Supporting information

A robust data analytical method to investigate sequence-dependence in flow-based peptide synthesis

Bálint Tamás,[#] Pietro Luigi Willi,[#] Héloïse Bürgisser, Nina Hartrampf[*]

Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[#] these authors contributed equally to the project

[*] correspondence to nina.hartrampf@chem.uzh.ch

Table of contents

1 Material and general methods

1.1 Reagents and solvents

Fmoc- and side chain-protected L-amino acids (Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH) were purchased from Bachem; O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) and (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) were purchased from Advanced ChemTech; N,N-diisopropylethylamine (iPr$_2$NEt, DIPEA, 99.5%) was purchased from Sigma-Aldrich; trifluoroacetic acid (TFA, for HPLC, ≥99.0%), triisopropylsilane (TIPS, 98%) and 3,6-dioxa-1,8-octane-dithiol (DODT, 95%) were purchased from Sigma-Aldrich. N,N-Dimethylformamide (DMF) was purchased from the Supelco-line from Sigma-Aldrich Canada Ltd.; dichloromethane (DCM, ≥99.8%) was purchased from Fisher Scientific Ltd.; diethyl ether was purchased from Honeywell Riedel-de Haën; acetonitrile (MeCN, for HPLC gradient grade, ≥99.9%) was purchased from Sigma-Aldrich. NovaPEG Rink Amide resin (0.41 or 0.20 mmol/g loading) was purchased from the Novabiochem-line from Sigma-Aldrich Canada Ltd.

1.2 Automated flow-based peptide synthesis (AFPS)

Peptides were synthesized on an automated flow system built in the Hartrampf lab, which is similar to the published AFPS system.[S1] Capitalized letters refer to L-amino acids. Unless otherwise noted, the following settings were used for peptide synthesis: flow rate = 20 mL/min for coupling and deprotection steps, temperature = 90 °C (loop) for all canonical amino acids except histidine and cysteine, those were coupled on room temperature and 90 °C (reactor). The standard synthetic cycle involves a first step of prewashing the resin at 90 °C for 60 s at 40 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid, and a 5.0 mL/min pump head pumps iPr$_2$NEt (*neat*). The 50 mL/min pump head pumps delivered 0.398679 mL of liquid per pump stroke, the 5.0 mL/min pump head pumps 3.9239×10^{-2} mL of liquid per pump stroke.

All peptides were prepared by AFPS on NovaPEG Rink Amide resin (0.41 or 0.20 mmol/g) and standard Fmoc/tBu protected amino acids (0.40 M in DMF) were coupled using HATU (0.38 M in DMF) or PyAOP (0.38 M in DMF) with DIPEA (neat, 3.0 mL/min) at a total flow rate of 20 mL/min. For amino acids D, E, F, G, I, K, L, M, P, S, W, and Y, a total volume of 6.4 mL of the "coupling solution" (i.e. amino acid (0.20 M), HATU or PyAOP (0.19 M), and DIPEA in DMF) was applied for each coupling. For amino acids A, C, H, N, Q, R, S, T, and V, a total of 10.4 mL of "coupling solution" was applied for each coupling. All amino acids except C and H were preheated at 90 °C during the activation step with HATU or PyAOP, whereas C and H were preactivated with PyAOP at room temperature. Removal of the N^α-Fmoc group was achieved using 20% piperidine with 1% formic acid in DMF at a flow rate of 20 mL/min and a total volume of 6.4 mL at 90 °C. Between each coupling and deprotection step, the resin was washed with DMF (32 mL) at 90 °C with a flow rate of 40 mL/min. After completion of the peptide sequence, the resins were manually washed with DCM (3 × 5 mL) and dried under reduced pressure.

1.3 Batch Synthesis

Unless otherwise noted, pre-functionalized NovaPEG Rink Amide resin (0.41 or 0.20 mmol/g loading) was used in all experiments. The amino resin (1 eq., 21 μmol.) was swelled with DCM (1 × 5 mL) for

1 min. and washed with DMF (1 × 5 mL), then the solvent was removed by filtration under reduced pressure. For each coupling: a solution of Fmoc- and sidechain-protected amino acid (0.5 mL, 0.2 M in DMF, 5.0 eq.) and HATU (0.5 mL, 0.19 M in DMF, 4.8 eq.) was prepared. To this solution, DIPEA (10 eq.) was added, and the solution was gently agitated at 23 °C for 1 min. The solution was then added to the resin, and the reaction was gently stirred for 20 sec., then left to react at 23 °C for 30 min. Afterwards, the resin was drained, washed with DMF (3 × 5 mL) and DCM (3 × 5 mL). For each deprotection: 2 mL of solution of 20% piperidine in DMF (*v/v*) was then added to the reactor, the resin was gently stirred for 20 sec. then left to react at 23 °C for 20 min. Afterwards, the resin was drained, then washed with DMF (3 × 5 mL) and DCM (3 × 5 mL).

1.4 Peptide cleavage and deprotection

The cleavage of the peptides from the resin was initiated by weighing 13 mg of the dry peptidyl resin. All peptides were cleaved using a solution of TFA/TIPS/DODT/H_2O (94:1:2.5:2.5, *v/v/v/v*, 1–3 mL) for 2 h at room temperature with gentle mixing. TFA was then removed by evaporation under a light stream of N_2, and the peptides were precipitated and isolated by centrifugation from ice-cold diethyl ether (14–45 mL), twice. The resulting peptide pellets were then briefly dried under a light stream of N_2, then dissolved in 2 mL of an aqueous solution containing 50% MeCN and 0.1 % TFA and lyophilized. Crude peptides were then analyzed by LC–MS and UHPLC (214 nm). Pure peptide samples were obtained using RP-HPLC and were analyzed for purity by LC–MS and UHPLC (214 nm).

1.5 Analytical Ultra-High Performance Liquid Chromatography (UHPLC)

For determination of purity by UHPLC, the filtered peptide solution was diluted in 10–50% acetonitrile (MeCN) in water with 0.1% TFA (500 µL) to a final concentration of approximately 1.0 mg/mL. The samples were analyzed on Agilent 1290 Infinity II Series using Agilent OpenLab CDS and ChemStation software.

For standard analysis of all peptide samples, analytical UHPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C18 Narrow-Bore Rapid Resolution HD column (2.1 mm × 150 mm, 5.0 µm particle size) at a flow rate of 0.8 mL/min with UV detection at 214 nm. A binary solvent system was used, wherein Solvent A was 5% MeCN in 95% water with 0.1% TFA, and Solvent B was 95% MeCN containing 5% water and 0.1% TFA. After an isocratic period at 0% Solvent A for 3 min, a linear gradient of 0–100% Solvent B (corresponding to 5–95% MeCN) over 20 min was used (*ca.* 4.5% MeCN/min). The total method time was 23.1 min. Then, the column was re-equilibrated using a post-run method at 0% Solvent B for 2 min. Purities of the crude and purified peptides were determined by ChemStation integration of all UHPLC signals at 214 nm in the area of 3–18 min.

For Barstar[75–90] the crude purities were determined through the area under the curves between 9–13 min. and for NBDY[53–68] area between 6–9 min.

1.6 Liquid Chromatography with High-Resolution Electrospray Ionization Mass Spectrometry (LC–MS)

For determination of peptide masses and purity by LC–MS, the filtered peptide solution was diluted in 10–50% acetonitrile (MeCN) in water with 0.1% TFA (60–500 µL) to a final concentration of approximately 0.1 mg/mL. The samples were analyzed on an Acquity UPLC (Waters, Milford, USA) connected to an Acquity eλ diode array detector and a Synapt G2HR-ESI-QTOF-MS (Waters, Milford, USA).

For standard analysis of all peptide samples, LC–MS spectra were recorded on an Acquity BEH C8 HPLC column (2.1 × 100 mm, 1.7 µm particle size, Waters) kept at 30 °C at a flow rate of 0.4 mL/min with UV detection at 190–300 nm. A binary solvent system was used, wherein Solvent A was water

containing 0.02% formic acid and 0.04% TFA, and Solvent B was MeCN containing 0.04% formic acid and 0.02% TFA. The LC–MS gradient used was as follows: isocratic at 3% Solvent B for 3 min, then linear gradient of 3–95% Solvent B over 9 min, followed by isocratic at 95% Solvent B for 1 min.

UV spectra recorded at 1.2 nm resolution and 20 points s^{-1}; ESI: positive ionization mode, capillary voltage 3.0 kV, sampling cone 40V, extraction cone 4V, N_2 cone gas 4 L h^{-1}, N_2 desolvation gas 800 L min^{-1}, source temperature 120 °C; mass analyzer in resolution mode: mass range 150–3000 m/z with a scan rate of 1 Hz; mass calibration to <2 ppm within 50–2500 m/z with a 5.0 mM aq. soln. of HCO_2Na, lock masses: m/z 195.0882 (caffein, 0.7 ng mL^{-1}) and 556.2771 (Leucine-enkephalin, 2 ng mL^{-1}).

All mass spectra show deconvoluted masses from the raw m/z values, calculated using Mestrelab Research S.L.© MestReNova v. 14.1 Mnova MS Suite. Purity based on LC–MS was calculated by calculating the Area Under the Curve (AUC) of desired product peak as a percentage of the AUC of all peaks (within 2–8 min) of the absorbance chromatogram ($\lambda = 214$ nm).

2   Data processing and analysis

The program is composed of four python scripts: The main script (**afps_dataiter.py**) iterates through the files containing the synthesis data (.pep files) and extracts the synthesis data. The output of the main script is plotted using the **plotting.py** script. Using the areas computed by the **afps_dataiter.py** script, the **permutator.py** script combinatorically iterates through all the deletion permutations and calculates their likelihood based on the areas. The **aggregationfinder.py** script uses the output of the main script to find the position and magnitude of aggregation based on the change in peak angle. The three scripts all contain a class called by an IPython notebook, **AFPSDataExport.ipynb**.

2.1   Peak isolation and computation of area, height, and width

The afps_dataiter contains the synthesis class, which consists of various methods enabling it to open the raw synthesis (.pep) files, isolate the deprotection peaks and then calculate a set of summary statistics. First, the baseline is corrected by subtracting a fitted third-order polynomial. The deprotection peaks are then isolated by querying the steps that correspond to deprotection. New deprotection starts every 22 steps and the first deprotection step occurs at step 36. The summary statistics that are then calculated from the deprotection peaks are area, maximum height, full width at half maximum and angle. The area is calculated by taking the Riemann sum under the peak using the following formula:

$$a = \sum \frac{0.5(A_{n+1} + A_n)}{t_{n+1} - t_n}$$

Where a stands for the area under the peak, A for absorption and t for time (ms). The height of the peak is simply computed by just taking the maximum value of the peak. The width is found by calculating the difference in time between the two half maxima at either side of the peak.

2.2   Angle computation

The angle computation differs in complexity from the previous summary statistics in 2.1. Its complexity is derived from the fitting of a Gaussian function onto the peak and from the need to increase the robustness of the method. The whole angle computation is split into two main methods, process_peak and get_angle. The process_peak method is given the raw deprotection peak as input. The peak is first split into front and tail at its maximum by calling the static method split_peak. A threshold is set so that any absorbance values larger than the threshold (0.90 AU) are set to the threshold + 0.001. This ensures that all the oversaturated parts of the peak have the same values. Then the middle of the peak is found by finding the median time of the maximum absorbance. The thresholding makes it easier to find the middle of the peak in case of oversaturation. The peak is then split into front and tail along the median of the maxima. Using the static method mirror_peak, the two half peaks are then individually mirrored

and concatenated with the original half peak, resulting in two full Gaussian-like peaks. Finally, the static method trim_peak, removes the oversaturated part of the peak by simply deleting the points with an absorbance higher than the threshold by their index. In addition, it also trims the peaks at their minima, making the peak more Gaussian-like. These three static methods are all called by the process_peak method in the order described to yield the coordinates of the front and tail peaks.

The get_angle method then calls the process_peak method and fits a parametrized Gaussian function with the following formula:

$$A = ae^{\frac{-(t-b)^2}{2c^2}}$$

Where t stands for time and A for absorption. The curve fitting yields the parameters a, b and c for the front and the tail peak. The parameters are then used to find the maximum gradient of the function G. The maximum gradient corresponds to the gradient at the point of inflection. The gradient of the function G is defined by its first order derivative, $\frac{dA}{dt}$:

$$\frac{dA}{dt} = \frac{-(t-b)}{c^2}ae^{\frac{-(t-b)^2}{2c^2}}$$

The arctangent of one over the gradient at the inflection will yield the angle at the top of the peak. After performing the angle computation both on the front and tail peak, the two resultant angles are summed together to yield the peak angle using the following formulas:

$$\alpha_{front} = \arctan\left(\frac{1}{max\left(\frac{dA}{dt}^{front}\right)}\right)$$

$$\alpha_{tail} = \arctan\left(\frac{1}{max\left(\frac{dA}{dt}^{tail}\right)}\right)$$

$$\alpha = \alpha_{front} + \alpha_{tail}$$

The get_integrals method iterates through the steps that are defined as deprotections and computes the various summary statistics.

2.3   Angle normalization using the mass

The angle is standardized using the resin mass. The mass normalization is performed to remove the peak angle's dependence on resin mass (but not resin loading). The standardization normalizes the peak angle to a defined standard resin mass of $m_{st} = 150$ mg using the following formula:

$$\alpha_{st} = 180° - \frac{m_{st}}{m}(180° - \alpha)$$

Where m is mass of the resin used during the synthesis in mg, m_{st} the standard resin mass in mg, α the measured angle and α_{st} is the standardized angle.

2.4   Trimming of outlying temperatures

The trim_peak method performs a temperature correction. To reduce outliers, points belonging to deprotections that have been performed outside of a +/-20% temperature range compared to the median temperature of the synthesis are set to the average of the previous and next points carried out within the temperature range. If a temperature anomaly occurs as the first or last deprotection this point is set equal to only one neighbor.

2.5   Aggregation Detection

After the listed descriptors of all deprotection peaks have been calculated and added to a pandas dataframe, an object of the AggregationFinder class contained by the aggregationfinder.py script is instantiated using the dataframe created by the Synthesis class and the name of the parameter onto which the method of the AggregationFinder class is to be applied. The methods were designed to be used with the peak angle, but they were also designed to work with other parameters. This class executes the automated detection of the position and magnitude of the aggregation, based on the detection of a sudden permanent increase of the peak angle. Two different methods of aggregation detection are applied. The cumulative slope method computes the average slope of the summary statistic trace between a point and all the other point for every point, using the following formula:

$$C_k = \frac{\sum_{\substack{i=1 \\ i \neq k}}^{n} \frac{y_i - y_k}{x_i - x_k}}{n}$$

Where y_k is the angle belonging to the peak of choice, y_i the angle belonging to every other peak, x_k is the position of the peptide in a reversed C→N peptide, and x_i is the position of all the other amino acids. n is the length of the synthesized peptide, to normalize amongst different peptide lengths. The maximum of the cumulative slope corresponds to the position and magnitude of aggregation for a particular synthesis.

The other method fits a parametrized sigmoid curve onto the summary statistic trace using the following formula and uses the point of inflection of the sigmoid as the position of aggregation.

$$\sigma(x) = \frac{a}{1 + e^{dx - b}} + c$$

The $\frac{a}{c}$ indicates the magnitude of aggregation and $\frac{b}{d}$ the position of it.

The cumulative slope is calculated by the accumulate_slope method, yielding the gradient array, while the sigmoid is fit by the fit_sigmoid method.

The find_aggregation method then iterates through a dataframe containing the synthesis summary statistics for one or multiple syntheses. It differentiates between different syntheses by querying them by their serial numbers sequentially. It first calls the trim_peak method, then the add_sequence method. The add_sequence method computes the full sequence and the growing peptide chain sequence of the synthesis. By iterating through the unique serial numbers of the syntheses, the add_sigmoid method and the accumulate_slope method are clled on each synthesis individually.

The afps_dataiter script, more specifically the parts performing the parsing of the .pep files were adapted from the literature.[SI1,2]

2.6   Loading determination

Using the areas of the deprotection peaks, the loading of the solid support can be determined. Loading, L, is defined as mmol of binding points, n, per mass unit of resin, m. In most techniques to determine the loading, the number of binding points are approximated with the amount of Fmoc removed after the first coupling. Using this assumption, the determination is done using the following formula:

$$L = \frac{n}{m}$$

$$n = V \int_{t_{start}}^{t_{end}} cdt$$

$$A = l\varepsilon c$$

Combining these formulas results in:

$$L = \frac{\dot{V}}{ml\varepsilon} \int_{t_{start}}^{t_{end}} Adt$$

In this final equation V stands for flowrate, l for part length in the flow cell of the UV-detector, ε is the extinction coefficient of the deprotection solution and the integral of the absorption (A) over time is the area of the peak. Using this simple formula, we can automatically determine the loading of the resins with ease, eliminating the cumbersome extra step its measurement required.

2.7   Deletion Computation using peak integrals

The deletion probability calculation assumes that the first amino acid was coupled with 100% coupling efficiency. Under this assumption we can calculate the coupling efficiency of all the other amino acids by dividing their deprotection peak area by the area of the first deprotection. The different deletion permutations are enumerated by using the combinatorial property of binary numbers. The number of different deletion combinations is equal to $(n - 1)^2 - 1$, where n is the length of the sequence. By counting down from the maximum number of combinations in binary, all the various deletion permutations are enumerated. In the binary number the 0 digits correspond to a deletion at that particular position. This same property that is used to enumerate the permutations is also used to calculate the probability. The probability of a particular deletion permutation can be calculated using the following formula:

$$P = \prod_{i=0}^{n} \left(b_n A_n + b_n^{-1} A_n^{-1} \right)$$

Where b is the binary array that represents a particular deletion permutation, b_n is the digit at position n of the binary array, A_n is the area of the n-th deprotection normalized by the area of the first deprotection, $A_n^{-1} = 1 - A_n$, and b_n^{-1} is the ones' complement of the binary array with the first digit always set to 1 as it corresponds to the first deprotection. This is all implemented in the DeletionPermutator class of the permutator.py script, which is provided with the pandas dataframe containing the processed data from the synthesis class and the serial number of the synthesis of interest. The class has three methods: the convert_to_binary_with_leading_zeros static method converts numbers into binary with leading zeros and one in the first position, the permute_sequence method iterates enumerates the deletion permutations by calling the convert_to_binary_with_leading_zeros and the compute_mass calculates the mass for every deletion permutation.

2.8   Comparison of the efficiency of data analytical methods for oversaturated signals

To investigate the capability of the data analytical methods in capturing aggregation in signals that have oversaturated peaks we generated the oversaturation *in silico*. Taking the standard synthesis (150 mg, normal loading, Barstar[75–90] synthesis) we gradually cut down the signal. Once the value of the cutting down reached the tallest peak, that is where oversaturation was considered 0%. For every value of oversaturation, we analyzed the traces using both the peak angle and aggregation factor and compared the similarity of the plots to the original analysis using R^2. An example for this can be seen in **Figure S1** at approximately 50% oversaturation.
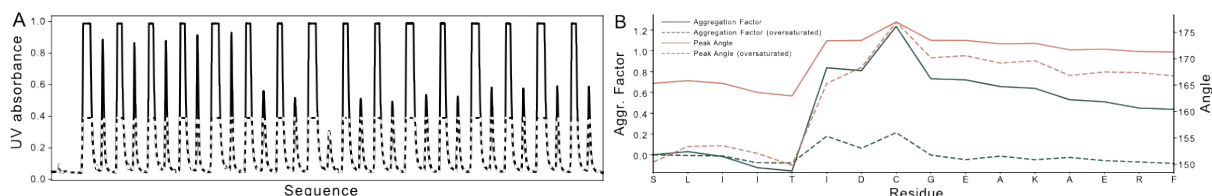
Figure S1: A) An example for the artificially generated oversaturated plot at the UV absorption value of 0.4, which corresponds to an approximate 50% oversaturation, compared to the original plot **B)** The analysis of the oversaturated plot using both peak angle and aggregation factor. It is clearly visible that while peak angle maintains its characteristics and covers the aggregation point accurately aggregation factor loses accuracy. By calculating and plotting the R2 we can compare the capability of the two methods in handling oversaturation

3 Screening of synthesis parameters and their effect on aggregation

3.1 Resin Mass

3.1.1 NBDY [53–68]

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Synthesis conditions: standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for resin mass. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

Comparison of both data processing methods:
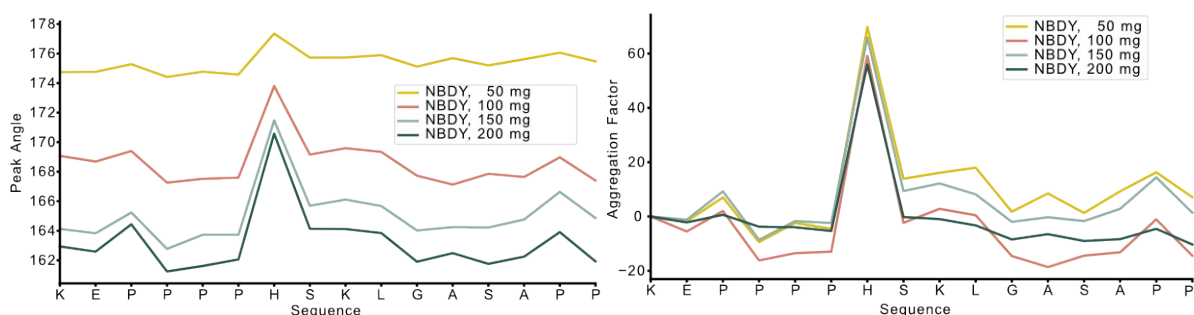
Figure S2: A) Mass screening of NBDY analyzed with the peak angle B) Mass screening of NBDY analyzed with the aggregation factor

50 mg:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 50 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~37 mg

Crude purity (UHPLC): 96%

Synthesis time: 45 min.
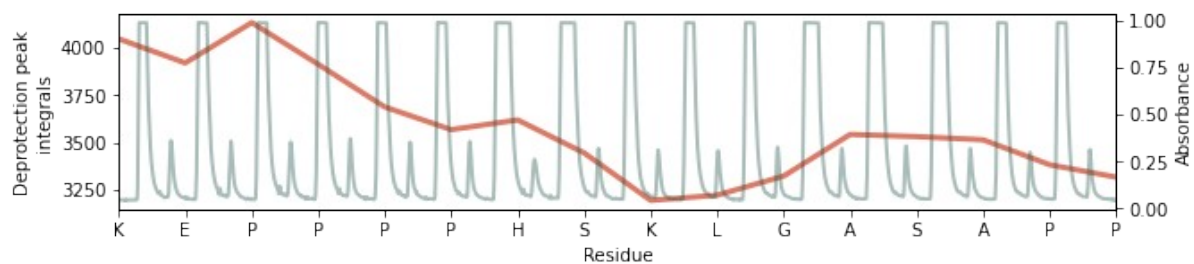
Figure S3: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized on 50 mg of resin but otherwise under standard AFPS conditions.
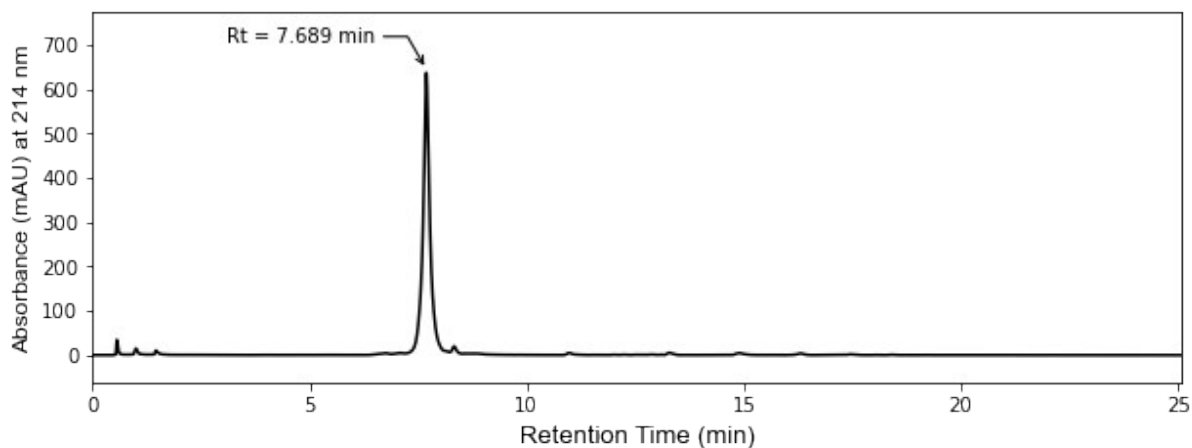
UHPLC analysis of crude NBDY[53–68]
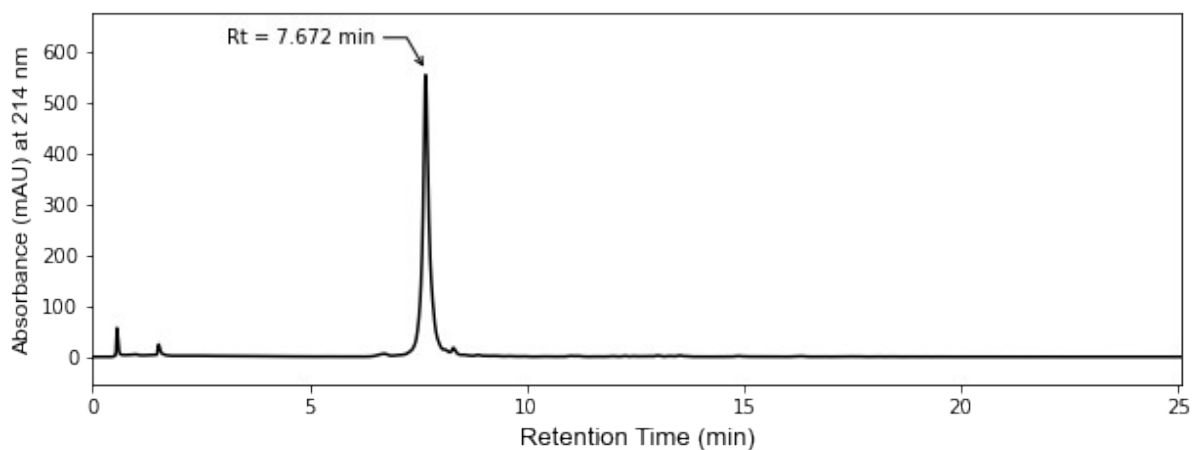
Figure S4: UHPLC of NBDY[53–68] synthesized with 50 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.
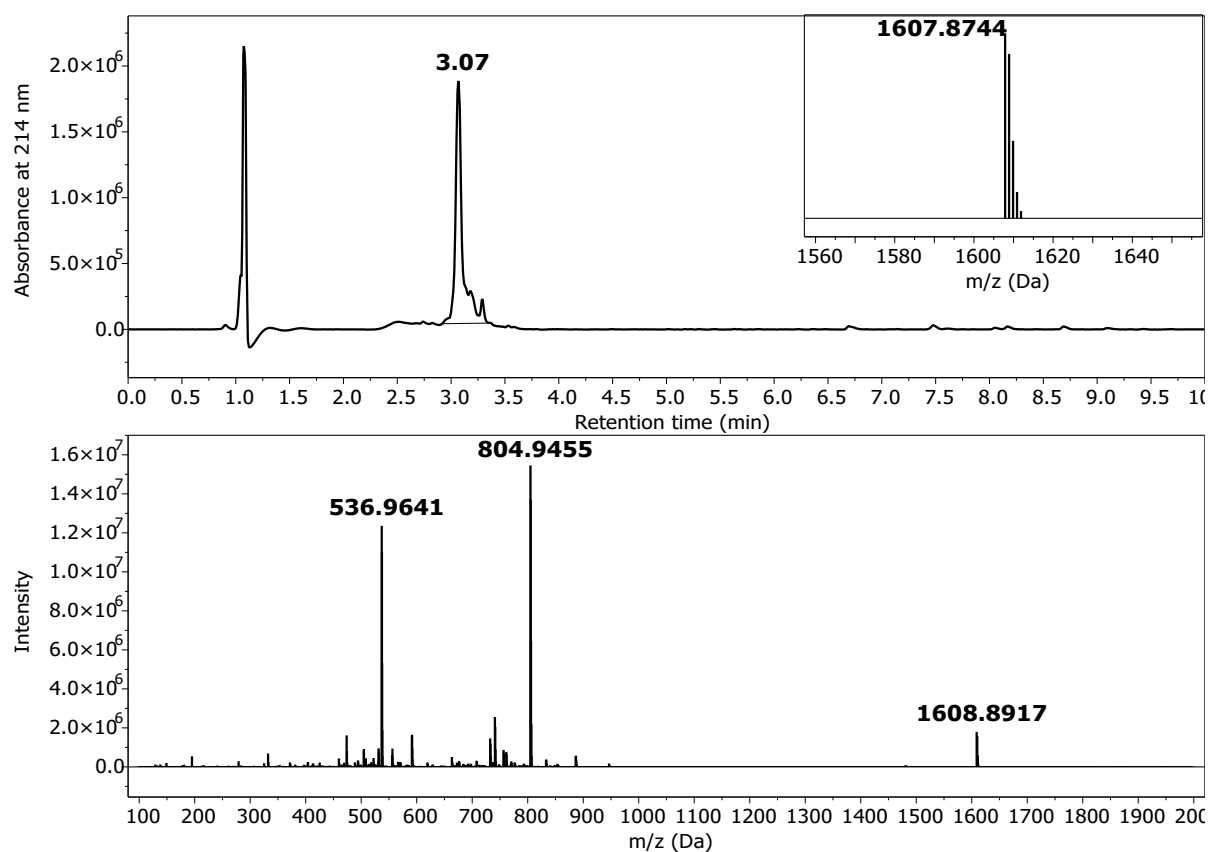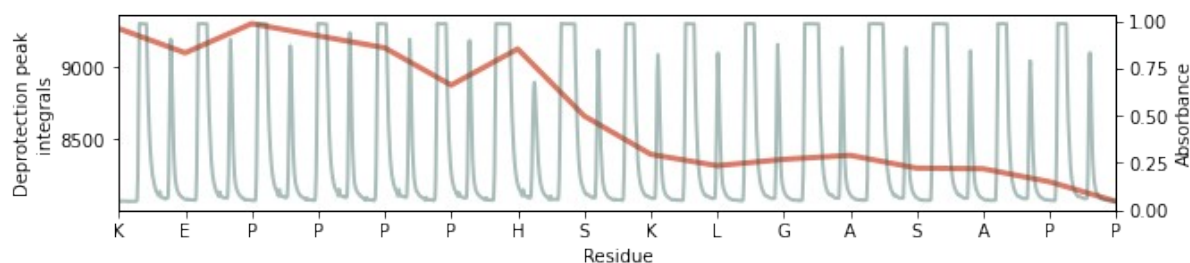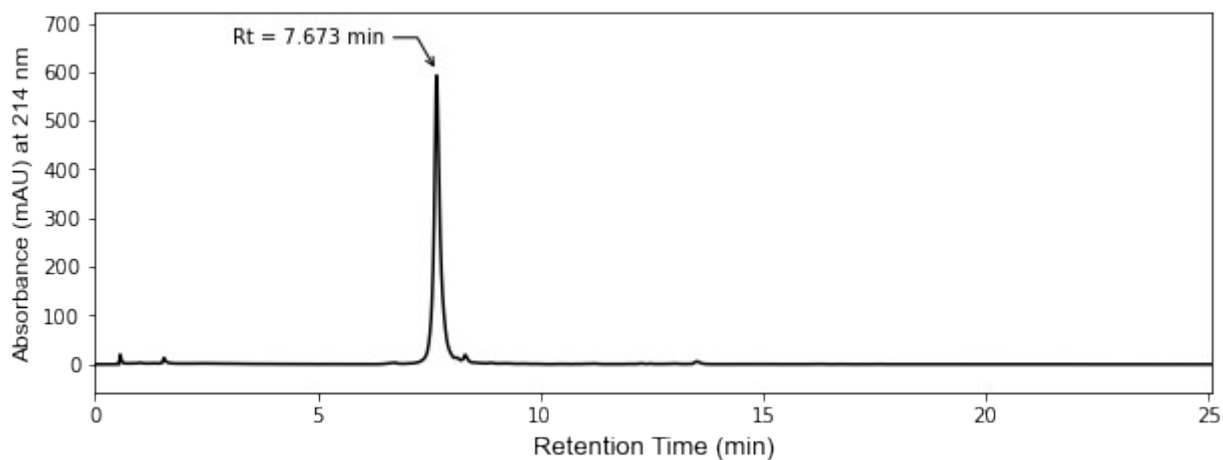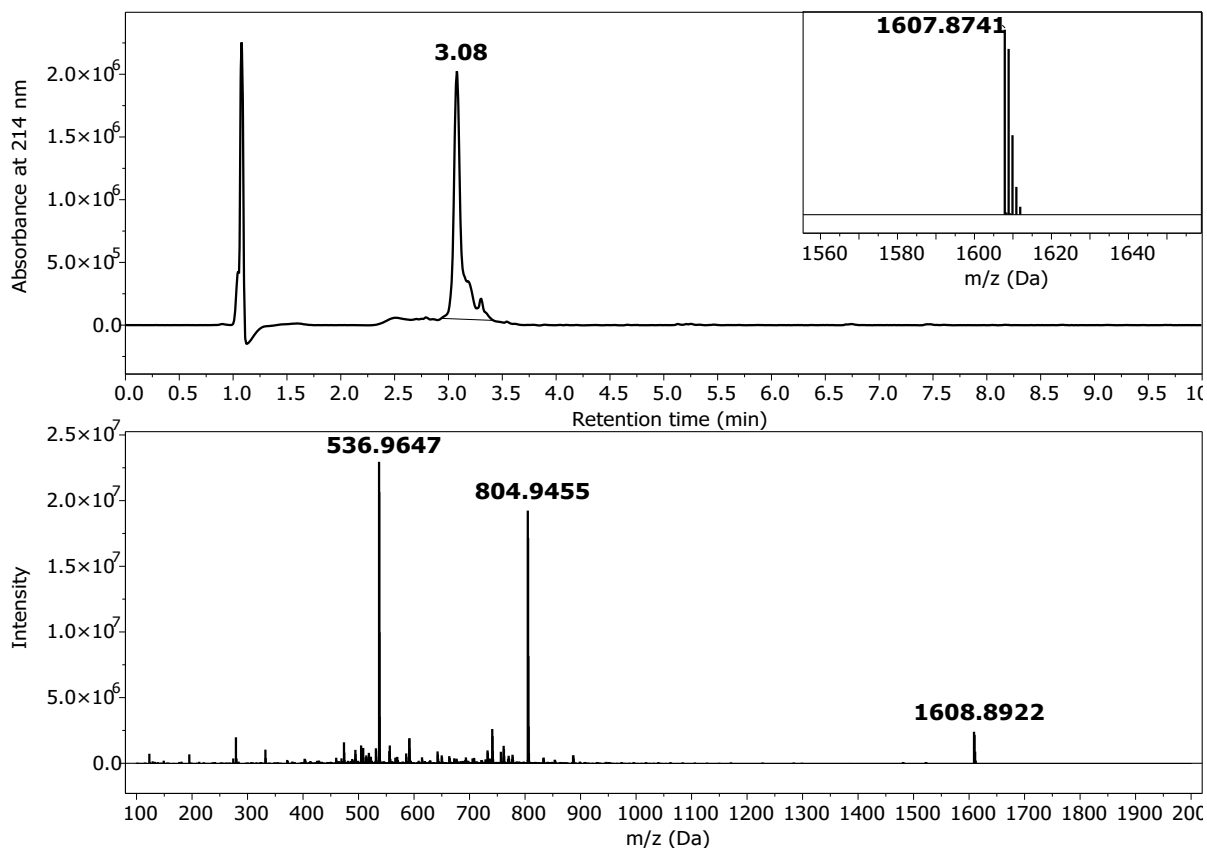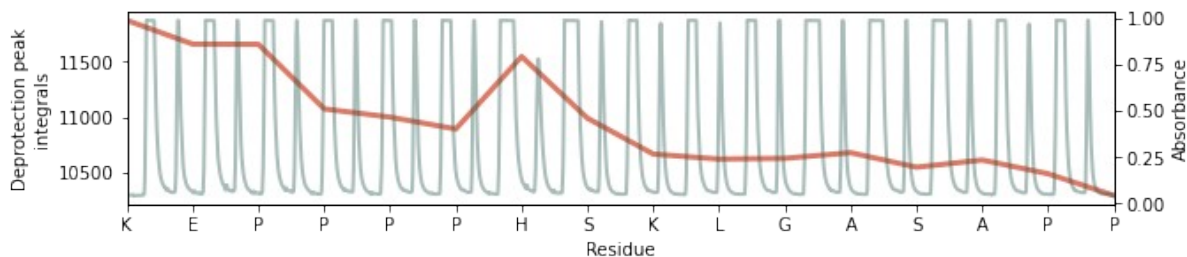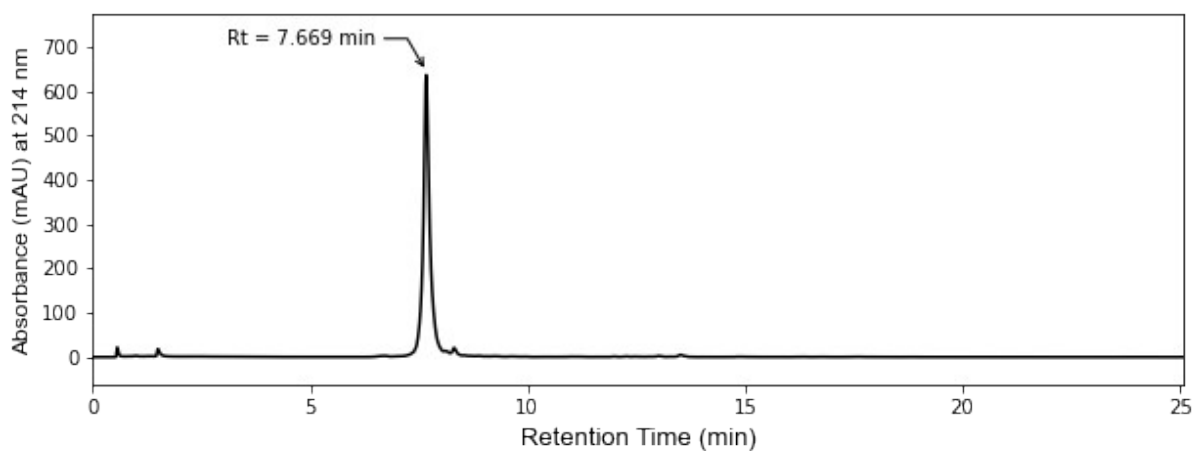
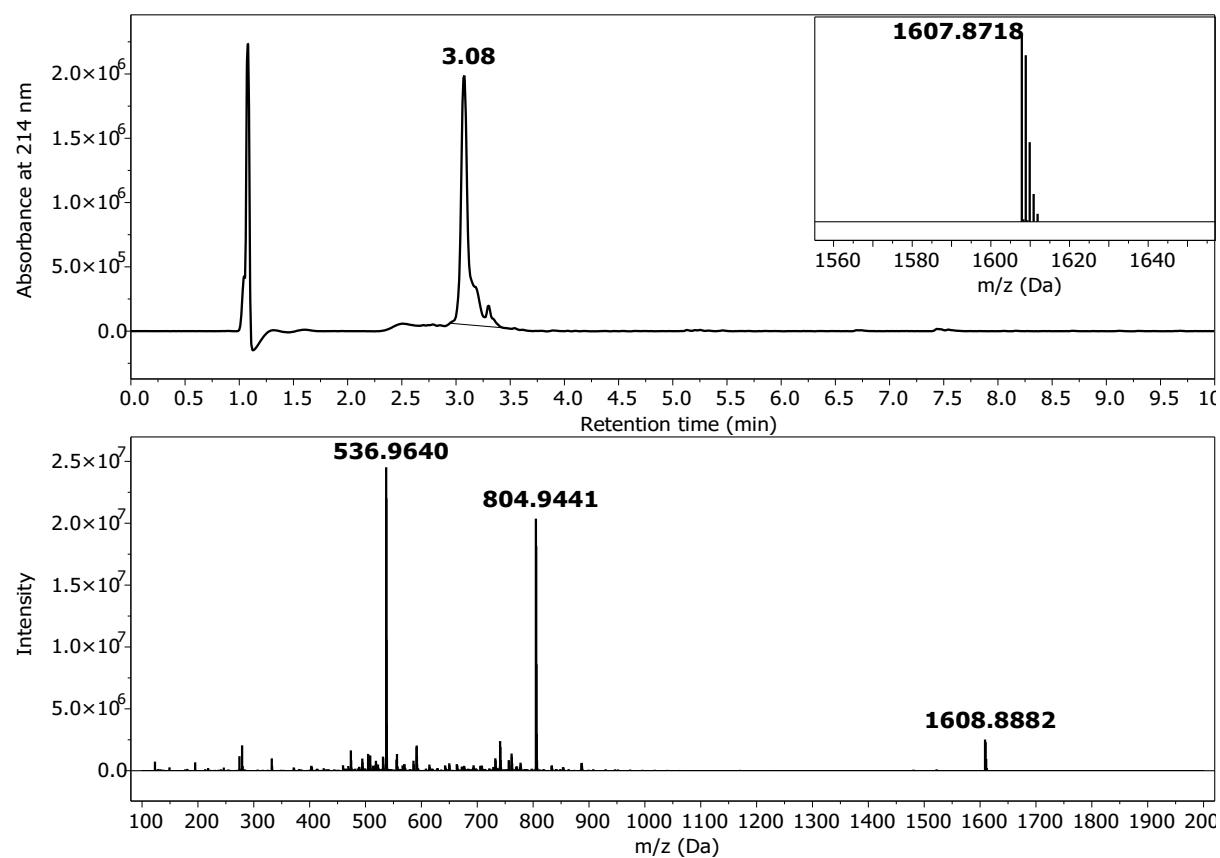## LC–MS analysis of crude NBDY[53–68]
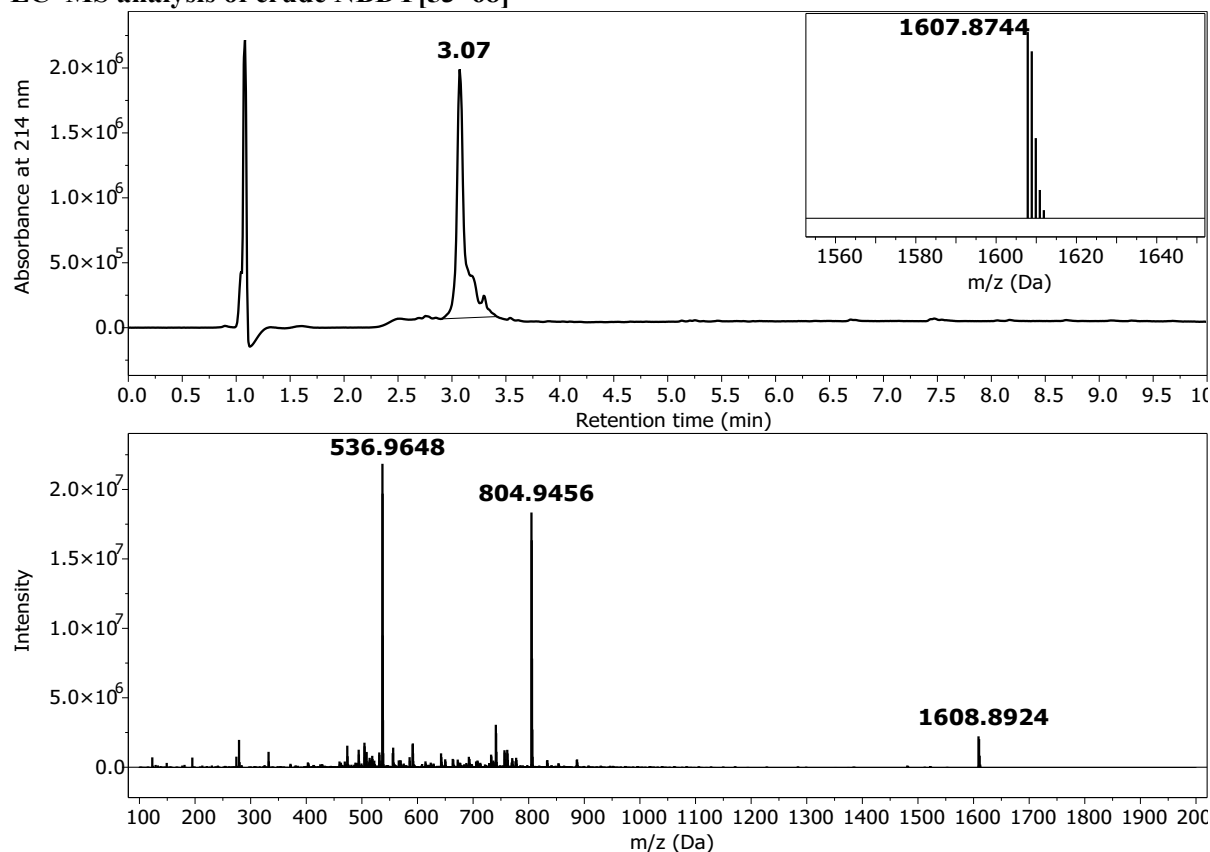
Figure S5: LC–MS profile of crude NBDY[53–68] synthesized with 50 mg of resin. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8744.

100 mg:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 100 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 41.0 μmol)

Crude peptide mass: ~70 mg

Crude purity (UHPLC): 98%

Synthesis time: 45 min.
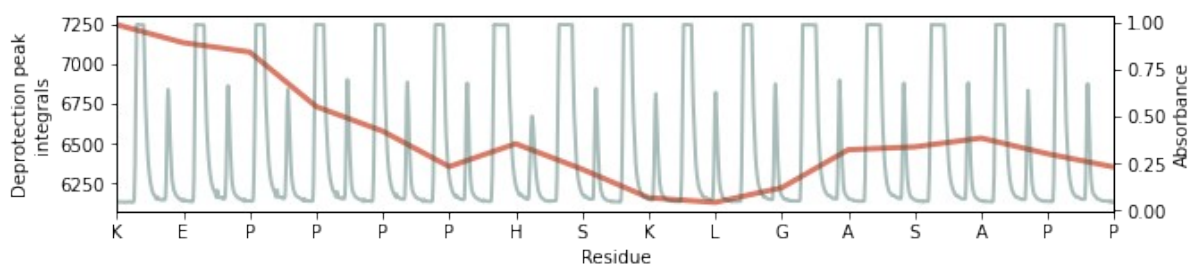
Figure S6: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized on 100 mg of resin but otherwise under standard AFPS conditions.

UHPLC analysis of crude NBDY[53–68]

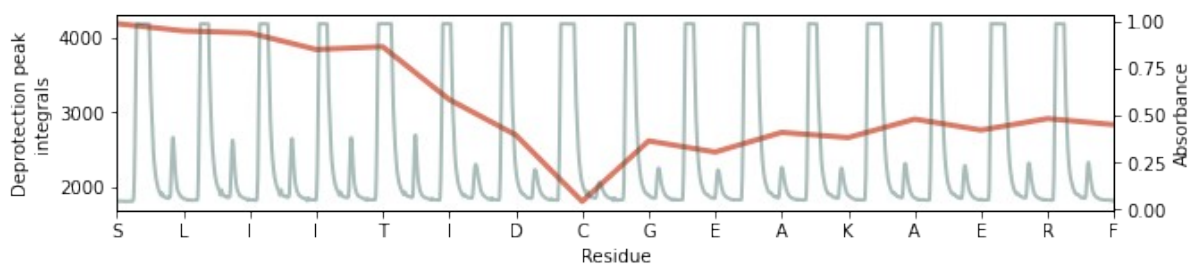**Figure S7:** UHPLC of NBDY[53–68] synthesized with 100 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

11

LC–MS analysis of crude NBDY[53–68]

Figure S2: LC–MS profile of crude NBDY[53-68] synthesized with 100 mg of resin. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8744.

150 mg:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 μmol)

Crude peptide mass: ~119 mg

Crude purity (UHPLC): 96%

Synthesis time: 45 min.

Figure S9: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized under standard AFPS conditions

UHPLC analysis of crude NBDY[53–68]

Figure S10: UHPLC of NBDY[53–68] synthesized under standard AFPS conditions. The standard UHPLC procedure was used.

LC–MS analysis of crude NBDY[53–68]

Figure S11: LC–MS profile of crude NBDY[53–68]. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8741.

13

200 mg:

Sequence: `PPASAGLKSH PPPPEK-CONH`$_2$ (16 amino acids)

Resin: 200 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 82.0 μmol)

Crude peptide mass: ~147 mg

Crude purity (UHPLC): 96%

Synthesis time: 45 min.

Figure S12: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized on 200 mg of resin but otherwise under standard AFPS conditions.
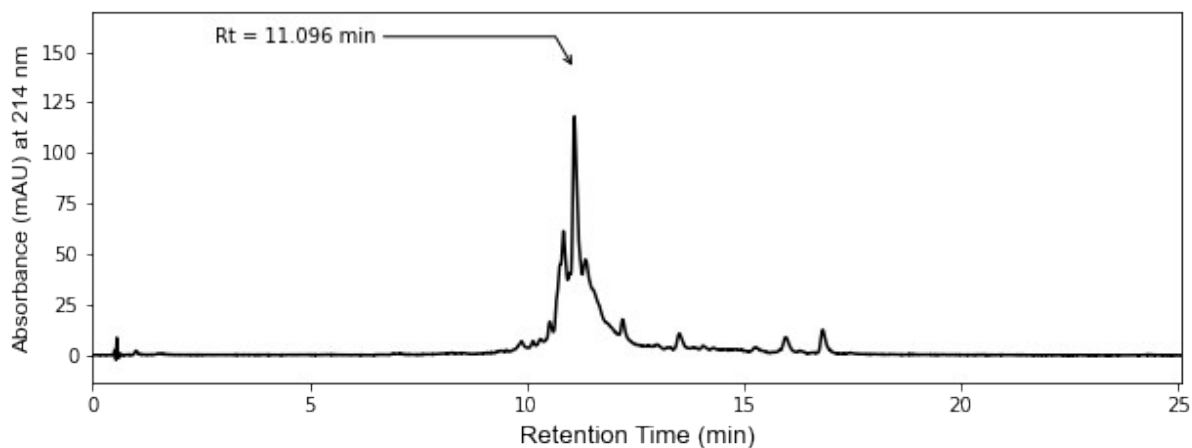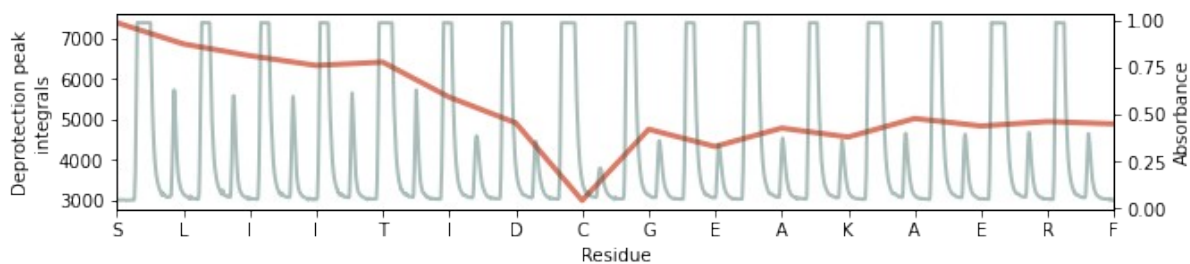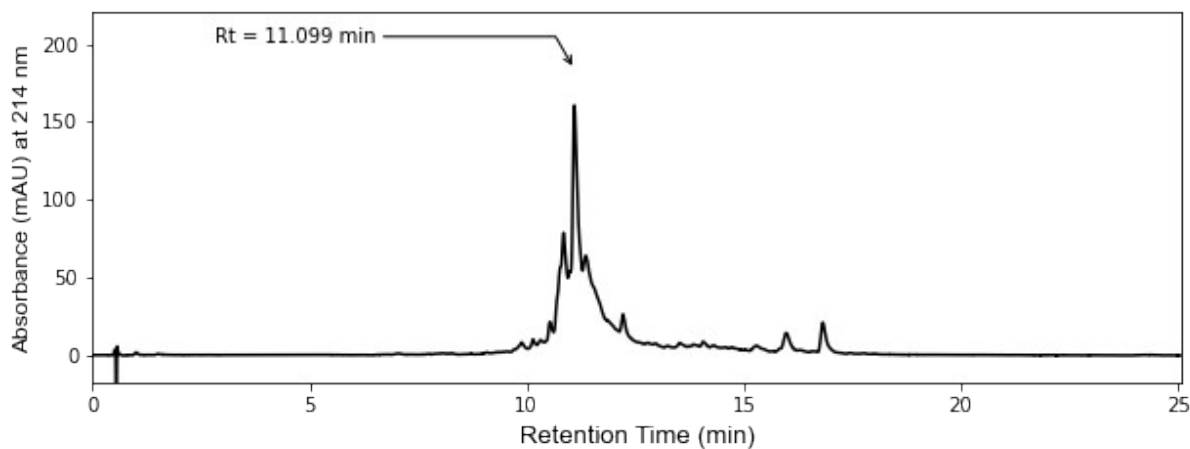
UHPLC analysis of crude NBDY[53–68]

Figure S13: UHPLC of NBDY[53–68] synthesized with 200 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

LC–MS analysis of crude NBDY[53–68]

Figure S34: LC–MS analysis of crude NBDY[53–68] synthesized on 200 mg of resin. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8718.

3.1.2 Barstar[75–90]

Sequence: `FREAKAEGCD ITIILS`-CONH$_2$ (16 amino acids)

Synthesis conditions: standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for resin mass. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

Comparison of both data processing methods:
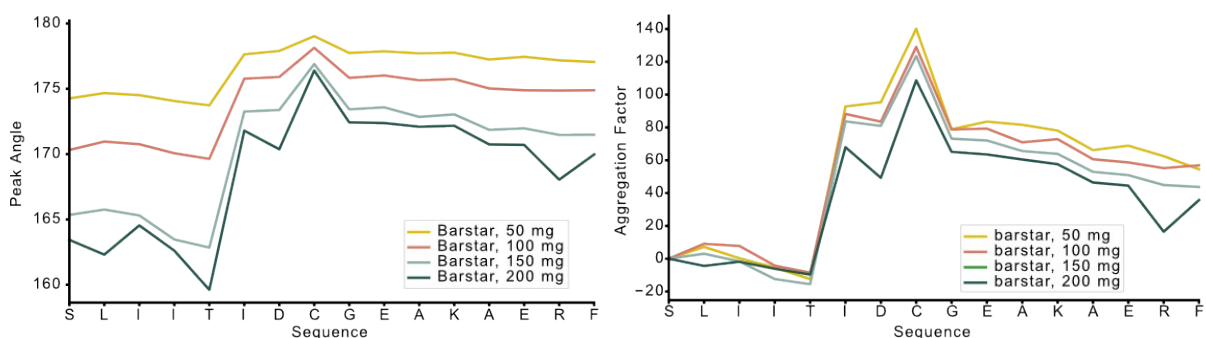
Figure S15: A) Mass screening of Barstar[75–90] analyzed with the "peak angle" data processing method. B) Mass screening of Barstar analyzed with the "aggregation factor" data processing method.

50 mg:

Sequence: `FREAKAEGCD ITIILS`-CONH$_2$ (16 amino acids)

Resin: 50 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~32 mg

Crude purity (UHPLC): 36%

Synthesis time: 45 min.

Figure S46: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized on 50 mg of resin but otherwise under standard AFPS conditions.

UHPLC analysis of crude Barstar[75–90]

Figure S17: UHPLC of Barstar[75–90] synthesized with 50 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.
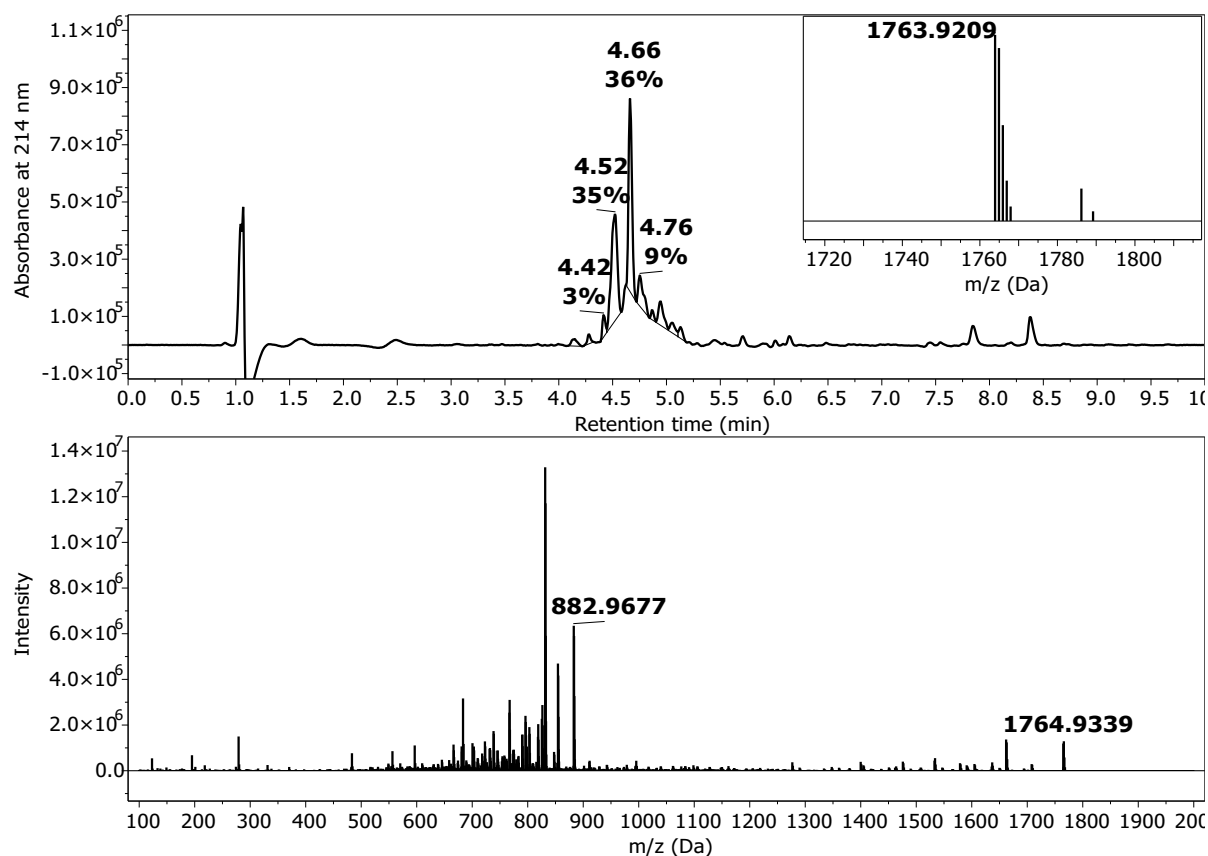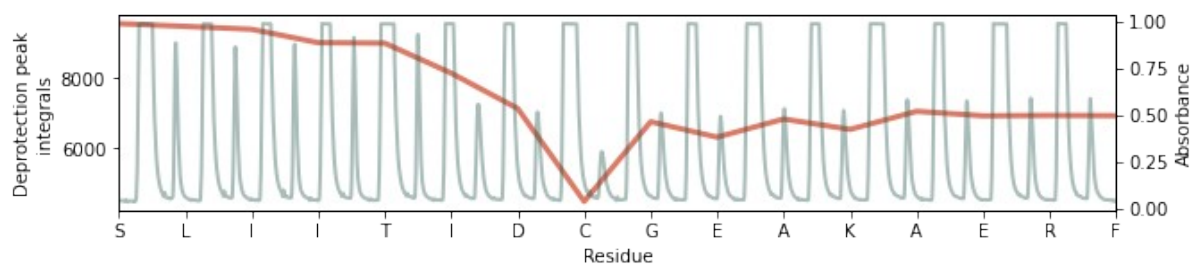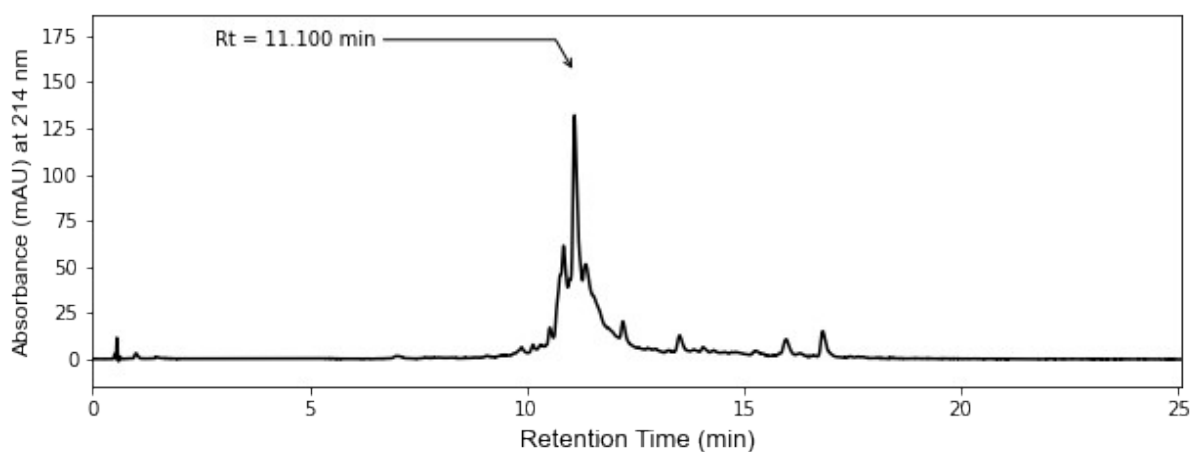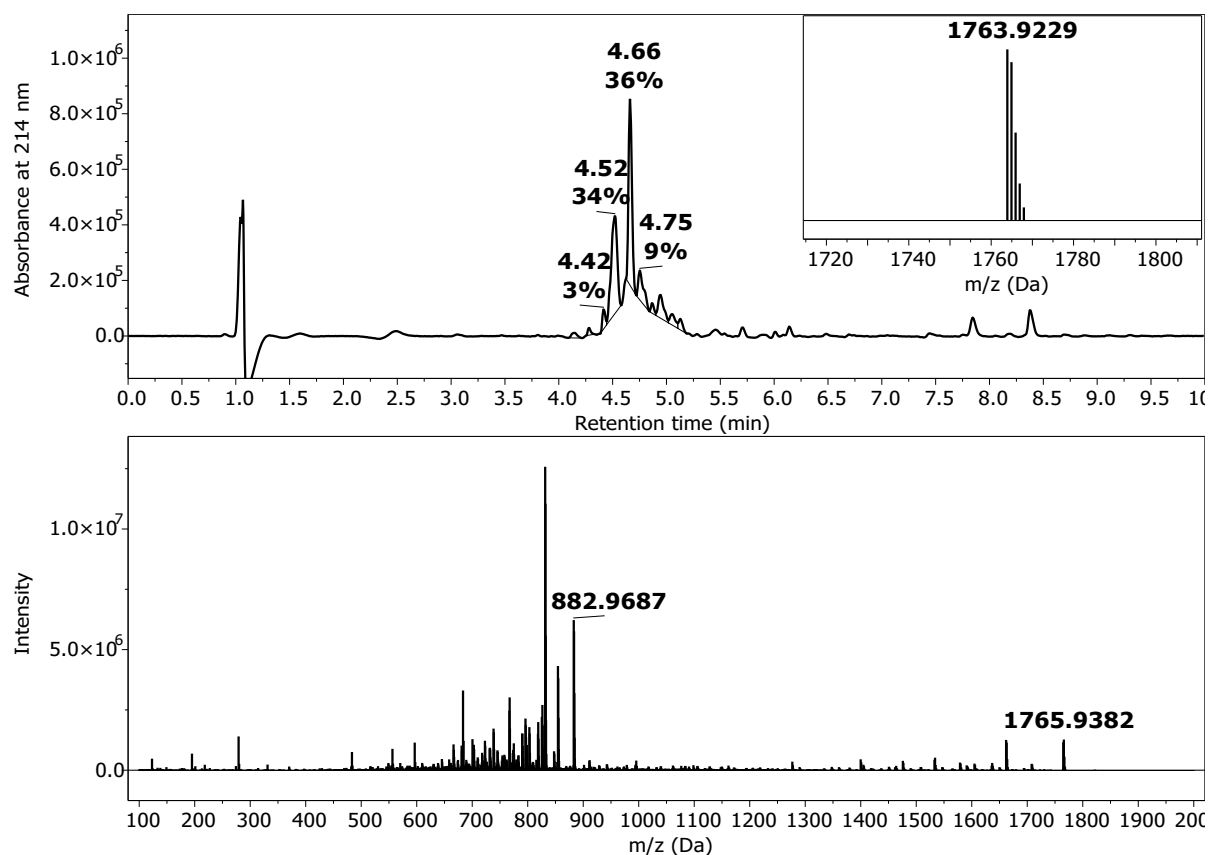
LC–MS analysis of crude Barstar[75–90]
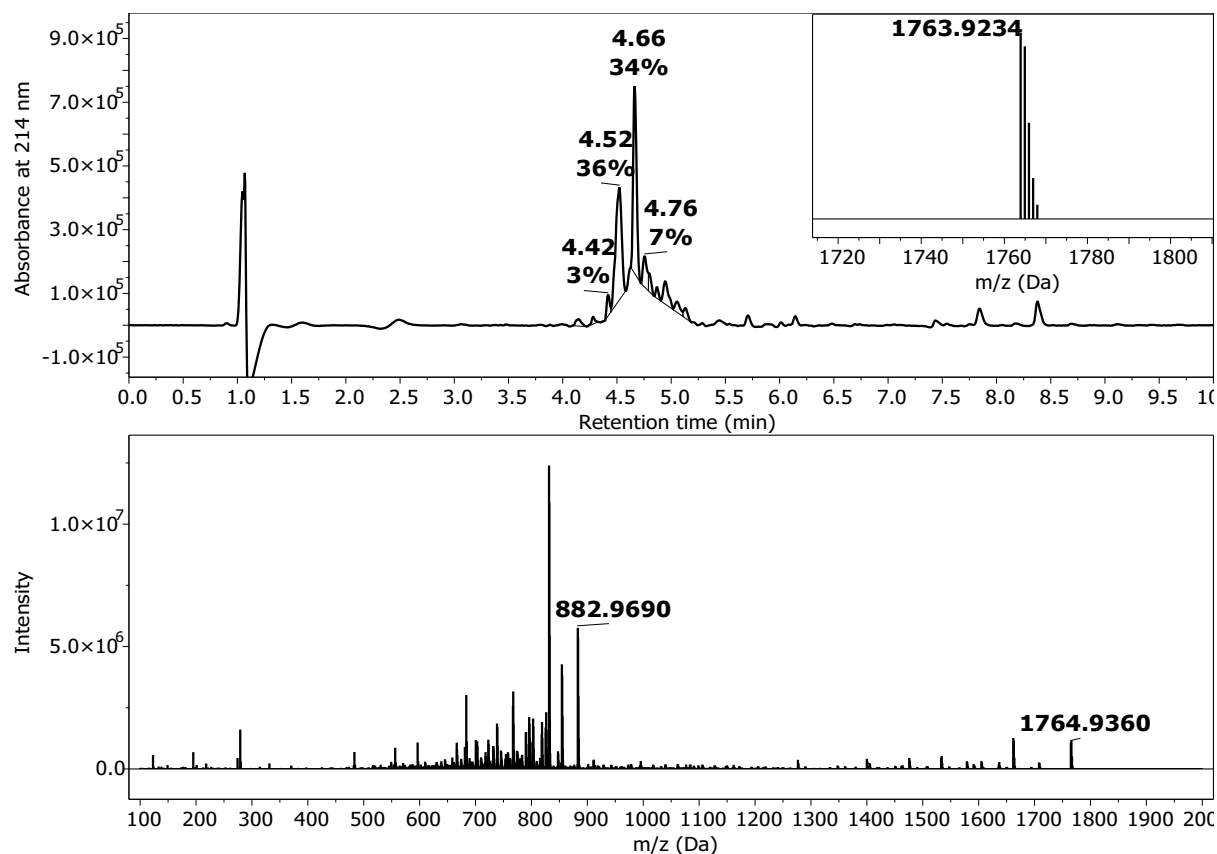
Figure S5: LC–MS profile of crude Barstar[75–90] synthesized with 50 mg of resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9234.

100 mg:

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 100 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 41.0 μmol)

Crude peptide mass: ~66 mg

Crude purity (UHPLC): 35%
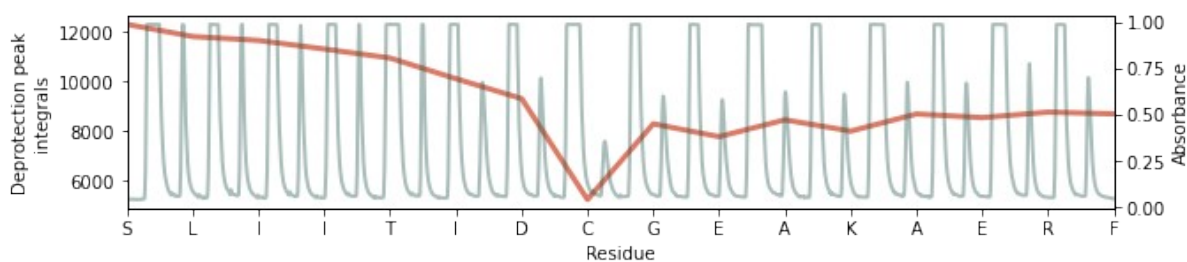
Synthesis time: 45 min.

Figure S6: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized on 100 mg of resin but otherwise under standard AFPS conditions.

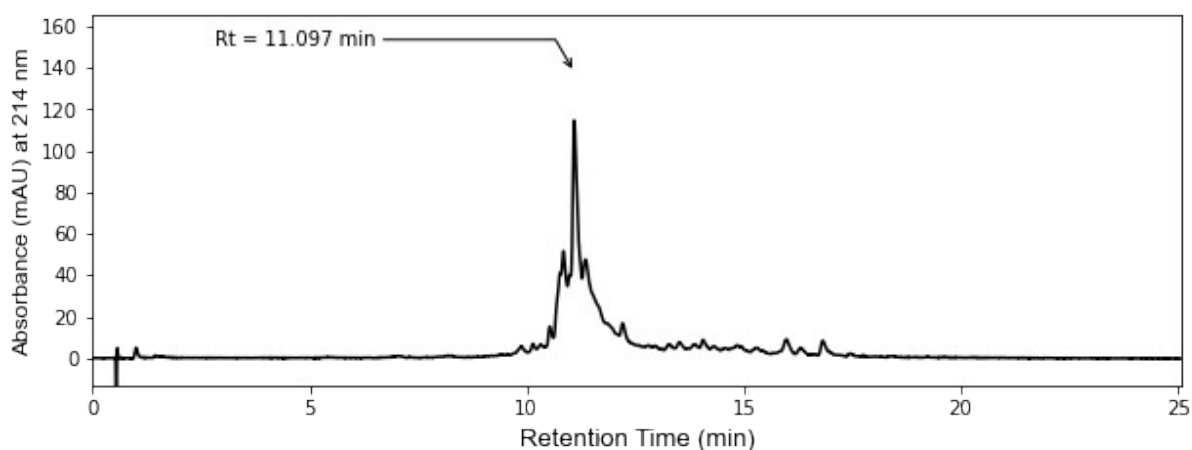## UHPLC analysis of crude Barstar[75–90]

Figure S20: UHPLC of Barstar[75–90] synthesized with 100 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.
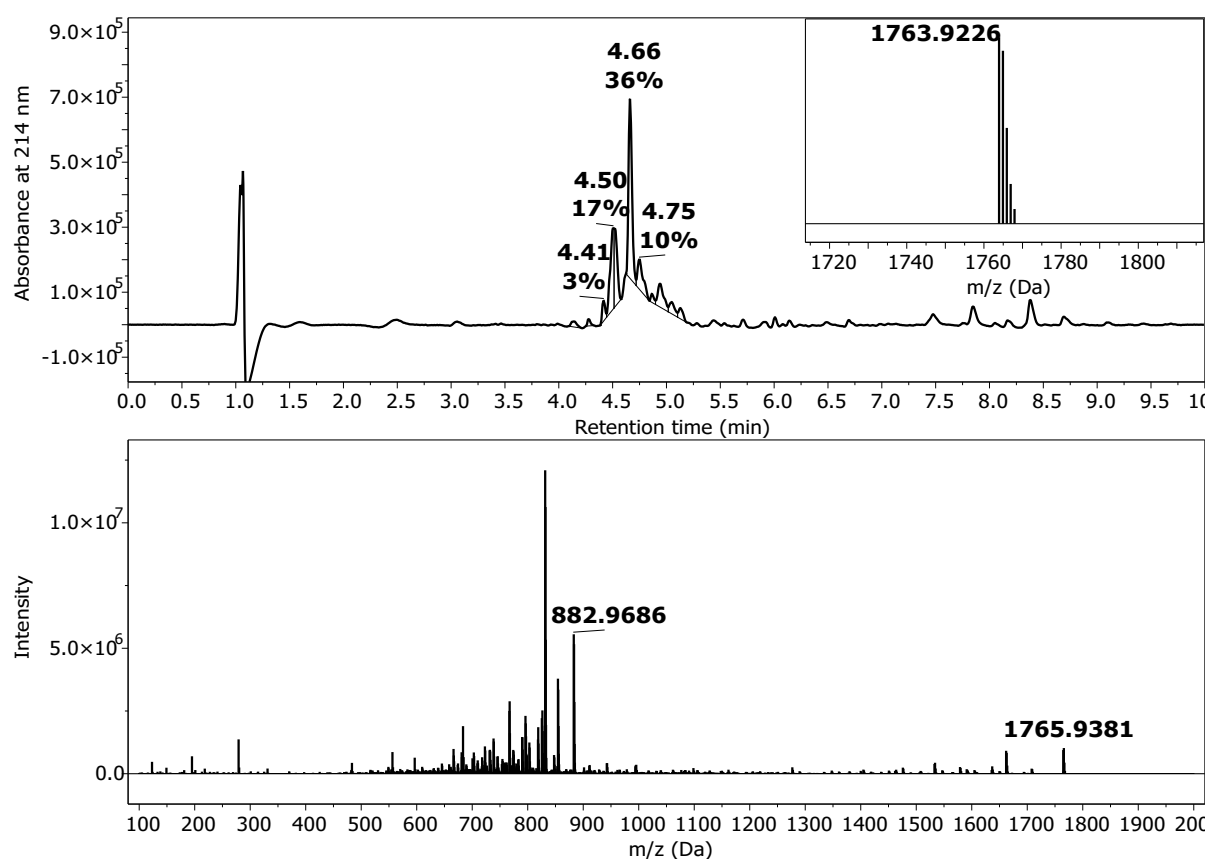
LC–MS analysis of crude Barstar[75–90]

Figure S21: LC–MS profile of crude Barstar[75–90] synthesized with 100 mg of resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9209.

150 mg:

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 μmol)

Crude peptide mass: ~89 mg

Crude purity (UHPLC): 37%

Synthesis time: 45 min.

Figure S22: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized under standard AFPS conditions.

UHPLC analysis of crude Barstar[75–90]

Figure S23: UHPLC of Barstar[75–90] synthesized with 150 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

LC–MS: analysis of crude Barstar[75–90]

Figure S24: LC–MS profile of crude Barstar[75–90] synthesized with 150 mg of resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9229.

200 mg:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 200 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 82.0 µmol)

Crude peptide mass: 125 mg approx.

Crude purity (UHPLC): 35%

Synthesis time: 45 min.

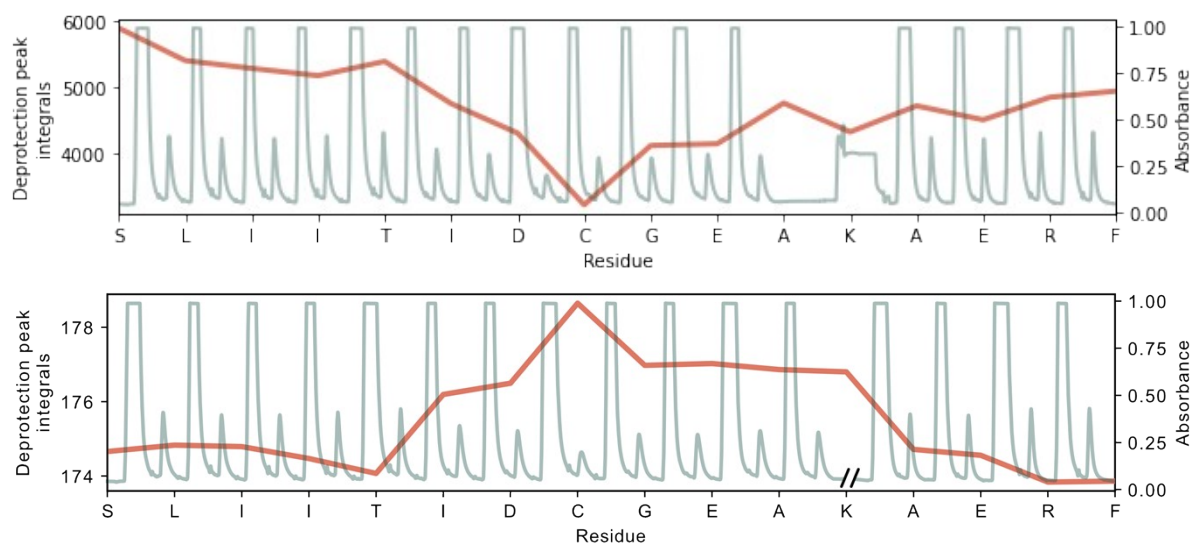**Figure S7:** AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized on 200 mg of resin but otherwise under standard AFPS conditions.

UHPLC analysis of crude Barstar[75–90]

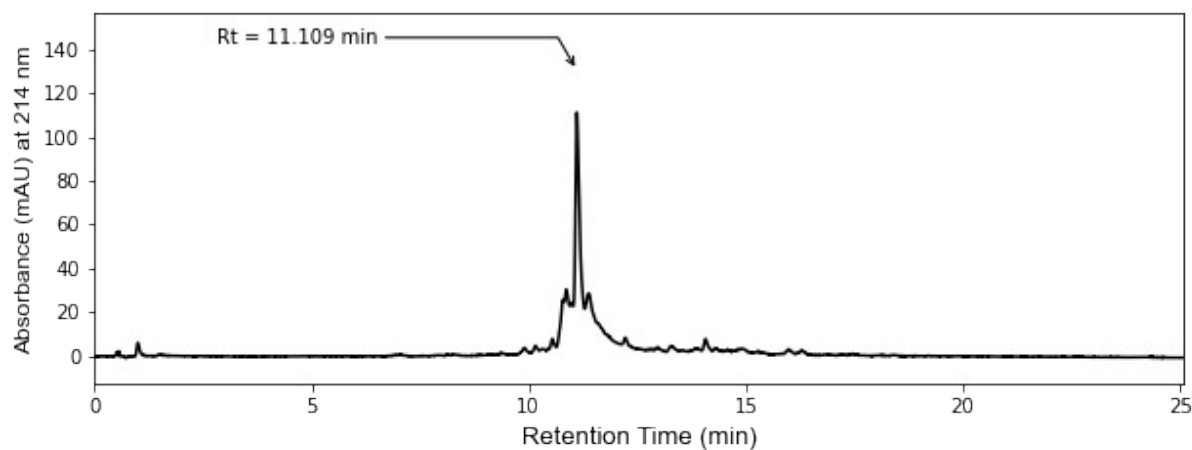**Figure S26:** UHPLC of Barstar[75–90]synthesized with 200 mg of resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.
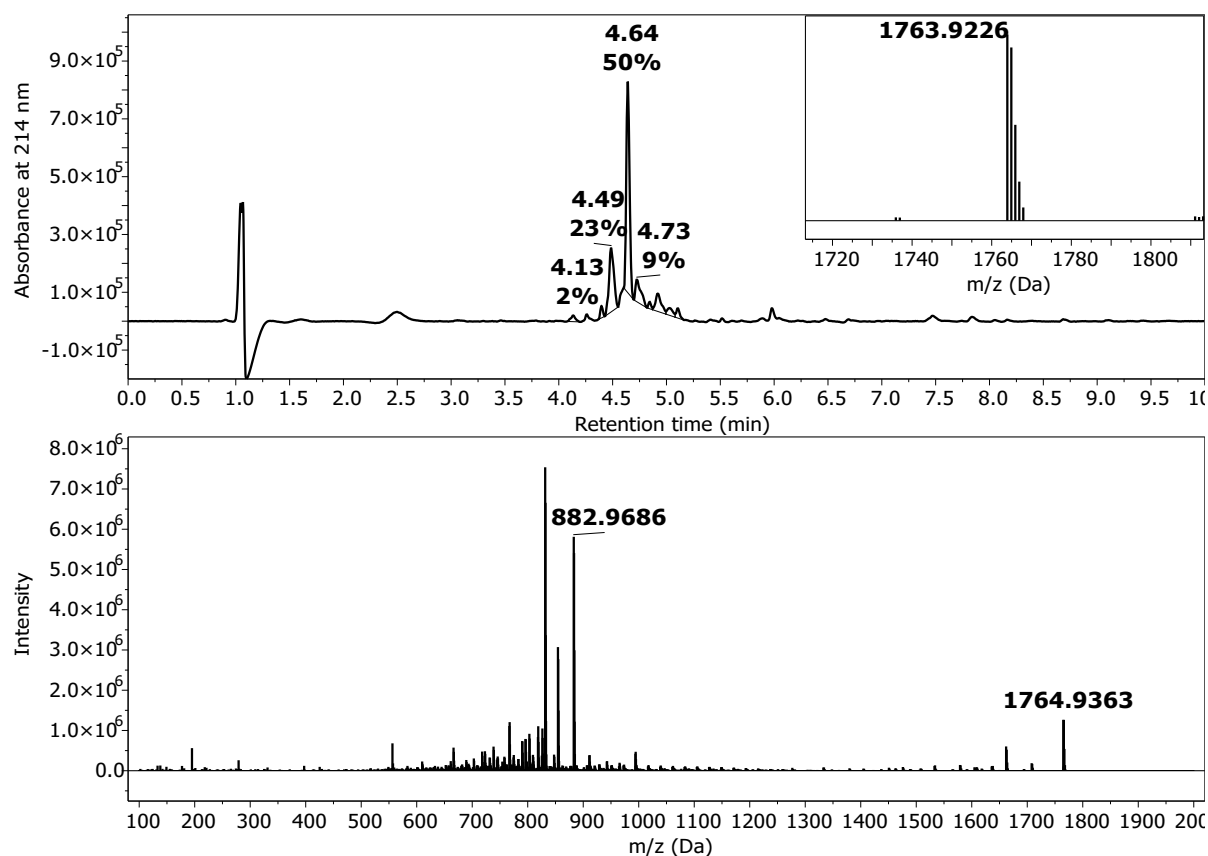
LC–MS analysis of crude Barstar[75–90]

Figure S27: LC–MS profile of crude Barstar[75–90] synthesized with 200 mg of resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9226.

3.2 Resin Loading

3.2.1 Barstar[75–90]

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Synthesis conditions: standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for resin loading and mass. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

Comparison of both data processing methods:
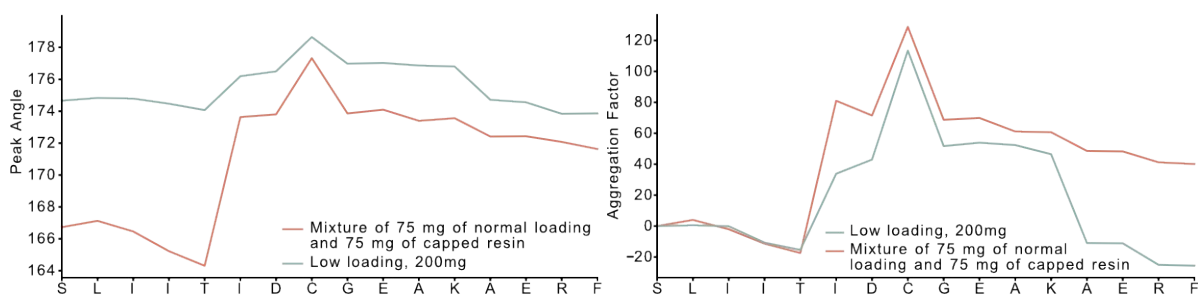
Figure S28: A) Experiments to investigate the effect of loading with Barstar[75–90] analyzed with the "peak angle" data processing method B) Experiments to investigate the effect of loading with Barstar[75–90] analyzed with the "aggregation factor" data processing method.

Low loading 200 mg:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 200 mg of low-loading NovaPEG Rink Amide resin (0.20 mmol/g loading, 40.0 µmol)

Crude peptide mass: ~41 mg
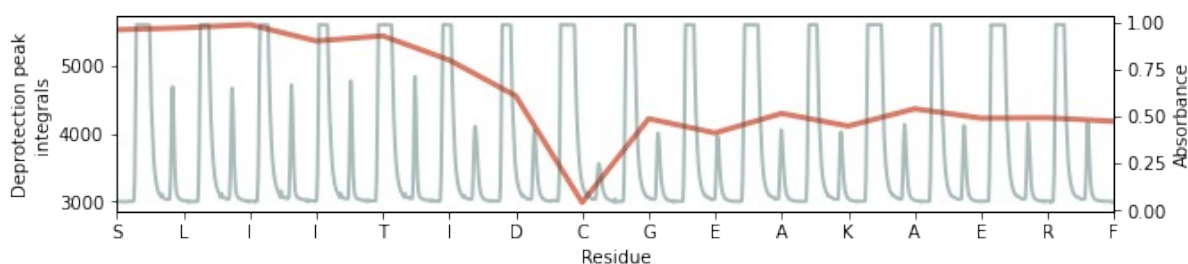
Crude purity (UHPLC): 42%

Synthesis time: 45 min.

Figure S29: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized on 200 mg of low loading resin but otherwise under standard AFPS conditions. The synthesis was paused after completion of the washing step of lys[79] and the syringe exchanged

UHPLC analysis of crude Barstar[75–90]

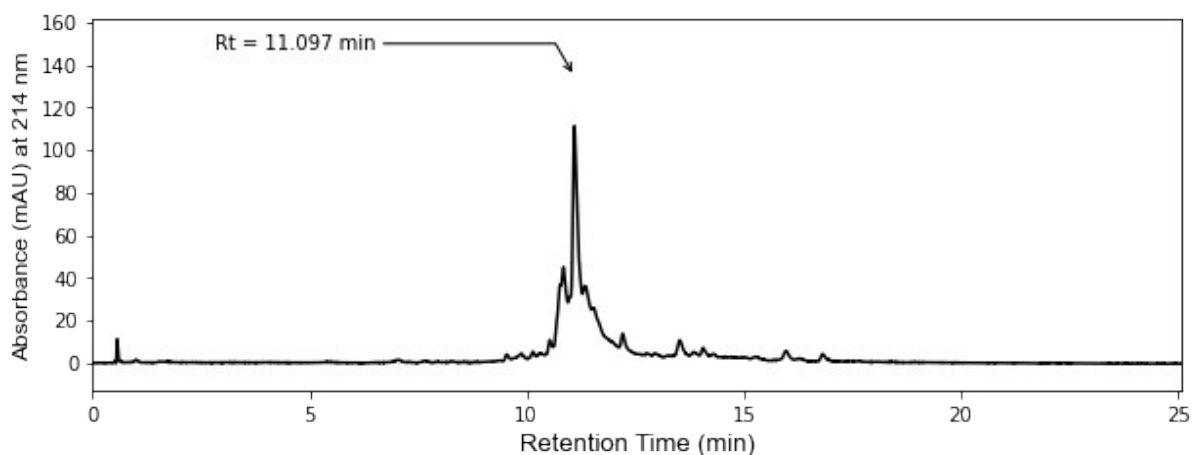**Figure S30:** UHPLC of Barstar[75–90] synthesized with 200 mg of low loading resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

LC–MS analysis of crude Barstar[75–90]

Figure S31: LC–MS profile of crude Barstar[75–90] synthesized with 200 mg of resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9226.

75 mg of normal loading resin mixed with 75 mg of capped resin:

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 75 mg of normal loading NovaPEG Rink Amide resin (0.41 mmol/g loading, 30.75 µmol) and 75 mg of capped NovaPEG Rink Amide resin (0 mmol/g loading, 0 µmol)

Crude peptide mass: ~43 mg

Crode purity (UHPLC): 39%

Synthesis time: 45 min.

Figure S32: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized on 75 mg capped and 75 mg of normal loading resin but otherwise under standard AFPS conditions.

UHPLC analysis of crude Barstar[75–90]

Figure S33: UHPLC of Barstar[75–90] synthesized with 75 mg capped and 75 mg of normal loading resin but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

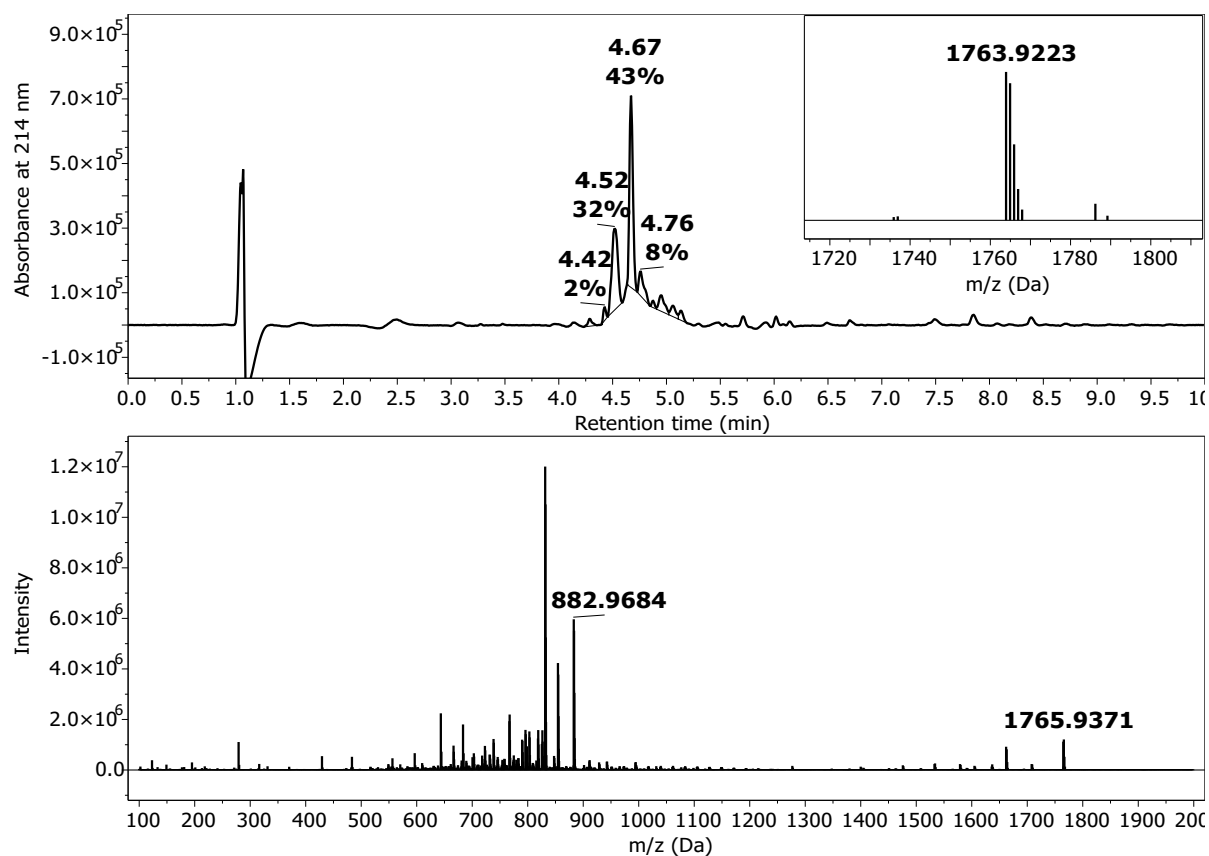## LC–MS analysis of crude Barstar[75–90]

Figure S34: LC–MS profile of crude Barstar[75–90] synthesized with 75 mg capped and 75 mg of normal loading resin but otherwise under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9223.

3.2.2 Capping experiment

200 mg of pre-functionalized NovaPEG Rink Amide resin (0.41 mmol/g loading) was used in all experiments. The amino resin (1 eq., 82 µmol.) was swelled with DMF (1 × 5 mL) for 1 min., then the solvent was removed by filtration under reduced pressure. For the capping acetic anhydride (50 eq., 387 µL) and pyridine (50eq., 330 µL) were mixed in 3 ml of DMF and added to the resin. The solution was then added to the resin, and the reaction was gently stirred for 20 sec., then left to react at 23 °C overnight to ensure high coupling efficiency. Afterwards, the resin was drained, washed with DMF (3 × 5 mL) and DCM (3 × 5 mL).

3.3 Temperature

3.3.1 NBDY [53–68]

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Synthesis conditions: Standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for activation loop temperature. The activation loop was set to the designated temperature (only histidine was coupled at room temperature to prevent racemization). Standard cleavage (**section 2.4**), LC–MS (**section 2.6**), and UHPLC (**section 2.5**) procedures were used.

Comparison of both data processing methods:
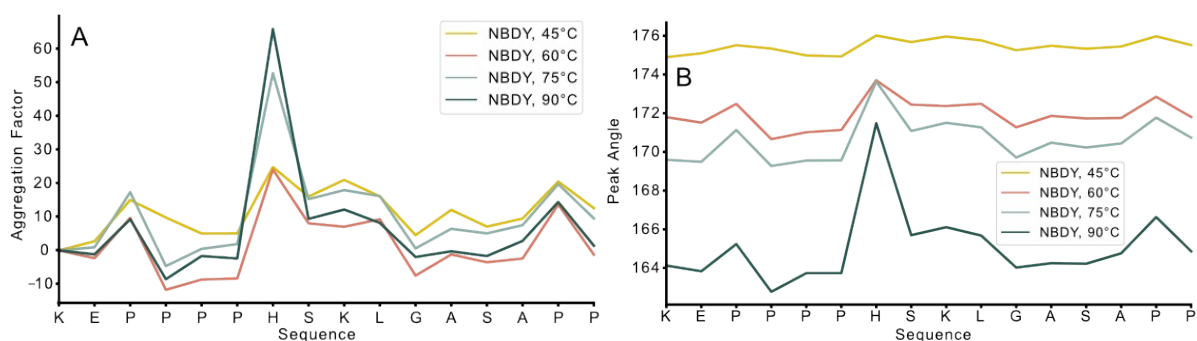
Figure S35: **A)** Temperature screening of NBDY analyzed with the peak angle **B)** Temperature screening of NBDY analyzed with the aggregation factor

90 °C:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 µmol)

Crude peptide mass: ~119 mg

Crude purity (UHPLC): 96%

Synthesis time: 45 min.

See section 3.1.1 for AFPS, LC–MS and UHPLC trace.

75 °C:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 µmol)

Crude peptide mass: ~122 mg

Crude purity (UHPLC): 96%
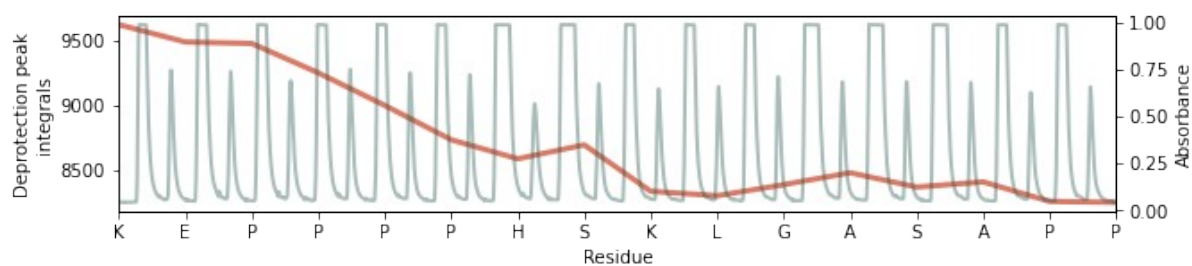
Synthesis time: 45 min.

Figure S36: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized at 75 °C but otherwise under standard AFPS conditions.
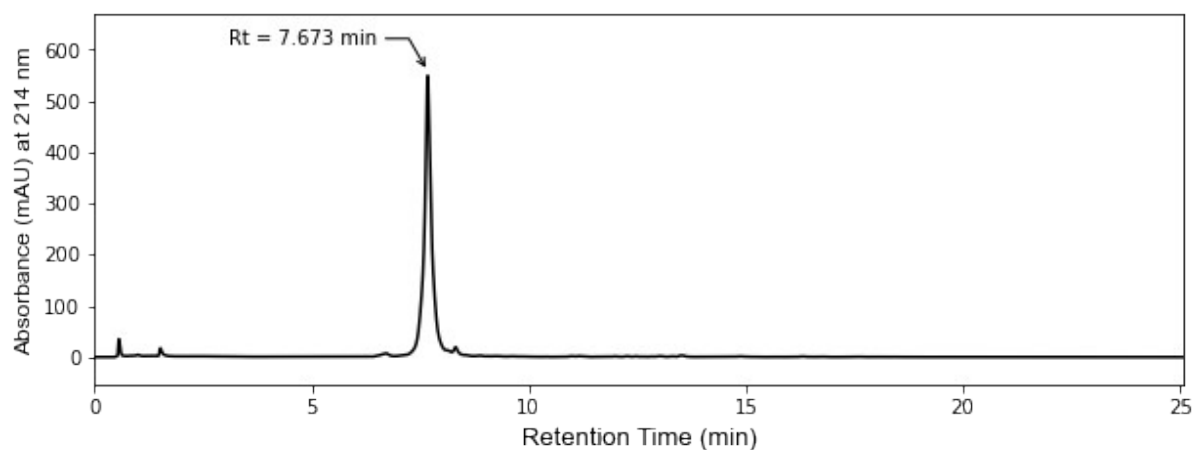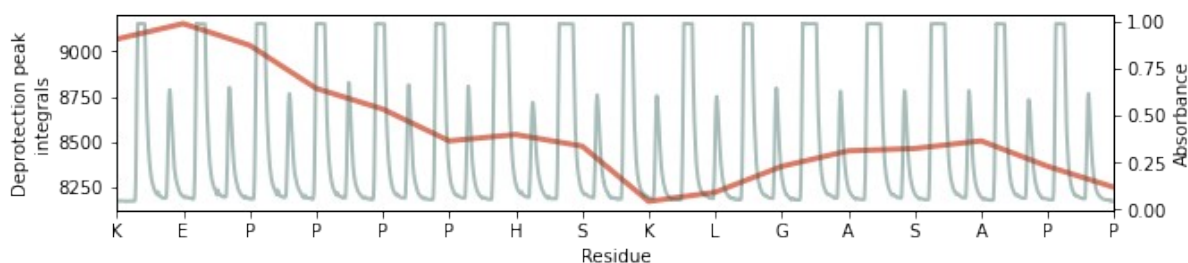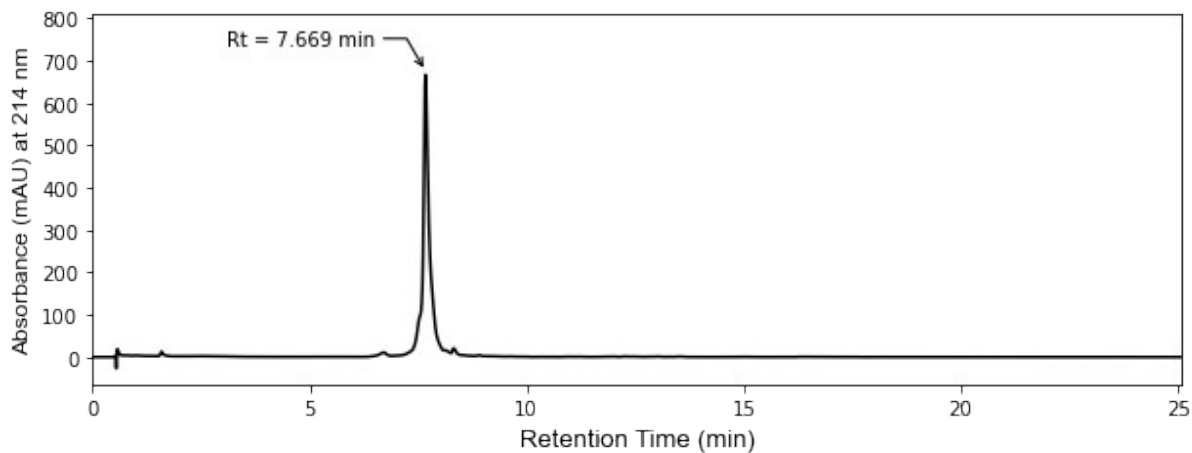
UHPLC analysis of crude NBDY[53–68]

Figure S37: UHPLC of NBDY[53–68] synthesized at 75 °C but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.
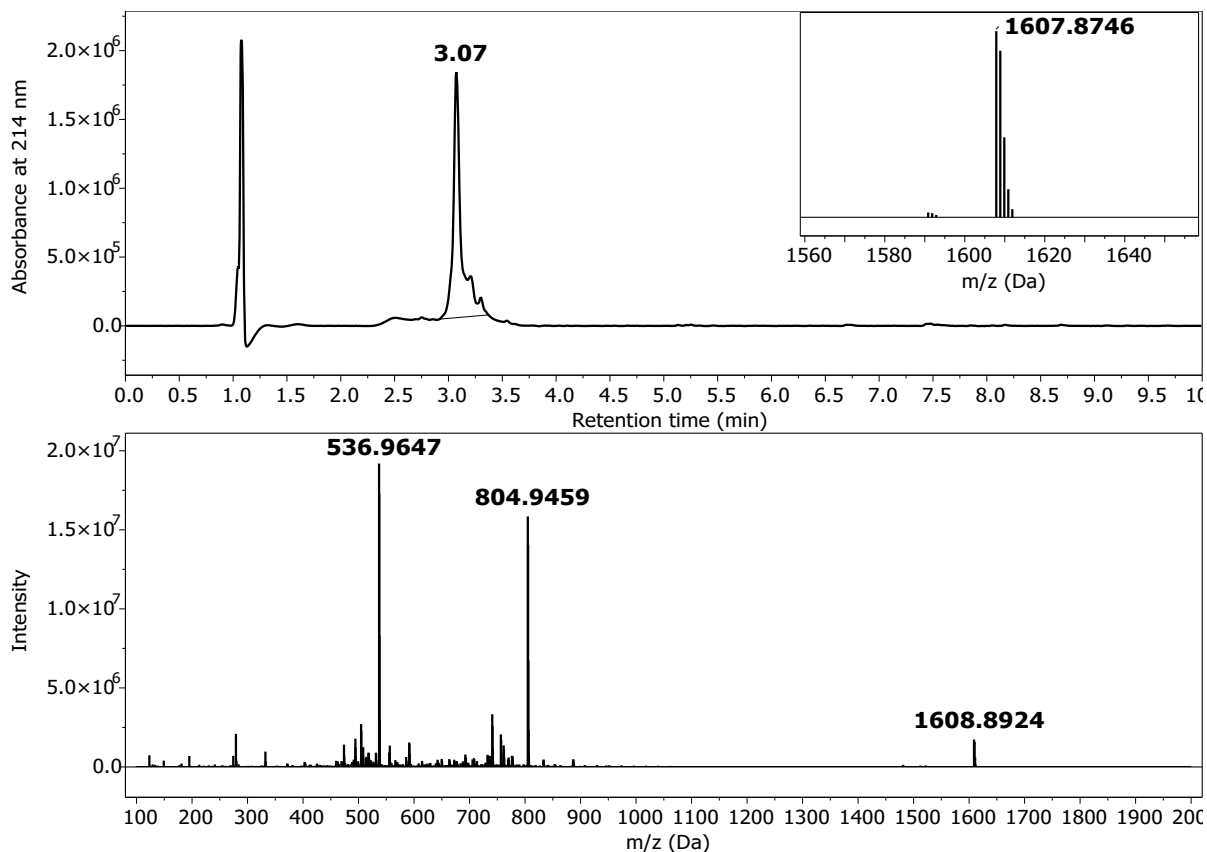
LC–MS analysis of crude NBDY[53–68]
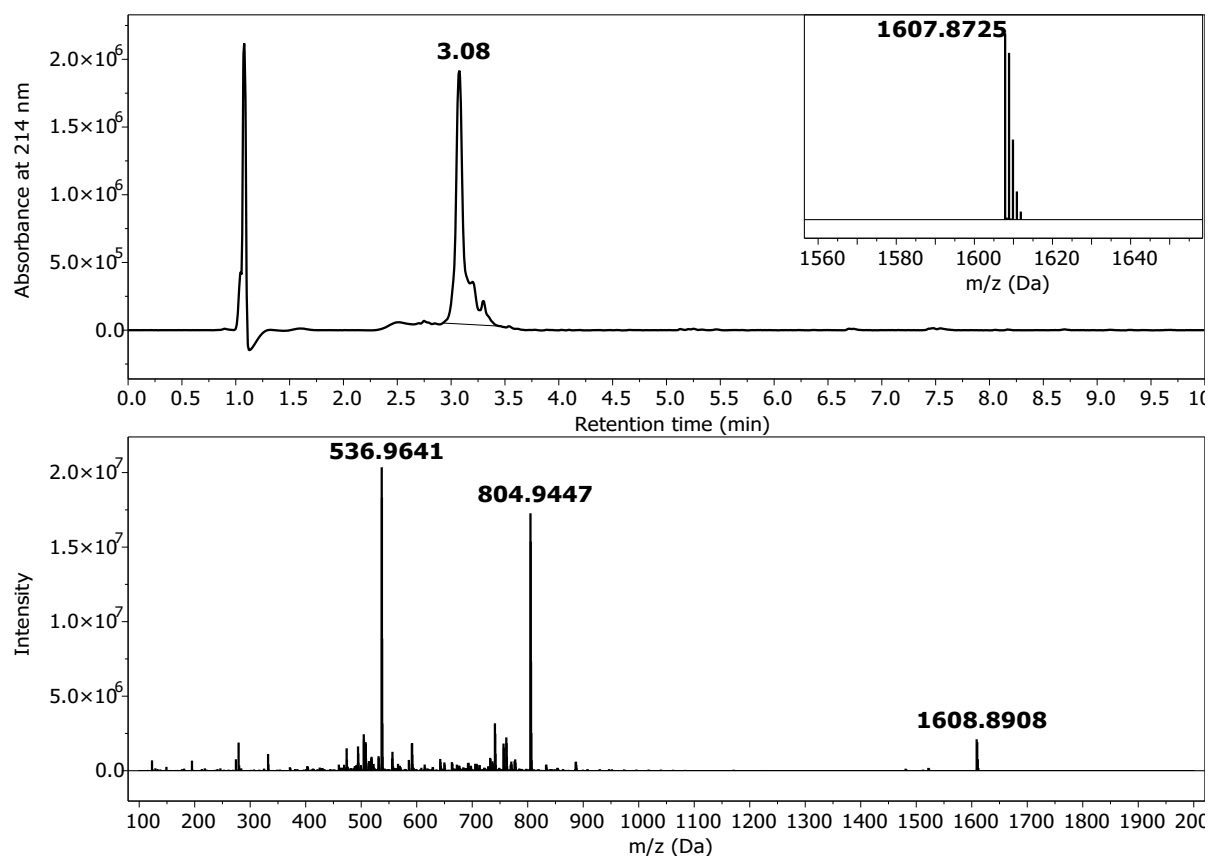
Figure S38: LC–MS profile of crude NBDY[53–68] synthesized at 75 °C. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8725.

60 °C:

Sequence: PPASAGLKSH PPPPEK-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 μmol)

Crude peptide mass: ~113 mg

Crude purity (UHPLC): 96%

Synthesis time: 45 min.

Figure S39: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY[53–68] synthesized at 60 °C but otherwise under standard AFPS conditions.

UHPLC analysis of crude NBDY[53–68]

Figure S40: UHPLC of NBDY[53–68] synthesized at 60 °C but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

LC-MS analysis of crude NBDY[53–68]

Figure S41: LC–MS profile of crude NBDY[53–68] synthesized at 60 °C. Monoisotopic mass (ESI+) calcd. for $C_{73}H_{117}N_{21}O_{20}$ 1607.8784, found 1607.8746.

29

3.3.2 Barstar[75–90]

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Synthesis conditions: Standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for activation loop temperature. Activation loop was set to the designated temperature (only histidine was coupled under standard conditions to prevent racemization). Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

Comparison of both data processing methods:
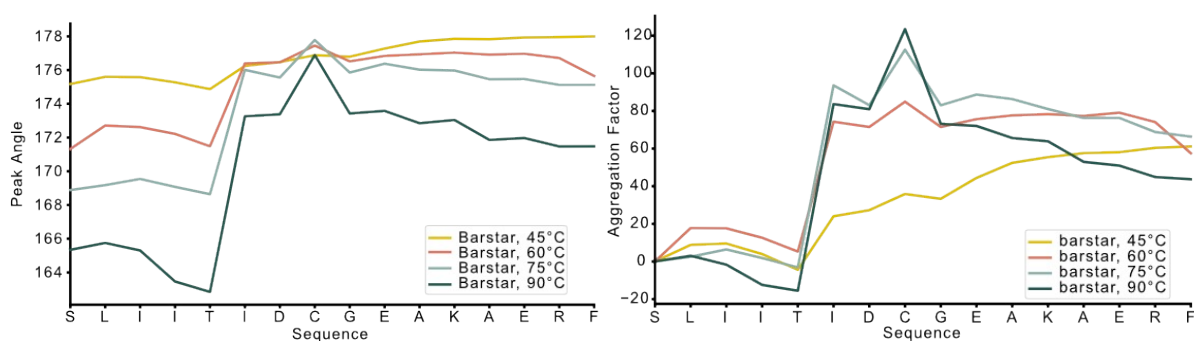
Figure S2: A) Temperature screening of Barstar[75–90] analyzed with the peak angle. B) Temperature screening of Barstar[75–90] analyzed with the aggregation factor.

90 °C:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 µmol)

Crude peptide mass: ~89 mg

Crude purity (UHPLC): 37%

Synthesis time: 45 min.

See section 3.1.2 for AFPS, LC–MS and UHPLC trace.

75 °C:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 µmol)

Crude peptide mass: ~82 mg

Crude purity (UHPLC): 25%
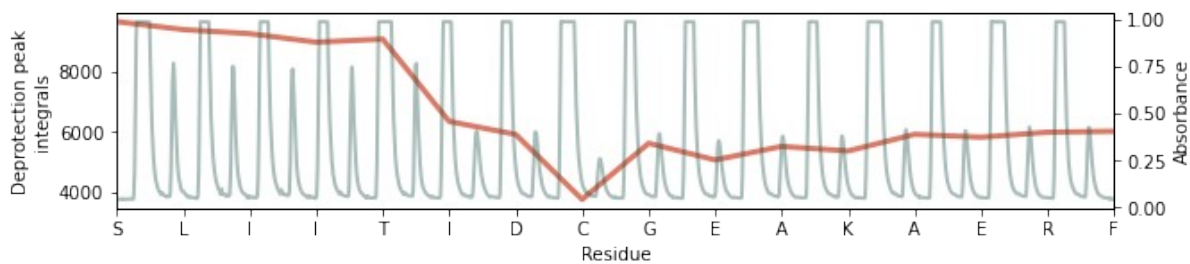
Synthesis time: 45 min.

Figure S43: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized at 75 °C but otherwise under standard AFPS conditions.
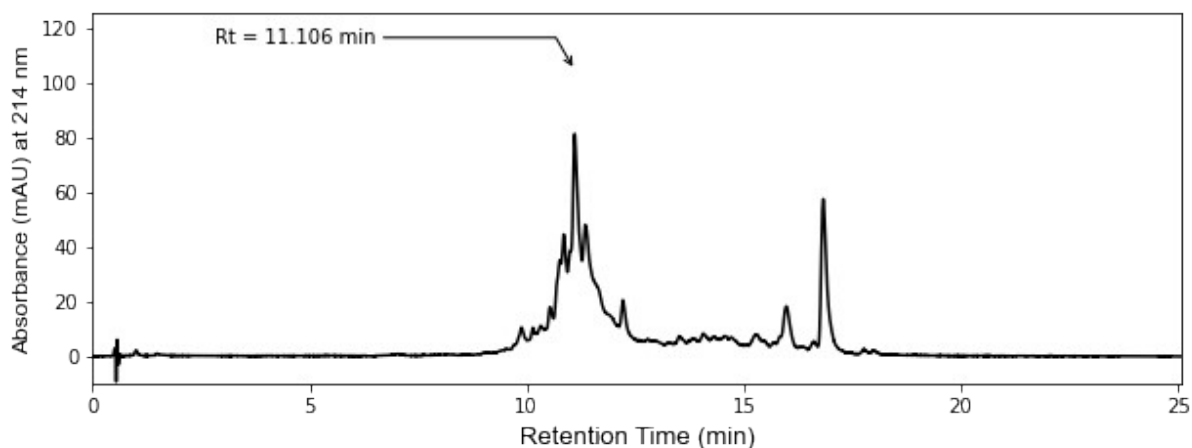
UHPLC analysis of crude Barstar[75–90]

Figure S44: UHPLC of Barstar[75–90] synthesized at 75 °C but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

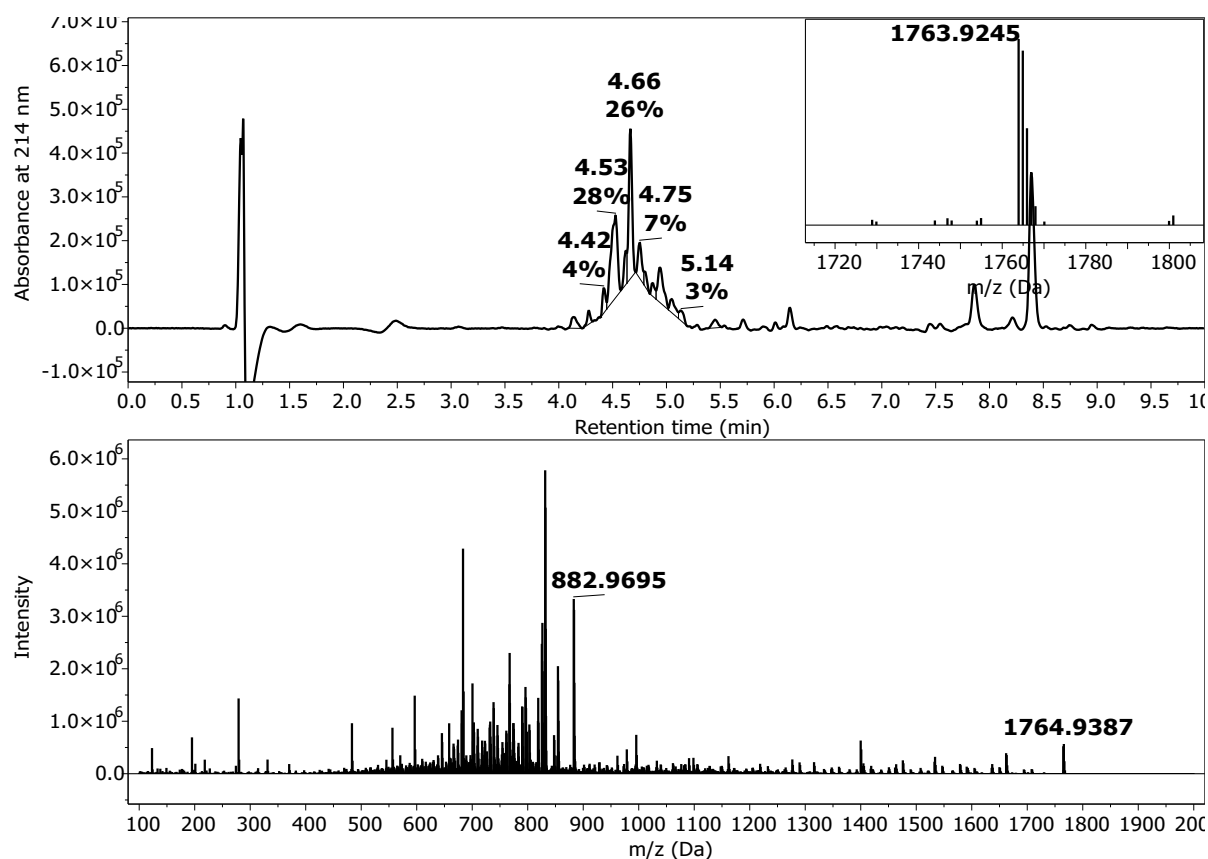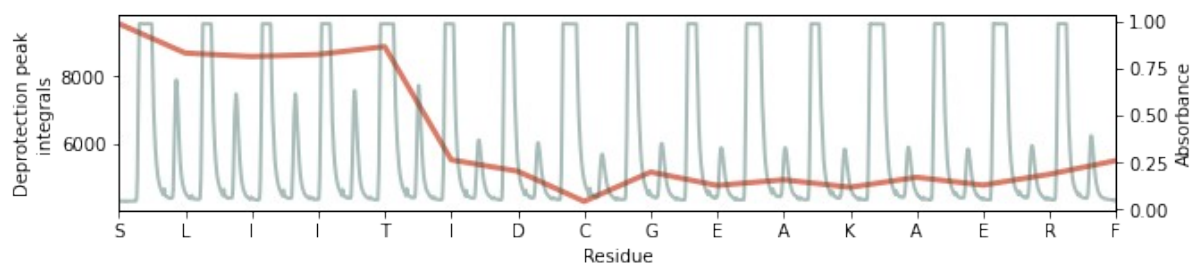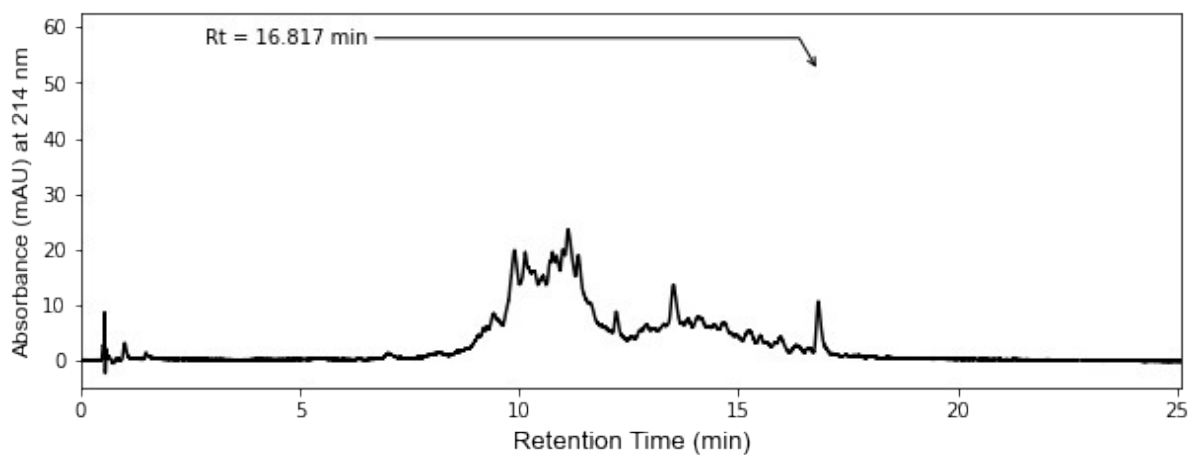## LC–MS analysis of crude Barstar[75–90]

Figure S45: LC–MS profile of crude Barstar[75-90] synthesized at 75 °C. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9245.

60 °C:

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 61.5 μmol)

Crude peptide mass: ~79 mg

Crude purity (UHPLC): not possible to determine.

Synthesis time: 45 min.

Figure S46: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized at 60 °C but otherwise under standard AFPS conditions.

32

UHPLC analysis of crude Barstar[75–90]

Figure S47: UHPLC of Barstar[75–90] synthesized at 60 °C but otherwise under standard AFPS conditions. The standard UHPLC procedure was used.

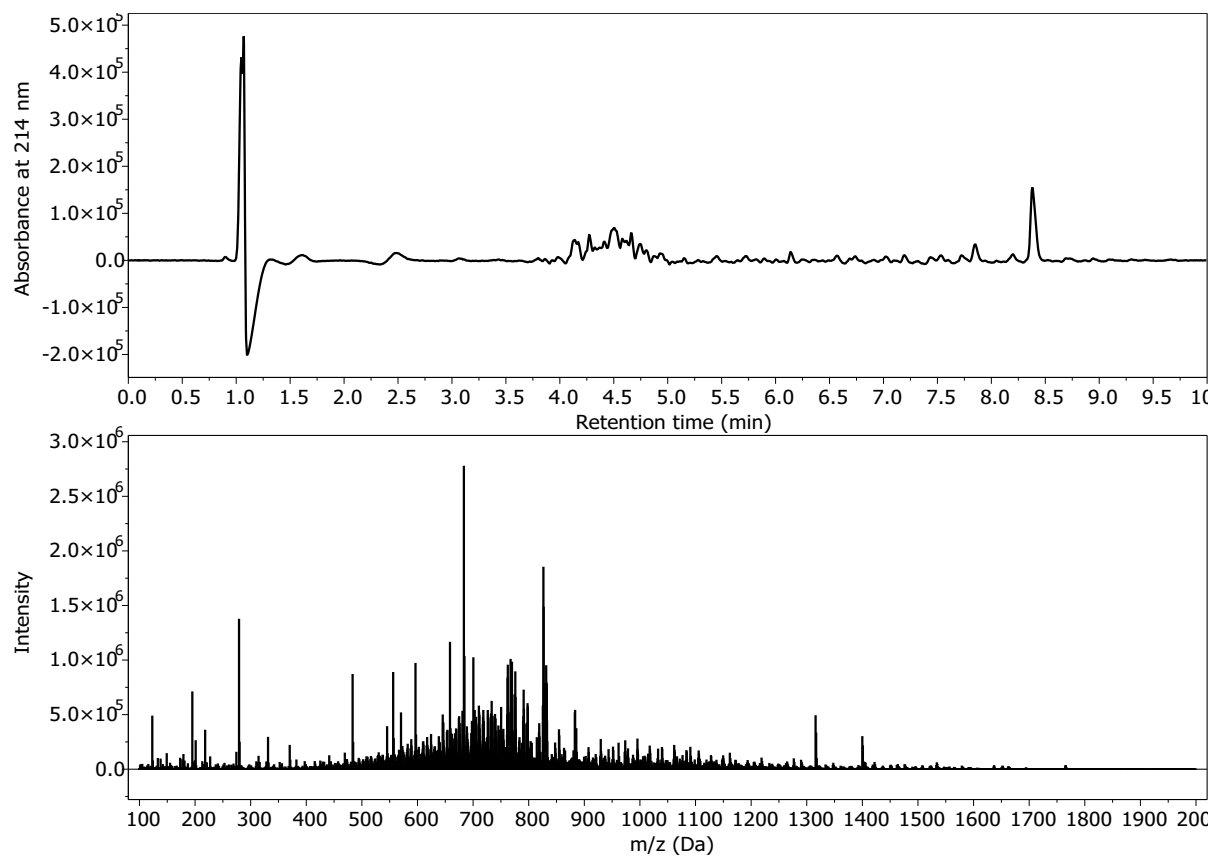## LC–MS analysis of crude Barstar[75–90]

Figure S48: LC–MS profile of crude Barstar[75-90] synthesized at 60 °C. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92 not found.

3.4   Batch-SPPS vs. flow-SPPS (AFPS)

3.4.1   NBDY – batch-SPPS

Sequence: MGDQPCASGR STLPPGNARE AKPPKKRCLL APRWDYPEGT PNGGSTTLPS APPPASAGLK SHPPPPEK-CONH$_2$ (68 amino acids)

Conditions: standard batch conditions (**section 2.3**, 20 mL/min flowrate) except for resin mass. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used. The peptides was synthesized starting on 400 mg of NovaPEG Rink Amide resin (0.20 mmol/g loading). Upon every 10 amino acid coupling, a portion of the resin was saved (approx. 25%) and the synthesis was carried with the rest of the resin.

Resin: Approx. 64 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading)

Crude peptide mass: ~9 mg

Crude purity (UHPLC): 66%

Synthesis time: approximately 68 hours

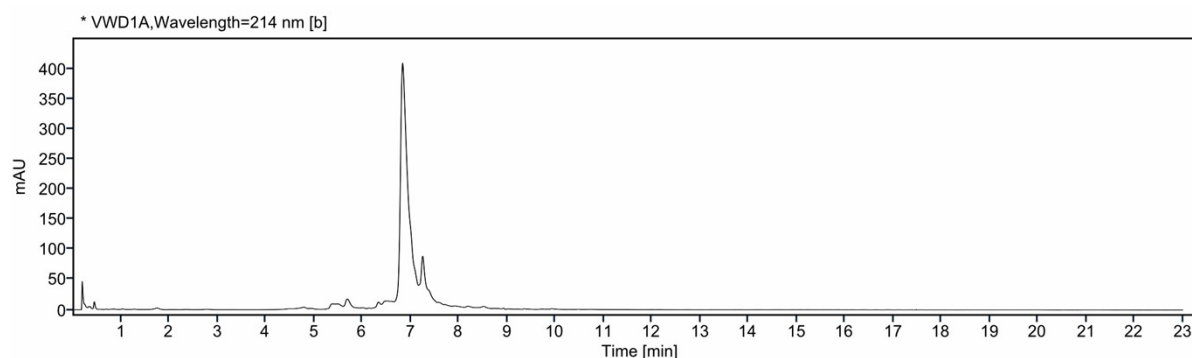**UHPLC analysis of crude NBDY**

Figure S55: UHPLC of NBDY synthesized under standard batch-SPPS conditions. RT: 6.826 min. The standard UHPLC method was used on an Agilent Poroshell 300SB-C8 2.1x75mm 5um column.
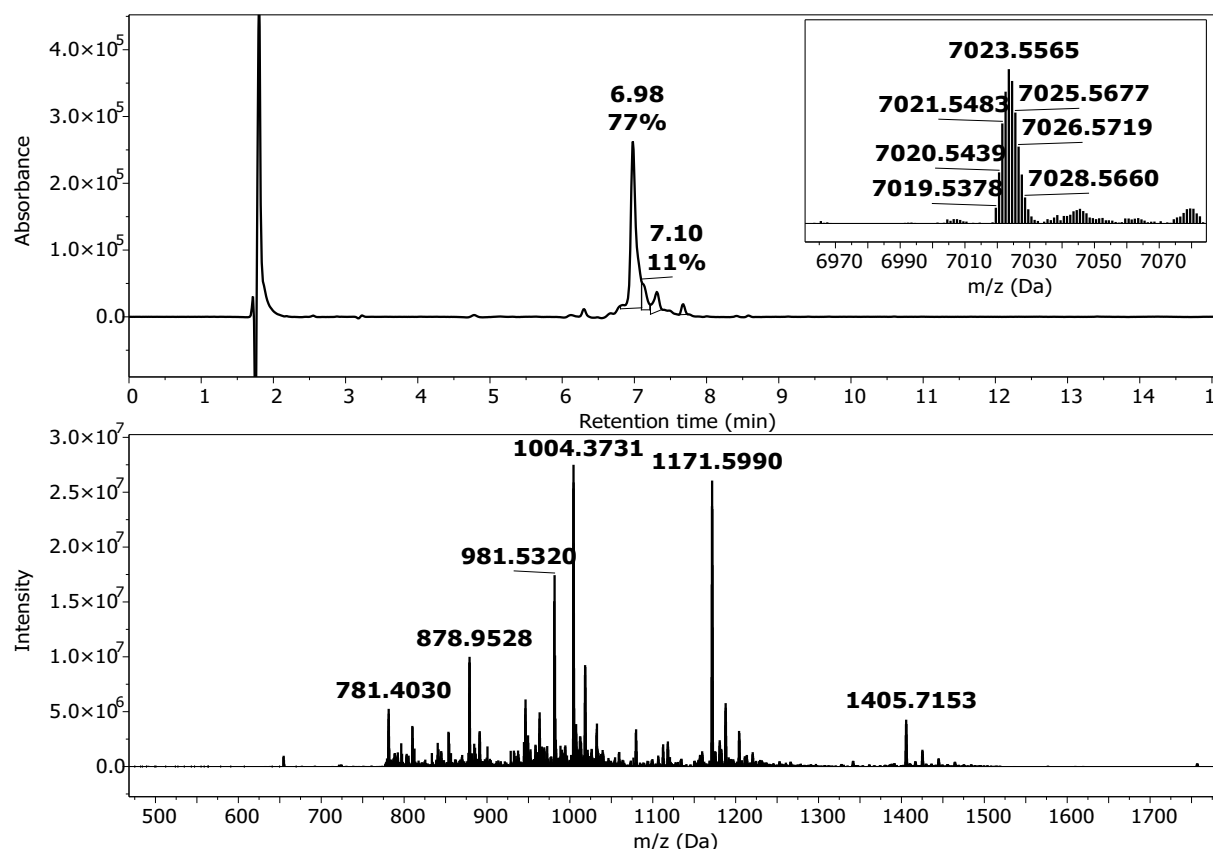
LC–MS analysis of crude NBDY

Figure S56: LC–MS profile of crude NBDY synthesized under standard batch-SPPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 7019.5498, found 7019.5378. Sample was injected into an ACQUITY UPLC@BioResolve-RP-mAb 2.7 um 2.1 mm x150 mm, 450 A(Waters, USA) column. For desalting resp. separation on an Acquity UPLC station, a gradient buffer A (0.1% DFA in water)/ buffer B (0.1% DFA in AN/75% 2-PrOH) at a flow rate 200 ul/min at 60°C over 30min was applied. The analysis was performed on a Synapt G2-Si mass spectrometer directly coupled to the UPLC station.Mass spectra were acquired in the positive-ion mode by scanning the m/z range from 400 to 5000 Da with a scan duration of 1 s and an interscan delay of 0.1s. The spray voltage was set to 3 kV, the cone voltage to 50 V, and the source temperature to 100°C. The data were recorded with the MassLynx 4.2 Software (both Waters, UK).

3.4.2   NBDY – AFPS

Sequence: MGDQPCASGR STLPPGNARE AKPPKKRCLL APRWDYPEGT PNGGSTTLPS APPPASAGLK SHPPPPEK- $CONH_2$ (68 amino acids)

Synthesis conditions: standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for resin mass. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

Resin: 149 mg of NovaPEG Rink Amide resin (0.20 mmol/g loading)

Crude peptide mass: mass was not measured

Crude purity (UHPLC): 63%
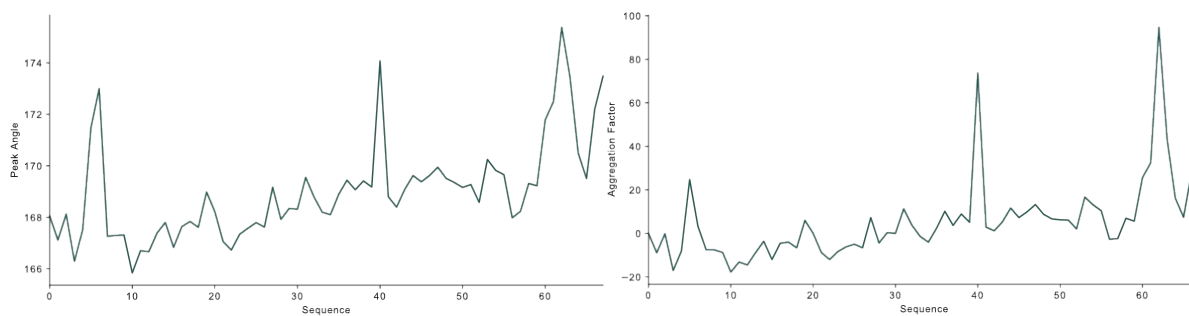
Synthesis time: 3 hours 12 minutes

Figure S57: A) NBDY synthesis analyzed with the peak angle B) NBDY synthesis analyzed with the aggregation factor
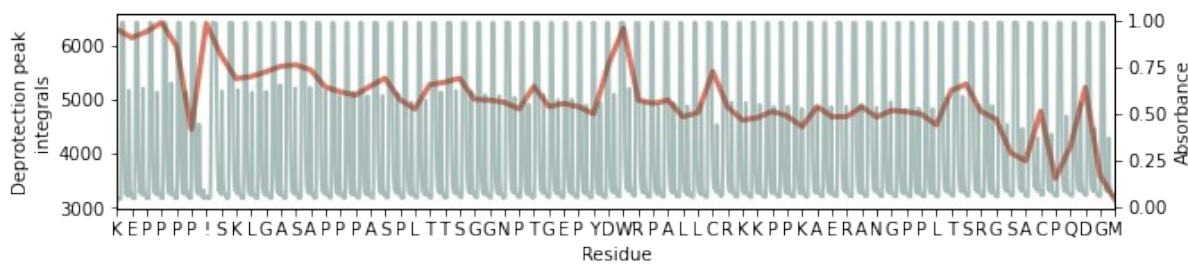

Figure S58: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of NBDY synthesized at under standard AFPS conditions, except for cysteine which was also coupled on 90 °C.
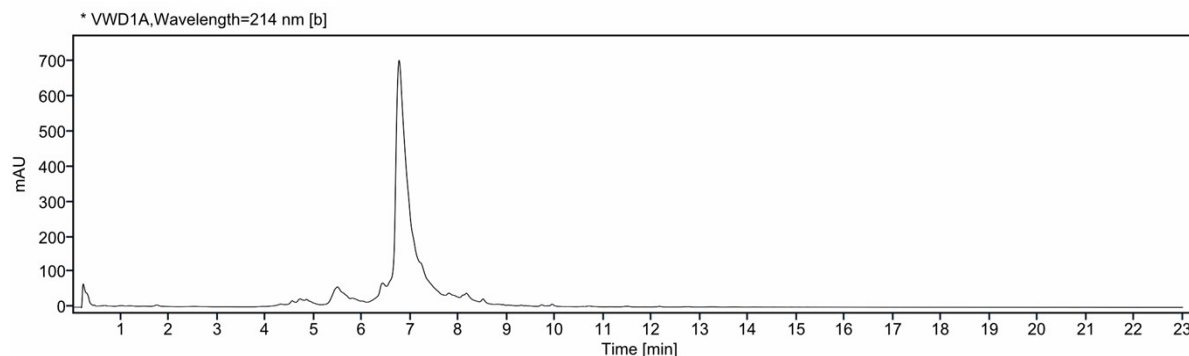
UHPLC analysis of crude NBDY

Figure S59: UHPLC of NBDY synthesized under standard AFPS conditions. RT: 6.756 min. The standard UHPLC method was used on an Agilent Poroshell 300SB-C8 2.1x75mm 5um column.
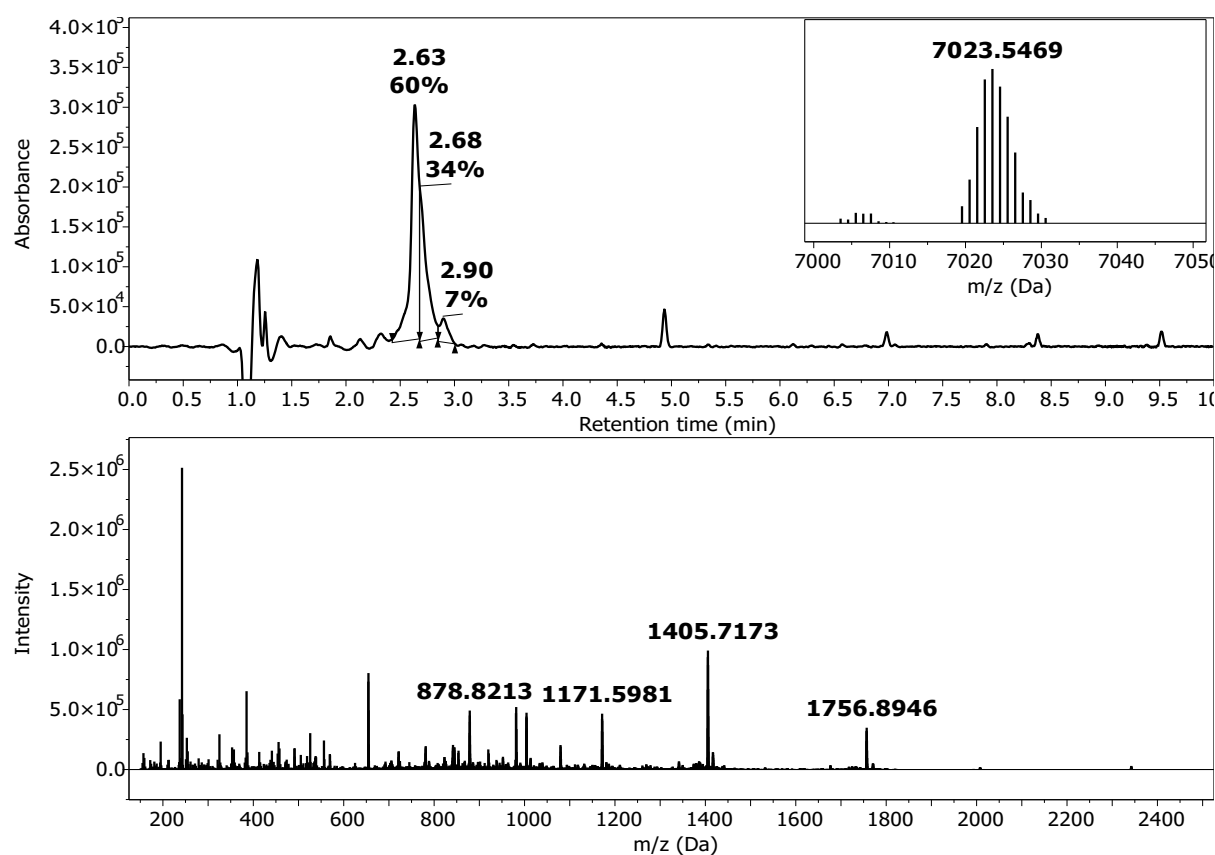
LC–MS analysis of crude NBDY:

Figure S60: LC–MS profile of crude NBDY synthesized under standard AFPS conditions. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 7019.5498, found 7019.5387.

3.5   Washing solvent (DMF) reduction

3.5.1   Barstar[75–90]

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Synthesis conditions: standard AFPS conditions (**section 2.2**, 20 mL/min flowrate) except for washing volumes. The change in washing volume is effected by changing the washing strokes which in turn change the total synthesis time. Standard cleavage (**section 2.4**), LC–MS (**section 2.6**) and UHPLC (**section 2.5**) procedures were used.

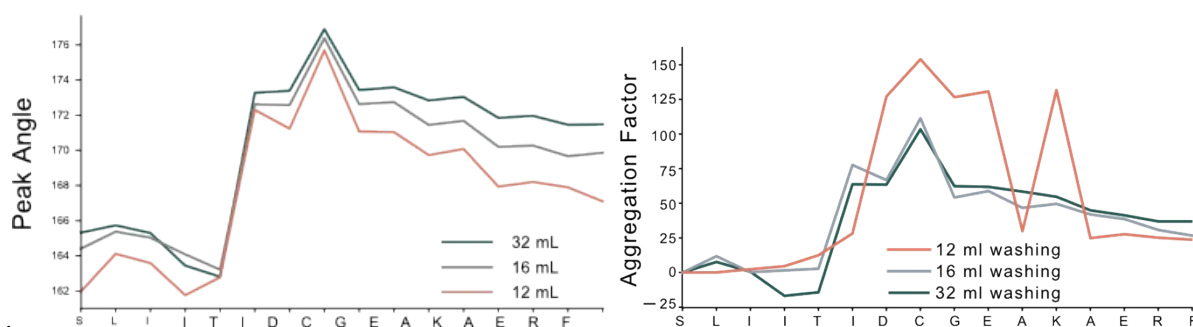## Comparison of both data processing methods:

Figure S61: A) Washing volume reduction screening of Barstar analyzed with the peak angle B) Washing volume reduction screening of Barstar analyzed with the aggregation factor

Synthesis UV trace from the AFPS:

32 mL/coupling cycle:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~89 mg approx.

Crude purity (UHPLC): 35%

Synthesis time: 45 min.

See section 3.1.2 for AFPS, LC–MS and UHPLC trace.

16 mL/coupling cycle:

Sequence: `FREAKAEGCD ITIILS-CONH`$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~98 mg

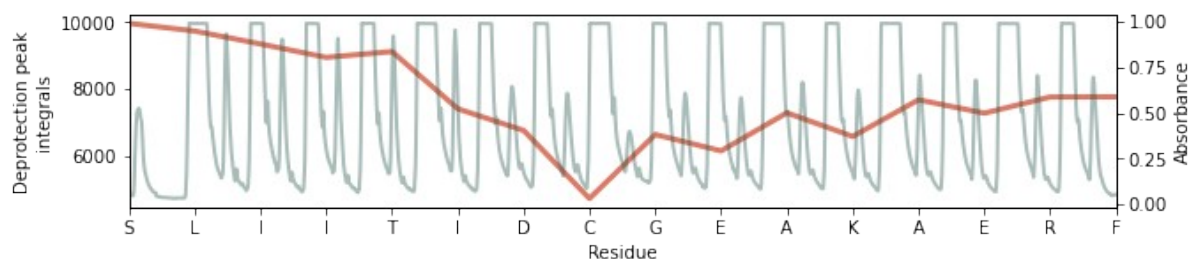**Crude purity (UHPLC)**: 36%

Synthesis time: 32 min.

Figure S62: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized under standard AFPS conditions using reduced washing volumes of 16 mL/cycle.

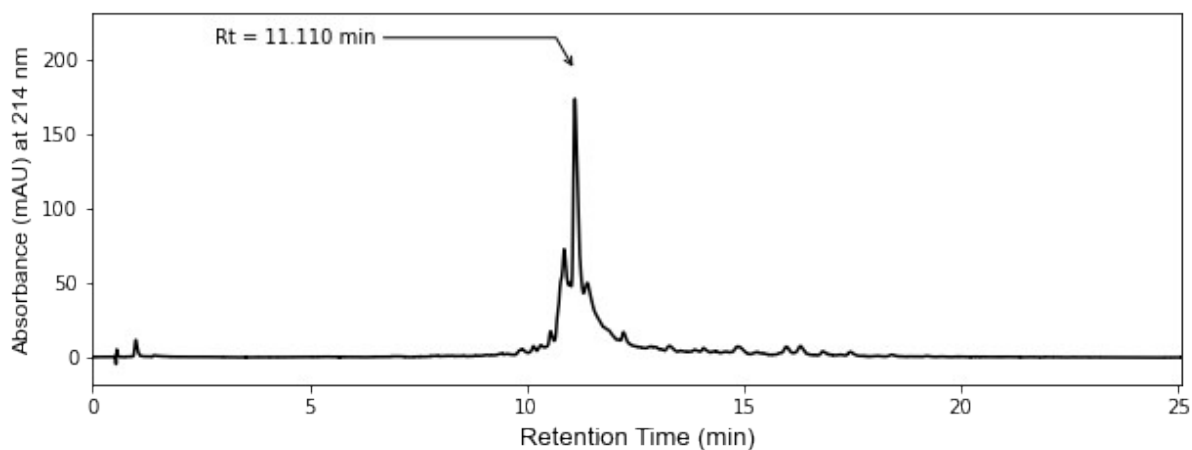## UHPLC analysis of crude Barstar[75–90]

Figure S63: UHPLC of Barstar[75–90] synthesized under standard batch-SPPS conditions using reduced washing volumes of 16 mL/cycle. The standard UHPLC procedure was used.

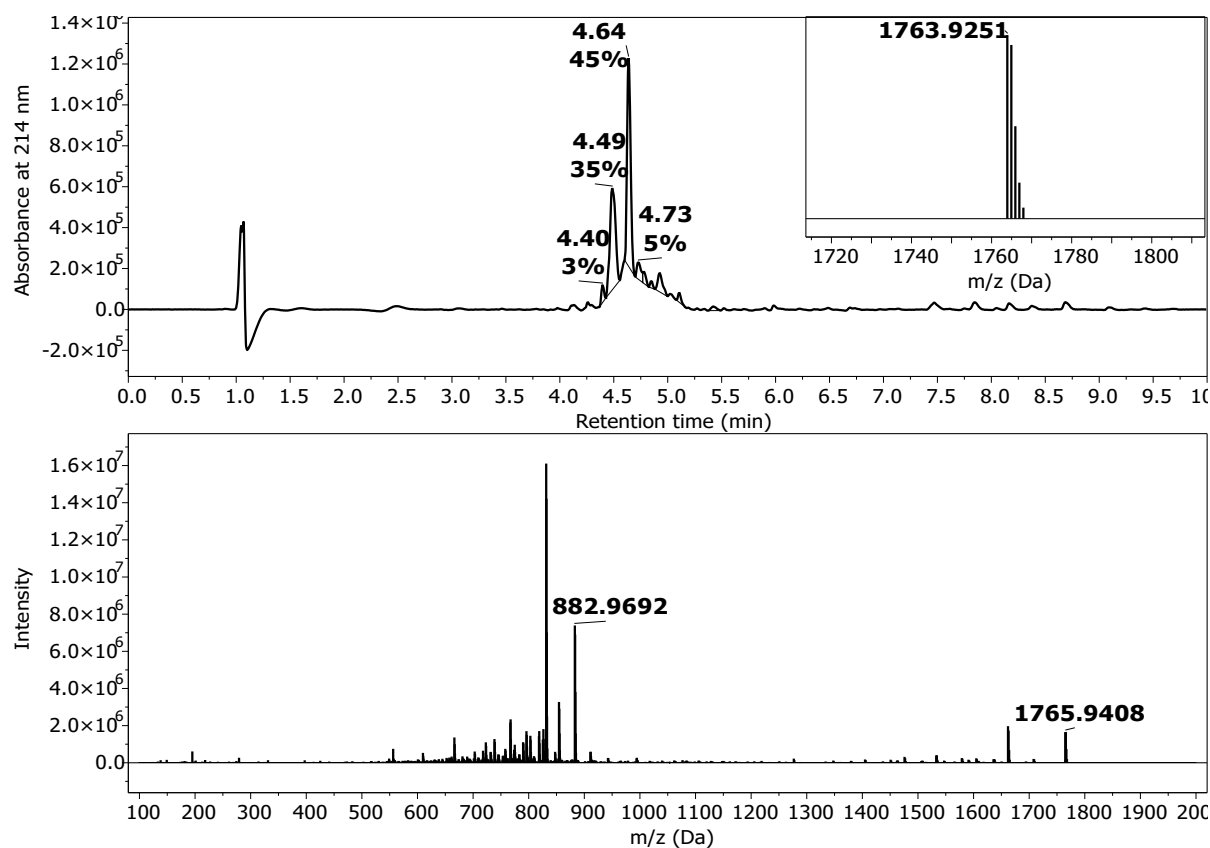## LC–MS analysis of crude Barstar[75–90]

Figure S64: LC–MS profile of crude Barstar[75-90] synthesized under standard batch-SPPS conditions using reduced washing volumes of 16 mL/cycle. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9251.

12 mL/coupling cycle:

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~71 mg

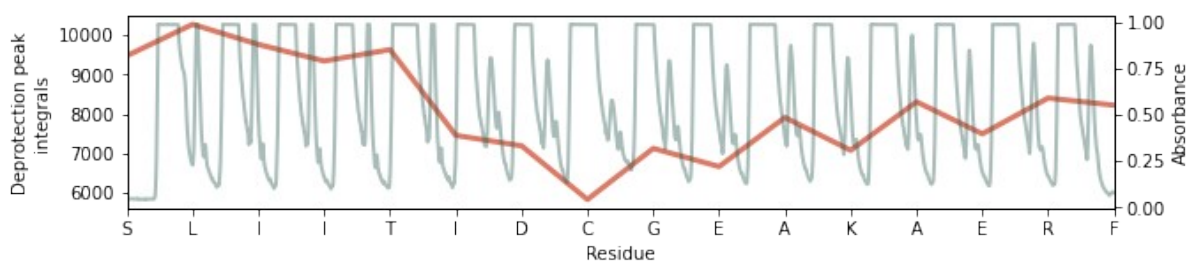**Crude purity (UHPLC):** 38%

Synthesis time: 29 min.

Figure S65: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized under standard AFPS conditions using reduced washing volumes of 12 mL/cycle.

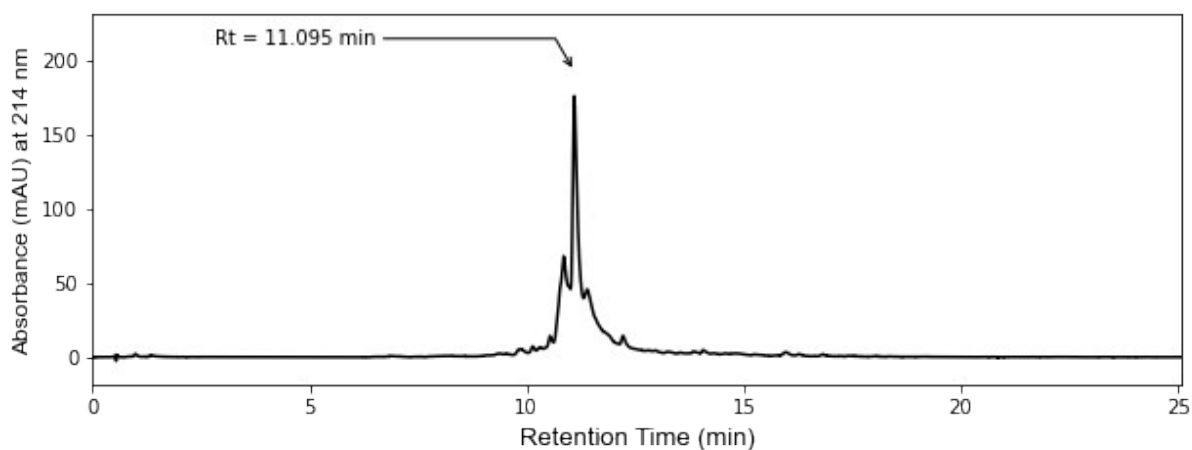## UHPLC analysis of crude Barstar[75–90]

Figure S66: UHPLC of Barstar[75–90] synthesized under standard AFPS conditions using reduced washing volumes of 12 mL/cycle. The standard UHPLC procedure was used.

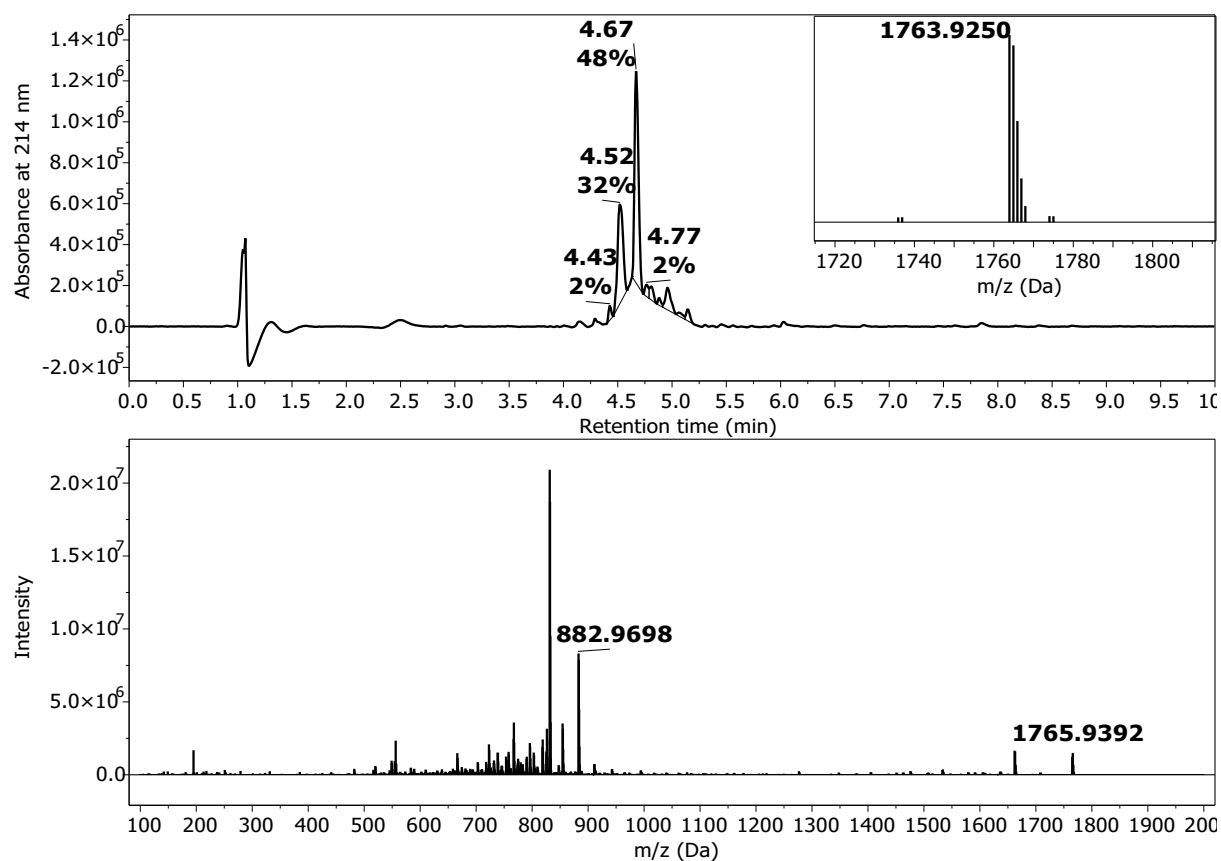## LC–MS analysis of crude Barstar[75–90]

Figure S67: LC–MS profile of crude Barstar[75-90] synthesized under standard AFPS conditions using reduced washing volumes of 12 mL/cycle. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9250.

8 mL/coupling cycle: (Due to the complete lack of baseline resolution the areas are not included)

Sequence: FREAKAEGCD ITIILS-CONH$_2$ (16 amino acids)

Resin: 150 mg of NovaPEG Rink Amide resin (0.41 mmol/g loading, 20.5 μmol)

Crude peptide mass: ~59 mg

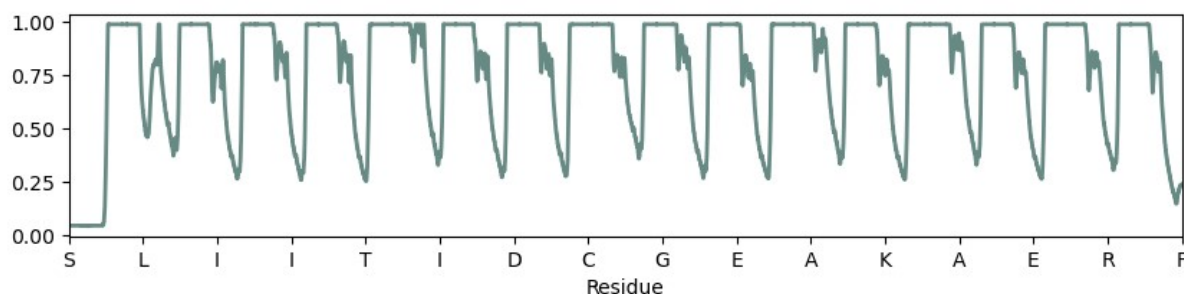**Crude purity (UHPLC):** 25%

Synthesis time: 26 min.

Figure S68: AFPS inline UV–Vis trace in green and deprotection peak integrals (AU) in red of Barstar[75–90] synthesized under standard AFPS conditions using reduced washing volumes of 8 mL/cycle.

41

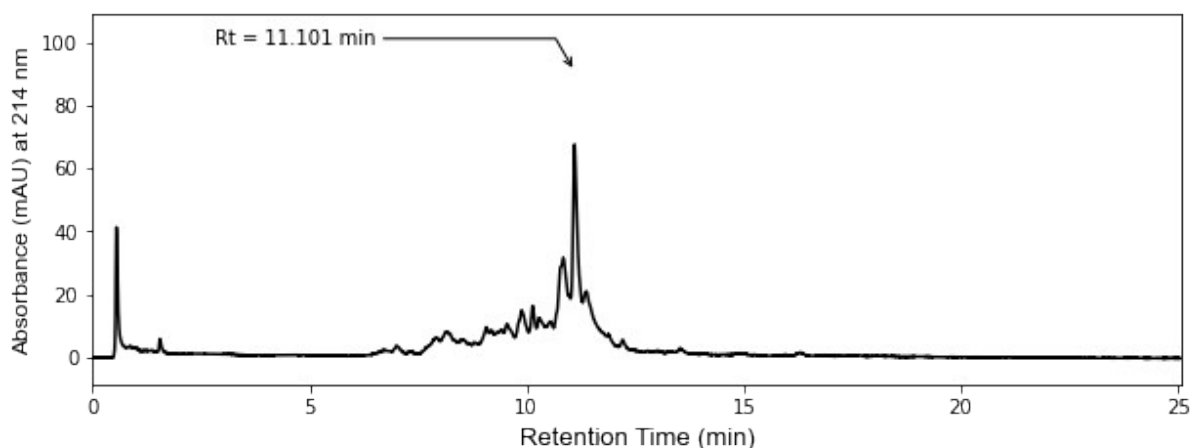## UHPLC analysis of crude Barstar[75–90]

Figure S69: UHPLC of Barstar[75–90] synthesized under standard AFPS conditions using reduced washing volumes of 8 mL/cycle. The standard UHPLC procedure was used.
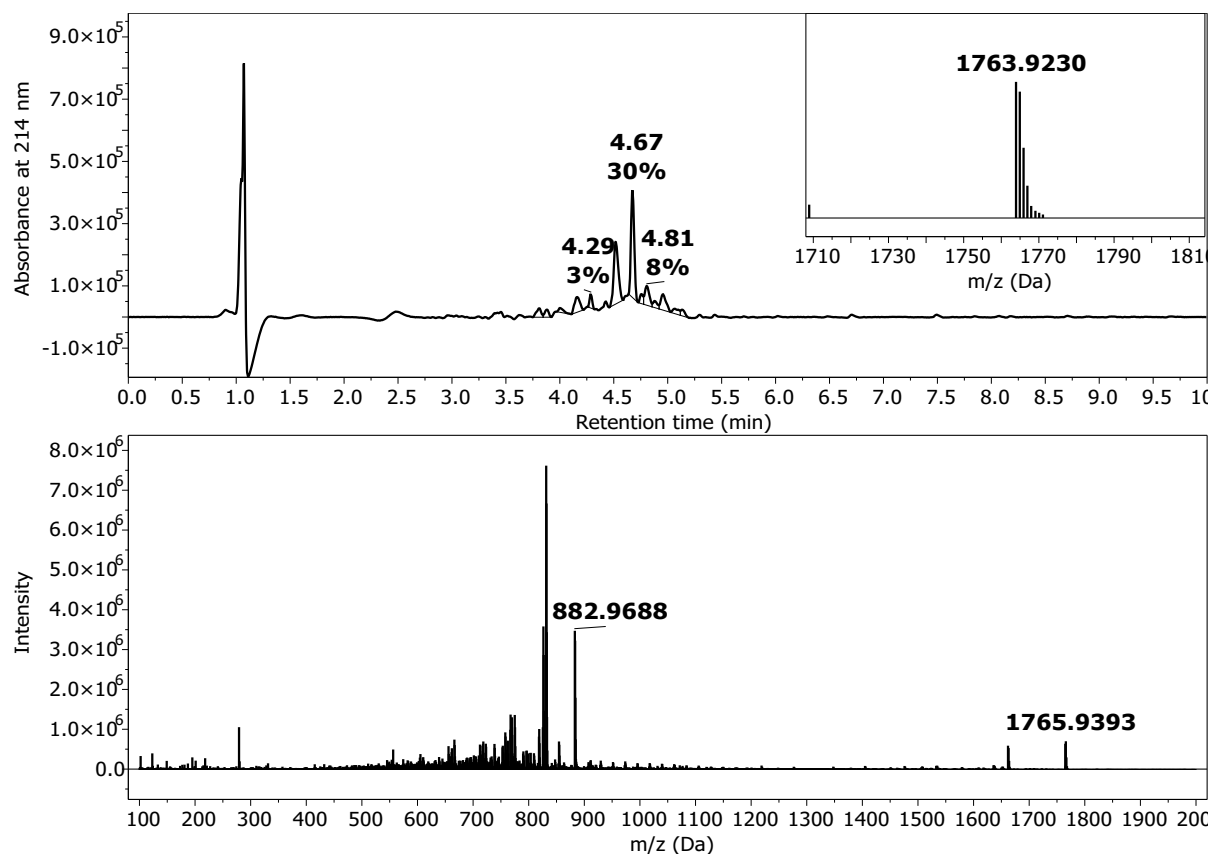
LC–MS analysis of crude Barstar[75–90]

Figure S70: LC–MS profile of crude Barstar[75-90] synthesized under standard AFPS conditions using reduced washing volumes of 8 mL/cycle. Monoisotopic mass (ESI+) calcd. for $C_{77}H_{129}N_{21}O_{24}S$ 1763.92, found 1763.9230.

4   References

S1   N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen and B. L. Pentelute, *Science*, 2020, **368**, 980–987.

S2   S. Mohapatra, N. Hartrampf, M. Poskus, A. Loas, R. Gómez-Bombarelli and B. L. Pentelute, *ACS Cent. Sci.*, 2020, **6**, 2277–2286.