Supporting information

Chimeric Streptavidins as Host Proteins for Artificial Metalloenzymes

Michela M. Pellizzoni,[1†] Fabian Schwizer,[1] Christopher W. Wood,[2] Valerio Sabatino,[1] Yoann Cotelle,[3‡] Stefan Matile,[3] Derek N. Woolfson,[2,4,5] Thomas R. Ward[1,*]*

[1]University of Basel, Department of Chemistry, Mattenstrasse 24a, BPR 1096, CH 4002 Basel, Switzerland.

[2]School of Chemistry, University of Bristol, Bristol, BS8 1TS, United Kingdom.

[3]School of Chemistry and Biochemistry, University of Geneva, Quai Ernest Ansermet 30, CH-1211 Geneva, Switzerland.

[4]School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD, UK

[5]BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK

Corresponding Author

*E-mail: thomas.ward@unibas.ch

D.N.Woolfson@bristol.ac.uk

Table of contents

S1. Supplementary natural loop design

Loops were extracted from the PDB (downloaded May 2016). Each structure was loaded into ISAMBARD[1] as an AMPAL object and secondary structure was assigned using the "tag_secondary_structure" method, which annotates the secondary structure of each residue using DSSP.[2-3] Loop regions were defined as any continuous region of backbone structure assigned 'B', 'T', 'S' or unlabelled, which correspond to isolated β-bridge, hydrogen-bonded turn, bend or random coil respectively. Data on each loop was stored in a SQLite database, including:

1. PDB code
2. Residue and chain identifiers for the entering residue (immediately before the loop) and the exiting residue (immediately after the loop)
3. Secondary structure assignments for the entering and exiting residues
4. Crossing angle of the exiting and entering regions of secondary structure
5. The resolution of the structure
6. Cα-Cα distance of the entering and exiting residue
7. Backbone coordinates for the loop region and the entering/exiting residues
8. Amino-acid sequence of the loop

The region between residues A46 and A50 of Sav were targeted to be replaced with a novel loop region. The Cα-Cα distance was calculated for residues A46 and A50, and this was used as the basis for the initial query of the loop database, collecting all loops with a Cα-Cα distance with 0.5 Å of the target, which returned 1946 potential loops. All loops were aligned by minimising the root mean square deviation of distances between corresponding atoms of the entering/exiting residues of the loop and A46 – A50 residues of Sav. Chimeric models of Sav with the candidate loop replacing residues A47 – A49, were created using ISAMBARD by replacing the residues in the polypeptide object representing chain A, with those of the aligned loop. As the number of potential loops was manageable, each model was visually examined for suitable candidate designs.

S2. Cloning procedure, Protein expression and purification

Cloning of $(GGS)_2$ loop mutants:

The gene sequences for streptavidin loop mutants containing one $(GGS)_2$-inserts at positions G48…N49, T66…D67, R84…N85 or A117…N118 were ordered as double stranded DNA-string fragments (GeneArt® Strings™ DNA Fragments, invitrogen™ by life technologies™) containing a four base pair overhang (**ATAT, TATA**) at both ends (Table S1). In addition, new unique restriction sites were introduced into the gene sequences before and after each $(GGS)_2$-insert (Figure S1). This allows a simple subsequent shuffling/combination of the loops (i.e. streptavidin mutants containing two $(GGS)_2$-inserts at different positions; Table S8, entries 5-10).

Table S1: Sequences of streptavidin loop mutants containing a $(GGS)_2$-insert.

Entry	Position of insert	Sequence
1	G48…N49	**ATAT**CATATGGCAAGCATGACGGGTGGCCAGCAGATGGGTCGTGATCAGGCAGGTATTACCGGCACC TGGTATAATCAGCTGGGTAGCACCTTTATTGTTACCGCGGGCGCAGATGGTGCACTGACCGGCACGT ACGAAAGCGCAGTTGGTGGCGGCAGCGGCGGCAGCAATGCAGAAAGCCGTTATGTTCTGACCGGTC GTTATGATAGCGCACCGGCAACCGATGGTAGCGGTACCGCACTGGGTTGGACCGTTGCATGGAAAAA TAACTATCGTAATGCACATAGCGCAACCACGTGGTCAGGTCAGTATGTTGGTGGTGCAGAAGCACGC ATTAACACCCAGTGGCTGCTGACCAGCGGCACCACCGAAGCAAATGCCTGGAAAAGCACCCTGGTTG GTCATGATACCTTTACCAAAGTTAAACCGAGCGCAGCATCAATTGATGCAGCAAAAAAAGCCGGTGT GAATAATGGTAATCCGCTGGATGCAGTTCAGCAGTAATAGGGATCC**TATA**
2	T66…D67	**ATAT**CATATGGCAAGCATGACGGGTGGCCAGCAGATGGGTCGTGATCAGGCAGGTATTACCGGCACC TGGTATAATCAGCTGGGTAGCACCTTTATTGTTACCGCGGGCGCAGATGGTGCACTGACCGGCACGT ACGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTCTGACCGGTCGTTATGATAGCGCACCGGC AACCGGCGGCAGCGGCGGCAGCGATGGTAGCGGTACCGCACTGGGTTGGACCGTTGCATGGAAAAA TAACTATCGTAATGCACATAGCGCAACCACGTGGTCAGGTCAGTATGTTGGTGGTGCAGAAGCACGC ATTAACACCCAGTGGCTGCTGACCAGCGGCACCACCGAAGCAAATGCCTGGAAAAGCACCCTGGTTG GTCATGATACCTTTACCAAAGTTAAACCGAGCGCAGCATCAATTGATGCAGCAAAAAAAGCCGGTGT GAATAATGGTAATCCGCTGGATGCAGTTCAGCAGTAATAGGGATCC**TATA**
3	R84…N85	**ATAT**CATATGGCAAGCATGACGGGTGGCCAGCAGATGGGTCGTGATCAGGCAGGTATTACCGGCACC TGGTATAATCAGCTGGGTAGCACCTTTATTGTTACCGCGGGCGCAGATGGTGCACTGACCGGCACGT ACGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTCTGACCGGTCGTTATGATAGCGCACCGGC AACCGATGGTAGCGGTACCGCACTGGGTTGGACCGTTGCATGGAAAAATAACTATCGTGGCGGCAGC GGCGGCAGCAATGCACATAGCGCAACCACGTGGTCAGGTCAGTATGTTGGTGGTGCAGAAGCACGC ATTAACACCCAGTGGCTGCTGACCAGCGGCACCACCGAAGCAAATGCCTGGAAAAGCACCCTGGTTG GTCATGATACCTTTACCAAAGTTAAACCGAGCGCAGCATCAATTGATGCAGCAAAAAAAGCCGGTGT GAATAATGGTAATCCGCTGGATGCAGTTCAGCAGTAATAGGGATCC**TATA**
4	A117…N118	**ATAT**CATATGGCAAGCATGACGGGTGGCCAGCAGATGGGTCGTGATCAGGCAGGTATTACCGGCACC TGGTATAATCAGCTGGGTAGCACCTTTATTGTTACCGCGGGCGCAGATGGTGCACTGACCGGCACGT ACGAAAGCGCAGTTGGTAATGCAGAAAGCCGTTATGTTCTGACCGGTCGTTATGATAGCGCACCGGC AACCGATGGTAGCGGTACCGCACTGGGTTGGACCGTTGCATGGAAAAATAACTATCGTAATGCACAT AGCGCAACCACGTGGTCAGGTCAGTATGTTGGTGGTGCAGAAGCACGCATTAACACCCAGTGGCTGC TGACCAGCGGCACCACCGAAGCAGGCGGCAGCGGCGGCAGCAATGCCTGGAAAAGCACCCTGGTTG GTCATGATACCTTTACCAAAGTTAAACCGAGCGCAGCATCAATTGATGCAGCAAAAAAAGCCGGTGT GAATAATGGTAATCCGCTGGATGCAGTTCAGCAGTAATAGGGATCC**TATA**

The sequences collected in Table S1 were amplified by PCR (New England BioLabs® Q5® Hot Start HF DNA Polymerase; forward primer: 5'-ATATCATATGGCAAGCATGACGGGTGGCC-3'; reverse primer: 5'-

TATAGGATCCCTATTAAAAACGCGCCAGCACG-3'). Amplified DNA was purified by precipitation from ethanol. DNA pellets were dissolved in water followed by digestion with restriction enzymes (New England BioLabs® NdeI + BamHI-HF, 37°C, 60 min; removal of the ATAT/TATA overhangs). Simultaneously, an empty pET30a vector (Novagen) was digested applying the same conditions. All digested samples were loaded onto an agarose gel (1% agarose) and the desired bands were cut out and purified (Macherey-Nagel NucleoSpin® Gel Clean-up kit). The inserts were then ligated into the digested pET30a vector (New England BioLabs® T4-DNA Ligase, 16°C, overnight, 10-fold molar excess of insert compared to vector). Ligated products were transformed into electro-competent TOP10 (DE3) *E.coli* cells and plated onto LB-agar plates containing kanamycin (50 μg/mL). Colonies were picked and overnight cultures were prepared (LB-medium, 37°C). Cells were harvested and plasmids were isolated (Macherey-Nagel NucleoSpin® Plasmid kit) and sequenced (Microsynth AG).

In order to introduce two $(GGS)_2$ loops into the same gene, the previously obtained single $(GGS)_2$ loop mutants were digested applying different restriction enzymes (Figure S1, Table S2). The digested backbones and inserts were combined in the desired way followed by re-ligation The ligation products were then transformed and sequenced as described for the single $(GGS)_2$ mutants. Overexpression was performed using the *E.coli* TOP10 (DE3) strain (see annexes for SDS gel electrophoresis).

Table S2: Digestion of streptavidin loop sequences applying different restriction enzymes.

entry	$(GGS)_2$ loop at position	restriction enzyme 1	restriction enzyme 2	used fragment
1	G48…N49	AgeI	BamHI	backbone
2	T66…D67	AgeI	BamHI	insert
3	R84…N85	AgeI	BamHI	insert
4	A117…N118	AgeI	BamHI	insert
5	T66…D67	KpnI	BamHI	backbone
6	R84…N85	KpnI	BamHI	insert
7	A117…N118	KpnI	BamHI	insert
8	A117…N118	Pm1I	NdeI	backbone
9	R84…N85	Pm1I	NdeI	insert

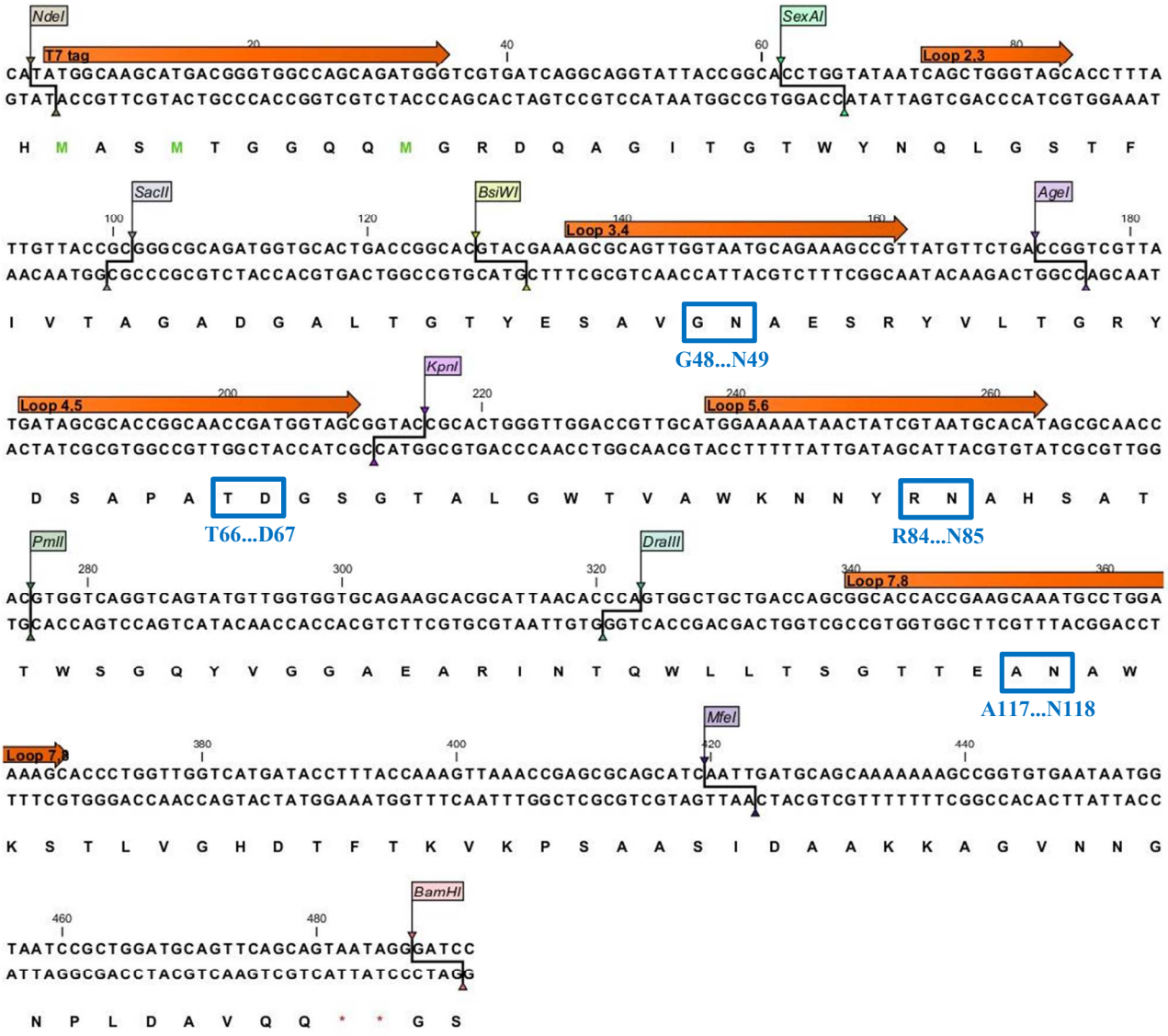


Figure S1: Designed streptavidin sequence containing unique restriction sites before and after the loops. Streptavidin sequence: double stranded DNA (upper lines), amino acids (lower line). Streptavidin loops are displayed as orange arrows. Positions for $(GGS)_2$-inserts are labelled in blue.

In order to expand the diversity of the streptavidin loop mutants $(GGX)_n$, the first serine residue in the loop …G_{48}-G-G-S-G-G-S-N_{49}… was selected for mutagenesis. A primer containing the degenerate "NDT" codon was used, which can encode for 12 different amino acids including aliphatic, aromatic, charged, small and large residues (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Gly and Ser). PCR was performed (New England BioLabs® Phusion® HF DNA polymerase; forward primer: 5'-GGTGGCGGCNDTGGCGGCAG-3'; reverse primer: 5'-

CTGCCGCCAHNACCGCCACC-3') and mutants were analysed as described for the single $(GGS)_2$ mutants. Due to the high GC-content of the streptavidin sequence and the similarity of the GGX and the GGS motif, primers likely annealed at different position (3 amino acids shifted), leading to an elongation of the loop (Table S8, entries 11-18). Overexpression was performed using *E.coli* BL21 (DE3) strain (see annexes for SDS gel electrophoresis).

Expression procedure for $(GGX)_n$ loop mutants

The preculture (10 mL LB-medium, 50 µg/mL kanamycin) was inoculated from glycerol stocks and incubated overnight (37°C, 200 rpm). The main culture (1000 mL LB-rich medium containing: 5.35 g/L yeast extract, 10.70 g/L bactotryptone, 1.77 g/L Na_2HPO_4, 1.70 g/L KH_2PO_4, 1.34 g/L NH_4Cl, 0.36 g/L Na_2SO_4, 0.24 g/L $MgSO_4$, 0.5 g/L glucose, 6.31 g/L glycerol and 50 µg/mL kanamycin) was inoculated with the preculture to a starting OD_{600} of 0.1 and incubated for 2.5 h (37°C, 200 rpm). The temperature was then set to 20 °C and the culture was incubated for ~1 h. At an OD_{600} of ~1.0, the main culture was induced with IPTG (final concentration: 40 µM) and incubated for an additional 20 h (20 °C, 200 rpm). The cell culture was centrifuged (4600 g, 4 °C, 10 min), the obtained cell pellet was resuspended in 20 mL lysis buffer (PBS-buffer (1x, pH 7.45) + 0.5mg/mL DNAse-I + 1mg/mL lysozyme + 0.1 mM phenylmethylsulfonylfluoride) and incubated for 90 min (37 °C, 200rpm). The mixture was frozen (-20 °C, overnight), thawed again and incubated for 3 h (DNAseI digest, 25 °C, 200 rpm). The sample was centrifuged (4600 g, 4 °C, 40 min) and the supernatant (= cell free extract) was analyzed by SDS gel electrophoresis and purified using AKTA *prime* equipped with a sepharose iminobiotin coloumn.[4-5]

Cloning, expression and purification for chimeric 2D_Savs and MP_Sav

pET24(+) plasmid encoding for T7-Tag 2D_Sav's and MP_Sav's were purchased from GenScript USA Inc. and used to transform *E. coli* BL21 (DE3) as expression strain. The consensus sequences of the 2D_Savs are collected in Figure S2.

Name	Consensus sequence
PPR	-GNS**VTYNTLISGLG**KAGR**LEEALELFEEMKE**KGIVPDV-
SH3	-GRVR**WARAL**YDFEALEEDELGFRSGE**VVEVLD**SSNPS**WWTGRL**HN**KLGLFP**ANY**VAPM**MR-
AR	-NGRT**PLHLAARN**GH**LEVVKLLLEA**GADVNAKDK-
HP	-LS**DEDFKAVF**GMTRSAFANLPL**WKQQHLKKE**KGLF-
FDP	-**SPLSEALT**KAN**SPAEAYKASRGAG**-

Figure S2. Summary of the consensus sequences used in this study to generate chimeric 2D_Sav family members. The secondary structures in the motifs (α-helix and β-sheet) are highlighted in bold and underlined-bold, respectively.

Protein overexpression was performed using the autoinduction medium protocol described elswhere.[6-7] After overexpression, the cell were harvested and lysed as described above. The supernatant, and the pellet were analysed by SDS gel electrophoresis, to determine the correct folding of the constructs using B4F (biotin-4-fluorescein),[4] Figure S3.

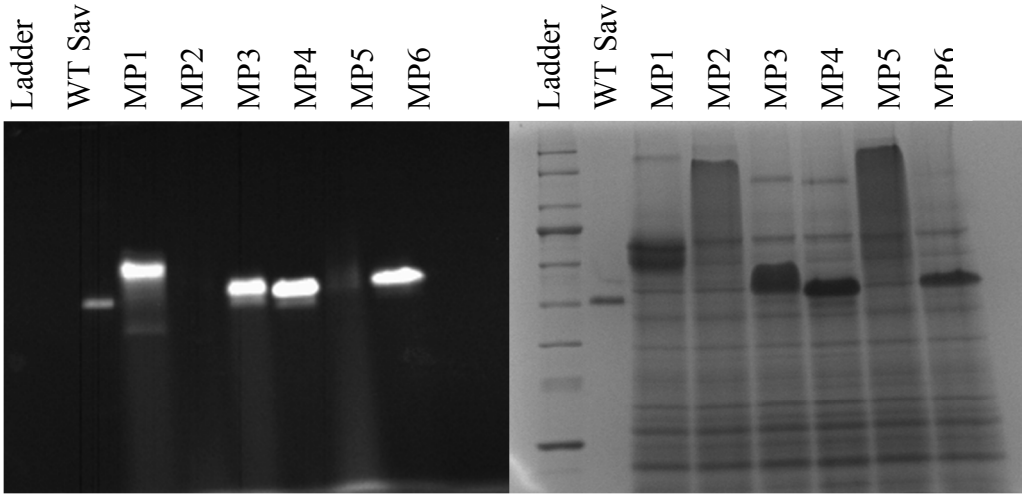


Figure S3. Example of SDS gel electrophoresis with MP_Savs constructs. On the left, acrylamide gel exposed to UV-light. The fluorescent spots (left gel) highlight the biotin-binding capability of the chimeric Savs towards B4F. Comassie blue-stained gel highlighting the efficient overexpression of various MP_Savs (right gel).

Based on the folding behaviour of chimeric Savs, two purification protocols were applied (Table S3).

Protocol 1: The standard purification method, as reported elswhere,[5] was applied for mutants overexpressed as soluble proteins. Briefly, samples were introduced in dialysis bags and equilibrated overnight in guanidinium chloride at 4 °C, then in TRIS 20 mM, pH = 7.4 for 12 h and finally dialyzed against the imminobiotin binding buffer. After 12 h, samples were purified with AKTA *prime*.[5]

Protocol 2: This procedure was applied for mutants that were overexpressed as inclusion bodies. After cell lysis, pellet (700 mg, wet weight) was dissolved in urea (3.4 ml of 8 M stock solution) for 1 h at 4 °C and then heated at 70 °C for 20 minutes. The protein solution was slowly added into 200 ml of refolding buffer (MES 50 mM, pH = 6.0, NaCl 9.6 mM, KCl 0.4 mM, $MgCl_2 \cdot 6H_2O$ 2 mM, $CaCl_2$ 2 mM, arginine 0.5 M and 0.05% polyethylene glycol 3.400) while gently vortexing and incubated (1 h at 4 °C). The resulting solution was dialyzed overnight versus imminobiotin binding buffer and purified with AKTA *prime*.[5]

Table S3. Expression level and purification protocol used for chimeric Savs.

entry	Chimera Sav[a,b]	purification method	mg/l	MW_{th} (g/mol)[c]	$MW_{measured}$ (g/mol)[c]
1	PPR_159	1	60,2	20515.5	20515.8
2	SH3_46-52	2	nd	22877.2	nd
3	SH3_63_70	nd	nd	22819.2	nd
4	SH3_82_87	nd	nd	22843.1	nd
5	SH3_113_117	2	9,4	23016.4	23016.7
6	SH3_159	1	42	23347.7	23347.9
7	AR_46_52	2	3,8	19457.4	nd
8	AR_64_70	2	4	19496.5	19539.4
9	AR_81_84	nd	nd	19650.5	nd
10	AR_115_117	2	1,5	19798.8	nd
11	HP_46_52	1	40	20061.0	20061.6
12	HP_63_70	2	1	19960.0	nd
13	HP_64_70	1	9,5	20045.2	20045.0
14	HP_82_84	1	1,9	20325.4	nd
15	HP_113_117	nd	nd	20157.3	nd
16	FDP_46_52	1	66	18313.0	18313.3
17	FDP_64_70	2	3,5	18352.1	18394.8
18	FDP_81_84	1	4,2	18506.2	nd
19	FDP_115_117	2	1,5	18654.4	nd
20	MP 1	1	13,5	16484.0	16483.5
21	MP 2	not tested	nd	16526.0	nd
22	MP 3	1	16,0	16645.1	16645.0
23	MP 4	1	27,4	16625.1	16625.0
24	MP 5	not tested	nd	16754.3	nd
25	MP 6	1	10	16792.1	16792.6
26	MP 7	1	16,1	17064.6	17065.0
27	MP 8	not tested	nd	17081.6	nd
28	MP 9	1	27,3	17098.6	17099.5
29	MP 10	not tested	nd	17046.7	nd
30	MP 11	1	25,2	17101.6	17102.0
31	MP 12	1	24	17260.7	17261.2
32	MP 1_*K-A_K-F*	not tested	nd	16445.9	nd
33	MP 3_*K-A*	1	12,9	16588.0	16588.1
34	MP 3_*K-F*	1	11,5	16664.1	16663.6
35	MP 6_*K-A_R-F*	1	9,0	16726.0	16725.9
36	MP 9_*K-F*	1	14,0	17117.0	17117.5
37	MP 12_*R-A_K-F*	1	12,2	17194.0	17195.0
38	MP 3_K121F	1	12,1	16664.0	16664.1
39	MP 3_*K-A*-K121F	1	18,4	16607.0	16607.1

[a] The numbering of the residues of Sav was kept identical to the Sav WT numbering, even though the insertion may be placed before the position of the mutation. [b] Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details). [c] Calculated molecular weight of the monomer.

The constructs were characterized by ESI-TOF mass-spectroscopy (see Annexes).

S3. General procedures and analytical methods

Asymmetric transfer hydrogenation

Procedure. 10 µl of proteins stock solution in Milli-Q H_2O (200 µM free binding site), was added to 185 µl of buffer (MOPS 0,66 M, pH = 7,5 and formate 3,33 M) followed by the addition of 5 µl of the biotinylated metal complex [Cp*Ir(biot-L)Cl] 1 from a stock solution (0.4 mM in DMSO). The solution was mixed for 20 min at 37 °C and 800 rpm in a thermo-mixer for precomplexation. Finally, 5 µl of substrate **4** or **6** stock solution (400 mM in DMSO) was added and the mixture was stirred at 37 °C for 16 h. Subsequently, 100 µl of an NaOH 20 % solution was added to the reaction mixture, followed by the addition of 1 ml of CH_2Cl_2. After mixing, the organic phase was separated and dried with anhydrous Na_2SO_4. Solids were separated through centrifugation (2 min at 21 000 g), and the supernatant was analyzed by using HPLC or GC.

Catalytic reactions performed with substrate **4** were analysed by HPLC using Chiral stationary phase (CHIRALPAK IC 250 x 4.6 mm, 5 µm; CH_2Cl_2/*i*-PrOH/DEA 99,5:0,5:0.1, 1 ml/min, 25 °C, detection at 265 nm). t_{ret} = 7.5 min ((*S*)-**5**), t_{ret} = 8.4 min (**4**), t_{ret} = 11.6 min ((*R*)-**5**) Conversions for substrate **4** were determined by comparison of the signal area of product and substrate peaks under consideration of the experimentally determined response factor (1.26).[8]

The conversion of substrate **6** was determined by GC-FID (Agilent J&W CAM, 30 m × 0.32 mm, 0.25 µm; 150 °C isotherm, 1.7 ml He/min; injector: 250 °C, split 100; detector: 240 °C). t_{ret} = 3.64 min (amine **7**), t_{ret} = 3.30 min (imine **6**).

For ee determination, the product **7** was converted to the corresponding trifluoroacetamide. Trifluoroacetic anhydride (TFAA, 200 µl) was added to the GC-samples used for the determination of the conversion and the volatiles were removed to near dryness using a gentle stream of N_2. The residue was dissolved in a small amount of CH_2Cl_2 (30-100 µl) and analyzed by GC-FID on a chiral stationary phase (Agilent CP-Chirasil-DEX CB, 25 m × 0.25 mm, 0.25 µm; 140 °C isothermal, 1.7 ml He/min; injector: 300 °C, split 100; detector: 275 °C). Product **7**: t_{ret} = 9.1 min ((*S*)-**7**), t_{ret} = 9.4 min ((*R*)-**7**). Assignment of absolute configuration was based on literature data.[9]

Table S4. ATHase results obtained for the 2D_Sav combined with [Cp*Ir(biot-L)Cl] **1.**[a] Error margins in square brackets.

entry	2D_Sav	substrate	ee (*R*)	TON
1	No protein	**4**	0	197 [0.2]
2	No protein	**6**	0	0
3	Wt Sav	**4**	45 [0.5]	170 [2.0]
4	Wt Sav	**6**	76 [1.2]	22 [0.7]
5	HP_46-52	**4**	59 [0.6]	223 [2.2]
6	FDP_46-52	**4**	51 [1.0]	340 [5.1]
7	HP_46-52	**6**	80 [1.8]	162 [2.3]
8	FDP_46-52	**6**	82 [2.7]	158 [2.5]
9	HP_64-70	**4**	22 [2.0]	14 [1.2]
10	SH3_113-117	**4**	18 [0.7]	12 [0.6]
11	SH3_159	**4**	52 [1.1]	50 [0.3]
12	PPR_159	**4**	48 [0.4]	55 [1.7]
13	AR_64-70	**4**	3 [nd]	5 [nd]

[a] reaction conditions: substrate 10 mM, 16 h, 37°C, [Cp*Ir(biot-L)Cl] **1** 10 µM, Sav's Free Binding Site (FBS) 20 µM, MOPS 0,6 M, pH = 7,5, formate 3 M, V_{tot} = 200 µL.

Table S5. ATHase results obtained for the MP_Sav introduced between position A46 and A50 combined with [Cp*Ir(biot-L)Cl] **1**.[a] Error margins in square brackets.

entry	MP_Sav[b,c]	substrate	ee (*R*)	TON
1	No protein	**4**	0	197[0.2]
2	No protein	**6**	0	0
3	Wt Sav	**4**	45 [0.5]	170 [2.0]
4	Wt Sav	**6**	76 [1.2]	22 [0.7]
5	Sav K121F	**4**	-13 *(S)* [0.1]	970 [8.7]
6	MP 1	**4**	48 [1.1]	9 [0.1]
7	MP 3	**4**	53 [2.3]	6 [0.2]
8	MP 4	**4**	42 [0.1]	67 [2.5]
9	MP 9	**4**	53 [0.5]	82 [3.4]
10	MP 11	**4**	45 [0.4]	61 [5.5]
11	MP 12	**4**	62 [0.2]	15 [1.6]
12	MP 3_*K-A*	**4**	46 [0.9]	192 [3.0]
13	MP 3_*K-F*	**4**	33 [2.1]	168 [5.5]
14	MP3_*K-A*-K121F	**4**	14 [0.2]	462 [8.7]
15	MP6_*K-A,R-F*	**4**	46 [1.5]	146 [1.7]
16	MP 9_ *K-F*	**4**	49 [0.4]	198 [2.1]
17	MP12_*R-A-K-F*	**4**	53 [0.6]	56 [1.4]
18	MP 3_K121F	**4**	15 [0.6]	568 [9.3]
19	MP 3_ *K-A*	**6**	79 [1.6]	57 [2.5]
20	MP 6_*K-A_R-F*	**6**	77 [1.5]	100 [1.9]
21	MP 9_*K-F*	**6**	80 [2.0]	71 [1.0]

[a] reaction conditions: substrate 10 mM, 16 h, 37°C, [Cp*Ir(biot-L)Cl] **1** 10 µM, Sav's Free Binding Site (FBS) 20 µM, MOPS 0,6 M, pH = 7,5, formate 3 M, V_{tot} = 200 µL. [b] The numbering of the residues of Sav was kept identical to the WT numbering, even though the insertion may be placed before the position of the mutation. [c] Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details).

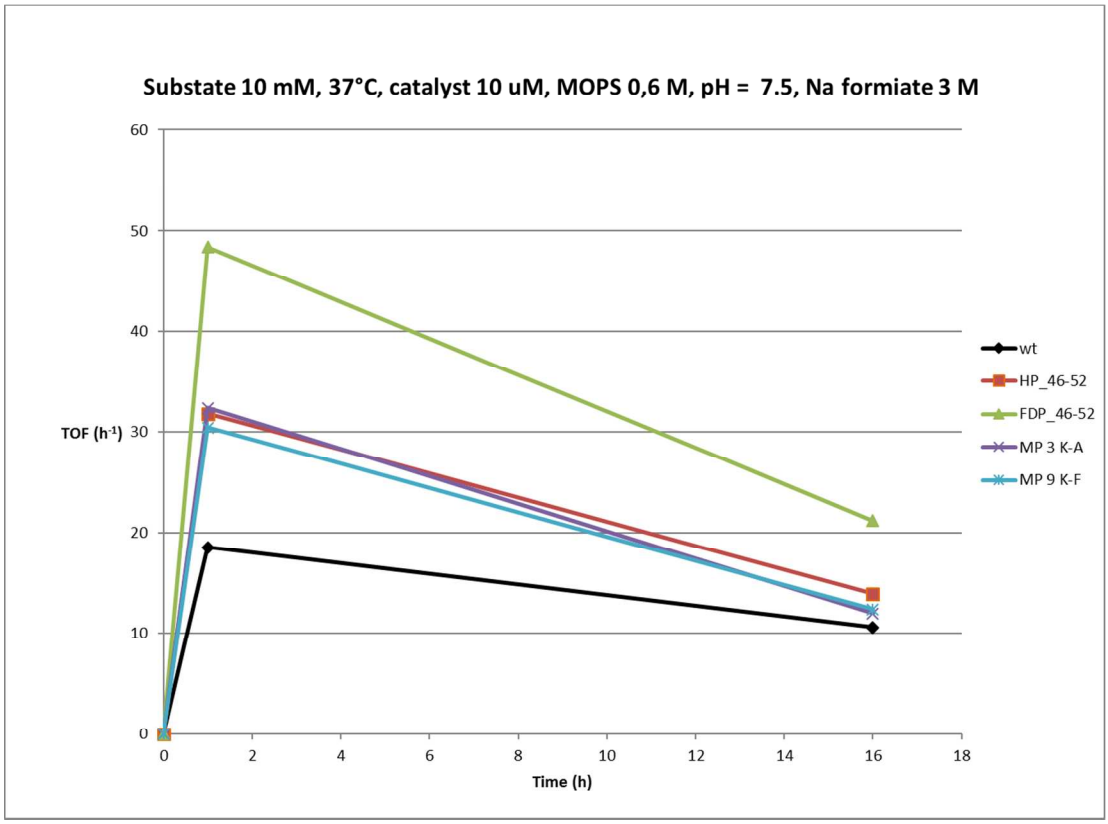


Figure S4. Qualitative kinetic data in ATH of substrate **4**.

Ring closing metathesis

Procedure. In a 1.2 mL glass vial, 10 µl of protein stock solution (400 µM in MQ water) and 5 µl **biot Ru 2** (400 µM in DMSO) were added to 85 µl of reaction buffer (100 mM AcOH/AcONa, 0.5 M $MgCl_2$ pH = 4.0) and incubated at 37 °C for 20 min. After incubation, the substrate **8 (**100 µl of a 20 mM stock solution) was added to the reaction mixture and the reaction was stirred at 37 °C for 16 hours at 1000 rpm. After reaction, methanol (750 µl) and the internal standard (50 µl of a 100 µM stock solution) were added to the reaction mixture and transferred to Eppendorf tubes for centrifugation (10 min, 10000 g). The supernatant (20 µl) was then transferred into HPLC vials containing MQ water (980 µl) and the sample analysed by UPLC-MS for the quantification of product **9**.

Samples were analysed by UPLC-MS for the quantification of product **9**, t_{ret} = 0.96 min.

The calibration curve was established using **9**-d_9 (i.e containing a perdeuterated $–N(CD_3)_3$ moiety) as internal standard. Samples were prepared in triplicate using a concentration range of 0-1 µM of the metathesis product **9** and a fixed concentration of internal standard **9**-d_9 (500 nM) and analysed by UPLC-MS. The linear regression was obtained by plotting the product/standard ratio integration as a function of the product concentration as reported in Figure S5.

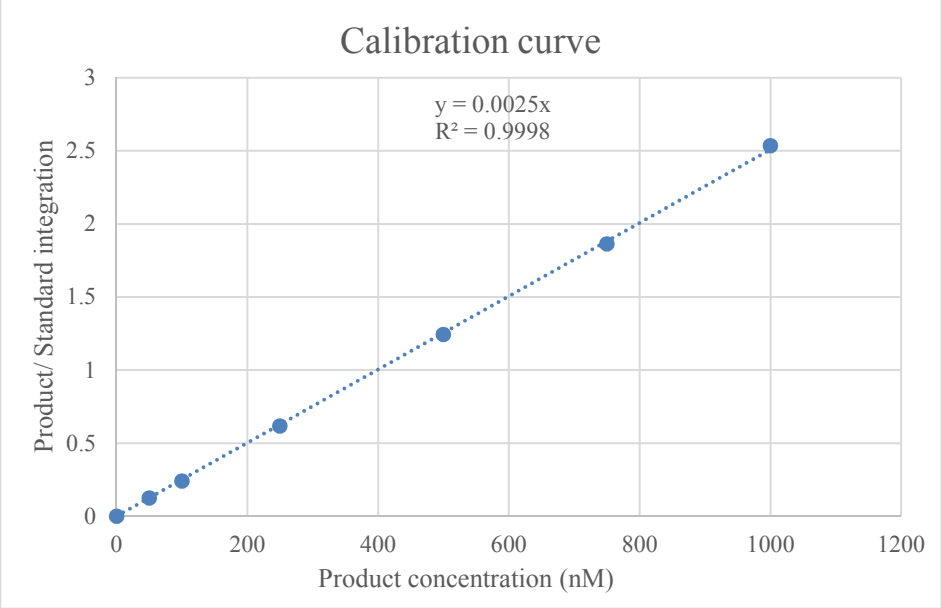


Figure S5. Calibration curve for the quantification of the ring closing metathesis reaction using **9**-d_9 as internal standard.

Table S6. Results for ring closing metathesis reaction using 2D_Sav and MP_Sav.[a] Error margins in square brackets.

entry	protein[b,c]	TON
1	No protein	38 [1.5]
2	WT Sav	87 [1.1]
3	K121F	105 [0.1]
4	HP_64-70	56 [0.5]
5	PPR_159	64 [0.7]
6	HP_46-52	18 [0.6]
7	AR_64-70	14 [0.9]
8	SH3_113-117	28 [1.7]
9	SH3_159	22 [0.3]
10	FDP_46-52	18 [0.3]
11	MP 1	19 [0.3]
12	MP 3	49 [0.8]
13	MP 3_*K-A*	61 [0.1]
14	MP 3_*K-F*	59 [1.2]
15	MP 4	42 [0.2]
16	MP 6_*K-A_R-F*	16 [0.8]
17	MP 9	47 [1.5]
18	MP 9_*K-F*	40 [1.2]
19	MP 11	49 [0.5]
20	MP 12	21 [0.2]
21	MP 3_*K-A*-K121F	98 [3.0]
22	MP 3_K121F	88 [1.6]
23	MP 12_*R-A_K-F*	25 [0.4]

[a] reaction conditions: substrate **8** 10 mM, **biot-Ru 2** 10 µM, 16 h, 37°C, Chim_Sav and MP_Sav 20 µM free binding site, acetate buffer 100 mM, $MgCl_2$ 0.5 M pH = 4.0, V_{tot} = 200 µL. [b] The numbering of the residues of Sav were kept as the WT numbering, even though the insertion may be placed before the position of the mutation. [c] Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details).

Anion-π catalysis

Procedure. Stock solutions of substrates **10** (40 mM), **11** (400 mM) and **biot NDI 3** (2 mM) were prepared in CD_3CN. Solutions of substrates **10** should be freshly prepared as decarboxylation to afford **13 D** build up upon aging. Solutions were prepared by mixing successively streptavidin WT or mutants (100 µL, 2 mM, glycine buffer pH 3, 0.2 µmol), biotinylated ligand **3** (50 µL, 0.1 µmol), substrates **10** (25 µL, 1 µmol) and **11** (25 µL, 10 µmol)

and stirred at 20 °C. After 24h, the mixture was extracted with $CDCl_3$ (0.7 mL), dried over Na_2SO_4, filtered and analyzed by ^{1}H-NMR Figure S6.

Integrals associated with the protons alpha to the carbonyl group of the addition product **12 A** (δ 3.01 ppm; d, 2H) were compared to the combined integration of all –OCH_3 protons present in the substrates and products (i.e., from δ 3.84–3.79 ppm) to determine conversion of the reaction. The spectroscopic data obtained for product **12 A** were identical to the ones reported in the literature.[10] Crude mixtures were analysed by chiral HPLC, (*i.e.*, column: CHIRALPAK ID column; mobile phase: *n*-Hexane/*i*-PrOH 60/40, 0.5 mL/min, rt; detection: 254 nm, Figure S7).

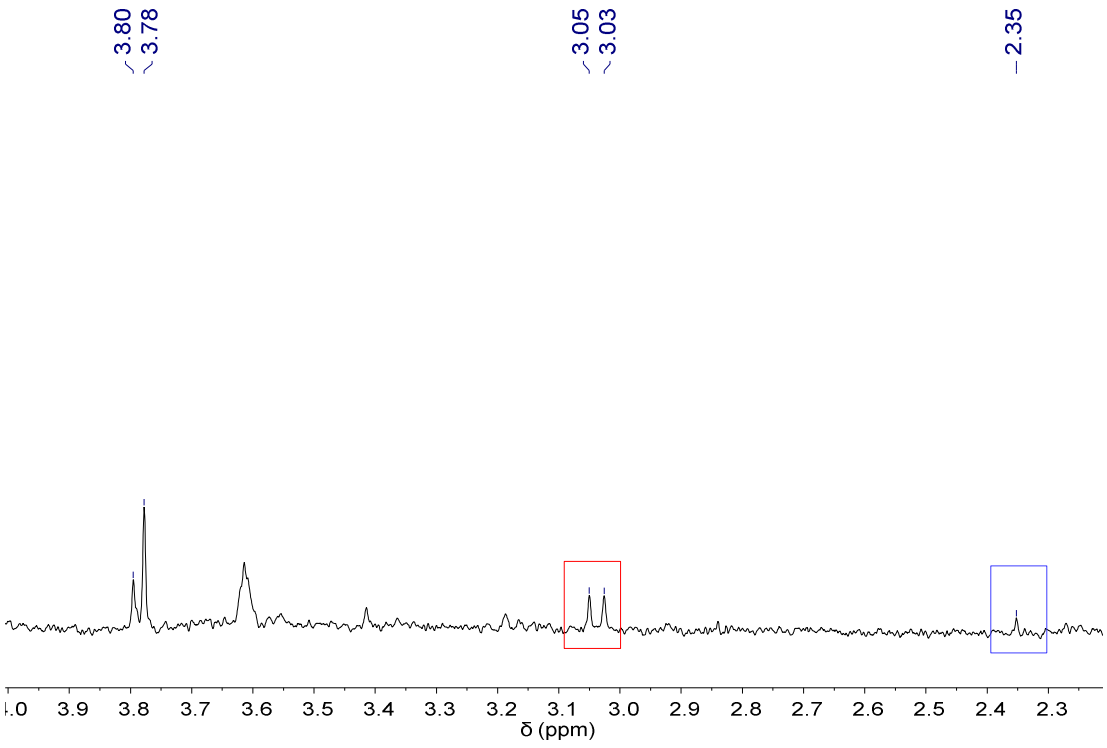


Figure S6. ^{1}H NMR spectra of a mixture of **10** (5 mM), **11** (50 mM) and PPR_159 · **biot-NDI 3** (10% mol) in glycine buffer/CD_3CN 1:1 at 20 °C extracted in $CDCl_3$. The red rectangle highlights the diagnostic peak corresponding to the protons alpha to the carbonyl group of the addition product **12 A**, the blue rectangle highlights the methyl protons from the decarboxylation product **13 D**.

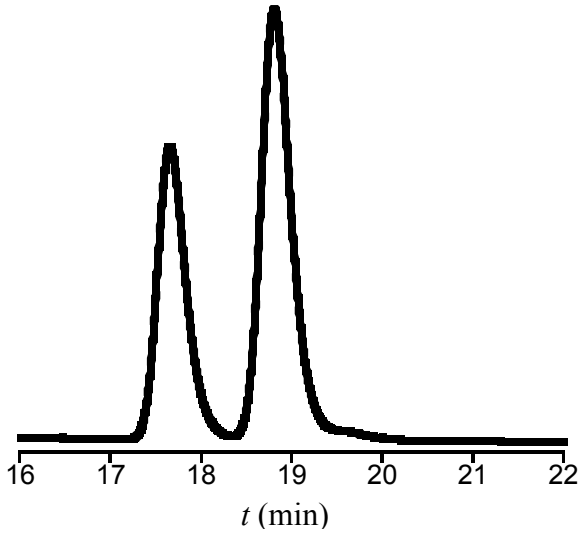


Figure S7. Chiral HPLC traces obtained using CHIRALPAK ID column (*n*-hexane/*i*-PrOH 60:40, rt, 0.5 mL/min, 254 nm) for crude mixture PPR_159 · **biot-NDI 3** (22 % *ee*).

Table S7. Results obtained for anion-π catalysis using **biot-NDI 3** · 2D_Sav or **biot-NDI 3** · MP_Sav.[a]

entry	Sav[b,c]	Yield %	12A/13D	ee (%)
1	WT Sav	60	>30	41
2	FDP_46-52	0	nd	nd
3	HP_46-52	0	nd	nd
4	PPR_159	57	>30	22
5	MP 3_*K-A*	20	>30	0
6	MP 3_*K-A*_K121F	55	>30	0
7	MP 3_K121F	51	>30	0
8	MP 4	12	>30	0
9	MP 6_*K-A_R-F*	7	nd	0
10	MP 9	23	>30	0
11	MP 9_*K-F*	45	>30	0
12	MP 11	24	>30	0
13	MP 12	traces	nd	0
14	MP 12_*R-A_K-F*	61	>30	0

[a]Reaction conditions: **biot-NDI 3** 10 mol%, Sav's FBS 20 mol %, CD_3CN/Glycine buffer pH = 3, substrate **10** 5 mM, substrate **11** 50 mM. [b] The numbering of the residues of Sav were kept as that of WT Sav, even though the insertion may be placed before the position of the mutation. [c] Italicized one letter aminoacid abbreviations designate cationic aminoacids within the inserted loops that were mutated to either A or F to probe the effect of the charge on the catalytic performance (See Table 3 for details).

S4. Annexes

Table S8: Overview of the streptavidin mutants containing $(GGX)_n$-inserts and their expression yields.

Entry	Name	Position(s)	Sequence of insert(s)	Expression	Mass of monomer[a] [Da]		Cell free extract		Purified protein	
					Calculated	Determined[b]	FBBS[f] [nmol/mg]	Yield[c] [mg/l]	FBBS/Tetramer	Yield[d] [mg/l]
1	48	G48…N49	GGSGGS	Soluble tetramer	16827.3	16828.1	2.7	549.6	nd.	13.8
2	66	T66…D67	GGSGGS	Soluble tetramer	16827.3	16827.7	2.7	557.0	nd.	16.3
3	84	R84…N85	GGSGGS	Soluble tetramer	16827.3	16828.0	2.8	529.4	nd.	13.0
4	117	A117…N118	GGSGGS	Soluble tetramer	16827.3	-	2.6	235.4	-	-
5	48+66	G48…N49 T66…D67	GGSGGS GGSGGS	Soluble tetramer	17229.7	[e]	6.2	338.0	nd.	1.5
6	48+84	G48…N49 R84…N85	GGSGGS GGSGGS	Soluble tetramer	17229.7	[e]	5.6	489.0	nd.	0.9
7	48+117	G48…N49 A117…N118	GGSGGS GGSGGS	Soluble tetramer	17229.7	-	0.2	363.0	-	-
8	66+84	T66…D67 R84…N85	GGSGGS GGSGGS	Insoluble monomer	17229.7	-	-	-	-	-
9	66+117	T66…D67 A117…N118	GGSGGS GGSGGS	Soluble tetramer	17229.7	-	0.2	383.0	-	-
10	84+117	R84…N85 A117…N118	GGSGGS GGSGGS	Insoluble monomer	17229.7	-	-	-	-	-
11	Loop 1	G48…N49	GGDGGNGGSGGLGGCGGS	Soluble tetramer	17729.2	-	19.4	678.0	-	-
12	Loop 2	G48…N49	GGNGGNGGGGGVGGS	Soluble tetramer	17466.9	17468.3	22.0	580.0	nd.	55.0
13	Loop 3	G48…N49	GGIGGSGGGGGHGGRGGGGGVGGS	Soluble tetramer	18131.6	-	12.3	619.0	-	-
14	Loop 4	G48…N49	GGNGGSGGGGGGGGSGGSGGS	Soluble tetramer	17800.2	-	12.3	596.0	-	-
15	Loop 5	G48…N49	GGRGGGGGHGGCGGVGGS	Soluble tetramer	17749.3	-	17.0	1039.0	-	-
16	Loop 7	G48…N49	GGDGGS	Soluble tetramer	16855.3	-	15.8	630.0	-	-
17	Loop 8	G48…N49	GGCGGSGGGGGGGGGGCGGS	Soluble tetramer	17775.2	-	10.8	544.0	-	-
18	Loop 9	G48…N49	GGCGGIGGS	Soluble tetramer	17070.6	-	15.4	519.0	-	-

[a]Mass of the streptavidin monomer without the N-terminal methionine. [b]Mass was only determined for purified proteins. [c]Dried cell free extract (CFE) powder per 1 L culture. [d]Isolated purified protein per 1 L culture. [e]Peak corresponding to desired product mass was not detected in the MS analysis. [f]Amount of free biotin binding sites per mass CFE.

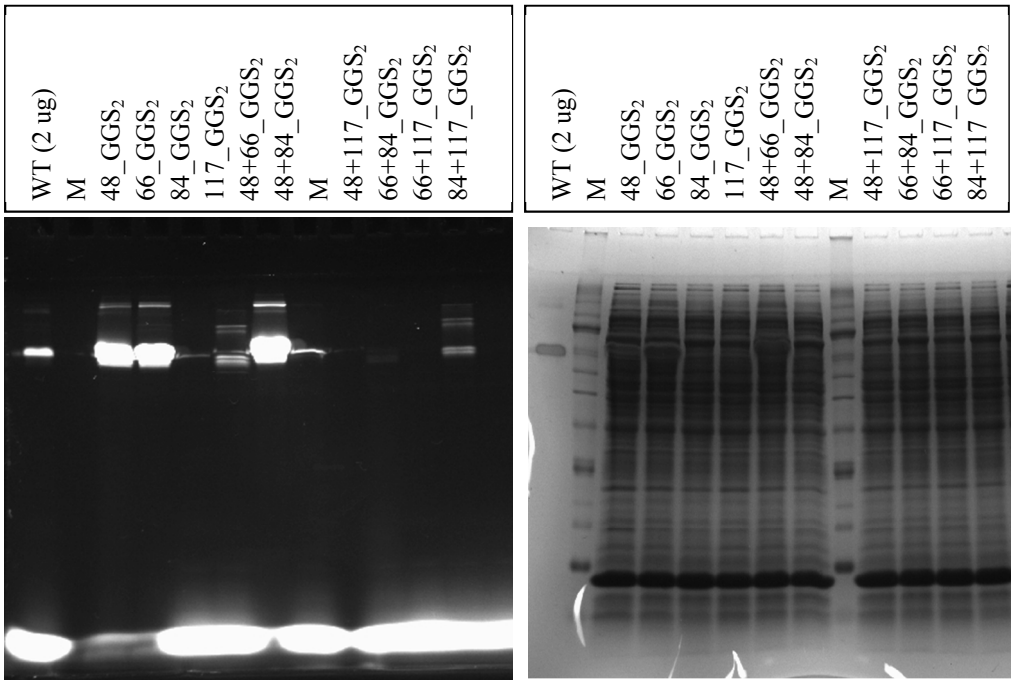
$(Gly\text{-}Gly\text{-}Ser)_2$ motif:

Example sequence of a streptavidin loop mutant with the $(GGS)_2$ motif between residues G48 and N49 (including T7-tag):

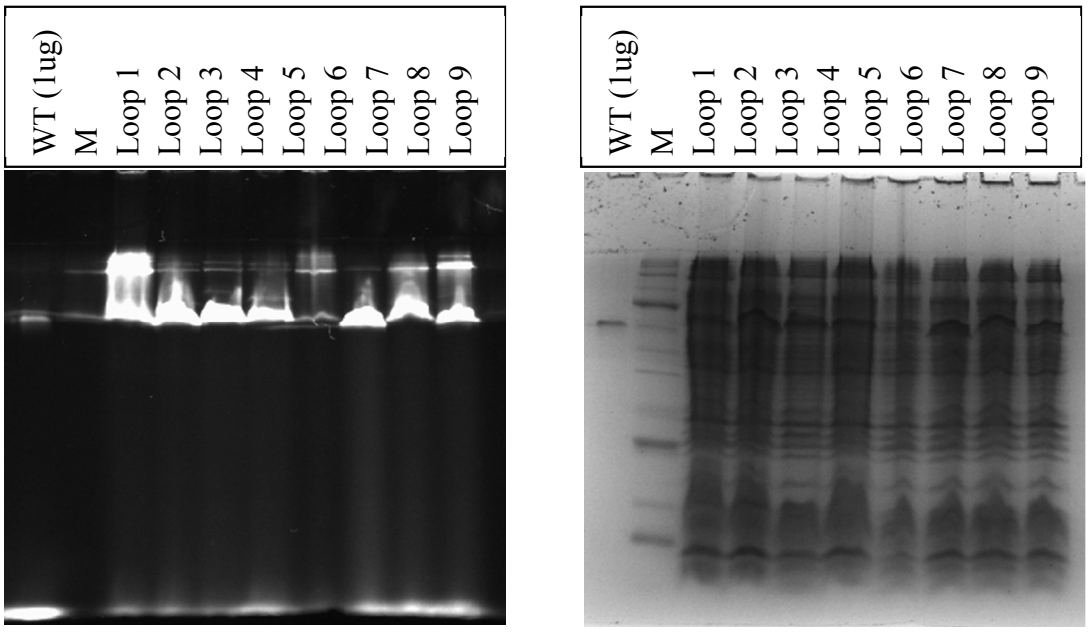
H_2N—

ASMTGGQQMGRDQAGITGTWYNQLGSTFIVTAGADGALTGTYESAVG$_{48}$GGSGGSN$_{49}$AESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHS ATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASIDAAKKAGVNNGNPLDAVQQ—**CO_2H**

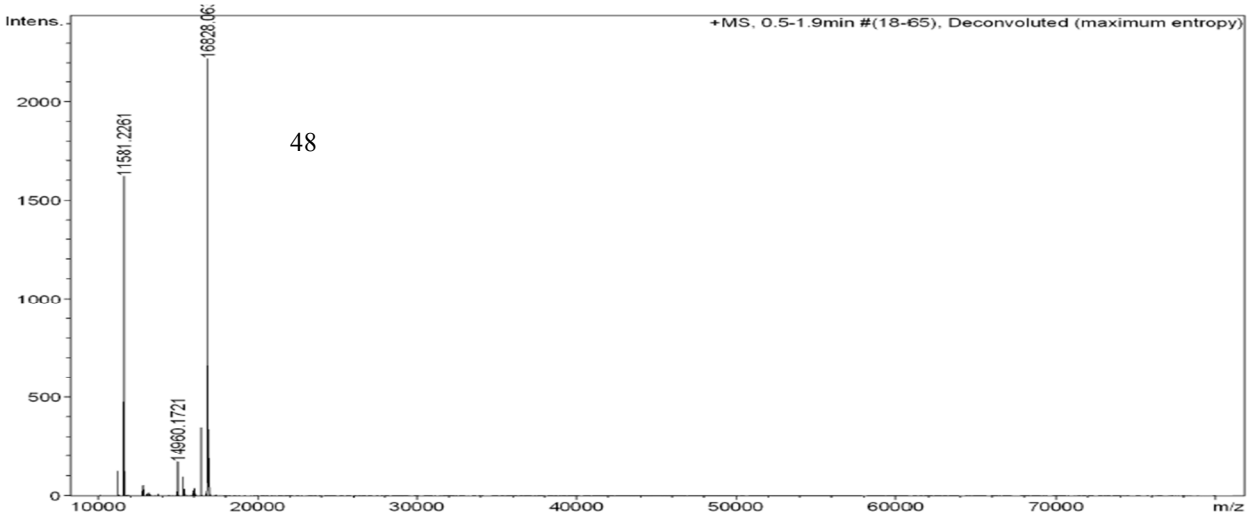
SDS gel electrophoresis of CFEs (soluble fraction) for $(GGS)_2$ mutants:

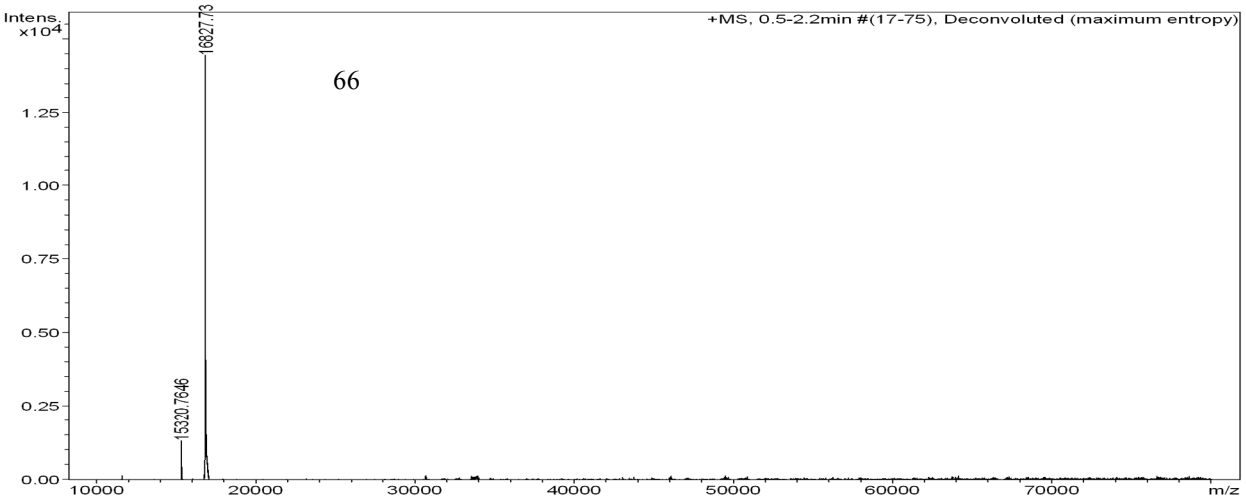


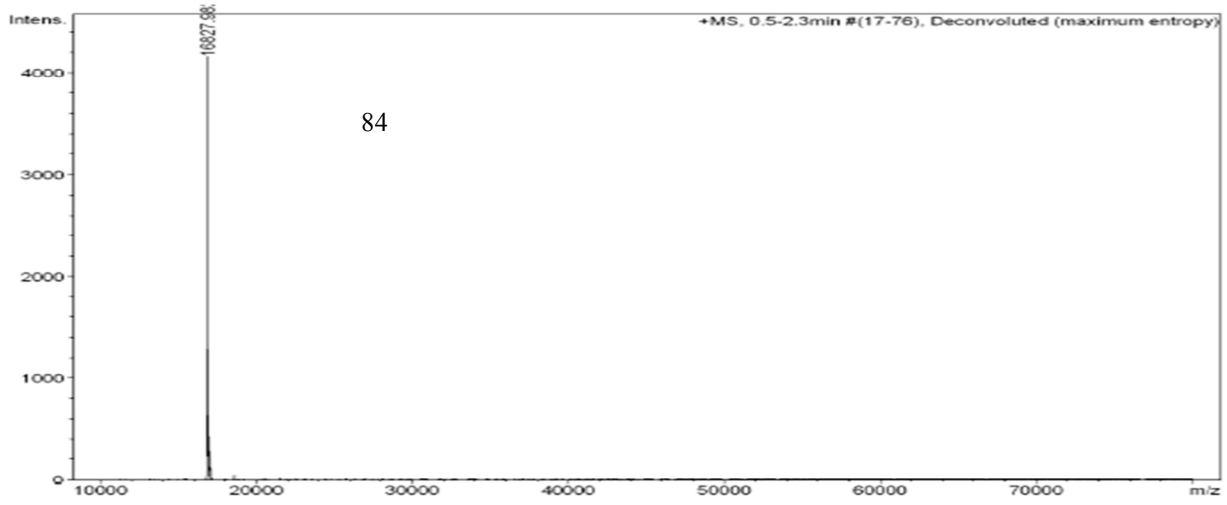
SDS gel electrophoresis of CFEs (soluble fraction) for $(GGS)_n$ mutants:

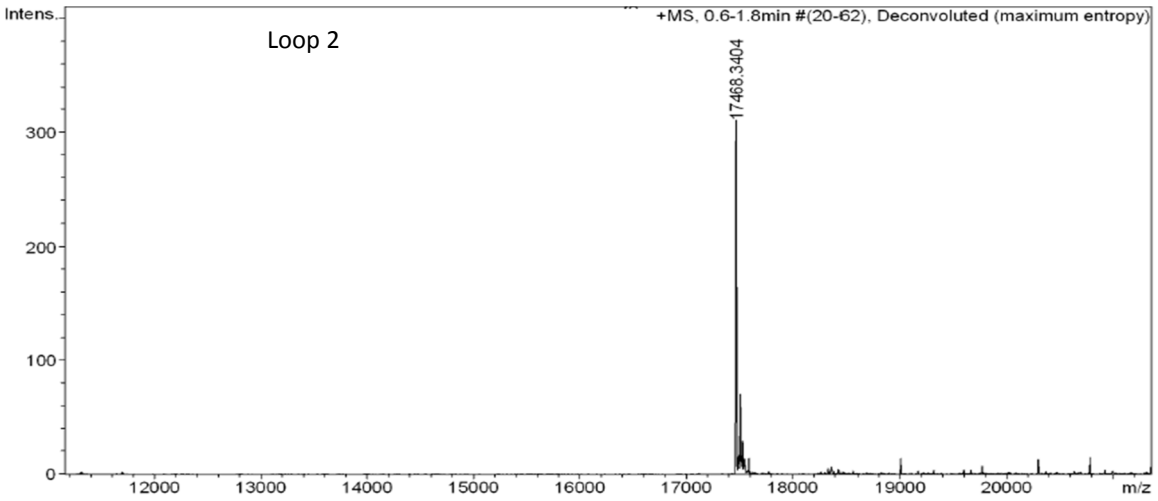


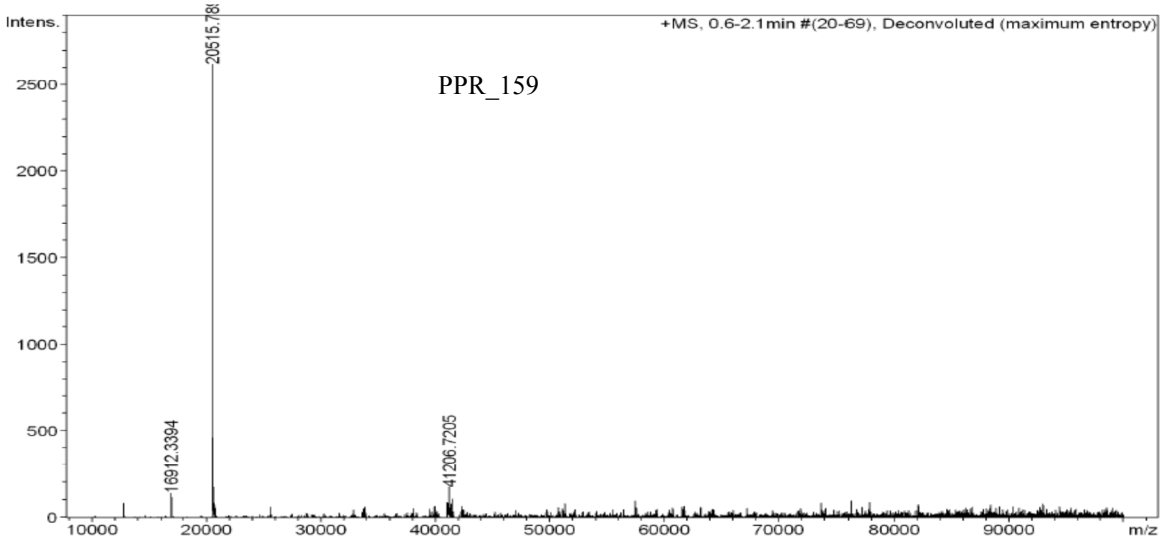
ESI-TOF mass spectrum (after deconvolution) of the purified chimeric streptavidins

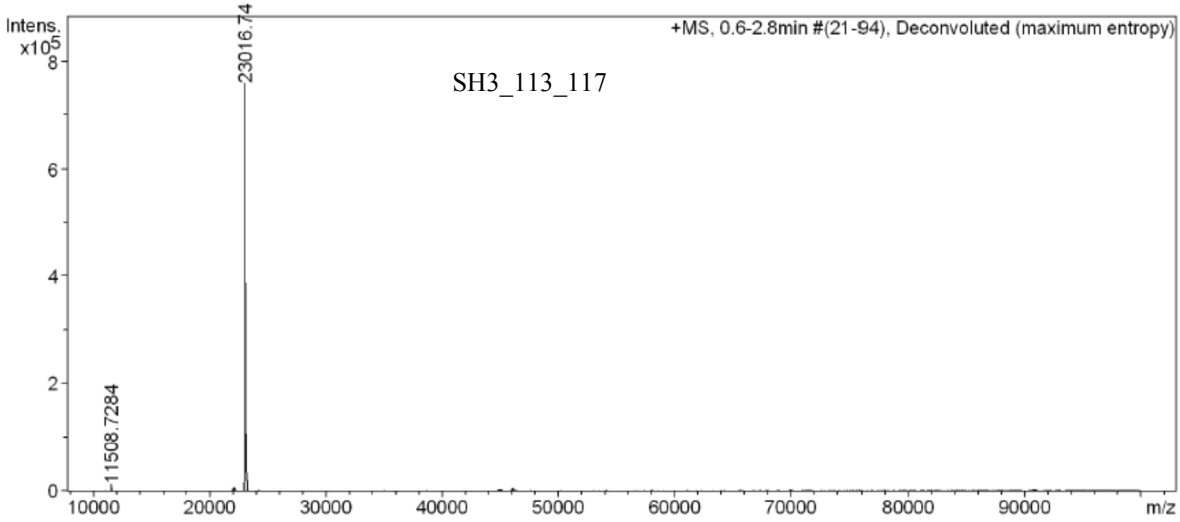


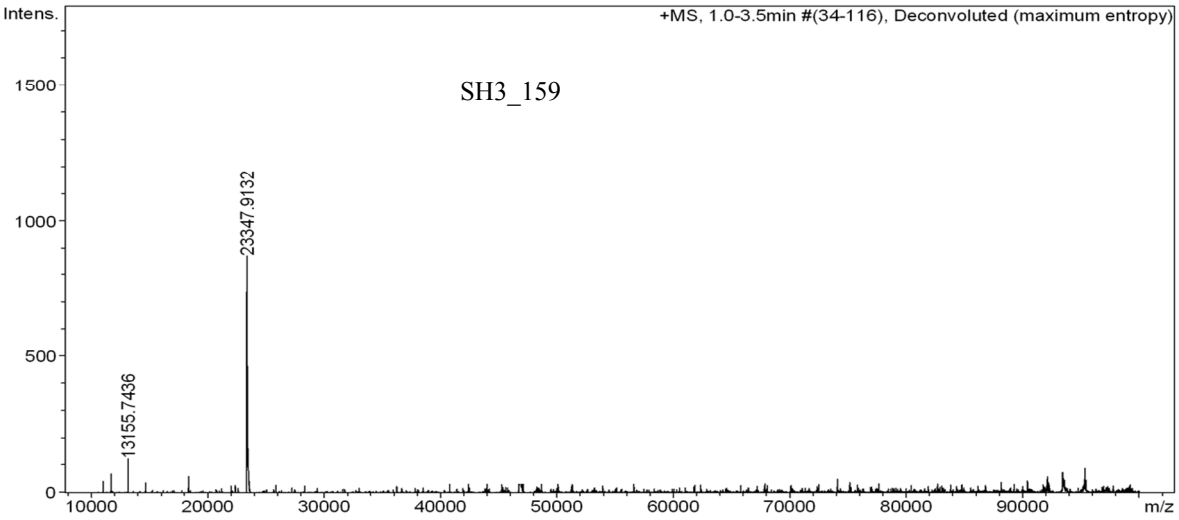


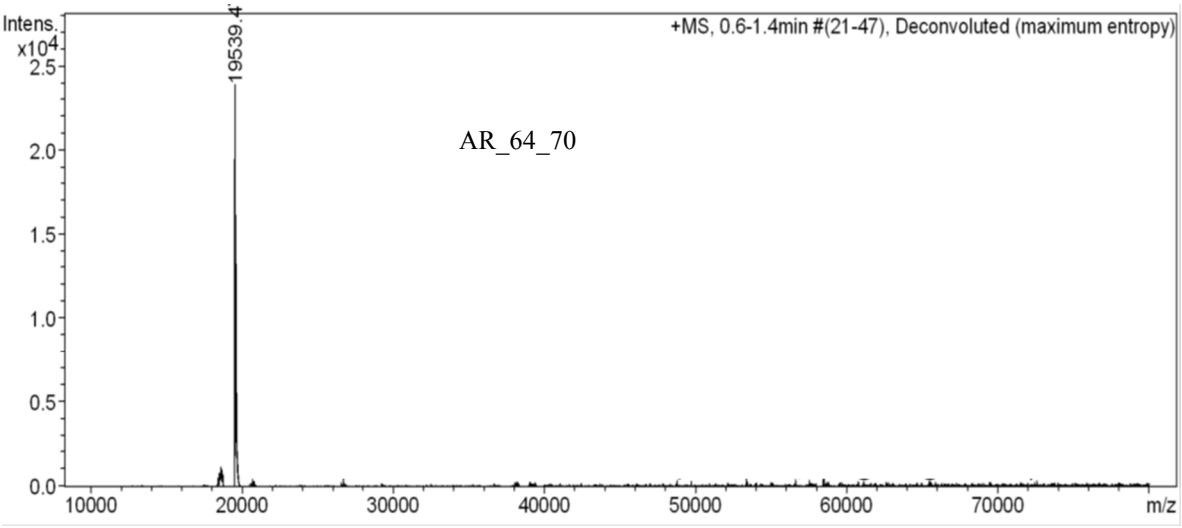


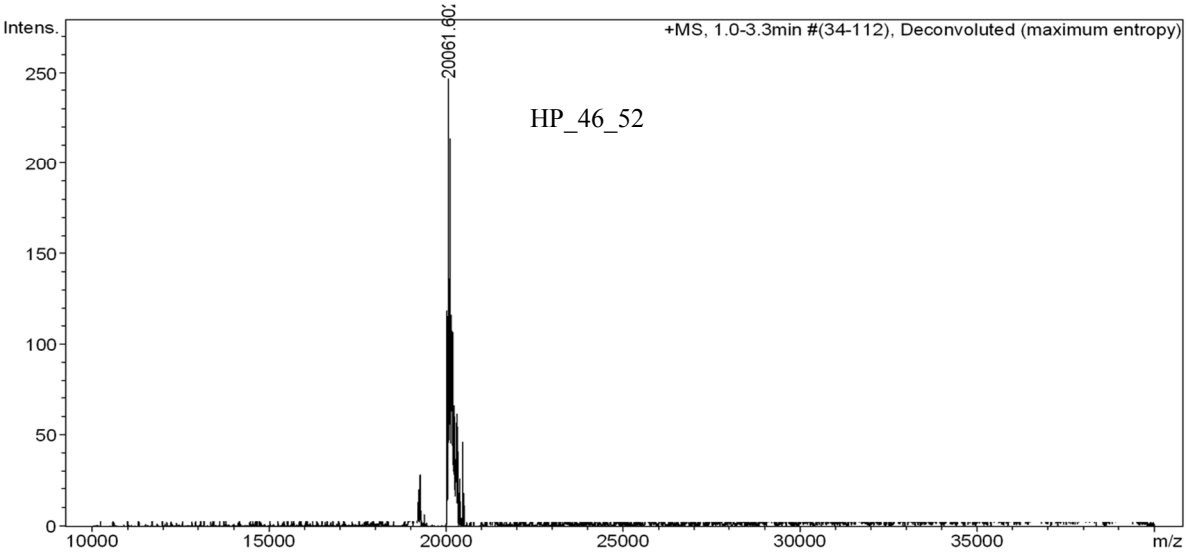

Intens.
+MS, 0.6-1.8min #(20-62), Deconvoluted (maximum entropy)
Loop 2
17468.3404
300
200
100
0
12000
13000
14000
15000
16000
17000
18000
19000
20000
m/z

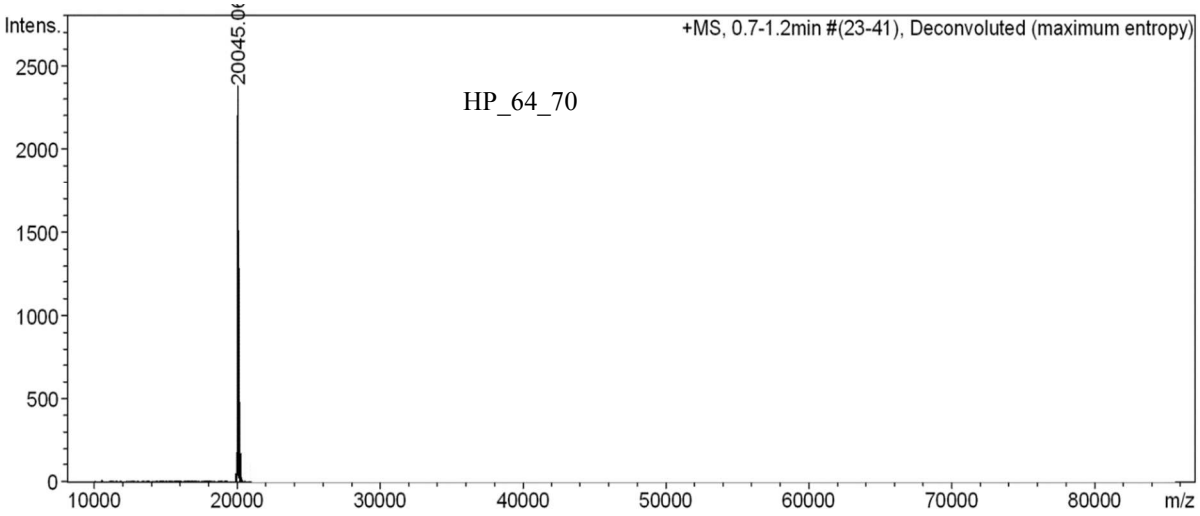

Intens.
+MS, 0.6-2.1min #(20-69), Deconvoluted (maximum entropy)
20515.78
PPR_159
2500
2000
1500
1000
500
0
16912.3394
41206.7205
10000
20000
30000
40000
50000
60000
70000
80000
90000
m/z

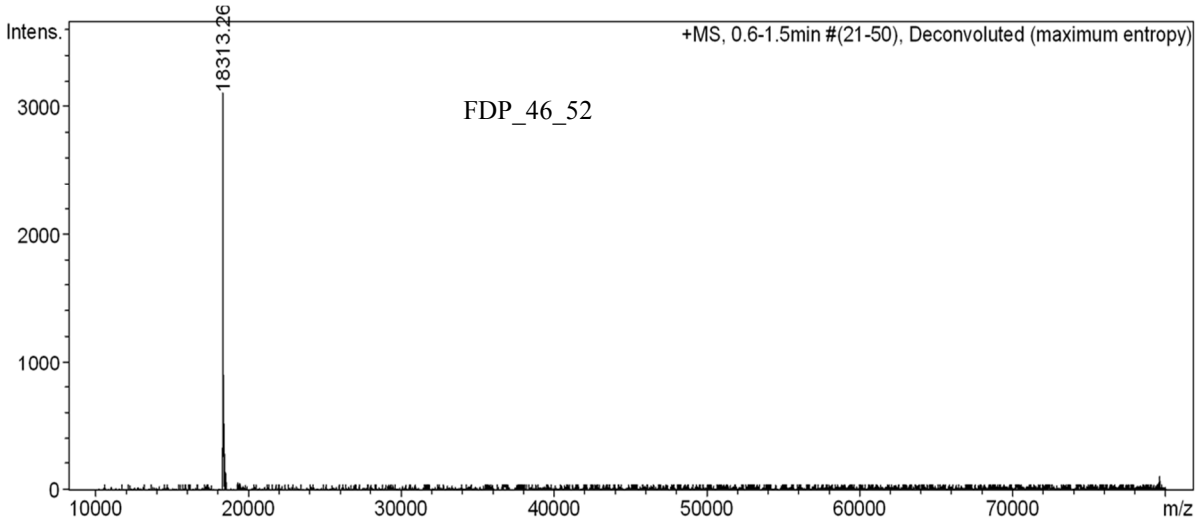

Intens.
x10^5
+MS, 0.6-2.8min #(21-94), Deconvoluted (maximum entropy)
23016.74
SH3_113_117
8
6
4
2
0
11508.7284
10000
20000
30000
40000
50000
60000
70000
80000
90000
m/z

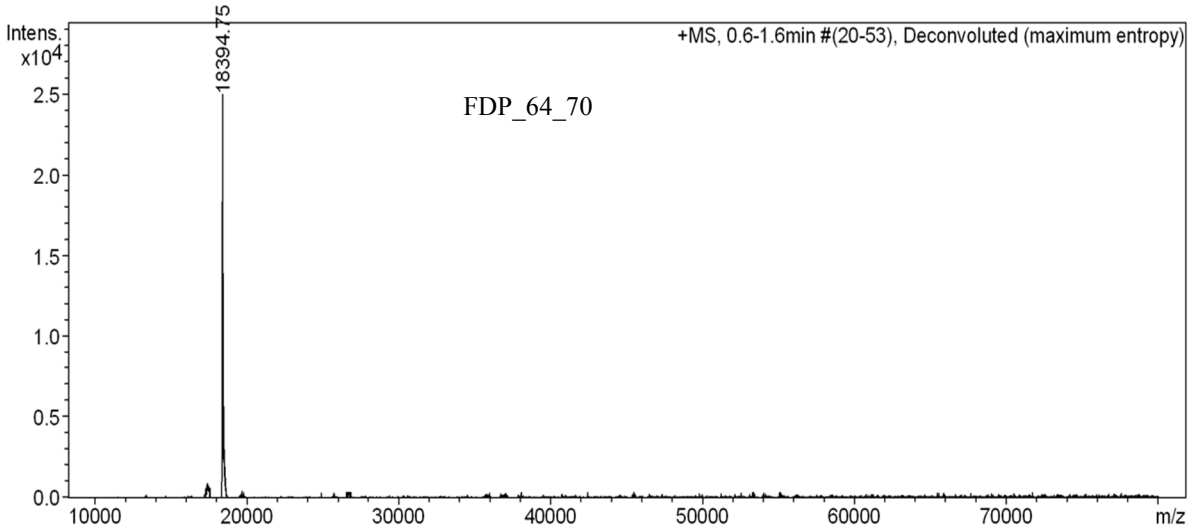

Intens.
+MS, 1.0-3.5min #(34-116), Deconvoluted (maximum entropy)
SH3_159
23347.9132
13155.7436
m/z

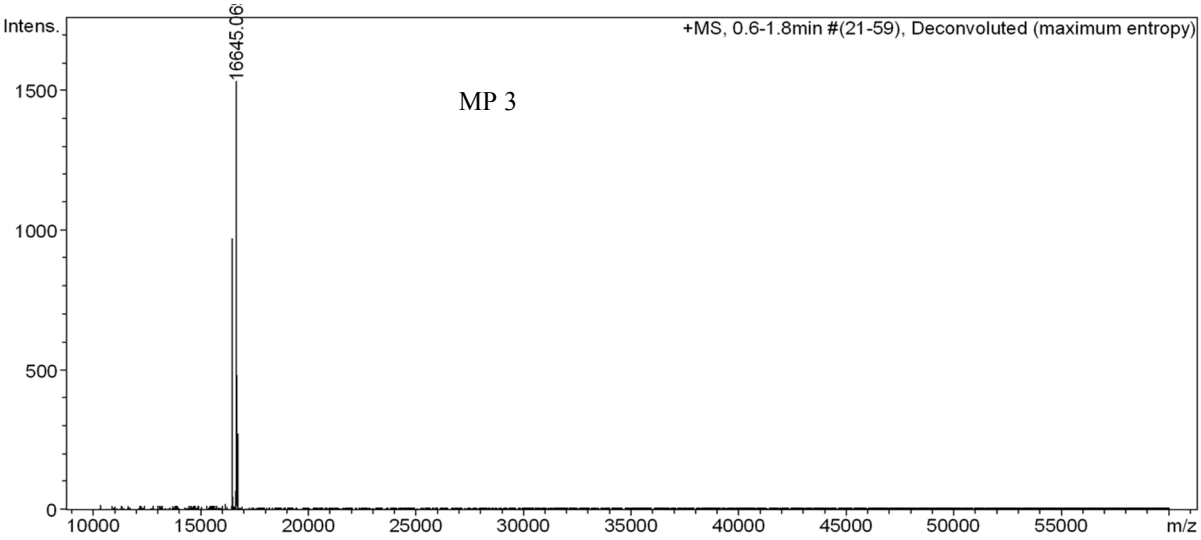

Intens.
x10^4
+MS, 0.6-1.4min #(21-47), Deconvoluted (maximum entropy)
AR_64_70
19539.4
m/z

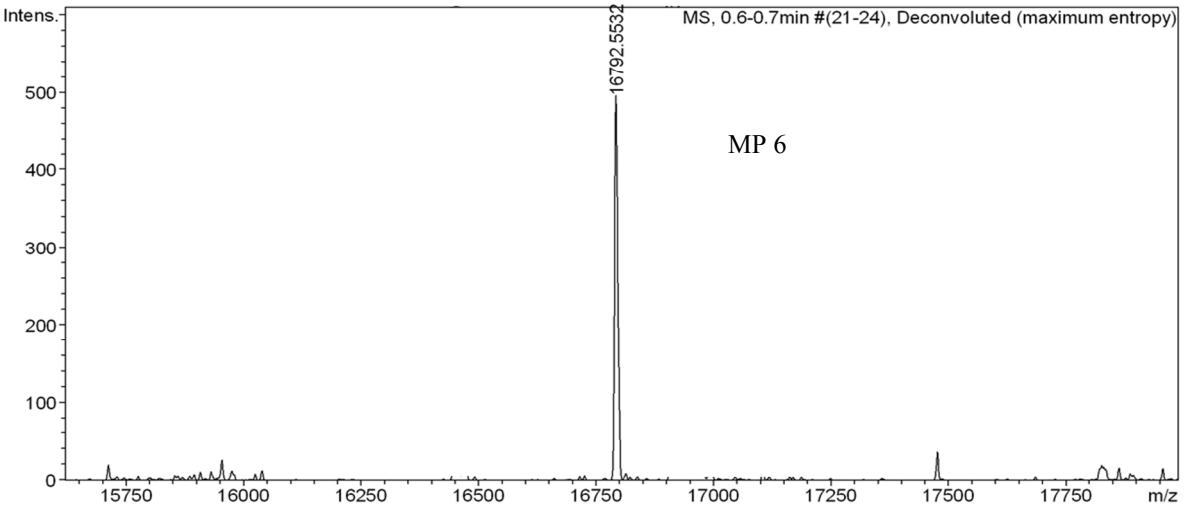

Intens.
+MS, 1.0-3.3min #(34-112), Deconvoluted (maximum entropy)
HP_46_52
20061.60
m/z

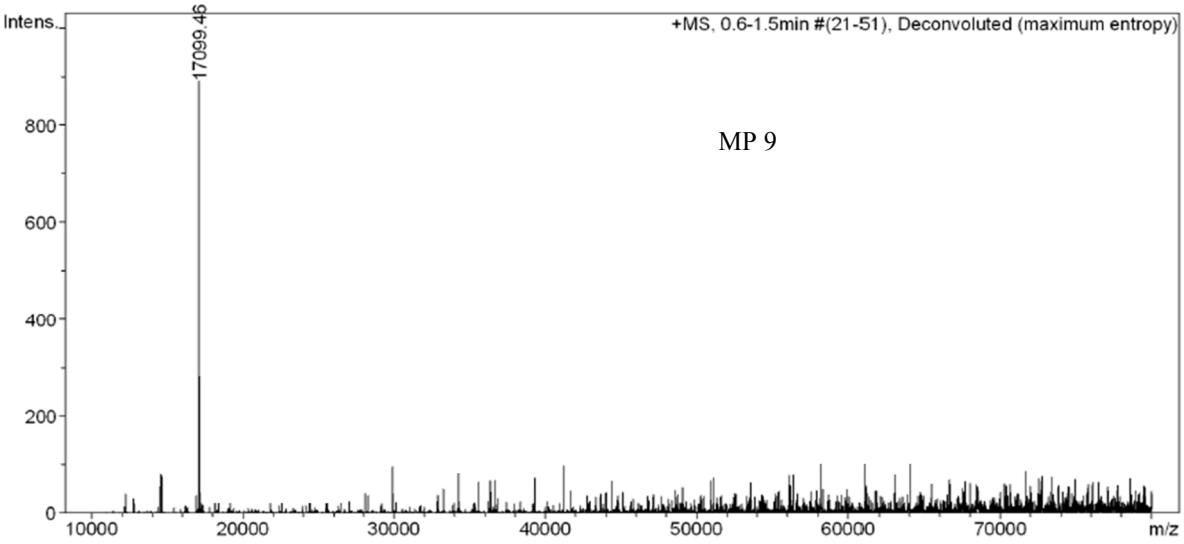

Intens.
+MS, 0.7-1.2min #(23-41), Deconvoluted (maximum entropy)
20045.0
HP_64_70
2500
2000
1500
1000
500
0
10000
20000
30000
40000
50000
60000
70000
80000
m/z

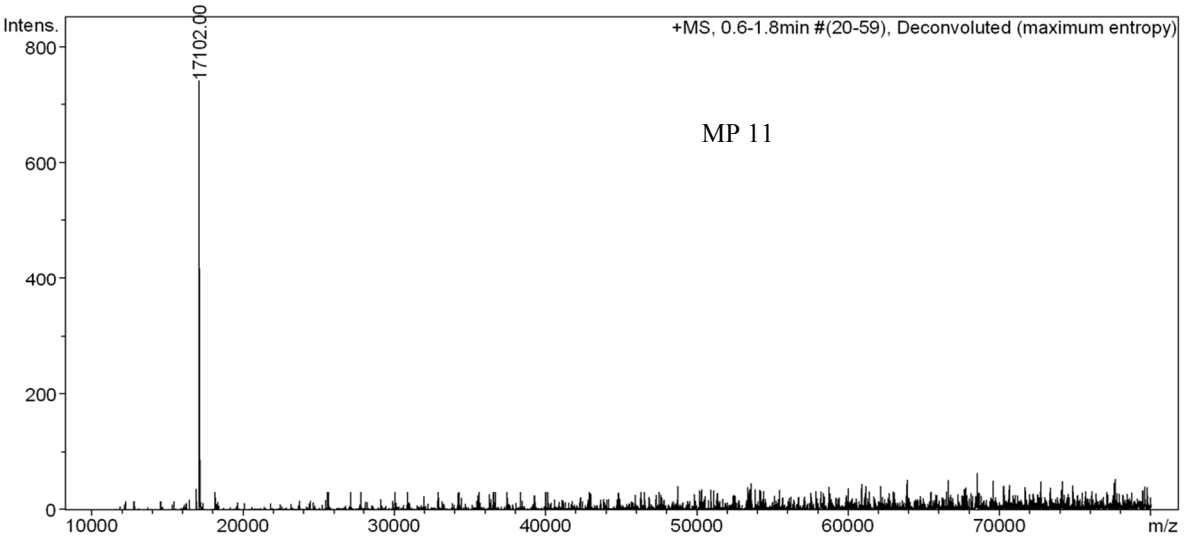

Intens.
+MS, 0.6-1.5min #(21-50), Deconvoluted (maximum entropy)
18313.26
FDP_46_52
3000
2000
1000
0
10000
20000
30000
40000
50000
60000
70000
m/z

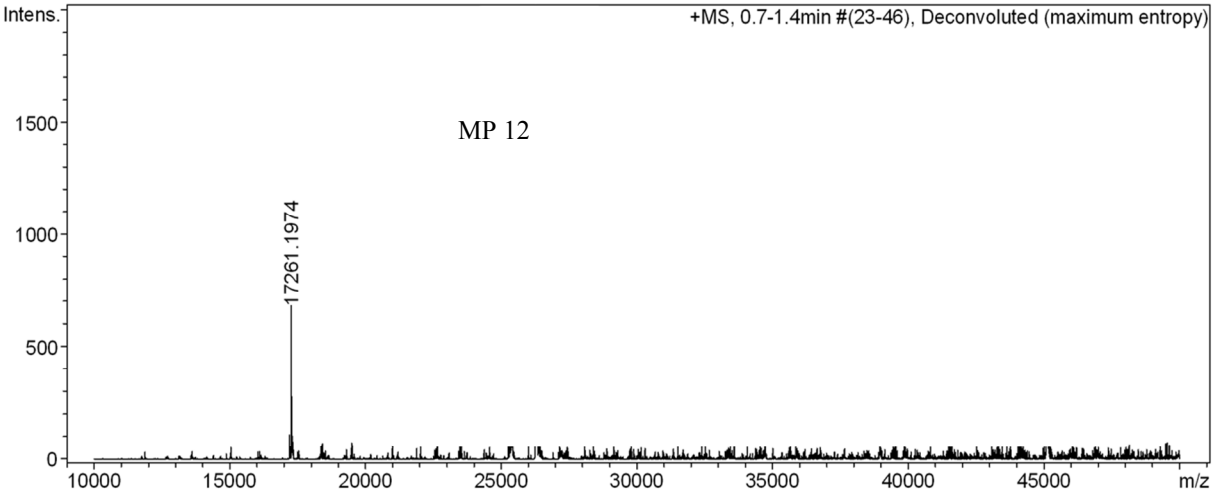

Intens.
x10^4
+MS, 0.6-1.6min #(20-53), Deconvoluted (maximum entropy)
18394.75
FDP_64_70
2.5
2.0
1.5
1.0
0.5
0.0
10000
20000
30000
40000
50000
60000
70000
m/z

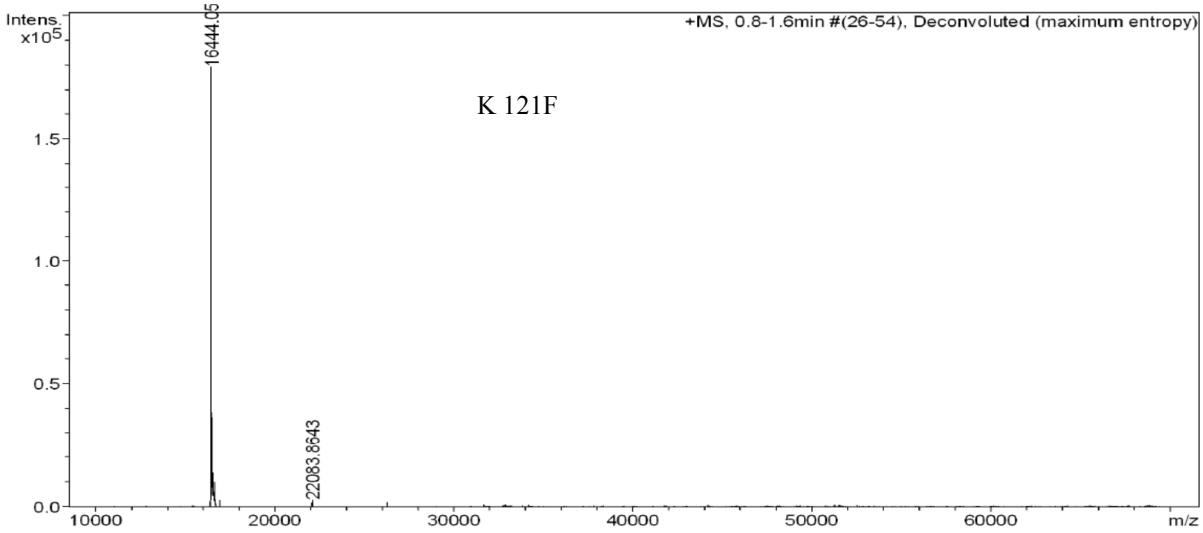

Intens.
+MS, 0.6-1.8min #(21-59), Deconvoluted (maximum entropy)
16645.06
MP 3
1500
1000
500
0
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
m/z

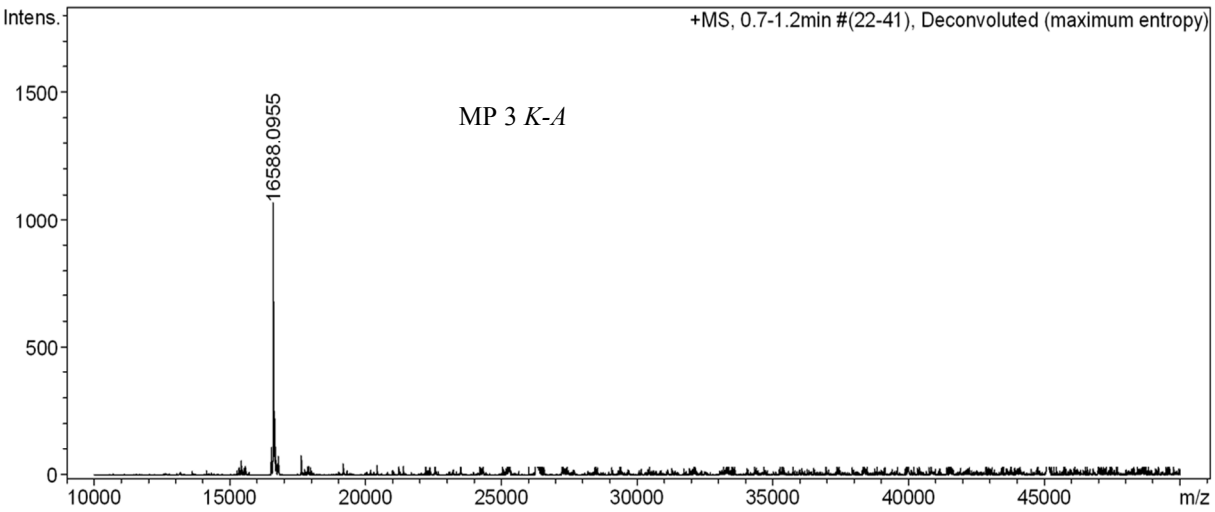

Intens.
MS, 0.6-0.7min #(21-24), Deconvoluted (maximum entropy)
16792.5532
MP 6
500
400
300
200
100
0
15750
16000
16250
16500
16750
17000
17250
17500
17750
m/z

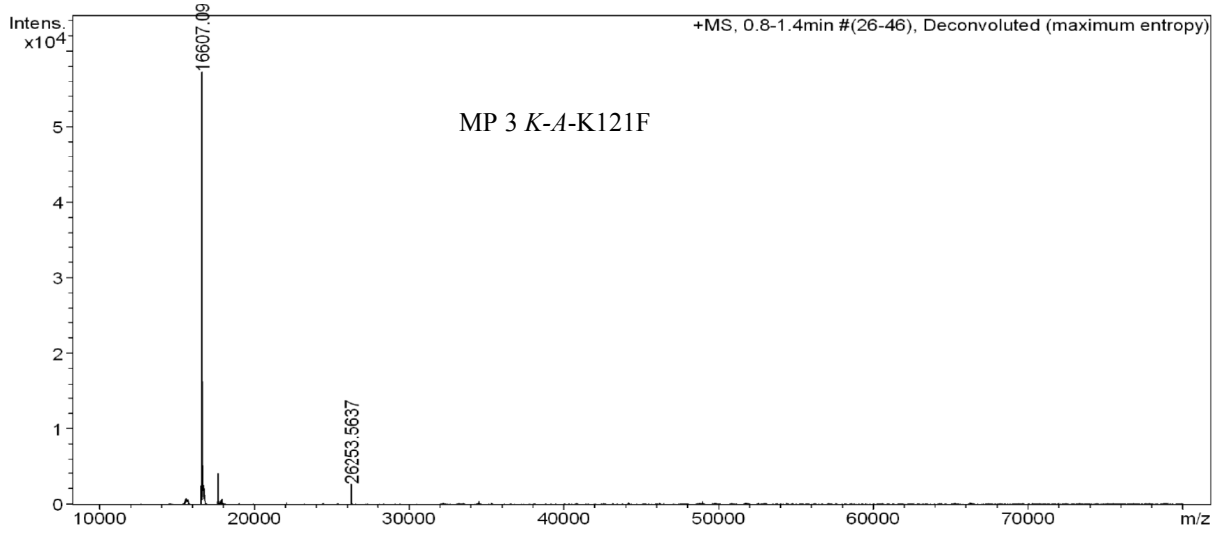

Intens.
+MS, 0.6-1.5min #(21-51), Deconvoluted (maximum entropy)
17099.46
MP 9
800
600
400
200
0
10000
20000
30000
40000
50000
60000
70000
m/z

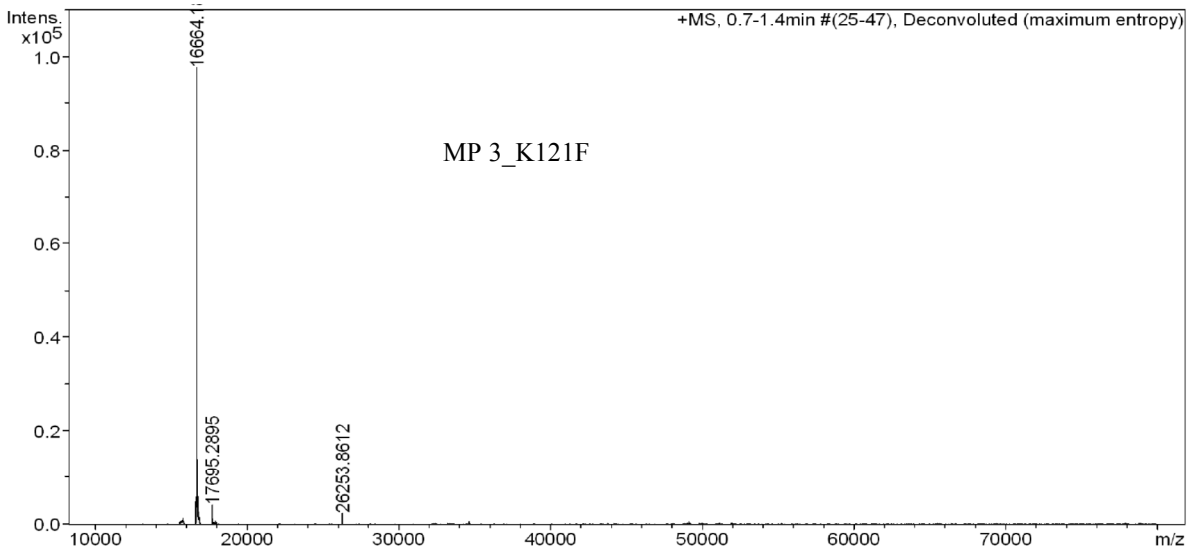

Intens.
+MS, 0.6-1.8min #(20-59), Deconvoluted (maximum entropy)
17102.00
MP 11
800
600
400
200
0
10000
20000
30000
40000
50000
60000
70000
m/z

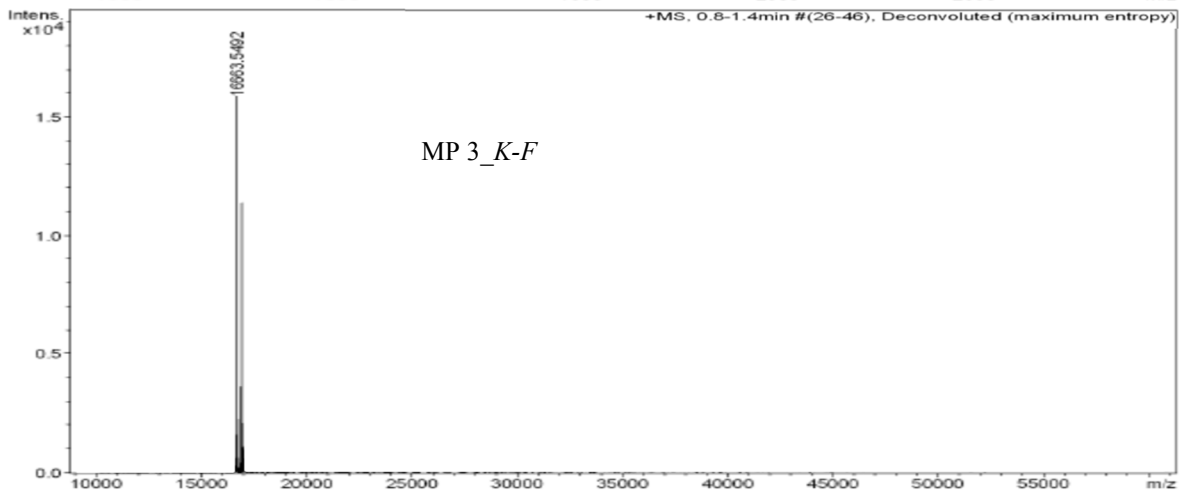

Intens.
+MS, 0.7-1.4min #(23-46), Deconvoluted (maximum entropy)
MP 12
17261.1974
1500
1000
500
0
10000
15000
20000
25000
30000
35000
40000
45000
m/z


Intens.
x10^5
+MS, 0.8-1.6min #(26-54), Deconvoluted (maximum entropy)
16444.05
K 121F
22083.8643
1.5
1.0
0.5
0.0
10000
20000
30000
40000
50000
60000
m/z

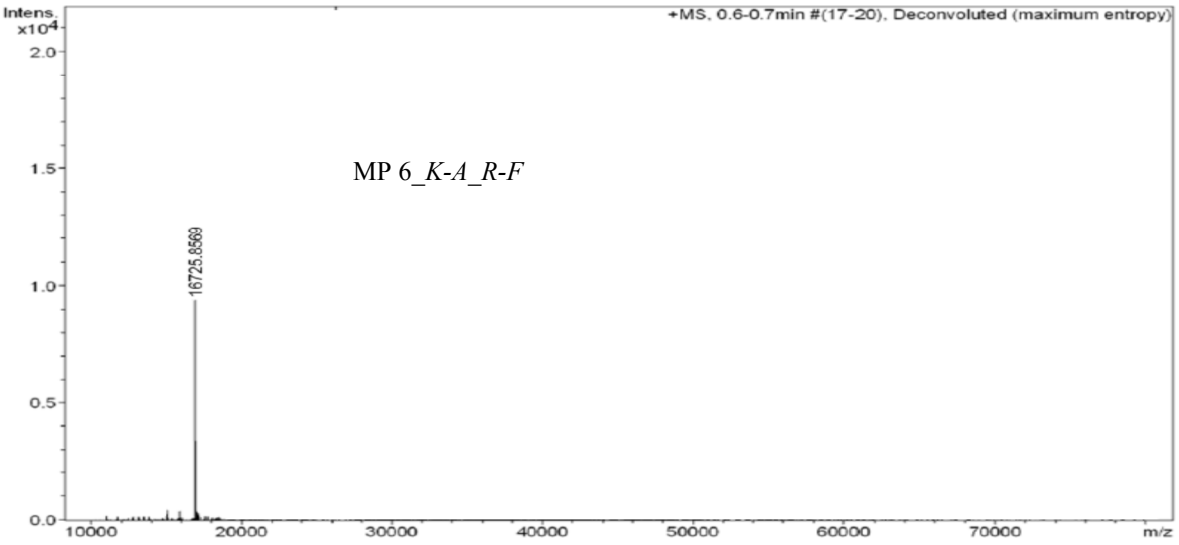

Intens.
+MS, 0.7-1.2min #(22-41), Deconvoluted (maximum entropy)
1500
1000
500
0
16588.0955
MP 3 *K-A*
10000
15000
20000
25000
30000
35000
40000
45000
m/z


Intens.
x10^4
+MS, 0.8-1.4min #(26-46), Deconvoluted (maximum entropy)
5
4
3
2
1
0
16607.09
26253.5637
MP 3 *K-A*-K121F
10000
20000
30000
40000
50000
60000
70000
m/z

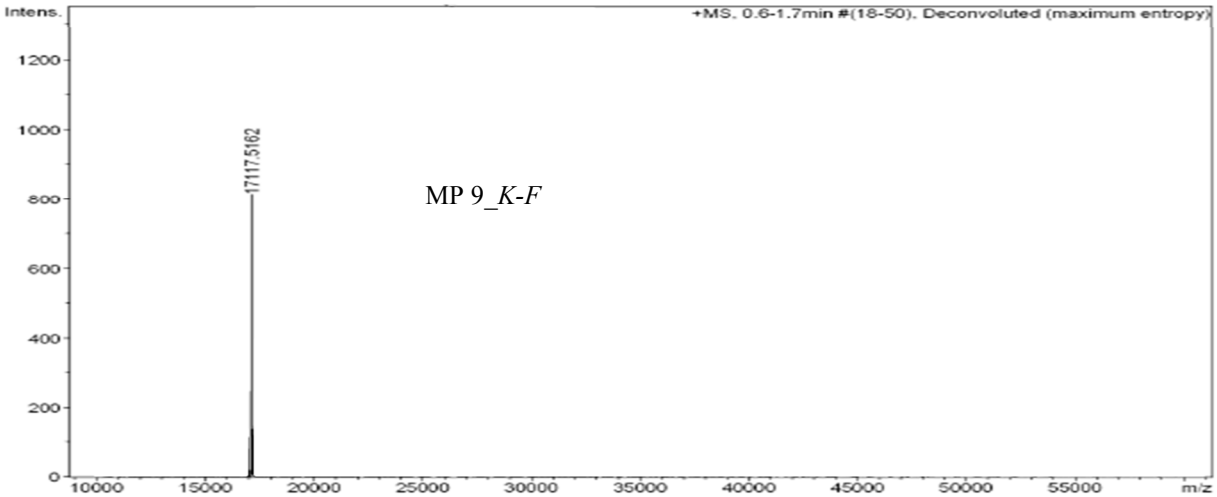

Intens.
x10^5
+MS, 0.7-1.4min #(25-47), Deconvoluted (maximum entropy)
1.0
0.8
0.6
0.4
0.2
0.0
16664.1
17695.2895
26253.8612
MP 3_K121F
10000
20000
30000
40000
50000
60000
70000
m/z



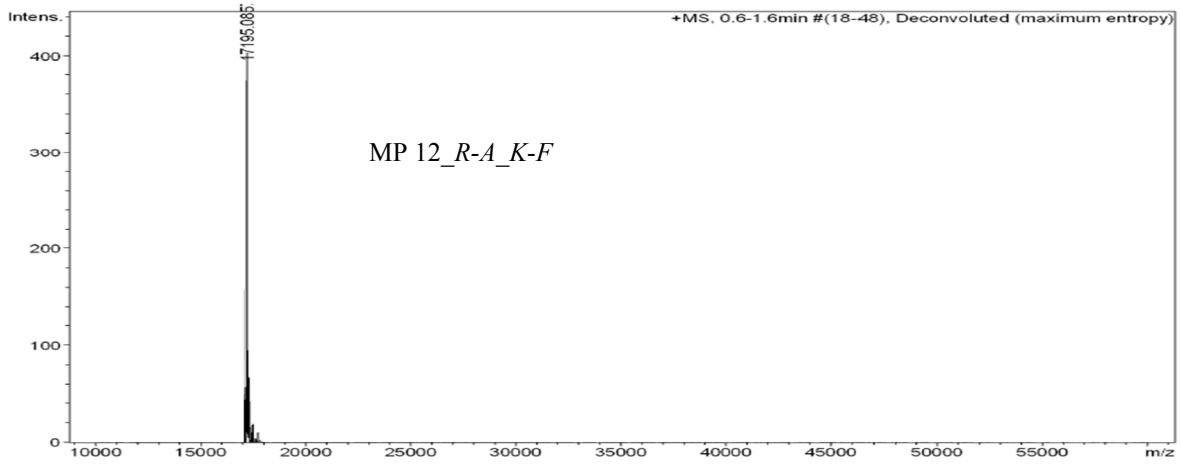
MP 3_*K-F*



MP 6_*K-A_R-F*



MP 9_*K-F*



S5. References

1. Wood, C. W.; Heal, J. W.; Thomson, A. R.; Bartlett, G. J.; Ibarra, A. A.; Brady, R. L.; Sessions, R. B.; Woolfson, D. N., *Bioinformatics* **2017,** *33*, 3043-3050.
2. Kabsch, W.; Sander, C., *Biopolymers* **1983,** *22*, 2577-2637
3. Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G., *Nucleic Acids Res.* **2015,** *43*, D364-D368.
4. Schwizer, F.; Kohler, V.; Durrenberger, M.; Knorr, L.; Ward, T. R., *ACS Catal.* **2013,** *3*, 1752-1755.
5. Koehler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knoerr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; Ward, T. R., *Angew. Chem., Int. Ed.* **2011,** *50*, 10863-10866.
6. Mallin, H.; Hestericova, M.; Reuter, R.; Ward, T. R., *Nat. Protoc.* **2016,** *11*, 835-852.
7. Studier, F. W., *Protein Expression Purif.* **2005,** *41*, 207-234.
8. Monnard, F. W.; Nogueira, E. S.; Heinisch, T.; Schirmer, T.; Ward, T. R., *Chem. Sci.* **2013,** *4*, 3269-3274.
9. Kohler, V.; Wilson, Y. M.; Durrenberger, M.; Ghislieri, D.; Churakova, E.; Quinto, T.; Knorr, L.; Haussinger, D.; Hollmann, F.; Turner, N. J., *Nat. Chem.* **2013,** *5*, 93-99.
10. Cotelle, Y.; Lebrun, V.; Sakai, N.; Ward, T. R.; Matile, S., *ACS Cent. Sci.* **2016,** *2*, 388-393.