Supporting information

Probing Ligand Binding Sites on Large Proteins by NMR Spectroscopy of Genetically Encoded Non-Canonical Amino Acids

Kasuni B. Ekanayake,[1] Mithun C. Mahawaththa,[1] Haocheng Qianzhu,[2] Elwy H. Abdelkader,[1] Josemon George,[2] Sven Ullrich,[2] Rhys B. Murphy,[2] Sarah E. Fry,[3] Jason Johansen-Leete,[3] Richard J. Payne,[3] Christoph Nitsche,[2] Thomas Huber,[2] and Gottfried Otting[1]*

[1] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[3] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science and School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia

[*] Corresponding author

Contents

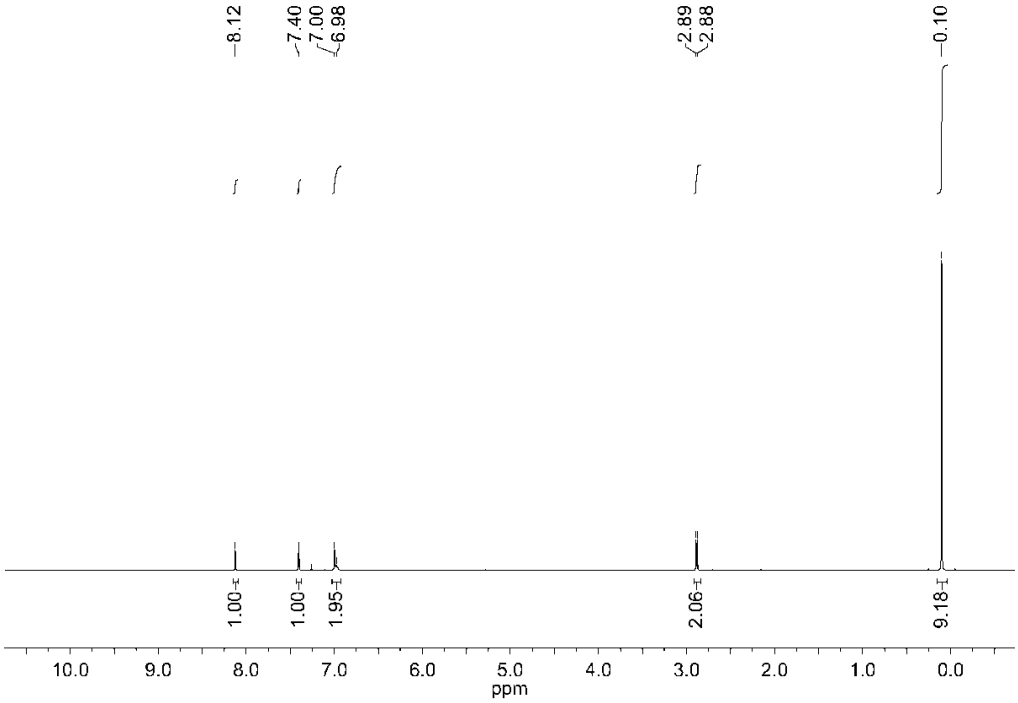

Figure S1. ^{1}H NMR spectrum of *N*-((trimethylsilyl)methyl)-1*H*-imidazole-1-carboxamide (**2**) (400 MHz, $CDCl_3$, 298 K).

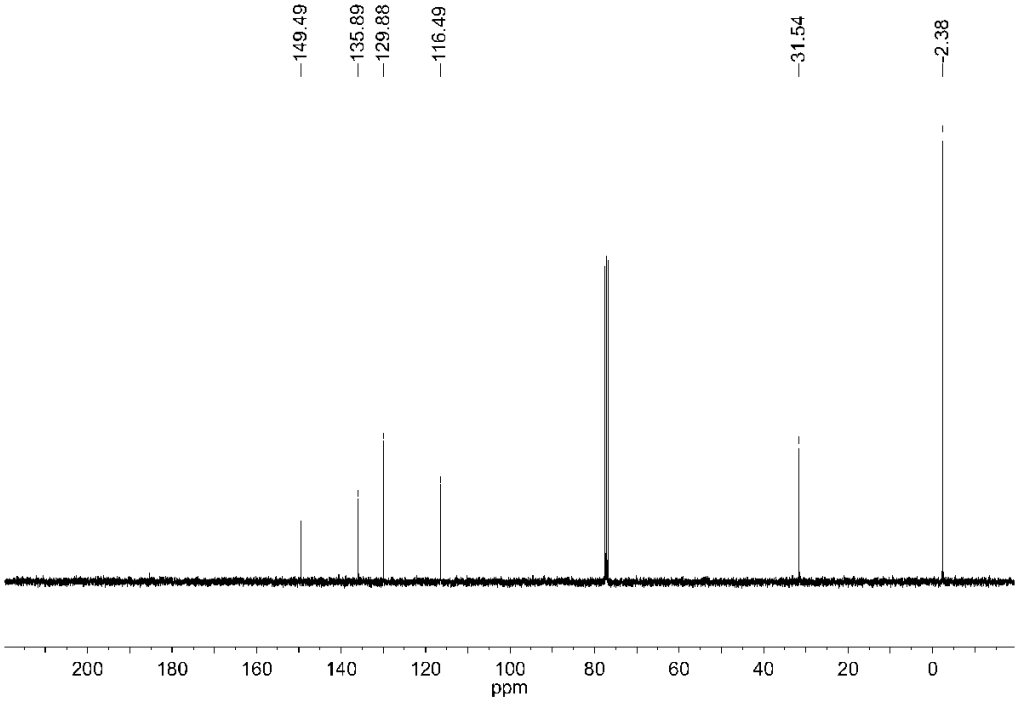

Figure S2. ^{13}C NMR spectrum of *N*-((trimethylsilyl)methyl)-1*H*-imidazole-1-carboxamide (**2**) (100 MHz, $CDCl_3$, 298 K).

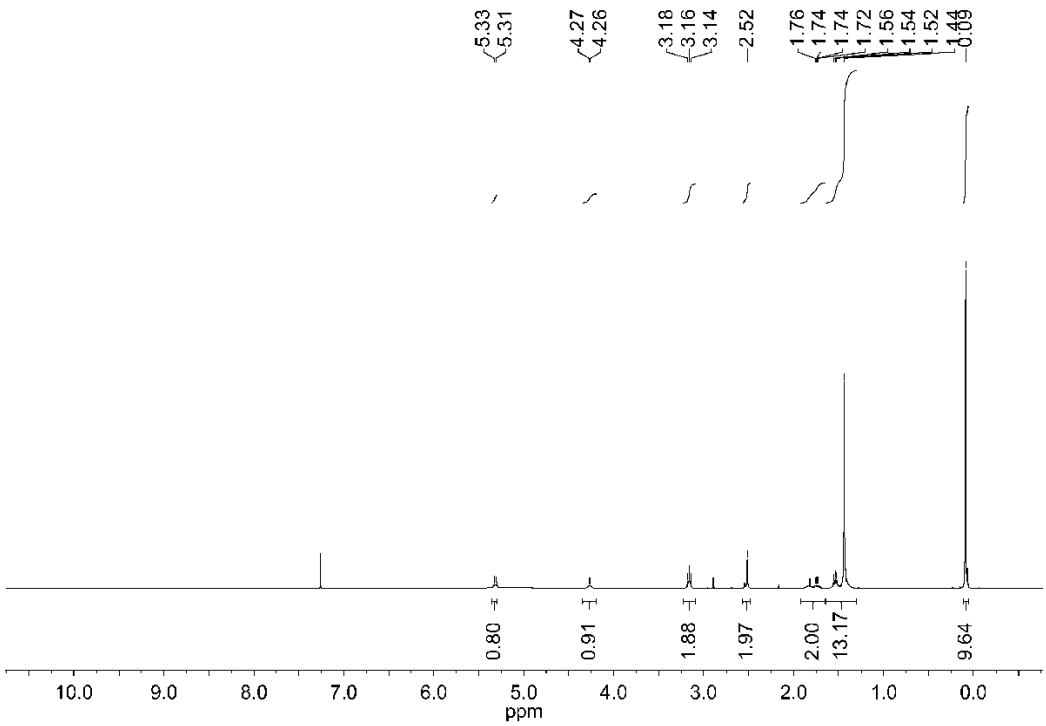

Figure S3. ^{1}H NMR spectrum of N^2-(*tert*-butoxycarbonyl)-N^6-(((trimethylsilyl)methyl) carbamoyl)-L-lysine (**3**) (400 MHz, $CDCl_3$, 298 K).

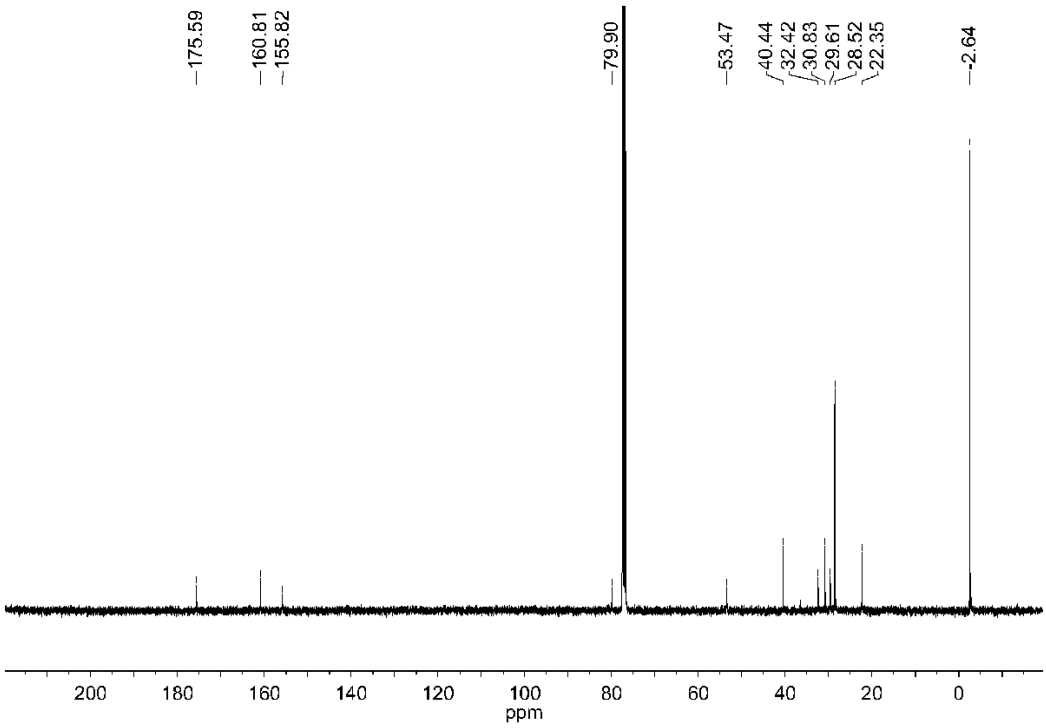

Figure S4. ^{13}C NMR spectrum of N^2-(*tert*-butoxycarbonyl)-N^6-(((trimethylsilyl)methyl) carbamoyl)-L-lysine (**3**) (100 MHz, $CDCl_3$, 298 K).

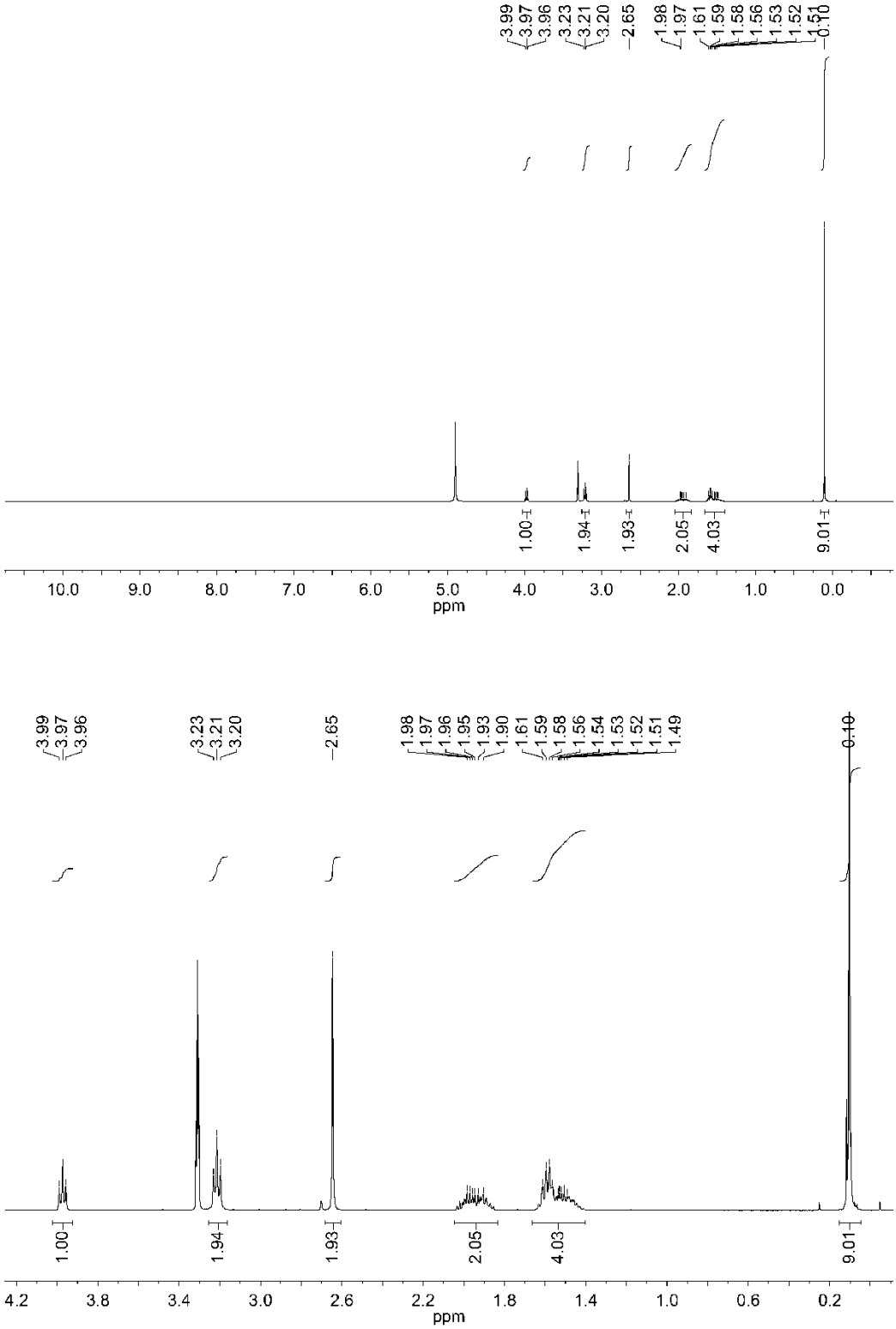

Figure S5. ^{1}H NMR spectrum of N^{6}-(((trimethylsilyl)methyl)carbamoyl)-L-lysine hydrochloride, TMSNK hydrochloride (**4**) (400 MHz, MeOD-d_4, 298 K).

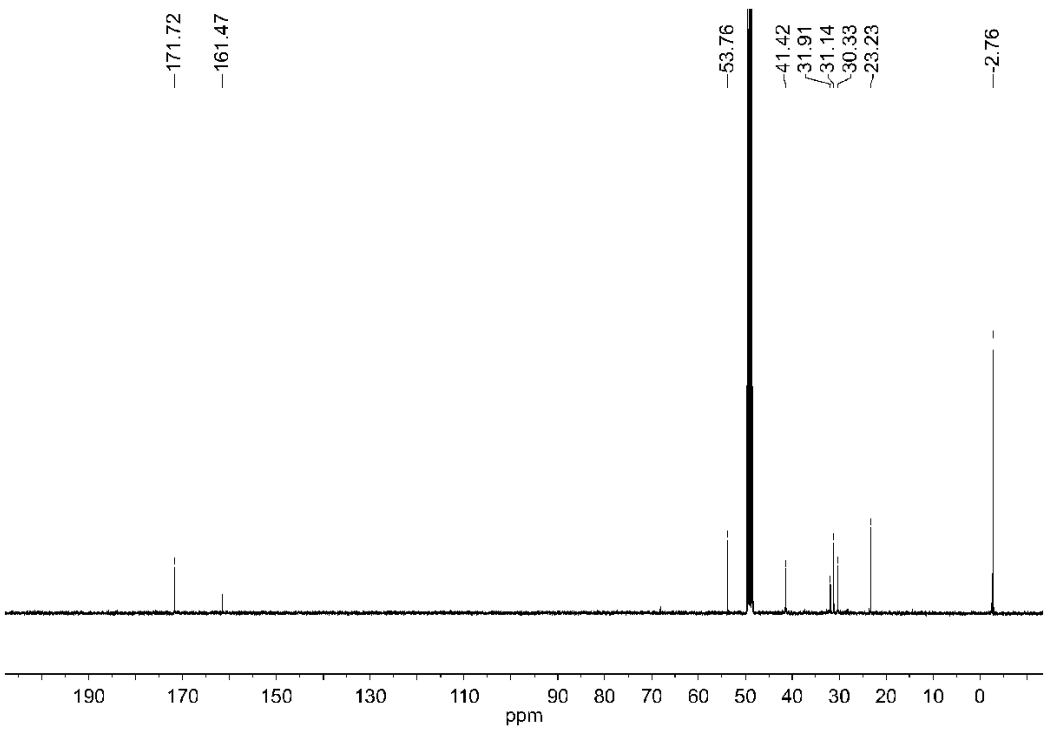

Figure S6. ^{13}C NMR spectrum of N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine hydrochloride, TMSNK hydrochloride (**4**) (100 MHz, MeOD-d_4, 298 K).

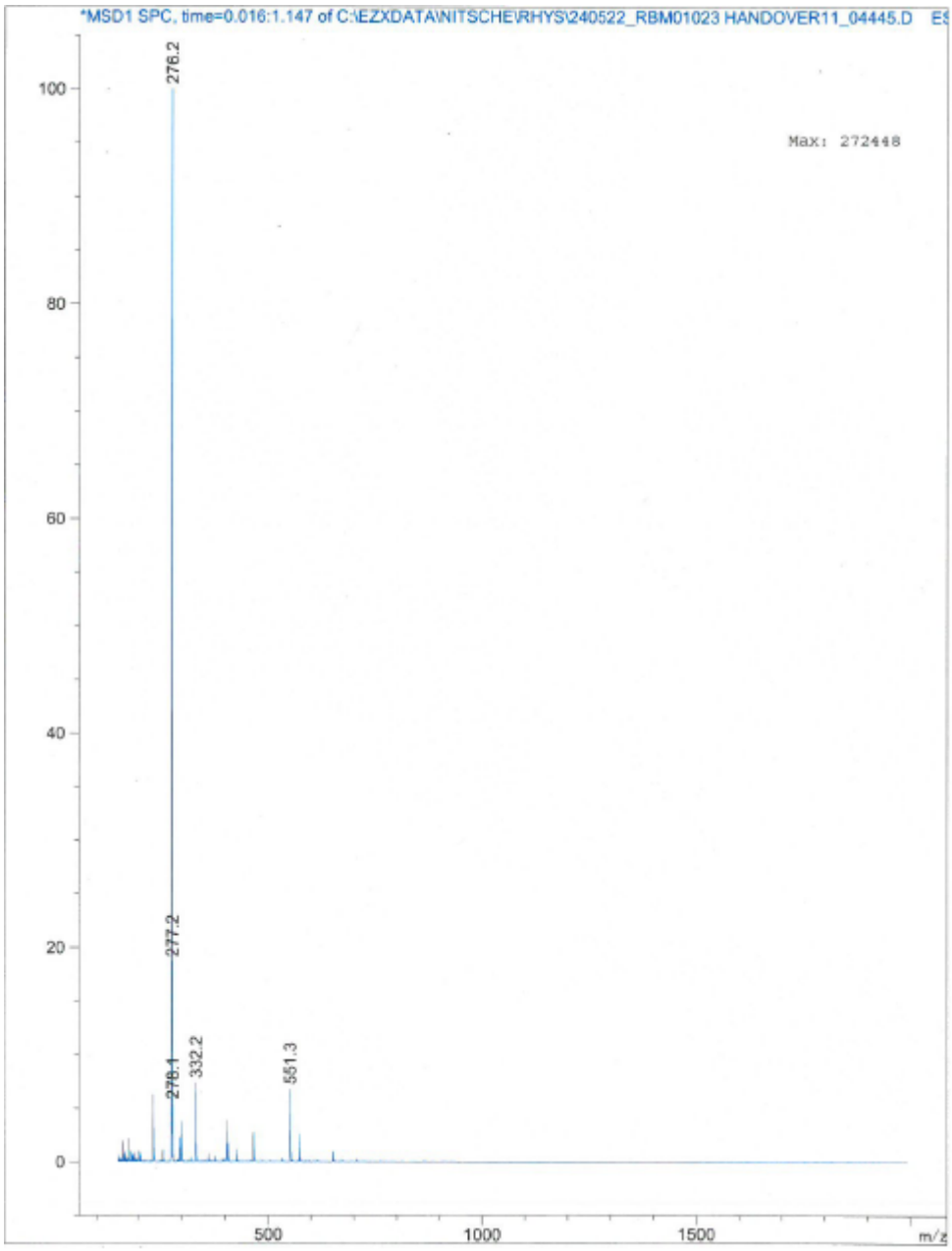

Figure S7. LR-ESI-MS (positive mode) of N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine hydrochloride, TMSNK hydrochloride (**4**). $[C_{11}H_{26}N_3O_3Si]^+$ $[M+H]^+$: Calc: 276.2. Found: 276.2.

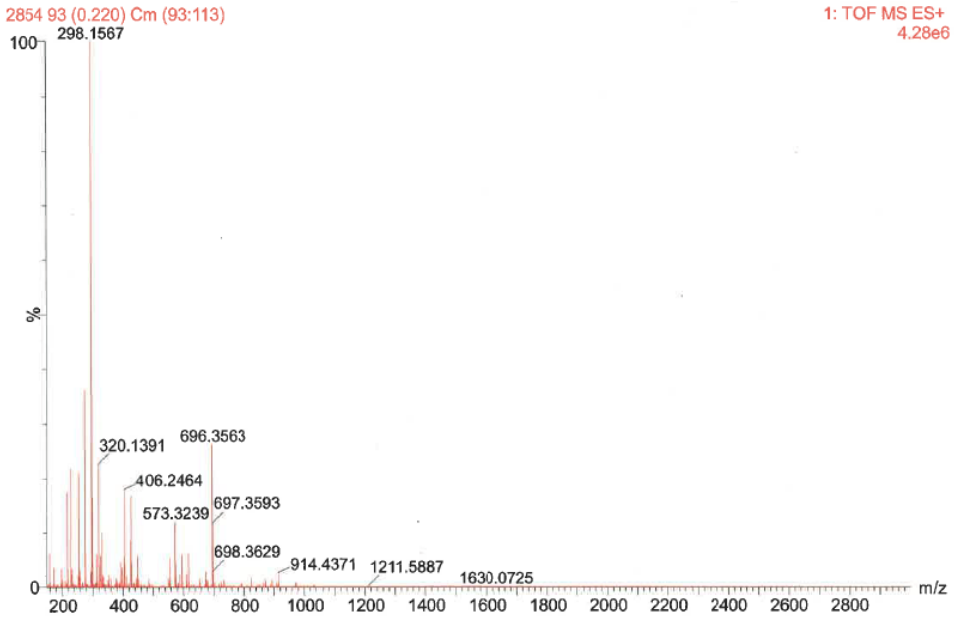

Figure S8. HR-ESI-TOF-MS (positive mode) of N^6-(((trimethylsilyl)methyl)carbamoyl)-L-lysine hydrochloride, TMSNK hydrochloride (**4**). $[C_{11}H_{25}N_3O_3SiNa]^+$ $[M+Na]^+$: Calc: 298.1563. Found: 298.1567.

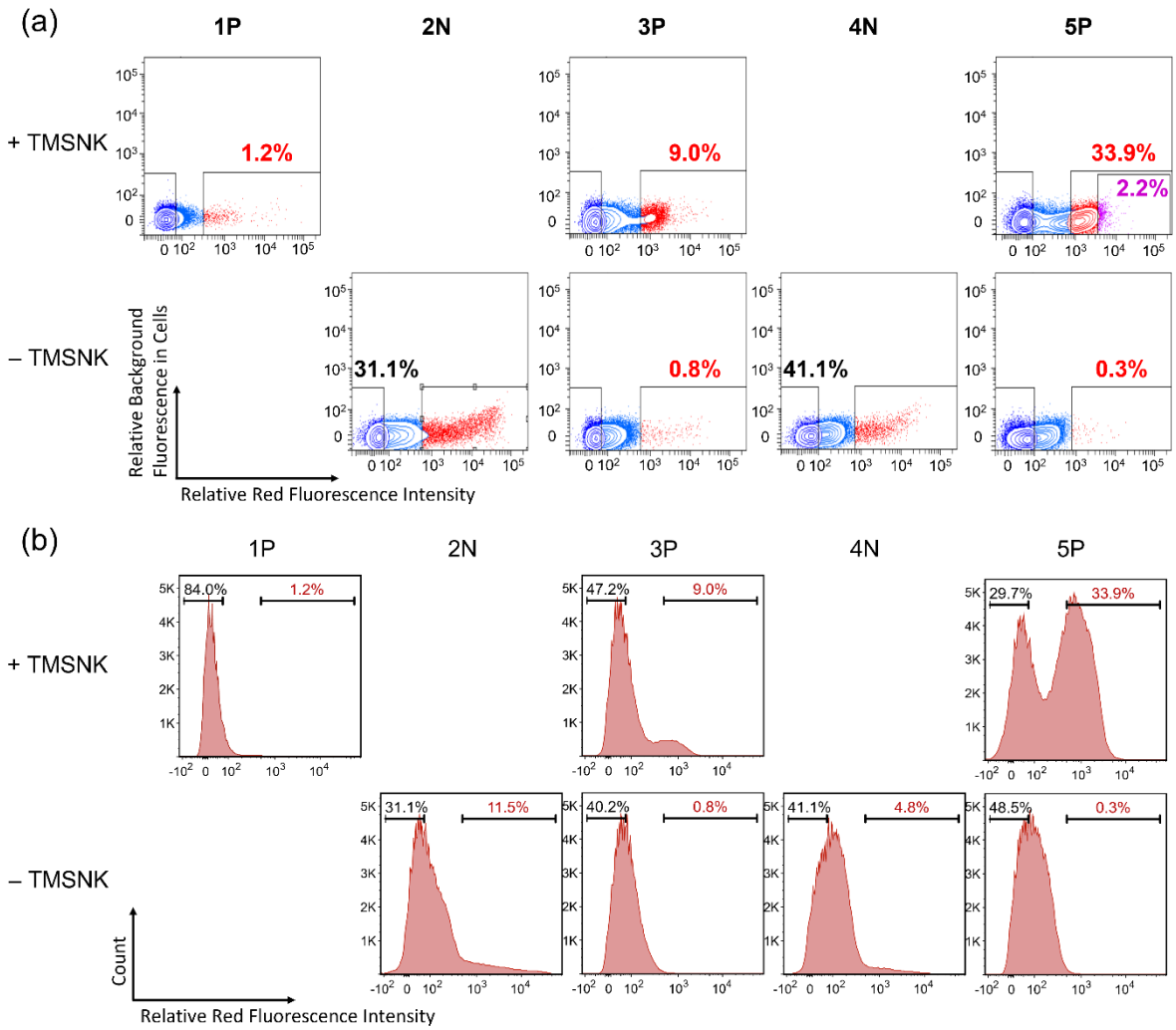

Figure S9. Selection of an RS enzyme specific for TMSNK from a G1 PylRS library. (a) FACS experiments for RS enzyme selection. The horizontal axis indicates the relative intensity of red fluorescence. The vertical axis plots the level of background fluorescence in cells excited at 488 nm. **P** stands for positive selection rounds where cells were cultured in the presence of TMSNK; **N** indicates negative rounds without ncAA supply. Cells with high level of RFP fluorescence in the **P** rounds were collected (indicated by boxes drawn) and re-cultured. Following **N** rounds, cells with low RFP expression were selected for use in subsequent rounds. Identical cell cultures omitting TMSNK (–TMSNK) in **3P** and **5P** rounds serve as control. Cells shown in violet in the **5P+** sample (2.2% of the total population) were collected and characterized further. (b) Histograms of the selection rounds to identify active RS enzymes for TMSNK via FACS screening. The difference in RFP fluorescence intensity of cells grown with or without TMSNK serves as an indicator of the presence of TMSNK-specific RS enzymes in the gene pool. The plots show the cell count versus relative intensity of red fluorescence.

Table S1. Mutations found in seven selected colonies with PylRS variants that recognize TMSNK.[a]

Mutant	Site						
G1PylRS wt	L124	Y125	N165	V167	Y204	A221	W237
TMSNK07	L	Y	N	G	F	C	W
TMSNK11	L	Y	N	G	W	A	W
TMSNK27	L	Y	N	G	W	C	W
TMSNK45	L	Y	N	G	W	C	W
TMSNK47	L	Y	N	G	W	A	W
TMSNK52	L	Y	N	G	F	C	W
TMSNK54	L	Y	N	G	W	A	W

[a] Mutation sets with white/grey/red background colour are of the same variant. The colony with mutation set TMSNK11 produced the strongest fluorescence and the RS isolated from this colony is referred to as G1TMSNKRS, which was used for installing TMSNK residues in M^{pro}.

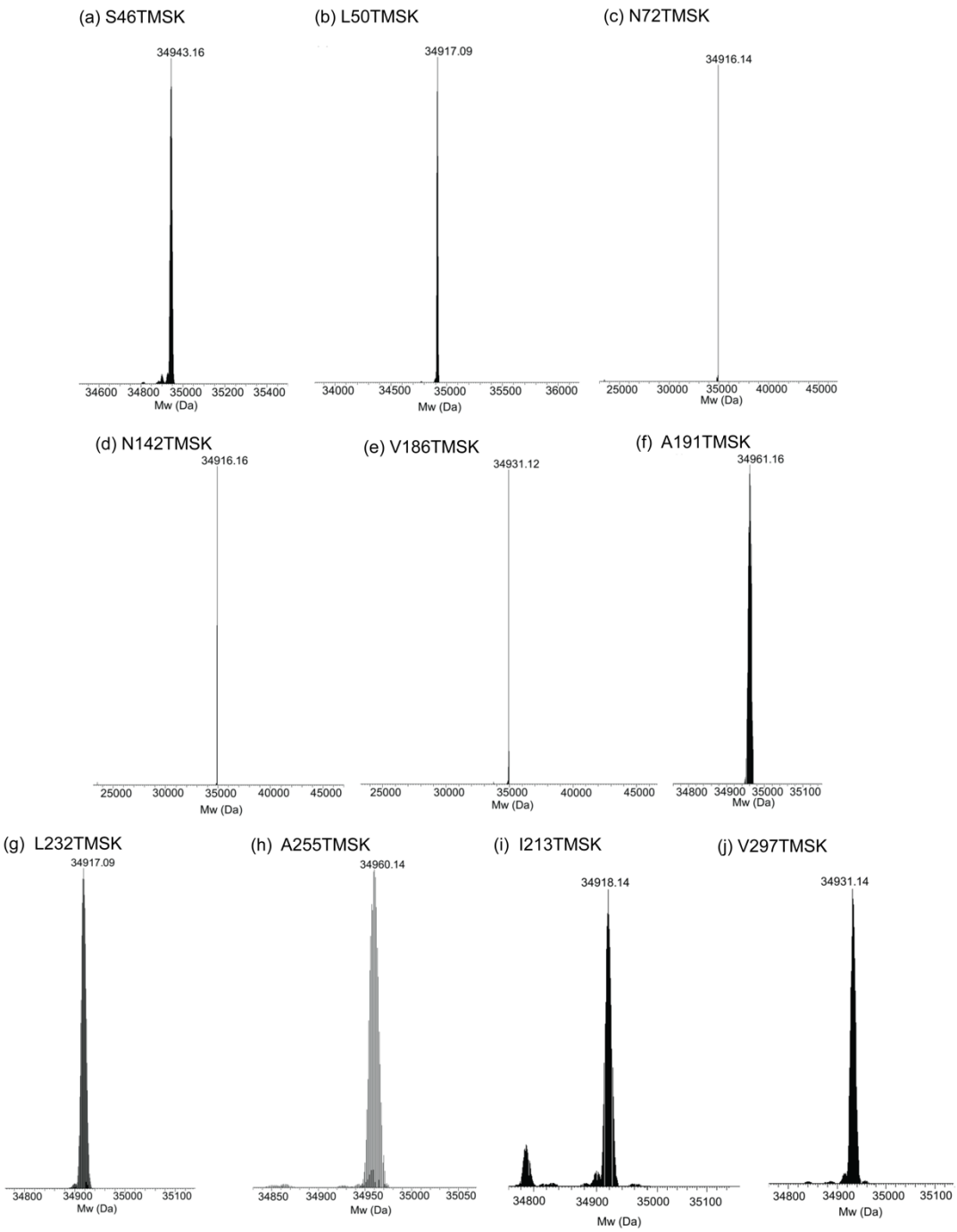

Figure S10. Intact protein mass spectrometric analysis of TMSK mutants. Calculated molecular weights in brackets. (a) S46TMSK (34,945.06 Da). (b) L50TMSK (34,918.98 Da). (c) N72TMSK (34,918.03 Da). (d) N142TMSK (34,918.03 Da). (e) V186TMSK (34,933.00 Da). (f) A191TMSK (34,961.06 Da). (g) L232TMSK (34,918.98 Da). (h) A255TMSK (34,961.06 Da). (i) I213TMSK (34,918.98 Da). (j) V297TMSK (34,933.00 Da).

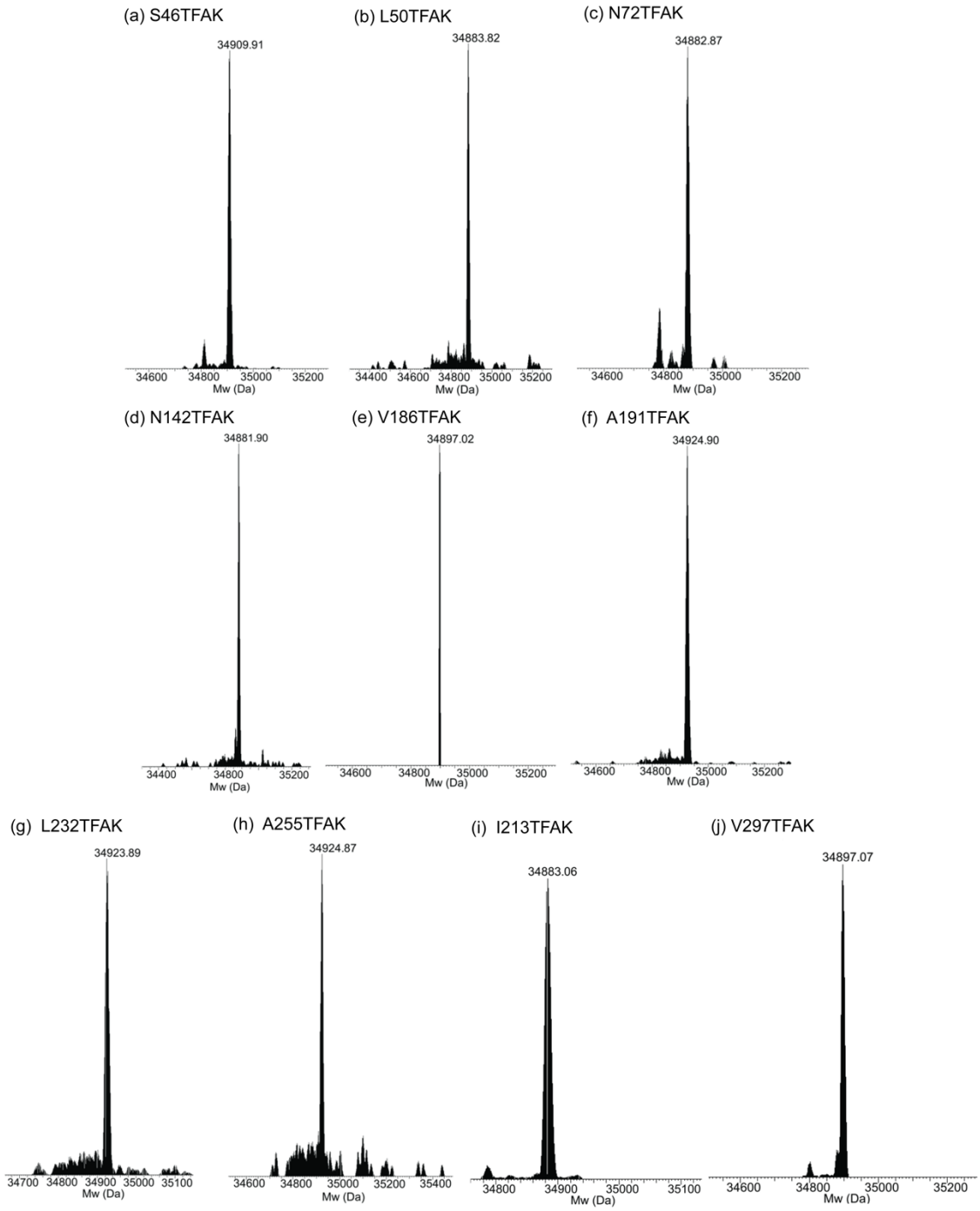

Figure S11. Intact protein mass spectrometric analysis of TFAK mutants. Calculated molecular weights in brackets. (a) S46TFAK (34,910.85 Da). (b) L50TFSK (34,884.77 Da). (c) N72TFAK (34,883.82 Da). (d) N142TFAK (34,883.82 Da). (e) V186TFAK (34,898.79 Da). (f) A191TFAK (34,926.85 Da). (g) L232TFAK (34,884.77 Da). (h) A255TFAK (34,926.85 Da). (i) I213TFAK (34,884.77 Da). (j) V297TFAK (34,898.79 Da)

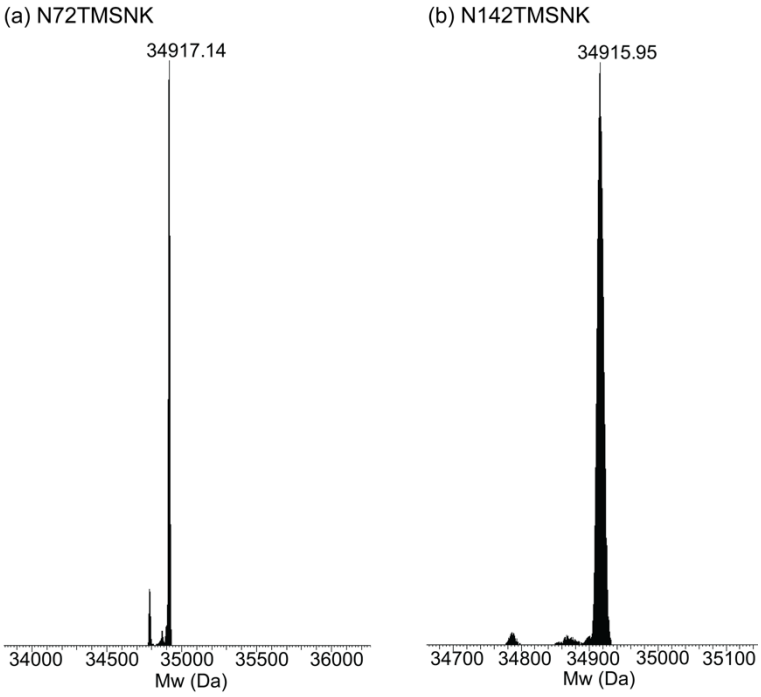

Figure S12. Intact protein mass spectrometric analysis of TMSNK mutants. Calculated molecular weights in brackets. (a) N72TMSNK (34,917.04 Da). (b) N142TMSNK (34,917.04 Da).

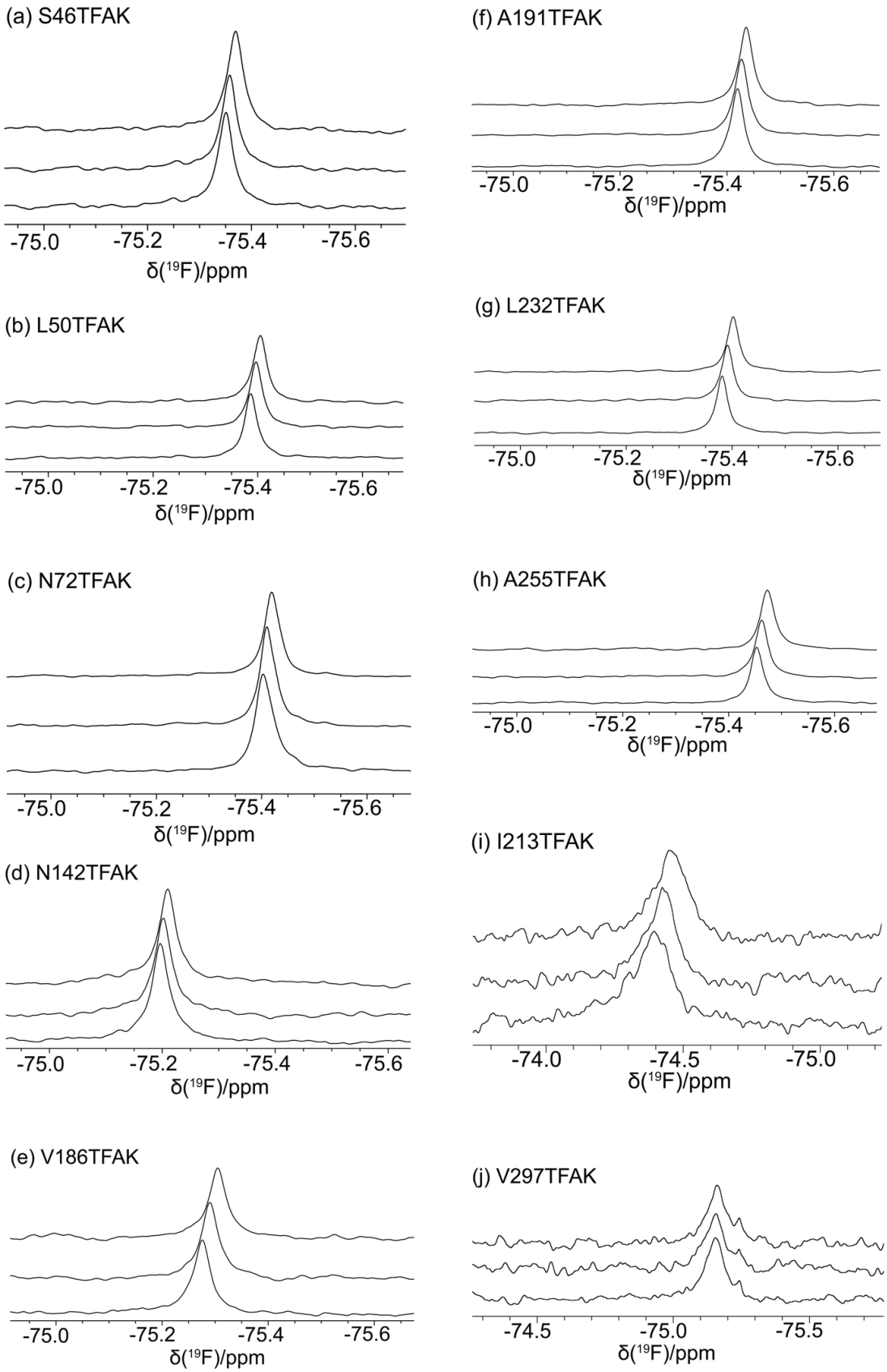

Figure S13. 1D ^{19}F-NMR spectra showing the TFAK resonance of different M^{pro} variants as a function of added DMSO. The protein concentration was 0.1 mM. The concentration of DMSO

was 0 (bottom spectrum), 2% v/v (middle spectrum), and 4% v/v (top spectrum). (a) – (e) Mutation sites close to the active site. (f) – (j) Mutation sites remote from the active site.

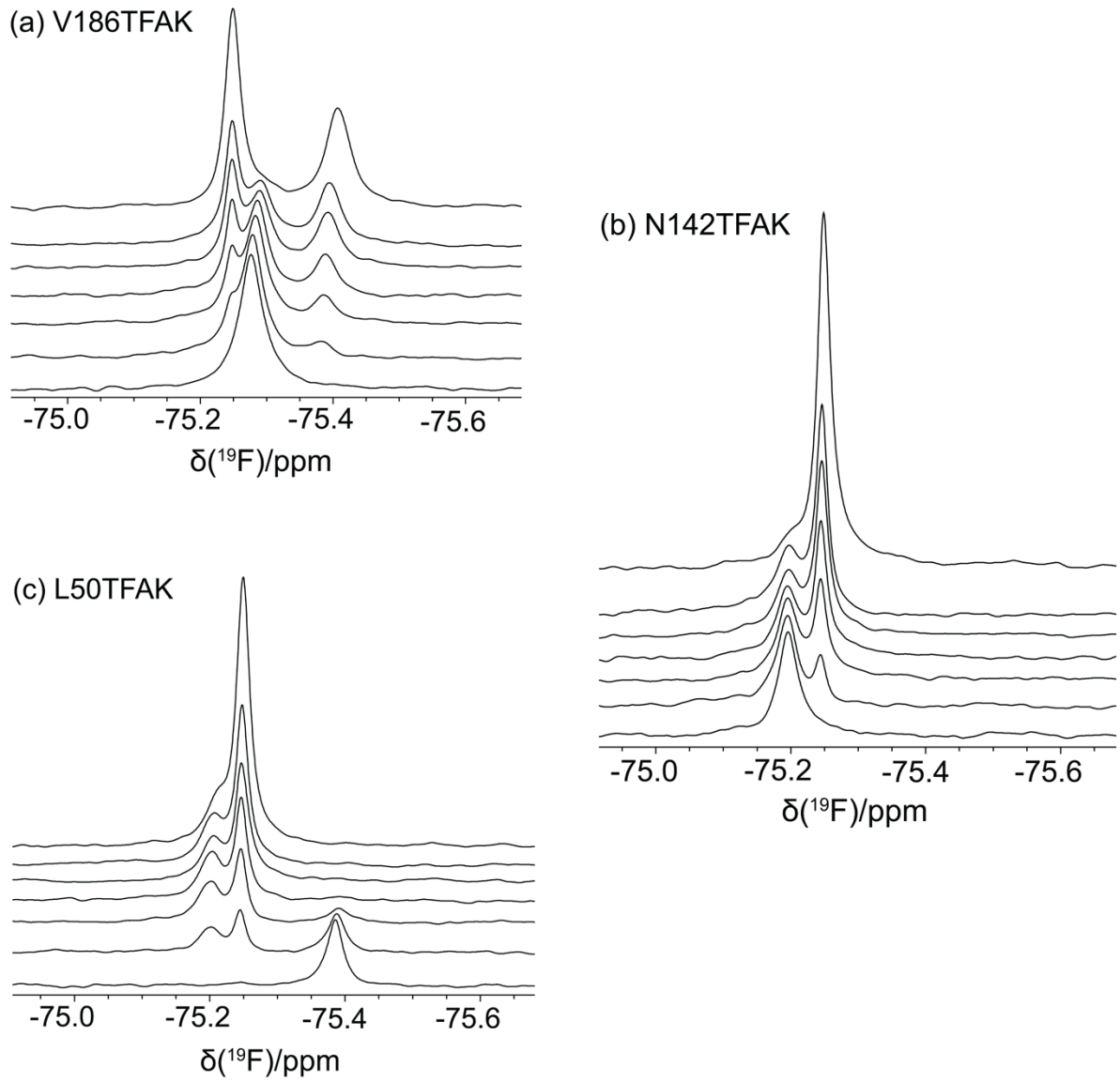

Figure S14. 1D ^{19}F-NMR spectra showing the TFAK resonance of three M^{pro} variants (0.1 mM solutions) as a function of added inhibitor **1**, which was purified using formic acid instead of TFA in the HPLC purification after peptide synthesis. The titration ratio of inhibitor to protein was, from bottom to top, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, and 1.0. The signal at -75.25 ppm is of spurious TFA in the inhibitor stock. (a) – (c) TFAK mutants of M^{pro} as indicated in the spectra. In (b), it is unclear whether the signal of the N142TFAK mutant in the complex with **1** is obscured by the signal of free TFA or broadened beyond detection, as N142 is centrally located in the inhibitor binding site.

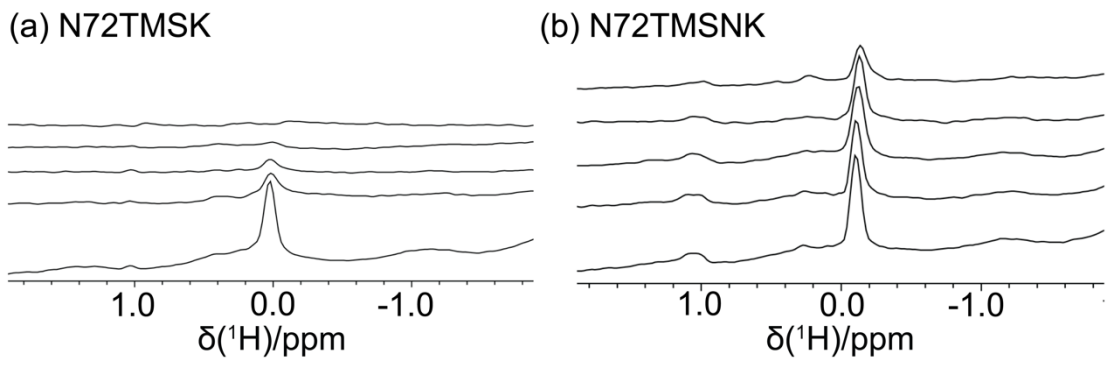

Figure S15. 1D ^{1}H NMR spectra showing the improved stability of M^{pro} samples prepared with TMSNK versus TMSK. The M^{pro} mutants N72TMSK and N72TMSNK were produced in parallel, and aliquots were frozen and stored at -20 ºC for 1, 2, 3, or 4 weeks, after which they were thawn and ^{1}H NMR spectra recorded of 30 μM solutions. (a) Spectra of N72TMSK. The bottom spectrum was recorded of the fresh sample without freezing. The spectra plotted above were recorded after 1, 2, 3, and 4 weeks of storage. (b) Same as (a), but for N72TMSNK.