Chem, Volume *10*

Supplemental information

Covalent bicyclization of protein complexes yields durable quaternary structures

George H. Hutchins, Sebastian Kiehstaller, Pascal Poc, Abigail H. Lewis, Jisun Oh, Raya Sadighi, Nicholas M. Pearce, Mohamed Ibrahim, Ivana Drienovská, Anouk M. Rijs, Saskia Neubacher, Sven Hennig, and Tom N. Grossmann

Supplementary Experimental Procedures

Molecular Biology

Wild-type PFE (**p1**) in a previously reported construct [1] with a C-terminal His-tag was used to prepare variants **p2**, **p3** and **p4** containing T3C and Q174C mutations by site-directed mutagenesis, via a modified QuikChange (Agilent) protocol as described previously [2], using *Pfu* Turbo polymerase and 27-mer overlapping primers (obtained from IDT). Residual template DNA in the completed PCR reactions was digested by addition of DpnI (NEB) at 37°C for 1 hour followed by transformation of 1–5 µL of the reaction into chemically competent *E. coli* DH5-α cells, mini-prep scale DNA purification and confirmation of successful mutagenesis by Sanger sequencing (Eurofins Genomics).

All other trimeric protein variants (Table S11) were synthesized and cloned by Twist Bioscience in a pET28a (+) expression vector (kanamycin resistant, IPTG inducible). Constructs for the suite of protein trimers were designed to match the exact expressed sequence of proteins deposited to the protein databank, containing either an N-terminal or C-terminal hexa-histidine tag and any additional cleavage sites and tags. Full-length sequences of all expressed proteins are included in Supplementary Figure S1 and Table S11.

Protein Expression and Purification

Protein expression vectors were transformed into *E. coli* BL21(DE3) cells by addition of 10–50 ng of purified plasmid DNA and heat-shock at 42°C for 45 seconds. Transformed colonies were inoculated in 100 mL of LB (with 100 mg/L ampicillin or 50 mg/L kanamycin dependent on resistance) and incubated overnight at 37°C, 130 RPM. 20 mL of overnight culture was transferred to 1 L volumes of LB (containing respective antibiotics) and cultured to an OD_{600nm} of 0.6–0.8 prior to induction by addition of 0.2% (m/v) *L*-rhamnose or 1 mM IPTG and expressed for 16–18 hours at 28°C, 130 RPM. Alternative trimers were expressed as above (50 mg/L kanamycin, 1 mM IPTG) in TB media. Cell pellets were harvested by centrifugation (4000 RPM, 20 minutes) and resuspended in lysis buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8).

Cells were lysed by four passes through a microfluidizer (Microfluidics LM20), and cell debris removed by centrifugation at 18 000 RPM for 30 minutes at 4°C. His-tagged proteins were purified by nickel affinity chromatography, flowing the clarified lysate through a 1 mL HisTrap FF column (Cytiva) equilibrated with lysis buffer at 3 mL/min., followed by further column washing with lysis buffer, and protein elution by increasing imidazole concentration to 250 mM. Protein samples were further purified by size exclusion chromatography with a 16/600 Superdex 75 or 200 pg column (Cytiva), buffered in 50 mM HEPES, 150 mM NaCl, at pH 8. Tris (2-carboxyethyl) phosphine (TCEP) at a concentration of 0.5 mM was added to the size exclusion buffer for purification of crosslinking variants to ensure reduction of exogenous solvent exposed cysteine residues. Successful purification of protein samples was confirmed by SDS-PAGE and LC-MS analysis, and samples were concentrated to 65–150 µM (Amicon centrifugal filtration) and flash frozen in liquid nitrogen for storage at –80°C prior to use. Extinction coefficients (ε) used to determine protein concentration were predicted using the Expasy ProtParam tool [3]. All protein concentrations are reported as *trimeric* concentration, regardless of crosslinking status (*e.g.*, PFE monomeric ε = 36 743 $M^{-1}cm^{-1}$, trimeric ε = 110 229 $M^{-1}cm^{-1}$) unless otherwise stated.

SDS-PAGE

SDS-PAGE samples were prepared by mixing a concentration of 5–20 µM protein with standard Coomassie loading dye and incubated at 95°C for 15 minutes, except (as noted in Figure S17) where a final concentration of 6 M urea was added in addition to loading dye, with no heating. Samples were analysed by a 4–20% gradient gel (GenScript) for 1 hour at 150 V and stained by Instant Blue Coomassie (Abcam).

Protein Liquid Chromatography/Mass Spectrometry

The molecular weight of variants and their crosslinked counterparts was assessed by LC/MS on different systems, indicated for each data collection in the respective supplementary figure. High-resolution MS of PFE variants and their crosslinked counterparts (Table S1) was performed using an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). The samples were injected on a 1.0 mm I.D. x 30 cm, 4 µM

TSKgel superSW3000 SEC column (Tosoh). The mobile phase comprised of 150 mM ammonium acetate (AA). The isocratic gradient applied was as follows: 0–16 min, 100% AA at a flow rate of 16 μL/min. The injection volume and column temperature were set at 1 μL and 23°C, respectively. This LC set up was coupled to a TIMS-Qq-TOF mass spectrometer (first generation, Bruker equipped with an electrospray-ionization (ESI) source [4]. All measurements were acquired for 16 min operating the mass spectrometer in positive mode with the following ESI and MS settings: capillary voltage of 4.2 kV, nebulizer gas of 0.8 bar, dry gas flow of 4.0 L/min, drying temperature of 200°C, ion energy of 5 eV, collision energy of 10 eV, and in-source collision-induced dissociation at 80 eV. The mass spectrometer was in TIMS-off mode (no ion mobility) using a m/z range of 500–8000. Ion funnels were set at values of 350 Vpp and 600 Vpp. Data analysis was performed using Bruker Compass DataAnalysis. Samples were calibrated externally using Agilent ESI tuning mix for ESI-ToF.

Analysis of PFE variant charge states (Figure 1C) was measured on an Agilent 6230 ESI-TOF LC/MS (3500 V capillary voltage), whilst molecular weight of all other trimeric variants (Figures S12-S16) was determined using either an Agilent 6230 ESI-TOF or LC/MSD XT ESI-Quadrupole LC/MS (3000 V capillary voltage). Both systems were used with the same two LC mobile phases; A (H_2O, 0.1% v/v formic acid) and B (80% isopropanol, 10% acetonitrile, 10% H_2O, 0.1% formic acid v/v), with injection of 0.01-0.1 nmol of protein sample via a 50 x 2.1 mm Agilent AdvanceBio RP mAB reverse-phase C4 column. A 10-minute gradient of 0–95% solvent B was applied with a flow rate of 0.3 mL/min. Mass spectra were recorded in the m/z 500–3000 range at a scan rate of 1.0 Hz. Deconvoluted masses were obtained using the maximum entropy algorithm in Agilent software with a mass range of 10,000–100,000 Da and H^+ as the amplifier. Tables for all LC/MS measurements with the calculated (m/z = [MW+Z]/Z) and found values for all predominant peaks are included in the Supplementary Information.

The calculated molecular weight of each protein trimer was derived from the Expasy ProtParam molecular weight [3], excluding the N-terminal methionine residue which is typically cleaved from expressed protein sequence [5]. An additional mass of approximately +178 Da was frequently observed for both unreacted and crosslinked proteins, corresponding to a previously reported N-linked gluconylation [6]. Crosslinked trimer molecular weights were calculated by addition of the mass of reacted Ta-I_3 (381.4 g/mol), minus the mass of three hydrogens per crosslinker (displaced from the reacted cysteine thiol groups).

PFE Activity Assays

Enzymatic activity was assessed by measuring turnover of *p*-nitrophenyl acetate (*p*NPA) to *p*-nitrophenolate (*p*NP). Guanidine hydrochloride dependent activity was determined by incubating PFE variants at a concentration of 50 nM with GuHCl (0, 0.5, 1, 1.5 and 2 M) for 10 minutes prior to the assay. The substrate stock (*p*NPA) was prepared daily in buffer (50 mM HEPES, 50 mM NaCl, pH 7.5) with 10% DMSO v/v at a concentration of 2 mM. Activity was measured at 20°C at a final trimeric concentration of 2 nM PFE in a 100 μL volume (5% DMSO v/v), in a half-area 96 well microplate, initiating the reaction by addition of a final substrate concentration of 1 mM *p*NPA. Absorbance at 410 nm was measured in a plate reader (TECAN Spark 20M) for up to 30 minutes at an interval of 20 seconds, with shaking at 270 RPM between measurements. An additional measurement without enzyme was taken to determine background hydrolysis of the substrate at a given concentration of GuHCl and subtracted from the data. All assays were carried out in 50 mM HEPES, 50 mM NaCl, at pH 7.5. Long term high temperature storage experiments (Figure S9) were performed using a final trimer assay concentration of 0.67 nM, after incubating the protein at 50°C at a concentration of 30 μM in 50 mM HEPES, 50 mM NaCl, pH 8.

Supplementary Figures

p1 (*PFE wild type*)
MSTFVAKDGTQIYFKDWGSGKPVLFSHGWLLDADMWEYQMEYLSSRGYRTIAFDRRGFGRSDQPWTGNDYDTFA
DDIAQLIEHLDLKEVTLVGFSMGGGDVARYIARHGSARVAGLVLLGAVTPLFGQKPDYPQGVPLDVFARFKTEL
LKDRAQFISDFNAPFYGINKGQVVSQGVQTQTLQIALLASLKATVDCVTAFAETDFRPDMAKIDVPTLVIHGDG
DQIVPFETTGKVAAELIKGAELKVYKDAPHGFAVTHAQQLNEDLLAFLKRGSHHHHHH

p2 (PFE T3C)
MS**C**FVAKDGTQIYFKDWGSGKPVLFSHGWLLDADMWEYQMEYLSSRGYRTIAFDRRGFGRSDQPWTGNDYDTFA
DDIAQLIEHLDLKEVTLVGFSMGGGDVARYIARHGSARVAGLVLLGAVTPLFGQKPDYPQGVPLDVFARFKTEL
LKDRAQFISDFNAPFYGINKGQVVSQGVQTQTLQIALLASLKATVDCVTAFAETDFRPDMAKIDVPTLVIHGDG
DQIVPFETTGKVAAELIKGAELKVYKDAPHGFAVTHAQQLNEDLLAFLKRGSHHHHHH

p3 (PFE Q174C)
MSTFVAKDGTQIYFKDWGSGKPVLFSHGWLLDADMWEYQMEYLSSRGYRTIAFDRRGFGRSDQPWTGNDYDTFA
DDIAQLIEHLDLKEVTLVGFSMGGGDVARYIARHGSARVAGLVLLGAVTPLFGQKPDYPQGVPLDVFARFKTEL
LKDRAQFISDFNAPFYGINKGQVVS**C**GVQTQTLQIALLASLKATVDCVTAFAETDFRPDMAKIDVPTLVIHGDG
DQIVPFETTGKVAAELIKGAELKVYKDAPHGFAVTHAQQLNEDLLAFLKRGSHHHHHH

p4 (PFE T3C+Q174C)
MS**C**FVAKDGTQIYFKDWGSGKPVLFSHGWLLDADMWEYQMEYLSSRGYRTIAFDRRGFGRSDQPWTGNDYDTFA
DDIAQLIEHLDLKEVTLVGFSMGGGDVARYIARHGSARVAGLVLLGAVTPLFGQKPDYPQGVPLDVFARFKTEL
LKDRAQFISDFNAPFYGINKGQVVS**C**GVQTQTLQIALLASLKATVDCVTAFAETDFRPDMAKIDVPTLVIHGDG
DQIVPFETTGKVAAELIKGAELKVYKDAPHGFAVTHAQQLNEDLLAFLKRGSHHHHHH

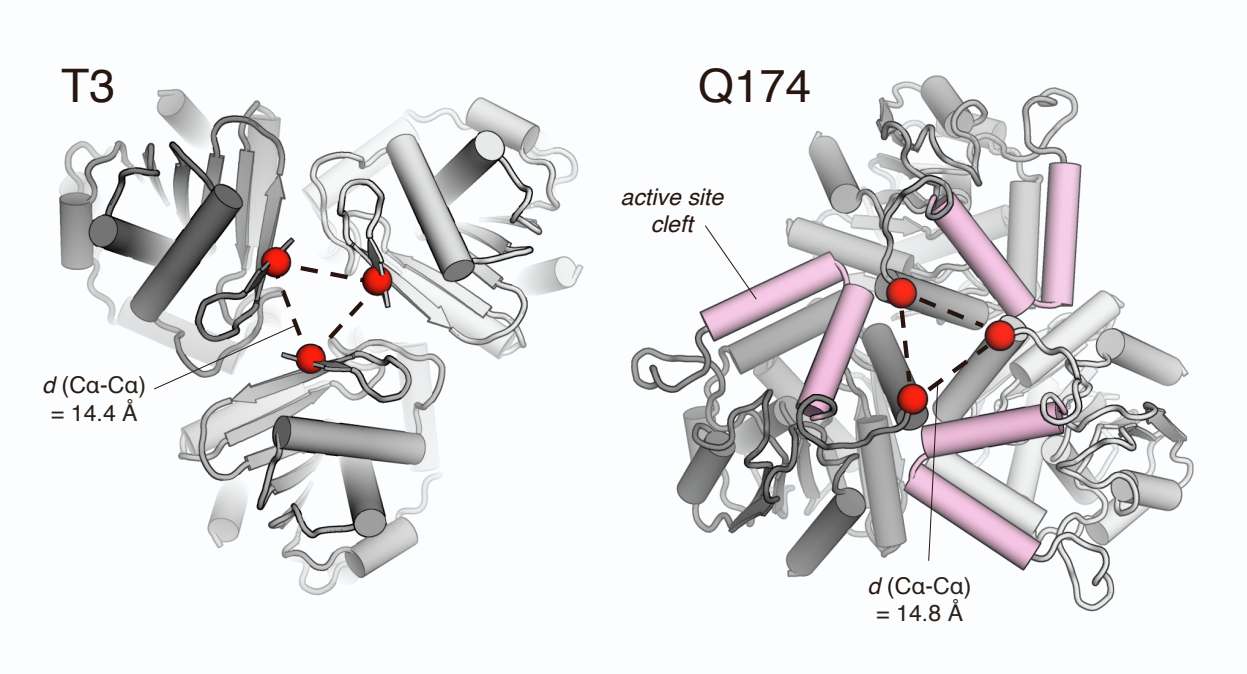


Figure S1. Design schema for PFE trimeric crosslinking (relating to Figure 1A). Top: Expressed sequences of PFE wild type (**p1**) and the three variants designed for trimer crosslinking (**p2**, **p3**, **p4**). Bottom: Dashed lines indicate the distance (*d*) between the C*α* atoms (red) of selected crosslinking sites. Substrate entry to the enzyme active site is via a V-shaped helical hairpin (pink, PDB ID: 1va4).

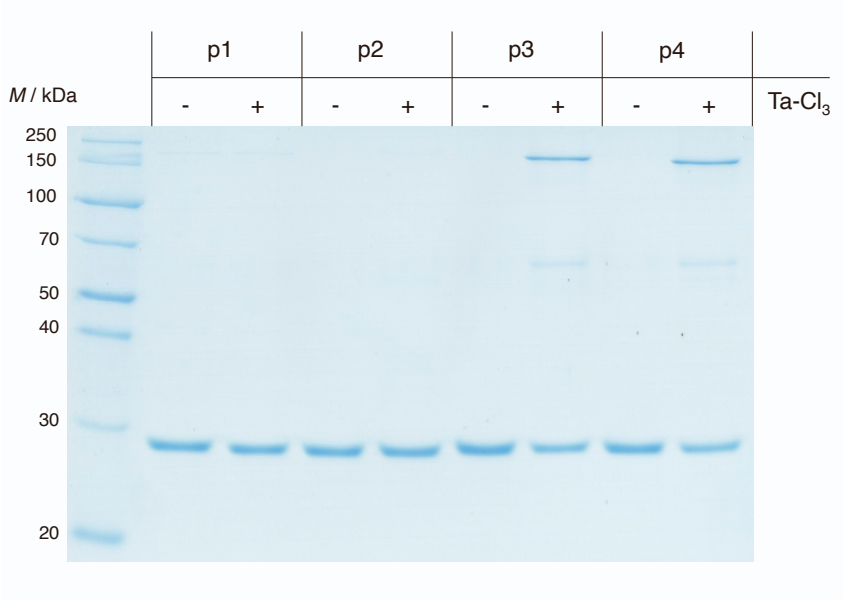


Figure S2. **Reactivity of PFE variants with Ta-Cl_3 (chemical structure in Figure 1B).** Reaction time was 4 hours at 21°C, 15 µM protein trimer, 300 µM Ta-Cl_3. No reactivity is observed with **p2** (T3C) whilst limited reactivity is observed with **p3** (Q174C) and **p4** (T3C + Q174C) resulting in partial formation of covalently linked trimers (top band) and a small proportion of dimer (middle band).

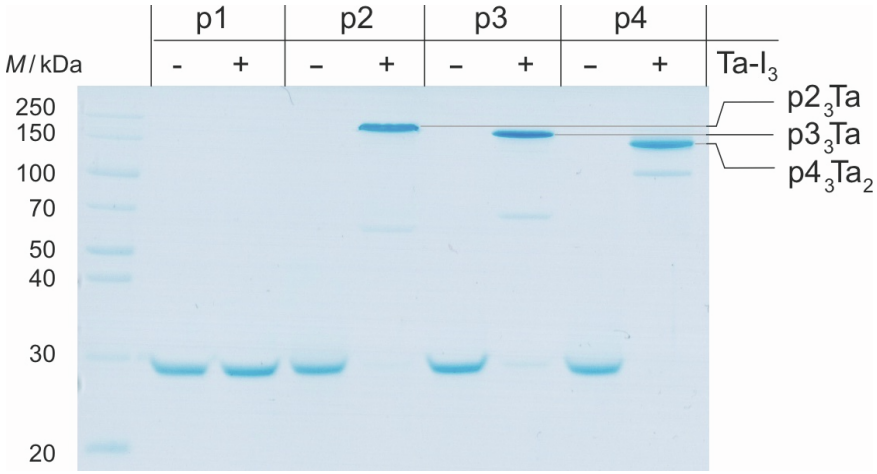


Figure S3. **Reactivity of PFE variants with Ta-I_3 (chemical structure in Figure 1B).** SDS-PAGE analysis of PFE variants prior and after incubation with Ta-I_3 demonstrates the formation of covalently linked, higher molecular weight species, and hinted at the modified chain topology with variable band migration (**p4$_3$Ta$_2$** > **p3$_3$Ta** > **p2$_3$Ta**) despite equivalent molecular weight.

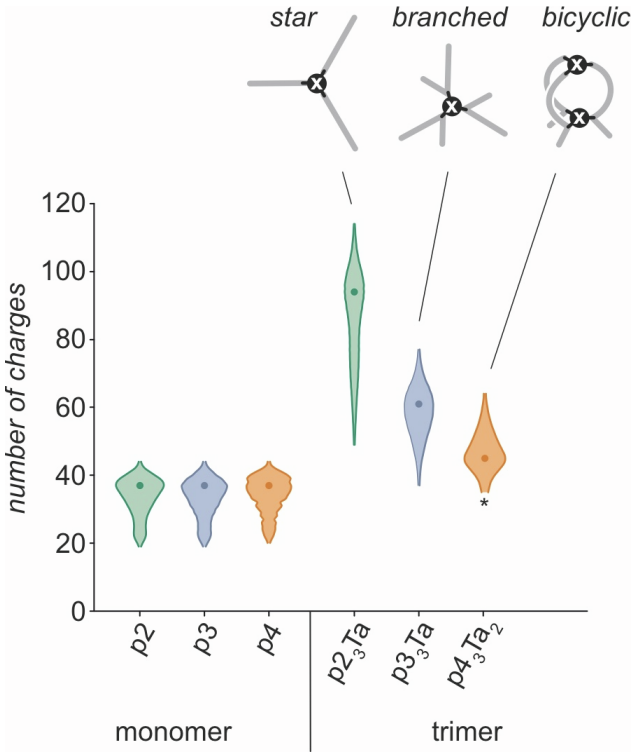


Figure S4. Violin plot showing distribution and range of charge states for PFE variants p2–p4 and crosslinked variants p2–p4 as observed by TOF-LC/MS, relating to Figure 1C. For MS signals and intensities see Tables S2–S5. *Due to the m/z range of the instrument lower charge states could not be detected. Raw data for **p2$_3$Ta**, **p3$_3$Ta** and **p4$_3$Ta$_2$** are presented in Figure 1C.

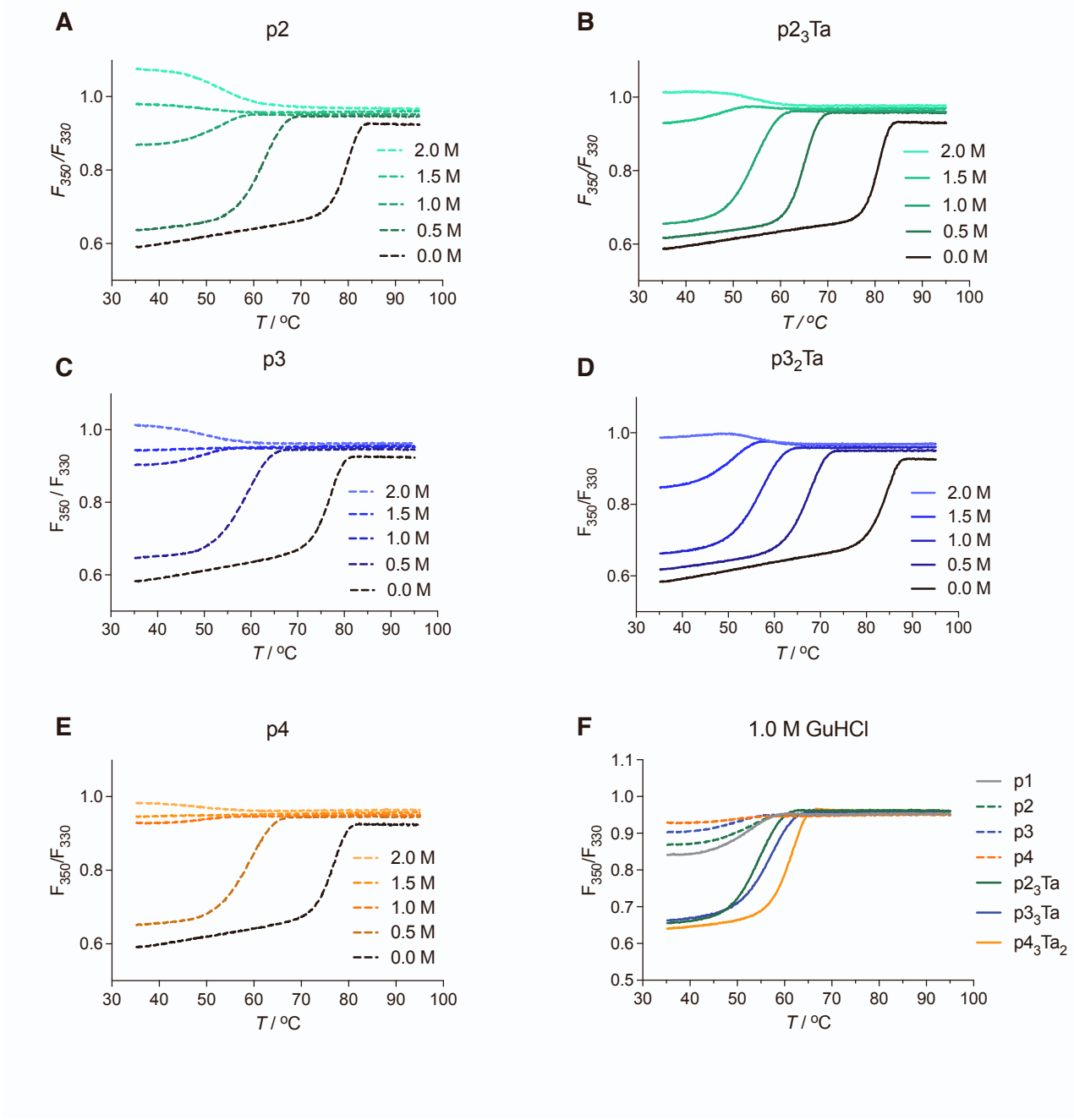


Figure S5. Thermal melt curves for PFE variants and their crosslinked counterparts. (A–E) Additional thermal denaturation curves (DSF measurements of tryptophan fluorescence ratio) for PFE variants **p2**–**p4** before and after crosslinking by Ta-I_3. Midpoint inflection temperatures (T_i) are reported in Figure 2A and Table S6. (F) Comparison of thermal denaturation curves for all variants at 1 M GuHCl further demonstrates the resistance of crosslinked PFE to chemical stress, an effect which is most pronounced for bicyclic **p4$_3$Ta$_2$**.

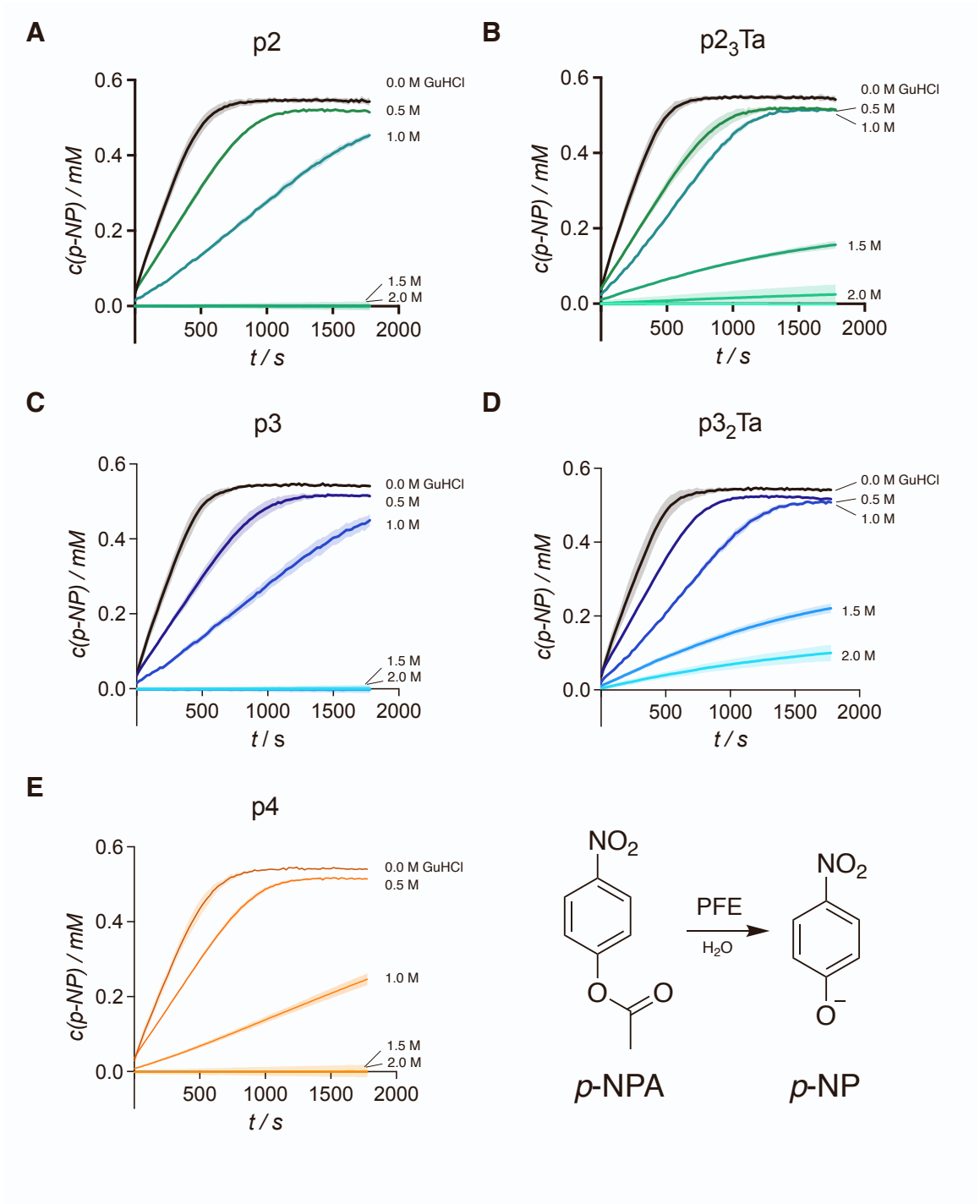


Figure S6. Enzymatic activity for PFE variants and their crosslinked counterparts. Activity assayed by hydrolysis of *para*-nitrophenyl acetate (*p*-NPA) to *para*-nitrophenolate (*p*-NP). Additional assay data shown for PFE variants **p2**–**p4** before and after crosslinking. The initial reaction rate relative to that of **p4$_3$Ta$_2$** for each concentration of GuHCl was used to compare variants reported in Figure 2D (for absolute values with errors see Table S7). Triplicate technical repeats, standard error shaded.

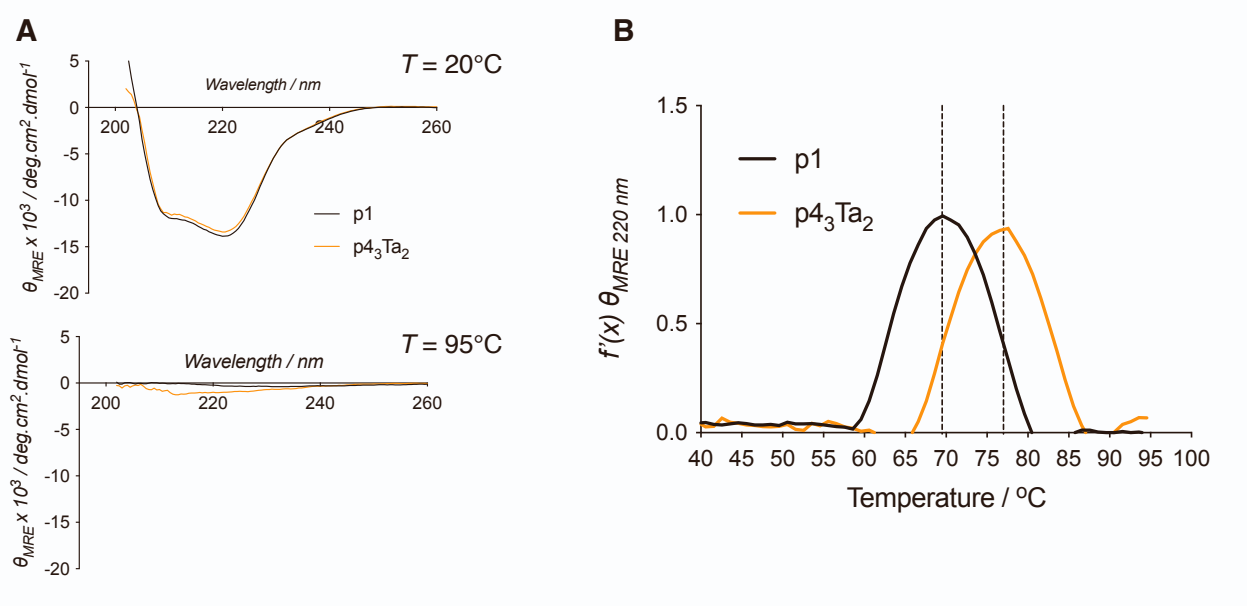


Figure S7. Circular dichroism spectroscopy of PFE variants (relates to Figures 3A and 3B). (A) Circular dichroism spectra of PFE variants **p1** and **p4$_3$Ta$_2$** at 20°C (also shown in Figure 3A) and 95°C at a 4 μM trimer concentration (in 50 mM potassium chloride, 50 mM potassium phosphate, pH 8). At 20°C, we obtained equivalent CD spectra for both **p1** and **p4$_3$Ta$_2$**, showing characteristic two minima around 210 and 220 nm corresponding to the predominantly helical fold of the trimer. The absence of signal at 95°C suggests complete precipitation of the protein. (B) First derivative of melt curves, T_m values (Figure 3B) were determined at the curve maximum.

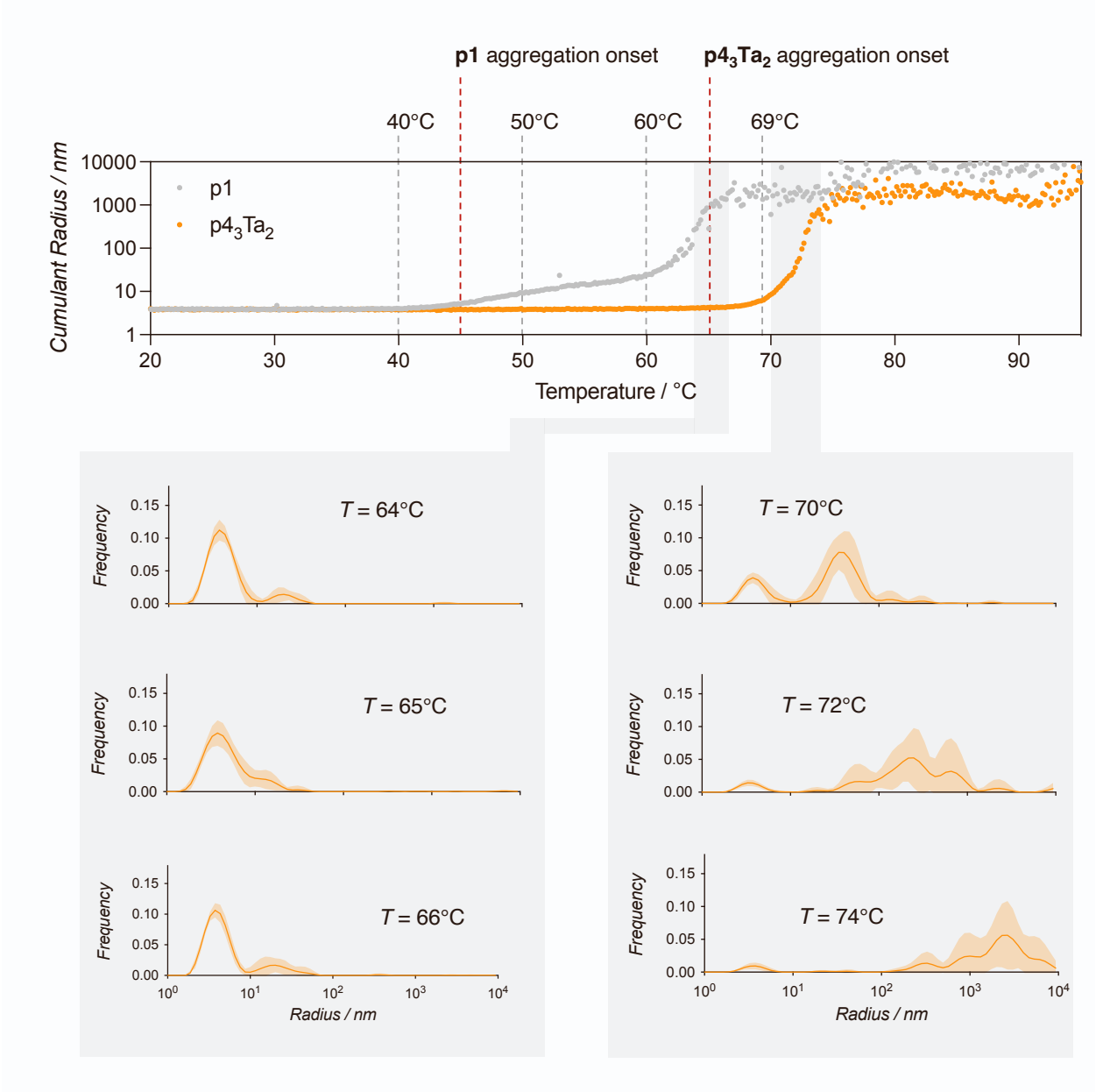


Figure S8. Crosslinked PFE DLS analysis shows aggregation resistance (relates to Figure 3C). Cumulant radius values determined by dynamic light scattering for **p1** (grey) and **p4₃Ta₂** (orange) demonstrate the predominant particle size of each protein between 20–95°C. Higher order **p1** oligomers begin to form above 45°C, whereas **p4₃Ta₂** remains monodisperse until 65°C (size distribution analysis for **p4₃Ta** at this temperature reveals some polydispersity that is not apparent from the cumulant radius value, which identifies only the dominant species). Temperatures for the initial onset of aggregation are indicated (dashed, red), as are the size distribution slices presented in Figure 3C (dashed, grey) with standard deviation shaded. Additional size distribution slices are shown for **p4₃Ta₂** at 64°C, 65°C, 66°C (showing initial aggregation onset) and at 70°C, 72°C, 74°C (showing complete aggregation). Each curve represents an average of six measurements within *T*+0.95°C of the given temperature.

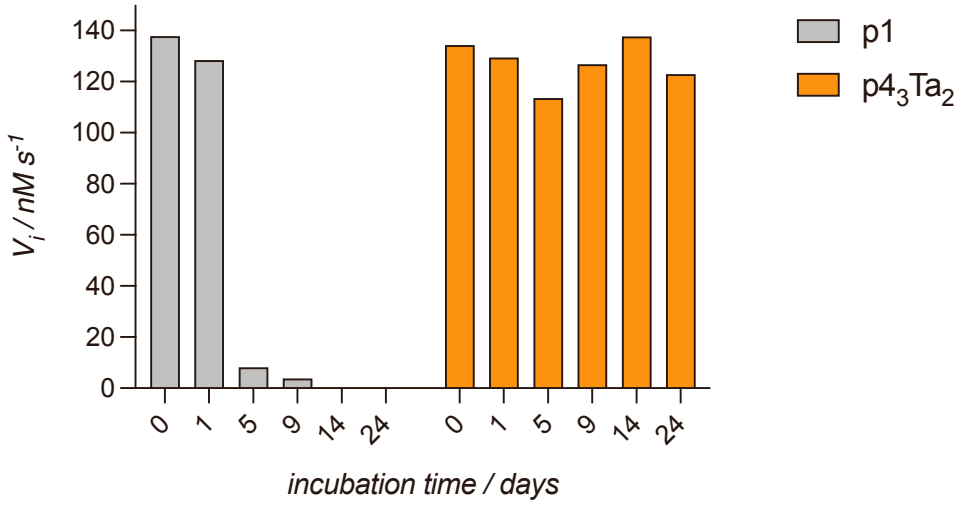


Figure S9. Long term temperature resistance of p4$_3$Ta$_2$. Hydrolysis activity of PFE variants **p1** and bicyclic **p4$_3$Ta$_2$** shows retained of activity of the crosslinked protein after more than three weeks of incubation at 50°C in 50 mM HEPES, 50 mM NaCl, pH 8.

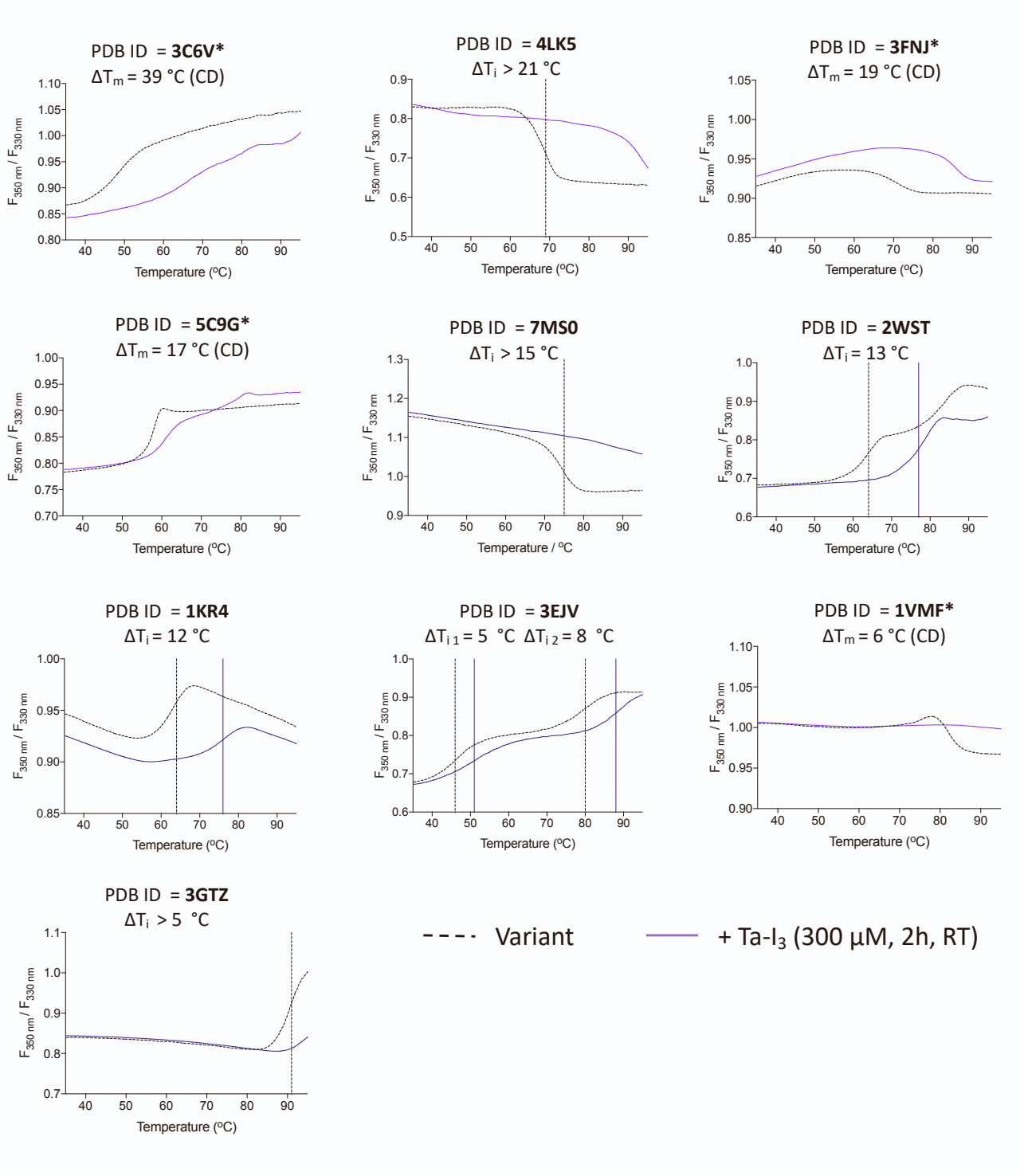


Figure S10. Widespread trimer stabilization by triselectrophile addition. DSF thermal denaturation measurements for the ten trimeric designs which demonstrated a stabilizing effect after addition of the crosslinker Ta-I_3 ($\Delta T_i > 5$°C). The bicyclized products of proteins derived from PDB IDs 3C6V, 5C9G, 3FNJ and 1VMF were further characterized and are presented in Figure 5. For these four examples, the T_m values indicated here were derived from circular dichroism (CD) measurements (Figure 5), whilst the T_i values for all other designs were determined by DSF.

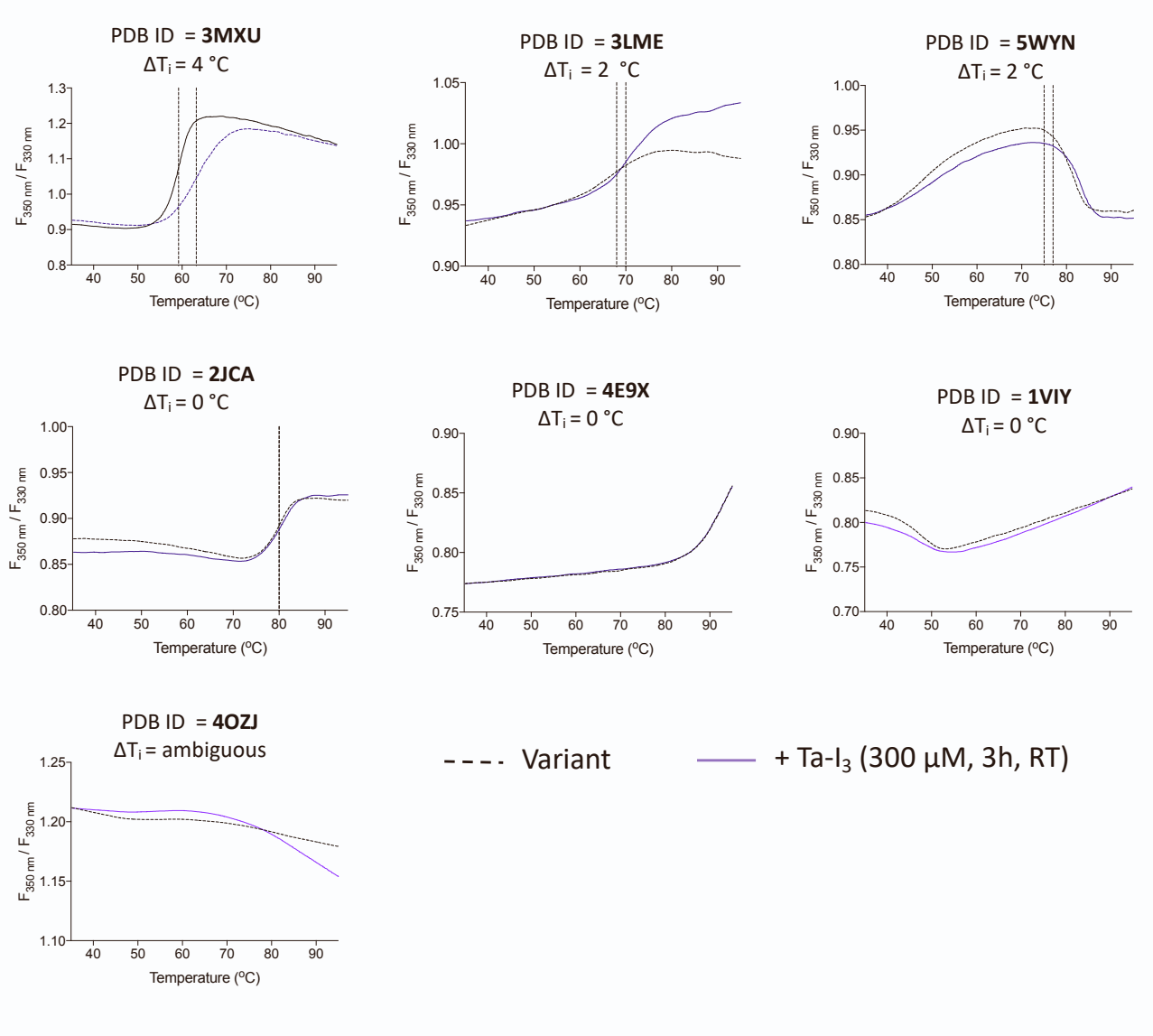
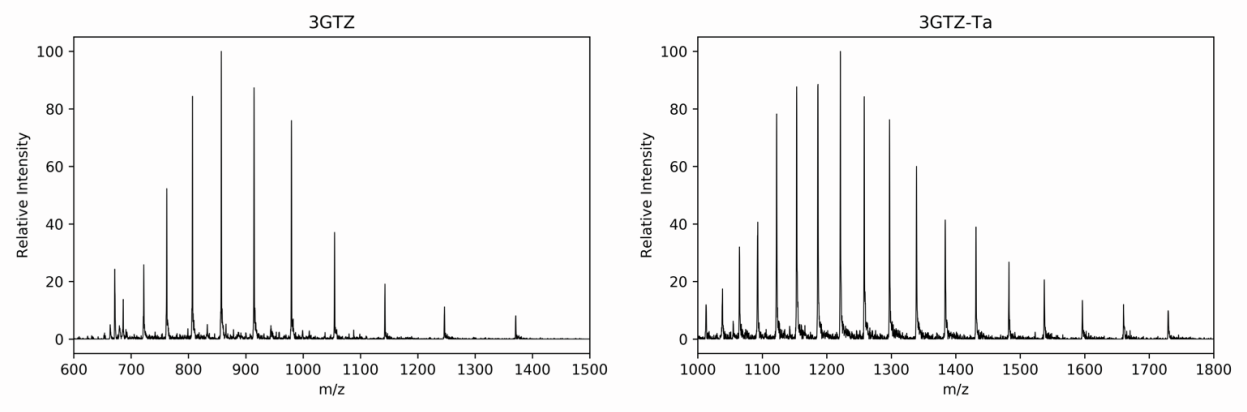


Figure S11. Unsuccessful stabilization of trimer variants. DSF thermal melt measurements of the seven tested designs for trimer crosslinking which failed to show a stabilization effect upon addition of the tris-electrophilic crosslinker Ta-I_3 (ΔT_i < 5°C).



3GTZ (N20C & Q96C)
MW calc. = 13 700.5 Da
found = 13 701.5 Da

Charge	m/z found $[M+H]^+$	m/z calc.
20	686.3	686.0
19	722.1	722.1
18	762.2	762.1
17	807.1	806.9
16	857.3	857.3
15	914.5	914.4
14	979.7	979.6
13	1054.9	1054.9
12	1142.8	1142.7
11	1246.7	1246.5
10	1371.0	1371.1

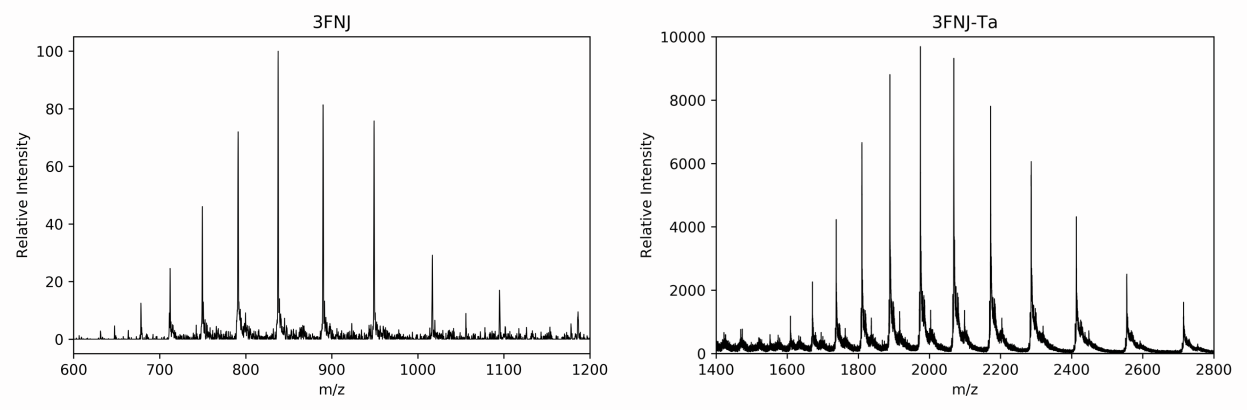
Un-reacted 3GTZ peak list

single-crosslinked*:* **$3GTZ_3Ta_1$**
MW calc. = 41 479.9 Da Da
found. = 41 482.7 Da

Charge	m/z found $[M+H]^+$	m/z calc.
39	1064.6	1064.6
38	1092.6	1092.6
37	1122.3	1122.1
36	1153.3	1153.2
35	1186.1	1186.1
34	1221.1	1221.0
33	1258.1	1258.0
32	1297.3	1297.2
31	1339.2	1339.1
30	1383.7	1383.7
29	1431.4	1431.3
28	1482.5	1482.4
27	1537.3	1537.3
26	1596.4	1596.4
25	1660.2	1660.2

$3GTZ_3Ta_1$ peak list

Figure S12. Mass spectra of 3GTZ (N20C & Q96C) before (left) and after (right) addition of the crosslinker Ta-I_3, and peak lists for the un-reacted protein and single-crosslinked covalent trimer formed ($3GTZ_3Ta_1$). The molecular weight of the crosslinked version corresponds to the mass of three monomers of the un-reacted protein variant (13,700.5 Da) plus only one molecule of reacted Ta-I_3 (381.37 Da), minus the hydrogen atoms of the three cysteine thiols. The failure of the second crosslinking site suggests possible solvent inaccessibility of the introduced cysteine side chains. Despite the failure to bicyclize, thermal stabilization (ΔT_i > 5°C, Figure S10) is observed for the mono-crosslinked trimer. Data collected on Agilent LC/MSD XT ESI-Quadrupole LC/MS.



I4 (3FNJ, E7C & E108C)
MW calc. = 14 221.4 Da
found = 14 222.8 Da

bicyclic **I4$_3$Ta$_2$**
MW calc. = 43 420.9 Da
found. = 43 421.1 Da

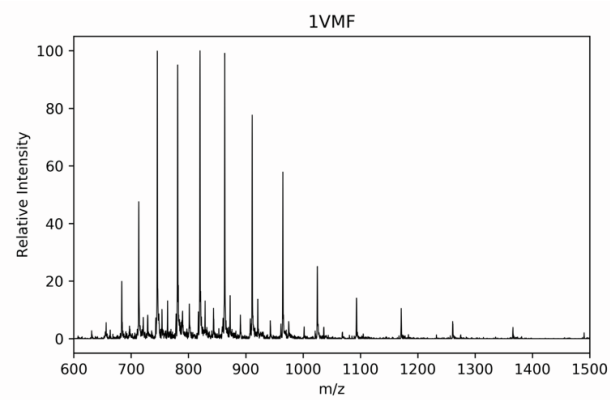
Charge	m/z found $[M+H]^+$	m/z calc.
21	678.2	678.3
20	712.1	712.2
19	749.6	749.6
18	791.2	791.2
17	837.7	837.6
16	889.9	889.9
15	949.2	949.2
14	1016.9	1016.9
13	1094.9	1095.1
12	1186.2	1186.3

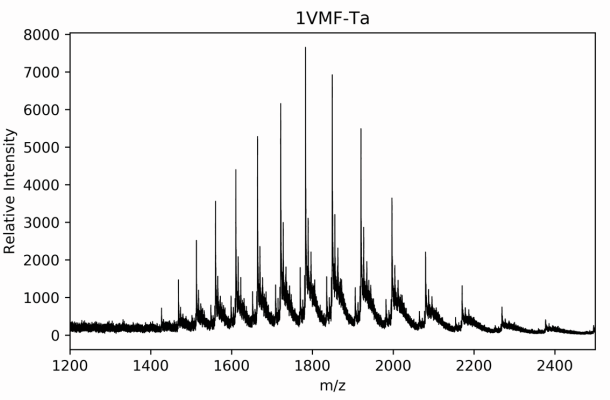
Un-reacted I4 peak list

Charge	m/z found $[M+H]^+$	m/z calc.
27	1609.19	1609.18
26	1671.11	1671.03
25	1737.85	1737.84
24	1810.22	1810.2
23	1888.88	1888.87
22	1974.7	1974.68
21	2068.68	2068.66
20	2171.94	2172.05
19	2286.32	2286.31
18	2413.33	2413.27
17	2555.25	2555.17
16	2714.85	2714.81

$I4_3Ta_2$ peak list (43 421.1 Da species)

Figure S13. Mass spectra and peak lists for I4 (PDB ID: 3FNJ, E7C & E108C) before (left) and after (right) crosslinking with two Ta-I$_3$ molecules, producing bicyclic **I4$_3$Ta$_2$** (Figure 5A). The found mass of **I4$_3$Ta$_2$** corresponds to the molecular weight of three copies of the un-reacted protein variant (14 221.4 Da) plus two molecules of reacted Ta (381.37 Da), minus six hydrogens. Data collected on Agilent LC/MSD XT ESI-Quadrupole LC/MS (**I4**) and Agilent 6230 ESI-TOF LC/MS (**I4$_3$Ta$_2$**).





b4 (1VMF, E66C & T132C)
MW calc. = 16 381.5 Da
found = 16 383.6 Da

Charge	m/z found $[M+H]^+$	m/z calc.
24	683.8	683.6
23	713.3	713.2
22	745.7	745.6
21	781.2	781.1
20	820.2	820.1
19	863.3	863.2
18	911.2	911.1
17	964.7	964.6
16	1024.9	1024.8
15	1093.2	1093.1
14	1171.2	1171.1
13	1261.0	1261.1
12	1366.1	1366.1

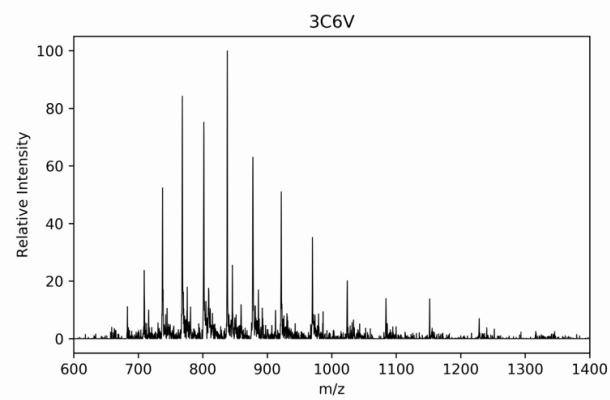
Un-reacted b4 peak list

bicyclic **$b4_3Ta_2$**
MW calc. = 49 901.1 Da
found. = 49 901.1 Da

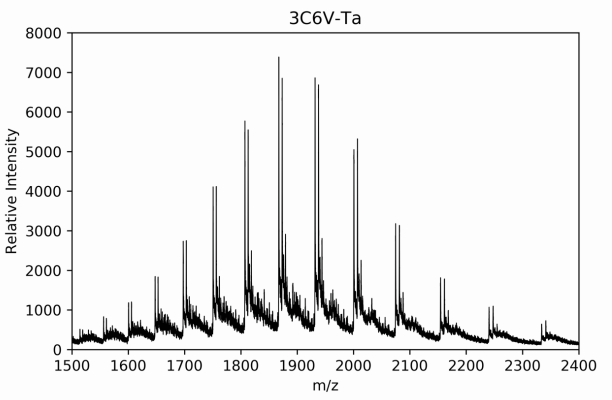
Charge	m/z found $[M+H]^+$	m/z calc.
35	1426.85	1426.75
34	1468.60	1468.68
33	1513.13	1513.15
32	1560.42	1560.41
31	1610.69	1610.71
30	1664.37	1664.37
29	1721.72	1721.73
28	1783.17	1783.18
27	1849.18	1849.19
26	1920.23	1920.27
25	1996.96	1997.04
24	2080.16	2080.21
23	2170.60	2170.61
22	2269.20	2269.23
21	2377.25	2377.24

$b4_3Ta_2$ peak list (49 901.1 Da species)

Figure S14. Mass spectra and peak lists for b4 (PDB ID: 1VMF, E66C & T132C) before (left) and after (right) crosslinking by two Ta-I_3 molecules, producing bicyclic **$b4_3Ta_2$** (Figure 5B). The found mass corresponds to the molecular weight of three copies of the un-reacted protein variant (16 381.5 Da) plus two molecules of reacted Ta (381.37 Da), minus six hydrogens. An additional species with a mass difference of approximately +178 Da (*MW* = 16 561 Da) was observed for the un-reacted protein, corresponding to a common post-translational gluconylation of the protein N-terminus [6]. This mass increase becomes more prominent in the subsequent spectra of the covalently cross-linked trimer due to the probability of at least one gluconylated monomer forming part of a trimer (*MW* = 50 079 Da). Data collected on Agilent LC/MSD XT ESI-Quadrupole LC/MS (**b4**) and Agilent 6230 ESI-TOF LC/MS (**$b4_3Ta_2$**).



a4 (3C6V, E45C & A70C)
MW calc. = 18 411.8 Da
found = 18 413.7 Da



bicyclic **a4$_3$Ta$_2$**
MW calc. = 55 992.1 Da
found. = 55 991.7 Da

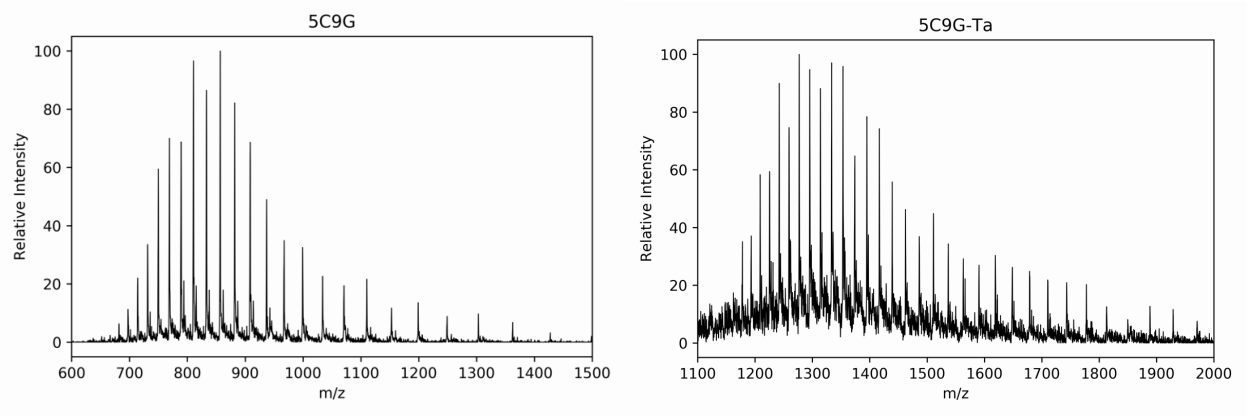
Charge	m/z found $[M+H]^+$	m/z calc.
27	683.3	682.9
26	709.2	709.1
25	737.7	737.5
24	768.4	768.2
23	801.7	801.5
22	838.0	837.9
21	877.7	877.8
20	921.7	921.6
19	970.0	970.0
18	1024.1	1023.9
17	1084.1	1084.0
16	1151.9	1151.7

Un-reacted a4 peak list

Charge	m/z found $[M+H]^+$	m/z calc.
36	1556.28	1556.34
35	1600.77	1600.77
34	1647.85	1647.83
33	1697.67	1697.73
32	1750.76	1750.75
31	1807.21	1807.20
30	1867.41	1867.40
29	1931.77	1931.76
28	2000.74	2000.72
27	2074.79	2074.78
26	2154.55	2154.54
25	2240.70	2240.68
24	2333.98	2334.00

$a4_3Ta_2$ peak list (55 991.7 Da species)

Figure S15. Mass spectra and peak lists for a4 (PDB ID: 3C6V, E45C & A70C) before (left) and after (right) crosslinking by two Ta-I$_3$ molecules, producing bicyclic **a4$_3$Ta$_2$** (Figure 5C). The found mass corresponds to the molecular weight of three copies of the un-reacted protein variant (18 411.8 Da) plus two molecules of reacted Ta (381.37 Da), minus six hydrogens. An additional species with a mass difference of approximately +178 Da (*MW* = 18 592 Da) was observed for the un-reacted protein, corresponding to a common post-translational gluconylation of the protein N-terminus [6]. This mass increase becomes more prominent in the subsequent spectra of the covalently cross-linked trimer due to the probability of at least one gluconylated monomer forming part of a trimer (*MW* = 56 170). Data collected on Agilent LC/MSD XT ESI-Quadrupole LC/MS (**a4**) and Agilent 6230 ESI-TOF LC/MS (**a4$_3$Ta$_2$**).



e4 (5C9G, S176C & G221C)
MW calc. = 29 951.1 Da
found = 29 952.9 Da

bicyclic **e4$_3$Ta$_2$**
MW calc. = 90 610.1 Da
found. = 90 606.7 Da

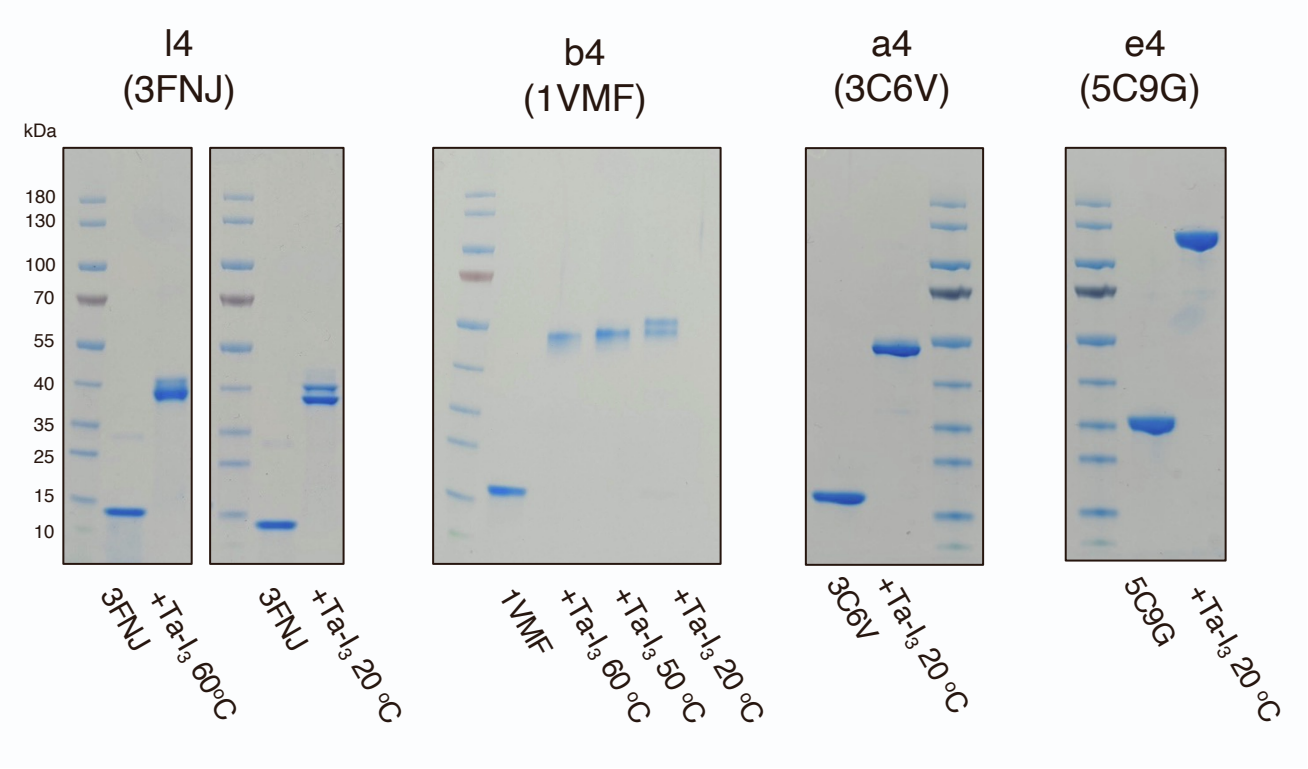
Charge	m/z found $[M+H]^+$	m/z calc.
41	731.4	731.5
40	749.8	749.8
39	769.1	769.0
38	789.3	789.2
37	810.6	810.5
36	833.0	833.0
35	856.9	856.7
34	882.0	881.9
33	908.8	908.6
32	937.1	937.0
31	967.4	967.2
30	999.4	999.4
29	1034.0	1033.8
28	1079.9	1070.7
27	1110.4	1110.3
26	1153.2	1153.0
25	1199.2	1199.0
24	1249.2	1249.0

Un-reacted e4 peak list

Charge	m/z found $[M+H]^+$	m/z calc.
74	1225.4	1225.5
73	1242.2	1242.2
72	1259.3	1259.5
71	1277.0	1277.2
70	1295.4	1295.4
69	1314.0	1314.2
68	1333.7	1333.5
67	1353.4	1353.4
66	1374.1	1373.9
65	1395.0	1395.0
64	1416.7	1416.8
63	1439.2	1439.3
62	1462.3	1462.5
61	1486.3	1486.4
60	1511.2	1511.2
59	1537.1	1536.8
58	1563.2	1563.2
57	1590.6	1590.7

$e4_3Ta_2$ peak list (90 606.7 Da species)

Figure S16. Mass spectra and peak lists for e4 (PDB ID: 5C9G, S176C & G221C) before (left) and after (right) crosslinking by two Ta-I$_3$ molecules, producing bicyclic **e4$_3$Ta$_2$** (Figure 5D). The found mass corresponds to the molecular weight of three copies of the un-reacted protein variant (29 951.1 Da) plus two molecules of reacted Ta (381.37 Da), minus six hydrogens. Data collected on Agilent LC/MSD XT ESI-Quadrupole LC/MS.



Variant	Expected MW monomer / kDa	Expected MW trimer / kDa
a4	18.4	56.0
e4	30.0	90.6
l4	14.2	43.4
b4	16.4	49.0

Figure S17. SDS-PAGE analysis of 2-hour reactions of Ta-l_3 with homotrimers shows formation of crosslinked species (relates to trimers presented in Figure 5). All four proteins demonstrate formation of a covalent trimer. In analogy to the bicyclic PFE variant (**$p4_3Ta_2$**), we observed migration behaviour deviating from the linear protein ladder (higher apparent masses) due to the altered protein chain topology. The reactions for **l4** (3FNJ) and **b4** (1VMF) at room temperature show evidence of two closely related species, however increasing the reaction temperature resolves both to a predominant single product. The crosslinked SDS-PAGE **b4** sample was prepared by addition of 6 M urea in addition to loading dye, and not heated to 95°C, as the standard protocol failed to migrate through the gel.

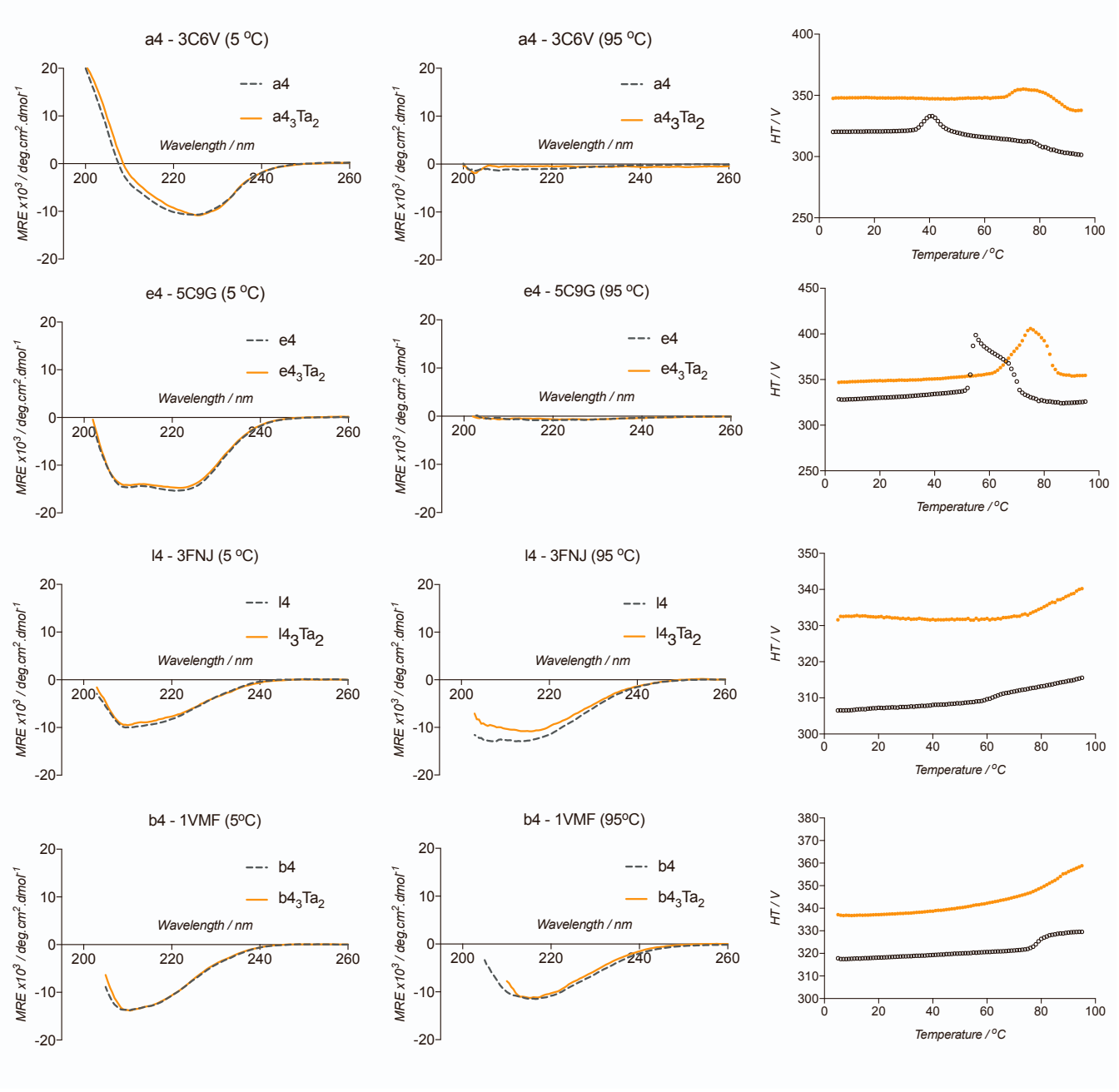


Figure S18. Circular dichroism validates stabilization of bicyclic trimers (relating to data presented in Figure 5). Circular dichroism spectra (λ = 200–260 nm) of trimeric variants before (black, dash) and after (orange) crosslinking, at 5°C (left) and 95°C (middle). Right: High tension (HT) recordings for thermal denaturation measurements (λ = 220).

Supplementary Tables

Table S1. High-resolution native MS of PFE trimers. Includes **p1** and crosslinked trimer variants **p2$_3$Ta**, **p3$_3$Ta** and **p4$_3$Ta$_2$**. Data collected on TIMS-Qq-TOF mass spectrometer with an ESI source.

Sample	Calculated mass trimer (*MW* / Da)	Found (*m*/*z*, with *z* = 20$^+$)	Deconvoluted (*MW* / Da)	Mass accuracy (ppm)
p1	92784.00	4640.12	92783.15	-9 ppm
p2$_3$Ta	93168.44	4659.73	93167.10	-14 ppm
p3$_3$Ta	93087.38	4655.59	93086.20	-9.4 ppm
p4$_3$Ta	93471.82	4674.83	93469.89	-20 ppm

Table S2. Peak list for MS charge state analysis of p2$_3$Ta. Most abundant charge state is Z = 91 (Figure 1C). Data collected on Agilent Technologies G6230A ESI-TOF LC/MS.

Charge	m/z found $[M+H]^+$	Relative Abundance	Charge	m/z found $[M+H]^+$	Relative Abundance
114	818.20	0.029	79	1180.27	0.522
113	825.42	0.041	78	1195.38	0.492
111	840.28	0.053	77	1210.89	0.482
110	847.90	0.074	76	1226.81	0.483
109	855.68	0.096	75	1243.15	0.473
108	863.59	0.111	74	1259.94	0.488
107	871.64	0.139	73	1277.19	0.470
106	879.86	0.172	72	1294.91	0.469
105	888.24	0.209	71	1313.14	0.461
104	896.76	0.237	70	1331.88	0.445
103	905.47	0.292	69	1351.17	0.443
102	914.33	0.355	68	1371.04	0.421
101	923.38	0.417	67	1391.47	0.406
100	932.60	0.485	66	1412.54	0.402
99	942.02	0.578	65	1434.26	0.377
98	951.62	0.661	64	1456.70	0.352
97	961.42	0.758	63	1479.76	0.344
96	971.43	0.819	62	1503.61	0.310
95	981.64	0.893	61	1528.25	0.293
94	992.07	0.943	60	1553.71	0.268
93	1002.74	0.996	59	1580.02	0.253
92	1013.62	0.970	58	1607.25	0.233
91	**1024.75**	**1.000**	57	1635.42	0.209
90	1036.13	0.961	56	1664.61	0.194
89	1047.76	0.951	55	1694.85	0.177
88	1059.65	0.873	54	1726.22	0.157
87	1071.82	0.831	53	1758.77	0.145
86	1084.28	0.794	52	1792.58	0.123
85	1097.02	0.724	51	1827.71	0.111
84	1110.06	0.680	50	1864.25	0.099
83	1123.43	0.631	49	1902.27	0.084
82	1137.12	0.600	48	1941.88	0.069
81	1151.14	0.558	47	1983.17	0.058
80	1165.52	0.525	46	2026.28	0.047

Table S3. Peak list for MS charge state analysis of p3$_3$Ta. Most abundant charge state is Z = 61 (Figure 1C). Data collected on Agilent Technologies G6230A ESI-TOF LC/MS.

Charge	m/z found $[M+H]^+$	Relative Abundance	Charge	m/z found $[M+H]^+$	Relative Abundance
85	1096.11	0.008	58	1605.81	0.975
84	1109.05	0.006	57	1633.96	0.945
83	1122.38	0.009	56	1663.12	0.884
82	1136.05	0.019	55	1693.34	0.832
81	1150.13	0.016	54	1724.68	0.777
80	1164.52	0.020	53	1757.20	0.711
79	1179.19	0.031	52	1790.98	0.631
78	1194.32	0.042	51	1826.07	0.562
77	1209.81	0.053	50	1862.57	0.504
76	1225.7	0.074	49	1900.57	0.444
75	1242.05	0.093	48	1940.14	0.393
74	1258.82	0.119	47	1981.40	0.340
73	1276.06	0.152	46	2024.46	0.291
72	1293.76	0.200	45	2069.42	0.239
71	1311.97	0.248	44	2116.43	0.192
70	1330.7	0.312	43	2165.63	0.159
69	1349.96	0.389	42	2217.16	0.131
68	1369.8	0.469	41	2271.22	0.108
67	1390.24	0.575	40	2327.97	0.090
66	1411.29	0.692	39	2387.64	0.074
65	1432.98	0.769	38	2450.44	0.062
64	1455.35	0.868	37	2516.64	0.052
63	1478.45	0.932	36	2586.52	0.039
62	1502.28	0.970	35	2660.40	0.027
61	**1526.88**	**1.000**	34	2738.63	0.018
60	1552.32	0.999	33	2821.58	0.011
59	1578.61	0.989	32	2909.88	0.006

Table S4. Peak list for MS charge state analysis of $p4_3Ta_2$. Most abundant charge state is Z = 42 (Figure 1C). Data collected on Agilent Technologies G6230A ESI-TOF LC/MS.

Charge	m/z found $[M+H]^+$	Relative Abundance
66	1417.20	0.023
63	1484.61	0.035
62	1508.50	0.042
61	1533.17	0.058
60	1558.73	0.064
59	1585.09	0.080
58	1612.39	0.104
57	1640.67	0.132
56	1669.95	0.156
55	1700.28	0.192
54	1731.75	0.233
53	1764.39	0.277
52	1798.31	0.326
51	1833.55	0.383
50	1870.20	0.446
49	1908.34	0.508
48	1948.08	0.574
47	1989.50	0.647
46	2032.74	0.722
45	2077.88	0.808
44	2125.09	0.889
43	2174.48	0.970
42	**2226.23**	**1.000**
41	2280.50	0.980
40	2337.49	0.888
39	2397.41	0.757
38	2460.47	0.621
37	2526.95	0.485
36	2597.12	0.381
35	2671.31	0.295
34	2749.85	0.246
33	2833.15	0.204
32	2921.66	0.170

Table S5. Peak lists for MS charge state analysis of un-reacted p2, p3 and p4. Most abundant charge states are 37, 37 and 38 respectively. Data collected on Agilent Technologies G6230A ESI-TOF LC/MS.

p2

Charge	m/z found $[M+H]^+$	Relative Abundance
47	659.0	1686
46	673.3	3526
45	688.3	7620
44	703.9	16883
43	720.2	34836
42	737.4	73319
41	755.3	139003
40	774.2	208155
39	794.0	276399
38	814.9	319631
37	**836.9**	**330838**
36	860.1	314184
35	884.6	283557
34	910.6	253840
33	938.2	222451
32	967.5	194658
31	998.6	169614
30	1031.9	143373
29	1067.4	122401
28	1105.5	102762
27	1146.4	90941
26	1190.5	85006
25	1238.1	81834
24	1289.6	83511
23	1345.6	85830
22	1406.8	80611
21	1473.7	62805
20	1547.3	37650
19	1628.7	17555
18	1719.2	8198
17	1820.2	4385
16	1933.9	2530

p3

Charge	m/z found $[M+H]^+$	Relative Abundance
44	703.3	310
43	719.6	549
42	736.7	1083
41	754.7	1871
40	773.5	2638
39	793.3	3720
38	814.2	4391
37	**836.2**	**4919**
36	859.4	4820
35	883.9	4535
34	909.9	4098
33	937.4	3641
32	966.7	3317
31	997.8	3012
30	1031.0	2446
29	1066.6	2211
28	1104.6	1756
27	1145.5	1574
26	1189.5	1419
25	1237.1	1422
24	1288.6	1401
23	1344.5	1464
22	1405.6	1365
21	1472.5	1084
20	1546.1	691
19	1627.4	325

p4

Charge	m/z found $[M+H]^+$	Relative Abundance
45	687.8	377
44	703.4	575
43	719.7	1139
41	754.8	2929
40	773.6	5755
39	793.4	8270
38	**814.3**	**9580**
37	836.3	9069
36	859.5	9763
35	884.0	9723
34	910.0	9134
33	937.5	8188
32	966.8	7341
31	997.9	5153
30	1031.2	5660
29	1066.7	4647
28	1104.7	3492
27	1145.6	3667
26	1189.7	2481
25	1237.2	3149
24	1288.7	2859
23	1344.7	2881
22	1405.8	2480
21	1472.6	1993
20	1546.2	1217
19	1627.6	619
18	1717.9	314

Table S6. **Thermal stabilization of crosslinked PFE variants.** Inflection melting temperatures (T_i) for PFE variants **p1**-**p4** and their crosslinked variants. Values are an average of triplicate DSF measurements, T_i values are not reported where the dynamic range of the measurement is less than 30% of the corresponding native melt (melting curves presented in Figure 2B, 2C and Figure S5).

	Native T_i / °C	**Denaturing T_i / °C**			
	0.0 M GuHCl	0.5 M GuHCl	1.0 M GuHCl	1.5 M GuHCl	2.0 M GuHCl
p1	79.6	62.0	52.6	-	-
p2	79.7	62.3	52.9	-	-
p3	76.9	59.3	-	-	-
p4	76.8	59.3	-	-	-
$p2_3Ta$	80.6	65.0	54.6	-	-
$p3_3Ta$	84.3	67.7	57.3	51.2	-
$p4_3Ta_2$	85.2	70.9	61.6	53.0	-

Table S7. Denaturant resistant activity of crosslinked PFE variants (relates to Figure 2D). Initial rates of *p*-NPA hydrolysis by PFE variants (errors account for 1σ). Initial rate values determined from curves presented in Figure 2E, 2F and Figure S6. The relative rates (relative to the initial **$p4_3Ta_2$** rate at each given GuHCl concentration) are presented in Figure 2D.

	p1		**$p2_3Ta$**		**$p3_3Ta$**		**$p4_3Ta_2$**	
Concentration of GuHCl / M	**Initial rate / mM.s^{-1}**	**Error**	**Initial rate / mM.s^{-1}**	**Error**	**Initial rate / mM.s^{-1}**	**Error**	**Initial rate / mM.s^{-1}**	**Error**
0.0	$1.06 \cdot 10^{-3}$	$0.16 \cdot 10^{-3}$	$1.28 \cdot 10^{-3}$	$0.14 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$0.17 \cdot 10^{-3}$	$1.25 \cdot 10^{-3}$	$0.13 \cdot 10^{-3}$
0.5	$5.57 \cdot 10^{-4}$	$0.13 \cdot 10^{-4}$	$5.48 \cdot 10^{-4}$	$0.12 \cdot 10^{-4}$	$6.27 \cdot 10^{-4}$	$0.093 \cdot 10^{-4}$	$5.97 \cdot 10^{-4}$	$0.15 \cdot 10^{-4}$
1.0	$2.78 \cdot 10^{-4}$	$0.053 \cdot 10^{-4}$	$3.65 \cdot 10^{-4}$	$0.088 \cdot 10^{-4}$	$3.38 \cdot 10^{-4}$	$0.068 \cdot 10^{-4}$	$3.45 \cdot 10^{-4}$	$0.19 \cdot 10^{-4}$
1.5	$7.08 \cdot 10^{-6}$	$0.79 \cdot 10^{-6}$	$9.89 \cdot 10^{-5}$	$0.091 \cdot 10^{-5}$	$1.52 \cdot 10^{-4}$	$0.027 \cdot 10^{-4}$	$2.03 \cdot 10^{-4}$	$0.052 \cdot 10^{-4}$
2.0	$5.74 \cdot 10^{-7}$	$0.40 \cdot 10^{-7}$	$1.36 \cdot 10^{-5}$	$0.18 \cdot 10^{-5}$	$5.36 \cdot 10^{-5}$	$0.14 \cdot 10^{-5}$	$7.52 \cdot 10^{-5}$	$0.30 \cdot 10^{-5}$

Table S8. Data collection and structural refinement statistics for bicyclic $p4_3Ta_2$. Values in parentheses represent the highest resolution shell (2.50–2.60 Å). Crystal structure presented in Figure 4.

	$p4_3Ta_2$ (PDB ID = 8pi1)
Data Collection	
Beamline	Diamond Light Source, I04
Space group	C 1 2 1
Unit cell dimensions (Å): a, b, c	254.80, 146.25, 154.59
Unit cell angles (°): a, b, g	90.00, 122.63, 90.00
Resolution (Å)	2.50
Number of unique reflections	164678
$CC_{1/2}$	99.7 *(52.1)*
I/σ	8.88 *(1.12)*
Completeness (%)	100.0 *(99.9)*
Refinement	
Resolution (Å)	2.50
No. unique reflections used in refinement	156444
R_{work}	0.182
R_{free}	0.224
No. protein atoms used in refinement	31 894
No. water molecules used in refinement	533
Average B-factors (Å^2)	45.25
R.M.S deviations – length (Å)	0.010
R.M.S deviations – angle (°)	1.320
Ramachandran favoured residues (%)	97
Ramachandran outlying residues (%)	0

Table S9. PDB ID codes of 119 proteins identified by search for non-redundant C3-symmetric protein trimers. Each contains up to one cysteine residue and at least one tryptophan per monomer, as well as a hexa-histidine tag in their PDB sequence.

No.	PDB	Cys	No.	PDB_ID	Cys	No.	PDB	Cys	No.	PDB	Cys
1	6T6J	0	**32**	4UMI	0	**63**	2JCA	0	**94**	5WYN	1
2	2CZ4	0	**33**	5BOH	0	**64**	3ZEZ	1	**95**	4F3J	1
3	3FNJ	0	**34**	4PXK	0	**65**	7MS0	1	**96**	4UAH	1
4	3LME	0	**35**	5KKH	0	**66**	4OZJ	1	**97**	6YZY	1
5	3H5I	0	**36**	4XXJ	0	**67**	4E9X	1	**98**	4EX8	1
6	2WDY	0	**37**	6XT4	0	**68**	3T9W	1	**99**	3ZE3	1
7	3MXU	0	**38**	4UC0	0	**69**	5TB7	1	**100**	6YSP	1
8	4D0V	0	**39**	3E35	0	**70**	4XI0	1	**101**	4LHR	1
9	3I7T	0	**40**	2ETV	0	**71**	7RFO	1	**102**	4OUS	1
10	4JCU	0	**41**	2WAM	0	**72**	4XC5	1	**103**	4OUL	1
11	3Q8U	0	**42**	6EUS	0	**73**	4NJN	1	**104**	1VMH	1
12	3C6V	0	**43**	5KA6	0	**74**	6PSP	1	**105**	4M1A	1
13	5B8F	0	**44**	3K9A	0	**75**	5C9G	1	**106**	3OFV	1
14	5XUB	0	**45**	5APZ	0	**76**	3GVF	1	**107**	3B8L	1
15	6B7C	0	**46**	2IBL	0	**77**	5WUF	1	**108**	2RDM	1
16	4GCY	0	**47**	5ZHY	0	**78**	2NT8	1	**109**	4DI1	1
17	3EJV	0	**48**	5LNL	0	**79**	4LK5	1	**110**	4NSN	1
18	3ZF1	0	**49**	4Y2L	0	**80**	3H0U	1	**111**	5N2C	1
19	1VIY	0	**50**	3LYB	0	**81**	2WST	1	**112**	4HZ5	1
20	7DSZ	0	**51**	2YO2	0	**82**	2DCH	1	**113**	3AM2	1
21	5WUC	0	**52**	3BK6	0	**83**	3GTZ	1	**114**	3ZJB	1
22	1PG6	0	**53**	3WPP	0	**84**	1VPH	1	**115**	3K12	1
23	4G9Q	0	**54**	3QV0	0	**85**	4XEL	1	**116**	4L8P	1
24	6ZLO	0	**55**	3LX2	0	**86**	1VMF	1	**117**	7E4G	1
25	3GKB	0	**56**	2XQH	0	**87**	2YKP	1	**118**	2B2H	1
26	2PBZ	0	**57**	6H21	0	**88**	2YKO	1	**119**	5KVB	1
27	5O68	0	**58**	2C3F	0	**89**	1KR4	1			
28	1LR0	0	**59**	4USX	0	**90**	1VHF	1			
29	4LKB	0	**60**	5HRZ	0	**91**	3RPX	1			
30	4EC6	0	**61**	4LGO	0	**92**	4LEH	1			
31	2F1V	0	**62**	5WTR	0	**93**	4L8O	1			

Table S10. Bicyclization via Ta-I_3 was attempted for 14 unique CATH domains, of which variants for 13 domains were successfully expressed and purified. Eight domains demonstrated significant thermal stabilization after addition of the crosslinker, while bicyclization was confirmed for four examples (*) by mass spectrometry (Figures S13–S16). **Indicates T_m values (equilibrium midpoint) obtained by circular dichroism measurements (Figure 5), all other values indicate T_i (inflection temperature) obtained by DSF measurements (Figures S10 and S11). Sequences are provided in (Table S11).

No.	CATH Domain	Domain name	PDB ID (variant)	Crosslinking Sites	Variant T_i / °C	Crosslinked T_i / °C	ΔT_i / °C
1	3.30.429.10	Macrophage Migration Inhibitory Factor	3C6V* (a4)	*E45C & A70C*	41	80	39**
			7MS0	*R10C & E107C*	75	>90	>15
2	3.90.226.10	2-enoyl-CoA Hydratase	5C9G* (e4)	*S176C & G221C*	54	71	17**
			4LK5	*A185C & E231C*	69	>90	>21
3	3.40.250.10	Rhodanese-like domain	3FNJ* (I4)	*E7C & E108C*	61	80	19**
4	2.60.120.460	YjbQ-like, Jelly Rolls, Sandwich	1VMF* (b4)	*E66C & T132C*	77	83	6**
5	2.60.90.10	Adenovirus pIV-related, attachment domain	2WST	*V117C & V245C*	64	77	13
6	3.30.70.120	Alpha-Beta Plaits	1KR4	*E34C & E86C*	64	76	12
			4OZJ	*N43C & E108C*	ambiguous	ambiguous	-
7	3.10.450.50	Nuclear Transport Factor 2	3EJV	*A4C & V126C*	46, 80	51, 88	5-8
8	3.30.1330.40	RutC-like	3GTZ	*N20C & Q96C*	90	>95	> 5
			3LME	*D53C & Q136C*	68	70	2
9	2.40.50.100	RNA polymers II, barrel-sandwich hybrid domain	3MXU	*D9C & A115C*	59	63	4
10	2.30.42.10	PDZ domain	5WYN	*Q146C & A327C*	76	78	2
11	3.40.50.300	P-loop containing NTP hydrolases	1VIY	*E79C & A167C*	ambiguous	ambiguous	0
12	3.90.470.20	4'-phosphopantetheinyl transferase domain	2JCA	*S2C & E12C*	80	80	0
13	2.60.40.420	Cupredoxins - blue copper proteins	4E9X	*M195&D254*	>90°C	>90°C	0
14	2.70.40.10	Deoxyuridine triphosphatase (dUTPase)	3ZEZ	*Y57C & K100C*	No expression	-	-

Table S11. Full-length sequences of the expanded suite of C3-symmetric trimer variants designed for bicyclization, including purification tags. All proteins were expressed in a pET28a (+) vector.

3FNJ_INCYPRO (l4)
MANDKKICLLTTYLSLYIDHHTVLADMQNATGKYVVLDVRNAPAQVKKDQIKGAIAMPAKDLATRIGELDPAKTYVV
YDWTGGTTLGKTALLVLLSAGFEAYELAGALCGWKGMQLPVETLADLEHHHHHH

3LME_INCYPRO
MASLKIIAPTDKTITPSGTWSIGARAGCFVFIGGMHGTDRVTGKMVDGDEARIRRMFDNMLAAAEAAGATKADAVRL
TVFVTDVAKYRPVVNKVQKDIWGDGPYPPRTVLCVPALDQGDIAEIDGTFYAPAEGHHHHHH

2JCA_INCYPRO
MGSSHHHHHHSSGLVPRGSHMCIIGVGIDVACVERFGAALERTPALAGRLFLESELLLPGGERRGVASLAARFAAKE
ALAKALGAPAGLLWTDAEVWVEAGGRPRLRVTGTVAARAAELGVASWHVSLSHDAGIASAVVIAEG

3MXU_INCYPRO
MAHHHHHHMGTLEAQTQGPGSMSKTYFTQCHEWLSVEGQVVTVGITDYAQEQLGDLVFIDLPQNGTKLSKGDAAAVV
ESVKAASDVYAPLDGEVVEINAALAESPELVNQKAETEGWLWKMTVQDETQLERLLDECAYKELIG

3C6V_INCYPRO (a4)
MGSSHHHHHHSSGRENLYFQGMPRWLIQHSPNTLTPEEKSHLAQQITQAYVGFGLPAFYVQVHFICQPAGTSFIGGE
QHPNFVALTIYHLCRTMTSDEQRQGFLKRIDAFLTPMFEPKGIDWEYFVTEAPRDLWKINGLAPPAAGSEEEKVWVR
ENRPVRF

3EJV_INCYPRO
MGSDKIHHHHHHENLYFQGMTMCDETIILNVLGQYTRAHDRRDPDAMAALFAPEATIEIVDAVGGASRSISRLEGRD
AIRVAVRQMMAPHGYRAWSQNVVNAPIIVIEGDHAVLDAQFMVFSILAAEVPDGGWPTGTFGAQGRICPIEAGQYRL
TLRTVADGWVISAMRIEHRLPMAFG

3ZEZ_INCYPRO
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMTNTLQVKLLSKNARMPERNHKTDAGYDIFSAETVVLEPQEKA
VIKTDVAVSIPEGCVGLLTSRSGVSSKTHLVIETGKIDAGYHGNLGINIKNDHEDDCMQTIFLRNIDNEKIFEKERH
LYKLGSYRIEKGERIAQLVIVPIWTPELKQVEEFESVSERGEKGFGSSGV

1VIY_INCYPRO
MASLRYIVALTGGIGSGKSTVANAFADLGINVIDADIIARQVVEPGAPALHAIADHFGANMIAADGTLQRRALRERI
FANPCEKNWLNALLHPLIQQETQHQIQQATSPYVLWVVPLLVENSLYKKANRVLVVDVSPETQLKRTMQRDDVTREH
VEQILAAQATREARLCVADDVIDNNGAPDAIASDVARLHAHYLQLASQFVSQEKPEGGSHHHHHH

7MS0_INCYPRO
MAPTYTCWSQCIRISREAKQRIAEAITDAHHELAHAPKYLVQVIFNEVEPDSYFIAAQSASENHIWVQATIRSGRTE
KQKEELLLRLTQEIALILGIPNEEVWVYITCIPGSNMTEYGRLLMEPGEEEKWFNSLPEGLRERLTELEGSSEENLY
FQGLEHHHHHH

4OZJ_INCYPRO
MGSSHHHHHHSSGLVPAGSHMSDADLPNDGGIKLVMAIIRPDKLADVKTALAEVGAPSLTVTCVSGRGSQPAKKSQW
RGEEYTVDLHQKVKVECVVADTPAEDVADAIADAAHTGEKGDGKIFILPVCNAIQVRTGKTGRDAV

4E9X_INCYPRO
MAEREFDMTIEEVTIKVAPGLDYKVFGFNGQVPGPLIHVQEGDDVIVNVTNNTSLPHTIHWHGVHQKGTWRSDGVPG
VTQQPIEAGDSYTYKFKADRIGTLWYHCHVNVNEHVGVRGMWGPLIVDPKQPLPIEKRVTKDVIMMMSTWESAVADK
YGEGGTPCNVADYFSVNAKSFPLTQPLRVKKGDVVKIRFFGAGGGIHAMHSHGHDMLVTHKDGLPLCSPYYADTVLV
SPGERYDVIIEADNPGRFIFHDHVDTHVTAGGKHPGGPITVIEYDGVPVDDWYVWKDKDYDPNFFYSESLKQGYGMF
DHDGFKGEFEQRQRRPGRKLAAALEHHHHHH

5C9G_INCYPRO (e4)
MGHHHHHHSSGVDLGTENLYFQSMTLPIRLDIAAPLAEIVLNKPERRNALSVDMWAAIPGLVAEANANPDVKLILIH
GGDAGAFAAGADISEFETIYATEDAAKASGQRIAQALDAIENSEKPVIAAIEGACVGGGVSLAMAADLRVAGEGAKF
GVTPGKLGLVYPAGDTRRLLAAVGPGATKDILFTGRIFTAGEAKCLGLIDRLVEKGTALEAARVWAGEIAAISQWSV
RATKRMIRGLQTCWTDETPEAQSLFLNGFANEDFKEGYRAFLDKRPAKFTYR

4LK5_INCYPRO

MGHHHHHHSSGVDLGTENLYFQSMPSSAIATLAPVAGLDVTLSDGVFSVTINRPDSLNSLTVPVITGIADAMEYAAT
DPEVKVVRIGGAGRGFSSGAGISADDVSDGGGVPPDEIILEINRLVRAIAALPHPVVAVVQGPAAGVGVSIALACDV
VLASENAFFMLAFTKIGLMPDGGASALVAAAVGRIRAMQMALLPERLPAAEALCWGLVTAVYPADEFEAEVDKVIAR
LLSGPAVAFAKTKLAINAATLTCLDPALQREFDGQSVLLKSPDFVEGATAFQQRRTPNFTDR

2WST_INCYPRO

MGSSHHHHHHSSGLVPRGSHMASMTGGQQGRILCYPTLWTGPAPEANVTFSGENSPSGILRLCLSRTGGTVIGTLSV
QGSLTNPSTGQTLGMNLYFDADGNVLSESNLVRGSWGMKDQDTLVTPIANGQYLMPNLTAYPRLIQTLTSSYIYTQA
HLDHNNSCVDIKIGLNTDLRPTAAYGLSFTMTFTNSPPTSFGTDLVQFGYLGQD

3GTZ_INCYPRO

MASLSIVRIDAEDRWSDVVIYNCTLWYTGVPENLDADAFEQTANTLAQIDAVLEKQGSSKSRILDATIFLSDKADFA
AMNKAWDAWVVAGHAPVRCTVCAGLMNPKYKVEIKIVAAVEGHHHHHH

1VMF_INCYPRO (b4)

MGSDKIHHHHHHMKTFHLTTQSRDEMVDITSQIETWIRETGVTNGVAIVSSLHTTAGITVNENADPDVKRDMIMRLD
CVYPWHHENDRHMEGNTAAHLKTSTVGHAQTLIISEGRLVLGTWQGVYFCEFDGPRTNRKFVVKLLCD

1KR4_INCYPRO

MGSSHHHHHHSSGREALYFMGHMILVYSTFPNEEKALEIGRKLLEKRLIACFNAFCIRSGYWWKGEIVQDKEWAAIF
KTTEEKEKELYEELRKLHPYETPAIFTLKVCNILTEYMNWLRESVLGS

5WYN_INCYPRO

MAVPSPPPASPRSCYNFIADVVEKTAPAVVYIEILDRHPFLGREVPISNGSGFVVAADGLIVTNAHVVADRRRVRVR
LLSGDTYEAVVTAVDPVADIATLRIQTKEPLPTLPLGRSADVRQGEFVVAMGSPFALQNTITSGIVSSAQRPARDLG
LPQTNVEYIQTDAAIDFGNSGGPLVNLDGEVIGVNTMKVTCGISFAIPSDRLREFLHRGEKKNSSSGISGSQRRYIG
VMMLTLSPSILAELQLREPSFPDVQHGVLIHKVILGSPAHRAGLRPGDVILAIGEQMVQNAEDVWEAVRTQSQLAVQ
IRRGRETLTLYVTPEVTEHHHHHH

Table S12. Enhanced thermal stability of bicyclic trimer designs. Circular dichroism derived T_m values for trimer designs before and after bicyclization, presented in Figure 5.

Protein	T_m / °C		
	Unreacted	Bicyclic	ΔT_m / °C
l4 (3FNJ)	61	80	19
b4 (1VMF)	77	83	6
a4 (3C6V)	41	80	39
e4 (5C9G)	54	71	17

Supplementary References

[1] Drienovská, I., Gajdoš, M., Kindler, A., Takhtehchian, M., Darnhofer, B., Birner-Gruenberger, R., Dörr, M., Bornscheuer, U.T., and Kourist, R. (2020). Folding Assessment of Incorporation of Noncanonical Amino Acids Facilitates Expansion of Functional-Group Diversity for Enzyme Engineering. Chemistry - A European Journal 26, 12338–12342. 10.1002/chem.202002077.

[2] Fürst, M.J.L.J., Martin, C., Lončar, N., and Fraaije, M.W. (2018). Experimental Protocols for Generating Focused Mutant Libraries and Screening for Thermostable Proteins. Methods Enzymol 608, 151–187. 10.1016/BS.MIE.2018.04.007.

[3] Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112, 531–552. 10.1385/1-59259-584-7:531.

[4] Depraz Depland, A., Stroganova, I., Wootton, C.A., and Rijs, A.M. (2023). Developments in Trapped Ion Mobility Mass Spectrometry to Probe the Early Stages of Peptide Aggregation. J Am Soc Mass Spectrom 34, 193–204. 10.1021/jasms.2c00253.

[5] Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R.C., Giglione, C., and Meinnel, T. (2006). The proteomics of N-terminal methionine cleavage. Molecular and Cellular Proteomics 5, 2336–2349. 10.1074/mcp.M600225-MCP200.

[6] Geoghegan, K.F., Dixon, H.B.F., Rosner, P.J., Hoth, L.R., Lanzetti, A.J., Borzilleri, K.A., Marr, E.S., Pezzullo, L.H., Martin, L.B., Lemotte, P.K., et al. (1999). Spontaneous α-N-6-phosphogluconoylation of a “His tag” in Escherichia coli: The cause of extra mass of 258 or 178 Da in fusion proteins. Anal Biochem 267, 169–184. 10.1006/abio.1998.2990.