Supporting information

Total Chemical Synthesis of Murine ISG15 and an Activity-based Probe with Physiological Binding Properties

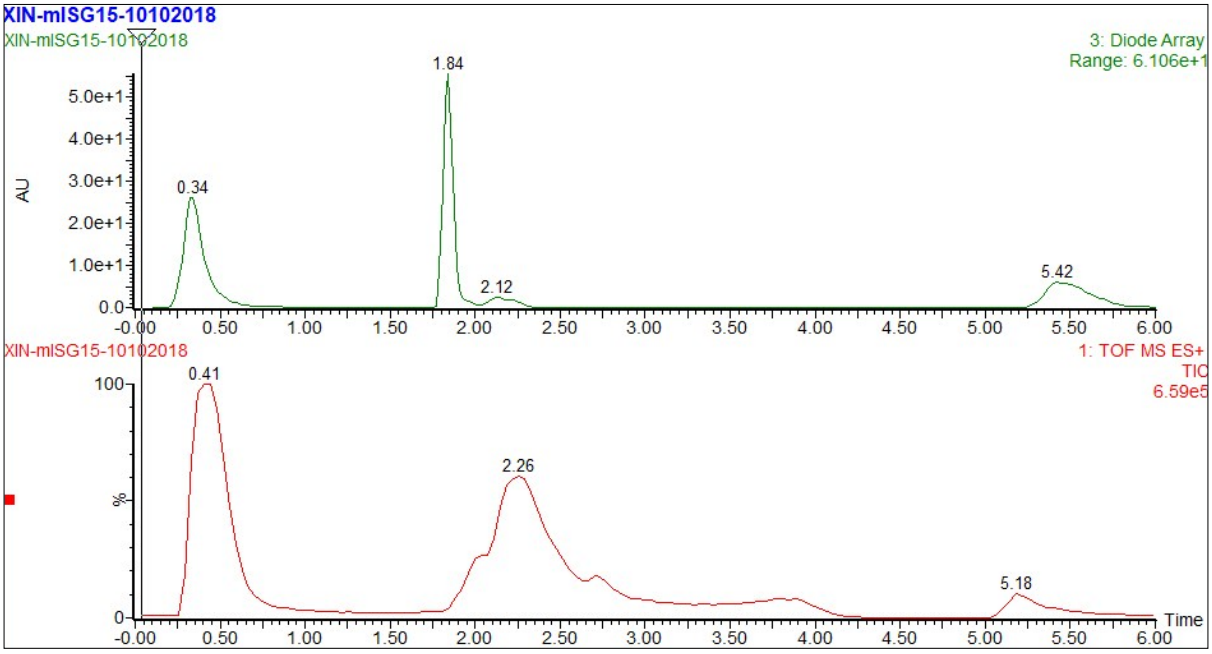
Bo-Tao Xin,[a‡] Jin Gan, [a‡] Daniel J. Fernandez,[b] Klaus-Peter Knobeloch,[b] Paul P. Geurink,[a] and Huib Ovaa.[a*]

[a] Oncode Institute and Department of Cell and Chemical Biology, Leiden University Medical Center (LUMC), 2333 ZC Leiden, The Netherlands;

[b] Institute of Neuropathology, Faculty of Medicine, University of Freiburg, Freiburg, Germany, Breisacherstraße 64, 79106 Freiburg, Germany.

Table of Content

Figure S1. LC-MS data of crude reaction mixture (Protein **1**) (LCT)



MS spectrum (Rt= 2.26 min)

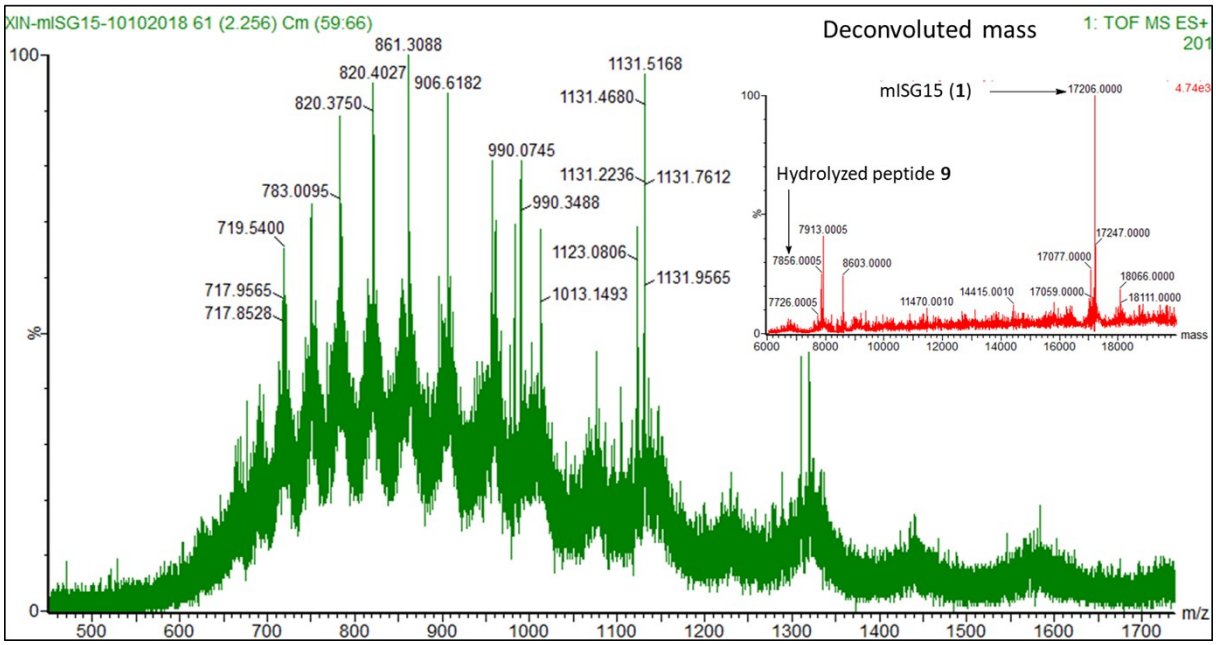
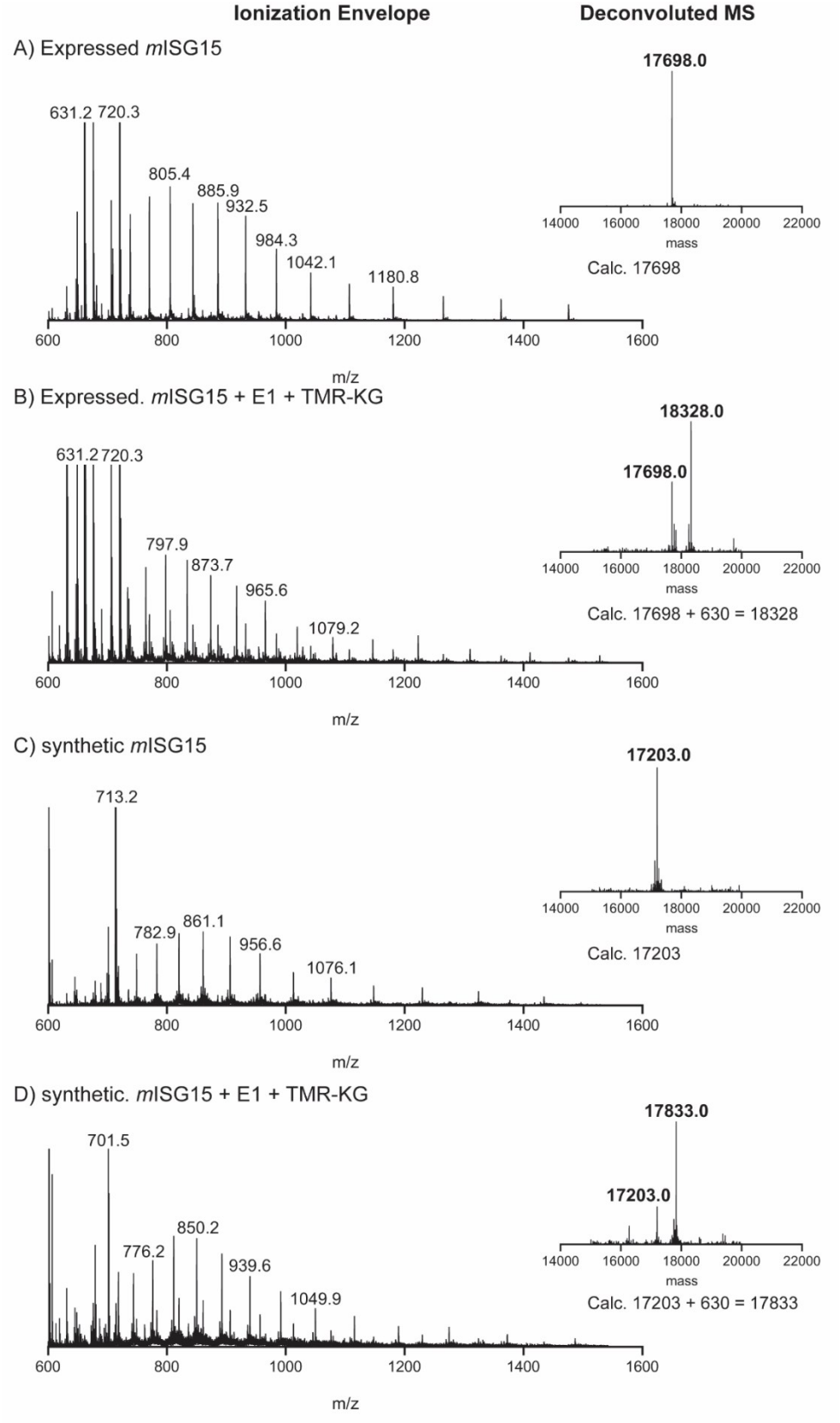
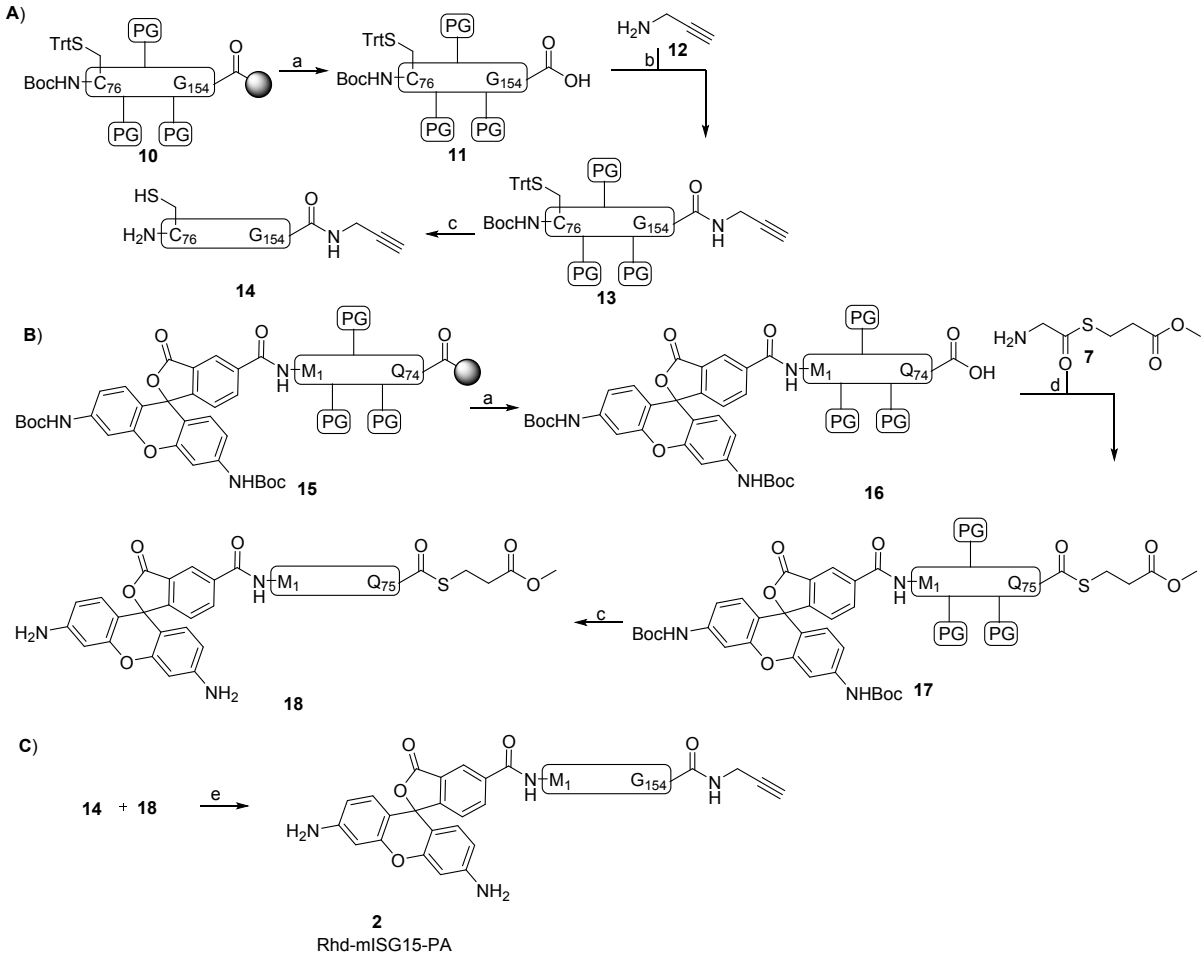


Figure S2. MS analysis of 5-TAMRA-Lys(mISG15)-Gly-OH (XEVO). LC-MS analysis of the reaction between expressed mISG15 with TMR-KG substrate in the absence of ISG15 E1 (A), or in the presence of ISG15 E1 (B); LC-MS analysis of the reaction between synthetic mISG15 with TMR-KG substrate in the absence of ISG15 E1 (C), or in the presence of ISG15 E1 (D).

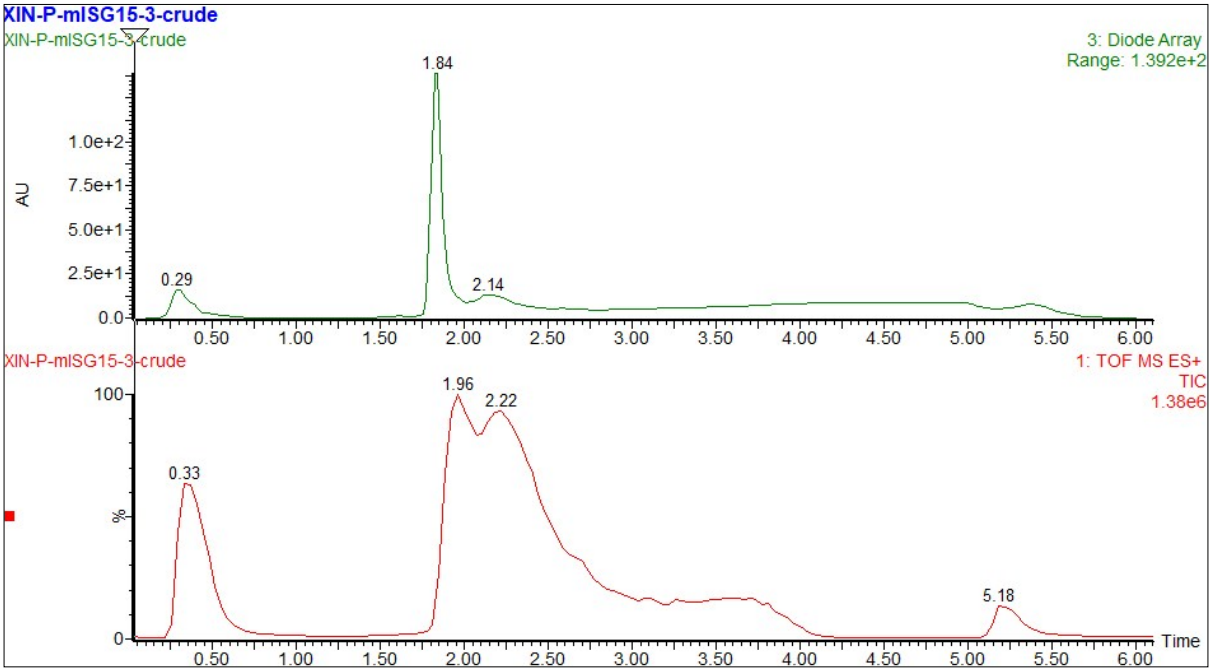


Scheme S1. Synthesis of Rho-mISG15-PA probe (2). (A) Synthesis of the N-terminal domain; (B) synthesis of the C-terminal domain; (C) synthesis of **2** via NCL.



Reagents and conditions: (a) 20% HFIP, DCM (v/v); (b) EDC·HCl, HOBt, $CHCl_3$ (c) TFA, H_2O, TIPS, phenol (90.5/5/2.5/2, v/v/v/v); (d) EDC·HCl, HOBt, $CHCl_3/CF_3CH_2OH$, -18 ºC; (e) 6M Gnd·HCl, 0.15M sodium phosphate, 0.25M MPAA, pH 7.5, 37 ºC.

Figure S3. LC-MS data of crude reaction mixture (Protein 2) (LCT)



MS spectrum (Rt= 2.22 min)

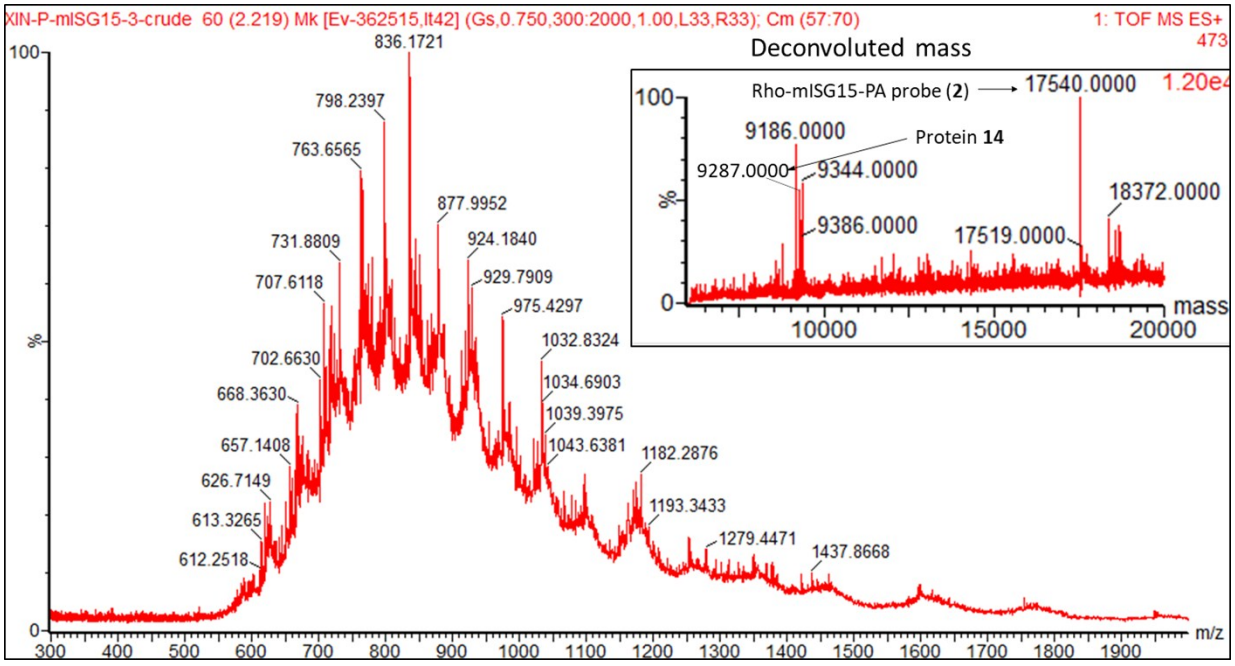
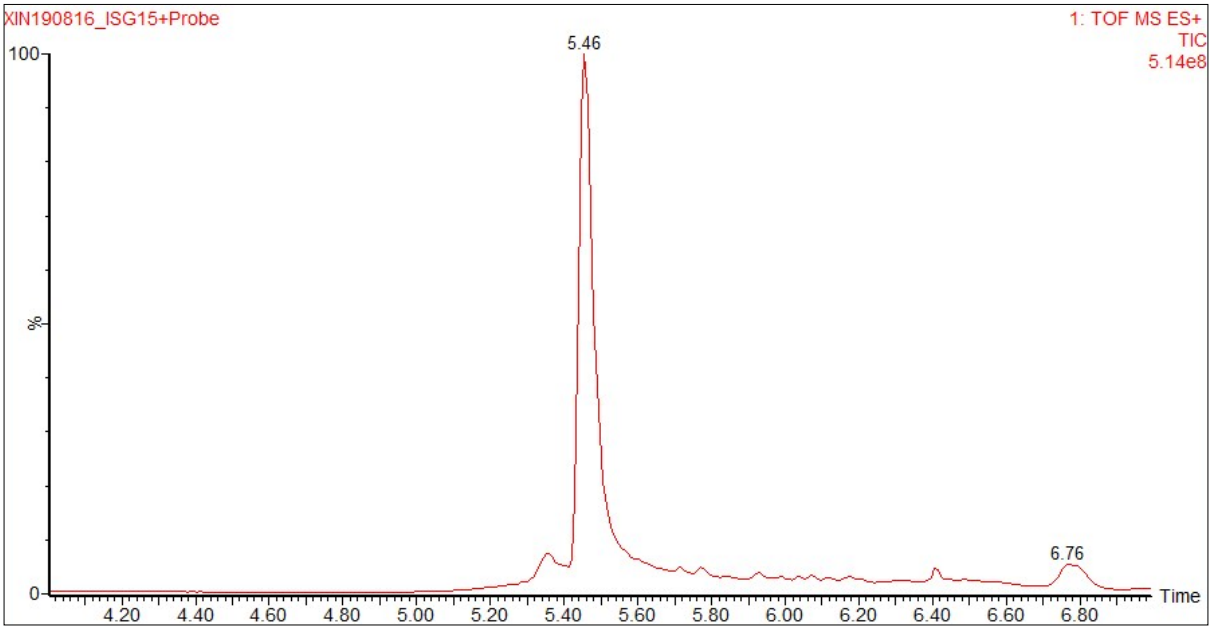


Figure S4. MS data of Rho-mISG15-PA probe (**2**) (XEVO)



MS spectrum (Rt= 5.46 min)

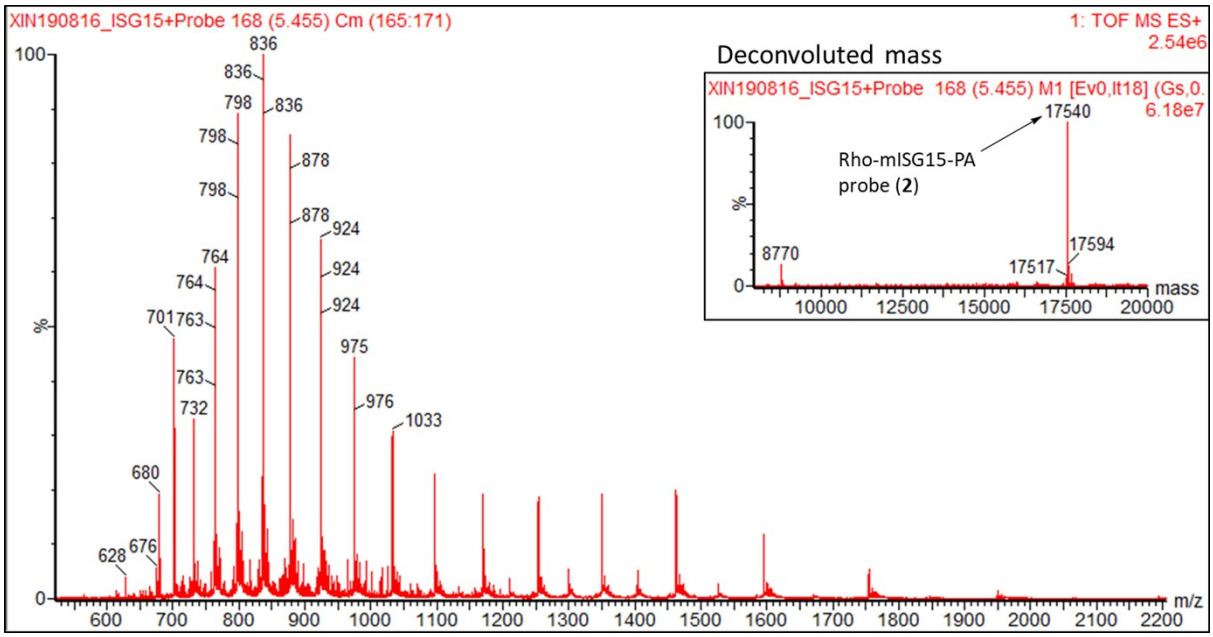


Figure S5. Assessment of Rhodamine-mISG15-PA reactivity against mUSP18. (A) Corresponding immunoblots for labelling of endogenous enzymes in mouse NMuMG cell lysates. (B) Corresponding fluorescence scan and full immunoblots for time-dependent labelling of overexpressed Flag-mUSP18 in human HEK293T cell lysates.

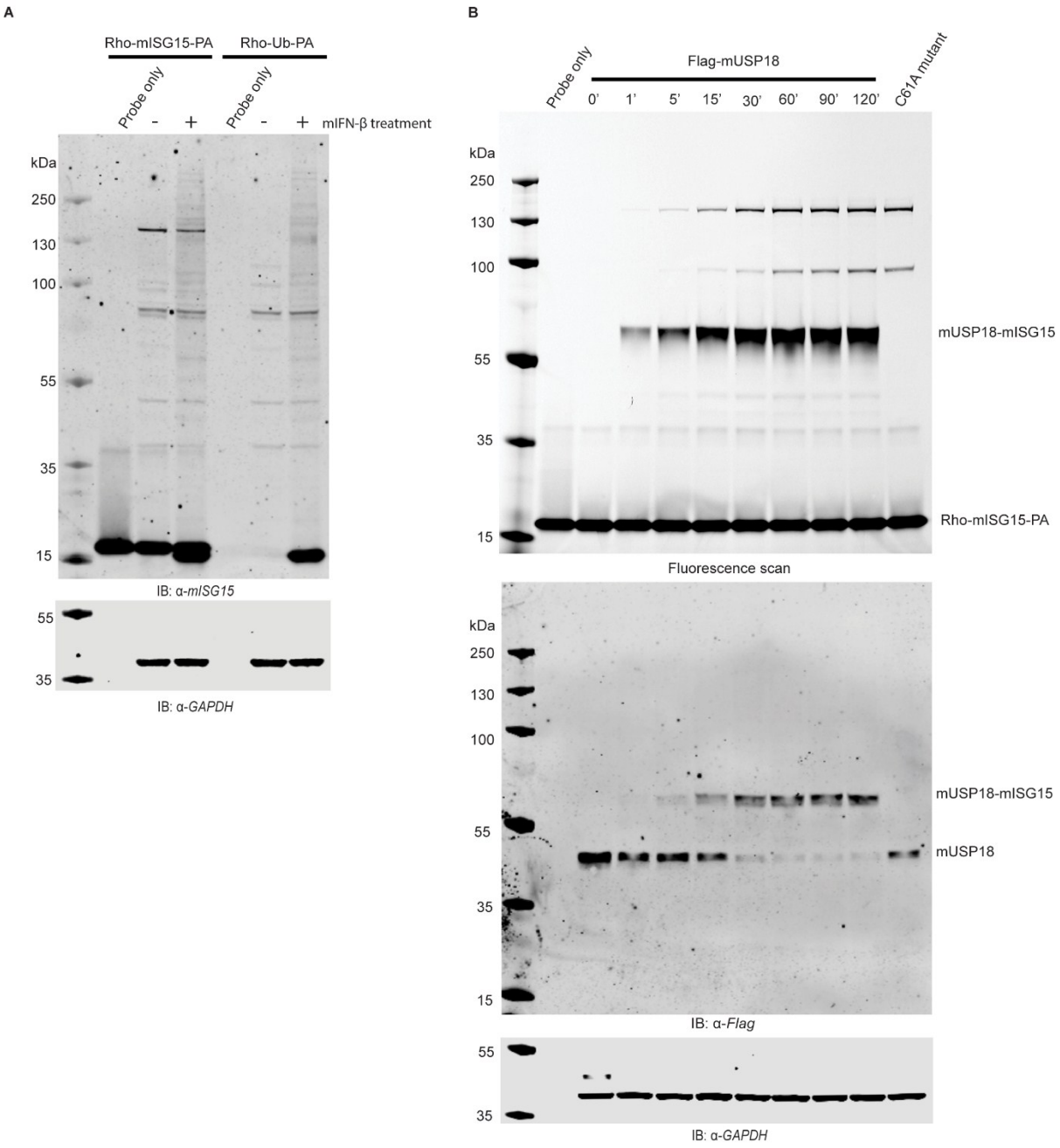


Figure S6. Assessment of Rhodamine-mISG15-PA reactivity against endogenous mUSP18. Corresponding fluorescence scan, full immunoblots for labelling of endogenous mUSP18 in mouse NMuMG cell lysates. The labelled samples were resolved in SDS-PAGE gel, and scanned in Rhodamine channel for probe signal, then the gel was transferred to Nitrocelluse membrane, and immunobloted against anti-mUSP18, anti-mISG15 and anti-β-Actin antibodies sequentially.

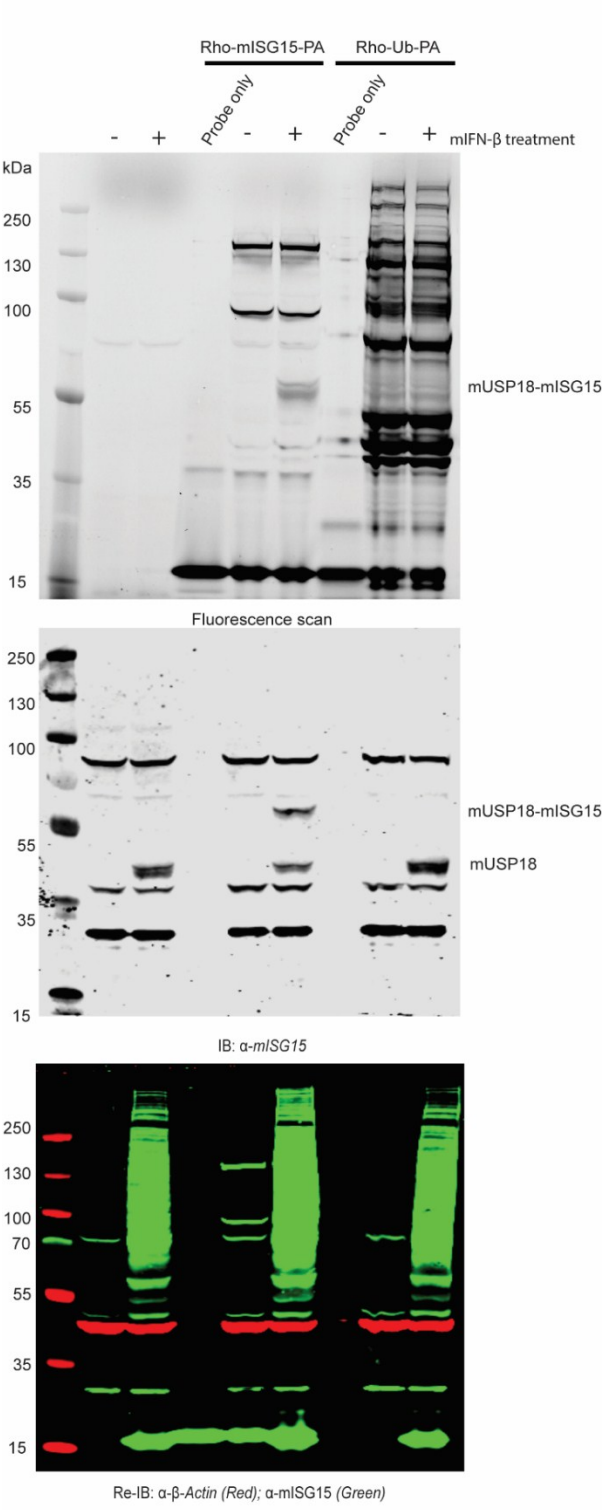
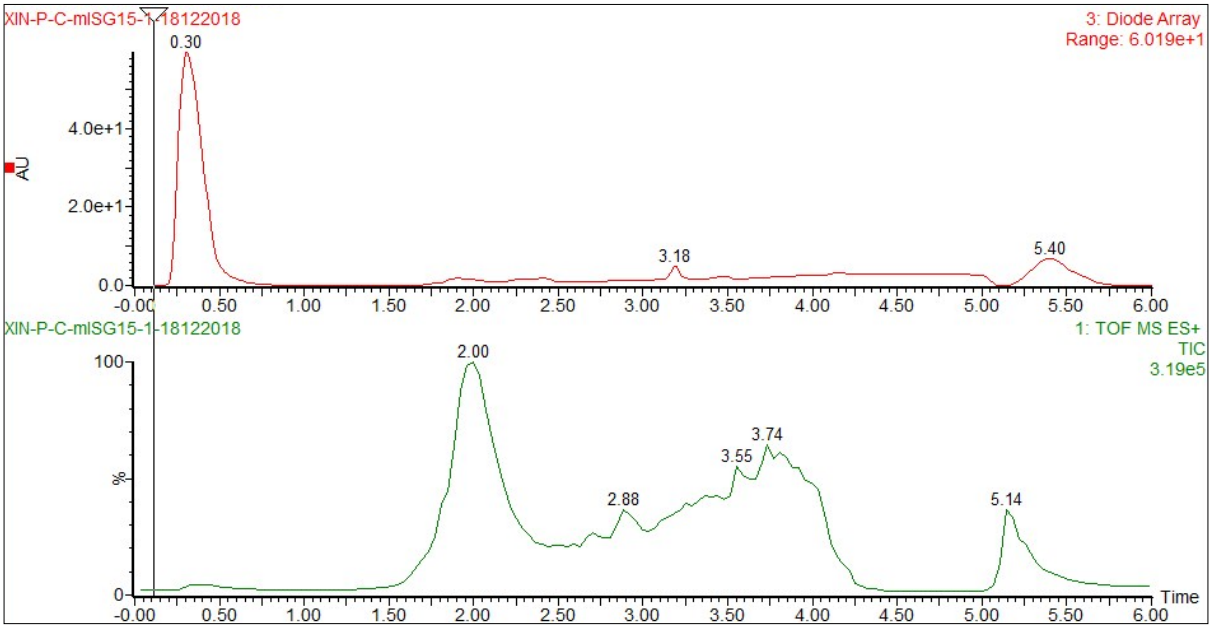


Figure S7. LC-MS data of Protein **4** (LCT)



MS spectrum (Rt= 2.00 min)

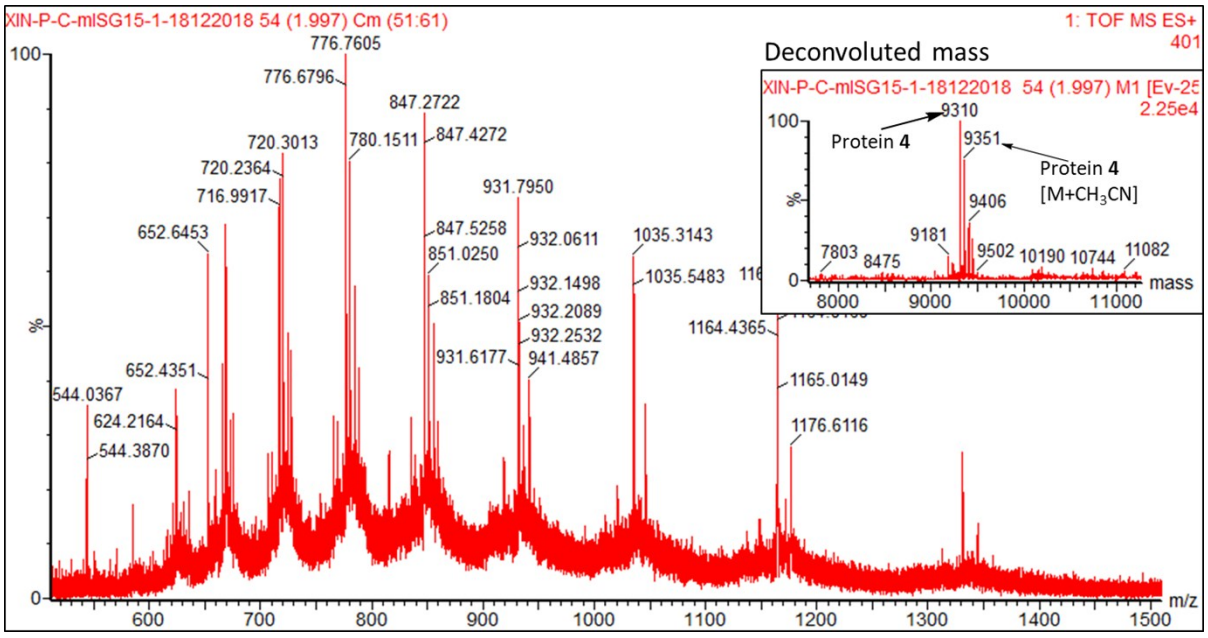
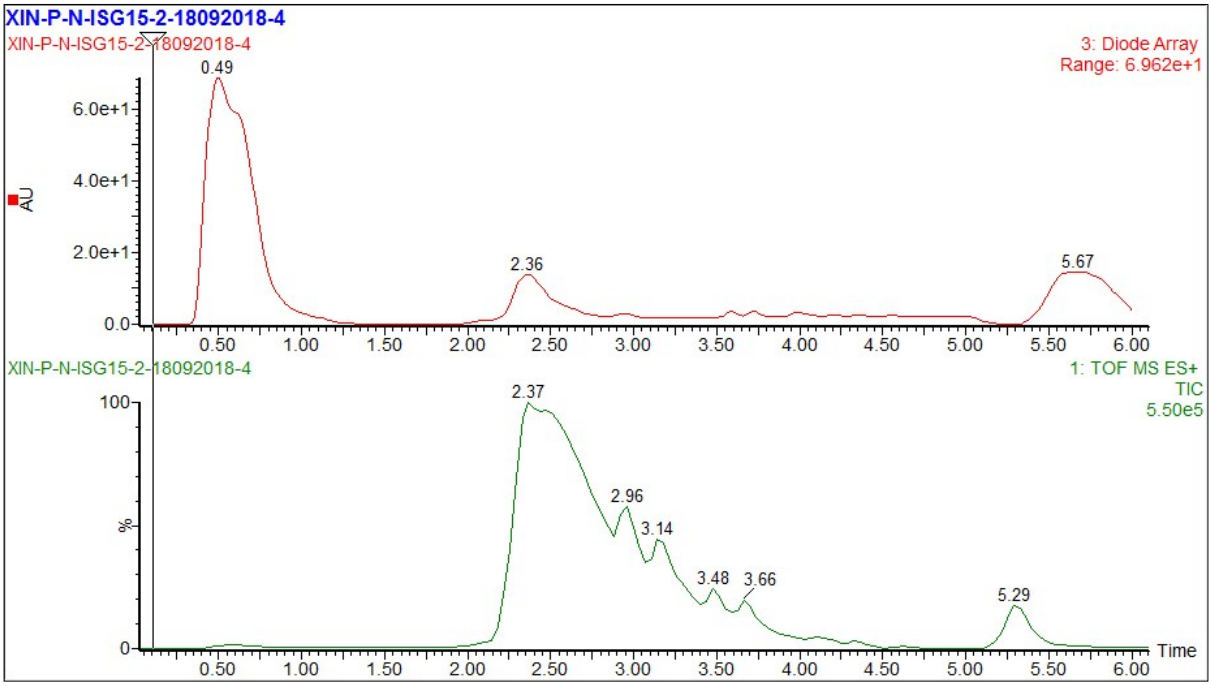


Figure S8. LC-MS data of Protein **6** (LCT)



MS spectrum (Rt= 2.37 min)

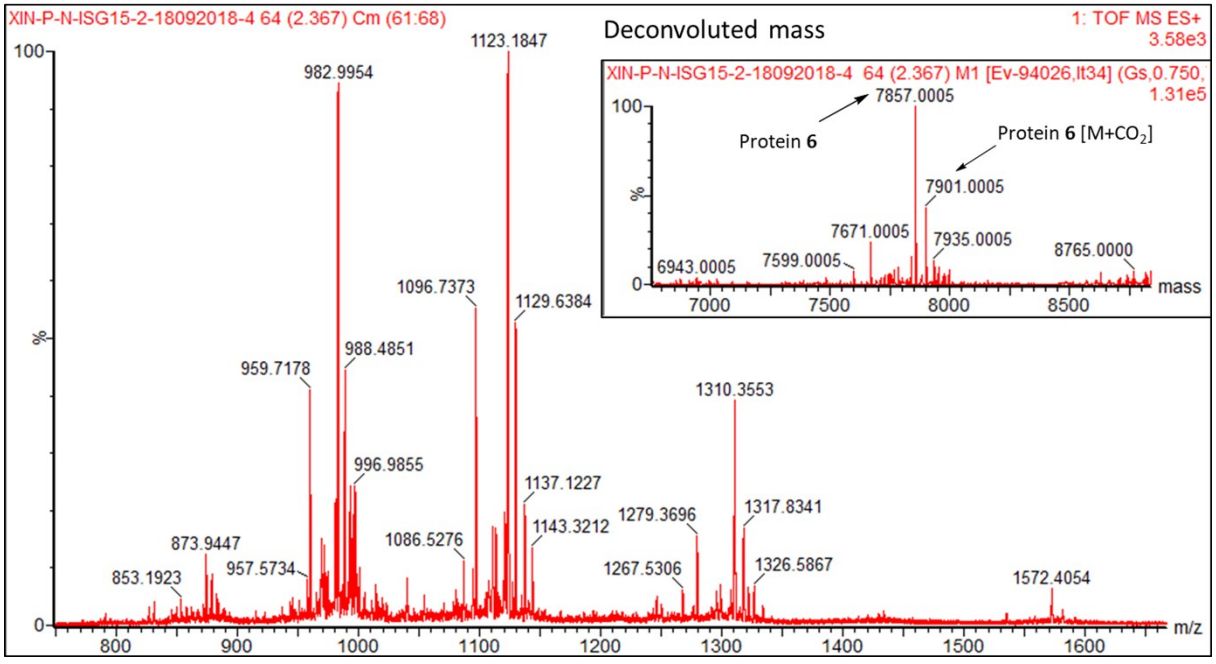
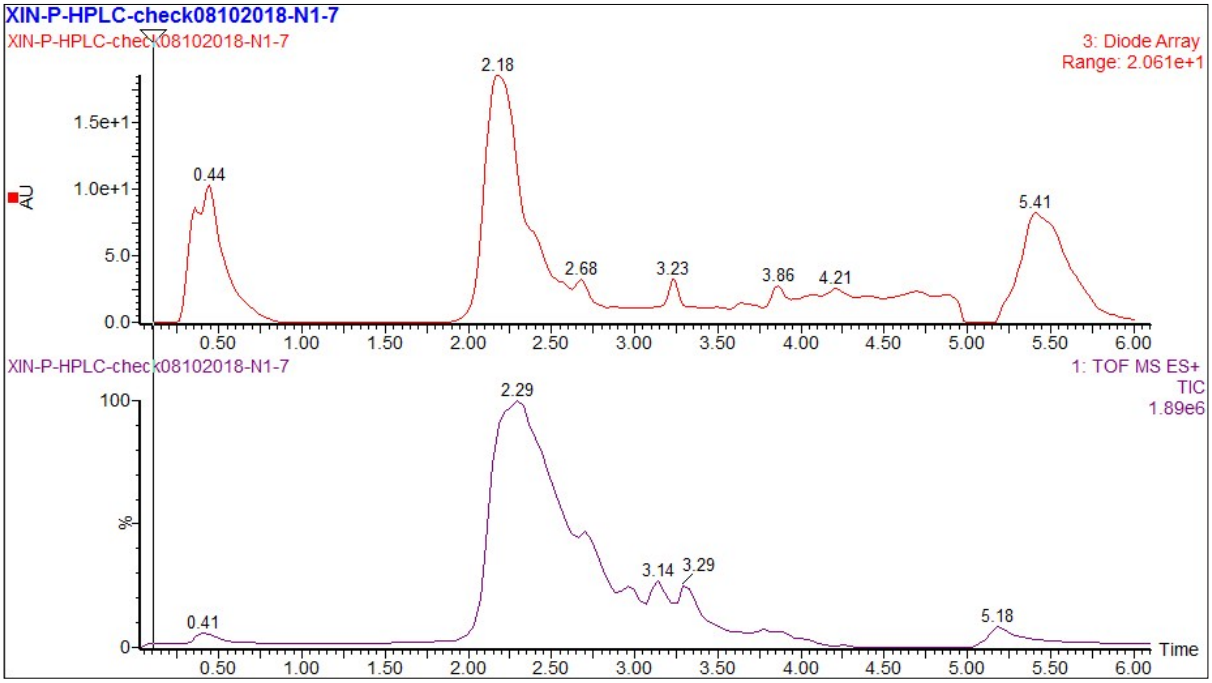


Figure S9. LC-MS data of Protein **9** (LCT)



MS spectrum (Rt= 2.29 min)

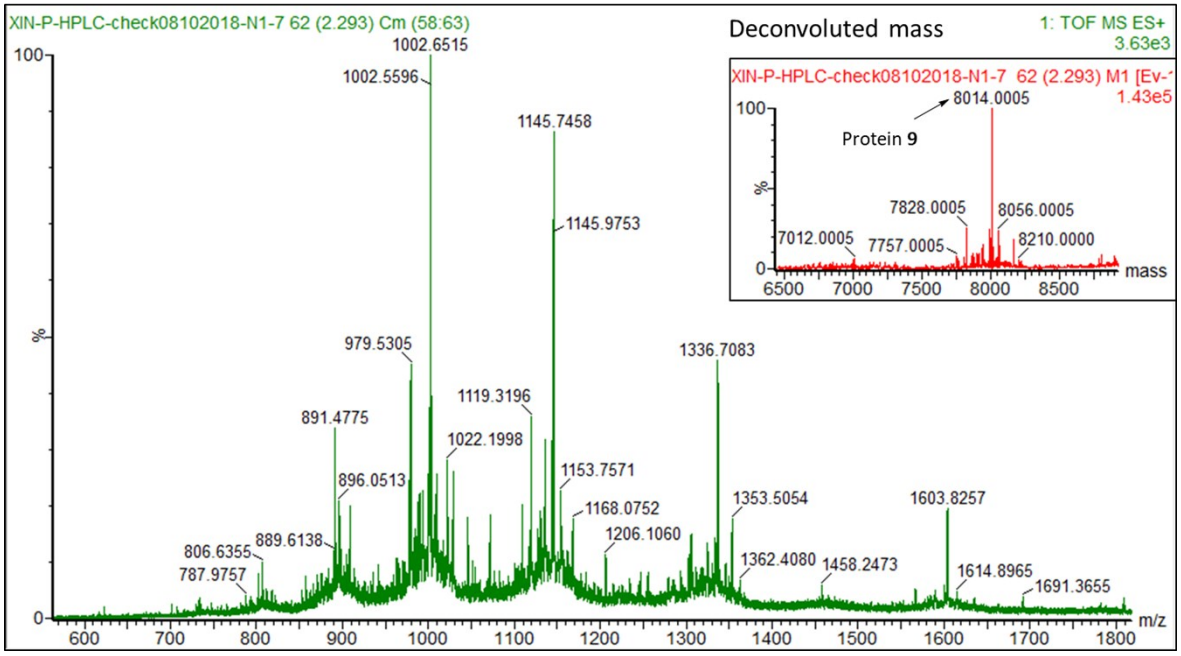
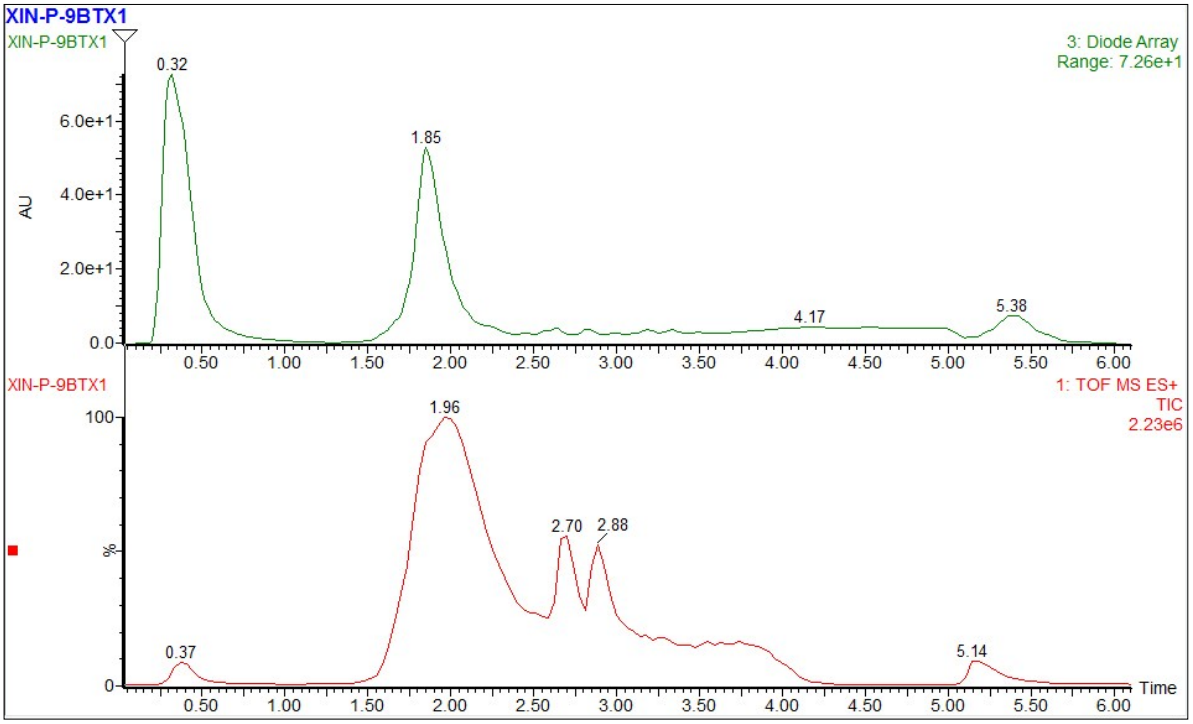


Figure S10. LC-MS data of Protein **11** (LCT)



MS spectrum (Rt= 1.96 min)

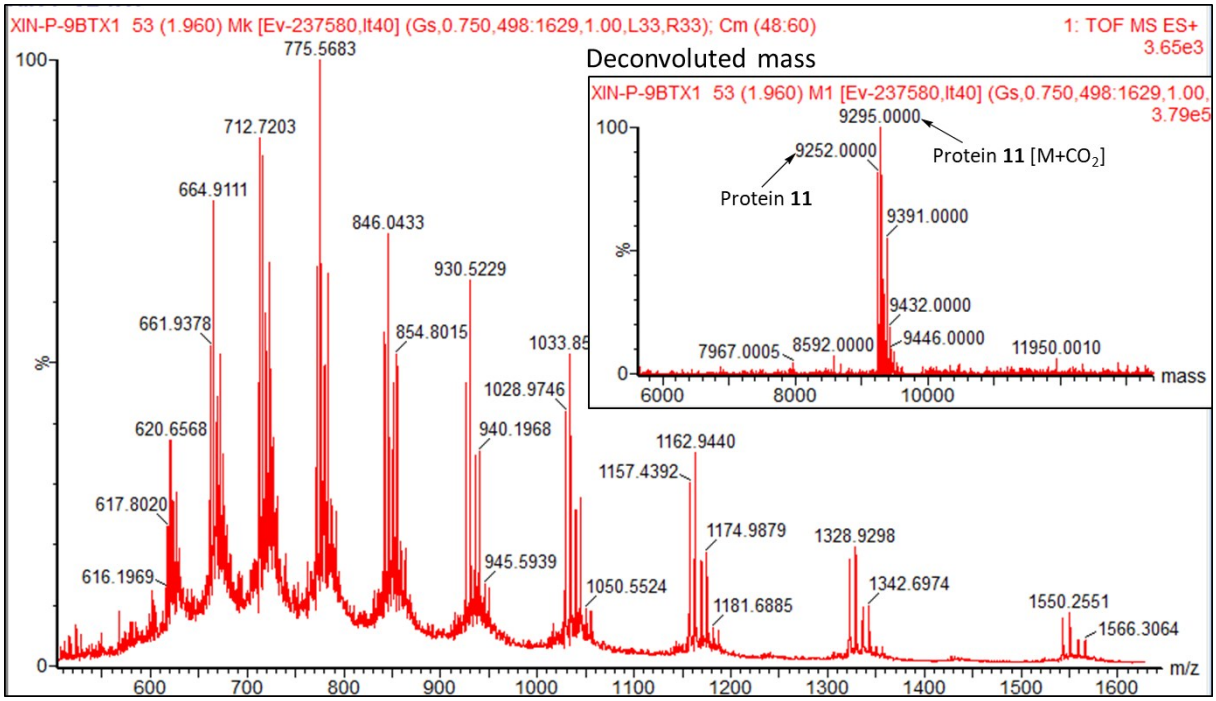
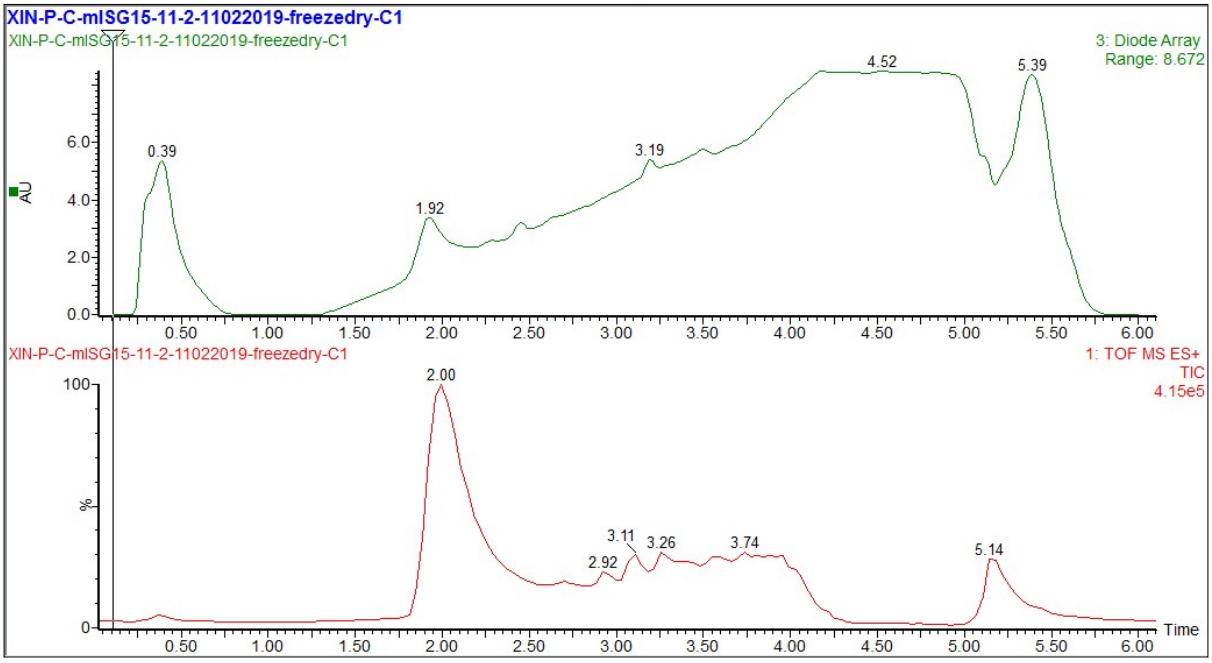


Figure S11. LC-MS data of Protein **14** (LCT)



MS spectrum (Rt= 2.00 min)

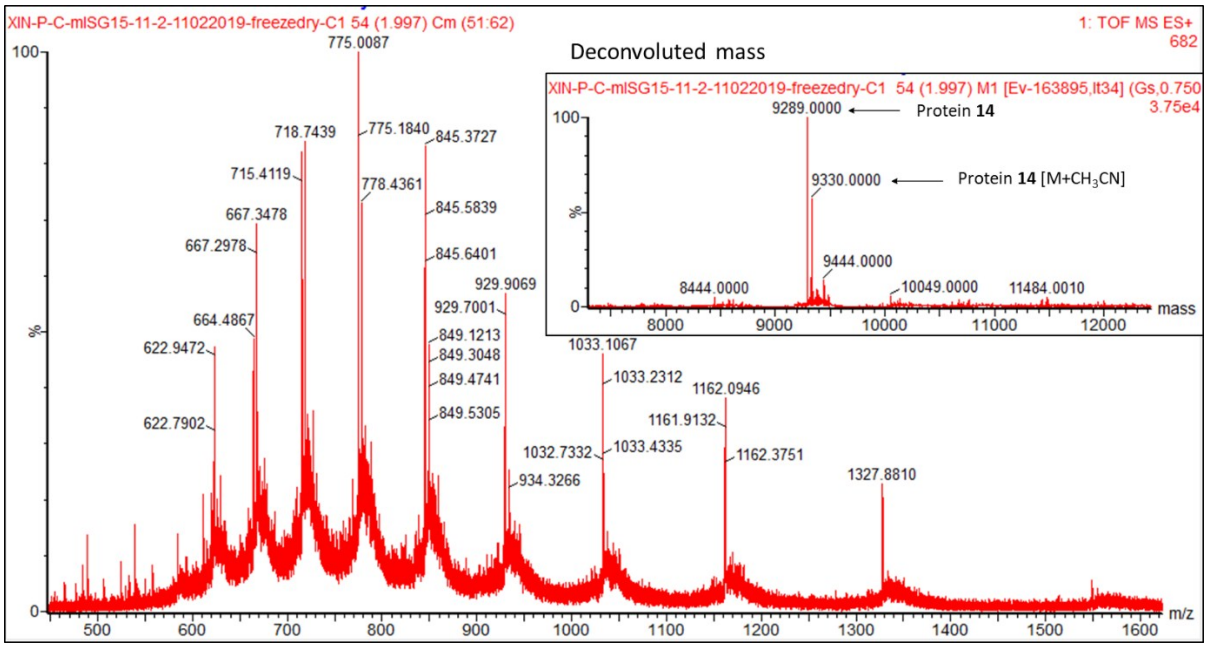
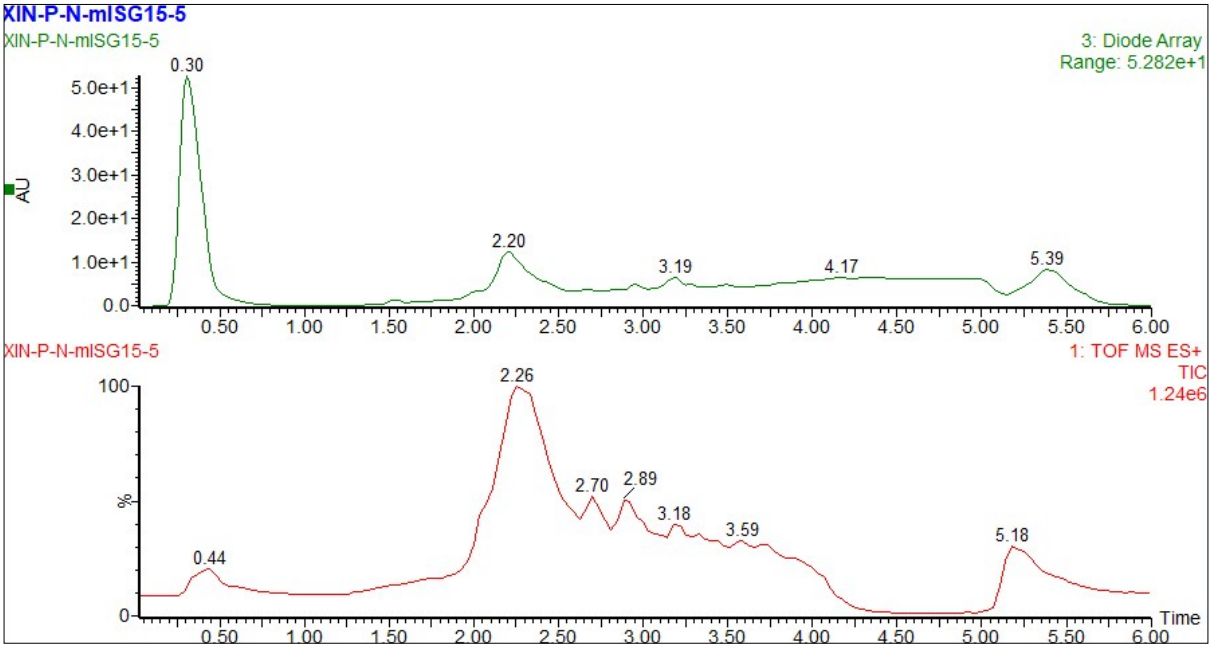


Figure S12. LC-MS data of Protein **16** (LCT)



MS spectrum (Rt= 2.26 min)

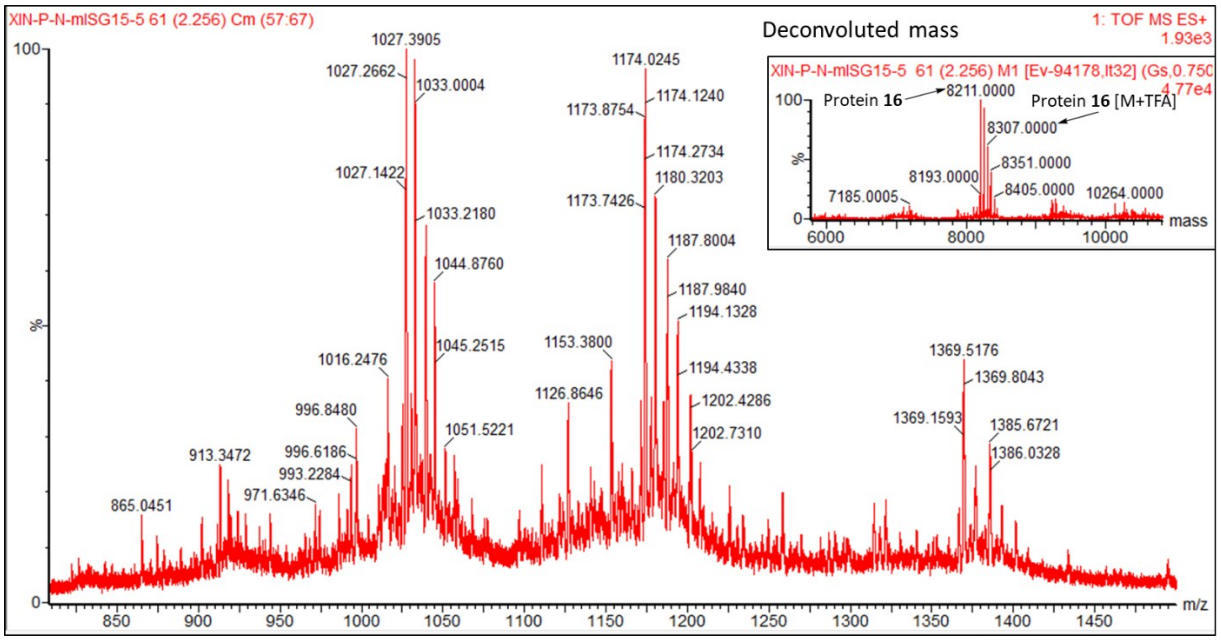
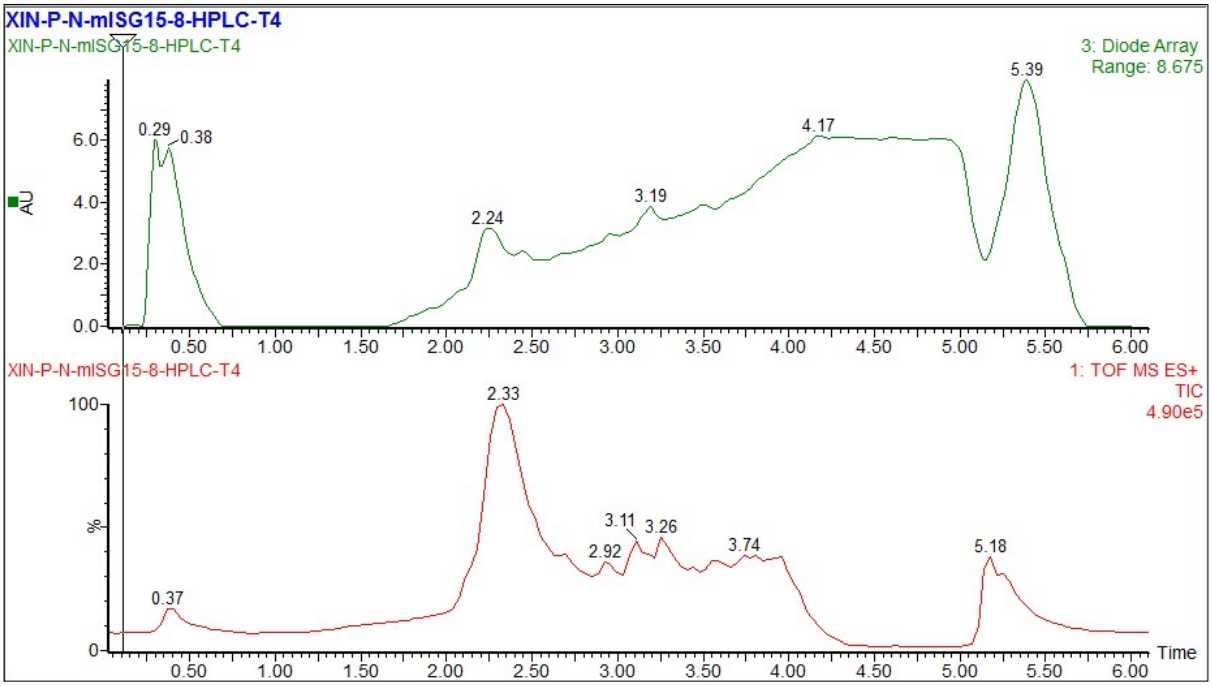
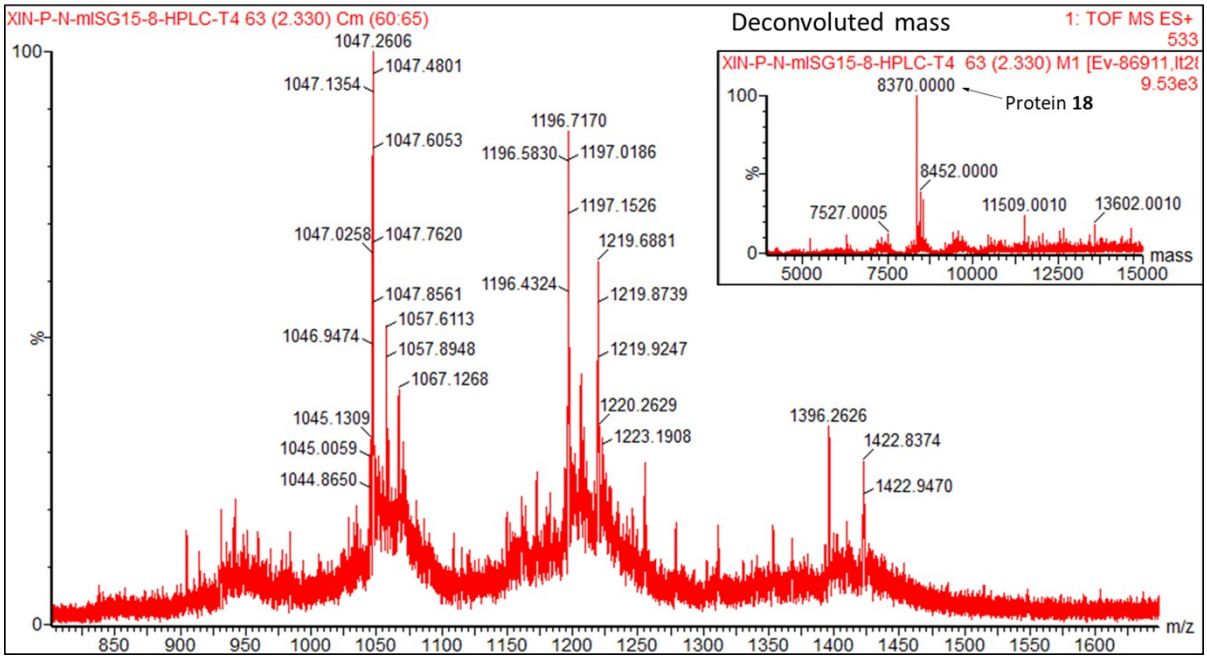


Figure S13. LC-MS data of Protein **18** (LCT)



MS spectrum (Rt= 2.33 min)



Chemical Synthesis

LC-MS analysis (LCT)

LC-MS data was recorded on a Waters LC-MS system equipped with a Waters 2795 Seperation Module (Alliance HT), a Waters 2996 Photodiode Array Detector (190-750 nm), a Waters Xbridge C18 column (2.1 x 30 mm, 3.5 µm) and a LCT ESI-Orthogonal Acceleration Time of Flight Mass Spectrometer. Samples were run using 2 mobile phases: A = 1% CH_3CN, 0.1% formic acid in H_2O and B = 1% H_2O, 0.1 % formic acid in CH_3CN, at a flow rate of 0.8 mL/min. Gradient: 0-0.2 min, 5 % B; 0.2-3.2 min, 5 → 95 % B; 3.2-4.2 min 95% B; 4.2-4.4 95% → 5% B; 4.4-6.2 min: 5% B. Data was processed using Waters MassLynx Mass Spectrometry Software 4.1 (deconvulation with Maxent1 function).

HRMS-measurements(XEVO)

High resolution mass spectra were recorded on a Waters Acquity H-class UPLC with XEVO-G2 XS Q-TOF mass spectrometer equipped with an electrospray ion source in positive mode (source voltage 3.0 kV, desolvation gas flow: 900 L/h, temperature 60 ºC) with resolution R = 22000 (mass range m/z = 50-2000) and 200 pg/uL Leu-Enk (m/z = 556.2771) as a lock mass. Samples were run using 2 mobile phases: A = 0.1% formic acid in water and B = 0.1% formic acid in CH3CN on a Waters Acquity UPLC BEH C18 column (2.1 × 50 mm, 1.7 µm); flow rate = 0.6 mL min−1, runtime = 3.00 min, column T = 60 °C, mass detection: 50–1500 Da. Data processing was performed using Waters MassLynx Mass Spectrometry Software 4.1.

Program A: Gradient: 0 – 4.15 min: 2% B; 4.15 – 5.85 min: 2% → 100% B; 5.85 – 6.05: 100% B; 6.05 – 7.00 min: 100% → 2% B. (The first 4 minutes were diverted to waste to get rid of salts)

Program B: Gradient: 0 – 0.15 min: 2% B; 0.15 – 1.85 min: 2% → 100% B; 1.85 – 2.05: 100% B; 2.05 – 2.10 min: 100% → 2% B; 2.10–3.00 min: 100% B.

HPLC purification

HPLC purifications were carried out on a Waters HPLC equipped with a Waters 2489 UV/Vis detector, Waters fraction collector III and Waters XBridge prep C8 OBD column (250 × 22 mm, 10 µm). Flow rate = 37.5 mL min^{-1}. Mobile phase: A = H_2O, B = CH_3CN and C = 1% TFA in H_2O. Gradient: 0-5 min: 90% A, 5% B, 5%C; 5-7 min :5 → 20% B, 5% C; 7-18 min: 20 → 45% B, 5% C; 18-18.5 min: 45 → 95% B, 5% C; 18.5-21.6 min: 95% B, 5% C; 21.6-25 min: 95% → 5% B, 5% C.

Solid Phase Peptide Synthesis

Synthesis of the C-terminal domain

SPPS was carried out on a Syro II MultiSyntech Automated Peptide synthesizer using standard 9-fluorenylmethoxycarbonyl (Fmoc)-based solid phase peptide chemistry at 25 µmol scale. The coupling condition is fourfold excess of amino acids relative to pre-loaded Fmoc-L-Gly-PEG-PS resin (0.18 mmol/g) or Fmoc-L-Gly-Trityl-TentaGel resin (0.18 mmol/g). Double couplings were performed in NMP for 45 min with PyBOP (4 eq.) and DiPEA (8 eq.). The following protected pseudoproline building blocks were incorporated during the synthesis as indicated in Figure 1: Fmoc-L-Gln(Trt)-L-Thr(psiMe,Mepro)-OH, Fmoc-L-Val-L-Ser(psiMe,Mepro)-OH and Fmoc-L-Leu-L-Ser(psiMe,Mepro)-OH. All amino acid and dipeptide building blocks were dried overnight under high vacuum before usage. Fmoc was removed with 20% piperidine in NMP (3×1.2 mL, 2×2 and 1×5 min). After the synthesis was finished, the resin was washed with Et_2O, dried under high vacuum and stored for further use. An initial trial cleavage using a small amount of the resin was performed to access the quality of the crude proteins (analysis data can be found in Figure S7 & S10).

Synthesis of the N-terminal domain

SPPS was carried out on an Intavis MultiPep CF peptide synthesizer using standard 9-fluorenylmethoxycarbonyl (Fmoc)-based solid phase peptide chemistry on a 50 μmol scale and the synthesis was optimized by following the Fmoc cleavage with the Intavis. The coupling condition is fourfold excess of amino acids relative to pre-loaded Fmoc-L-Gln(Trt)-Trityl-TentaGel resin (0.16 mmol/g). Double couplings were performed in NMP for 45 min with PyBOP (4 eq.) and DiPEA (16 eq.). The following protected pseudoproline building blocks were incorporated during the synthesis as indicated in Figure 1: Fmoc-L-Gly-L-(Dmb)Gly-OH, Fmoc-L-Val-L-Ser(psiMe,Mepro)-OH and Fmoc-L-Ser(tBu)-L-Ser(psiMe,Mepro)-OH. All amino acid and dipeptide building blocks were dried overnight under high vacuum before usage. Fmoc was removed with 20% piperidine (3× 4 min or 15 min) at r.t. After the synthesis was finished, the resin was washed with Et_2O, dried under high vacuum and stored for further use. An initial trial cleavage using a small amount of the resin was performed to access the quality of the crude proteins (analysis data can be found in Figure S8 & S12).

C-terminal part of mISG15 (**4**)

The protein was cleaved from the resin (25 μmol) and deprotected by treatment with TFA/H_2O/TIPS/Phenol (90:5:2.5:2.5 v/v/v/v) for 3.5 h at r.t., followed by precipitation with cold Et_2O/n-pentane (3:1; v/v). The precipitated protein was further washed with Et_2O (2×). The pellet was dissolved in H_2O/CH_3CN/formic acid (65/25/10; v/v/v) and lyophilized. The protein was purified by HPLC. Yield: 38.1 mg, 4.1 μmol, 16.4%. LC-MS: Rt = 1.96 min; MS: ESI MS+: calculated: 9308.49 $[M]^+$, found (deconv) 9310.0.

N-terminal domain of mISG15 (**9**)

The resin (**5**, 25 μmol) was rinsed two times with DCM and then treated with 20% hexafluoroisopropanol (HFIP) in DCM (v/v; 2× 20 min). The combined filtrates were concentrated under reduced pressure and the residue was co-evaporated with DCE (3×) and lyophilized overnight to give crude protein **6**. The protected protein (**6**) was dissolved in chloroform (30 mL) and the reaction mixture was cooled to -18 ºC. The 2,2,2-trifluroethanol (10 mL) was added, followed by the addition of HCl·H-Gly-S$(CH_2)_2CO_2$Me (**7**) (3 eq.), EDC (3.0 eq.) and HOBt (3.0 eq.) and the reaction was stirred overnight. An initial trial cleavage using a small amount of the reaction mixture was performed to check the reaction and LC/MS data shows a good conversion of the starting material. The crude protein **8** was concentrated under reduced pressure and treated with TFA/H_2O/phenol/TIPS (90.5/5/2.5/2; v/v/v/v;) for 3.5 h at r.t. The protein was precipitated from ice-cold Et_2O/n-pentane (3/1; v/v). The solution was centrifuged and the pellet was further washed with Et_2O (3×). The pellet was dissolved in H_2O/CH_3CN/formic acid (65/25/10; v/v/v) and lyophilized. The protein was purified by HPLC. Yield: 19.5 mg, 2.4 μmol, 9.6%. LC-MS: Rt = 2.18 min; MS: ESI MS+: calculated: 8012.9 $[M]^+$, found (deconv) 8014.0.

C-terminal part of mISG15 probe (**14**)

Product **10** was obtained from SPPS and protein **14** was prepared in the same fashion as described above for the preparation of protein **9** on 25 μmol scale. The protein was purified by HPLC. Yield: 30.1 mg, 3.2 μmol, 12.8%. LC-MS: Rt = 2.18 min; MS: ESI MS+ (amu) calculated: 9288.49 $[M]^+$, found (deconv) 9289.0.

N-terminal part of mISG15 probe (**18**)

Product **15** was obtained from SPPS and protein **18** was prepared in the same fashion as described above for the preparation of protein **9** on 25 μmol scale. The protein was purified by HPLC. Yield: 16.1 mg, 4.1 μmol, 7.6 %. LC-MS: Rt = 2.24 min; MS: ESI MS+ (amu) calculated: 8369.06 $[M]^+$, found (deconv) 8370.0.

Native chemical ligation

Full-length mISG15 (**1**)

The mISG15-thioester peptide (**9**) (10.5 mg) was first dissolved in 200 μL of buffer containing 8.0 M Gnd·HCl and 0.2 M Na_2HPO_4 at pH 7.5. The C-terminal peptide (**4**) (15.2 mg) was dissolved in 300 μL of the same buffer, followed by the addition of 100 μL MPAA (1M) and this solution was pre-incubated for 5 minutes. These two

solutions were properly mixed and the pH of the reaction mixture was adjusted to pH 7.5 by the addition of 10% Na_2CO_3 solution. The reaction mixture was shaken overnight at 37 °C. LC-MS analysis showed complete consumption of the peptide **9**. The crude product was purified by Waters HPLC. Pure fractions were combined and lyophilized. The obtained protein was further dissolved in H_2O/CH_3CN/formic acid (65/25/10; v/v/v) and lyophilized three times. The desired product was obtained as a white solid. The protein was further purified by gel filtration on a Sephadex S75 16/60 PG column (GE Healthcare), using a 100 mM MES, 200 mM NaCl, 10% Glycerol buffer at pH 6.5. The pure protein was kept in the buffer for further experiments. Protein concentration was determined (method is described in biochemical experiments part) and then the total amount of protein was calculated. (Yield: 1.2 mg; 0.070 µmol; 5.3%.). LC-MS: Rt = 2.26 min; MS: ESI MS+ (amu) calculated: 17202.4 $[M]^+$, found (deconv) 17206.0. LC-MS (XEVO): Rt = 5.45 min, MS: ESI MS+ (amu) calculated: 17202.4 $[M]^+$, found (deconv) 17203.0.

Rho-mISG15-PA probe (**2**)

This probe was prepared in the same way as described for the synthesis of full length mISG15 (**1**), using Rho-mISG15-thioester peptide (**18**, 5.2 mg) and C-terminal peptide (**14**, 11.1 mg). After purification by Waters HPLC, a solid was obtained. The product was further purified by gel filtration and kept in buffer for further experiments. Protein concentration was determined (method is described in biochemical experiments part) and then the total amount of protein was calculated. (Yield: 0.21 mg, 0.012 µmol, 1.9%). LC-MS: Rt = 2.13 min; MS: ESI MS+ (amu) calculated: 17538.6 $[M]^+$, found (deconv) 17541.0. LC-MS (XEVO): Rt = 5.45 min, MS: ESI MS+ (amu) calculated: 17538.6 $[M]^+$, found (deconv) 17540.0.

Biochemical experiments:

Cell culture

HEK293T and NMuMG cells were obtained from the ATCC and cultured under standard conditions in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 8% FCS (biowest) at 37 °C and 5% CO_2.

Protein concentration measurements

The protein concentration of synthetic mISG15 and its probe was quantified in SDS-PAGE gel bands using ImageJ software (NIH), with expressed ubiquitin (a range from 0.1 µg/µL to 1 µg/µL) being used for the standard curve.

E1-mediated native chemical ligation of mISG15 withthiolysine modified peptide

The expressed mISG15 wildtype protein was purified from *E.coli* as described previously.[1] 2 µg of mISG15 and 1 µg of ISG15 E1/UBE1L (E-309, BostonBiochem) were reacted in 50 µL native chemical ligation buffer (100 mM MESNa, 20 mM $MgCl_2$, 20 mM ATP, 50 µM TAMRA-thiolysine-glycine peptide, pH 8.0) at 37 °C for 18 hours. Next, 10 µL of the crude ligation mixture was added to LDS (lithium dodecyl sulfate) sample buffer (Invitrogen Life Technologies, Carlsbad, CA, USA) containing 2.5% β-mercaptoethanol, followed by boiling for 7 minutes. Samples were resolved on a 12% Bis-Tris NuPAGE Gel (Invitrogen) using MES buffer (Invitrogen Life Technologies, Carlsbad, CA, USA). Labelled mISG15 was visualized by in-gel fluorescence using a Typhoon FLA 9500 imaging system (GE Healthcare Life Sciences) (TAMRA channel for labelled mISG15, Cy5 channel for protein marker). The gel was then stained with InstantBlue (Expedeon). For LC-MS detection, the crude ligation mixture was pre-treated with TCEP (final concentration 2 mM) for 10 minutes.

Labelling of overexpressed mUSP18 in HEK 293T cell lysates

HEK293T cells were transfected with plasmids overexpressing Flag-mUSP18 wildtype (Flag-mUSP18 wt) and Flag-mUSP18 catalytic cysteine-to-alanine mutant (Flag-mUSP18 C61A)[2] separately in 6-well plates using polyethylenimine (PEI, Polysciences, lnc.) according to manufacturer's instructions. After 24 hours, cells were harvested and resuspended in HR buffer (50 mM Tris-HCl, 5 mM $MgCl_2$, 250 mM sucrose, 1 mM DTT and a Protease inhibitor tablet (Roche), pH7.4). Cell lysis was achieved by sonication (Bioruptor, Diagenode, high intensity for 10 minutes with an ON/OFF cycle of 30 seconds) at 4°C. After a centrifugation step (14,000 Rpm for 15 minutes) to remove cell debris, the protein concentration of the supernatant was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific) by measuring the absorbance at 280 nm. 40 µg of each protein sample was treated with Rhodamine-mISG15-propargylamide probe **2** (final conc. 500 nM) at 37 °C for indicated incubation time. Reactions were stopped by the addition of LDS (lithium dodecyl sulfate) sample buffer (Invitrogen Life Technologies, Carlsbad, CA, USA) containing 2.5% β-mercaptoethanol, followed by boiling for 7 minutes. Samples were resolved on a 4-12% Bis-Tris NuPAGE Gel (Invitrogen) using MOPS buffer (Invitrogen Life Technologies, Carlsbad, CA, USA). Labelled enzymes were visualized by in-gel fluorescence using a Typhoon FLA 9500 imaging system (GE Healthcare Life Sciences) (Rhodamine channel for probe, Cy5 channel for protein marker). The gel was then transferred to Nitrocellulose membranes and immunoblotted using mouse anti-flag (M2, Sigma-Aldrich) and mouse anti-GAPDH (1D4, Enzo Lifesciences). The following fluorescent secondary antibody (from LICOR) was used: anti-mouse-680. Immunoblots were visualized using a LICOR Odyssey system.

Labelling of endogenous enzymes in mouse NMuMG cell lysates

Mouse NMuMG cells were stimulated with 300 U/mL mouse IFN-β[3] for 24 hours to induce ISGylation. Cell lysates were prepared in the same way as described above. 40 µg of each protein sample was treated with Rhodamine-mISG15-propargylamide probe **2** (final conc. 500 nM) at 37 °C for one hour. Reactions were stopped by the addition of LDS (lithium dodecyl sulfate) sample buffer (Invitrogen Life Technologies, Carlsbad, CA, USA) containing 2.5% β-mercaptoethanol, followed by boiling for 7 minutes. Samples were resolved on a 4-12% Bis-Tris NuPAGE Gel (Invitrogen) using MOPS buffer (Invitrogen Life Technologies, Carlsbad, CA, USA). Labelled enzymes were visualized by in-gel fluorescence using a Typhoon FLA 9500 imaging system (GE Healthcare Life Sciences) (Rhodamine channel for probe, Cy5 channel for protein marker). The gel was then transferred to Nitrocellulose membranes and immunoblotted using rabbit anti-mISG15[4], rabbit anti-mUSP18,[5] mouse anti-GAPDH (1D4, Enzo Lifesciences) or mouse anti-*β*-actin (clone AC-15, Sigma-Aldrich). The following fluorescent secondary antibodies were used: anti-mouse-680 and anti-rabbit-800. Immunoblots were visualized using a LICOR Odyssey system.

References:

1. A. Basters, P. P. Geurink, F. El Oualid, L. Ketscher, M. S. Casutt, E. Krause, H. Ovaa, K. P. Knobeloch and G. Fritz, *The FEBS journal*, 2014, **281**, 1918-1928.
2. L. Ketscher, R. Hannss, D. J. Morales, A. Basters, S. Guerra, T. Goldmann, A. Hausmann, M. Prinz, R. Naumann, A. Pekosz, O. Utermohlen, D. J. Lenschow and K. P. Knobeloch, *Proc Natl Acad Sci U S A*, 2015, **112**, 1577-1582.
3. A. Basters, P. P. Geurink, A. Rocker, K. F. Witting, R. Tadayon, S. Hess, M. S. Semrau, P. Storici, H. Ovaa, K. P. Knobeloch and G. Fritz, *Nat Struct Mol Biol*, 2017, **24**, 270-278.
4. A. Osiak, O. Utermohlen, S. Niendorf, I. Horak and K. P. Knobeloch, *Mol Cell Biol*, 2005, **25**, 6338-6345.
5. K. N. Swatek, M. Aumayr, J. N. Pruneda, L. J. Visser, S. Berryman, A. F. Kueck, P. P. Geurink, H. Ovaa, F. J. M. van Kuppeveld, T. J. Tuthill, T. Skern and D. Komander, 2018, **115**, 2371-2376.