Supplementary Information

Construction of Lasso Peptide Fusion Proteins

Chuhan Zong[1], Mikhail O. Maksimov[2], A. James Link[2,3]*
Departments of [1]Chemistry, [2]Chemical and Biological Engineering, and [3]Molecular Biology, Princeton University, Princeton, NJ 08544
*to whom correspondence should be addressed, ajlink@princeton.edu

Contents

Table S1: Plasmids used in this study

Plasmid Name	Description	Primer
pCA14	AtxB1C1	NA
pMM32	AtxA1-B1C1	NA
pQE9-A1	A1	NA
pDA38	sfGFP	NA
pJP47	$(GSSG)_5$	NA
pCZ1	Atx1-A1-BC	1,2,3,4
pCZ5	Atx1-A1-B	1,5
pCZ2	Atx1-A1-woBC	1,6
pCZ16	Atx1-$(GSSG)_5$-A1-BC	1,4,7,8,9,10
pCZ20	Atx1-$(GSSG)_5$-A1-B	1,5
pCZ21	Atx1-$(GSSG)_5$-A1-woBC	1,6
pCZ43	Atx1-$(GSSG)_5$-Thb-A1-BC	1,4,11,12
pCZ44	Atx1-$(GSSG)_5$-Thb-A1-B	1,5
pCZ45	Atx1-$(GSSG)_5$-Thb-A1-woBC	1,6
pCZ25	Atx1-$(GSSG)_2$-A1-BC	1,4,13,14
pCZ26	Atx1-$(GSSG)_2$-A1-B	1,5
pCZ27	Atx1-$(GSSG)_2$-A1-woBC	1,6
pCZ29	Atx1-$(GSSG)_2$-Thb-A1-BC	1,4,15,16
pCZ40	Atx1-$(GSSG)_2$-Thb-A1-B	1,5
pCZ39	Atx1-$(GSSG)_2$-Thb-A1-woBC	1,6
pCZ22	Atx1-$(GSSG)_5$-sfGFP-BC	1,4,17,18
pCZ46	Atx1-$(GSSG)_5$-Thb-sfGFP-BC	1,4,18,19

Table S2: Sequences of primers used in this study

Primer	Sequence
Primer 1	5- AAT CGA ATT CTA AAT GAA AGG AGA GTT CGA AAT GC -3
Primer 2	5- GTC CTG GTT GAT GCG AGA CTC T -3
Primer 3	5- TCG CAT CAA CCA GGA CGG ATC CAT GGC TAG CGG TG -3
Primer 4	5- GGA TAG ATC TTC AGT GAT GGT GAT GGT GAT GAT TAA GCT TGG CTG CAG GTC G -3
Primer 5	5- CTA CCG ATC GGT ACC TCA GAC GAC CAG -3
Primer 6	5- CCG ATC GGT ACC TCA GTG ATG GTG ATG -3
Primer 7	5- GCA TCA ACC AGG ACG GCA GCT CCG GCG G -3
Primer 8	5- CCG CCG GAG CTG CCGTCCTGGTTG ATG C -3
Primer 9	5- GGG GTA GCT CAG GCA TGG CTA GCG GTG AC -3
Primer 10	5- GTC ACC GCT AGC CAT GCC TGA GCT ACC CC -3
Primer 11	5- CTG GTT CCG CGT GGA TCT ATG GCT AGC GG -3
Primer 12	5- AGA TCC ACG CGG AAC CAG GCC TGA GCT AC -3
Primer 13	5- GCA GCT CCG GCG GTA GCA GCG GAA TGG CTA GCG -3
Primer 14	5- CCG CTG CTA CCG CCG GAG CTG CCG TCC TGG TTG -3
Primer 15	5- CAG CGG ACT GGT TCC GCG TGG ATC TAT GGC TAG CGG -3
Primer 16	5- CTA GCC ATA GAT CCA CGC GGA ACC AGT CCG CTG CTA C -3
Primer 17	5- CGG GGG TAG CTC AGG CAT GAG CAA AGG AGA AG -3
Primer 18	5- CTT CTC CTT TGC TCA TGC CTG AGC TAC CCC CG -3
Primer 19	5- CTG GTT CCG CGT GGA TCT ATG AGC AAA GGA GAA G -3

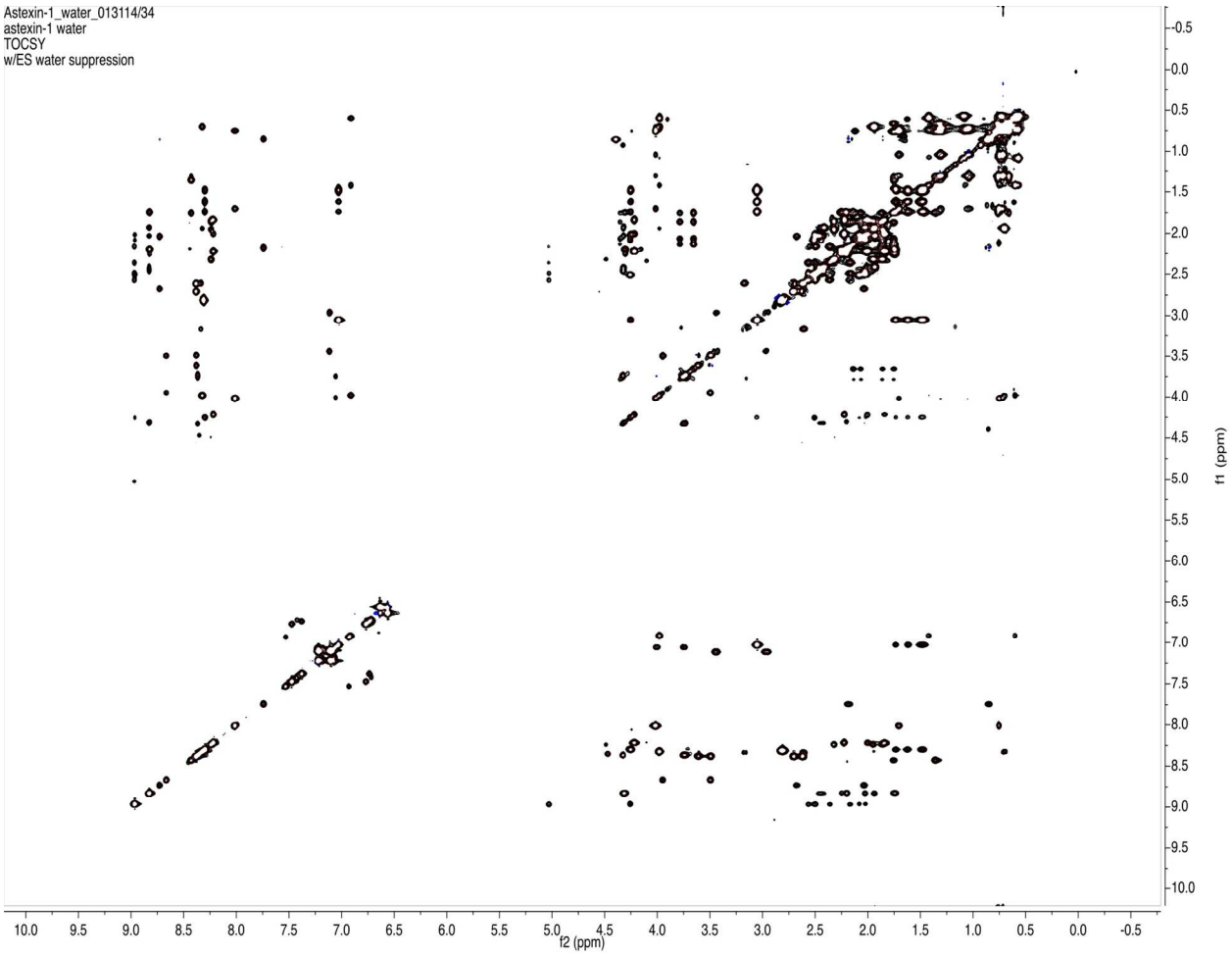

Figure S1: TOCSY spectrum of astexin-1 in H_2O/D_2O.

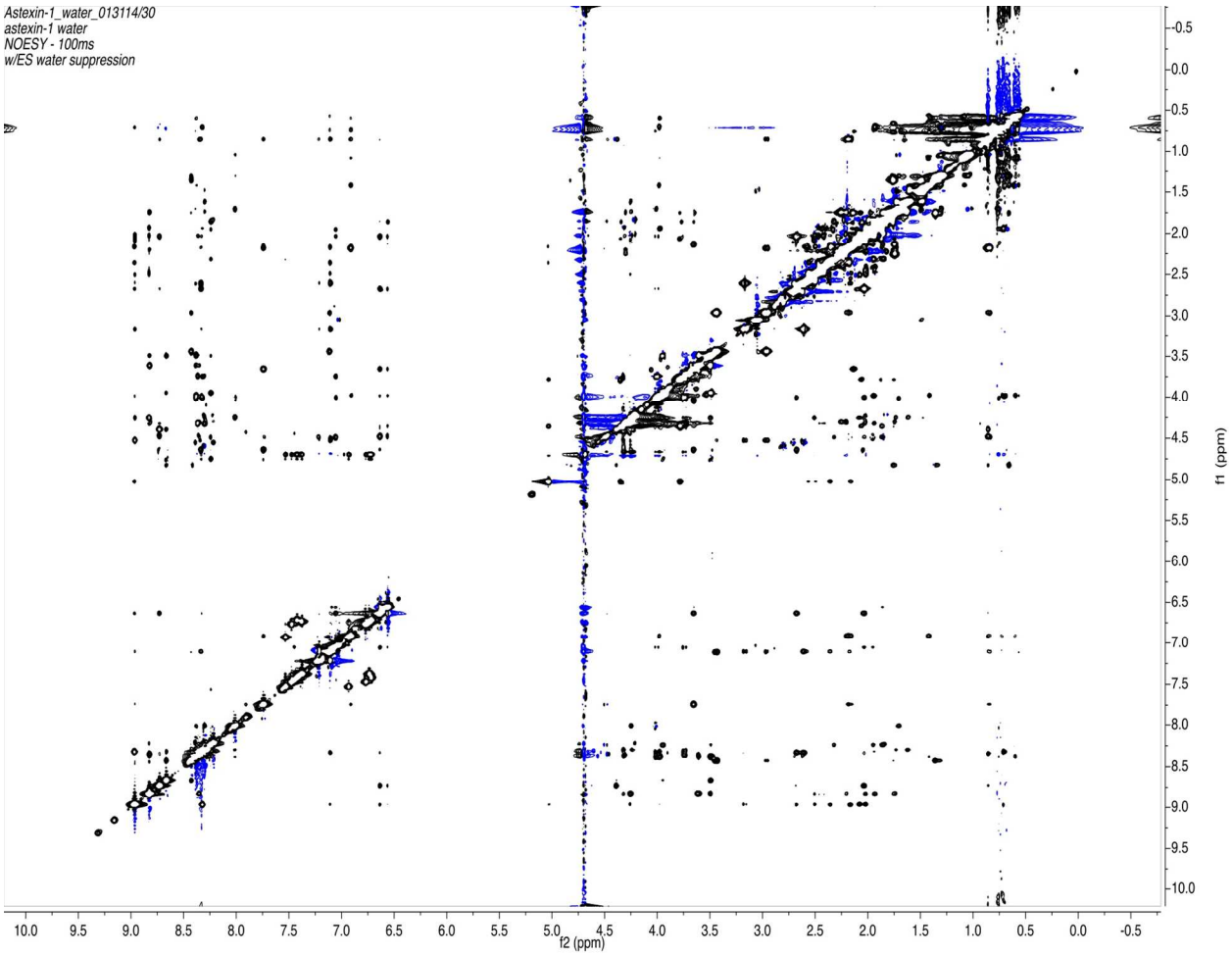

Figure S2: NOESY spectrum of astexin-1 in H_2O/D_2O, 100 ms mixing time

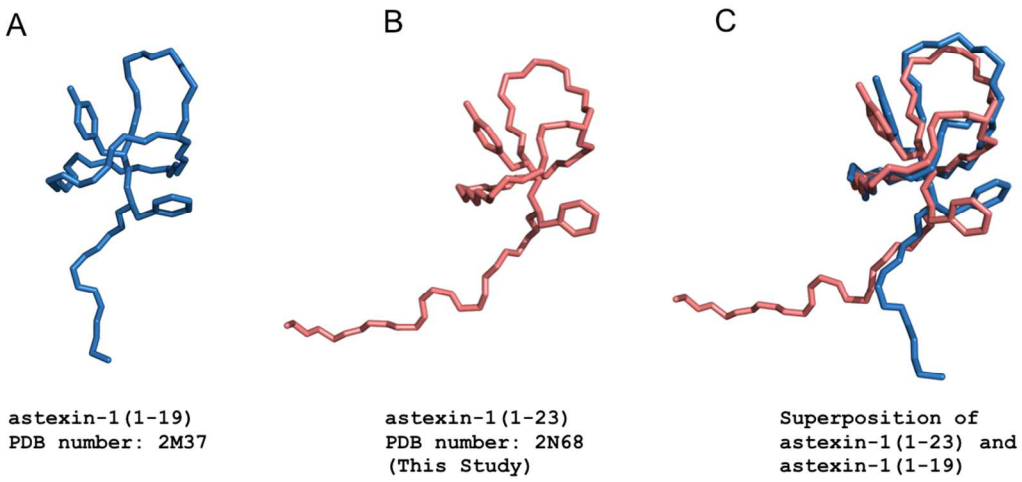

Figure S3: Comparison of astexin-1(1-23) structure to previously published astexin-1(1-19) structure. A: Lowest energy conformer of astexin-1(1-19) from PDB file 2M37. B: Lowest energy conformer of full-length astexin-1 from this work, PDB file 2N68. C: Superposition of the two structures showing their similarity.

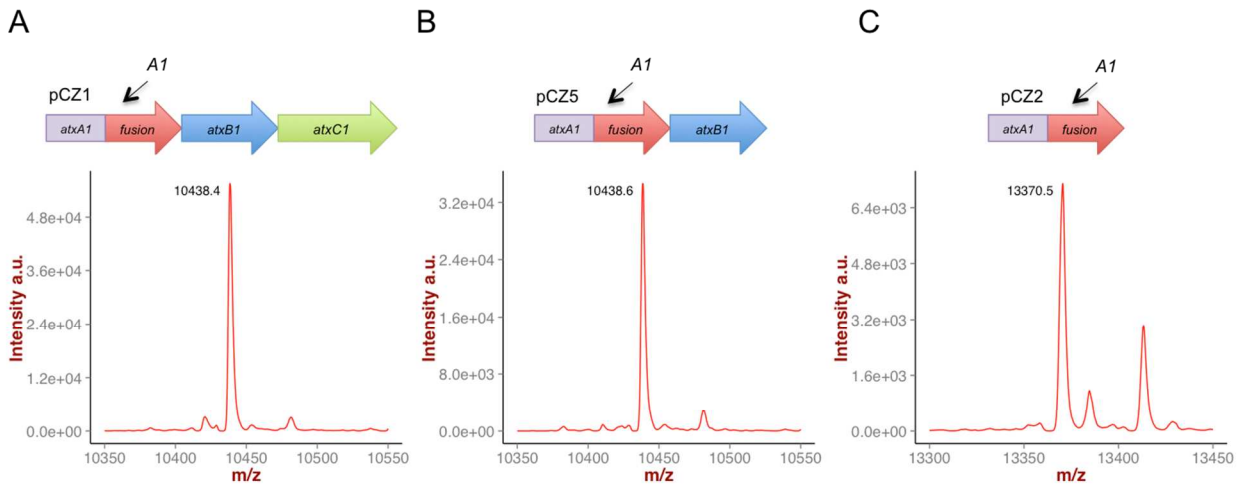

Figure S4: Electrospray mass spectra of direct fusions of astexin-1 to A1 leucine zipper. A: Deconvoluted spectrum of unlassoed astexin-1-A1 protein produced by full gene cluster in plasmid pCZ1. B: Spectrum of unlassoed protein produced by gene cluster lacking the *atxC1* gene. The leader peptide was still correctly processed but lasso cyclization did not occur. C: Spectrum of unprocessed, unlassoed protein produced by gene cluster lacking both the *atxB1* and *atxC1* genes. Refer to Table 1 for expected masses of each construct.

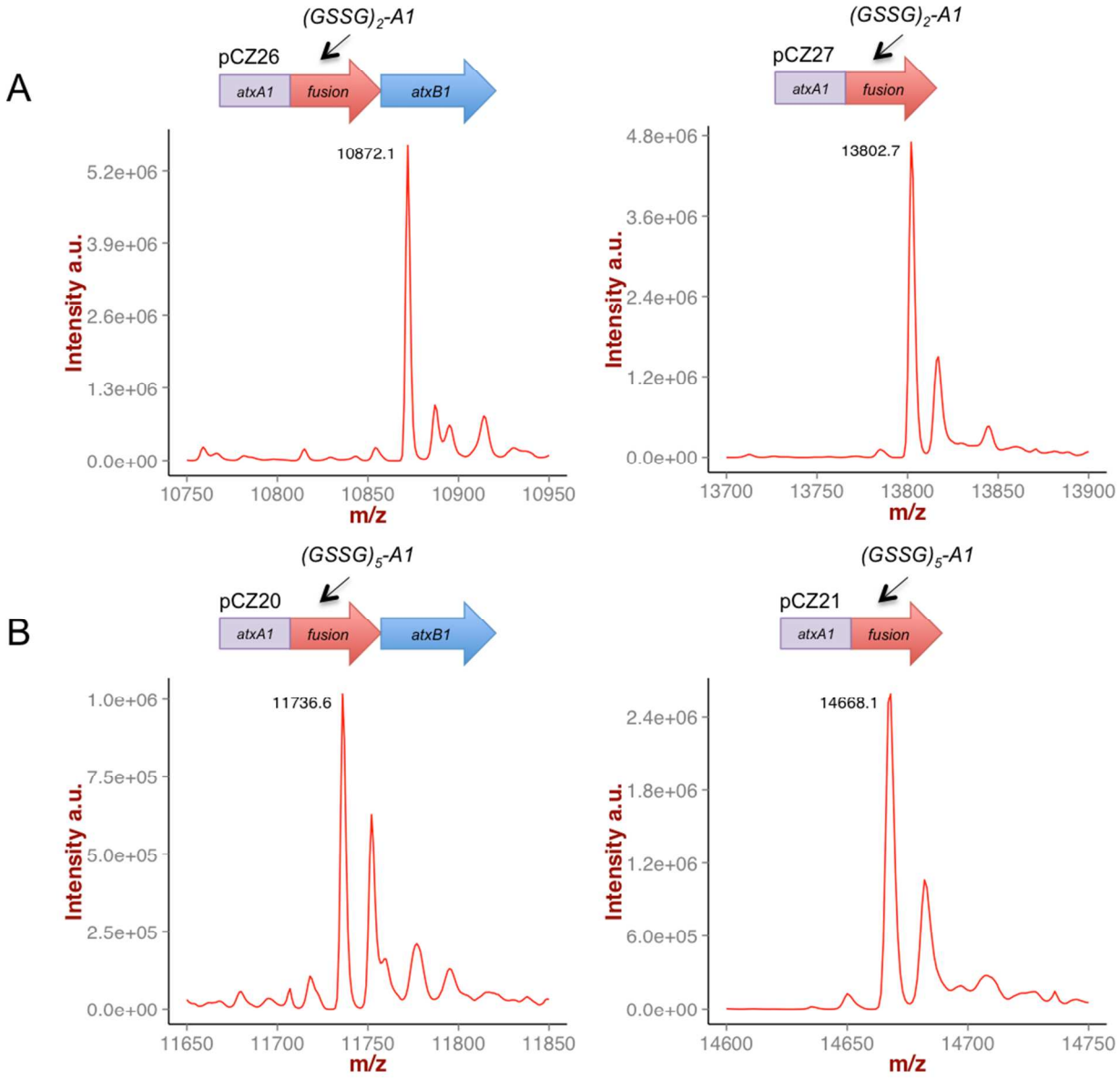

Figure S5: Deconvoluted mass spectra of astexin-1-$(GSSG)_x$-A1 fusion proteins. A: Spectra of the astexin-1-$(GSSG)_2$-A1 protein product of gene clusters lacking *atxB1* (left) or both *atxB1* and *atxC1* (right). B: As in panel A, but for the astexin-1-$(GSSG)_5$-A1 protein.

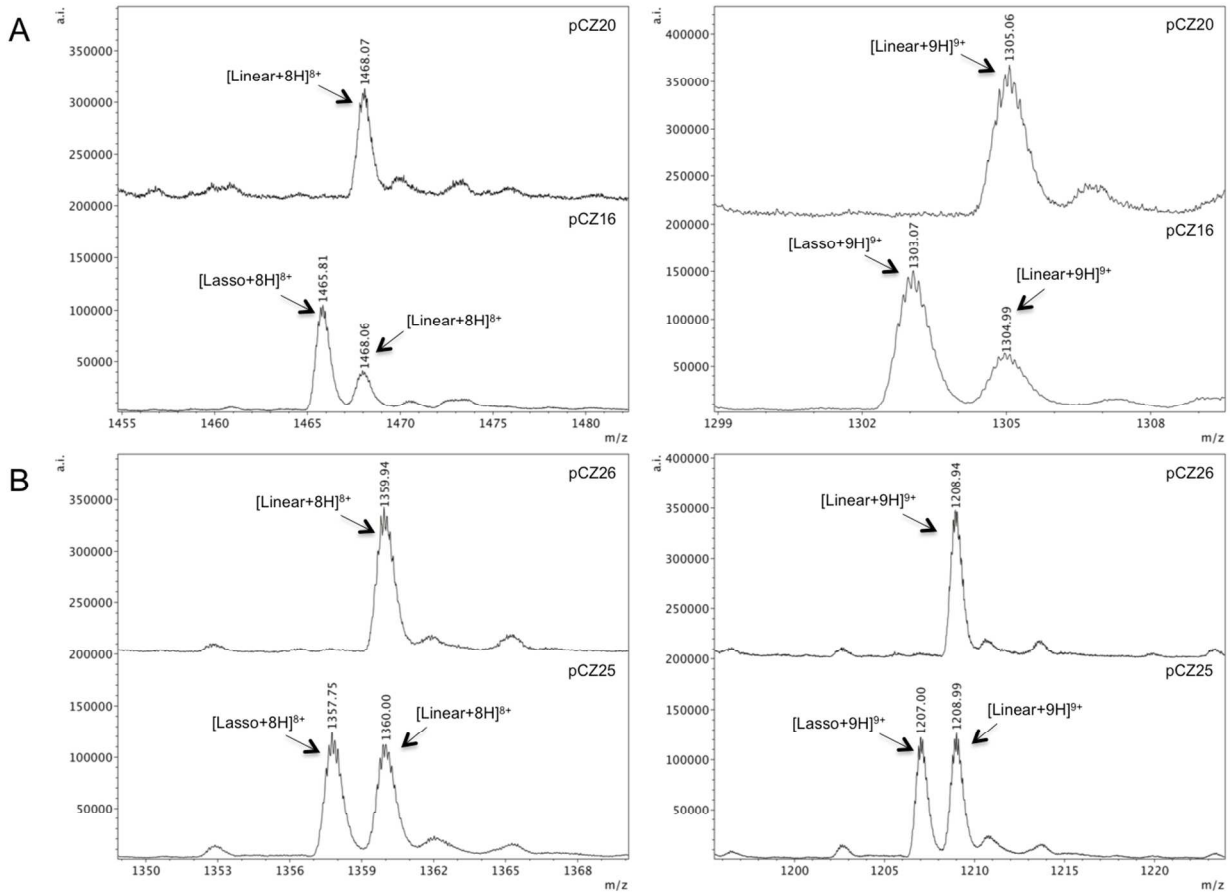

Figure S6: Raw mass spectra (before deconvolution) of the astexin-1-$(GSSG)_x$-A1 proteins. A: Spectra of intact astexin-1-$(GSSG)_5$-A1 protein in the +8 (left) and +9 (right) charge states. The top spectrum is of the protein product of the gene cluster lacking *atxC1* (plasmid pCZ20) while the bottom spectrum corresponds to the full gene cluster, housed on plasmid pCZ16. The difference between the mass of the lassoed and linear proteins are clearly discernible. B: As in panel A, but for the astexin-1-$(GSSG)_2$-A1 protein. The plasmid pCZ26 is the gene cluster lacking *atxC1* and pCZ25 is the full gene cluster. This data was used to generate the deconvoluted spectra in Figures 2 and S5.

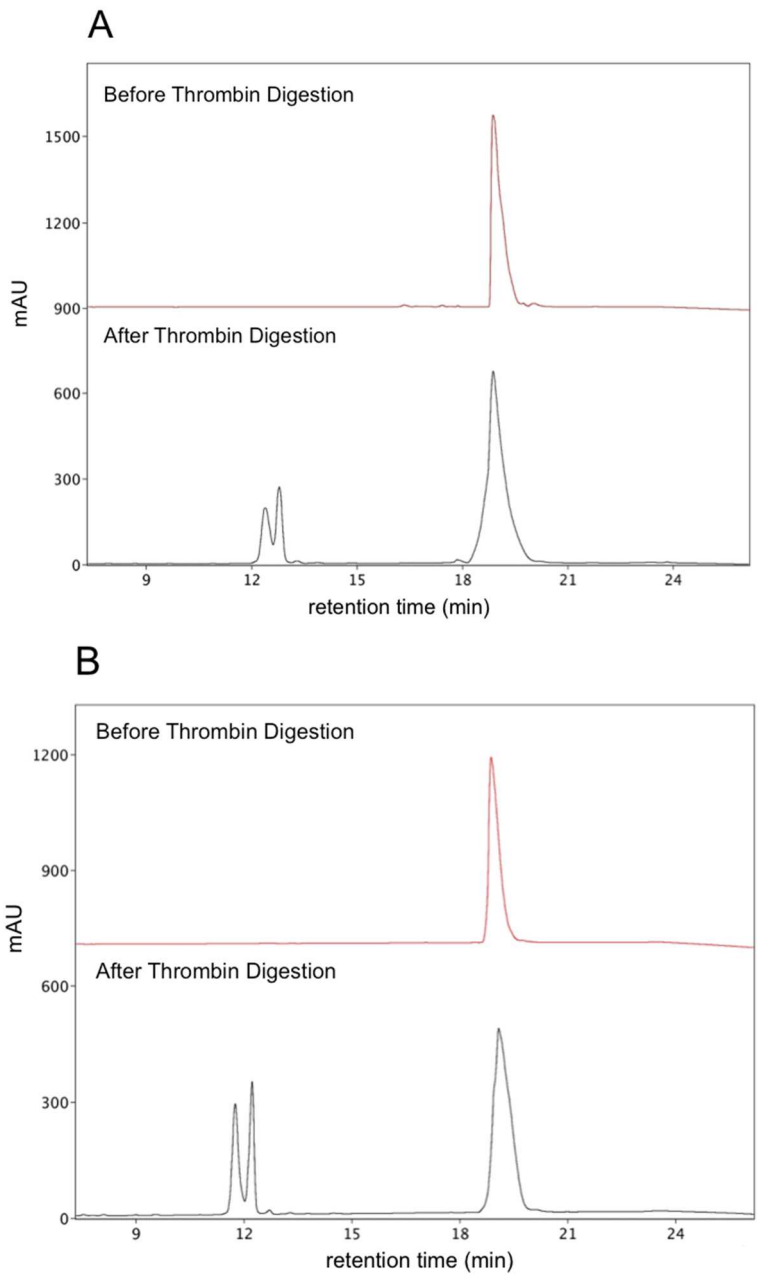

Figure S7: Thrombin digestion of astexin-1-$(GSSG)_x$-thrombin-A1 proteins. A: HPLC traces of astexin-1-$(GSSG)_2$-thrombin-A1 protein before (top) and after (bottom) thrombin digestion. The peaks for the lasso peptide and its linear counterpart are readily resolved. B: As in part A, but for the astexin-1-$(GSSG)_5$-thrombin-A1 protein.

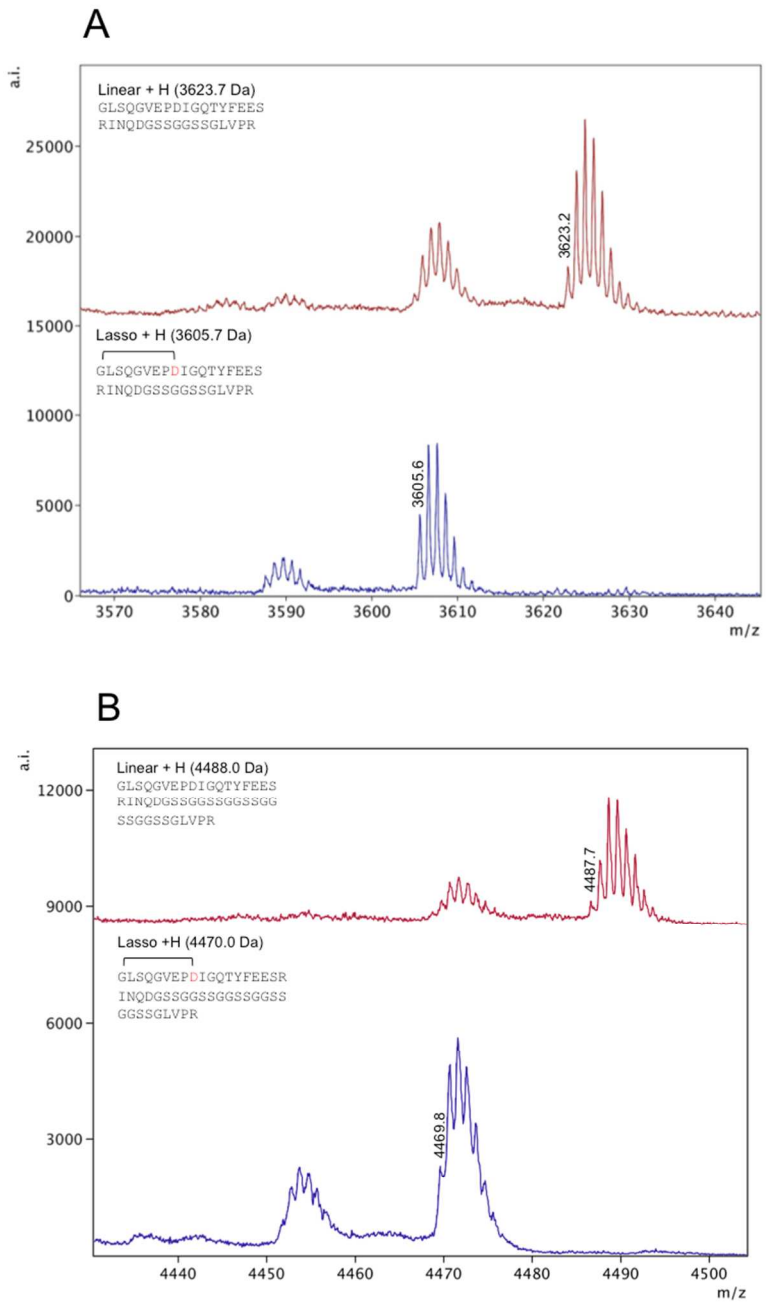

Figure S8: MALDI-MS analysis of thrombin-cleaved lasso and linear peptides. A: Spectrum of purified linear peptide (top) and lasso peptide (bottom) after thrombin digestion of the astexin-1-$(GSSG)_2$-thrombin-A1 construct. B: As in A, but for the astexin-1-$(GSSG)_5$-thrombin-A1 protein. The lower intensity peaks 18 mass units below the labeled peaks are from MALDI-induced dehydration due to the high laser intensity needed to analyze these high molecular weight species.

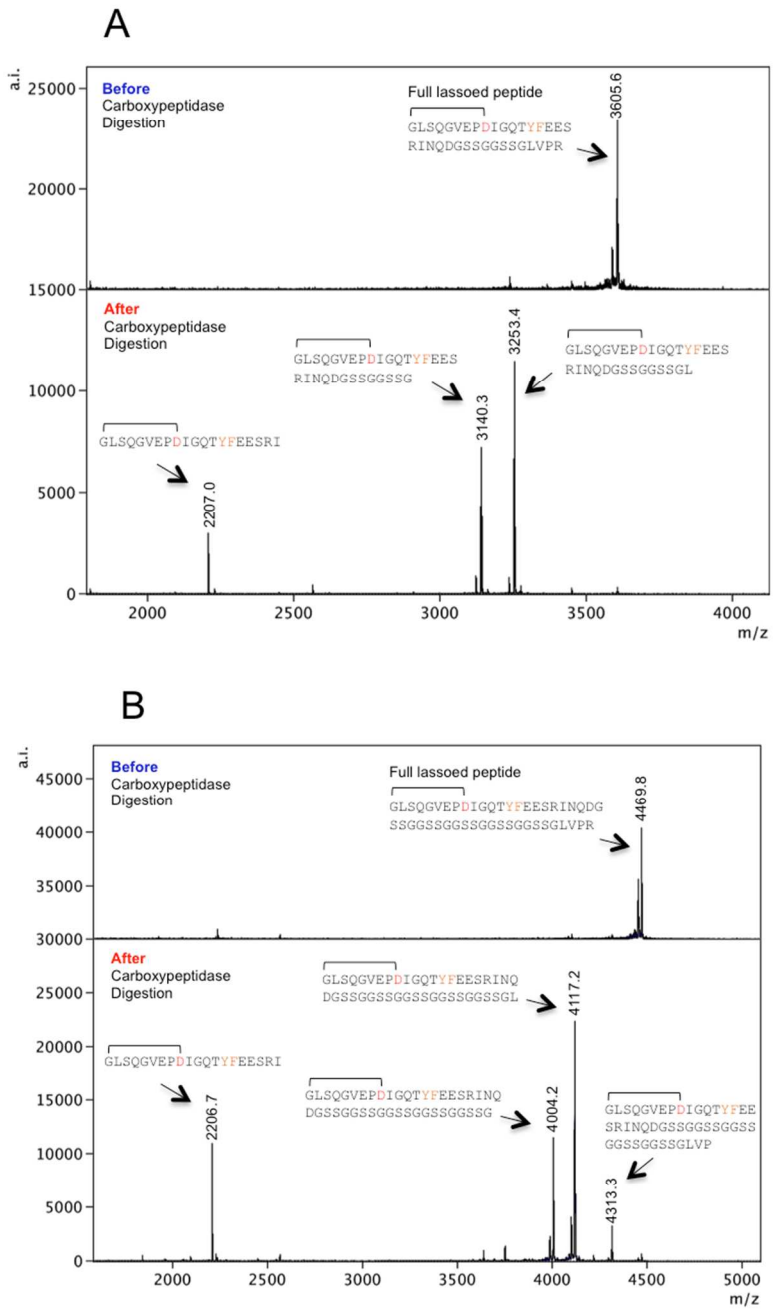

Figure S9: MALDI-MS analysis of carboxypeptidase-digested lasso peptides. Following thrombin cleavage and HPLC purification, the tail-extended astexin-1 peptides were digested with carboxypeptidase B and Y to confirm formation of the lasso structure. A: Lasso peptide derived from the astexin-1-$(GSSG)_2$-thrombin-A1 protein B: Lasso peptide derived from the astexin-1-$(GSSG)_5$-thrombin-A1 protein. In both cases, the carboxypeptidase digestion stops at residue Ile-20 due to steric hindrance of the lasso fold.

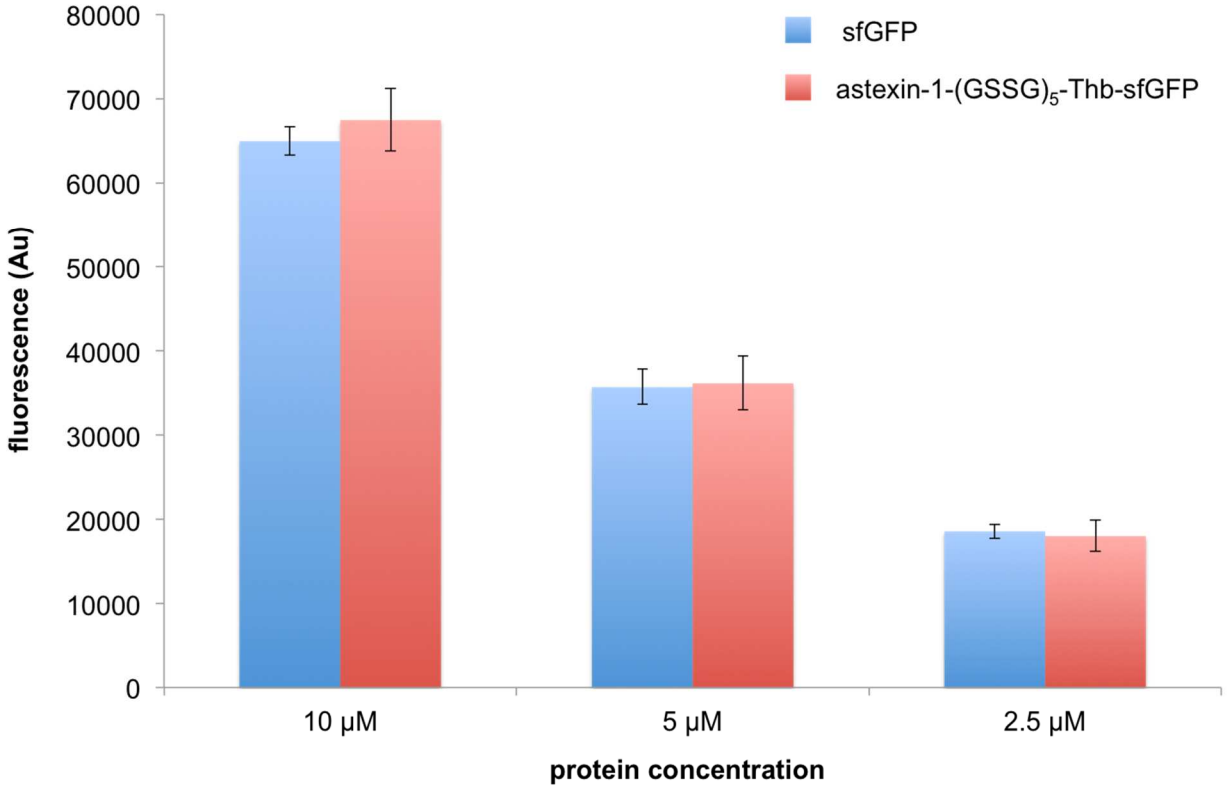

Figure S10: Comparison of the fluorescence of astexin-1-$(GSSG)_5$-thrombin-superfolder GFP (red) to native superfolder GFP (blue).