Supporting Information for
Molecular Basis of Broad-Substrate Selectivity of a Peptide Prenyltransferase

Yue Hao[1,#], Elizabeth Pierce[2,#], Daniel R Roe[2], Maho Morita[2], John McIntosh[2], Vinayak Agarwal[1,3], Thomas E Cheatham, III[2], Eric W Schmidt[2,*] and Satish K Nair[1,3,4,*]

Supporting Materials and Methods

Substrates

Dimethylallylpyrophosphate was synthesized using an established procedure (1, 2). Cyclic peptide substrates were synthesized at the University of Utah DNA/peptide synthesis core facility. L-Tyr, N-boc-LTyr and Tyr-Tyr-Tyr peptide were purchased from Sigma Aldrich. N-acetyl-L-Tyr was purchased from Acros Organics. Linear peptides YLYEIAR and YLYQWLGAPV were synthesized from fmoc-protected amino acids using a standard protocol on an Aapptek Endeavor peptide synthesizer.

Protein expression and purification

The gene encoding PagF from *Planktothrix agardhii* was cloned into NdeI and EcoRI sites of pET28, in frame with the N-terminal his-tag. Amino acid substitutions in PagF were made by site-directed mutagenesis in the same vector and mutations were confirmed by sequencing. Plasmids encoding PagF variants were transformed into BL21(DE3) *E. coli*. For each expression, several colonies were grown separately overnight at 30 ^{o}C and used inoculate 1 L cultures (LB medium with 50 μg/mL kanamycin, in 2.8 L Fernbach flasks). These were grown at 37 ^{o}C until OD600 = 1.5, the temperature was lowered to 18 ^{o}C, 0.3 mM IPTG was added, and cultures were grown for 20 hours more. Cells were pelleted by centrifugation and stored at -80 ^{o}C until use.

For PagF purification, cells were thawed and suspended in 5-10 volumes of lysis buffer (0.5 M NaCl, 10 % (v/v) glycerol, 20 mM tris-Cl, pH 8.0). Cell suspensions were supplemented with 1 mM phenylmethanesulfonyl fluoride and 0.6 mg/mL lysozyme, stirred at 4 ^{o}C for 1 hour, sonicated on ice for two minutes, and centrifuged at 11,000 rpm in a JA-20 rotor for 1 hour at 4 ^{o}C. The remainder of the purification was done at room temperature. The supernatant was passed through a 0.45 μm filter and transferred to fresh tubes containing Ni-NTA resin (Qiagen) and incubated on an inversion mixer at for 30 minutes. Lysates and resin were applied to gravity columns, and the resin was washed with 20 volumes of wash buffer (0.5 M NaCl, 20 mM imidazole, 2 mM ATP, 20 mM tris-Cl pH 8.0). PagF was eluted with a linear gradient from 20 to 200 mM imidazole in (0.5 M NaCl, 20 mM tris-Cl, pH 8.0). Fractions from the gradient were analyzed by SDS-PAGE (Figure S3A), and those containing PagF were pooled and supplemented with ~ 60 μg/mL thrombin (Fisher) to remove the N-terminal 6x-his tag and dialyzed at 4 ^{o}C against two changes of buffer containing 0.1 M NaCl and 20 mM tris-Cl, pH 8.0. Thrombin-cleaved PagF was separated from uncleaved protein by passing it over fresh Ni-NTA resin in buffer containing 0.3 M NaCl and 20 mM tris-Cl, pH 8.0 (Figure S3B). This protein solution was supplemented with 10 % (v/v) glycerol, flash frozen and stored at -80 ^{o}C.

Conditions for optimal activity

We found that PagF requires a divalent cation for activity. The Mg^{2+} concentration needed for half-maximal activity, $K_{Mg}{}^{2+}$, is 45 ± 9 mM (Figure S4). Mn^{2+} (50 mM) can substitute for Mg^{2+} at the same concentration, giving nearly equal activity. No activity was detected in the absence of Mg^{2+} with 50 mM Ca^{2+}, Cu^{2+} or Zn^{2+}.

PagF is most active between pH 7 and 8.5 (Figure S5). Addition of the reductant tris-carboxyethyl-phosphine had no apparent effect on activity. We observed a slight increase in activity (from ~ 4 min^{-1} to around 7 min^{-1}) as the NaCl concentration was increased from 5 mM to 1.7 M, and KCl could be substituted for NaCl with no decrease in activity.

Enzyme assays

Unless otherwise indicated, enzyme reactions for analysis of different tyrosine substrates contained NaCl (1.1 M), bis-tris-propane, pH 7.5 (100 mM), $MgCl_2$ (200 mM), and 4 % (v/v) DMSO, to solubilize some substrates. The Km for DMAPP was determined by varying the DMAPP concentration in assays with 20 mM *N*-boc-L-Tyr (Figure S6). Kinetic parameters for various Tyr substrates were determined from reactions containing 50 mM DMAPP. Reactions were pre-heated to 37 ^{o}C and were initiated by PagF addition. Aliquots for HPLC and LC/MS analysis (20 μL) were quenched in 180 μL methanol. The assay with *N*-boc-L-Tyr was scaled up and sufficient material for analysis of the product by NMR (Figure S7) was purified by HPLC.

Assays were analyzed by HPLC as follows: 90 μL quenched sample was injected onto a C4 column (Phenomenex Jupiter C4, 5 μ, 4.6 x 150 mM) or on a C18 column (Agilent Eclipse XDB-C18, 5 μm, 4.6 x 150 mm) and run on a linear gradient from 1 to 100 percent acetonitrile over 24 minutes in water/0.05 % (v/v) TFA at 0.7 mL per minute. Substrate and product peaks were integrated and product formation was calculated as a fraction of the total substrate and product peak areas, in order to smooth variations caused by the large assay quench volume and variations in HPLC injection volume. This treatment assumes that the extinction coefficients of substrate and product are essentially the same and that the total peak area is constant over the course of the assay (Figure S8). Some assays were additionally analyzed by HPLC-ESI-MS or ESI-MS/MS on a Waters Micromass Q-Tof mass spectrometer. Samples were directly injected onto the mass spectrometer or were run on the Agilent Eclipse column with a linear gradient from 99% A (H_2O with 0.1 % formic acid) and 1 % B (acetonitrile) to 100 % B over 24 minutes at 0.7 mL per minute.

PagF prenylates L-tyrosine in a range of contexts, including the free amino acid and *N*-modified tyrosines, a Tyr-Tyr-Tyr tripeptide, and one of its native substrates, cyclic[INPYLYP], which is the precursor of prenylagaramide B (3). Prenylation of cyclic[INPYLYP] is specific for one tyrosine (the sequence of the natural product is INP-pTyr-LYP), but after extended incubation under forcing conditions the second tyrosine is partly prenylated (Figure 5). In contrast we observed efficient addition of a second prenyl group and extremely slow addition of a third prenyl group to the Tyr-Tyr-Tyr peptide (Figure S9, Tables S1-S2). MS/MS analysis of products of PagF assays with Tyr-Tyr-Tyr, YLYEIAR and YLYQWLGAPV peptides supports the hypothesis that in each linear peptide the N-terminal Tyr can be prenylated first, since in all three fragmentations of singly prenylated products, singly prenylated a1 and a2 ions are observed (Figure S9, Table S2). Our MS/MS data does not rule out the possibility that the singly prenylated ions observed are actually mixtures of peptides prenylated at different positions.

Kinetic data analysis

Assays for analysis of progress curves used 50 mM DMAPP (roughly 7-fold more than the Km) and various substrate concentrations. Progress curves from individual reactions were fit using the Runge-Kutta algorithm in Berkeley-Madonna 8.3 to the Michaelis-Menten equation [d/dt (P) = V_{max}*(S0-P)/(K_m+(S0-P))] where P is product concentration and S0 is the initial substrate concentration (Figure S10). In most of the fits, S0 was allowed to vary because while many reactions appeared essentially irreversible (final substrate amounts were below the detection limit in the assay), in other assays, the reaction stopped before all substrate was consumed. Inorganic pyrophosphatase (2-5 units per unit of PagF) was added to some assays to test whether inhibition by the product pyrophosphate prevented assays from going to completion.

In no case did this addition make a difference to the rate of the reaction or to the final substrate and product concentrations. Attempts to fit these assays with a reversible kinetic model [d/dt (P) = V_{max}*(S0-P)/(K_m+(S0-P))-V_{max}r*(S0-P)/(K_mr+(S0-P))] were unsuccessful.

Kinetic parameters with N-acetyl-L-tyrosine were additionally determined from a plot of initial rates vs. substrate concentration in a series of reactions (Figure S11). K_m and V_{max} determined from these assays were on the same order of magnitude as expected from fits to single assays.

Molecular dynamics simulations

All simulations were performed with a development version of pmemd.cuda from Amber 14(4, 5) with support for center-of-mass distance restraints. All molecular dynamics (MD) simulations were run in the canonical (NVT) ensemble at 300 K using a Berendsen(6) thermostat with coupling time of 5.0 ps. Long range electrostatic interactions were handled using the particle mesh Ewald(7) scheme with a cutoff of 8.0 Å and default parameters. Bonds to hydrogen were constrained using SHAKE,(8) and a MD time step of 2 fs was used.

All analysis was performed with CPPTRAJ(9) from AmberTools 14. The Amber 14SB and GAFF(10) force fields were used. Parameters for dimethylallyl pyrophosphate (DMP) were generated using the antechamber program from AmberTools 14. Charges for DMP were obtained by RESP-fitting the electrostatic potential obtained from Gaussian 09(11) using Hartree-Fock geometry optimization at the 6-31G* level.

MD simulations were conducted in which either Tyr4 or Tyr6 of cyc[INPYLYP] (referred to hereafter as npPB) were pulled into the active site of PagF via harmonic distance restraints. Two harmonic distance restraints were used: one between all alpha carbons of PagF and all alpha carbons of npPB centered at a target distance of 7.2 Å (referred to as Pull1), and the other between the heavy atoms of PagF residues 222, 235, 271, and 290 of PagF and all heavy side chain atoms of either Tyr4 or Tyr6 of npPB centered at a target distance of 3.0 Å (referred to as Pull2-Y4 and Pull2-Y6 respectively) in order to properly orient the target tyrosine with the active site. Both restraints were harmonic for 1.0 Å in each direction and linear thereafter and used a force constant of 5.0 kcal/mol*Å^2. The coordinates for PagF and DMP were taken from the crystal structure of PagF + dimethylallyl S-thiolodiphosphate + L-boc-Tyr. The initial structure for the Tyr4-Pull simulation was generated by manually positioning npPB outside PagF using CPPTRAJ in the following manner. First, the center of mass of PagF (residues 1 to 297) was shifted to the coordinate origin. Second, the entire system was aligned along the principal axes of PagF – this had the effect of aligning the beta barrel along the first (X) principal axis. The npPB was then translated 15 Å along the X axis so it cleared the beta barrel. The system was then solvated with 9870 TIP3P(12) waters, neutralized with 5 sodium cations using Joung-Cheatham ion parameters,(13) and minimized and equilibrated using a previously described protocol.(14) The initial distances for Tyr4-Pull were 22.5 Å for Pull1 and 16.6 Å for Pull2-Y4. The initial structure for the Tyr6-Pull simulation was created in a similar fashion, but had to be by 20 Å along the X axis to ensure no clashes with PagF when rotating Tyr6 towards the active site. This was followed by a short 100 ps MD simulation in which the Pull1 distance was set to 24 Å and Pull2-Y6 was set to 16.6 Å in order to orient the Tyr6 closer to the active site than Tyr4; the initial distances for Tyr6-Pull were 24.0 Å for Pull1 and 17.0 Å for Pull2-Y6. MD simulations for each pull type were repeated 10 times with different initial velocity distributions (for a total of 20 simulations), and each simulation was 1 ns in length (see Figure S14 for more detailed results from pulling simulations).

Umbrella sampling (US) simulations were also conducted in which either Tyr4 or Tyr6 of npPB was guided into initial contact with PagF. Two distance restraints slightly different from the ones used in the pulling simulations were used: one between all heavy atoms of residues 51, 67, 69, 138, 170, 222, 271, and 292 of PagF (collectively referred to hereafter as the 'Cage' because they surround tyrosine in the active site when bound) and the heavy side chain atoms of Tyr4 or Tyr6 from npPB, and the other between the Cage and all backbone (BB) atoms of npPB. The new restraints provided more stable trajectories in the US simulations than the pulling restraints. The US simulations for Y4 and Y6 started from the same respective starting structures as the pulling simulations, and each window was conducted in serial (i.e. the final coordinates of a window were initial coordinates for the next window). Harmonic distance restraints centered on each windows target restraint values with a width of 1.0 Å in each direction and a force constant of 5.0 kcal/mol*Å^2 were used. The Y4-US simulation windows started from Cage-Y4 distance of 14.0 Å and Cage-BB distance of 19.0 Å down to 5.5 Å and 11.75 Å respectively for a total of 18 windows. The Y6-US simulation windows started from Cage-Y6 distance of 14.5 Å and Cage-BB distance of 18.8 Å down to 5.5 and 11.75 Å respectively for a total of 19 simulation windows. Each window was simulated for 1 ns using the same conditions as in the pulling simulations. Restraint values were written every 10 steps. Potentials of mean force using both restraint distances as reaction coordinates were calculated using the weighted histogram analysis method (15) as implemented by A. Grossfield.(16) Figure S13 shows PMFs calculated from all data, Figure S15 shows PMFs calculated from 1-500 ps and 1-1000 ps.

Crystallization, phasing and structure determination

Crystals were PagF and all relevant complexes were grown using the hanging drop vapor diffusion method. Briefly, a 1 μL drop of protein at a concentration of 20-25 mg/mL (in a buffer of 20 mM Tris-HCl pH=7.5, 150 mM KCl) was mixed with an equivalent amount of precipitant (20 mM Tris-HCl pH=7.5, 200 mM NaCl, 20-25% PEG 3350). Ligand complexes were obtained by soaking crystals with the appropriate ligand (4 mM DMSPP, 10 mM $MgCl_2$, and either 10 mM boc-L-Tyr, 20 mM Tyr-Tyr-Tyr or 100 μM cyclic[INPYLYP].Crystals were immersed into a crystallization media supplemented to a final PEG 3350 concentration of 40% before vitrification by direct immersion into liquid nitrogen.

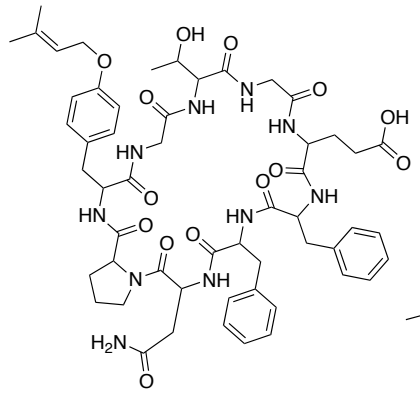
Crystallographic phases were determined by single wavelength anomalous diffraction data collected on crystals soaked with trimethyl lead acetate. A 12-fold redundant data set was collected near the lead absorption edge, to a limiting resolution of 2.65 Å (overall R_{merge} = 0.090, I/σ (I) = 10 in the highest resolution shell) utilizing a Mar 300 CCD detector (LS-CAT, Sector 21 ID-D, Advanced Photon Source, Argonne, IL, USA). Data were indexed and scaled using the HKL2000 package (17). The initial heavy atom substrate was determined using HySS (18), and heavy atom parameters were refined using Phaser as implemented in the Phenix suite of programs (19) to yield a figure of merit of 0.462. The resultant map permitted most of the main-chain and more than half of the side-chain residues to be automatically built using ARP/wARP. This initial model was refined against higher resolution data using REFMAC5 (20), and further improved through iterative rounds of manual rebuilding using Coot (21). Cross validation, using 5% of the data for the calculation of the free R factor, was utilized throughout the model-building process in order to monitor building bias (22). X-ray diffraction data from crystals of the ligand complexes were collected and processed in a similar manner. These structures were determined by the molecular replacement method using the refined coordinates of the unliganded PagF as a search model (23). The resultant solutions were subject to initial rounds of automated

model building using ARP/wARP, followed by subsequent rounds of manual rebuilding using Coot, interspersed with crystallographic refinement using REFMAC5.

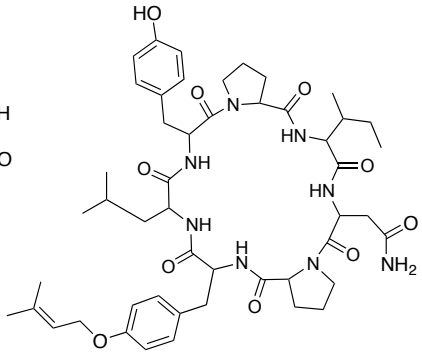
Sequence similarity network analysis

The PagF protein sequence was used to query the EFI-EST server (24) as the query sequence to achieve the all-by-all blast result of sequences sharing a similarity with an E-value threshold of 10^{-5}. The result were displayed in Cytoscape (25) with a sequence-length limit (between 280aa and 330aa in length) set to avoid fragmented sequences or sequences from multi-domain proteins. 26 sequences were displayed and E-value cutoff was set to 10^{-80} for further analysis.

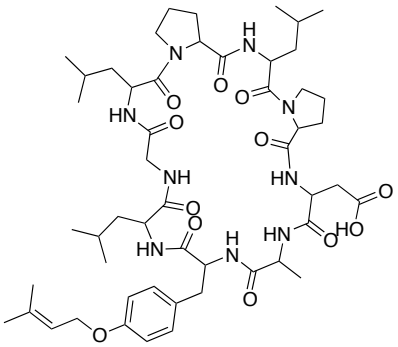
Figure S1. Prenylated cyanobactins from *tru*, *pag*, and *lyn* pathways.



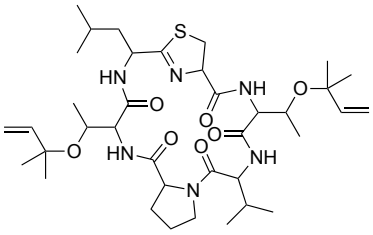
prenylagaramide A



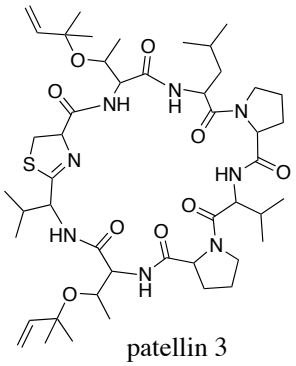
prenylagaramide B



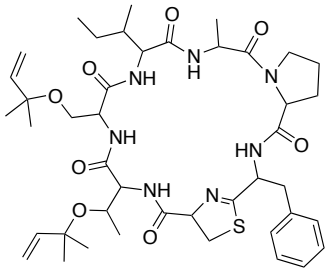
prenylagaramide C



patellin 2



patellin 3



trunkamide

Figure S2. Comparison of the hydrophobic active site barrel of related small molecule prenyltransferases such as (A) FgPT2) from *Apergillus fumigatus* (PDB 3I4Z), (B) NphB (PDB 1ZCW), and (C) CloQ (PDB 2XLQ) reveals that the base of the barrel is enclosed either with long loops or with secondary structural elements. By comparison, the active site of PagF is open and is encapsulated only upon binding of substrate peptide.

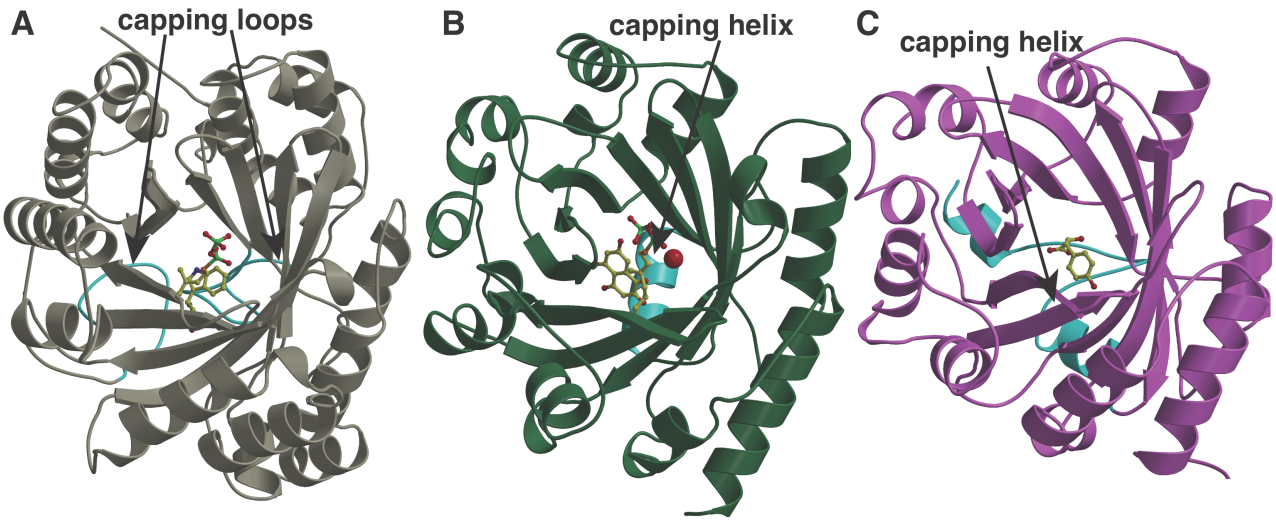


Figure S3. PagF purification by Ni-NTA (A) before and (B) after thrombin cleavage. (A) I: insoluble proteins from cell lysis; S: soluble proteins from cell lysis; M: broad range protein marker (NEB P7702); FT: Ni-NTA flowthrough; W: 20 mM imidazole wash; E1-E14: elution fractions from imidazole gradient. (B) T: thrombin; M: protein marker; FT-P: precipitated proteins from Ni-NTA flowthrough; FT-S: soluble protein from Ni-NTA flowthrough; W: no imidazole wash; E: 250 mM imidazole elution.

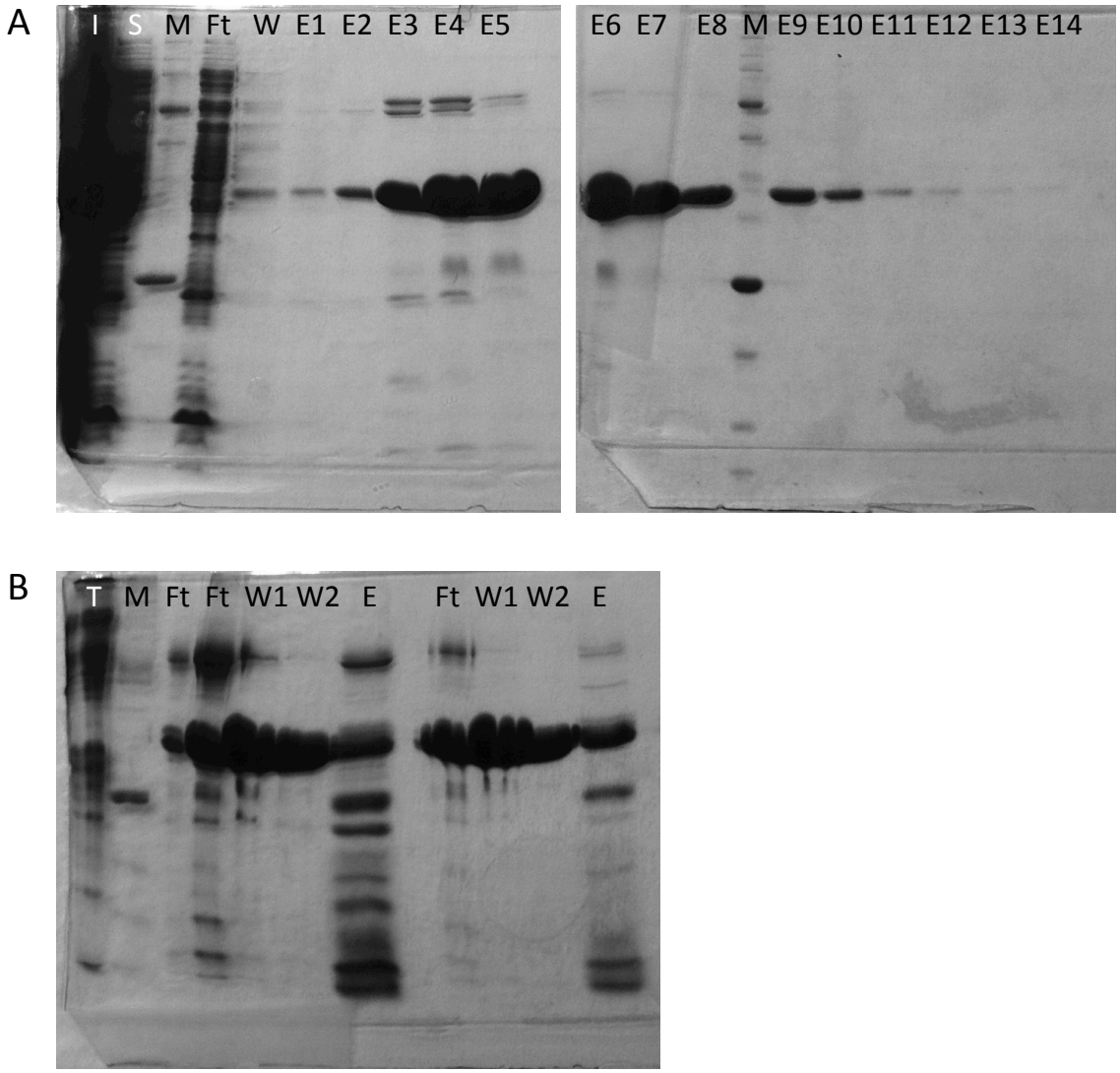


Figure S4. Dependence of activity on $MgCl_2$ concentration. Reactions contained bis-tris-propane, pH 7.5 (100 mM), DMAPP (10 mM), boc-L-Tyr (10 mM) and varying concentrations of NaCl to make an ionic strength of 1.7. Circles represent rates of individual reactions. The line is the best fit to a hyperbola [$k_{obs} = k^{Mg}_{max}/(1+K_{Mg}/[Mg^{2+}])$]. Parameters from that fit are k^{Mg}_{max} = 11.3 ± 0.7 min^{-1} and K_{Mg} = 44.6 ± 8.5 mM.

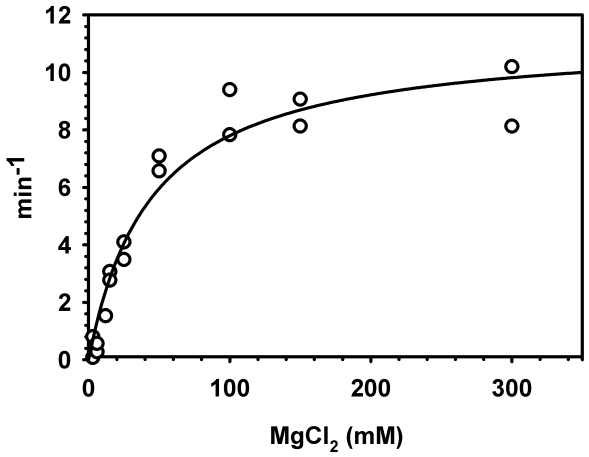


Figure S5. Dependence of PagF activity on pH. Reactions contained bis-tris-propane (40 mM), DMAPP (10 mM), boc-L-Tyr (20 mM), $MgCl_2$ (200 mM) and NaCl (1095 mM). Open circles represent average rates from three assays.

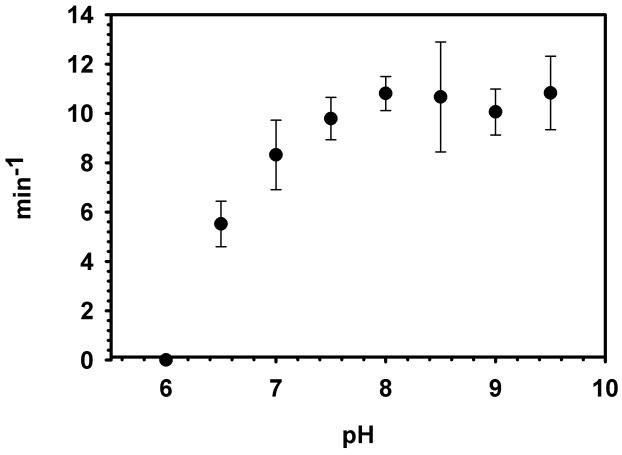


Figure S6. Dependence of PagF activity on DMAPP concentration. Reactions contained NaCl (950 mM), glycylglycine pH 9.0 (40 mM), $MgCl_2$ (12 mM), tris(2-carboxyethyl)phosphine (3 mM) and boc-L-Tyr (3 mM). Reactions were incubated at 37 ^{o}C and were initiated by PagF addition. Samples were taken at various time points and analyzed for O-prenyl-boc-L-Tyr formation by HPLC. Circles represent averages of three reactions and error bars are standard deviations. The line is the best fit to the Michaelis-Menten equation. Parameters from that fit are V_{max} = 12.6 ± 0.6 min^{-1} and K_m = 0.52 ± 0.10 mM.

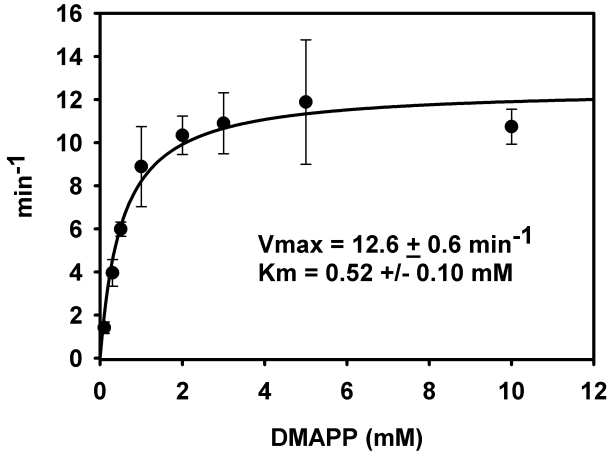
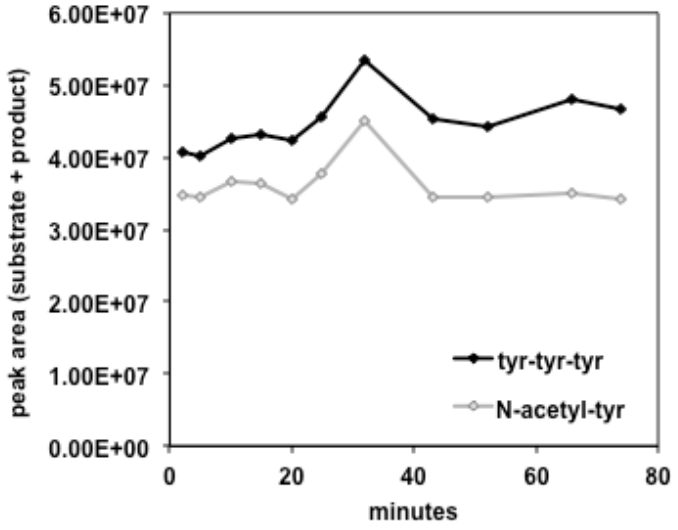


Figure S7. Sums of substrate and product HPLC peak areas. Data in progress curve reactions were analyzied by dividing HPLC product peak area by total peak area (substrate + product), in order to smooth variations caused by the large assay quench volume and variations in HPLC injection volume. This treatment assumes that the extinction coefficients of substrate and product are essentially the same.



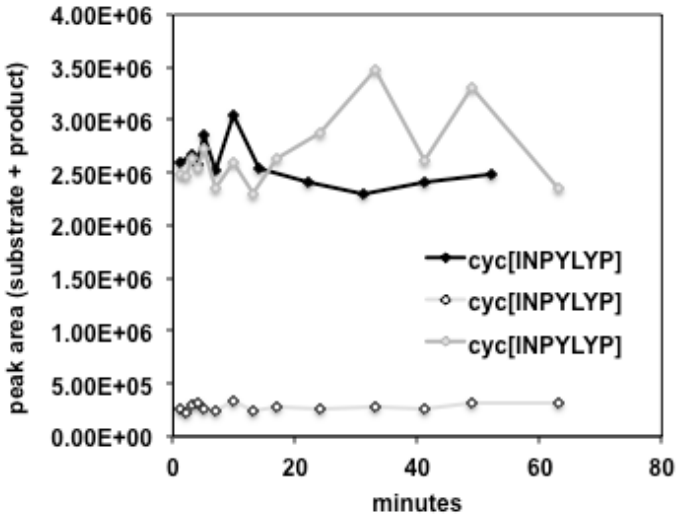


Figure S8. NMR of PagF-treated N-boc-L-tyrosine. PagF was assayed with N-boc-L-Tyr and DMAPP. The tyrosine product was HPLC purified and analyzed in $CDCl_3$ on a 500 MHz Varian NMR.

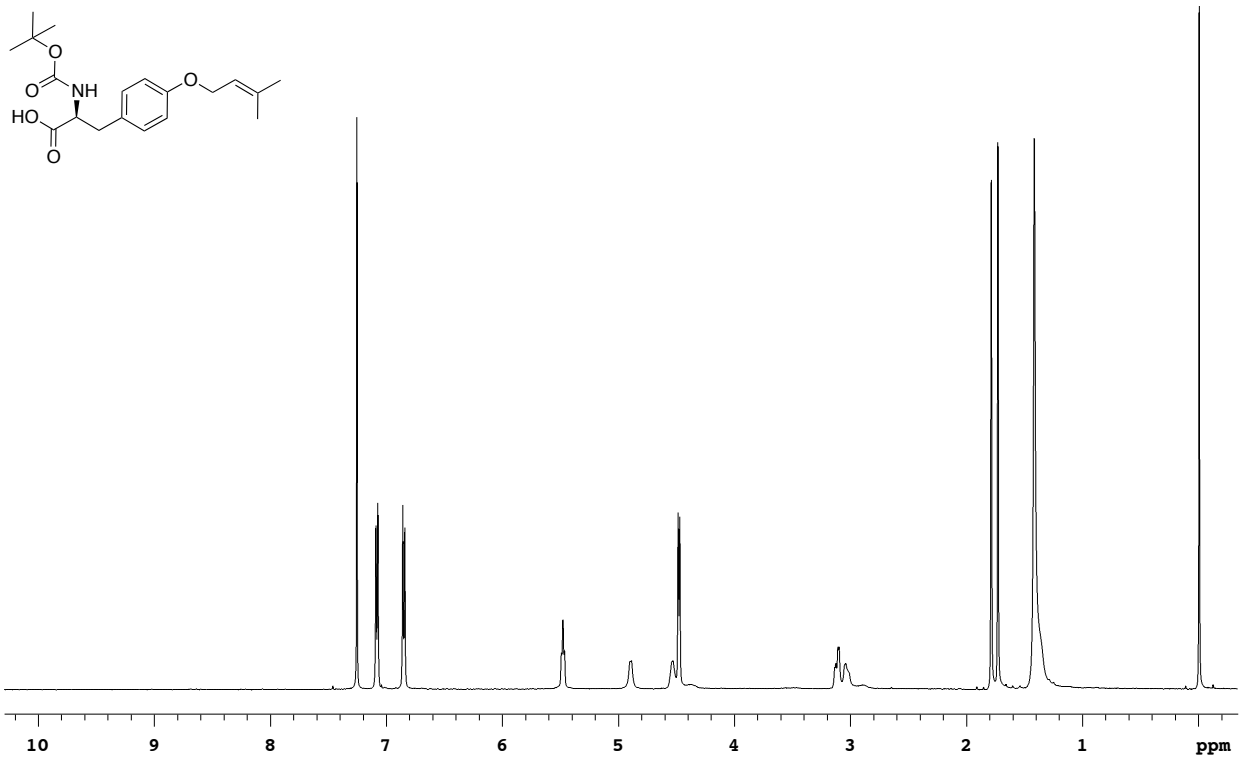
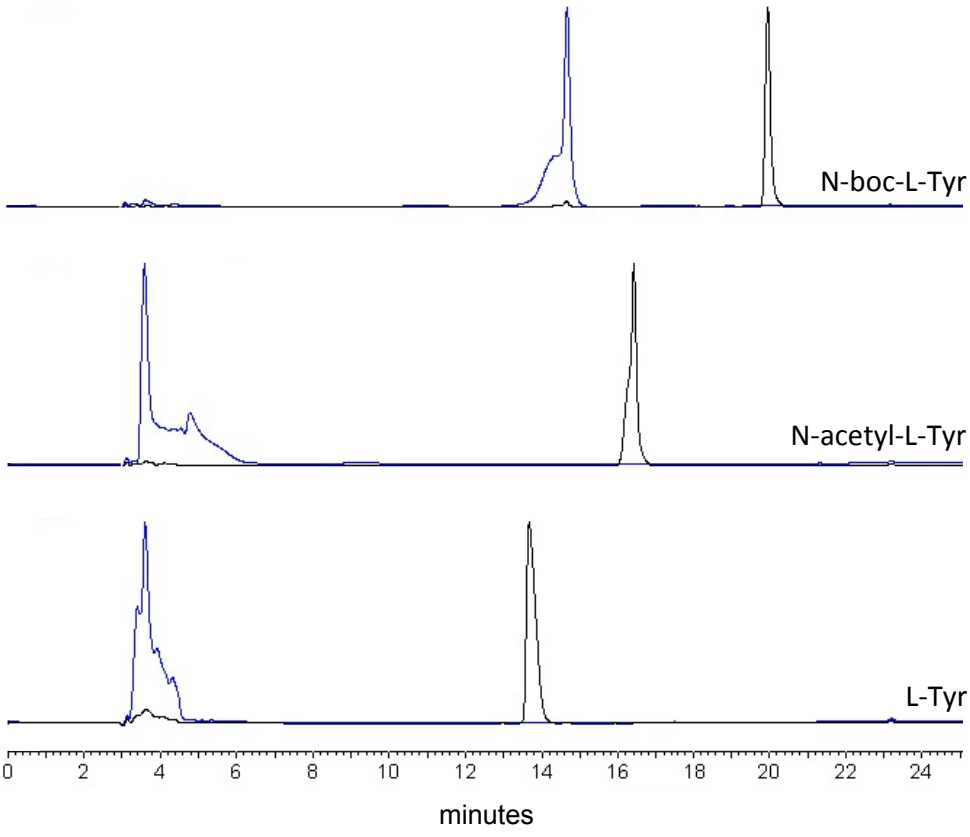
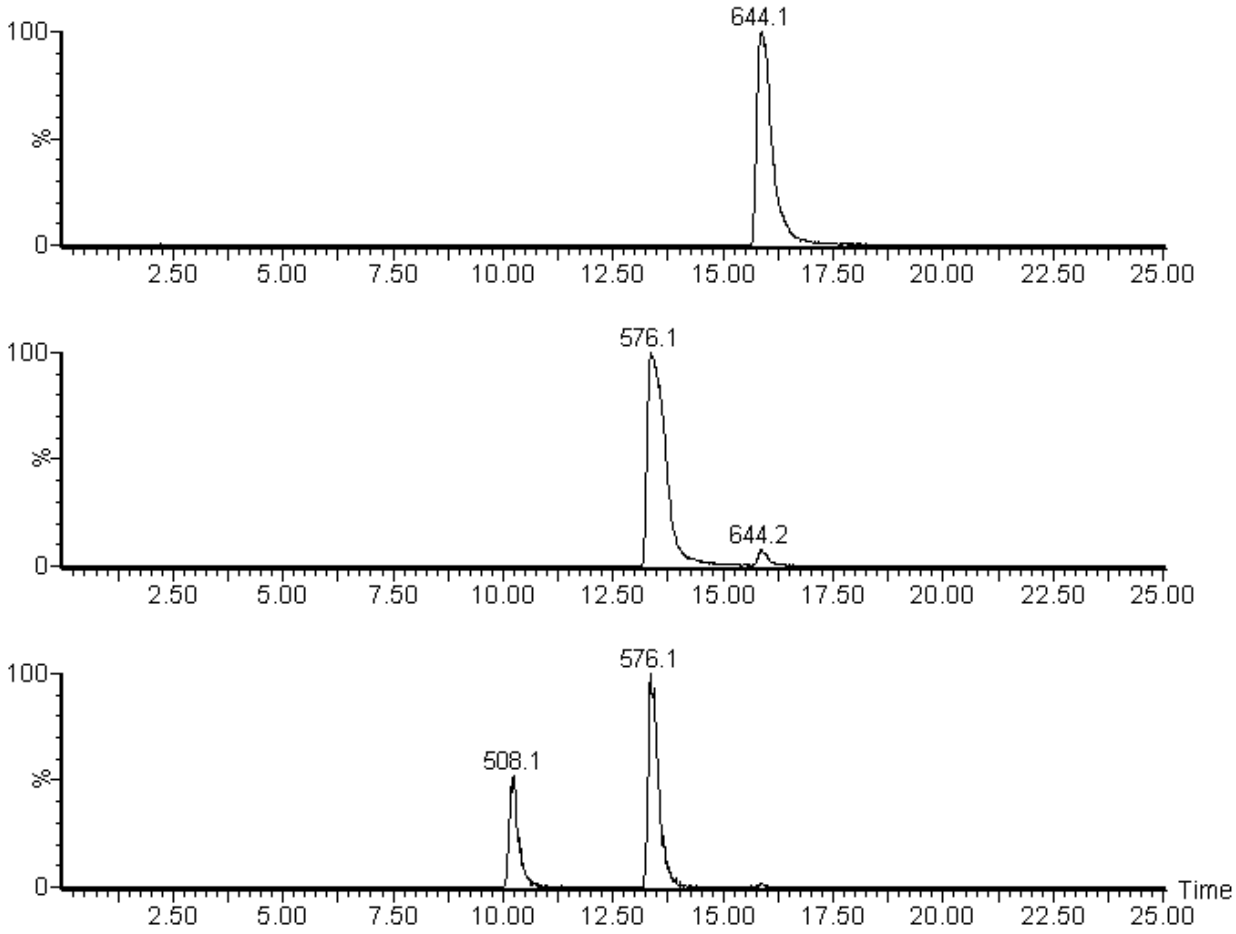


Figure S9. HPLC, LC/MS and MS/MS characterization of PagF substrates and products including (A) L-tyrosine, N-acetyl-L-tyrosine and N-boc-L-Tyr (B-D) Tyr-Tyr-Tyr, (E-F) cyclic[INPYLYP], (G-I) YLYEIAR, and (J-L) YLYQWLGAPV and their prenylated forms. HPLC data were collected on a Hitachi Primaide HPLC. Compounds were separated on a Phenomenex Jupiter C4 column (5 μ, 4.6 x 150 mM) with a gradient from 1 to 100 percent acetonitrile over 24 minutes in water/0.05 % (v/v) TFA. LC/MS data were collected on a Waters Micromass Q-Tof ESI spectrometer connected to an Alliance HP separations module. Compounds were separated on an Agilent Eclipse C18 column (5 μ, 4.6 x 150 mm) with a gradient from 1 to 100 percent acetonitrile over 24 minutes in water/0.1 % (v/v) formic acid. Each y-axis is normalized to the maximum in that particular trace; different y-axes within each figure are not scaled to match each other.

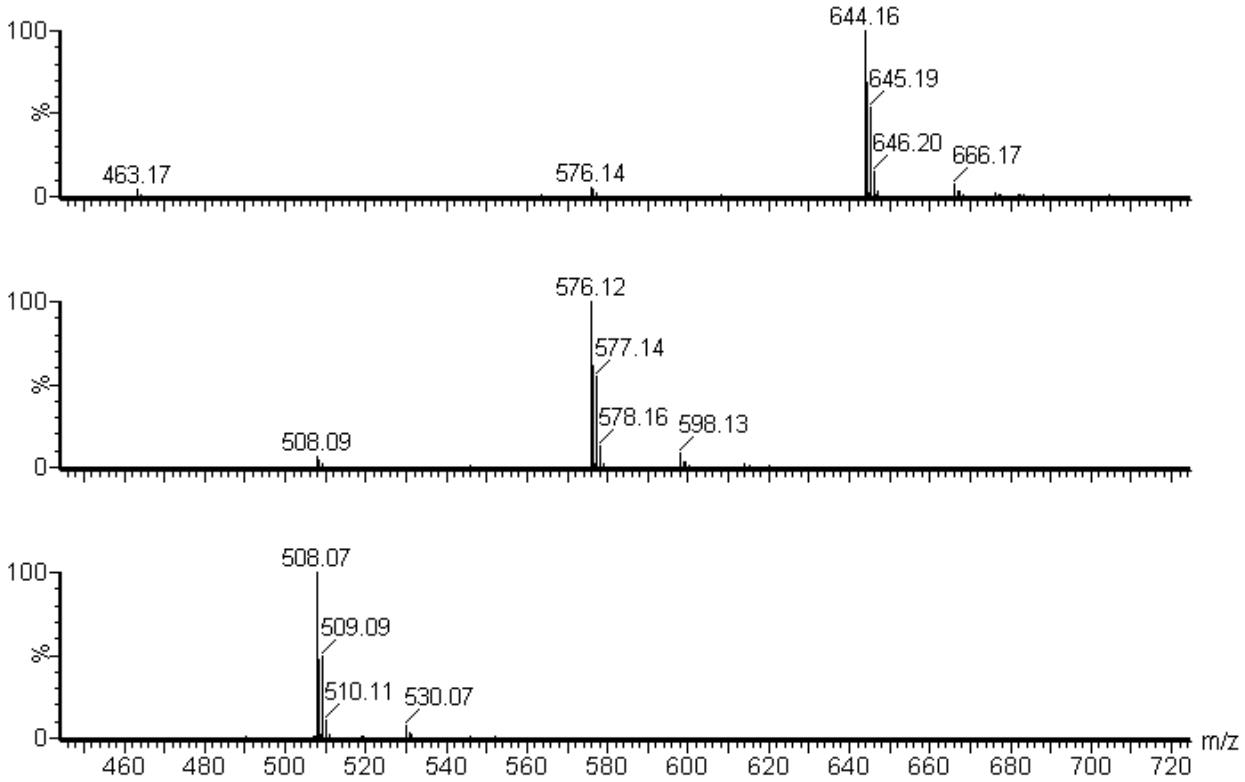
(A) HPLC separation of L-Tyr, N-acetyl-L-Tyr and N-boc-L-Tyr from prenylated products. Substrate traces are shown in blue, prenylated product traces are black.



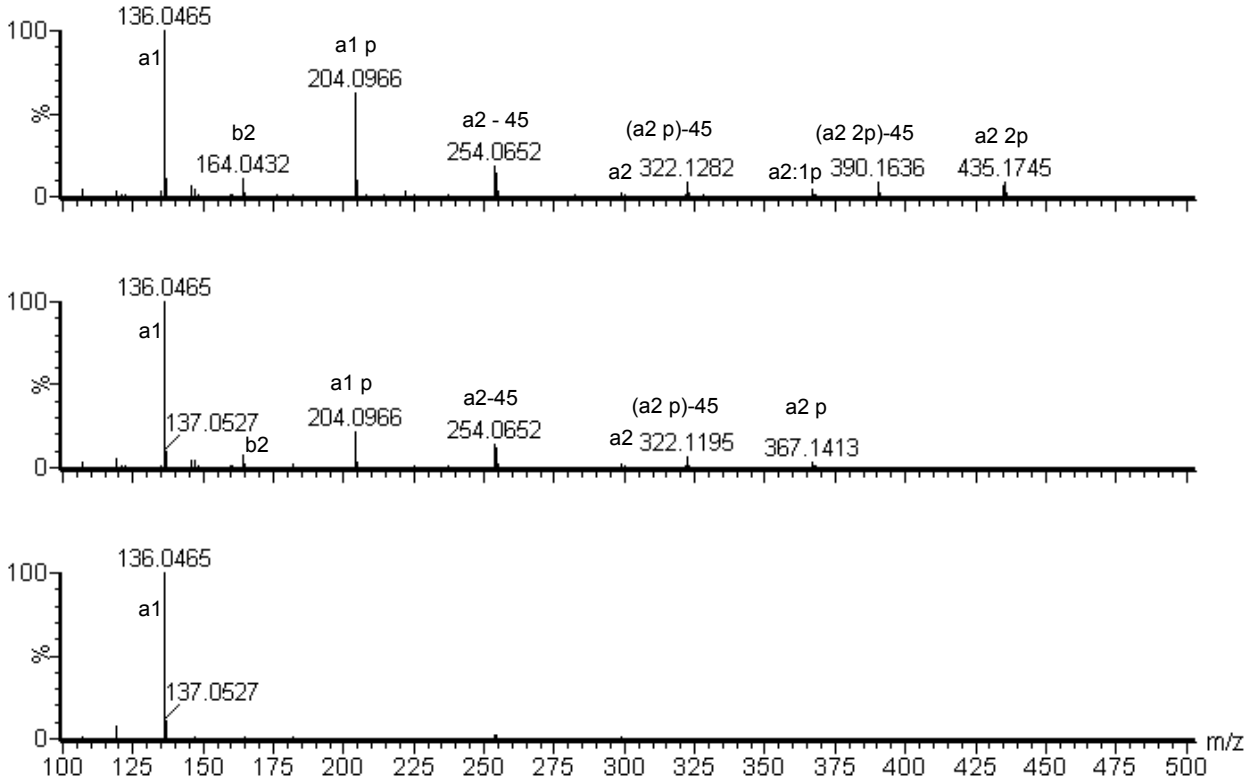
(B) LC/MS chromatograms of unprenylated Tyr-Tyr-Tyr (508.2 Da), singly prenylated Tyr-Tyr-Tyr (576.3 Da) and doubly prenylated Tyr-Tyr-Tyr (644.3 Da). The apparent elution of 508.2 Da peaks at 13.4 and 15.9 minutes and the 576.3 Da peak at 15.9 minutes results from fragmentation of prenylated peptides in the mass spec.



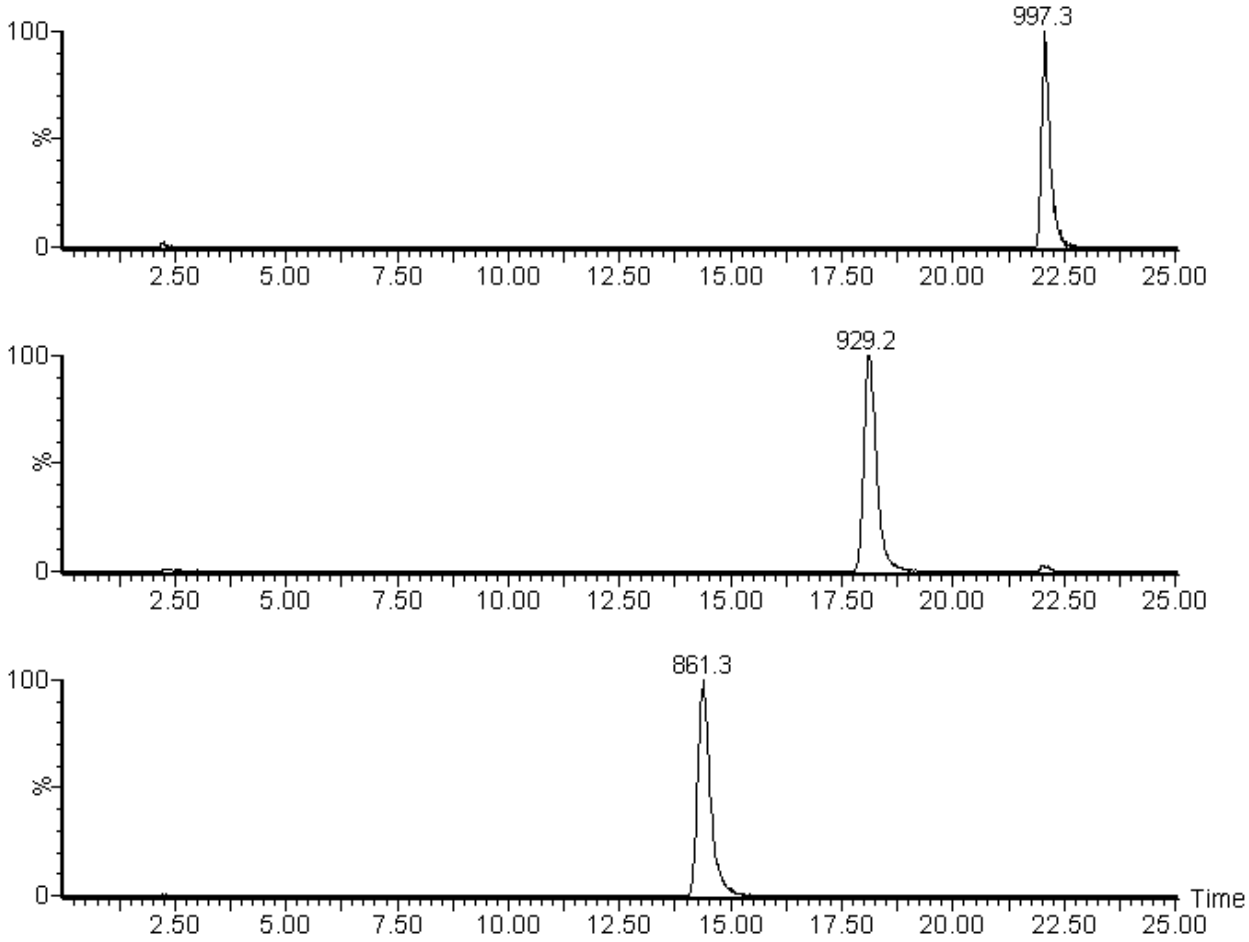
(C) mass spectra of compounds eluting at 10.2 (bottom), 13.4 (middle) and 15.9 minutes shown in figure S9B.



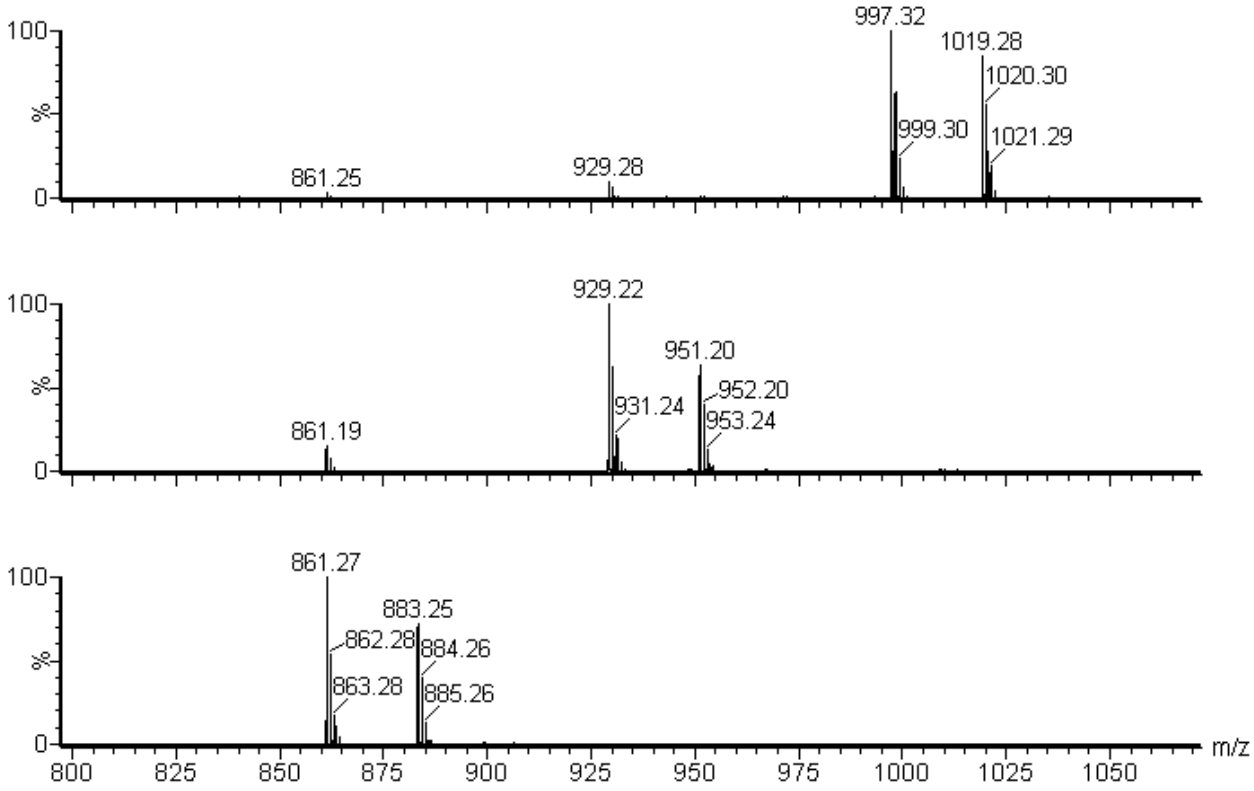
(D) top: fragmentation of doubly prenylated Tyr-Tyr-Tyr (*m/z* = 644.3), middle: fragmentation of singly prenylated tyr-tyr-tyr (*m/z* = 576.3), bottom: fragmentation of unprenylated Tyr-Tyr-Tyr (*m/z* = 508.2).



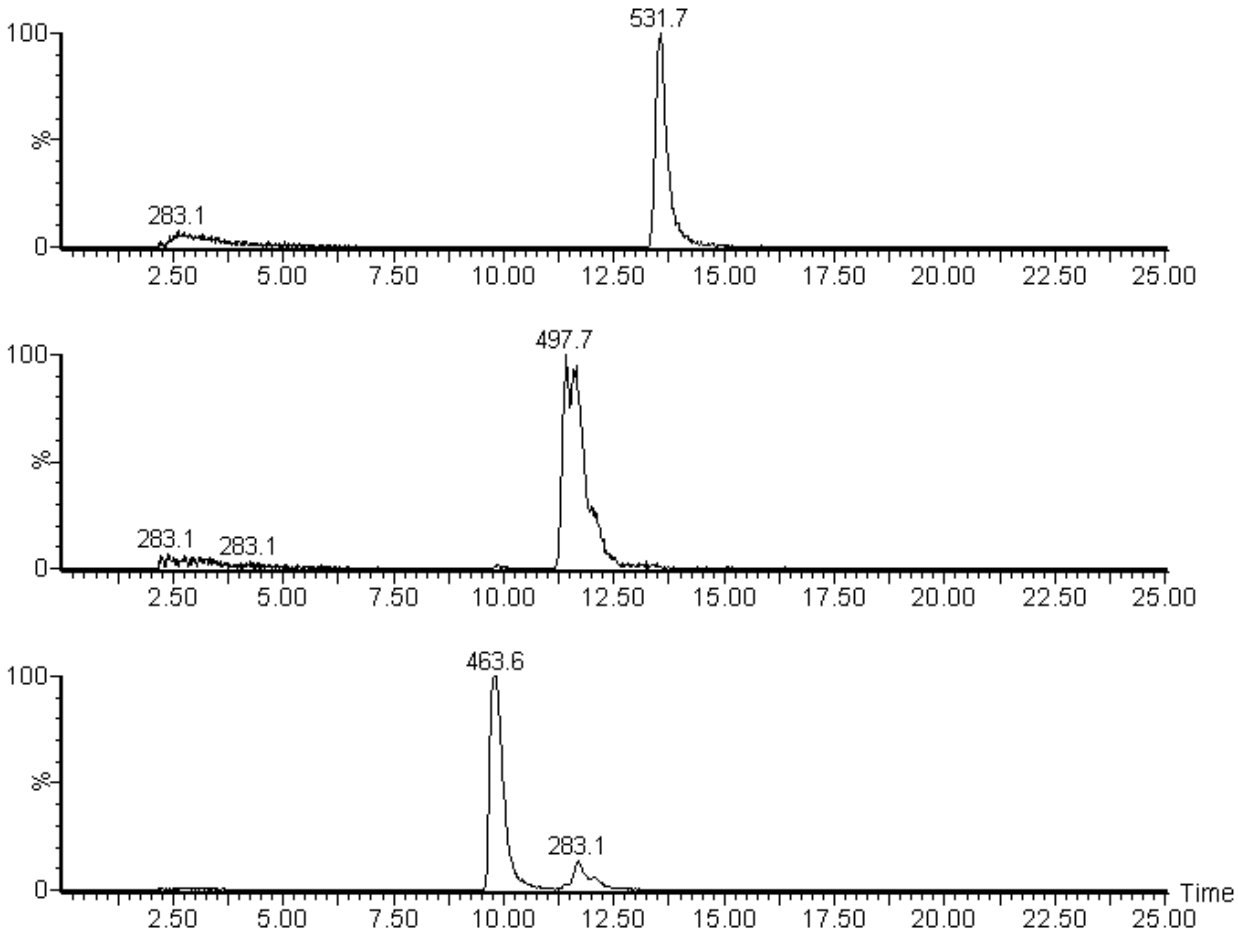
(E) LC/MS chromatograms of unprenylated cyclic[INPYLYP] (861.5 Da), singly prenylated cyc[INPYLYP] (929.5 Da) and doubly prenylated cyc[INPYLYP] (997.6 Da).



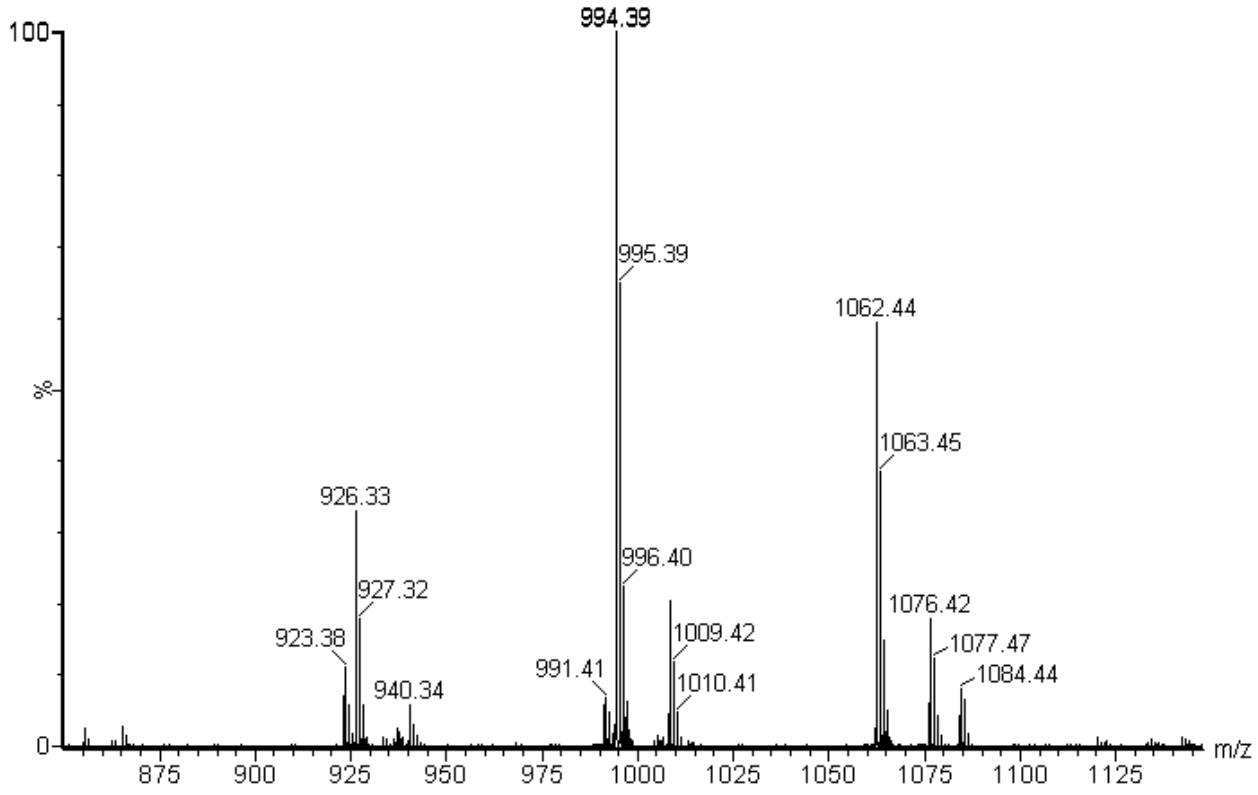
(F) mass spectra of compounds eluting at 14.4 minutes (unprenylated), 18.1 minutes (singly prenylated) and 29.3 minutes (doubly prenylated) shown in figure S5E.

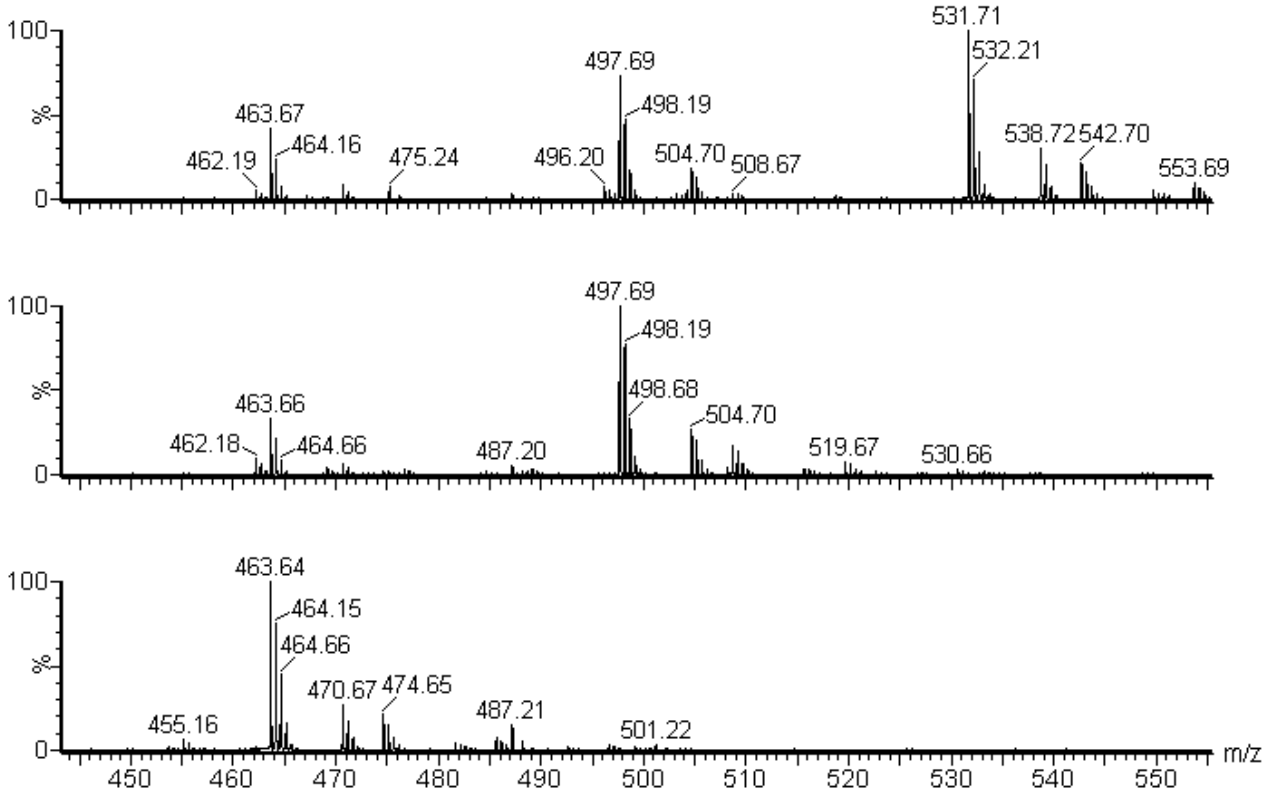


(G) LC/MS chromatograms of unprenylated YLYEIAR ($[M+H]^+$ = 926.5, $[M+2H]^{2+}$ = 463.6), singly prenylated YLYEIAR ($[M+H]^+$ = 994.5, $[M+2H]^{2+}$ = 497.7) and doubly prenylated YLYEIAR ($[M+H]^+$ = 1062.6, $[M+2H]^{2+}$ = 531.7).

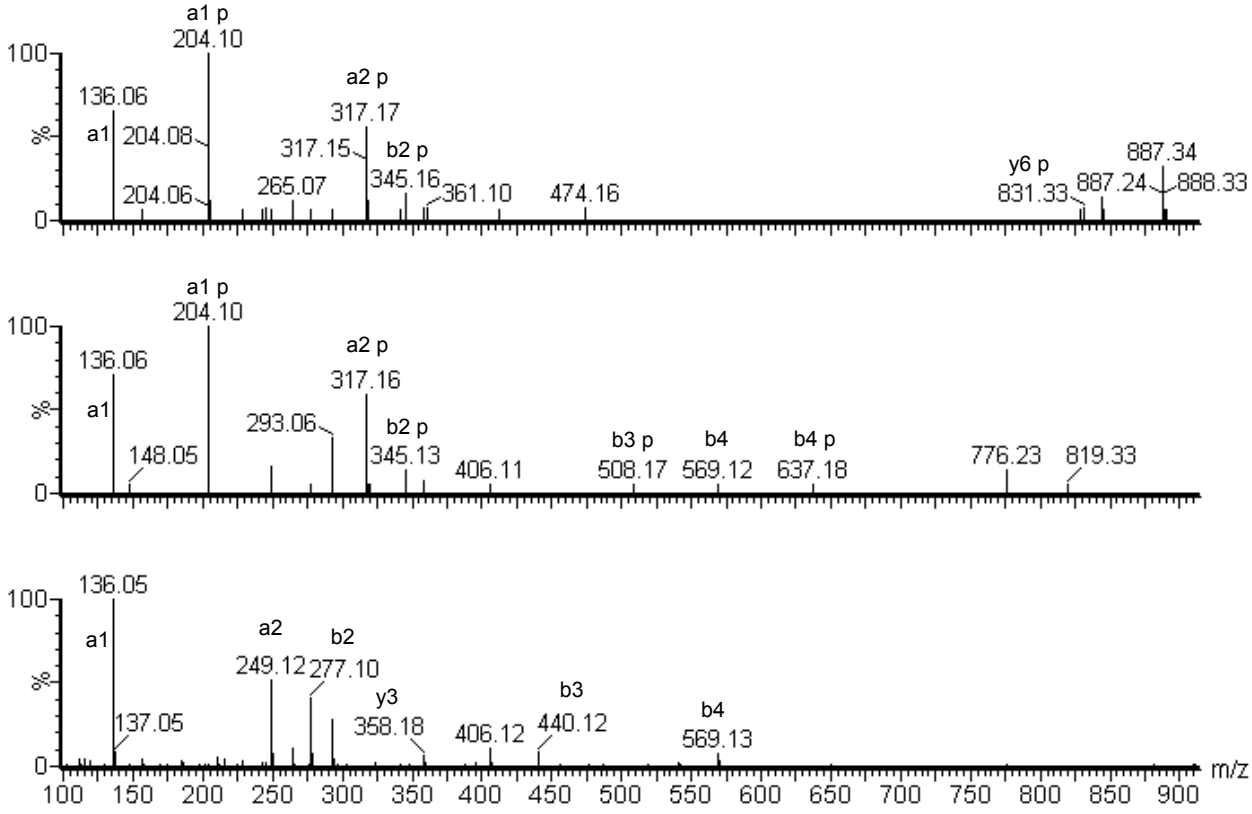


(H) mass spectra of compounds eluting at 10.0 minutes (unprenylated), 11.6 minutes (singly prenylated) and 13.6 minutes (doubly prenylated) shown in figure S9G. Spectral regions with singly and doubly charged ions are shown.

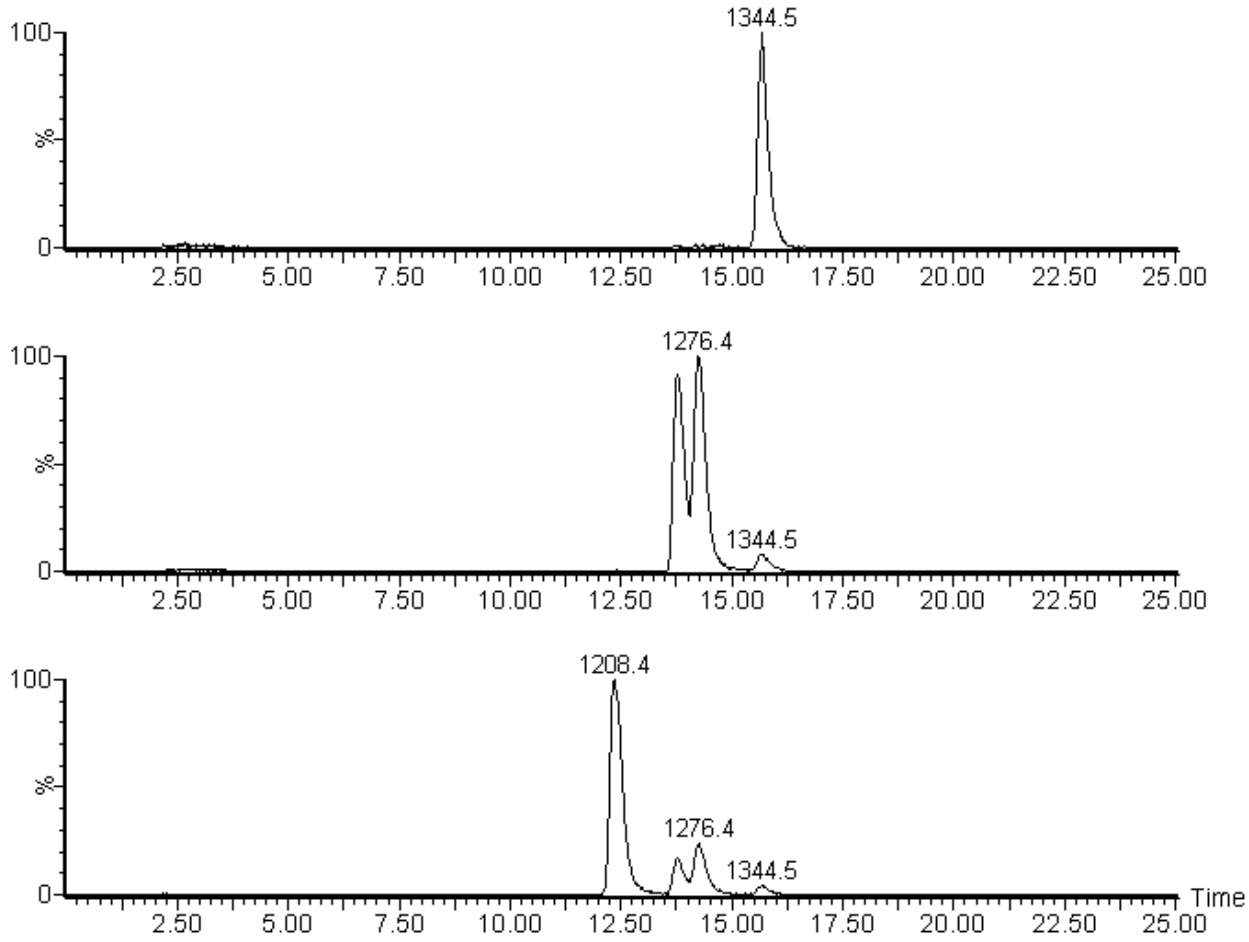



100
%
0
531.71
532.21
497.69
498.19
463.67
464.16
462.19
475.24
496.20
504.70
508.67
538.72
542.70
553.69
100
%
0
497.69
498.19
463.66
498.68
462.18
464.66
487.20
504.70
519.67
530.66
100
%
0
463.64
464.15
464.66
470.67
474.65
455.16
487.21
501.22
m/z
450
460
470
480
490
500
510
520
530
540
550

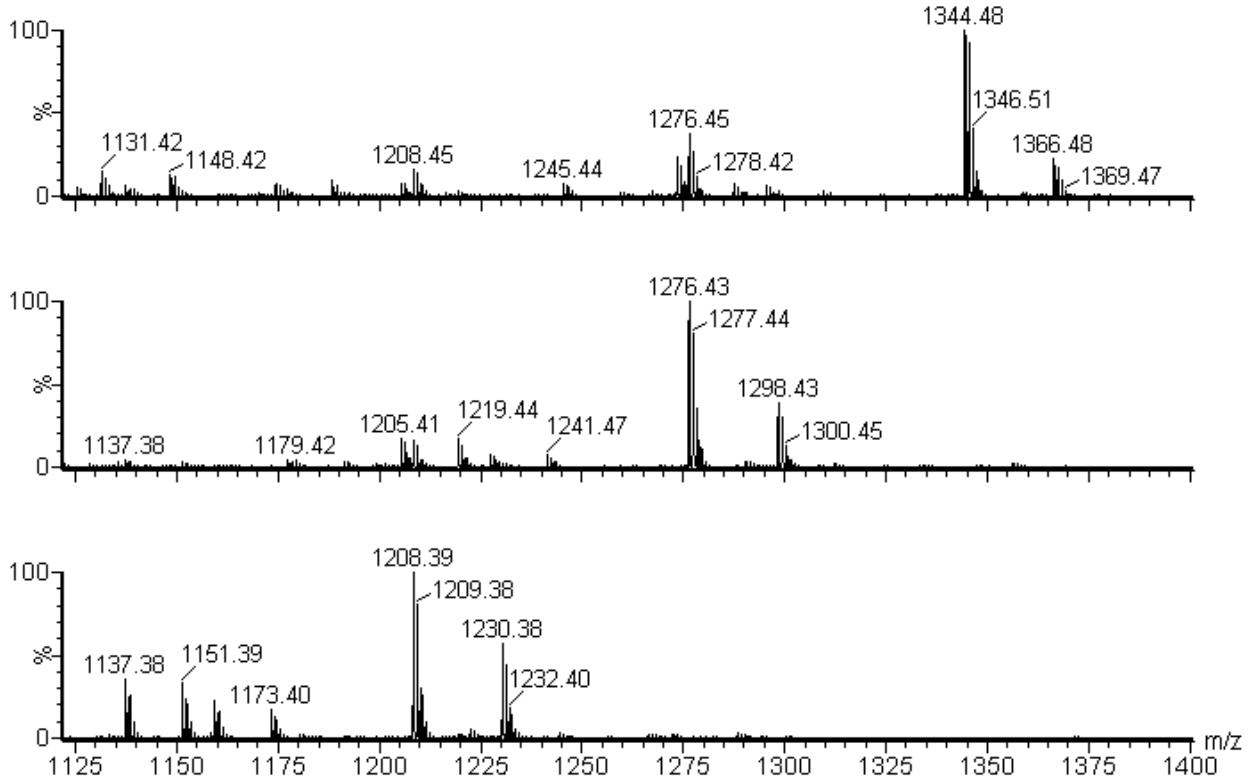
(I) top: fragmentation of doubly prenylated YLYEIAR (*m/z* = 1062.6), middle: fragmentation of singly prenylated YLYEIAR (*m/z* = 994.5), bottom: fragmentation of unprenylated YLYEIAR (*m/z* = 926.5).



(J) LC/MS chromatograms of unprenylated YLYQWLGAPV (1208.6 Da), singly prenylated YLYQWLGAPV (1276.7 Da) and doubly prenylated YLYQWLGAPV (1344.8 Da).



(K) mass spectra of compounds eluting at 12.5 minutes (unprenylated), 14.3 minutes (singly prenylated) and 15.7 minutes (doubly prenylated) shown in figure S9J.



(L) top: fragmentation of doubly prenylated YLYQWLGAPV (*m/z* = 1344.8), middle two traces: fragmentation of singly prenylated YLYQWLGAPV (*m/z* = 1276.7), which elutes from HPLC in two peaks, bottom: fragmentation of unprenylated YLYQWLGAPV (*m/z* = 1208.6).

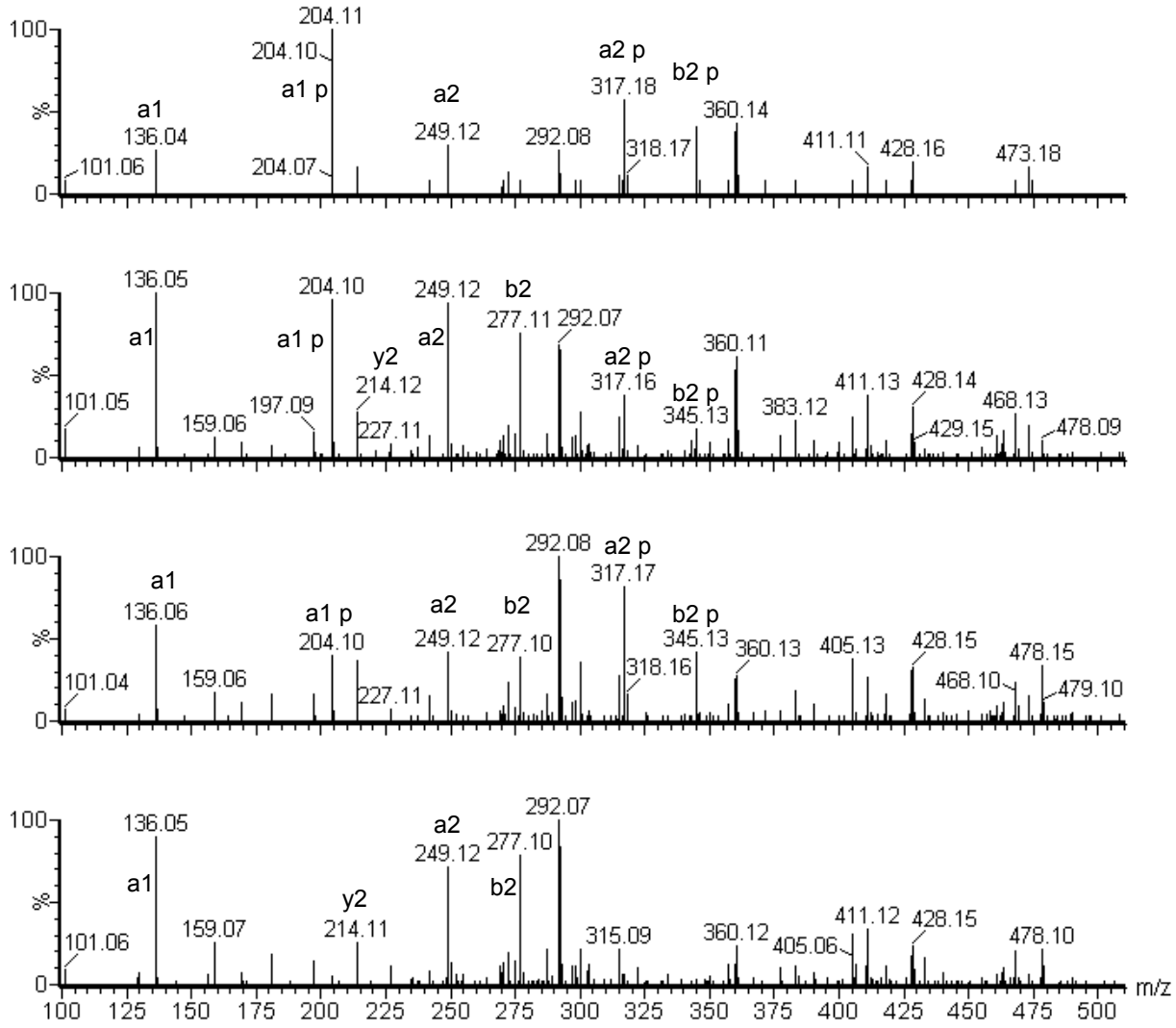


Table S1. Expected and observed masses for peptides used in this study. Data were collected on a Waters Micromass Q-Tof ESI spectrometer in positive mode. The mass spectrometer was calibrated with a Leucine Enkaphalin standard (Sigma-Aldrich). Lower-case y denotes prenylated tyrosine. YLYQWLGAPV and YLYEIAR peptides are C-terminally amidated.

Substrate or product	Expected	Observed
YYY	508.2084	508.2075
yYY	576.2710	576.2717
yyY	644.3336	644.3326
YLYQWLGAPV	1208.6468	1208.6582
YLYQWLGAPV	1208.6468	1208.6631
yLYQWLGAPV	1276.7094	1276.7200
yLYQWLGAPV	1276.7094	1276.7151
yLyQWLGAPV	1344.7720	1344.7806
yLyQWLGAPV $[M+Na]^+$ ion	1366.7539	1366.7628
YLYEIAR	926.4862	926.5126
yLYEIAR	994.5488	994.5855
yLyEIAR	1062.6114	1062.6394
cyc[INPYLYP]	861.4505	861.4534
cyc[INPYLYP] - monoprenyl	929.5131	929.5161
cyc[INPYLYP] - diprenyl	997.5757	997.5787

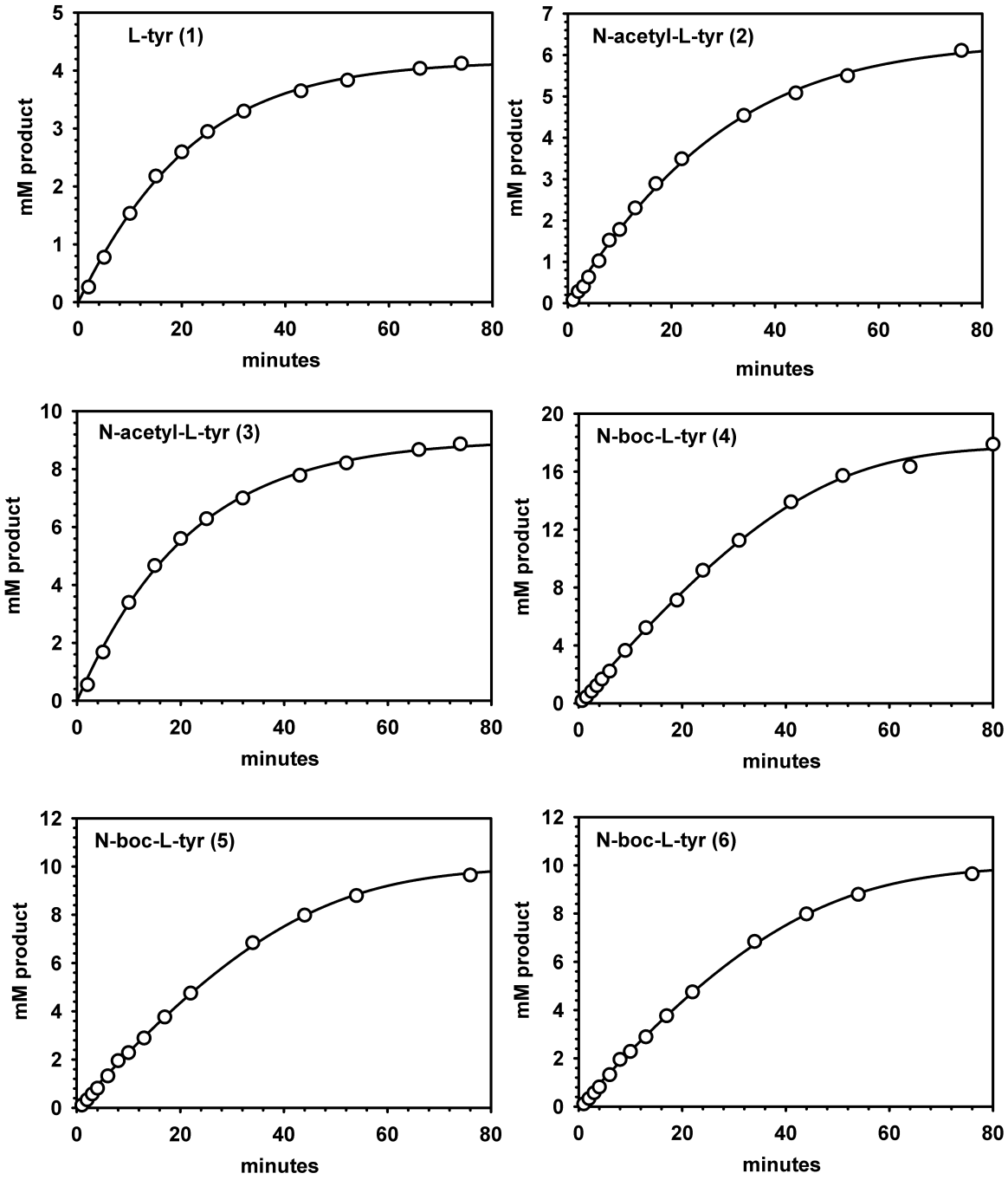
Table S2. Expected and observed masses for peptide fragments. Singly and doubly prenylated ions are denoted by a+p, a+2p, etc. YLYQWLGAPV and YLYEIAR peptides are C-terminally amidated. For tyr-tyr-tyr peptide, unprenylated, singly and doubly prenylated (a2-45)-ions were also observed. This could correspond to loss of (NH_3 + CO) which has been attributed to cyclization of a2 ions (26).

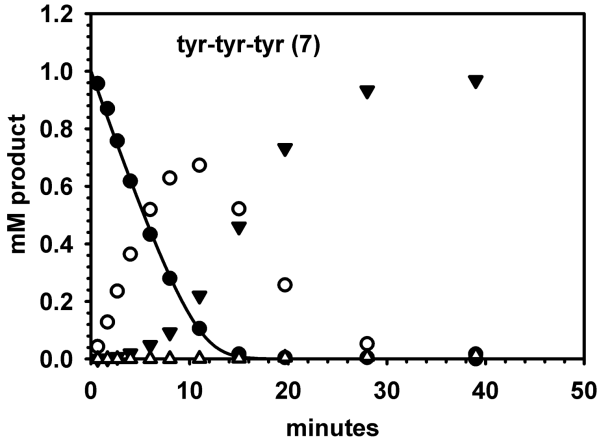
	a-ions	a+p	a+2p	b-ions	b+p	b+2p	y-ions	y+p	y+2p	
unprenylated tyr-tyr-tyr										
Y	136.08			-			-			YY
YY	-			-			-			Y
singly prenylated tyr-tyr-tyr										
Y	136.08	204.14		-	-		-	-	-	YY
YY	299.13	367.20		327.13	-		-	-	-	Y
doubly prenylated tyr-tyr-tyr										
Y	136.08	204.14		-	-		-	-	-	YY
YY	299.13	367.20	435.26	327.13	-	-	-	-		Y

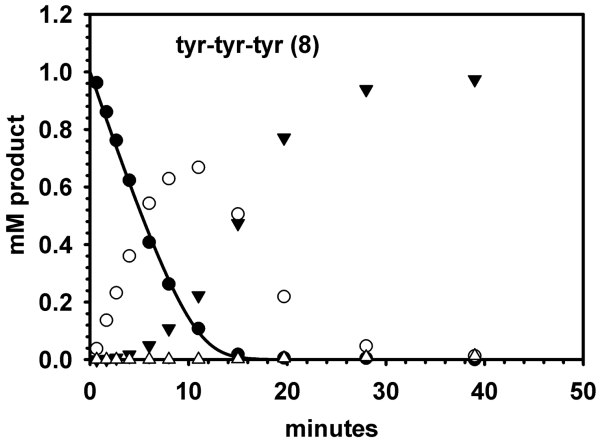
	a-ions	a+p	a+2p	b-ions	b+p	b+2p	y-ions	y+p	
unprenylated YLYEIAR									
Y	136.08			-			-		LYEIAR
YL	249.16			277.15			-		YEIAR
YLY	-			440.22			-		EIAR
YLYE	-			569.26			358.26		IAR
YLYEI	-			-			-		AR
YLYEIA	-			-			-		R
singly prenylated YLYEIAR									
Y	136.08	204.14		-	-		-	-	LYEIAR
YL	249.16	317.22		-	345.22		-	-	YEIAR
YLY	-	-		-	508.28		-	-	EIAR
YLYE	-	-		569.26	637.32		-	-	IAR
YLYEI	-	-		-	-		-	-	AR
YLYEIA	-	-		-	-		-	-	R
doubly prenylated YLYEIAR									
Y	136.08	204.14		-	-		-	831.51	LYEIAR
YL	-	-		277.15	345.22		-	-	YEIAR
YLY	-	-	-	-	-	-	-	-	EIAR
YLYE	-	-	-	-	-	-	-	-	IAR
YLYEI	-	-	-	-	-	-	-	-	AR
YLYEIA	-	-	-	-	-	-	-	-	R

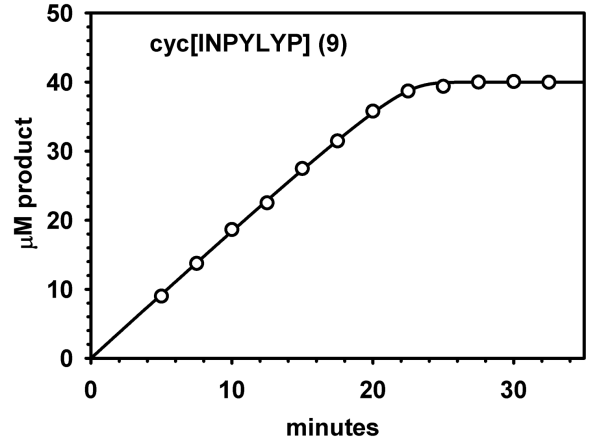
	a-ions	a+p	a+ 2p	b-ions	b+p	b+ 2p	y-ions	y+p	
unprenylated YLYQWLGAPV									
Y	136.08			-			-		LYQWLGAPV
YL	249.16			277.15			-		YQWLGAPV
YLY	-			-			-		QWLGAPV
YLYQ	-			-			-		WLGAPV
YLYQW	-			-			-		LGAPV
YLYQWL	-			-			-		GAPV
YLYQWLG	-			-			-		APV
YLYQWLGA	-			-			214.16		PV
YLYQWLGAP	-			-			-		V
singly prenylated YLYQWLGAPV									
Y	136.08	204.14		-	-		-	-	LYQWLGAPV
YL	249.16	317.22		277.15	345.22		-	-	YQWLGAPV
YLY	-	-		-	-		-		QWLGAPV
YLYQ	-	-		-	-		-		WLGAPV
YLYQW	-	-		-	-		-		LGAPV
YLYQWL	-	-		-	-		-		GAPV
YLYQWLG	-	-		-	-		-		APV
YLYQWLGA	-	-		-	-		214.16		PV
YLYQWLGAP	-	-		-	-		-		V
doubly prenylated YLYQWLGAPV									
Y	136.08	204.14		-	-		-	-	LYQWLGAPV
YL	249.16	317.22		277.15	345.22		-	-	YQWLGAPV
YLY	-	-	-	-	-	-	-		QWLGAPV
YLYQ	-	-	-	-	-	-	-		WLGAPV
YLYQW	-	-	-	-	-	-	-		LGAPV
YLYQWL	-	-	-	-	-	-	-		GAPV
YLYQWLG	-	-	-	-	-	-	-		APV
YLYQWLGA	-	-	-	-	-	-	-		PV
YLYQWLGAP	-	-	-	-	-	-	-		V

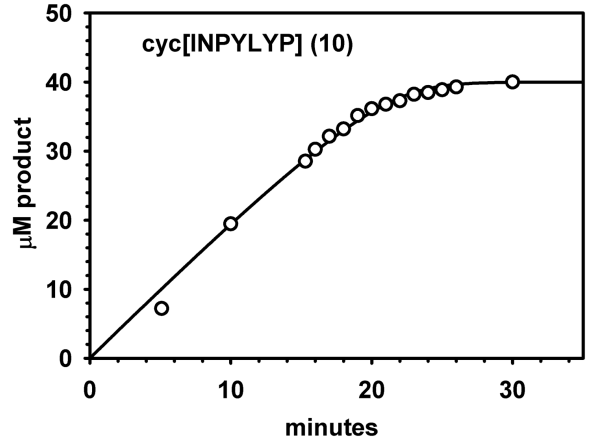
Figure S10. Fits of the Michaelis-Menten equation to progress curves of PagF reactions. Each assay contained NaCl (1100 mM), $MgCl_2$ (200 mM), bis-tris propane, pH 7.5 (100 mM), and DMAPP (50 mM). Concentrations of Tyr substrate and PagF are given in the table. Open circles represent substrate concentrations, lines are best fits. In the Tyr-Tyr-Tyr assays, closed circles represent substrate concentration, open circles represent singly prenylated product, closed triangles represent doubly prenylated product and open triangles represent triply prenylated product.




tyr-tyr-tyr (7)
mM product
minutes
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
10
20
30
40
50


tyr-tyr-tyr (8)
mM product
minutes
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
10
20
30
40
50


cyc[INPYLYP] (9)
µM product
minutes
0
10
20
30
40
50
0
10
20
30


cyc[INPYLYP] (10)
µM product
minutes
0
10
20
30
40
50
0
10
20
30

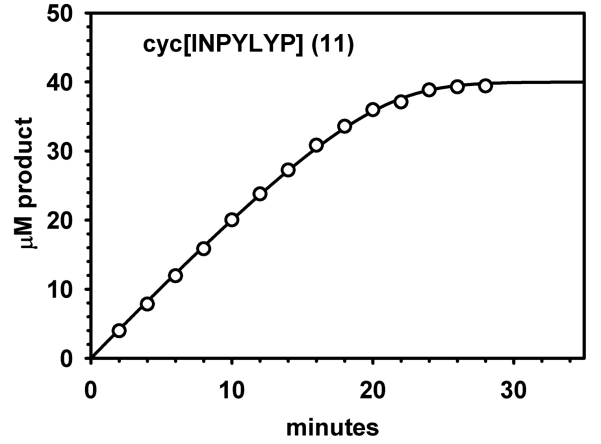

cyc[INPYLYP] (11)
µM product
minutes
0
10
20
30
40
50
0
10
20
30

Table S3. Full kinetic parameters from fitting models to progress curves of single assays of PagF with various substrates.

substrate	PagF (μM)	S0 (mM)	Maximum rate from fit (mM min^{-1}) [a]	V_{max} (min^{-1})	K_m (mM)	RMS dev. from fit	V/K (min^{-1} mM^{-1})
4.6 mM L-Tyr (1)[b]	50	4.2	0.87	17	16	0.05	1.1
10 mM N-acetyl-L-Tyr (2)	50	6.3	0.57	11	12	0.09	1.0
10 mM N-acetyl-L-Tyr (3)	50	9.0	4.6	92	92	0.11	1.0
N-acetyl-L-Tyr[c]	5	N/A[c]	N/A[c]	11	3.8	N/A[c]	2.8
20 mM N-boc-L-Tyr (4)	50	18	0.56	11	6.2	0.23	1.8
10 mM N-boc-L-Tyr (5)	50	9.8	0.33	6.5	3.7	0.085	1.8
10 mM N-boc-L-Tyr (6)	50	N/A[d]	0.35	7.1	4.8	0.092	1.5
1 mM Tyr-Tyr-Tyr (7)	20	N/A	0.11	5.6	0.14	0.013	39
1 mM Tyr-Tyr-Tyr (8)	20	N/A	0.12	5.8	0.16	0.016	37
0.4 mM cyc[INPYLYP] (9)	5	0.37	0.037	7.4	$2.2x10^{-4}$	$9.1x10^{-3}$	$3.4x10^{4}$
0.4 mM cyc[INPYLYP] (10)	1	0.37	$9.0x10^{-3}$	9.0	$5.2x10^{-5}$	$8.0x10^{-3}$	$1.7x10^{5}$

[a] The maximum rate given by the fit depends on the concentration of enzyme in the assay. V_{max} was calculated from this maximum rate and the enzyme concentration.
[b]Numbers correspond to assays shown in Figure S10
[c]These parameters are from initial rates in a series of reactions with different substrate concentrations.
[d]S0 was not allowed to vary in these fits.

Figure S11. N-acetyl-L-Tyr concentration vs. activity in initial rate assays. Each assay contained NaCl (1100 mM), $MgCl_2$ (200 mM), bis-tris propane, pH 7.5 (100 mM), and DMAPP (10 mM). Reactions were incubated at 37 ^{o}C and were initiated by PagF addition. Open circles represent rates determined in single assays; the line is the best fit to the Michaelis-Menten equation (V_{max} = 10.6 ± 0.7 min^{-1}, K_m = 3.8 ± 0.8 mM).

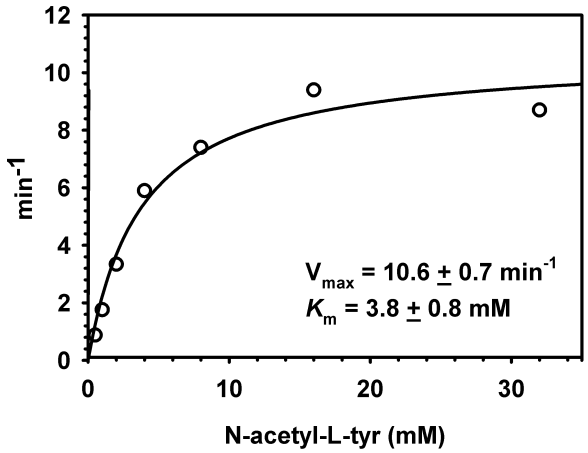
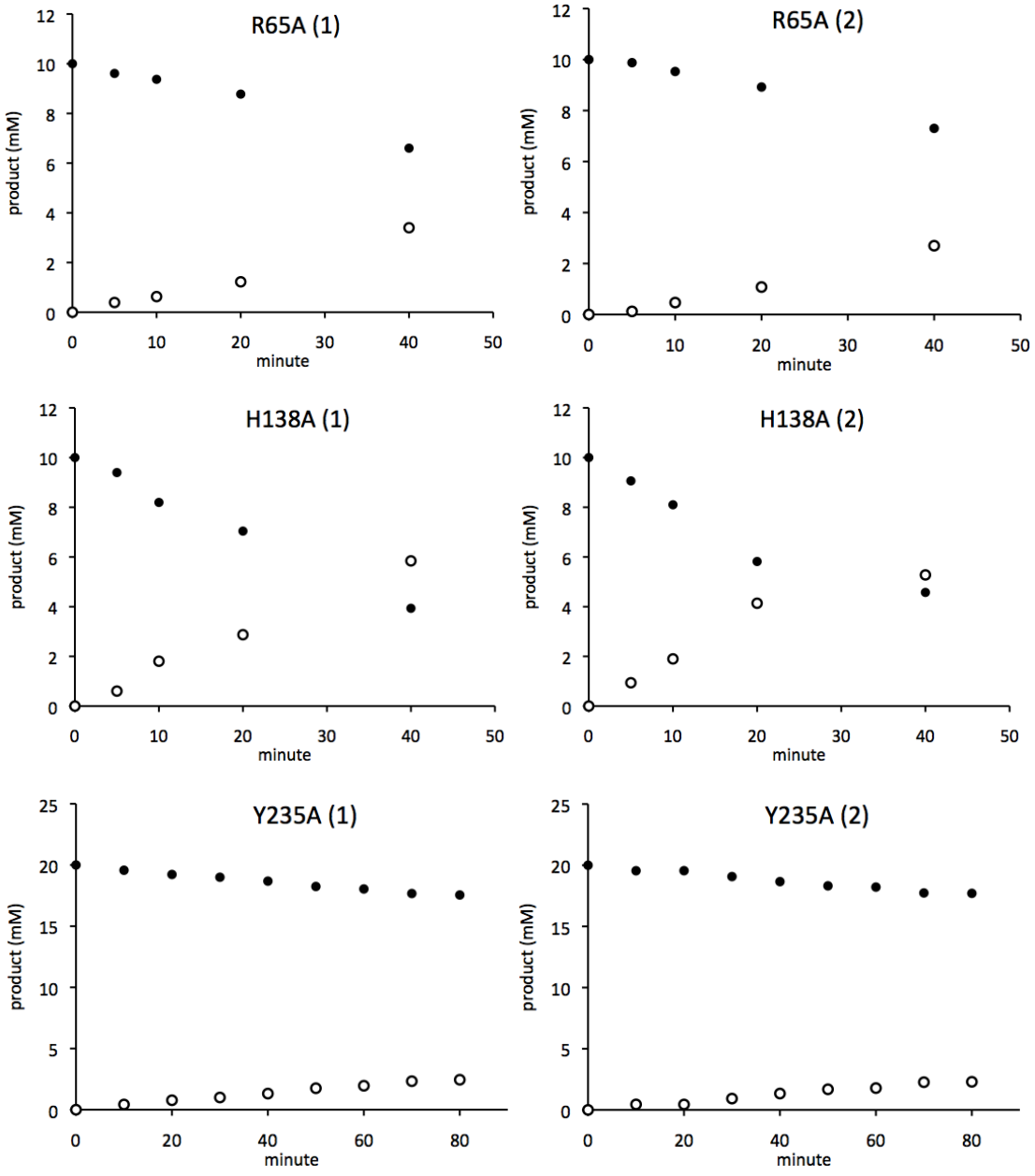


Figure S12. Progress curves of reaction of PagF variants using Tyr_3 tripeptide as a substrate. Each assay contained NaCl (1100 mM), $MgCl_2$ (200 mM), bis-tris propane, pH 7.5 (100 mM), and DMAPP (50 mM). Concentrations of substrate and PagF are given in the table. In the assays, closed circles represent substrate concentration, open circles represent singly prenylated product, closed triangles represent doubly prenylated product and the lines represent the best fit to the Michaelis-Menten equation.



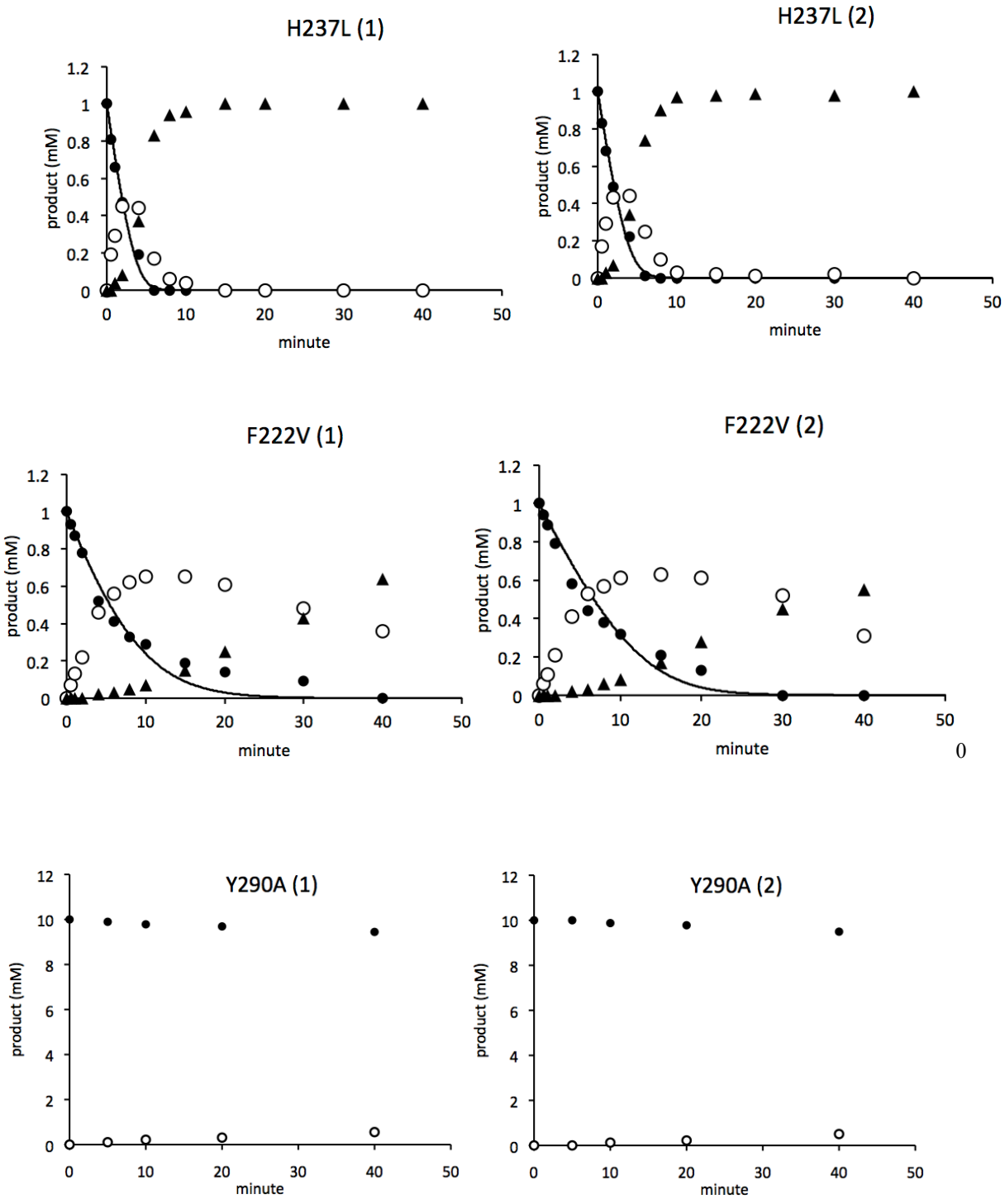


Figure S13. Potential of mean force determined from umbrella sampling MD simulations modelling the initial approach of cyc[INPYLYP] towards PagF leading with either Tyr4 (red crosses) or Tyr6 (green Xs). See Methods for exact definitions of the distances. Figure generated with Gnuplot 4.6.

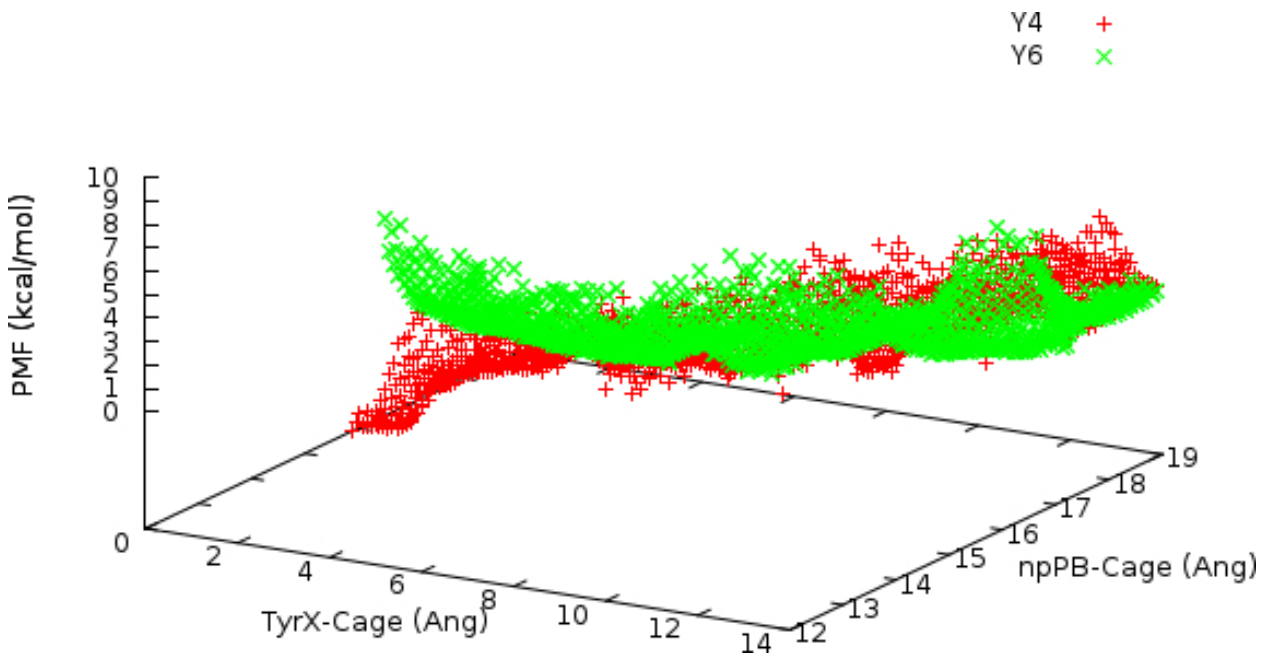

Y4
Y6
PMF (kcal/mol)
0
1
2
3
4
5
6
7
8
9
10
TyrX-Cage (Ang)
0
2
4
6
8
10
12
14
npPB-Cage (Ang)
12
13
14
15
16
17
18
19

Figure S14. Left: Average distance versus time over all pulling simulations for guiding Tyr4 (black) or Tyr6 (red) of cyc[INPYLYP] into the active site of PagF (each line is average of 10 simulations, error bars represent standard deviation). Right: Total restraint energy for each individual simulation for pulling Tyr4 (black dots) or Tyr6 (red dots) into the active site of PagF (1 frame = 1 ps).

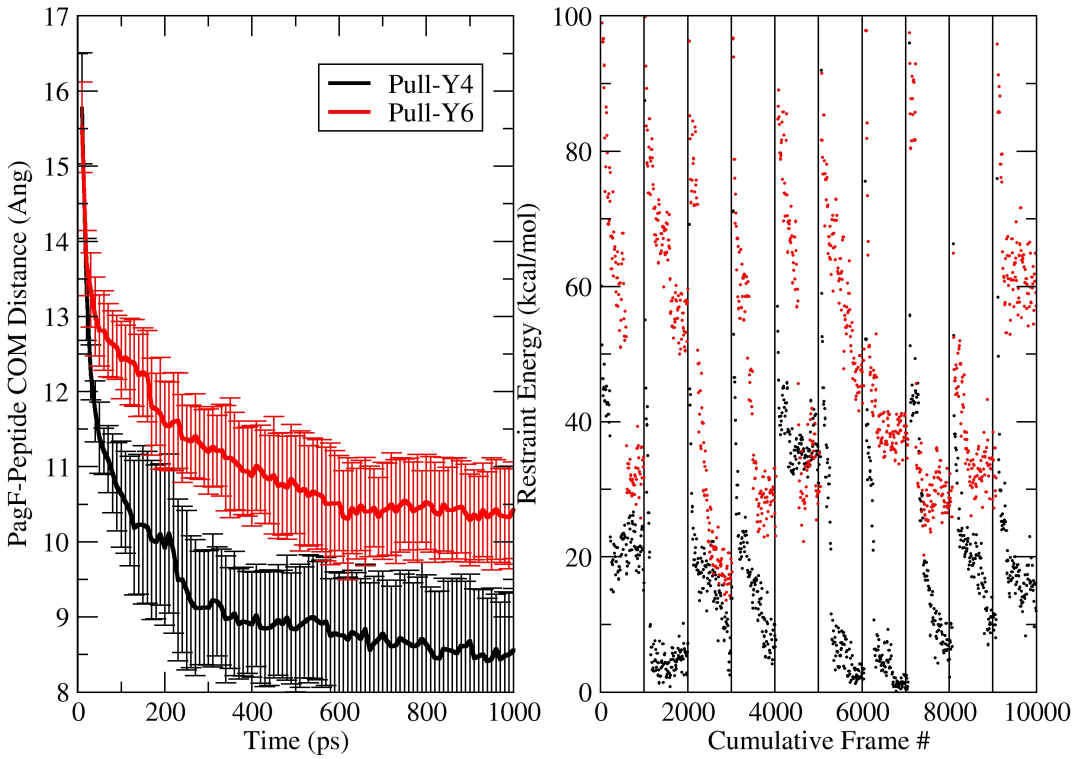


Figure S15. PMFs generated using data from 1-500 ps (black) and 1-1000 ps (red) from the umbrella sampling simulations for guiding either Tyr4 (top) or Tyr6 (bottom) of cyc[INPYLYP] into the active site of PagF. The 2D PMFs are shown flattened along the second restraint (the aromatic cage to peptide backbone) to facilitate comparison.

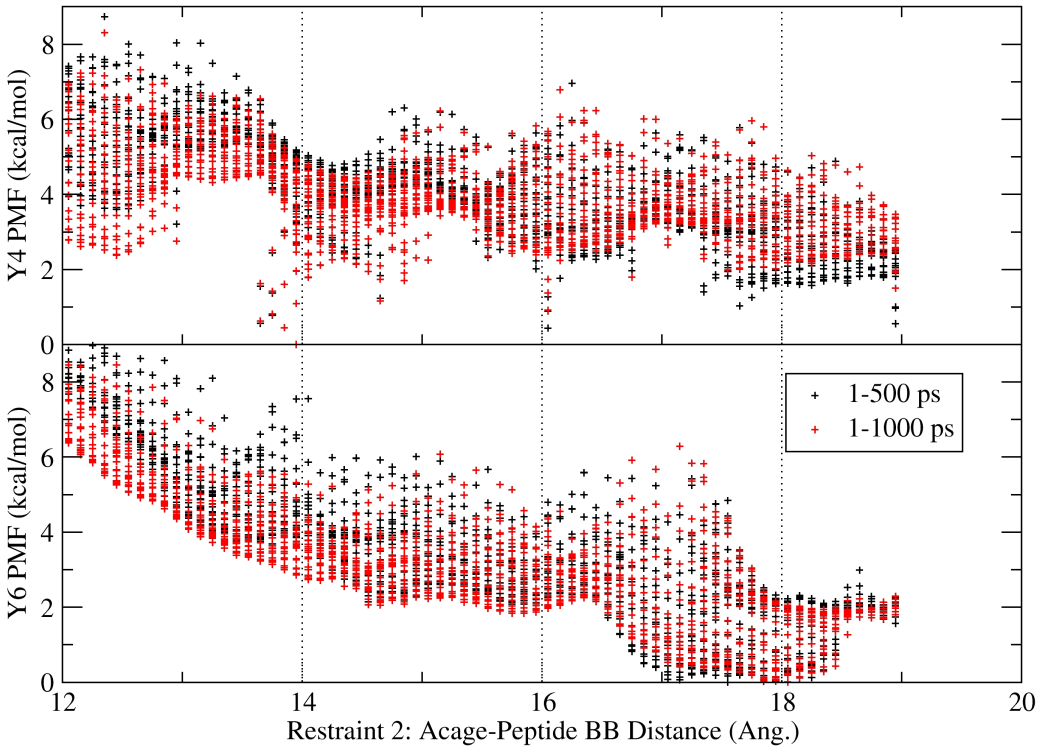


Table S3. Data collection, phasing and refinement statistics

	Native	Lead Acetate	bocTyr	$(Tyr)_3$	cyclic[INPYLYP]
Data collection					
Space Group	$P6_5$	$P6_5$	$P6_5$	$P6_5$	$P6_5$
a, b, c (Å)	116.9, 116.9, 43.6	116.1, 116.1, 43.5	116.4, 116.4, 43.7	117.1, 117.1, 43.8	116.5, 116.5, 43.6
Resolution (Å)[1]	50-1.8 (1.83-1.8)	50-2.65 (2.7-2.65)	50-2.1 (2.14-2.1)	50-1.9 (1.93-1.9)	50-2.2 (2.3-2.2)
R_{sym} (%)[2]	9.0 (61.2)	9.0 (27.9)	8.3 (67.3)	9.1 (67.5)	9.1 (67.3)
I/σ(I)	14.1 (2.0)	36.8 (8.5)	16.5 (2.4)	15.6 (2.3)	12.7 (2.0)
Completeness (%)	99.8 (100)	99.7 (100)	97.0 (93.0)	99.7 (98.4)	93.5 (69.9)
Redundancy	4.4 (3.3)	12.5 (11.7)	6.7 (5.4)	5.3 (4.8)	3.7 (3.4)
Accession Codes	5TTY		5TU4	5TU5	5TU6
Phasing					
FOM		0.461			
Refinement					
Resolution (Å)	25.0-1.8		25.0-2.1	25.0-1.9	25.0-2.22
No. reflections	30,175		19,016	25,668	16,035
R_{work} / R_{free}[4]	19.9/22.5		18.4/25.8	20.7/23.4	18.2/23.9
Number of atoms					
Protein	2465		2468	2460	2468
Peptide	-		21	36	62
DMSPP	-		14	14	14
Water	277		199	136	112
B-factors					
Protein	20.7		27.0	21.9	33.4
Peptide	-		40.4	29.1	66.0
DMSPP	-		42.7	28.5	40.4
Water	30.8		33.6	28.9	36.5
R.m.s deviations					
Bond lengths (Å)	0.005		0.012	0.006	0.010
Bond angles (°)	1.04		1.45	1.1	1.44

1. Highest resolution shell is shown in parenthesis.
2. $R_{sym} = \Sigma |(I_i - \langle I_i \rangle| \Sigma I_i$ where I_i = intensity of the *i*th reflection and $\langle I_i \rangle$ = mean intensity.
3. Mean figure of merit (acentric/centric)
4. 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.

References

1. McIntosh JA, Donia MS, Nair SK, & Schmidt EW (2011) Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *J Am Chem Soc* 133:13698-13705.
2. Woodside AB, Huang Z, & Poulter CD (1988) Trisammonium Geranyl Diphosphate. *Organic Syntheses* 66:211.
3. Murakami M, Itou Y, Ishida K, & Shin HJ (1999) Prenylagaramides A and B, New Cyclic Peptides from Two Strains of Oscillatoria agardhii. *Journal of Natural Products* 62:752-755.
4. Case DA*, et al.* (2014) AMBER 14 (University of California, San Francisco, San Francisco, CA).
5. Case DA*, et al.* (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry* 26:1668-1688.
6. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, & Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81:3684-3690.
7. Darden T, York D, & Pedersen L (1993) Partical mesh Ewald - an Nlog(N) method for Ewald sums in large systems. *Journal of Chemical Physics* 98:8577-8593.
8. Ryckaert JP, Ciccotti G, & Berendsen HJC (1977) Numerical integration of cartesian equations of motion of a system with constraints - molecular dynamics of N-alkanes. *Journal of Computational Physics* 23:327-341.
9. Roe DR & Cheatham III TE (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 9:3084-3095.
10. Wang J, Wolf RM, Caldwell JW, Kollman PA, & Case DA (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry* 25:1157-1174.
11. Frisch MJ*, et al.* (2009) Gaussian 09 (Gaussian, Inc., Wallingford, CT, USA).
12. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein M (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79:926-935.
13. Joung IS & Cheatham III TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B* 112:9020-9041.
14. Ashton NN, Roe DR, Weiss RB, Cheatham TE, & Stewart RJ (2013) Self-Tensioning Aquatic Caddisfly Silk: Ca2+-Dependent Structure, Strength, and Load Cycle Hysteresis. *Biomacromolecules* 14:3668-3681.
15. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, & Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* 13:1011-1021.
16. Grossfield A (WHAM: the weighted histogram analysis method), 2.0.9.
17. Otwinowski Z, Borek D, Majewski W, & Minor W (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr A* 59:228-234.
18. Zwart PH*, et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419-435.

19. Adams PD*, et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948-1954.
20. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
21. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.
22. Kleywegt GJ & Brunger AT (1996) Checking your imagination: applications of the free R value. *Structure* 4:897-904.
23. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
24. Gerlt JA*, et al.* (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* 1854:1019-1037.
25. Shannon P*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504.
26. Simon-Manso Y, Neta P, Yang X, & Stein SE (2011) Loss of 45 Da from a2 ions and preferential loss of 48 Da from a2 ions containing methionine in peptide ion tandem mass spectra. *J Am Soc Mass Spectrom* 22:280-289.