

# Generation of Proteins with Free N-Terminal Cysteine by Aminopeptidases

Jordan P. Hempfling, Emily R. Sekera, Amar Sarkar, Amanda B. Hummon, and Dehua Pei\*

## Supporting Information

### Table of Contents

Table S1.....	S2
Table S2.....	S2
Table S3.....	S2
Figure S1.....	S3
Figure S2.....	S4
Figure S3.....	S18

**Table S1.** Peptide Sequences Derived from Intensely Colored Beads during Library Screening against *A. sobria* ProAP (Most Preferred Substrates; total 28 sequences)<sup>a</sup>

GNRS	GFKG	AAGQ	ALKH	EFAR
GHKF	GGKU	AAAR	ALKN	FGSV
GHYA	GFSF	AURU	AAKF	RNYN
GGKA <sup>b</sup>	AALK	ASKA <sup>b</sup>	SFKG	MIHE
GGRA	AHKA <sup>b</sup>	AFKN	SHKG <sup>b</sup>	
GNFK	ALAR	AGVR	EFTR <sup>b</sup>	

<sup>a</sup>Selected from three screening experiments at 100 nM, 300 nM, and 1.0  $\mu$ M *A. sobria* ProAP. U,  $\alpha$ -L-aminobutyric acid; M, norleucine.

<sup>b</sup>Peptides selected for kinetic analysis.

**Table S2.** Peptide Sequences Derived from Medium Colored Beads during Library Screening against *A. sobria* ProAP (Preferred Substrates; total 56 sequences)<sup>a</sup>

GAGK	GRAU	GFHG	AAGK	PARV
GYLS	GHNK	GQUY	UAKS	ERKG
GAUR	GYGI	GAFI	UUKN	EIGR
GYAF	GNKY	GNKG	UASR	RHTR
GHFG	GLKI	GHAG	UMKG	RHFI
GFGY	GGRV	GVYG	SLKM	RTNL
GMLY	GYUN	GALK	SLYP	KPAK
GSSK <sup>b</sup>	GFIG	AHKK	SUKG	YEML
GAIK	GGKF	AHKN	SFKN	YRDU
GUGR	GRAL	AUKL	SAKK	FRGK
GYSK	GRGA	AMIK	SNKG	LUFN
GMGK				

<sup>a</sup>Selected from three screening experiments performed at 100, 300, and 1000 nM *A. sobria* ProAP. U,  $\alpha$ -L-aminobutyric acid; M, norleucine.

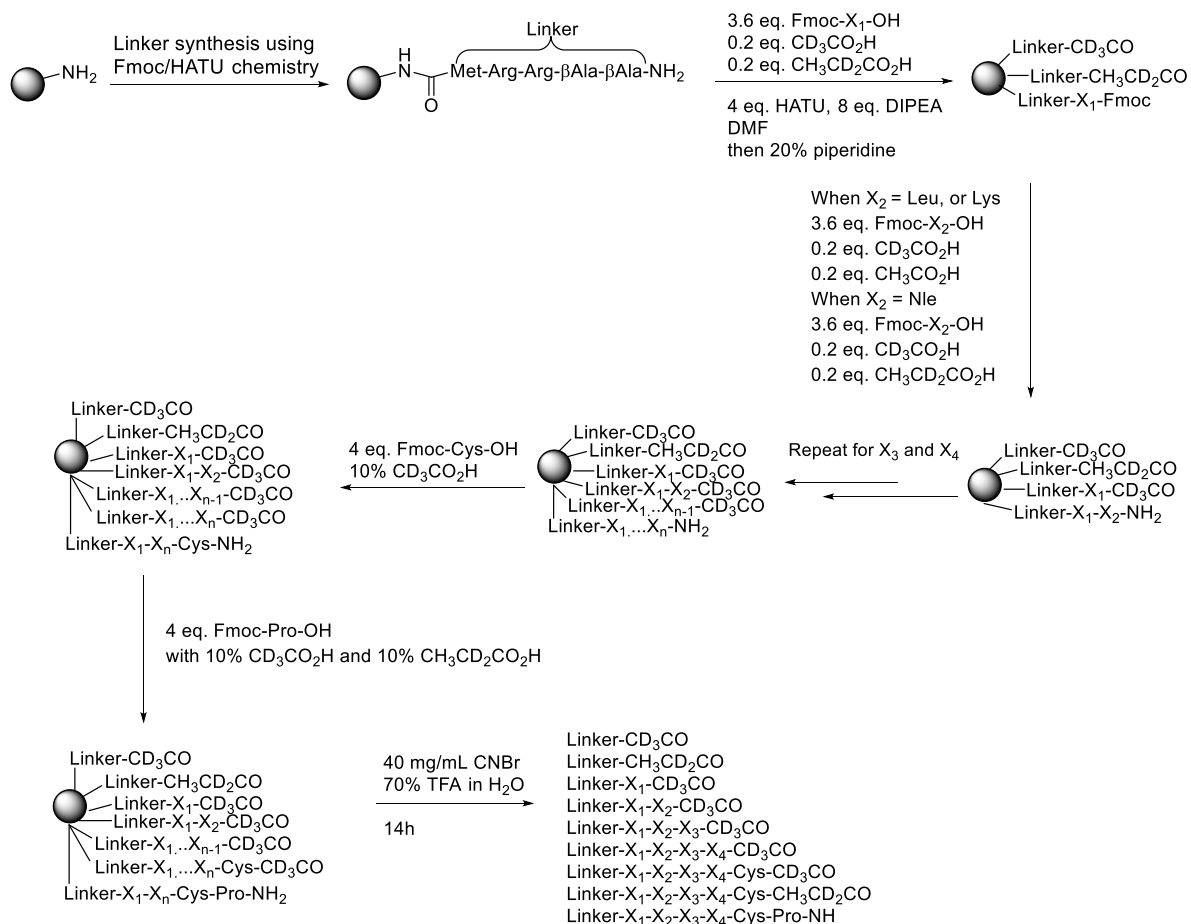
<sup>b</sup>Peptide selected for kinetic analysis.

**Table S3.** Peptide Sequences Derived from Randomly Selected Colorless Beads during Library Screening against *A. sobria* ProAP (Poor Substrates; total 25 sequences)<sup>a</sup>

ADDR	EUVD	HDLA	LQKI	YENI
APPE	ELLU	IDIS	VQNI	
AMVE	HDET <sup>b</sup>	IUAU	VNAK	
UVPN	HIDS	ITAU	KEUT	
DKPK	HAYV	ITNF	QUUI	
DFVD	HPQD	LGSP	TUVV	

<sup>a</sup>Selected from the screening experiment with 1000 nM *A. sobria* ProAP. U,  $\alpha$ -L-aminobutyric acid; M, norleucine.

<sup>b</sup>Peptide selected for kinetic analysis.

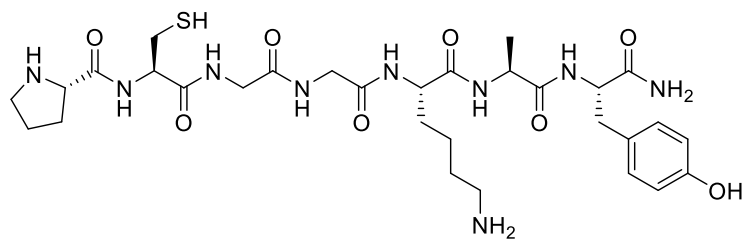


**Figure S1.** Scheme showing the synthesis of resin-bound ProAP substrate peptide library. HATU, (1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate; DIPEA, diisopropylethylamine.

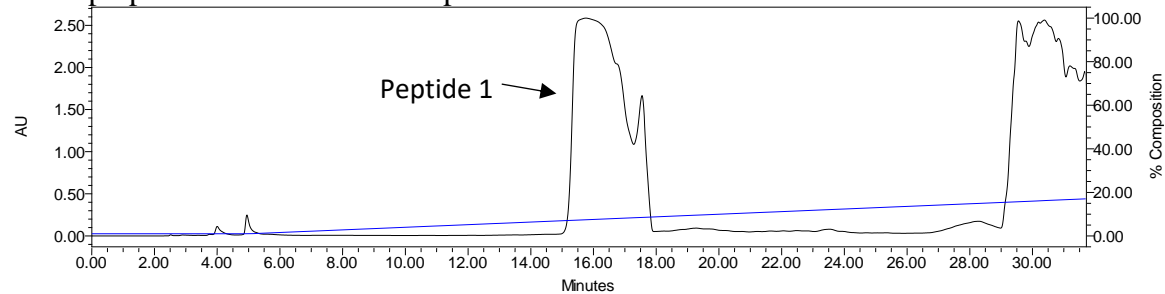
**Figure S2.** Structures, preparative HPLC chromatograms, UPLC chromatograms, and mass spectra of compounds used in this work. HRMS analysis was performed by MALDI FT-ICR mass spectrometry.

Compound ID	Molecular formula [M+H] <sup>+</sup>	Purity (%)	Expected Mass [M+H] <sup>+</sup>	Observed Mass [M+H] <sup>+</sup>
<b>1</b>	C <sub>30</sub> H <sub>48</sub> N <sub>9</sub> O <sub>8</sub> S <sup>+</sup>	95	694.3341	694.3338
<b>2</b>	C <sub>31</sub> H <sub>50</sub> N <sub>9</sub> O <sub>10</sub> S <sup>+</sup>	96	740.3396	740.3391
<b>3</b>	C <sub>32</sub> H <sub>52</sub> N <sub>9</sub> O <sub>9</sub> S <sup>+</sup>	96	738.3603	738.3596
<b>4</b>	C <sub>34</sub> H <sub>51</sub> N <sub>11</sub> O <sub>9</sub> S <sup>+</sup>	97	790.3665	790.3664
<b>5</b>	C <sub>41</sub> H <sub>60</sub> N <sub>11</sub> O <sub>11</sub> S <sup>+</sup>	100	914.4189	914.4194
<b>6</b>	C <sub>35</sub> H <sub>54</sub> N <sub>11</sub> O <sub>8</sub> S <sup>+</sup>	100	788.3872	788.3868
<b>7</b>	C <sub>37</sub> H <sub>59</sub> N <sub>12</sub> O <sub>8</sub> S <sup>+</sup>	98	831.4294	831.4291
<b>8</b>	C <sub>36</sub> H <sub>54</sub> N <sub>11</sub> O <sub>8</sub> S <sup>+</sup>	98	800.3872	800.3882
<b>9</b>	C <sub>34</sub> H <sub>52</sub> N <sub>11</sub> O <sub>8</sub> S	99	774.3716	774.3713
<b>10</b>	C <sub>37</sub> H <sub>56</sub> N <sub>11</sub> O <sub>8</sub> S <sup>+</sup>	96	814.4029	814.4025
<b>11</b>	C <sub>36</sub> H <sub>51</sub> N <sub>10</sub> O <sub>13</sub> S <sup>+</sup>	98	863.3352	863.3351
<b>Dabcyl-CBT</b>	C <sub>25</sub> H <sub>22</sub> N <sub>7</sub> O <sub>2</sub> S <sup>+</sup>	-	484.1550	484.1547
<b>CPP12-CBT</b>	C <sub>84</sub> H <sub>116</sub> N <sub>27</sub> O <sub>15</sub> S <sup>+</sup>	97	1774.8859	1774.8911

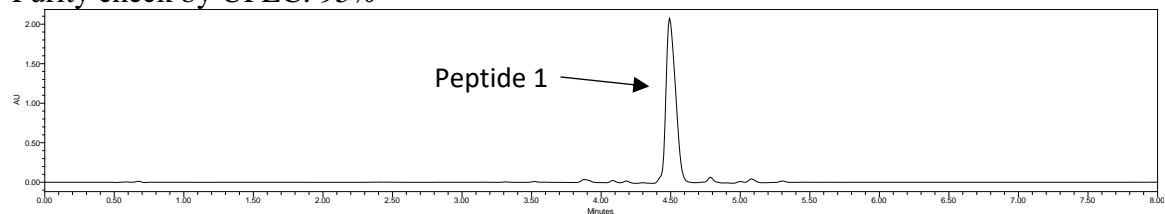
## Peptide 1



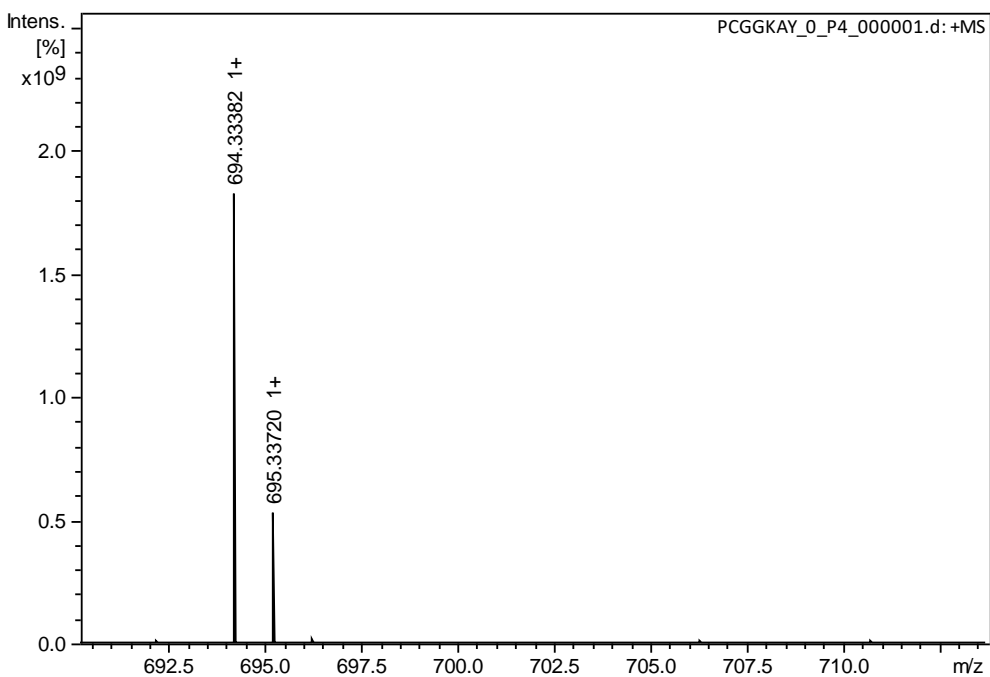
## Semi-preparative HPLC of crude product



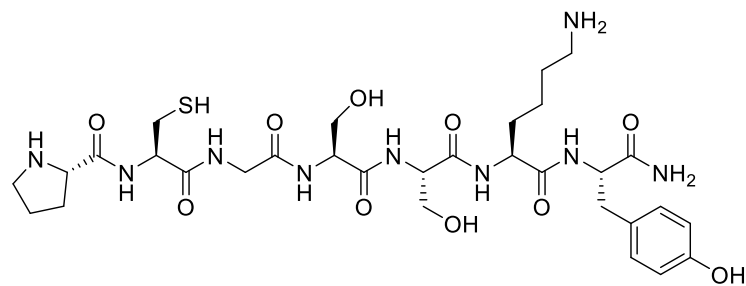
## Purity check by UPLC: 95%



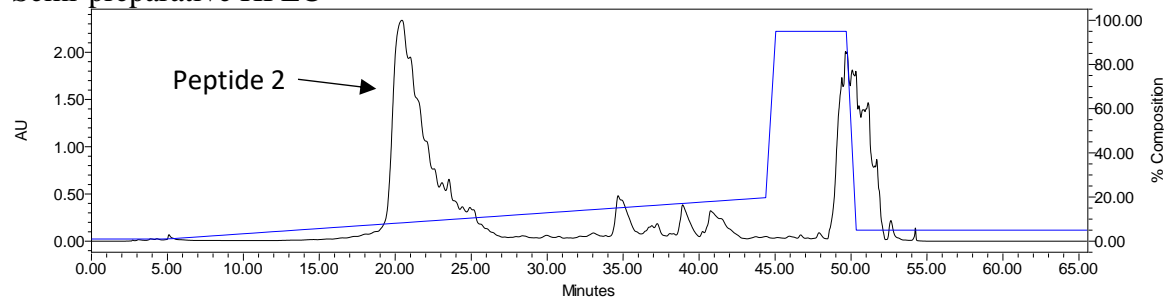
HRMS: Calculated  $[M+H]^+$  694.3341, observed 694.3338



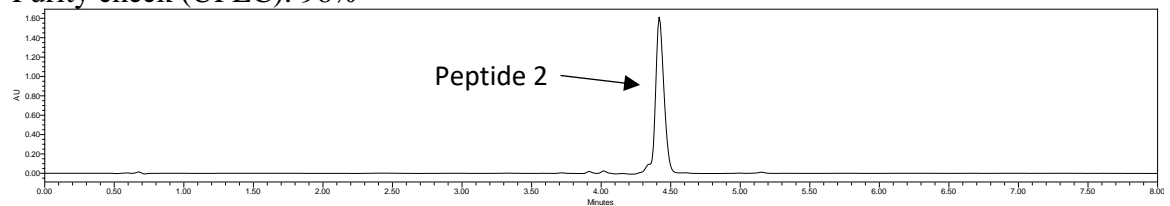
## Peptide 2



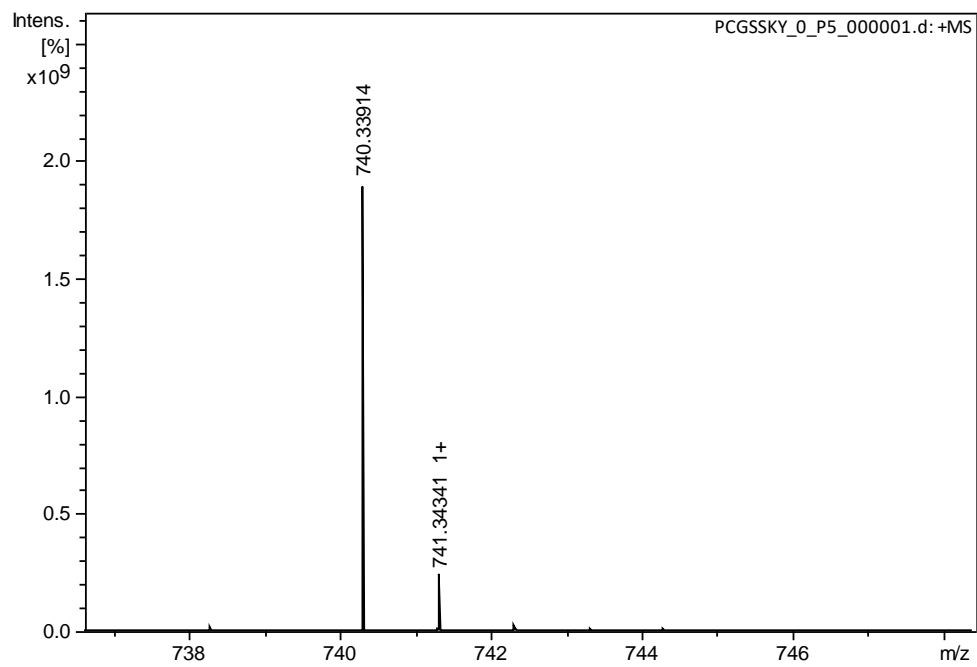
### Semi-preparative HPLC



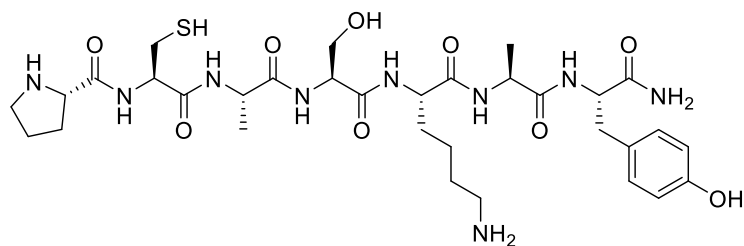
### Purity check (UPLC): 96%



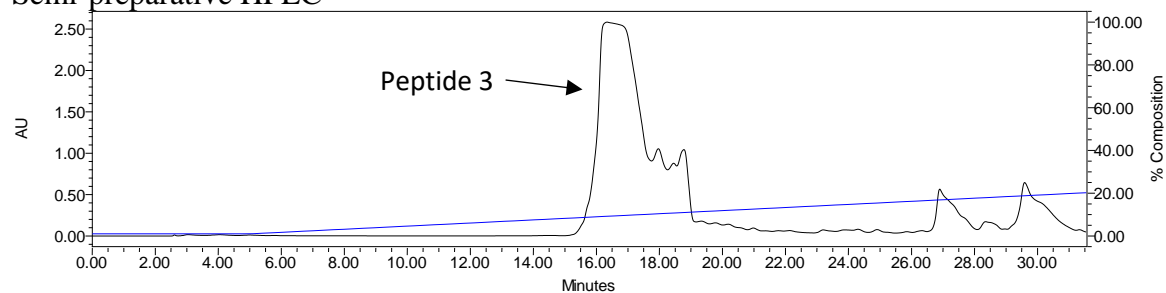
HRMS: Calculated  $[M+H]^+$  740.3396, observed 740.3391



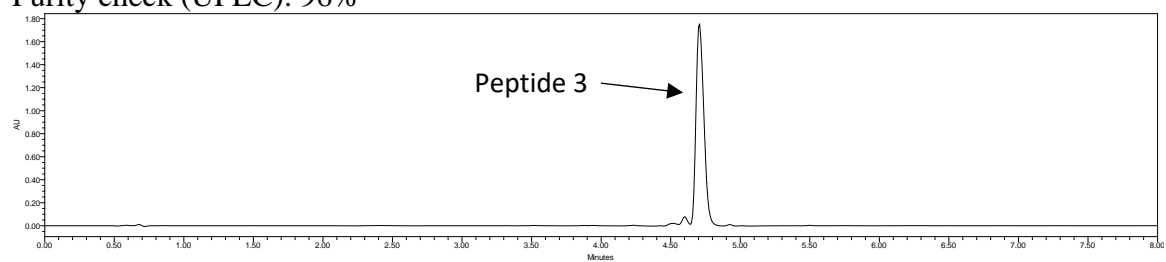
## Peptide 3



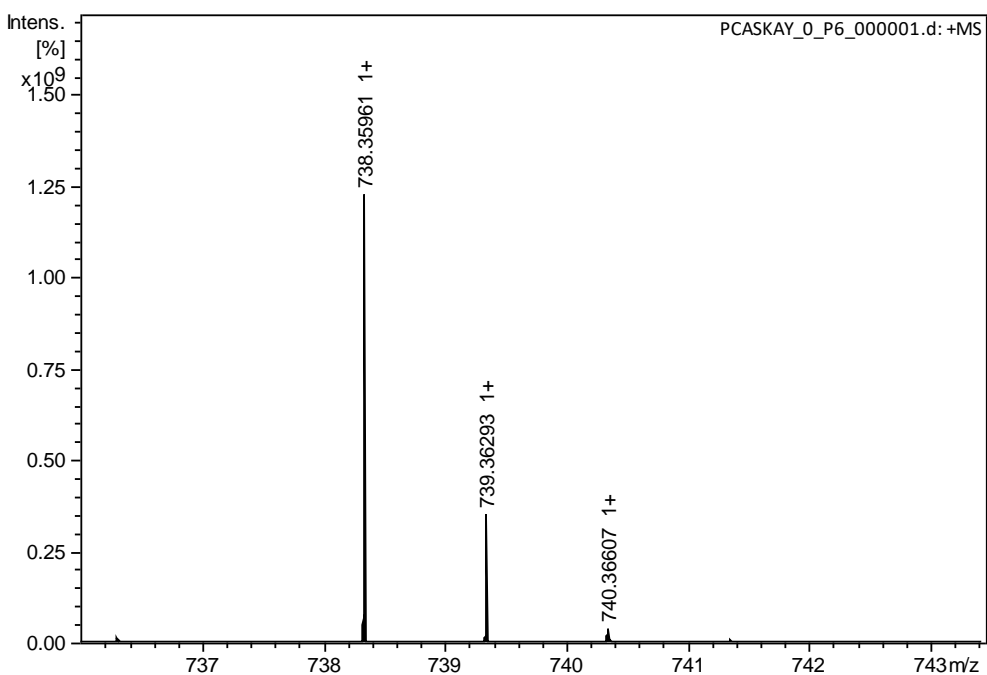
## Semi-preparative HPLC



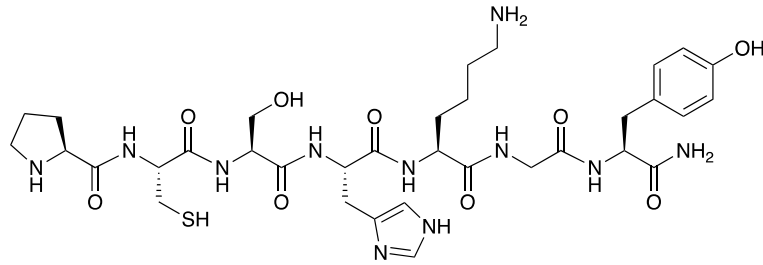
## Purity check (UPLC): 96%



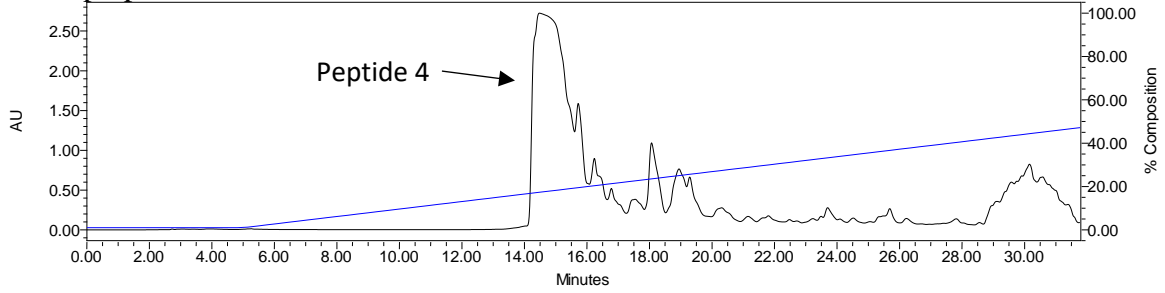
HRMS: Calculated  $[M+H]^+$  738.3603, observed 738.3596



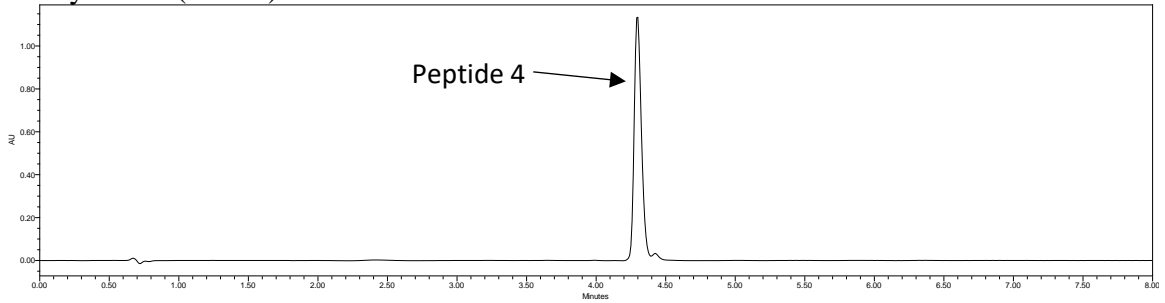
# Peptide 4



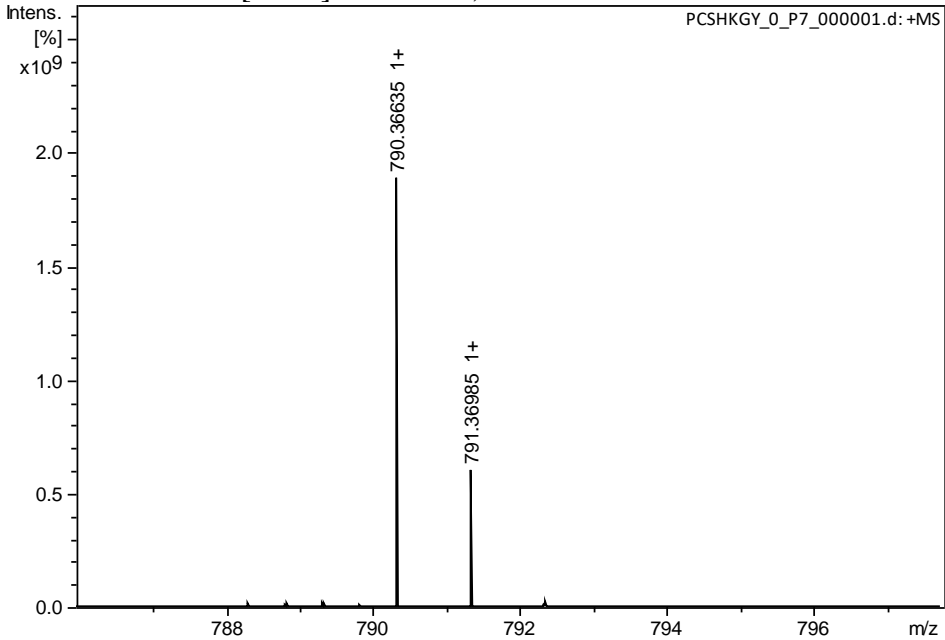
## Semi-preparative HPLC



## Purity check (UPLC): 97%

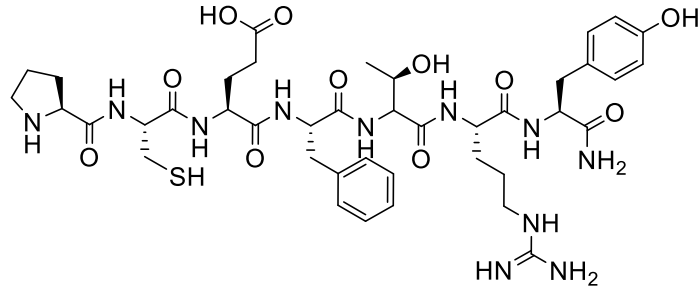


## HRMS: Calculated $[M+H]^+$ 790.3665, observed 790.3664

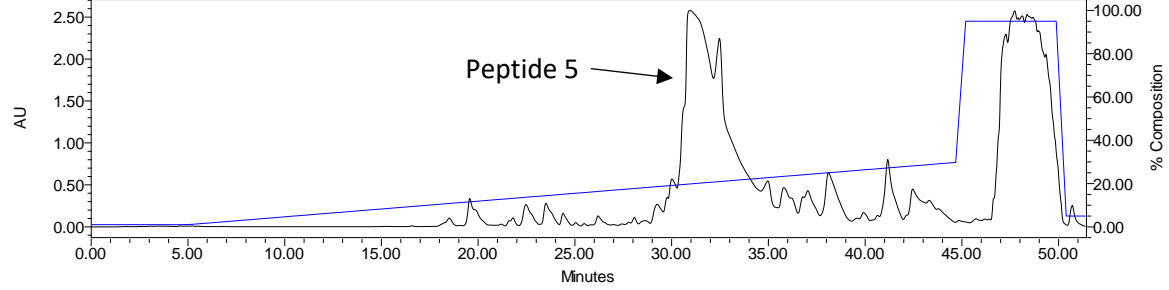




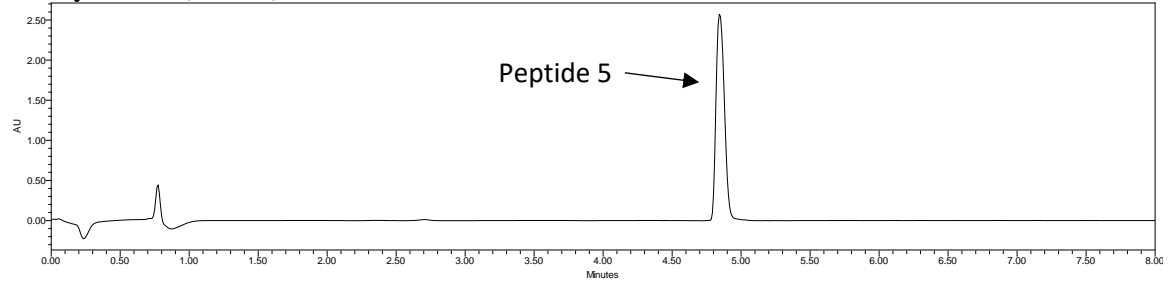
## Peptide 5



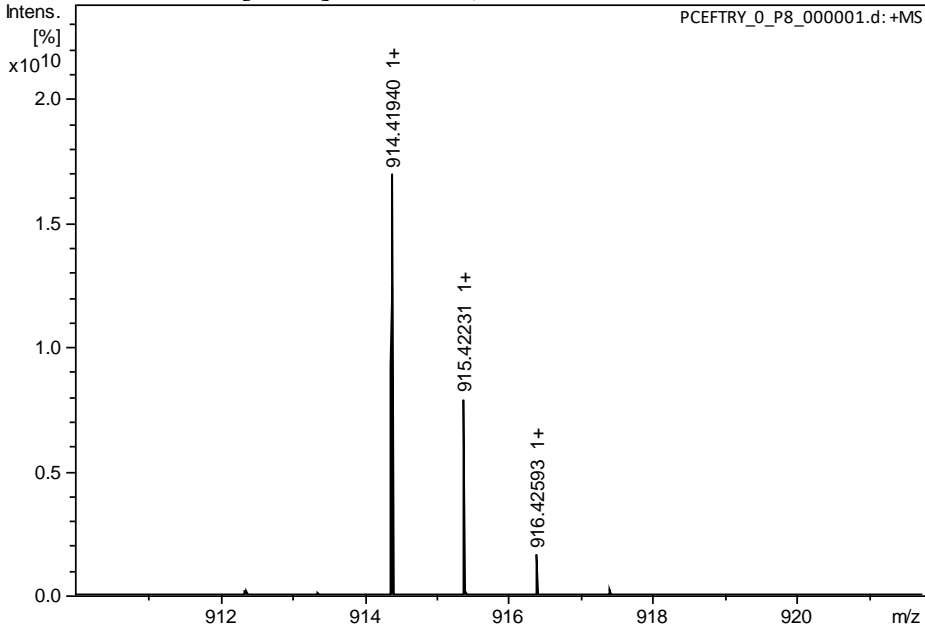
## Semi-preparative HPLC



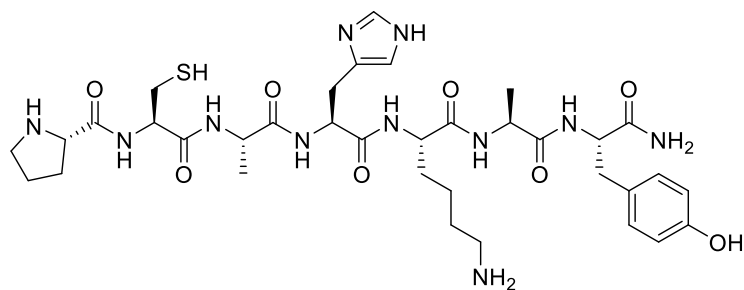
## Purity check (UPLC): 100%



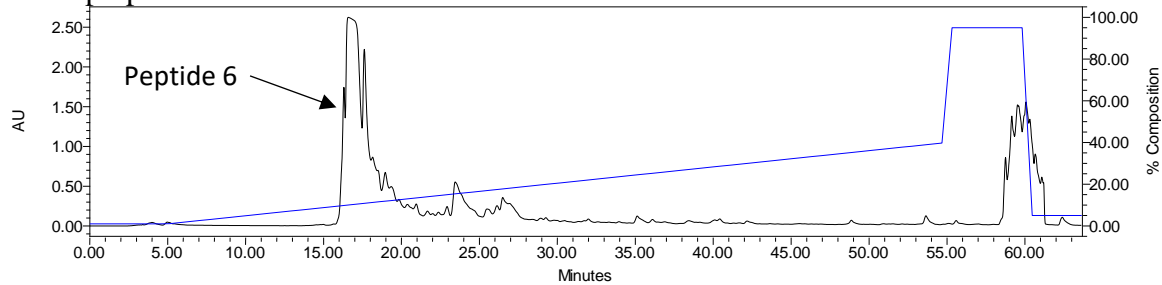
## HRMS: Calculated $[M+H]^+$ 914.4189, observed 914.4194



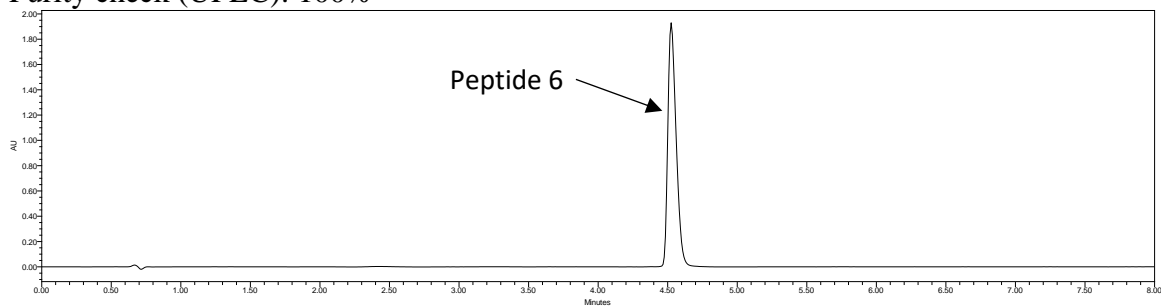
## Peptide 6



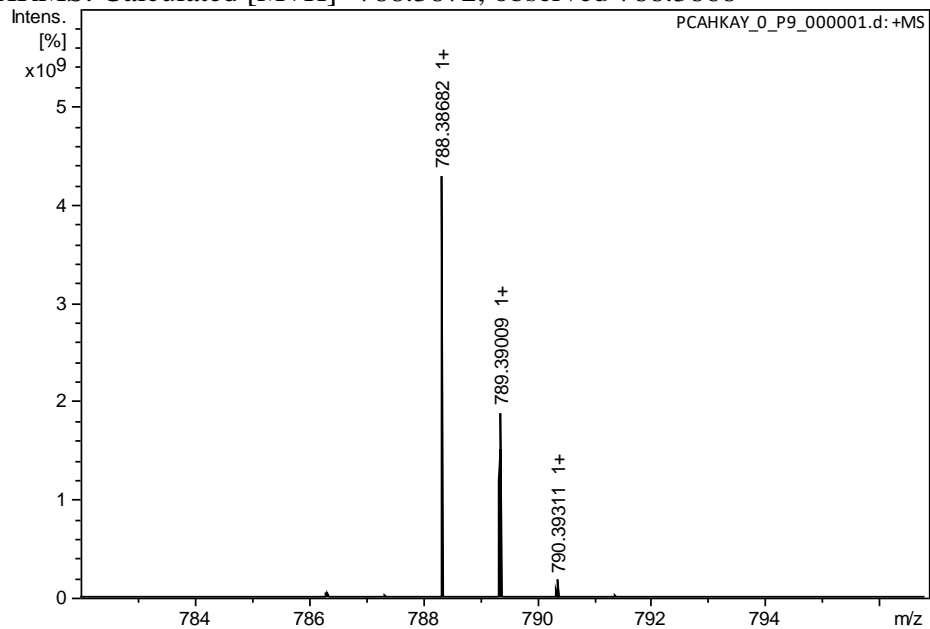
## Semi-preparative HPLC



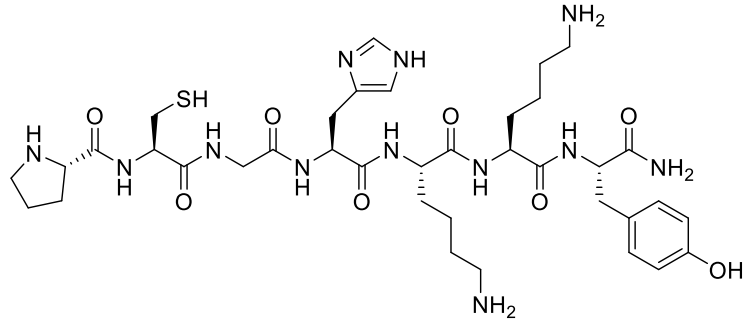
## Purity check (UPLC): 100%



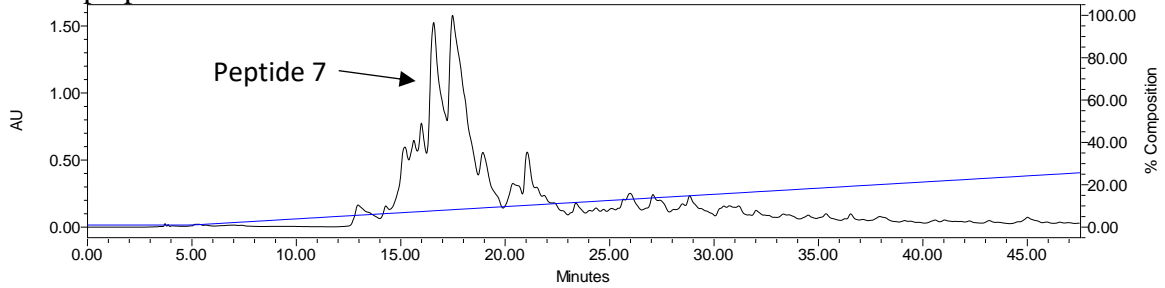
## HRMS: Calculated $[M+H]^+$ 788.3872, observed 788.3868



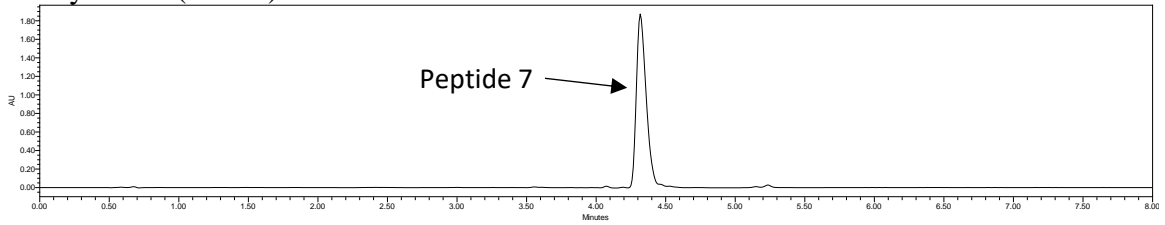
# Peptide 7



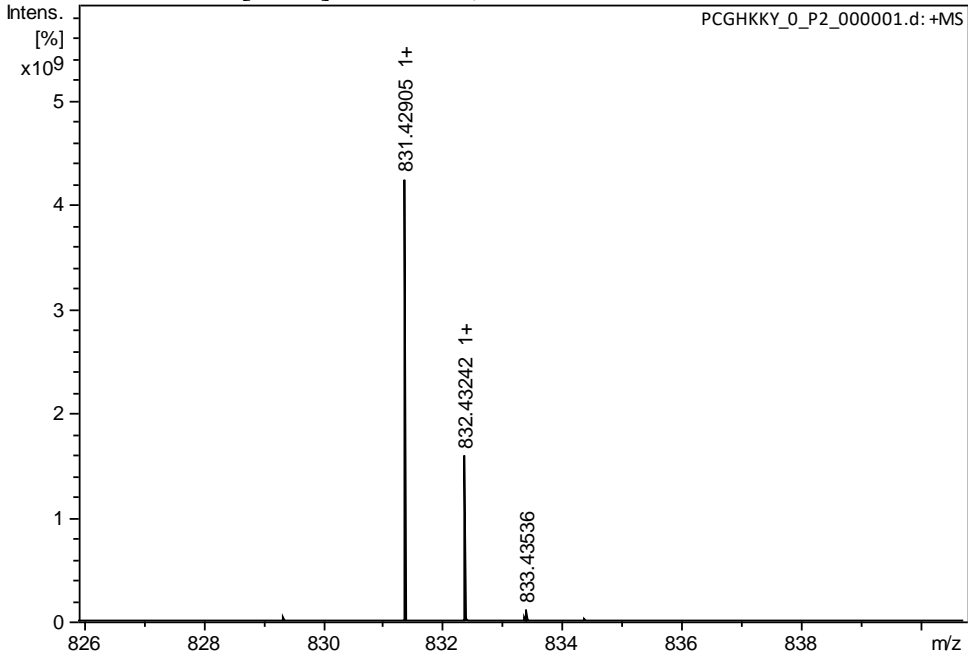
## Semi-preparative HPLC



## Purity check (UPLC): 98%

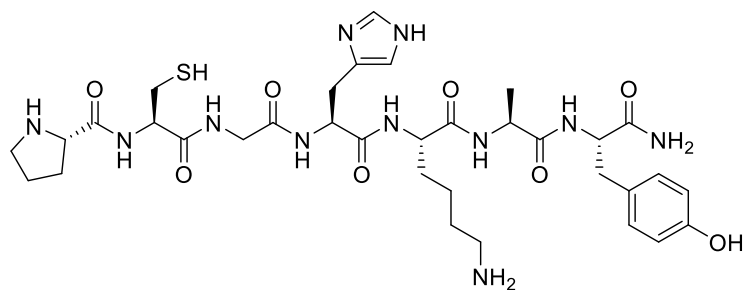


## HRMS: Calculated $[M+H]^+$ 831.4294, observed 831.4291

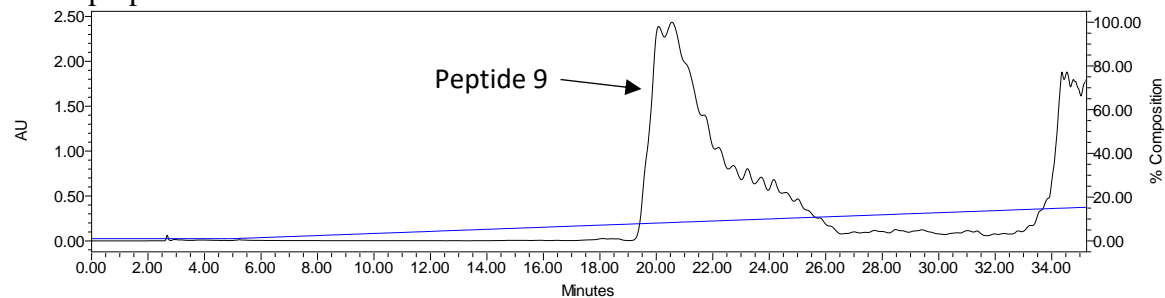




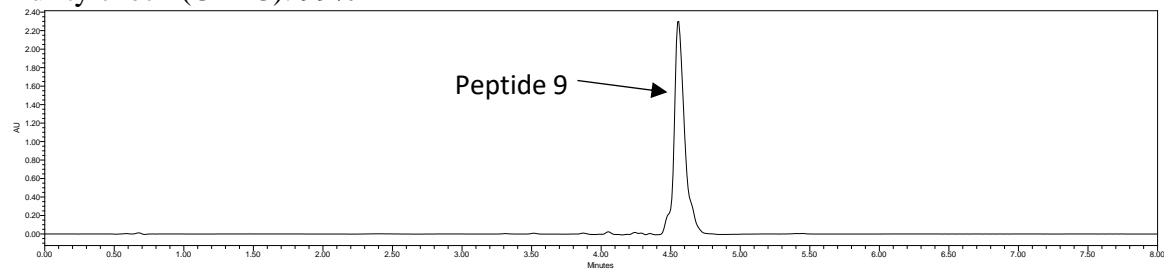
## Peptide 9



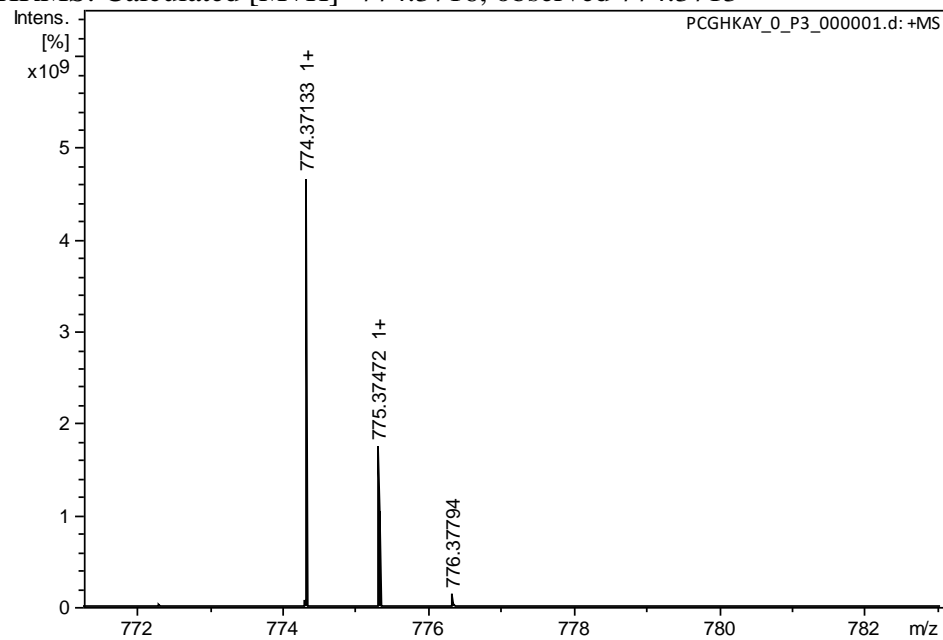
### Semi-preparative HPLC



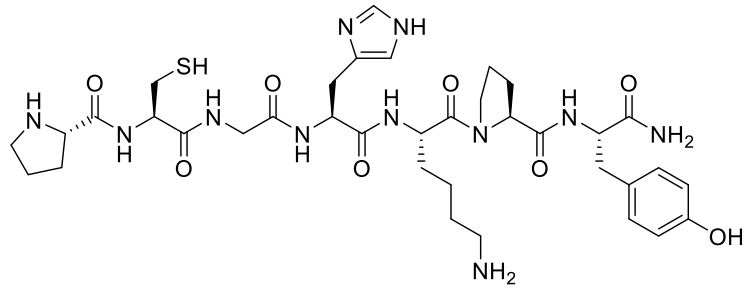
### Purity check (UPLC): 99%



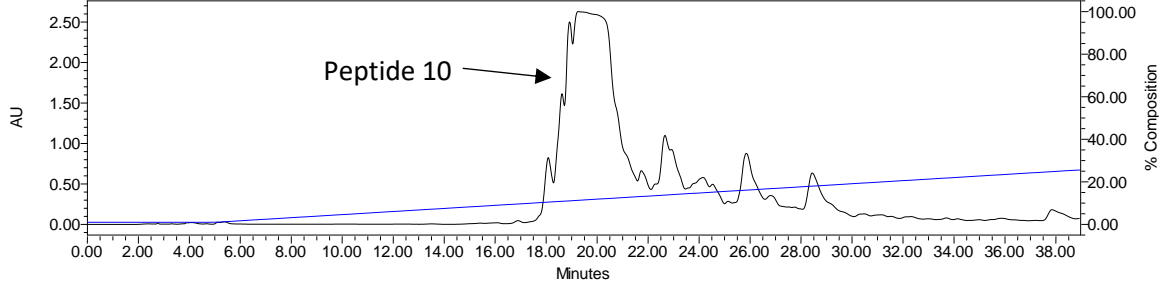
### HRMS: Calculated $[M+H]^+$ 774.3716, observed 774.3713



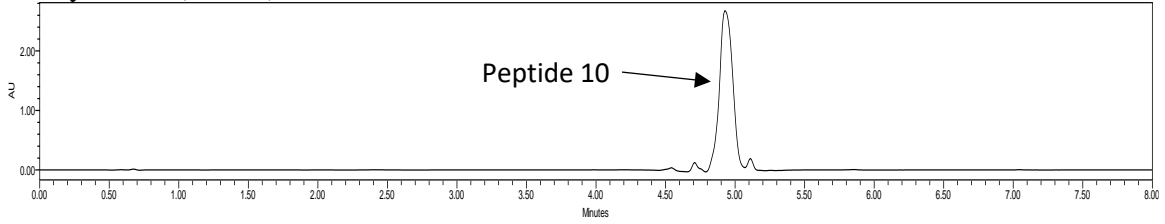
# Peptide 10



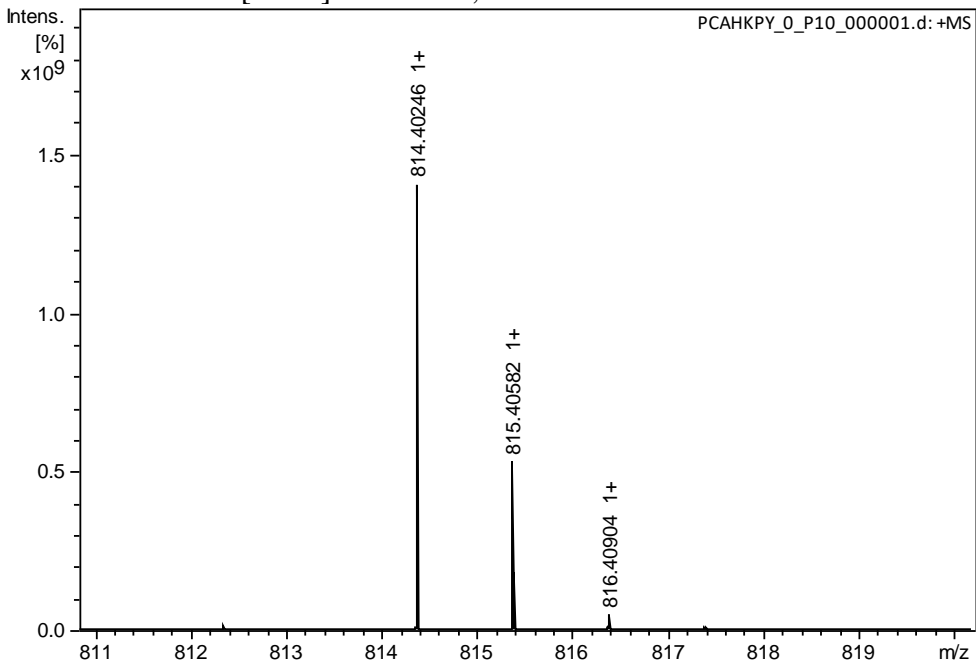
## Semi-preparative HPLC



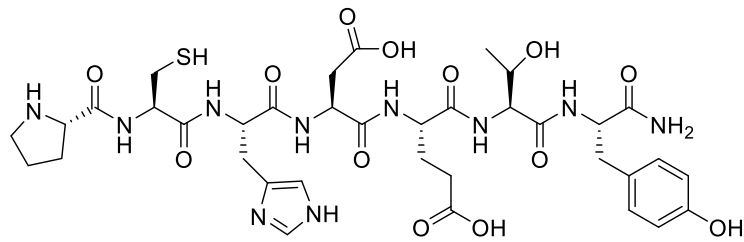
## Purity check (UPLC): 96%



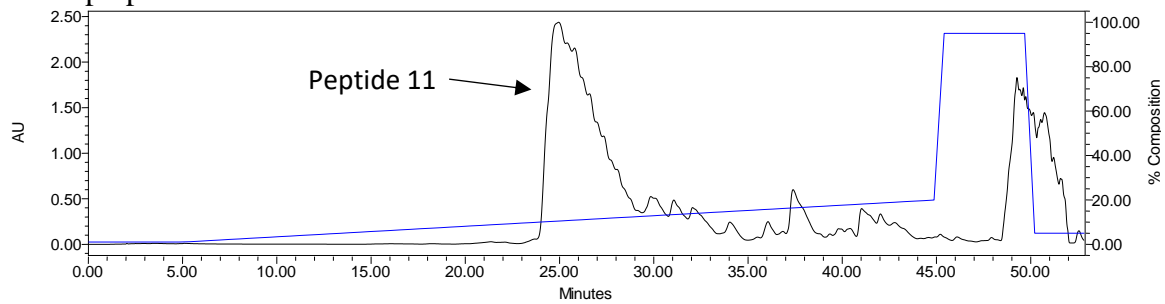
## HRMS: Calculated $[M+H]^+$ 814.4029, observed 814.4025



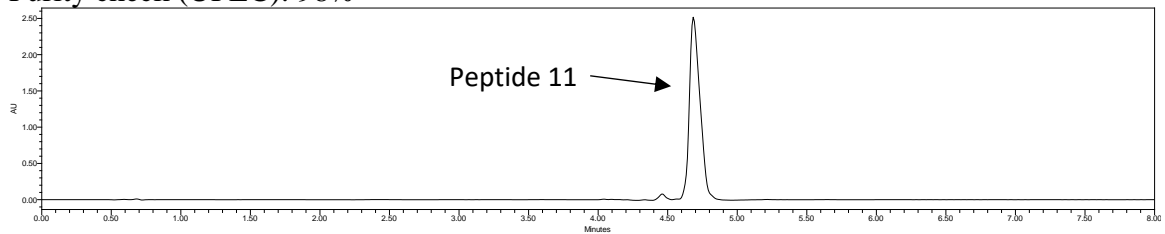
## Peptide 11



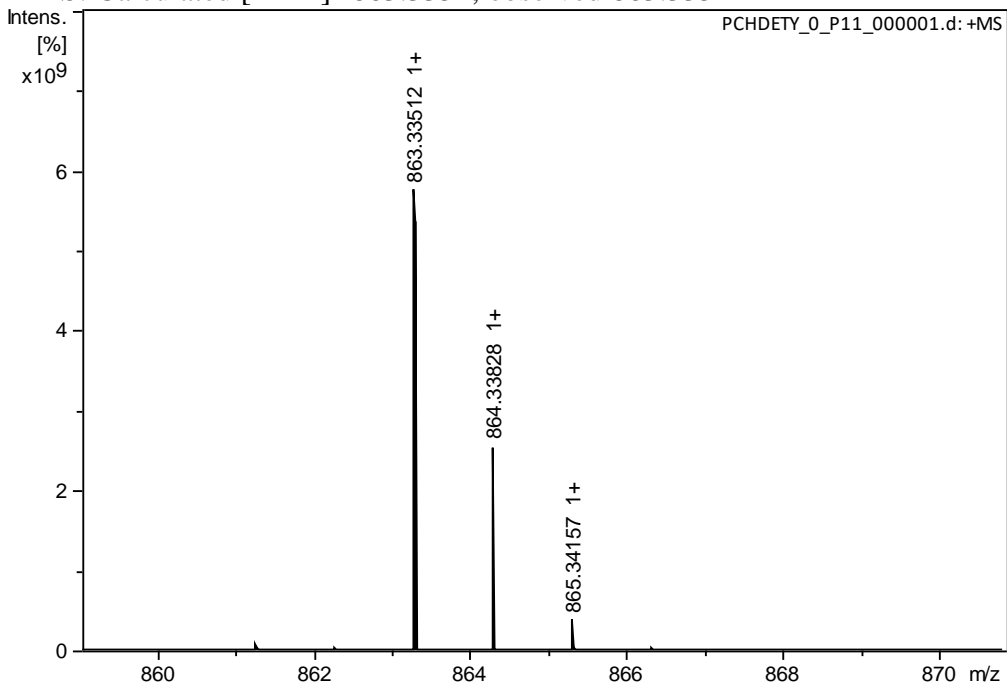
### Semi-preparative HPLC



### Purity check (UPLC): 98%



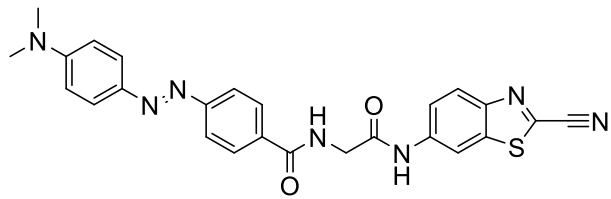
### HRMS: Calculated $[M+H]^+$ 863.3352, observed 863.3351



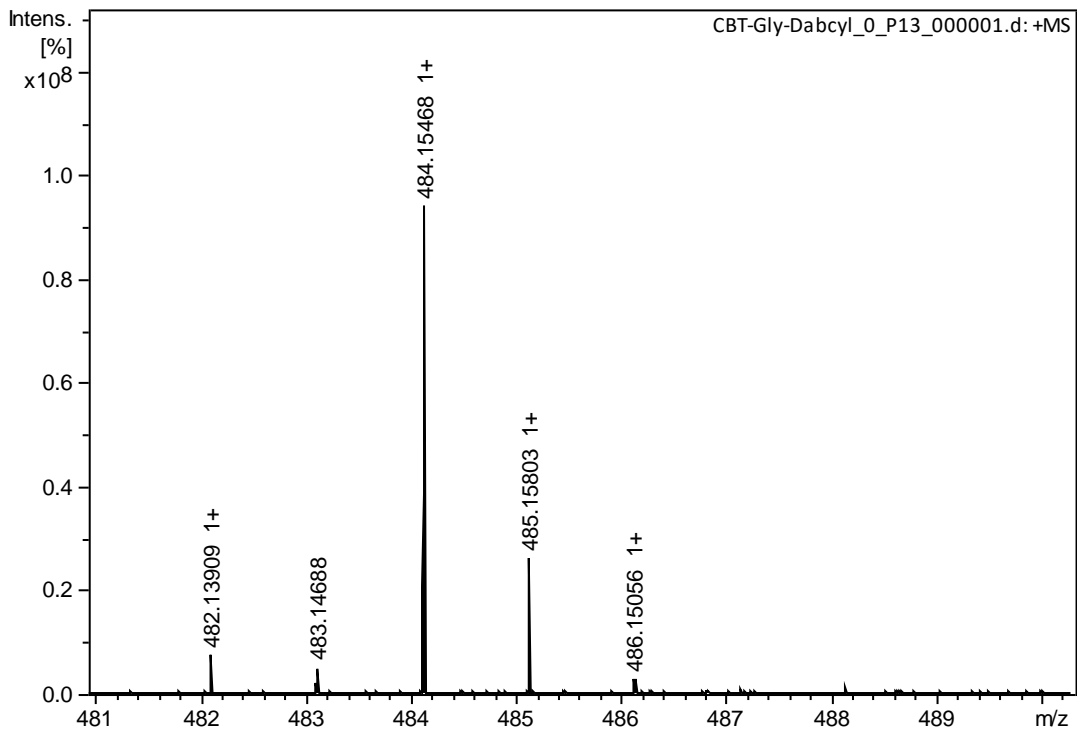




## Dabcyl-CBT

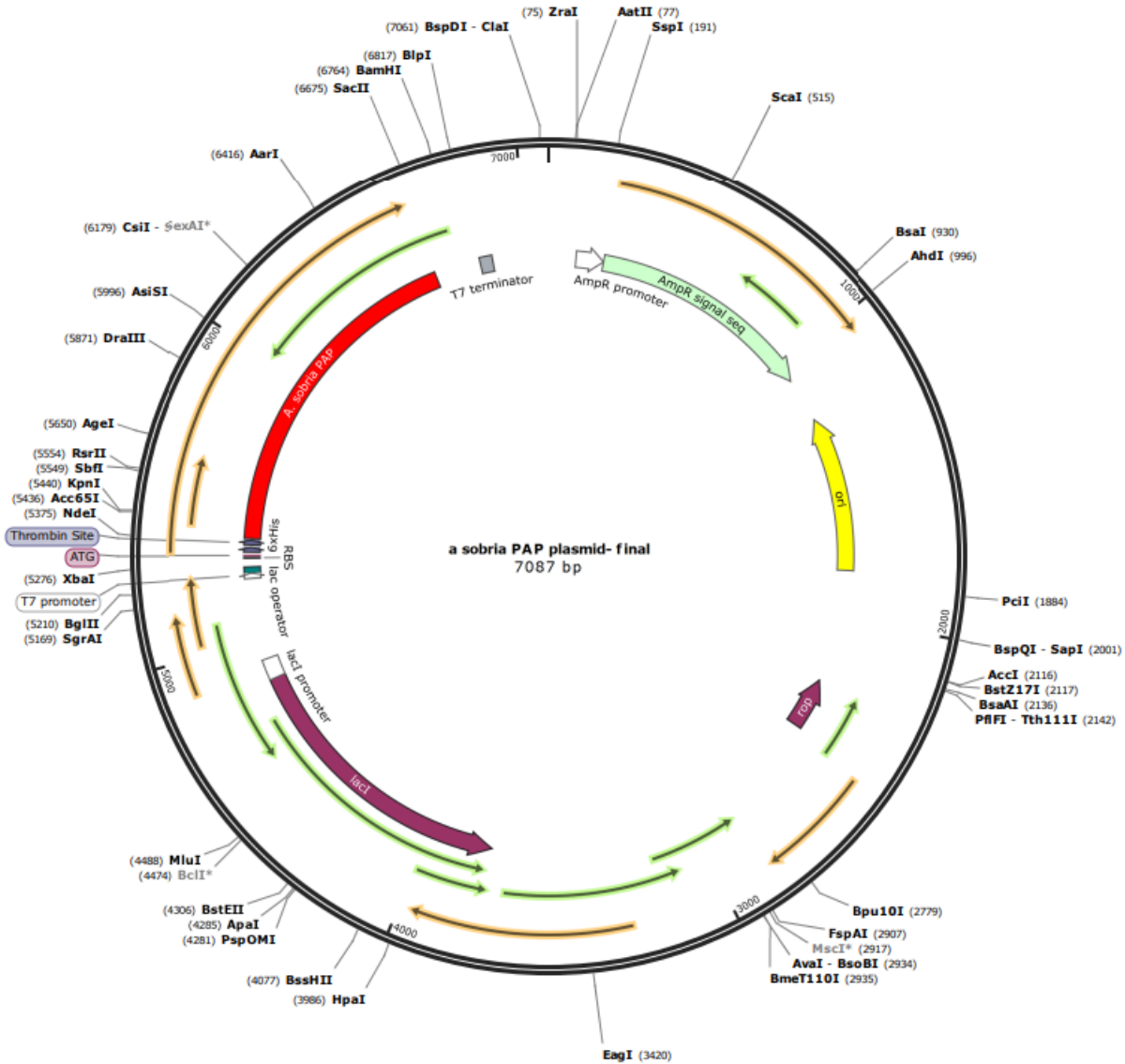


HRMS: Calculated  $[M+H]^+$  484.1550, observed 484.1547



**Figure S3.** Plasmids used for bacterial expression of ProAP and RBDV in this work. *(a)* Restriction map of plasmid pET-15b(+)-His<sub>6</sub>-ProAP. The plasmids for RBDV expression have a similar restriction map. *(b)* Amino acid sequences for *A. sobria* ProAP and RBDV proteins.

*a*



***b***

*A. sobria* ProAP:

MGSSHHHHHSSGLVPRGSHMSSPLHYVLDGIHCEPHFFTVPDLHQQPDEETITLFGRT  
LCRKDRLDDELPWLLYLQGGPGFGAPRPSANGGWIKRALQEFRVLLLDQRGTGHSTPIH  
AELLAHLNPRQQADYLSHFRAADSIVRDAELIREQLSPDHPWSLLGQSFGGFCSLTYLSLF  
PDSLHEVYLTGGVAPIGRSADEVYRATYQRVADKNRAFFARFPHAQAIANRLATHLQR  
HDVRLPNGQRLTVEQLQQQGLDLGASGAFEELYLLEDAFIGEKLNPAFLYQVQAMQP  
FNTNPVFAILHELIYCEGAASHWAAERVRGEFPALAWAQGKDFAFMGEMIFPWFQF  
RELIPLKEAAHLLAEKADWGPLYDPVQLARNKVPVACAVYAEDMYVEFDYSRETLKGL  
SNSRAWITNEYEHNGLRVDGEQILDRLIRLNRDC\*

MPC-RBDV:

MPCKTSNTIRVLLPNQEWTVVKVRNGMSLHDSLMKALKRHGLQPESAVFRLLEHKG  
KKARLDWNTDAASLIGEELQVDFLDHVPLTTHNFARKTFLKLGHRDGSSGSSLEHHHH  
HH\*

MPCGHKP-RBDV:

MPCGHKPKTSNTIRVLLPNQEWTVVKVRNGMSLHDSLMKALKRHGLQPESAVFRLLEH  
EHKGGKKARLDWNTDAASLIGEELQVDFLDHVPLTTHNFARKTFLKLGHRDGSSGSSLE  
HHHHHH\*

MPCGHKP-(GSS)<sub>2</sub>-RBDV:

MPCGHKPGSSGSSKTSNTIRVLLPNQEWTVVKVRNGMSLHDSLMKALKRHGLQPESAV  
VFRLLEHKGKKARLDWNTDAASLIGEELQVDFLDHVPLTTHNFARKTFLKLGHRDGS  
SGSSLEHHHHHH\*