

Electronic Supporting Information  
for  
**Controlling Macromolecular Topology with Genetically Encoded  
SpyTag-SpyCatcher Chemistry**

Wen-Bin Zhang,<sup>†</sup> Fei Sun,<sup>†</sup> David A. Tirrell,\* and Frances H. Arnold\*

Division of Chemistry and Chemical Engineering, California Institute of Technology,

Pasadena, California 91125, United States

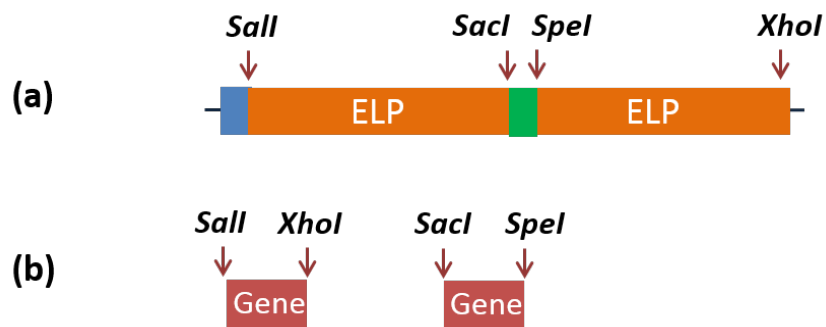
<sup>†</sup> These two authors contributed equally.

E-mail: [tirrell@caltech.edu](mailto:tirrell@caltech.edu); [frances@cheme.caltech.edu](mailto:frances@cheme.caltech.edu)

## Table of Contents

	Page
1. Figure S1. (a) Cartoon illustration of the original ELP construct. It contains four restriction sites: Sall (GTCGAC), SacI (GAGCTC), SpeI (ACTAGT), and XhoI (CTCGAG) as shown. (b) The gene inserts can be made in a way that is either flanked by Sall and XhoI for insertion into the N-terminus or C-terminus or by SacI and SpeI for insertion into the middle of the chain. The restriction sites Sall and XhoI share the same sticky ends. Therefore, the insert can be used to ligate the vector that is digested by Sall or XhoI. The inserts can be SpyTag or SpyCatcher.....	S4
2. Figure S2. Full amino acid sequence of AB20D (351 a.a. MW = 34002). The reactive aspartic acid (20D) is shown in the box at the first line. It is changed to nonreactive alanine (20A) in AB20A. The reactive lysine (255K) is shown in the box at the 7 <sup>th</sup> line. The segments highlighted in red, yellow, and purple are the SpyTag, TEV protease recognition sequence, and SpyCatcher, respectively.....	S5
3. Figure S3. MALDI-TOF mass spectra of AB20D expressed at 37 °C with induction: (a) monomer; (b) dimer; (c) trimer; and (d) tetramer.....	S6
4. Figure S4. MALDI-TOF mass spectrum of AB20A .....	S7
5. Figure S5. MALDI-TOF mass spectrum of relinearized AB20D.....	S8
6. Figure S6. Full amino acid sequence of EAEB (332 a.a. MW = 32172). The reactive aspartic acid (102D) and lysine (236K) are shown in the box at the 3 <sup>th</sup> and 6 <sup>th</sup> lines, respectively. The segments highlighted in red, yellow, and purple are the SpyTag, TEV protease recognition sequence, and SpyCatcher, respectively.....	S9
7. Figure S7. MALDI-TOF mass spectra of EAEB expressed at 37 °C with 1 mM IPTG induction: (a) monomer; (b) dimer; and (c) trimer.....	S10
8. Figure S8. Full amino acid sequence of (a) EA and (b) EB. The reactive aspartic acid (203D) and lysine (240K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.....	S11
9. Figure S9. MALDI-TOF mass spectra of EA (a) and EB (b). .....	S12
10. Figure S10. Full amino acid sequence of (a) EAE and (b) EBE. The reactive aspartic acid (102D) and lysine (139K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.....	S13
11. Figure S11. MALDI-TOF mass spectra of EAE (a) and EBE (b). .....	S14

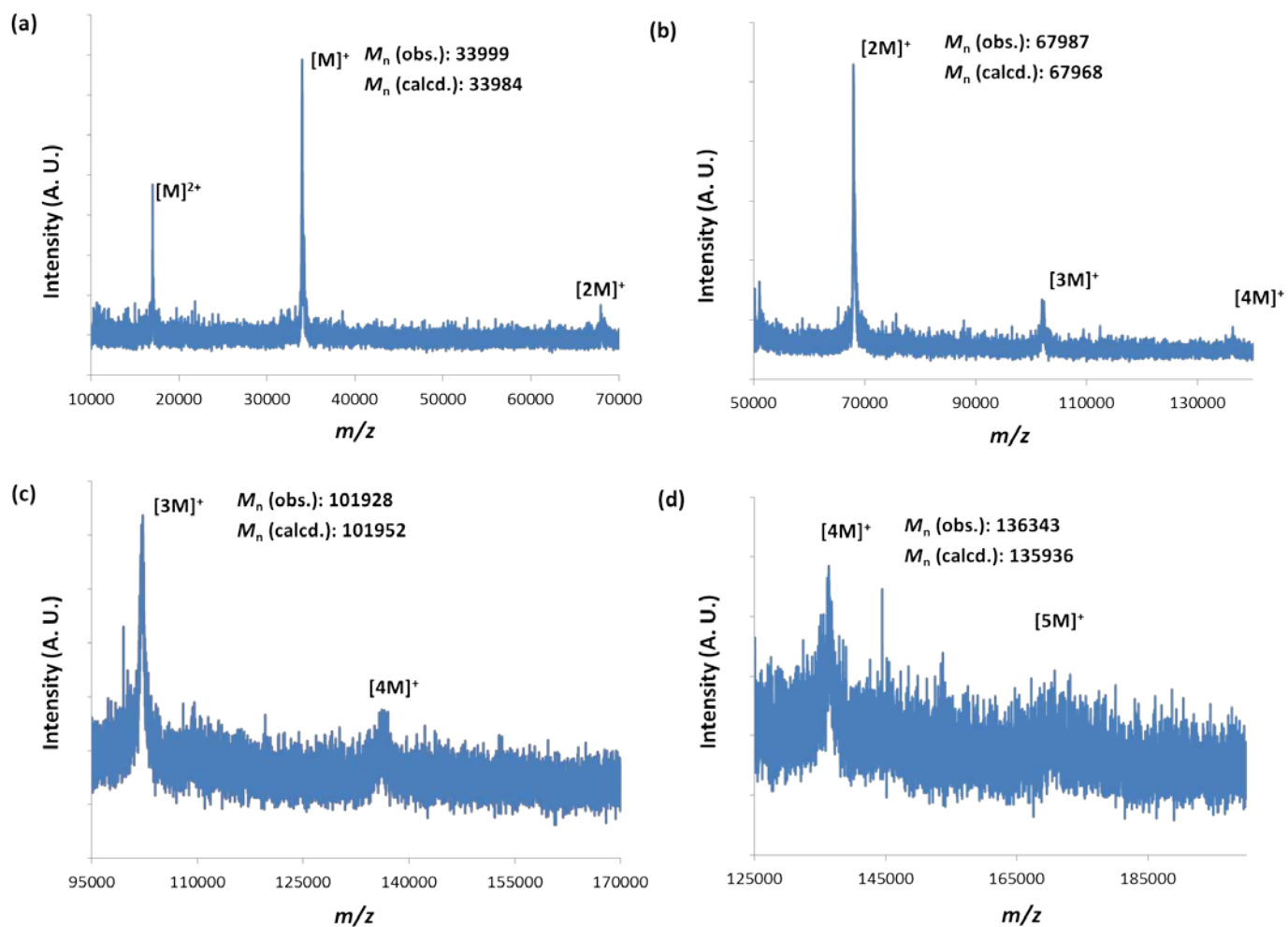
12.	Figure S12. Full amino acid sequence of (a) AA and (b) BB. The reactive aspartic acid (20D and 218D) and lysine (57K and 377K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.....	S15
13.	Figure S13. MALDI-TOF mass spectra of AA (a) and BB (b).....	S16
14.	Figure S14. MALDI-TOF mass spectrum of the block protein (EA+EB). ....	S17
15.	Figure S15. MALDI-TOF mass spectra of the 3-arm star proteins EA+EBE (a) and EAE+EB (b) and the 4-arm star protein EAE+EBE (c).....	S18
16.	Figure S16. MALDI-TOF mass spectra of the H-shaped proteins AA+2EBE (a) and BB+2EAE (b) and the corresponding intermediate 3-arm proteins AA+EBE (c) and BB+EBE (d). ....	S19
17.	Table S1. Summary of bacterial strains and plasmids used in this study .....	S20



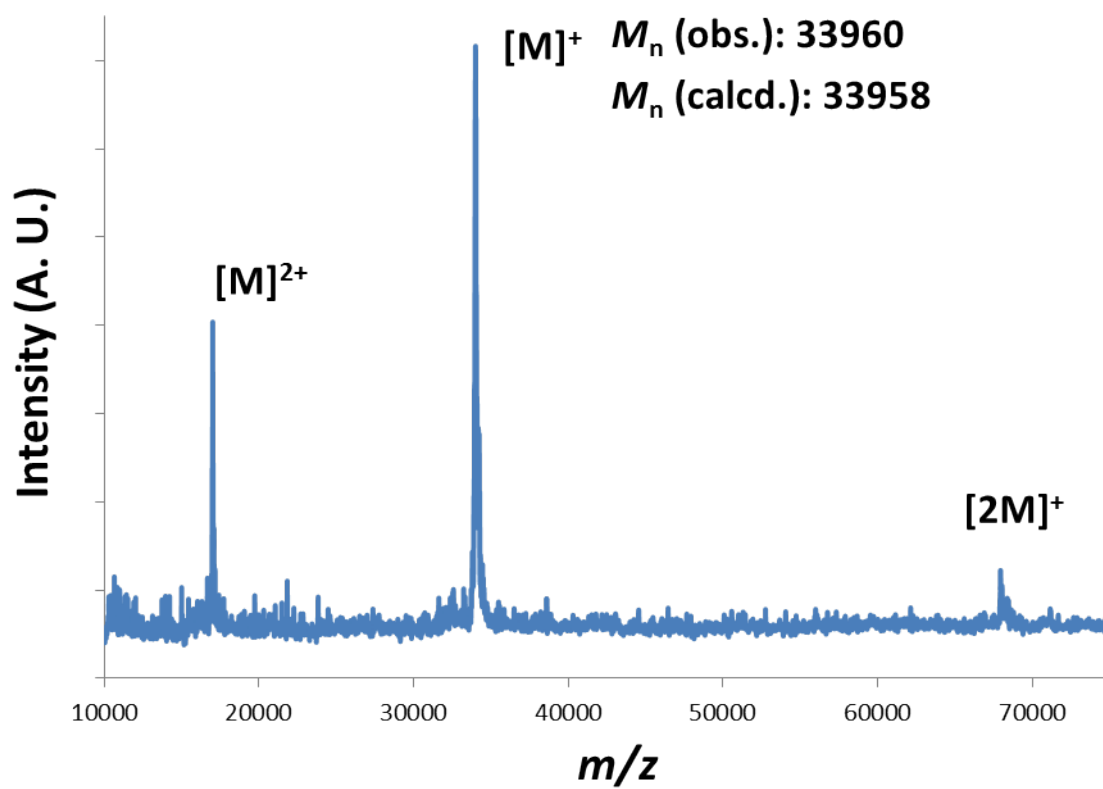
**Figure S1.** (a) Cartoon illustration of the original ELP construct. It contains four restriction sites: *Sall* (GTCGAC), *SacI* (GAGCTC), *SpeI* (ACTAGT), and *XhoI* (CTCGAG) as shown. (b) The gene inserts can be made in a way that is either flanked by *Sall* and *XhoI* for insertion into the N-terminus or C-terminus or by *SacI* and *SpeI* for insertion into the middle of the chain. The restriction sites *Sall* and *XhoI* share the same sticky ends. Therefore, the insert can be used to ligate the vector that is digested by *Sall* or *XhoI*. The inserts can be SpyTag or SpyCatcher.

1 MKGSSHHHHHHVD **AHIVMV** **D** **AYKPTK**LDGHGVGVPGVGVP  
 41 GVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGV  
 81 GVGVPGVGVPGVGVPGEGVPGVGVPGVGVGELYAVTGRGDS  
 121 ASSAPIATSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGV  
 161 PGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGV  
 201 PGVGVPGLLDIPTT **ENLYFQG**AMVDTLSGLSSEQQSGD  
 241 **MTIEEDSATHIKFS** **K** **RDEDGKELAGATMELRDSSGKTIST**  
 281 **WISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNE**  
 321 **QGQVTVNGKATKGAHID**GPQGIWGQLEWKK

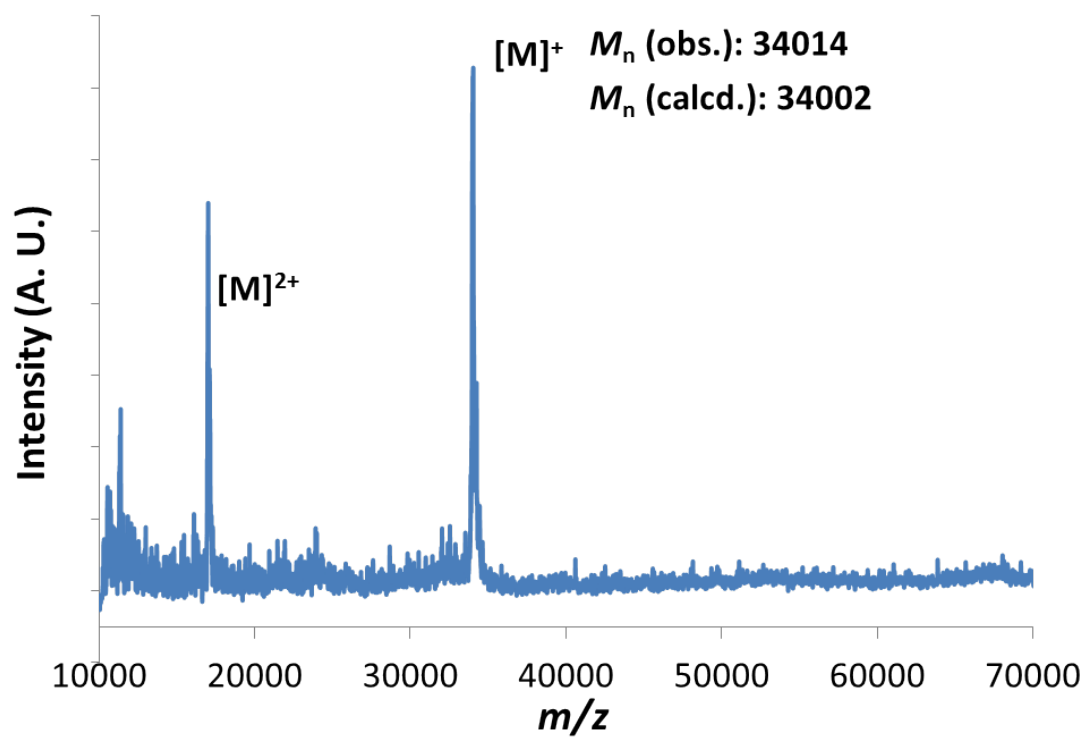
**Figure S2.** Full amino acid sequence of AB20D (351 a.a. MW = 34002). The reactive aspartic acid (20D) is shown in the box at the first line. It is changed to nonreactive alanine (20A) in AB20A. The reactive lysine (255K) is shown in the box at the 7<sup>th</sup> line. The segments highlighted in red, yellow, and purple are the SpyTag, TEV protease recognition sequence, and SpyCatcher, respectively.



**Figure S3.** MALDI-TOF mass spectra of AB20D expressed at 37 °C with induction: (a) monomer; (b) dimer; (c) trimer; and (d) tetramer.



**Figure S4.** MALDI-TOF mass spectrum of AB20A.

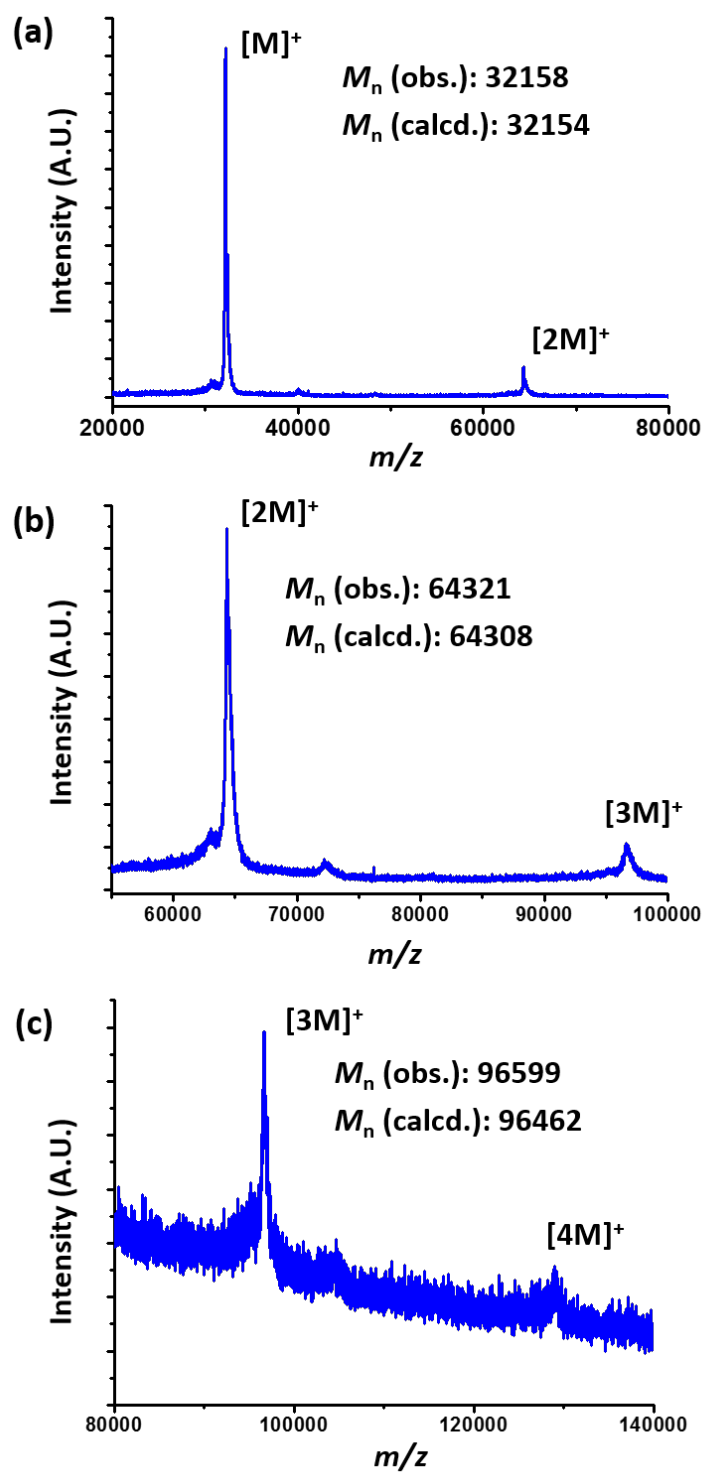


**Figure S5.** MALDI-TOF mass spectrum of relinearized AB20D.



1 MKGSSHHHHHHVDGHGVPGVGVPGVGVPGEGVPGVGV  
 41 GVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGV  
 81 GEGVPGVGVPGVGVGELAHIVMV D AYKPTKTSVPGVGVPGV  
 121 VPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGV  
 161 VPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLDIPTTENLY  
 201 FQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFS K RDED  
 241 GKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTF  
 281 VETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIDG  
 321 PQGIWGQLEWKK

**Figure S6.** Full amino acid sequence of EAEB (332 a.a. MW = 32172). The reactive aspartic acid (102D) and lysine (236K) are shown in the box at the 3<sup>th</sup> and 6<sup>th</sup> lines, respectively. The segments highlighted in red, yellow, and purple are the SpyTag, TEV protease recognition sequence, and SpyCatcher, respectively.



**Figure S7.** MALDI-TOF mass spectra of EAEB expressed at 37 °C with 1 mM IPTG induction: (a) monomer; (b) dimer; and (c) trimer.

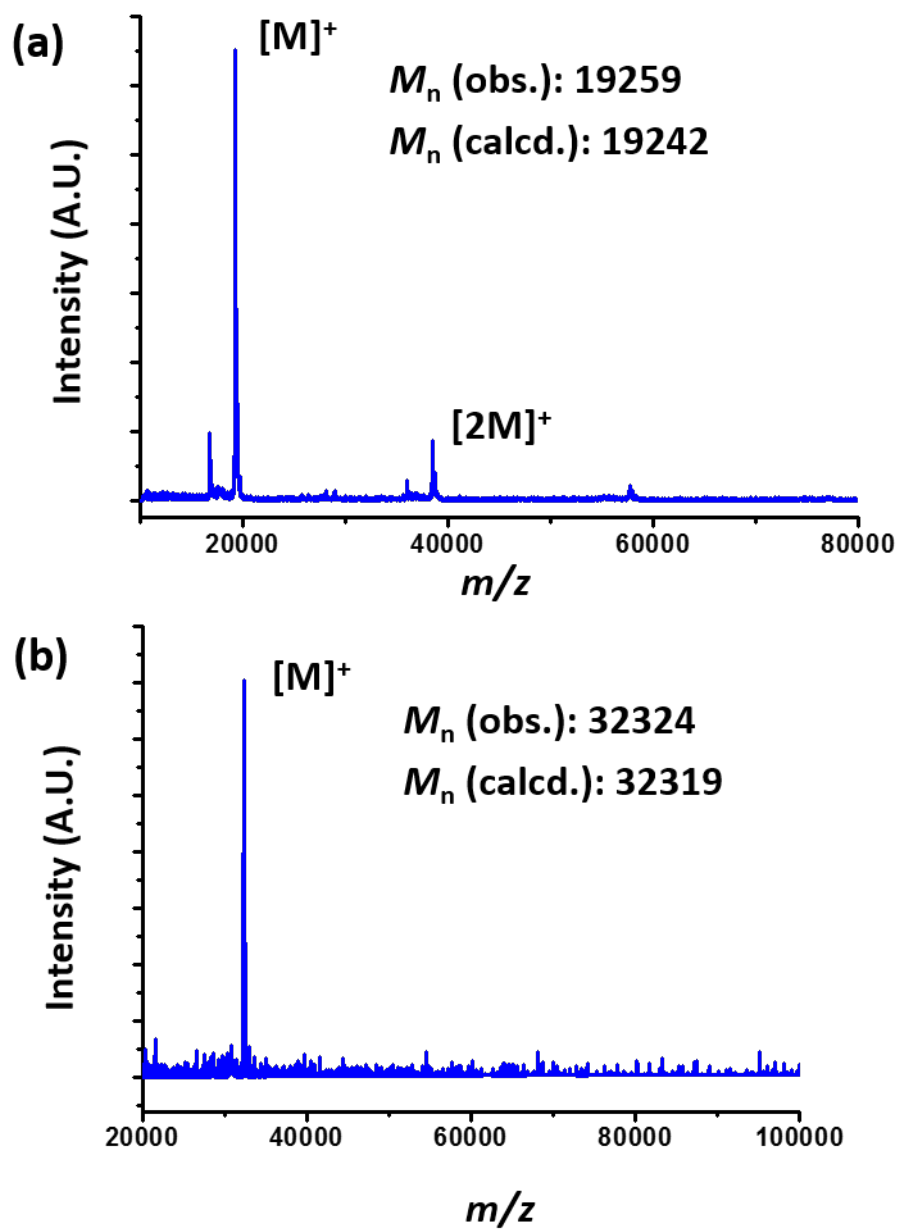
(a) EA: 214 a.a. MW = 19242

1 MKGSSHHHHHHVDGHGVGVPGVGVPGVGVPGEVPGVGVV  
41 GVGVPGVGVPGVGVPGEVPGVGVVPGVGVVPGVGVVPGVGVV  
81 GEGVPGVGVVPGVGVVAVTGRGDSPASSAPIATSVPGVGV  
121 PGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVV  
161 PGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVV  
201 MV D AYKPTKLEWKK

(b) EB: 336 a.a. MW = 32319

1 MKGSSHHHHHHVDGHGVGVPGVGVPGVGVPGEVPGVGVV  
41 GVGVPGVGVPGVGVPGEVPGVGVVPGVGVVPGVGVVPGVGVV  
81 GEGVPGVGVVPGVGVVAVTGRGDSPASSAPIATSVPGVGV  
121 PGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVV  
161 PGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVVPGVGVV  
201 ENLYFQGA MVDTL SGLSSEQQSGDMTIEEDSATHIKFS K  
241 RDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPG  
281 KYTFVETAAPDGYEVATAITFTVNEQQQVTVNGKATKGDA  
321 HID GPQGIWGQLEWKK

**Figure S8.** Full amino acid sequence of (a) EA and (b) EB. The reactive aspartic acid (203D) and lysine (240K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.



**Figure S9.** MALDI-TOF mass spectra of EA (a) and EB (b).

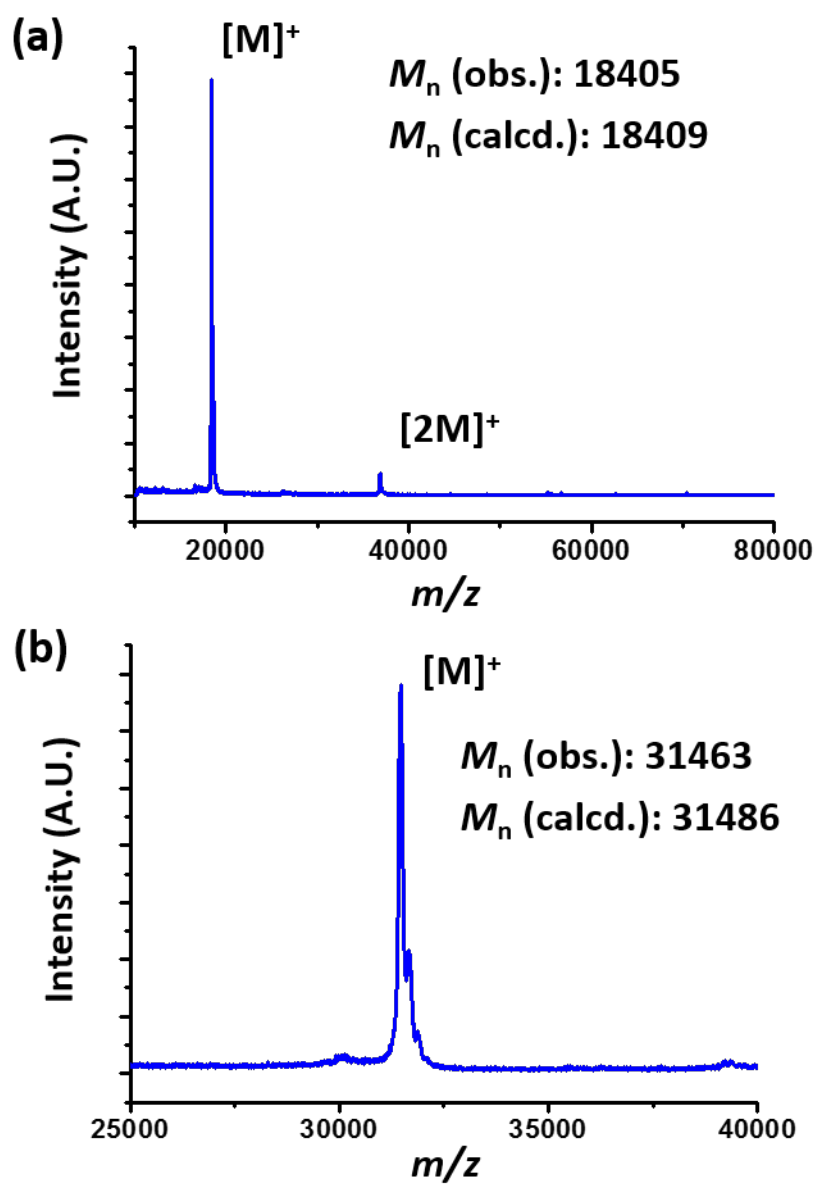
(a) EAE: 205 a.a. MW = 18409

1 MKGSSHHHHHHVDGHGVPGVGVPGVGVPGEGVPGVGVV  
41 GVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVV  
81 GEGVPGVGVPGVGVGELAHIVMV D AYKPTKTSVPGVGVPGVGV  
121 VPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGV  
161 VPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLDGPQGIWGQ  
201 LEKKM

(b) EBE: 327 a.a. MW = 31486

1 MKGSSHHHHHHVDGHGVPGVGVPGVGVPGEGVPGVGVV  
41 GVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVV  
81 GEGVPGVGVPGVGVGELIPTTENLYFQG AMVDTLSGLSSEQG  
121 QSGDMTIEEDSATHIKFS K RDEDGKELAGATMELRDSSGK  
161 TISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITF  
201 TVNEQGQVTVNGKATKGDAHIDGPQGIWGQTSVPGVGVVPG  
241 VGVPGEGVPGVGVPGVGVPGVGVPGEGVPGVGVVPGVGVV  
281 VGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLDGPQGIW  
321 GQLEKKM

**Figure S10.** Full amino acid sequence of (a) EAE and (b) EBE. The reactive aspartic acid (102D) and lysine (139K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.



**Figure S11.** MALDI-TOF mass spectra of EAE (a) and EBE (b).

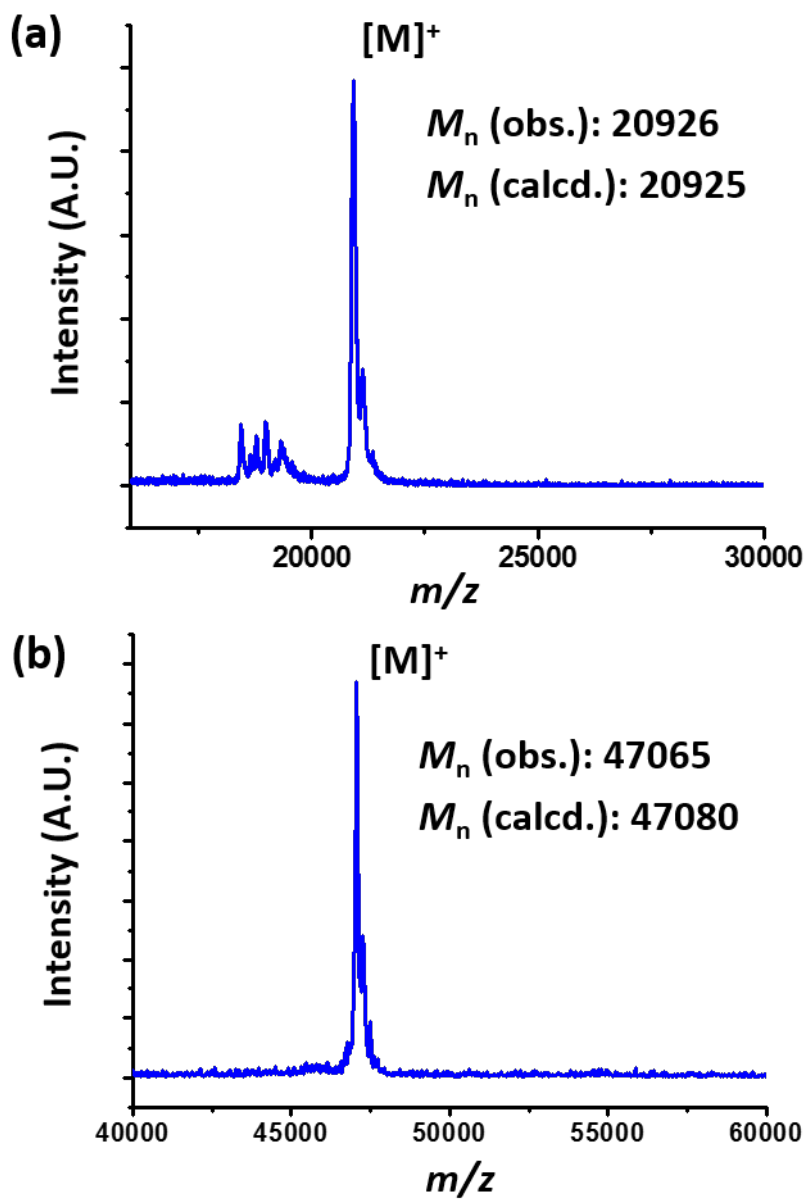
(a) AA: 229 a.a. MW = 20925

1 MKGSSHHHHHHVD **AHIVMV** **D** **AYKPTK** LDGHG VGVPGVGV  
41 GVGVP GEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGV  
81 GVGVP GGVPGVGVPGEGVPGVGVPGVGVGELYAVTGRGDSP  
121 ASSAPIATSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGV  
161 PGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGV  
201 PGVGVPGGLLD **AHIVMV** **D** **AYKPTK** LEWKK\*

(b) BB: 473 a.a. MW = 47080

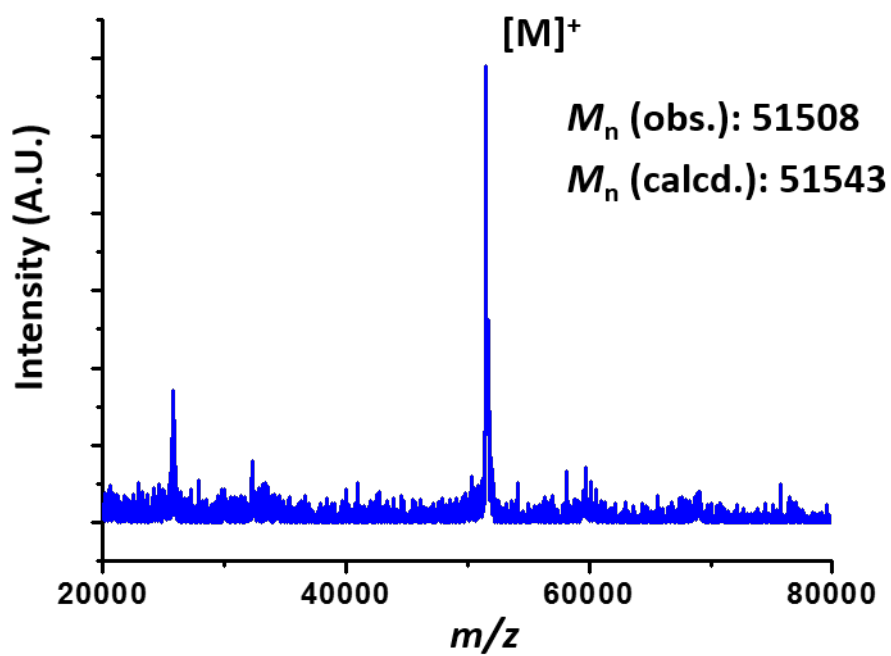
1 MKGSSHHHHHHVDIPTTENLYFQG **AMVDTLSGLSSEQQS**  
41 **GDMTIEEDSATHIKFS** **K** **RDEDGKELAGATMELRDSSGKTI**  
81 **STWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV**  
121 **NEQQQVTVNGKATKGAHID** GPQGIWGQLDGHG VGVPGVGV  
161 VPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGV  
201 VPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVGELYAVTGRGD  
241 SPASSAPIATSVPGVGVPGVGVPGEGVPGVGVPGVGVPGV  
281 GVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGV  
321 GVPGVGVPGGLLDIPTTENLYFQG **AMVDTLSGLSSEQQS**  
361 **GDMTIEEDSATHIKFS** **K** **RDEDGKELAGATMELRDSSGKTI**  
401 **STWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV**  
441 **NEQQQVTVNGKATKGAHID** GPQGIWGQLEWKK\*

**Figure S12.** Full amino acid sequence of (a) AA and (b) BB. The reactive aspartic acid (20D and 218D) and lysine (57K and 377K) are shown in the box. The segments highlighted in red and purple are SpyTag and SpyCatcher, respectively.

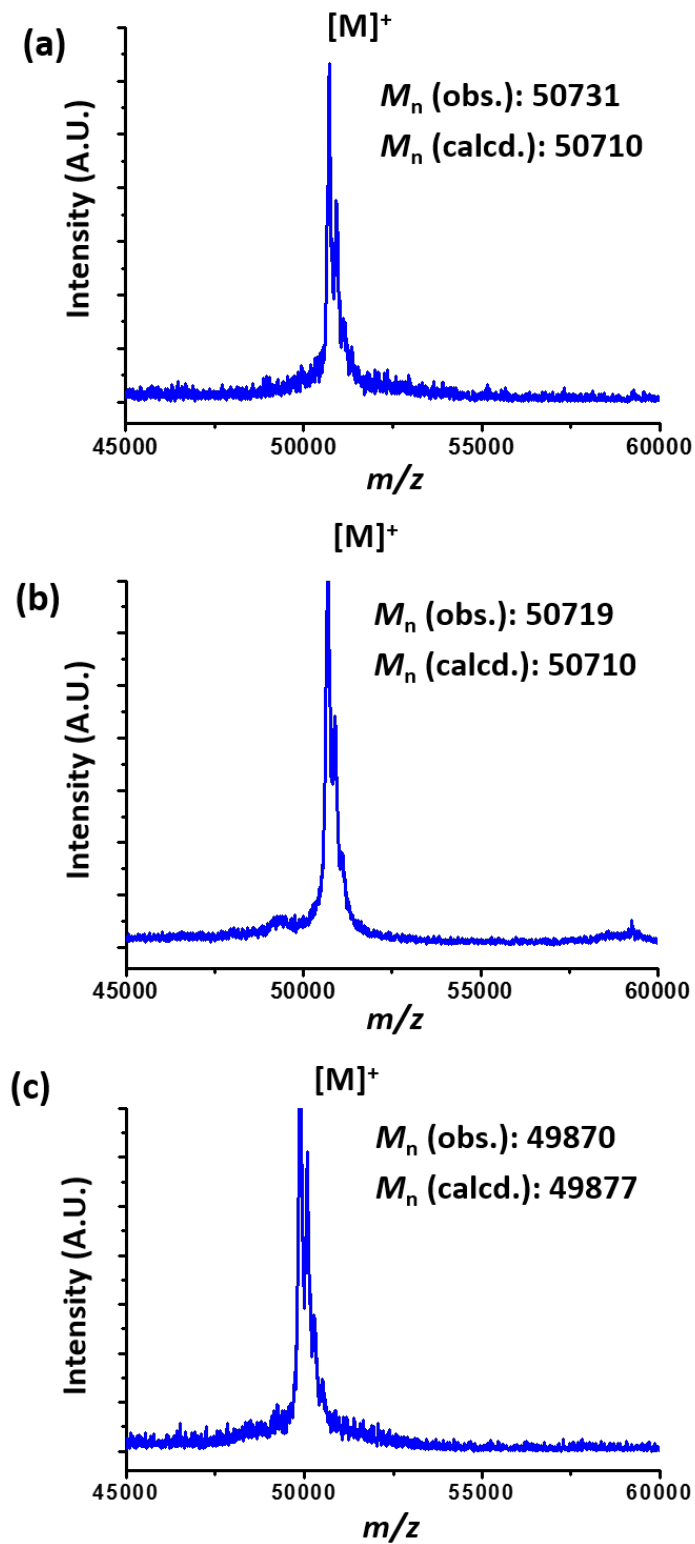


**Figure S13.** MALDI-TOF mass spectra of AA (a) and BB (b).

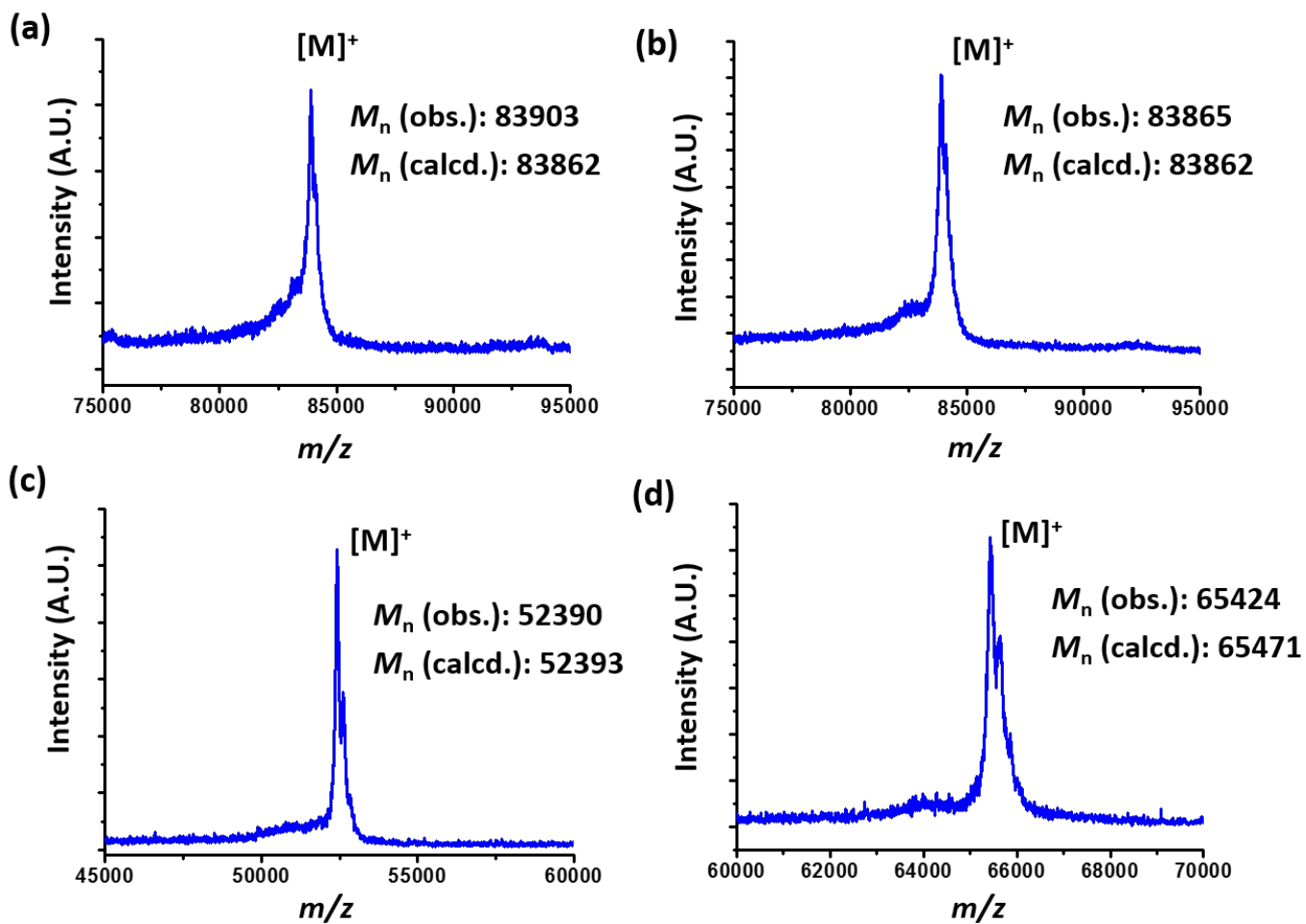




**Figure S14.** MALDI-TOF mass spectrum of the block protein (EA+EB).



**Figure S15.** MALDI-TOF mass spectra of the 3-arm star proteins EA+EBE (a) and EAE+EB (b) and the 4-arm star protein EAE+EBE (c).



**Figure S16.** MALDI-TOF mass spectra of the H-shaped proteins AA+2EBE (a) and BB+2EAE (b) and the corresponding intermediate 3-arm proteins AA+EBE (c) and BB+EBE (d).

**Table S1.** Summary of bacterial strains and plasmids used in this study

Strains	Relevant Characteristics	Sources
<i>E. coli</i>		
DH5 $\alpha$		Stratagene
BL21 star(DE3)		Invitrogen
Plasmids		
pQE-80L	T5 promoter-operator, N-terminal His tag, Amp <sup>r</sup>	Qiagen
pQE- <i>ELP</i>	The pQE-80L plasmid containing the gene encoding elastin	The starting vector for all the constructions of recombinant proteins in this study
pQE- <i>AB20D</i>	The plasmid for the expression of circular protein	This study
pQE- <i>AB20A</i>	The plasmid mutant for the expression of the control linear protein	This study
pQE- <i>EAE</i> B	The plasmid for the expression of tadpole protein	This study
pQE- <i>EA</i>	The plasmid for the expression of Elastin-SpyTag	This study
pQE- <i>EB</i>	The plasmid for the expression of Elastin-SpyCatcher	This study
pQE- <i>EAE</i>	The plasmid for the expression of Elastin-SpyTag-Elastin	This study
pQE- <i>EBE</i>	The plasmid for the expression of Elastin-SpyCatcher-Elastin	This study
pQE- <i>AA</i>	The plasmid for the expression of SpyTag-Elastin-SpyTag	This study
pQE- <i>BB</i>	The plasmid for the expression of SpyCatcher-Elastin-SpyCatcher	This study