**Supporting Information for:**

*In Vitro* Selection of Highly Modified Cyclic Peptides That Act as Tight Binding Inhibitors

Yollete V. Guillen Schlippe, Matthew C.T. Hartman[†], Kristopher Josephson[††] and Jack W. Szostak*

## Materials and Methods

**Amino Acids.** DL-7-azatryptophan ($W_a$), DL-hydroxy-norvaline ($T_a$), L-canavanine ($R_a$), quisqualic acid ($E_a$), *p*-bromo-DL-phenylalanine ($F_a$) were purchased from Sigma, natural amino acids, 3-fluoro-L-tyrosine ($Y_a$), 1-amino-1-cyclopentane carboxylic acid ($V_a$) were purchased from Fluka, and trans-4,5-dehydro-DL-lysine ($K_a$), L-threo-β-hydroxy aspartic acid ($D_a$), L-thiazolidine-4-carboxylic acid ($P_a$), and β-t-butyl-DL-alanine ($L_a$) were purchased from Bachem, and Tocris, Acros, and ChemImpex respectively. DL-2-amino-hex-5-ynoic acid ($M_a$) used in translation was kindly provided by J. Link and D. Tirrell (Caltech), and T. Kawate. All amino acids were dissolved in water and the pH was adjusted to 7.0-7.5 with KOH and filtered. Isotopically labeled $^{35}$S-Cys was from Perkin-Elmer.

**Translation Factors, Enzymes, and Ribosomes.** Hexa-histidine tagged IF1, IF2, IF3, EF-Tu, EF-G, EF-Ts, RF1, RF3, RRF, MTF, MetRS, GluRS, PheRS, AspRS, SerRS, and ThrRS, ArgRS, GlnRS, IleRS, LeuRS, TrpRS, AsnRS, HisRS, TyrRS, ValRS, ProRS, AlaRS, CysRS, TrpRS, GlyRS, and PheRS Ala294Gly mutant were expressed and purified as previously described.[1,2] Enzyme concentrations were determined from the UV absorbance at 280 nm and extinction coefficients calculated from amino acid compositions.

MetRS required storage in buffer containing 50% glycerol at -20 C. Under these conditions MetRS was active for ~6 months.

**Selection Reagents.** Biotinylated thrombin was purchased from Novagen. SuperScript III Reverse Transcriptase and RNaseOUT were purchased from Invitrogen. TURBO DNase was purchased from Ambion. Taq DNA Polymesase was purchased from Roche. Oligonucleotides were purchased from the Keck facility at Yale. α,α'-dibrom-m-xylol (dibromoxylene) was purchased from Fluka.

**Components for mRNA-display.** Unless otherwise specified, primers were ordered from our in house DNA core facility. The DNA library was ordered as the bottom strand from Keck (Yale): 5'-CTAGCTACCTATAGCCGGTGGTGATGGTGATGGTGGCCTAAGCTACCGGAGCCGCAWNNWNNWNNWNNWNNWNNWNNWNNWNNWNNGCACATTTAGCTGTCCTCCTTACTAAAGTTAACCCTATAGTGAGTCGTATTA-3' and gel purified. RNA was produced from the DNA library with T7 RNA polymerase and the primer containing the T7 transcription start site (5'-TAATACGACTCACTATAGGGTTAACTTTAG-3'). After purification by denaturing PAGE the RNA was photo-crosslinked with the XL-PSO oligonucleotide (5'-X(uagccggug) AAA AAA AAA AAA AAA ZZ ACC P-3'; X=psoralen C6, lower case=2'OMe, Z=spacer 9, P=puromycin, stretch of A's and ACC is DNA, from Keck) with ~50 % efficiency and was then ethanol precipitated [3]. The sequence of the RT-primer was 5'-TTTTTTTTTTTTTTTGTGATGGTGATGGTGGCCTAAGC-3'. PCR

primers used during selection were CX10 FWD (5'-TAATACGACTCACTATAGGGTTAACTTTAGTAAGGAGG-3') and CX10 REV (5'-CTAGCTACCTATAGCCGGTGGTGATG-3').

**PURE Translation System.** The PURE translation system was used as previously described with slight modification.[1] Freshly prepared standard polymix buffer was supplemented with 2 mM ATP, 2 mM GTP, 10 mM creatine phosphate, and 30 μM 10-formyl-5,6,7,8-tetrahydrofolic acid. Translation reactions (50 μL) contained final concentrations of 0.2 μM MTF, 1.0 μM IF1, 0.3 μM IF2, 0.7 μM IF3, 3.2 μM EF-Tu, 0.6 μM EF-Ts, 0.5 μM EF-G, 0.3 μM RF1, 0.4 μM RF3, 0.1 μM RRF, 0.5 μM ribosomes, and 0.05 A260 units/L total tRNA. In addition, the reactions contained the amino acids (as specified) and AARSs (0.1 μM MetRS, 0.3 μM LeuRS, 0.6 μM GluRS, 0.2 μM ProRS, 1.0 μM GlnRS, 1.0 μM HisRS, 0.3 μM PheRS, 1.5 μM TrpRS, 0.2 μM SerRS, 0.2 μM IleRS, 0.4 μM ThrRS, 0.6 μM AsnRS, 0.6 μM AspRS, 0.5 μM TyrRS, 0.5 μM LysRS, 0.4 μM ArgRS) necessary to translate each peptide. Reactions were assembled on ice and started by addition of mRNA to 1.0 μM followed by incubation for 1 h at 37 C. Peptides were isotopically labeled by including $^{35}$S-Cys.

**Peptide Translation Assay.** Translated peptides were purified from 50 μL reactions using Ni-NTA agarose (Qiagen). Following two buffer washes (50 mM Tris-HCl, 300 mM NaCl, pH 8.0), the peptides were eluted with 0.1% TFA, and the peptide yield was determined by liquid scintillation counting and the specific activity of $^{35}$S-Cys. For mass analysis, peptides

were concentrated and desalted by reverse phase micro-chromatography using C18 zip tips (Millipore) and eluted with 50% acetonitrile ($CH_3CN$), 0.1% trifluoroacetic acid (TFA) saturated with the matrix α-cyano-4-hydroxycinnamic acid. Mass measurements were made using an Applied Biosystems Voyager MALDI-TOF with delayed extraction operated in the positive mode and calibrated with the insulin B chain (3494.65 Da) standard (Sigma) unless otherwise noted.

**Ribosomal synthesis, cyclization and purification of selection winners for $K_d$ determination and LC-MS analysis.** Translation reactions were set up as before, with a few modifications. Reactions were 250 μL and a minimal mix of the synthetases and amino acids needed for translation was used. For example to translate peptide U1 the minimal mix contained ($M_a$/MRS, $L_a$/LRS, $Y_a$/YRS, $T_a$/TRS, $K_a$/KRS, Cys/CRS, Gln/QRS, Asn/NRS, Ser/SRS, Gly/GRS, His/HRS). Amino acid concentrations were slightly adjusted to improve yield and fidelity (0.08 mM $M_a$, 100 μM Gln, 100 μM Cys). For radiolabeled peptides, amino acid concentrations were slightly further adjusted to improve fidelity and to increase the specific radioactivity of the peptides (10 μM Gln, 10 μM Cys). Peptides were cyclized on the Ni-NTA resin in cyclization buffer (20 mM Tris pH 8.0, 100 mM NaCl, 0.2 mM TCEP, 5 mM dibromoxylene, 50% $CH_3CN$) for 30 min at 25°C and eluted with a 1:1 mixture of 0.1% TFA and $CH_3CN$ in four 250 μL fractions. Fractions with the highest amount of peptide were combined and spin-filtered through a YM-10 spin-filter to remove larger His tagged components of the pure system and concentrated to remove the $CH_3CN$

and TFA. Peptides were analyzed by MALDI-TOF and spectra calibrated with the P14R (1532.8582 Da) and insulin B chain (3493.6513 Da) standards (Sigma) internally.

**Ribosomal synthesis, iodoacetamide modification and purification of selection winners with His tag deletions for LC-MS/MS analysis.** Translations were set up as before, with a few modifications. Reactions were 250 μL and the same minimal mix of the synthetases and amino acids for full-length peptides was used. Amino acid concentrations were slightly adjusted to improve yield and fidelity further (0.08 mM $M_a$, 200 μM Gln, 400 μM Cys). Translations were stopped by adjusting MgOAc to10 mM and NaCl to 0.4 M. The solution was acidified with 2 volumes 0.2 % TFA, filtered through a 0.22 μm Durapore PVDF centrifugal filter (Millipore) and desalted on a C18 ZipTip (Millipore) and eluted with 70 % $CH_3CN$, 0.1 % FA. Solvent was evaporated using a speedvac and peptides were resuspended in 10 mM DTT, 100 mM Tris pH 8.0, 30% $CH_3CN$ and incubated for 1 hr. Cys residues were modified by the addition of iodoacetamide (25 mM final concentration) in the dark for 30 min. Samples were desalted on a G-10 column (GE Healthcare) equilibrated with 50 mM ammonium bicarbonate in 30 % $CH_3CN$. Buffer was removed using the SPD 1010 SpeedVac System (Thermo Savant) and dissolved in 30 μL 25 % $CH_3CN$, 0.1 % FA.

**$K_d$ determination of translated peptides.** Peptides (0.2 nM, determined by scintillation counting, unless otherwise indicated) were incubated with different concentrations of human thrombin from plasma (CalBiochem) in selection buffer (50 mM Tris, pH 7.8, 100

mM NaCl, 4 mM $MgCl_2$, 0.25%Triton X-100, 200 μL) for 1 hr at 25°C. The samples were spun in a YM-30 filter at 13000 rpm for 15 sec to wet the filter, then transferred to a fresh collection tube. Bound peptides were separated from free peptides by spinning samples through the filter at 13000 rpm for 2 more min. After this spin ~half of the volume had been filtered into the bottom chamber and the radioactivity in each chamber was determined by removing 75 μL from each to give bound and free peptide or total peptide ("top") and free peptide ("bottom") concentrations. The fractional saturation at different protein concentrations was calculated from the formula $f_a$ = [bound peptide]/[total peptide], where [bound peptide] is the concentration of peptide bound to thrombin and [total peptide] is the total peptide concentration, by substituting top cpm minus bottom cpm for [bound peptide] and top cpm for [total peptide]. These experiments were done at least in duplicate.

**Ribosomal synthesis, cyclization and purification of selection winners for Inhibition Assays.** Translation reactions were set up as before, with a few modifications. Reactions were 200 μL and a minimal mix of the synthetases and amino acids needed for translation was used. All natural amino acid concentrations were 300 μM each and unnatural amino acid concentrations are as described above, with the exception of Gln (0.2 mM) and $M_a$ (100 μM). Peptides were cyclized on the Ni-NTA resin in cyclization buffer (20 mM Tris pH 8.0, 100 mM NaCl, 0.2 mM TCEP, 5 mM dibromoxylene, 50% $CH_3CN$) for 30 min at 25°C, washed 3 times with 10 column volumes of wash buffer 1 (20 mM Tris pH 8.0, 5 mM BME) followed by 3 10 column washes of 50 mM Ammonium bicarbonate and eluted with a 1:1 mixture of 0.2% TFA and $CH_3CN$ in four 200 μL fractions. Fractions with the

highest amount of peptide were combined $CH_3CN$ and TFA. Peptides were analyzed by MALDI-TOF and spectra calibrated with the P14R (1532.8582 Da) and insulin B chain (3493.6513 Da) standards (Sigma) internally. Buffer was removed using the SPD 1010 SpeedVac System (Thermo Savant) and dissolved in assay buffer 35-150 μL (AnaSpec). The concentration of peptide in solution was determined by scintillation counting and ranged between 0.5-0.7 μM. eptide solutions were serial diluted into assay buffer. Assays were preformed according to the established procedures. Briefly, varying concentrations of peptides were preincubated with thrombin 10.2 nM final concentration) in assay buffer for 1 hr at 25ºC. Reaction were initiated with the addition of 10 μM 5-FAM/QXL FRET substrate and the increase in fluorescence (Ex/Em = 490 nm/520 nm) monitored for 1 hr on a SpectraMax Gemini EM multiwall plate reader at 25ºC. Initial rates were determined at various peptide inhibitor concentrations. The *in vitro* selected peptides had no effect on the quenched fluorescence of the thrombin substrate at the highest concentration tested. The fractional activity was plotted against peptide inhibitor concentrations and fitted to the Morrison tight binding equation to determine the apparent inhibition constants $K_i^{app}$.

## Supplementary Figures and Tables:

Supplementary Figures:

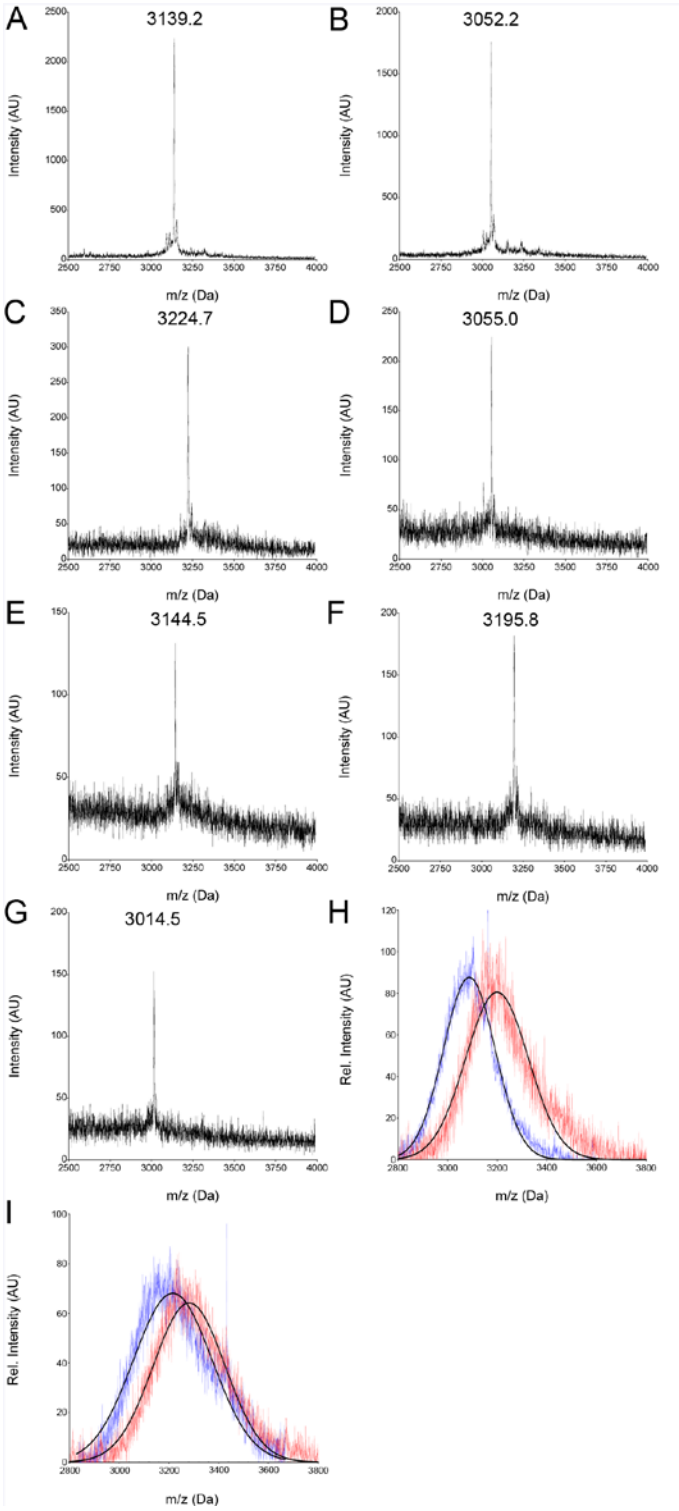


Figure S1. Cyclization of individual library members and the total library mixture. MALDI-TOF spectra of individual library members cyclized with dibromoxylene. Sequences are shown in Table S2. A. 3A, obs. mass 3139.2 (calc.: 3141.5). B. 4A, obs. mass 3052.2 (calc.: 3054.4). C. 6A, obs. mass of 3224.7 (calc.: 3226.7). D. 8A, obs. mass of 3055.0 (calc.: 3056.5). E. 9A, obs. mass of 3144.5 (calc.: 3146.5). F. 12A, obs. mass of 3195.8 (calc.: 3198.6). G. 14A, obs.  mass of 3014.5 (calc.: 3017.3). MALDI-TOF spectra of the linear peptide library (blue) and the dibromoxylene cyclized peptide library (red). Black lines represent fits of the data to Gaussian distribution curves. H. peptide library translated with natural amino acids I. peptide library translated with unnatural amino acids.

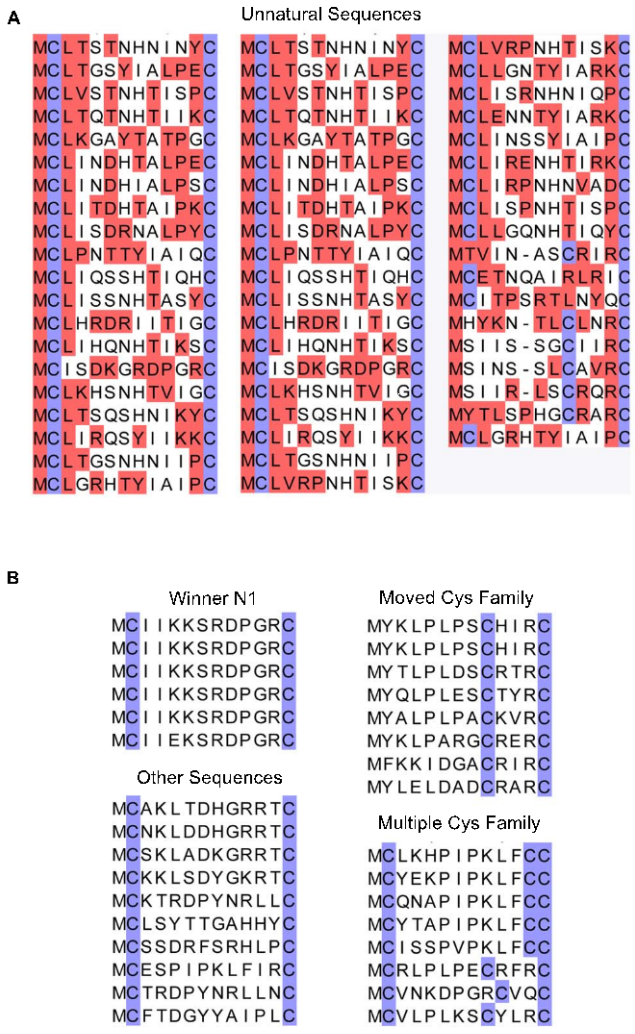



Figure S2. Sequences after round 7. A. Unnatural peptides. Unnatural amino acids are highlighted in red and Cys used for cyclization in blue. For simplicity, only the random region is shown. B. Natural peptides (only random region is shown).

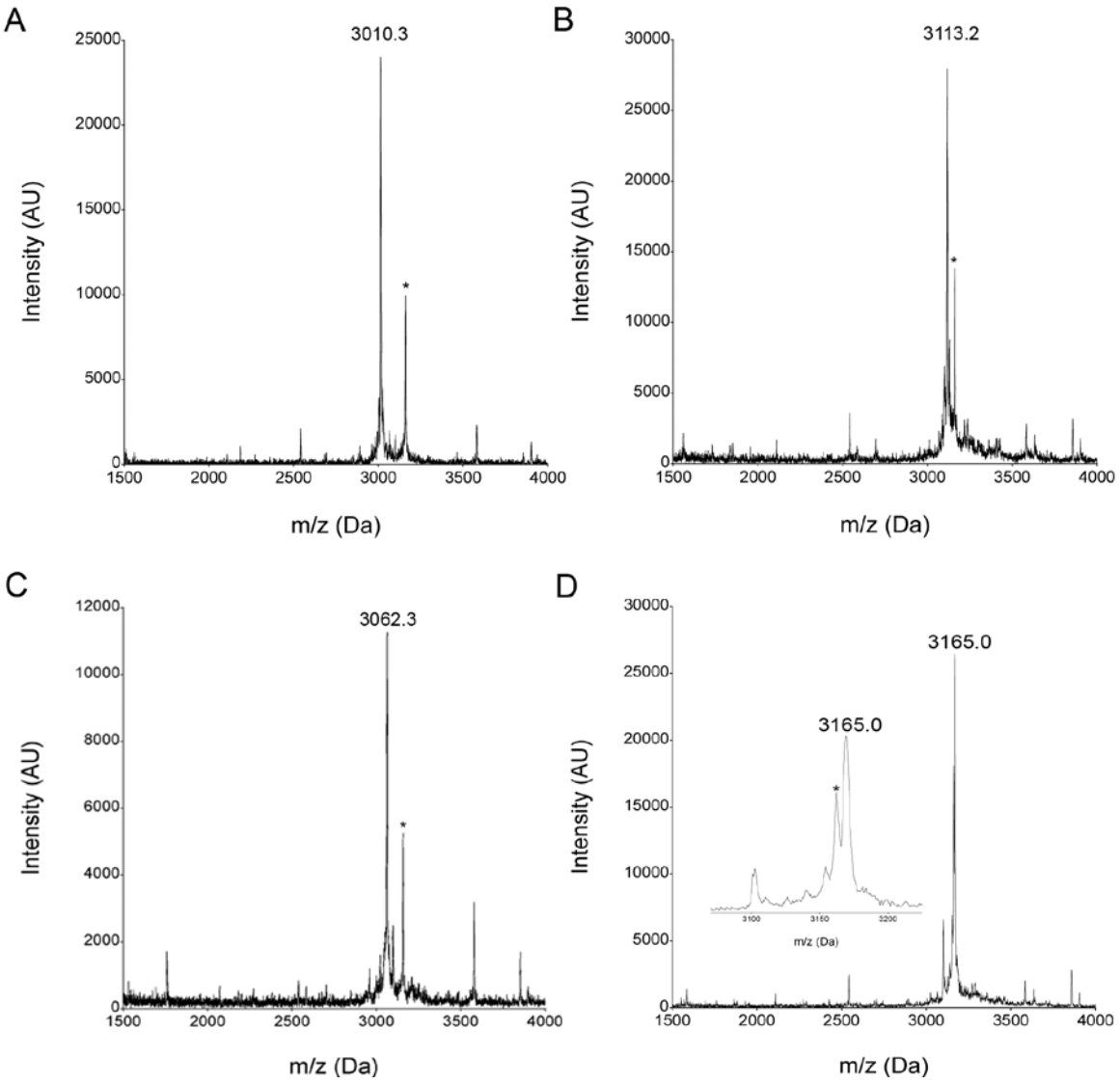



Figure S3. MALDI-TOF spectra of unnatural peptide U1 ($M_aCL_aQNSY_aIAT_aK_aGCGSGSL_aGHHHHHHR_aL_a$, where unnatural amino acids are labelled with the subscript a). A. U1 translated as the natural linear peptide, obs. mass 3010.3 (calc.: 3011.4). B. U1 translated as the natural cyclic peptide, obs. mass of 3113.2 (calc.: 3113.4). C. Unnatural linear peptide U1, obs. mass of 3062.3 (calc. 3063.6). d. Unnatural cyclic peptide U1, obs. mass of 3165.0 (calc. 3165.6). The asterisk (obs. 3158.7)

denotes an impurity from the PURE system which co-purifies with the translated peptides and is present in the negative control reactions containing no mRNA.

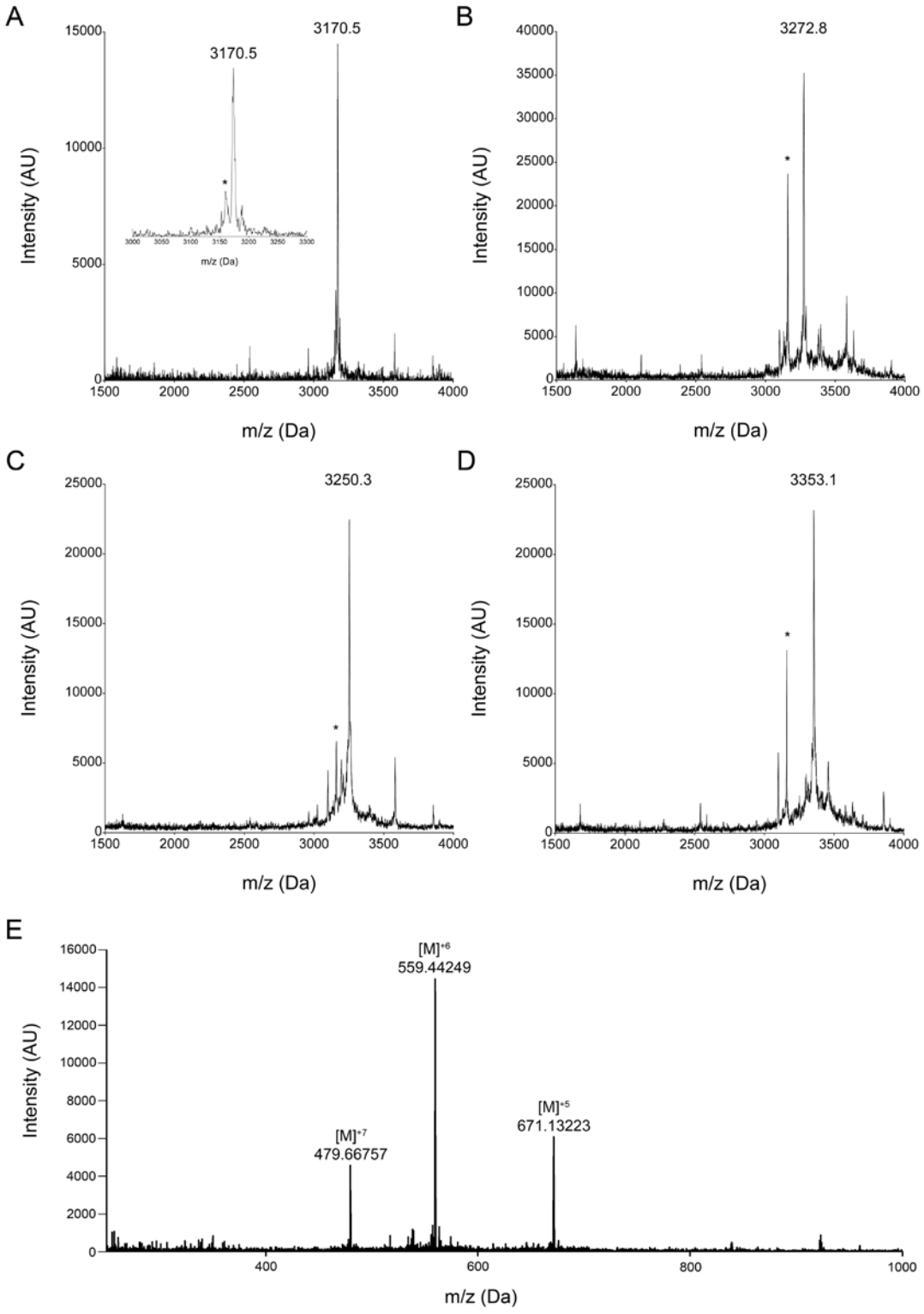

A
3170.5
3170.5
Intensity (AU)
m/z (Da)
B
3272.8
Intensity (AU)
m/z (Da)
C
3250.3
Intensity (AU)
m/z (Da)
D
3353.1
Intensity (AU)
m/z (Da)
E
[M]+6
559.44249
[M]+5
671.13223
[M]+7
479.66757
Intensity (AU)
m/z (Da)

Figure S4. MALDI-TOF spectra of *in vitro* translated unnatural peptide U2 ($M_aCE_aT_aNL_aAIK_aQR_aICGSGR_aL_aGHHHHHHR_aL_a$, where unnatural amino acids are labelled with the subscript a). A. U2 translated as the natural linear peptide, obs. mass 3170.5 (calc.: 3171.7). B. U2 translated as the natural cyclic peptide, obs. mass 3273.7 (calc.: 3273.7). C. U2 translated as the unnatural linear peptide, obs. mass 3250.3 (calc.: 3252.9). D. U2 translated as the unnatural cyclic peptide, obs. mass 3353.1 (calc.: 3353.9). The asterisk (obs. 3158.7) denotes the same impurity noted in Fig S4. E. LC-QTOF-MS confirms formation of full-length cyclic unnatural peptide U2 (calc. $[M]^{+5}$ m/z = 671.12604 (9.2 p.p.m.), $[M]^{+6}$ m/z = 559.43968 (5.0 p.p.m.), $[M]^{+7}$ m/z = 479.66370 (8.1 p.p.m.).

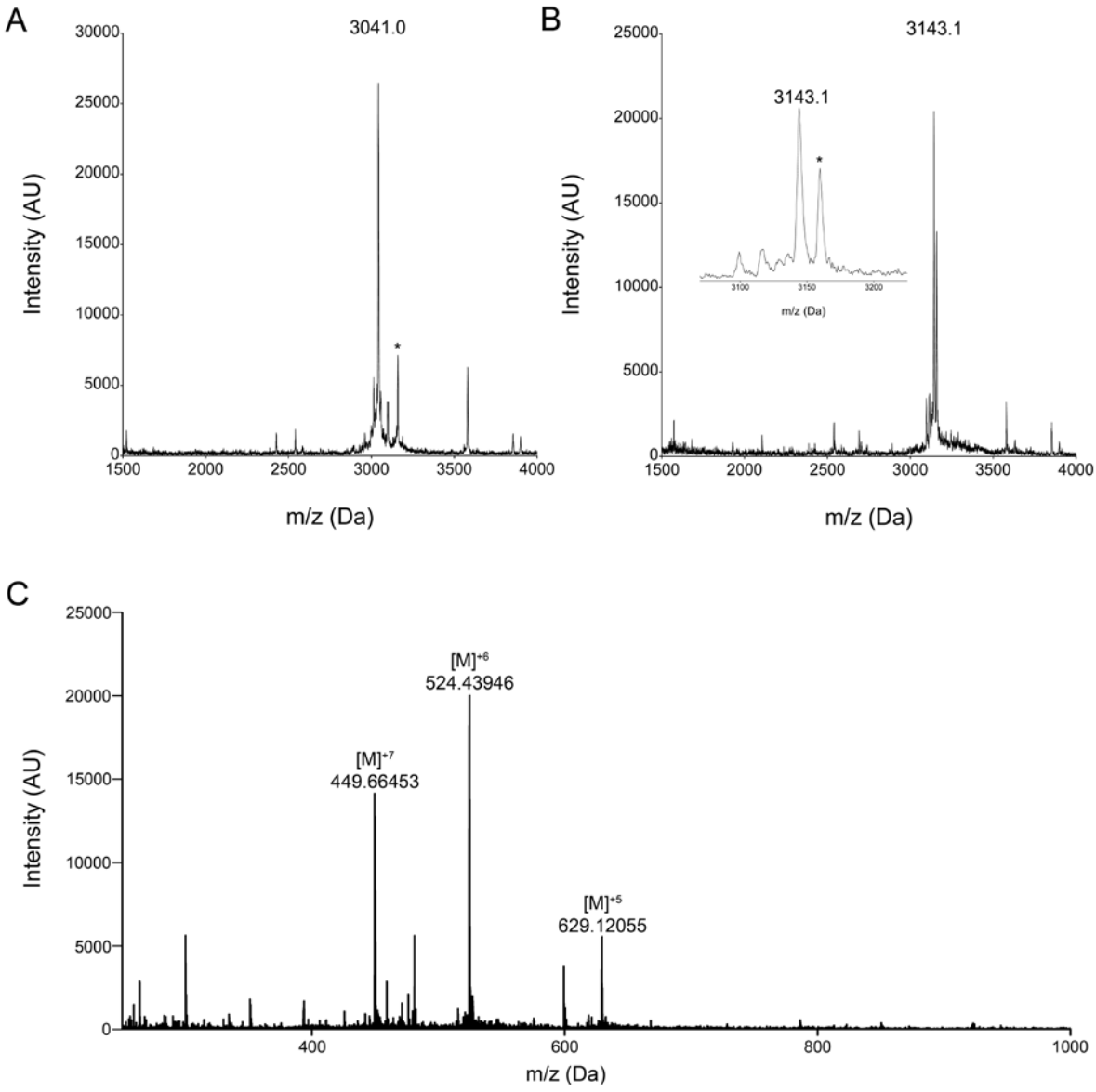



Figure S5. MALDI-TOF spectra of natural peptide N1 (MCIIKKSRDPGRCVGSLGHHHHHHRL). A. Linear peptide N1, obs. mass 3041.0 (calc.: 3041.6). B. Cyclic peptide N1, obs. mass 3143.1 (calc.: 3143.8). The asterisk (obs. 3158.7) denotes the impurity noted in Fig. S4. C. LC-QTOF-MS confirms formation of full-length cyclic natural peptide N1 (calc. $[M]^{+5}$ m/z = 629.12199 (2.3 p.p.m.), $[M]^{+6}$ m/z = 524.43630 (6.0 p.p.m.), $[M]^{+7}$ m/z = 449.66080 (8.3 p.p.m.).

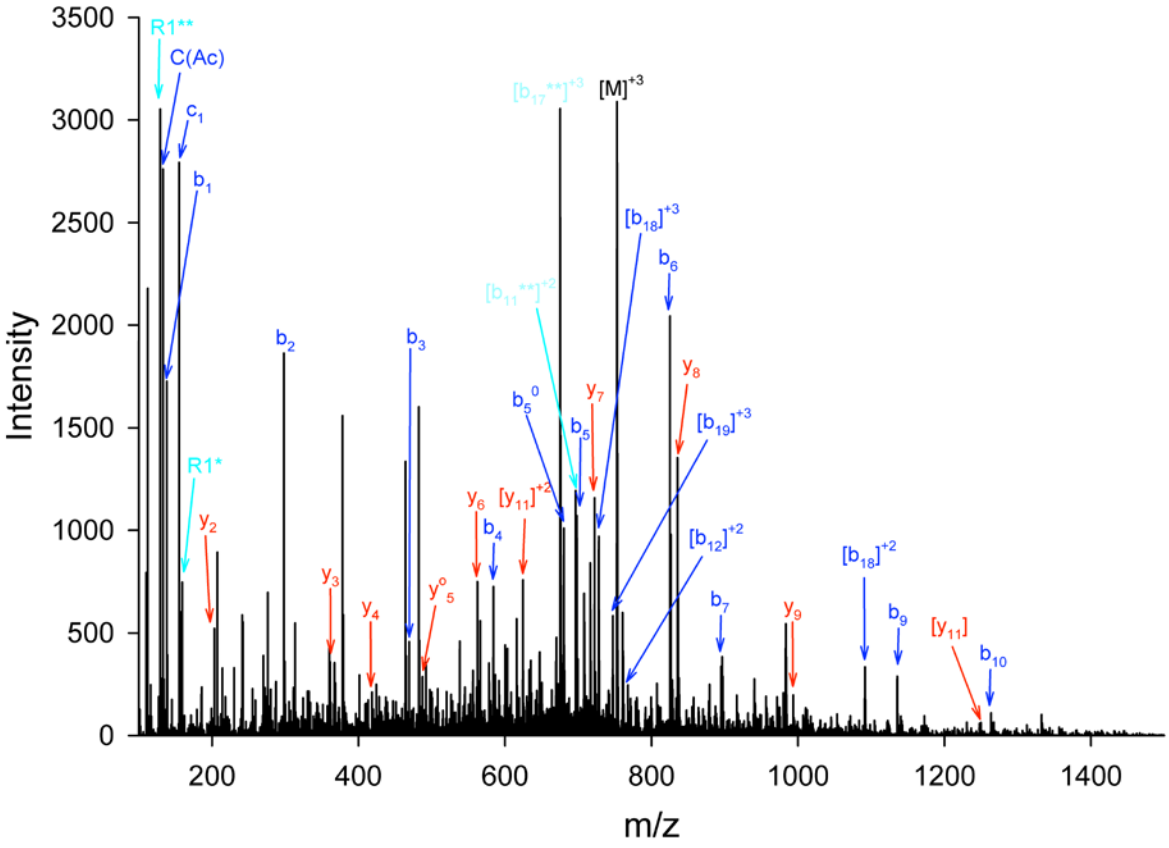



Figure S6. Verification of structure of unnatural peptide U2 by MS/MS sequencing. For simplicity we translated a peptide lacking the $His_6$-tag and modified the two cysteines with iodoacetamide to prevent disulfide formation. LC-MS/MS analysis confirms site specific incorporation of each of the unnatural amino acids into peptide U2 with the sequence $M_a\underline{C}E_aT_aNLAIK_aQR_aI\underline{C}GSGR_aL_aG$, (unnatural amino acids are labelled with the subscript a and the iodoacetamide modified Cys residues are underlined). Predicted and observed ions are summarized in Tables S6-S8.

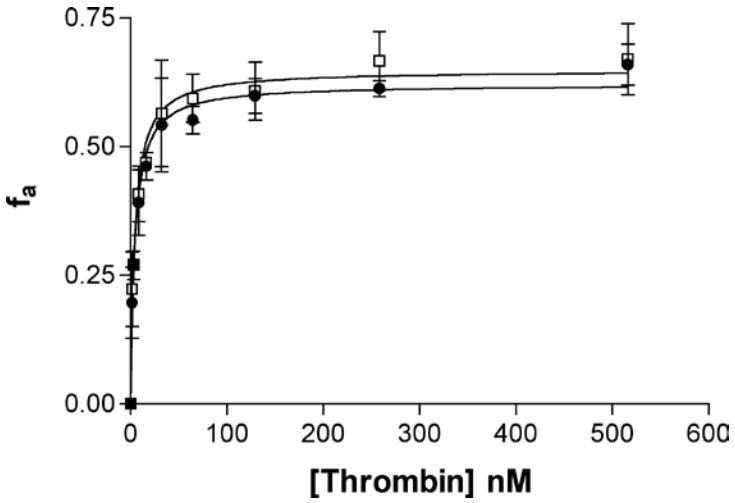



Figure S7. Affinity of translated U1 peptide for thrombin, as determined by equilibrium ultrafiltration binding assay. $^{35}$S-Cys labeled peptides were incubated with varying concentrations of thrombin and incubated for 1hr. Bound and free peptides were separated using a 30-MW cut-off spin-filter and their concentrations determined to calculate $f_a$ ($f_a$ = [bound peptide]/[total peptide], by substituting top cpm minus bottom cpm for [bound peptide] and top cpm for [peptide] as described in the method section.

The concentration of thrombin was plotted against $f_a$ and fit to a simple hyperbola to obtain the $K_d$ values. U1 unnatural cyclic with TCEP (open squares) $K_d = 4.4 \pm 0.8$ nM, U1 unnatural cyclic without TCEP (closed circles) $K_d = 4.5 \pm 1$ nM. The error is a standard error.

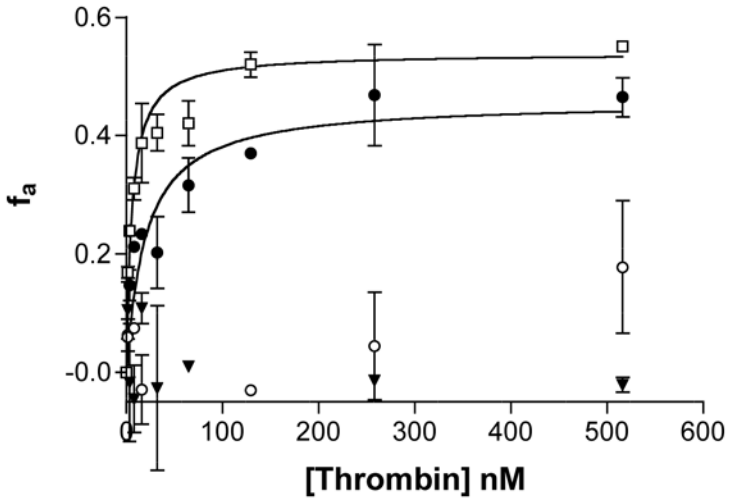



Figure S8. Affinity of translated U2 peptide for thrombin, as determined by equilibrium ultrafiltration binding assay. $K_d$ values were determined as described in Fig. S8. U2 unnatural cyclic with TCEP (open squares) $K_d = 5.7 \pm 1.3$ nM, U2 unnatural cyclic peptide, without TCEP (closed circles) $K_d = 20 \pm 7$ nM, U2 unnatural linear peptide with TCEP (open circles) $K_d > 500$ nM, U2 natural cyclic peptide without TCEP (closed triangles) $K_d > 500$ nM. The error is a standard error.

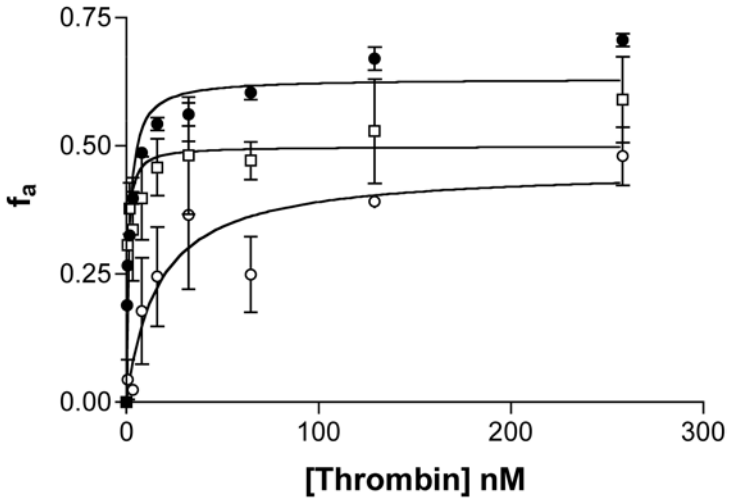


Figure S9. Affinity of translated N1 peptide for thrombin, as determined by equilibrium ultrafiltration binding assay. $K_d$ values were determined as described in Fig. S8. N1 natural cyclic peptide with TCEP (open squares) $K_d = 0.7 \pm 0.3$ nM, N1 natural cyclic peptide without TCEP (closed circles) $K_d = 1.5 \pm 0.2$ nM, N1 natural linear peptide with TCEP (open circles) $K_d = 17 \pm 8$ nM.

## Supplemental Tables:

### Table S1. Summary of unoptimized and optimized amino acid concentrations:

| Amino acid | Unoptimized concentration mM | Optimized concentration mM |
|---|---|---|
| DL-2-Amino-hex-5-ynoic acid ($M_a$) | 0.6 | 1.2 |
| L-canavanine ($R_a$) | 0.2 | 0.4 |
| 3-fluoro-L-tyrosine ($Y_a$) | 0.8 | 1.6 |
| *p*-bromo-DL-phenylalanine ($F_a$) | 0.8 | 1.6 |
| L-quisqualic acid ($E_a$) | 0.8 | 4 |
| trans-4,5-dehydro-DL-lysine ($K_a$) | 0.4 | 0.8 |
| DL-hydroxy-norvaline ($T_a$) | 0.3 | 0.6 |
| 1-amino-1-cyclopentane carboxylic acid ($V_a$) | 0.8 | 1.6 |
| L-thiazolidine-4-carboxylic acid ($P_a$) | 0.3 | 0.02 |
| DL-7-azatryptophan ($W_a$) | 0.4 | 0.8 |
| L-threo-β-hydroxy aspartic acid ($D_a$) | 0.4 | 0.8 |
| β-t-butyl-DL-alanine ($L_a$) | 0.8 | 6.6 |
| Natural amino acids G, A, H, S, Q, N, I | 0.05 | 0.05 |
| Natural amino acid C | 0.01 | 0.01 |

### Table S2 Sequences of peptides in SF2

| Peptide sequence | Name |
|---|---|
| MCVFGNRGTQPFCGSGSLGHHHHHHRL | 3A |
| MCSGTLYSLPTPCGSGSLGHHHHHHHRL | 4A |
| MCEFFDKKILAPCGSGSLGHHHHHHRL | 6A |
| MCPRTPILTSPLCSGSLGHHHHHHRL | 8A |

| | |
|---|---|
| MCFLQLVAQGDHCGSGSLGHHHHHHRL | 9A |
| MCRLQNVPRIALCGSGSLGHHHHHHRL | 12A |
| MCLRGVHSGTTACGSGSLGHHHHHHRL | 14A |

**Table S3. Observed and calculated b-ions for peptide U1 with the sequence $M_aCL_aQNSY_aIAT_aK_aGCGSGSL_aG$.**

| Sequence | # | b-ion | |
|---|---|---|---|
| | | Obs. | Calc. |
| $M_a$ | 1 | N.O. | 138.05 |
| C | 2 | 298.08 | 298.09 |
| La | 3 | 425.19 | 425.19 |
| Q | 4 | 553.24 | 553.25 |
| N | 5 | 667.28 | 667.29 |
| S | 6 | 754.31, 736.31 ($b_6^0$) | 754.32 |
| $Y_a$ | 7 | 935.37, 917.35 ($b_7^0$) | 935.37 |
| I | 8 | 1048.44 | 1048.46 |
| A | 9 | 1119.49 | 1119.50 |
| $T_a$ | 10 | 1234.57 | 1234.56 |
| $K_a$ | 11 | 1360.61 | 1360.64 |
| G | 12 | 1417.64 | 1417.66 |
| C | 13 | 1577.68 | 1577.69 |
| G | 14 | N.O. | 1634.71 |
| S | 15 | 861.39 $[b_{15}]^{+2}$, 852.41 $[b_{15}^0]^{+2}$ | 1721.75 |
| G | 16 | 1778.79, 889.87 $[b_{16}]^{+2}$, 881.37 $[b_{16}*]^{+2}$ | 1778.77 |
| S | 17 | 933.40 $[b_{17}]^{+2}$ | 1865.80 |

| | | | |
|---|---|---|---|
| $L_a$ | 18 | 996.93 $[b_{18}]^{+2}$, 987.95 $[b_{18}{}^{0}]^{+2}$ | 1992.90 |
| $b^0$ is the corresponding b ion -$H_2O$; b* is the corresponding b ion -$NH_3$;<br>N.O. not observed | | | |

**Table S4. Observed and calculated y-ions for peptide U1 with the sequence $M_aCL_aQNSY_aIAT_aK_aGCGSGSL_aG$.**

| Sequence | # | y-ion | |
|---|---|---|---|
| | | Obs. | Calc. |
| $M_a$ | 18 | N.O. | 1930.87 |
| C | 17 | N.O. | 1770.84 |
| La | 16 | 1643.76 | 1643.74 |
| Q | 15 | 1515.66 | 1515.68 |
| N | 14 | 1401.62 | 1401.64 |
| S | 13 | 1314.61 | 1314.61 |
| $Y_a$ | 12 | 1133.55 | 1133.55 |
| I | 11 | 1020.47 | 1020.47 |
| A | 10 | 949.44 | 949.43 |
| $T_a$ | 9 | 834.36 | 834.37 |
| $K_a$ | 8 | 708.29 | 708.29 |
| G | 7 | 651.27 | 651.27 |
| C | 6 | 491.24 | 491.24 |
| G | 5 | 434.21 | 434.21 |
| S | 4 | 347.19 | 347.19 |
| G | 3 | 290.16 | 290.16 |
| S | 2 | 203.14 | 203.13 |

| | | | |
|---|---|---|---|
| $L_a$ | 1 | N.A. | 76.03 |
| N.A. not applicable; N.O. not observed | | | |

**Table S5. Other observed and calculated ions for peptide U1 with the sequence $M_aCL_aQNSY_aIAT_aK_aGCGSGSL_aG$.**

| ion | mass | |
|---|---|---|
| | Obs. | Calc. |
| $[M]^{+2}$ | 1034.47 | 1034.46 |
| $[M^0]^{+2}$ | 1025.46 | 1025.46 |
| $[M]^{+2}$ is the doubly charged parent ion; $[M^0]^{+2}$ is the doubly charged parent ion –H2O | | |

**Table S6. Observed and calculated b-ions for peptide U2 with the sequence $M_aCE_aT_aNL_aAIK_aQR_aICGSGR_aL_aG$.**

| Sequence | # | b-ion | |
|---|---|---|---|
| | | Obs. | Calc. |
| $M_a$ | 1 | 138.05 | 138.05 |
| C | 2 | 298.08 | 298.09 |
| $E_a$ | 3 | 469.11 | 469.11 |
| $T_a$ | 4 | 584.17, 566.17 ($b_4^0$) | 584.18 |
| N | 5 | 698.22, 680.22 ($b_5^0$) | 698.22 |
| $L_a$ | 6 | 825.31, 807.32 ($b_6^0$) | 825.32 |
| A | 7 | 896.36 | 896.36 |
| I | 8 | N.O. | 1009.44 |
| $K_a$ | 9 | 1135.51 | 1135.52 |

| | | | |
|---|---|---|---|
| Q | 10 | 1263.58 | 1263.58 |
| $R_a$ | 11 | 696.32 $[b_{11}**]^{+2}$ | 1421.66 |
| I | 12 | 767.87 $[b_{12}]^{+2}$ | 1534.74 |
| C | 13 | 847.90 $[b_{12}]^{+2}$ | 1694.78 |
| G | 14 | N.O. | 1751.80 |
| S | 15 | N.O. | 1838.83 |
| G | 16 | 948.43 $[b_{16}]^{+2}$ | 1895.85 |
| $R_a$ | 17 | 685.33 $[b_{17}]^{+3}$, 675.31 $[b_{17}**]^{+3}$ | 2053.93 |
| $L_a$ | 18 | 1091.00 $[b_{18}]^{+2}$, 727.68 $[b_{18}]^{+3}$ | 2181.03 |
| G | 19 | 746.69 $[b_{19}]^{+3}$ | 2238.05 |
| $b^0$ is the corresponding b ion $-H_2O$; b* is the corresponding b ion $-NH_3$; b** is the corresponding b-ion $–COH_2$; N.O. not observed | | | |

**Table S7. Observed and calculated y-ions for peptide U2 with the sequence $M_aCE_aT_aNL_aAIK_aQR_aICGSGR_aL_aG$.**

| Sequence | # | y-ion | |
|---|---|---|---|
| | | Obs. | Calc. |
| $M_a$ | 19 | 752.68 $[y_{19}]^{+3}$ | 2256.05 |
| C | 18 | N.O. | 2119.005 |
| $E_a$ | 17 | 979.99 $[y_{17}]^{+2}$ | 1958.97 |
| $T_a$ | 16 | N.O. | 1787.94 |
| N | 15 | 836.94 $[y_{15}]^{+2}$ | 1672.88 |
| $L_a$ | 14 | 779.92 $[y_{14}]^{+2}$ | 1558.84 |
| A | 13 | 716.38 $[y_{13}]^{+2}$ | 1431.74 |
| I | 12 | 1360.71 | 1360.7 |

| | | | |
|---|---|---|---|
| $K_a$ | 11 | 1247.65, 624.31 $[y_{11}]^{+2}$ | 1247.62 |
| Q | 10 | 1121.58, 561.26 $[y_{10}]^{+2}$ | 1121.54 |
| $R_a$ | 9 | 993.47, 497.25 $[y_9]^{+2}$ | 993.48 |
| I | 8 | 835.40 | 835.40 |
| C | 7 | 722.32 | 722.32 |
| G | 6 | 562.29 | 562.29 |
| S | 5 | 505.28 | 505.27 |
| G | 4 | 418.23 | 418.23 |
| $R_a$ | 3 | 361.22 | 361.21 |
| $L_a$ | 2 | 203.14 | 203.13 |
| G | 1 | N.A. | 76.03 |
| N.A. not applicable; N.O. not observed | | | |

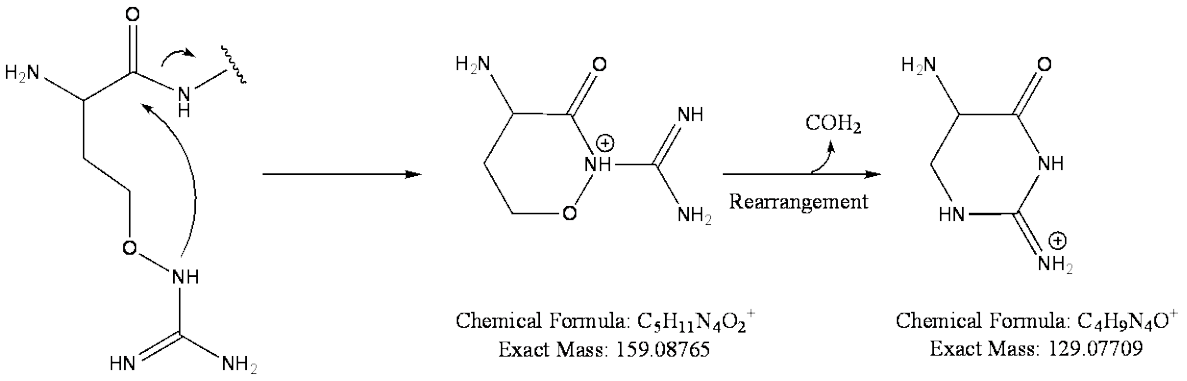
**Table S8. Other observed ions for peptide U2 with the sequence**

**$M_aCE_aT_aNL_aAIK_aQR_aICGSGR_aL_aG$.**

| Ion | Molecular Formula | m/z | | Diff. (ppm) |
|---|---|---|---|---|
| | | Obs. | Calc. | |
| $R_a$* | C5 H11 N4 O2 | 159.0867 | 159.08765 | 6 |
| $R_a$** | C4 H9 N4 O | 129.07653 | 129.07709 | 4.3 |
| C(Ac) | C4 H9 N2 O S | 133.04318 | 133.04301 | 1.28 |
| $[b_{11}**]^{+2}$ | C57H91N20O19S | 696.32442 | 696.32814 | 5.4 |
| $[b_{17}**]^{+3}$ | C80H131N30O28S2 | 675.31055 | 675.31119 | 0.9 |
| $[M]^{+3}$ | C90H150N32O32S2 | 752.6914 | 752.6926 | 2.6 |
| $R_a$* is formed from $y_3$ or $y_9$ ion by elimination of the remaining peptide chain via cyclization. *Ra*** is Ra* - $COH_2$. C(Ac) is the immonium ion of alkylated Cys; b** is the corresponding b-ion $–COH_2$; $[M]^{+3}$ is the triply charged parent ion; | | | | |

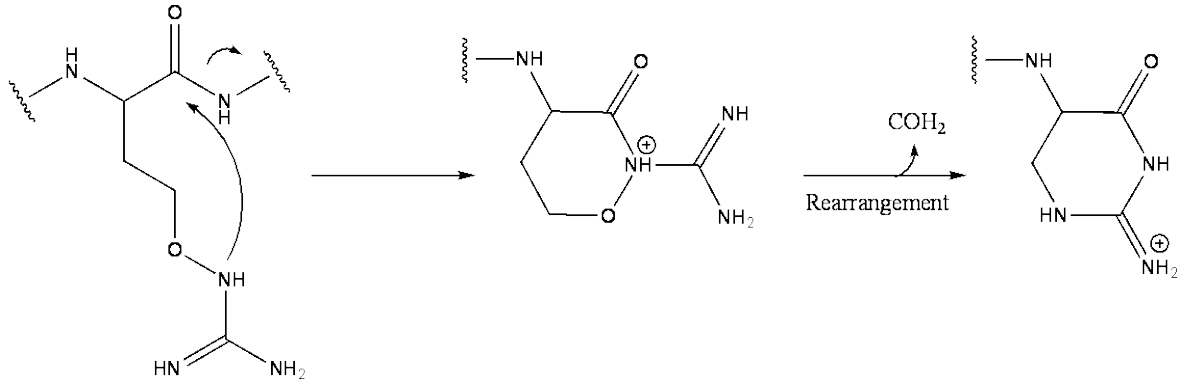
Scheme 1:

Possible mechanism for $R_a$* and $R_a$** formation:




Scheme 2:

Possible mechanism for $b_{11}$** and $b_{17}$** formation:



Commentary:

The use of racemic amino acids in translations:

Half of the unnatural amino acids used in this selection are racemic mixtures and we do not know what fraction of the enantiomers are substrates for the amino acid tRNA synthetases (aaRS). It is also unknown if the D-aa-tRNAs are substrates for the translational machinery which could allow for the incorporation of D-amino acids into peptides. However several lines of evidence suggests that the incorporation of D-amino acids in the presence of equal amounts of L-amino acids is very unlikely. Although some D-amino acids are substrates for some aaRS, the charged D-amino acyl tRNAs are poor substrates for the translational machinery.[4-8] We have observed missincorporation of near cognate tRNAs, and truncations before or after D-amino acid incorporation, when we have substituted D-amino acids for L-amino acids using N-terminally tagged peptides (unpublished results). In addition our specific ion- extracted LC-MS analysis of full-length cyclic peptides show only one large peak (data not shown) suggesting the presence of a single homogeneous species. The only case where we cannot completely rule out D-amino acid incorporation is at the initiation of peptide synthesis. Initiator tRNAs precharged with some D-amino acids are substrates for the initiation of peptide synthesis.[9] However, the presence of some L-amino acyl initiator tRNA significantly decreases incorporation of D-amino acids.[9] $M_a$ is an enantiomeric mixture and it is possible that some of the peptides synthesised contain some D-$M_a$, but we believe that this would represent only a very minor fraction of peptides, which we cannot detect. The inclusion of only L-amino acids could remove such uncertainties.

References:

(1) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.

(2) Josephson, K.; Hartman, M. C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 11727.

(3) Kurz, M.; Gu, K.; Lohse, P. A. *Nucleic Acids Res* **2000**, *28*, E83.

(4) Yamane, T.; Miller, D. L.; Hopfield, J. J. *Biochemistry* **1981**, *20*, 7059.

(5) Soutourina, J.; Plateau, P.; Blanquet, S. *J. Biol. Chem.* **2000**, *275*, 32535.

(6) Bain, J. D.; Diala, E. S.; Glabe, C. G.; Wacker, D. A.; Lyttle, M. H.; Dix, T. A.; Chamberlin, A. R. *Biochemistry* **1991**, *30*, 5411.

(7) Calendar, R.; Berg, P. *Journal of molecular biology* **1967**, *26*, 39.

(8) Dedkova, L. M.; Fahmi, N. E.; Golovine, S. Y.; Hecht, S. M. *Journal of the American Chemical Society* **2003**, *125*, 6616.

(9) Goto, Y.; Murakami, H.; Suga, H. *RNA* **2008**, *14*, 1390.