

Supplementary Materials for

Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds

Authors: Tanaya Bose¹, Gil Fridkin^{1,2}, Chen Davidovich^{1†}, Miri Krupkin^{1‡}, Nikita Dinger^{1§}, Alla Falkovich¹, Yoav Peleg¹, Ilana Agmon^{3,4}, Anat Bashan¹, Ada Yonath^{1*}

Correspondence to: ada.yonath@weizmann.ac.il

This PDF file includes:

Supplementary Text
Figs. S1 to S21
Tables S1

Figure S1: A two-dimensional representation of the RNA constructs designed based on part of *T. thermophilus* 23S rRNA around the PTC. The nucleotides highlighted in yellow were used as the sequences of the RNA constructs tt_A1, tt_P1, tt_A1P1, tt_A2, tt_P2 and tt_A2P2, respectively. Green stretches indicate direct connection between nucleotides and green lines represents positions where the hyper stable CUUCGG hairpin was inserted. rRNA helices and nucleotide numbers (*E. coli* numbering) are designated in blue and black, respectively. A-loop and P-loop are pointed with red arrows.

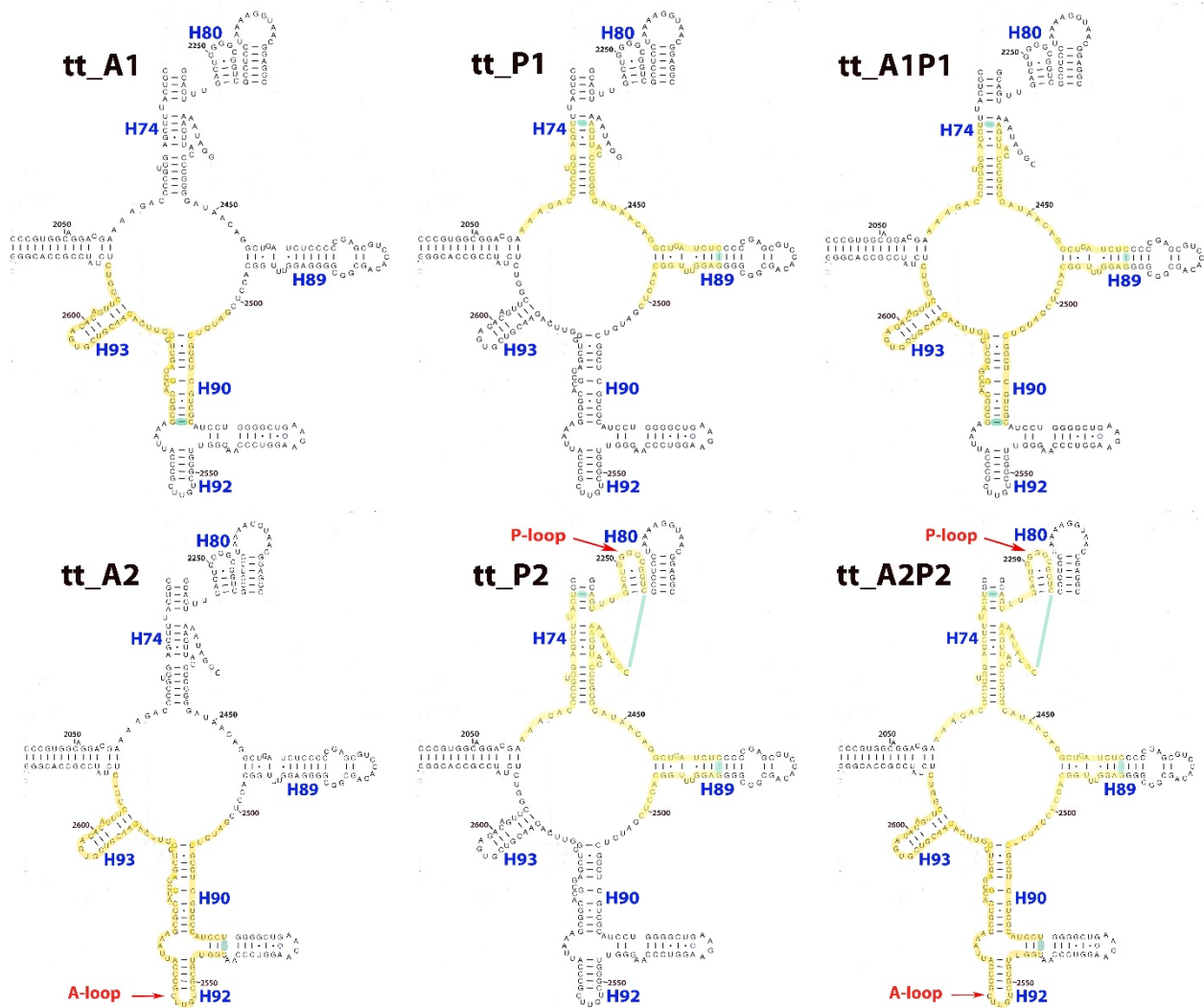


Figure S2: RNA constructs sequence alignment (G- red, A- green, C- orange and U- blue). **A.** The alignment of A-reg short and long sequences. **B.** The alignment of P-reg minimal and extended sequences. GUGA addition to tt_P1c and tt_P2c is marked with a yellow box. **C.** The alignment of long sequences which represent the full proto-ribosome. **D.** The alignment of various P-reg minimal sequences. rRNA sequences were aligned using Clustal omega server (Sievers, Wilm et al. 2011) and imaged using Jalview (Waterhouse, Procter et al. 2009).

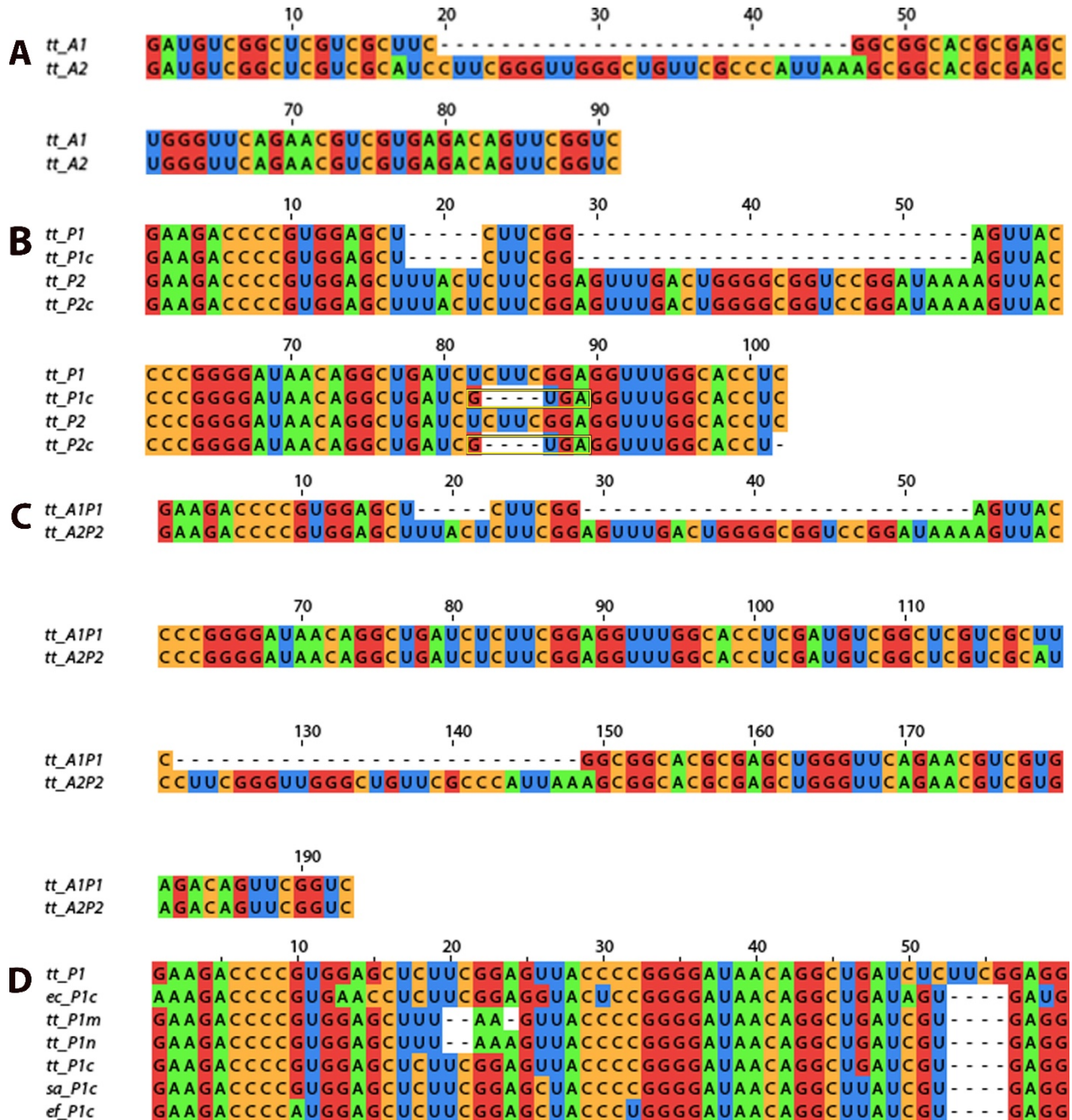


Figure S3: Design schemes of the constructs tt_P1c, tt_P1n and tt_P1m. **A.** Incorporation of a GNRA tetraloop in construct tt_P1 afforded tt_P1c. **B.** Constructs tt_P1n and tt_P1m design. Natural shorter UUAA and UUA loops were added to H74 as shown in the upper image.

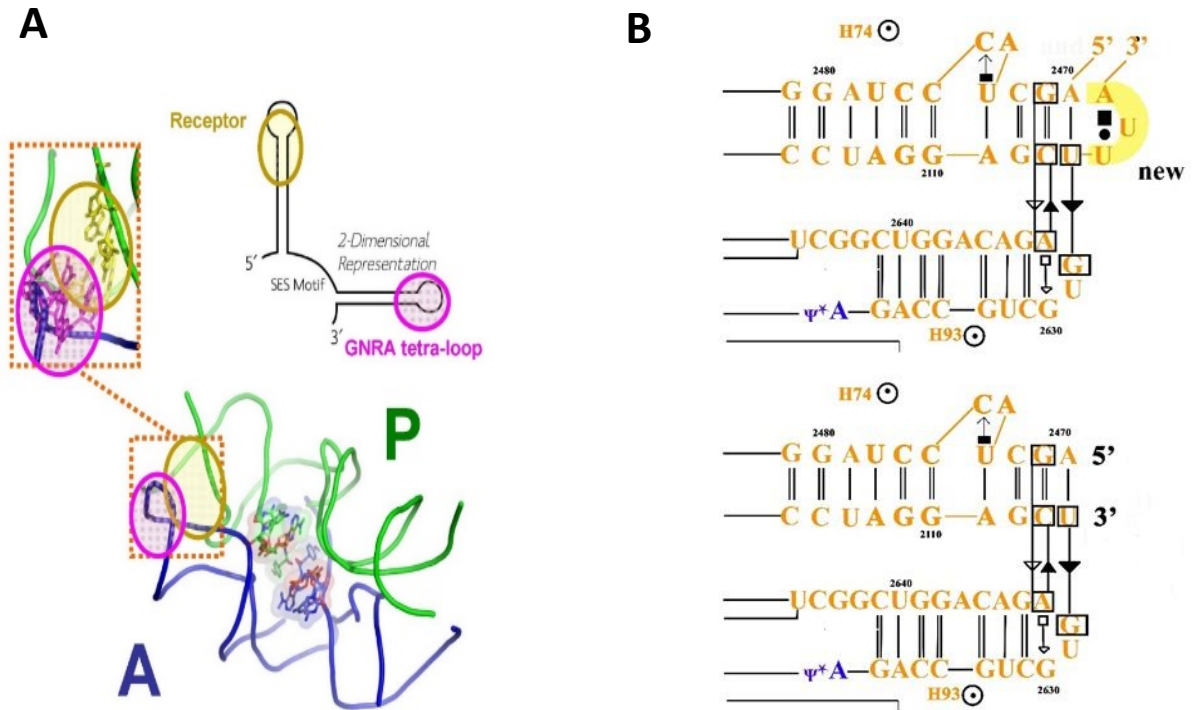


Figure S4: 2D structure predictions of the protoribosome constructs. All predictions were calculated using mfold (66) where ΔG values are refined at 37⁰C.

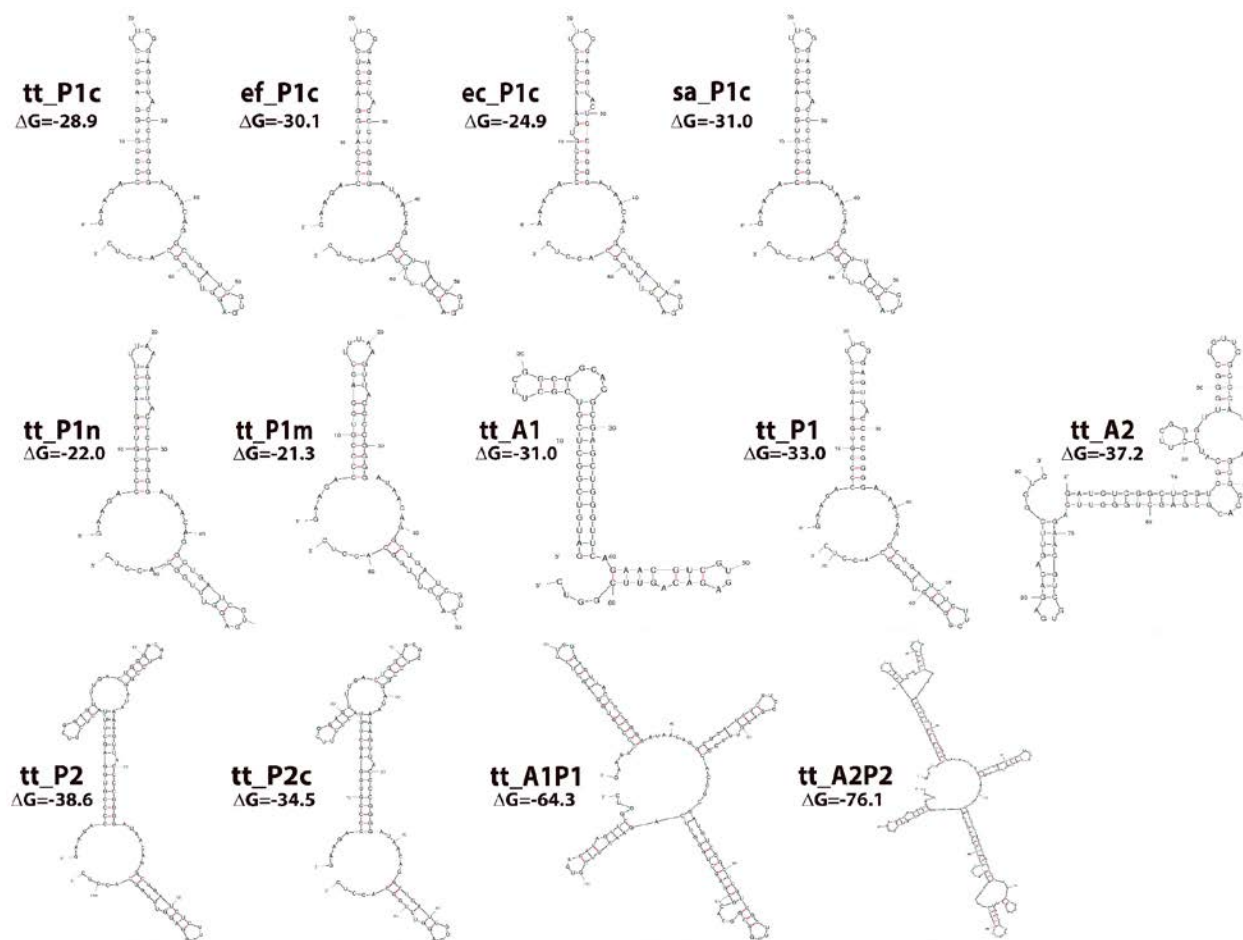


Figure S5: 2D structure predictions of the protoribosome constructs. All predictions were calculated using RNA structure (67) where ΔG values are refined at 37°C.



Figure S6. Denaturing gels of the RNA constructs: A. Purified tt_P1, tt_P1c, tt_P2, tt_P2c, tt_A1P1 and tt_A2P2 RNA constructs (lanes 4-9). **B.** Purchased ec_p1c, hm_p1c, tt_p1m, tt_p1n, tt_A1 and tt_A2 RNA constructs (lanes 4-9). **C.** Purified sa_p1c and ef_p1c RNA constructs (lanes 4-5). **D.** Denaturing gel of the active constructs after fragment reaction of 24h at 37⁰C (lanes 4-8). A single strand RNA (ssRNA) ladder and two DNA ladders (ultralow range (ULR) and 50 base pair (50BP) were used as reference (lanes 1-3). 10% polyacrylamide was used for all gels presented.

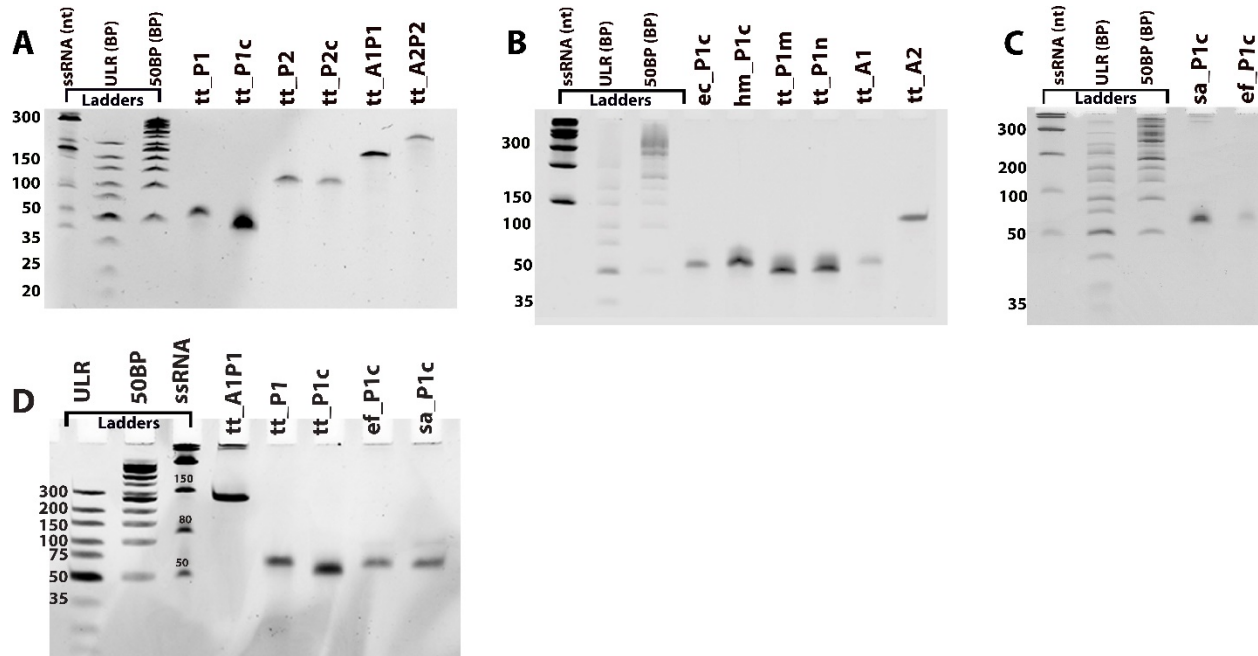


Figure S7: Labeling of RNA oligonucleotides with Fluorecein-5-thiosemicarbazide (FTSC). Sodium periodate was used to cleave the cis diol (3' termini) of RNA into 2',3'-dialdehyde and then, fluorescein-5-thiosemicarbazide, was added to the quenched reaction mixture to label the 3' terminus of RNA through a condensation reaction between the carbazide and the aldehyde.

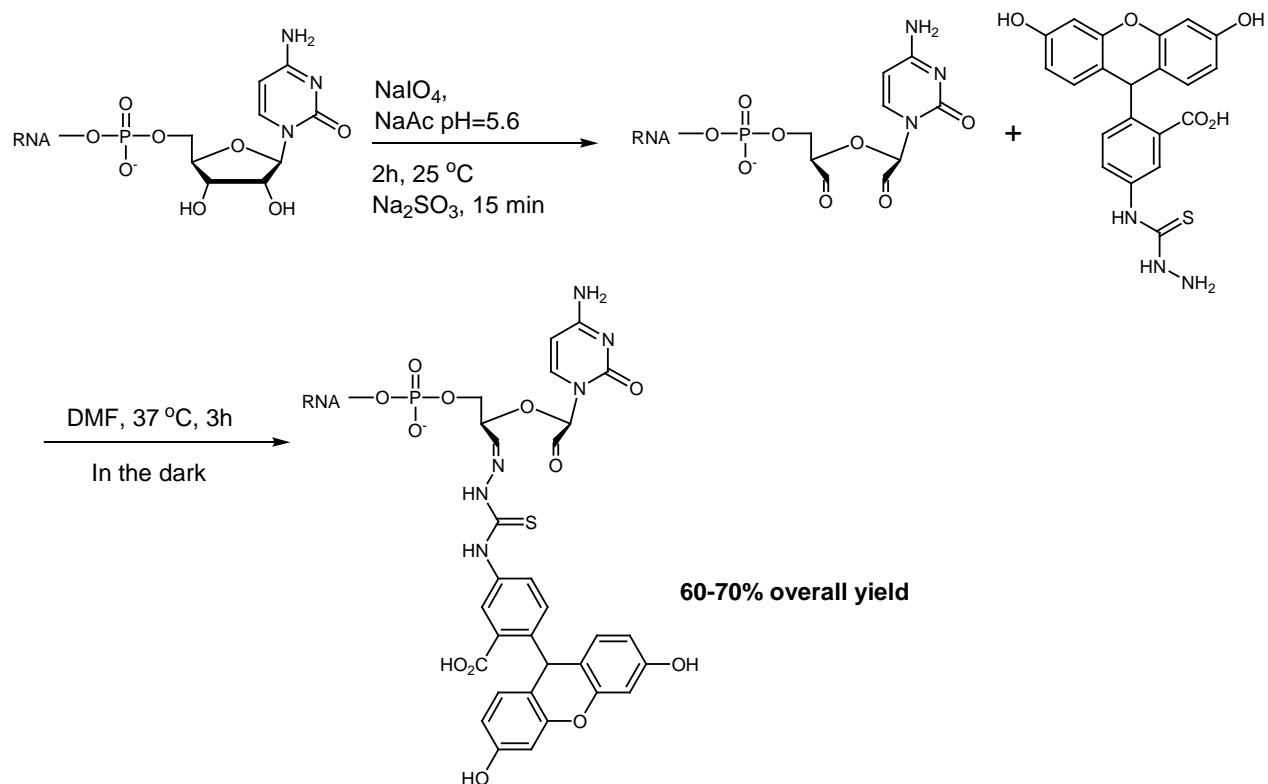
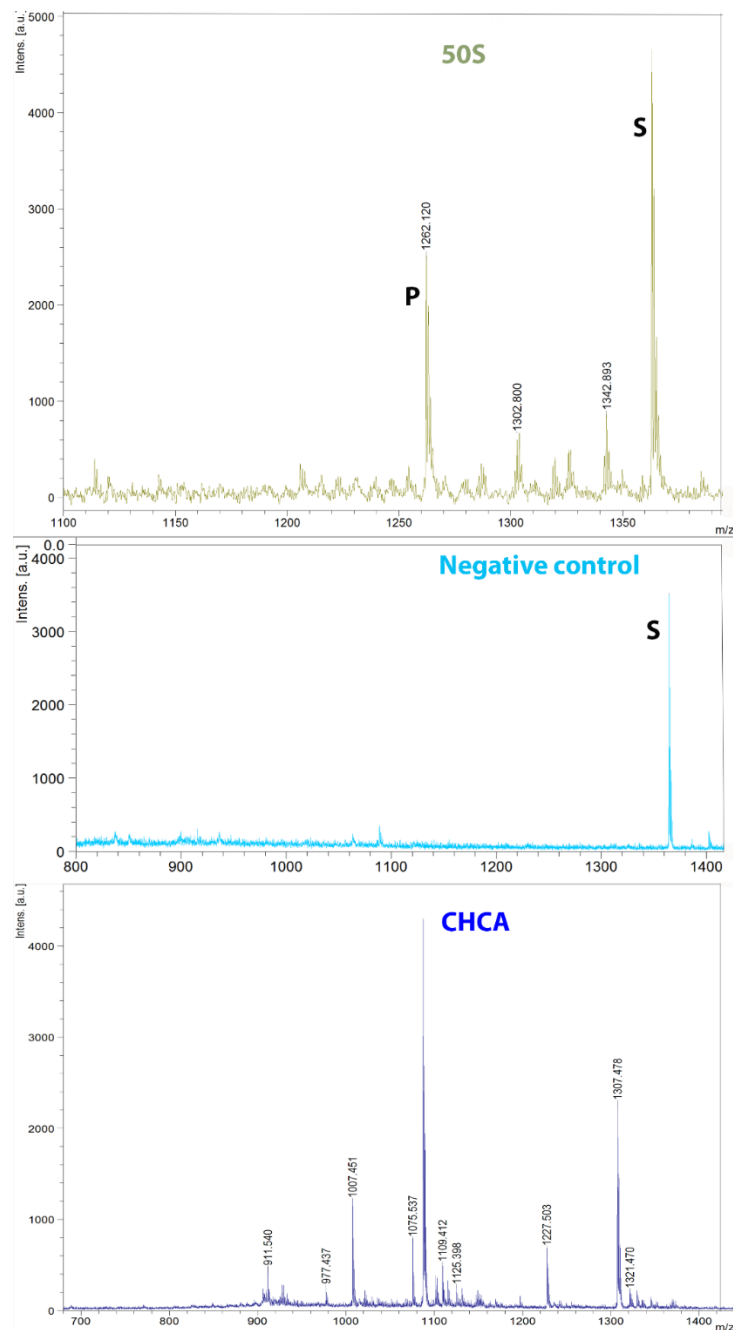
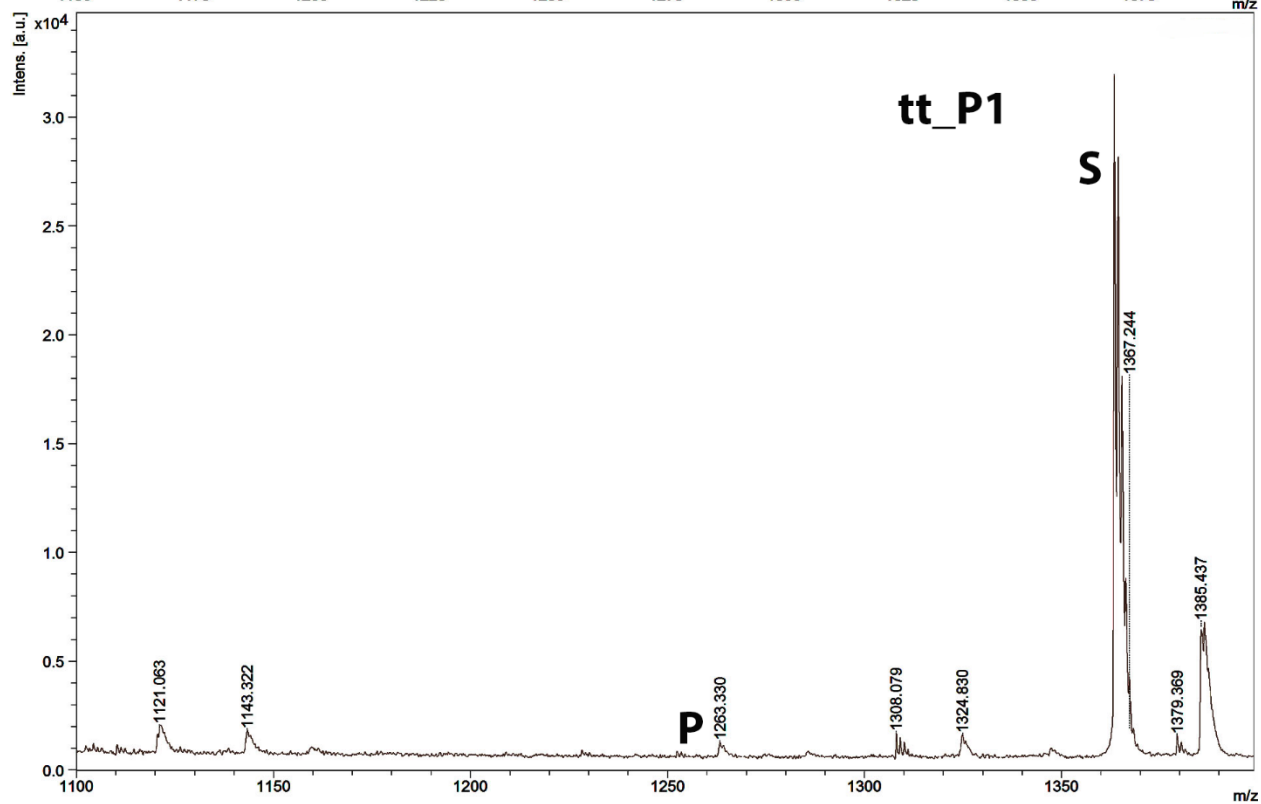
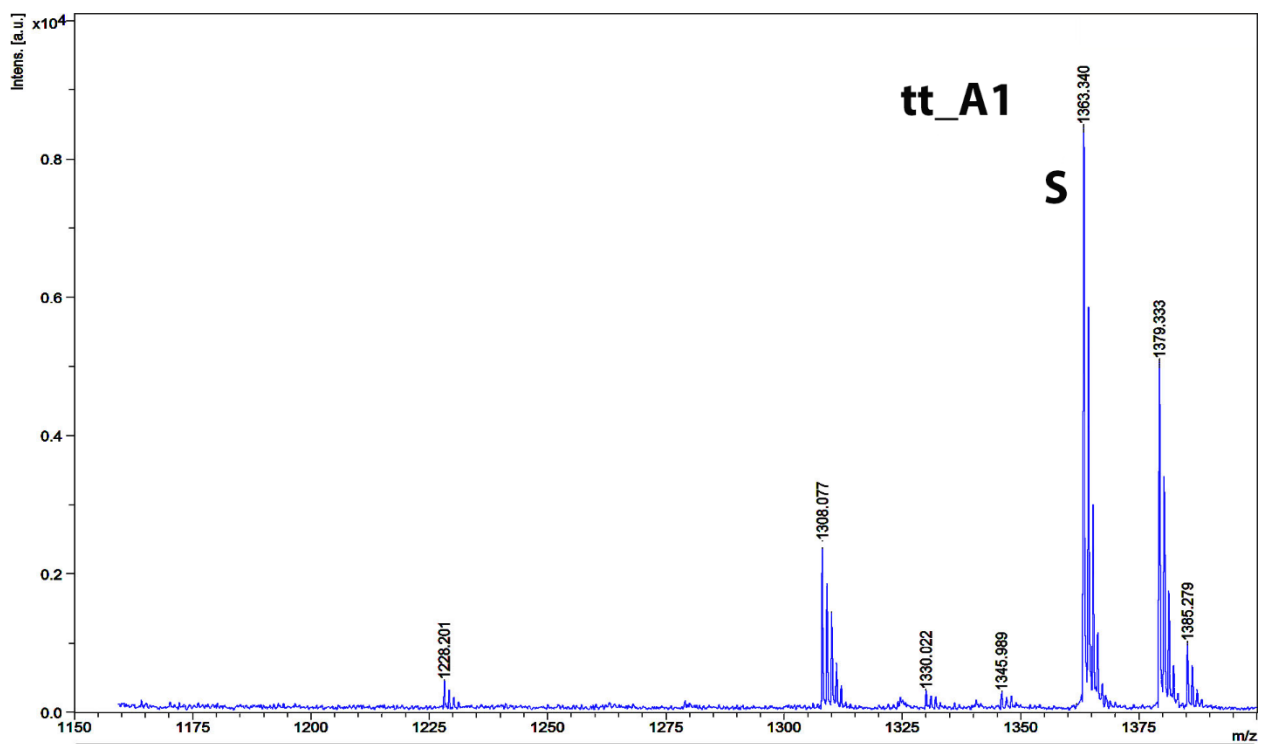


Figure S8: MALDI spectra of the fragment reaction between CCA- pcb and C-Pmn in the presence of the RNA constructs at 37°C. A. Positive (50S) and negative controls (no RNA) and CHCA matrix. **B.** Constructs tt_A1, tt_P1. **C.** Constructs sa_P1c, tt_P1c. **D.** Constructs tt_A1P1, tt_A2P2. **E.** Constructs tt_A2, ec_P1c. **F.** Constructs tt_P1n and tt_P1m. **G.** Constructs tt_P2 and tt_P2c. **H.** Construct cm_tt_P1c. The starting material CCA- pcb and the product C-Pmn-pcb are designated by S and P throughout.

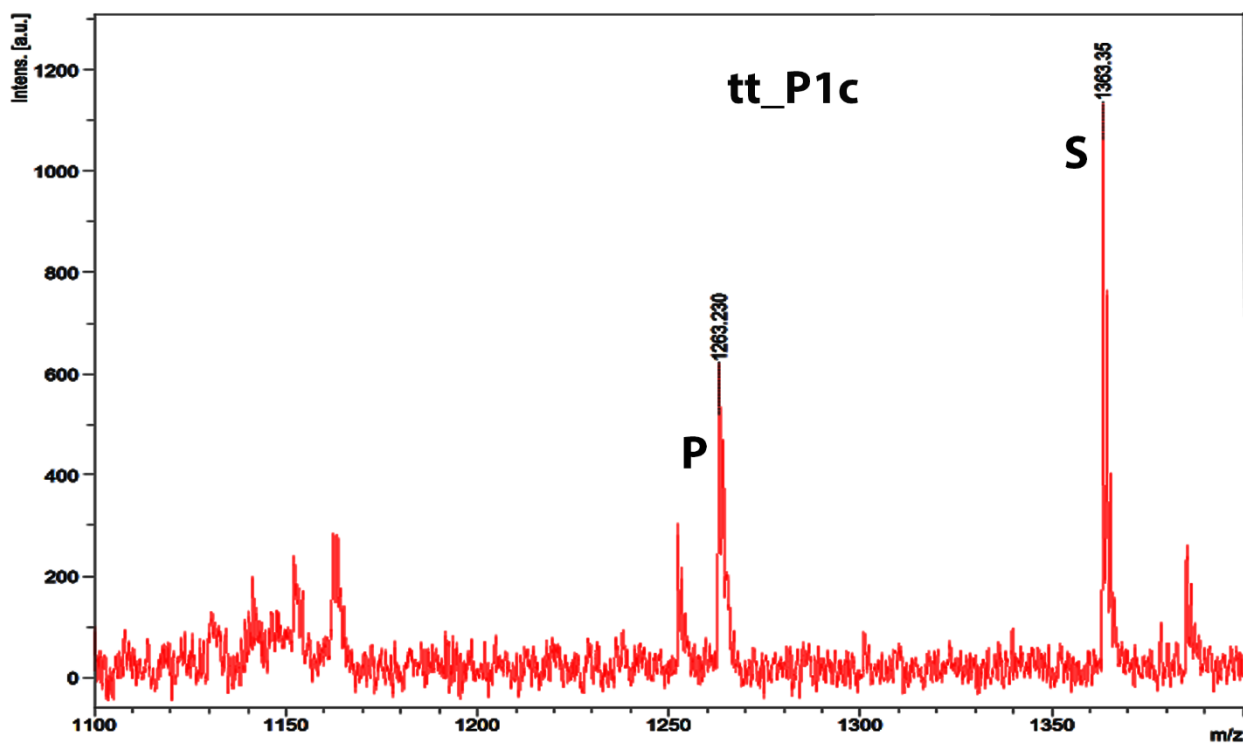
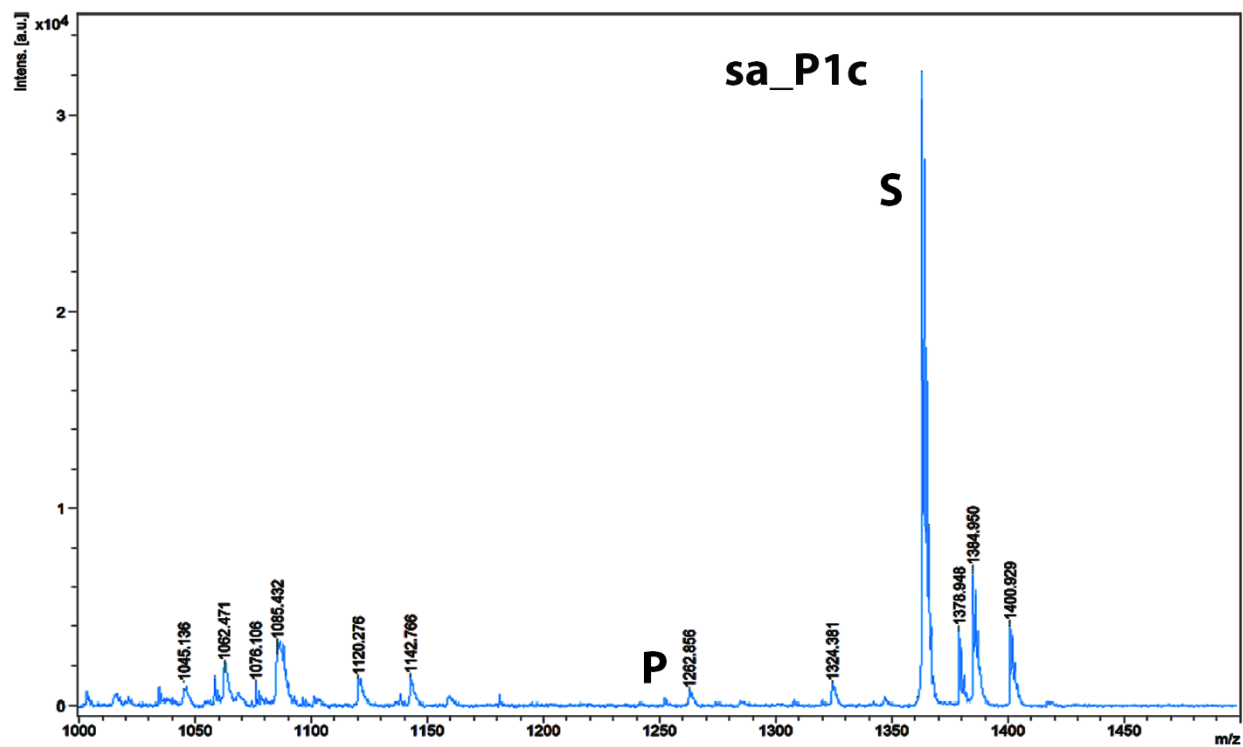
A



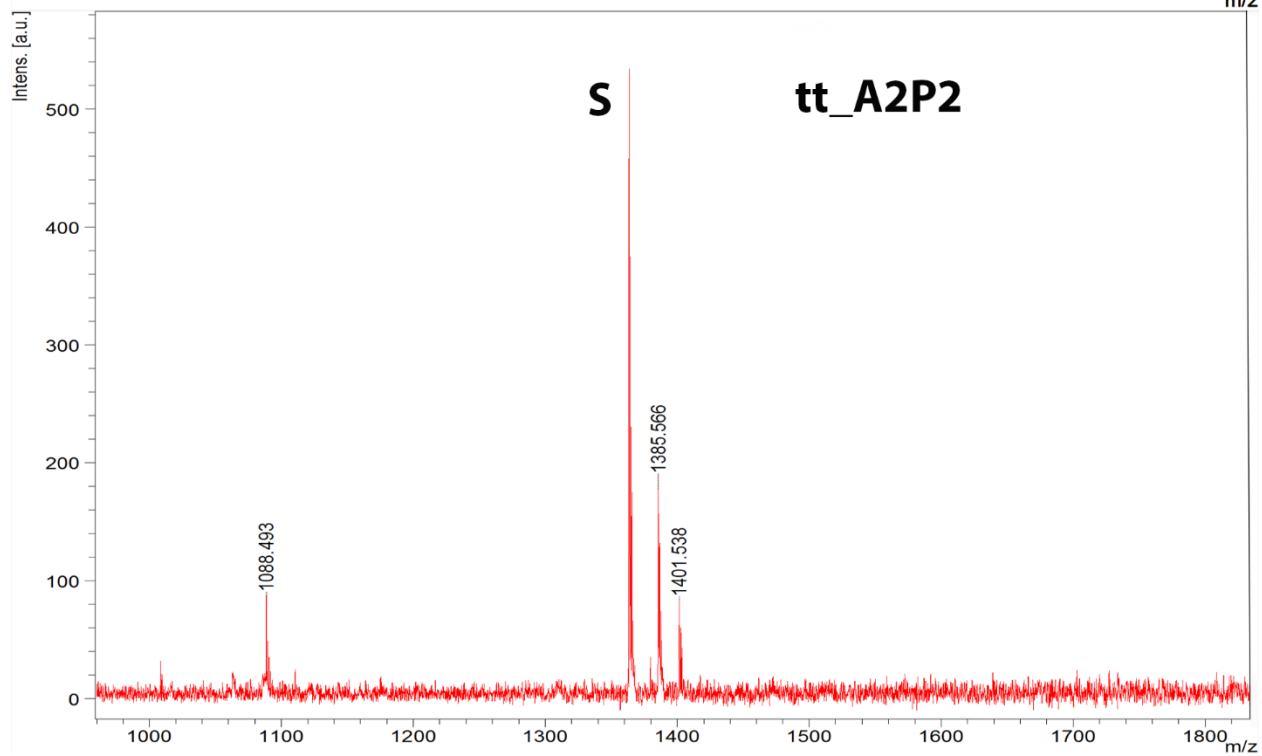
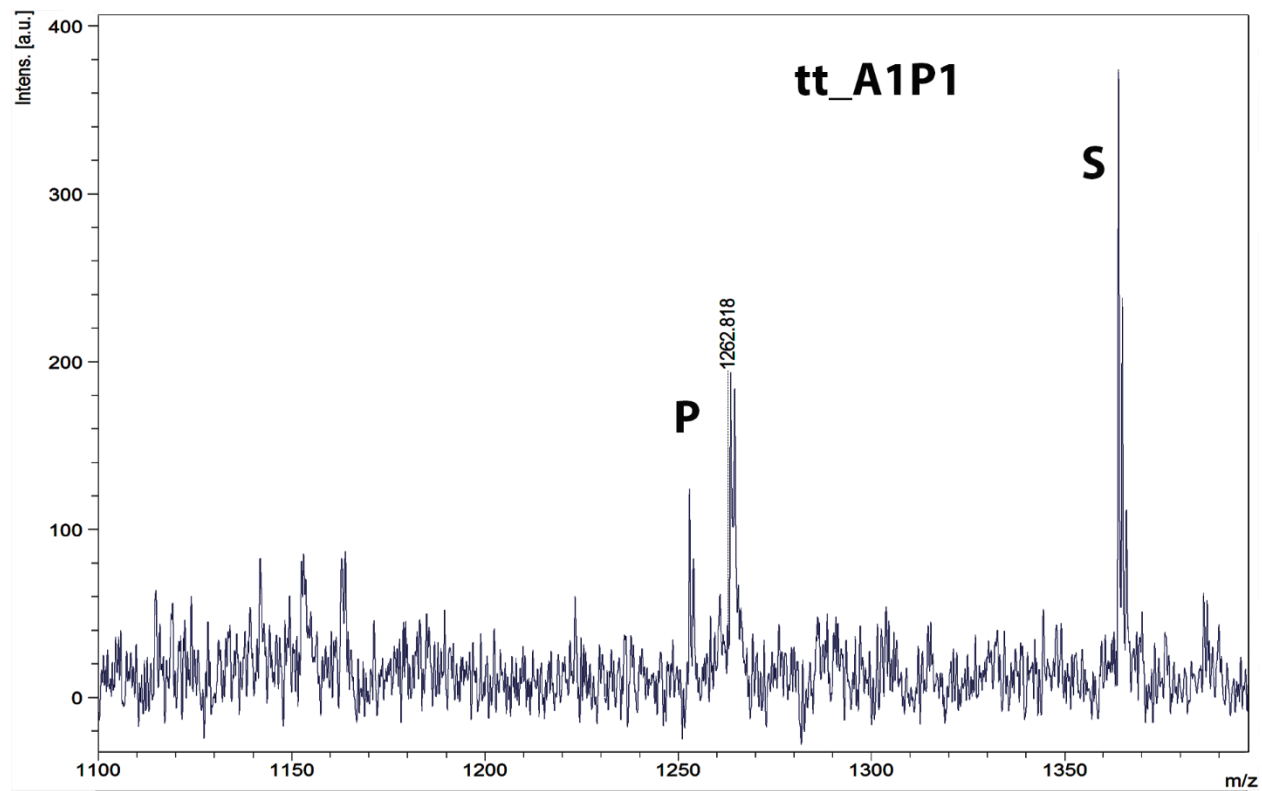
B



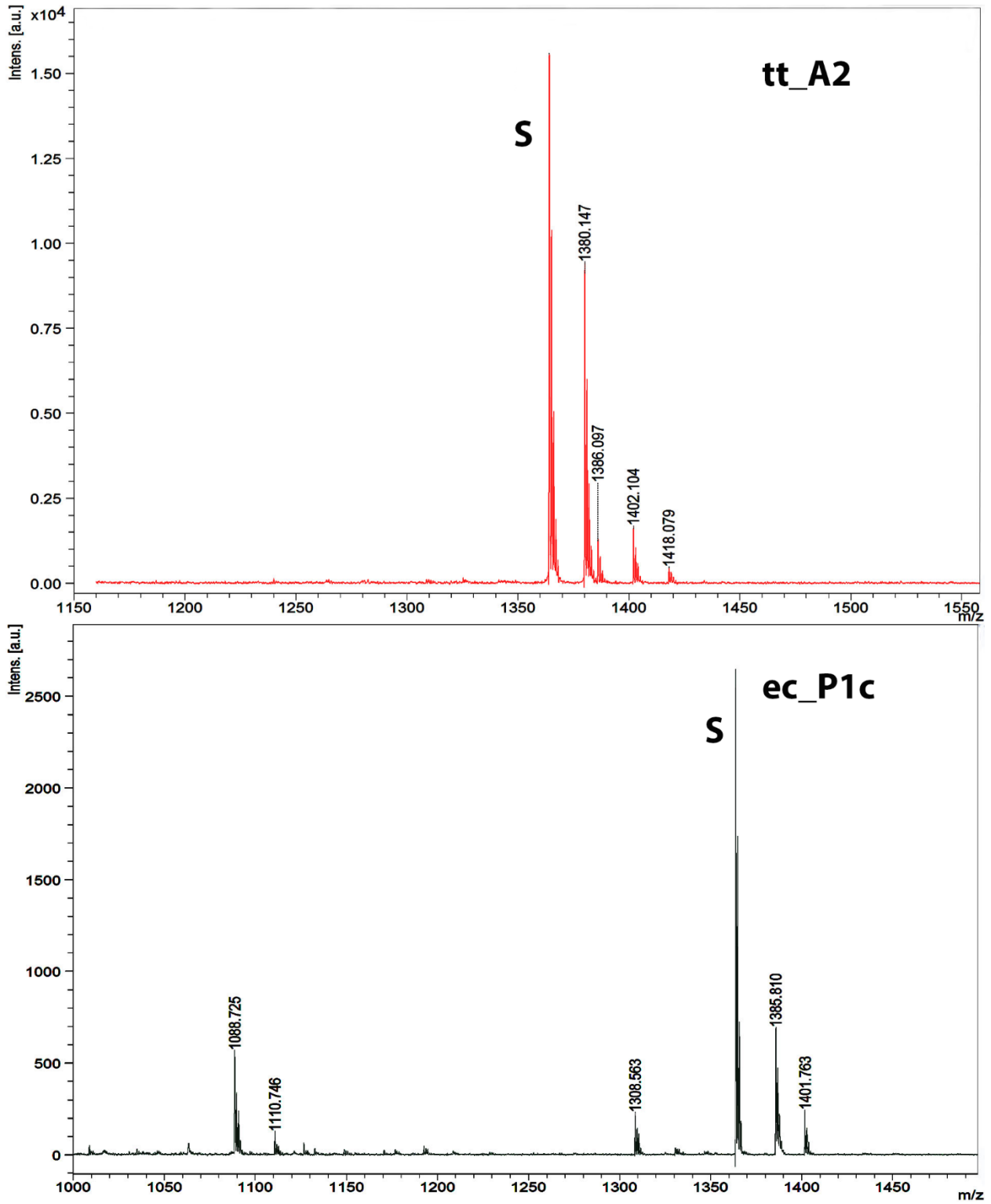
C

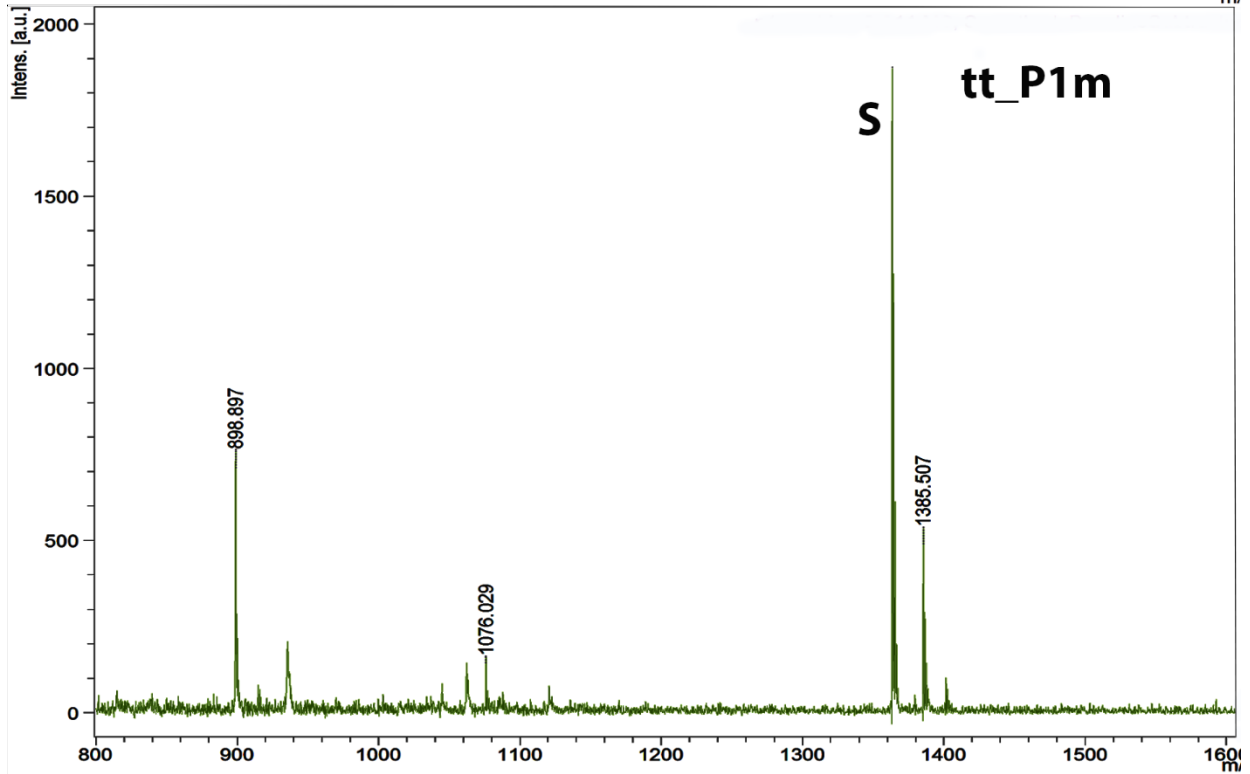
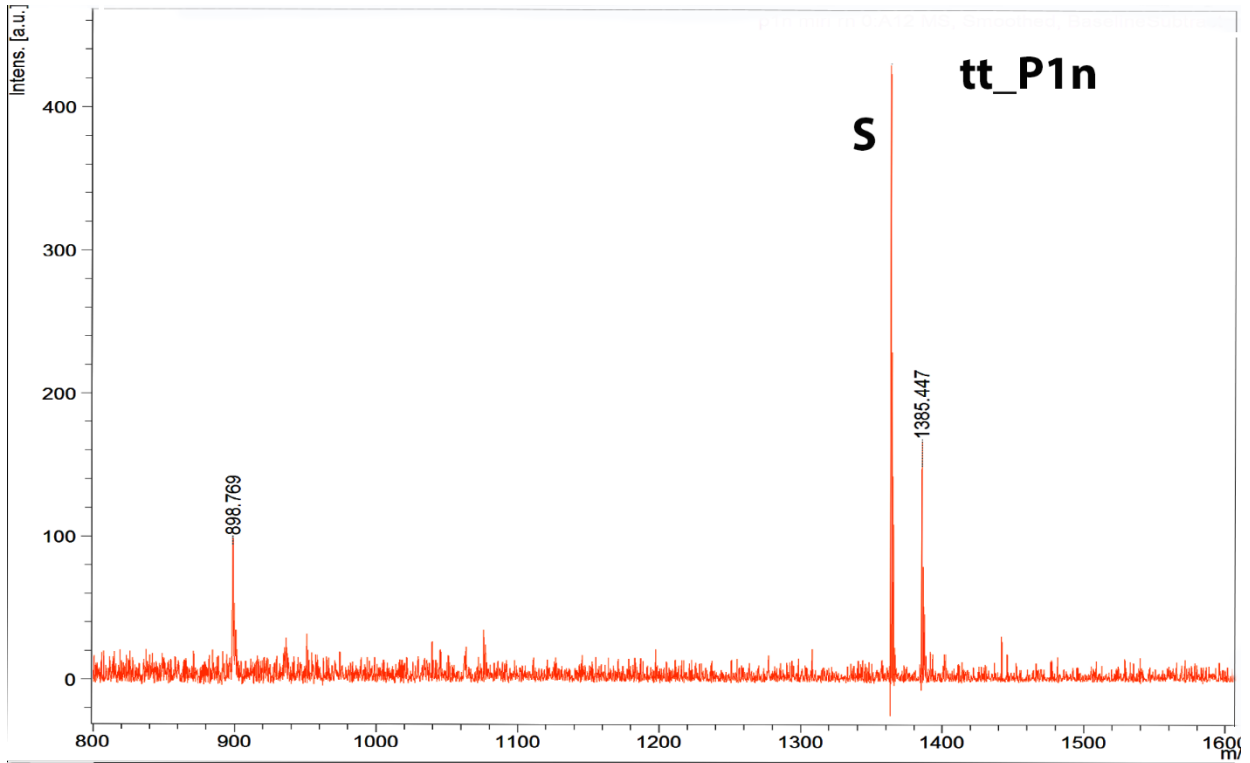


D

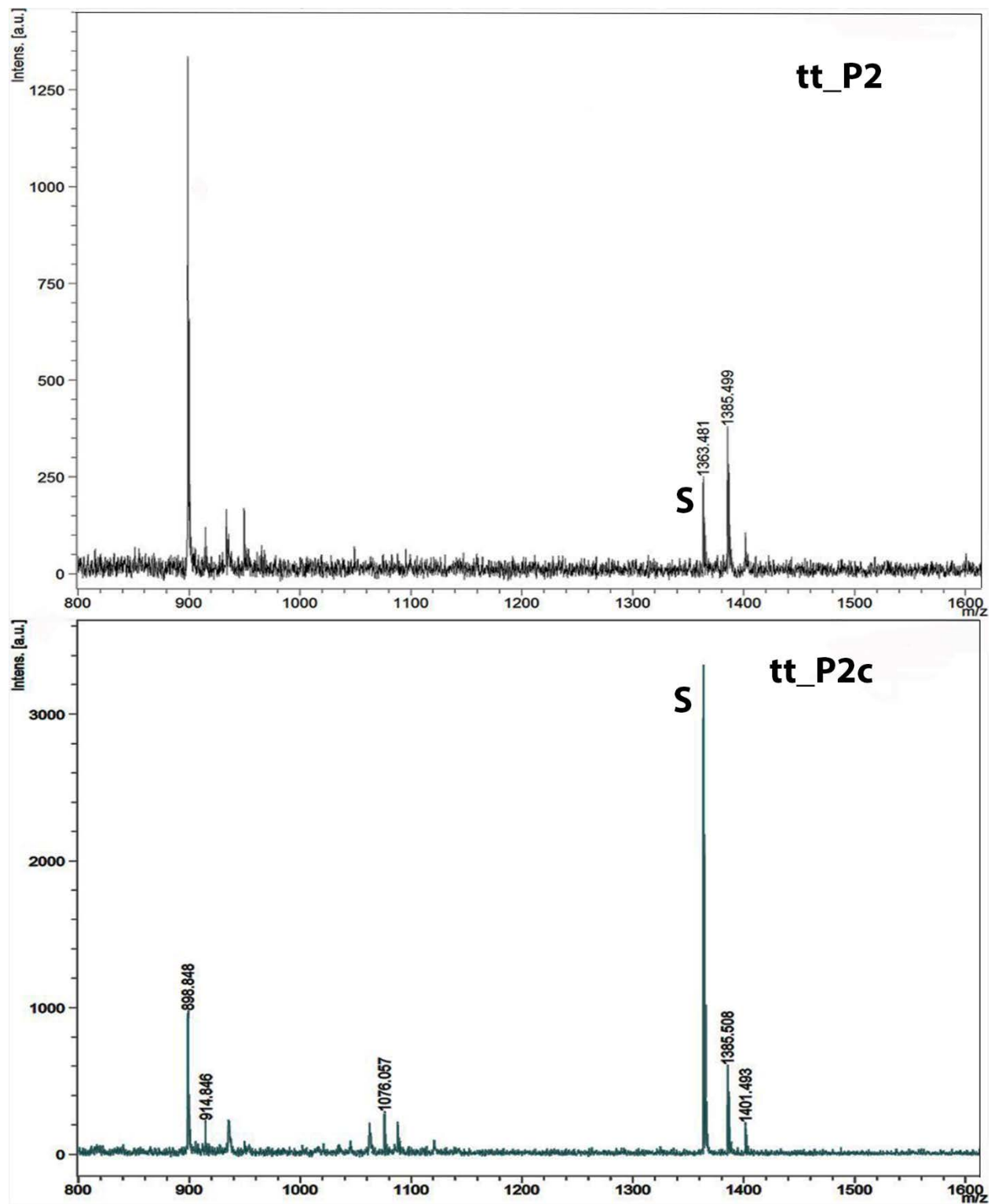


F



F

G



H

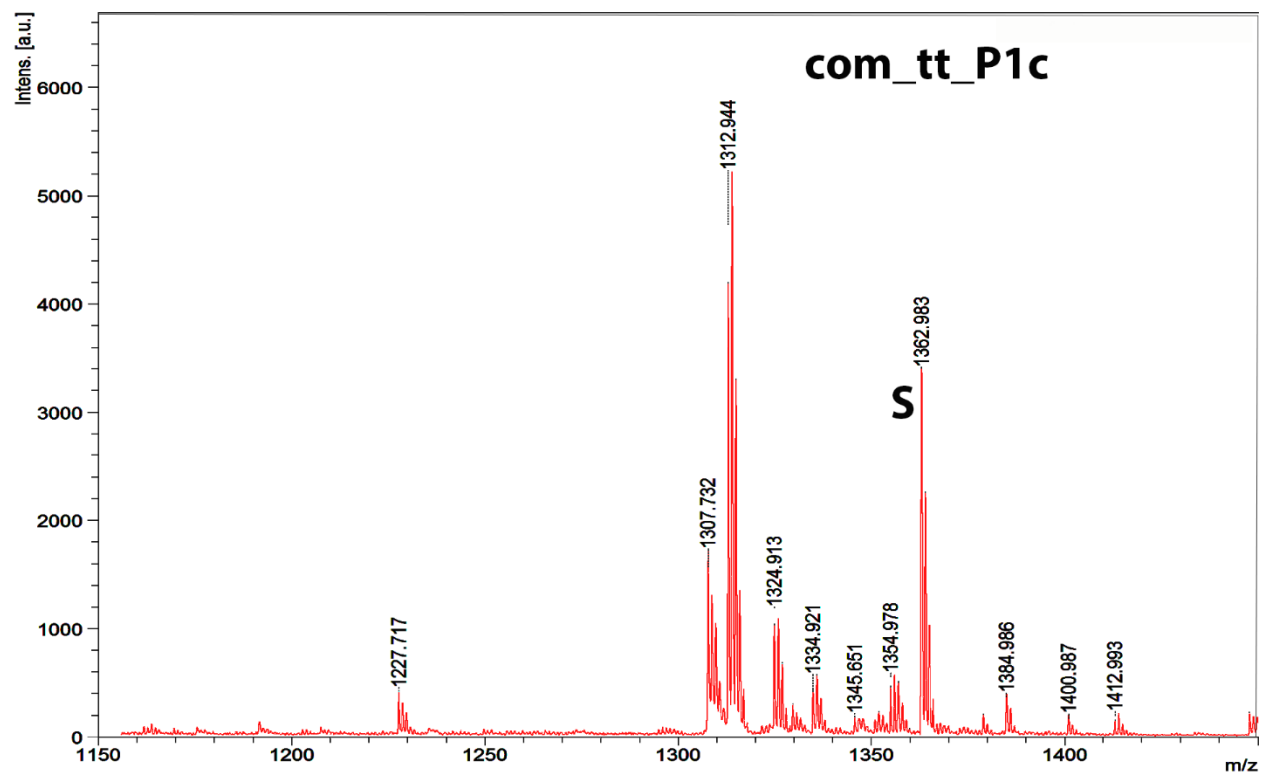
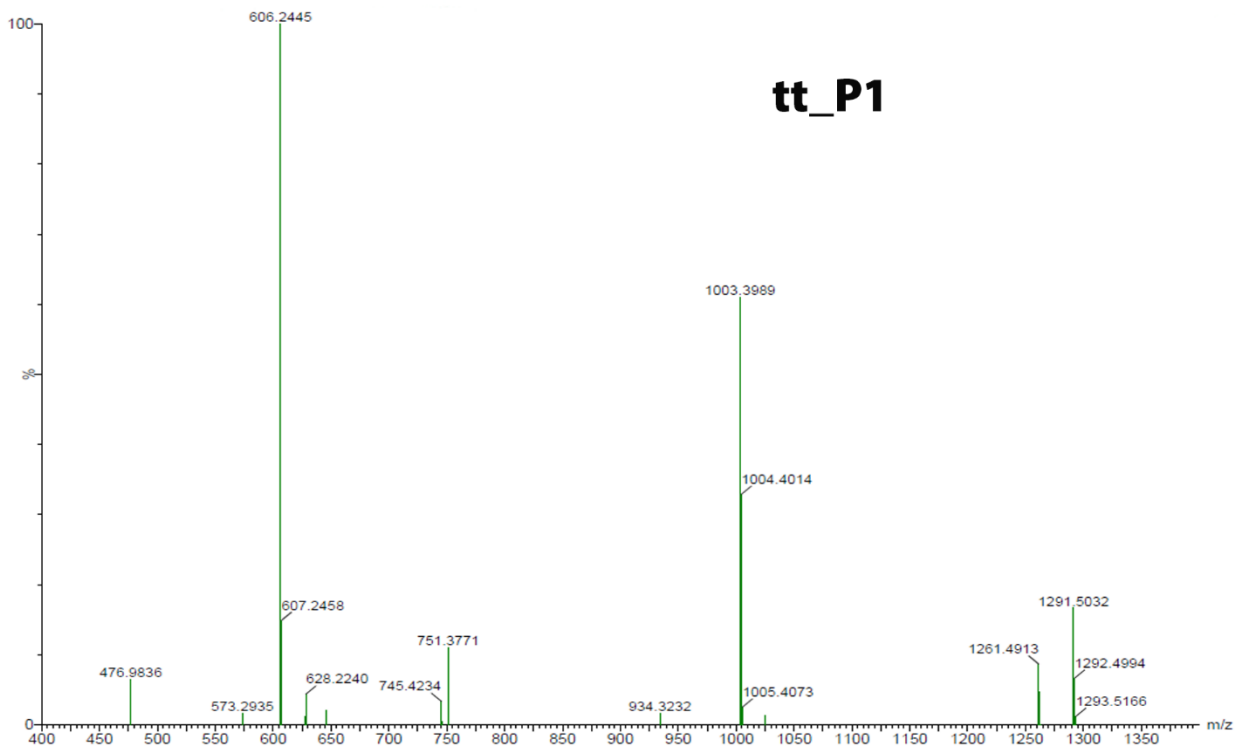
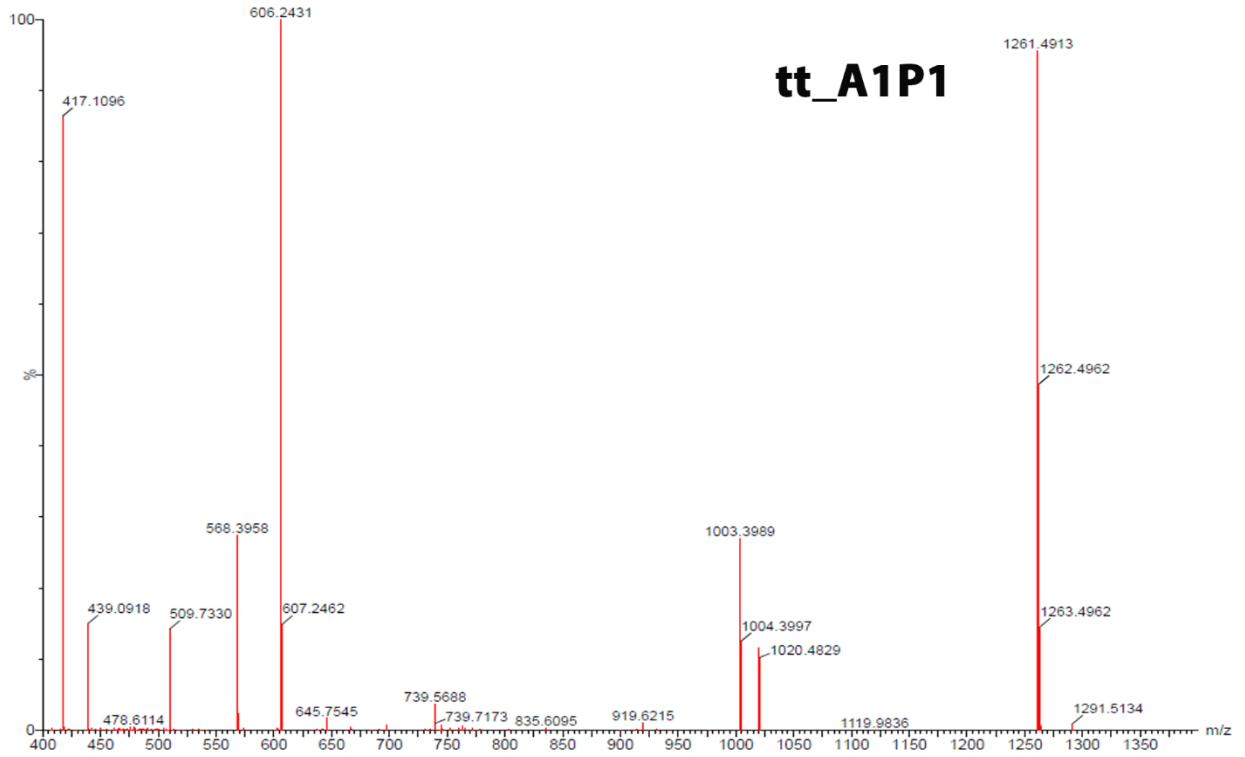


Figure S9: LC-MS spectra of A. tt_A1P1 and tt_P1 B. sa_P1c and tt_P1c

A



B

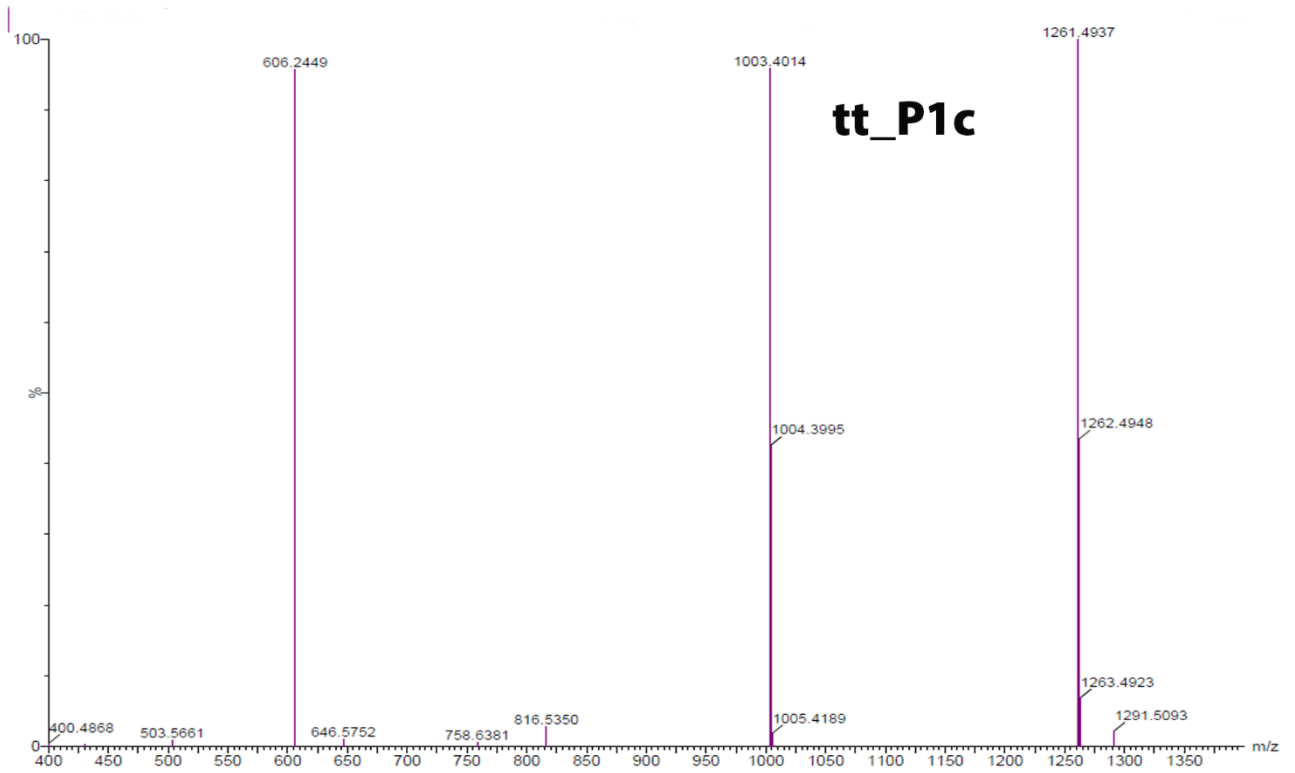
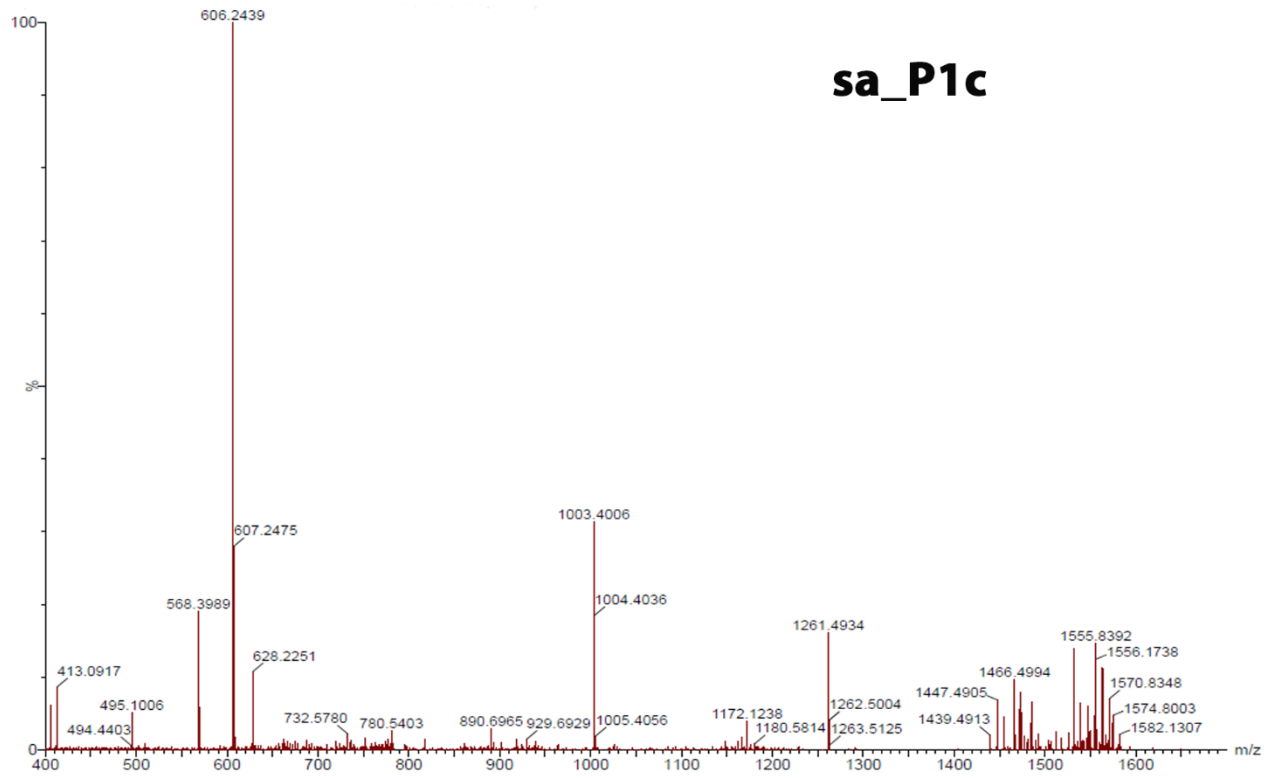
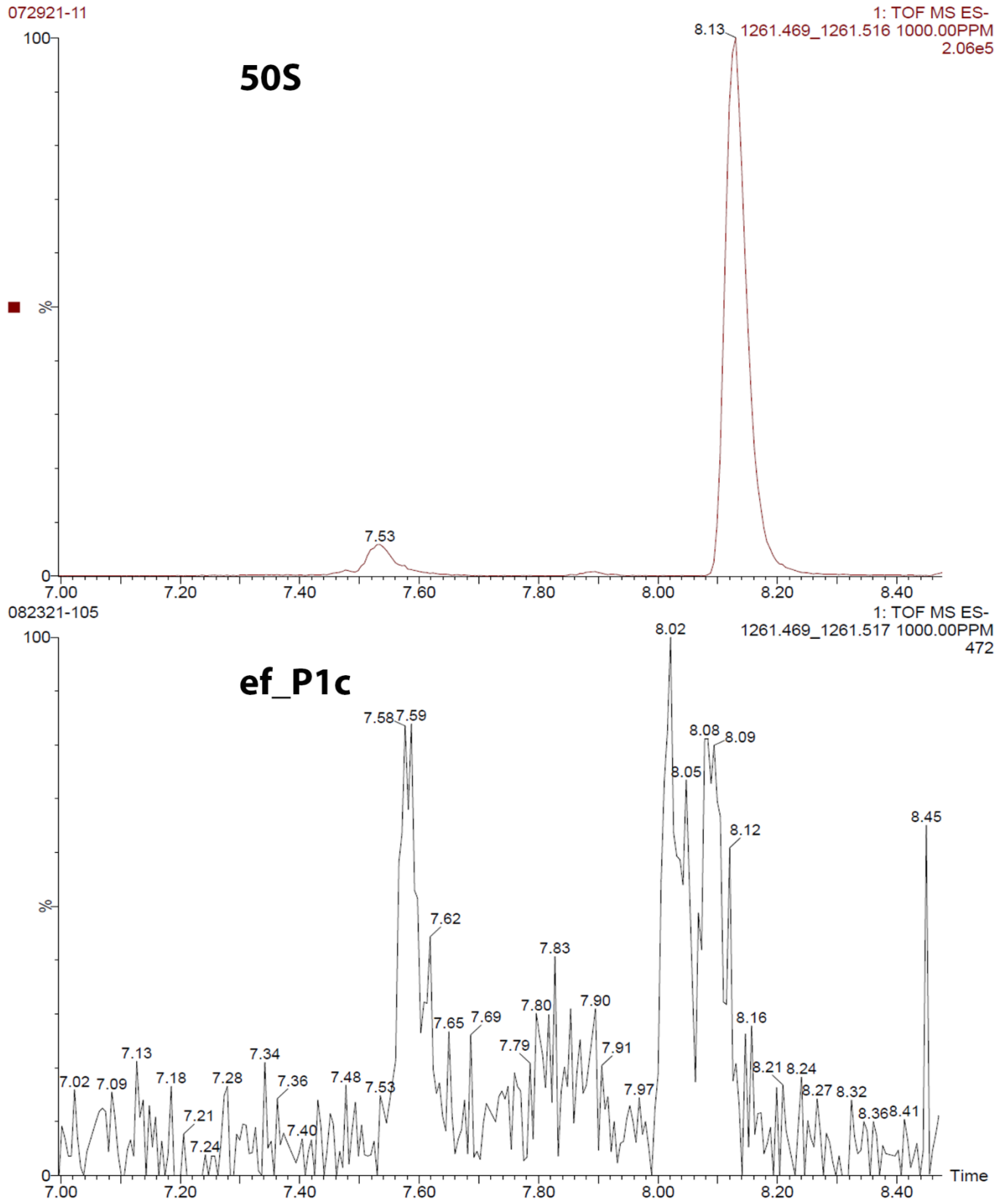


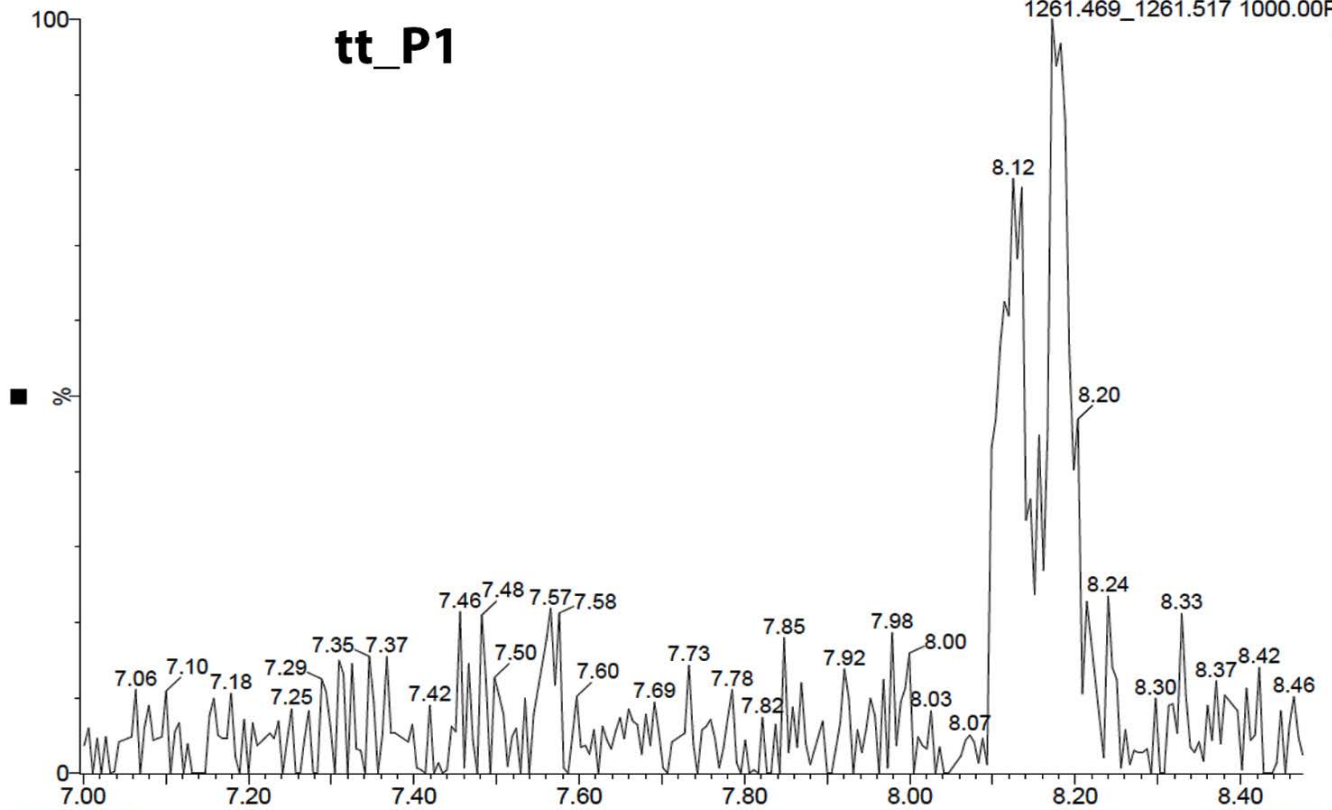
Figure S10: UPLC chromatograms of 50S, ef_P1c, tt_P1, sa_P1c, tt_A1P1, tt_P1c



081621-101

tt_P1

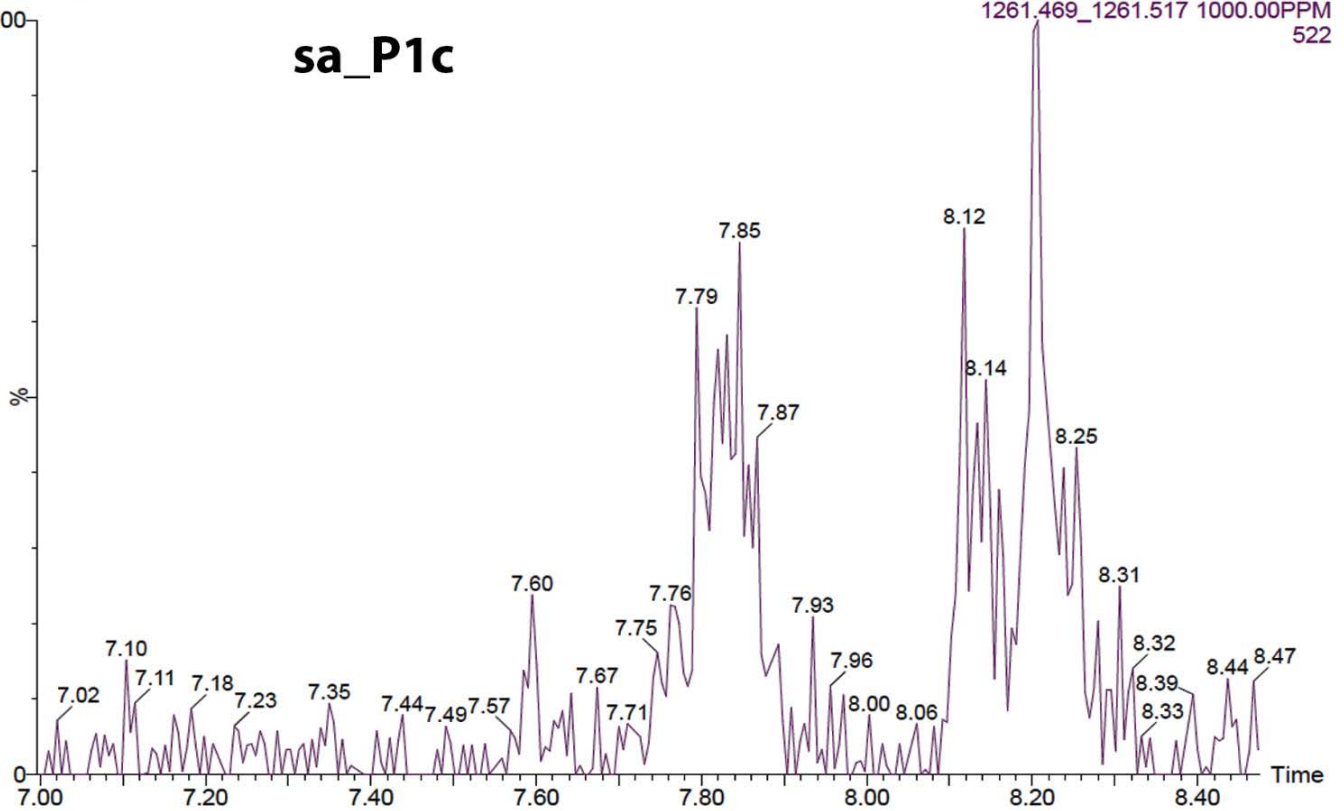
1: TOF MS ES-
1261.469_1261.517 1000.00PPM
435



082621-102

sa_P1c

1: TOF MS ES-
1261.469_1261.517 1000.00PPM
522



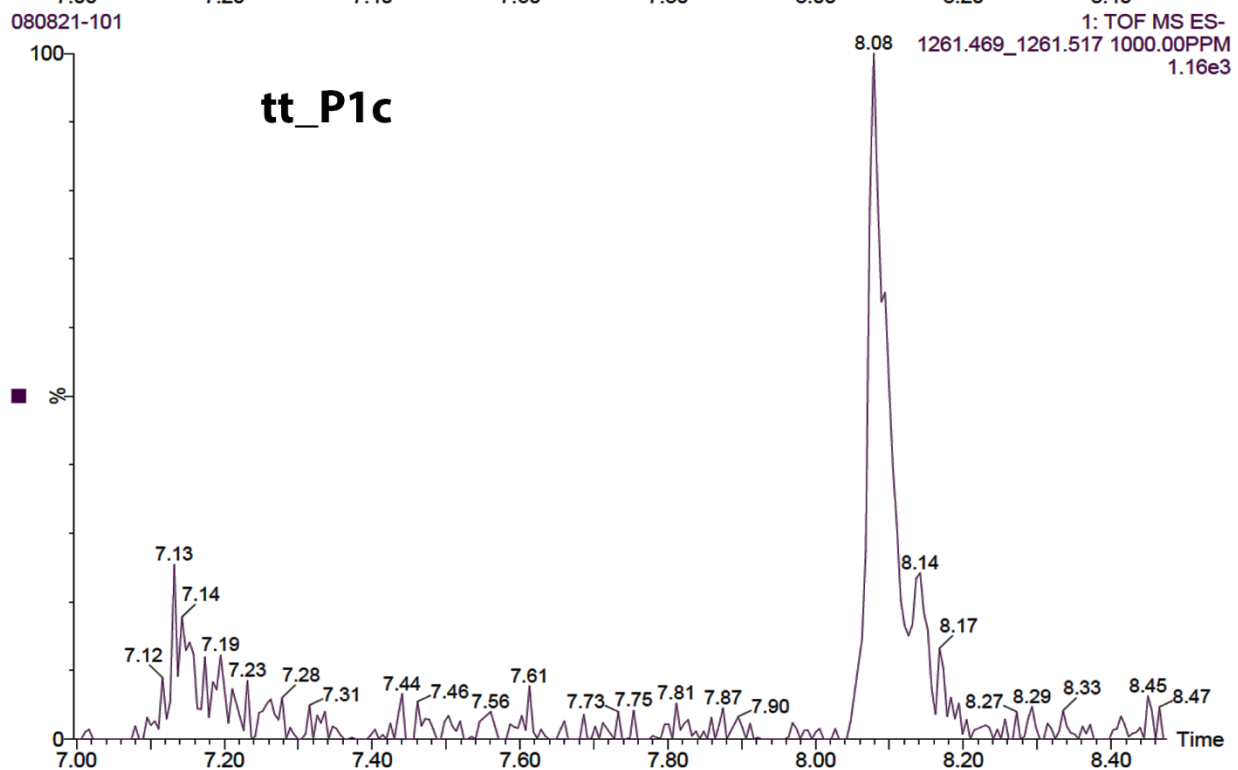
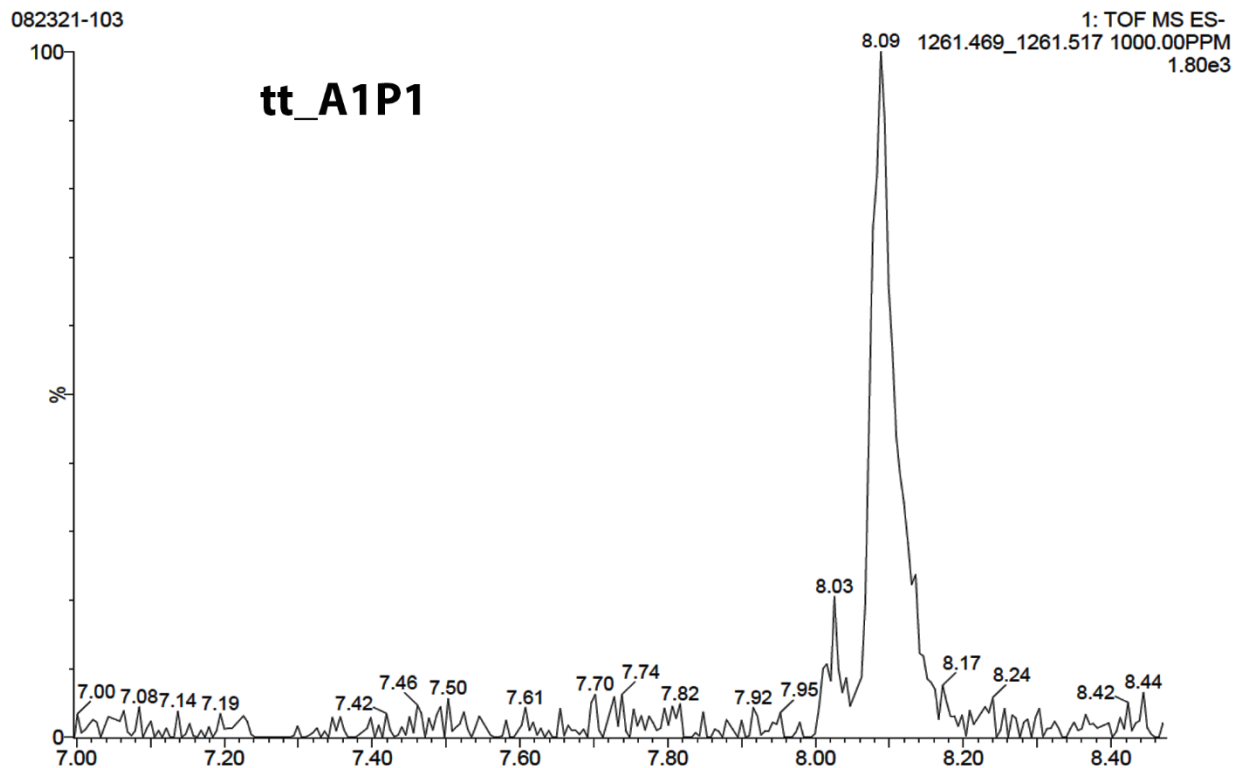
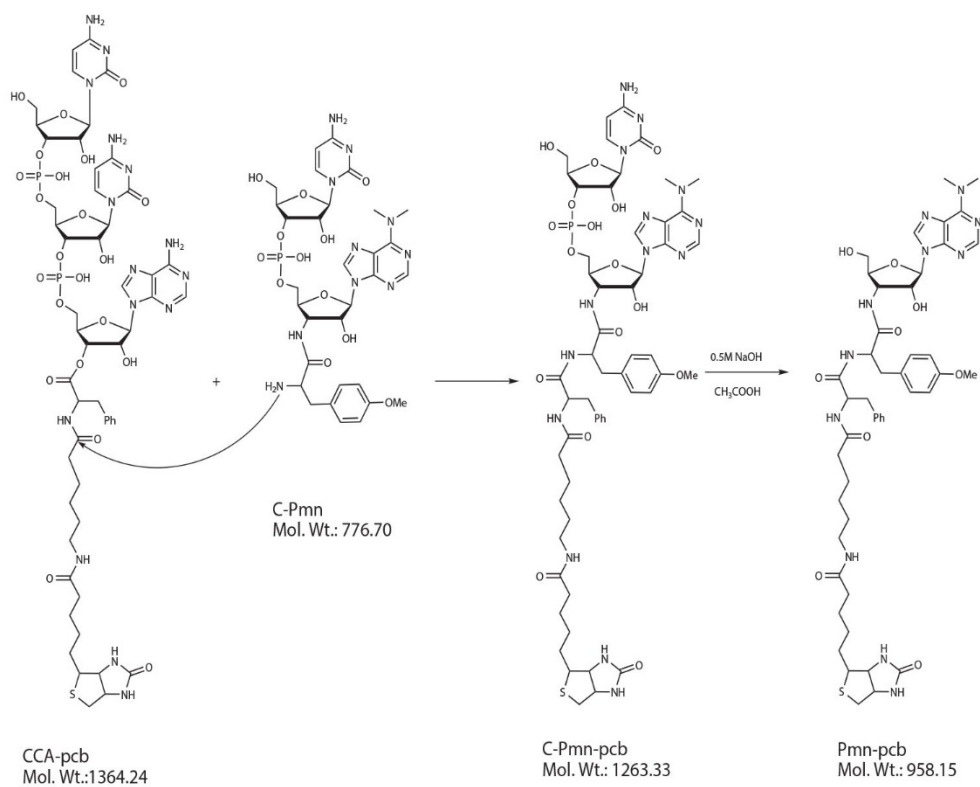


Figure S11: Verification of the product identity; presence of a peptide bond. A. Schematic representation of the product basic hydrolysis. B. MALDI spectra of the reaction mixture of construct tt_P1c before and after base treatment.

A



B

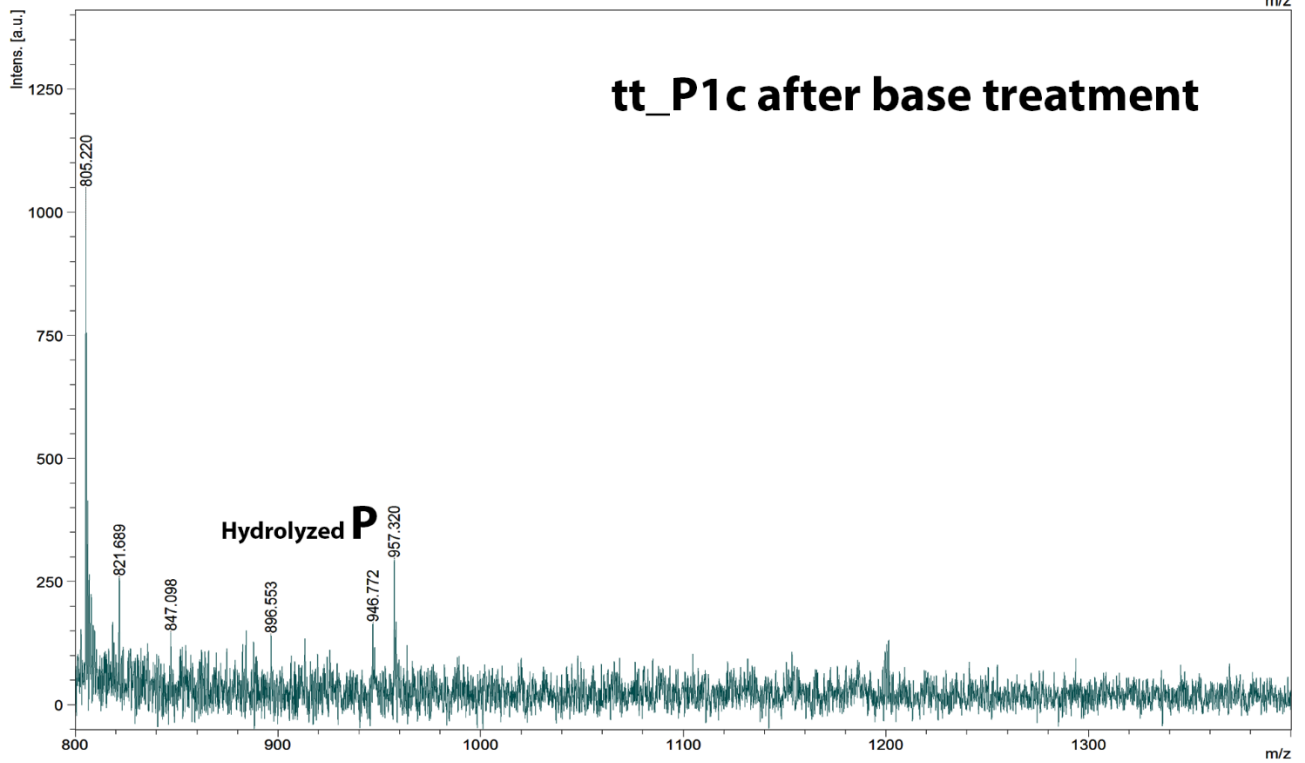
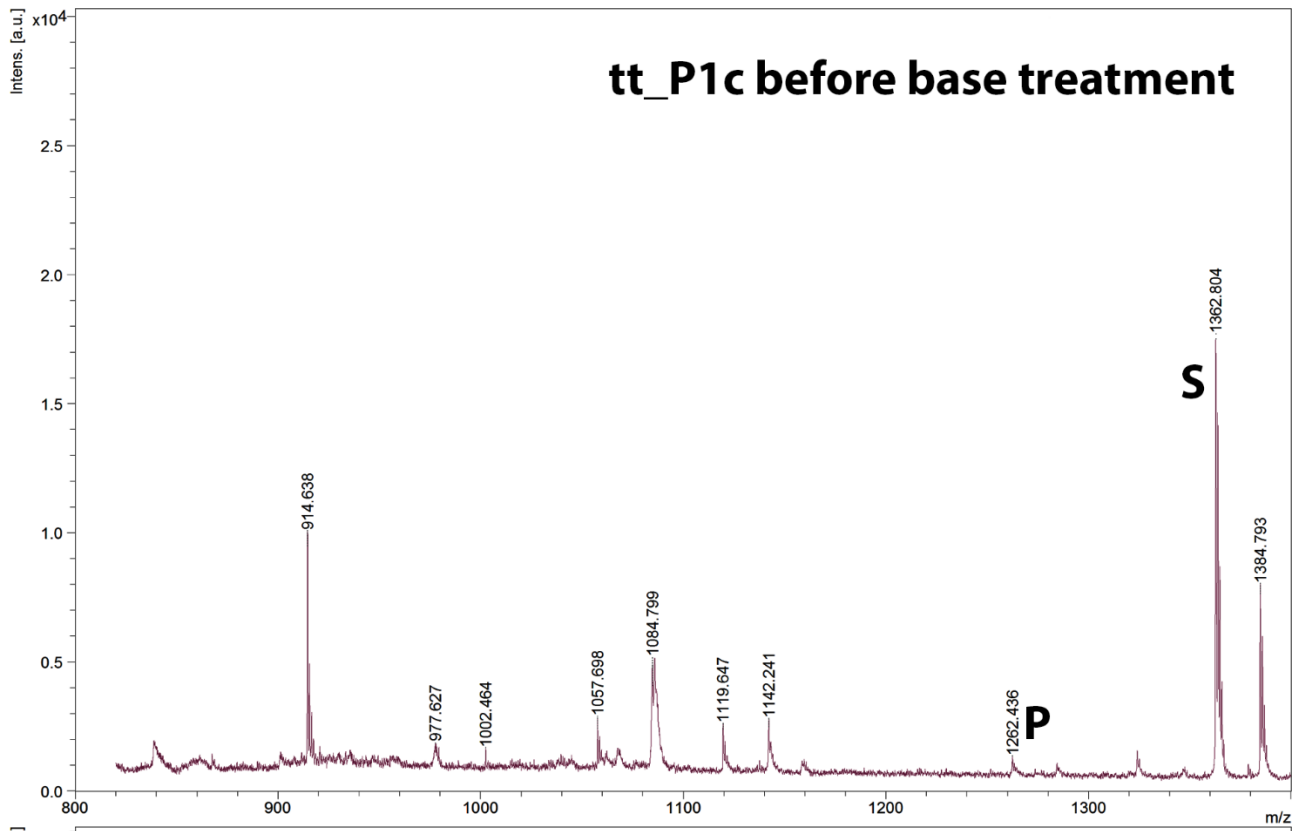
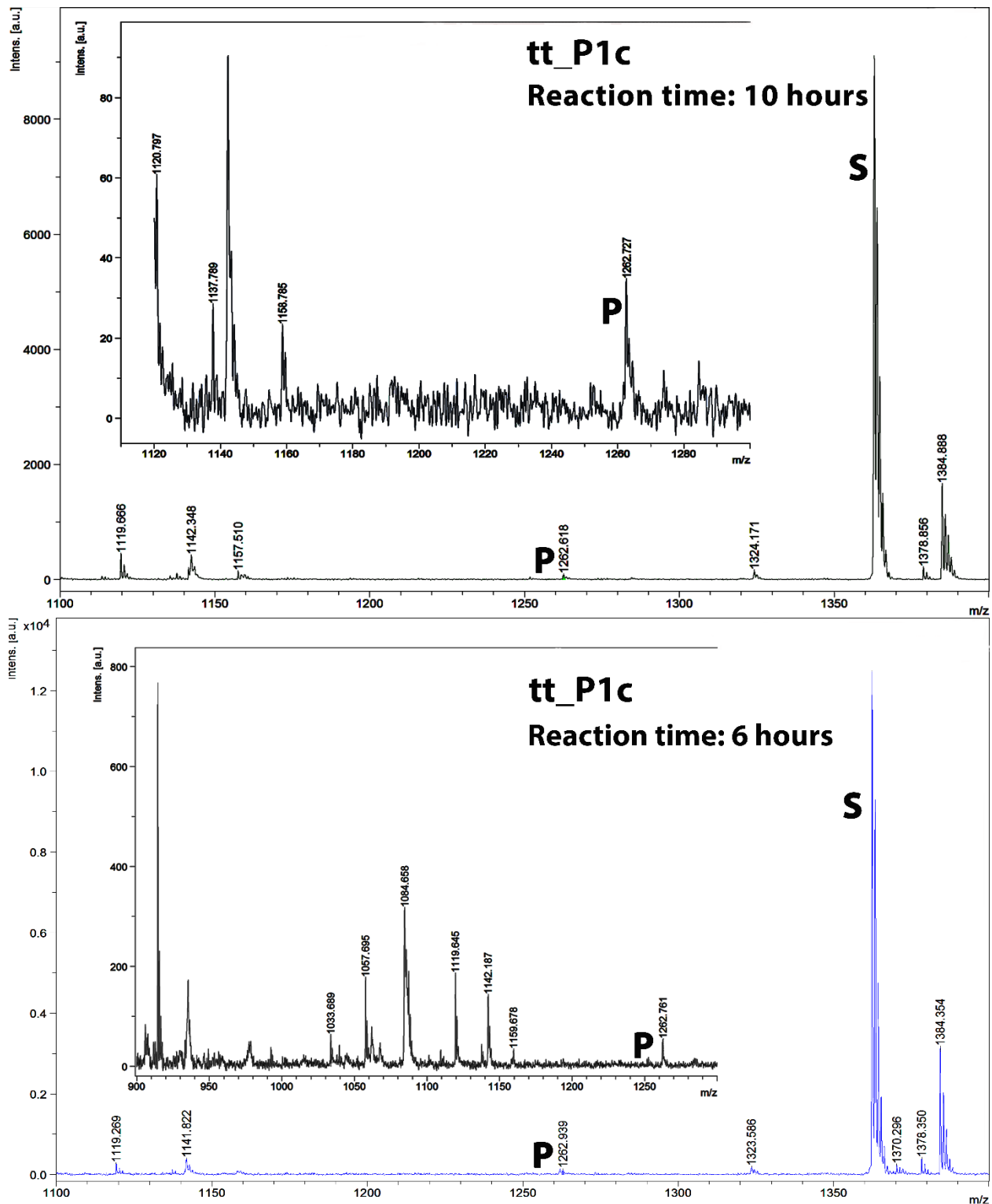
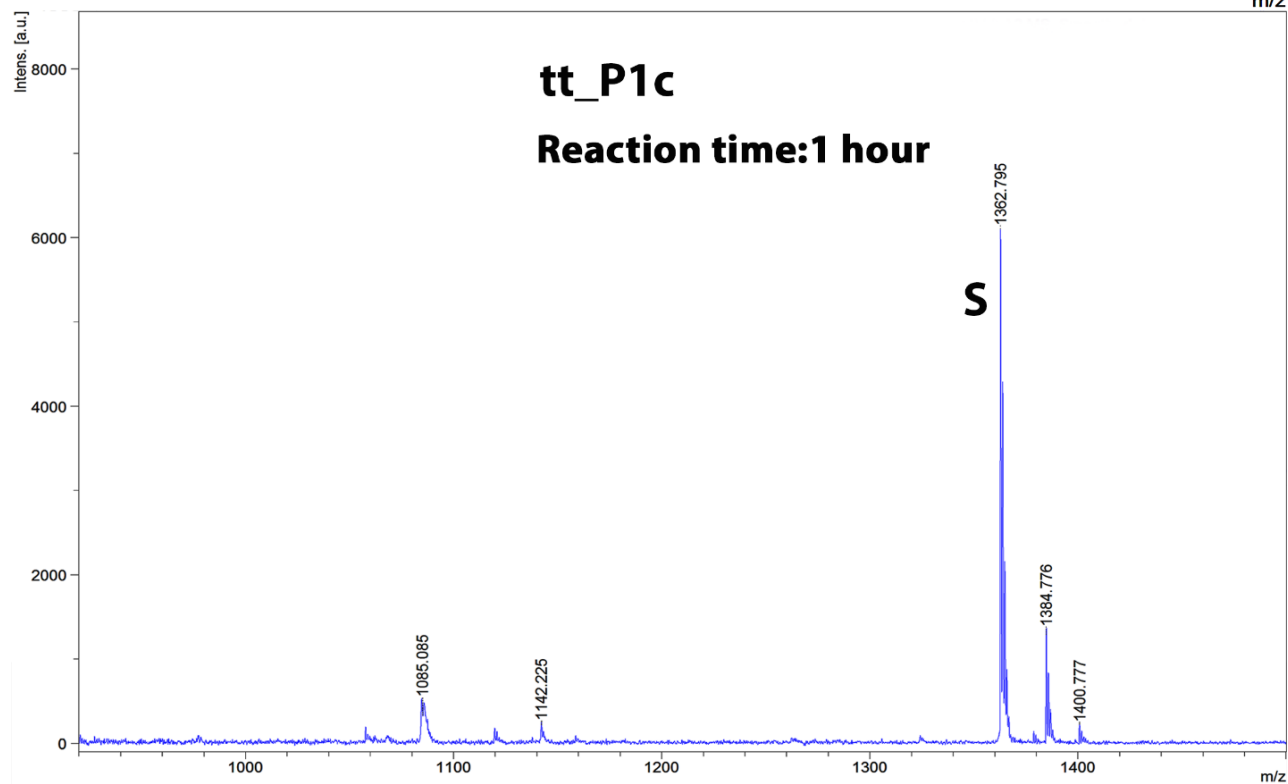
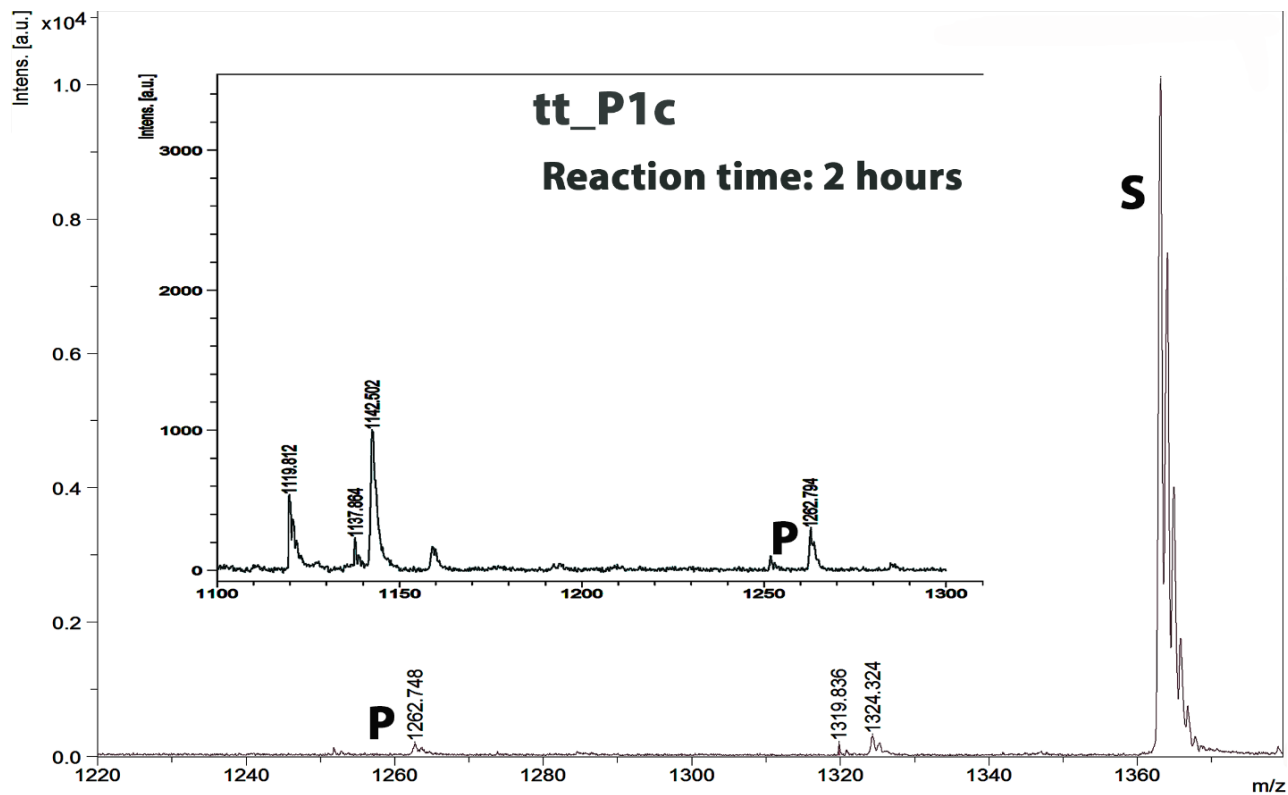
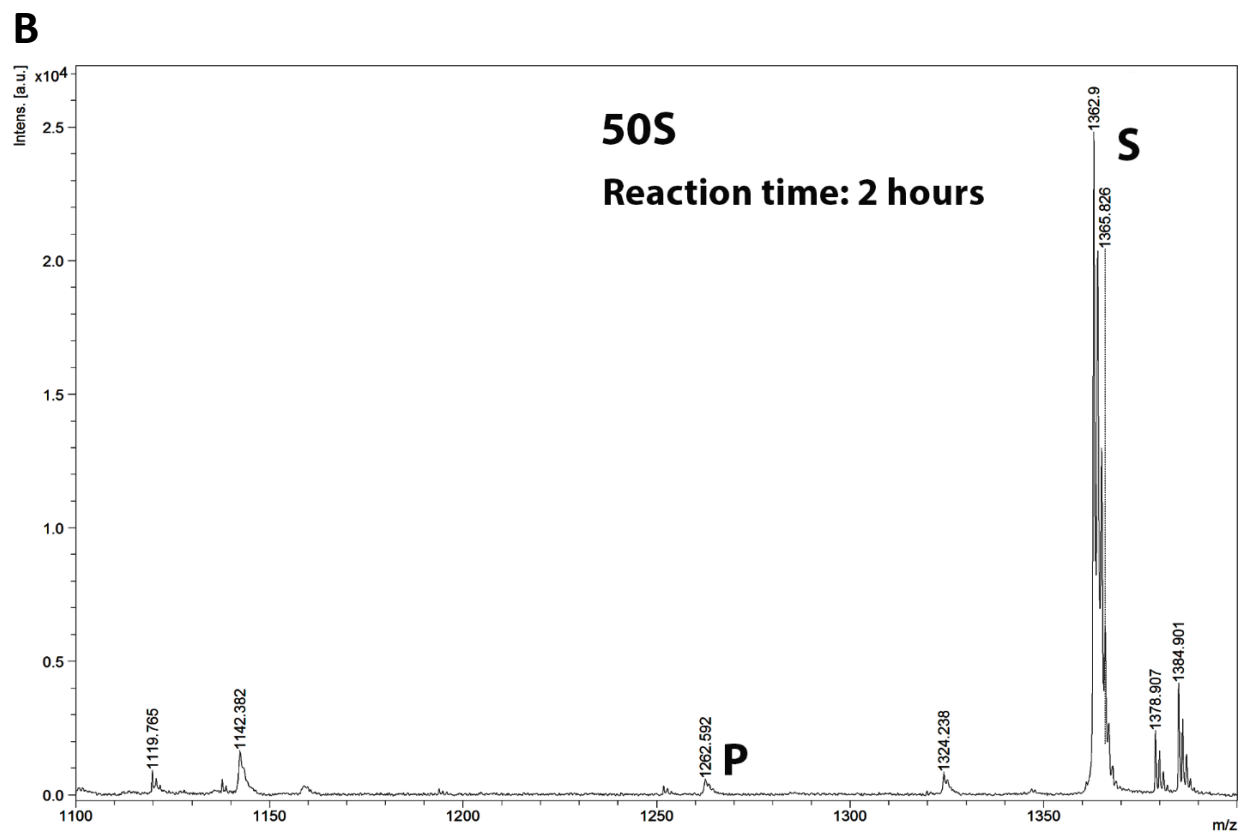
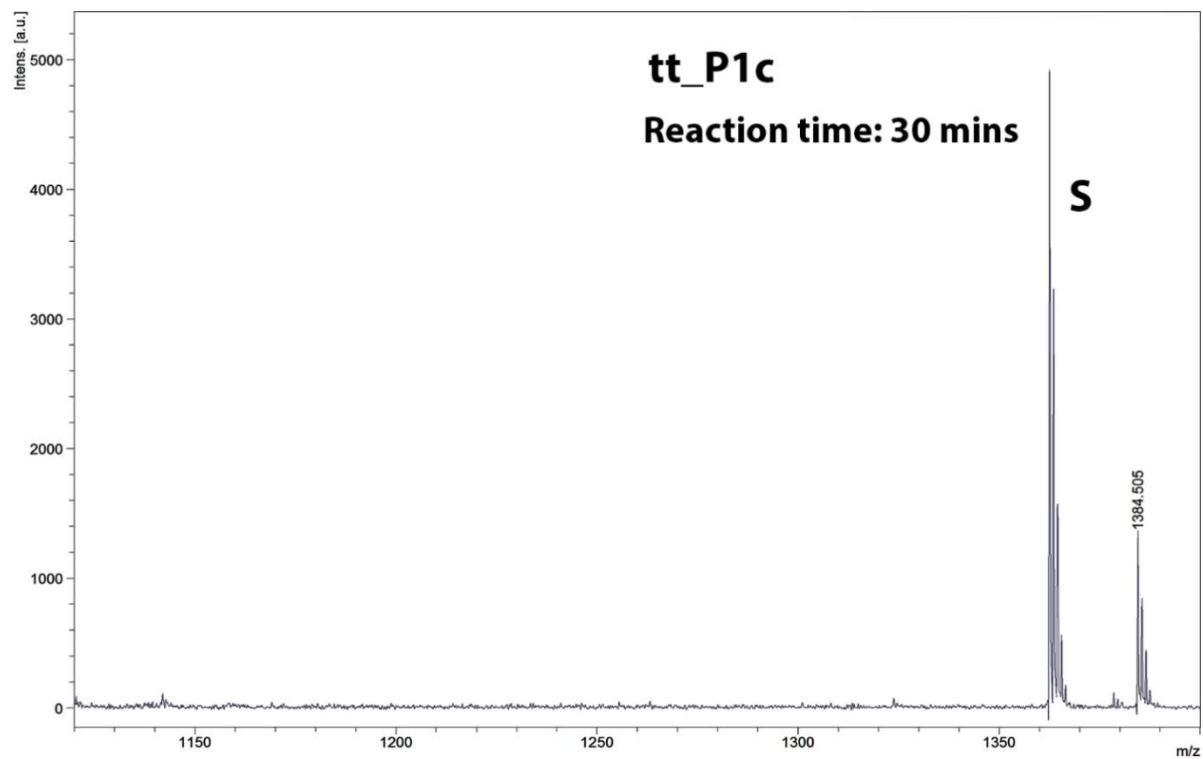


Figure S12: MALDI spectra of the fragment reaction for various time durations A. tt_P1c B. 50S.

A







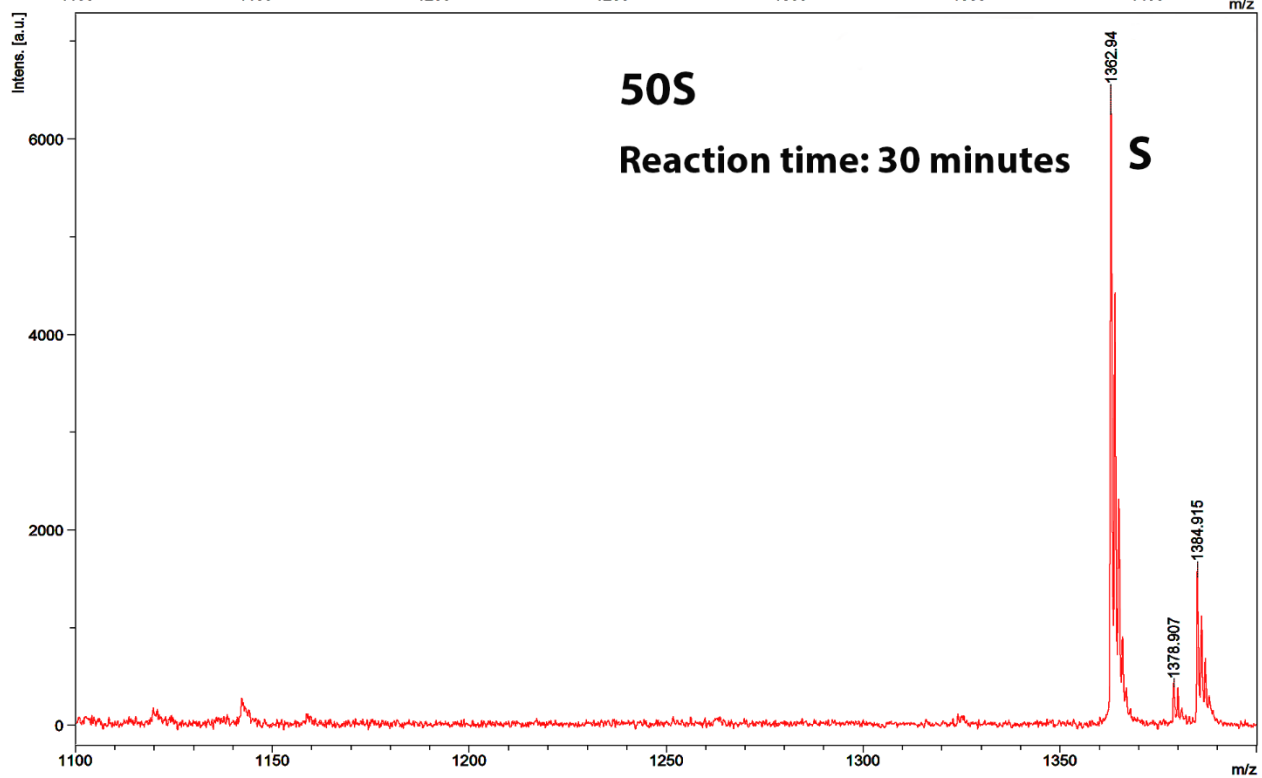
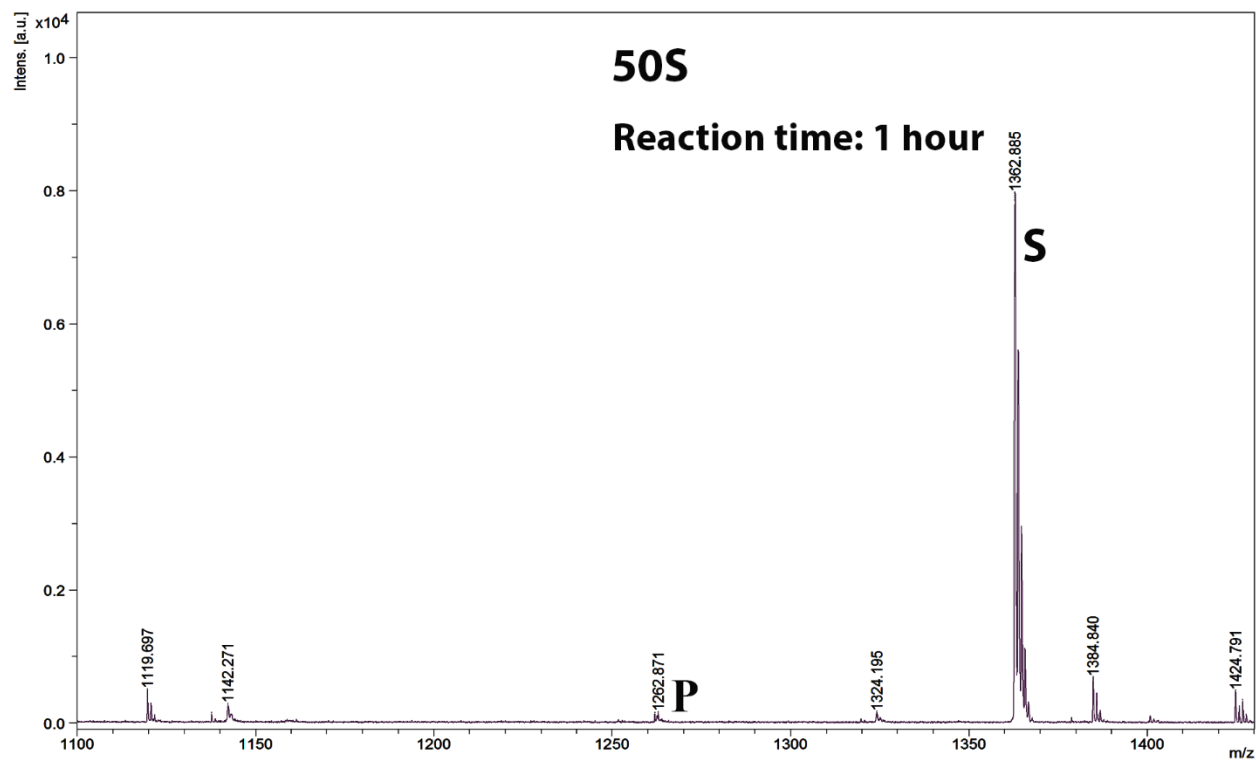
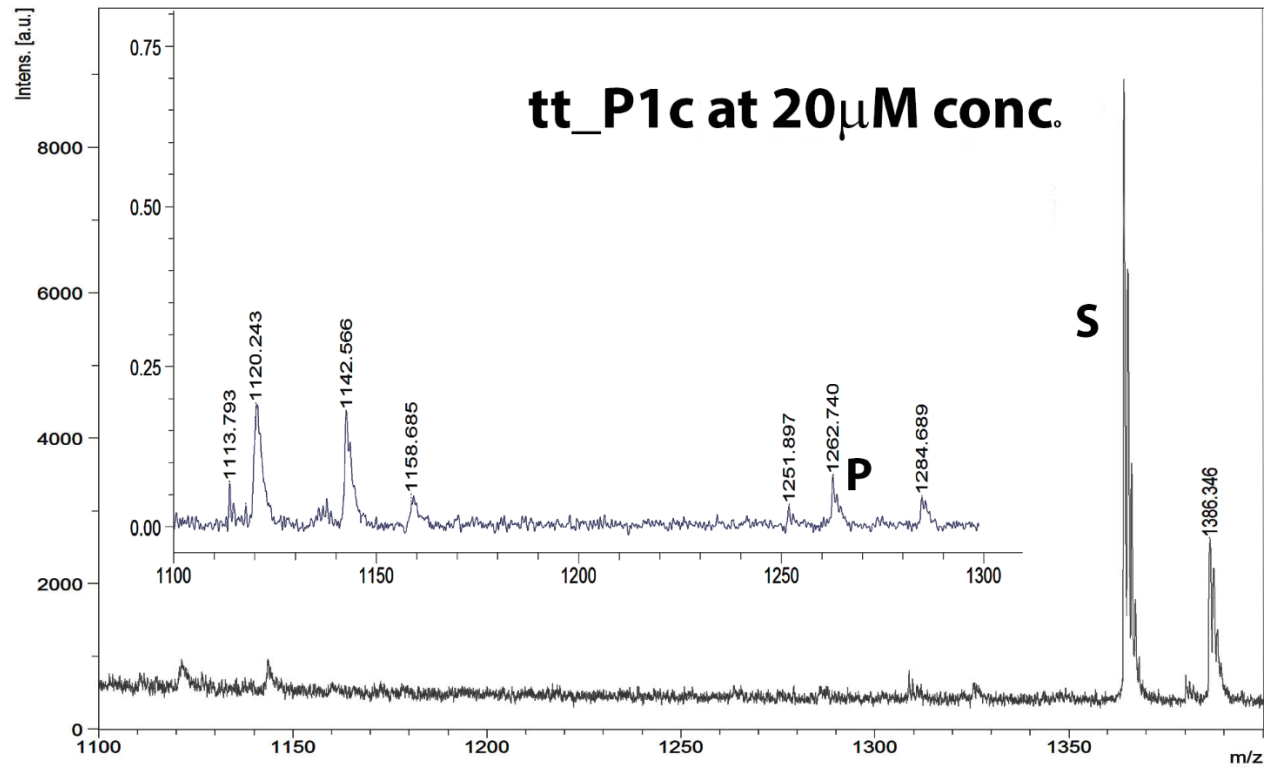
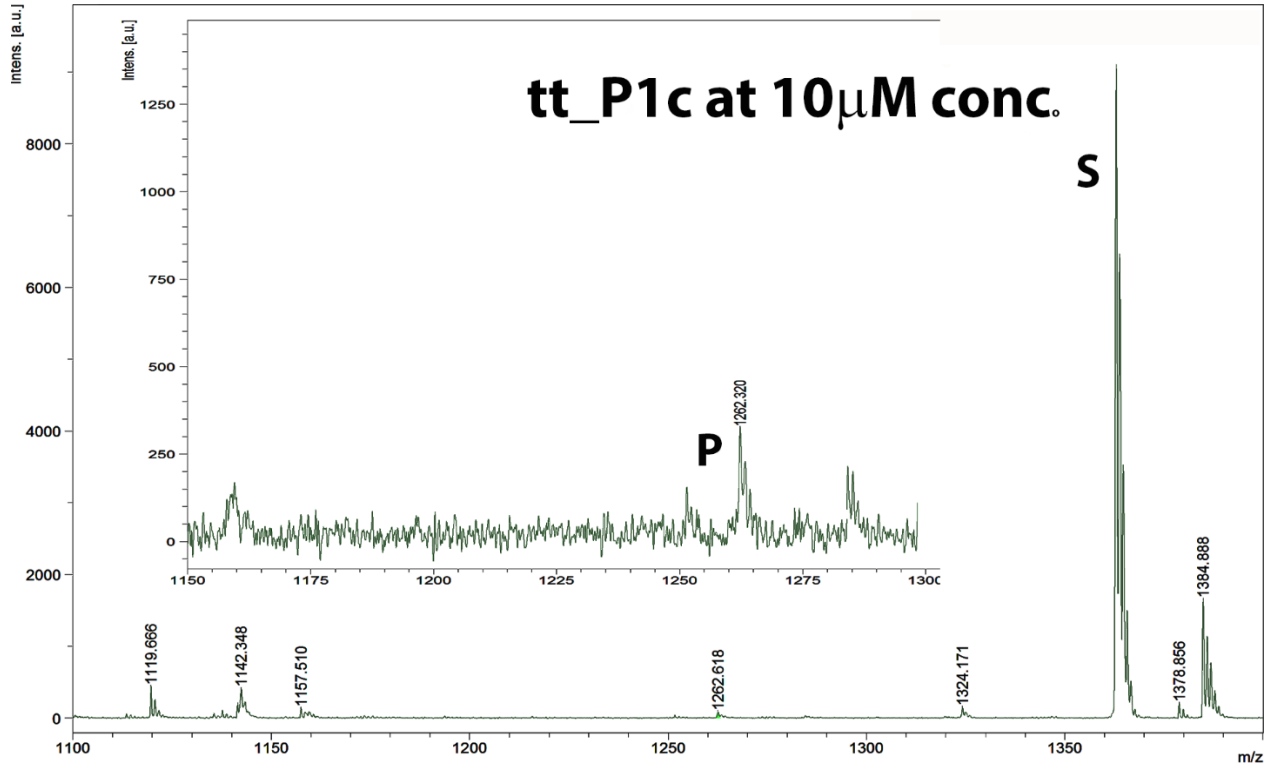
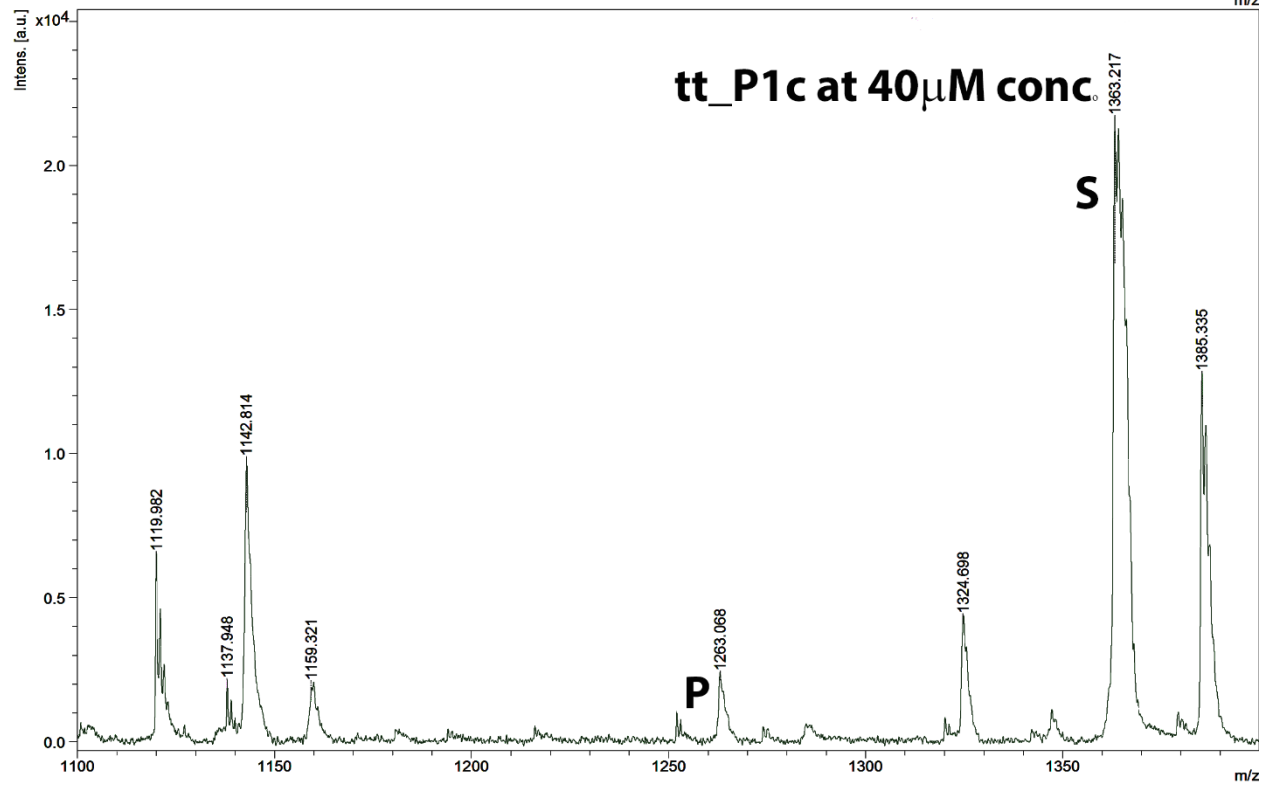
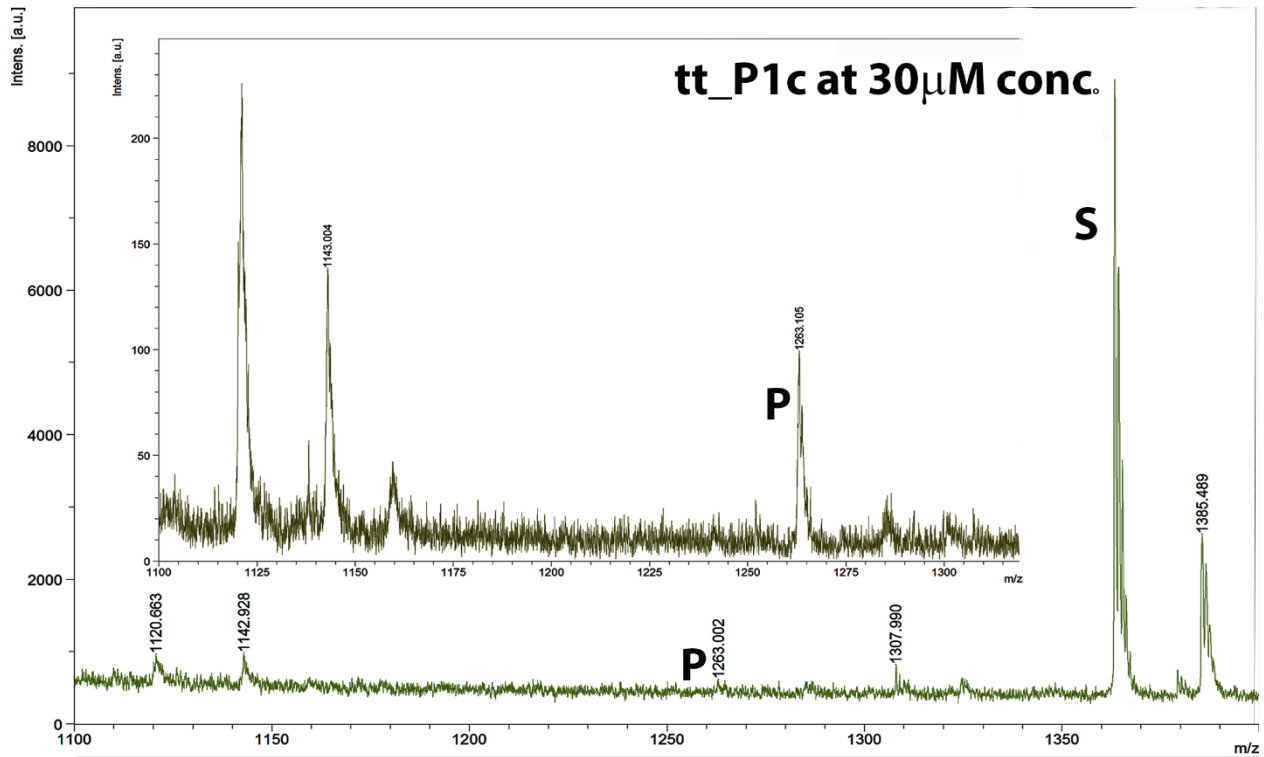


Figure S13: MALDI spectra of fragment reactions using 10, 20, 30, 40, 50 and 100 μM concentration of construct tt_P1c.





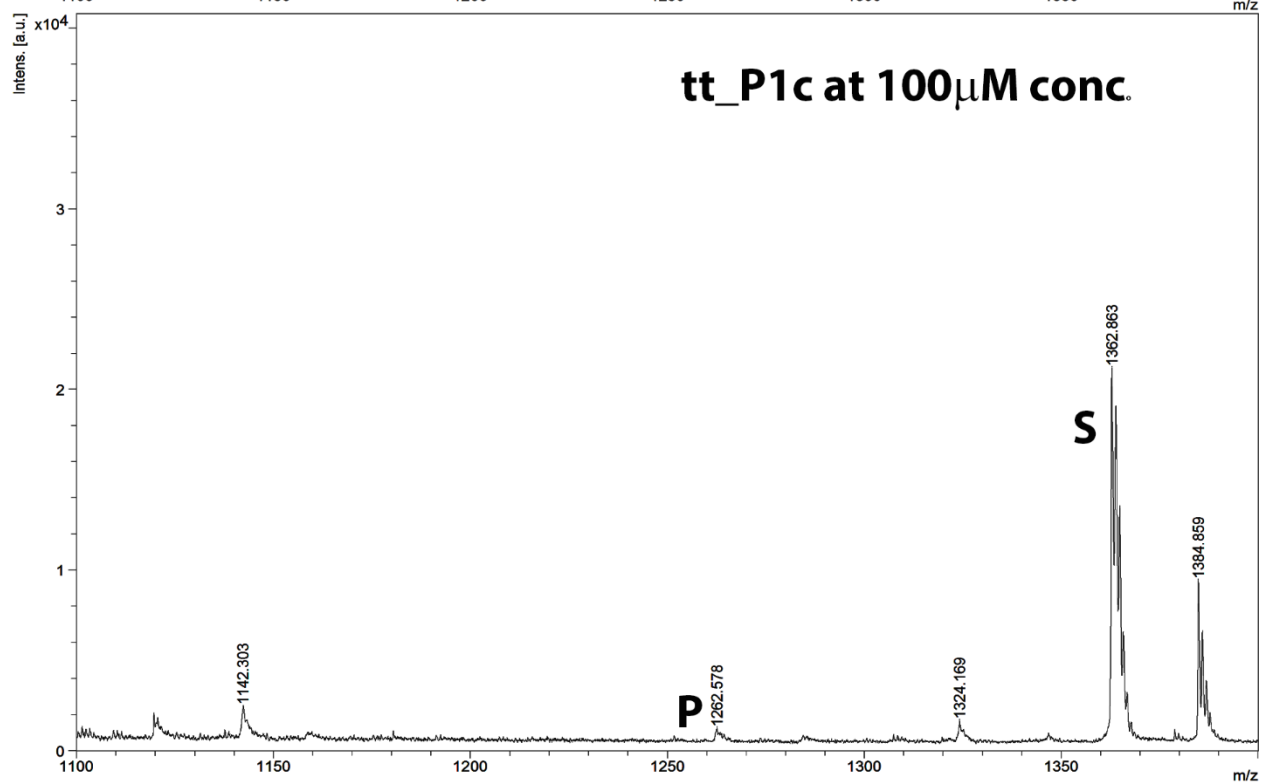
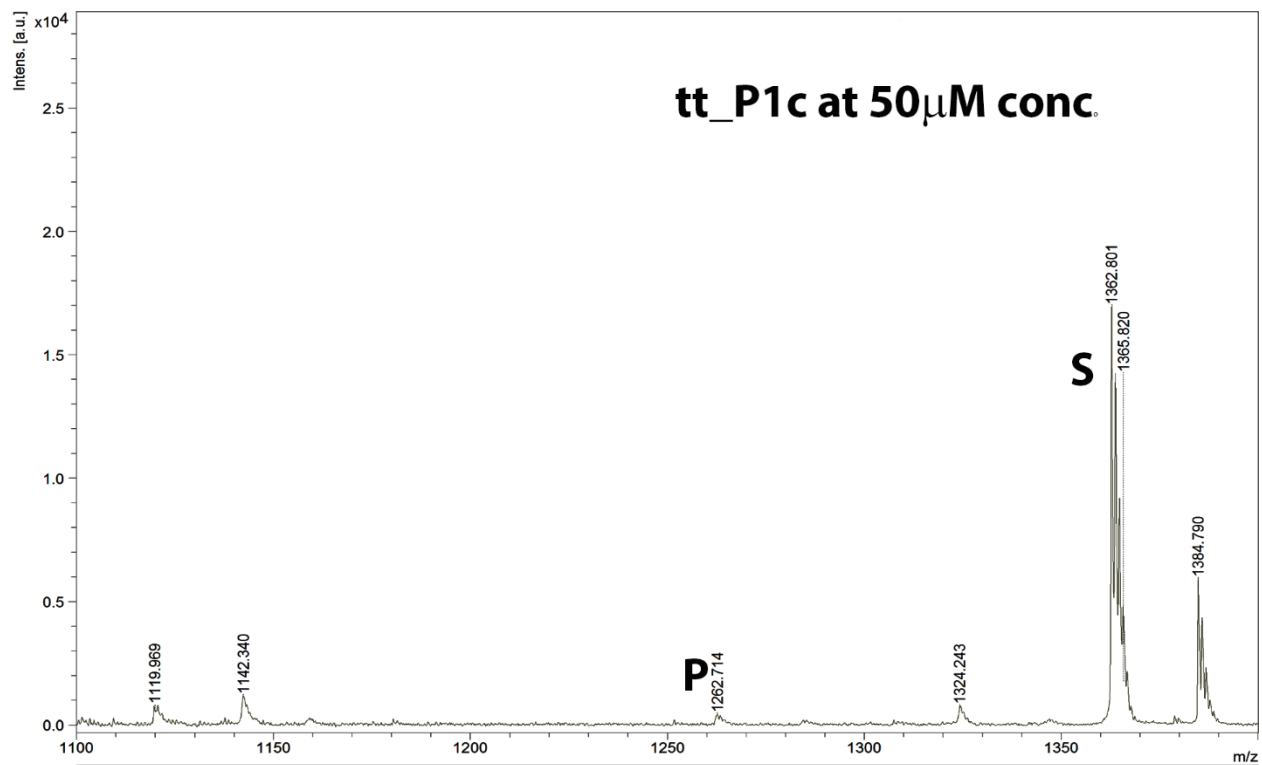


Figure S14: MALDI spectra of the reactions at 2.8 μM concentration of tt_P1c and different substrate concentrations.

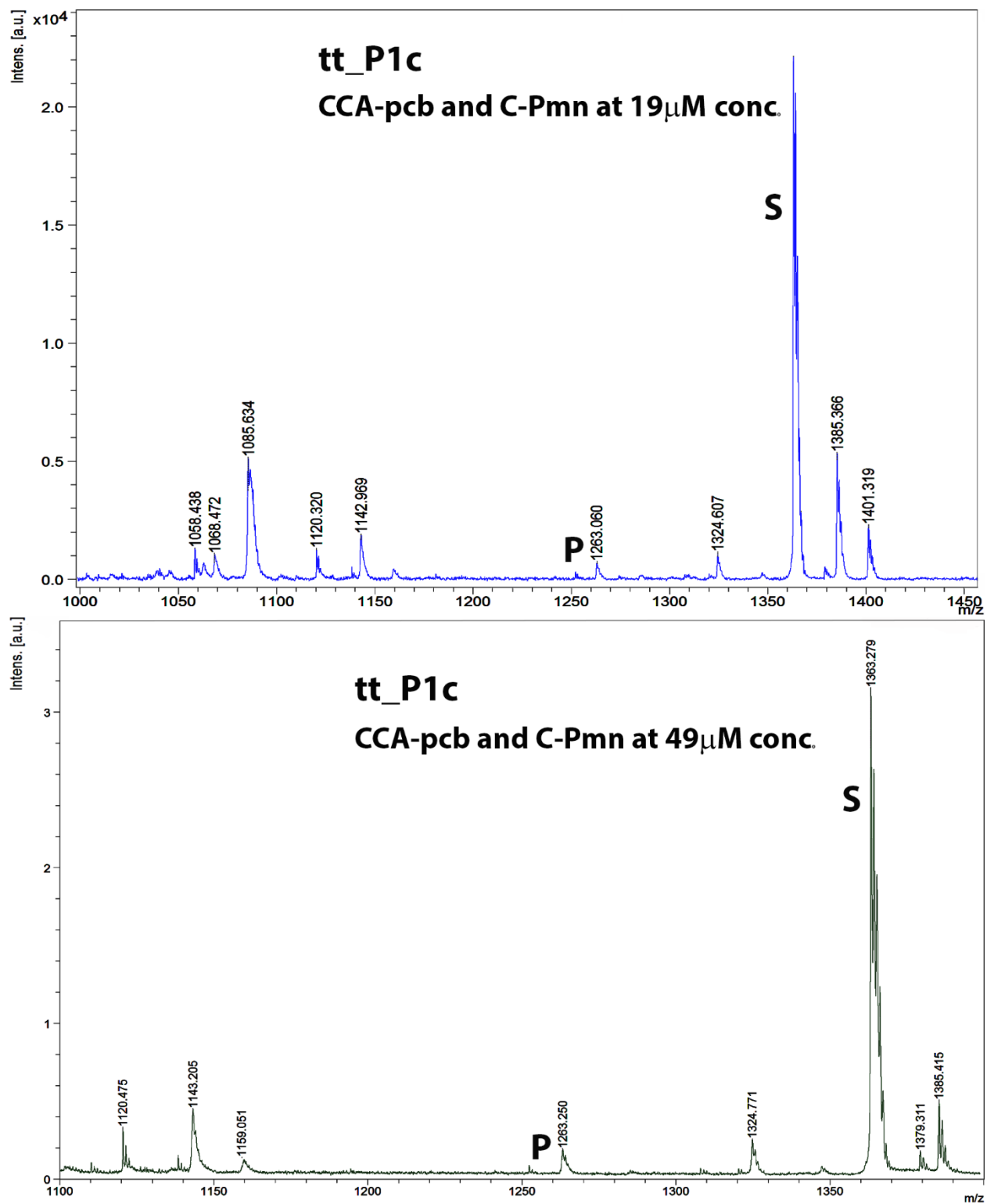
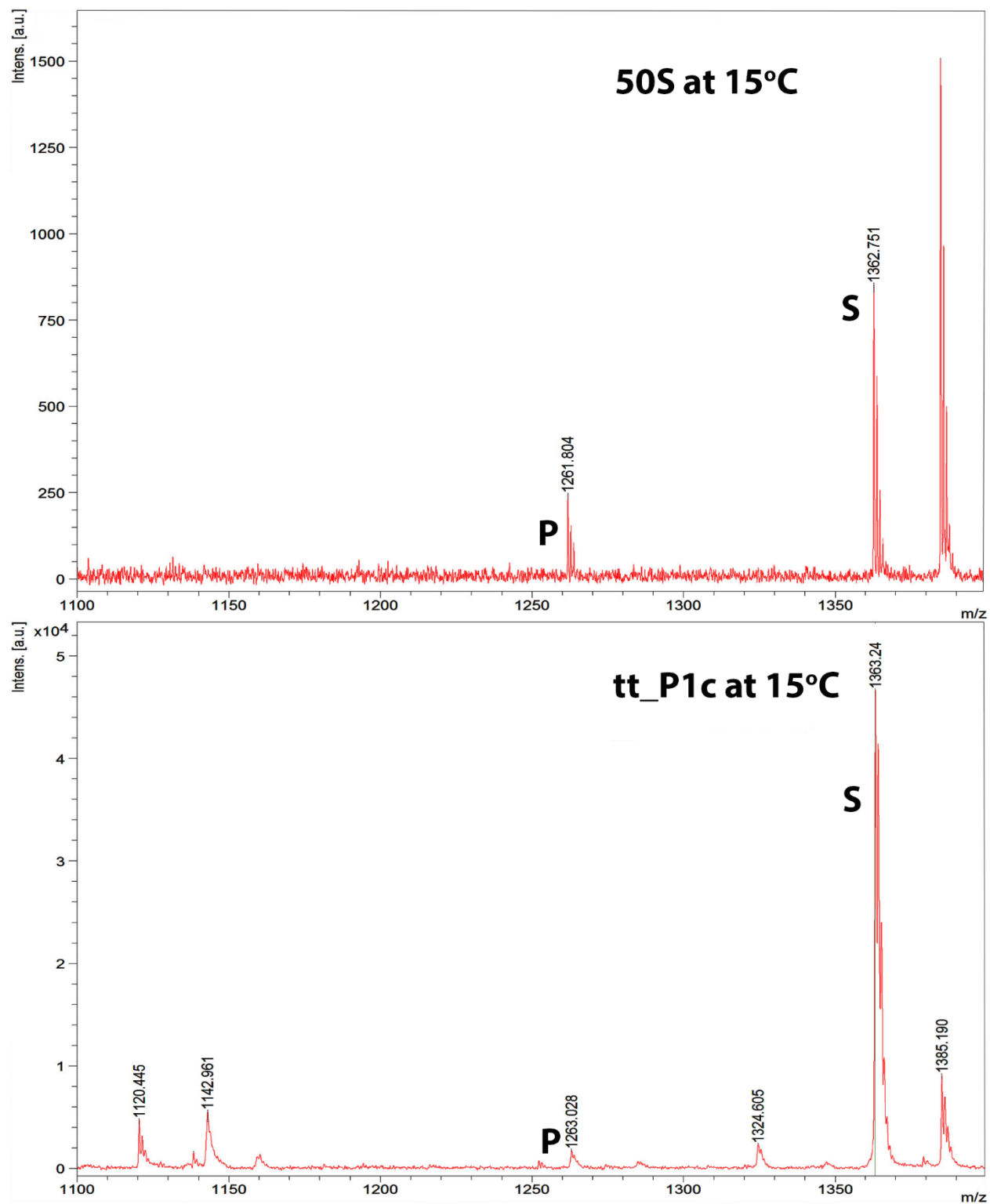
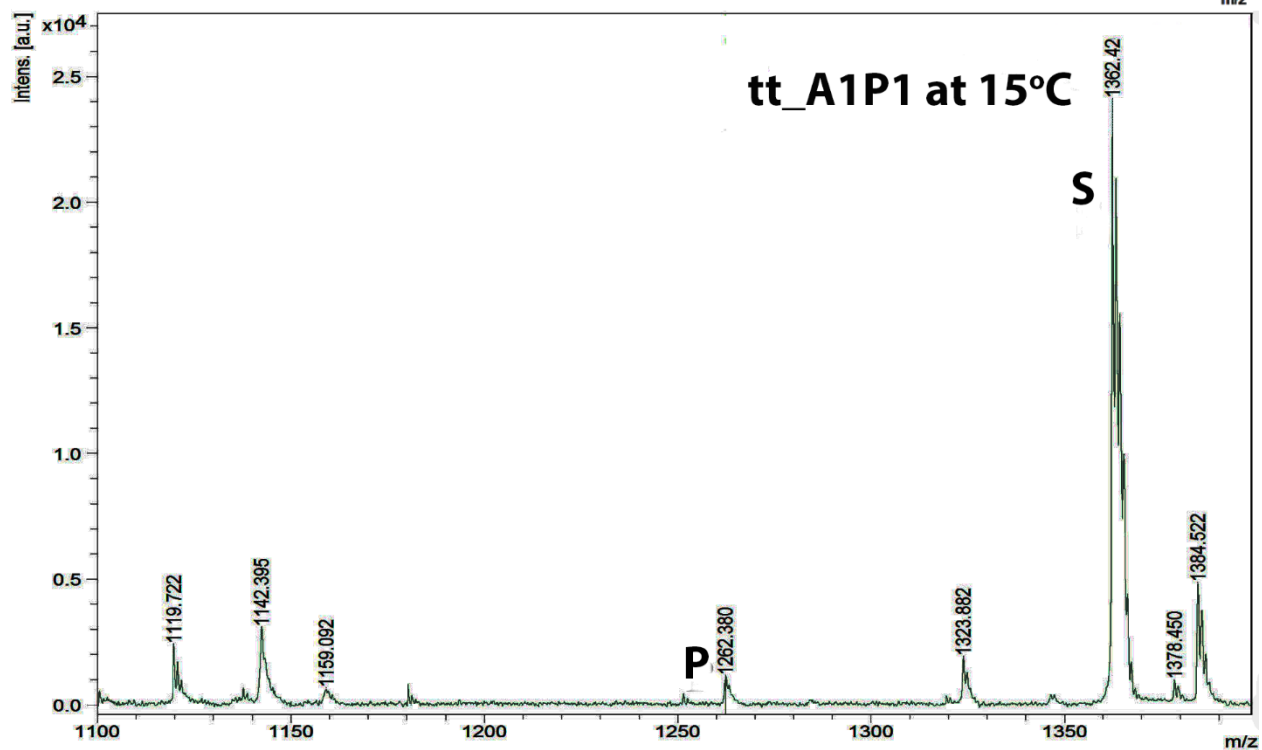
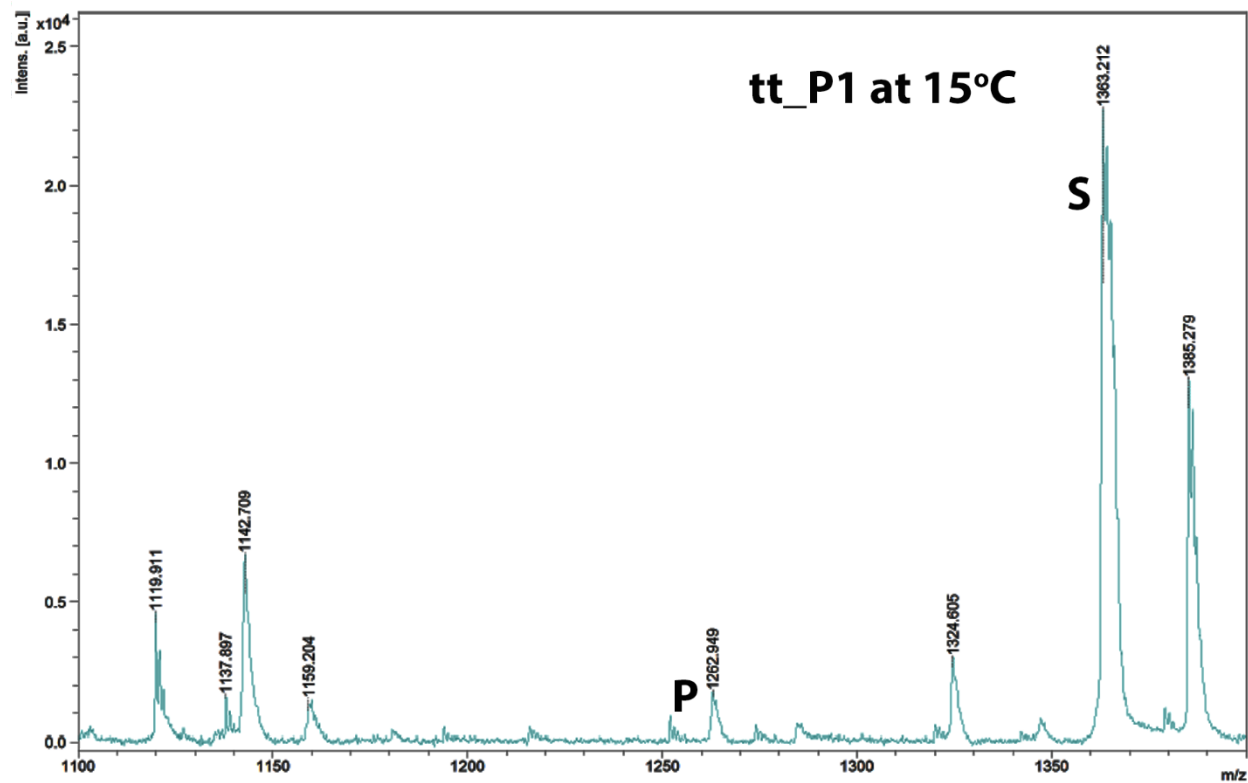


Figure S15: MALDI spectra of the reaction mixtures at 15°C. A. Positive control (50S) and the construct tt_P1c. **B.** The constructs tt_P1, tt_A1P1. **C.** The constructs ef_P1c and sa_P1c.

A



B



C

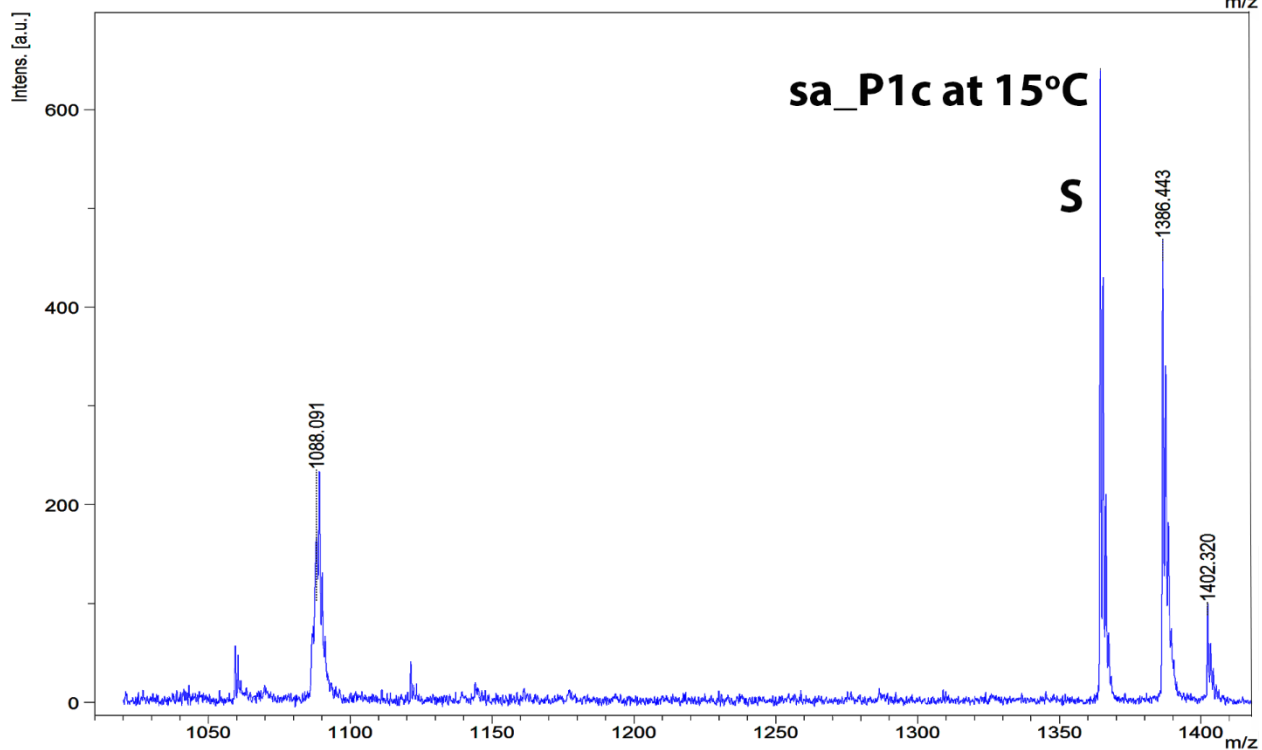
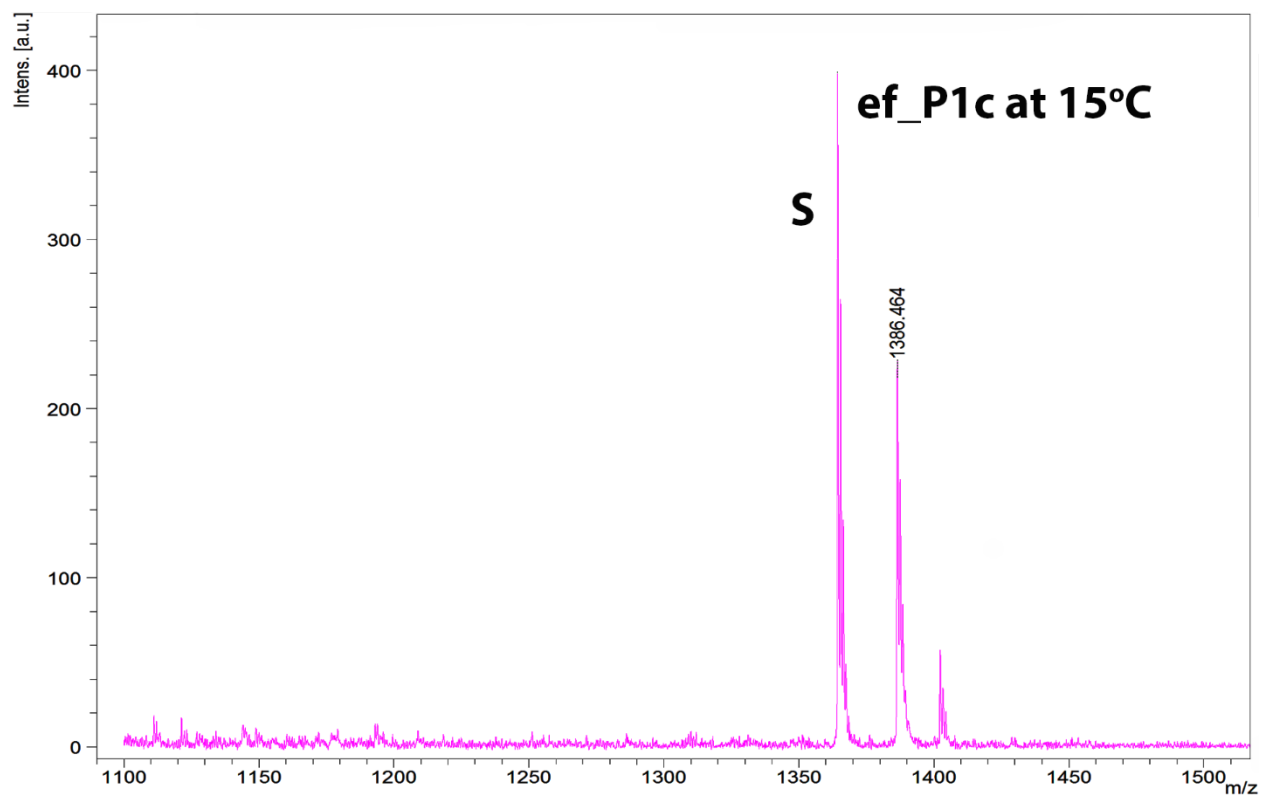
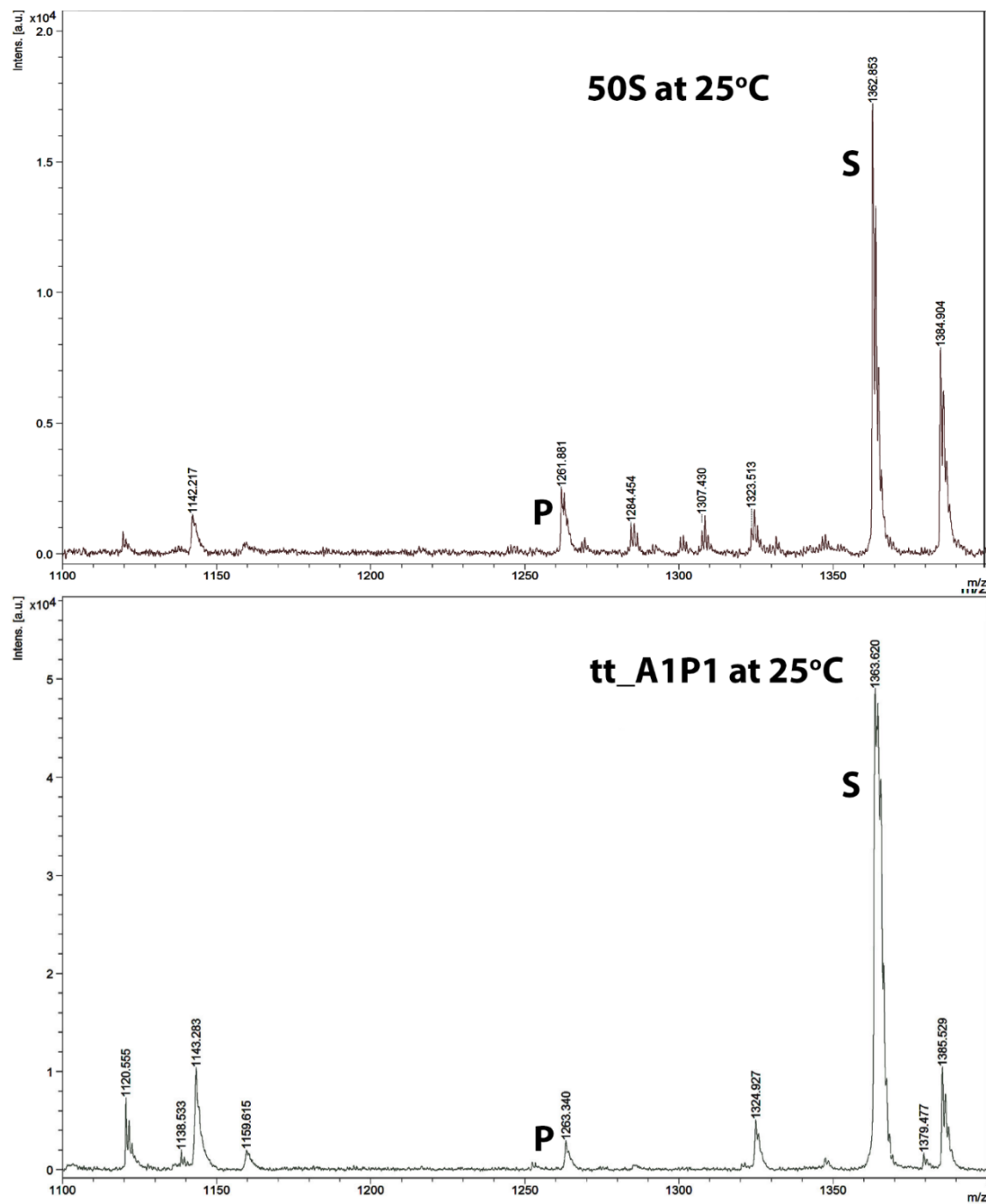
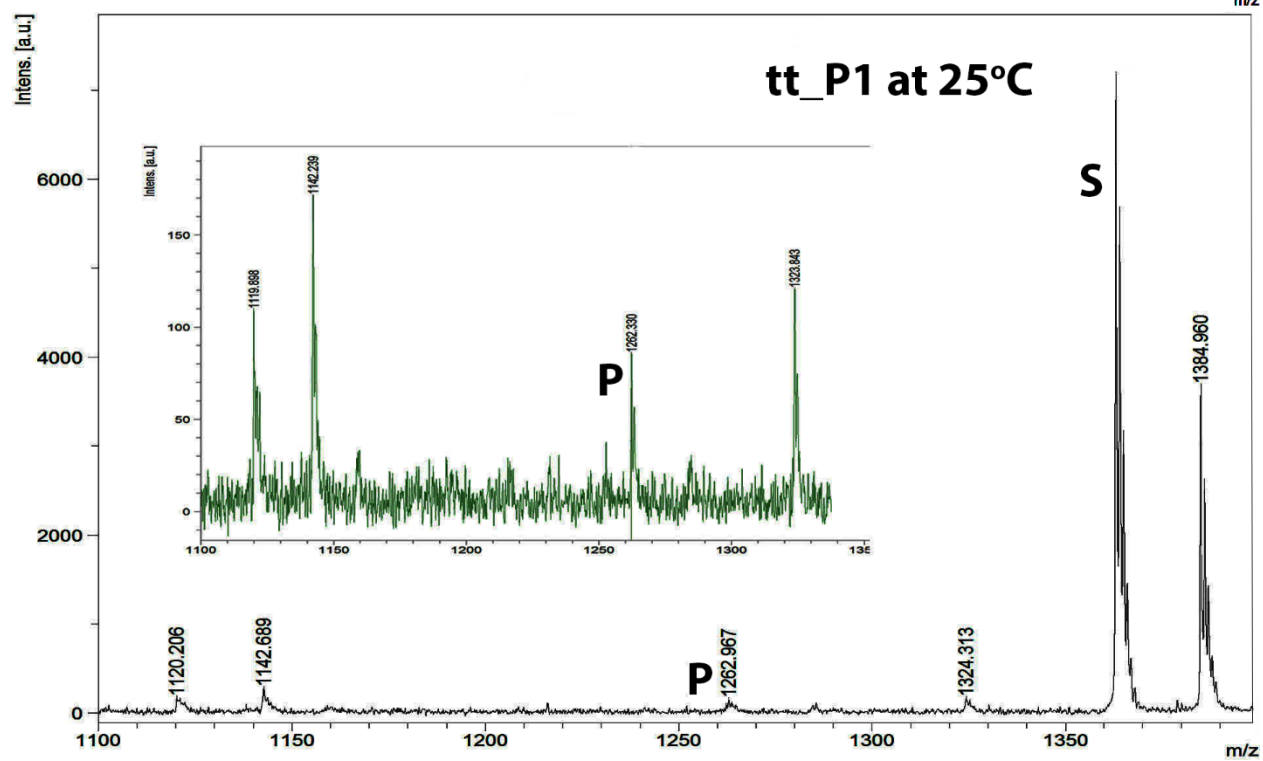
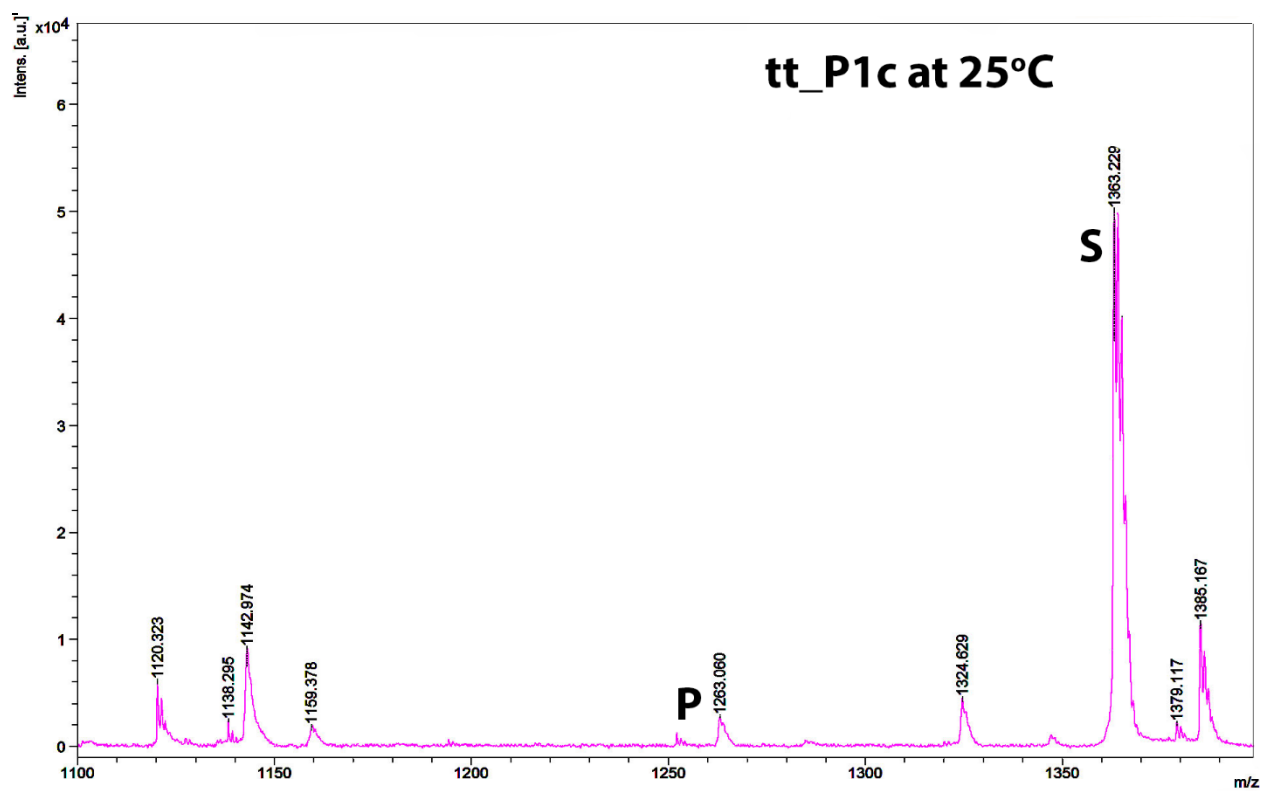


Figure S16: MALDI spectra of the reaction mixtures at 25°C. A. Positive control (50S) and the construct tt_A1P1. **B.** The constructs tt_P1c and tt_P1. **C.** The constructs of ef_P1c and sa_P1c.

A



B

C

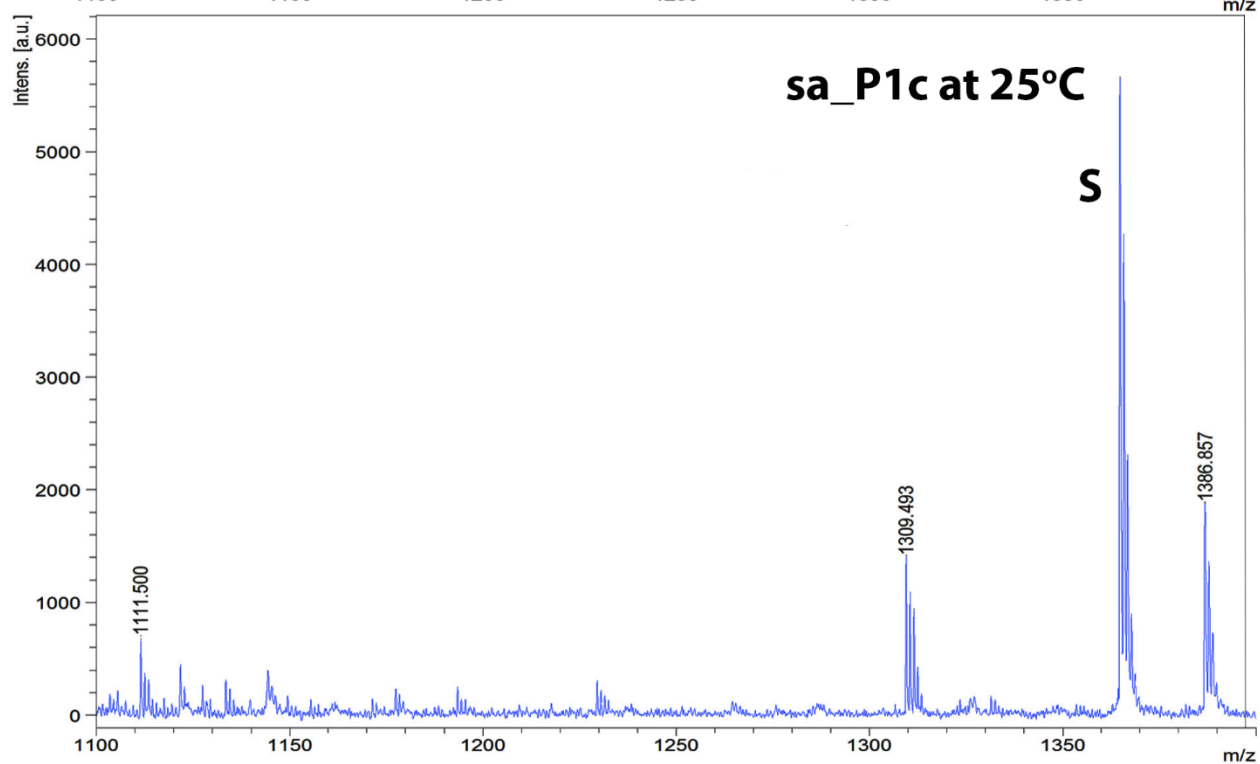
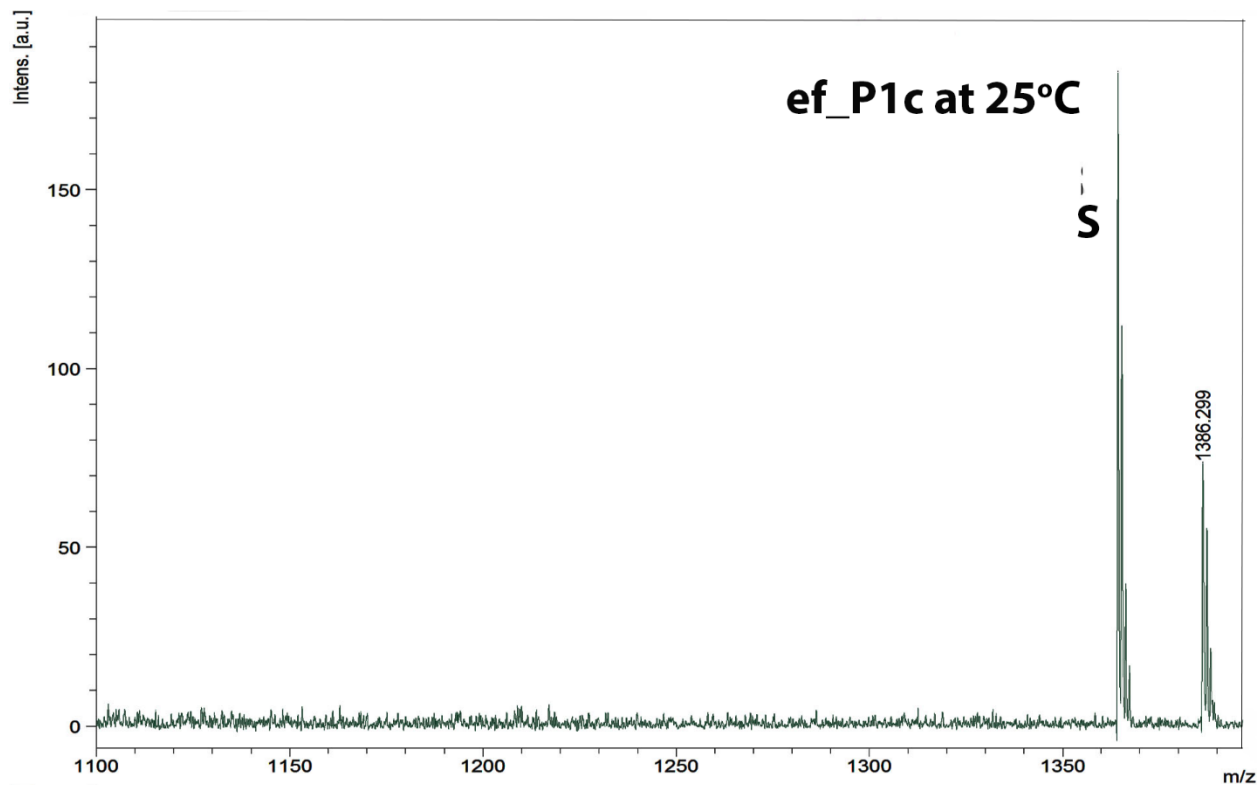
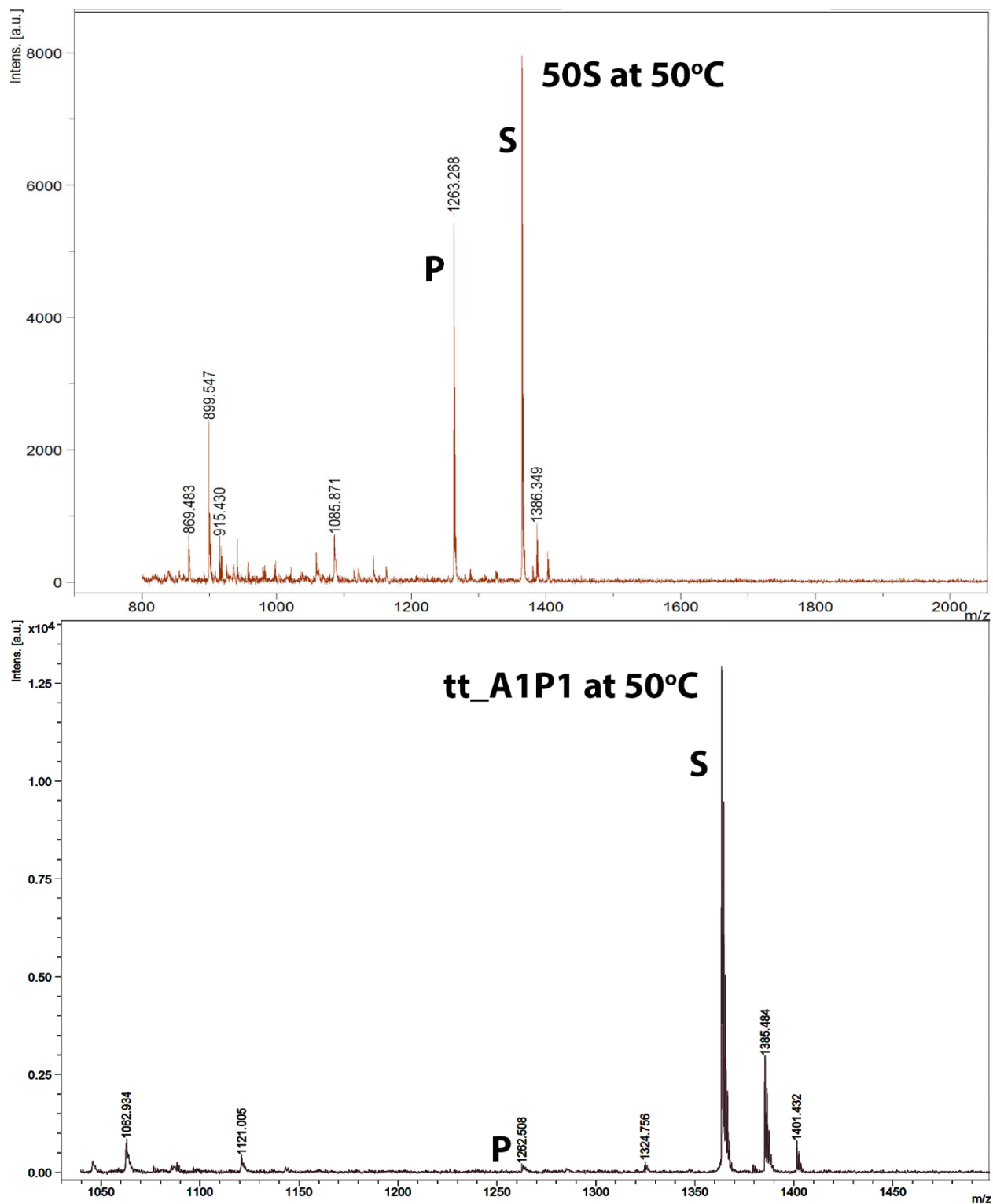
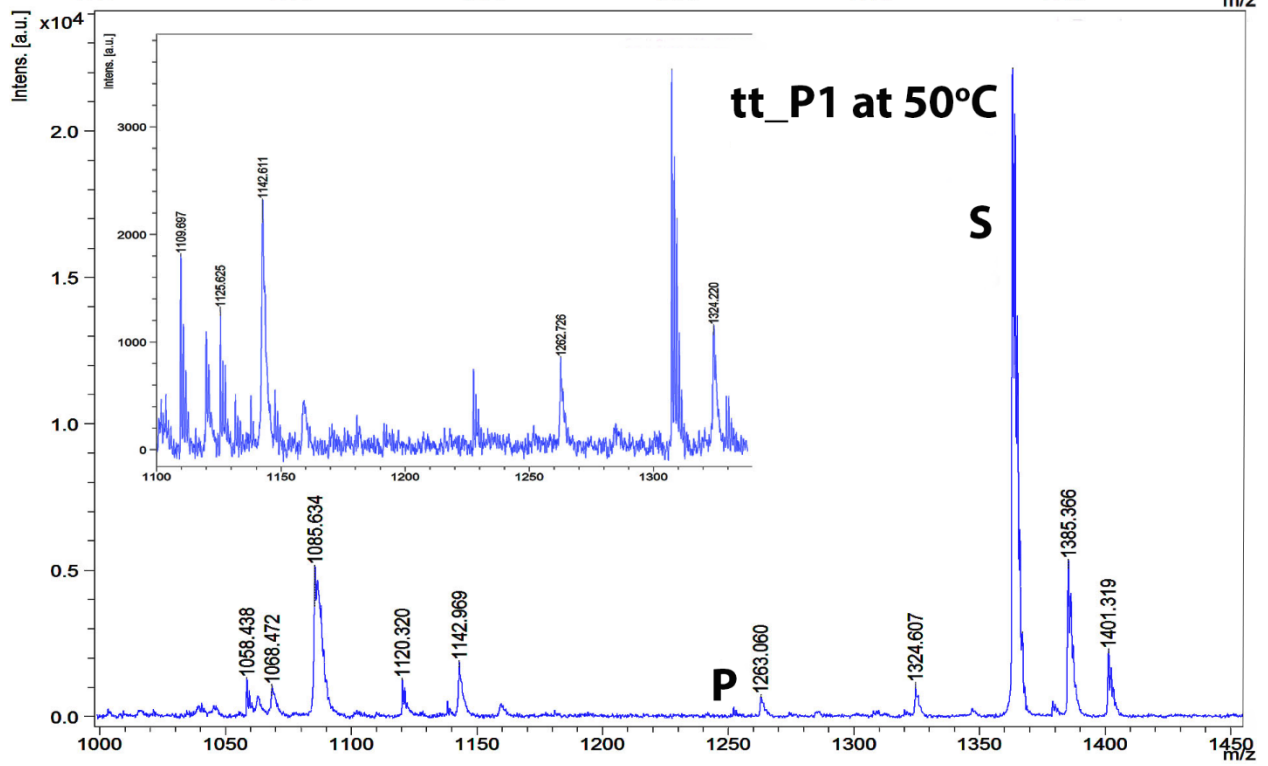
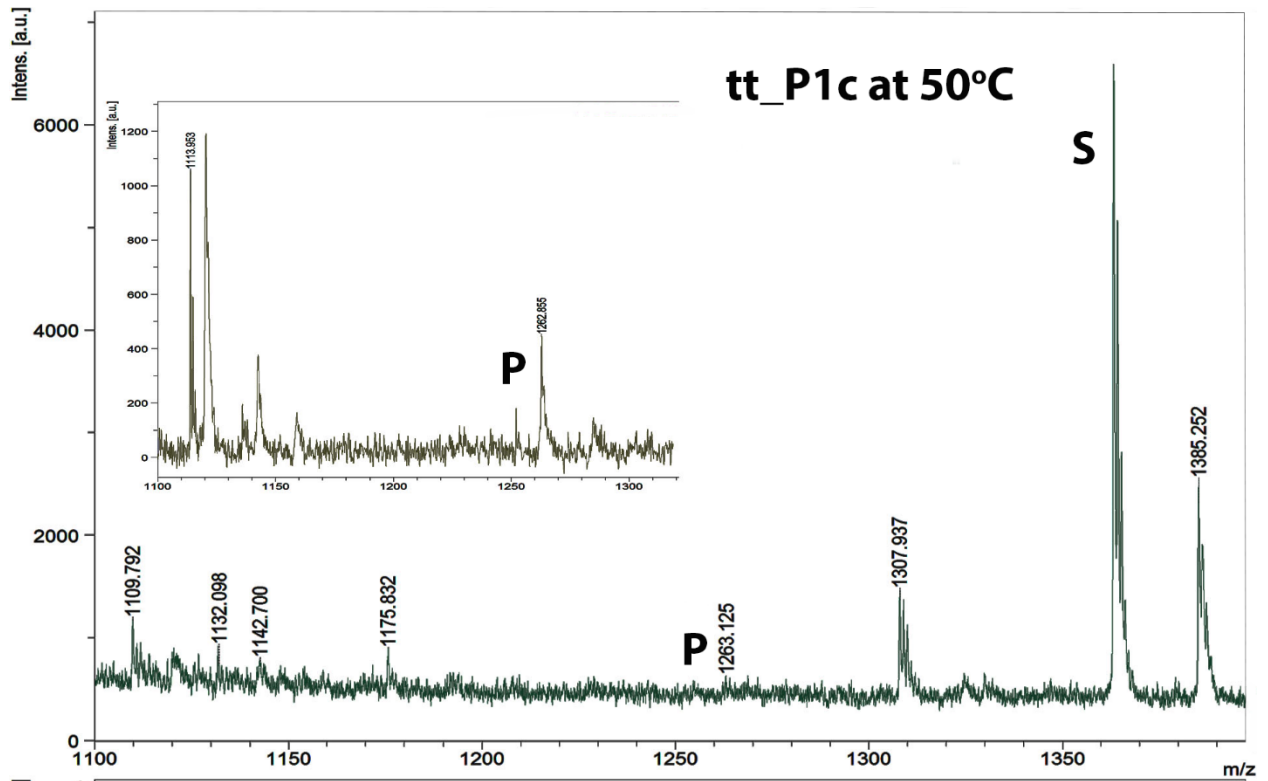


Figure S17: MALDI spectra of the reaction mixtures at 50°C. A. positive control (50S) and the construct tt_A1P1 **B.** The constructs tt_P1c and tt_P1. **C.** The constructs sa_P1c and ef_P1c.

A



B



C

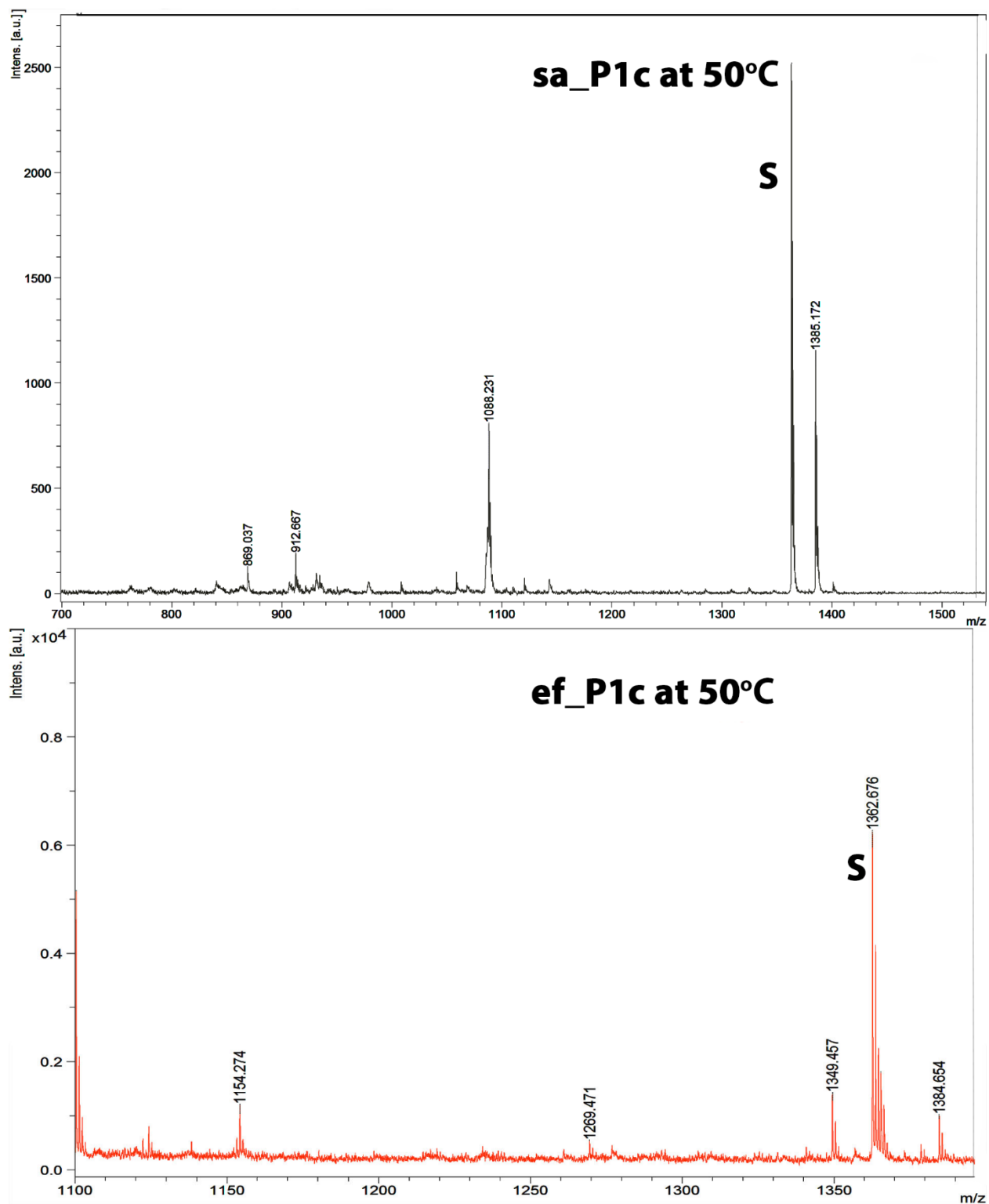


Figure S18: EtBr stained native gel of RNA constructs under various controlled conditions of dimerization. **A-C.** tt_P1c, tt_P1, and tt_A1P1, respectively in the absence of salts. **D.** tt_P1, tt_P1c, and tt_A1P1 with 30 mM KCl and no Mg²⁺ **E.** tt_P1c, tt_P1, and tt_A1P1 with no salts but in the presence of the substrates CCA-pcb and C-Pmn. **F.** tt_A1, tt_A2 and ec_P1c at 20 μM RNA concentration. Gels were imaged at the wavelength of 254nm.

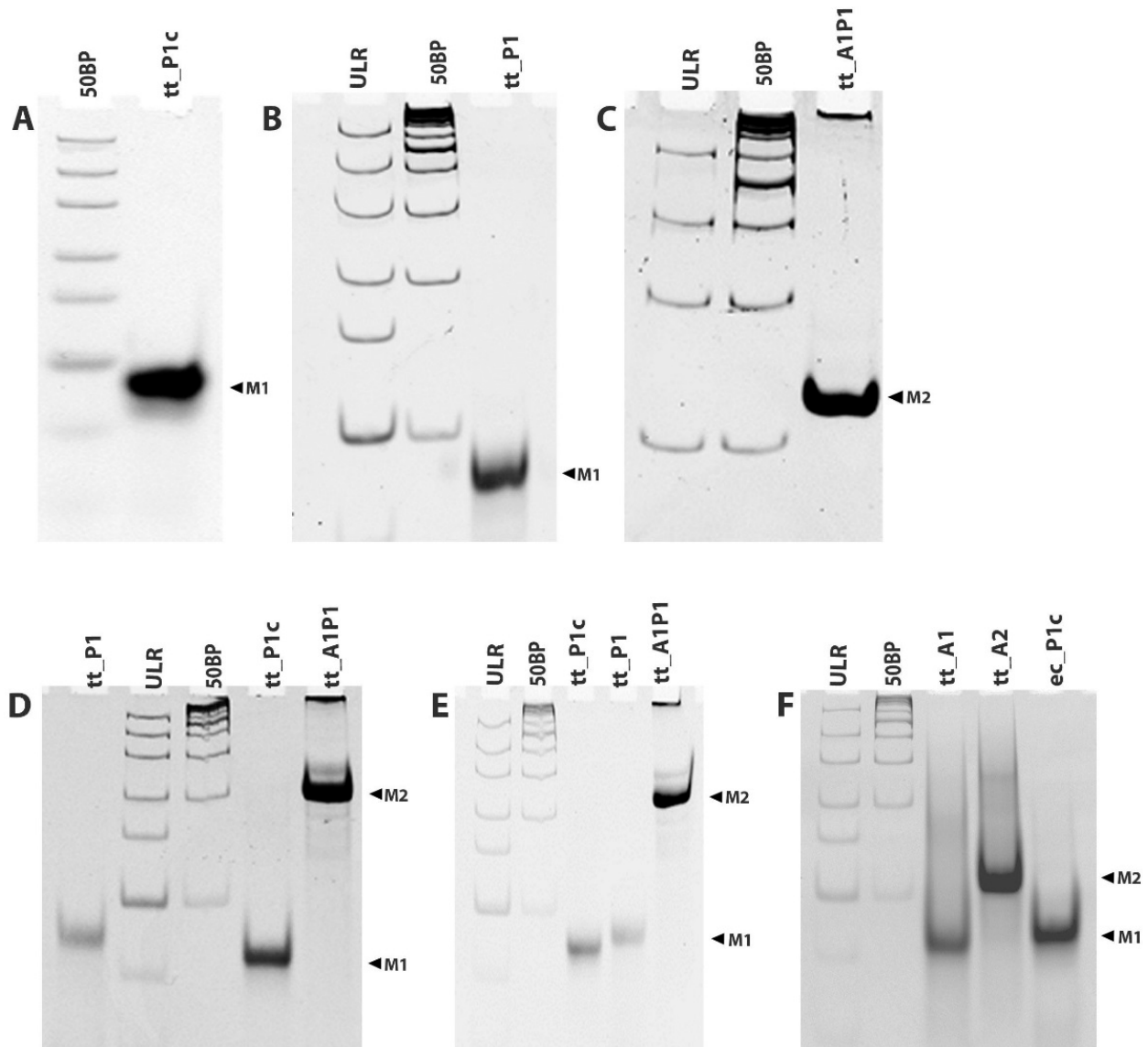
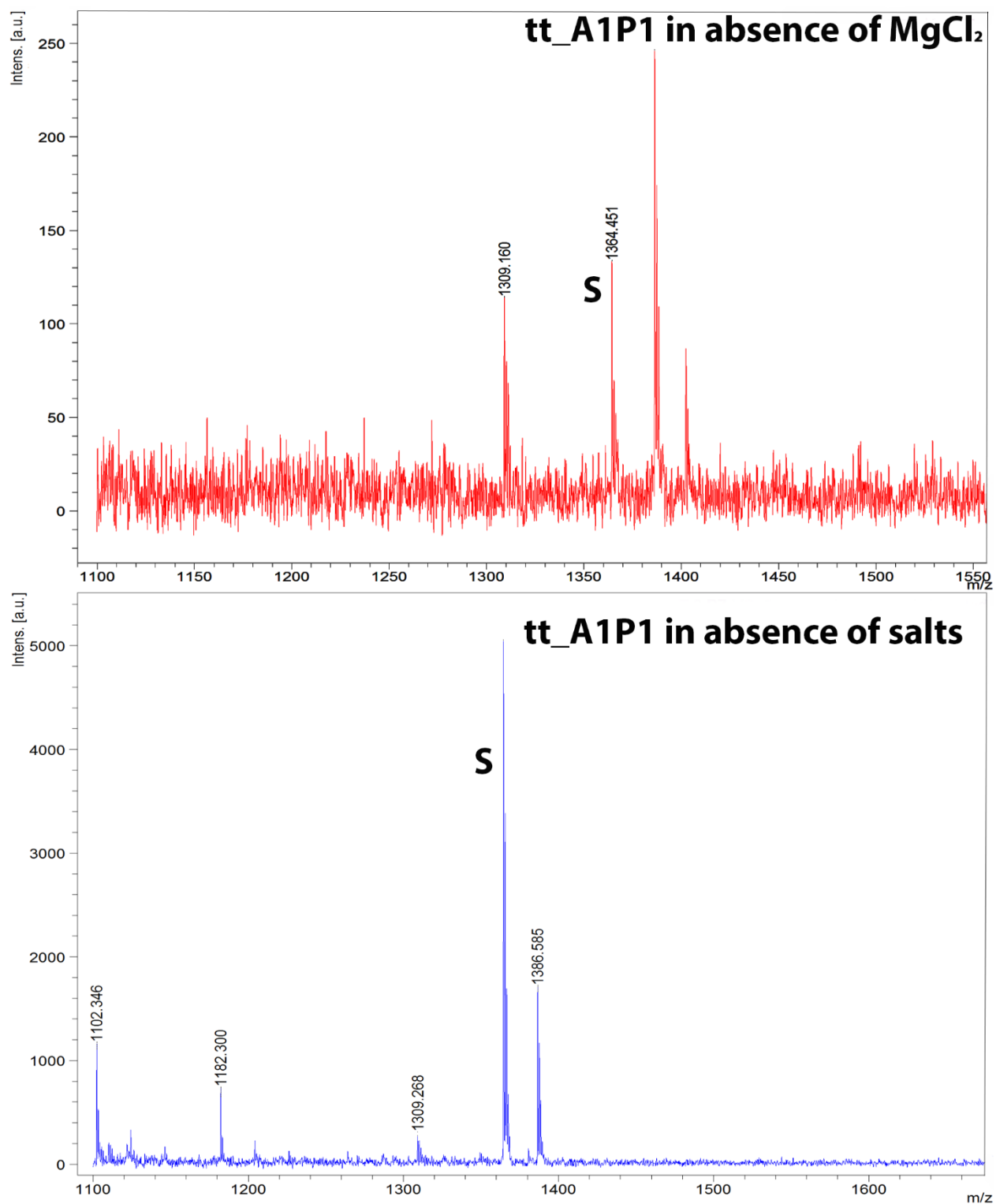
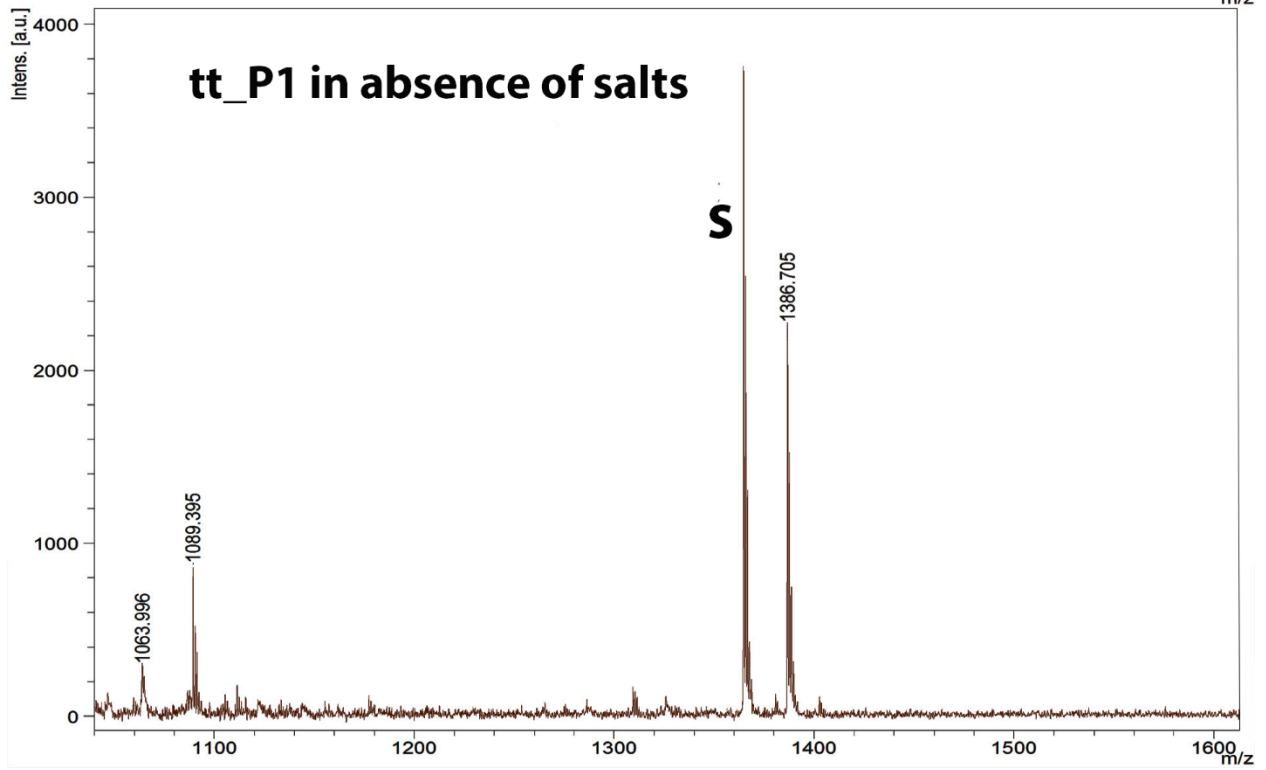
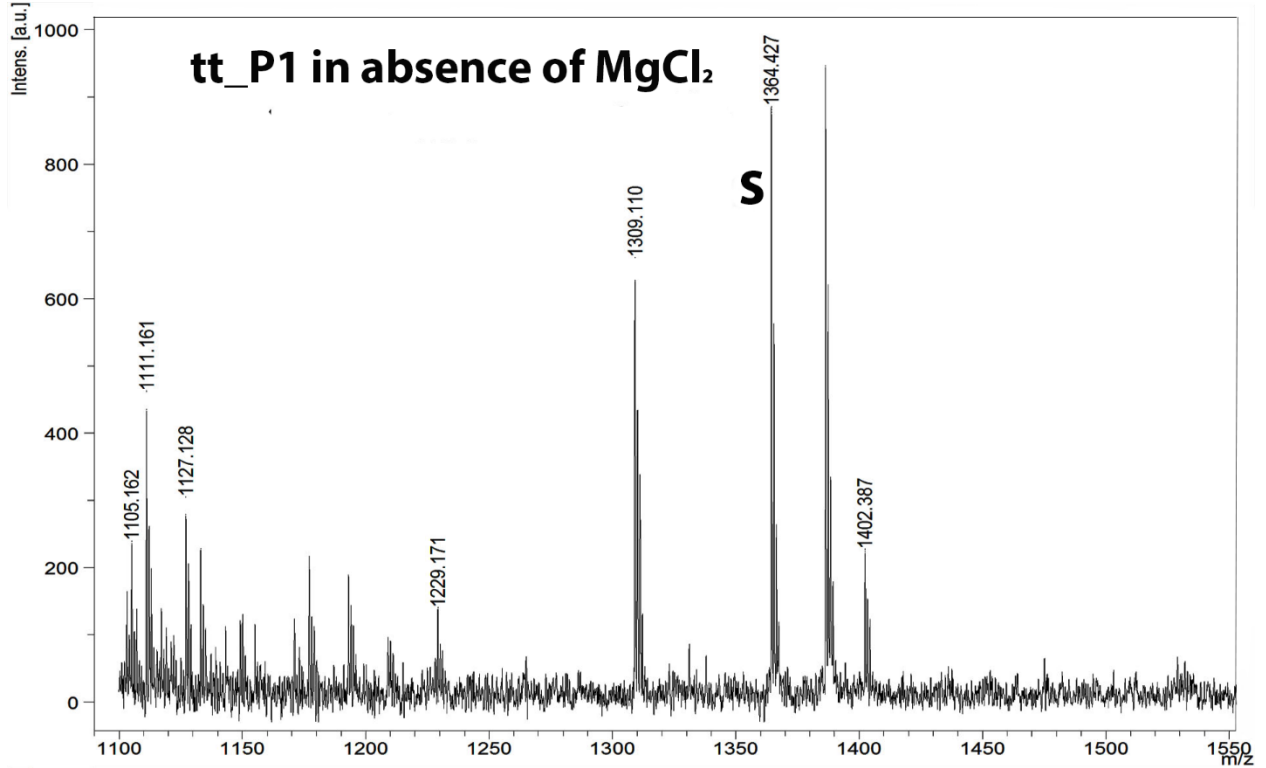


Figure S19: MALDI spectra of tt_A1P1, tt_P1 and tt_P1c reactions performed in the absence of KCl and MgCl₂ (salts) or in the absence of MgCl₂.

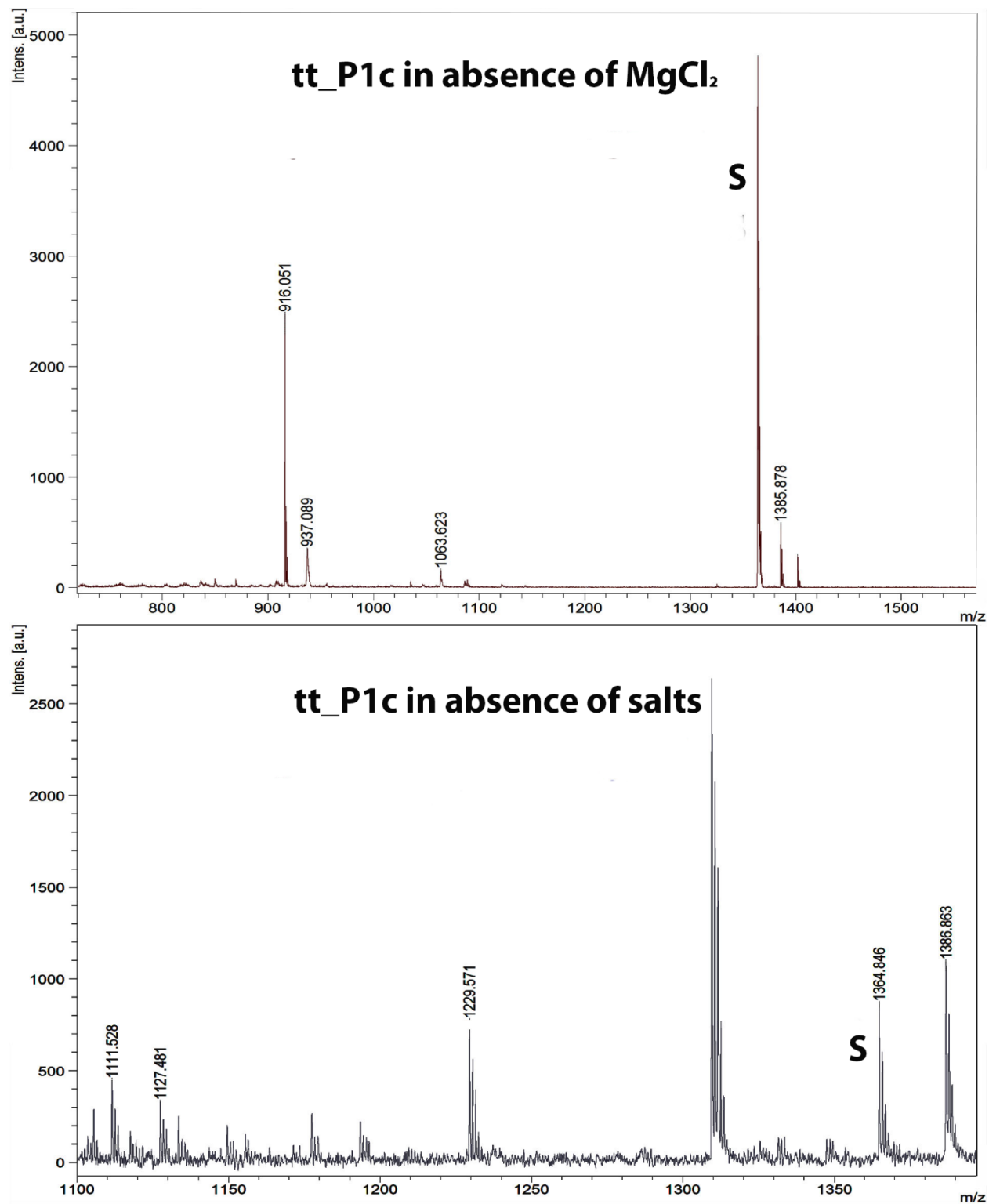
A



B

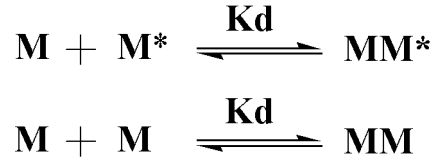


C



Derivation of Equation used to calculate the dimerization constant (K_d)

The reaction between a labelled and unlabelled oligomer can be represented as follows:



[M = Unlabelled oligomer, M* = Labelled oligomer]

Since, the labelled and unlabelled oligomers are same RNA sequences, we assume that they have same K_d

$$\mathbf{K}_d = \frac{[\mathbf{M}][\mathbf{M}^*]}{[\mathbf{MM}^*]} = \frac{[\mathbf{M}][\mathbf{M}]}{[\mathbf{MM}]}$$

$$\mathbf{M}_0^* = [\mathbf{M}^*] + [\mathbf{MM}^*] \quad \mathbf{M}_0^* = \text{Total Concentration of labelled oligomer}$$

$$\mathbf{M}_0 = [\mathbf{M}] + 2[\mathbf{MM}] + [\mathbf{MM}^*] \quad \mathbf{M}_0 = \text{Total concentration of unlabelled oligomer}$$

$$\mathbf{Y} = \frac{[\mathbf{MM}^*]}{[\mathbf{MM}^*] + [\mathbf{M}^*]} * \mathbf{100} \quad \text{where Y = Dimer Fraction \%}$$

Substituting the values in the above equation,

$$\mathbf{Dimer \%} = \left\{ \frac{\sqrt{\mathbf{K}_d(\mathbf{K}_d + 8\mathbf{x})} - \mathbf{K}_d - 2\mathbf{x}}{2(\mathbf{K}_d - \mathbf{x})} \right\} * \mathbf{100} \quad (1)$$

Where X = M_0 or concentration of the unlabelled oligomers

Y = Dimer %

$$\mathbf{where, Dimer}^0\% = \frac{[\mathbf{Dimer}]}{[\mathbf{Dimer} + \mathbf{Monomer}]} * \mathbf{100} \quad (2)$$

Figure S20: Determination of the dimerization constant (K_d) of the RNA constructs using fluorescently imaged EMSA native gels. A-F. Dimerization was assayed at increasing RNA concentrations from 0.0 μ M (left) to 42 μ M (right) (lanes 1-12 RNA concentrations were 0, 0.08, 0.16, 0.24, 0.4, 0.72, 1.36, 2.46, 5.2, 10.4, 21 and 42 μ M respectively), with constant (0.15 μ M) Fluorecein-5-thiosemicarbazide labelled RNA concentration in all lanes. EMSA was performed for tt_P1c, tt_A1P1, tt_P1, tt_P1m, sa_P1c and ef_P1c constructs, respectively.

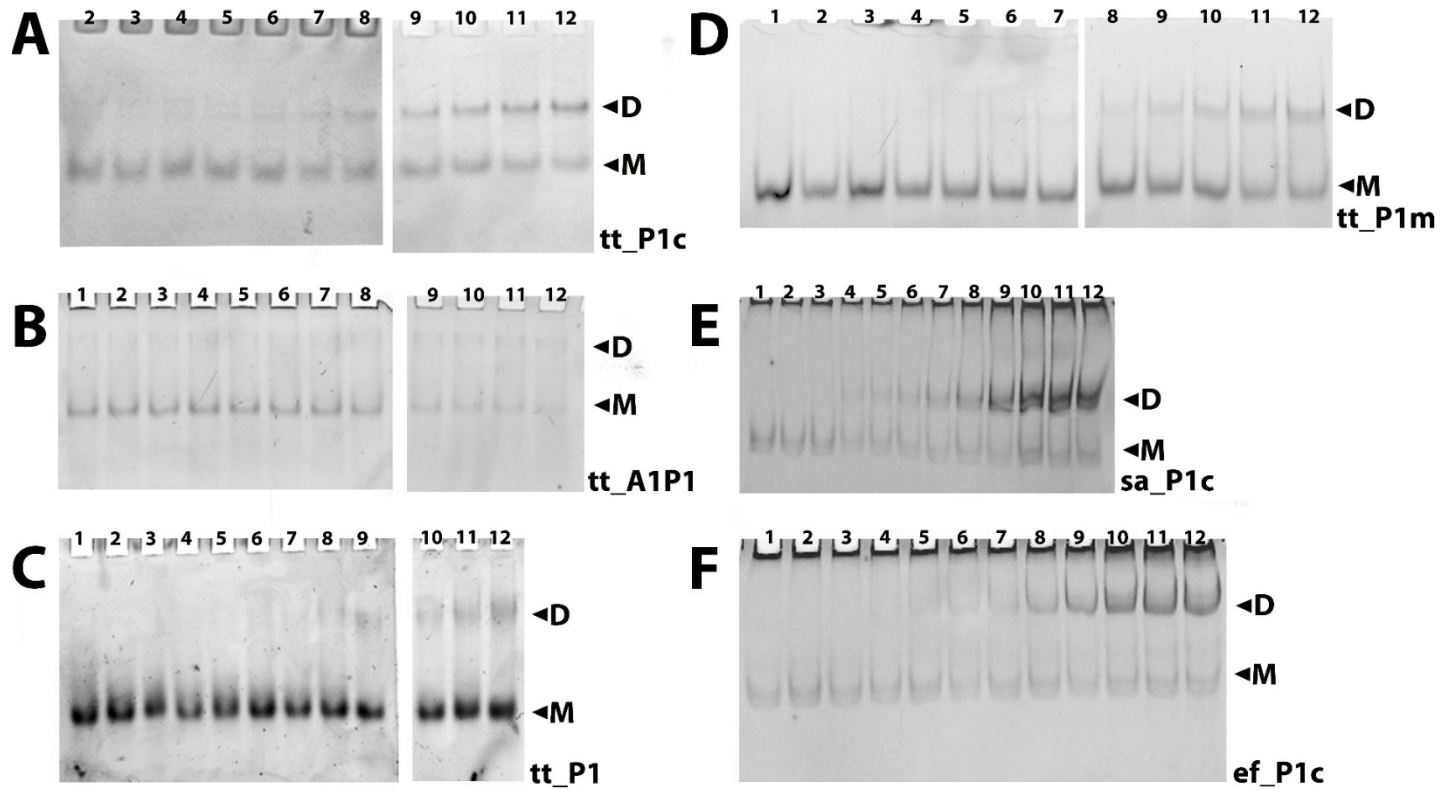


Figure S21: EtBr staining of the native gels used for the determination of the K_d of the RNA constructs. A-G. RNA construct dimerization is shown for tt_P1c, tt_A1P1, tt_P1, tt_P1m, tt_P1n, sa_P1c and ef_P1c, respectively. Dimerization was assayed at increasing RNA concentrations from 0.0 μ M (left) to 42 μ M (right), with constant (0.15 μ M) labelled RNA concentration in all lanes (lanes 1-12 RNA concentrations are: 0, 0.08, 0.16, 0.24, 0.4, 0.72, 1.36, 2.46, 5.2, 10.4, 21 and 42 μ M, respectively). Gels were imaged at a wavelength of 254nm.

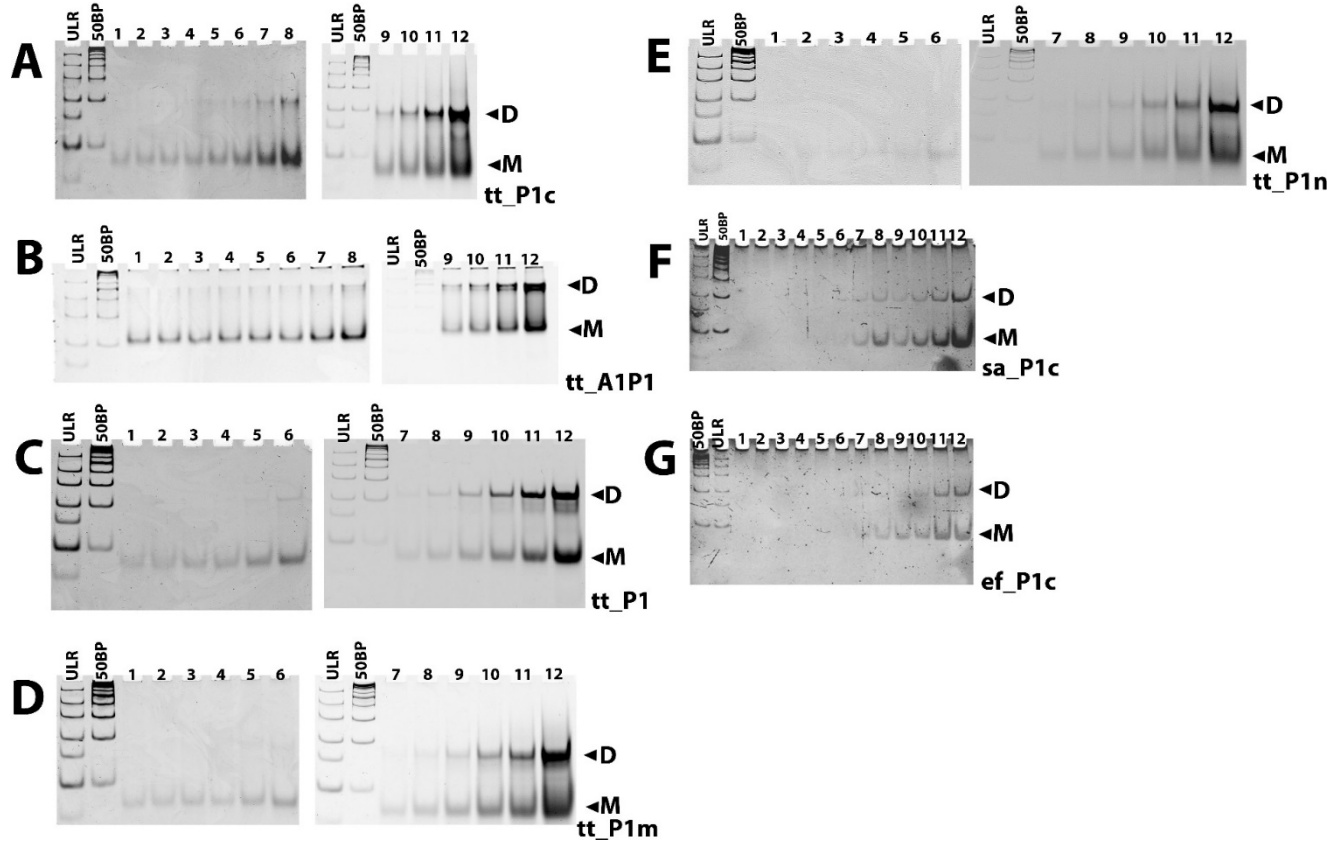


Table S1: Sequences of the RNA constructs

RNA Construct	Sequence*
tt_A1	GAUGUCGGCUCGUCG CUUCGG CGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC
tt_P1	GAAGACCCCGUGGAGCU CUUCGG AGUUACCCCGGGGAU AACAGGCU GAUCU CUUCGG AGGUUUGGCACCUC
tt_A1P1	GAAGACCCCGUGGAGCU CUUCGG AGUUACCCCGGGGAU AACAGGCU GAUCU CUUCGG AGGUUUGGCACCUCGAUGUCGGCUCGUCG CUUCGG CGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC
tt_A2	GAUGUCGGCUCGUCGCAUC CUUCGG GUUGGGCUGUUCGCCAUUAAAGCGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC
tt_P2	GAAGACCCCGUGGAGCUUUACU CUUCGG AGUUUGACUGGGGCGGUCCGGAUAAAAGUUACCCCGGGGAU AACAGGCUGAUCU CUUCGG AGGUUUGGCACCUC
tt_A2P2	GAAGACCCCGUGGAGCUUUACU CUUCGG AGUUUGACUGGGGCGGUCCGGAUAAAAGUUACCCCGGGGAU AACAGGCUGAUCU CUUCGG AGGUUUGGCACCUCGAUGUCGGCUCGUCGCAUC CUUCGG GUUGGGCUGUUCGCCAUUAAAGCGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC
tt_P2c	GAAGACCCCGUGGAGCUUUACU CUUCGG AGUUUGACUGGGGCGGUCCGGAUAAAAGUUACCCCGGGGAU AACAGGCUGAUC GUGA GGUUUGGCACCUC
tt_P1c	GAAGACCCCGUGGAGCU CUUCGG AGUUACCCCGGGGAU AACAGGCU GAUC GUGA GGUUUGGCACCUC
ec-P1c	AAAGACCCCGUGAACCU CUUCGG AGGUACUCCGGGGAU AACAGGCU GAUC GUGA UGUUUGGCACCUC
sa_P1c	GAAGACCCCGUGGAGCU CUUCGG AGCUACCCCGGGGAU AACAGGCU UAUC GUGA GGUUUGGCACCUC
ef_P1c	GAAGACCCCAUGGAGCU CUUCGG AGCUACCCUGGGGAU AACAGGCU UAUC GUGA GGUUUGGCACCUC
tt_P1m	GAAGACCCCGUGGAGCUU UAAG UUACCCCGGGGAU AACAGGCUGAUC GUGA GGUUUGGCACCUC
tt_P1n	GAAGACCCCGUGGAGCU UUAAG UUACCCCGGGGAU AACAGGCUGAUC GUGA GGUUUGGCACCUC
com_tt_P1c	GAGGUGCCAAACCUACGAUCAGCCUGUUAUCCCGGGGUAACUCCGAAGAGCUCCACGGGGUCUUC

* Closing inserted loops are colored; hyper stable CUUCGG loop is colored red, GNRA (GUGA) tetraloop is colored cyan and H74 natural loop is colored green.