in Supporting Table S1. The two gene amplifications were combined using an overlap PCR. Using In-Fusion® HD cloning, the *paaR2-paaA2-parE2* operon was then cloned in a pET15b plasmid which had been digested with BamHI and NcoI. $CaCl_2$-competent BL21 (DE3) *E. coli* cells were transformed with pET15b-PaaR2His-PaaA2-ParE2. The cysteine at position 120 of PaaR2 in the pET15b-PaaR2His-PaaA2-ParE2 plasmid was mutated to a serine using the Phusion site-directed mutagenesis kit (Thermo Fisher Scientific) and primers PaaR2HisC120SFwd and PaaR2HisC120SRev (Supporting Table S1). The mutation was confirmed by sequencing, after which the resulting plasmid, pET15bR2HisC120SA2E2, was transformed into expression strain *E. coli* BL21 (DE3) using the $CaCl_2$ method.

Expression and purification

A colony of BL21 (DE3) (pET15bR2HisC120SA2E2) was grown overnight at 37°C in 300 mL LB medium supplemented with ampicillin (100 µg mL^{-1}). The overnight culture was diluted fifty

times in twelve 1 L flasks of LB and grown at 37°C while shaking. When the OD reached 0.6-0.8, the cells were induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After further incubation for 4 hours at 37°C while shaking, the cells were spun down for 13 minutes using the JLA 8.1000 rotor at 5000 rpm (6238 g) and 4°C. Each pellet was resuspended in lysis buffer (20 mM Tris-HCl pH 7.3, 500 mM NaCl, 20 mM $MgCl_2$, 0.1 mg mL^{-1} 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 1 µg mL^{-1} leupeptine, 50 µg mL^{-1} DNase I). This suspension was then left to stir for 30 minutes at 4°C. Lysis occurred by sonicating three times for 1 minute. The lysate was centrifuged for 45 minutes using the JA 20 rotor at 18000 rpm (39191 g) and loaded onto a 5 mL HisTrap^{TM} HP Ni^{2+}-Sepharose column (GE Healthcare) that had been pre-equilibrated for at least one column volume with buffer A (20 mM Tris-HCl pH 7.3, 500 mM NaCl, 5 mM imidazole). After a wash period, buffer B (20 mM Tris-HCl pH 7.3, 500 mM NaCl, 1 M imidazole) was added to generate a linear gradient of 0 – 1 M imidazole over 50 column volumes. The fractions containing the protein of interest, PaaR2, were pooled and concentrated to a volume of 2 mL. The Ni-NTA-purified protein was subsequently loaded on a Superdex 200 16/90 SEC column (GE Healthcare), which had been washed and pre-equilibrated with SEC buffer (20 mM Tris-HCl pH 7.3, 500 mM NaCl) for at least one column volume. The buffers for the purification of the wild type PaaR2 are the same with the addition of 1 mM TCEP.

Electrophoretic mobility shift assays

Primers were designed, using the DNA sequence of *E. coli* O157:H7 Str. EDL933, to encompass the promoter operator region FL1 and FL2 or a random intergenic region Neg1 and Neg2, whilst maintaining a similar GC content and equal length (97 bp). The sequences are given in Supporting Table S1. All protein samples were concentrated and dialyzed to 1x PBS pH 8.0 (1 mM TCEP was added in the case of PaaR2 wild type). Binding reactions were performed by adding 1 µL of DNA at a final concentration of 0.25 µM of DNA to 9 µL of a protein solution. The mixture was left to incubate for 30 minutes at 20°C. After addition of 2 µL of retardation dye (25% ficoll, 0.1% xylenecyanol, 0.1% bromophenol blue), the samples were loaded in a 6% polyacrylamide gel. The electrophoresis was performed for 10 minutes at 180 V until the two dyes were clearly separated. Consequently, the voltage was diminished to 120 V for 40 minutes. The gel was then stained using EtBr.

Native mass spectrometry

Because of the changing oligomeric states of the protein, in all experiments the protein concentrations are expressed in molar concentrations of monomer equivalents. A sample of PaaR2 was prepared at different concentrations (from 65 to 0.33 µM) by overnight dialysis against 150 mM ammonium acetate at pH 7.3. Using in-house prepared, gold-coated borosilicate glass needles, the sample was introduced into the vacuum of the mass spectrometer via nano-electrospray ionization with a voltage of +1.6 kV. Spectra were recorded on a traveling wave ion mobility quadrupole time-of-flight instrument (Synapt G2 HDMS, Waters). Critical voltages throughout the instrument were 40 V for the sampling cone, 1 V for the extraction cone, 10 V trap collision voltage, 45 V for the trap DC bias and 0.5 V transfer collision voltage. Pressures throughout the instrument were 6 mbar, 4.3 E-2 mbar 3 mbar and 3.9 E-2 mbar for the Source, Trap collision cell, ion mobility cell, transfer collision cell, respectively. Collision cross sections (CCS) were obtained after calibration with proteins of known CCS as reported elsewhere (30): concanavalin A, alcohol dehydrogenase, glutamate dehydrogenase and avidin. Mass spectrometry and ion mobility data were analyzed using MassLynx 4.1 and Driftscope 2.3 (Waters). Experimental values for the CCS of PaaR2 were compared to theoretical models. The model for spherical growth was estimated based on the relevant equations by assuming a typical density of 0.44 Da per Å^3, as previously reported elsewhere (31, 32). Based on the obtained volume, a dense sphere was assumed, and the radius calculated. CCS were subsequently determined by calculating the cross section of a sphere where the final radius was $r_{PaaR2}+r_{N2}$ to compensate for the use of N_2 as the drift gas. The isotropic model was generated using the obtained experimental CCS of the dimer as this appeared to be the smallest stable building block of PaaR2 and scaled according to $\Omega=\Omega_{dimer}*N^{2/3}$, with N the number of building blocks present in the oligomer.

Size exclusion chromatography and multi-angle light scattering

Analytical SEC was performed using high performance liquid chromatography (HPLC) on a high molecular weight Shodex KW404-4F column (Showa Denko K.K.) pre-equilibrated for at least one column volume with a high salt running buffer (20 mM Tris-HCl pH 7.3, 500 mM NaCl). Samples were concentrated to 104 µM and 50 µL was injected. Every buffer was filtered three times through a 0.1 µm filter (Sartorius).

The MALLS set-up consisted of an on-line UV detector (Shimadzu), a DAWN® HELEOS® MALLS

detector (Wyatt Technology) and an Optilab® T-rEX refractive index detector (Wyatt Technology) attached to the Shodex KW404-4F column. Before sample injection, the inter-detector delay volumes were determined using a monomeric sample of BSA at 2.0 mg mL^{-1} (Pierce, Thermo Fisher scientific), which was buffer exchanged to the running buffer. Molecular weight calculations were performed using the ASTRA V software.

Dynamic Light Scattering

Dynamic light scattering was performed on a DynaPro Nanostar (Wyatt Technology). The sample was spun down for 15 minutes at 16200 g and degassed for 10 minutes. The curves were recorded in a plastic micro-cuvette at room temperature in 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$, pH 7.4) using a protein concentration of 26 µM. For each measurement ten acquisitions were taken (with 10 seconds for each acquisition). Thermal unfolding experiments were performed by increasing the temperature from 25°C to 80°C at a constant ramping speed of 1°C min^{-1}. Reversibility of the unfolding was checked by measuring the sample again after doing thermal unfolding. The DLS data were analyzed using the Dynamics V7 software package.

Small-angle X-ray scattering

The small-angle X-ray scattering (SAXS) curves were recorded at the SOLEIL synchrotron on the SWING beam line in Gif-sur-Yvette, France. PaaR2 was used at a concentration of 541 µM and the sample was run in HPLC mode at 15°C using the running buffer 20 mM Tris-HCl pH 7.3, 500 mM NaCl. The buffer scattering curves were averaged and subtracted from the sample scattering curves using the in-house Foxtrot platform (SWING beamline). The data were further processed using PRIMUS from the ATSAS software package (see Supporting Table S2) (33, 34). Different techniques were used to calculate the mass. The first method used is based on the Guinier analysis and uses equation 1 (35):

$$M_{exp} = \frac{I_{exp}(0).1385}{c} \qquad (1)$$

With $I_{exp}(0)$ and c respectively the extrapolated scattering intensity at zero angle and the concentration at the tip of the peak in the chromatogram corresponding to PaaR2 (36). Using GNOM, a P(r) (particle distance distribution) function was created which also gave an

estimate of R_G but on the basis of this P(r) function (37). With the Q_R method (38), the mass was calculated using the following formula:

$$MW = \frac{Q_R}{0.1231} \qquad (2)$$

In which $Q_R = V_C^2/R_G$. V_C is the ratio of the particle's *I(0)* to its total scattered intensity. Furthermore, using SAXS MoW, a molecular weight could be estimated using the P(r) function (39). Finally, with the Porod-Debye plot in scatter, the Porod volume was calculated (40, 41). When this volume is divided by 1.7, an approximation for the molecular weight is obtained (33).

CD spectroscopy

Far-UV CD spectra were recorded using a Jasco J-715 spectropolarimeter. The protein sample was spun down for 15 minutes at 16200 g and degassed for 10 minutes. Circular dichroism spectra were measured between 200-250 nm, with a scan rate of 50 nm min^{-1}, bandwidth of 1.0 nm and a resolution of 0.5 nm. Five accumulations were taken in PBS buffer with samples of PaaR2 at 1.3 and 13 µM concentrations placed in a 0.5 and 0.1 cm cuvettes, respectively. Thermal unfolding experiments were performed by increasing the temperature from 15°C to 95°C using different heating rates 1, 2 or 4 °C/min. Thermal unfolding appeared independent of heating rate and showed complete recovery of a signal after renaturation suggesting that the process is not kinetically limited and appears to be reversible. For further analysis of thermal unfolding scans with 1 °C/min were used measured at 222 nm. The mean residue ellipticity ($[\theta]$ in deg.cm^2.dmol^{-1}) was calculated from the raw CD data by normalizing for the concentration of protein and the number of residues:

$$[\theta] = \frac{\theta \cdot MM}{n \cdot C \cdot l} \qquad (3)$$

With *MM*, *n*, *C* and *l* representing the molecular weight (Da), the number of amino acids, the protein concentration (mg mL^{-1}) and the length of the cuvette (cm). Secondary structure content was estimated using BeStSEL(42, 43) and the programs provided on DichroWeb (44): K2D (45), CDSSTR (46) and Contin-LL (47) and compared to the sequence-derived secondary structure predictions using JPRED (48) and PSIPRED (49).

Fluorimetry

Intrinsic protein fluorescence emission spectra were measured between 300 and 420 nm using the Perkin Elmer LS 50 spectrofluorimeter (Perkin Elmer, USA) and an excitation wavelength of 290 nm. To avoid signal overload, protein samples at 13 and 1.3 µM were measured in 0.3 cm and 1 cm cuvettes, respectively. Emission spectra were recorded as a function of temperature in the 15-85 °C interval with the applied heating rate being 1 °C min^{-1}. For the analysis of thermal unfolding, fluorescence intensity at 355 nm (at the tryptophan emission maximum) as a function of temperature was used.

Differential scanning calorimetry

The calorimetric scans were performed using a MicroCal VP-DSC high-sensitivity differential scanning micro-calorimeter. The samples were spun down at 16200 g and degassed for 10 min. Thermal unfolding was performed with two different scanning rates: 1.5 and 1 °C min^{-1}, yielding very similar unfolding profiles. The reversibility of the thermal unfolding (checked by reheating the solution in the calorimeter cell after cooling from the up-scan run) was at least 95%. The sample consisted of PaaR2 at 23 µM in 1x PBS. The measured signal was corrected for the buffer contribution and normalized per mole of protein in the measuring cell to obtain the partial molar heat capacity ($\overline{C}_{p,2}$) of protein as a function of temperature. For further analysis $c_{p,2}$ was expressed as the excess heat capacity, $\Delta c_p = \overline{C}_{p,2} - \overline{C}_{p,int}$, where $\overline{C}_{p,int}$ represents an intrinsic heat capacity of protein. The total intrinsic heat capacity is the sum of the intrinsic heat capacities of the protein in different states: $\overline{C}_{p,int} = \sum \alpha_i \overline{C}_{p,i}$, where α_i represents a molar fraction of the protein in state *i.* In the measured temperature interval, $\overline{C}_{p,int}$ was approximated by the second order polynomial on T and fitted to the pre- and post-transition parts of the experimental $\overline{C}_{p,2}$.

Results and discussion

PaaR2 can be produced in an active form

Initially we expressed wild-type PaaR2 with a His-tag placed at its N-terminus. This construct, however, led to a protein preparation that when eluted from the Ni-NTA column was highly aggregated and carried a DNA contamination that could not be removed by additional chromatographic steps or by the addition of DNase K. We then designed a second construct where the histidine tag was moved to the C-terminus. No DNA contamination was observed here after Ni-NTA purification, but the protein showed a tendency towards aggregation, which was largely but not fully prevented by including 1 mM TCEP to all buffers. We reasoned that this could be caused by the presence of the single cysteine at position 120 and therefore constructed the mutant Cys120Ser (PaaR2C120S).

PaaR2C120S can be produced with significantly higher yields and, in contrast to wild-type PaaR2, does not aggregate over time. Both proteins migrate as a single band on SDS-PAGE (Supporting Figure S1a), show an identical CD spectrum (Supporting Figure S1b) when freshly purified and bind to the 97 bp intergenic region preceding the *paaR2-paaA2-parE2* operon with essentially the same affinity while no binding is detected for a random DNA fragment of the same length and with the same % GC contents (Figure 1). For these reasons, we chose to continue with PaaR2C120S for most of our experiments.

PaaR2 is a hierarchically organized octamer in solution

The molecular weight of PaaR2C120S, as determined by mass spectrometry is 15295.4 ± 2.0 Da, in close agreement with the theoretical molecular weight of 15296.3 Da. To assess the oligomeric state of PaaR2C120S, its absolute molecular weight was determined using SEC-MALLS. PaaR2C120S migrates as a single peak on a Shodex KW404-4F column in high salt conditions when injected at a concentration of 105 μM. The MALLS-estimated molecular weight of the corresponding molecular species (at an elution volume of 3.36 mL) is 126 kDa, corresponding closely to the theoretical mass of 122.37 kDa expected for an octamer (Figure 2a). It should be noted at this point that in SEC the C120S mutant and freshly prepared wild-type PaaR2 elute at essentially the same volume, indicating that the Cys120Ser mutation does not affect the oligomeric state of the protein (Supporting Figure S1c).

The octameric assembly of PaaR2C120S and its potential concentration dependency were further assessed by native mass spectrometry. We used protein concentrations over two

orders of magnitude ranging from 0.33 to 65 µM (Figures 2b and c). At concentrations over 3.3 µM a particle corresponding to an octamer was shown to be the most prevalent species (Figure 2b; MW_{exp} = 122.780 kDa vs MW_{theory} = 122.368 kDa), while lower concentrations were hallmarked by a mixture of PaaR2C120S monomers, dimers, and tetramers (Figure 2c). Additionally, the concentration analysis shows that the octamer to dimer ratio is rapidly increasing with concentration. To address the situation in solution, we performed additional dilution experiments using SEC. At 9 µM concentration, the sample elutes as an octamer. By decreasing the concentration to 0.5 µM, we observe that the elution volume starts to show a slight shift towards a larger elution volume indicating a dynamic equilibrium between octamer and oligomers of lower molecular weight (Supporting Figure S2a). The native MS also reveals that, once formed, the PaaR2C120S octamer is a rather stable and compact particle. At a concentration of 65 µM, the mass spectrum of octameric PaaR2C120S shows a single, compact charge state distribution, which is characteristic for a well-folded protein complex, with little to no significant structure disorder (50, 51).

The oligomeric state of PaaR2C120S in solution was also investigated using SAXS (Supporting Figure S3a and Supporting Table S2). PaaR2C120S was injected for HPLC-SAXS at a concentration of 541 µM, and the analysis was performed on scattering data from the top of the peak (127 µM). The $R_G/I(0)$ plot (Supporting Figure S3b) yields a very stable R_G as a function of the frame number which supports the above-mentioned data and suggests that the collected SAXS curves are characteristic of a single type of particle. Guinier analysis (Supporting Figure S3c) reveals that the particle has an R_G of around 51 Å, which is consistent with R_G-values obtained from p(r) and Guinier peak analysis (Supporting Table S2). Additionally, molecular mass estimations based on I(0) extrapolation and the Porod volume are consistent with a PaaR2C120S octamer and are in agreement with the results from SEC-MALS and native MS.

The PaaR2 octamer has a compact non-spherical structure

Further analysis of the HPLC-SAXS data indicates that the PaaR2C120S octamer is compact. The normalised Kratky plot is characteristic of a well-folded particle with a low degree of flexibility (Figure 3a). Moreover, the Porod-Debye plot illustrates that a clear Porod plateau is reached and the Porod exponent is estimated at 3.9, which is consistent with a high level of compactness (Supporting Figure S3d). Together with the native MS data mentioned earlier,

these findings support that the PaaR2C120S octamer is a stable, compact particle displaying very little flexibility.

More insights in the overall architecture of the PaaR2C120S octamer can be obtained from the R_G power law for protein oligomers (52). However, it seems that the PaaR2 octamer does not obey this power law suggesting that the particle has a peculiar architecture. Particles typically deviating from the R_G power law for protein oligomers are either very elongated or display a dihedral or cyclic symmetry. While the R_G power law was derived for hexamers with dihedral and cyclic symmetries (52), this remains unexplored for octamers with this type of symmetry. Based on the P(r) function, it is clear that the PaaR2C120S octamer is not elongated (Figure 3b). This can be quantified by the elongation ratio (E.R.) proposed by Putnam (53). Typically, symmetric objects have an E.R. around 1.0, while very extended particles possess very high E.R. values (10 or higher). The PaaR2 octamer has an E.R. of 1.3, which is reminiscent of a symmetric particle but not consistent with a perfect sphere (E.R. = 0.94). For the PaaR2 octamer, the power law exponent is caculated to be 0.467, which deviates substantially from the exponent generally found for protein monomers and oligomers (0.40). This would point towards an octamer with a cyclic symmetry given that, for a given number of residues, cyclic oligomers tend to have larger R_G values than other symmetries (52).

Additional support that PaaR2C120S has a symmetric, non-spherical shape comes from analytical SEC and DLS data and their respective comparison with the SEC-MALLS and SAXS data. The analytical SEC profile was measured at concentrations that correspond to a population consisting of an octameric complex (104 µM injected concentration). The molecular mass extrapolated from the elution volume of PaaR2C120S on the Shodex KW404-4F column is significantly higher (215 kDa) than that obtained from SEC-MALLS (110 to 126 kDa) (Figure 2a). Importantly, it should be noted that SEC-MALLS provides an absolute measurement of the molecular weight and is not influenced by the shape of the molecule. Therefore, this discrepancy between the molecular weight estimates derived from analytical SEC and SEC-MALLS could be caused by the PaaR2C120S octamer possessing a non-spherical shape. This is further advocated by a comparison of the radius of gyration (R_G; obtained via SAXS) of PaaR2C120S and its hydrodynamic radius (R_H; determined by DLS and analytical SEC). The R_H of PaaR2C120S as calculated from the SEC data is 55 Å. This is consistent with DLS measurements at a concentration of 29 µM, which provide an R_H estimate of 54 Å (Figure 2c). The R_G/R_H ratio is 0.775 and 1.1 for perfect spheres and anisometric polymers, respectively.

For PaaR2C120S, the R_G/R_H ratio is 0.94. Together, these data indicate that the PaaR2C120S octamer is compact and possesses a symmetric, non-spherical shape. This peculiar architecture most likely explains why there is such a discrepancy between the different methods for SAXS-based molecular mass determination (Supporting Table S2) (54).

Finally, we turned to ion mobility-mass spectrometry to gain additional structural information about the global shape of the molecule. Ion mobility in conjunction with mass spectrometry allows for separation of conformations based on mass, charge and global shape, by separating ions in a gas filled tube under the influence of an electric field (55). For PaaR2C120S we obtained collision cross sections (CCS) of 2472 Å^2, 3876 Å^2 and 6319 Å^2 for the dimer, tetramer and octamer, respectively. It is possible to predict the collision cross sections for different shapes of a molecule as a function of increasing size. Figure 3d shows the obtained experimental CCS values for PaaR2C120S in relation to a linear, spherical and isotropic (equal in all directions) growth models. It becomes apparent that both the linear and spherical models do not match the observed experimental CCS values, but that PaaR2C120S oligomerization follows isotropic growth. These results further reinforce the conclusion that PaaR2C120S is a symmetric non-spherical compact particle with little flexibility.

Thermal unfolding of PaaR2 is reversible and displays several transitions suggesting a complex unfolding mechanism

Circular dichrosim (CD) spectra of both wild-type PaaR2 and PaaR2C120S bear a shape characteristic of α-helical proteins with minima at 208 and 222 nm (Figure 4a, Supporting Figure S1b). Deconvolution of the CD spectrum using different methods suggests an α-helical content of about 67% (Table 1). This compares reasonably well to values of α-helical content determined from amino acid sequence-based secondary structure prediction (JPRED 55% and PSIPRED 56% - Table 1 and Supporting Figure S4). In addition, these algorithms predict that the protein contains an N-terminal likely helix-turn-helix motif and a C-terminal helix predicted to form a coiled coil.

Increasing the temperature leads to the unfolding of the protein, with CD spectra at 92°C indicating that the majority of secondary structure is random-coil and that two thirds of the α-helical structure are lost (Figure 4a). Cooling the sample back to 25°C leads to complete recovery of signal, suggesting a reversible unfolding/refolding process for both wild-type and mutant PaaR2 (Figure 4a, Supporting Figure S5). This was also confirmed for PaaR2C120S

while monitoring DLS unfolding (Supporting Figure S6). As the temperature is increased (until 85°C), a homogeneous population of particles showing a larger decay time from the autocorrelation curve is observed. When the sample is cooled down, even after several heating cycles, the autocorrelation curve at room temperature again corresponds to the initial curve, indicating a reversible transition. Following the change in ellipticity as a function of temperature and protein concentration reveals three transitions, of which the first and third are concentration-dependent (Figure 4b). The second, sharper, transition at around 67°C appears to be concentration-independent.

Given the complexity of the process, we gathered additional information by following the tryptophan fluorescence as a function of temperature. Sequence analysis reveals that one tryptophan residue is located in the predicted helix-turn-helix domain, while the other residue is part of the C-terminal coiled coil (Supporting Figure S4). The native protein shows a fluorescence spectrum with an emission maximum at 345 nm (Figure 4c). Thermal unfolding leads to a strong decrease of the fluorescence emission intensity and a red-shift of the emission maximum (Figure 4c). Both observations indicate exposure of tryptophan residues to a more polar surrounding. Again, refolding leads to the recovery of the fluorescence emission spectrum, suggesting reversibility of the unfolding/refolding process. Change of the emission intensity as a function of temperature and protein concentration also suggests a multi-step process, but with some differences with respect to what was observed in CD. In fluorescence, an additional low-temperature concentration-dependent transition is observed while the last transition corresponds to the second concentration-independent transition observed in CD (Figure 4d).

These observations suggest that the native octamer dissociates without any loss of secondary structure but with some changes in the polarity of tryptophan surrounding. This is followed by the second dissociation event observed by both CD and fluorescence. A third transition recorded by CD and fluorescence is the most cooperative and does not involve a change of oligomeric state (concentration independent). The final transition again involves dissociation and slight loss of secondary structure, but no further change in the fluorescence signal.

A global thermodynamic model describes a five-state unfolding process

Unfolding of PaaR2C120S was further followed using differential scanning calorimetry (DSC) (Figure 5) and the DSC data were, together with the CD and fluorescence data, integrated into

a model of the reversible unfolding of PaaR2C120S. Based on the available information, the simplest model for the thermal unfolding of PaaR2C120S that best describes the data from CD, fluorescence and DSC seems to involve four distinct transitions as shown schematically in Figure 5A. Additional models of varying complexity were also tested, however none of them can adequately describe the FL, CD and DSC simultaneously (Supporting Figure S7, Supporting Table S5). A detailed explanation on model analysis and model selection is described in the Supporting methods section. Briefly, for each transition i (equations 4-6 (see Supporting Materials and Methods) the model parameters ΔG_i, ΔH_i and $\Delta c_{p,i}$ (standard Gibbs free energy, enthalpy and heat capacity) given per mole of monomer at 25°C were first obtained by fitting the model function (equations 7 and 8 (see Supporting Materials and Methods)) to each dataset (CD, fluorescence and DSC) separately, providing reasonable initial estimates of the model parameters. However, given the number of parameters required to describe this complex unfolding process, some of them are strongly correlated. This can be partially avoided by fitting the model function to all data simultaneously (equation 9 (see Supporting Materials and Methods), Supporting Figure S8). The final set of thermodynamic parameters from the global fitting of FL, CD and DSC data is shown in Figure 6b and Supporting Table S3. The validity and robustness of model parameters was further assessed using Monte Carlo error analysis providing estimated parameter distributions and correlations (Supporting Figure S9 and Supporting Tables S3 and S4). It should be noted that the global fit shows some systematic discrepancies and that goodness-of-fit statistics indicate that the model is not perfect (Supporting Table S5). This is likely due to certain simplifications aimed to reduce the number of fitting parameters (further discussed in Supporting Information).

The DSC model function (eq. 8 (see Supporting Materials and Methods)) provides a good description of the DSC experiment (Figure 5) which, unlike spectroscopic experiments, detects all unfolding transitions. The first endothermic peak involves transition of the octamer to tetramer, dimer and dimer intermediate, while the second peak at high temperature corresponds to transition of dimer intermediate into unfolded monomer (Figure 5b). At 25°C the dissociation of the octamer (P_8) to the tetramer (P_4) and then to the dimer (P_2) does not involve significant changes in enthalpy (Figure 6b), likely because exposure of polar and apolar surfaces to water have opposing effects on the enthalpy change, resulting in an overall enthalpy compensation. On the other hand, conversion of the folded dimer (P_2) into a largely unfolded dimer intermediate (I_2) is enthalpically unfavorable and results in a species with little

native secondary structure. This dimeric intermediate with solvent exposed tryptophan residues dissociates into a final monomeric unfolded form, a process which is also associated by a positive enthalpy change. The total enthalpy change for unfolding is larger than expected for an average globular protein from the dataset of Robertson and Murphy (145 vs. 90 kcal mol^{-1} at the dataset reference temperature of 60 °C) (56). The Robertson and Murphy dataset deals mainly with monomeric proteins and therefore a larger enthalpy change might reflect favorable interactions between PaaR2C120S monomers in the octamer. The estimated heat capacity contributions of each unfolding step are subjected to considerable uncertainties due to strong parameter correlations (in Supporting Table S4) (57). However, a reasonable estimate can be obtained for the overall heat capacity change for the total unfolding. This overall heat capacity change is positive (indicating a general exposure of apolar surface upon unfolding) but smaller in value compared to the average for globular proteins in the Robertson and Murphy dataset (0.5 vs. 1.7 kcal mol^{-1}K^{-1}). This can be explained through the compensating effect of heat capacity contributions: exposure of polar surfaces, which probably accompanies oligomer dissociation, results in negative heat capacity changes while the change is positive for the exposure of apolar surface during the later steps of unfolding (58).

A stability phase diagram provides insights into the behavior of PaaR2 under physiological conditions and its putative role as a transcriptional regulator

Based on the model parameters, we calculated which species is predominant as a function of protein concentration and temperature resulting in a protein stability phase diagram presented on Figure 6c. Regions of the phase diagram are colored according to which molecular species has the highest molar fraction at a given temperature and protein concentration (e.g. a region with 40% tetramer, and 30% of each octamer and dimer has an orange color). Below 60°C, the protein exists in a wide concentration range in an octamer-tetramer-dimer equilibrium, with dimer as the predominant species. Although native mass measurements are not expected to correlate quantitatively with a situation in solution, we observe a reasonable agreement with the protein stability phase diagram. The octameric population is dominant at 10 μM and above but decreases rapidly at lower concentrations. Furthermore, the SEC dilution experiment shows that samples at 10 μM concentration contain only octamers, while samples at 0.5 μM concentration are a mixture of oligomers

(octamer, tetramer, dimer), in good agreement with the model-predicted species populations (Supporting Figure S2).

Of interest is also the observation that the octamer dissociates to dimer with decreasing temperature (Figure 6b). This is reminiscent of the cold denaturation phenomenon observed for some monomeric proteins and more commonly for multimeric proteins and supramolecular structures (59). It appears that the positive heat capacity values combined with a relatively low enthalpy change for the first two transitions (P_8>P_4>P_2, ΔH = 2 kcal/mol, Δc_P>0) leads to dissociation of the octamer with decreasing temperature. The physiological concentration of the PaaR2 protein is not known, however a recent absolute quantification of *E. coli* proteins reports a median value of 67 protein copies per cell (around 10^{-7} M) (60). At physiological temperature 37 °C and assuming 100 PaaR2 copies in the cell this corresponds to an area of the phase diagram (indicated in Figure 6c) with highest dimer fraction but also a significant fraction of tetramer and octamer (the corresponding fractions are: α_{P2}=0.5, α_{P4}=0.4, α_{P8}=0.1). In this area, a shift in the protein concentration by an order of magnitude results in a complete redistribution of protein species which may be relevant for protein function. Given that the protein should bind to the operator DNA sequence, the binding could promote the oligomerization due to the increase in local protein concentration (61). It has been shown that DNA binding by monomeric transcription factors and subsequent oligomerization offers a kinetic discrimination of specific DNA sites in the large pool of nonspecific DNA (62).

Conclusion

A biophysical analysis through a combination of SAXS, native mass spectrometry and SEC-MALLS has shown that the PaaR2 regulator from the *paaR2-paaA2-parE2* operon in *E. coli* O157:H7 forms a complex consisting of eight subunits in solution. This complex falls apart in populations consisting predominantly of a tetramer and finally a dimer at lower concentrations, suggesting that this is the pathway in which the octamer is built. This hierarchical structure is corroborated by a complex five-state thermal unfolding pathway that involves monomer, dimer, tetramer and octamer species of variable secondary structure contents. Although there is no sequence similarity, this architecture is reminiscent to that of the Leucine Responsive Protein family of transcription regulators.

Supporting citations

References (63–67) appear in the Supporting Material.

References

1. Hayes, F., and L. Van Melderen. 2011. Toxins-antitoxins: diversity, evolution and function. Crit. Rev. Biochem. Mol. Biol. 46: 386–408.

2. Page, R., and W. Peti. 2016. Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat. Chem. Biol. 12: 208–214.

3. Harms, A., D.E. Brodersen, N. Mitarai, and K. Gerdes. 2018. toxins, targets, and triggers: An overview of toxin-antitoxin biology. Mol. Cell. 70: 768–784.

4. Gerdes, K., P.B. Rasmussen, and S. Molin. 1986. Unique type of plasmid maintenance function: Postsegregational killing of plasmid-free cells. Proc. Natl. Acad. Sci. 83: 3116–3120.

5. Melderen, L. Van, P. Bernard, and M. Couturier. 1994. Lon-dependent proteolysis of

CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. Mol. Microbiol. 11: 1151–1157.

6.   Tsilibaris, V., G. Maenhaut-Michel, N. Mine, and L. Van Melderen. 2007. What is the benefit to *Escherichia coli* of having multiple toxin-antitoxin systems in its genome? J. Bacteriol. 189: 6101–6108.

7.   Gerdes, K., S.K. Christensen, and A. Løbner-Olesen. 2005. Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol. 3: 371–382.

8.   Amitai, S., Y. Yassin, and H. Engelberg-kulka. 2004. MazF-mediated cell death in *Escherichia coli* : A point of no return. J. Bacteriol. 186: 8295–8300.

9.   Korch, S.B., and T.M. Hill. 2006. Ectopic overexpression of wild-type and mutant *hipA* genes in *Escherichia coli*: Effects on macromolecular synthesis and persister formation. 188: 3826–3836.

10.   Pedersen, K., S.K. Christensen, and K. Gerdes. 2002. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. Mol. Microbiol. 45: 501–510.

11.   Harms, A., C. Fino, M.A. Sørensen, S. Semsey, and K. Gerdes. 2017. Prophages and growth dynamics confound experimental results with antibiotic-tolerant persister cells. MBio. 8: 1–18.

12.   Goormaghtigh, F., N. Fraikin, M. Putrinš, T. Hallaert, V. Hauryliuk, A. Garcia-Pino, A. Sjödin, S. Kasvandik, K. Udekwu, T. Tenson, N. Kaldalu, and L. Van Melderen. 2018. Reassessing the role of type II toxin-antitoxin systems in formation of *Escherichia coli* type II persister cells. MBio. 9: 1–14.

13.   Tripathi, A., P.C. Dewan, B. Barua, and R. Varadarajan. 2012. Additional role for the *ccd* operon of F-plasmid as a transmissible persistence factor. Proc. Natl. Acad. Sci. U. S. A. 109: 12497–12502.

14.   Tripathi, A., P.C. Dewan, S.A. Siddique, and R. Varadarajan. 2014. MazF-induced growth inhibition and persister generation in *Escherichia coli*. J. Biol. Chem. 289: 4191–4205.

15.   Helaine, S., A.M. Cheverton, K.G. Watson, L.M. Faure, S.A. Matthews, and D.W. Holden. 2014. Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. Science (80-.). 343: 204–208.

16.   Loris, R., and A. Garcia-Pino. 2014. Disorder- and dynamics-based regulatory

mechanisms in toxin-antitoxin modules. Chem. Rev. 114: 6933–6947.

17. Overgaard, M., J. Borch, M.G. Jørgensen, and K. Gerdes. 2008. Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. Mol. Microbiol. 69: 841–857.

18. De Jonge, N., A. Garcia-Pino, L. Buts, S. Haesaerts, D. Charlier, K. Zangger, L. Wyns, H. De Greve, and R. Loris. 2009. Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. Mol. Cell. 35: 154–163.

19. Garcia-Pino, A., S. Balasubramanian, L. Wyns, E. Gazit, H. De Greve, R.D. Magnuson, D. Charlier, N. a J. van Nuland, and R. Loris. 2010. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell. 142: 101–111.

20. Schureck, M.A., T. Maehigashi, S.J. Miles, J. Marquez, S.E. Cho, R. Erdman, and C.M. Dunham. 2013. Structure of the *Proteus vulgaris* HigB-(HigA) 2 -HigB toxin-antitoxin complex. J. Biol. Chem. 289: 1060–1070.

21. Talavera, A., H. Tamman, A. Ainelo, A. Konijnenberg, S. Hadži, F. Sobott, A. Garcia-Pino, R. Hõrak, and R. Loris. 2019. A dual role in regulation and toxicity for the disordered N-terminus of the toxin GraT. Nat. Commun. 10: 1–13.

22. Meinhart, A., J.C. Alonso, and N. Stra. 2003. Crystal structure of the plasmid maintenance system ε/ζ: Functional mechanism of toxin ζ and inactivation by $\varepsilon_2\zeta_2$ complex formation. Proc. Natl. Acad. Sci. 100: 1661–1666.

23. Khoo, S.K., B. Loll, W.T. Chan, R.L. Shoeman, L. Ngoo, C.C. Yeo, and A. Meinhart. 2007. Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. J. Biol. Chem. 282: 19606–19618.

24. de la Hoz, A.B., F. Pratto, R. Misselwitz, C. Speck, W. Weihofen, K. Welfle, W. Saenger, H. Welfle, and J.C. Alonso. 2004. Recognition of DNA by ω protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: Analysis of binding to operator DNA with one to four heptad repeats. Nucleic Acids Res. 32: 3136–3147.

25. Weihofen, W.A., A. Cicek, F. Pratto, J.C. Alonso, and W. Saenger. 2006. Structures of ω repressors bound to direct and inverted DNA repeats explain modulation of transcription. Nucleic Acids Res. 34: 1450–1458.

26. Hallez, R., D. Geeraerts, Y. Sterckx, N. Mine, R. Loris, and L. Van Melderen. 2010. New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in

Escherichia coli O157:H7. Mol. Microbiol. 76: 719–732.

27. Sterckx, Y.G.J., A.N. Volkov, W.F. Vranken, J. Kragelj, M.R. Jensen, L. Buts, A. Garcia-Pino, T. Jové, L. Van Melderen, M. Blackledge, N. a J. van Nuland, and R. Loris. 2014. Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. Structure. 22: 854–865.

28. Sterckx, Y.G.-J., T. Jové, A. V. Shkumatov, A. Garcia-Pino, L. Geerts, M. De Kerpel, J. Lah, H. De Greve, L. Van Melderen, and R. Loris. 2016. A unique hetero-hexadecameric architecture displayed by the *E. coli* O157 PaaA2-ParE2 antitoxin-toxin complex. J. Mol. Biol. 428: 1589–1603.

29. Dalton, K.M., and S. Crosson. 2010. A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. Biochemistry. 49: 2205–2215.

30. Bush, M.F., I.D.G. Campuzano, and C. V Robinson. 2012. Ion mobility mass spectrometry of peptide ions: Effects of drift gas and calibration strategies. Anal. Chem. 84: 7124–7130.

31. Knapman, T.W., V.L. Morton, N.J. Stonehouse, P.G. Stockley, and A.E. Ashcroft. 2010. Determining the topology of virus assembly intermediates using ion mobility spectrometry – mass spectrometry. Rapid Commun Mass Spectrom. 24: 3033–3042.

32. Young, L.M., J.C. Saunders, R.A. Mahood, C.H. Revill, R.J. Foster, A.E. Ashcroft, and S.E. Radford. 2016. ESI-IMS – MS : A method for rapid analysis of protein aggregation and its inhibition by small molecules. Methods. 95: 62–69.

33. Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P. V. Konarev, and D.I. Svergun. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. 45: 342–350.

34. Konarev, P. V., V. V. Volkov, A. V. Sokolova, M.H.J. Koch, and D.I. Svergun. 2003. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36: 1277–1282.

35. Mylonas, E., and D.I. Svergun. 2007. Accuracy of molecular mass determination of proteins in solution by small-angle x-rqay scattering. J. Appl. Crystallogr. 40: 245–249.

36. Orthaber, D., and O. Glatter. 2000. Synthetic phospholipid analogs : a structural investigation with scattering methods. Chem. Phys. Lipids. 107: 179–189.

37. Svergun, D.I. 1992. Determination of the regularization parameter in indirect-

transform methods using perceptual criteria. J. Appl. Crystallogr. 25: 495–503.

38. Rambo, R.P., and J. a Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. Nature. 496: 477–481.

39. Fischer, H., M. de Oliveira Neto, H.B. Napolitano, I. Polikarpov, and a. F. Craievich. 2010. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J. Appl. Crystallogr. 43: 101–109.

40. Porod, G. 1952. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. Kolloid-Zeitschrift. 125: 51–57.

41. Rambo, R.P., and J.A. Tainer. 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers. 95: 559–571.

42. Micsonai, A., F. Wien, L. Kernya, Y. Lee, Y. Goto, M. Réfrégiers, and J. Kardos. 2015. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc. Natl. Acad. Sci. 112: 3095–3103.

43. Micsonai, A., F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos. 2018. BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Res. 46: W315–W322.

44. Whitmore, L., and B.A. Wallace. 2004. DICHROWEB , an online server for protein secondary structure analyses from circular dichroism spectroscopic data. 32: 668–673.

45. Andrade, M., P. Chacón, J.J. Merelo, and F. Morán. 1993. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. Protein Eng. 6: 383–390.

46. Sreerama, N., and R.W. Woody. 2000. Estimation of protein secondary structure from circular dichroism spectra : Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal. Biochem. 287: 252–260.

47. van Stokkum, I.H.M., H.J.W. Spoelder, M. Bloemendal, R. van Grondelle, and F.C.A. Groen. 1990. Estimation of protein secondary structure analysis from circular dichroism spectra and error. Anal. Biochem. 191: 110–118.

48. Cole, C., J.D. Barber, and G.J. Barton. 2008. The Jpred 3 secondary structure

prediction server. Nucleic Acids Res. 36: 197–201.

49.　Buchan, D.W.A., F. Minneci, T.C.O. Nugent, K. Bryson, and D.T. Jones. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 41: 349–357.

50.　Testa, L., S. Brocca, and R. Grandori. 2011. Charge-surface correlation in electrospray ionization of folded and unfolded proteins. Anal. Chem. 83: 6459–6463.

51.　Testa, L., S. Brocca, C. Santambrogio, A. D'Urzo, J. Habchi, S. Longhi, V.N. Uversky, and R. Grandori. 2013. Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. Intrinsically Disord. Proteins. 1: 1–7.

52.　Tanner, J.J. 2016. Empirical power laws for the radii of gyration of protein oligomers. Acta Crystallogr. Sect. D Biol. Crystallogr. D72: 1119–1129.

53.　Putnam, C.D. 2016. Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data research papers. J. Appl. Crystallogr. 49: 1412–1419.

54.　Korasick, D.A., and J.J. Tanner. 2018. Determination of protein oligomeric structure from small-angle X-ray scattering. Protein Sci. 27: 814–824.

55.　Konijnenberg, A., A. Butterer, and F. Sobott. 2013. Native ion mobility-mass spectrometry and related methods in structural biology. Biochim. Biophys. Acta. 1834: 1239–1256.

56.　Robertson, A.D., and K.P. Murphy. 1997. Protein structure and the energetics of protein stability. Chem. Rev. 97: 1251–1268.

57.　Combet, C., C. Blanchet, C. Geourjon, and G. Deléage. 2000. NPS@: Network Protein Sequence Analysis. Trends Biochem. Sci. 0004: 147–150.

58.　Murphy, K.P., and E. Freire. 1992. Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv. Protein Chem. 43: 313–361.

59.　Privalov, P.L. 1990. Cold denaturation of protein. Crit. Rev. Biochem. Mol. Biol. 25: 281–306.

60.　Wiśniewski, J.R., and D. Rakus. 2014. Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the Escherichia coli proteome. J. Proteomics. 109: 322–331.

61.　Oehler, S., and B. Müller-Hill. 2010. High local concentration: A fundamental strategy of life. J. Mol. Biol. 395: 242–253.

62. Kohler, J.J., S.J. Metallo, T.L. Schneider, and A. Schepartz. 1999. DNA specificity enhanced by sequential binding of protein monomers. Proc. Natl. Acad. Sci. U. S. A. 96: 11735–11739.

63. Wales, D.J., and J.P.K. Doye. 1997. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A. 101: 5111–5116.

64. Burnam, K.P., and D.R. Anderson. 2002. Model selection and inference: A practical information-theoretic approach. 2nd edition. Springer-Verlag.

65. Drobnak, I., G. Vesnaver, and J. Lah. 2010. Model-based thermodynamic analysis of reversible unfolding processes. J. Phys. Chem. B. 114: 8713–8722.

66. Powell, M.J.D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput. J. 7: 155–162.

67. Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 2002. Numerical recipes in C++: the art of scientific computing, 2nd Edit. Cambridge Univ. Press. New York, NY.

Table 1. Analysis of secondary structure content of PaaR2 determined by CD spectroscopy and amino acid sequence analysis.

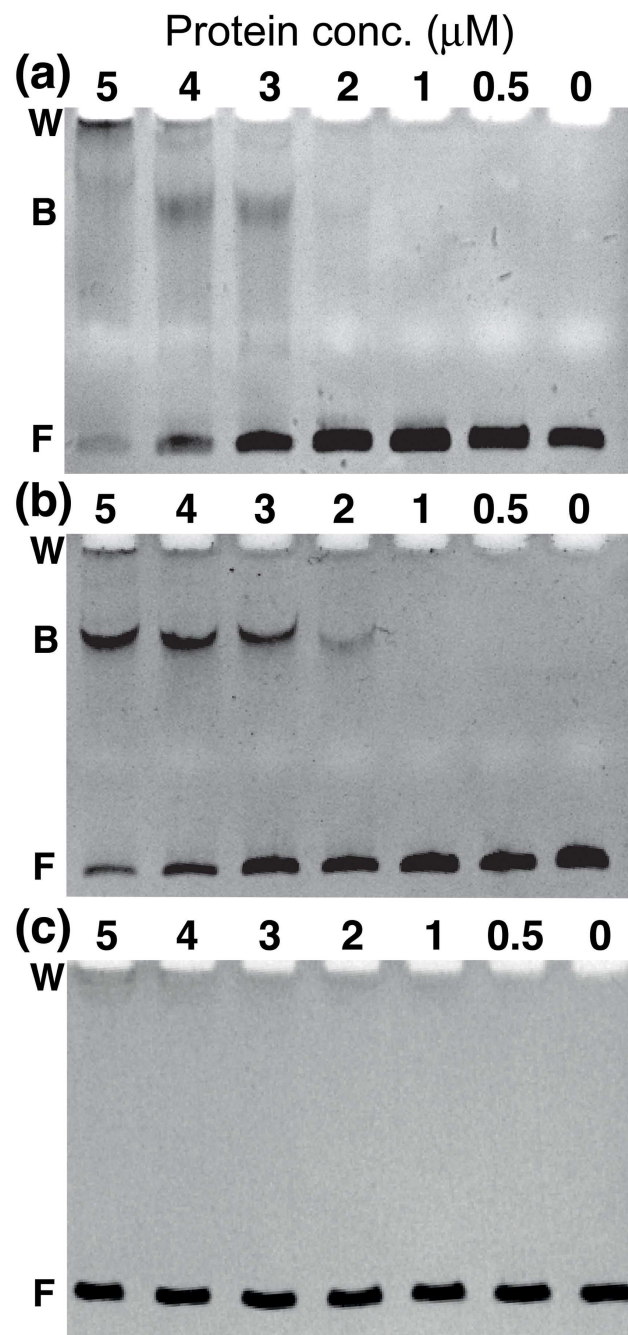| Analysis	α (%)	β (%)	coil (%)	NRMSD
K2D	68.0	4.0	28.0	0.069
Contin-LL	67.1	4.3	28.3	0.344
CDSSTR	68.6	10.8	20.6	0.004
BeStSel	62.6	0.0	37.3	0.004
Average	66.6	4.8	28.6	-
JPRED	54.6	0.0	45.4	-
PSIPRED	57.6	0.0	42.4	-

Figures

Figure 1. EMSA binding studies of PaaR2. EMSA assays were performed in 1x PBS (1 mM TCEP was added for wild-type PaaR2). On top of each lane, the protein concentration is given in μM while the DNA concentration is fixed to 0.25 μM. Free DNA is labeled F while B corresponds to the bound species and W to the bottom of the wells. **(a)** Titration of wild-type PaaR2 to the putative operator region. **(b)** Identical titration of the Paar2C120S mutant to the same DNA fragment. **(c)** Titration of Paar2C120S to a control DNA fragment of the same length and with the same GC contents as the operator fragment. Both wild-type and mutant PaaR2 start to show binding from 2 μM (monomer equivalent concentration) onwards while no binding is observed for the control segment over the full concentration range.
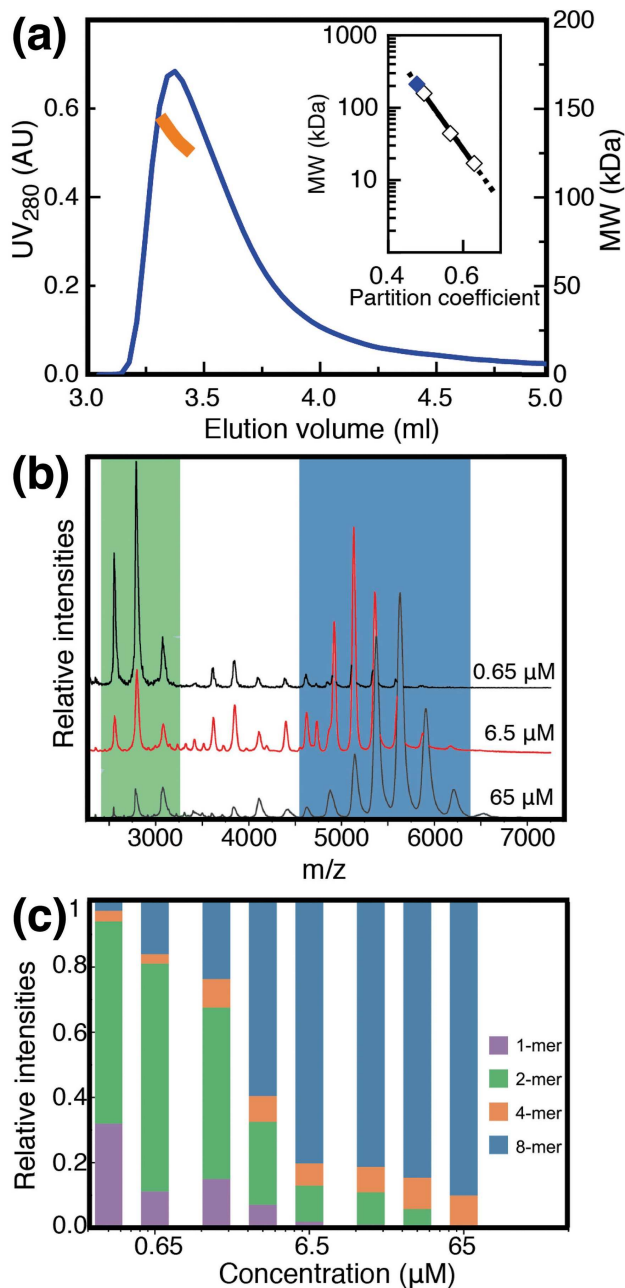
Figure 2. Oligomeric composition of PaaR2. (a) SEC-MALLS profile of PaaR2. SEC measurements were performed on a Shodex KW404-4F column (104 μM in 20 mM Tris-HCl pH 7.3, 500 mM NaCl). The elution profile of PaaR2 is shown in blue while the molecular weight curve determined by MALLS is shown in orange. The inset shows the molecular weight estimation using the elution volumes and molecular weights of the Bio-Rad Size Exclusion standard, from which the molecular weights of bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa) and horse myoglobin (17 kDa) are within the linear range of the column and indicated on the figure (white diamonds). The elution volume of PaaR2 was plotted on this curve (blue diamond), leading to a MW estimate of 215 kDa. **(b)** Native mass spectra for PaaR2 at three different concentrations (65 μM, 6.5 μM and 0.65 μM). Peaks corresponding to dimer and octamer are indicated respectively in green and blue. **(c)** Concentration-dependent species distribution as determined by native mass spectrometry. At high concentrations, the octamer is dominant while below 2 μM the dimer becomes the dominant species.
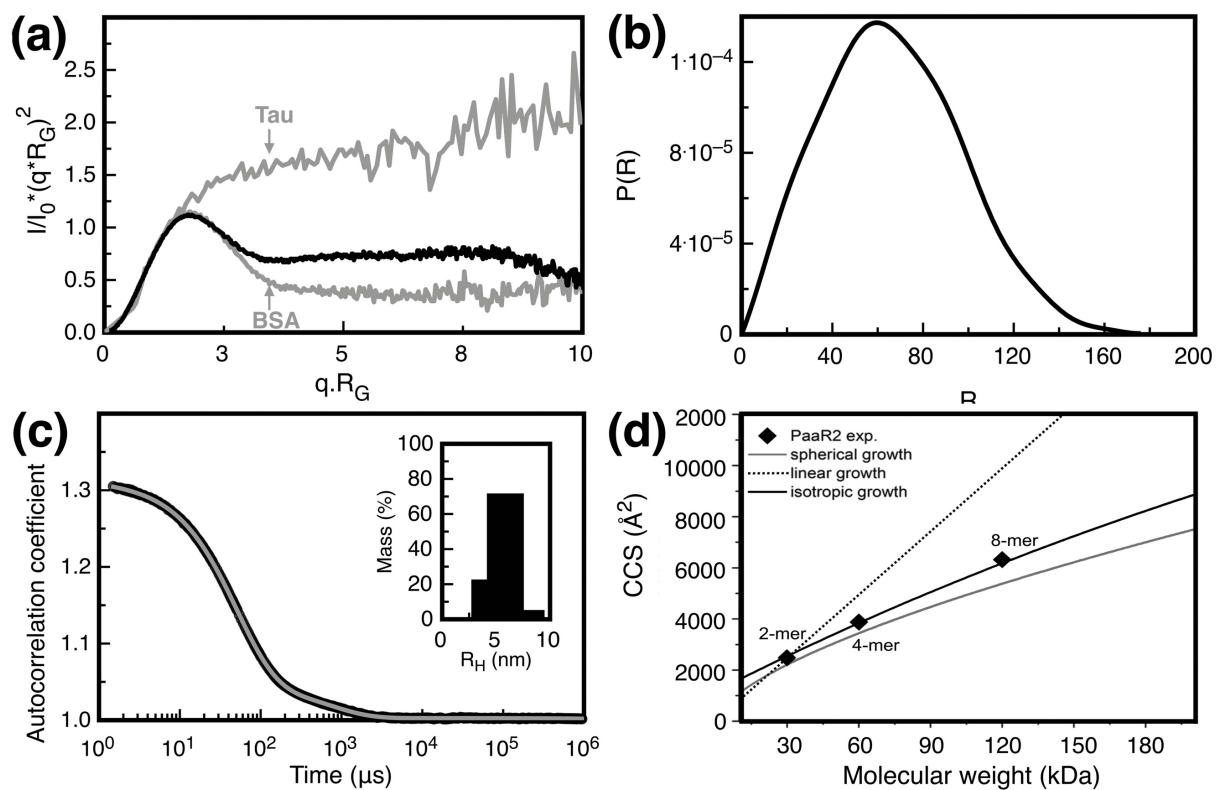
Figure 3. Non-spherical behavior of PaaR2. (a) Kratky plot of the SAXS curve for PaaR2 (black). As a reference, the Kratky plots from Tau (IDP) and BSA (spherical protein) (both in grey) are also given. **(b)** Distance distribution function (P(r)) at the maximum dimension value (D_{max}) of 175 Å. **(c)** DLS autocorrelation curve of PaaR2 at room temperature (black). The regularization fit is shown in grey. The inset shows the regularization graph in mass percentage for the native protein prior to unfolding. The values for R_H and the % of polydispersity of the protein at room temperature are 5.3 nm and 14.8% respectively. **(d)** Experimental collision cross sections (CCS) obtained for the PaaR2 dimer, tetramer and octamer (black diamonds) in relation the theoretical models of a linear (dotted line), isotropic (black) or spherical (gray) growth model.
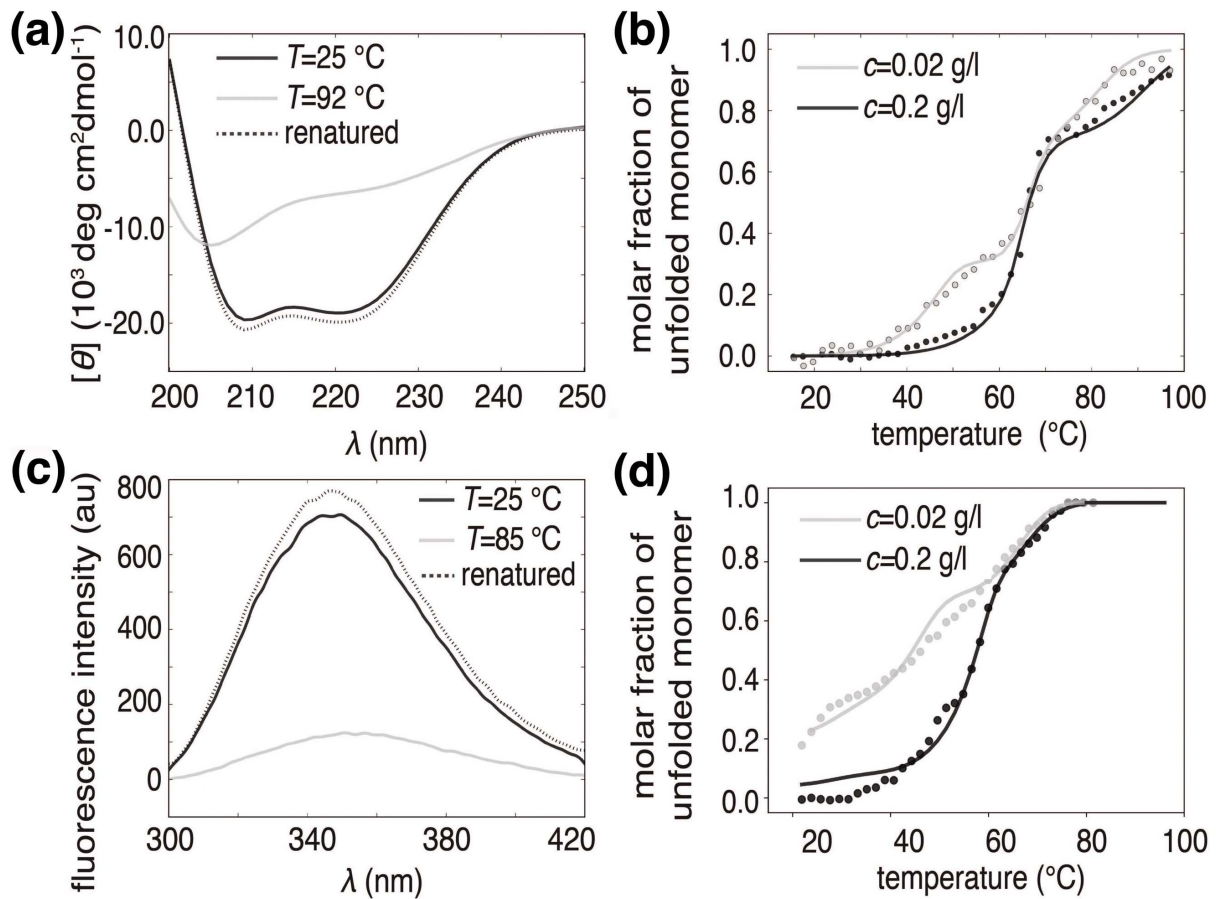
Figure 4. Thermal unfolding followed by spectroscopic methods. (a) Circular dichroism spectra of PaaR2 at 13 μM in PBS buffer. The protein has a characteristic α-helical CD spectrum at 25 °C. Around two thirds of native secondary structure are lost upon increasing the temperature to 92 °C, which is completely recovered after cooling back to 25°C. **(b)** Normalized ellipticity measured at 222 nm as a function of temperature (dots) and fits of the model function to the CD data (solid lines). Melting curves for two protein concentrations (13 and 1.3 μM, dark gray and light gray) are not identical indicating a change in the oligomerization state. **(c)** Tryptophan fluorescence of PaaR2 at 13 μM in PBS buffer. A loss of intensity and red-shift of emission maximum is observed in heat-unfolded sample (gray curve, 85°C) compared to native protein (black curve, 25°C). **(d)** Normalized tryptophan fluorescence from the emission maximum (355 nm) as a function of temperature (dots) and fits of the model function to the fluorescence (solid line).
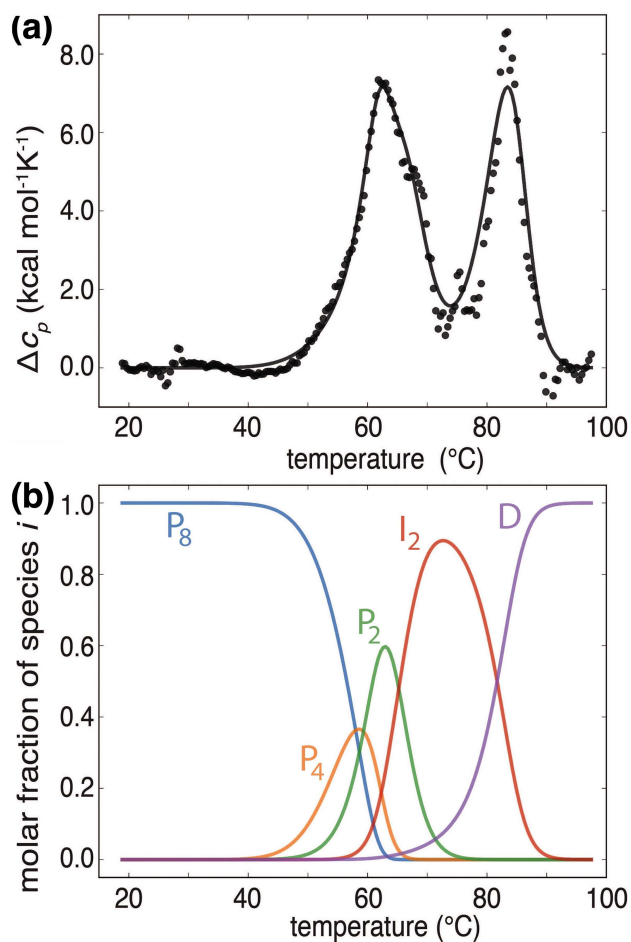
Figure 5. **Thermal unfolding followed by differential scanning calorimetry. (a)** Excess heat capacity of PaaR2 at 23 µM in PBS (dots) and fits of the model function (solid line). **(b)** Speciation diagram corresponding to the DSC scan in panel above is based on the best fit model parameters obtained from fitting the model function to DSC data.
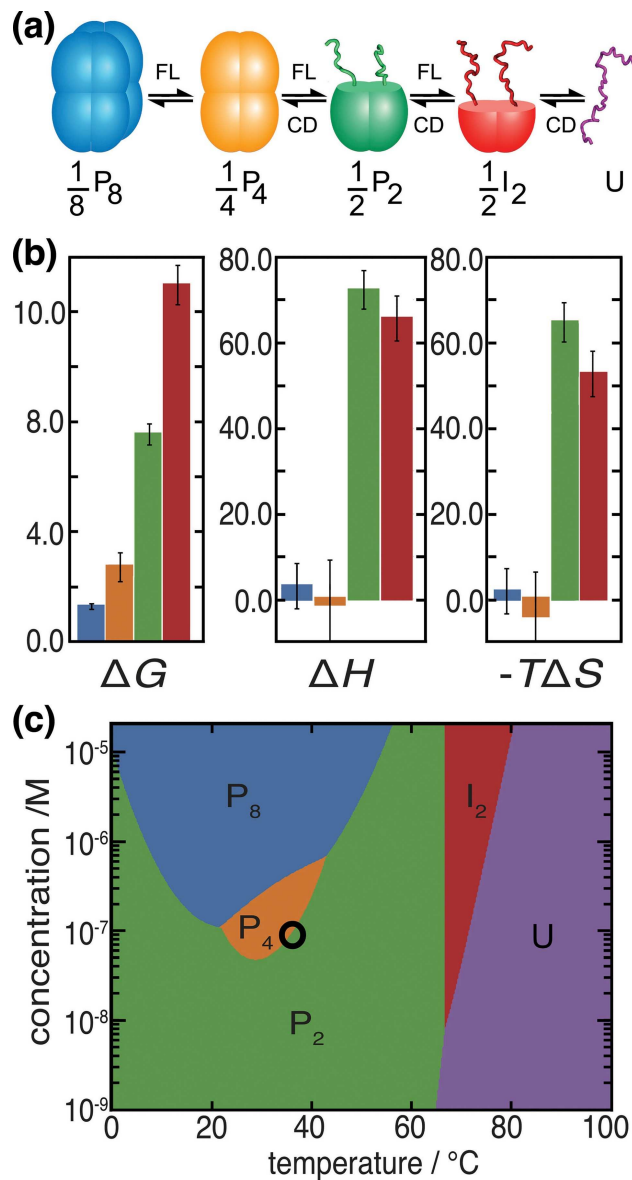
Figure 6. Global model for thermal unfolding of PaaR2. (a) Schematic representation of the five-state denaturation mechanism. Transitions that are visible by fluorimetry (FL) or circular dichroism spectroscopy (CD) are indicated, while all transitions are detected by DSC. **(b)** Values of thermodynamic parameters (in kcal mol^{-1} at 25 °C) accompanying each sub-process (colored according to the starting state as in panel (a)). Values are obtained as average values of the parameter distribution obtained from a global fit of the model functions to CD, fluorescence and DSC data. Error bars represent two standard deviations. **(c)** Phase diagram showing the most prevalent protein species as a function of protein concentration (per mole of monomer) and temperature. At each T and c_{TOT} the fraction of protein species i was calculated using eqs. 5-7. The black circle lies at 37 °C and 0.1 µM concentration (100 protein copies in cell), a situation which may be encountered *in vivo*.

Supporting material

Thermodynamic stability of the transcription regulator PaaR2 from *Escherichia coli* O157:H7

Running title: Thermodynamic stability of PaaR2

Pieter De Bruyn[1,2]*, San Hadži[1,2,3]*, Alexandra Vandervelde[1,2]*+, Albert Konijnenberg[1,2,4], Maruša Prolič-Kalinšek[1,2], Yann G.-J. Sterckx[1,2,5], Frank Sobott[4,6,7], Jurij Lah[3], Laurence Van Melderen[8], Remy Loris[1,2]

[1] Structural Biology Brussels, Department of Biotechnology, Vrije Universiteit Brussel, B-1050 Brussel, Belgium

[2] Center for Structural Biology, Vlaams Instituut voor Biotechnologie, B-1050 Brussel, Belgium

[3] Department of Physical Chemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, 1000 Ljubljana, Slovenia

[4] Biomolecular and Analytical Mass Spectrometry group, Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

[5] Laboratory of Medical Biochemistry, University of Antwerp (UA), Campus Drie Eiken, Universiteitsplein 1, 2610 Wilrijk, Belgium

[6] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

[7] School of Molecular and Cellular Biology, University of Leeds, Leeds, UK

[8] Cellular and Molecular Microbiology, Faculté des Sciences, Université Libre de Bruxelles (ULB), rue des Professeurs Jeener et Brachet 12, B-6041 Gosselies, Belgium

Correspondence to: Remy Loris (reloris@vub.ac.be; Remy.Loris@VIB-VUB.be)
     Tel. 0032 2 6291989
     Fax 0032 2 6291963

* These authors contributed equally to this work and should be considered joint first author.
+ Present address: KULeuven Laboratory of Dynamics in Biological Systems, O&N I Herestraat 49 - box 802, 3000 Leuven

Supporting Materials and Methods

Model selection

Different experimental techniques suggested that PaaR2 exists as an octamer in solution and that denaturation is a multistate process. This starting point already suggests many possible ways on how the denaturation process could proceed. Mass spectrometry identified the presence of dimer and tetramer species. Therefore, we decided to include them in the model mechanism and disregard some other model mechanisms involving heptamers, pentamers, etc. The concentration dependent melting curves further suggest a change in the oligomeric states during the melting. With this model-independent information in mind, we developed 8 models of the PaaR2 denaturation process (listed in Supporting Table S5). For each model the best-fit model parameters were obtained by nonlinear-fitting of the merit function (eq. 9) using the basin-hopping global optimization algorithm with randomly initiated starting parameters (1). The corresponding fits of the model function to the experimental data are shown in Figure S7. Correspondence between model function and the data was determined using chi-squared statistics (Table S5). To address the question of the model selection the Akaike information criterion (AIC) was used, which weights the simplicity of the model with the quality of description and chooses a model with the least information loss (2). The relative difference in the AIC values for the different models can be used to compute the relative likelihood of the models.

The simplest model, Model 1, involves only three states, a minimum required to describe the observed two endothermic DSC peaks. Models 2 and 3 involve four states, but the concentration dependence is treated differently (Model 2 involves one concentration-independent step). These two models show that using four states leads to better description of DSC data (Supporting Figures S7 and S8, Supporting Table S5). The shoulder in the first peak is now more accurately described, however the temperature dependence of the spectroscopic signals is not well captured. In Model 4, the dissociation steps from octamer to dimer are treated using the same affinity constant, assuming that the oligomerization is reminiscent of the binding to independent sites without any cooperative effects. This reduces the number of parameters although the whole transition involves 6 states. Model 5 (the original model) and 6 differ in the treatment of the last transition step, which is concentration-independent for the Model 6. This model captures the CD data well but fails in describing the DSC data. Model 7 involves an additional transition compared to the original model, but the resulting fit to the spectroscopic data is poor. Finally, Model 8 was developed, however the resulting fits are unstable and fitting runs into problems due to difficulties in the root-solving procedure. Based on low χ^2 values and AIC values (Supporting Table S5), it appears that Model 5 (the one presented in our manuscript) is the most probable among candidate models. The large values in the difference AIC values, $\Delta > 100$, ($\Delta = AIC_i - AIC_{min}$), suggest that other candidate models are very unlikely. In particular, the relative model probabilities for other candidate models are low ($p > 10^{-7}$), showing that Model 5 is significantly more probable compared to other models. The

validity of the model parameters for the selected model have been assessed using Monte Carlo error estimation analysis (Supporting Figure S9), showing that the parameters are reliable within the presented error margins and are not strongly correlated (Supporting Tables S4 and S5).

Nevertheless, the model is not perfect (χ^2 is 4.6 and the model function shows some systematic deviations with respect to the data, specially the transition monitored by CD at 350 K, see Supporting Figure S8). One likely possibility for the observed discrepancy is the rather simple treatment of spectroscopic signals. Analysis of the spectroscopic data of protein denaturation usually involves normalization of the signal with respect to the spectroscopic properties of the initial and final states (f_I, f_F) which may be estimated from the pre- and post-transition baselines (described in eq. 7). This procedure is employed as it reduces the number of fitting parameters. The spectroscopic properties f_I and f_F are temperature dependent quantities and their temperature dependence can be estimated from the baseline. However, when the description of the process involves more than two optically active states, additional parameters f_i are employed, but their temperature-dependence is neglected. This is because it is not possible to estimate their temperature dependence from the data and even assuming a simple linear temperature dependence would increase the number of fitting parameters considerably. Thus, we believe that because the temperature dependencies of the spectroscopic properties of the intermediate states were neglected, this leads to systematic errors in the normalized spectroscopic data. These errors increase through extrapolation of each f_I with temperature, which might explain why at higher temperature the model function deviates more from the CD data. Nevertheless, the model appears to give sensible predictions as observed by independent experiments using SEC (Supporting Figure S2).

Model analysis of thermal unfolding

Each transition step can be defined by the corresponding equilibrium constant expressed as:

$$K_1 = \left(\frac{\alpha_{P4}{}^2 c_{TOT}}{2\alpha_{P8}}\right)^{\frac{1}{8}}; \ K_2 = \left(\frac{\alpha_{P2}{}^2 c_{TOT}}{\alpha_{P4}}\right)^{\frac{1}{4}}; \ K_3 = \sqrt{\frac{\alpha_{I2}}{\alpha_{P2}}}; \ K_4$$

$$= \frac{\alpha_D \sqrt{2c_{TOT}}}{\sqrt{\alpha_{I2}}}$$

(4)

where α_i represents the molar fractions of the protein in its different states i ($\alpha_i = j[I_j]/c_{TOT}$) and c_{TOT} is the total protein concentration in monomer equivalents. The total protein concentration during the unfolding is constant, therefore:

$$\alpha_{P8} + \alpha_{P4} + \alpha_{P2} + \alpha_{I2} + \alpha_D = 1 \quad (5)$$

Each α_i in equation 5 can be expressed as a function of the equilibrium constants K_i, α_D and c_{TOT} using equation 4. Next, for a given set of K_i (specified by equation 6 - see below), c_{TOT} and T, values of α_D from equation 5 are calculated using a root solving routine by applying the

criterion $0 < \alpha_D < 1$. The obtained α_D is then used for the calculation of α_i via equation 4. The molar fractions α_i change with temperature according to the temperature dependence of the corresponding equilibrium constant as defined by the Gibbs-Helmholtz equation and Kirchhoff's law:

$$-RTlnK_i(T) = \Delta G_i(T) =$$

$$\Delta G_{i,T0}\left(\frac{T}{T_0}\right) + \Delta H_{i,T0}\left(1 - \frac{T}{T_0}\right) + \Delta c_{p,i}\left(T - T_0 - Tln(\frac{T}{T_0})\right) \quad (6)$$

Thus, at any T and c_{TOT}, the adjustable parameters (standard free energy ($\Delta G_{i,T0}$), standard enthalpy ($\Delta H_{i,T0}$) and standard heat capacity ($\Delta c_{p,i}$) change (defined at the reference temperature T_0=298.15 K) define the equilibrium constant K_i. A set of equilibrium constants at a given T and c_{TOT} is used for the calculation of the model-dependent molar fractions α_i using equations 4-6.

Based on the assumed model mechanism, we express the measured quantity f (ellipticity, fluorescence intensity or partial molar enthalpy of the protein ($\overline{H}_2$)) as a sum of contributions f_i each representing the property f of the pure state i:

$$f = \sum f_i \alpha_i \quad (7)$$

In the case of spectroscopic techniques, the signal can more conveniently be expressed in the terms of a normalized signal, $\Delta f = \left.(f - f_I)\middle/(f_F - f_I)\right.$, where f_I and f_F correspond to the properties of the initial and final states of a transition and can be estimated from the pre- and post-unfolding baselines. The normalized signal can then be expressed as sum $\Delta f = \sum Y_i \alpha_i$ where Y_i represent spectroscopic properties of the pure states i normalized with respect to initial and final state and are treated as model parameters. The DSC signal was expressed as partial molar heat capacity of the protein relative to the protein in native state (Δc_P) as described above. In that case, the property f from equation 7 is the partial molar enthalpy of the protein sample $\overline{H}_2$ which is the sum of the partial molar enthalpies of all species (expressed per mole of monomeric unit) $\overline{H}_2 = \sum \overline{H}_i \alpha_i$. The DSC model function was obtained by taking the partial derivative of $\overline{H}_2$ on temperature at constant pressure to obtain $\overline{C}_{p,2}$ and then formally subtracting $\overline{C}_{p,int}$:

$$\Delta c_P = \sum \Delta H_i \left(\frac{\partial \alpha_i}{\partial T}\right)_P \quad (8)$$

In equation 8, ΔH_i are the standard enthalpies (given per mole of monomeric subunits) associated with the transition of the protein from its native octameric state P_8 to a given state i. After the data have been normalized as described, a single set of parameters can be used to describe all experimental CD, fluorescence, and DSC unfolding curves simultaneously. Thus,

for each dataset given the experimental data y_i (Δf and Δc_P) measured as a function of variables x_i (T and c_{TOT}), we compute the sum (S) of residuals with the described model functions f (equation 7 and 8) using a set of parameters P: $S = \sum(y_i - f(x, P))^2$. Finally, the sum of residuals S calculated for each dataset were combined into a merit function $\mathcal{F}$ that was subjected to the minimization algorithm:

$$\mathcal{F} = w_{CD}S_{CD} + w_{FL}S_{FL} + w_{DSC}S_{DSC} \qquad (9)$$

The residuals from different experimental methods are multiplied by a factor w, such that each dataset has approximately the same weight in the fitting. In other words, the factors w correct for different number of experimental points in each dataset and for the higher numeric values of Δc_P than Δf. The applied global fitting procedure provides a much higher number of experimental points per fitted parameter as opposed to fitting each set of parameters to the individual experiments (3). The merit function $\mathcal{F}$ was minimized by modifying the fitting parameters using the Powel algorithm (4). The reliability of the estimated model parameters was assessed using a Monte Carlo simulation analysis. We generated 250 pseudo-experimental datasets by adding random errors obtained from a normal distribution. This distribution is specified by the mean centred around the values predicted by the best fit model function to the original experimental dataset and by the width corresponding to the estimated experimental error (δ=0.05 for Δf from CD and fluorescence and δ=800 cal mol^{-1}K^{-1} for DSC). For each pseudo experimental dataset (CD, fluorescence, DSC) we obtain a set of parameters by minimizing the merit function $\mathcal{F}$ (equation 9) (Supporting Figure S8), resulting in the distribution of parameter values. This distribution should correspond to a distribution obtained by performing a large number of repeated measurements on the same experimental system (5). The final set of thermodynamic parameters reported in Figure 6b and Supporting Table S3 corresponds to the means and standard deviations of the normal distribution function that best-fits the Monte Carlo derived parameter distributions shown in Supporting Figure S9. Additionally, such error analysis provides information on the parameter correlations showing the extent by which the determination of one parameter determines the value of another (shown in Supporting Table S4).

Supporting references

1.   Wales, D.J., and J.P.K. Doye. 1997. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A. 101: 5111–5116.

2.   Burnam, K.P., and D.R. Anderson. 2002. Model Selection and Inference: A Practical Information-Theoretic Approach. 2nd edit. Springer-Verlag. New York.

3.   Drobnak, I., G. Vesnaver, and J. Lah. 2010. Model-Based Thermodynamic Analysis of Reversible Unfolding Processes. : 8713–8722.

4.   Powell, M.J.D. 1964. An efficient method for finding the minimum of a function of

several variables without calculating derivatives. Comput. J. 7: 155–162.

5.   Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 2002. Numerical recipes in C++: the art of scientific computing, 2nd Edit. Cambridge Univ. Press. New York.

Supporting tables

Supporting Table S1. Primers used for cloning and mutagenesis.

Primer	Sequence
PaaR2_1	5'-AGGAGATATACCATGCAAAAAAAAGAAATTCGC-3'
PaaR2_2	5'-TCAGTGATGATGATGATGATGGCTGCTGCCGGCGCGGCGGCATTTTTG-3'
PaaR2_3	5'-GGCAGCAGCCATCATCATCATCATCACTGAGTTATAAAACCGGAGGAAAC-3'
PaaR2_4	5'-GTTAGCAGCCGGATCTTAGGGAAACTGGCGTCTTG-3'
PaaR2HisC120SFwd	5'-CGGCGCAAAAAAGCCGCCGCGCT-3'
PaaR2HisC120SRev	5'-CAATCCATCTCGCTATGAGATCTTC-3'
FL1	5'-ACCTTCCTCGGTTTAGTGTT-3'
FL2	5'-GTGGTAATTATCTTTAGTAATC-3'
Neg1	5'-GATGAATACCTTGACTGCGA-3'
Neg2	5'-GCTTTTGCGCAGAAATTTCG-3'

Supporting Table S2. SAXS data collection parameters. Abbreviations used are: |q| = 4π sin $(\theta)/\lambda$, Momentum transfer; I(0), extrapolated scattering intensity at the zero angle; R_G, radius of gyration calculated using the Guinier approximation (AUTORG) or the indirect Fourier transform package GNOM (P(r)); D_{max}, maximal particle dimension; V_{Porod}, Porod volume.

Data-collection parameters	
Beam line	SEC-SAXS at SWING beamline in SOLEIL
HPLC column	Shodex KW404-4F
Wavelength (Å)	1.03
q range (Å^{-1})	0.0028 - 0.31
Exposure time (ms) / Dead time (ms)	1500 / 500
Injected concentration	70 µl at 541 µM
Temperature (°C)	15
Secondary standard used	Water I(0)
Structural parameters	
I(0) (cm^{-1}/absorbance) [from Guinier]	0.1401 $\pm$ 0.0001
R_G (Å) [using AUTORG]	51.28 $\pm$ 0.06
I(0) (cm^{-1}/absorbance) [from P(r)]	0.1399 $\pm$ 0.0001
R_G (Å) [from P(r)]	51.83 $\pm$ 0.06
D_{max} (Å)	175.03
V_{Porod} estimate (Å^3)[1]	217997
Porod exponent[1]	3.9
Molecular mass determination	
Using the Q_R method (kDa)[2]	200.86
From Porod Volume (V_{Porod}/1.7) (kDa)	128.23
SAXS MoW (kDa)[2]	172.60
From Guinier analysis (kDa)[3]	101
Molecular mass from sequence (kDa)	15.2 (octamer: 122.3)
Software employed	
Data processing and analysis	Foxtrot, DataSW, Primus, Scatter

[1]Porod volume and exponent determined using the Porod-Debye plot in Scatter.
[2]Determined using a q_{max} of 0.275 Å^{-1}.
[3]Calculated using 0.7425 cm^3/g for the partial specific volume and a protein concentration of 127 µM at the top of the peak.

Supporting Table S3. Standard thermodynamic parameters derived from global fit. Reported values are means and standard deviations obtained from the Monte Carlo parameter distributions (Figure S9). The subscript index of a parameter denotes the corresponding transition. Values are given at 25 °C in kcal mol^{-1} and kcal mol^{-1} K^{-1} for the changes in heat capacity.

| Parameter	Transition	Value	Standard error
ΔG_1	$\frac{1}{8}P_8 \rightarrow \frac{1}{4}P_4$	1.28	0.02
ΔH_1		3.0	0.9
$\Delta c_{p,1}$		-0.16	0.08
ΔG_2	$\frac{1}{4}P_4 \rightarrow \frac{1}{2}P_2$	2.7	0.1
ΔH_2		-1	2
$\Delta c_{p,2}$		0.98	0.08
ΔG_3	$\frac{1}{2}P_2 \rightarrow \frac{1}{2}I_2$	7.5	0.1
ΔH_3		72.3	0.8
$\Delta c_{p,3}$		-0.54	0.05
ΔG_4	$\frac{1}{2}I_2 \rightarrow U$	13.0	0.1
ΔH_4		65.6	0.9
$\Delta c_{p,4}$		0.17	0.02

Supporting Table S4. Correlations of parameters derived from Monte Carlo error simulation analysis. Only pairs with correlation coefficient > 0.5 are reported.

| Parameter 1	Parameter 2	Correlation coefficient
ΔG_4	ΔH_4	0.75
ΔH_1	$\Delta c_{p,2}$	0.79
ΔH_3	$\Delta c_{p,1}$	0.71
$\Delta c_{p,3}$	$\Delta c_{p,4}$	0.53

Supporting Table S5: Candidate models and the corresponding goodness-of-fit statistics. The selected model presented in the manuscript is highlighted in bold.

| Model	Model mechanism	χ^2 per degree of freedom	number of variables	AIC value
1	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons D$	16.6	8	972
2	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons \frac{1}{2}I_2 \rightleftharpoons D$	15.1	13	939
3	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons D$	10.4	13	810
4	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{6}P_6 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons \frac{1}{2}I_2 \rightleftharpoons D$	7.1	13	680
5	$\mathbf{\frac{1}{8}P_8 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons \frac{1}{2}I_2 \rightleftharpoons D}$	**4.6**	**16**	**531**
6	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons I \rightleftharpoons D$	18.0	16	1004
7	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons \frac{1}{2}I_2 \rightleftharpoons J \rightleftharpoons D$	12.2	20	864
8	$\frac{1}{8}P_8 \rightleftharpoons \frac{1}{4}P_4 \rightleftharpoons \frac{1}{2}P_2 \rightleftharpoons \frac{1}{2}I_2 \rightleftharpoons \frac{1}{2}J_2 \rightleftharpoons D$	does not converge	20	/

Supporting figures
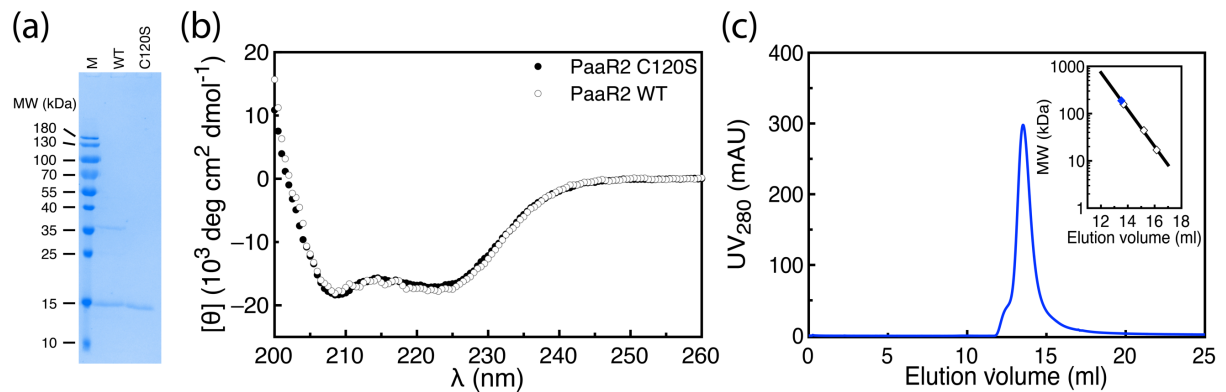
Supporting Figure S1. Characterization of PaaR2C120S and PaaR2 wild type. (a) SDS-PAGE analysis of purified PaaR2 C120S and PaaR2 wild type. PaaR2 occurs as a single band with apparent molecular weight slightly below 15 kDa. PaaR2 wild type shows an additional band at 35 kDa because of boiling-induced oligomerization. **(b)** CD spectra of PaaR2 wild type and PaaR2 C120S at 15°C. **(c)** Analytical SEC of PaaR2 wild type which was run immediately after Ni-NTA purification on a BioRad Enrich SEC650 column (20 mM Tris pH 7.3, 500 mM NaCl, 1 mM TCEP). The inset shows the molecular weight estimation using the elution volumes and molecular weights of the Bio-Rad Size Exclusion standard. The molecular weights of bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa) and horse myoglobin (17 kDa) are within the linear range of the column and indicated on the figure (white diamonds). The elution volume of PaaR2 wild type was plotted on this curve (blue diamond), leading to a MW estimate of 188 kDa.
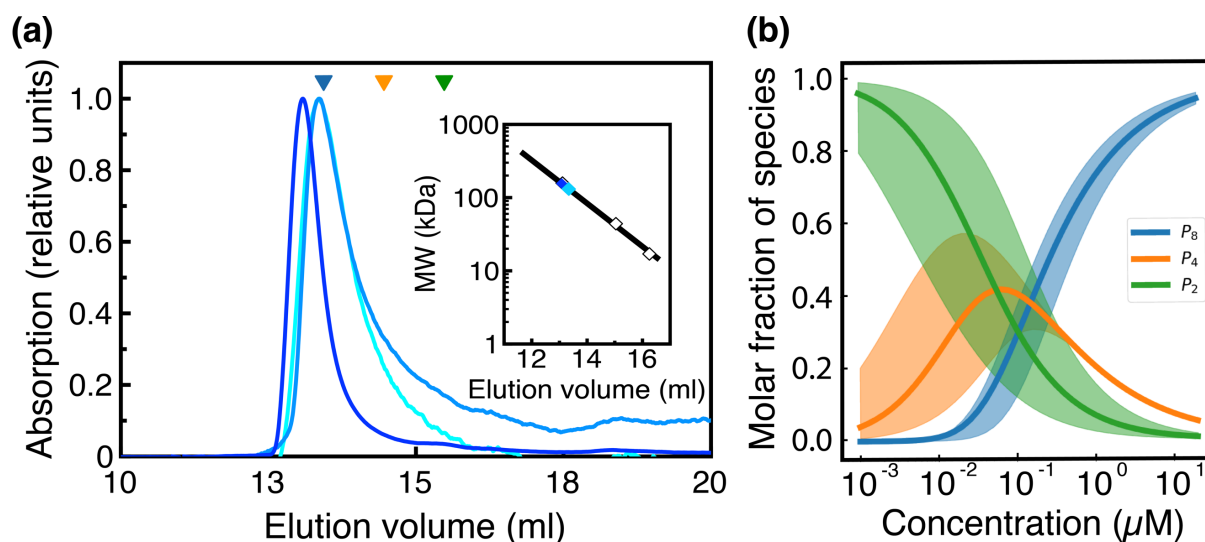
Supporting Figure S2. Comparison of model predictions with experimental data. The normalized analytical SEC profiles (run on a BioRad Enrich SEC 650 column) of PaaR2His(C120S) at different concentrations is shown in panel **(a)**. The inset shows the relevant concentrations were estimated from the top of the peak 9 μM (dark blue), 0.8 μM (light blue) and 0.5 μM (cyan). The theoretical elution volumes for octamer, tetramer and dimer are indicated in blue, orange and green arrows respectively. The model predicted molar fraction as a function of concentration at 25°C is shown in panel **(b)**. Model predicted fractions are shown as solid lines (based on the best fit parameters given in Table S3). The shaded area represents variation of the model-predicted fraction when the best-fit parameters lie within one standard deviation from the mean.
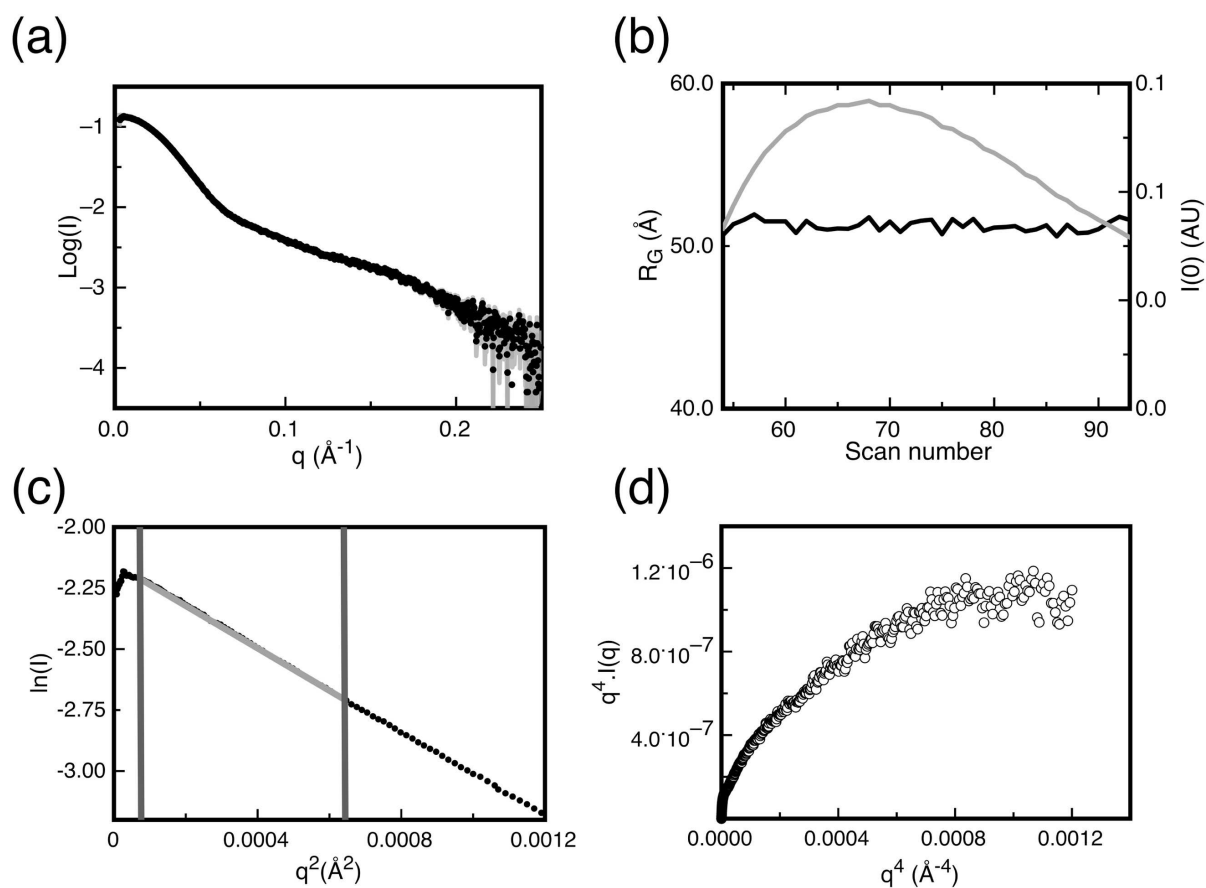
Supporting Figure S3. Small-angle X-ray scattering. (a) SAXS curve of PaaR2 after buffer subtraction and averaging of individual SAXS curves from the top of the HPLC peak. The scattering intensities are indicated in black and the errors are presented in grey. **(b)** $R_G/I(0)$ function. Plot of the R_G (black) and I(0) (grey) as a function of the scan number. **(c)** Guinier approximation of the SAXS curve (grey). **(d)** Porod-Debye plot used to calculate the Porod volume.
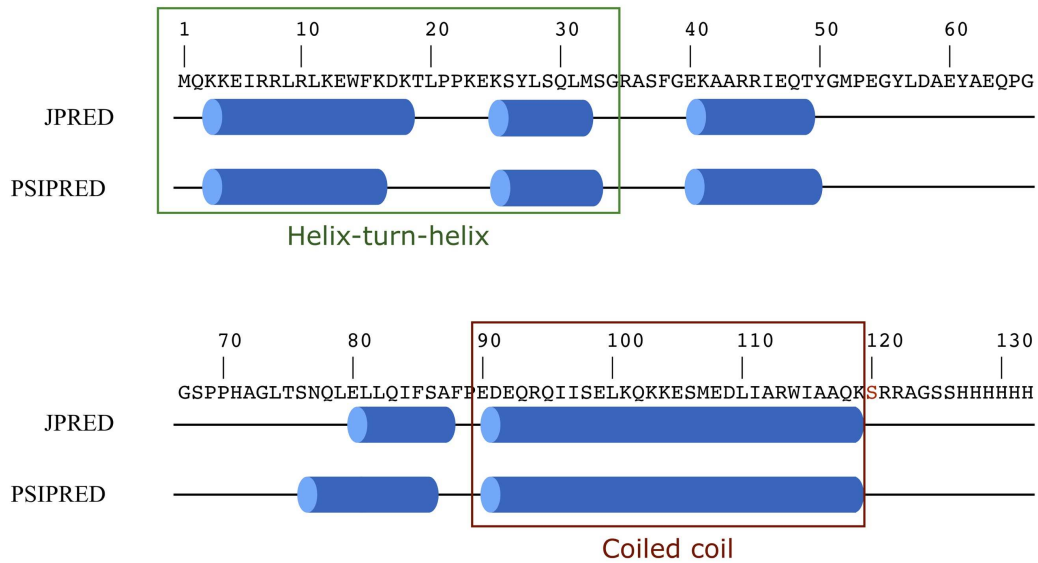
Supporting Figure S4. Sequence-based secondary structure prediction. The predictions by JPRED and PSIPRED are shown graphically. Blue bars represent α-helices, the black line represents random coil. The C120S mutation is marked red in the sequence. The HTH-motif is indicated by the green box, the coiled coil (predicted with the Coiled Coil webserver tool) is indicated by the red box.
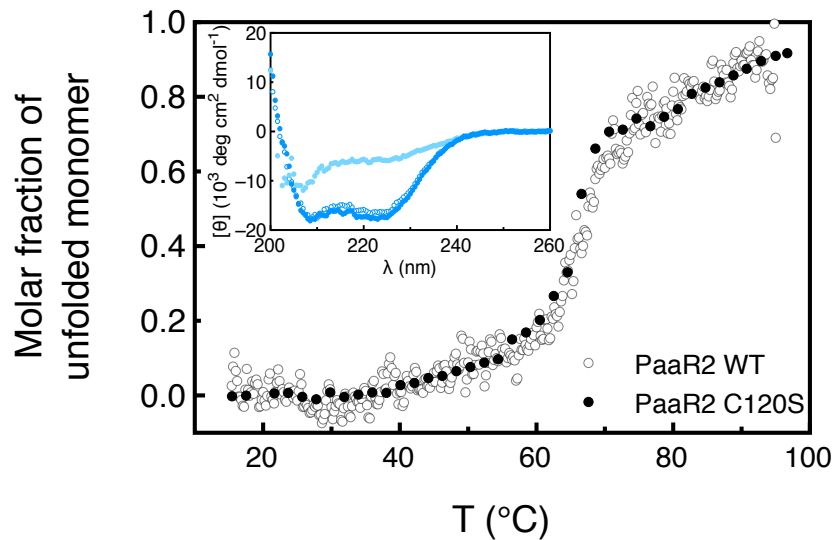
Supporting Figure S5. Thermal unfolding of PaaR2 wild type and PaaR2C120S followed by CD spectroscopy. The superposition of the PaaR2 wild type and PaaR2 C120S thermal unfolding is shown. The inset shows the spectra of PaaR2 wild type at 15°C (blue full circles), 95°C (light blue full circles) and renatured (blue empty circles).
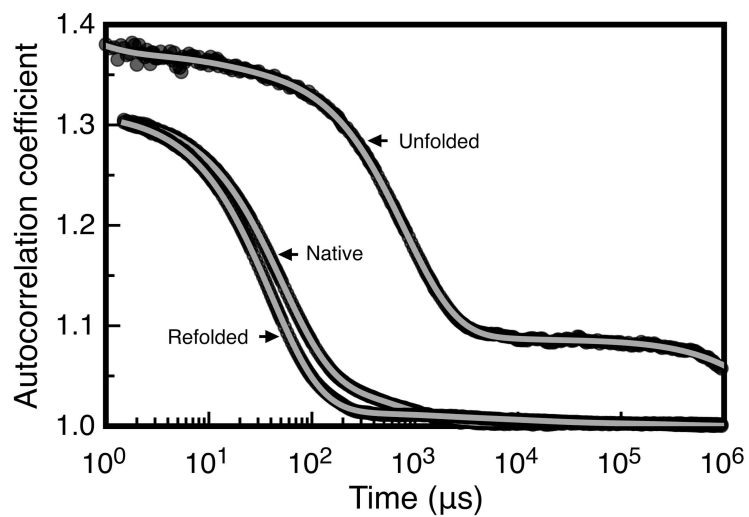
Supporting Figure S6. DLS profile of PaaR2. DLS autocorrelation curves are shown of PaaR2 before unfolding (native), at 80°C (unfolded) and after refolding (refolded). The regularization fits are shown in grey.
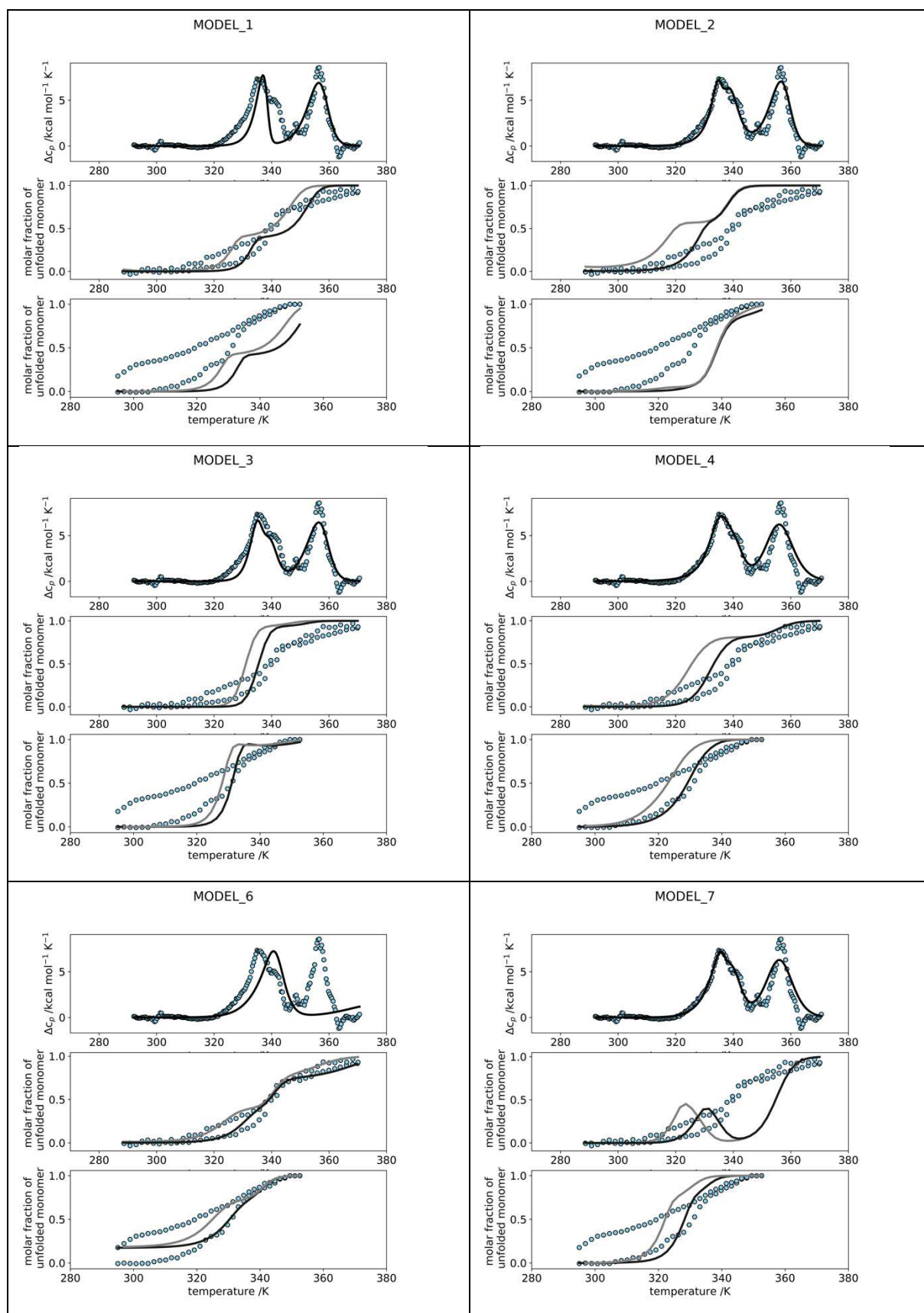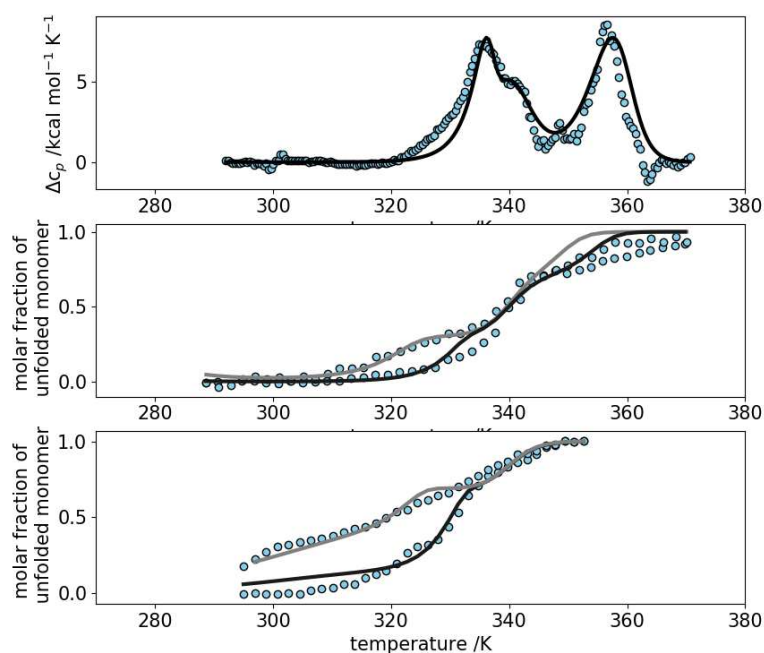
Figure S7. Experimental and model-derived values calculated using several model mechanisms. The experimental data are shown as blue circles while the theoretical fits correspond to full lines with grey lines corresponding to a lower concentration. In each panel the DSC data are shown on top, fluorescence data in the middle and CD data below.

Supporting Figure S8. An example of global fit for the selected model (Model 5). An example of global fit of the model function (eqs. 8 and 9, solid lines) to CD, FL and DSC data (circles). Black solid lines correspond to samples at 13 μM, gray solid lines correspond to samples at lower (0.13 μM) concentration.
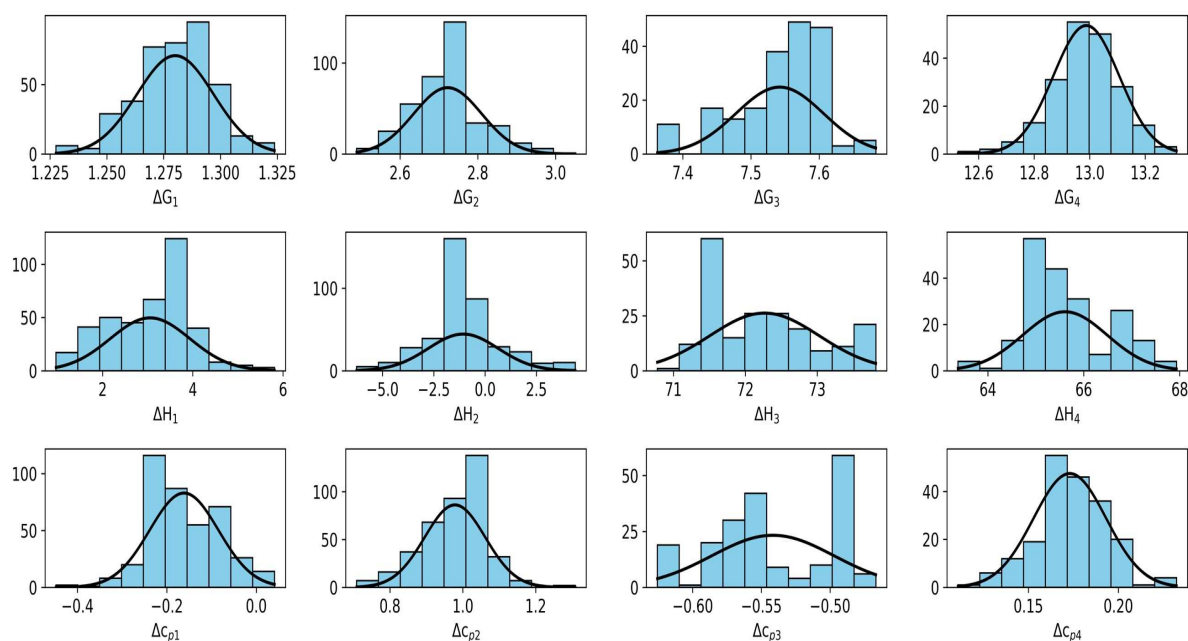
Supporting Figure S9. Monte Carlo error simulation analysis. Histograms show distributions of parameter values obtained from global fits to 250 pseudo-experimental datasets each consisting CD, FL and DSC data. Final set of thermodynamic parameters reported in Figure 6b and Supporting Table S3 are means and standard deviations of normal distribution function (black curve) that best-fits the Monte Carlo derived parameter distributions (bars).