Supporting Information:

Site-directed cysteine coupling of disulfide-containing non-antibody carrier proteins (THIOCAPs)

Ariana Rueda, Julian I. Mendoza, Lorena Alba-Castellon, Eloi Parladé, Eric Voltà-Durán, David Paez, Anna Aviño, Ramon Eritja, Esther Vázquez, Antonio Villaverde, Ramón Mangues, Ugutz Unzueta.

Supplementary Methods:

Transmission Electron Microscopy

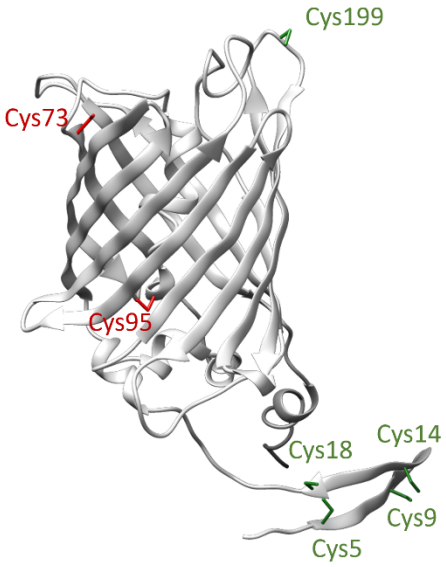
High-resolution electron microscopy images of the nanomaterials were obtained by transmission electron microscopy (TEM) using the Jeol 1400 transmission electron microscope (Jeol Ltd.) set to operate at 80 kV and equipped with a Gatan Orius 8 9 SC200 CCD (Gatan Inc.). Sample preparation consisted in placing 5 µL of protein nanoconjugate at 0.1 mg/ml on glow-discharged 200 mesh carbon-coated copper grids (Electron Microscopy Science) for 1 minute. Excess liquid was removed by blotting with a Whatman filter paper, the protein was then negatively stained with a 5 µL droplet of 1% uranyl acetate (Polysciences Inc.) for another minute and blotted again. Lastly, the grids were left to air-dry at room temperature for at least 10 minutes. The presented images were captured from different fields at 12,000x and 25,000x magnifications. Average nanoparticles size distribution was calculated using Image J 1.54d software from representative images.

Dynamic Light Scattering

Size distribution of THIO-T22-GFP-H6-MMAE and T22-HSNBT-H6-MMAE nanoconjugates was determined by Dynamic Light Scattering (DLS) in a Zetasizer Ultra Red (Malvern Instruments) at 633nm and backscattered detector (173º). SDS 4% (v/v) was used for the disassembly of the nanoparticles to their building blocks. Samples were measured in triplicate (n = 3) and the average intensity size values expressed as mean ± standard error.

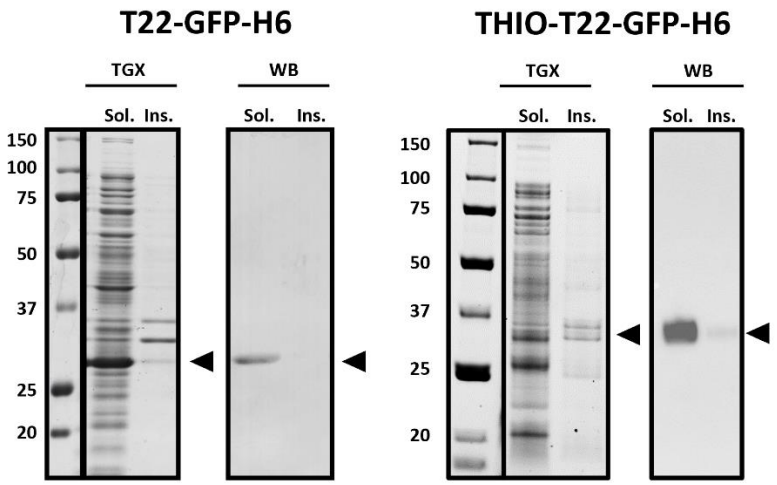
Micotoxin conjugation

Maleimide functionalized Micotoxin was purchased at Levena Biopharma (SET0206) and resuspended in anhydrous dimethyl sulfoxide (DMSO). THIO-T22-GFP-H6 was then incubated in presence of 1:2 protein:drug molar ratios for 4 h, in aqueous solution (166mM $NaCO_3H$, 333mM NaCl, 0.8mM EDTA pH=8) at room temperature. Excess of non-conjugated drug was finally completely removed by 5 mL HighTrap Desalting column (Cytiva) equilibrated with sodium bicarbonate with salt solution (166 mM $NaCO_3H$, 333mM NaCl, pH 8) in an ÄKTA pure (Cytiva) chromatography system.



Residues		RSA
Cys5		0.73
Cys9	S-S	0.82
Cys14	S-S	0.75
Cys18		0.74
Cys73		0.12
Cys95		0
Cys199		0.70

Figure S1: In-silico analysis of relative solvent accessibility (RSA) of cysteines in THIO-T22-GFP-H6 protein.



PROTEIN PRODUCTION	Cell soluble fraction	Cell insoluble fraction
T22-GFP-H6	94.83 %	5.17%
THIO-T22-GFP-H6	93.22%	6.78%

Figure S2: T22-GFP-H6 (30.69 kDa) and THIO-T22-GFP-H6 (30.79 kDa) protein production in *E.coli* origami B upon induction with IPTG over night at 20ºC. Cell soluble (Sol.) and insoluble (Ins.) fractions were analyzed by SDS-PAGE (TGX) and Western blot immunodetection (WB) using an anti-His monoclonal antibody.

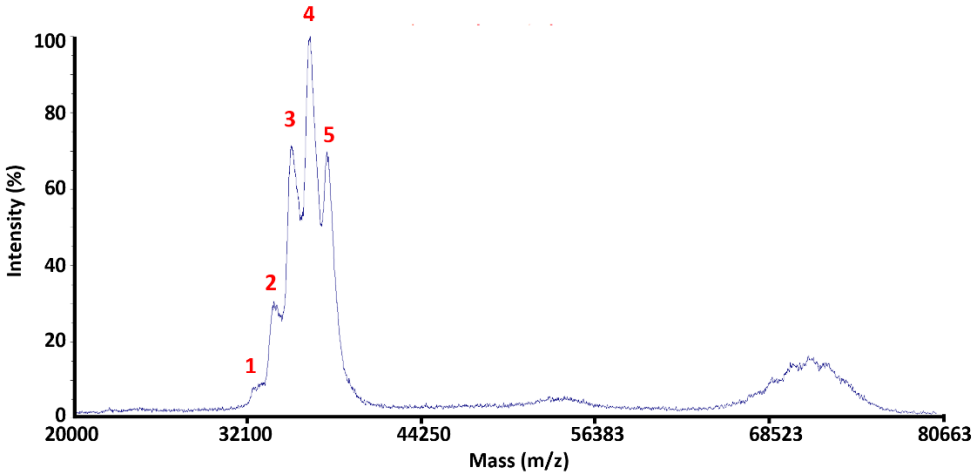


Figure S3: MALDI-TOF spectrum of THIO-T22-GFP-H6-MMAE nanoconjugates obtained from the cysteine-selective conjugation of TCEP-resuspended THIO-T22-GFP-H6 Carrier protein with MMAE at low protein:drug molar ratio (1:5 ratio). Numbers above each peak indicates the amount of MMAE molecules incorporated over the protein (0: 30.79kDa, 1: 32,11 kDa, 2: 33,42 kDa, 3: 34,73 kDa, 4: 36,05 kDa, 5: 37,37 kDa). Proteins with up to 5 MMAE molecules can be detected indicating the presence of 5 available uncapped cysteines (Cys5, Cys9, Cys14, Cys18 from the disulfudes-disrupted T22 peptide plus the engineered Cys199).

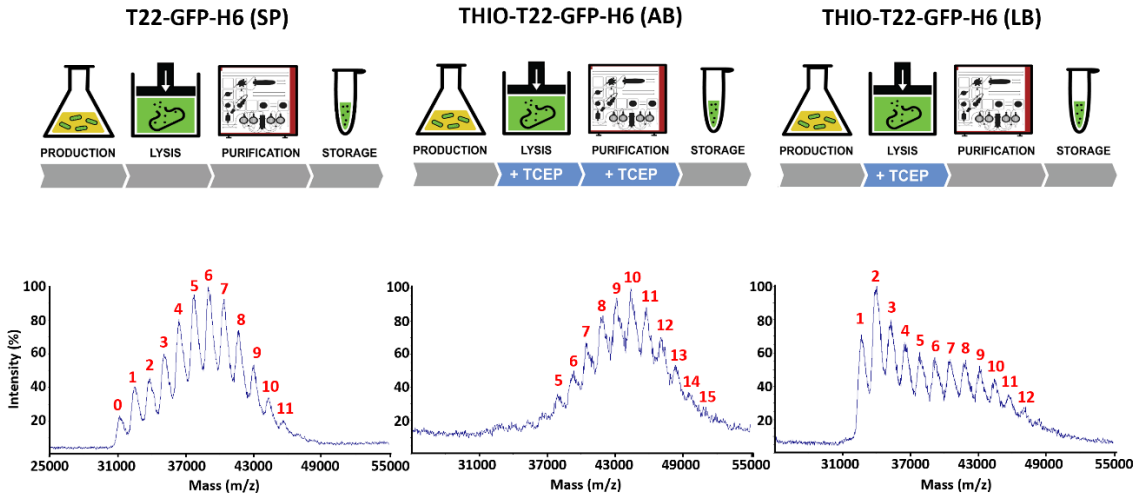


Figure S4: MALDI-TOF spectrum of T22-GFP-H6-MMAE and THIO-T22-GFP-H6-MMAE nanoconjugates generated by unspecific nucleophilic attack of protein lysine-amines at high protein: MMAE molar ratio (1:50 ratio). Numbers above each peak indicates the amount of MMAE molecules incorporated over the protein. T22-GFP-H6 standard protocol (SP), 0: 30.69 kDa, 1: 32.00 kDa, 2: 33.32 kDa, 3: 34.64 kDa, 4: 35.95 kDa, 5: 37.27 kDa, 6: 38.59 kDa, 7: 39.90 kDa, 8: 41.21 kDa, 9: 42.54 kDa, 10: 43.85 kDa, 11: 45.17 kDa. THIO-T22-GFP-H6 all buffers protocol (AB) and THIO-T22-GFP-H6 lysis buffer protocol (LB), 0: 30,79 kDa, 1: 32.11 kDa, 2: 33.42 kDa, 3: 34.74 kDa, 4: 36.05 kDa, 5: 37.37 kDa, 6: 38.68 kDa, 7: 40.00 kDa, 8: 41.31 kDa, 9: 42.63 kDa, 10: 43.95 kDa, 11: 45.27 kDa, 12: 46.58 kDa, 13: 47.89 kDa, 14: 49.21 kDa, 15: 50.53 kDa.

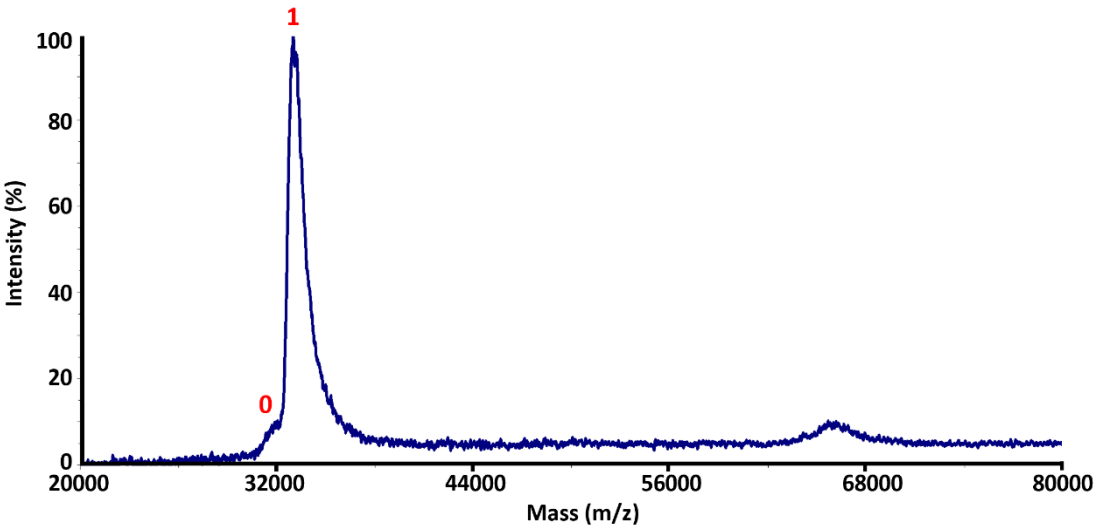


Figure S5: MALDI-TOF spectrum of THIO-T22-GFP-H6-Micotoxin conjugated at 1:2 molar ratio of maleimide-functionalized drug. Numbers above each peak indicates the amount of drug molecules (1,6kDa) incorporated over the protein (0: 30.79kDa, 1: 32,4 kDa).

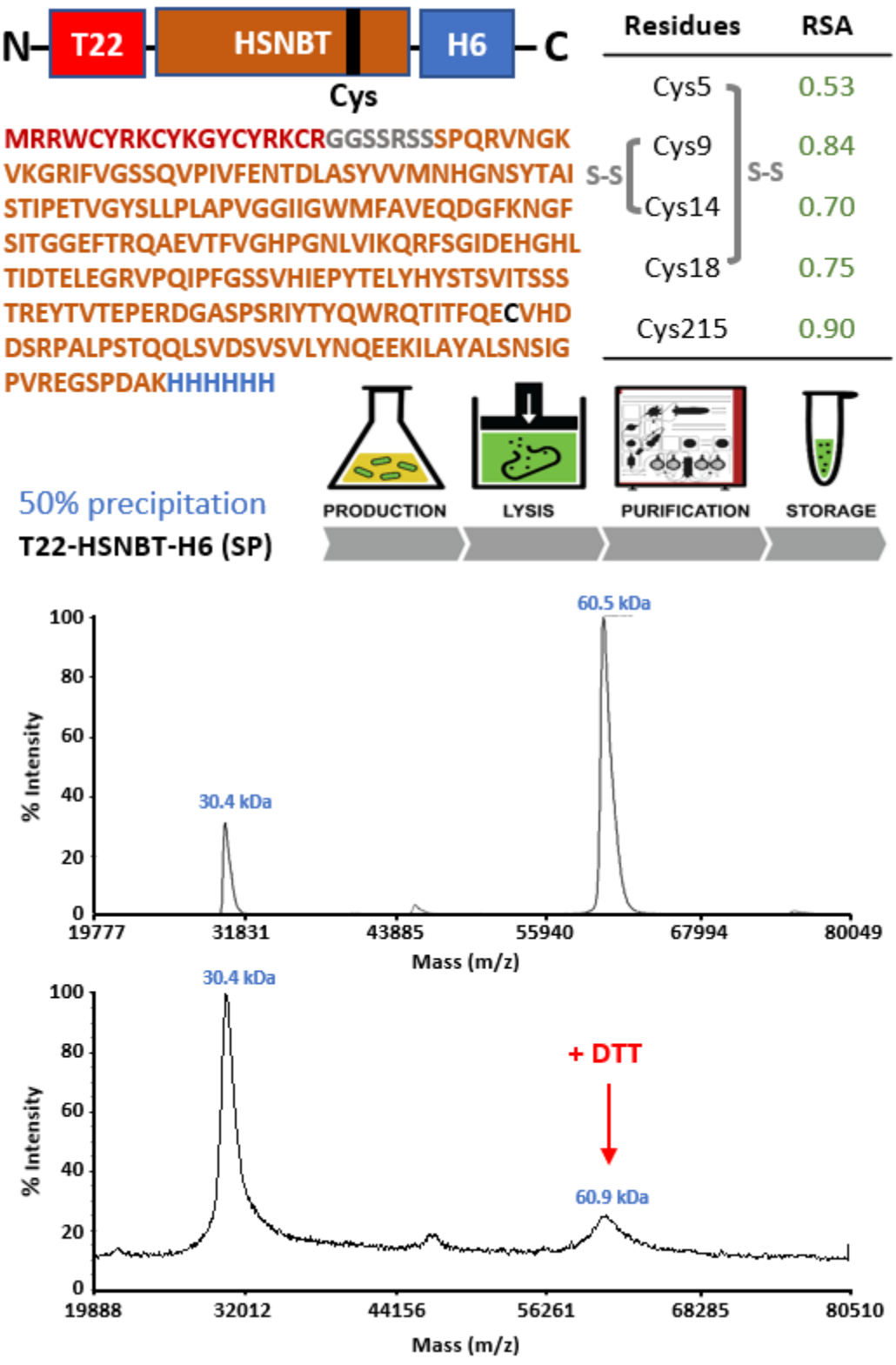
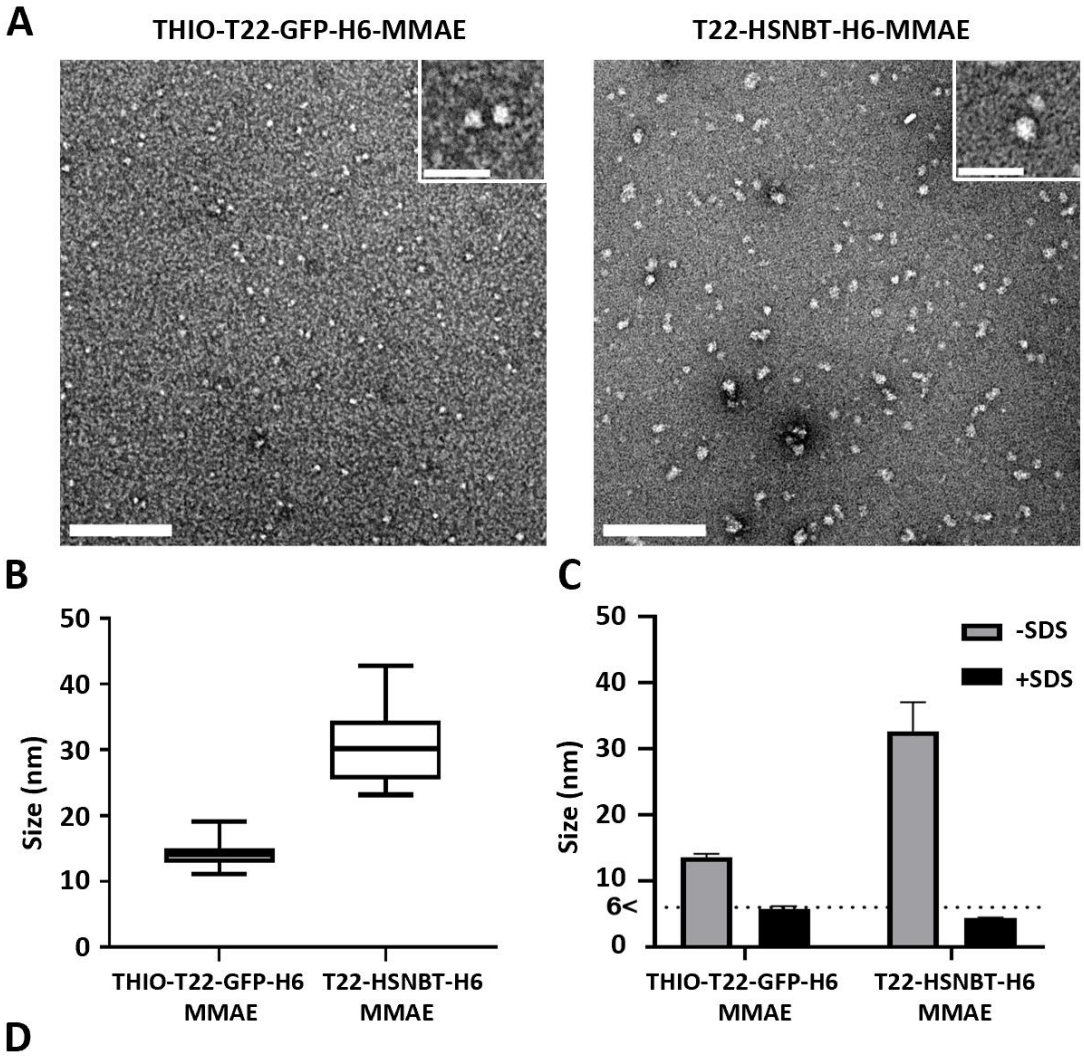


Figure S6: Modular protein design and amino acid sequence of T22-HSNBT-H6 THIOCAP. Box sizes of the modules are only indicative. Table shows, in-silico analysis of relative solvent accessibility (RSA) of cysteines. Below, MALDI-TOF spectrum of T22-HSNBT-H6 in the cell soluble fraction upon purification with the standard protocol. A mixture of protein monomers (30.3 kDa) and disulfide dimers (60.6 kDa) are observed which disappear upon addition of a disulfide reducing agent (+DTT).



D

	DLS (nm)	TEM (nm)
THIO-T22-GFP-H6-MMAE	**13,61** ± 0,5	**14,25** ± 1,9
T22-HSNBT-H6-MMAE	**32,65** ± 4,4	**30,36** ± 5,1

Figure S7: A) Transmission electron microscopy (TEM) images of THIO-T22-GFP-H6-MMAE and T22-HSNBT-H6-MMAE nanoconjugates. White scale bar indicates 200nm, 50nm in the inset. B) Size distribution of THIO-T22-GFP-H6-MMAE and T22-HSNBT-H6-MMAE nanoparticles calculated from TEM image analysis. C) Size distribution of THIO-T22-GFP-H6-MMAE and T22-HSNBT-H6-MMAE nanoparticles determined by DLS. SDS was used for nanoparticles disassembly into their building blocks. D) Side-to-side size distribution comparison from DLS and TEM values.

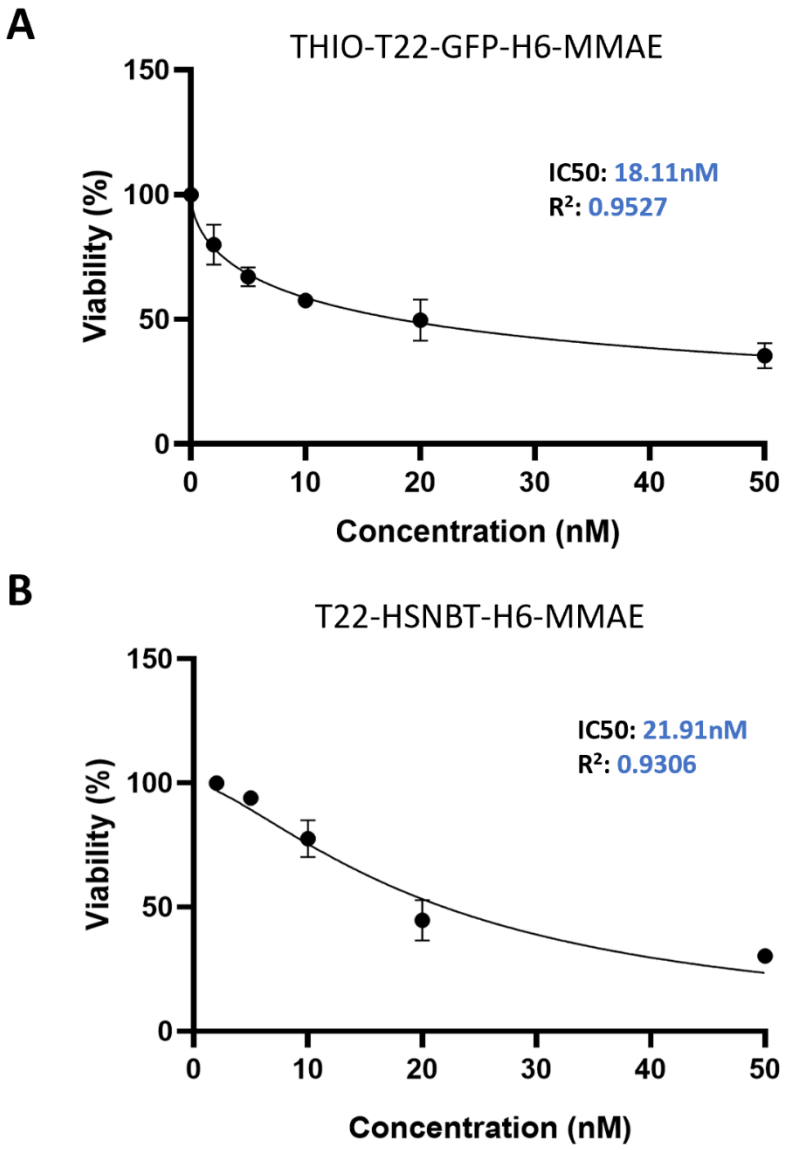


Figure S8: Dose-response trend representation and IC50 determination over UM-SCC-22A-CXCR4+ cancer cells upon 48 h incubation at different THIO-T22-GFP-H6-MMAE (A) and T22-HSNBT-H6-MMAE (B) concentrations. Data are represented as mean viability ±standard error.