Supporting Information for:

Functional Elucidation of TfuA in Peptide Backbone Thioamidation

Andi Liu[1,2], Yuanyuan Si[2,3], Shi-Hui Dong[4,5], Nilkamal Mahanta[2,3,6], Haley N. Penkala[1,2], Satish K. Nair[2,3,4], and Douglas A. Mitchell[1,2,3,*]

[1]Department of Microbiology, [2]Carl R. Woese Institute for Genomic Biology, [3]Department of Chemistry, [4]Department of Biochemistry, University of Illinois, Urbana, IL, 61801, USA; [5]current address: State Key Laboratory of Applied Organic Chemistry, College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou, 730000, P.R. China; [6]current address: Department of Chemistry, Indian Institute of Technology Dharwad, Karnataka, 580011, India

*Author to whom correspondence should be addressed: douglasm@illinois.edu

Supplementary Table 1. Oligonucleotide primers used in this study. Nucleotide sequences are given in the 5' to 3' direction. f, forward primer; r, reverse primer; capital letters, mutagenized codon; capitalized and italicized, restriction enzyme recognition site. Mt, *Methanothermobacter* sp. CaT2; Azo, *Azospirillum* sp. B510.

Primer name	Oligonucleotide sequence
*Mt*ThiS_BamHI	ata*GGATCC*atggtgatagggatgaagtttac
*Mt*ThiS_HindIII	aaa*AAGC*ttaaccgccgtagataacc
His-*Mt*TfuA_SalI_f	aaa*GTCGAC*atgcatggtaagaaaataatcatc
His-*Mt*TfuA_NotI_r	ata*GCGGCCGC*ctattctgtatatgcaagtctc
*Mt*TfuA_HindIII_f	aa*AAGCTT*aatgcatggtaagaaaataatcatcttcacg
*Mt*TfuA_NotI_r	aaa*GCGGCCGC*ctattctgtatatgcaagtctctttatatgacgg
*Mt*YcaO_HindIII_f	aa*AAGCTT*aatgttccgggatattcctgtcag
*Mt*YcaO_NotI_r	aaa*GCGGCCGC*tcagatggaggacctgatgcg
*Mt*ThiF_HindIII_f	aa*AAGCTT*aatgcctgagagatatgagggtatg
*Mt*ThiF_NotI_r	aaa*GCGGCCGC*tcaggatagttcaacaattttaaatggttcgc
*Azo*TfuA_BamHI_f	aaa*GGATCC*aatgaagatctgtgttttccttggtcc
*Azo*TfuA_NotI_r	aaa*GCGGCCGC*tcacgcctgcatctcctcttc
*Azo*YcaO_NdeI_f	tta*CATATG*atgcgagcttcgacaaatc
*Azo*YcaO_KpnI_r	att*GGTACC*tcacgccacgcctgcc
*Mt*ThiS_E47Q/E51D_f	aaaagaacggcCAGatagtcataGATgaagaggagatcttcgatggcg
*Mt*ThiS_E47Q/E51D_r	atctcctcttcATCtatgactatCTGgccgttctttttcacgacaaccg
*Mt*TfuA_R85A_f	tgggggccctgGCAgcctcagagctcagtgaccttgg
*Mt*TfuA_R85A_r	agctctgaggcTGCcagggcccccatgctggcccc
*Mt*TfuA_D51A_f	taggcataataGCCggtgttttccaccagagccctgctg
*Mt*TfuA_D51A_r	tggaaaacaccGGCtattatgcctattatgtcagggttctccttc
*Mt*TfuA_D112A_f	aaatcgaatcgGCTgatgacgttgcagttgcctttaaccc
*Mt*TfuA_D112A_r	gcaacgtcatcAGCcgattcgatttcaccctcaaggtatgatc
*Mt*TfuA_K198M_f	gcagggaccttATGagggaggacgcactggaggtcatc
*Mt*TfuA_K198M_r	gcgtcctccctCATaaggtccctgccctctgattcaagg
*Mt*TfuA_D201N_f	ttaaaagggagAACgcactggaggtcatccgtcatataaag
*Mt*TfuA_D201N_r	acctccagtgcGTTctccctttaaggtccctgccctc
*Mt*TfuA_Y104A_f	tcttcagatcaGCActtgagggtgaaatcgaatcggatgatg
*Mt*TfuA_Y104A_r	tcaccctcaagTGCtgatctgaagatgcggccaacacc
*Mt*TfuA_S80A_f	ttggaggggccGCCatgggggccctgagggcctcagagc
*Mt*TfuA_S80A_r	agggccccccatGGCggcccctccaaccaccttcacacccctc
*Mt*TfuA_S80C_f	ttggaggggccTGCatgggggccctgagggcctcagag
*Mt*TfuA_S80C_r	agggccccccatGCAggcccctccaaccaccttcacacccctc
*Mt*TfuA_S80T_f	ttggaggggccACCatgggggccctgagggcctcagag
*Mt*TfuA_S80T_r	agggccccccatGGTggcccctccaaccaccttcacacccctc
*Mt*TfuA_E64A_f	ggtcacagggcgatcattgatgccatcagg

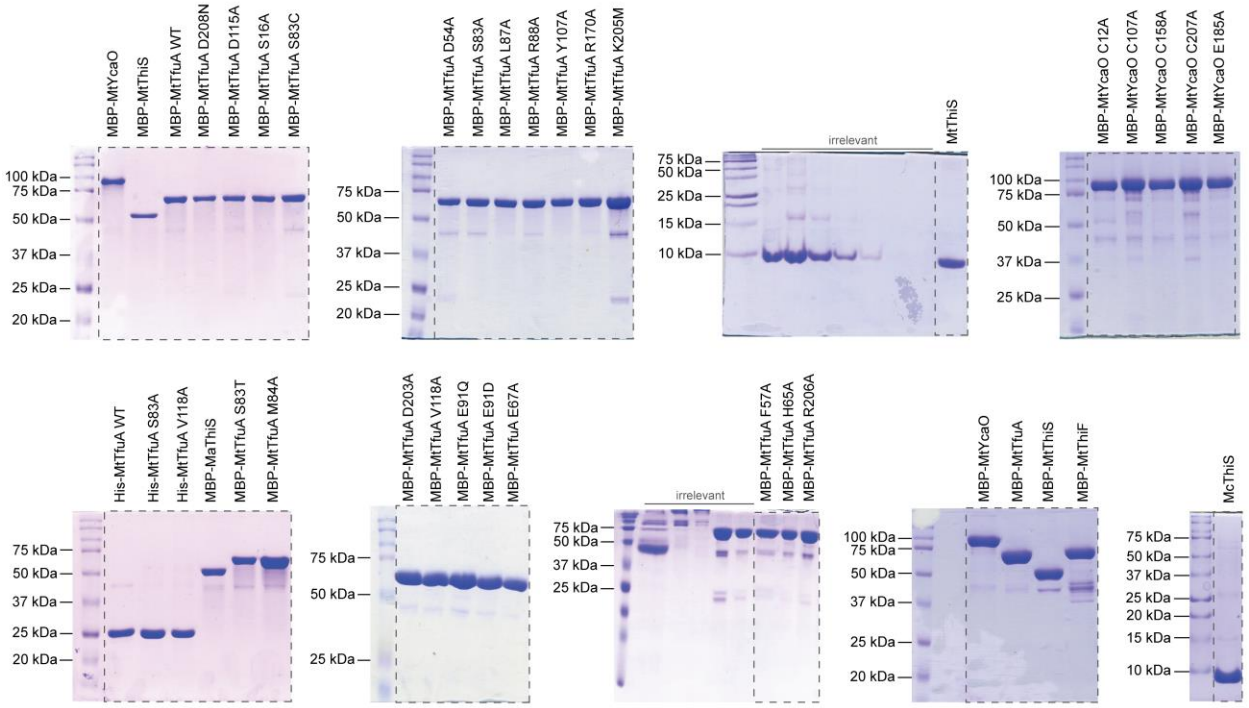
*Mt*TfuA_E64A_r	atcaatgatcgccctgtgaccaacagcagg
*Mt*TfuA_M81A_f	ggggccagcgcgggggccctgagggcc
*Mt*TfuA_M81A_r	cagggccccgcgctggcccctccaaccac
*Mt*TfuA_L84A_f	atgggggccgcgagggcctcagagctcag
*Mt*TfuA_L84A_r	tgaggccctcgcggcccccatgctggc
*Mt*TfuA_V115A_f	gatgatgacgcggcagttgccctttaaccctg
*Mt*TfuA_V115A_r	ggcaactgccgcgtcatcatccgattcgatttc
*Mt*TfuA_R167A_f	taccccctcgcgaattacaggaggatcctc
*Mt*TfuA_R167A_r	cctgtaattcgcgagggggtagaagaggttc
*Mt*TfuA_F54A_f	gacggtgttGCGcaccagagccctgctgttggtcac
*Mt*TfuA_F54A_r	ggctctggtgCGCaacaccgtctattatgcctattatg
*Mt*TfuA_D201A_f	cttaaaagggagGCGgcactggaggtcatccgtcatataaag
*Mt*TfuA_D201A_r	cctccagtgcCGCctccctttttaaggtccctgccctctg
*Mt*TfuA_E88Q_f	agggcctcaCAGctcagtgaccttggaatg
*Mt*TfuA_E88Q_r	gtcactgagCTGtgaggccctcagggc
*Mt*TfuA_E88D_f	agggcctcaGATctcagtgaccttggaatg
*Mt*TfuA_E88D_r	gtcactgagATCtgaggccctcagggc
*Mt*TfuA_H62A_f	gctgttggtGCAagggagatcattgatgcc
*Mt*TfuA_H62A_r	gatctccctTGCaccaacagcagggctc
*Mt*TfuA_D198A_f	gagggcaggGCActtaaaagggaggac
*Mt*TfuA_D198A_r	ccttttaagTGCcctgccctctgattc
*Mt*TfuA_R201A_f	gaccttaaaGCAgaggacgcactggag
*Mt*TfuA_R201A_r	tgcgtcctcTGCtttaaggtccctgcc
*Mt*YcaO_C12A_f	ggtacattggaGCCacccacagggccgtgaggcc
*Mt*YcaO_C12A_r	gccctgtgggtGGCtccaatgtacctgacaggaatatcccg
*Mt*YcaO_C107A_f	agccagaggacGCTgacggccttgaccctgagtcactg
*Mt*YcaO_C107A_r	tcaaggccgtcAGCgtcctctggctgtgctgtgaatgtttc
*Mt*YcaO_C158A_f	ctccggaaggcGCCatgagtctgttccgatcaaacacc
*Mt*YcaO_C158A_r	aacagactcatGGCgccttccggagggttgtatgg
*Mt*YcaO_C207A_f	tggaggttgacGCCtcagggacggataatgatataatatc
*Mt*YcaO_C207A_r	tccgtccctgaGGCgtcaacctccactttgggac
*Mt*YcaO_E185A_f	acgggctgatgGCGgtgattgaaagggatgcctggagc
*Mt*YcaO_E185A_r	ctttcaatcacCGCcatcagcccgtggaatatggcc

Supplementary Table 2. Data collection, phasing and refinement statistics

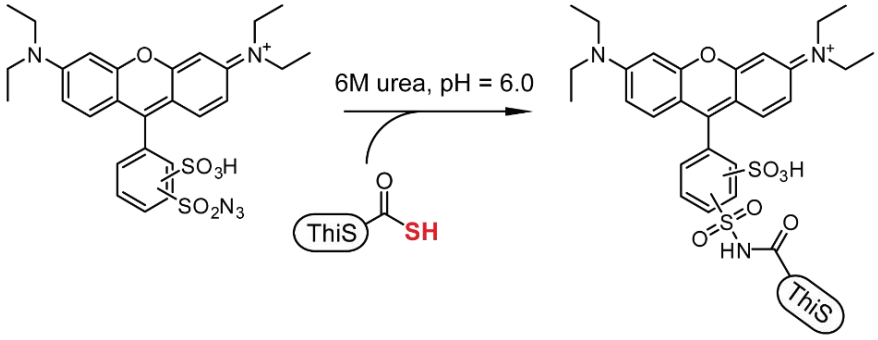
	SeMet	Native
Data collection		
Space group	$P2_12_12_1$	C2
Unit cell dimensions (Å/º)	46.3, 77.8, 129.9	103.5, 36.8, 59.6 / 121.7
Resolution (Å)	50.0 - 1.70 (1.73 - 1.70)	50.9 - 1.65 (1.66 - 1.65)
Total reflections	433,274	172,632
Unique reflections	48,427	23,292
R_{sym} (%) [1]	0.050 (0.740)	0.055 (0.868)
$I/\sigma I$ [1]	28.2 (2.0)	21.4 (2.2)
Completeness (%) [1]	98.7 (88.5)	99.8 (95.9)
Redundancy	9.0 (7.1)	7.4 (7.4)
CC (1/2)	0.997 (0.839)	0.999 (0.787)
Figure of merit	0.322	
Refinement		
Resolution (Å)		25.0 - 1.65
No. reflections		22,284
R_{work} / R_{free} [2]		16.8/19.5
Number of atoms		
Protein		1,681
Solvent		254
B-factors		
Protein		23.3
Solvent		34.5
Ramachandran statistics		
Favored/Allowed		97.67/2.33
Outliers		0
R.m.s deviations		
Bond lengths (Å)		0.746
Bond angles (°)		0.006

1. Highest resolution shell is shown in parenthesis.

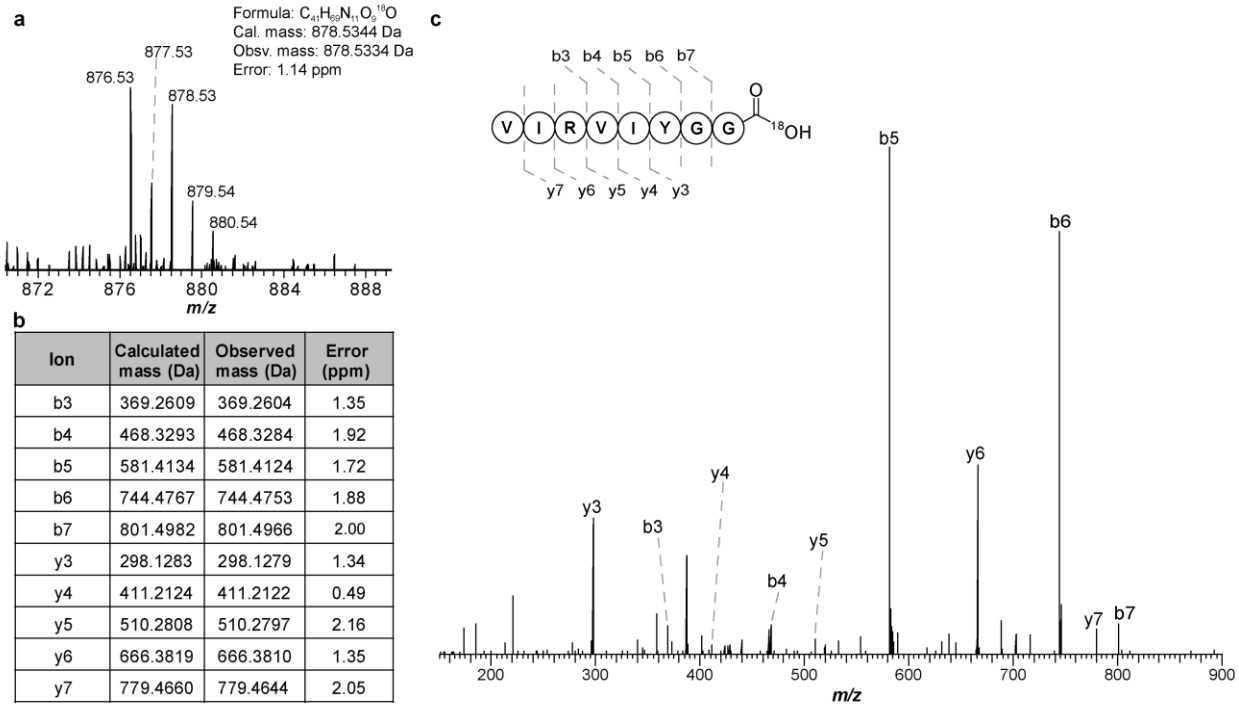
2. R-factor = $\Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma |F_{obs}|$and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.



Supplementary Figure 1. SDS-PAGE analysis of the proteins used in the study. Protein purity was assessed visually using Coomassie staining. Abbreviations: Mt, *Methanothermobacter* sp. CaT2; Mc, *Methanothermococcus thermolithotrophicus*; Ma, *Methanosarcina acetivorans*; MBP, maltose-binding protein. Each protein was purified in at least two independent experiments and similar purity was observed. Boxed area denotes proteins used in this study.



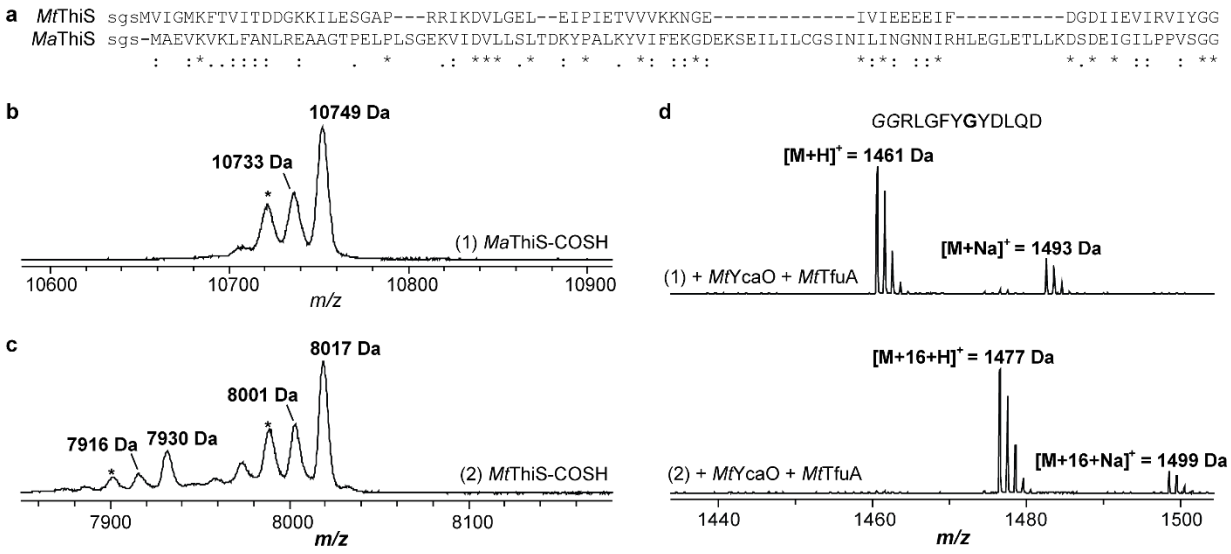
Supplementary Figure 2. Reaction scheme of ThiS-COSH reacting with lissamine rhodamine sulfonyl azide (LRSA)[1]. This reaction was used to quantify ThiS-COSH consumption during thioamidation reactions.



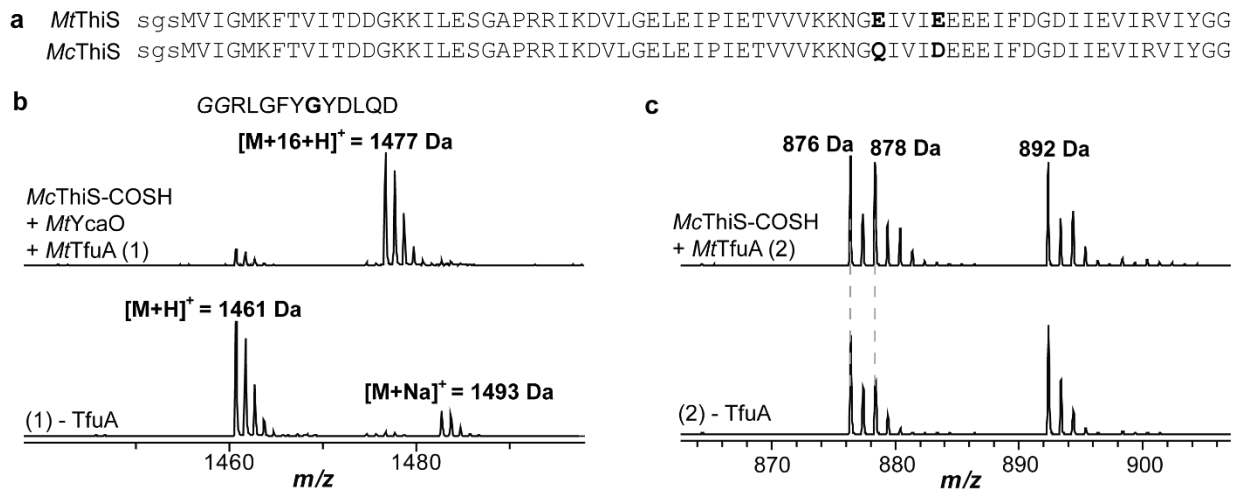
b

Ion	Calculated mass (Da)	Observed mass (Da)	Error (ppm)
b3	369.2609	369.2604	1.35
b4	468.3293	468.3284	1.92
b5	581.4134	581.4124	1.72
b6	744.4767	744.4753	1.88
b7	801.4982	801.4966	2.00
y3	298.1283	298.1279	1.34
y4	411.2124	411.2122	0.49
y5	510.2808	510.2797	2.16
y6	666.3819	666.3810	1.35
y7	779.4660	779.4644	2.05

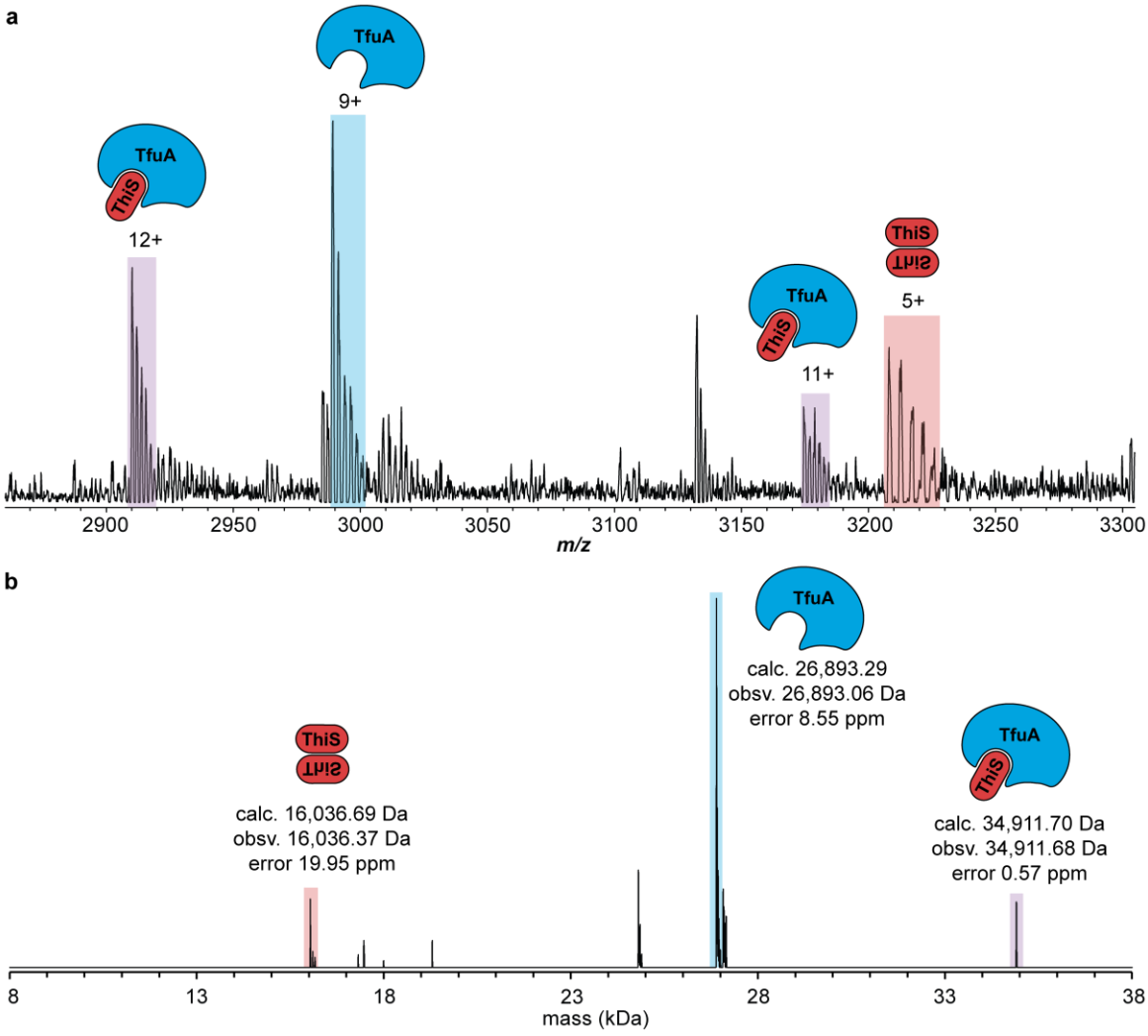
Supplementary Figure 3. High-resolution and tandem MS of the *Mt*ThiS-COOH C-terminal fragment. (**a**) The C-terminal GluC peptide fragment of *Mt*ThiS containing a [^{18}O] label observed in HRMS. (**b**) *m/z* 878.53 was subjected to CID with assigned ions indicated in tabular form. (**c**) MS/MS spectrum showing the +2 Da mass change observed for all y-ions but not observed for any b-ion, localizing the mass increase to the C-terminus.



Supplementary Figure 4. *Ma*ThiS-COSH is not a sulfur donor for thioamidation catalyzed by the *Mt* enzymes. (**a**) Sequence alignment of *Mt*ThiS (WP_048176273.1) and *Ma*ThiS (WP_011023978.1). The N-terminal SGS is derived from the cloning vector. (**b**) MALDI-TOF mass spectrum of *Ma*ThiS-COSH (*m/z* 10,749 Da) and *Ma*ThiS-COOH (*m/z* 10,733 Da). (**c**) MALDI-TOF mass spectrum of *Mt*ThiS-COSH (*m/z* 8,017 Da). *m/z* 7,930 represents *Mt*ThiS-COSH with the N-terminal Ser (from cloning) proteolytically removed. Asterisks indicate laser-induced deamination. (**d**) *Top*, MALDI-TOF mass spectrum of the McrA peptide after reaction with *Mt*YcaO, *Mt*TfuA, and non-cognate *Ma*ThiS-COSH, yielding essentially no product. *Bottom*, an identical reaction using cognate *Mt*ThiS-COSH, yielding fully thioamidated McrA peptide.



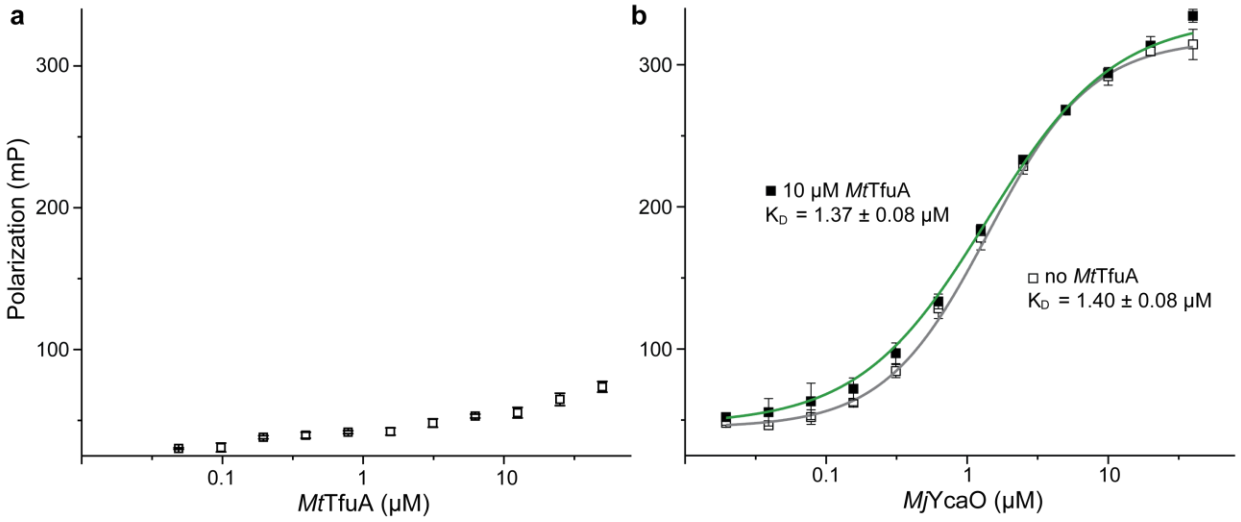
Supplementary Figure 5. *Mc*ThiS-COSH as the substrate for *Mt*TfuA-catalyzed hydrolysis and thioamidation. (**a**) Sequence alignment of *Mt*ThiS (WP_048176273.1) and *Mc*ThiS (AAB86213.1) with the two non-identical residues in bold. The N-terminal SGS is derived from the cloning vector. (**b**) *Top*, MALDI-TOF mass spectrum of the McrA peptide reacted with *Mt*YcaO, *Mt*TfuA, and *Mc*ThiS-COSH showing the thioamidated product (*m/z* 1,477 Da). *Bottom*, control with TfuA omitted. (**c**) *Top*, MALDI-TOF mass spectrum of the C-terminal GluC peptide fragment of *Mc*ThiS-COSH (*m/z* 892 Da). Upon treatment with TfuA in [^{18}O]-H_2O, *Mc*ThiS-COSH undergoes hydrolysis to *Mc*ThiS-COOH with incorporation of [^{18}O] (*m/z* 878 Da). *Bottom*, control with TfuA omitted.



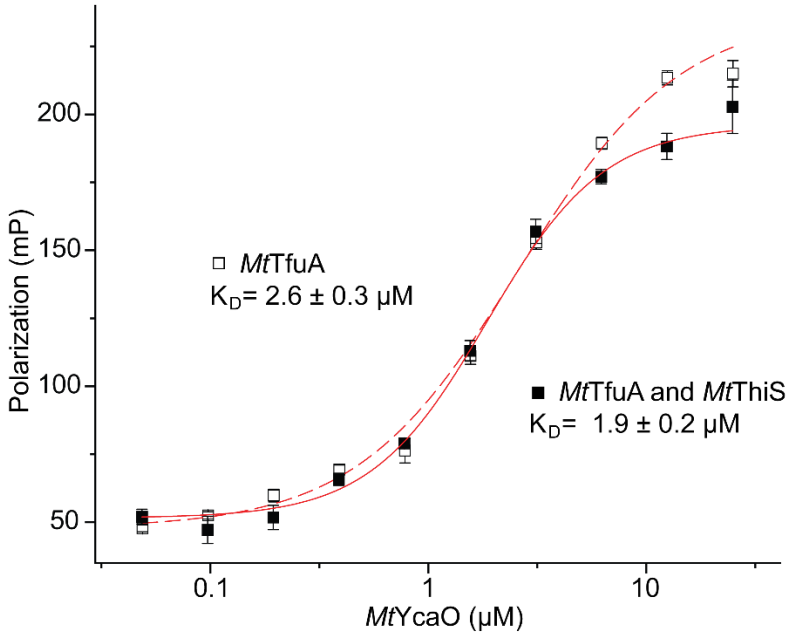
c

Mass (Da)	Relative Abundance	Fractional Abundance	Number of Charge States	Charge State Distribution
26893.06	100.0	26.9	8	6 - 13
26914.52	43.2	11.6	7	6 - 12
24809.04	26.4	7.1	6	5 - 10
26936.30	26.3	7.1	7	6 - 12
27071.14	21.3	5.7	8	6 - 13
16036.37	18.5	5.0	4	4 - 7
34911.68	17.7	4.8	9	7 - 15
27093.55	16.4	4.4	7	6 - 12
27152.70	13.9	3.7	6	6 - 11
26957.75	13.3	3.6	7	6 - 12

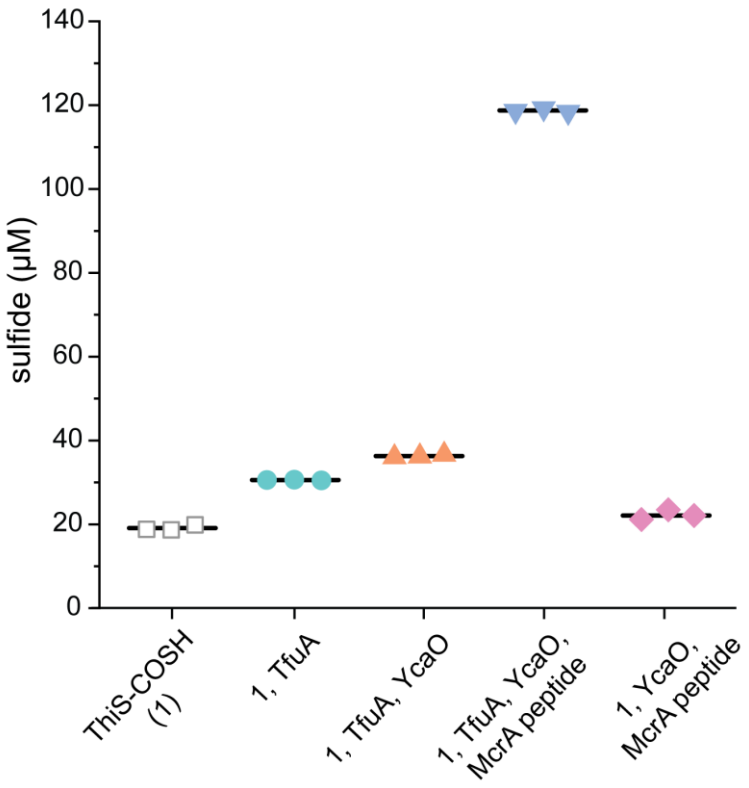
Supplementary Figure 6. Native mass spectra of *Mt*TfuA and *Mt*ThiS-COSH. (a) The native mass spectrum acquired using a ThermoFisher Q Exactive Ultra High Mass Range (UHMR) mass spectrometer showing the *Mt*TfuA monomer (blue), homodimeric *Mt*ThiS-COSH (red), and *Mt*TfuA:*Mt*ThiS-COSH heterodimer (purple) at various charge states. **(b)** Deconvoluted mass spectrum from data in panel **a** using BioPharma Finder v3.1. The experimental and calculated masses for *Mt*TfuA (His-tagged and N-terminal Met removed), *Mt*ThiS-COSH homodimer, and *Mt*TfuA:*Mt*ThiS-COSH heterodimer are indicated. (**c**) Mass table of the top ten species by relative abundance.



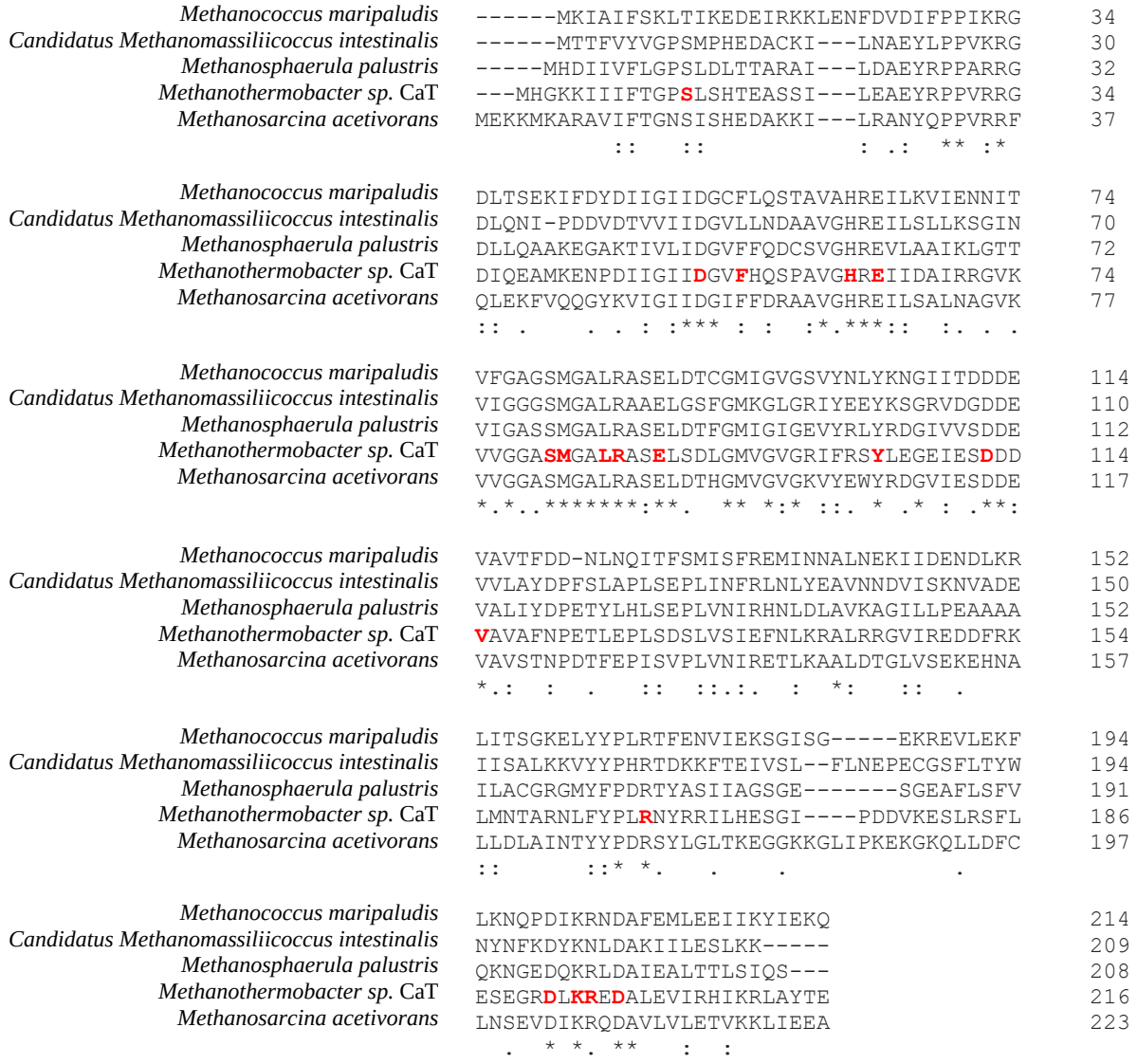
Supplementary Figure 7. *Mt*TfuA does not bind the McrA peptide or change its binding to *Mj*YcaO. (**a**) Fluorescence polarization (FP) titration of TfuA with FITC-labeled McrA peptide (*GG*-RLGFY**G**YDLQD). No significant polarization was observed. Polarization values are reported as mean values ± SD (*n* = 3 independent experiments). (**b**) FP titration of the same peptide with *Mj*YcaO in the presence (filled squares) and absence (open squares) of *Mt*TfuA. Polarization values are reported as mean values ± SD (*n* = 3 independent experiments) and analyzed through non-linear regression (dose-response model). The resultant estimates for K_D ± SE are reported.



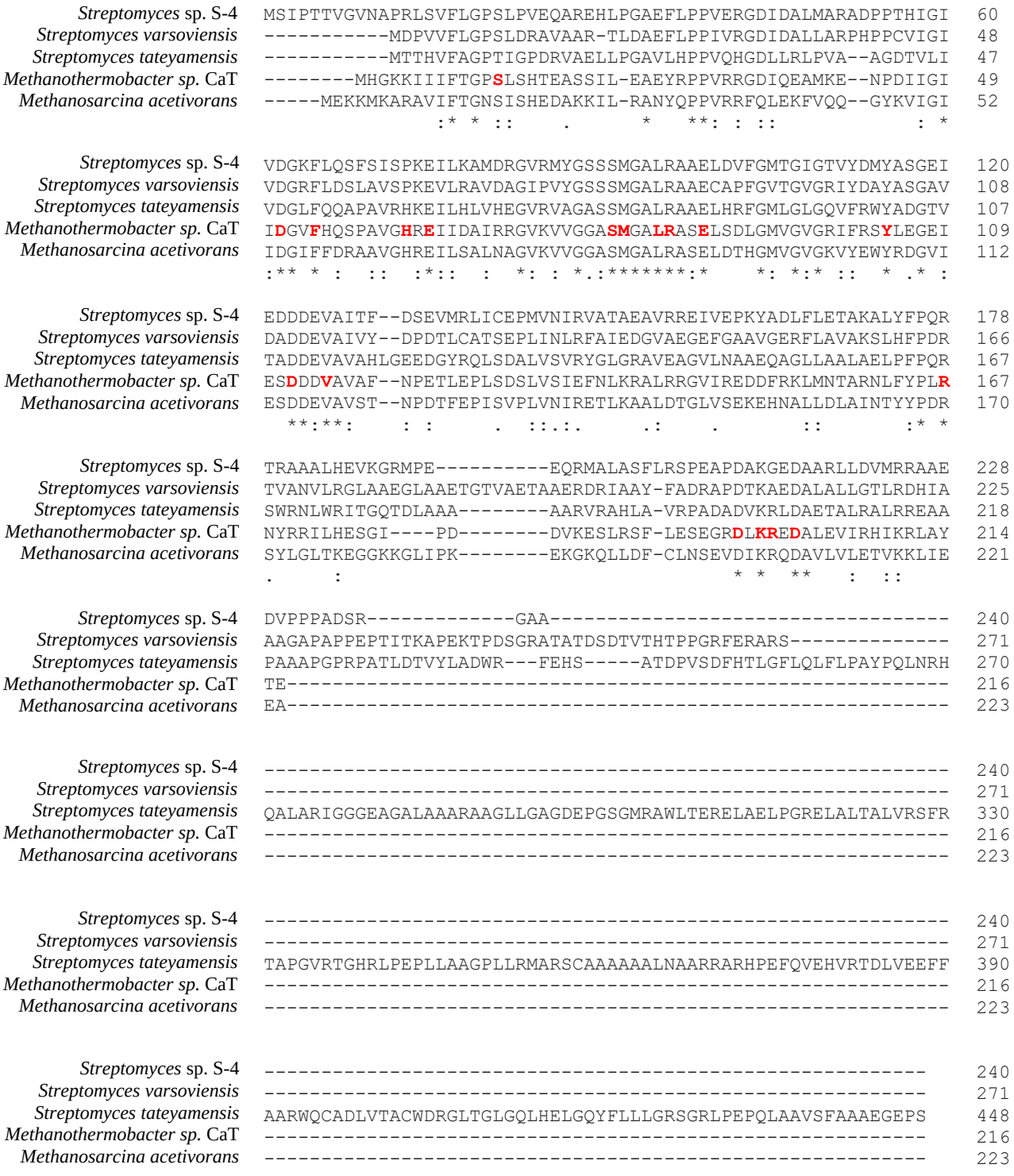
Supplementary Figure 8. *Mt*ThiS does not alter the *Mt*YcaO-McrA peptide binding. Fluorescence polarization (FP) titration of *Mt*YcaO with FITC-labeled McrA peptide (*GG*-RLGFY**G**YDLQD) in the presence of *Mt*TfuA (10 μM, open squares) or *Mt*TfuA and *Mt*ThiS (10 μM each, filled squares). Polarization values are reported as mean values ± SD (*n* = 3 independent experiments) and analyzed through non-linear regression (dose-response model). The resultant estimates for K_D ± SE are reported.



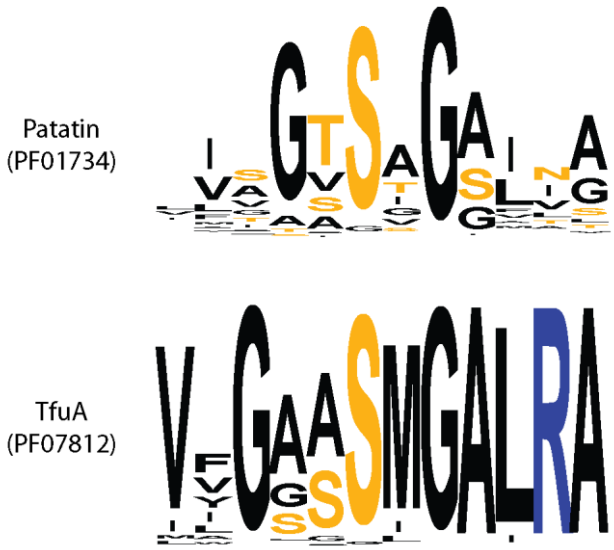
Supplementary Figure 9. Quantification of sulfide released from ThiS-COSH hydrolysis. Fluorescence quantification of sulfide production as a result of *Mt*ThiS-COSH (150 μM) hydrolysis after a 2 h reaction in the presence of different components (10 μM *Mt*TfuA, *Mt*YcaO, and 50 μM McrA peptide). ATP was omitted to prevent thioamidation of the peptide substrate. Individual data points (n = 3 independent experiments) and the mean values (lines) are presented.



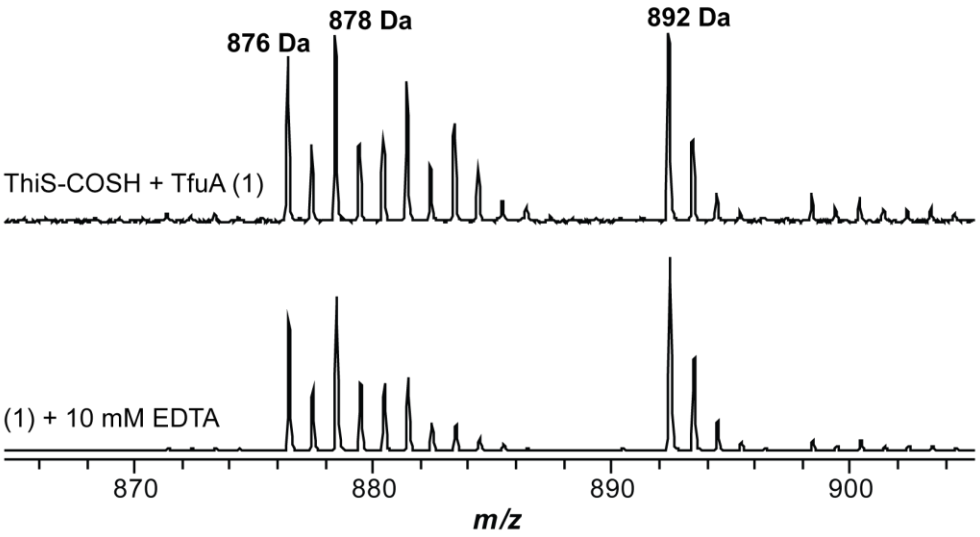
Supplementary Figure 10. Sequence alignment of diversity-maximized methanogenic TfuA proteins. TfuA protein sequences from *Methanothermobacter sp.* CaT2 (WP_048175617.1), *Methanococcus maripaludis* (AVB77113.1), *Candidatus Methanomassiliicoccus intestinalis* (WP_020449400.1), *Methanosphaerula palustris* (ACL16548.1), *Methanosarcina acetivorans* (WP_011020222.1) are aligned with Clustal Omega[2]. Residues highlighted in red were substituted for various aspects of the current study.



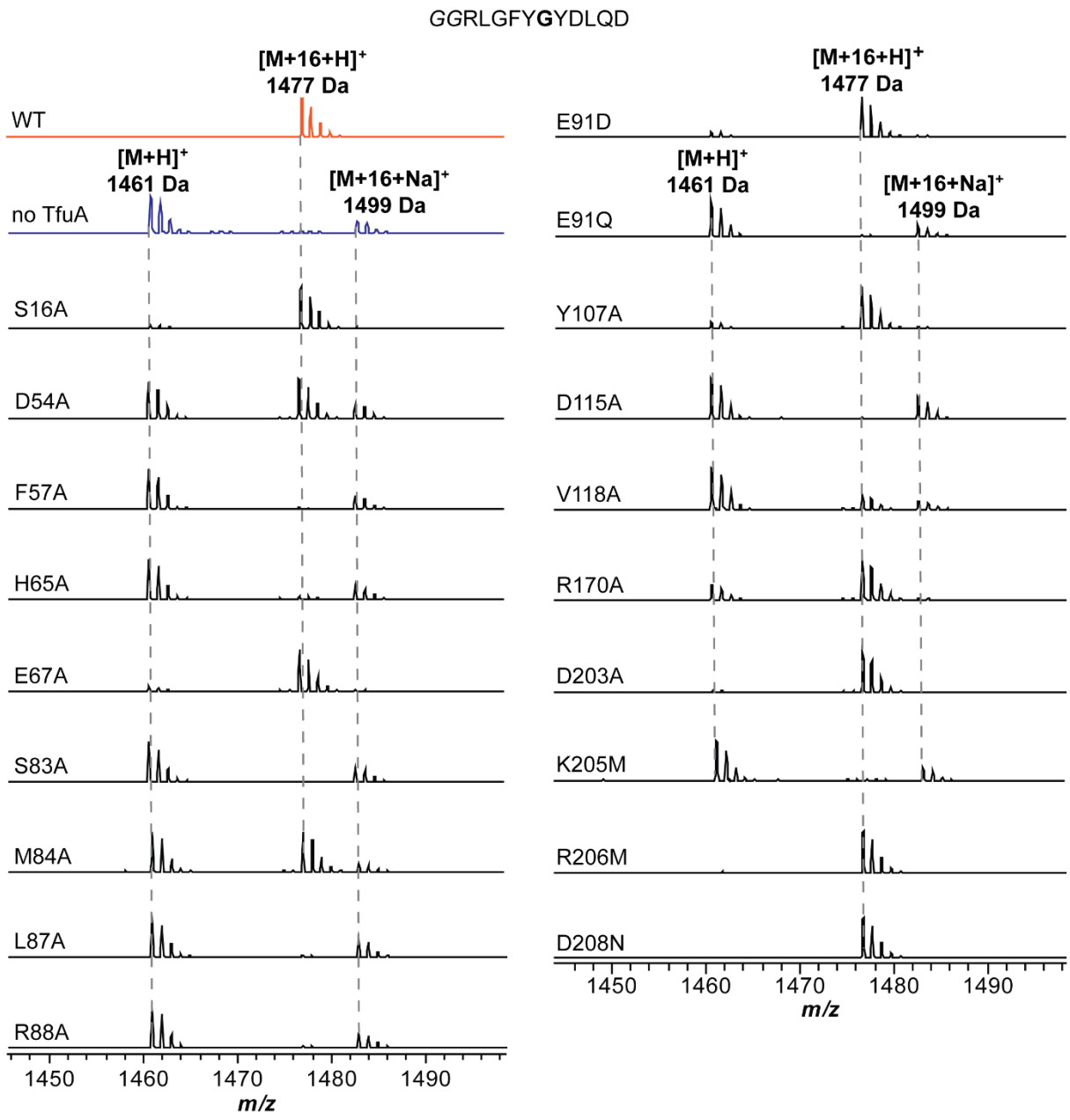
Supplementary Figure 11. Sequence alignment for TfuA proteins that have been linked to peptide thioamidation. TfuA protein sequences from *Methanothermobacter sp.* CaT2, *M. acetivorans*, *Streptomyces varsoviensis* (WP_063763981.1, thiovarsolin), *S. tateyamensis* (WP_110669062.1, thiopeptin), and *Streptomyces* sp. NRRL S-4 (WP_051834498.1, thiostreptamide) are aligned with Clustal Omega[2]. Residues highlighted in red were substituted for various aspects of the current study.



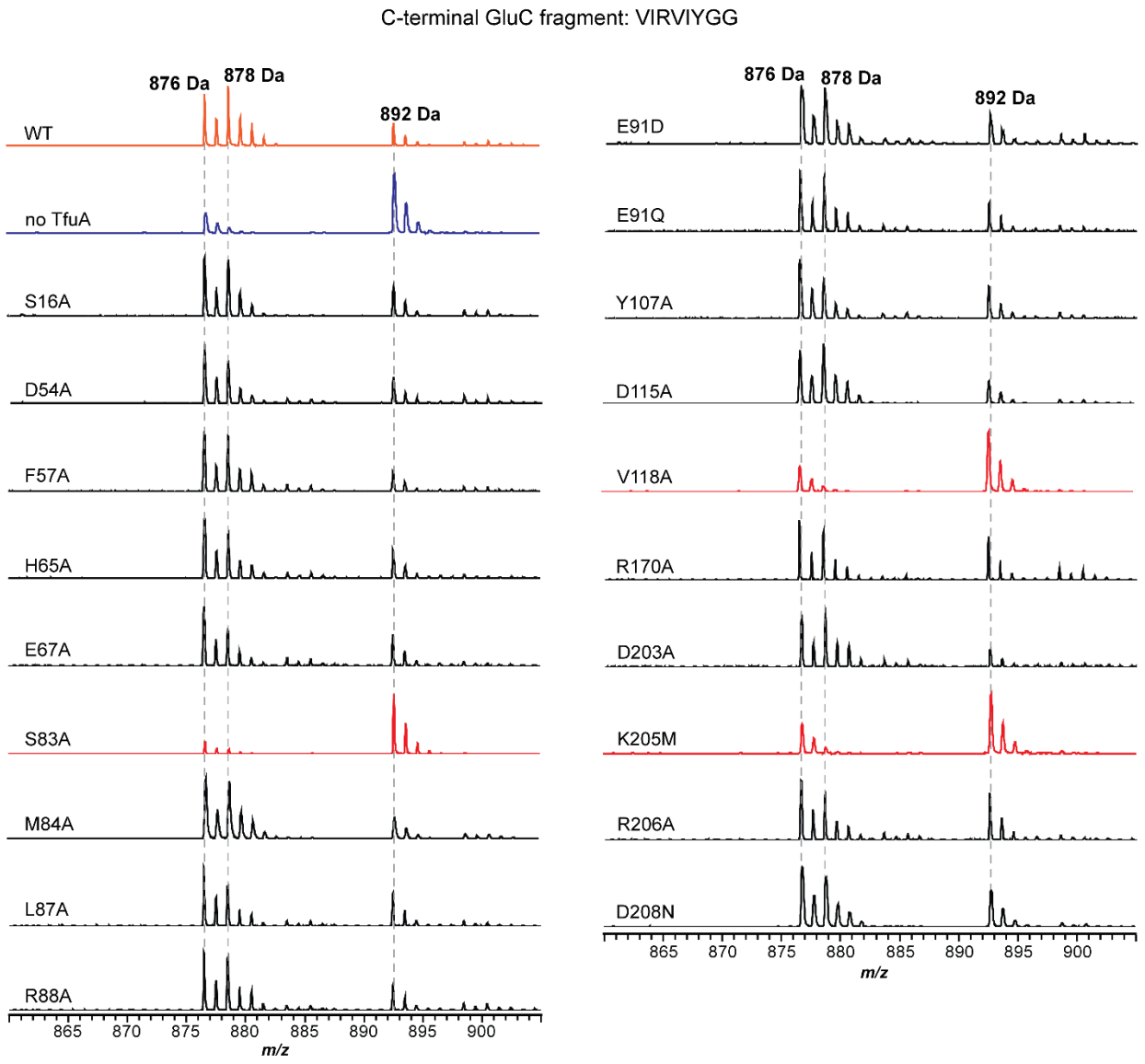
Supplementary Figure 12. Sequence logos of TfuA and patatin active sites. Sequence logos were generated for the Patatin and the TfuA protein families using WebLogo[3]. The Gly-(Xaa)$_{1\text{-}2}$-Ser-Xaa-Gly motifs from both families contain the active site Ser and are aligned in the HHpred analysis of TfuA[4]. Basic and uncharged polar residues are colored blue and orange, respectively.



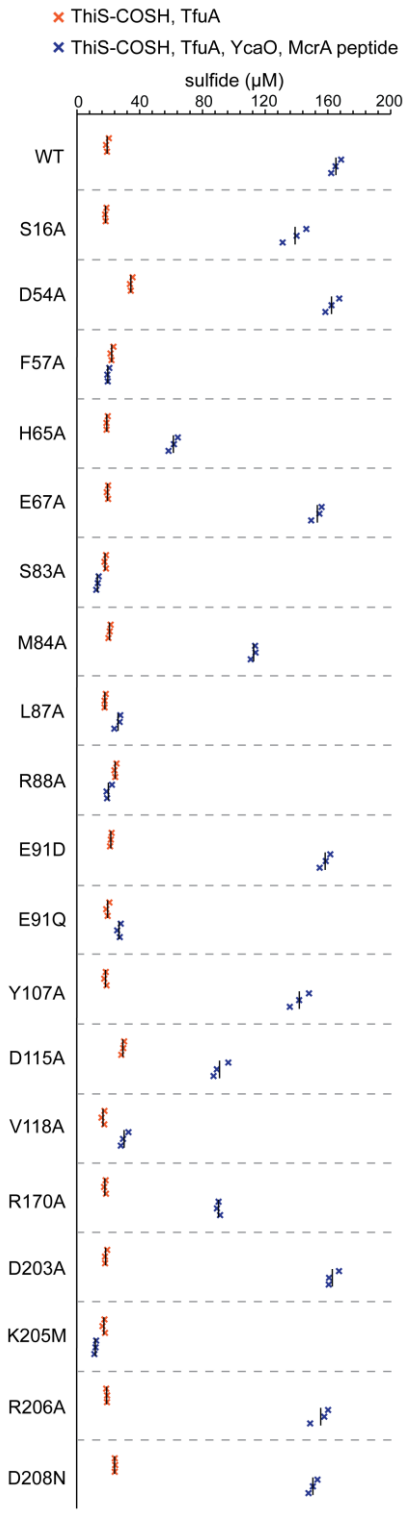
Supplementary Figure 13. *Mt*TfuA-catalyzed *Mt*ThiS-COSH hydrolysis is unaffected by EDTA. *Top*, MALDI-TOF mass spectrum of the C-terminal GluC peptide fragment of *Mt*ThiS-COSH (*m/z* 892 Da). Upon treatment with *Mt*TfuA in [^{18}O]-H_2O, *Mt*ThiS-COSH undergoes hydrolysis to *Mt*ThiS-COOH with incorporation of [^{18}O] (*m/z* 878 Da). *Bottom*, identical experiment with 10 mM EDTA added.



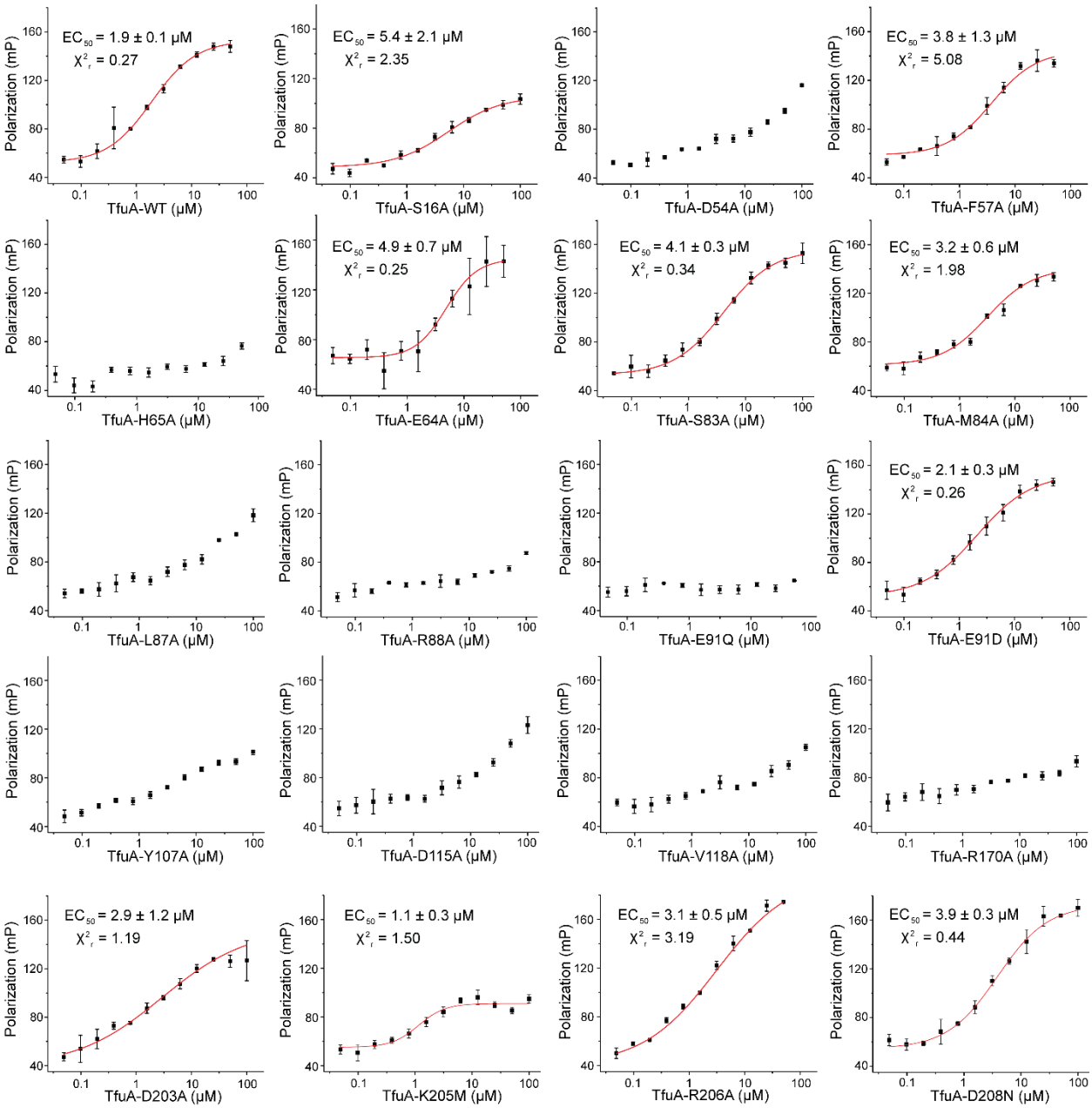
Supplementary Figure 14. MALDI-TOF-MS analysis of McrA peptide thioamidation using *Mt*TfuA variants. MALDI-TOF mass spectra of the McrA peptide (40 μM) after reaction with *Mt*YcaO (2 μM), *Mt*ThiS-COSH (50 μM), and *Mt*TfuA (2 μM) variants under identical conditions for 1 h.



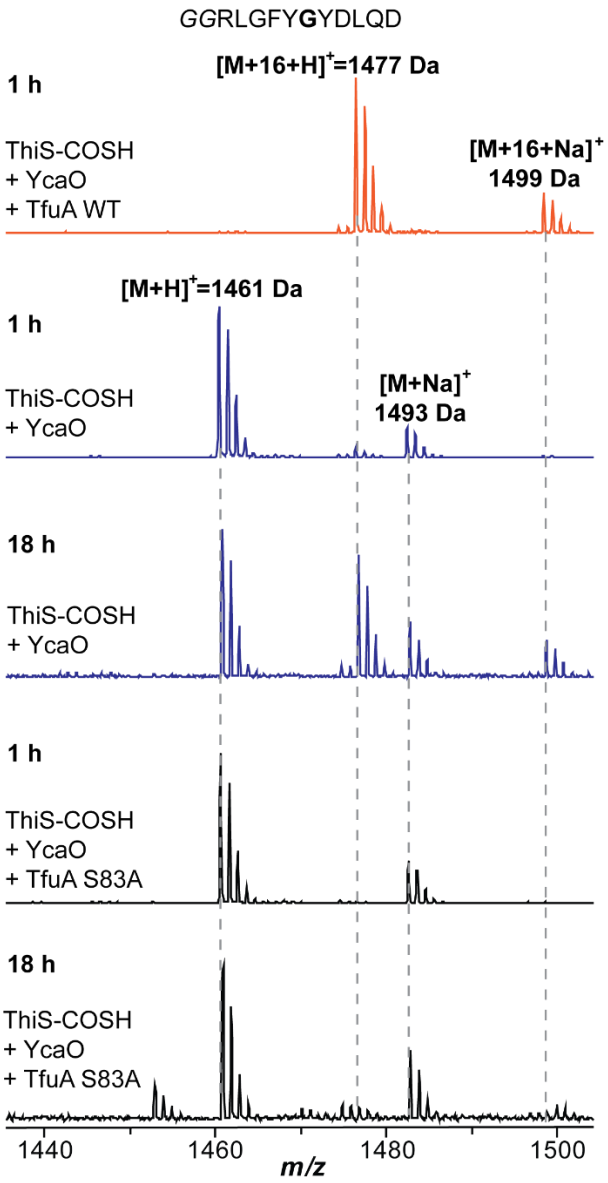
Supplementary Figure 15. MALDI-TOF-MS analysis of *Mt*ThiS-COSH hydrolysis using *Mt*TfuA variants. MALDI-TOF mass spectra of the C-terminal GluC peptide fragment of *Mt*ThiS (45 μM, carboxylate, *m/z* 876 Da; [^{18}O]-carboxylate, *m/z* 878 Da; thiocarboxylate, *m/z* 892 Da) after reaction with *Mt*TfuA variants (5 μM) under identical conditions. Control reactions with wild-type *Mt*TfuA (orange, 5 μM) and no TfuA (blue) are highlighted. Three *Mt*TfuA variants gave background-levels of *Mt*ThiS-COSH hydrolysis (S83A, V118A, and K205M, red).



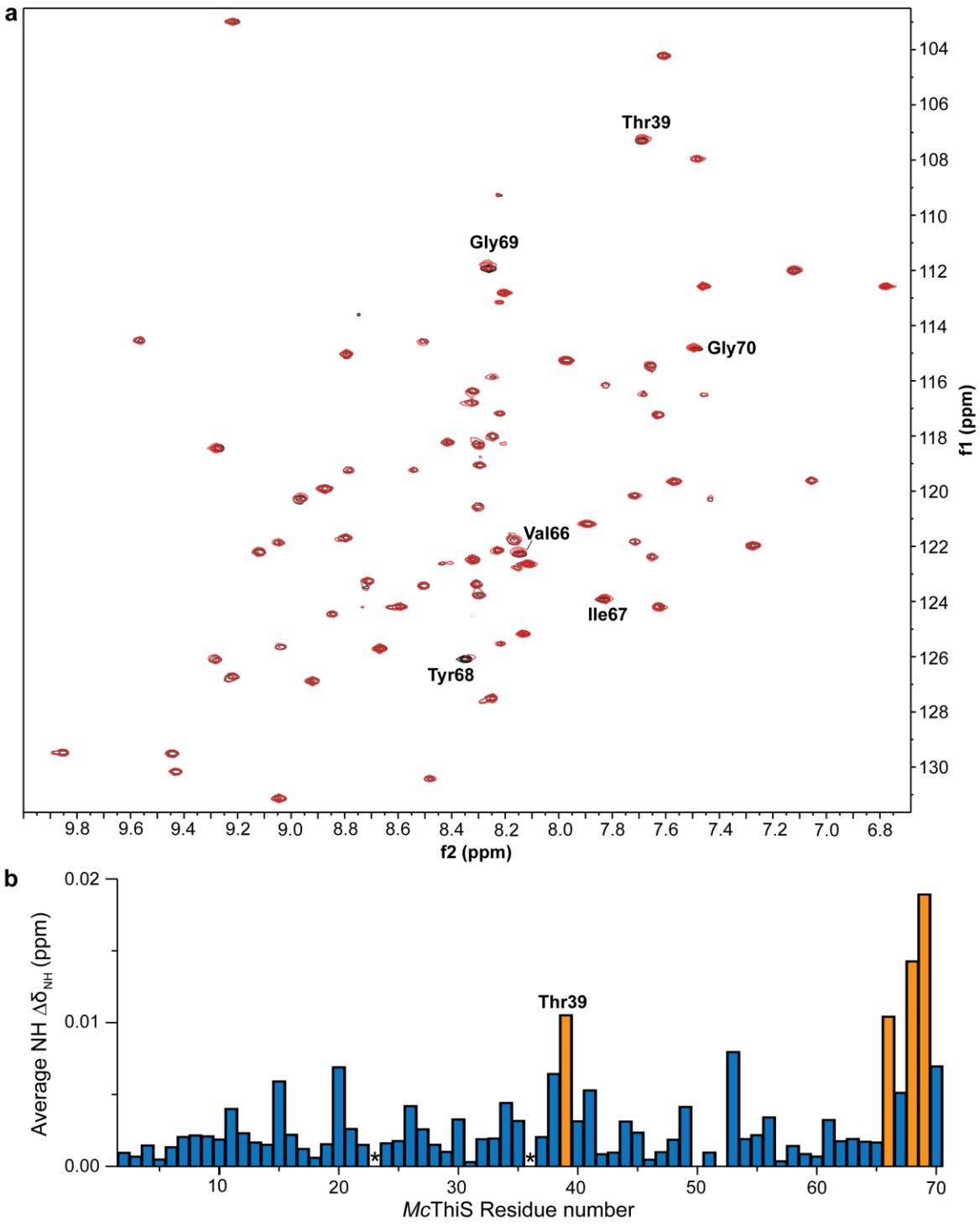
Supplementary Figure 16. Sulfide release from *Mt*ThiS-COSH hydrolysis using *Mt*TfuA variants. Endpoint fluorescence quantification of sulfide produced from *Mt*ThiS-COSH (150 μM) hydrolysis upon reaction with the *Mt*TfuA variant alone (10 μM), and *Mt*TfuA variant (10 μM), *Mt*YcaO (10 μM), and McrA peptide (50 μM). ATP was absent in all reactions. Individual data points (n = 3 independent experiments) and the mean values (lines) are presented.



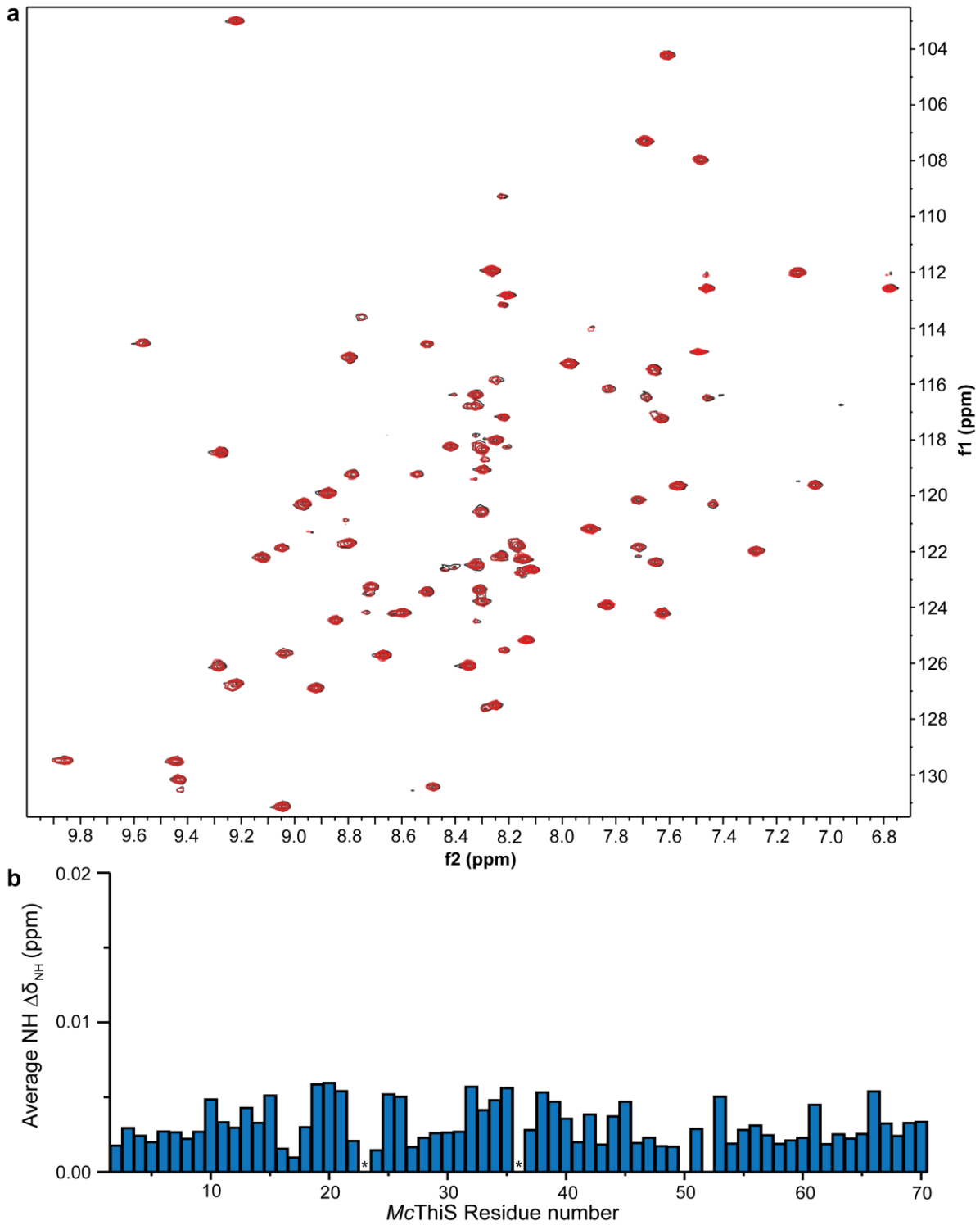
Supplementary Figure 17. Fluorescence polarization analysis of *Mt*TfuA variants. FP titrations of *Mt*TfuA variants using FITC-labeled McrA peptide (*GG*-RLGFY**G**YDLQD) and 1.5 μM *Mt*YcaO. Polarization values are presented as the mean values ± SD (n = 3 independent experiments) and fitted into a dose-response model via non-linear regression analysis. The resultant estimates for EC_{50} ± SE and reduced chi-square values are reported.



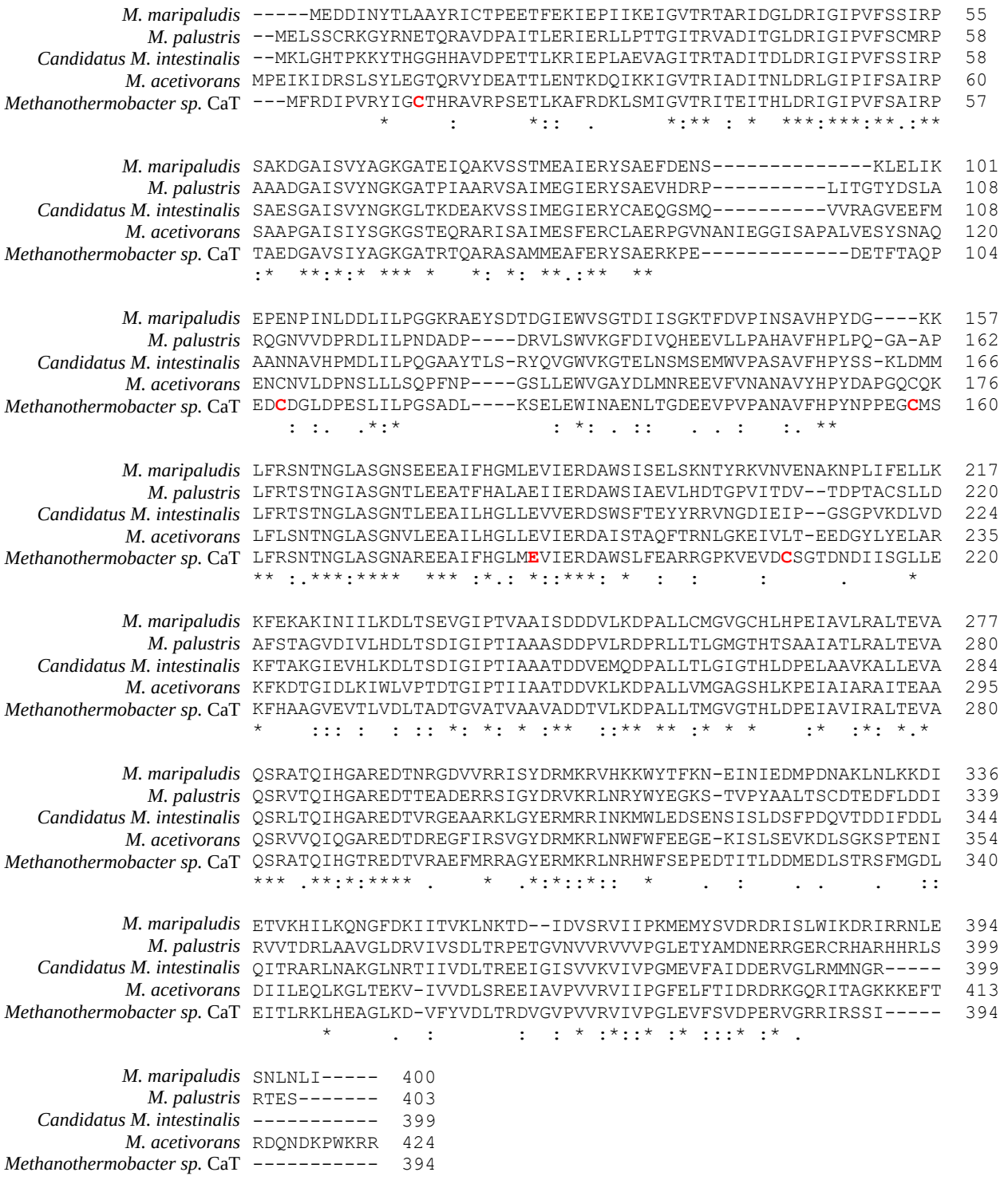
Supplementary Figure 18. MALDI-TOF-MS analysis of reactions using *Mt*TfuA-S83A variant. *Top*, MALDI-TOF mass spectrum of the McrA peptide after reaction with *Mt*YcaO, *Mt*TfuA, and *Mt*ThiS-COSH showing thioamidated product (*m/z* 1,477 Da). Identical reactions with TfuA omitted (blue) or substituted by TfuA-S83A (black) at 1 h and 18 h are shown. Reactions containing all components were completed within 1 h, whereas reactions lacking TfuA remained incomplete even after 18 h. Reactions containing the TfuA-S83A were severely inhibited, likely from binding to ThiS-COSH and hindering hydrolysis.



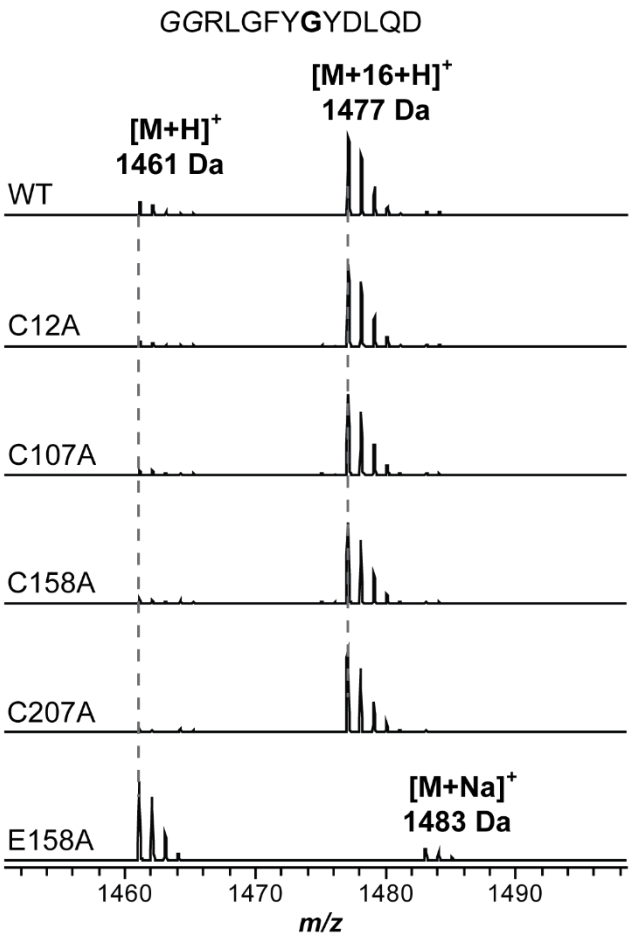
Supplementary Figure 19. *Mc*ThiS ^{1}H,^{15}N-HMQC spectra upon *Mt*TfuA-S83A variant titration. (**a**) An overlay of *Mc*ThiS ^{1}H,^{15}N-HMQC spectra in the presence (red) and absence (black) of *Mt*TfuA-S83A (2 equiv.). Residues with significant chemical shift perturbations are labeled. (**b**) Chemical shift perturbation for each residue. Residues with the largest chemical shift perturbation upon titration with *Mt*TfuA-S83A are highlighted in orange. Asterisks, Pro residues. *Mt*TfuA-S83A binds to *Mc*ThiS in a manner similar to wild type *Mt*TfuA.



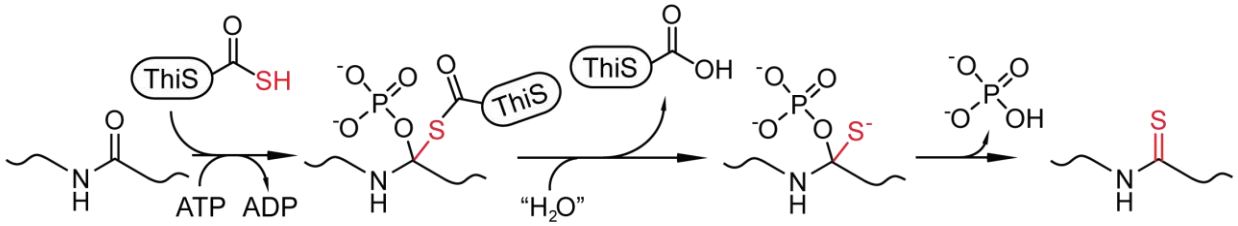
Supplementary Figure 20. *Mc*ThiS ^{1}H,^{15}N-HMQC spectra upon *Mt*TfuA-V118A variant titration. (**a**) An overlay of *Mc*ThiS ^{1}H,^{15}N-HMQC spectra in the presence (red) and absence (black) of *Mt*TfuA-V118A (6 equiv.). (**b**) Chemical shift perturbation for each residue. Asterisks, Pro residues. *Mt*TfuA-V118A does not bind to ThiS in a wild type-like manner.



Supplementary Figure 21. Sequence alignment of representative TfuA-dependent methanogenic YcaO proteins. YcaO protein sequences from *Methanothermobacter sp.* CaT2 (WP_048175616.1), *M. maripaludis* (AVB77164.1), *Candidatus M. intestinalis* (WP_020449401.1), *M. palustris* (ACL16549.1), *M. acetivorans* (WP_011020223.1) are aligned with Clustal Omega[2]. Residues highlighted in red were substituted in the current study.



Supplementary Figure 22. MALDI-TOF-MS analysis of McrA peptide thioamidation using *Mt*YcaO variants. MALDI-TOF mass spectra of the McrA peptide after reaction with *Mt*TfuA, *Mt*ThiS-COSH, and *Mt*YcaO variants under identical conditions for 1 h. *Mt*YcaO Cys variants showed wild-type-like activity. As a control, the presumed ATP-binding deficient *Mt*YcaO variant (E158A) exhibited abolished activity.



Supplementary Figure 23. Alternative sulfur transfer mechanism. ThiS-COSH attacks activated McrA peptide to form a thioester-hemiorthoamide intermediate. Subsequent hydrolysis aided by TfuA leads to the formation of thioamide and release of ThiS-COOH.

References

1. Krishnamoorthy, K. & Begley, T. P. Reagent for the Detection of Protein Thiocarboxylates in the Bacterial Proteome: Lissamine Rhodamine B Sulfonyl Azide. *J. Am. Chem. Soc.* **132**, 11608–11612 (2010).
2. F, M. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
3. Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
4. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).