

## **Studies on a Landscape of Perfluoroaromatic-Reactive Peptides**

**Authors:** Ethan D. Evans, Bradley L. Pentelute

**Affiliations:** Department of Chemistry, Massachusetts Institute of Technology, 77  
Massachusetts Avenue, Cambridge, MA, 02139

\*correspondence to:

Bradley L. Pentelute

Email: [blp@mit.edu](mailto:blp@mit.edu)

Tel. (+1) 617 324 0180

## Table of Contents

1. Materials .....	3
2. Methods .....	3
2.1. Solvent mixtures .....	3
2.2. Liquid chromatography-mass spectrometry (LCMS) .....	3
2.3. Peptide synthesis, cleavage, purification and refolding.....	3
2.4. Capture Agent (CA) synthesis .....	4
2.5. Kinetics Analysis .....	4
2.6. Rosetta structure prediction and clustering .....	4
2.7. Circular Dichroism .....	6
3. Experimental data, tables and figures .....	6
3.1. Sequences and masses of peptides used .....	6
3.2. LCMS kinetics analysis .....	7
3.3. CD reaction analysis .....	4
3.4. CD characterization of additional MPs without CA reaction .....	33
3.5. Rosetta ab initio structure prediction .....	52
4. References .....	53

## 1. Materials

Pentafluorophenyl sulfide was from Santa Cruz Biotechnology. 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), Fmoc-L-Ala-OH, Fmoc-L-Cys(trt)-OH, Fmoc-L-Asp(tBu)-OH, Fmoc-L-Glu(tBu)-OH, Fmoc-L-Phe-OH, Fmoc-Gly-OH, Fmoc-L-His(Boc)-OH, Fmoc-L-Ile-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Leu-OH, Fmoc-L-Met-OH, Fmoc-L-Asn(Trt)-OH, Fmoc-L-Pro-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Ser(tBu)-OH, Fmoc-Val-OH, Fmoc-L-Thr(tBu)-OH, Fmoc-L-Trp(Boc)-OH, Fmoc-L-Tyr(tBu)-OH, Fmoc-L-Lys(biotin)-OH were from Chem-Impex International. H-rink-amide chemmatrix Hyr resin was from PCAS BioMatrix, Inc. (7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) was from P3 BioSystems. Tris(2-carboxyethyl)phosphine hydrochloride was from Hampton Research. *N, N*-dimethylformamide (DMF), acetonitrile (ACN), diethyl ether were from VWR. Trifluoroacetic acid (TFA) and formic acid (FA) were obtained from Sigma-Aldrich. Other chemicals listed were purchased from either Sigma-Aldrich or VWR and used as received.

## 2. Methods

### 2.1. Solvent mixtures

Solvent A is water with 0.1% (v/v) TFA, solvent B is acetonitrile with 0.1% (v/v) TFA. Solvent C is water with 0.1% (v/v) FA and solvent D is acetonitrile with 0.1% (v/v) FA.

### 2.2. Liquid chromatography-mass spectrometry (LCMS)

TIC refers to total ion chromatogram of an LCMS trace. LCMS chromatograms and mass spectra were obtained using either an Agilent 6520 ESI-Q-TOF mass spectrometer or an Agilent 6550 iFunnel Q-TOF mass spectrometer. Method 1 and 3 were used on the 6550, while method 2 was used on the 6520. LCMS analysis was done using the Agilent MassHunter package. General LCMS methods were as follows:

#### Method 1:

LC method: 0-4 minutes 99% C and 5% D, 4-17.5 minutes 1-61% D linear ramp, 17.5-18 minutes 61% D, 0.4mL/min flow rate.

Column: Poroshell 300SB-C<sub>3</sub> column (1 x 75mm, 5μm), 40°C

MS parameters: positive electrospray ionization (ESI), MS off at 10 minutes.

#### Method 2:

LC method: 0-2 minutes 5% D, 2-11 minutes 5-65% D linear ramp, 11-12 minutes 65% D, 0.8mL/min flow rate.

Column: Zorbax 300SB C<sub>3</sub> column (2.1 x 150mm, 5μm), 40°C

MS parameters: positive electrospray ionization (ESI).

#### Method 3:

LC method: 0-2 minutes 1% D, 2-12 minutes 5-61% D linear ramp, 12-16 minutes 61-91% D, 0.1mL/min flow rate.

Column: Phenomenex Jupiter C<sub>4</sub> (1 x 150mm, 5μm), 40°C

MS parameters: positive ESI

### 2.3. Peptide synthesis, cleavage, purification and refolding

Peptides were synthesized, cleaved and purified as previously described.<sup>1,2</sup> Peptides that required chemical refolding were diluted in a minimal volume of 0.5x buffer (12.5 mM pH 7.45 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 50 mM NaCl, 2.5 mM CaCl<sub>2</sub> and 2.5

mM MgCl<sub>2</sub>) to which solid guanidinium chloride was added until the solution went clear. The solution was then filtered through a 0.22 μm filter and then diluted to 50 mL with water. Solutions were then concentrated on a 3 kDa spin filter (Amicon Ultra, Millipore-Sigma) and washed with water until the absorbance reading at 214 nm was indistinguishable from the 0.5x buffer.

#### 2.4. Capture Agent (CA) synthesis

The CA was synthesized and purified as previously described.<sup>1</sup>

#### 2.5. Kinetics Analysis

Kinetics analysis was done as previously reported and is summarized here due to its importance.<sup>1</sup> All reactions used 100 μM peptide, 500 μM CA, 5 mM TCEP, 1x buffer (25 mM pH 7.45 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 100 mM NaCl, 5 mM CaCl<sub>2</sub> and 5 mM MgCl<sub>2</sub>) at room temperature. A 5 μl reaction aliquot was quenched with either 95 μl (for 6520 LCMS analysis) or 195 μl (for 6550 LCMS analysis) of 49.75% H<sub>2</sub>O, 49.75% ACN, 0.5% TFA. The integrated peak areas for the peptide in the linear range of the instrument were used for kinetics analysis. This data was then fit to a second order rate constant model:

$$k_2(A_0 - B_0)t = \ln\left(\frac{B_0A_t}{A_0B_t}\right)$$

$B_0$  is the initial MP concentration and  $B_t$  is its concentration at time  $t$ .  $A_0$  is the initial capture agent concentration,  $A_t$  is the CA's concentration at a given time which was determined using  $A_0 - B_t$  as the CA peak intensity was outside the linear range (and often the mass spectrum was turned off prior to its elution),  $k_2$  was the second order rate constant.

#### 2.6. Rosetta structure prediction and clustering

Structure prediction was not performed for MP02 because it was not reactive. For each peptide the following protocol was used (this same protocol is presented in Evans, E.D. et al. Conformational stabilization and rapid labeling of a 29-residue peptide by a small molecule reaction partner, *submitted*):

1. launch 100 of the following Rosetta protocol (40,000 total predictions):
 

```
AbinitioRelax.default.linuxgccrelease -in:file:fasta ./seq_name.fasta -in:file:frag3
./aat000_03_05_name.200_v1_3 -in:file:frag9 ./aat000_09_05_name.200_v1_3
-abinitio:relax -dump_connect_info -connect_info_cutoff 0 -
use_truncated_termini false -nstruct 400 -out:file:silent ./name.silent -out:path
/out/file/path/ -relax::fast -abinitio::increase_cycles 10 -abinitio::rg_reweight 0.5 -
abinitio::rsd_wt_helix 0.5 -abinitio::rsd_wt_loop 0.5 -use_filters true -psipred_ss2
./t000_name.psipred_ss2 -kill_hairpins ./t000_name.psipred_ss2
```
2. Following the structure prediction, the output file would be cleaned to remove any lines that could not be handled by the clustering program (lines when the .silent file was being written to by multiple copies of the same program). We would then run the following protocol to generate the true clustering input (the Rosetta clustering protocol was not actually used to supply any information to the actual clustering program, it simply got the input in the correct form):
 

```
cluster.linuxgccrelease -in:file:silent ./path/to/file/file.silent -in:file:fullatom
-out:file:silent ./out/path/clustered_name.silent -cluster:radius -1 -score:weights
score3
```

3. All of the pdb files were released with the following and moved into their own directory:  

```
extract_pdbs.linuxgccrelease -in:file:fullatom -in:file:silent ./clustering_file.silent
```
4. These were then clustered with a custom program (`single_batch_full_cluster.sh` on the github page below) that performed k-means clustering where  $k=40$ . This program used pymol for structure alignment and is easily customizable to cluster based on subsequences in the structure. This program is available via the following github page: [https://github.com/ethanev/Structure\\_clustering](https://github.com/ethanev/Structure_clustering). This program was customized for a 64 core compute node using a Slurm job manager but can easily be tuned for alternative workstations. Generally, it launches a program that assigns a data chunk (structures to be assigned to a center) to each cpu along with the current centers. Each chunk is then independently clustered and each clustering program writes its output. Then a center update program is launched to update the means and the process repeated. This gives a number of useful outputs including: pdb files of the cluster centers (these are average structures so atom placement may be nonsensical but strong structural elements of clusters are visually quite apparent), a file concerning the amount the centers moved in a given cycle, a file of text mapping the center structure to the size of its cluster, another file mapping (for each cycle) the center to a list of all pdb's in the cluster (used for energy analysis later on).
5. Next, the individual structural averages were clustered because due to initial starting conditions two (or more) clusters may actually be the 'same'. To accomplish this, the 40 centers were hierarchically clustered using the `cluster_means_2.sh` script. To visualize this result required the following two pieces of data: the resulting pickle file from the `cluster_means_2` script along with a \*.txt file containing the dictionary of the final rounds' center-to-cluster size (taken from the final dictionary of the `clusters_new.txt` file generated from the clustering program). Visualization was then performed with `plot_dendrogram.py` to make a dendrogram of the centers.
6. Further visualization and sequence-to-sequence analysis was performed using heatmaps. To make a heatmap, all center distances (for either the same sequence or two different sequences) were calculated with the `two_structure_distance.sh` script. This script is used by providing the path to the two directories containing the centers (note only the centers should be in this directory) along with the output files generated by the `plot_dendrogram.py` program. The distance pickle file is then plotted with `plot_heatmap.py`
7. Manual inspection of the dendrogram and heatmap allow for cluster family determination as typically many clusters actually represent the same overall structure. Once families are determined, structures that best represent the family are chosen by energy and root-mean-square deviation (RMSD) analysis.
8. Energy analysis is performed with the `energy_analysis.sh` program. Generally, this script takes a text file that lists all the centers that represent a given structure for each family and outputs the energy and **in-family** RMSD between the five lowest energy structures in each cluster and the rest of the family. In the following example there are two families where the first family is represented by center\_1 and 2 from the initial clustering and family two was represented by a single center. Thus this program would located the 5 lowest energy structures for center\_1, 2 and 3 and calculate the in family RMSD for each – so 10 possible structures for family 1 and five for family 2.

Example:

center\_1, center\_2  
center\_3

This program would then also release the .pdbs for each of the structures analyzed (15 in total here) and create a summary csv file.

9. Following this analysis the best representative structures were determined by a combination of lowest energy and smallest in-family RMSD (with this taking precedence)

### 2.7. Circular Dichroism

All experiments were performed at 24 °C on a Jasco J-1500 spectrometer at MIT's Biophysical Instrumentation Facility (BIF). In general, spectra were obtained in 10 mM sodium phosphate buffer at pH 7.45 with 1 mM TCEP and 50 μM peptide. Spectra of the reacted MPs had 250 μM CA in addition to the previous conditions and were reacted ~24 hours prior to measurement. The CA labeled versions were not purified post reaction and were analyzed in their reaction mixture. Several samples were not soluble in this phosphate buffer and thus their spectra are not presented.

## 3. Experimental data, tables and figures

### 3.1. Sequences and masses of peptides used

Table S1. Names, sequences and masses of peptides used

Name	Sequence	Calculated mass	Observed mass
MP02	MPNYGPLSPSQPSHGYTFWMVPIWDNSHNAAGSGSLGHHHHHRL	5049.3	5049.3
MP03	MTSVTASLLMHFCPIRAHITNKPSFNPSG	3155.6	3155.7
MP04	MRTPIKFAPRLSQPFPCFRKQHLHLPLIEG	3821.1	3821.2
MP05	MRPCARRDRTLWCDFDPAWFLLSGFSCG	3373.6	3373.7
MP06	MGIVHNATRFKRCFYFIAIRQSKNSIRVSG	3669.9	3670.0
MP07	MKTFSSDQRFKCKYRIYFHLRQRNHNTSVG	3961.0	3961.1
MP08	MQHEDLCTWYGFCPSGNFTPRNLRGDSGDG	3301.4	3301.5
MP09	MRYYIVLRLKSWCGGASARSSPRSCATKLLG	3428.8	3428.8
MP10	MHNAYLRKSMRQLCYFRRTLHNIHVMSHRG	3753.9	3754.0
MP11	MSTGDCHIQLRPFHGNIAWMRSGNMDG	3166.4	3166.4
MP12	MVKLSGKERTTRNCFSSFLASRRTKKFNNLSG	3722.0	3722.0
MP13	MGHLHICMVRVNTSGHILSVGHKSYSSHKTG	3557.8	3557.8
MP14	MSSGTHYGILNMVIRCHLVKNQTSQMVVLTG	3516.8	3516.8
MP15	MHHYCSKMKRRILMHYLFANTMAHRDLGTNG	3730.8	3730.9
MP16	MHLRMIRYLNRRRHLCVVEIRHGLFASREIG	3967.2	3967.2
MP17	MNGHYPCYLITSVLVGATTSGVPVVVHLRVG	3237.7	3237.8
MP18	MRHYHLTCFQGFRIFRRTVDSLEMEISLG	3540.8	3540.8
MP19	MHMHKTTSYRIRVLVGVVDYRMSHTCLTSSSG	3650.9	3650.9
MP20	MHTSLRSRAKSHRSFSGKASIYTRYLKMG	3457.8	3457.9
MP21	MQNSKHRPRRCLRLPLLRGLHRMFRERG	3762.1	3762.1
MP22	MRSTHQVRRPRNLCSFKHKWLKFLKTLTG	3836.2	3836.2
MP23	MRFFAHCLSIDSSYMWANFVSDRQTRG	3223.5	Could not purify
MP24	MRRTPSTRARGRVLLPTLRFITLCLNLNG	3505.0	3505.0
MP25	MNRIFHKRSTYQMVFGRCSDFTSTYHVLISYG	3842.9	3842.9
MP26	MTATSSSTSRGCRPSTAQVVQRLRGLLLVVG	3229.7	3229.8
MP27	MLFMRLTKKTMATKFCPFRRKRKRERRALYG	4055.2	4955.4

### 3.2. LCMS kinetics analysis

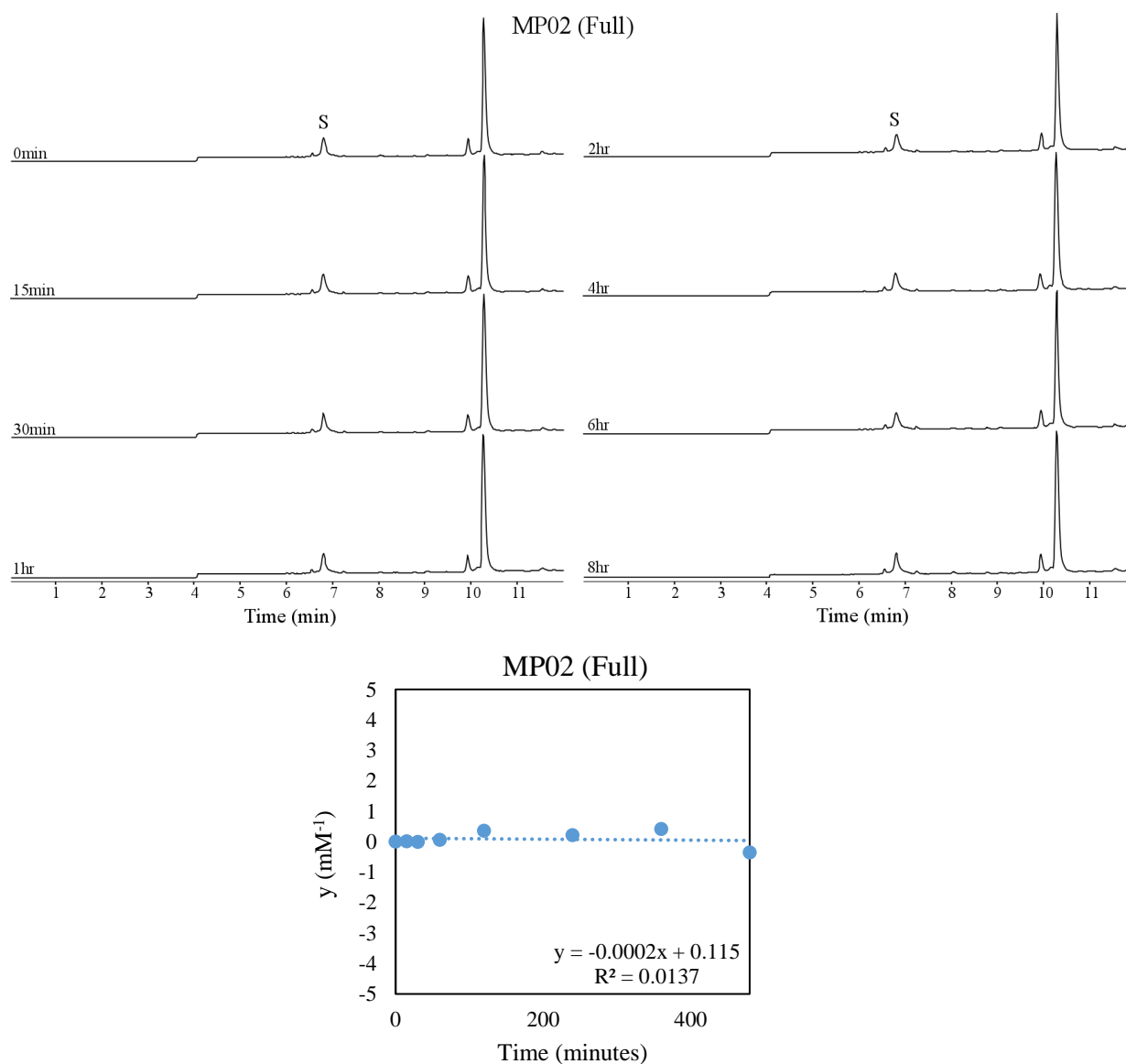


Figure S1. LCMS TIC traces for the kinetics analysis of full length (C-terminal constant region containing) MP02 (top). S – starting material, no CA labeled product was observed. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

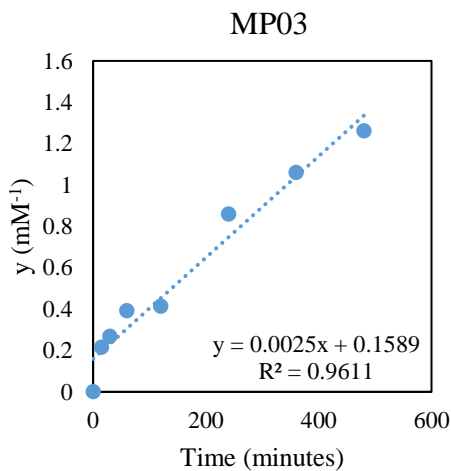
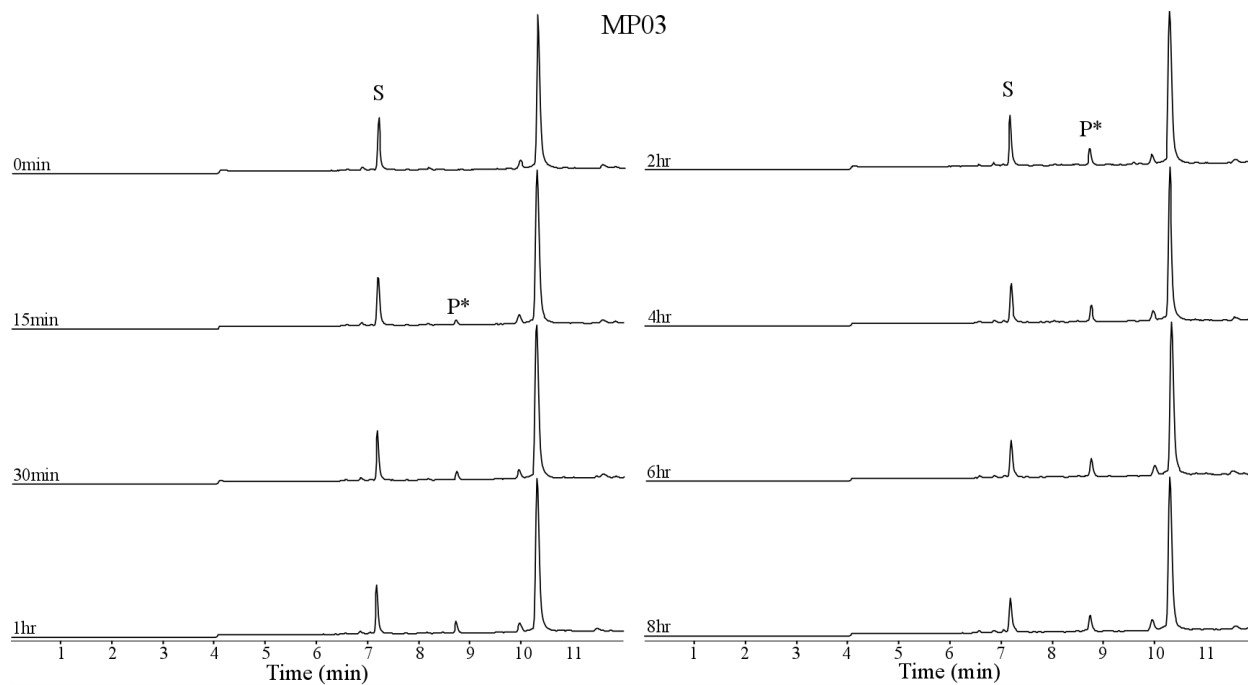


Figure S2. LCMS TIC traces for the kinetic analysis of MP03 (top). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).



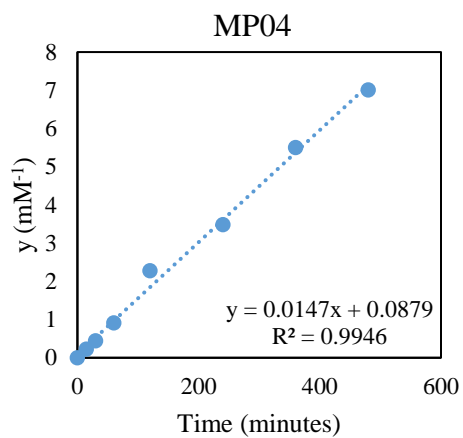
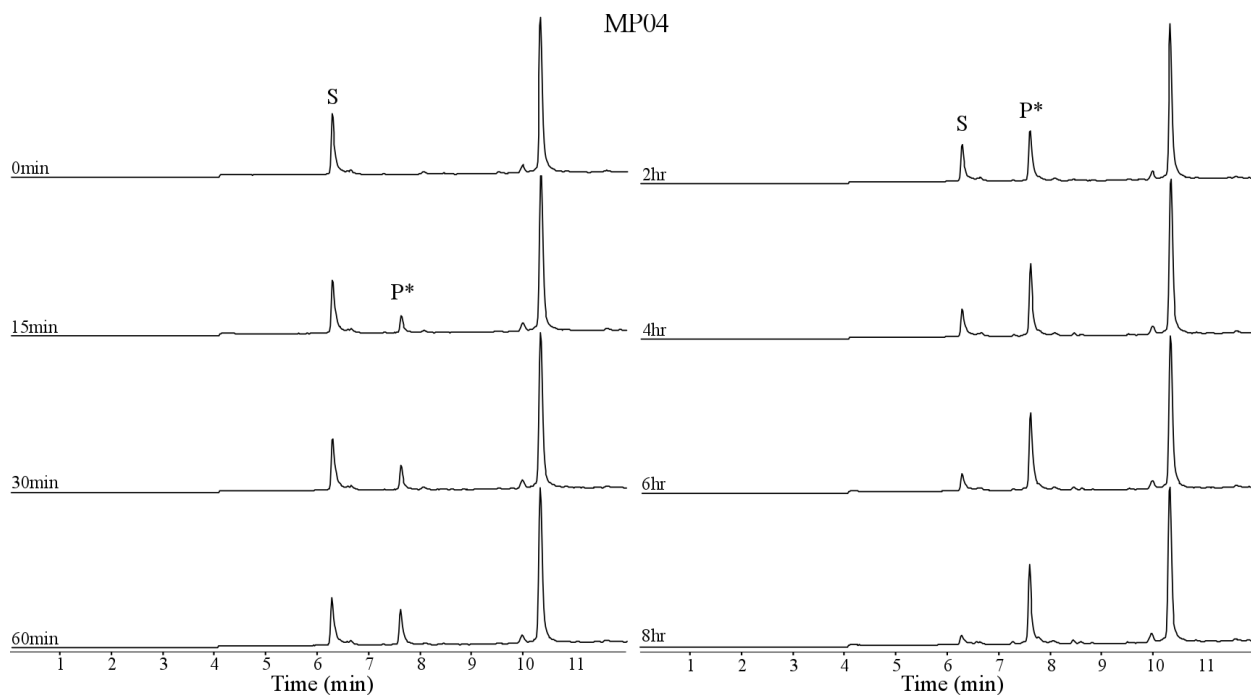


Figure S3. LCMS TIC traces for the kinetic analysis of MP04 (top). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

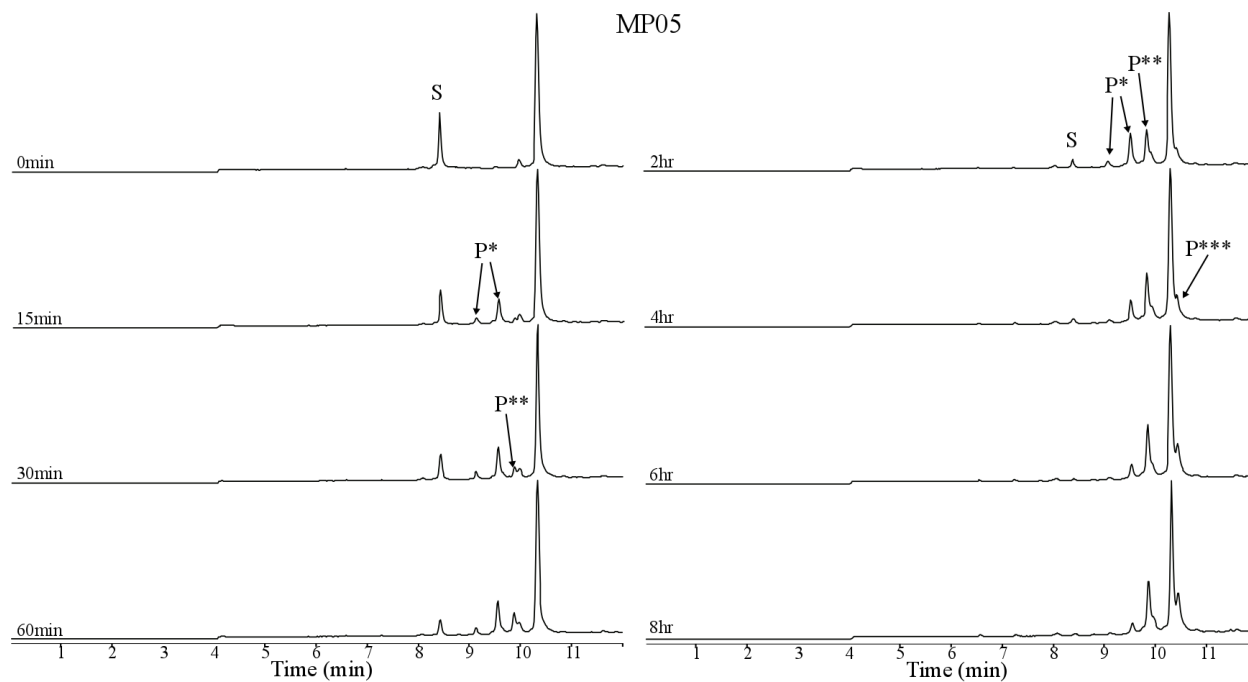


Figure S4. LCMS TIC traces for the labeling of MP05. S – starting material, P\* – singly labeled product, P\*\* – doubly labeled product, P\*\*\* – triply labeled product. Kinetics analysis is not performed to obtain an initial rate on this data set, it is purely used to show the three labeling states. The peaks near 10 minutes (in the 0 min trace and subsequent trace unless noted) are the oxidized CA (small peak) and CA (method 2).

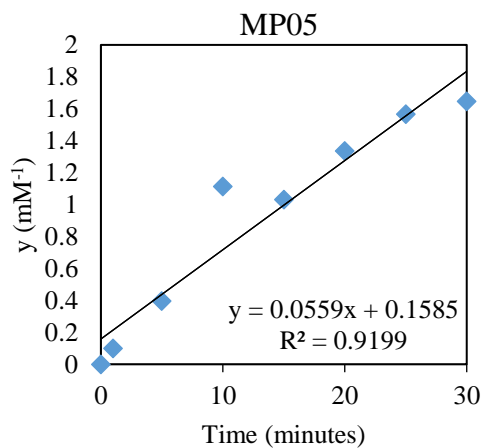
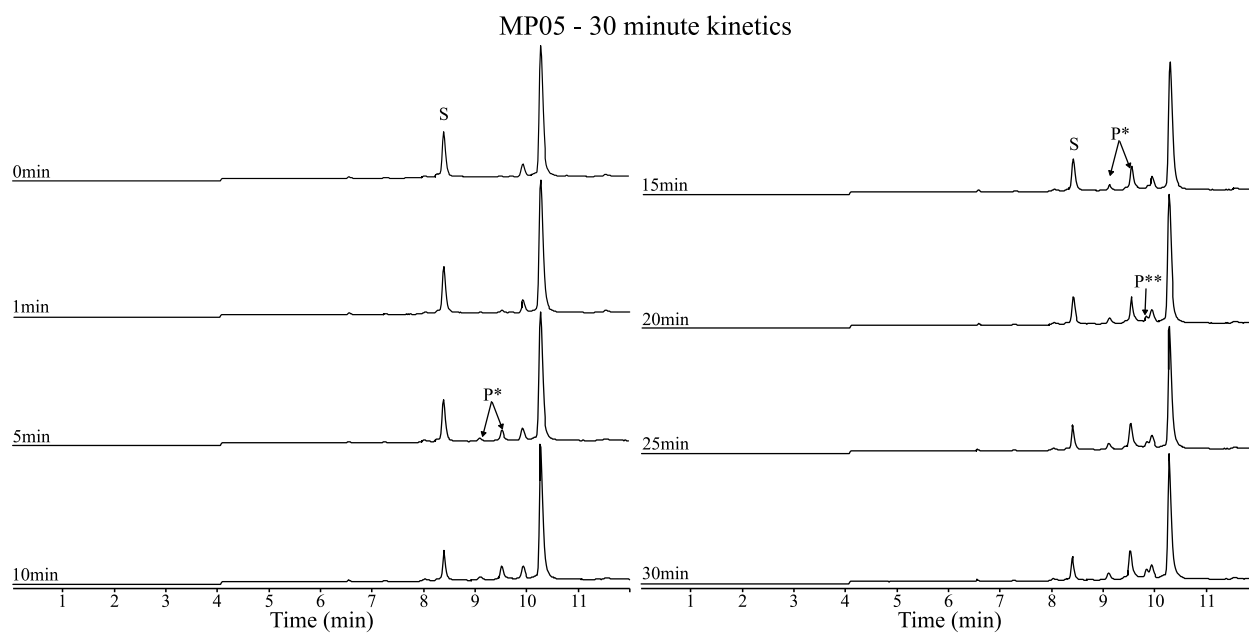


Figure S5. LCMS TIC trace (top) and kinetic analysis (bottom) for the labeling of MP05. S – starting material, P\* – singly labeled product, P\*\* – doubly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes (in the 0min trace and subsequent trace unless noted) are the oxidized CA (small peak) and CA (method 2).

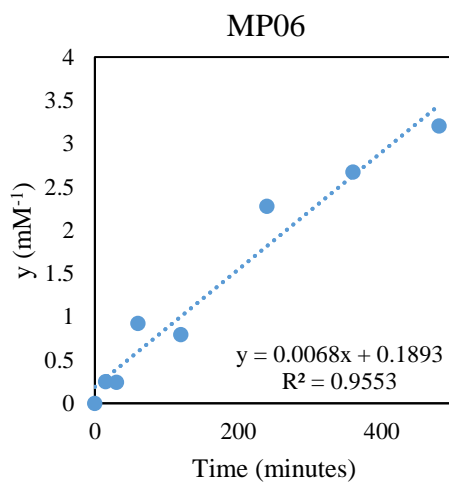
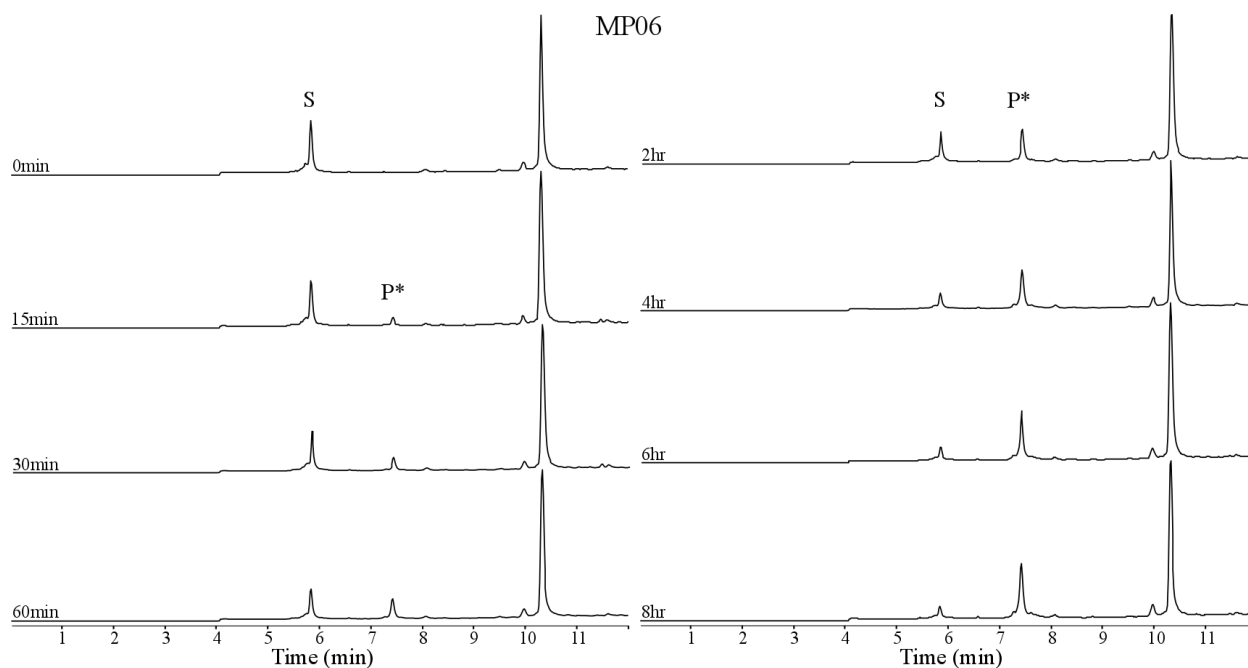


Figure S6. LCMS TIC traces for the kinetic analysis of MP06 (top). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

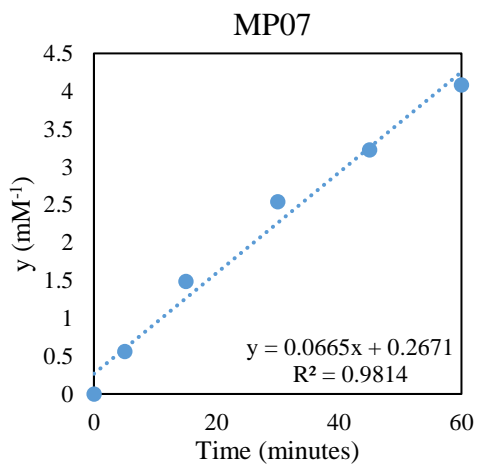
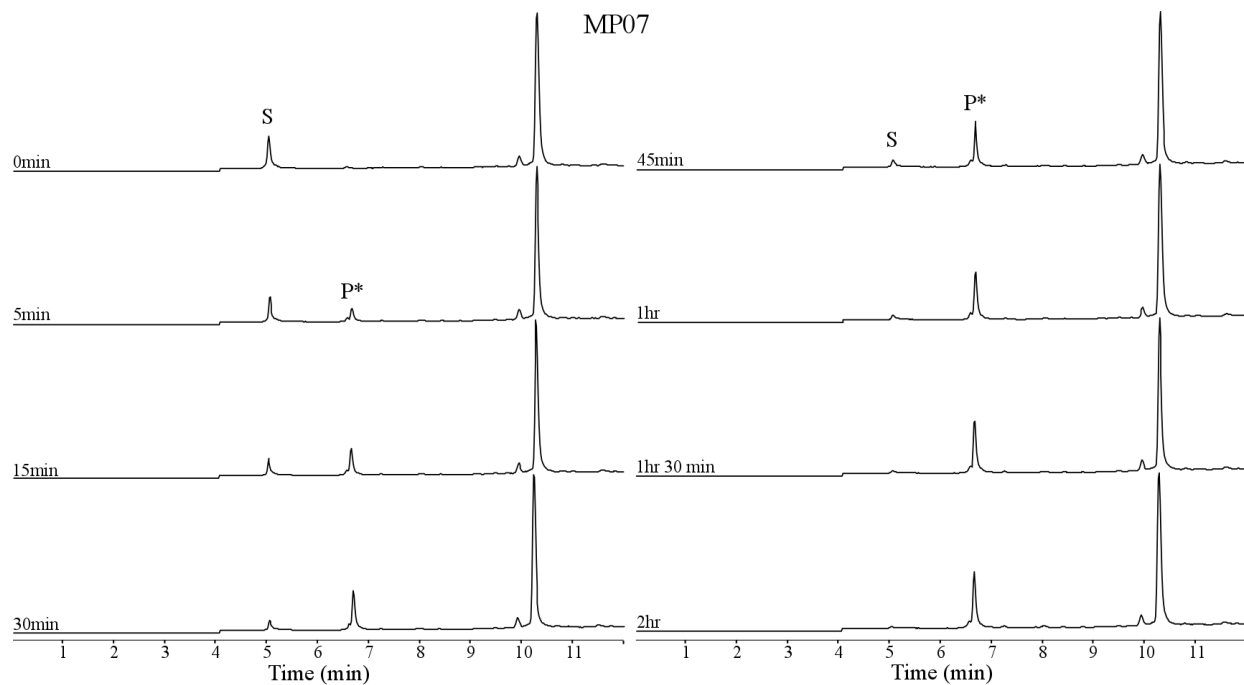


Figure S7. LCMS TIC traces for the kinetic analysis of MP07 (top). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The final two points were outside the linear range of the instrument and were not used (not integration during automatic integration). The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

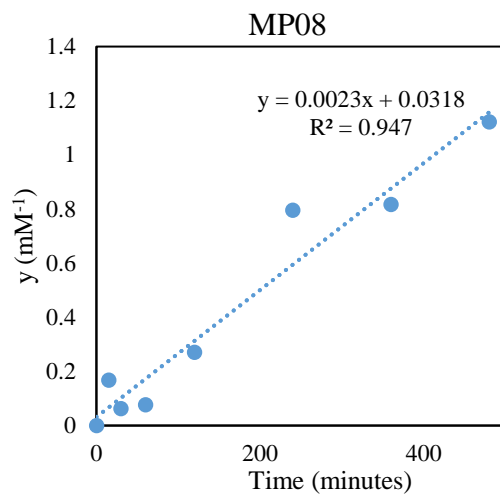
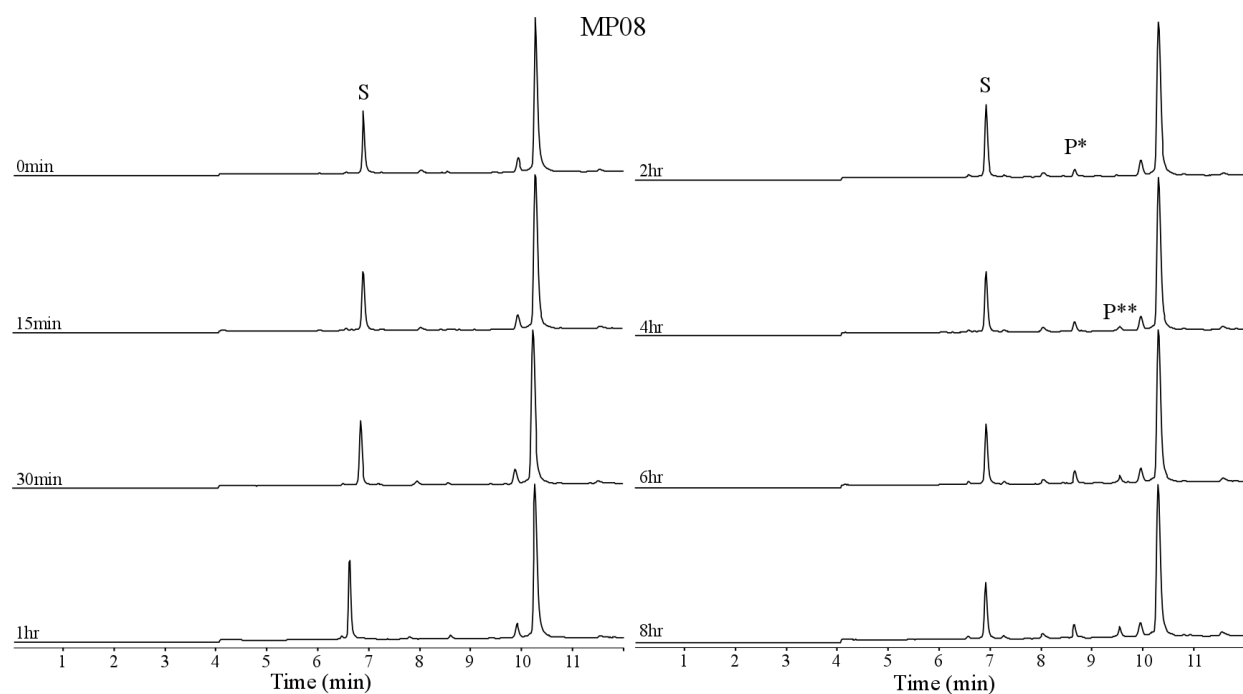


Figure S8. LCMS TIC traces for the kinetic analysis of MP08 (top). S – starting material, P\* – singly labeled product, P\*\* – doubly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

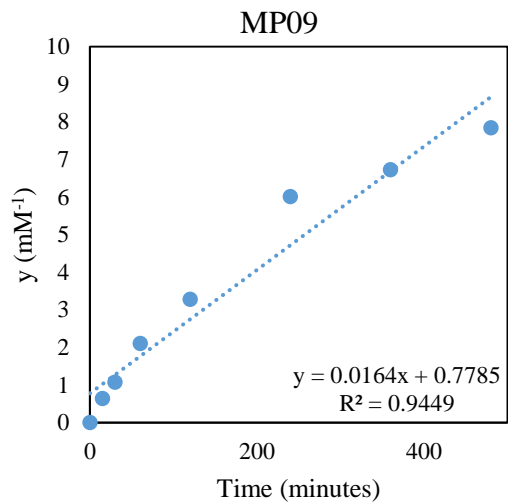
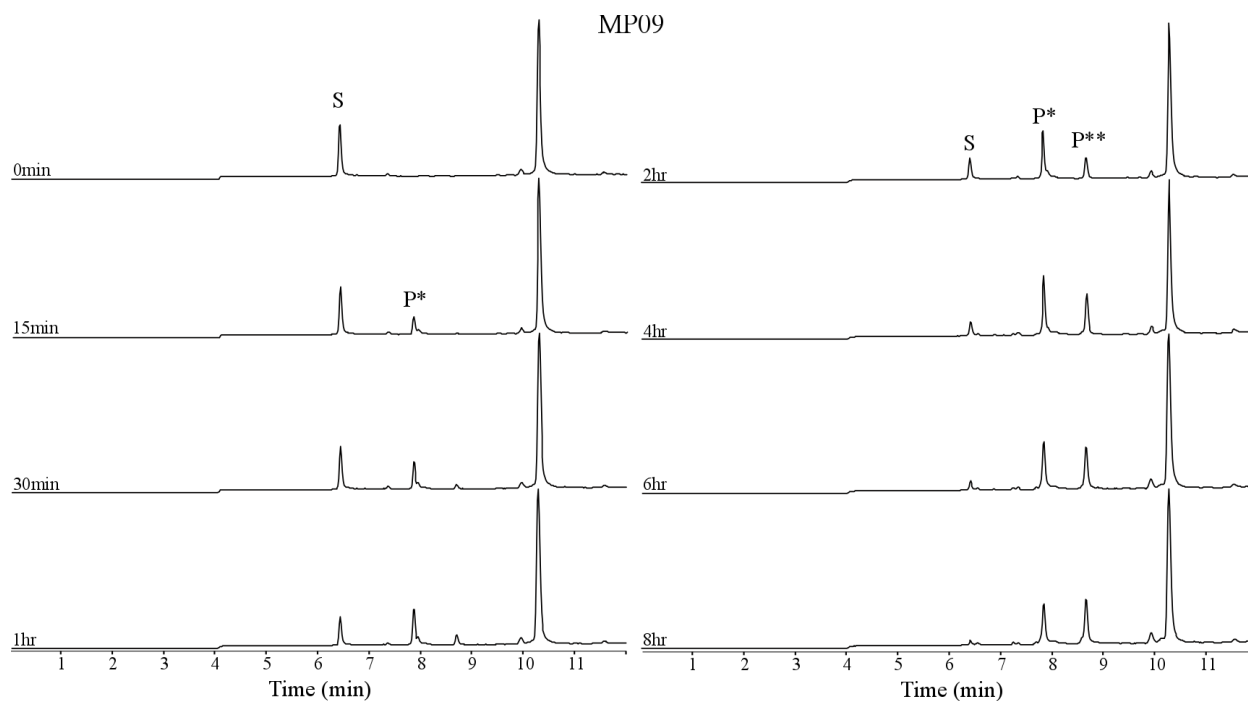


Figure S9. LCMS TIC traces for the kinetic analysis of MP09 (top). S – starting material, P\* – singly labeled product, P\*\* – doubly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).

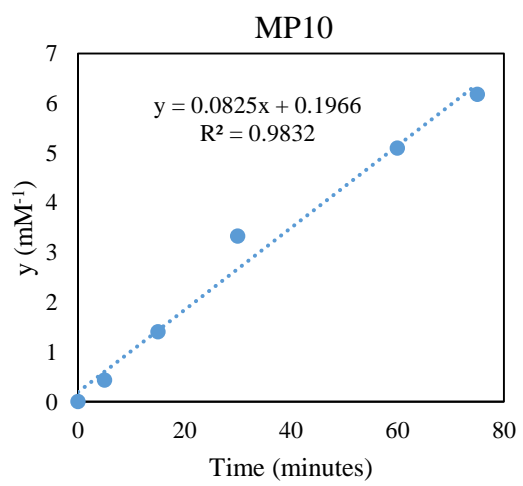
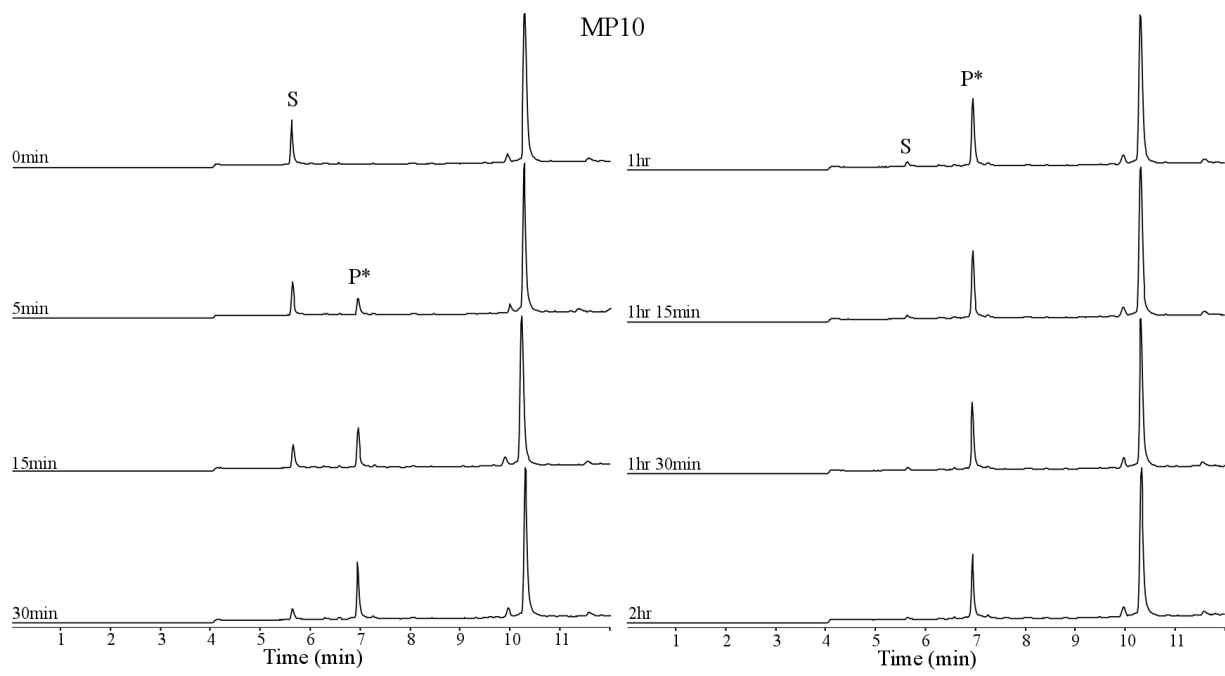


Figure S10. LCMS TIC traces for the kinetic analysis of MP10 (top). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant. The final point was outside the linear range of the instrument and was thus not used (not integration during automatic integration). The peaks near 10 minutes are the oxidized CA (small peak) and CA (method 2).



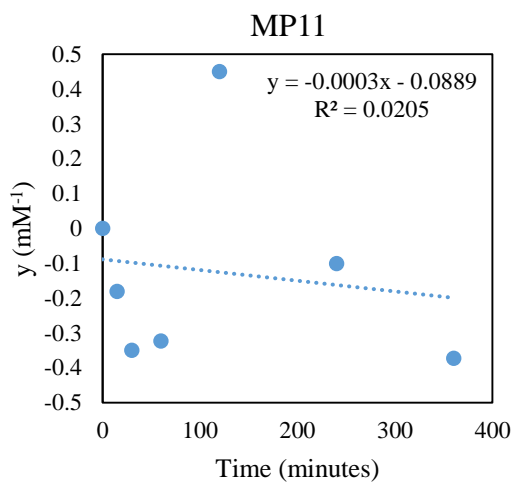
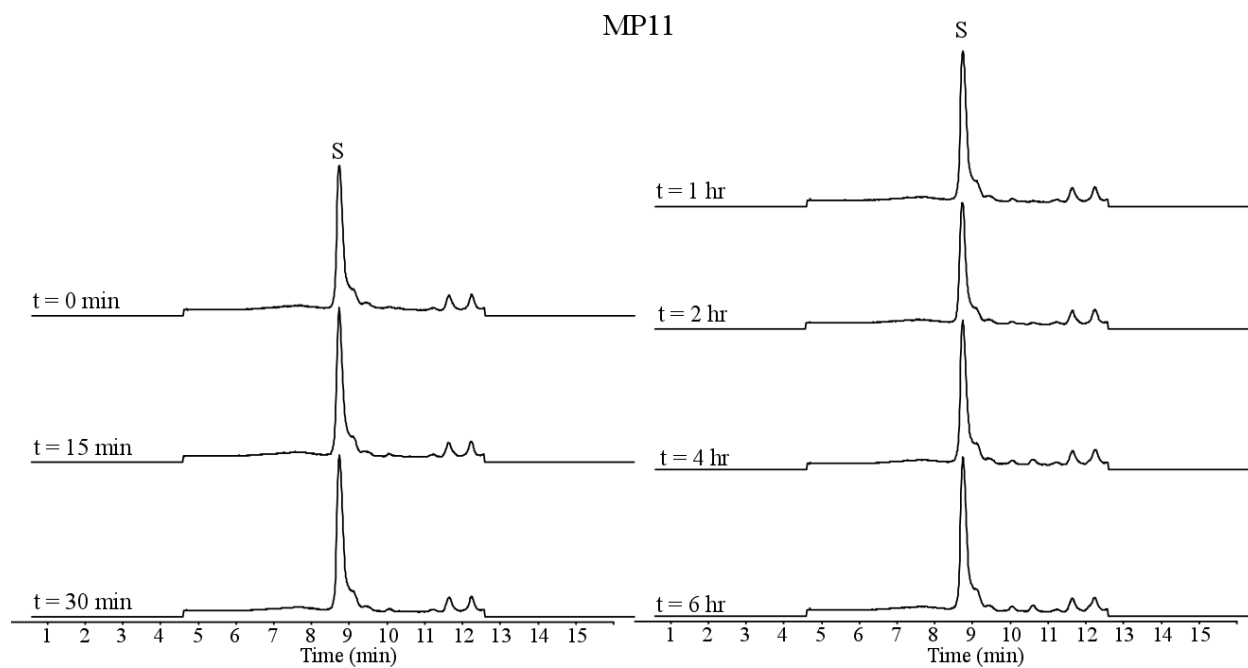


Figure S11. LCMS TIC traces for the kinetic analysis of MP11 (top, method 3 with the mass spectrometer off at 12.5 minutes). S – starting material. Below is the plot of integrated TIC peak area to determine a second order rate constant.

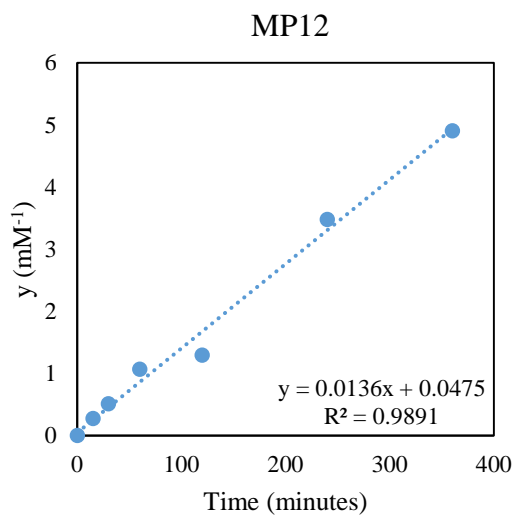
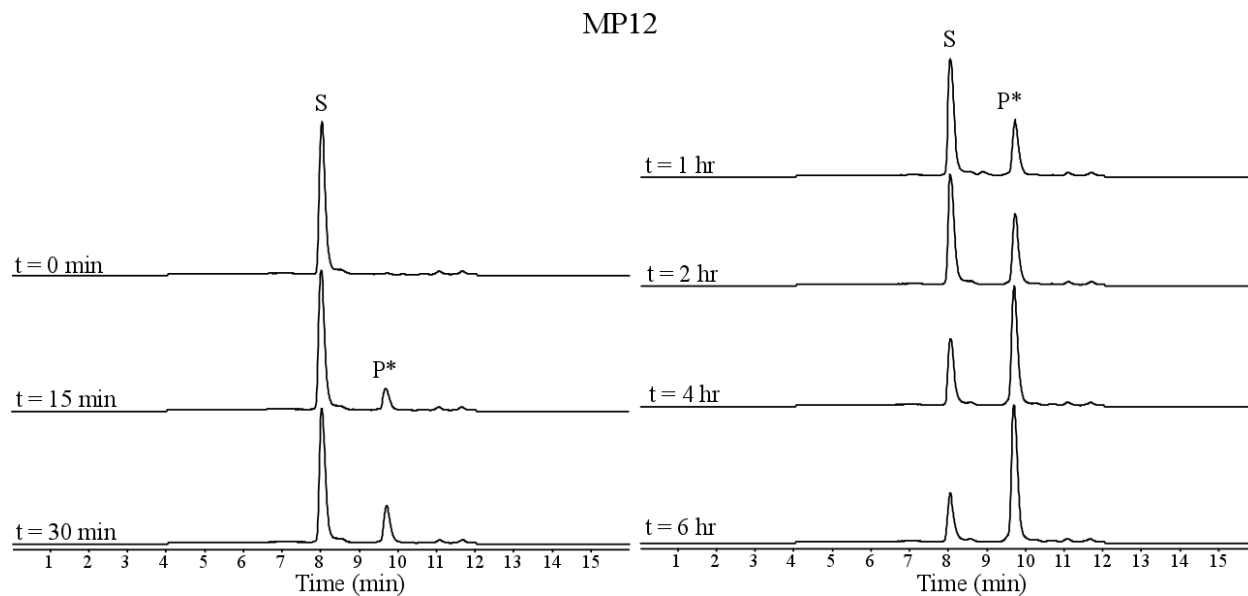


Figure S12. LCMS TIC traces for the kinetic analysis of MP12 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

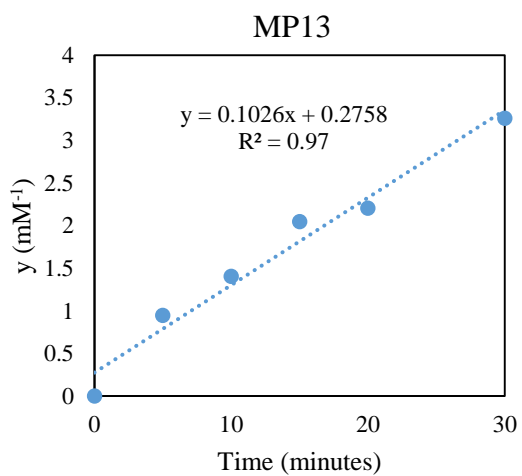
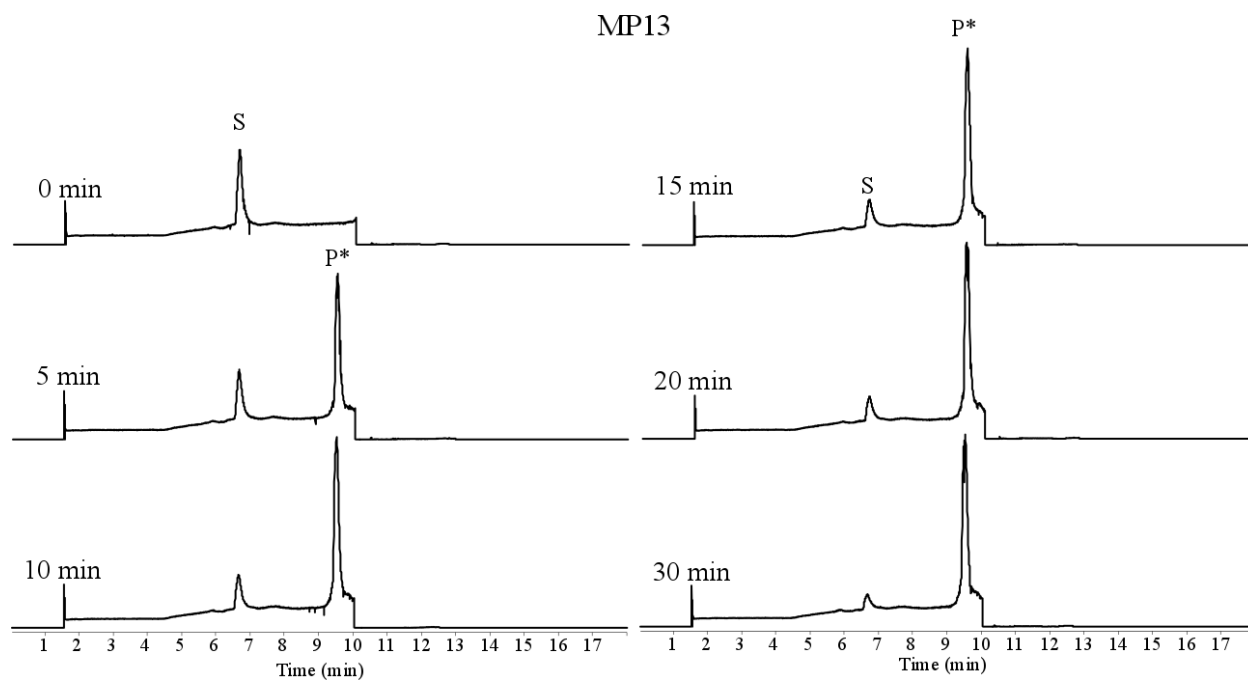


Figure S13. LCMS TIC traces for the kinetic analysis of MP13 (top, method 1). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

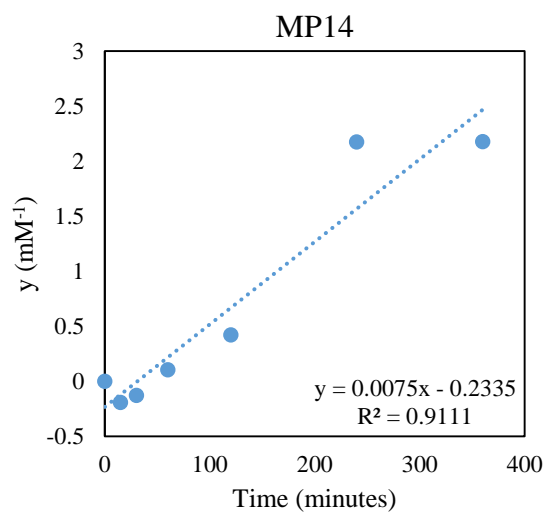
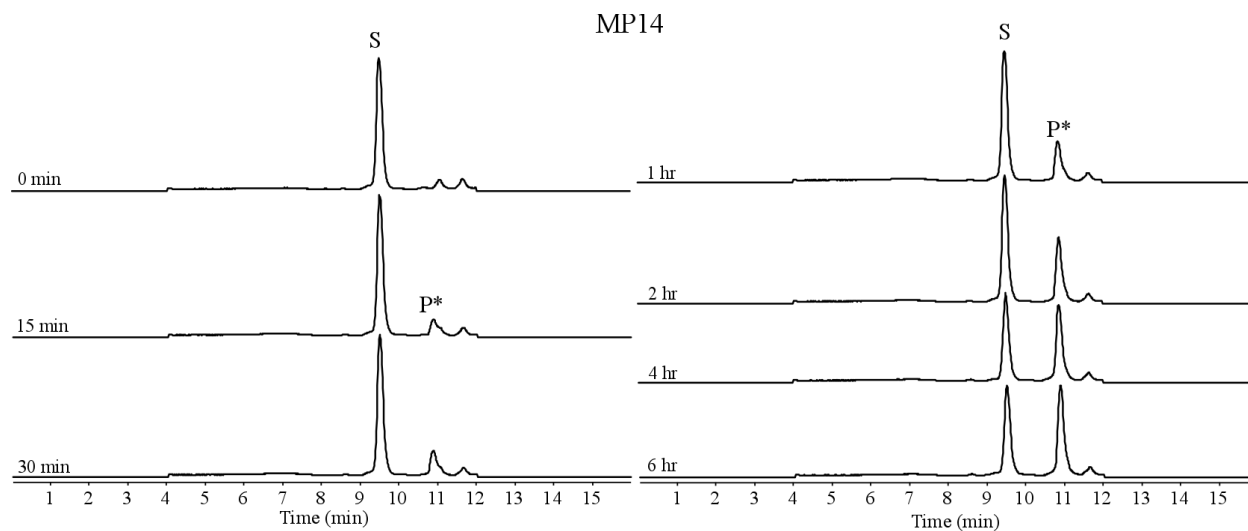


Figure S14. LCMS TIC traces for the kinetic analysis of MP14 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

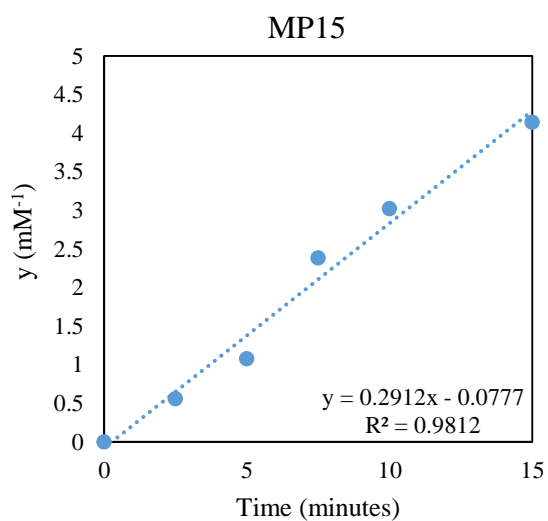
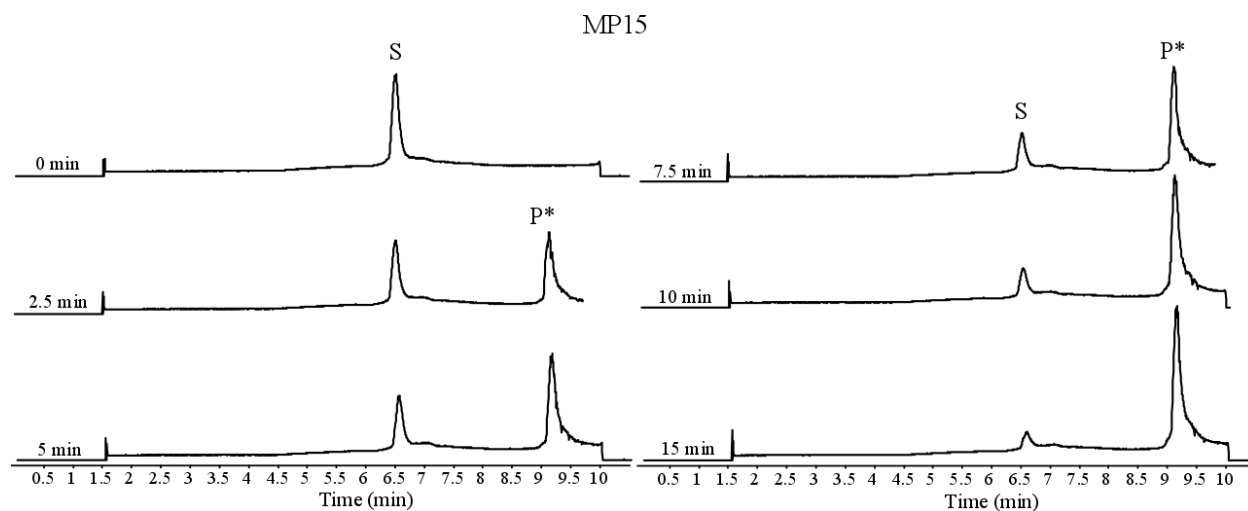


Figure S15. LCMS TIC traces for the kinetic analysis of MP15 (top, method 1). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

MP16

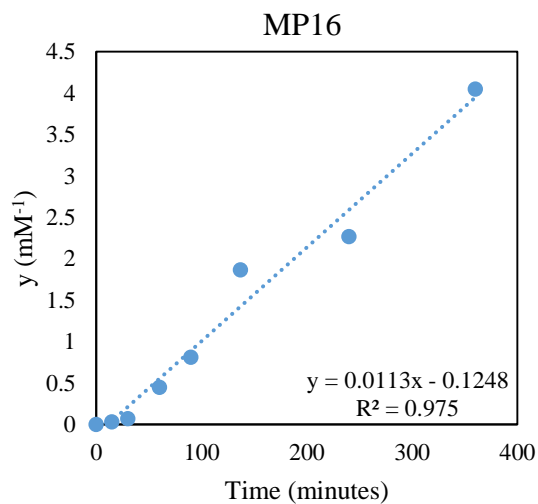
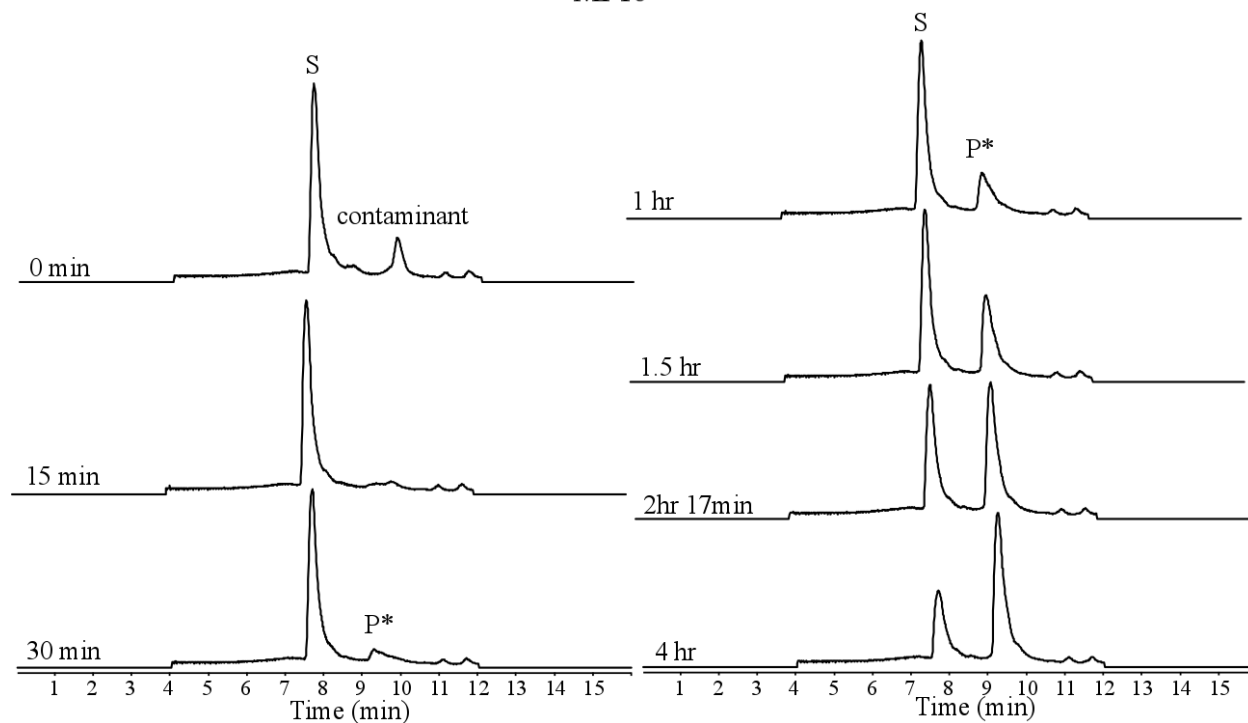


Figure S16. LCMS TIC traces for the kinetic analysis of MP16 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

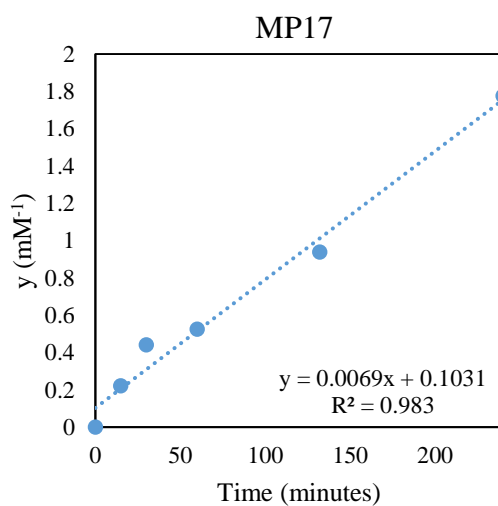
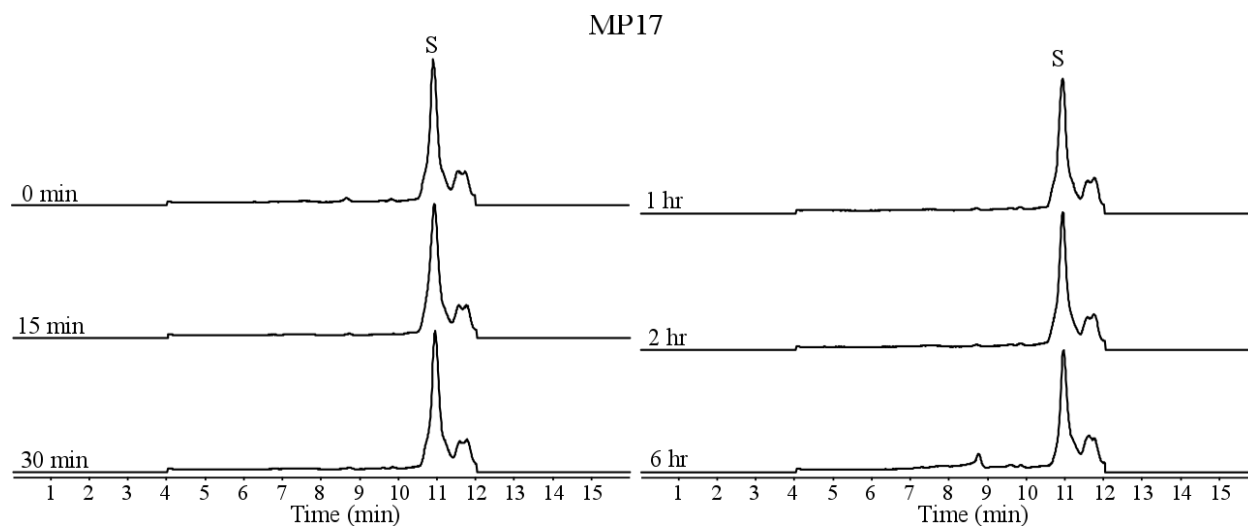


Figure S17. LCMS TIC traces for the kinetic analysis of MP17 (top, method 3). S – starting material. Below is the plot of integrated extracted ion chromatogram peak area (for the most abundant charge state of the starting material) to determine a second order rate constant.

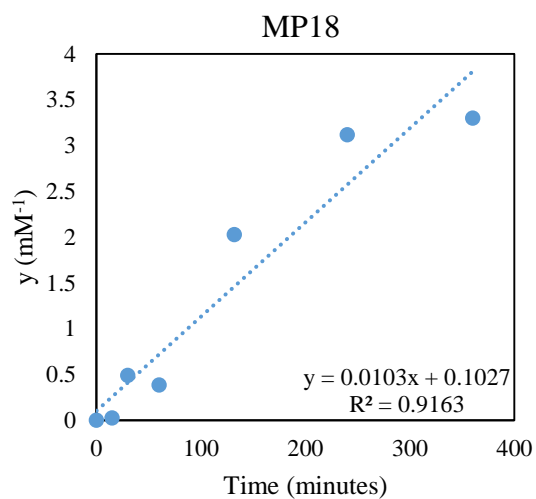
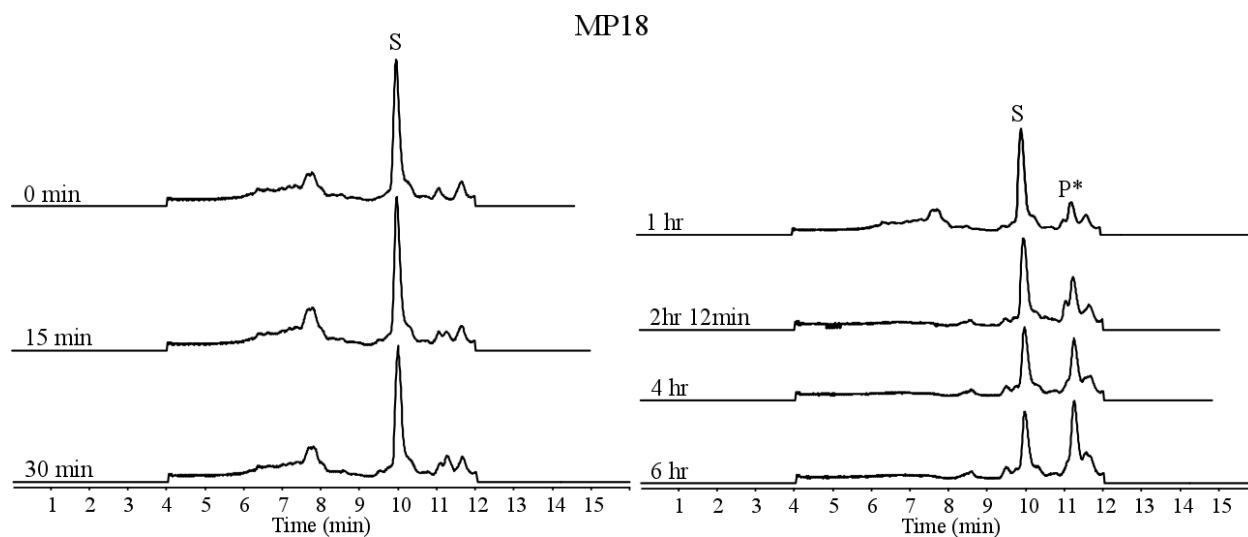


Figure S18. LCMS TIC traces for the kinetic analysis of MP18 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.



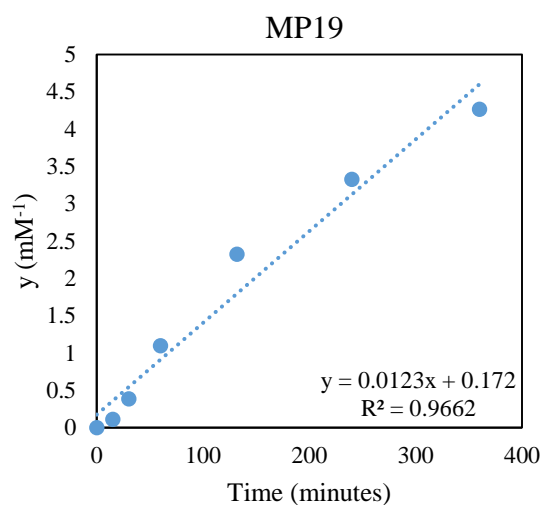
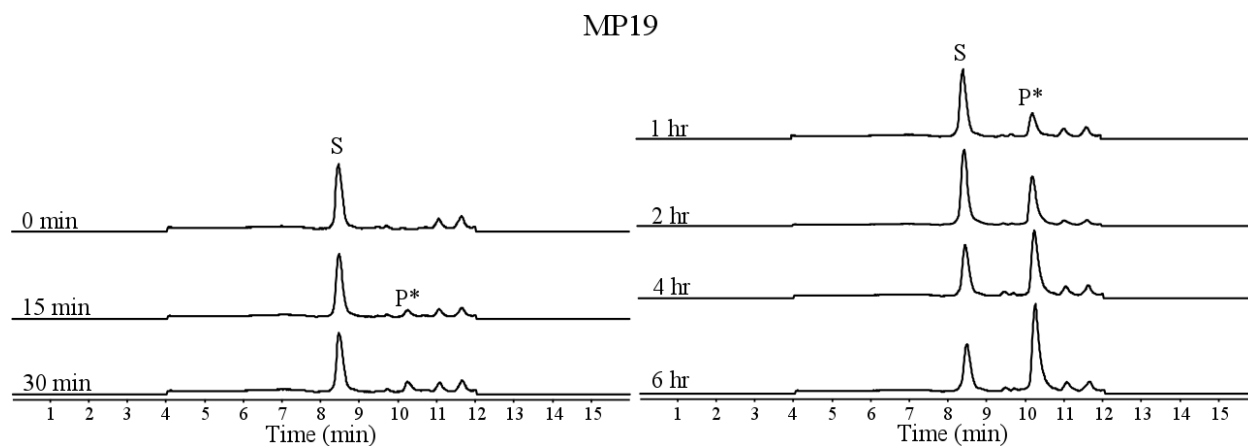


Figure S19. LCMS TIC traces for the kinetic analysis of MP19 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

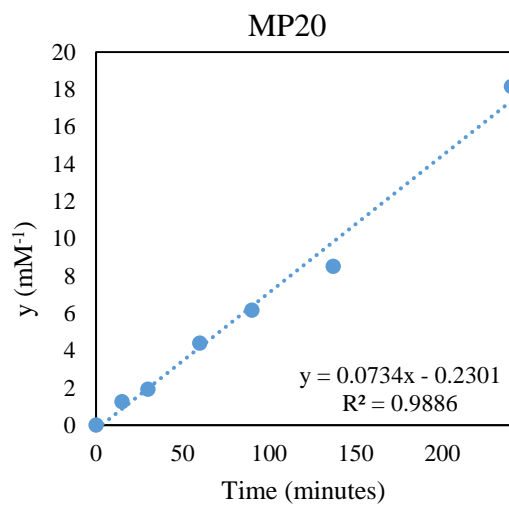
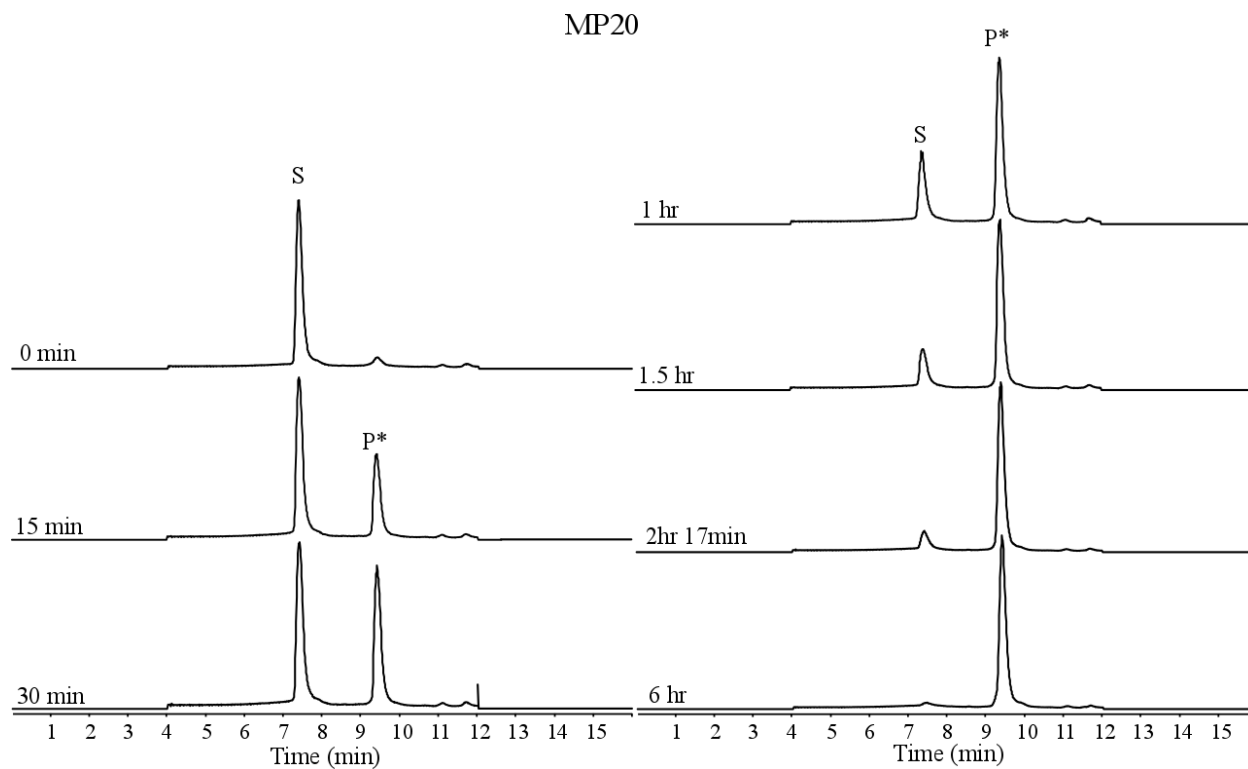


Figure S20. LCMS TIC traces for the kinetic analysis of MP20 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

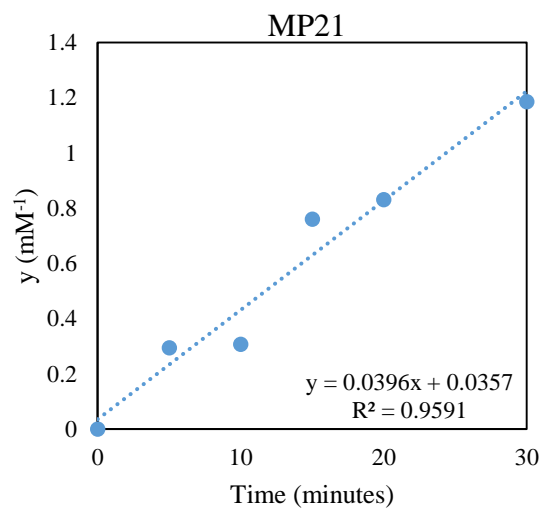
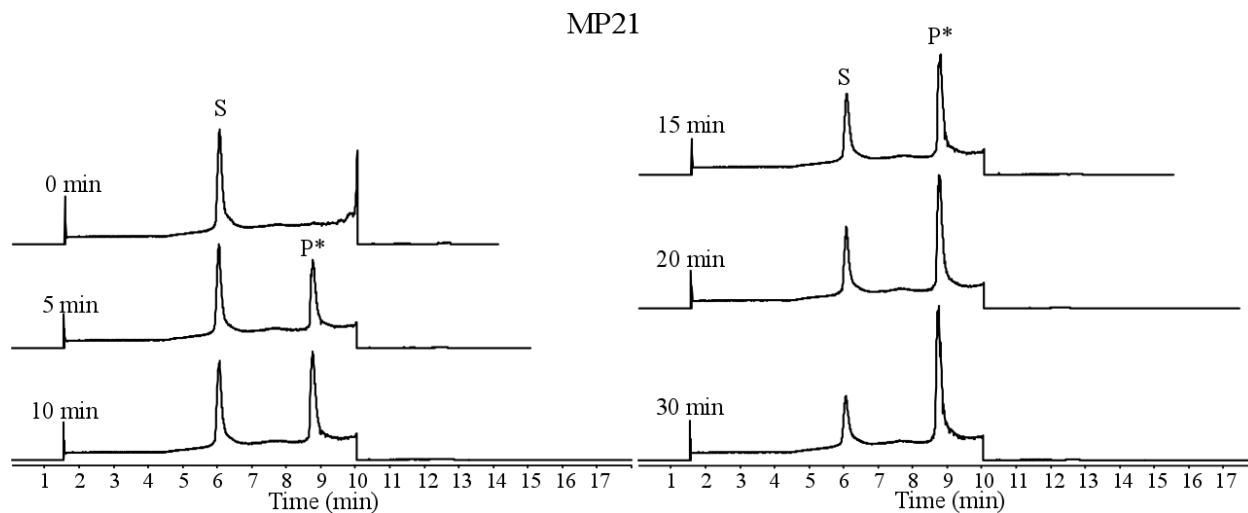


Figure S21. LCMS TIC traces for the kinetic analysis of MP21 (top, method 1). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

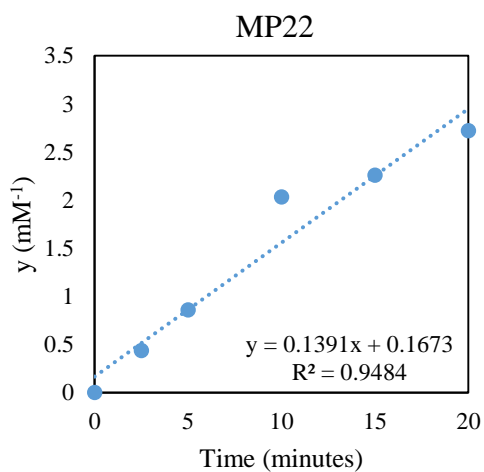
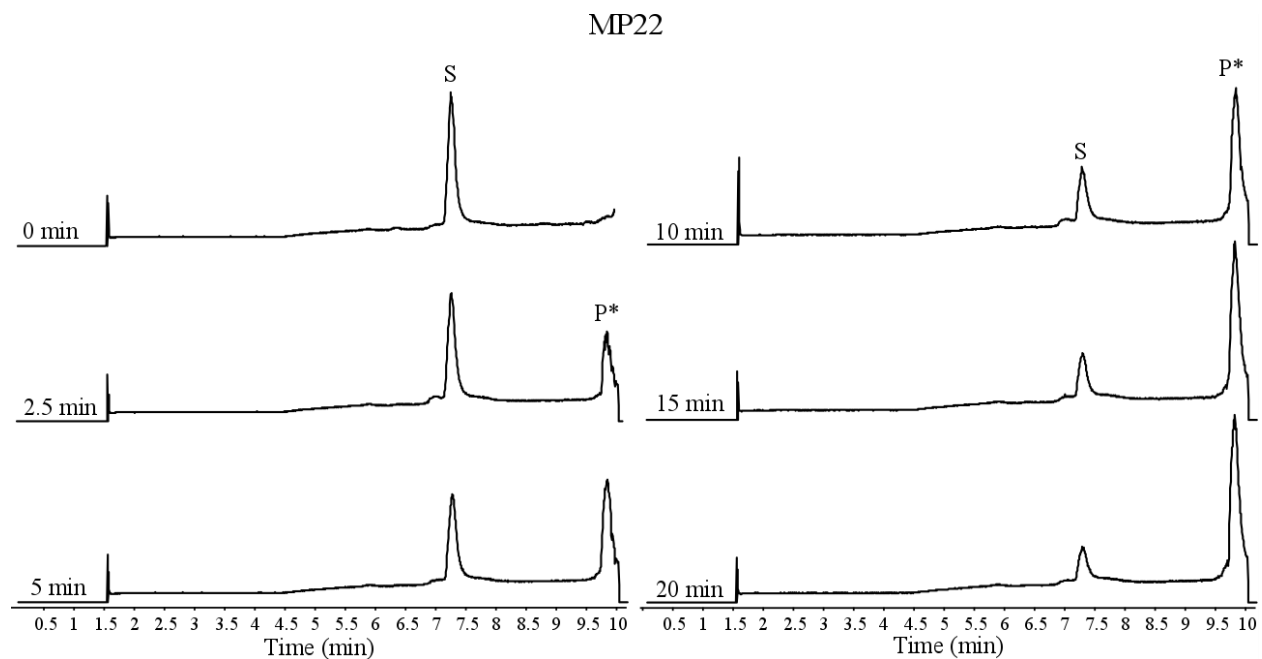


Figure S22. LCMS TIC traces for the kinetic analysis of MP22 (top, method 1). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

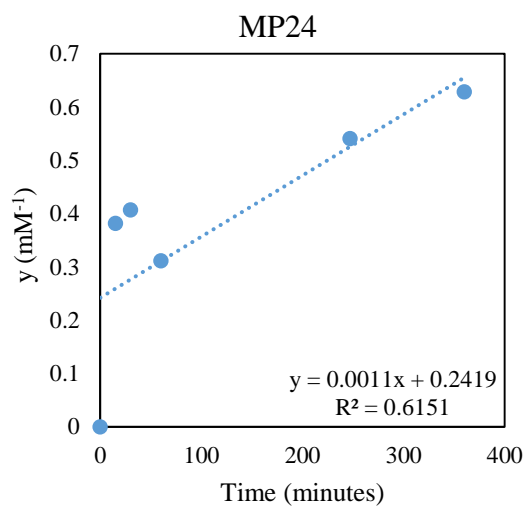
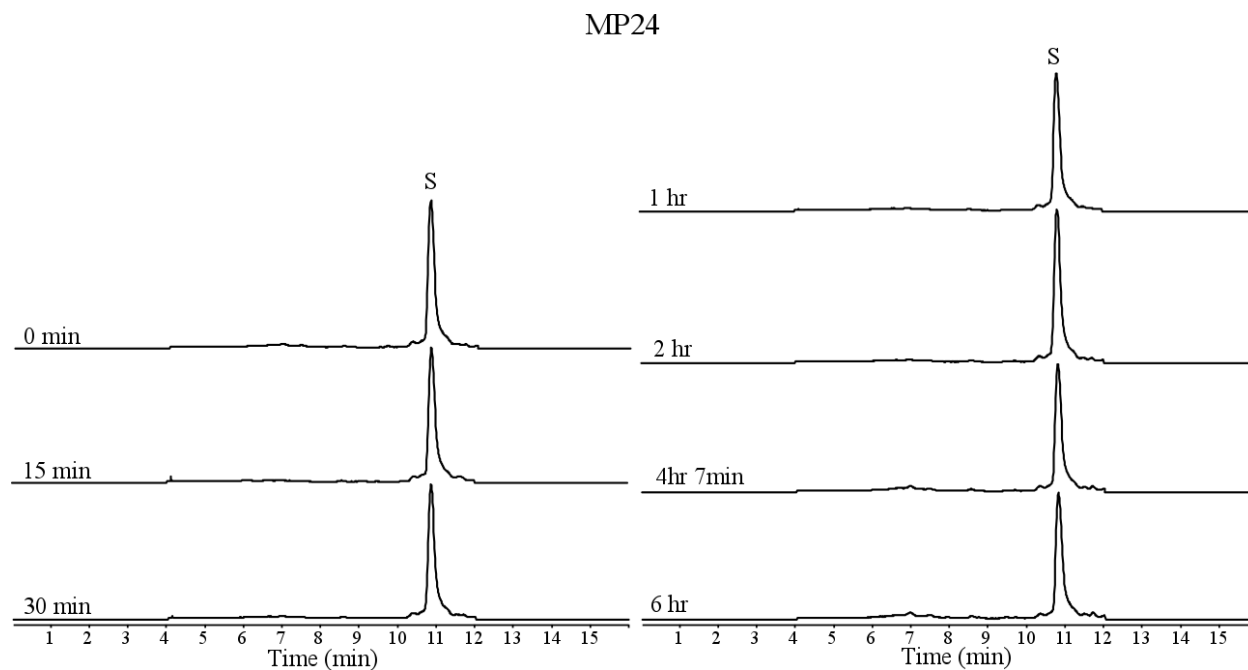


Figure S23. LCMS TIC traces for the kinetic analysis of MP24 (top, method 3). S – starting material. Below is the plot of integrated TIC peak area to determine a second order rate constant.

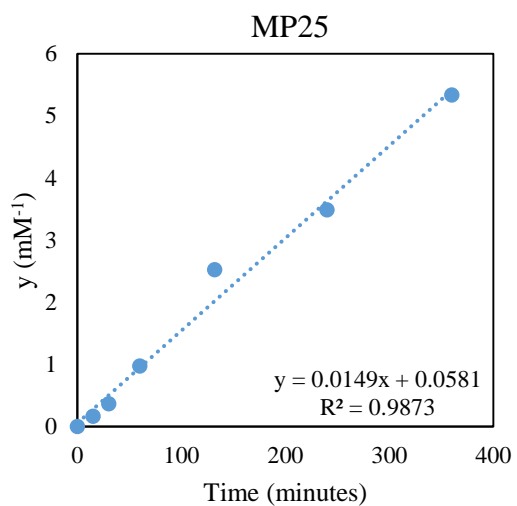
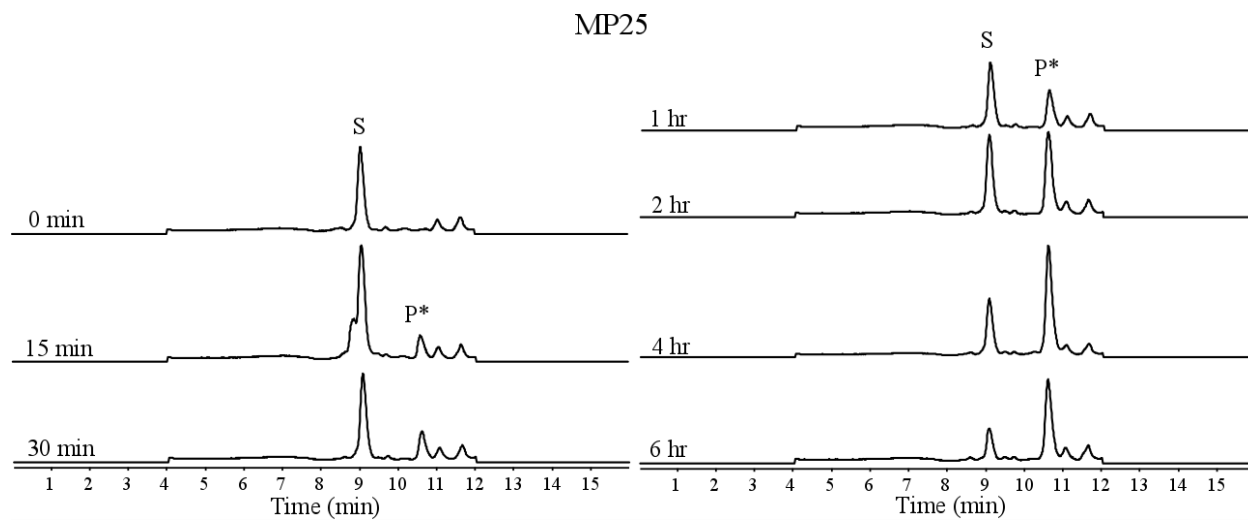


Figure S24. LCMS TIC traces for the kinetic analysis of MP25 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

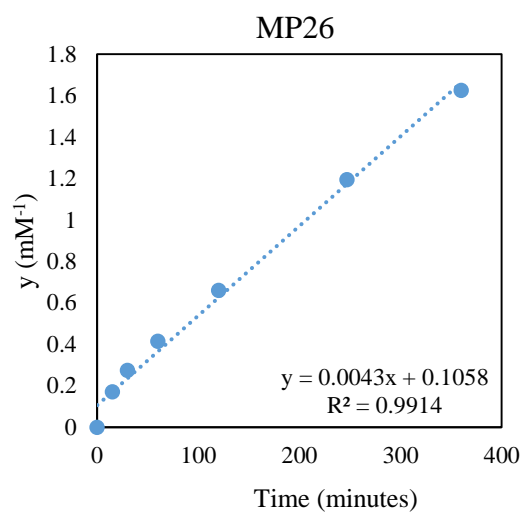
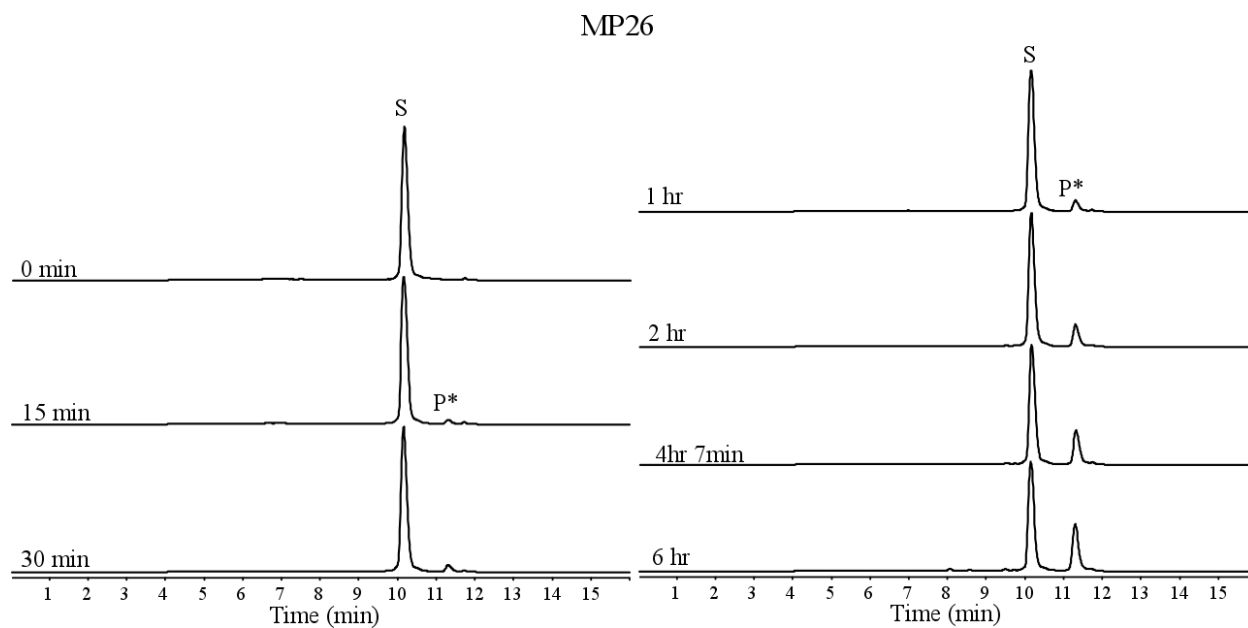


Figure S25. LCMS TIC traces for the kinetic analysis of MP26 (top, method 3). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.

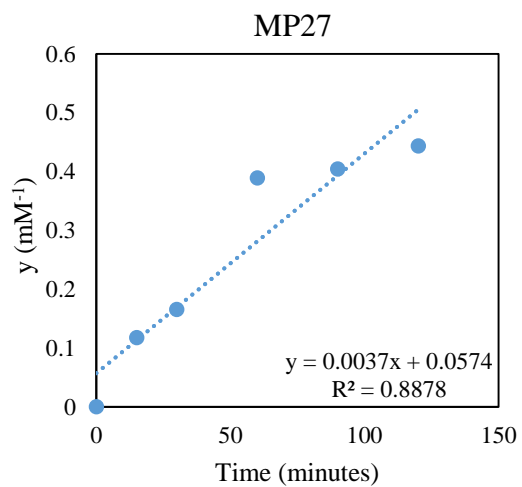
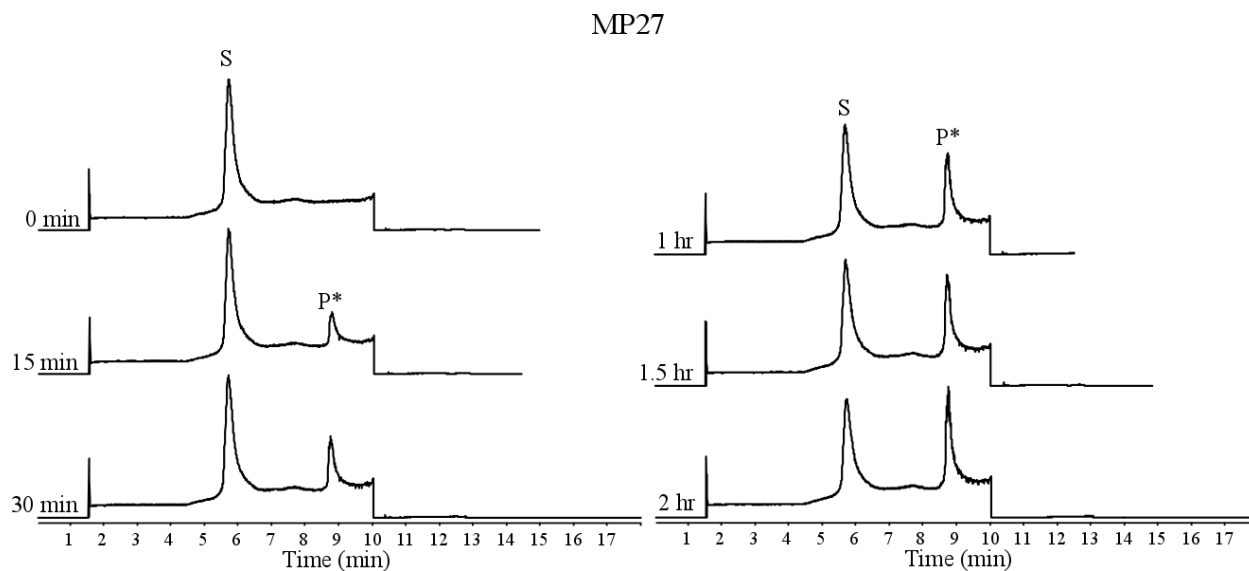


Figure S26. LCMS TIC traces for the kinetic analysis of MP27 (top, method 1). S – starting material, P\* – singly labeled product. Below is the plot of integrated TIC peak area to determine a second order rate constant.



### 3.3. CD reaction analysis MP04

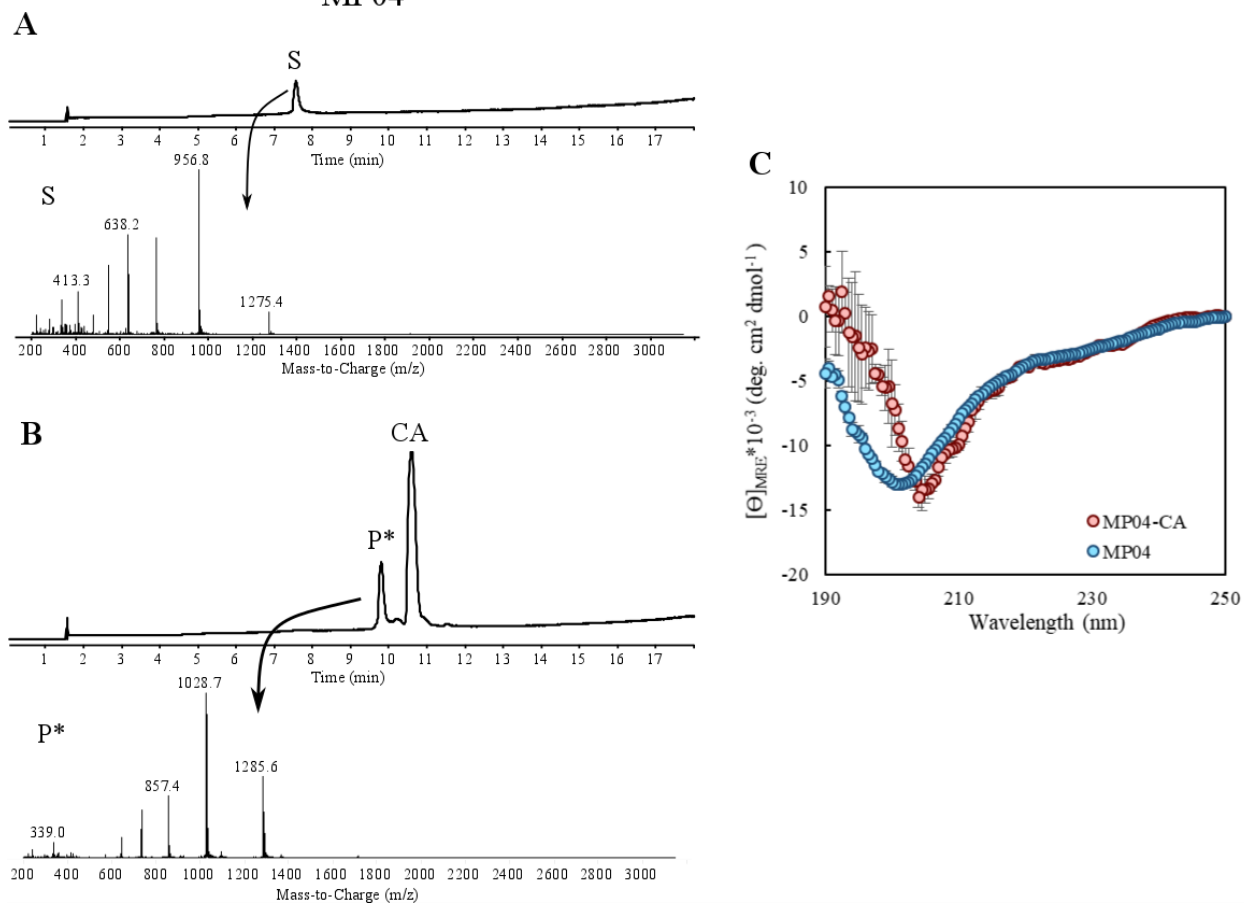


Figure S27. Characterization of MP04 and MP04-CA. (A) LCMS TIC trace for MP04 (50 μM) with mass spectrum and (B) following a reaction between MP04 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP04 samples.

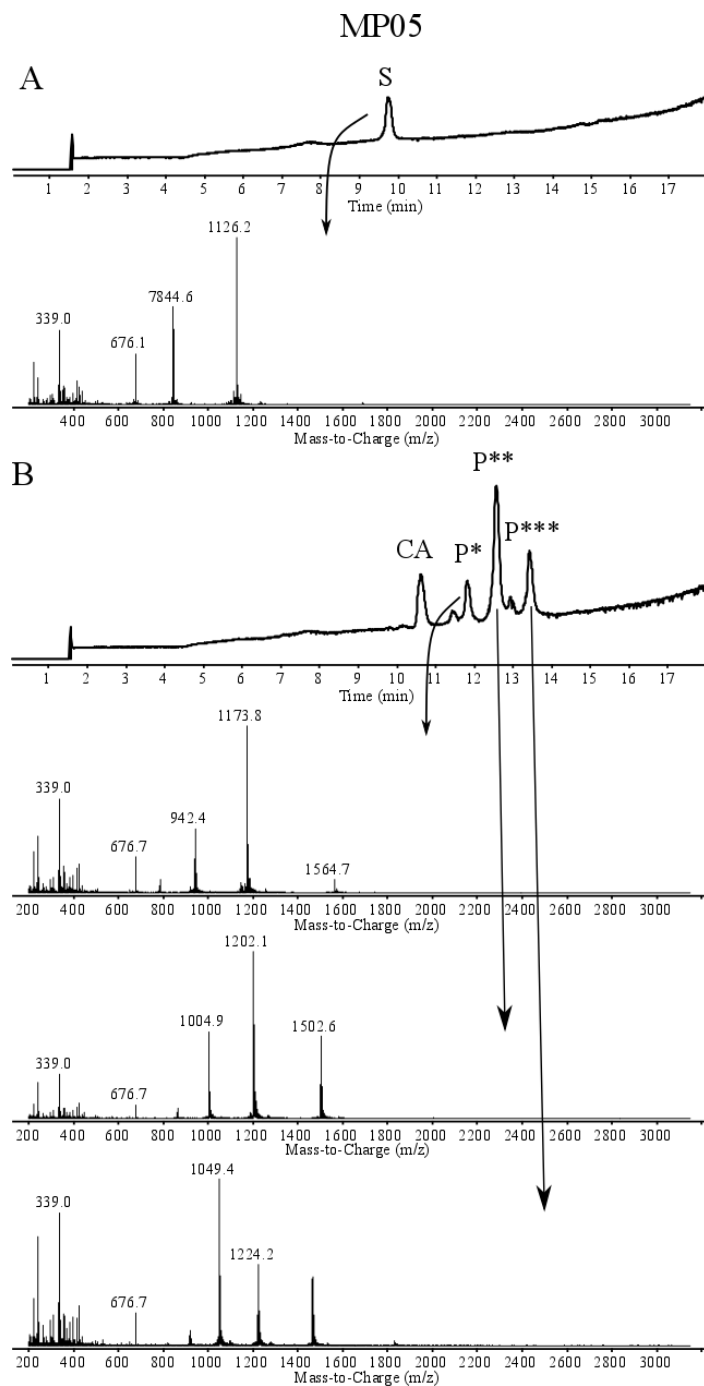


Figure S28. Characterization of MP05 and MP05-CA. (A) LCMS TIC trace for MP05 (50  $\mu$ M) with mass spectrum and (B) following a reaction between MP05 (50  $\mu$ M) and CA (250  $\mu$ M). Method 1 was used without turning the mass spectrometer off at 10 minutes. Both samples were not soluble in phosphate buffer so CD measurements were not obtained.

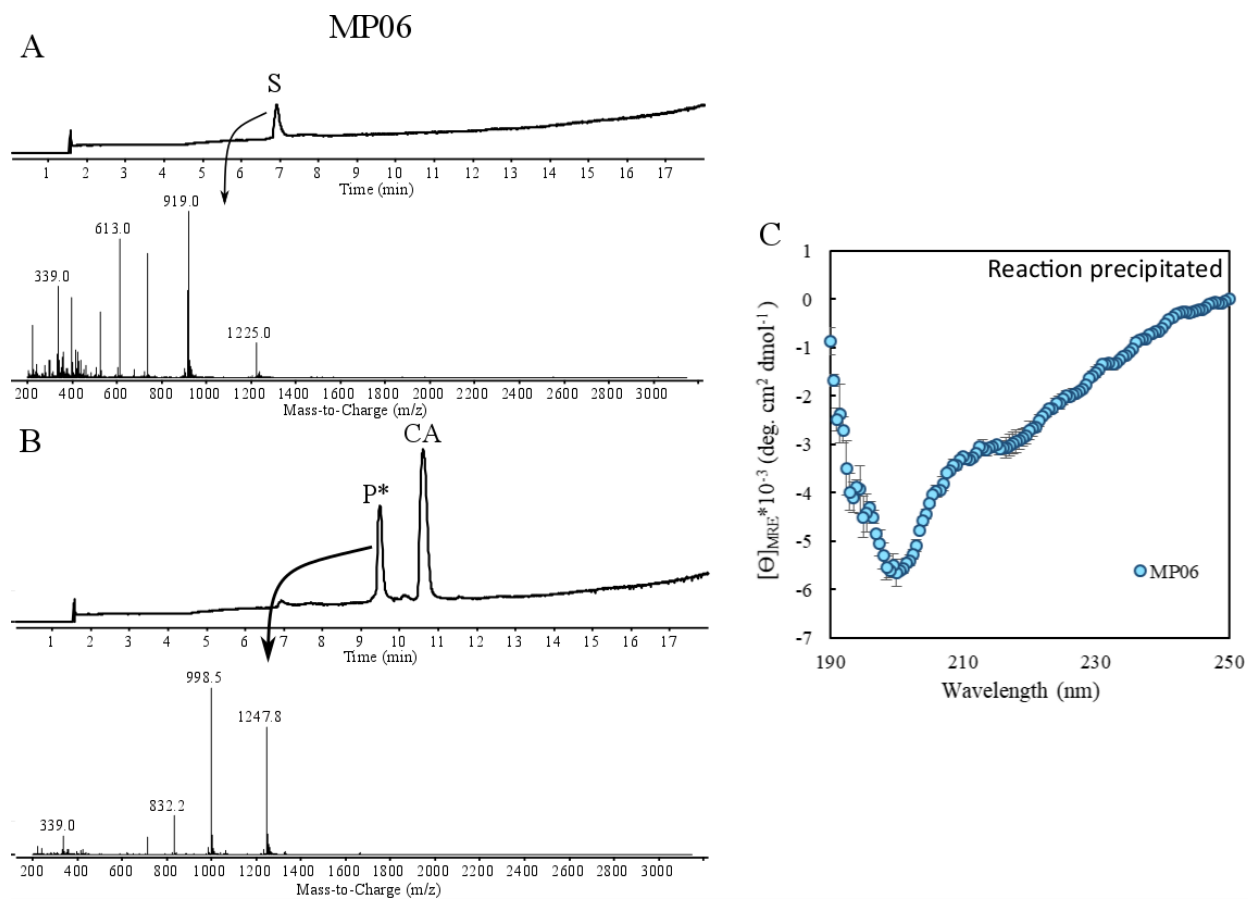


Figure S29. Characterization of MP06 and MP06-CA. (A) LCMS TIC trace for MP06 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP06 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP06, the CA reacted version was not soluble.

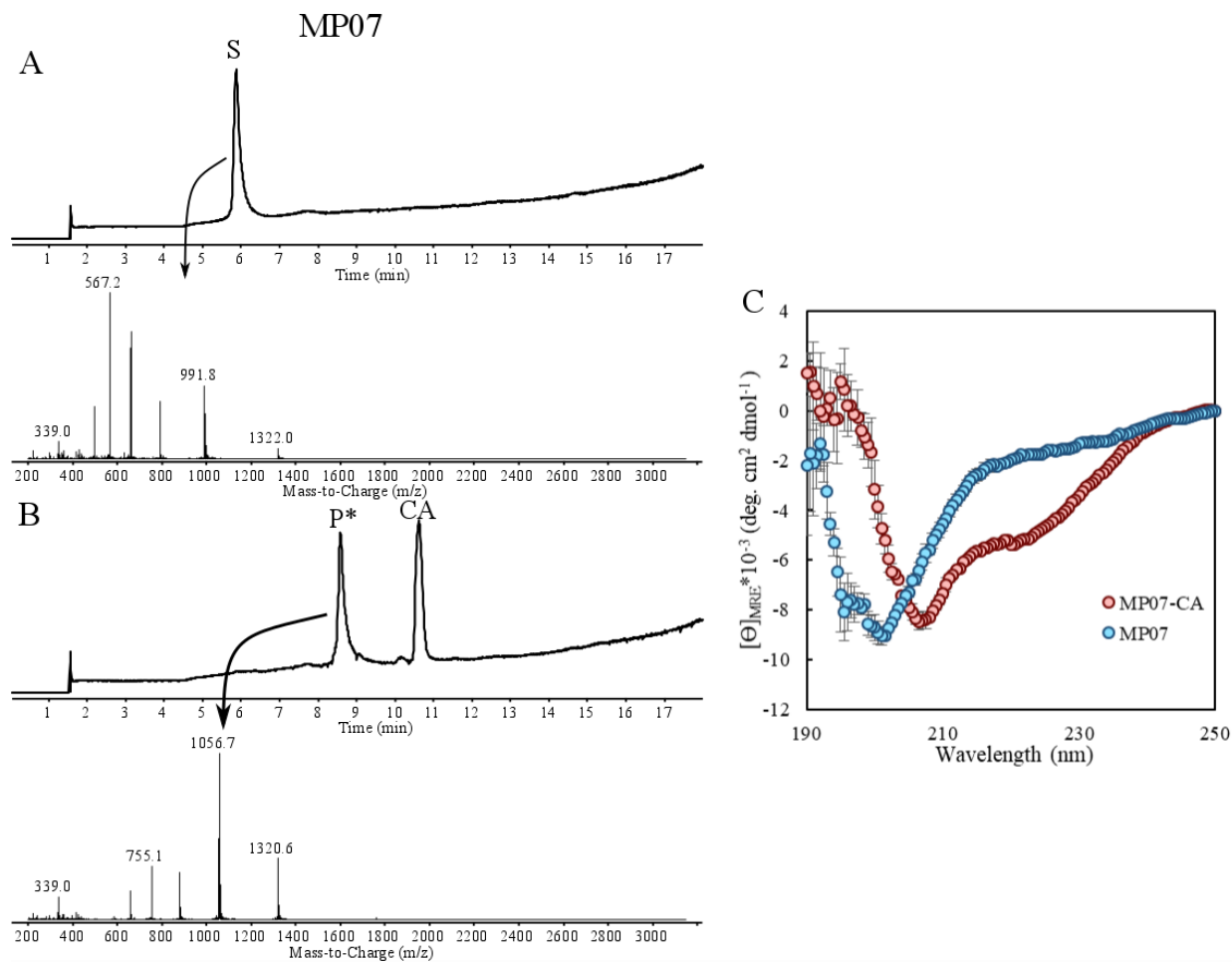


Figure S30. Characterization of MP07 and MP07-CA. (A) LCMS TIC trace for MP07 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP07 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP07 samples.

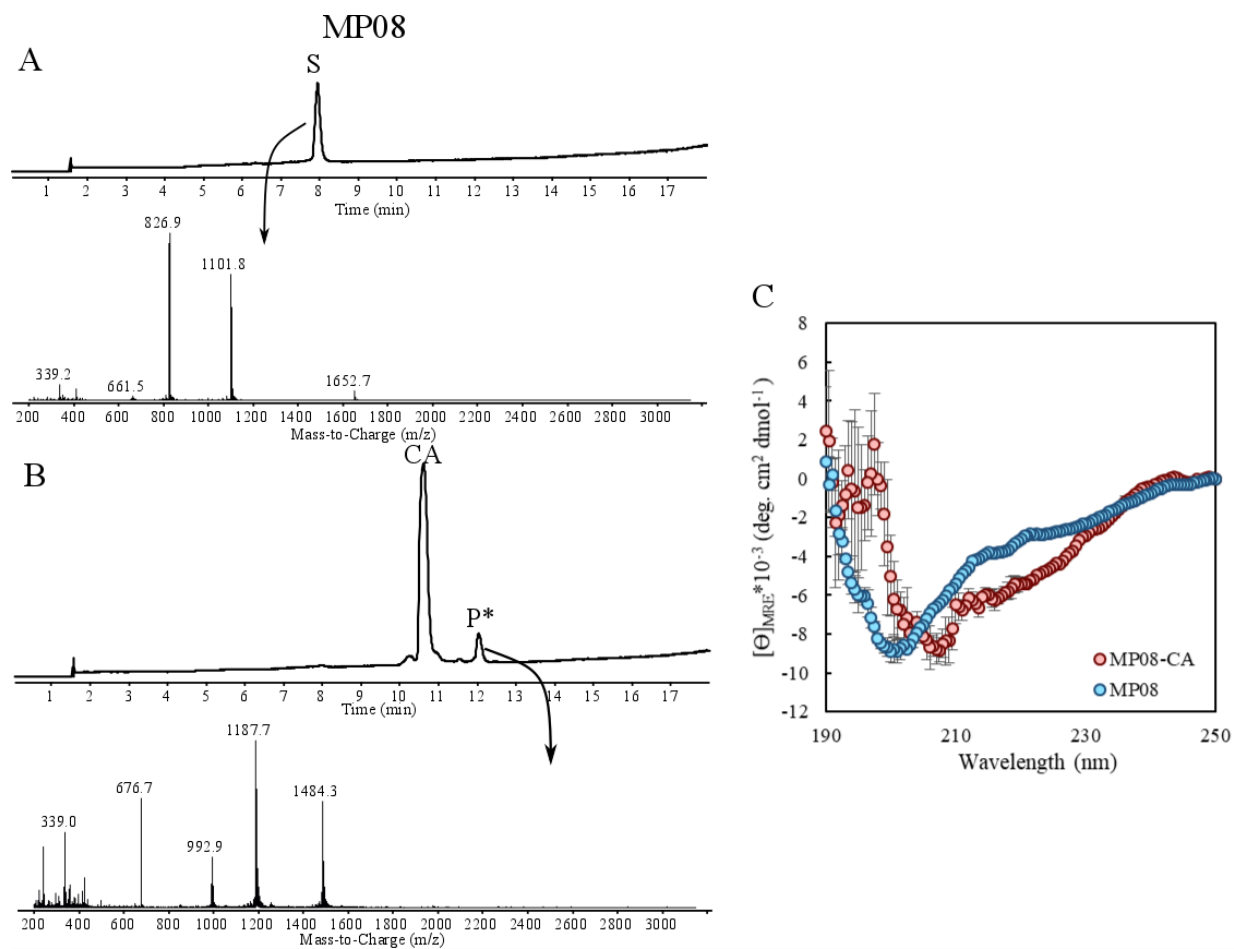


Figure S31. Characterization of MP08 and MP08-CA. (A) LCMS TIC trace for MP08 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP08 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP08 samples.

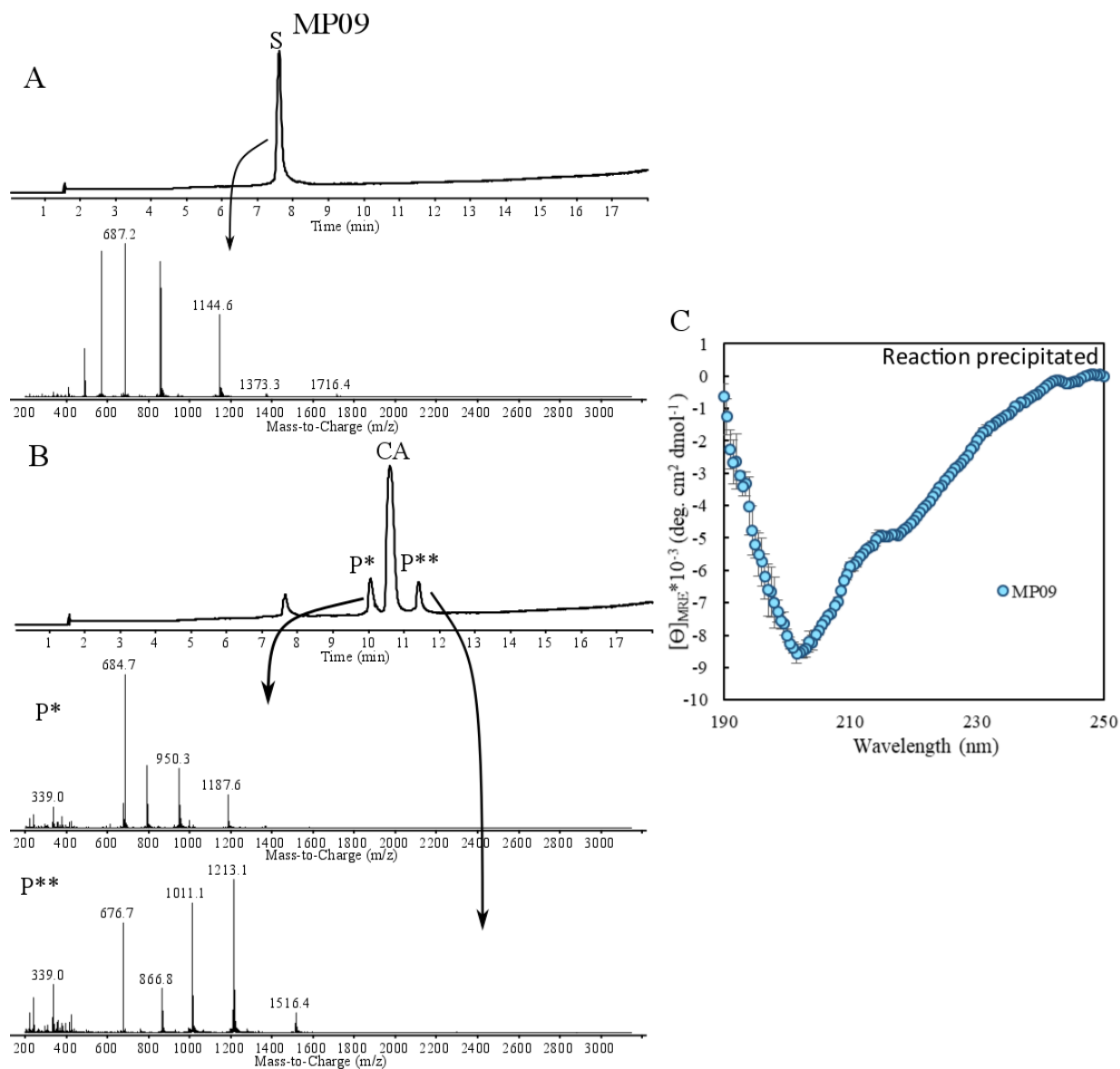


Figure S32. Characterization of MP09 and MP09-CA. (A) LCMS TIC trace for MP09 (50  $\mu$ M) with mass spectrum and (B) following a reaction between MP09 (50  $\mu$ M) and CA (250  $\mu$ M). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP09, the CA reacted version was not soluble.

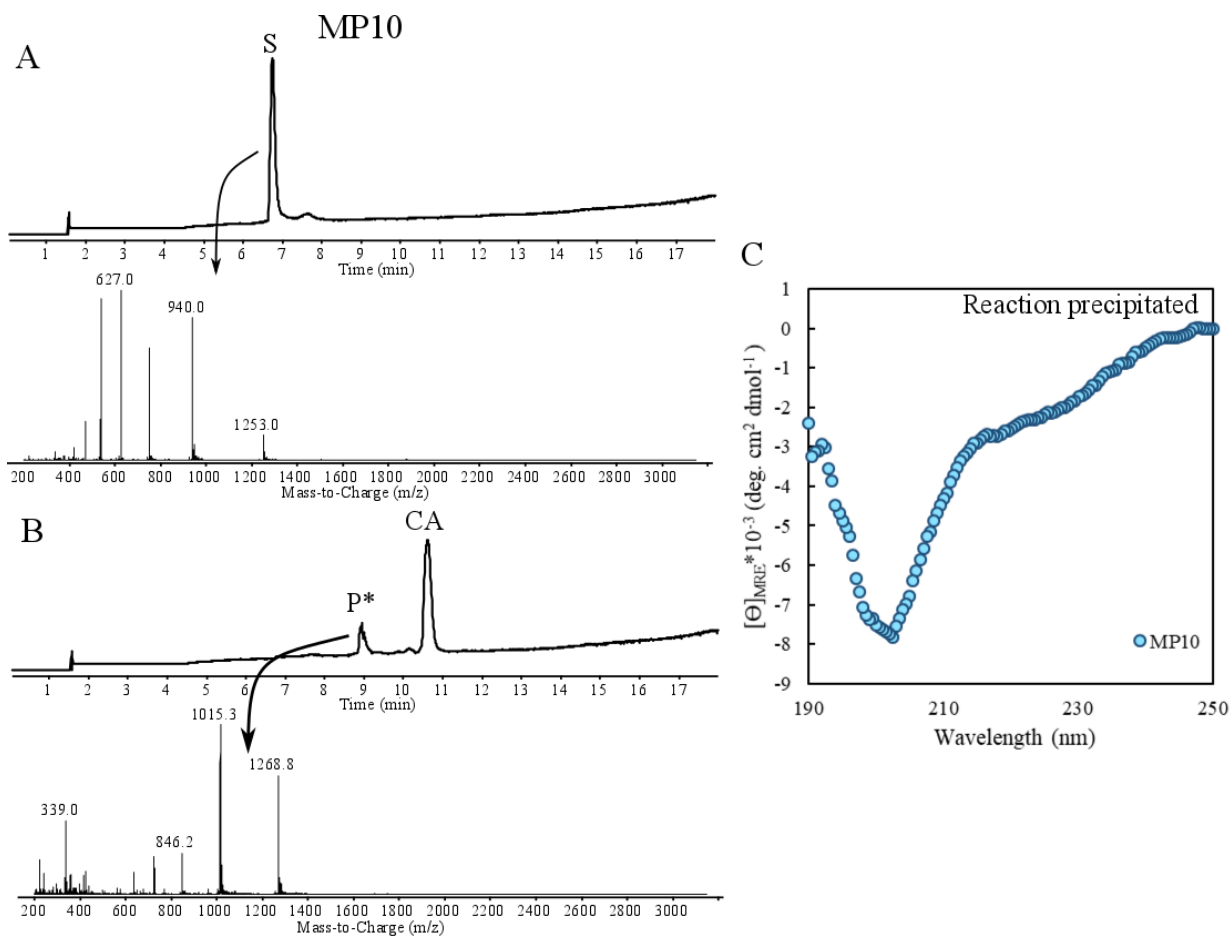


Figure S33. Characterization of MP10 and MP10-CA. (A) LCMS TIC trace for MP10 (50 μM) with mass spectrum and (B) following a reaction between MP10 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP10, the CA reacted version was not soluble.

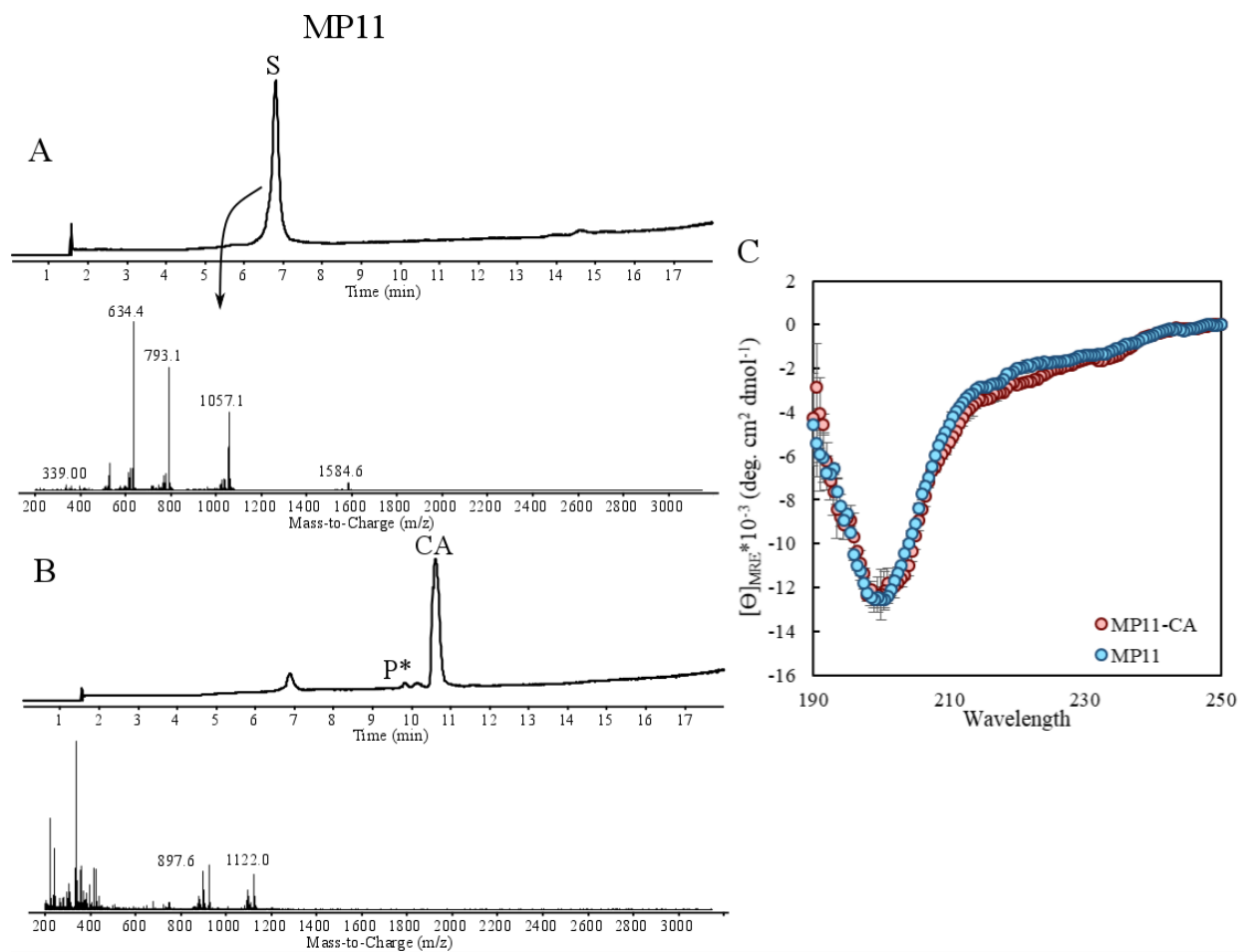


Figure S34. Characterization of MP11 and MP11-CA. (A) LCMS TIC trace for MP11 (50 μM) with mass spectrum and (B) following a reaction between MP11 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP11 samples.



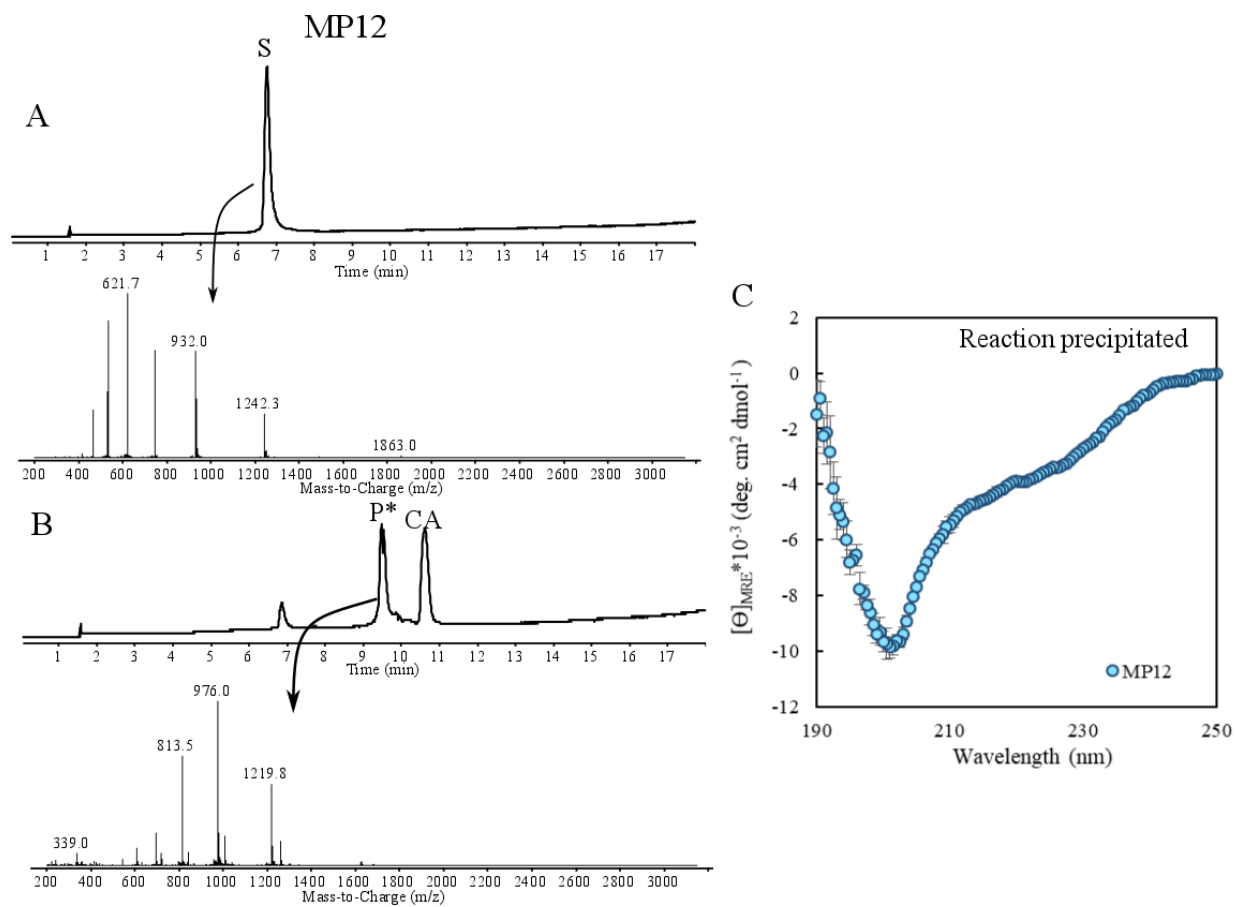


Figure S35. Characterization of MP12 and MP12-CA. (A) LCMS TIC trace for MP12 (50 μM) with mass spectrum and (B) following a reaction between MP12 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP12, the CA reacted version was not soluble.

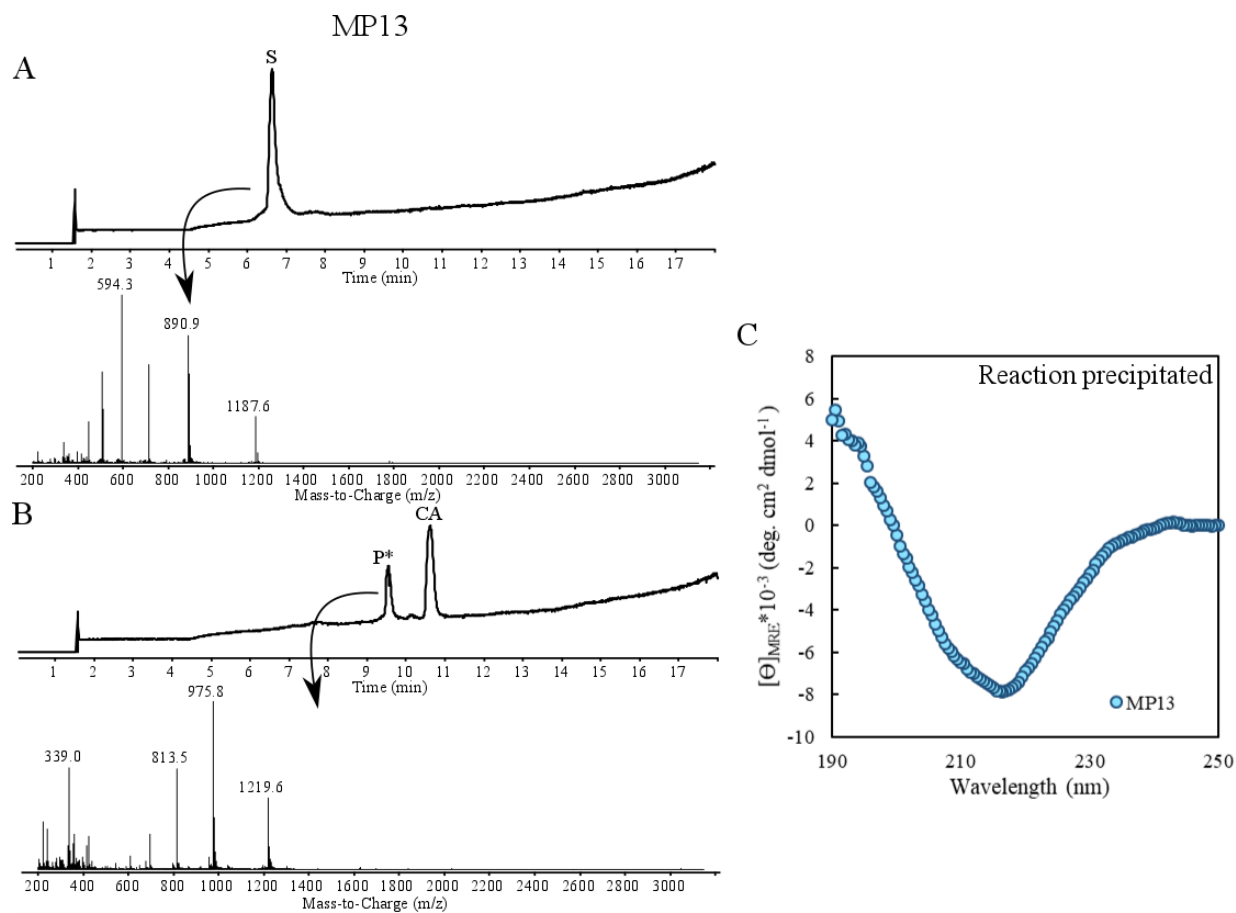


Figure S36. Characterization of MP13 and MP13-CA. (A) LCMS TIC trace for MP13 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP13 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP13, the CA reacted version was not soluble.

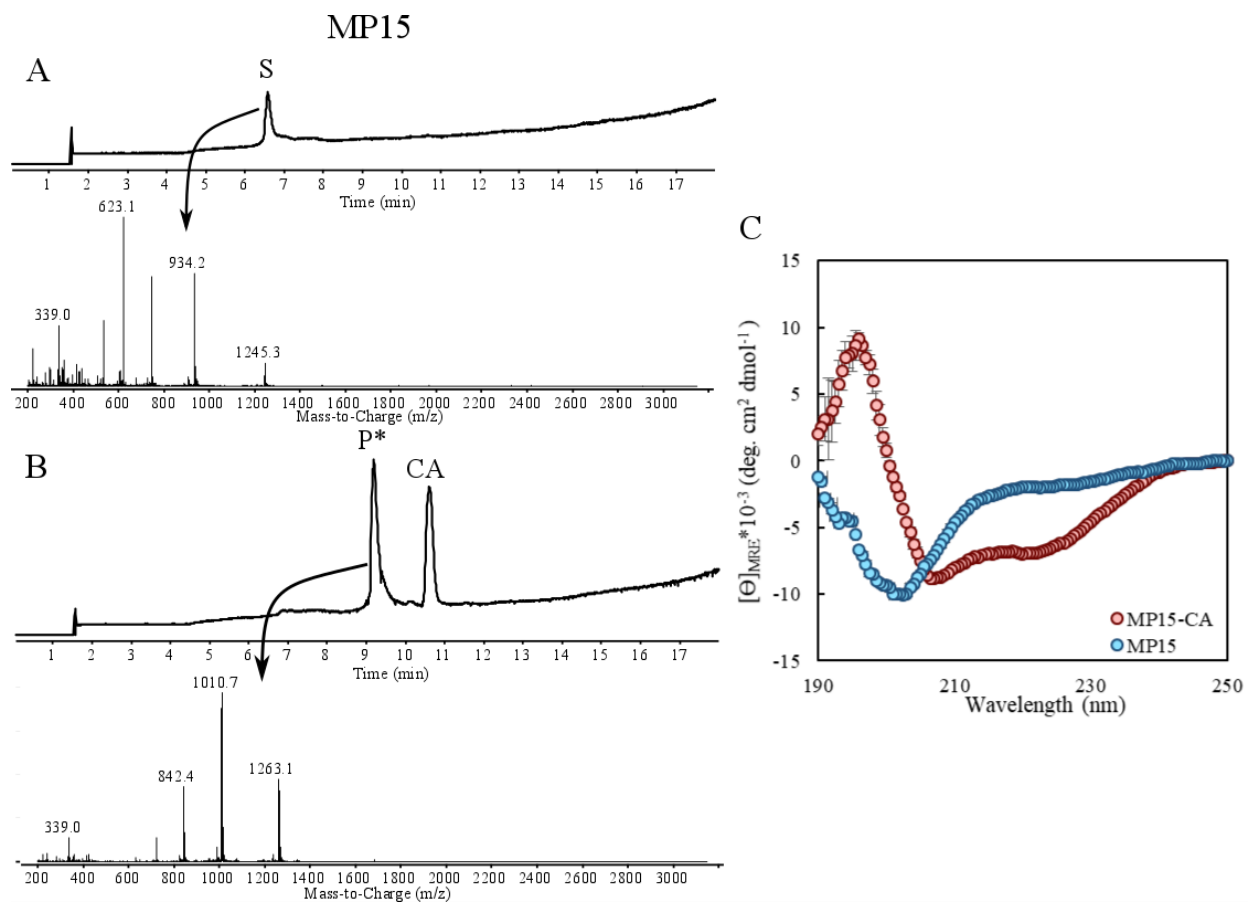


Figure S37. Characterization of MP15 and MP15-CA. (A) LCMS TIC trace for MP15 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP15 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP15 samples.

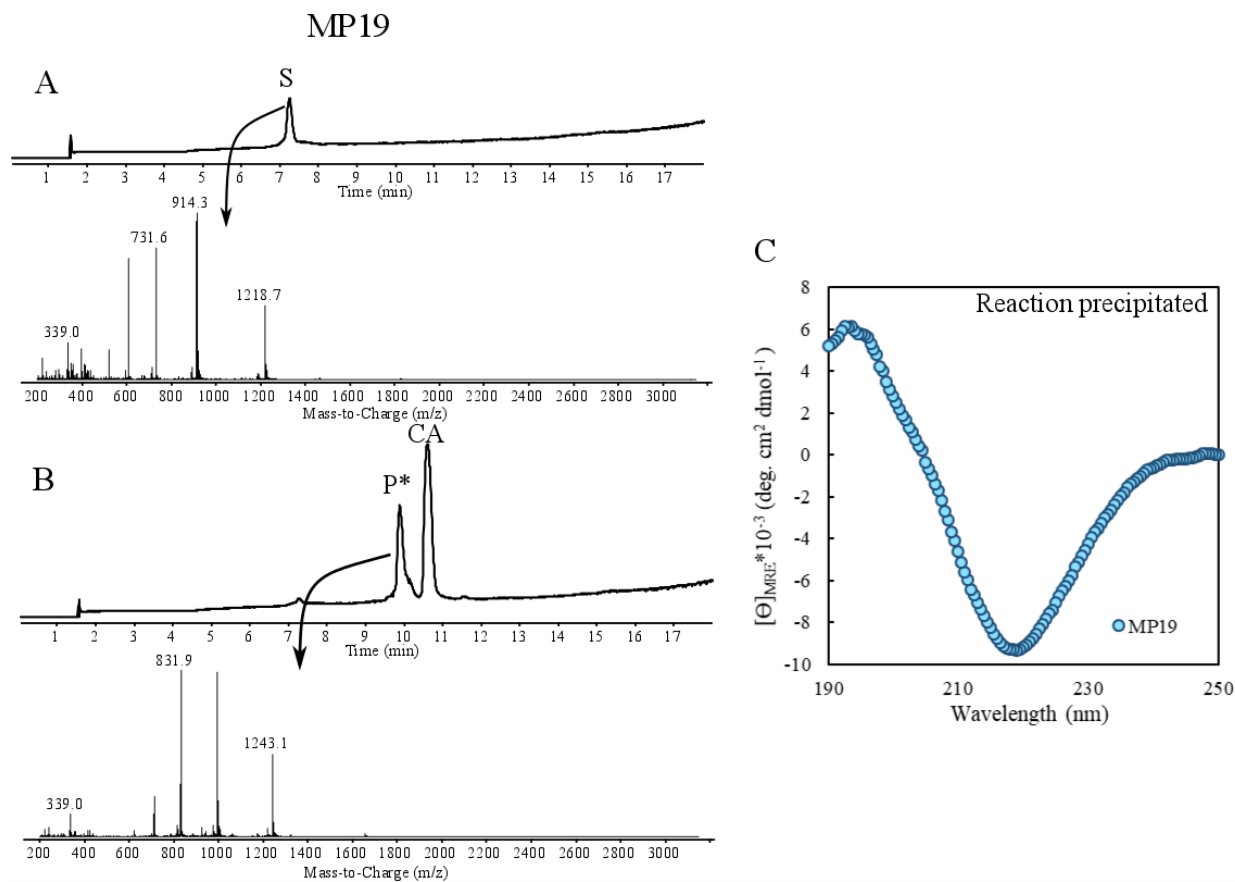


Figure S38. Characterization of MP19 and MP19-CA. (A) LCMS TIC trace for MP19 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP19 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP19, the CA reacted version was not soluble.

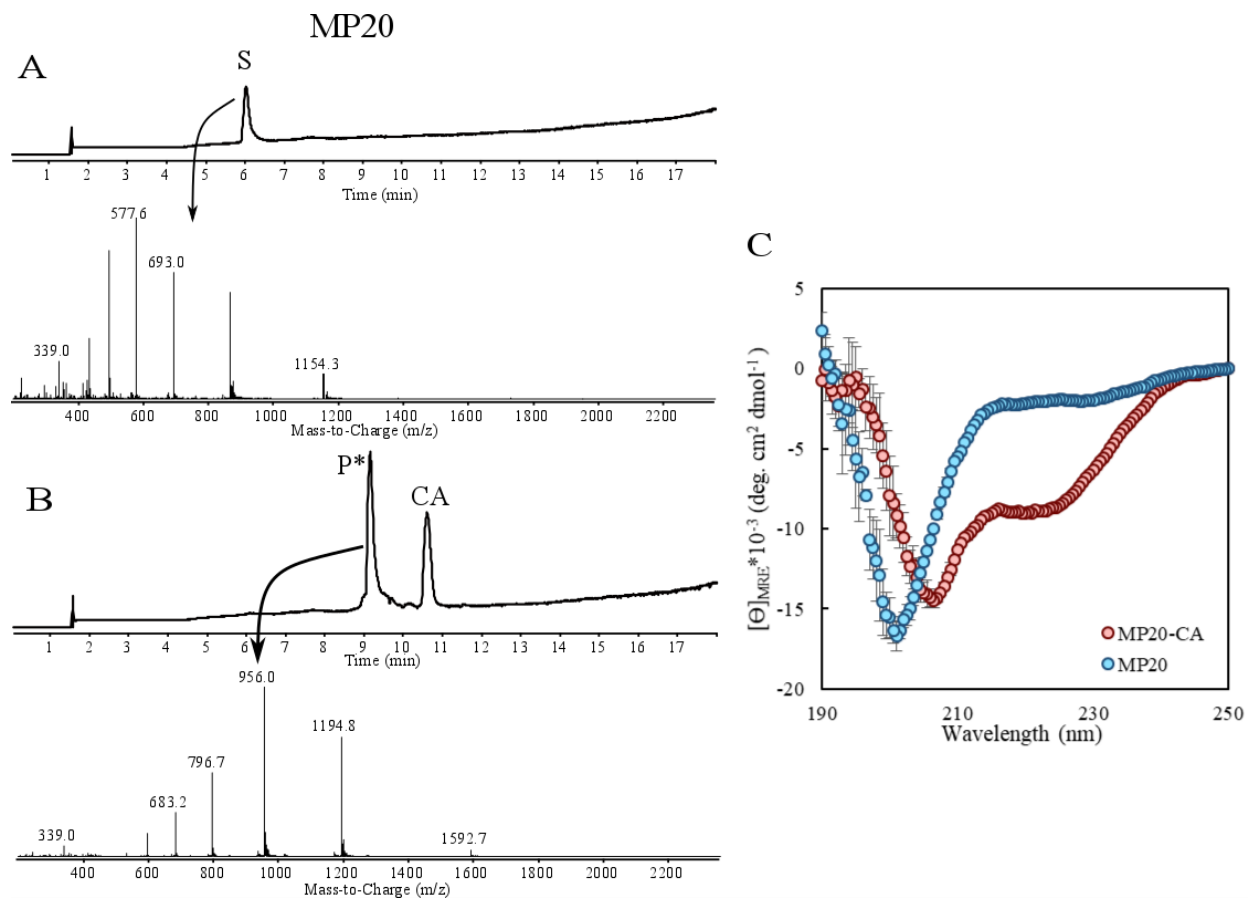


Figure S39. Characterization of MP20 and MP20-CA. (A) LCMS TIC trace for MP20 (50 μM) with mass spectrum and (B) following a reaction between MP20 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP20 samples.

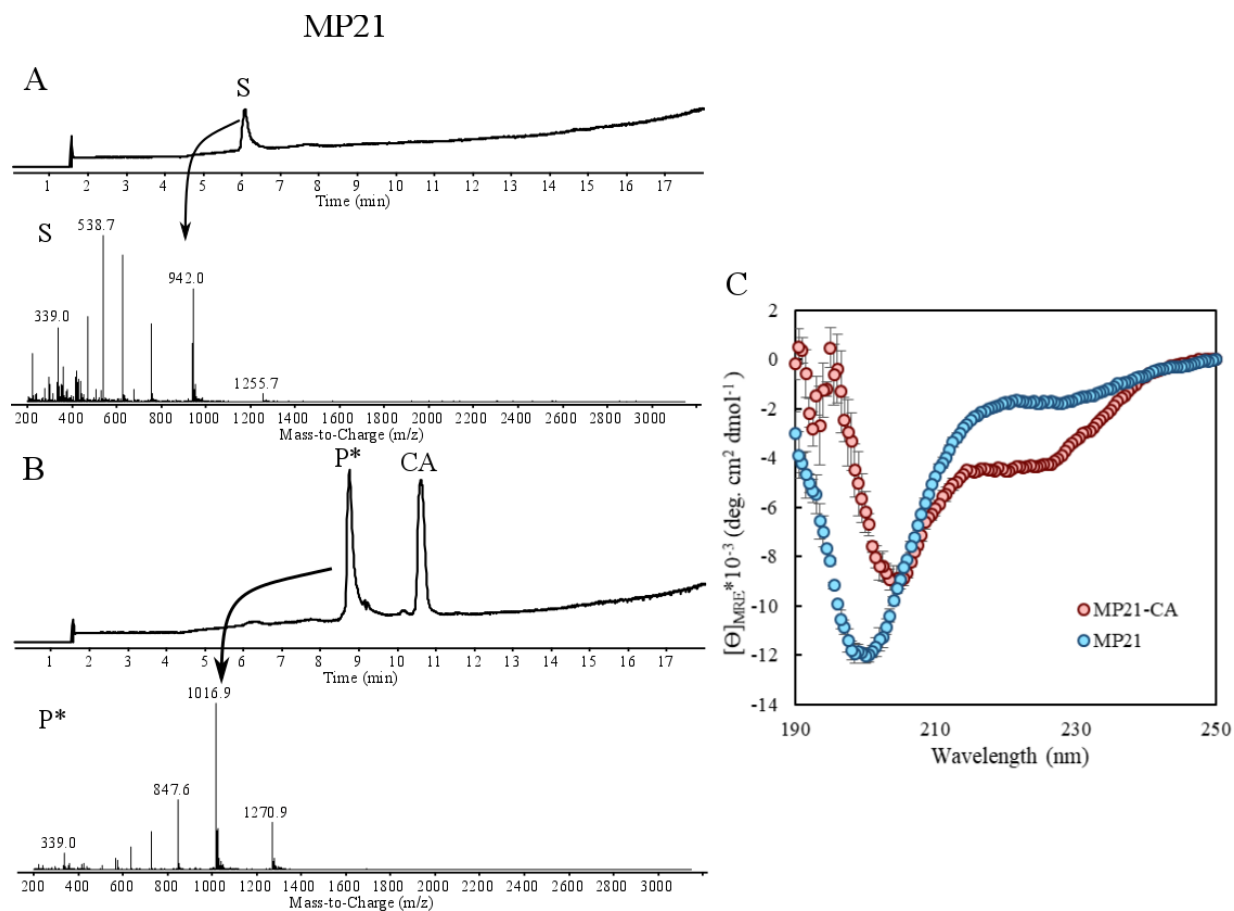


Figure S40. Characterization of MP21 and MP21-CA. (A) LCMS TIC trace for MP21 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP21 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP21 samples.

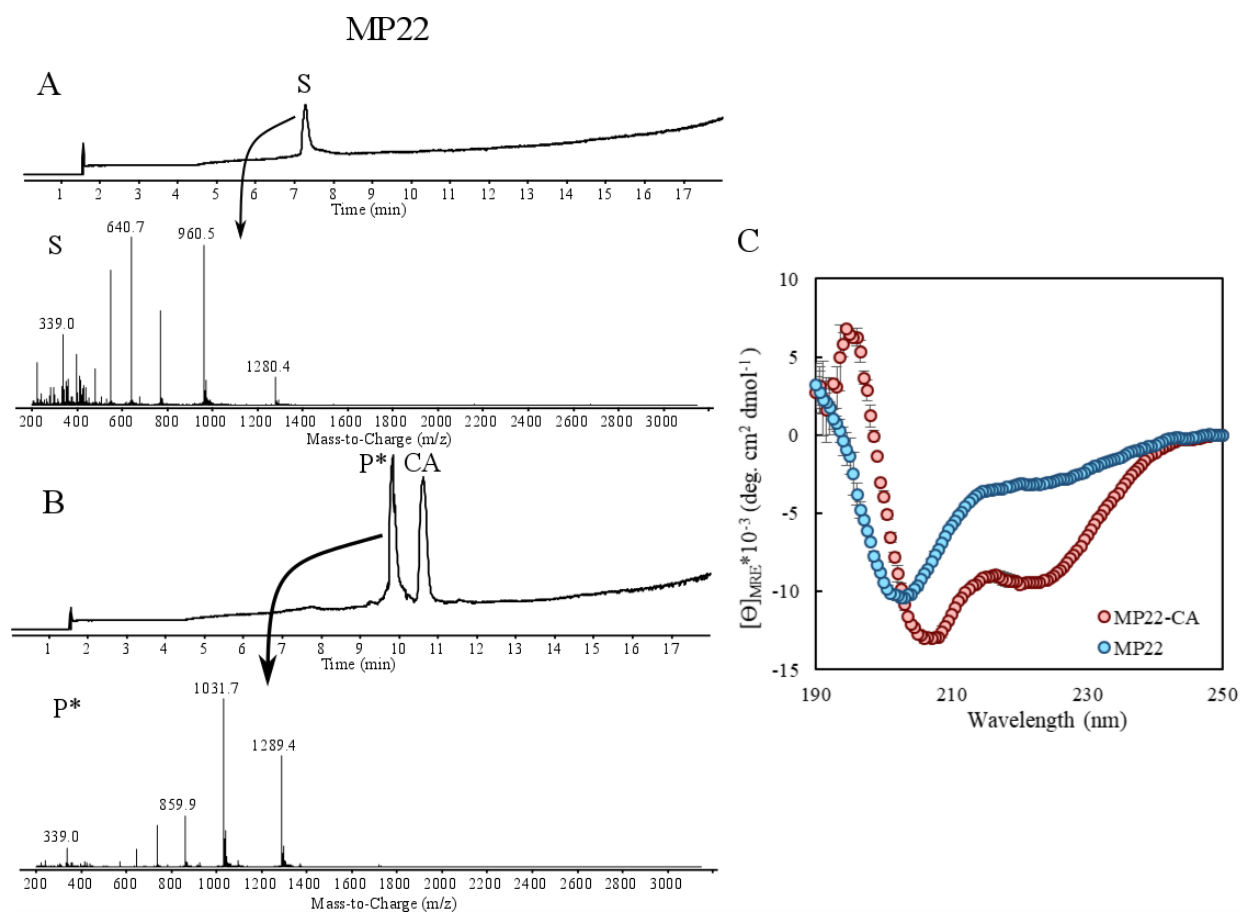


Figure S41. Characterization of MP22 and MP22-CA. (A) LCMS TIC trace for MP22 (50 μM) with mass spectrum and (B) following a reaction between MP22 (50 μM) and CA (250 μM). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP22 samples.

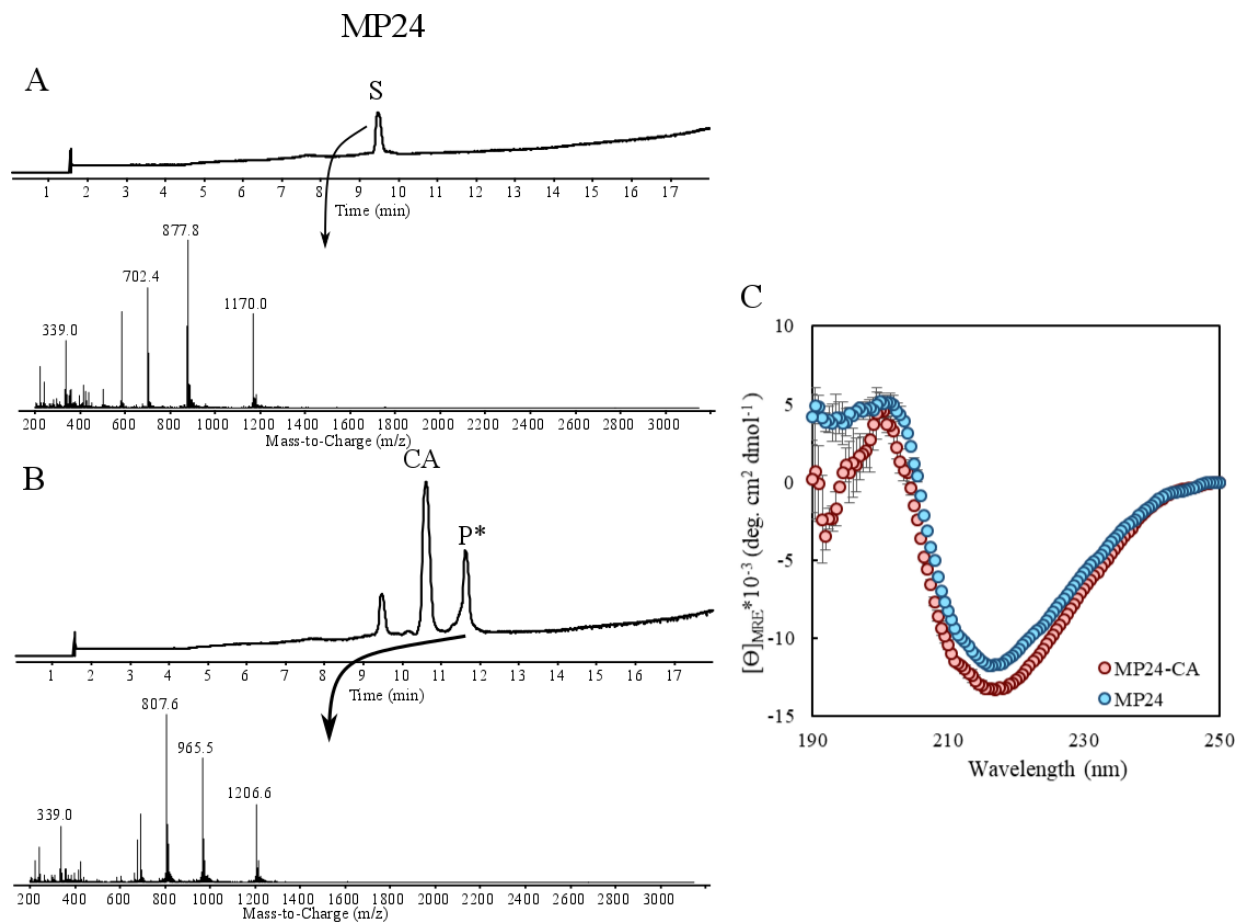


Figure S42. Characterization of MP24 and MP24-CA. (A) LCMS TIC trace for MP24 (50  $\mu$ M) with mass spectrum and (B) following a reaction between MP24 (50  $\mu$ M) and CA (250  $\mu$ M). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP24 samples.



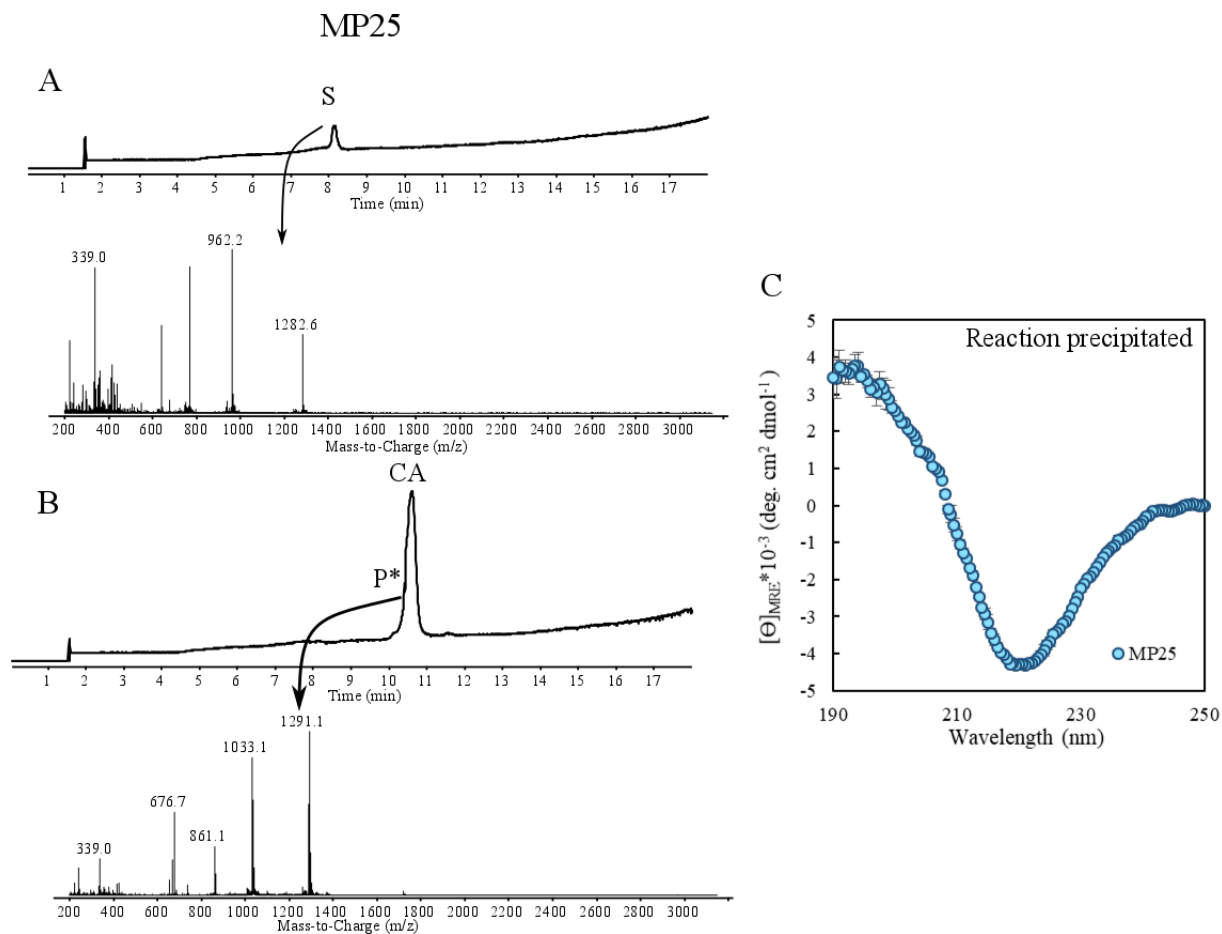


Figure S43. Characterization of MP25 and MP25-CA. (A) LCMS TIC trace for MP25 (50  $\mu$ M) with mass spectrum and (B) following a reaction between MP25 (50  $\mu$ M) and CA (250  $\mu$ M). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP25, the CA reacted version was not soluble.

## MP26

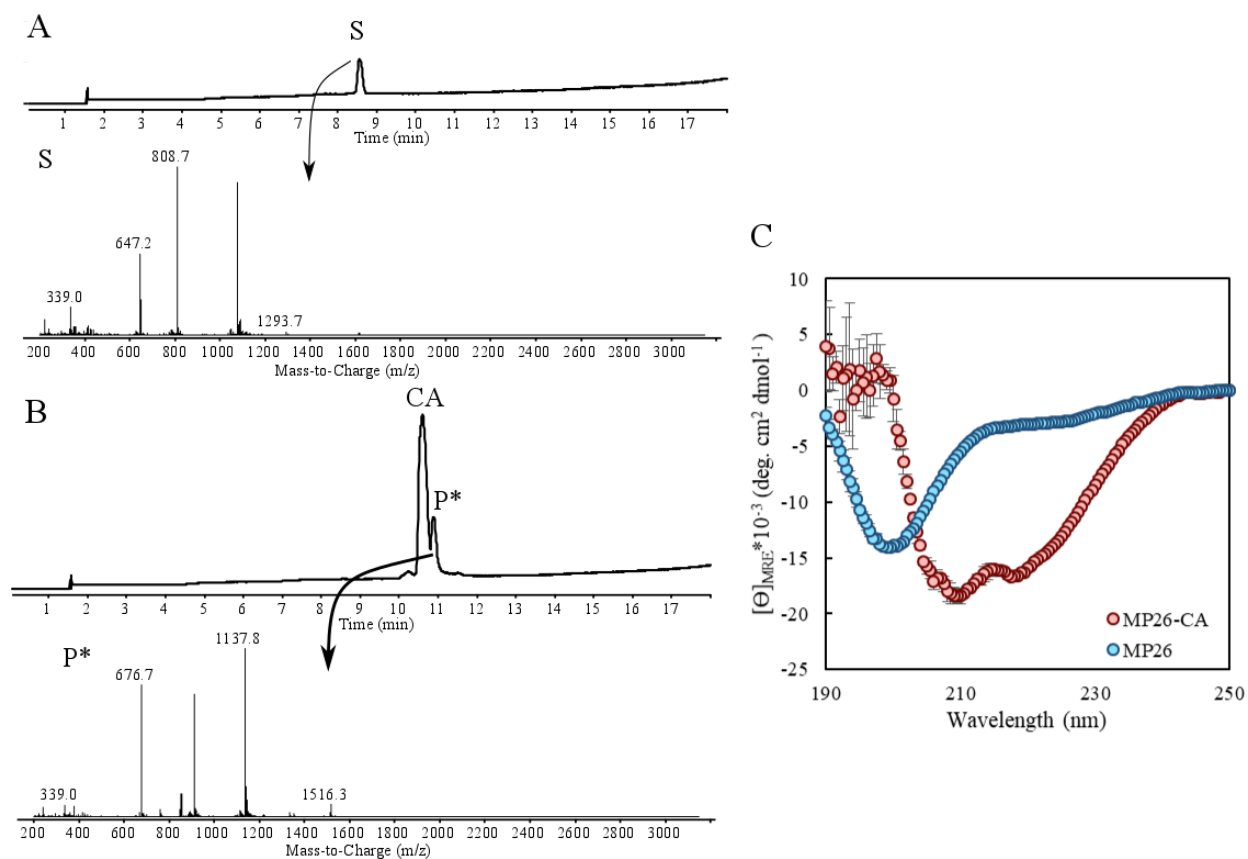


Figure S44. Characterization of MP26 and MP26-CA. (A) LCMS TIC trace for MP26 (50  $\mu$ M) with mass spectrum and (B) following a reaction between MP26 (50  $\mu$ M) and CA (250  $\mu$ M). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP26 samples.

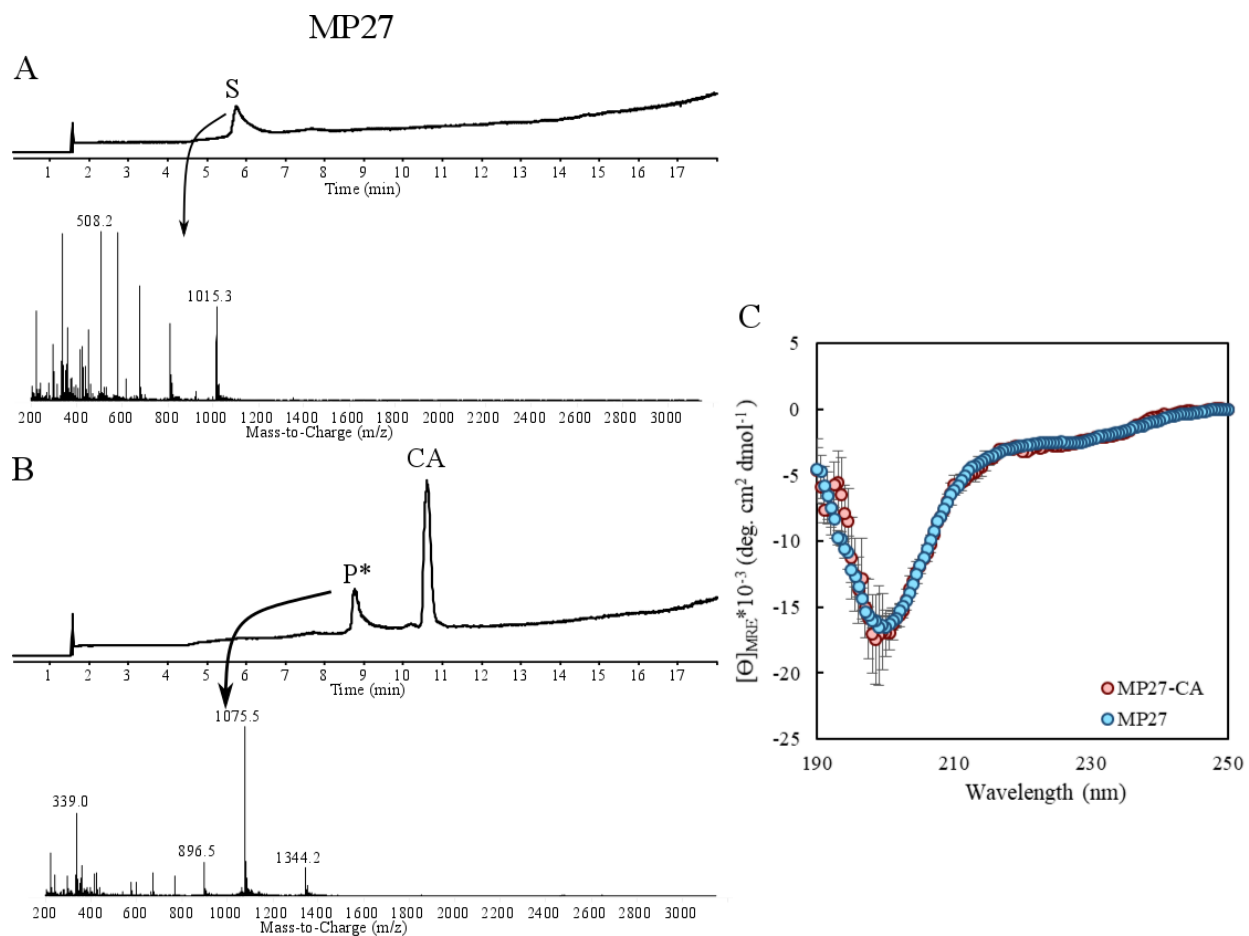


Figure S45. Characterization of MP27 and MP27-CA. (A) LCMS TIC trace for MP27 (50  $\mu\text{M}$ ) with mass spectrum and (B) following a reaction between MP27 (50  $\mu\text{M}$ ) and CA (250  $\mu\text{M}$ ). Method 1 was used without turning the mass spectrometer off at 10 minutes. (C) CD characterization of MP27 samples.

### 3.4. CD characterization of additional MPs without CA reaction

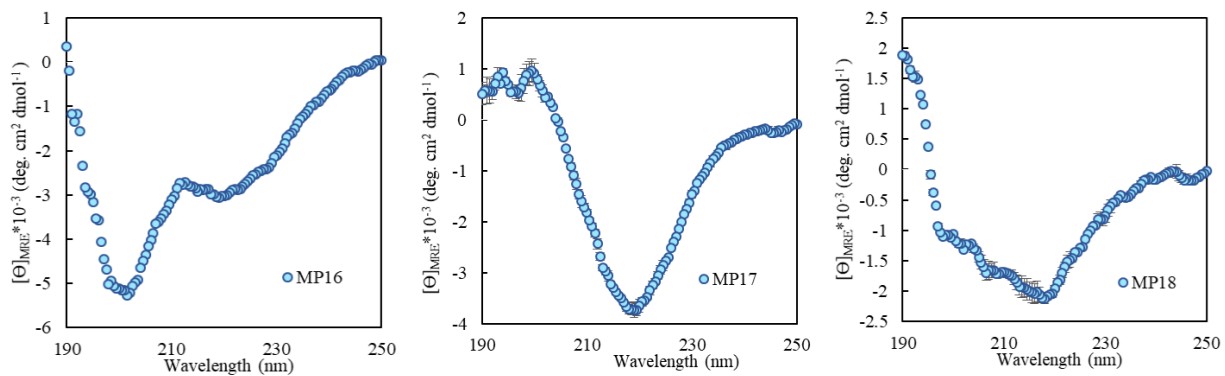


Figure S46. CD characterization of additional MPs. All peptides were at 50  $\mu\text{M}$  in standard conditions as discussed in the methods section.

### 3.5. Rosetta *ab initio* structure prediction

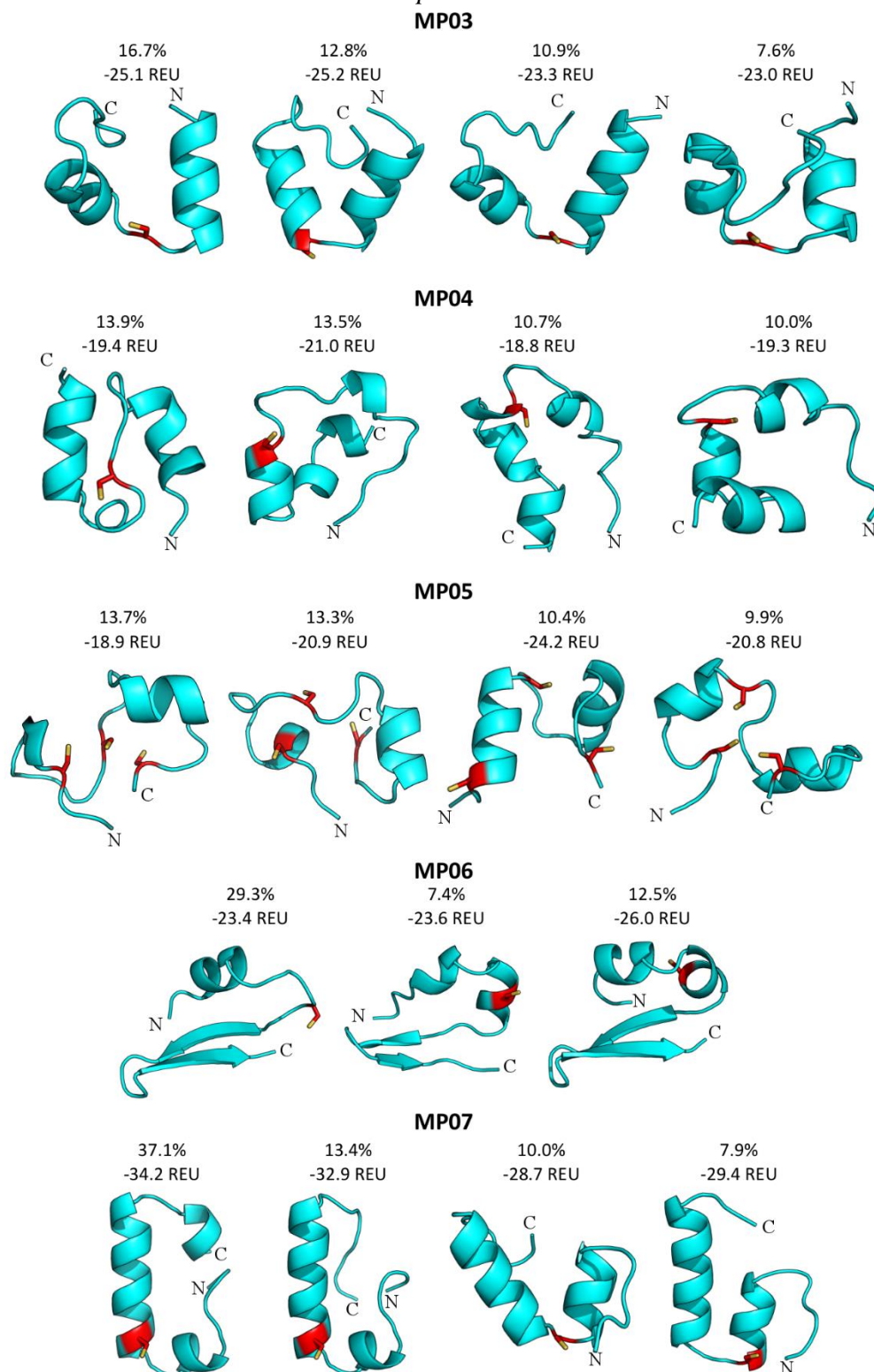


Figure S47. MP structure prediction analysis. Shown are the representative structures of the dominant families for MP03-07 under the family size (percentage) and structure energy in Rosetta energy units (REU). The active site cysteine is colored in red.

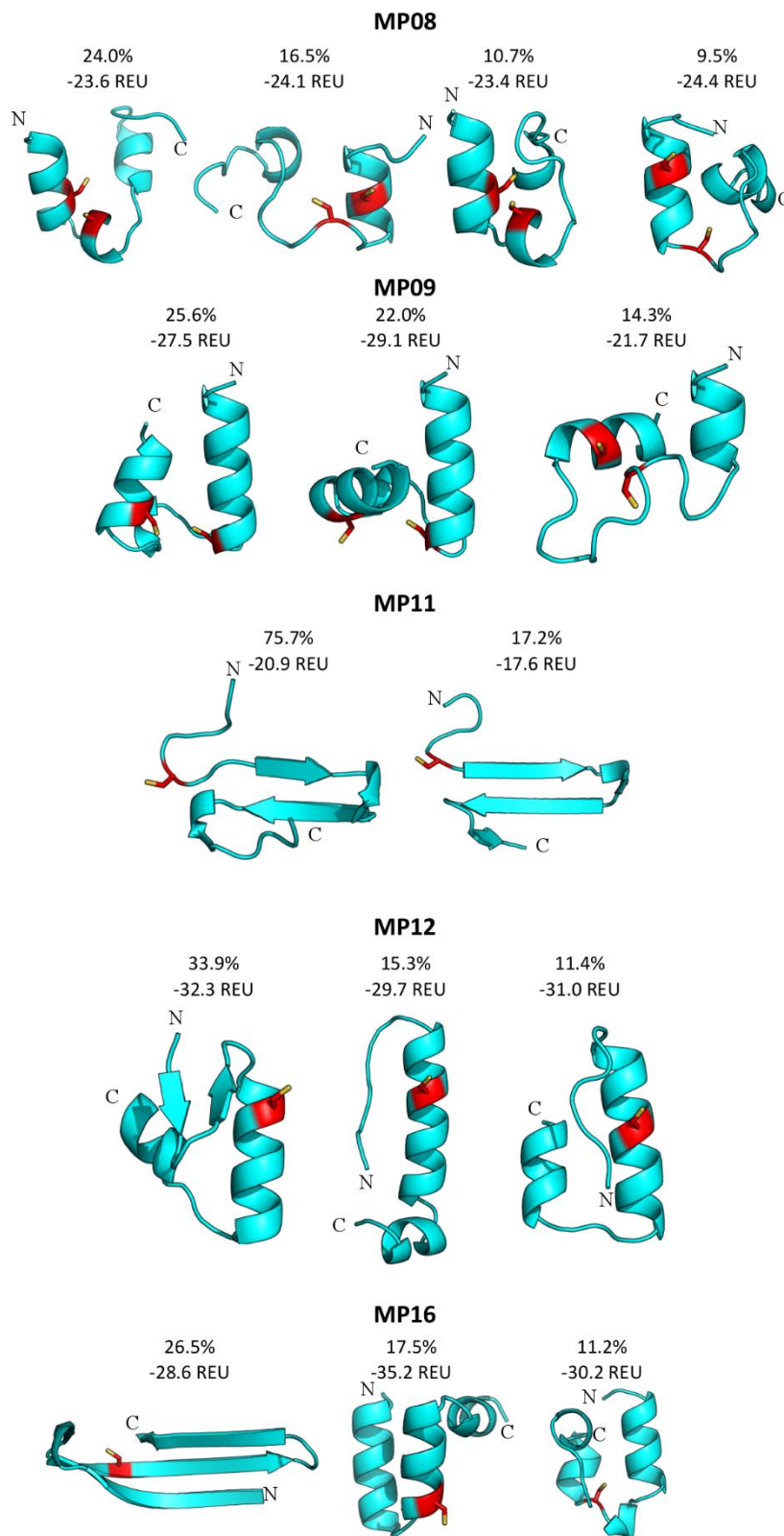


Figure S48. MP structure prediction analysis. Shown are the representative structures of the dominant families for MP08-16 under the family size (percentage) and structure energy in Rosetta energy units (REU). The active site cysteine is colored in red.

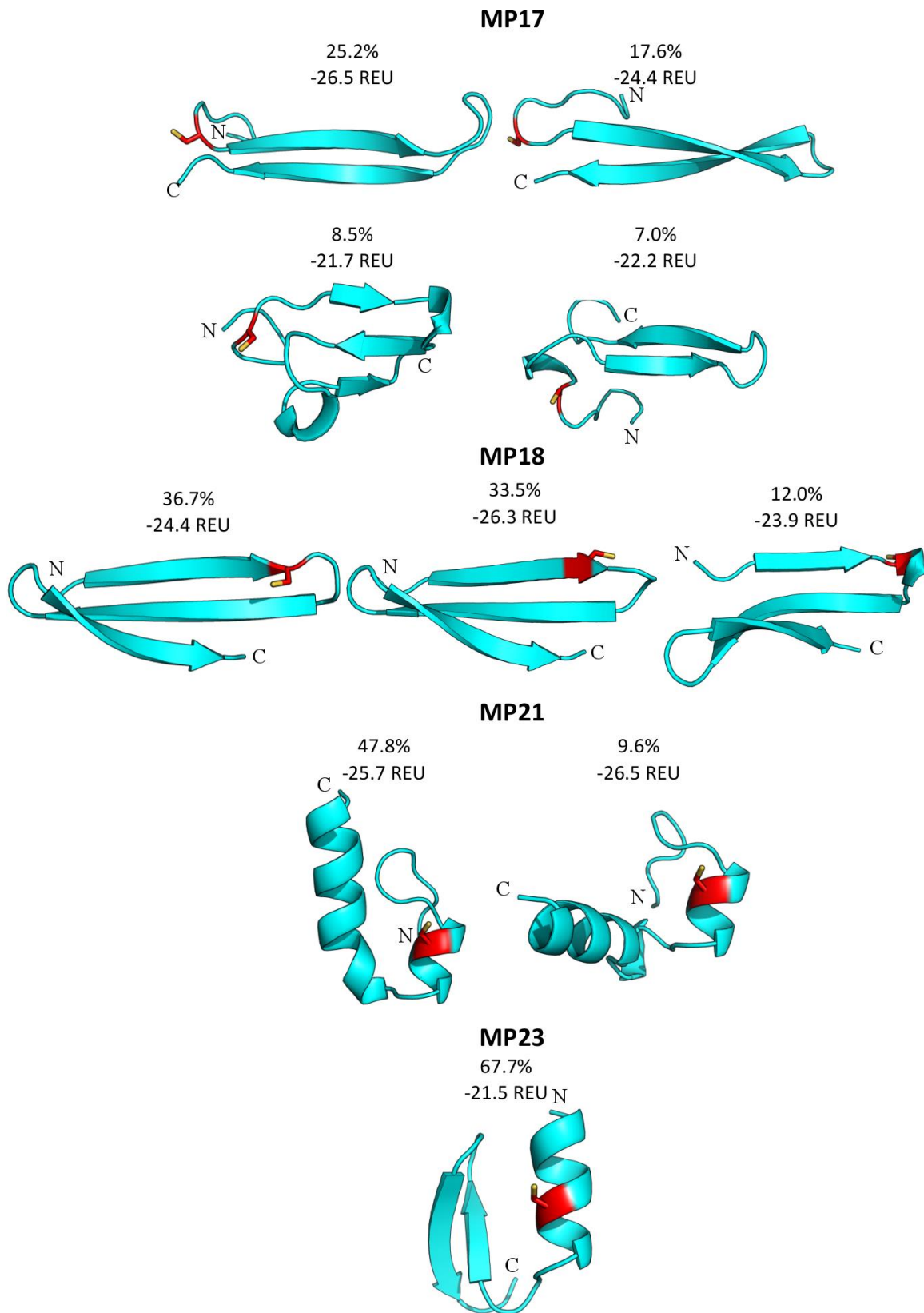


Figure S49. MP structure prediction analysis. Shown are the representative structures of the dominant families for MP17-23 under the family size (percentage) and structure energy in Rosetta energy units (REU). The active site cysteine is colored in red.

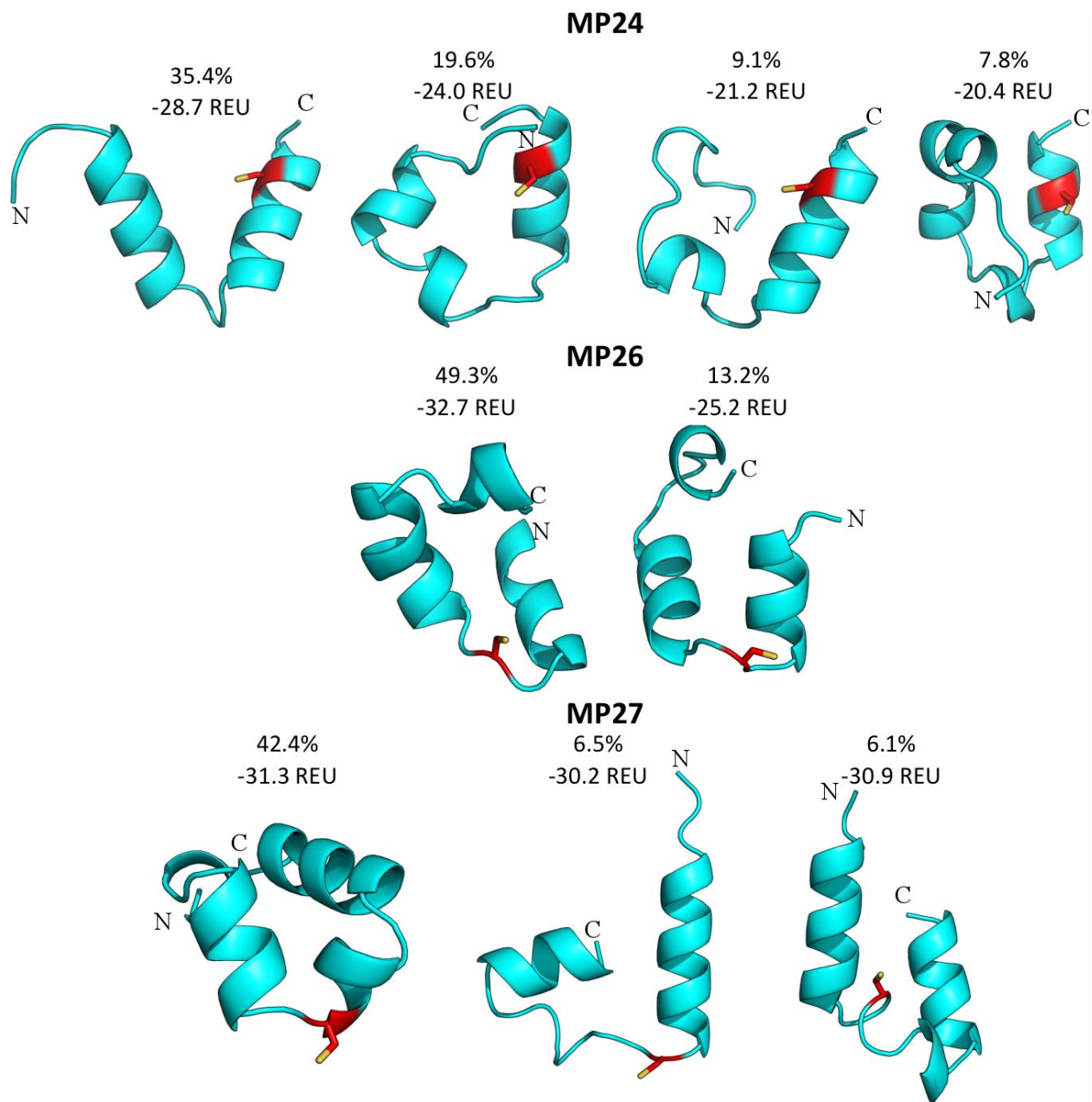


Figure S50. MP structure prediction analysis. Shown are the representative structures of the dominant families for MP24-27 under the family size (percentage) and structure energy in Rosetta energy units (REU). The active site cysteine is colored in red.



#### **4. References**

- 1 E. D. Evans and B. L. Pentelute, *ACS Chem. Biol.*, 2018, **13**, 527–532.
- 2 A. J. Mijalis, D. A. Thomas Iii, M. D. Simon, A. Adamo, R. Beaumont, K. F. Jensen and B. L. Pentelute, *Nat. Chem. Biol.*, 2017, **13**, 464–466.