EXPERIMENTAL SUPPORTING INFORMATION

Furan warheads for covalent trapping of weak protein-protein interactions: cross-linking of thymosin β4 to actin

Laia Miret-Casals*[a], Willem Vannecke[a,b], Kurt Hoogewijs[a], Gianluca Arauz-Garofalo[c], Marina Gay[c], Mireia Díaz-Lobo[c], Marta Vilaseca[c], Christophe Ampe[b], Marleen Van Troys*[b], and Annemieke Madder †[a]

a. Organic and Biomimetic Chemistry Research Group, Department of Organic and Macromolecular Chemistry, Faculty of Science, Ghent University, Krijgslaan 281 S4, B-9000 Ghent, Belgium
b. Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, Technologiepark 75, B-9052 Ghent-Zwijnaarde, Belgium
c. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain
† Annemieke.Madder@UGent.be

Table of Contents

1. MATERIALS AND METHODS

1.1. Materials

All-natural amino acids, coupling reagent 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), and trifluoroacetic acid (TFA) were purchased from Iris Biotech. Trityl alcohol ChemMatrix resin was obtained from Biotage. Fmoc-β-(2-furyl)-Ala-OH was purchased from Peptech. Peptide synthesis grade dimethylformamide (DMF) was purchased from Biosolve. Dichloromethane (DCM), methanol (MeOH), *N*,*N*-Diisopropylethylamine (DIPEA), *N*-methylmorpholine (NMM), and triisopropylsilane (TIS) were obtained from Sigma Aldrich. *N*-bromo succinimide was obtained from Sigma Aldrich. Rabbit alpha-skeletal muscle actin protein (95 or 99% pure) and pyrene-actin which is rabbit skeletal muscle actin labelled with N-(1-pyrene) iodoacetamide on Cys374 (0.6 label/actin labelling stoichiometry) were produced by Cytoskeleton, Inc. and purchased from Tebu-Bio.

1.2. Background on the used published crystal structures of the Tβ4-actin complex (PDB ID 4PL7 and 4PL8)

In this paper, we have used two different previously published crystal structures of the Tβ4-actin complex (PDB ID:4PL7 and 4PL8)[1] to decide on the positions to insert a furan moiety in Tβ4. The 4PL7 structure is a hybrid protein that consists of full-length human Tβ4 fused covalently through a flexible peptide linker to yeast *Pichia* (*Komagataella*) pastoris actin. In this structure Tβ4 makes extensive contacts with the actin protein over Tβ4's entire length involving both an N-terminal α-helix (Asp5 to Lys11) and a C-terminal α-helix (Lys31 to Gln39) (Figure 1B in main text). The 4PL8 structure is a non-covalent complex between rabbit skeletal muscle actin and a hybrid peptide comprising the N-terminal part of human Tβ4 before the LKKTET motif[2] (Ser1-Lys16), the lysine-rich region (Phe1095-Asp1106) of cordon blue, and full-length human Tβ4 (Ser1-Ser43). Cordon blue is a fruit fly protein that contains three WH2-repeats that each are in part similar to Tβ4[3]. In PDB ID 4PL8 only the N-terminal half of the full length Tβ4 in contact with actin was visible and not the more flexible transiently associating C-terminal half. The complex and/or hybrid nature of the co-crystalized proteins is needed for elucidating the structure of the Tβ4-actin complex since a non-covalent actin-thymosin 1:1 complex is not sufficiently stable to obtain crystals. The residues identified in the actin protein as possible nucleophiles to interact with Tβ4-E24 or Tβ4-M6 (respectively Lys18/61/213/215/336 in 4PL7 and Cys374/Lys291/Lys373 in 4PL8) (see §1.3) are conserved between human, *Pichia* and rabbit skeletal muscle actin (see sequence alignment of actins in[4]).

1.3. Analyzing and visualizing the modified 4PL7 and 4PL8 structures of the Tβ4-actin complex (Figure 1)

Schrödinger's Protein Preparation Wizard (PrepWizard) was used to prepare the structures 4PL7 and 4PL8 (downloaded from the Protein Data Bank (PDB), only one of the two identical chains present in the reported 4PL7 and 4PL8 were imported) for modeling and visualization. The PrepWizard adds missing hydrogen atoms, corrects metal ionization states, assigns bond orders, highlights residues with missing atoms, predicts protonation states, flips Asn/Gln and His residues to optimize the H-bond network, removes select crystallographic waters, optimizes the protein´s hydrogen bond network, and minimizes the protein structure. All resolved crystal water molecules were maintained. Maestro Version 12.3.013, MMshare Version 4.9.013, Release 2020-12, Platform Windows-x64 Schrödinger was used to substitute the Tβ4 amino acids at positions 6(Met) and 24(Glu) by 2-

furyl-*L*-alanine. This was followed by an additional optimization using PrepWizard. To optimally visualize the close proximity of WT-Tβ4/Fua-Tβ4 at positions 6 and 24 to nucleophilic amino acids of monomeric actin (less than 13 Å apart for 4PL7 and less than 15 Å apart for 4PL8) the side chains of the amino acids were rotated around selected dihedral angles. The number of possible dihedral angles depends on the residue. For example, lysine can be rotated by changing the various angles around the C_α-C_β bond, C_β-C_γ bond, C_γ–C_δ bond, and C_δ–C_ε bond. For each residue involved in the contact sites of Tβ4-Fua on actin, the rotamer with the nucleophilic group in closest proximity was selected. We note that this still gives a static view based on the crystal structures used. However, it is known for example from NMR experiments[5, 6] that actin-thymosin is a very dynamic complex.

1.4. Thymosin β4 Synthesis

Wild-type human Tβ4 and Fua-Tβ4 protein analogues with furyl-alanine at position 6, position 6 and 24 or position 24 (combined with either lysine or arginine at position 25) were synthesised in parallel on a MultiPep RSi (Intavis) automated peptide synthesiser. If a furylalanine was not incorporated at position 6, a norleucine residue was introduced at this position to avoid Met-oxidation. The synthesis protocol used is described here. Trityl alcohol ChemMatrix was first chlorinated. 100 mg resin was swelled in 1.5 mL of dry DCM. 30 µL of thionyl chloride ($SOCl_2$) were added to a total of 2% in dry DCM. The vessel was shaken overnight. The resin was subsequently drained and rinsed with DCM and 2% NMM in DCM. A solution of the Fmoc-L-serine(tBu)-OH (3 eq) with 4 eq of NMM in DCM was added to the resin, and the mixture was shaken for 6 hours. The resin was drained and rinsed with DCM. A solution of 25% NMM in MeOH was added and the resin was shaken for 1 h. The resin was drained and rinsed with DCM, DMF, MeOH and diethylether and subsequently dried *in vacuo*. Each Tβ4 protein analogue was subsequently prepared on a 10 µmol scale. Usually, synthesis with double coupling steps was performed as following: a mixture of 5 eq. amino acid in DMF (0.5 M), 5 equivalents HBTU in DMF (0.5 M) and 10 eq. DIPEA in NMP (2 M) was added to the resin, with subsequent reaction for 40 minutes at room temperature. The Fmoc-3-(2-furyl)-*L*-alanine was incorporated as a standard amino acid. A capping step using 10% NMM and 10% acetic anhydride in DMF was performed after each double coupling. The first 9 C-terminal amino acids (positions 43 - up to Glu35) to be incorporated were first coupled manually to exclude diketopiperazine formation at the C-terminus. To prevent degradation of the reagents, the peptide was further synthesised using a peptide synthesize in two steps (Ile34 → Lys18 and Lys17 → Ser1). The N-terminus was acetylated after Fmoc-deprotection. Cleavage was performed during 2h at room temperature using the following cleavage cocktail: 95% TFA, 2.5% TIS, and 2.5% H_2O. After reaction, the resin was removed by filtering and the majority of the cleavage cocktail was removed by nitrogen evaporation. Cold Et_2O was added in excess to further precipitate the peptide, followed by sonication and centrifugation (5', 7500 rpm at 2°C). The supernatant was discarded and a fresh volume of MTBE or Et_2O was added to repeat sonication and centrifugation. This process was repeated three times. The residual Tβ4 protein or analogue was dried with a gentle stream of nitrogen and then dissolved in H_2O/ACN to be analyzed by reversed phase high-performance liquid chromatography (RP-HPLC). The Tβ4 protein analogues were subsequently purified using prep-HPLC (0-60% ACN in 20 minutes), see §1.7.

1.5. Reversed Phase High-Performance Liquid Chromatography

RP-HPLC analysis was performed on an Agilent 1100 Series HPLC instrument equipped with a Phenomenex Luna C18(2) column (250 mm x 4.6 mm, 5 µM) at 35°C. UV detection was at 214 nm and the system was run at a flow rate of 1.0 mL/min. The column was eluted with a gradient, starting with 100% H_2O containing 0.1% TFA up to 100% acetonitrile for 20 minutes.

1.6. MALDI-TOF Mass Spectrometry

MALDI-TOF analysis of Tβ4 protein, actin protein, and Fua-Tβ4-actin cross-linked-complex was performed using an Applied Biosystems – 4800 Plus MALDI TOF/TOF™ Analyzer, with α-cyano-4-hydroxycinnamic acid as matrix.

1.7. Semi-Preparative RP-HPLC purification

The purification of thymosin β4 protein or its analogues was performed in a semi-preparative RP-HPLC (Agilent 1100 Series) using a Phenomenex Luna C18(2) column (250 mm x 10 mm, 5 µM) at 35°C. UV detection was at 214 nm and the system was run at a flow rate of 4.5 mL/min. The column was eluted with a linear gradient from 100% H_2O containing 0.1% TFA up to 60% acetonitrile for 20 minutes. The fractions were collected.

1.8. Freeze drying

Tβ4 protein fractions were lyophilized in a Heto Drywinner freeze dryer in combination with a Thermoelectron corporation Savat SPD111V Speedvac concentrator or in a RVC 2-18 CDplus (Christ) in combination with an Alpha 2-4 LDplus (Christ) lyophilisator.

1.9. Liquid Chromatography-Electrospray ionization-Mass Spectrometry (LC-ESI-MS)

LC-ESI-MS analysis of Tβ4 protein analogues was performed on an Agilent 1100 Series HPLC with DAD detector coupled to an Agilent G1946C single quadrupole mass detector equipped with ESI ionization source. The column used is a Phenomenex Kinetex C18 column (150 mm x 4.6 mm, 5 µm) at a flow rate of 1.5 mL/min at 35°C. UV detection was at 214 nm. The column was eluted with a gradient, starting with 100% H_2O containing 5 mM NH_4OAC up to 100% acetonitrile.

1.10. Protein concentration determination and storage.

The lyophilized Tβ4 protein was reconstituted in 10 mM $NaHCO_3$ in H_2O and set to neutral pH. The Tβ4 stock concentrations were determined using a Varian Cary 3E UV-VIS spectrophotometer at 205 nm. All Tβ4 stock solutions approximated ~ 1 mM and were stored at -20 or -80°C.

Rabbit skeletal muscle actin (1 mg, monomeric, G-actin) and pyrene-actin (1 mg, monomeric, G-actin) were reconstituted in aqueous solution as recommended by the manufacturer so that the proteins were in 5 mM Tris-HCl pH 8, 0.2 mM $CaCl_2$, 0.2 mM ATP, 5 % (w/v) sucrose and 1 % dextran. These solutions were aliquoted, snap frozen in liquid nitrogen and stored at -70 °C. Absorbance of the actin solutions was measured at 290 nM on a Nanodrop ND-1000 spectrophotometer using the ATP-containing buffer as blank and the concentration calculated using as molar extinction coefficient $\varepsilon = 2.591 \times 10^4\ M^{-1}\ cm^{-1}$.

1.11. SDS-PAGE analysis and detection

For separation of actin and the Tβ4-actin cross-linked complex based on their difference in molecular weight, denaturing gel electrophoresis in a Mini Protean II gel electrophoresis apparatus

(Bio-Rad) was used. SDS-PAGE gels were prepared according to a standard protocol with a 10% separation and 5% stacking gel. Gels were run at constant current (25 mA for one gel) for circa 1.5 h. Bromophenol blue present in the Laemmli sample buffer was used as tracking dye. After running the gel, the gel was stained using a standard Coomassie Blue R-250 solution for 1-2 h and destained in an ethanol/acetic acid solution for at least 2h. Gels were scanned using an Odyssey Infrared Imaging System (LI-COR Biosciences) and the result analyzed using Image Studio software (LI-COR Biosciences).

1.12. Cross-linking furan-containing Tβ4 to monomeric actin (Figure 3)

The concentrated monomeric G-actin protein was 25-fold diluted in a final buffer containing 5 mM Tris-HCl pH 7.7, 0.2 mM $CaCl_2$, 0.2 mM ATP and stored on ice. For cross-linking, a total volume of 30 µL was usually used. Tβ4 protein (wild type or Fua analogues) was added to the final buffer containing G-actin (final G-actin concentration 12 µM). The used concentration of Tβ4 protein (wild type or Fua analogues) is indicated in figures or figure legends. After incubation of G-actin with Tβ4 proteins at room temperature for 20' (binding step), the photosensitizers Rose Bengal (RB) or Rhodamine B (Rhd B) were added to a final concentration of 5 or 10 µM, respectively (or as indicated in figures). The samples were irradiated with a 110 WATT (12 VOLT) halogen lamp coupled with an optical fiber arm at room temperature during different times (see figures). This produces singlet oxygen that activates the furan moiety and allows the cross-linking of actin and Tβ4. After irradiation, Laemmli sample buffer was added, and the samples were boiled for 5 minutes prior to SDS-PAGE gel analysis. No background cross-linked complex is formed when Fua-Tβ4 analogue is added to actin in the absence of photosensitizer and irradiation (not shown).

Furan is alternatively activated through chemical oxidation by adding N-bromosuccinimide (NBS) at different concentration (see Sup. Fig. 4) and incubating for 2 h at room temperature.

Cross-linking of actin and thymosin β4 was also achieved using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and *N*-hydroxysulfosuccinimide (Sulfo-NHS) as described in Van Troys et al[2].

1.13. Characterization of the cross-linking sites between furan-containing Tβ4 and monomeric actin: crosslinking, sequential protein digestion and mass spectrometry and downstream data analysis (Figure 4 and Sup. Fig. 7)

The cross-link reaction was performed as described in §1.12 using 12 µM G-actin, 48 µM E24FuaTβ4 and 5 µM Rose Bengal as photosensitizer. After cross-linking, the proteins were alkylated for 30 min in the dark with 30 mM IAM (iodoacetamide) with previous treatment with 15 mM TCEP for 15 min at 30 °C. After SDS-PAGE, the band corresponding to the E24FuaTβ4-actin cross-link was excised from the Coomassie-stained gel and the gel pieces were submitted to in-gel tryptic digestion and subsequently to in solution digestion by AspN or chymotrypsin. This double digest was performed to selective shorten long tryptic (cross-linked) peptides as these may pose a potential problem regarding their detection[7]. In more detail, the excised gel bands were washed with 50 mM NH_4HCO_3 and acetonitrile. Then, in-gel digestion was performed with trypsin (0.1 µg/µL) (Sequencing Grade Modified Trypsin, Promega Cat# V511) in 50 mM NH_4HCO_3 at 37°C overnight. The digestion was stopped with 50 µL 5% formic acid (FA) and the tryptic peptides were extracted with 50 µL acetonitrile followed by 50 µL 5 % FA aqueous solution. The sample volumes were reduced in a SpeedVac to 3 µL after which 57 µL of 50 mM NH_4HCO_3 was added. Samples were additionally

digested in solution with AspN (0.1 µg) (Sequencing Grade Asp-N, Promega Cat# V1621) or chymotrypsin (0.1 µg) (Sequencing Grade Chymotrypsin, Promega Cat# V1062) at 37°C overnight.

Peptides resulting from the digestion procedures were diluted in 1% FA, 3% ACN in H_2O. Samples were loaded to a 100 µm × 2 cm Acclaim PepMap100, 5 µm, 100 Å, C18 (Thermo Scientific) at a flow rate of 15 µL/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column (Acclaim PepMap® RSLC (50 µm × 50 cm, nanoViper, C18, 2 µm, 100Å) (Thermo Scientific)) in a 90 min run, comprising three consecutive steps with linear gradients from 1 to 35% B in 63 min, from 35 to 50% B in 5 min, and from 50% to 85% B in 2 min. This was followed by isocratic elution at 85% B for 5 min and stabilization to initial conditions (A= 0.1% FA in water, B= 0.1% FA in CH3CN). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. MS1 scans were acquired in the Orbitrap at 120 k resolution (defined at 200 m/z). For MS2, we set the top speed fragmentation with a cycle time of 3 s. We used a 28% **H**igher-energy **C**ollisional **D**issociation (HCD) collision energy and detection in the orbitrap at 30 k. The ion count target value was 400,000 for the survey MS1 scans and 10,000 for the MS2. Target ions already selected for MS/MS were dynamically excluded for 15 s. Spray voltage in the NanoMate source was set to 1.70 kV. The radio frequency lens was tuned to 30%. The spectrometer was working in positive polarity mode and single charge state precursors were rejected for fragmentation.

Analysis of the MS-data from the cross-linked sample and searches for cross-linked peptides were performed with xiSEARCH v1.7.6.1[7]. This software allows the user to define any type of cross-linker, modification, digestion, and fragmentation method. This search had to be adapted to our system because we are not using a conventional cross-linker nor a standard amino acid. In the Tβ4-actin cross-linked complex, Tβ4 contains a furan moiety, which was incorporated using 2-furyl-*L*-alanine (Fua), in position 24 instead of a glutamic acid. We predicted that the Tβ4-Fua24 moiety after oxidation can react with proximal actin nucleophilic side chains, such as lysine, linking both proteins (Note that Lys residues are the only nucleophiles that are in close proximity to Tβ4-Fua24 (<13Å, Figure 1B, 4PI7). The expected nature of the chemical structure covalently linking the oxidized furan moiety of E24FuaTβ4 and Lys residue of actin is shown in Sup. Fig. 6. panel A.

To identify the crosslinked peptides in the mass data using xiSEARCH software we needed to deal with the fact that the xiSEARCH software does not allow to include non-natural amino acids such as Fua. We consequently artificially (and for data search proteins only) approached the mass of the Fua-Lys crosslink as shown in Sup. Fig. 6 panel B. The structure is split in three parts: an alanine part of Fua (in red), the anticipated covalent link (in blue) and the lysine part (in green from actin). Suppl. Figure 6B also shows the composition and mass that was searched for as extra mass in addition to the alanine-part and lysine-part. This can be different depending on whether the hydrated form of the cross-link was considered: $C_4H_2O_2$ (82.00548 Da) or the dehydrated form C_4O (63.99491 Da) (see Sup. Fig. 6, panel B). In xiSEARCH v1.7.6.1, this was implemented as follows. Only the molecular formula is needed to search for the crosslinked species. Since, as explained above, we assumed in the analysis that in the crosslinked complex E24FuaTβ4 is seen as having Ala + $C_4H_2O_2$ or Ala + C_4O at position 24, we modified the human fasta data accordingly, so that it contained P62328 (UniProt ID of human Thymosin β4) with an alanine at position 24 instead of the normal glutamic acid. We also expanded the fasta data to contain protein P68133 (rabbit actin, alpha skeletal

muscle). The structures $C_4H_2O_2$ (82.00548 Da) and the dehydrated form C_4O (63.99491 Da) defined above as indicating a crosslink (see Sup. Fig. 6, panel B) were added as modification in xiSEARCH. The additional xiSEARCH parameters were: Trypsin/AspN (T+A) or Trypsin/Chymotrypsin (T+C) as digesting enzymes allowing three missed cleavage sites and a minimum peptide length of four residues. We also included carbamidomethyl in C as static modification. As dynamic modifications, we included acetylation at the N-terminus as well as oxidation of M, W and Y, double oxidation of W and modification of W to kynurenine since singlet oxygen produced by Rose Bengal photosensitizer was used to activate the furan moiety. 10 ppm and 20 ppm were set as MS1 and MS2 mass tolerances, respectively. We deposited all MS-related files (*.raw, *.mzid) to the ProteomeXchange Consortium via the PRIDE[8] partner repository with the dataset identifier PXD025066 and 10.6019/PXD025066.

We launched 12 crosslinking searches arising from x2 search nodes (MeroX and xiSEARCH) x2 crosslinkers (C4H2O2 and C4O) and x3 type of digestion (Trypsin, Trypsin + Asp-N and Trypsin + Chymotrypsin). We developed a custom Python script called pyXL (commit 163ed1c) in order to unify these 12 search results in just two harmonized datasets, one for each crosslinker. Detected crosslinked peptides for each dataset where further filtered considering only those peptides with scores higher than those of the decoy to avoid false positive hits. 173 cross-linked peptides (with 300 cross-linked peptide spectrum matches (CSMs) i.e. a match to cross-linked products of a peptide pair consistent with a given precursor mass of the cross-linked peptide) were identified in the database search. Only cross-linked peptides that were heterodimers between thymosin and actin and that had the assigned Ala24 of thymosin β4 in their sequences, were considered as plausible hits. The furan moiety in E24FuaTβ4 is a chemical modification that was introduced using the commercially available 2-furyl-*L*-alanine amino acid (Fua) during solid phase peptide synthesis, thus, all the identified cross-linked peptides that did not contain E24Fua, defined as Ala24 + the modification $C_4H_2O_2$ or C_4O in our search, were false positives. This resulted in 17 cross-linked peptides and 26 cross-linked peptide spectrum matches (CSMs). All of them were identified using $C_4H_2O_2$ as cross-linker structure. No plausible cross-linked peptides between E24FuaTβ4 and actin were detected when using the dehydrated form C_4O as a cross-linker. Among this selection of 17 cross-linked peptides (not shown), only three cross-linked peptides passed a manual inspection. These three final hits were detected with seven cross-linked peptide spectrum matches (CSMs) as shown in Table 1. The seven CSM that passed the manual inspection all pointed to the same cross-linked site between E24FuaTβ4 and K61 of actin (see Table 1). We thus identified the same cross-linked residues in the two differently digested samples (Trypsin combined with either AspN or chymotrypsin), with three different Tβ4 peptide lengths and one of them with multiple charge states. The three identified cross-linked peptides contained the same tryptic actin peptide and thymosin tryptic peptides with different missed or partial cleavages. The three most representative identified CSMs are documented in more detail in Sup. Fig. 7. They showed a similar fragmentation pattern regarding the y-ions in the actin peptide, which included the thymosin β4 peptide in all cases.

Cross-linker	Digestion	m/z	z	M_{exp} (Da)	M_{theo} (Da)	Δm (ppm)	Score[1]	Sequence E24FuaTβ4[2]	Sequence Actin	Site Tβ4	Site Actin
$C_4H_2O_2$	T+A[2]	720.111	4	2877.421	2877.422	-0.34753	7.446	19-KTETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**
$C_4H_2O_2$	T+A	576.29	5	2877.421	2877.422	-0.34753	6.983	19-KTETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**

$C_4H_2O_2$	T+A	720.112	4	2877.425	2877.422	1.042599	6.28	19-KTETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**
$C_4H_2O_2$	T+C[3]	917.114	3	2749.326	2749.327	-0.36373	11.028	20-TETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**
$C_4H_2O_2$	T+C	720.112	4	2877.425	2877.422	1.042599	8.068	20-KTETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**
$C_4H_2O_2$	T+C	561.021	4	2241.062	2241.063	-0.44622	4.551	19-KTETQ**A***K-25	51-DSYVGDEAQS**K**R-62	**24**	**K-61**
$C_4H_2O_2$	T+C	959.813	3	2877.425	2877.422	1.042599	3.821	19-KTETQ**A***KNPLPSK-31	51-DSYVGDEAQS**K**R-62	**24**	**K-61**

Table 1: CSMs detected by the software xiSEARCH in both trypsin/AspN and trypsin/chymotrypsin digestions for cross-link between position 24 in Tβ4 of E24FuaTβ4 and K61 of actin.

[1] See text above the table. [2] A*: Hypothetical Ala residue at position 24 in Tβ4, refers to only the alanine part of the Fua mass, with the rest of the activated Fua mass incorporated in the mass of the cross-linker $C_4H_2O_2$ (82.00548 Da) (see Sup. Fig. 6 and ESI §1.13). In bold, cross-linked amino acids with crosslinker 1. [3] T+A: trypsin + AspN; T+C: trypsin + Chymotrypsin

1.14. Sequestering assay; determination of K_D for actin-Tβ4 complexes (Figure 2 and Sup. Fig. 1)

In this assay the measurement of fluorescence correlates with the F-actin concentration as described in Kouyama & Mihashi[9]. The actin used is composed of 80% unlabeled actin and 20% pyrene labeled actin. Pyrene labelled actin (PA) only displays a high fluorescence when it is present in the polymerized F-actin form with $\lambda_{Ex\ max}$ 360nm and $\lambda_{Em\ max}$ = 388nm. Pyrene fluorescence and F-actin concentration are linearly correlated. Fluorescence was measured on a SFM 25 fluorometer (Kontron Instruments, Zurich).

To measure the sequestering capacity of Tβ4 wild type and Tβ4-caged variants (see Sup. Fig. 1B), samples containing 1 µM F-actin (20% pyrene labelled) and increasing amounts of Tβ4 analogues (Wild type or Fua-Tβ4) in actin polymerizing buffer (5 mM Tris-HCl pH 7.7, 0.2 mM $CaCl_2$, 0.2 mM ATP, 0.1 mM DTT supplemented with 100 mM KCl and 1 mM $MgCl_2$) were incubated overnight at room temperature and allowed to reach equilibrium before measuring. By sequestering G-actin monomers, Tβ4 reduces the equilibrium amount of F-actin and consequently the pyrene fluorescence signal at equilibrium. From the measured fluorescence, the F-actin level is calculated and plotted versus the total thymosin concentration.

As detailed in [2, 9] the observed decrease in fluorescence is a measure for the concentration of actin-Tβ4 complex at equilibrium and the slope of these plots allow calculating the affinity using the CMC and the slope as follows:

The equilibrium dissociation constant of the actin-Tβ4 complex

$$K_D = \frac{[A]_{eq} * [T]_{eq}}{[TA]_{eq}} = \frac{[A]_{eq}([T]_0 - [TA]_{eq})}{[TA]_{eq}} \quad (1)$$

with $[A]_{eq}$, $[T]_{eq}$, $[TA]_{eq}$ the equilibrium concentrations of monomeric G-actin, Tβ4 and G-actin-Tβ4 complex. $[T]_0$ is the concentration of Tβ4 initially added. Equation (1) can be written as

$$[TA]_{eq} = \frac{[A]_{eq}}{K_D + [A]_{eq}} [T]_0 \quad (2)$$

with $[A]_{eq}$, the concentration of free monomeric actin, identical to the critical monomer concentration (CMC) for actin polymerization. Following the procedure of Pantaloni and Carlier[10] we determined, in a separate experiment that is presented in Suppl Fig 1A , this CMC under the experimental conditions to be 0.23 µM. This experiment also gives the linear correlation between measured pyrene fluorescence and the concentration of F-actin expressed as:

Fluorescence = 101 [F-actin] + 0.07 (3)

Furthermore, the total amount of actin ([A]t) in the absence [eq.(4a)] or presence [eq.(4b)] of Tβ4 can be written as:

$[A]_t = [FA]_{eq} + 0.23\ \mu M$ (4a)

$[A]_t = [FA]_{eq.T} + 0.23\ \mu M + [TA]_{eq}$ (4b)

with $[FA]_{eq}$ and $[FA]_{eq.T}$ representing equilibrium concentrations of polymerized actin in the absence or presence of Tβ4, respectively. The total actin concentration $[A]_t$ (1μM) is constant, so:

$[FA]_{eq} = [FA]_{eq.T} + [TA]_{eq}$ (5)

By introducing equation (2) in equation (5) we get:

$$[FA]_{eq.T} = [FA]_{eq} - \frac{0.23\mu M}{K_D + 0.23\mu M}[T]_0 \quad (6)$$

Thus, from equation (6), the K_D of the complex TA can be derived.

1.15. Actin polymerization assay (Sedimentation assay, Sup. Fig. 8 and 9)

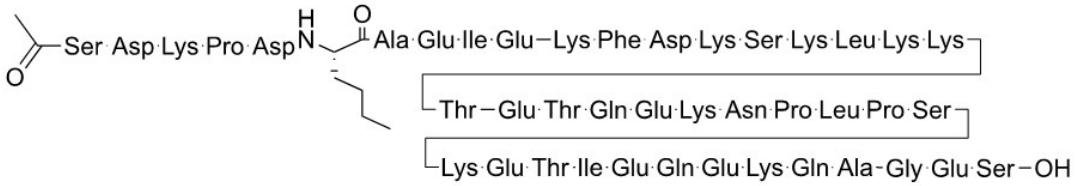
In vitro, actin can be maintained in its monomeric form (G-actin) in low salt buffer (5 mM Tris-HCl pH 7.7, 0.2 mM $CaCl_2$, 0.2 mM ATP). Its polymerization can be induced by adding KCl to 100 mM and $MgCl_2$ to 1 mM (below referred to as 'salts'). Monomeric rabbit alpha-skeletal muscle actin (G-actin, 12 μM) was left untreated or was treated with Rhd B (10 μM) or with RB (5 μM) and was irradiated with visible light for 2 h or 5 minutes, respectively. Subsequently, salts were added to induce the polymerization. The samples were left overnight at room temperature and equilibrium was reached. Then, samples were centrifuged in a Beckman Airfuge at 200.000 g for 15 minutes at room temperature. The supernatant (containing monomeric G-actin) was collected and the pellet (polymerized actin, F-actin) was resuspended in the original volume in the buffer indicated above supplemented with salts. Equal amount of supernatant and pellet were analyzed on SDS-PAGE. This experiment was performed in presence and absence of 12 or 24μM of Met6/Fua24 Tβ4 (added at start to G-actin).

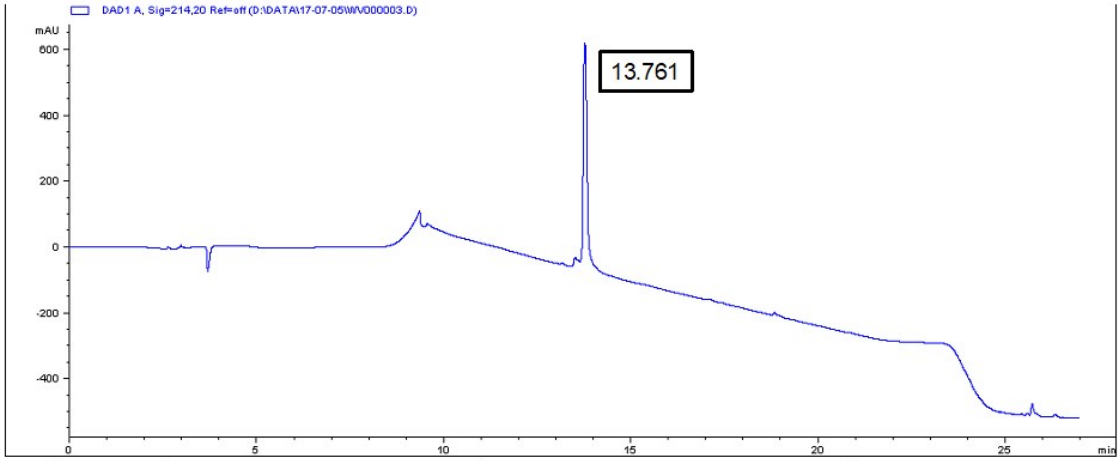
2. THYMOSIN ANALOGUE CHARACTERIZATION

2.1. Wild-type Tβ4 protein

The human thymosin β4 sequence (NCBI resSeq database protein ID: NP_066932 or UniProt code: Human: UniProtKB - P62328 (TYB4_HUMAN)) was used as basis. The initiator Met is removed and replaced by an acetyl group on Ser as in the natural product. In the chemical synthesis, we replaced Met6 with norleucine (unless Fua is at this position). Met6 is prone to oxidation to the sulfoxide reducing its affinity for actin by 20-fold[11]. The norleucine variant has a similar affinity as reported for WT Tβ4 (see Fig 2, main text).

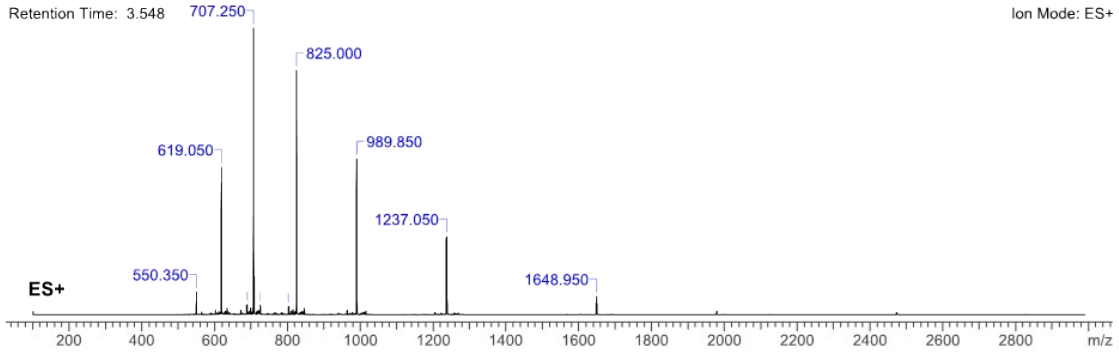
Wild-typeTβ4





Column: Phenomenex Luna C18 (5 µm × 4.6 mm × 250 mm)

RP-HPLC, t_r (0% to 100% of ACN over 20 min): 13.761 min

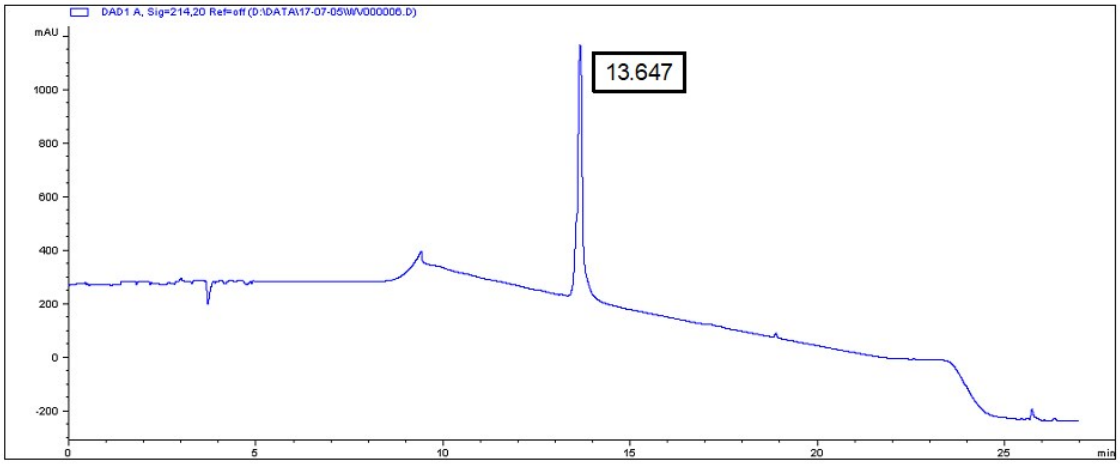


LC-ESI-MS *m/z* at t_r = 3,548: $[M+3H]^{3+}$= 1648,950; $[M+4H]^{4+}$= 1237,050; $[M+5H]^{5+}$= 989,850; $[M+6H]^{6+}$= 825,00; $[M+7H]^{7+}$= 707,250; $[M+8H]^{8+}$= 619,050; $[M+9H]^{9+}$= 550,350.

m/z calculated for $C_{213}H_{352}N_{56}O_{78}$ $[M+H]^+$: 4943,530 Da; found 4943,850 Da.

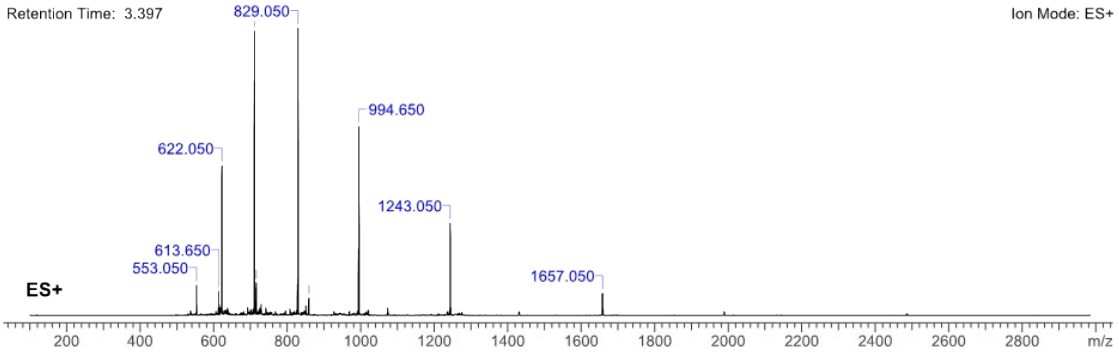
2.2. M6FuaTβ4 protein

M6FuaTβ4



Column: Phenomenex Luna C18 (5 µm × 4.6 mm × 250 mm)

RP-HPLC, t_r (0% to 100% of ACN over 20 min): 13.647 min

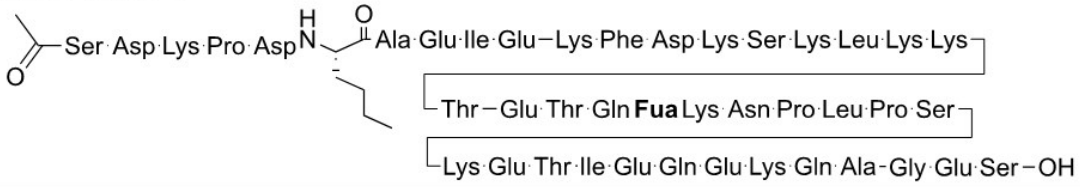


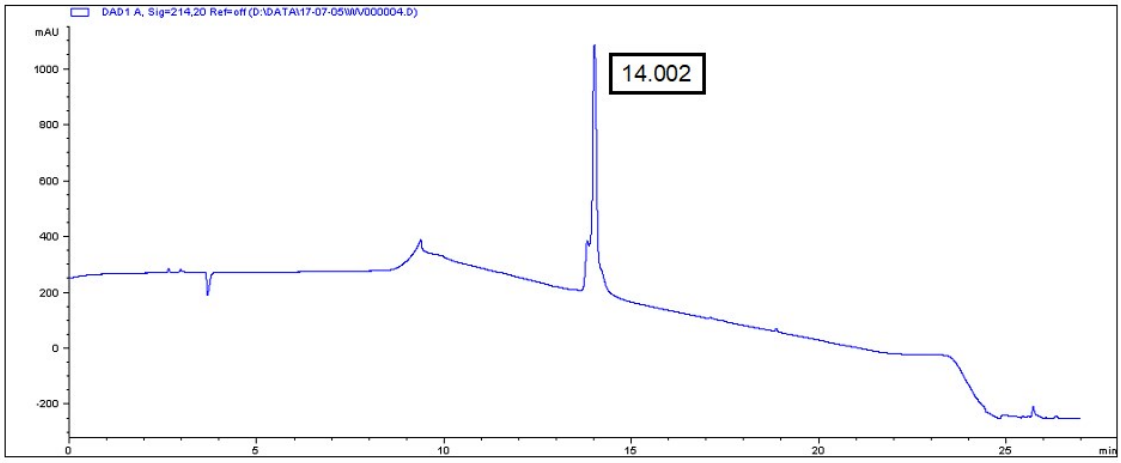
LC-ESI-MS *m/z* at t_r = 3,397: $[M+3H]^{3+}$= 1657,050; $[M+4H]^{4+}$= 1243,050; $[M+5H]^{5+}$= 994,650; $[M+6H]^{6+}$= 829,050; $[M+7H]^{7+}$= 710,400; $[M+8H]^{8+}$= 622,050; $[M+9H]^{9+}$= 553,050.

m/z calculated for $C_{214}H_{348}N_{56}O_{79}$ $[M+H]^+$: 4967,494 Da; found 4968,150 Da.

2.3. E24FuaTβ4 protein

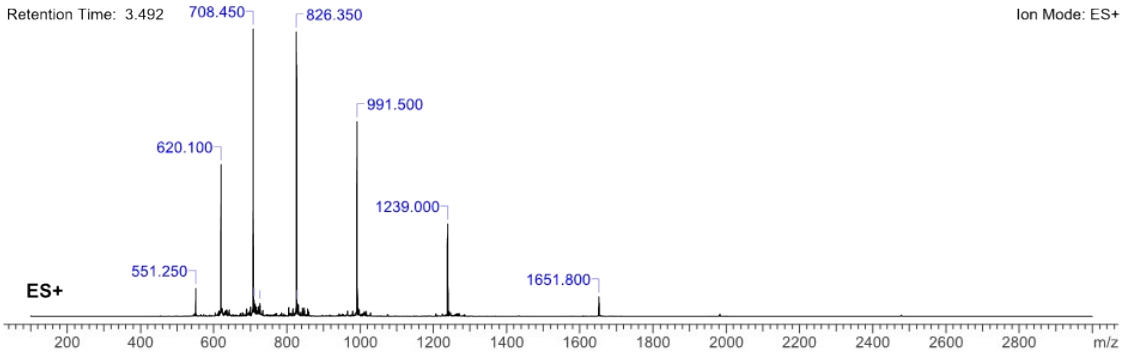
E24FuaTβ4





Column: Phenomenex Luna C18 (5 µm × 4.6 mm × 250 mm)

RP-HPLC, t_r (0% to 100% of ACN over 20 min): 14.002 min

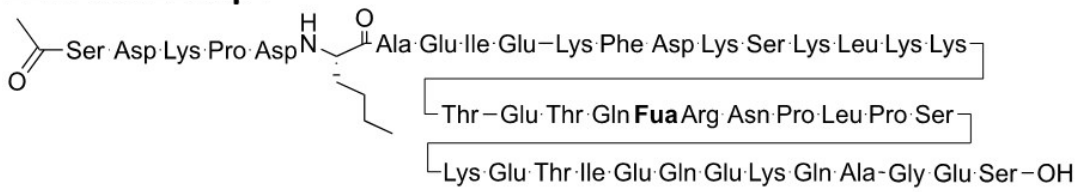


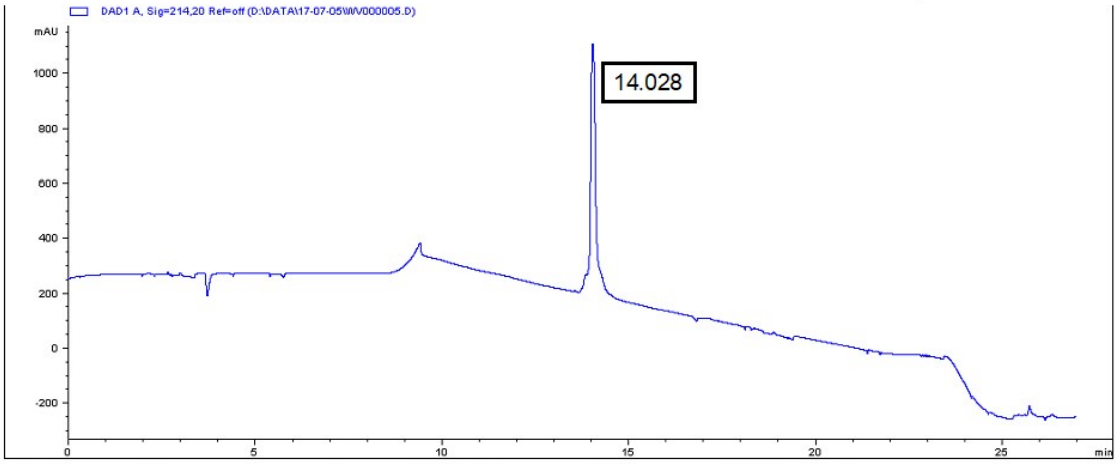
LC-ESI-MS *m/z* at t_r = 3,492: $[M+3H]^{3+}$= 1651,800; $[M+4H]^{4+}$= 1239,000; $[M+5H]^{5+}$= 991,500; $[M+6H]^{6+}$= 826,350; $[M+7H]^{7+}$= 708,450; $[M+8H]^{8+}$= 620,100; $[M+9H]^{9+}$= 551,250.

m/z calculated for $C_{215}H_{352}N_{56}O_{77}$ $[M+H]^+$: 4951,535 Da; found 4952,400 Da.

2.4. E24FuaK25RTβ4 protein

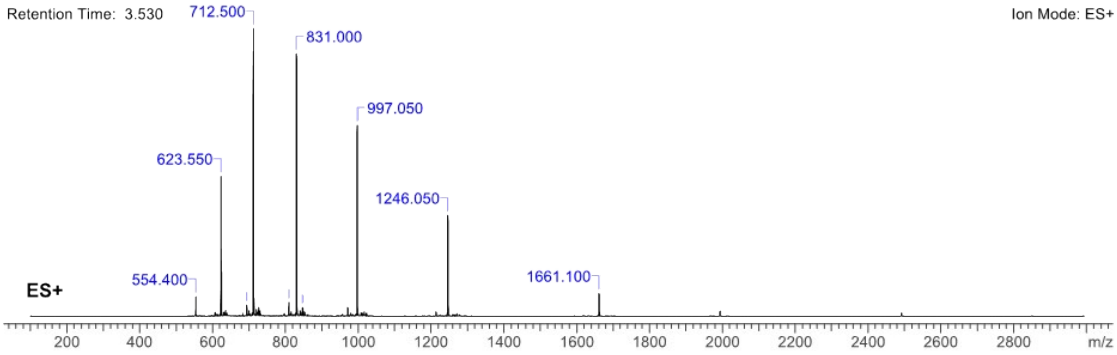
E24FuaK25RTβ4





Column: Phenomenex Luna C18 (5 µm × 4.6 mm × 250 mm)

RP-HPLC, t_r (0% to 100% of ACN over 20 min): 14.028 min

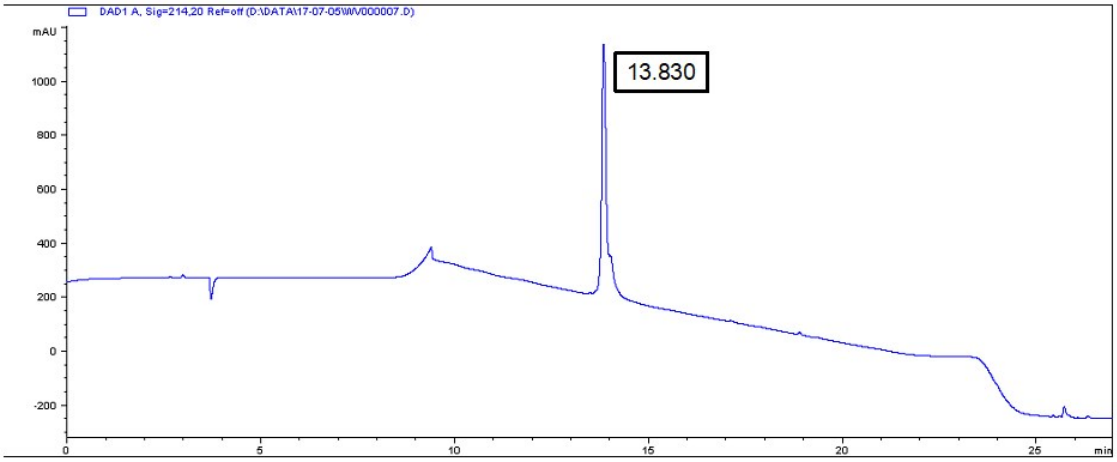


LC-ESI-MS *m/z* at t_r = 3,530: $[M+3H]^{3+}$= 1661,100; $[M+4H]^{4+}$= 1246,050; $[M+5H]^{5+}$= 997,050; $[M+6H]^{6+}$= 831,000; $[M+7H]^{7+}$= 712,500; $[M+8H]^{8+}$= 623,550; $[M+9H]^{9+}$= 554,400.

m/z calculated for $C_{215}H_{352}N_{58}O_{77}$ $[M+H]^+$: 4979,541 Da; found 4980,300 Da.

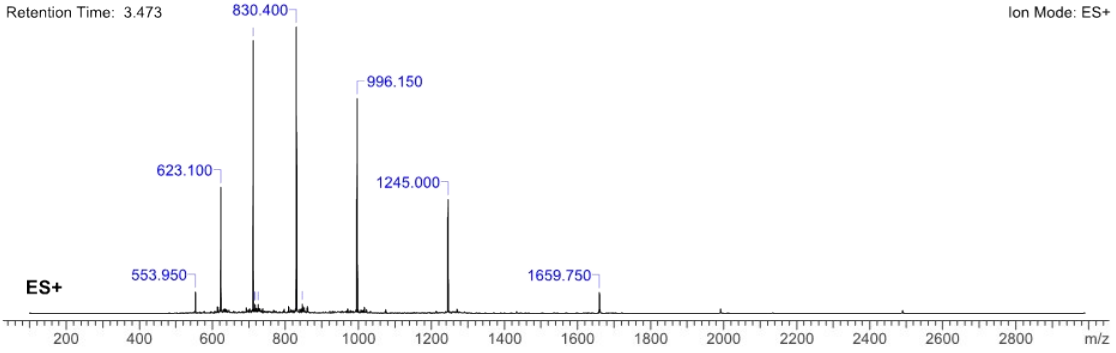
2.5. M6/E24FuaTβ4 protein

M6/E24FuaTβ4



Column: Phenomenex Luna C18 (5 µm × 4.6 mm × 250 mm)

RP-HPLC, t_r (0% to 100% of ACN over 20 min): 13.830 min

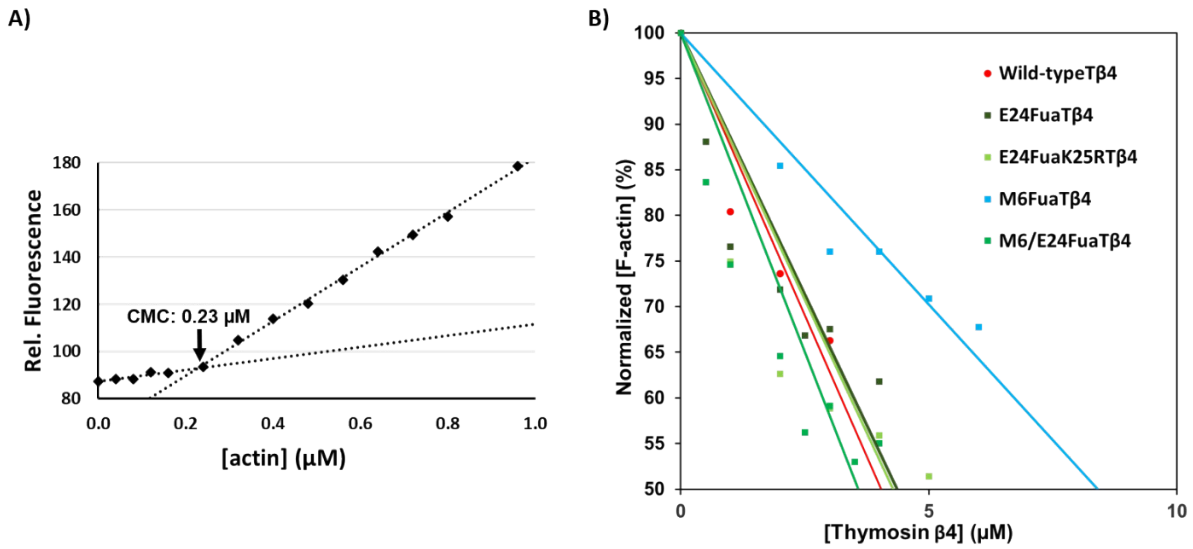


LC-ESI-MS *m/z* at t_r = 3,473: $[M+3H]^{3+}$= 1659,750; $[M+4H]^{4+}$= 1245,000; $[M+5H]^{5+}$= 996,150; $[M+6H]^{6+}$= 830,400; $[M+7H]^{7+}$= 711,900; $[M+8H]^{8+}$= 623,100; $[M+9H]^{9+}$= 553,950.

m/z calculated for $C_{216}H_{348}N_{56}O_{78}$ $[M+H]^+$: 4975,499 Da; found 4976,250 Da.

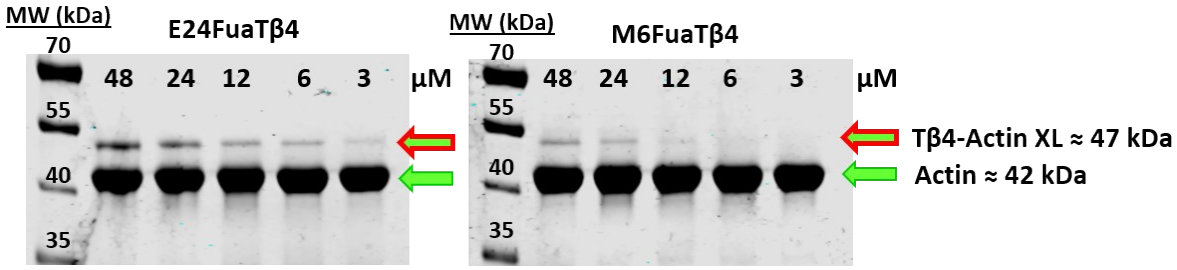
3. SUPPLEMENTARY FIGURES

3.1. Supplementary Figure 1



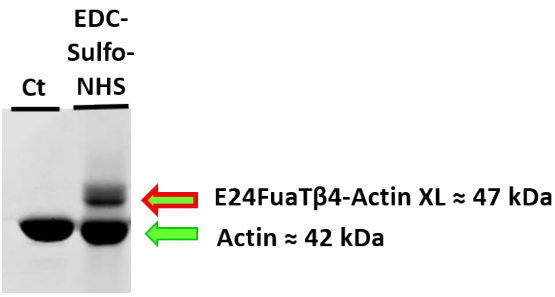
Supplementary Figure 1: A) Relative fluorescence of polymerized actin (F-actin) (20% pyrene-labelled) at equilibrium in function of total actin concentration. The critical monomer concentration (CMC) is derived from the intersection of the two linear fittings shown. The linear fit of the steep curve allows deriving a correlation between the measured fluorescence and the concentration of F-actin expressed as: [F-actin] = [actin] – CMC. This value was used in the calculations in section 1.14. B) Actin sequestering capacity of Fua-modified Tβ4 analogues. F-actin levels at different doses of Tβ4 wild-type or caged fua-modified analogue (for details see ESI 1.14). The dose-dependent reduction in F-actin reflects the G-actin sequestering activity (and thus F-actin depolymerization activity). Note that all Fua-analogues have wild-type activity except for M6FuaTβ4. Linear fits are shown on the data. From the decrease in F-actin (slope of the linear fitting), the K_D values of Tβ4-actin complex formation for WT and caged Fua-Tβ4 variants can be derived. These are presented in Fig 2 (main text).

3.2. Supplementary Figure 2



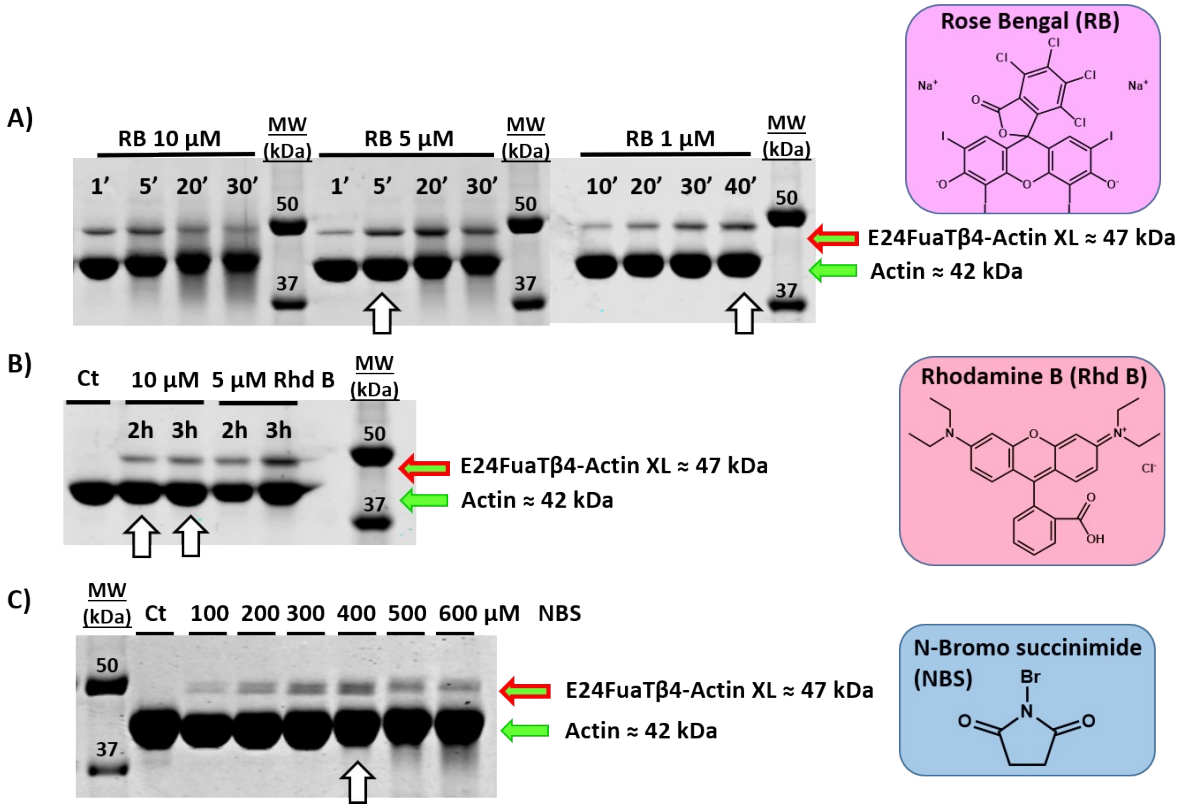
Supplementary Figure 2: Coomassie staining of SDS-PAGE after cross-linking actin (12µM) with E24FuaTβ4 and M6FuaTβ4 protein analogues at the different indicated concentrations (48, 24, 12, 6, and 3 µM) to optimize the dose of Fua-Tβ4 to be used in further experiments. Rose Bengal (10 µM) was used as a photosensitizer to generate singlet oxygen, and the mixture was irradiated for 1 minute with a white light source. First lanes are marker proteins with indicated molecular weight in kDa, XL refers to cross-linked complex.

3.3. Supplementary Figure 3



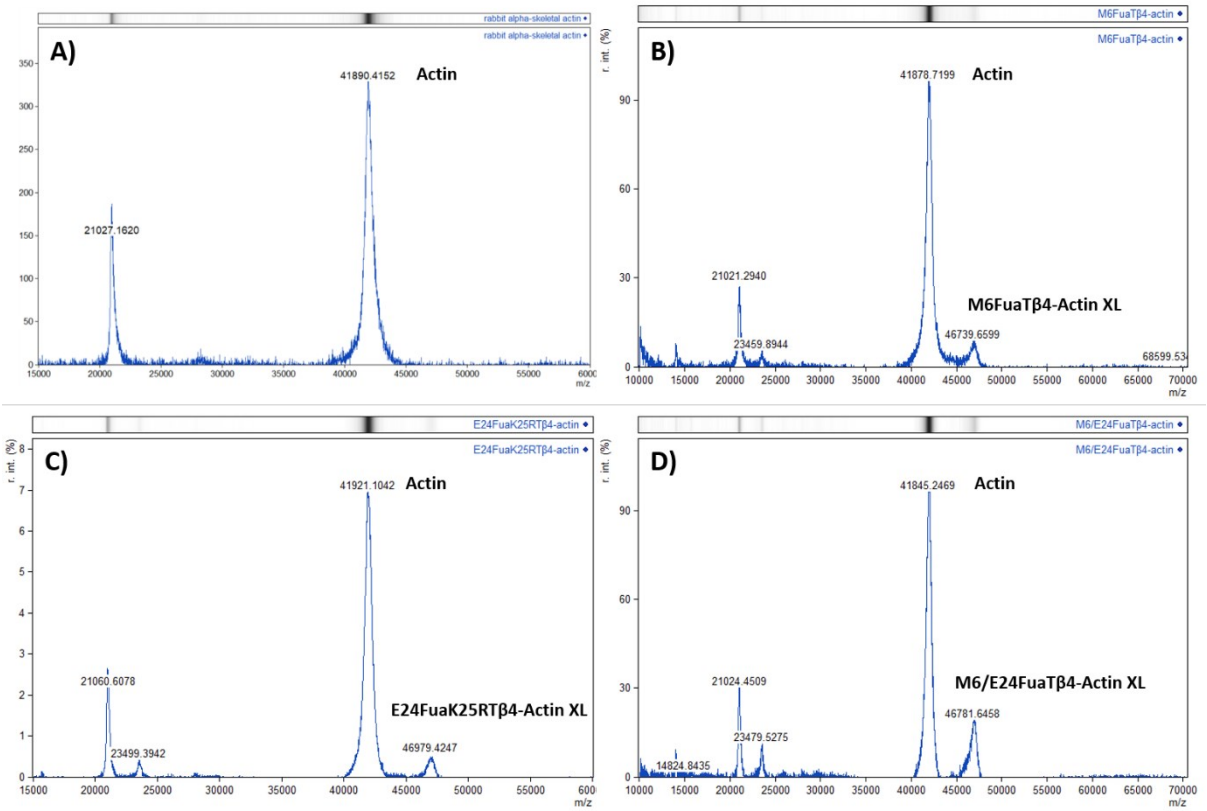
Supplementary Figure 3: SDS-PAGE analysis after cross-linking actin (12 µM) and E24FuaTβ4 (48 µM) using the non-site selective bifunctional cross-linker carbodiimide EDC combined with N-hydroxysulfosuccinimide (Sulfo-NHS) for 45 minutes. In control (Ct) only actin is present. EDC and Sulfo-NHS in combination activate any carboxyl group in a protein or protein complex to form an amine-reactive sulfo-NHS ester bond. The result is a less homogeneous cross-linked species (multiple bands)) and a total level of cross-link product lower than 50%. XL is cross-linked.

3.4. Supplementary Figure 4



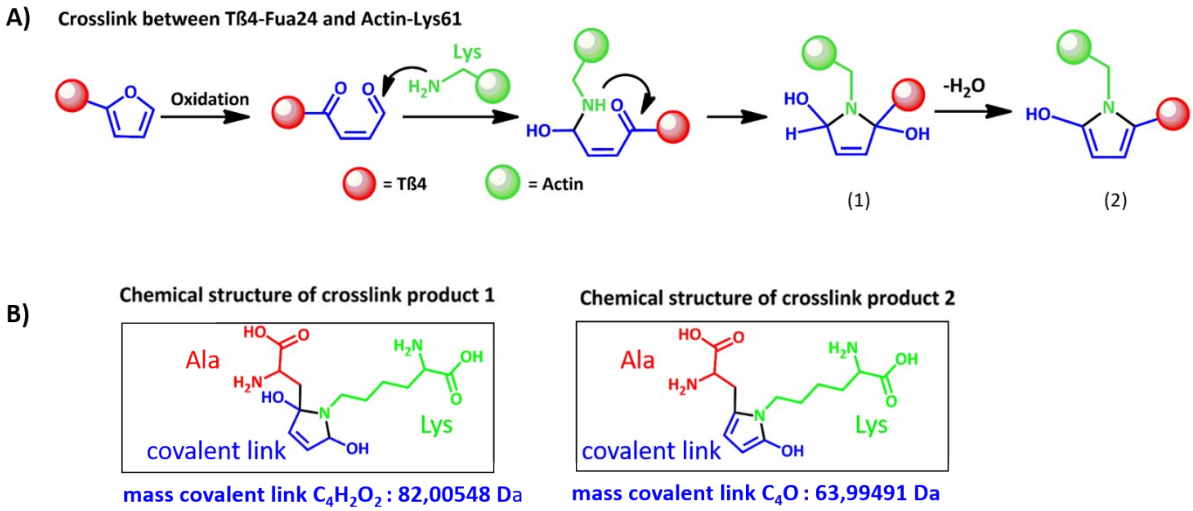
Supplementary Figure 4*:* Optimization of oxidation protocol. The effects of varying photosensitizer (PS) concentrations and irradiation times as well as of chemical oxidation of the furan moiety by the reagent N-bromo-succinimide (NBS) was tested. SDS-PAGE analysis is shown after cross-linking actin (12µM) with E24FuaTβ4 protein (48 µM) using A) Rose Bengal as a photosensitizer at different concentrations (10, 5 and 1 µM) to generate singlet oxygen upon irradiation at different irradiation times (1, 5, 20, 30 and 40 minutes) with a visible light source. b) Rhodamine B as a photosensitizer at different concentrations (10 and 5 µM), irradiation times (2 and 3 hours) with a white light source. c) N-Bromo succinimide (NBS) as an oxidizer at different concentrations (100, 200, 300, 400, 500 and 600 µM) for 2 hours. In control (Ct) only actin is present. Marker proteins indicate molecular weight in kDa, XL is cross-linked. Maximal specific furan cross-linking by E24FuaTβ4 was obtained using 5 µM RB for 5 min, 1 µM RB for 40 min, 10 µM of the more mildly acting Rhd B for 2-3h or with 400 µM NBS for 2 h. This demonstrates that the efficiency of furan cross-linking to proteins is easily tunable by the type and dose of the oxidizing system (NBS or PS-$^1O_2$) or by the PS properties (i.e. dose- and time-dependency of $^1O_2$ production). This is in line with what we previously showed for furan-based DNA- and peptide-cross-linking and other PS[12-15]. Note that too strong exposure to oxidizing conditions using long irradiation times in the presence of a PS or long incubation with high NBS concentration induces smear on the gel suggesting lower specificity of the cross-linking or degradation of the target protein actin.

3.5. Supplementary Figure 5



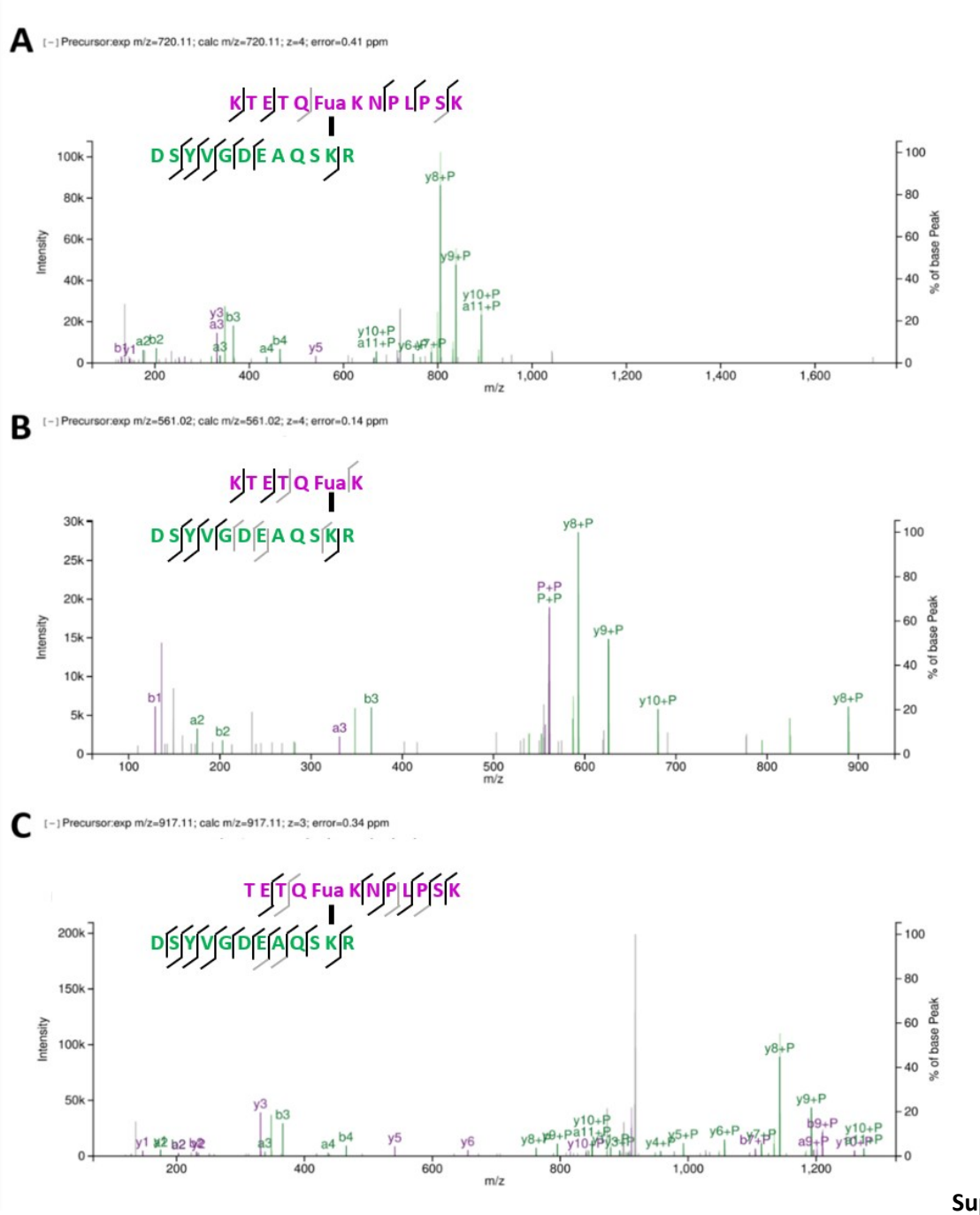
Supplementary Figure 5: MALDI-TOF Mass Spectrometry of monomeric G-actin and Fua-Tβ4-actin covalent complexes after treatment with 5 µM RB and 5 min irradiation. A) monomeric G-actin used as a control. B) G-actin and M6FuaTβ4-actin covalent complex. C) G-actin and E24FuaK25RTβ4-actin covalent complex. D) G-actin and M6/E24FuaTβ4-actin covalent complex. XL refers to cross-linked complex.

3.6. Supplementary Figure 6



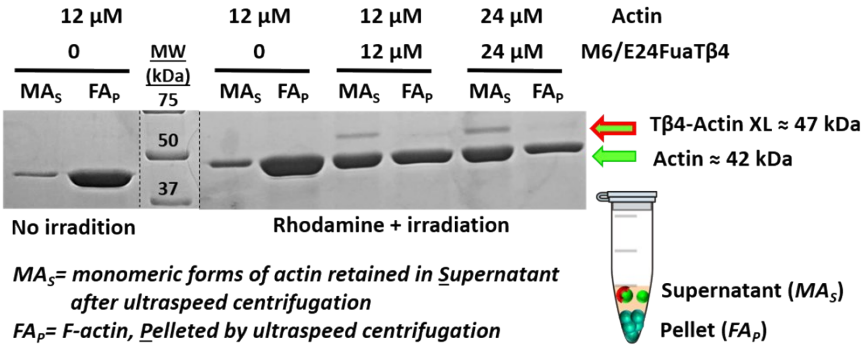
Supplementary Figure 6: A) Chemical structure of the plausible covalent bond that can be formed between E24FuaTβ4 and a lysine residue of monomeric actin (1) and the dehydrated form (2). B) The corresponding chemical formulas of the covalent links are $C_4H_2O_2$ (82.00548 Da) and, for the dehydrated form, C_4O (63.99491 Da), taking into account the atoms of the furan moiety involved (blue) and excluding the atoms of the remaining alanine part of the Fua residue (red) from E24FuaTβ4 and the atoms of the lysine residues from actin (green).

3.7. Supplementary Figure 7



Supplementary Figure 7: Identification of cross-linked peptides with a cross-link between Fua 24 in Thymosin β4 and Lys 61 in Actin. Precursor ions were fragmented by HCD in a Fusion Lumos Orbitrap. P denotes one of the peptide chains. See also Table 1 in ESI and ESI 1.13. A) Representative annotated MS/MS spectrum of the cross-linked (19-KTETQ**Fua**KNPLPSK-31) – (51-DSYVGDEAQS**K**R-62) peptides with the Fua –K61 link= Ala + $C_4H_2O_2$ (82.00548 Da), selecting +4 charge (m/z 720.11). B) Annotated MS/MS spectrum of cross-linked (19-KTETQ**Fua**K25) – (51-DSYVGDEAQS**K**R-62) peptides with the Fua –K61 link= Ala + $C_4H_2O_2$ (82.00548 Da) (m/z 561.02, z=4). C) Annotated MS/MS spectrum of the cross-linked (20-TETQ**Fua**KNPLPSK-31) – (51-DSYVGDEAQS**K**R-62) peptides with the Fua –K61 link= Ala + $C_4H_2O_2$ (82.00548 Da) (m/z 917.11, z=3). Pictures have been generated by xiVIEW[16].

3.8. Supplementary Figure 8

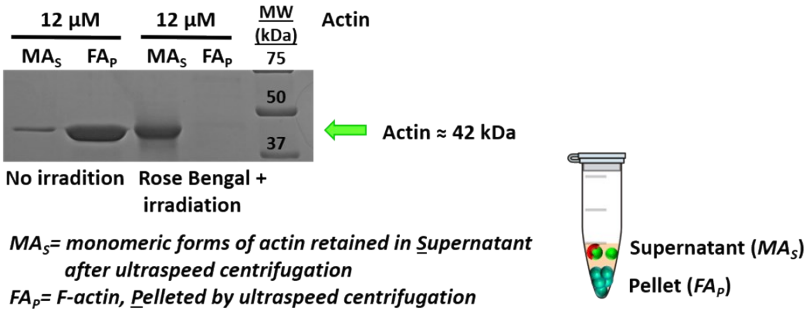


Supplementary Figure 8: Actin sedimentation assay. SDS-PAGE analysis of supernatant and pellet after the ultracentrifugation step in which polymerized actin (F-actin, FA) is collected in the pellet (sediments). Monomeric actin forms = free actin or covalently bound to M6/E24Fua Tβ4. Rhodamine B at 10 µM was used as a photosensitizer during 2 hours of irradiation. XL is cross-linked. The gel figure is assembled from two different gels.

Comparing lanes 1-2 and 3-4 indicates that under the mild oxidation condition used in which singlet oxygen is produced by Rhd B irradiation with white light source, the actin polymerization capacity of actin is not reduced: the amount of F-actin is comparable in lane 2 and 4.

Comparing lanes 3-4 to 5-6 or 7-8: the cross-linked M6/E24FuaTβ4-actin complex is in the supernatant. This indicates that Tβ4 covalently bound to actin using furan cross-linking prevents the bound actin to participate in actin-actin interaction (i.e. there is a shielding and inhibitory effect).

3.9. Supplementary Figure 9



Supplementary Figure 9: Actin sedimentation assay. SDS-PAGE analysis of supernatant and pellet after the ultracentrifugation step in which polymerized actin (F-actin, FA) is collected in the pellet and monomeric actin (MA) in the supernatant = free actin. No cross-linkable thymosin was present. For the sample on the right, prior to sedimentation, the actin was treated with Rose Bengal at 5 µM as a photosensitizer and irradiated for 5 minutes. Note these under these conditions, actin polymerization is no longer occurring.

4. AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: K.H. synthesized the Tβ4 protein analogues. K.H. and W.V. performed the initial XL experiments. L.M.C. and M.V.T designed and performed the experiments and optimized the XL methodology. L.M.C, M.V.T, and A.M. discussed the results, provided critical feedback, and helped shape the research. C.A., M.V.T. and A.M. conceived the presented idea. G.A.G., M.G., M.D.L., M.V. did the XL analysis and MS method development. L.M.C., C.A., M.V.T., and A.M. contributed to the final version of the manuscript. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. B. Xue, C. Leyrat, J. M. Grimes and R. C. Robinson, *Proc Natl Acad Sci U S A*, 2014, **111**, E4596-4605.
2. M. Van Troys, D. Dewitte, M. Goethals, M. F. Carlier, J. Vandekerckhove and C. Ampe, *EMBO J*, 1996, **15**, 201-210.
3. C. Husson, L. Renault, D. Didry, D. Pantaloni and M. F. Carlier, *Mol Cell*, 2011, **43**, 464-477.
4. B. Xue and R. C. Robinson, *Vitam Horm*, 2016, **102**, 55-71.
5. M. Domanski, M. Hertzog, J. Coutant, I. Gutsche-Perelroizen, F. Bontems, M. F. Carlier, E. Guittet and C. van Heijenoort, *J Biol Chem*, 2004, **279**, 23637-23645.
6. C. Simenel, M. Van Troys, J. Vandekerckhove, C. Ampe and M. Delepierre, *Eur J Biochem*, 2000, **267**, 3530-3538.
7. M. L. Mendes, L. Fischer, Z. A. Chen, M. Barbon, F. J. O'Reilly, S. H. Giese, M. Bohlke-Schneider, A. Belsom, T. Dau, C. W. Combe, M. Graham, M. R. Eisele, W. Baumeister, C. Speck and J. Rappsilber, *Mol Syst Biol*, 2019, **15**, e8994.
8. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma and J. A. Vizcaino, *Nucleic Acids Res*, 2019, **47**, D442-D450.
9. T. Kouyama and K. Mihashi, *Eur J Biochem*, 1980, **105**, 279-287.
10. D. Pantaloni and M. F. Carlier, *Cell*, 1993, **75**, 1007-1014.
11. D. Heintz, A. Reichert, M. Mihelic-Rapp, S. Stoeva, W. Voelter and H. Faulstich, *Eur J Biochem*, 1994, **223**, 345-350.
12. E. M. Llamas, J. P. C. Tome, J. M. M. Rodrigues, T. Torres and A. Madder, *Org Biomol Chem*, 2017, **15**, 5402-5409.
13. C. Veliz Montes, H. Memczak, E. Gyssels, T. Torres, A. Madder and R. J. Schneider, *Langmuir*, 2017, **33**, 1197-1201.
14. M. Op de Beeck and A. Madder, *J Am Chem Soc*, 2012, **134**, 10737-10740.
15. E. Antonatou, K. Hoogewijs, D. Kalaitzakis, A. Baudot, G. Vassilikogiannakis and A. Madder, *Chemistry*, 2016, **22**, 8457-8461.
16. M. Graham, C. W. Combe, L. Kolbowski and J. Rappsilber, *BioRxiv*, 2019.