SUPPORTING INFORMATION PARAGRAPH for

Semisynthesis of dimeric proteins by Expressed Protein Ligation

by Barbara Ziaco[a], Soccorsa Pensato[a], Luca D. D'Andrea[b], Ettore Benedetti[a], Alessandra Romanelli[a]*

General experimental

All chemicals and solvents are commercially available (Novabiochem, Sigma-Aldrich, LabScan) and were used without further purification. The pTXB1 vector and the chitin resin were purchased at New England Biolabs; pTrcHisA and Ni-NTA resin were from Invitrogen. Column chromatography was performed on Fluka silica gel 60 (size: 0.04-0.063mm). ^{1}H- and ^{13}C-NMR spectra were recorded on a Varian Innova instrument (600 MHz) at room temperature. All chemical shifts are expressed in ppm with respect to the signals of the residual protonated solvents ($CDCl_3$ or DMSO d_6). LCMS analyses of the proteins were run on a Thermo Finnigan instrument equipped with a LCQ Deca XPMax ES source using a Phenomenex Jupiter 5μ C4 300Å, 250x2.00 mm column. LCMS analyses of the linker were run on a Thermo Finnigan instrument equipped with a MSQ ES source using a Phenomenex Jupiter 5μ C18 300Å, 150x4.6 mm column. Preparative HPLC was performed on a Phenomex Jupiter 10μ Proteo 90Å 250x10.00 mm column.

Synthesis of the linker N-N' bis-cysteinyl- ethylendiamine To a solution of ethylendiamine (30.1μL, 0.45mmol) and DIPEA (485.5 μL, 2.7 mmol) in dry DCM (300 μL) was added a solution of BocCys(Trt)OH (500mg, 1.08 mmol), PyBOP (467 mg, 0.90 mmol) and HOBT (121.4 mg, 0.90 mmol) in dry DCM (2 mL). The reaction was stirred overnight at room temperature. The mixture was extracted with a 5% $NaHCO_3$ aqueous solution. The organic phase was dried over $MgSO_4$ and the solvent was evaporated under reduced pressure. The residue was purified by silica gel flash chromatography (ethyl

acetate/petroleum ether, 7/3 v/v) to afford 427 mg of protected linker N-N' bis[N-tert-butyloxycarbonyl-S-triphenylmethyl-cysteinyl] ethylendiamine as a yellow oil (99% Yield).

The Boc/Trt protected linker (159 mg, 0.16 mmol) was dissolved in CH_2Cl_2/TFA/TIS 50/47/3 v/v/v (4 mL). Deprotection of the Trt and Boc groups was complete after 5 minutes. The solution was concentrated under reduced pressure and diluted with cold diethyl ether. The precipitated crude product N-N' bis-cysteinyl- ethylendiamine was dissolved in water, analyzed by LCMS and purified by preparative HPLC with an increasing gradient of CH_3CN (0.1% TFA), in water (0.1% TFA) from 1 to 70% in 30 minutes.

Yield: quantitative

N-N' bis[N-tert-butyloxycarbonyl-S-triphenylmethyl-cysteinyl] ethylendiamine

ESMS: 1: calculated 952.30 u.m.a., found 952.30;

^{1}H NMR: ($CDCl_3$) δ: 7.45-7.20 (m, 30H Trt); 4.9 (m, 2H Cα); 3.1-3-3 (m, 4H CH_2-N); 2.75 (m, 2H CH_2-S); 2.55 (m, 2H CH_2-S); 1.6 (s, 18H Boc). ^{13}C NMR: ($CDCl_3$) δ 171.7 (C=O-Cα); 144.8 (C=O Boc); 130.0, 128.5, 127.4 (C Trt); 80.7 (C_q Boc); 67.6 (C_q Trt); 54.2 (Cα); 39.7 ($\underline{C}H_2$-N); 28.8 (Boc, CH_2-S).

N-N' bis-cysteinyl- ethylendiamine

ESMS : calculated 267.09, found 265.9

^{1}H NMR (DMSO d_6) δ:8.6 NH; 3.9 (m, 2H Hα); 3.2 (m, 4H CH_2-N); 2.9 (m, 2H CH_2-S). ^{13}C NMR (DMSO d_6) δ: 170.9 (C=O); 58.1 (Cα); 42.4 ($\underline{C}H_2$-N); 29.0 (CH_2-S).

Protein expression and purification

Escherichia coli BL21 (DE3) cells, transformed with the appropriate plasmid, were grown to mid-log phase at 37°C in LB medium. Protein expression was induced with 0.4 mM isopropyl β-D-

thiogalactoside (IPTG) at 37°C for 5 h, after which cells were harvested and lysed by sonication. Protein expression was followed by SDS-PAGE (15%). Protein A-MxeGyrA fusion was purified at room temperature by affinity chromatography on a Ni^{2+} NTA resin. Protein was eluted in 50 mM NaH_2PO_4 , 300 mM NaCl, 250 mM imidazole pH 8.

Protein B-MxeGyrA- fusion was immobilized and purified on chitin resin, following standard protocol.

Synthesis of the homodimer (A-linker-A). Purified protein A-MxeGyrA fusion (150μM) was reacted in Phosphate buffer 50mM, 0.2mM MESNA, 0.3mM EDT, pH 7 with 0.5 (or 3) equivalents of linker. Splicing and ligation reactions were performed simultaneously overnight at room temperature. The reaction leads to the homodimer in a 50% yield. The crude was analysed by LCMS with an increasing gradient of CH_3CN (0.1% TFA), in water (0.1% TFA) from 5 to 70% in 30 minutes. Mass spectrum shows the presence of the homodimer A-linker-A and the protein A (or protein A-linker derivative).

Peptide A sequence: MGGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDRWGSGHIE GR

ES-MS:

Peptide A: calculated 4566.8 Da, found 4567.2 - 1142.6 $[M+4H]^{4+}$; 1523.0 $[M+3H]^{3+}$.

A-linker –A: calculated (- Met) 9364.6 Da, found 9363.0- 1873.6 $[M+5H]^{5+}$; 1561.5 $[M+6H]^{6+}$.

A-linker calculated 4815.8 Da.; found 4815.6 - $[M+4H]^{4+}$ 1204.9 - $[M+3H]^{3+}$1606.2.

Synthesis of the heterodimer (A-linker-B). Purified protein A-MxeGyrA fusion (50μM) was reacted in Phosphate buffer 20mM , 0.18mM MESNA, 0.23mM EDT, pH 7 with a three fold excess of linker. Splicing and ligation reactions were performed simultaneously overnight at room temperature to give the protein A-linker derivative in quantitative yield. The crude was dyalized against deionized water; A-

linker was purified by preparative HPLC with an increasing gradient of CH_3CN (0.1% TFA), in water (0.1% TFA) from 10 to 45% in 38 minutes. Protein B-MxeGyrA fusion was immobilized on the chitin resin, splicing was induced incubating the resin in 20 mM Phosphate buffer containing 300 mM NaCl, 50 mM MESNA and 1 mM EDTA, pH7. The thioester was purified by preparative HPLC with an increasing gradient of CH_3CN (0.1% TFA), in water (0.1% TFA) from 10 to 40% in 30 minutes. Analysis by LCMS of the thioester revealed the presence of peptide B lacking the N-terminus Met and a small fraction lacking both N-terminus Met and Ala. These side products are probably generated during the protein work up, due to the presence of proteases. Furthermore coelution of oxidized peptide B (M+16) with peptide B was also observed. This product is generated during the splicing reaction, if the solvents were not properly degassed. The purified protein A-linker (0.36mM) derivative was reacted with one equivalent of thioester protein B (0.36mM). Reaction was carried out in 20 mM Phosphate buffer containing 0.18 mM MESNA and 0.23 mM EDT, pH 7 overnight. The crude was analyzed by LCMS with an increasing gradient of CH_3CN (0.1% TFA), in water (0.1% TFA) from 5 to 70% in 30 minutes. Protein A-linker was all converted into the homodimer A-linker-B, while the protein B excess was converted into the EDT-thioester.

Peptide B sequence: MASSRVDGGR EFLEGSS

ESMS:

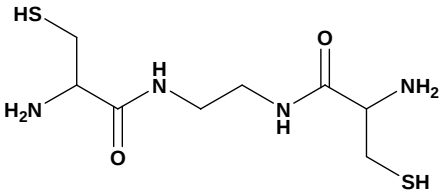
Peptide B: calculated (-Met) 1653.7, found 1653.7 -1653.7 $[M+H]^+$ -$[M+2H]^{2+}$ 827.9.

Peptide B MESNA thioester: calculated (-Met) 1776.8, found 1777.4 -1777,5 $[M+H]^+$- $[M+2H]^{2+}$ 889.7.

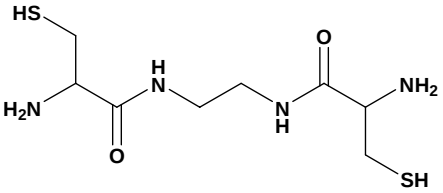
A-linker calculated 4815.8 Da., found 4815.6 - $[M+4H]^{4+}$ 1204.9 - $[M+3H]^{3+}$1606.2.

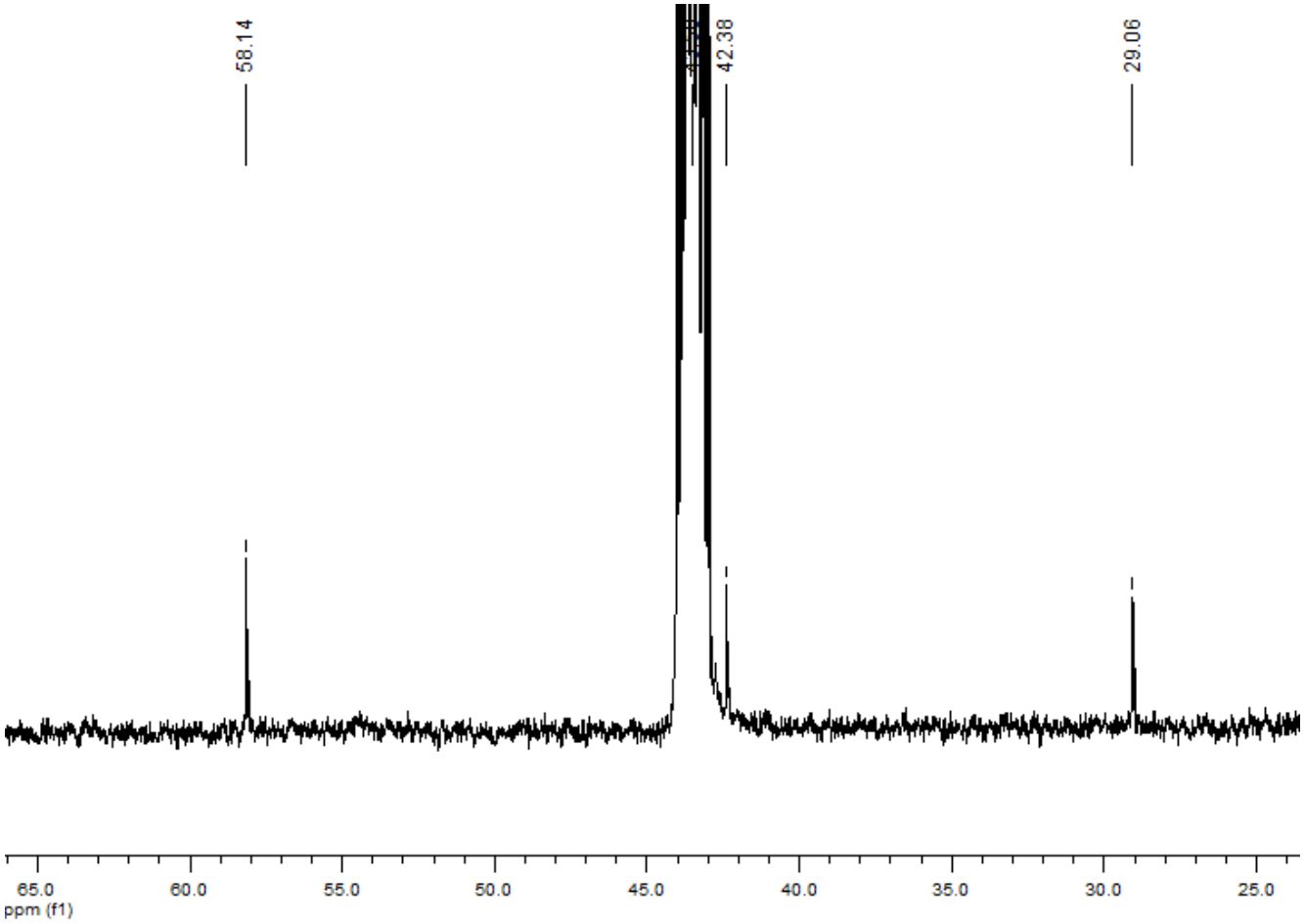
A-linker-B: calculated 6451.5 Da., found 6450.8 - $[M+5H]^{5+}$1291.2 - $[M+4H]^{4+}$ 1613.7 .

^{13}C of the linker (2)

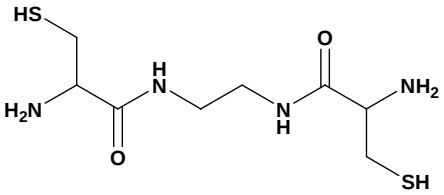


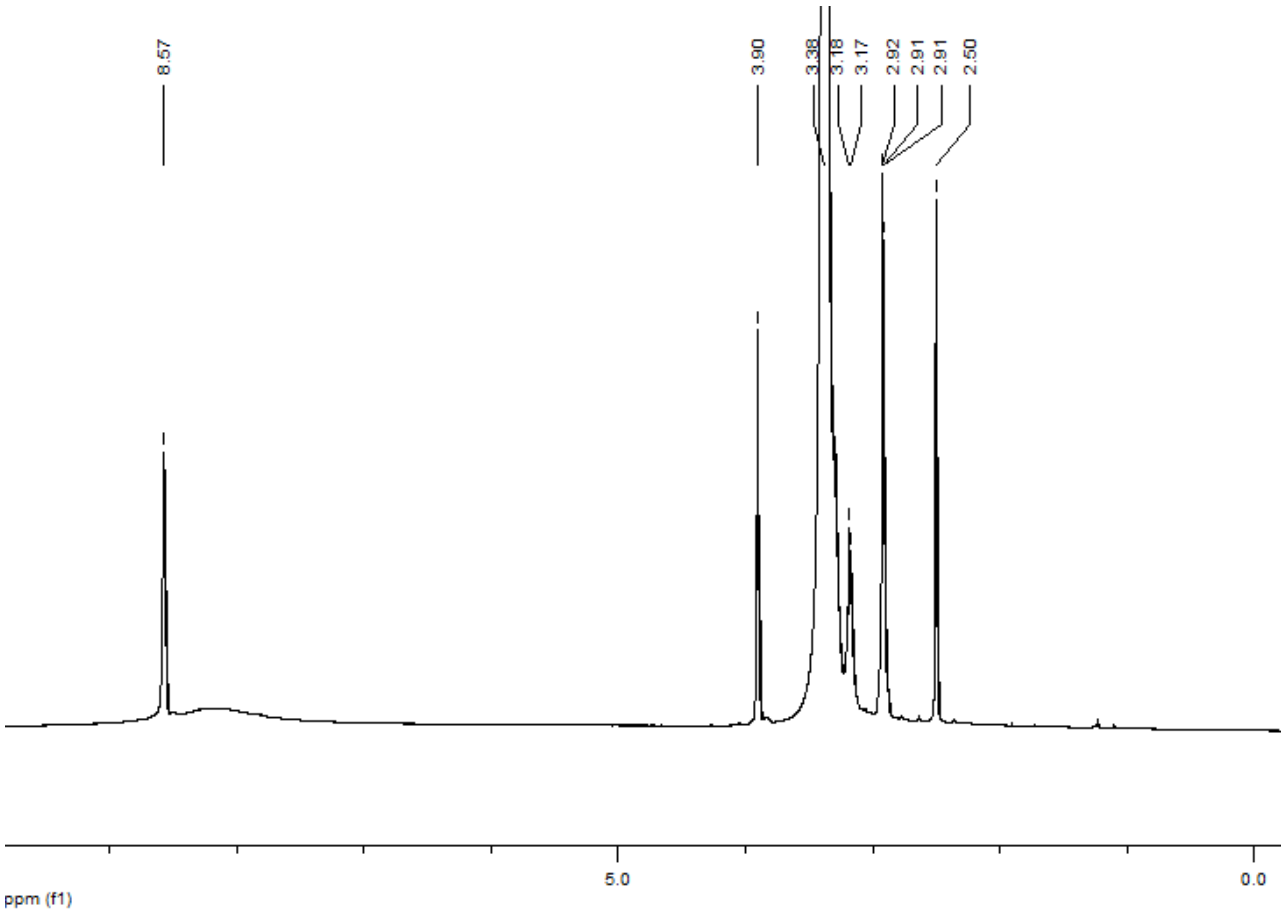
Expanded region of the ^{13}C of the linker (2)



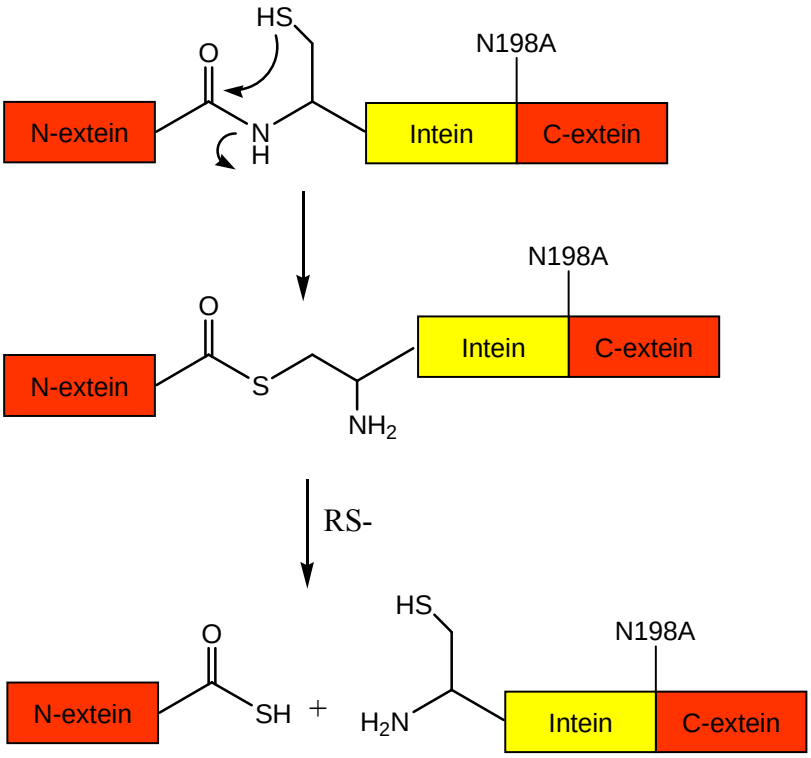


^{1}H NMR of the linker (2)



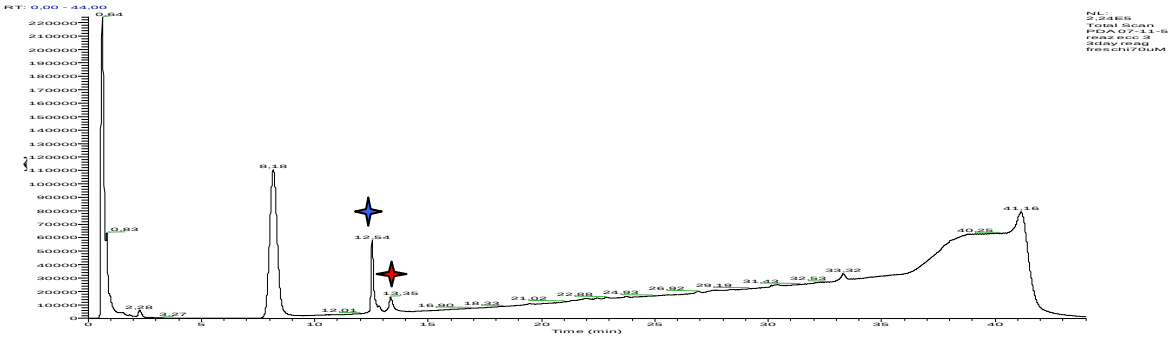


Splicing of the mutated N198A MxeGyrA



HPLC profiles of non optimized reactions for obtaining the homodimer A-linker-A, with different protein concentrations. (✦: A-linker; ✦ : A-linker-A). HPLC conditions are the same described in the **"Synthesis of the homodimer (A-linker-A)" paragraph**

1) protein concentration 70μM



2) protein concentration 110 μM