

MIT Open Access Articles

Synthesis of proteins by automated flow chemistry

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Hartrampf, N. et al. "Synthesis of proteins by automated flow chemistry." *Science* 368, 6494 (February 2020): 980-987 © 2020 American Association for the Advancement of Science

As Published: <http://dx.doi.org/10.1126/science.abb2491>

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: <https://hdl.handle.net/1721.1/128762>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Synthesis of Proteins by Automated Flow Chemistry

Nina Hartrampf, Azin Saebi, Mackenzie Poskus, Zachary P. Gates, Alexander J. Callahan, Amanda E. Cowfer, Stephanie Hanna, Sarah Antilla, Carly K. Schissel, Anthony J. Quartararo, Xiyun Ye, Alexander J. Mijalis, Mark D. Simon, **Andrei Loas**, Shunying Liu, Carsten Jessen, Thomas E. Nielsen, Bradley L. Pentelute

Submitted date: 11/02/2020 • Posted date: 11/02/2020

Licence: CC BY-NC-ND 4.0

Citation information: Hartrampf, Nina; Saebi, Azin; Poskus, Mackenzie; Gates, Zachary P.; Callahan, Alexander J.; Cowfer, Amanda E.; et al. (2020): Synthesis of Proteins by Automated Flow Chemistry. ChemRxiv. Preprint. <https://doi.org/10.26434/chemrxiv.11833503.v1>

Ribosomes produce most proteins of living cells in seconds. Here we report highly efficient chemistry matched with an automated fast-flow instrument for the direct manufacturing of peptide chains up to 164 amino acids over 328 consecutive reactions. The machine is rapid - the peptide chain elongation is complete in hours. We demonstrate the utility of this approach by the chemical synthesis of nine different protein chains that represent enzymes, structural units, and regulatory factors. After purification and folding, the synthetic materials display biophysical and enzymatic properties comparable to the biologically expressed proteins. High-fidelity automated flow chemistry is an alternative for producing single-domain proteins without the ribosome.

File list (2)

200210 Manuscript Final_ChemRxiv.pdf (2.19 MiB)

[view on ChemRxiv](#) • [download file](#)

200210 Supplementary Material_ChemRxiv.pdf (9.58 MiB)

[view on ChemRxiv](#) • [download file](#)

Synthesis of Proteins by Automated Flow Chemistry

Authors: N. Hartrampf¹, A. Saebi^{1†}, M. Poskus^{1†}, Z. P. Gates¹, A. J. Callahan¹, A. E. Cowfer¹, S. Hanna¹, S. Antilla¹, C. K. Schissel¹, A. J. Quartararo¹, X. Ye¹, A. J. Mijalis^{1,2}, M. D. Simon¹, A. Loas¹, S. Liu^{1,3}, C. Jessen⁴, T. E. Nielsen⁴ and B. L. Pentelute^{1*}

5

Affiliations:

¹ Massachusetts Institute of Technology, Department of Chemistry, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

10

² Current address: Harvard Medical School, Department of Genetics, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.

³ Current address: Department of Chemistry, East China Normal University, 3663 North Zhongshan Rd., Shanghai, 200062, China.

⁴ Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark.

*Correspondence to: blp@mit.edu

15

† authors contributed equally.

Abstract:

20

Ribosomes produce most proteins of living cells in seconds. Here we report highly efficient chemistry matched with an automated fast-flow instrument for the direct manufacturing of peptide chains up to 164 amino acids over 328 consecutive reactions. The machine is rapid - the peptide chain elongation is complete in hours. We demonstrate the utility of this approach by the chemical synthesis of nine different protein chains that represent enzymes, structural units, and regulatory factors. After purification and folding, the synthetic materials display biophysical and enzymatic properties comparable to the biologically expressed proteins. High-fidelity automated flow chemistry is an alternative for producing single-domain proteins without the ribosome.

25

One Sentence Summary:

A benchtop automated machine synthesizes protein chains in hours.

30

Main Text:

Mechanical pumps, valves, solid supports and computers have transformed the way we perform chemical reactions. In recent decades, the general availability of continuous multi-step flow technology enabled routine access to small molecules ranging from pharmaceutical ingredients to natural products and bulk commodities.(1) Select advantages of flow synthesis over batch methods are in-line spectroscopic monitoring, efficient mixing and precise control over the reaction parameters.(2) It is highly desirable to translate these capabilities and adapt flow chemistry to the total chemical synthesis of biopolymers, including peptides and proteins.

35

40 Sequence-controlled chemical synthesis of full-length protein chains holds the promise to deliver
the future generations of therapeutics. Protein production is an essential part of research in
academia and industry and can be accomplished by biological methods or chemical synthesis.(3)
The vast majority of proteins are obtained by biological expression, a process that limits their
chemical composition to the naturally occurring amino acids.(4) Advances in genetic code
45 expansion have allowed for the incorporation of single unnatural amino acids in the structures of
native proteins.(5) In contrast, chemical synthesis offers unmatched flexibility when incorporation
of multiple unnatural amino acids, post-translational modifications (PTMs) or artificial backbones
is desired.(4) Synthetic proteins have become accessible with a combination of solid-phase and
ligation methodologies. Yet, total chemical synthesis of proteins remains highly labor-intensive.

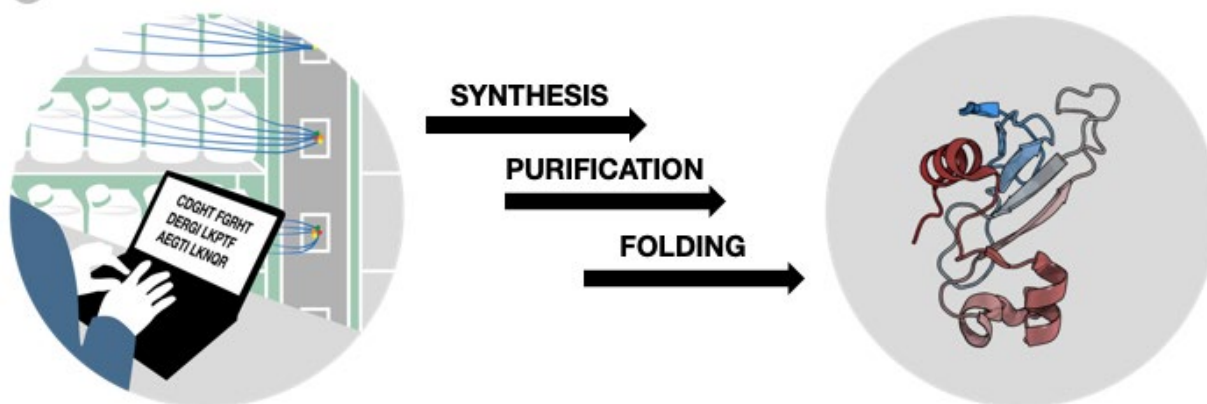
50 Solid-phase peptide synthesis (SPPS) is the foundation of chemical peptide and protein
production.(6) In SPPS, a protected amino acid is linked to an insoluble polymer support (resin)
and the polypeptide chain is elongated through iterative coupling and deprotection cycles. A final
deprotection and cleavage step releases the crude, unprotected peptide. However, despite decades
of optimization, peptides longer than 50 amino acids are difficult to synthesize with standard SPPS
instrumentation. New protocols that routinely overcome generation of by-products arising from
55 deletion, truncation and aggregation of the growing peptide chains are needed.(7, 8) Stepwise
SPPS was devised with the scope of simplifying and accelerating peptide synthesis,(9) however,
it was not until the development of native chemical ligation (NCL) that chemical synthesis of
protein chains became practical. Synthetic proteins can be produced by NCL of relatively long
peptide segments (>50 residues) obtained by SPPS.(10, 11) This methodology has considerably
60 expanded the sequence length of peptides and proteins accessible by chemical synthesis. Despite
the efforts dedicated to improving NCL techniques,(12) a major bottleneck resides in the absence
of a routine protocol to access the requisite peptide fragments.(10, 13) We set out to address this
problem by developing a reliable method to synthesize long peptides and protein chains using flow
chemistry.

65 Flow-based SPPS is gaining momentum owing to its advantageous features, e.g., control over
physical parameters and greatly reduced formation of side-products.(14–16) Studies carried out in
the 1980s found that *automation* and *high fidelity* of peptide synthesis could be achieved by
containing the solid support in a reactor and operating it as a fixed bed.(12) Instead of complex
systems for liquid handling to dispense reagents and wash the resin, high-performance liquid
70 chromatography (HPLC) pumps were employed to continuously deliver reagents, establishing the
principles of peptide synthesis in flow. Inspired by this early work, we developed over the past
five years rapid, automated fast-flow peptide synthesis (AFPS) instrumentation that incorporates
amino acid residues in as little as 40 seconds at temperatures up to 90 °C.(15)

75 Even though prior work by us and others on flow-based SPPS significantly reduced the total
synthesis time, the potential of flow chemistry to enable synthesis of peptide chains in the range
of single domain proteins has not been fully realized.(17–22) We set out to optimize our AFPS
technology to meet this challenge. Here we report a routine protocol that allows for the stepwise

chemical total synthesis of peptide chains exceeding 50 amino acids in length, with a cycle time of ~2.5 minutes per amino acid (**Fig. 1**). The optimized protocol was built on a collection of analytical data acquired with an AFPS system and shown to deliver products with high fidelity and of high chiral purity. Using this protocol, single domain protein chains ranging from barstar (90 amino acids) to sortase A₅₉₋₂₀₆ (sortase A*, 164 amino acids) were synthesized in 3.5–6.5 h. To demonstrate the application to the production of functional proteins, these sequences were folded and their biophysical properties and enzymatic activities were determined. This advancement brings the timescale of chemical protein synthesis on par with that of recombinant expression and therefore offers a practical alternative to biological methods, while opening up the chemical space beyond canonical amino acids.

A Fully Automated Synthesis of Functional Synthetic Proteins



B Automated flow peptide synthesis (AFPS)



Fig. 1 Automated fast-flow solid-phase peptide synthesis enables high-fidelity production of long amino acid sequences. A) Fully automated chemical flow synthesis yields peptide chains, which —after purification and folding— give functional proteins; B) Automated fast-flow peptide synthesizer (AFPS) and main advantages of flow chemistry for solid phase peptide synthesis (SPPS).

Results

Rapid and reproducible screening of reaction variables enables optimization of a general Fmoc chemistry-based AFPS protocol. We chose to first optimize coupling efficiency and later

investigate possible side-reactions induced by the optimized coupling conditions. On a benchmark AFPS instrument previously developed in our laboratory,(15, 16) reagents are mixed, heated and delivered onto a pre-tempered solid support using three HPLC pumps. In-line ultraviolet-visible (UV-Vis) detection of the reactor eluent is used to monitor removal of the *N*-terminal protecting group after each coupling cycle. Indirectly, this information reports on the efficiency of the preceding coupling step.

We first optimized general parameters including flow rate, reaction solvent, reagent concentration, temperature and coupling agents (**Table S1–S7**). Modifications to our original AFPS protocol included increasing reagent concentrations to 0.4 M,(23) the use of amine-free DMF and an increase of temperature to 85–90 °C for reagent activation and coupling.¹ The performance of different activators for the coupling step was also investigated, identifying the azabenzotriazol-reagents PyAOP and HATU as the optimal activators.

Automated collection of analysis data and synthesis parameters allowed for an optimization of residue-specific coupling conditions. By comparing data on amino acid deprotections, we were able to gain information on coupling efficiency for all canonical amino acids and generated a general amino-acid-specific recipe (**Table S8–S9**). Analytical comparison of the products obtained for GLP-1 is illustrative of the improvement in crude peptide quality achieved with the optimized synthesis conditions (**Fig. 2A**).

Conditions were modulated to suppress aspartimide formation, a major side reaction in SPPS and AFPS. As increased temperature leads to more aspartimide formation, various deprotection bases, additives and aspartic acid protecting groups were screened to minimize this unwanted side reaction.(24, 25) We found that milder deprotection bases (i.e., piperazine and HOBT/piperidine) and bulky aspartic acid protecting groups (i.e., 3-methyl-pent-3-yl esters, OMpe) decreased the rate of aspartimide formation (**Fig. S3 and Table S10**). The most effective strategies, however, were the addition of formic acid as a piperidine additive and backbone protection with dimethoxybenzyl glycine. Formic acid (1% in 20:80 v/v piperidine:DMF) was therefore used as an additive for deprotection, and backbone protection was applied for collagen and FGF1 syntheses.

The retention of chirality for amino acids at high risk of epimerization, cysteine and histidine, was confirmed in a final optimization step (**Fig S4-S9**).(26) The influence of temperature, time and activating agent, as well as different protecting groups were screened (**Fig. 2A,B**).(17) For both amino acids, epimerization significantly increases with activation time and temperature. The choice of protecting group proved to be critical for histidine. Ultimately, activation of Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH with PyAOP with a shorter activation time at 60 °C resulted in <2% D-epimer formation. Next, we determined that the amount of epimerization under these

¹ The flow rate was set to 40 mL/min to minimize the effect of pump cavitation.

optimized conditions does not increase over multiple coupling cycles (**Fig. 2D**). The amount of D-isomer did not change over 100 amino acid couplings, which verified that epimerization of cysteine and histidine only occurs during the activation step. Implementation of these conditions allowed us to finalize the general AFPS protocol, which was then applied to the production of sequences exceeding 50 amino acids (**Table S11 and SI section 3.10**).

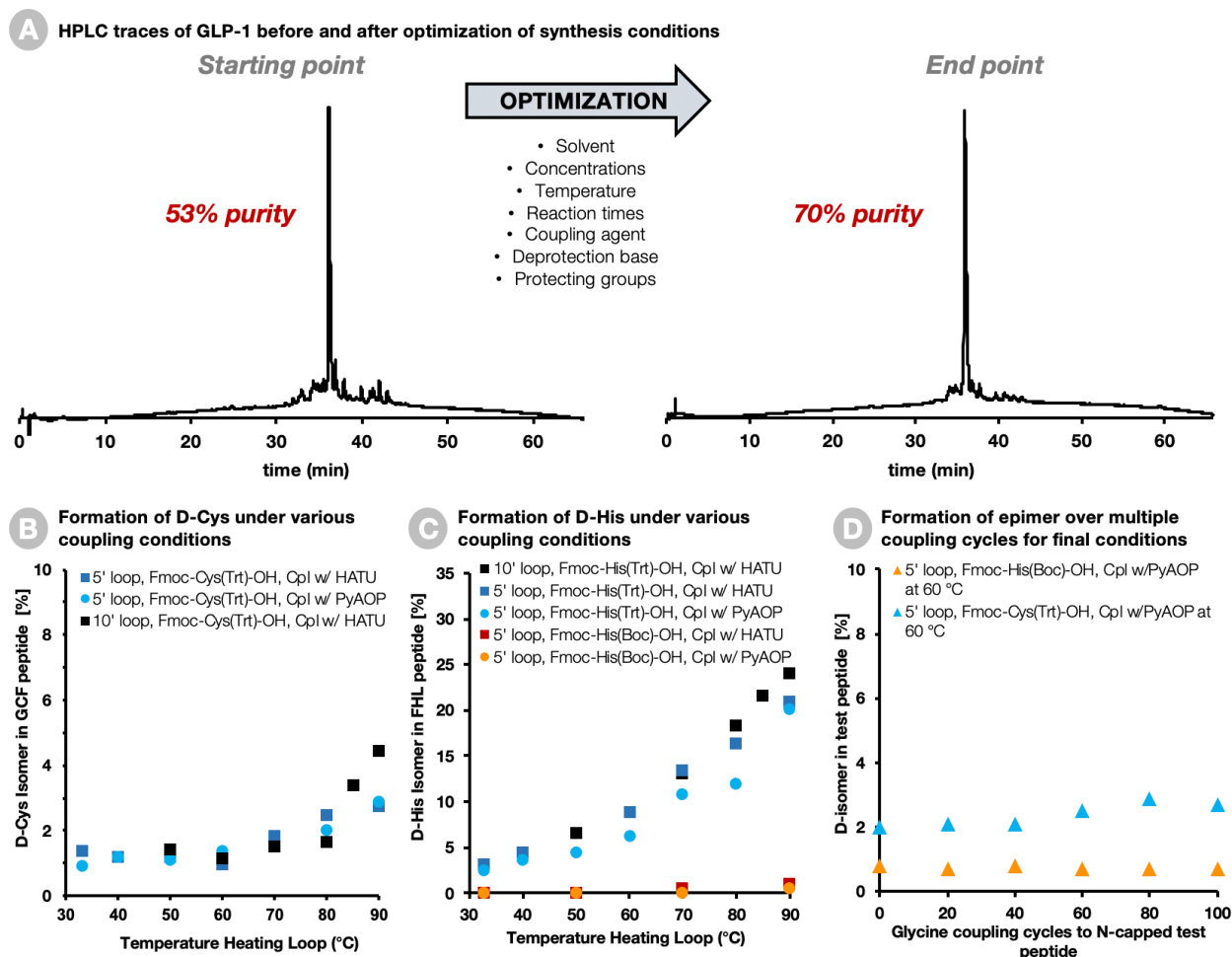


Fig. 2 Optimization of AFPS conditions improves synthesis outcome and reduces the amount of cysteine and histidine epimerization. A) Synthesis of GLP-1 using starting conditions and optimized conditions; B) Quantification of cysteine epimerization on activation temperature, heating time (5' loop and 10' loop) and activator in a GCF-test peptide, isomer was quantified from extracted ion chromatograms (EIC) on liquid chromatography–mass spectrometry (LC-MS) by comparison to reference peptides; C) Quantification of histidine epimerization on activation temperature, heating time (5' loop and 10' loop) and activator in a FHL-test peptide, the D-isomer was quantified by analytical HPLC by comparison to reference peptides; D) Quantification of epimerization over multiple coupling cycles. GCF and FHL were synthesized under optimized conditions and the N-terminus was manually capped with a *tert*-butyloxycarbonyl (Boc)-protecting group. 100 glycine couplings were executed, and a sample was taken out for analysis every 20 amino acid couplings. Cpl. w/ = coupled with.

140

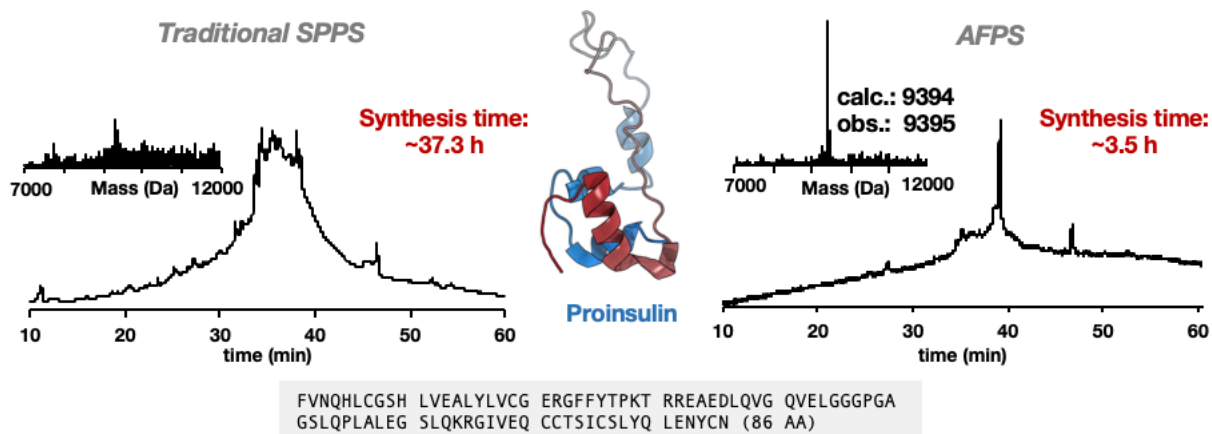
145

Optimized AFPS outperforms traditional synthesis methods. We investigated if our optimized AFPS conditions could facilitate the synthesis of longer sequences using proinsulin (86 amino acids) and HIV-1 protease (99 amino acids) as test sequences. The total synthesis of human proinsulin was previously reported using native chemical ligation of three peptide fragments

150 individually prepared by SPPS.(27) HIV-1 protease was also previously prepared using stepwise
and chemical ligation routes under Boc-SPPS conditions.(28, 29) Using our standard AFPS
protocol, the syntheses of proinsulin and HIV-1 protease were completed in 3.5 and 4.5 hours,
respectively. HPLC purification yielded 2.2 mg (1%) of purified proinsulin and 5.3 mg (1%) of
purified HIV-1 protease.

155 A comparison between AFPS and standard SPPS synthesis on commercially available synthesizers
performed at room temperature, 70 °C and 90 °C, indicated significantly improved synthetic
outcome for the optimized AFPS protocol (**Fig. 3** and **SI section 4**). To facilitate this comparison,
identical resin, scale, and amino acid equivalents were used for all sequences. On each instrument
machine-specific, optimized conditions were used to achieve the best synthesis outcome. For HIV-
160 protease and proinsulin, AFPS yielded the desired product as the major species along with minor
by-products of similar weight, as determined by analytical HPLC and liquid chromatography–
mass spectrometry (LC-MS). By contrast, synthesis on commercially available peptide
synthesizers took approximately five times longer and resulted in a complex compound mixture.
AFPS therefore offers a significant improvement when directly compared to traditional SPPS
165 methods, both with respect to time and performance.

Traditional SPPS vs. AFPS, analytical data for crude proinsulin



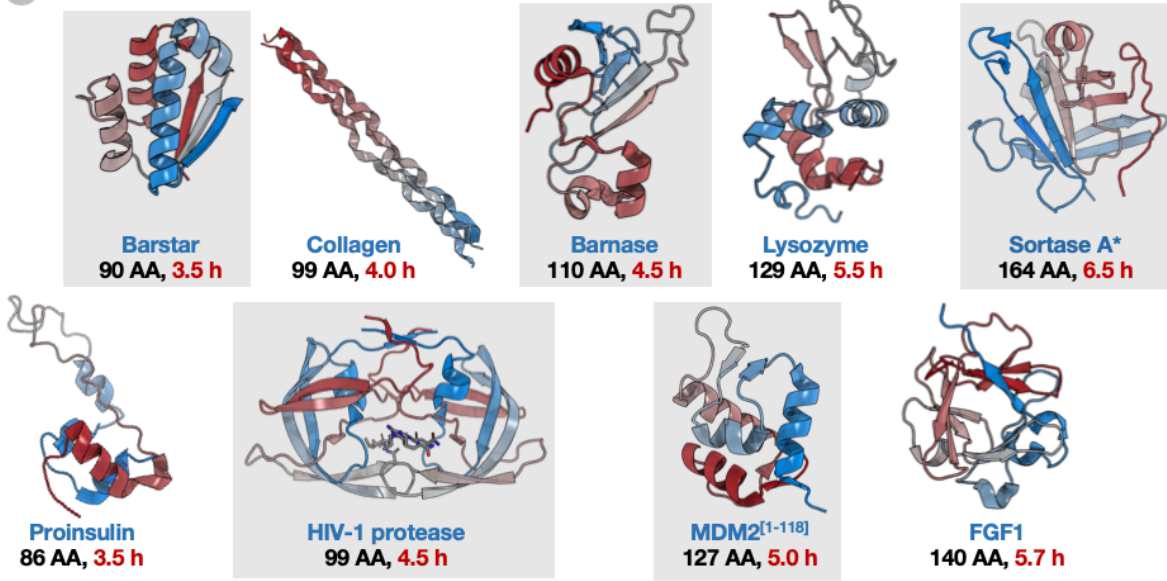
170 **Fig. 3 Synthesis of proinsulin demonstrates the advantage of AFPS over traditional SPPS methods.** Analytical HPLC data of
the crude proinsulin are presented as the main chromatographic trace with absorbance detection at 214 nm (additional details can be
found in the SI). Deconvoluted masses are displayed in the inset. Analytical data for the synthesis of crude proinsulin using SPPS on
a commercially available synthesizer at 70 °C with total cycle times of 26 min/amino acid and 40 eq. amino acid for each coupling are
displayed on the left; analytical data for the synthesis of crude proinsulin using AFPS are displayed on the right.

Optimized AFPS enables routine access to peptides in the size range of single domain proteins (~130 AAs).

To demonstrate general applicability of our synthesis protocol, the synthesis of additional protein chains ranging from ~70 to ~170 amino acids was performed (**Fig. 4 and SI section 5**). These sequences were chosen to enable comparison with literature data. We chose historically relevant targets for drug discovery, such as HIV-1 protease and MDM2,^(30, 31) but also proteins that serve as therapeutics themselves, such as FGF1 and proinsulin.^(32, 33) The ability of AFPS technology to rapidly incorporate non-canonical amino acids was tested by the synthesis of derivatives of barnase and HIV-1 protease. Barstar, barnase, lysozyme, MDM2 and sortase A* allowed for a direct comparison of recombinant and synthetic proteins. All sequences were successfully synthesized in 3.5 to 6.5 h of synthesis time using our optimized AFPS protocol.

The desired protein was the main product in every synthesis and HPLC purification yielded milligram quantities of product. Isolated yields following HPLC purification ranged from 2.2–19.0 mg (1–5%), a sufficient amount of material for folding and evaluation of tertiary structure and biological function (**Fig. 5 and SI section 5**). In conclusion, optimized AFPS allows for the routine stepwise chemical synthesis of peptide chains of up to ~170 amino acids and therefore significantly decreases time and labor associated with the production of single domain proteins.

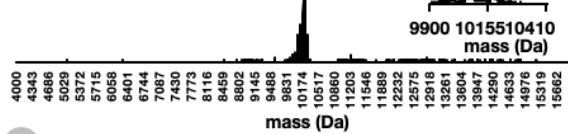
A Single domain proteins synthesized using AFPS



Length of amino acid sequences

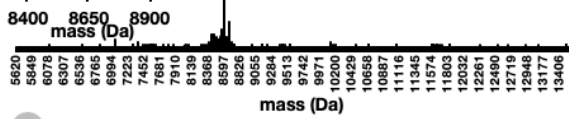
B Deconvoluted mass of crude barstar

Calc.: 10211
Obs.: 10211



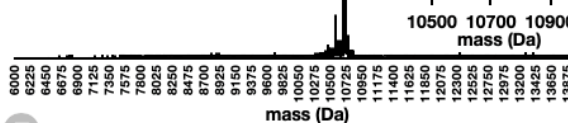
C Deconvoluted mass of crude collagen

Calc.: 8620
Obs.: 8620



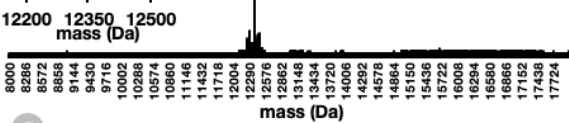
D Deconvoluted mass of crude HIV-1 protease

Calc.: 10703
Obs.: 10703



E Deconvoluted mass of crude barnase

Calc.: 12383
Obs.: 12384



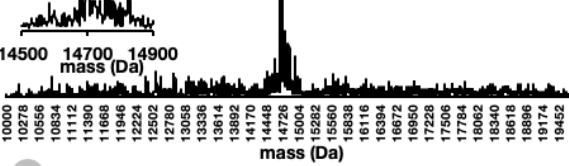
F Deconvoluted mass of crude MDM2

Calc.: 14557
Obs.: 14558



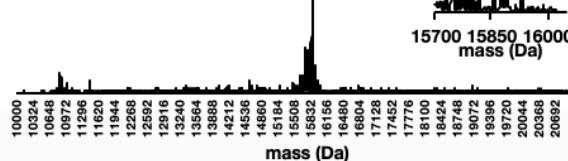
G Deconvoluted mass of crude lysozyme

Calc.: 14701
Obs.: 14701



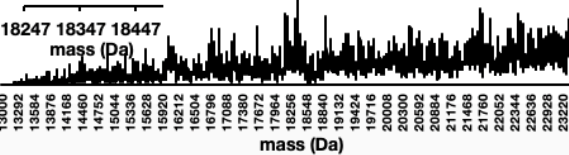
H Deconvoluted mass of crude FGF1

Calc.: 15875
Obs.: 15875



I Deconvoluted mass of crude sortase A*

Calc.: 18392
Obs.: 18393



190

Fig. 4 Automated fast-flow solid-phase peptide synthesis enables high-fidelity production of long amino acid sequences.

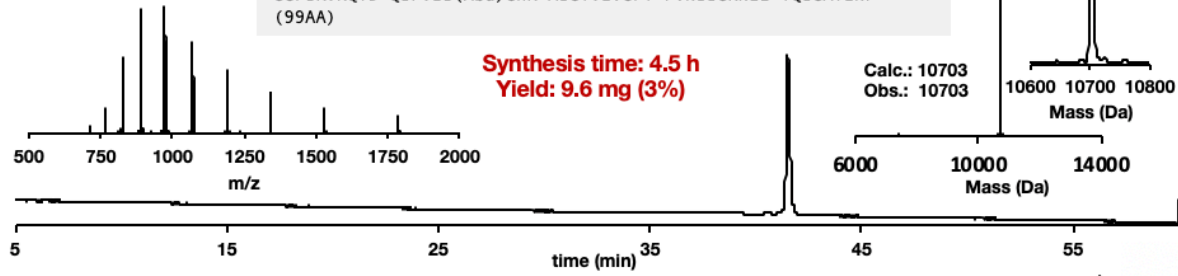
195

A) Sequences produced using an automated fast-flow peptide synthesis (AFPS) instrument; sequences highlighted in gray were folded, purified and their structure and biological activity was evaluated. All peptides and proteins were synthesized using a single standard recipe. PDB: 3IOL (GLP-1), 2ZA4 (barstar), 2KQP (proinsulin), 1CGD (collagen); 3HBO (HIV-1 protease dimer with inhibitor), 2ZA4 (barnase), 3G03 (MDM2), 1BB3 (lysozyme), 2J3P (FGF1), 2KID (sortase A). Deconvoluted mass spectra of crude B) barstar, C) collagen, D) HIV-protease (Kent sequence), E) barnase, F) MDM2, G) lysozyme, H) FGF1, and I) sortase A* obtained by summation of the respective entire liquid chromatography peaks. Additional analytical data for all syntheses can be found in the Supporting Information.

A HIV-1 protease

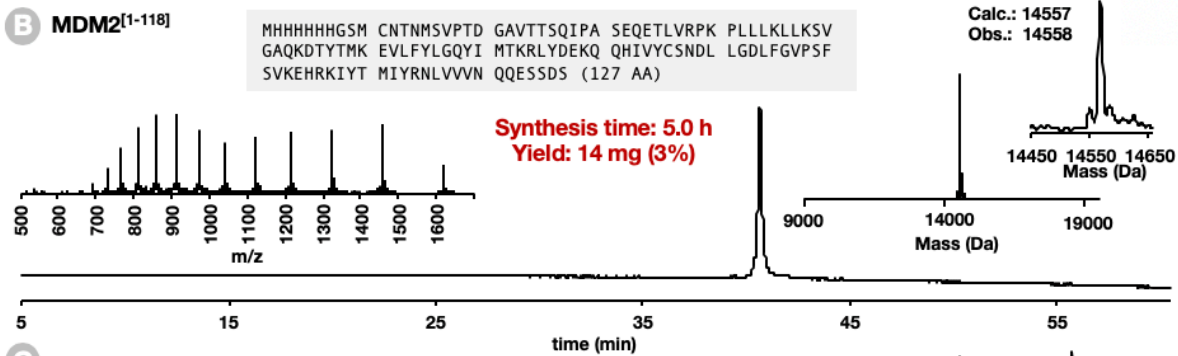
PQITLWKRPL VTIRIGGQLK EALLDTGADD TVIEE(N1e)NLPK KWKPK(N1e)IGGI
GGFIKVRQYD QIPVEI(Abu)GHK AIGTVLVGPT PVNIIGNLL TQIGATLNF
(99AA)

Synthesis time: 4.5 h
Yield: 9.6 mg (3%)

**B MDM2^[1-118]**

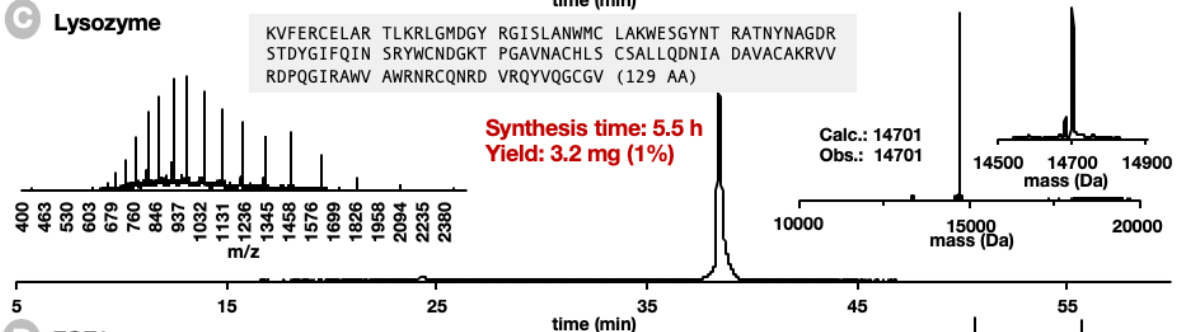
MHHHHHGM CNTNMSVPTD GAVTTSQIPA SEQETLVRPK PLLKLLKSV
GAQKDTYTMK EVLFYLGQYI MTKRLYDEKQ QHIVYCSNDL LGDLFGVPSF
SVKEHRKIYT MIYRNLVVVN QQESSDS (127 AA)

Synthesis time: 5.0 h
Yield: 14 mg (3%)

**C Lysozyme**

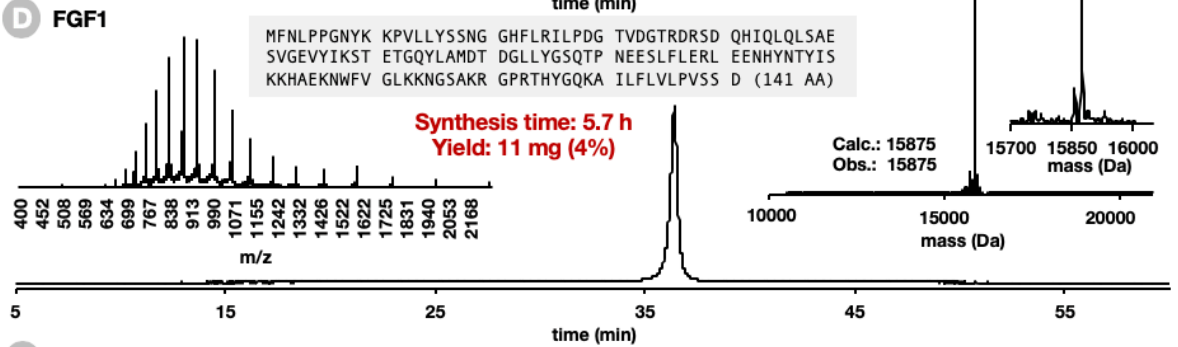
KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR
STDYGFIFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV
RDPQGIRAWW AWRNRCQNRD VRQYVQCGGV (129 AA)

Synthesis time: 5.5 h
Yield: 3.2 mg (1%)

**D FGF1**

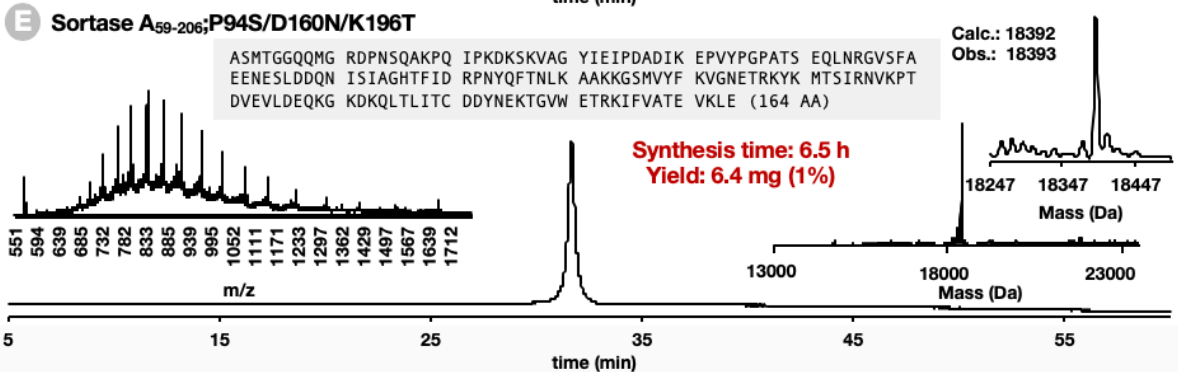
MFNLPPGNYK KPVLLYSSNG GHFLRILPDG TVDGTDRSD QHIQLQLSAE
SVGEVYIKST ETGQYLAMD T DGLLYGSQTP NEESLFLERL EENHYNTYIS
KKHAEKNWFV GLKKNQSAKR GPRTHYGQKA ILFLVLPVSS D (141 AA)

Synthesis time: 5.7 h
Yield: 11 mg (4%)

**E Sortase A₅₉₋₂₀₆:P94S/D160N/K196T**

ASMTGGQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGATS EQLNRGVSFA
EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKGSMVYF KVGNETRKYK MTSIRNVKPT
DVEVLDEQKG KDKQLTLITC DDYNEKTGVW ETRKIFVATE VKLE (164 AA)

Synthesis time: 6.5 h
Yield: 6.4 mg (1%)



200 **Fig. 5 AFPS yields milligram quantities of proteins in hours.** For all cases analytical HPLC data of the purified proteins are presented as the main chromatographic trace with absorbance detection at 214 nm (additional details are found in the SI). Electrospray ionization (ESI) mass spectrum (upper left) and deconvoluted mass spectrum (upper right) are also shown in each case. Both spectra were obtained by summation of the entire liquid chromatography peak; A) HIV-1 protease (99 amino acids), gradient for analytical HPLC: 5–65% B; B) MDM2^[1-118] with His-Tag (127 amino acids) gradient for analytical HPLC: 5–65% B, C) Lysozyme (129 amino acids) gradient for analytical HPLC: 5–65% B, D) FGF1 (141 amino acids) gradient for analytical HPLC: 5–65% B, E) sortase A (164 amino acids), gradient for analytical HPLC: 5–65% B. A linear gradient of acetonitrile with 0.08% trifluoroacetic acid (TFA) added (solvent B) in water with 0.1% TFA added (solvent A) was used in all cases.

205 **The structure and function of folded synthetic proteins are comparable to recombinant samples.** Determining the purity of long synthetic peptides is challenging because of difficulties associated with identification and quantification of by-products by standard analytical techniques.

210 In a physiological environment the native folded structure of a globular protein – which gives rise to its unique biological activity – is determined by its amino acid sequence.⁽³⁴⁾ As a consequence, the tertiary structure of a protein can be used as a measure of the chemical integrity of the primary amino acid sequence.⁽³⁵⁾

215 We folded and further purified select synthetic proteins by size exclusion chromatography and ion exchange chromatography and characterized their tertiary structure with biophysical and functional assays, alongside recombinant protein standards. Our goal was to demonstrate the fidelity of our AFPS protocol in delivering synthetic proteins of defined covalent structure and high chiral integrity. To this aim, we thoroughly characterized barnase, and further investigated barstar, sortase A, MDM2 and HIV-1 protease. Folding of the synthetic proteins was case-specific and was achieved either by following a literature protocol or by screening various folding conditions.

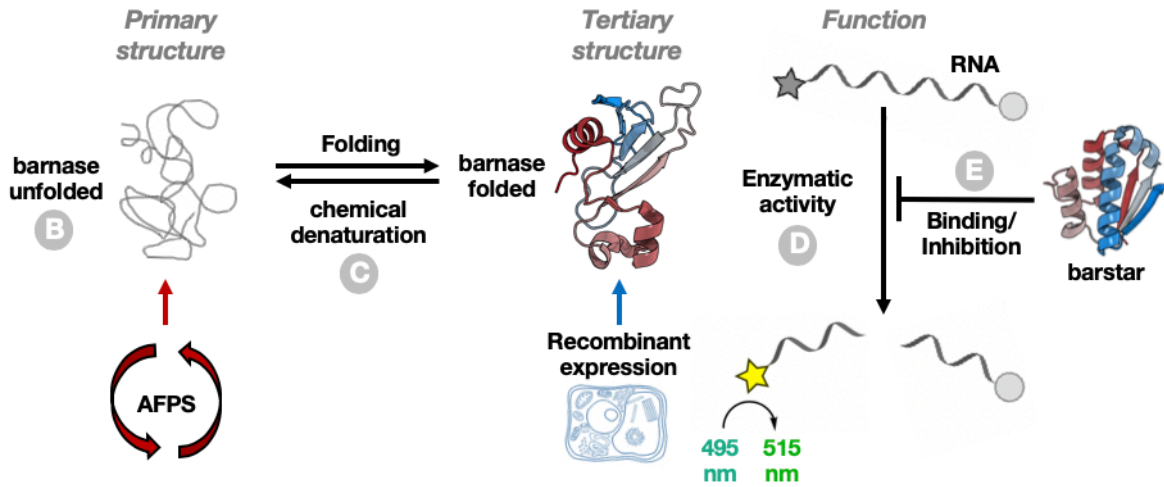
220 Chemical denaturation is diagnostic for assessing structural integrity and stability of synthetic proteins. The globular protein barnase, a bacterial RNase isolated from *Bacillus amyloliquefaciens*, is a model system to investigate protein folding, denaturation and binding to its inhibitor protein barstar (**Fig. 6A**).^(36, 37) The primary structures of synthetic and recombinant barnase were indistinguishable by LC-MS and HPLC methods (**Fig. 6B**). We then used a chemical denaturation fluorometric assay as a read-out for the integrity of the tertiary structure (**Fig. 6C**). In this assay, tyrosine fluorescence was used to monitor the folding equilibrium, as the concentration of urea was varied. Synthetic barnase exhibited a transition midpoint (the concentration at which half of the sample is unfolded) that compared well to both the authentic recombinant sample and literature value ($[D]_{50\%, \text{synthetic}} = 4.68 \pm 0.06 \text{ M}$; $[D]_{50\%, \text{recombinant}} = 4.63 \pm 0.04 \text{ M}$; $[D]_{50\%, \text{literature}} = 4.57 \text{ M}$).⁽³⁸⁾ More importantly, the m -values obtained in the experiment, which describe the slope of the unfolding transition and are a sensitive measure of structural homogeneity, were similar ($m_{\text{synthetic}} = 1.82 \pm 0.25 \text{ kcal mol}^{-1} \text{ M}^{-1}$, $m_{\text{recombinant}} = 1.88 \pm 0.21 \text{ kcal mol}^{-1} \text{ M}^{-1}$ and $m_{\text{literature}} = 2.06 \text{ kcal mol}^{-1} \text{ M}^{-1}$). If the synthetic protein were microheterogeneous (e.g., contained a distribution of isomers or deletion co-products), then the m -value may be altered due to the distribution of $[D]_{50\%}$ values represented within the mixture. Therefore, since the synthetic sample

exhibited an m -value within the error of the recombinant sample, we concluded that microheterogeneity was not significant.

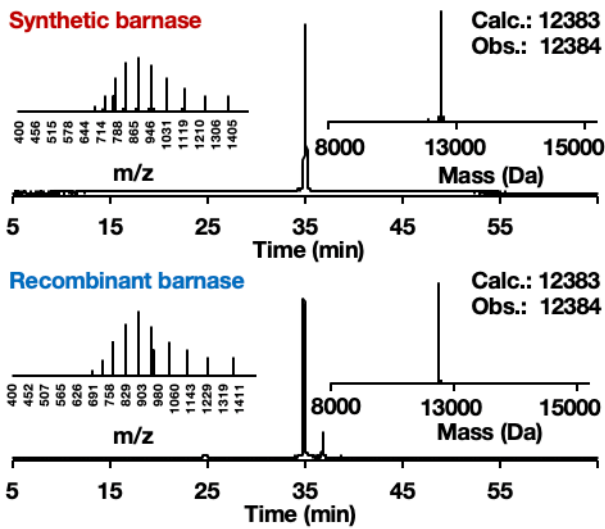
240 Enzymatic assays show comparable activity of synthetic proteins obtained by AFPS and their recombinant equivalents. Enzymatic catalysis is sensitive to minor changes in the enzyme's tertiary structure, for which even single point mutations can have a major impact.(39, 40) We evaluated the native activity of three synthetic variants of well-studied enzymes: barnase, HIV-1 protease and sortase A*. Barnase catalyzes hydrolysis at diribonucleotide GpN sites. Its specific activity can be measured by monitoring hydrolysis of a DNA/RNA hybrid containing a Förster Resonance Energy Transfer (FRET) fluorophore pair.(41) The enzymatic efficiency of synthetic barnase was $k_{cat}/K_M = (7.6 \pm 0.2) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, which is comparable to that of recombinant barnase (245 $k_{cat}/K_M = (9.0 \pm 0.3) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) as determined using the same assay (**Fig. 6D**). Further, HIV-1 protease hydrolyzes the peptides of HIV, and using a fluorogenic peptide allows for quantification of its proteolytic activity.(42) Synthetic HIV-1 protease displays a Michaelis constant of $K_M = 20.9 \pm 1.0 \text{ mM}$ and a turnover number of $k_{cat} = 29.6 \pm 4.1 \text{ s}^{-1}$, close to literature values for the recombinant sample (**Fig. S18 and S19**).(29) Furthermore, incubation of the synthetic protease with a model substrate peptide results in wild-type like specificity with exclusive cleavage at a single Phe/Pro site (**Fig. S20 and 21**).(28) Finally, sortase A₅₉₋₂₀₆ is a transpeptidase produced by Gram-positive bacteria that catalyzes a cell wall sorting reaction at a threonine-glycine bond in the LPXTG motif (43). We synthesized the 164-amino acid long sortase A* variant [P94S/D160N/K196T] to allow for direct comparison to a recombinant standard.(44, 45) At a concentration of 0.01 mg/mL, synthetic sortase A* led to 47% product formation by LC-MS within 24 h (starting from 0.2 mg/mL GGGGGLY and AQALPETGEE as test substrates) (**SI Fig. 23**). 250 This conversion value is comparable to that determined for the recombinant protein (50% product formation within 24 h). Enzymatic activity assays of synthetic proteins accessed by AFPS therefore confirmed both the high substrate specificity and comparable activity to recombinant enzymes and literature values.

265 Binding studies of synthetic MDM2 and barnase confirmed specific affinities for their respective substrates. Barnase binds selectively and with high affinity to its inhibitor barstar. In a gel-based assay recombinant barstar inhibited RNase activity of synthetic and recombinant barnase in a concentration-dependent manner (**Fig. 6E**).(42) In addition, synthetic barstar obtained with AFPS performed comparably to recombinant barstar. To quantify binding of a synthetic protein to a known ligand, we also characterized the N -terminal binding domain of MDM2^[1-118].(31) The binding of MDM2 to p53 is a key interaction in multiple pathways upregulated in cancer.(47, 48) 270 We folded milligram quantities of synthetic MDM2^[1-118] and characterized its binding to immobilized p53^[14-29] using biolayer interferometry (**Fig S24 and S25**). Synthetic MDM2^[1-118] displayed an affinity toward p53 ($K_d = 6.25 \text{ }\mu\text{M}$) comparable to the literature value ($K_d = 5.45 \text{ }\mu\text{M}$) obtained under the same folding conditions.

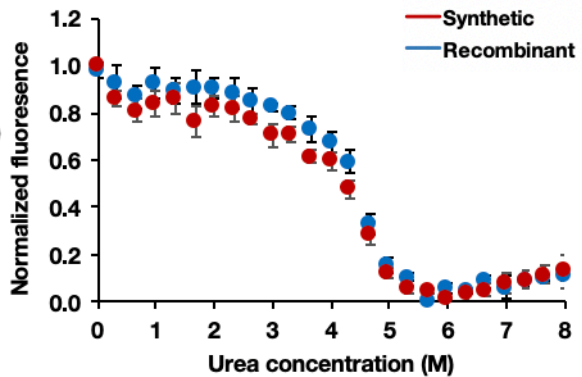
A Production, structure and function of barnase



B Analysis of primary structure



C Chemical denaturation of synthetic barnase

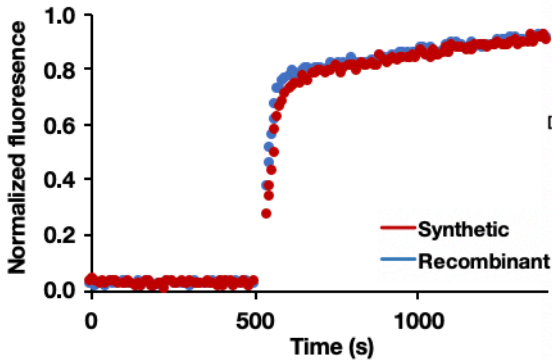


Synth. barnase: $m = 1.82 \pm 0.25 \text{ kcal mol}^{-1} \text{ M}^{-1}$
 $[D]_{50\%} = 4.68 \pm 0.06 \text{ M}$

Rec. barnase: $m = 1.88 \pm 0.21 \text{ kcal mol}^{-1} \text{ M}^{-1}$
 $[D]_{50\%} = 4.63 \pm 0.04 \text{ M}$

Literature: $m = 2.06 \text{ kcal mol}^{-1} \text{ M}^{-1}$; $[D]_{50\%} = 4.57 \text{ M}$

D Fluorogenic activity assay



Synthetic barnase: $k_{\text{cat}}/K_M = (7.6 \pm 0.2) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$

Recombinant barnase: $k_{\text{cat}}/K_M = (9.0 \pm 0.3) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$

Lit: $(1.3 \pm 0.4) \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$

E Barstar inhibits barnase activity

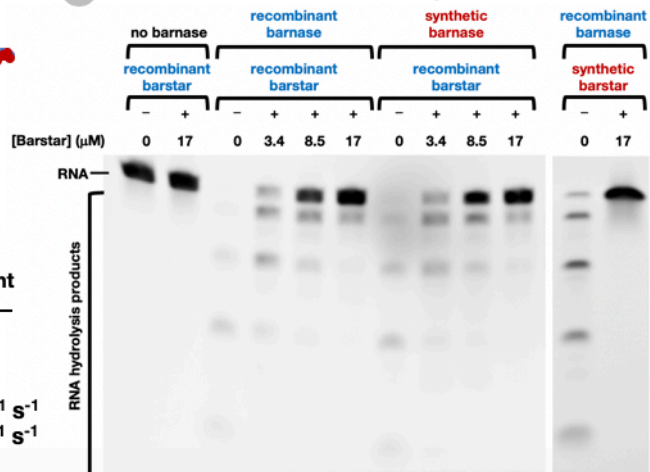


Fig. 6 Synthetic barnase and synthetic barstar fold into the native tertiary structure and display enzymatic activity comparable to recombinant samples. A) Conceptual overview of production and analysis methods; B) Comparison of primary sequence obtained from AFPS synthesis and recombinant expression. For both cases analytical HPLC data of the purified barnase are presented as the main chromatographic trace with absorbance detection at 214 nm (additional details can be found in the SI). Electro spray ionization (ESI) mass spectrum and deconvoluted mass spectrum of the purified peptide samples are displayed in the upper left and the upper right insets, respectively. Both spectra were obtained by summation over the entire liquid chromatography peak in the chromatogram; C) Structural evaluation of barnase in a chemical denaturation assay using urea as denaturant; D) Quantitative enzymatic activity assay for the determination of k_{cat}/K_M values; E) Barnase inhibition and binding assay using recombinant and synthetic barstar. 3.4 nM barnase was used in all experiments.

Discussion

AFPS technology enables rapid and high-fidelity production of peptide chains over three times longer than previously accessible by routine standard SPPS methods.⁽⁷⁾ We show here that solid-phase flow chemistry yields sequences up to 164 amino acids long of sufficient chiral purity and quality to facilitate production of functional single domain protein chains. Selected examples were folded and the integrity of the primary and tertiary structures, as well as their biological activity was confirmed. These advances provide a viable solution to reliably assemble long linear peptide chains, shifting the focus in the field of chemical protein synthesis to improving folding protocols and — most importantly — applications. We envision in future studies testing the generality of our flow synthesis method on a much larger pool of protein chains.

The optimized AFPS protocol demonstrates an advantage of flow chemistry over common batch chemistry methods for peptide synthesis. An improvement to existing flow protocols was achieved by rapid screening of variables in a reproducible reaction setup. Even though in this study AFPS yields superior results over traditional synthesis methods in terms of total synthesis time and crude product quality, general challenges associated with peptide synthesis, such as low atom economy and the use of DMF as a solvent, remain unsolved. Since we implemented AFPS, we have produced over 5000 peptides and automatically collected in-line analysis data for all syntheses. Moving forward, this extensive, high-quality data set could be leveraged to further improve peptide synthesis in flow using machine learning and other computational methods. Ultimately, we intend for this report to serve as a blueprint for the automated flow synthesis of other biopolymers and artificial sequence-defined polymers.⁽⁴⁹⁾

A robust, widely available routine method for chemical production of proteins is poised to have a strong impact on chemical biology and the development of new therapeutics. Combined with chemical ligation, rapid stepwise production of single domain proteins by AFPS technology will extend the practical applications of total chemical synthesis to the majority of human proteins (~30 kDa).^(10, 50) Additional research avenues opened by our method include rapid access to mirror-image proteins, post-translationally modified proteins and *de novo*-designed, abiotic proteins. Introduction of non-canonical amino acids as point mutations in native proteins will make accessible variants with significantly altered biological function, e.g., catalytic activity.^(51, 52)

315 Finally, AFPS has the potential to enable on-demand production of time-sensitive and potentially life-saving personalized medicine, such as for enzyme replacement therapy or neoantigen cancer vaccines.(53, 54) We are looking forward to unlocking the power of our method along these lines of exploration.

References and Notes:

- 320 1. J. Britton, C. L. Raston, Multi-step continuous-flow synthesis. *Chemical Society Reviews*. **46**, 1250–1271 (2017).
2. D. Webb, T. F. Jamison, Continuous flow multi-step organic synthesis. *Chemical Science*. **1**, 675–680 (2010).
- 325 3. B. G de la Torre, F. Albericio, The pharmaceutical industry in 2018. An analysis of FDA drug approvals from the perspective of molecules. *Molecules*. **24**, 809 (2019).
4. T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallée, A. Phantumrithiwath, O. D. Coleman, M.-L. Thézénas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, B. G. Davis, 330 Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science*. **354**, aag1465 (2016).
5. J. W. Chin, Expanding and reprogramming the genetic code. *Nature*. **550**, 53 (2017).
6. R. B. Merrifield, Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *Journal of the American Chemical Society*. **85**, 2149–2154 (1963).
- 335 7. S. B. H. Kent, Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–351 (2009).
8. A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, Manufacturing peptides as active pharmaceutical ingredients. *Future medicinal chemistry*. **1**, 361–377 (2009).
9. B. Merrifield, in *Methods in Enzymology* (Academic Press, 1997; <http://www.sciencedirect.com/science/article/pii/S0076687997890404>), vol. 289, pp. 3–13.
- 340 10. P. E. Dawson, S. B. H. Kent, Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem.* **69**, 923–960 (2000).
11. P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, Synthesis of proteins by native chemical ligation. *Science*. **266**, 776 (1994).
- 345 12. V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **119**, 7328–7443 (2019).

13. A. El-Faham, F. Albericio, Peptide Coupling Reagents, More than a Letter Soup. *Chem. Rev.* **111**, 6557–6602 (2011).
14. N. Ahmed, Peptide Bond Formations through Flow Chemistry. *Chemical Biology & Drug Design.* **91**, 647–650 (2018).
350
15. C. P. Gordon, The renaissance of continuous-flow peptide synthesis – an abridged account of solid and solution-based approaches. *Org. Biomol. Chem.* **16**, 180–196 (2018).
16. T. F. Jamison, G. Koch, Eds., in *Flow Chemistry in Organic Synthesis* (Georg Thieme Verlag, Stuttgart, ed. 2018, 2018; <http://www.thieme-connect.de/products/ebooks/lookinside/10.1055/sos-SD-228-00230>), *Science of Synthesis.*
355
17. A. J. Mijalis, D. A. Thomas III, M. D. Simon, A. Adamo, R. Beaumont, K. F. Jensen, B. L. Pentelute, A fully automated flow-based approach for accelerated peptide synthesis. *Nature Chemical Biology.* **13**, 464 (2017).
18. S. K. Mong, A. A. Vinogradov, M. D. Simon, B. L. Pentelute, Rapid Total Synthesis of DARPin pE59 and Barnase. *ChemBioChem.* **15**, 721–733 (2014).
360
19. S. Fuse, Y. Otake, H. Nakamura, Peptide Synthesis Utilizing Micro-flow Technology. *Chemistry—An Asian Journal.* **13**, 3818–3832 (2018).
20. S. Fuse, Y. Mifune, T. Takahashi, Efficient Amide Bond Formation through a Rapid and Strong Activation of Carboxylic Acids in a Microflow Reactor. *Angewandte Chemie International Edition.* **53**, 851–855 (2014).
365
21. L. K. Spare, V. Laude, D. G. Harman, J. R. Aldrich-Wright, C. P. Gordon, An optimised approach for continuous-flow solid-phase peptide synthesis utilising a rudimentary flow reactor. *Reaction Chemistry & Engineering.* **3**, 875–882 (2018).
22. E. T. Sletten, M. Nuño, D. Guthrie, P. H. Seeberger, Real-time monitoring of solid-phase peptide synthesis using a variable bed flow reactor. *Chem. Commun.* (2019),
370 doi:10.1039/C9CC08421E.
23. M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. Kent, In situ neutralization in Boc-chemistry solid phase peptide synthesis: Rapid, high yield assembly of difficult sequences. *International journal of peptide and protein research.* **40**, 180–193 (1992).
24. T. Michels, R. Dölling, U. Haberkorn, W. Mier, Acid-Mediated Prevention of Aspartimide Formation in Solid Phase Peptide Synthesis. *Org. Lett.* **14**, 5218–5221 (2012).
375
25. M. Mergler, F. Dick, B. Sax, P. Weiler, T. Vorherr, The aspartimide problem in Fmoc-based SPPS. Part I. *Journal of Peptide Science.* **9**, 36–46 (2003).
26. S. A. Palasek, Z. J. Cox, J. M. Collins, Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. *Journal of Peptide Science.* **13**,
380 143–148 (2007).

27. S. Luisier, M. Avital-Shmilovici, M. A. Weiss, S. B. H. Kent, Total chemical synthesis of human proinsulin. *Chem. Commun.* **46**, 8177–8179 (2010).
- 385 28. J. Schneider, S. B. Kent, Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. *Cell.* **54**, 363–368 (1988).
29. E. C. B. Johnson, E. Malito, Y. Shen, D. Rich, W.-J. Tang, S. B. H. Kent, Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease. *J. Am. Chem. Soc.* **129**, 11480–11490 (2007).
- 390 30. A. Brik, C.-H. Wong, HIV-1 protease: mechanism and drug discovery. *Organic & biomolecular chemistry.* **1**, 5–14 (2003).
31. J. D. Oliner, J. A. Pietenpol, S. Thiagalingam, J. Gyuris, K. W. Kinzler, B. Vogelstein, Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature.* **362**, 857–860 (1993).
- 395 32. D. Bresson, L. Togher, E. Rodrigo, Y. Chen, J. A. Bluestone, K. C. Herold, M. von Herrath, Anti-CD3 and nasal proinsulin combination therapy enhances remission from recent-onset autoimmune diabetes by inducing Tregs. *The Journal of clinical investigation.* **116**, 1371–1381 (2006).
33. E. Gasser, C. P. Moutos, M. Downes, R. M. Evans, FGF1—a new weapon to control type 2 diabetes mellitus. *Nature Reviews Endocrinology.* **13**, 599 (2017).
- 400 34. C. B. Anfinsen, Principles that Govern the Folding of Protein Chains. *Science.* **181**, 223 (1973).
35. S. B. Kent, P. F. Alewood, *Current Opinion in Chemical Biology*, in press, doi:10.1016/j.cbpa.2014.09.037.
- 405 36. R. W. Hartley, Barnase and barstar: two small proteins to fold and fit together. *Trends in biochemical sciences.* **14**, 450–454 (1989).
37. L. Serrano, J. T. Kellis Jr, P. Cann, A. Matouschek, A. R. Fersht, The folding of an enzyme: II. Substructure of barnase and the contribution of different interactions to protein stability. *Journal of molecular biology.* **224**, 783–804 (1992).
- 410 38. J. T. Kellis, K. Nyberg, A. R. Fersht, Energetics of complementary side chain packing in a protein hydrophobic core. *Biochemistry.* **28**, 4914–4922 (1989).
39. U. Arnold, M. P. Hinderaker, J. Köditz, R. Golbik, R. Ulbrich-Hofmann, R. T. Raines, Protein Prosthesis: A Nonnatural Residue Accelerates Folding and Increases Stability. *J. Am. Chem. Soc.* **125**, 7500–7501 (2003).
40. J. Knowles, Tinkering with enzymes: what are we learning? *Science.* **236**, 1252 (1987).

- 415 41. B. R. Kelemen, T. A. Klink, M. A. Behike, S. R. Eubanks, P. A. Leland, R. T. Raines, Hypersensitive substrate for ribonucleases. *Nucleic acids research*. **27**, 3696–3701 (1999).
42. M. V. Toth, G. R. Marshall, A simple, continuous fluorometric assay for HIV protease. *International Journal of Peptide and Protein Research*. **36**, 544–550 (1990).
43. H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, Purification and
420 characterization of sortase, the transpeptidase that cleaves surface proteins of
Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci USA*. **96**,
12424 (1999).
44. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao, B. L. Pentelute, Protein thioester
425 synthesis enabled by sortase. *Journal of the American Chemical Society*. **134**, 10749–10752
(2012).
45. R. L. Policarpo, H. Kang, X. Liao, A. E. Rabideau, M. D. Simon, B. L. Pentelute, Flow-
based enzymatic ligation by sortase A. *Angewandte Chemie International Edition*. **53**,
9203–9208 (2014).
46. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, Total synthesis and biochemical
430 characterization of mirror image barnase. *Chem. Sci*. **6**, 2997–3002 (2015).
47. F. Touti, Z. P. Gates, A. Bandyopdhyay, G. Lautrette, B. L. Pentelute, In-solution
enrichment identifies peptide inhibitors of protein–protein interactions. *Nature Chemical
Biology*. **15**, 410–418 (2019).
48. P. Chène, Inhibition of the p53-MDM2 Interaction: Targeting a Protein-Protein Interface.
435 *Mol Cancer Res*. **2**, 20 (2004).
49. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, Sequence-Controlled Polymers. *Science*.
341, 1238149 (2013).
50. S. J. Wheelan, A. Marchler-Bauer, S. H. Bryant, Domain size distributions can predict
domain boundaries. *Bioinformatics*. **16**, 613–618 (2000).
- 440 51. J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, Enhancing Protein Stability with
Genetically Encoded Noncanonical Amino Acids. *J Am Chem Soc*. **140**, 15997–16000
(2018).
52. H. Xiao, F. Nasertorabi, S. Choi, G. W. Han, S. A. Reed, R. C. Stevens, P. G. Schultz,
445 Exploring the potential impact of an expanded genetic code on protein function. *Proc Natl
Acad Sci USA*. **112**, 6961 (2015).
53. E. G. Baker, G. J. Bartlett, K. L. Porter Goff, D. N. Woolfson, Miniprotein Design: Past,
Present, and Prospects. *Acc. Chem. Res*. **50**, 2085–2092 (2017).
54. J.-P. Couso, P. Patraquim, Classification and function of small open reading frames. *Nature
Reviews Molecular Cell Biology*. **18**, 575–589 (2017).

- 450 55. A. L. Okorokov, R. W. Hartley, K. I. Panov, An Improved System for Ribonuclease Ba
Expression. *Protein Expression and Purification*. **5**, 547–552 (1994).
56. M. C. R. Shastry, J. B. Udgaonkar, The Folding Mechanism of Barstar: Evidence for
Multiple Pathways and Multiple Intermediates. *Journal of Molecular Biology*. **247**, 1013–
1027 (1995).
- 455 57. R. Loewenthal, J. Sancho, A. R. Fersht, Fluorescence spectrum of barnase: contributions of
three tryptophan residues and a histidine-related pH dependence. *Biochemistry*. **30**, 6775–
6779 (1991).
58. E. M. Nicholson, J. M. Scholtz, Conformational stability of the Escherichia coli HPr
protein: test of the linear extrapolation method and a thermodynamic characterization of
460 cold denaturation. *Biochemistry*. **35**, 11369–11378 (1996).
59. I. W. Windsor, R. T. Raines, Fluorogenic Assay for Inhibitors of HIV-1 Protease with Sub-
picomolar Affinity. *Cell*. **54**, 363–368 (1988).
60. C. Zhan, K. Varney, W. Yuan, L. Zhao, W. Lu, Interrogation of MDM2 Phosphorylation in
p53 Activation Using Native Chemical Ligation: The Functional Role of Ser17
465 Phosphorylation in MDM2 Reexamined. *J. Am. Chem. Soc.* **134**, 6855–6864 (2012).

Acknowledgments: We thank Prof. T. Jamison, Dr. H.U. Stilz, Dr. L.F. Iversen, Dr. K. Little and
Dr. D. Lundsgaard for productive discussions and administrative support. Furthermore, we
acknowledge the participants of the 26th American Peptide Symposium and the 8th Chemical
470 Protein Synthesis Meeting, especially Prof. P.D. Dawson, Prof. R.T. Raines and Prof. S.B.H. Kent,
for helpful discussions and for suggesting additional experiments. We also thank Dr. E.D. Evans,
Dr. F.W.W. Hartrampf and R.L. Holden for careful proofreading of the manuscript. We are
grateful to Dr. N. Truex for providing recombinant sortase A. Finally, we acknowledge C. M. T.
Hartrampf for designing Fig. 1A; **Funding:** Financial support for this project was provided by
475 Novo Nordisk. A.S., A.E.C., C.K.S. gratefully acknowledge support from the National Science
Foundation Graduate Research Fellowship under Grant No. 1122374; A.E.C. is additionally
supported by an MIT Dean of Science Fellowship. **Author contributions:** N.H., T.E.N. and B.L.P.
conceptualized the research; N.H., M.P. and S.L. optimized synthesis conditions; A.J.M. provided
the software for UV data analysis and M.S. built the AFPS used in this report; M.P., A.J.C. and
480 C.J. performed comparison of AFPS with traditional SPPS methods; N.H., A.S., M.P., A.J.C.,
A.E.C., S.H., S.A., C.K.S., and A.J.Q. synthesized, purified and analyzed protein samples; N.H.,
Z.P.G. and B.L.P. conceptualized folding and biological evaluation of the synthetic proteins; N.H.,
A.S., M.P., Z.P.G., A.J.C. and X.Y. performed biological evaluation and expression of
recombinant proteins; N.H., Z.P.G, A.L. and B.L.P wrote the manuscript with input of all co-
485 authors; **Competing interests:** B.L.P. is a co-founder of Amide Technologies and Resolute Bio.
Both companies focus on the development of protein and peptide therapeutics; **Data and
materials availability:** All data is available in the main text or the supplementary materials.

Supplementary Materials:

Materials and Methods

490 Supplementary Text
 Figs. S1 to S26
 Tables S1 to S16

495

200210 Manuscript Final_ChemRxiv.pdf (2.19 MiB)

[view on ChemRxiv](#) • [download file](#)

Supplementary Materials for

Synthesis of Proteins by Automated Flow Chemistry

Authors: N. Hartrampf¹, A. Saebi^{1†}, M. Poskus^{1†}, Z. P. Gates¹, A. J. Callahan¹, A. E. Cowfer¹, S. Hanna¹, S. Antilla¹, C. K. Schissel¹, A. J. Quartararo¹, X. Ye¹, A. J. Mijalis^{1,2}, M. D. Simon¹, A. Loas¹, S. Liu^{1,3}, C. Jessen⁴, T. E. Nielsen⁴ and B. L. Pentelute^{1*}

*Correspondence to: blp@mit.edu

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S25
Tables S1 to S14

1 Table of Contents

1	TABLE OF CONTENTS.....	1
2	MATERIALS AND GENERAL METHODS.....	3
2.1	REAGENTS AND SOLVENTS.....	3
2.2	AUTOMATED FLOW PEPTIDE SYNTHESIS SET-UP	3
2.3	CLEAVAGE PROTOCOL	4
2.4	LIQUID CHROMATOGRAPHY–MASS SPECTROMETRY (LCMS)	5
2.5	ANALYTICAL HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)	6
2.6	MASS-DIRECTED REVERSED-PHASE HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (RP-HPLC).....	6
2.7	DETERMINATION OF PROTEIN CONCENTRATIONS.....	7
3	DEVELOPMENT OF A GENERAL AFPS PROTOCOL.....	8
3.1	INITIAL AFPS CONDITIONS	8
3.2	TEMPERATURE.....	9
3.3	SOLVENTS AND REAGENT CONCENTRATIONS.....	11
3.4	COUPLING BASES.....	11
3.5	DEPROTECTION SOLUTIONS	13
3.6	ACTIVATORS.....	14
3.7	TREATMENT OF INDIVIDUAL AMINO ACIDS	15
3.8	ASPARTIMIDE FORMATION	16
3.9	EPIMERIZATION EXPERIMENTS.....	18
3.9.1	<i>Temperatures for activation in 10' heating loop</i>	<i>18</i>
3.9.2	<i>Temperatures for activation in 5' heating loop</i>	<i>19</i>
3.9.3	<i>Extended coupling time.....</i>	<i>22</i>
3.9.4	<i>Multiple coupling cycles.....</i>	<i>22</i>
3.10	FINAL SYNTHESIS PROTOCOL	25
3.10.1	<i>GLP-1</i>	<i>26</i>
3.10.2	<i>NN92</i>	<i>27</i>
3.10.3	<i>Amyloid beta [1-42]</i>	<i>28</i>
4	SYNTHESIS OF LONGER PEPTIDES - COMPARISON BETWEEN AFPS AND TRADITIONAL SYNTHESIS METHODS	29
4.1	SYNTHESIS ON AFPS.....	29
4.1.1	<i>GLP-1</i>	<i>29</i>
4.1.2	<i>Proinsulin.....</i>	<i>31</i>
4.1.3	<i>HIV protease (wild type sequence)</i>	<i>33</i>
4.2	SYNTHESIS ON COMMERCIALY AVAILABLE SYNTHESIZERS AT R.T., 70 °C AND 90 °C	35
4.2.1	<i>GLP-1</i>	<i>37</i>
4.2.2	<i>Proinsulin.....</i>	<i>38</i>

4.2.3	<i>HIV protease (wild type sequence)</i>	41
5	AFPS SYNTHESIS OF PEPTIDES IN THE LENGTH RANGE OF SINGLE DOMAIN PROTEINS...	43
5.1	MDM2 ^[25-109]	43
5.2	BARSTAR	45
5.3	COLLAGEN	47
5.4	HIV-1 PROTEASE (KENT SEQUENCE)	49
5.5	BARNASE	51
5.6	BARNASE (R110F _{BR})	53
5.7	MDM2 ^[1-118]	55
5.8	LYSOZYME	57
5.9	FGF1	59
5.10	SORTASE A ₅₉₋₂₀₆ ;P94S/D160N/K196T.....	61
6	FOLDING, PURIFICATION AND CHARACTERIZATION OF SYNTHETIC PROTEINS	63
6.1	BARNASE AND BARSTAR.....	63
6.1.1	<i>Expression and SEC purification of recombinant barnase</i>	63
6.1.2	<i>Folding and SEC purification of synthetic WT barnase</i>	64
6.1.3	<i>Folding and SEC purification of synthetic barnase R110F_{BR}</i>	65
6.1.4	<i>Folding and IEC purification of synthetic WT Barstar</i>	66
6.1.5	<i>RNA hydrolysis assay</i>	67
6.1.6	<i>Fluorogenic RNase activity assay</i>	69
6.1.7	<i>Chemical denaturation assay</i>	75
6.2	HIV-1 PROTEASE	77
6.2.1	<i>Folding of synthetic HIV-1 protease (Kent sequence)</i>	77
6.2.2	<i>Fluorogenic protease activity assay</i>	77
6.2.3	<i>Substrate specificity assay</i>	79
6.3	SORTASE A ^[59-206] ;P94S/D160N/K196T	81
6.3.1	<i>Expression and purification of recombinant sortase A[59-206];P94S/D160N/K196T</i> 81	
6.3.2	<i>Folding and SEC purification of synthetic sortase A[59-206];P94S/D160N/K196T ..</i>	82
6.3.3	<i>Semiquantitative activity assay</i>	83
6.4	MDM2 ^[1-118]	87
6.4.1	<i>Folding of synthetic MDM2^[1-118]</i>	87
6.4.2	<i>Preparation of biotinylated p53(15-29)</i>	87
6.4.3	<i>Assay of MDM2 binding activity, by biolayer interferometry</i>	87

2 Materials and general methods

2.1 Reagents and solvents.

All reagents were purchased and used as received. Fmoc-protected amino acids (Fmoc-Ala-OHxH₂O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp(Ot-Bu)-OH; Fmoc-Cys(Trt)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(Ot-Bu)-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(But)-OH; Fmoc-Thr(*t*-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(*t*-Bu)-OH; Fmoc-Val-OH) and Fmoc-His(Boc)-OH were purchased from the Novabiochem-line from Sigma Millipore; Fmoc-p-bromo-Phe-OH was purchased from Bachem (product number 4042637); O-(7-azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%), *N,N,N',N'*-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluoro-phosphate (HBTU, ≥97.0%), O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU, ≥97.0%) and (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexa-fluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade *N,N*-dimethylformamide (DMF) was purchased from EMD Millipore (DX1732-1). *N*-Methyl-2-pyrrolidone (NMP, ≥99.0%) was purchased from Sigma-Aldrich and dried over PPT Pure Process Technology solvent system. AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706), Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (HPLC grade, ≥99.0%), triisopropylsilane (≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), dimethyl sulfoxide (DMSO, HPLC grade, ≥99.7%), piperazine (≥99.0%) and 1,2-ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. Tetrafluoroethylene (TFE, extra pure, ≥99.8%) was purchased from Acros. Anisole (purum, GC grade, ≥99.0%) and 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU, GC grade, ≥99.0%) were purchased from Fluka. 1-Hydroxybenzotriazole hydrate (HOBt, ≥97.0%) was purchased from Peptides International. H-Rink Amide (0.49 mmol/g and 0.18 mmol/g loading) and HMPB ChemMatrix polyethylene glycol (0.45 mmol/g loading) resin were purchased from PCAS Biomatrix. Water was deionized using a Milli-Q Reference water purification system (Millipore). Nylon 0.22 μm syringe filters were TISCH brand SPEC17984.

2.2 Automated flow peptide synthesis set-up

All peptides were synthesized on an automated-flow system, which was built in the Pentelute lab (“Amidator”), which is similar to the published AFPS system.⁽¹⁷⁾ Capitalized letters refer to L-amino acids, uncapitalized letters refer to D-amino acids.

Unless otherwise noted, the following settings were used for peptide synthesis: flow-rate = 40 mL/min, temperature = 90 °C (loop) and 85–90 °C (reactor). The 50 ml/min pump head pumps 400 μL of liquid per pump stroke; the 5 mL/min pump head pumps 40 μL of liquid per pump stroke. The standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures 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 mL/min pump head pumps DIEA. The first two pumps are activated for 8 pumping strokes in order to prime the coupling agent and amino acid before the DIEA pump is activated. The three pumps are then actuated together for a period of 7 pumping strokes, after which the activating agent pump and amino acid pump are switched using a rotary valve to select DMF. The three pumps are actuated together for a final 8 pumping strokes, after which the DIEA pump is shut off and the other two pumps continue to wash the resin for another 40 pump strokes.

During the deprotection step, two HPLC pumps are used. Using a rotary valve, one HPLC pump selects deprotection stock solution and DMF. The pumps are activated for 13 pump strokes. Both solutions are mixed in a 1:1 ratio. Next, the rotary valves select DMF for both HPLC pumps, and the resin is washed for an additional 40 pump strokes. The coupling–deprotection cycle is repeated for all additional monomers.

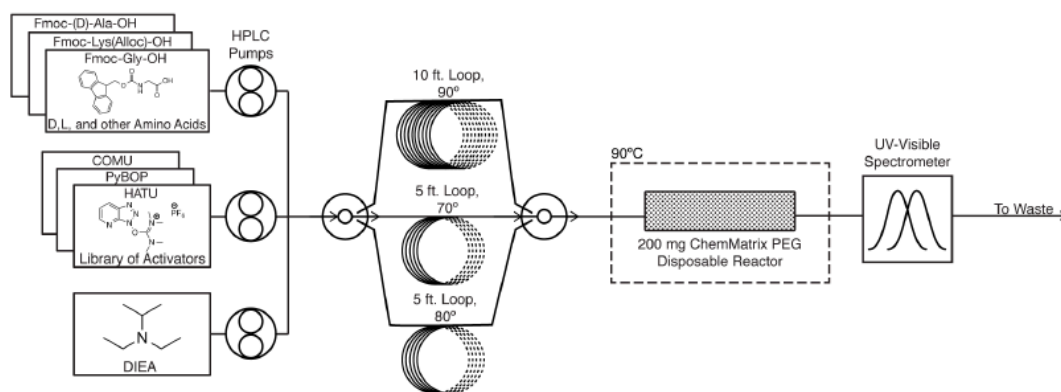


Fig. S1. Schematic AFPS setup used in this study. Modified from (17)

2.3 Cleavage Protocol

After synthesis, the peptidyl resin was washed with dichloromethane (3 x 5 mL), dried in a vacuum chamber, and weighed. 50% of the resin was transferred into a 50 mL conical polypropylene tube. For cleavage of peptides and proteins we used two different protocols:

- A) **Peptides¹**: Approximately 3 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The tube was kept at room temperature for 2 h.
- B) **Proteins and Cys-containing peptides (except for GCF peptide)**: Approximately 3 mL of cleavage solution (82.5% TFA, 5% water, 5% phenol, 5% thioanisole, 2.5% EDT) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The tube was kept on a nutating mixer at room temperature for 4 h.

Ice cold diethyl ether (45 mL) was added to the cleavage mixture and the precipitate was collected by centrifugation and triturated twice more with cold diethyl ether (45 mL). The supernatant was discarded. Residual ether was allowed to evaporate and the peptide was dissolved in 50% acetonitrile in water with 0.1% TFA (long peptides were dissolved 70% acetonitrile in water with 0.1% TFA). The peptide solution was filtrated with a Nylon 0.22 μm syringe filter and frozen, lyophilized until dry, and weighed.

¹ **Note:** for short peptides, which were soluble in ether (e.g. epimerization test peptides GCF and FHL), the trituration step was skipped and TFA was evaporated before addition of 50% acetonitrile in water with 0.1% TFA.

2.4 Liquid chromatography–mass spectrometry (LCMS)

For mass analysis, the filtered peptide solution (10 μ L of a 1mg/mL solution) was diluted in 50% acetonitrile in water with 0.1% TFA (90 μ L) to a final concentration approximately 0.1 mg/mL. LC-MS chromatograms and associated high resolution mass spectra were acquired using an Agilent 6520 Accurate-Mass Q-TOF LCMS (abbreviated as 6520) or an Agilent 6550 iFunnel Q-TOF LCMS system (abbreviated as 6550). Solvent compositions used in the LC-MS are 0.1% formic acid in H₂O (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS methods were used:

- **1-61% B over 9 min, Zorbax C3 column (6520)**
LC conditions: Zorbax 300SB-C3 column: 2.1 \times 150 mm, 5 μ m, column temperature: 40 $^{\circ}$ C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z .
- **1-91% B over 9 min, Zorbax C3 column (6520)**
LC conditions: Zorbax 300SB-C3 column: 2.1 \times 150 mm, 5 μ m, column temperature: 40 $^{\circ}$ C, gradient: 0-2 min 1% B, 2-11 min 1-91% B, 11-12 min 91-95% B; flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z .
- **1-61% B over 15 min, Phenomenex Jupiter C4 column (6550)**
LC conditions: Phenomenex Jupiter C4 column: 1.0 \times 150 mm, 5 μ m, column temperature: 40 $^{\circ}$ C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 20 min. MS is on from 4 to 12 min.
MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z .
- **1-61% B over 33 min, Phenomenex Jupiter C4 column (6550)**
LC conditions: Phenomenex Jupiter C4 column: 1.0 \times 150 mm, 5 μ m, column temperature: 40 $^{\circ}$ C, gradient: 0-2 min 1% B, 2-30 min 1-91% B, 30-34 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 38 min. MS is on from 4 to 30 min.
MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z .
- **1-61% B over 15 min, Zorbax C3 column (6550)**
LC conditions: Zorbax 300SB-C3 column: 2.1 \times 150 mm, 5 μ m, column temperature: 40 $^{\circ}$ C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z .

- **1-91% B over 20 min, Phenomenex Jupiter C4 column (6550)**

LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-18 min 1-91% B, 18-21 min 91% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 25 min. MS is on from 4 to 18 min.

MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 *m/z*.

- **1-61% B over 17 min, Luna C18 column (6550)**

LC conditions: Phenomenex Luna C18 column: 0.5 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-14 min 1-61% B, 14-18 min 61-91% B; flow rate: 0.1 mL/min. A final 5-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 23 min. MS is on from 4 to 14 min.

MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 *m/z*.

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.06.00 with BioConfirm Software.

2.5 Analytical high-performance liquid chromatography (HPLC)

For determination of purity by HPLC, the filtered peptide solution was diluted in 50% acetonitrile in water with 0.1% TFA (100 μL) to a final concentration of approximately 1.0 mg/mL. Peptide samples containing cysteines were diluted in 6M Guanidinium chloride containing 100 mM DTT². The samples were analyzed on Agilent Technologies 1200 Series, which was computer-controlled through Agilent ChemStation software.

For standard analysis of all peptide samples, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C3 column (2.1 mm × 150 mm, 5-μm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, gradients of 1% B per minute ramped up over 60 min at a flow rate of 0.4 mL/min. Gradients either started at 1% B (annotated as “1–61% B over 60 min”) or 5% B (annotated as “5–65% B over 60 min”). A final 3-min hold was performed. The total method time was 66 min. Crude HPLC purities were determined by manual integration of all signals in the area of 5–60 min.

For analysis of epimerization, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C18 column (2.1 mm × 150 mm, 5-μm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, a gradient of 2% B per minute ramped up over 20 min at a flow rate of 0.4 mL/min (annotated as “1–41% B over 20 min”). A final 3-min hold was performed. The total method time was 30 min. Epimer ratios were determined as described in section 2.9.

2.6 Mass-directed reversed-phase high performance liquid chromatography (RP-HPLC)

For RP-HPLC purification, the lyophilized peptide sample was dissolved in the gradient starting concentration (e.g. 5% acetonitrile in water with 0.1% TFA) or 6M Guanidinium chloride containing 100 mM DTT. All samples filtrated with a Nylon 0.22 μm syringe filter prior to

² Except for GCF peptides for epimerization studies

purification. For all HPLC purifications, a gradient of acetonitrile with 0.1 % TFA additive (solvent B) and water with a 0.1% TFA additive (solvent A) was used. All samples were purified on a semipreparative Agilent Zorbax 300SB-C3 column (9.4 mm × 250 mm, 5- μ m particle size) at a flow rate of 4 mL/min unless otherwise noted. Specific purification conditions such as column temperature and gradient are specified for each case.

2.7 Determination of protein concentrations.

The concentration of proteins in solution was determined by absorbance at 280 nm (A_{280}) or bicinchoninic acid (BCA) assay. The detailed protocol is as followed:

1. **A_{280} :** The concentration of proteins was calculated using the Beer-Lambert law by measuring the absorbance of the protein sample at 280 nm. The absorbance was measured by averaging at least two independent readings of the same sample on a BioTek Synergy HT plate reader outfitted with a BioTek Take 3 micro-volume plate. Unless indicated otherwise, the molar extinction coefficient of the protein of interest was estimated based on the sequence of the protein via ExPASy Swiss Institute of Bioinformatics - Bioinformatics Resource Portal.
2. **BCA assay:** The concentration of proteins was determined by the Pierce™ Rapid Gold BCA Protein Assay Kit from ThermoFisher Scientific (catalog number A53226) following the manufacturer's instructions. Briefly, nine standards of Albumin protein (BSA) at different concentrations was prepared in addition to at least two dilutions of the desired protein, in triplicates. In a 96-well plate, 10 μ L of each of the protein samples were added. 200 μ L of the Pierce Rapid Gold BCA Working Reagent was then added to the wells using a multichannel pipette. The plate was incubated at room temperature for 5 minutes. The absorbance at 480 nm was then measured using a BioTek Synergy HT plate reader. Finally, the concentration of the desired protein was calculated based on the standard curve generated with the BSA standards, averaged and reported.

3 Development of a general AFPS protocol

General notes: All peptides synthesized for the development of a general synthesis method were synthesized on ChemMatrix RINK amide resin and cleaved after synthesis using the cleavage protocol for peptides (**Section 1.3**). All peptides were analyzed by analytical HPLC and LC-MS.

3.1 Initial AFPS conditions

Parameter	Conditions
Temperature	85–90 °C in reactor, 70 °C in 10' activation loop
Flow Rate	40 mL/min
Coupling step	0.20 M amino acids stock solution in NMP 0.19 M activator stock solution in DMF DIEA (99.7%) as coupling base Coupling conditions: R was coupled with PyAOP and all other amino acids were coupled with HATU (13 pump strokes)
Deprotection step	40% piperidine in DMF (13 pump strokes)
Washing steps	DMF (40 pump strokes)

Table S1. Conditions that were used on the AFPS system “amidator” at the beginning of the optimization process.

NOTE: amino acid, activator and base are mixed in a 10:10:1 ratio in the coupling step and deprotection solution and DMF are mixed in a 1:1 ratio during the deprotection step.

Peptide sequences used for screening purposes:

GLP-1: HAEGTFTSDV SSYLEGQAAK EFIGLWKGR (30 AA)
GHRH: YADAI FTNSY RQVLG QLSAR KLLQD ILSA (29 AA)
NN92: RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)
Amyloid- β : DAEFRHDSGY EVHHQKLVFF AEDVGSNKGK IIGLMVGGVV IA (42 AA)

3.2 Temperature

Entry	Peptide	T (loop)	T (reactor)	Purity by HPLC	Crude Yield
1	GLP-1	70 °C	70 °C	46%	17%
2	GLP-1	70 °C	90 °C	52%	23%
3	GLP-1	90 °C	70 °C	53%	25%
4	GLP-1	90 °C	90 °C	60%	32%
5	GHRH	70 °C	90 °C	83%	60%
6	GHRG	90 °C	70 °C	78%	43%
7	GHRH	90 °C	90 °C	89%	50%

Table S2. Synthesis at 90 °C results in improved synthesis outcome for GLP-1 and GHRH. Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = as indicated in this table; flowrate = 40 mL/min; coupling step = 0.2 M amino acid stock solutions in NMP, 0.19 M activator in DMF, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes).

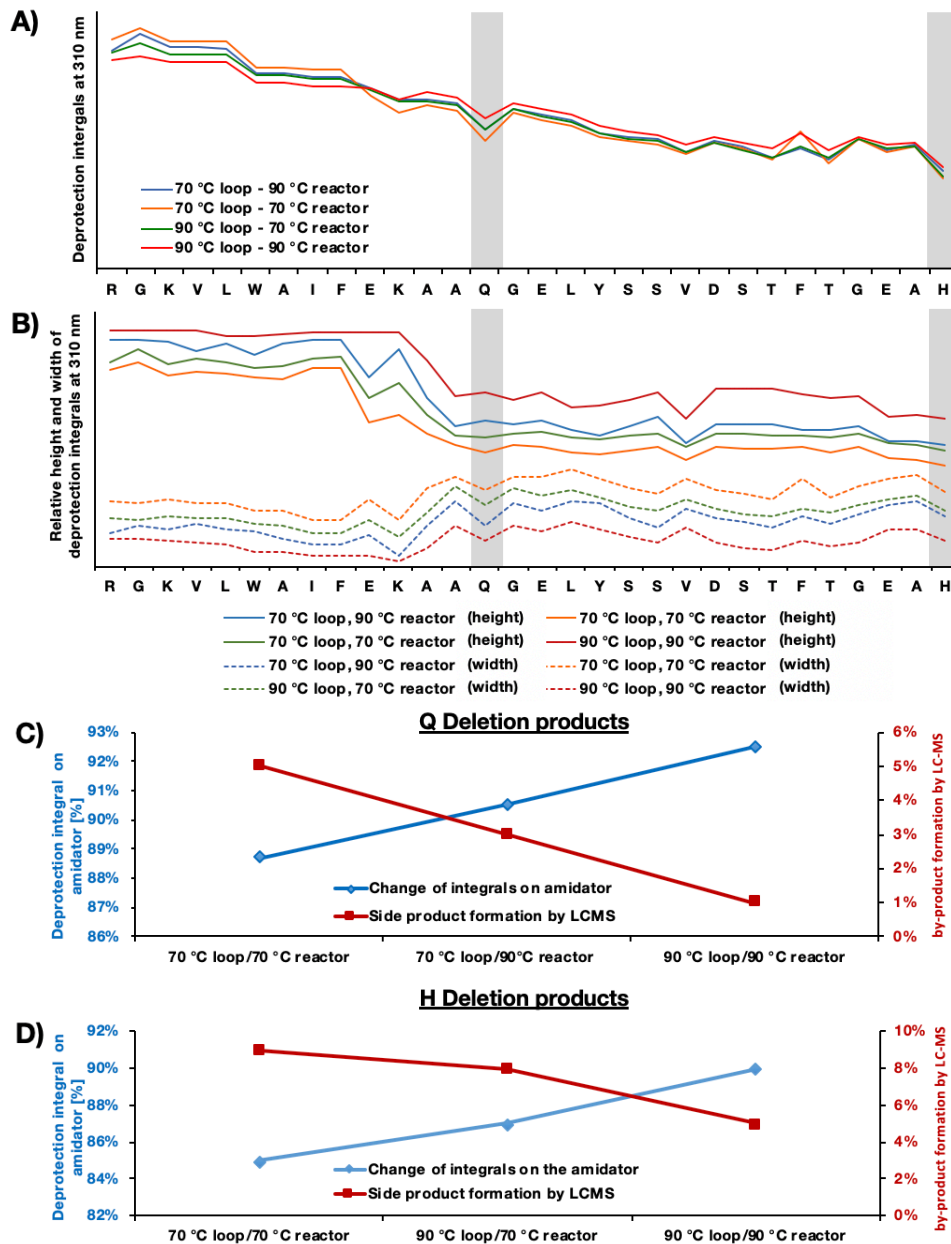


Fig. S2. Change of UV deprotection integrals during the synthesis of GLP-1 at different temperatures. A) Relative change of absolute deprotection signals. Higher temperatures lead to more consistent coupling efficiencies, which correlates to the observation of increased formation of deletion by-products at lower temperature. Highlighted in grey are integrals for glutamine (Q) and histidine (H), two major deletion by-products by LC-MS; B) The shape of the peaks at 90 °C is overall much sharper than at 70 °C, which indicates increased mass transfer. Highlighted in grey are integrals for glutamine (Q) and histidine (H), two major deletion by-products by LC-MS; D) glutamine (Q) deprotection integral versus by-product detected by LC-MS; C) histidine (H) deprotection integral versus by-product detected by LC-MS.

Conclusion: Higher temperature leads to increased coupling efficiency and less deletion by-products. 90 °C was therefore chosen as the optimal temperature for both activation loop and reactor.

3.3 Solvents and reagent concentrations

Different solvent and reagents were evaluated for two temperature settings:

Entry	peptide	Solvent	Amino acid conc.	HPLC purity	Crude yield
1	GLP-1	DMF	0.2M	54%	18%
2	GLP-1	NMP	0.4M	56%	15%
3	GLP-1	NMP	0.2M	46%	17%
4	GLP-1	amine-free DMF	0.4M	65%	13%
5	GLP-1	amine-free DMF	0.2M	61%	11%

Table S3. High reagent concentration lead to increased synthesis outcome for the synthesis of GLP-1. Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 70 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = concentrations and solvents as indicated in the table (“amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage), DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes).

Entry	peptide	Solvent	Amino acid conc.	HPLC purity	Crude yield
1	GLP-1	DMF	0.4M	69%	36%
2	GLP-1	DMF	0.2M	58%	– ³
3	GLP-1	NMP	0.4M	65%	21%
4	GLP-1	NMP	0.2M	53%	21%
5	GLP-1	amine-free DMF	0.4M	71%	13%
6	GLP-1	amine-free DMF	0.2M	55%	23%

Table S4. High reagent concentration lead to increased synthesis outcome for the synthesis of GLP-1. Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = concentrations and solvents as indicated in the table, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes). “Amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: A temperature of 90 °C in the heating loop, concentrations of 0.4 M (amino acids) and 0.38 M (activator) and amine-free DMF as a solvent significantly improve synthesis outcome for GLP-1. Temperature leads to the most significant improvement for GLP-1 synthesis, followed by the concentration of reagents.

3.4 Coupling bases

³ No yield due to partial sample loss during work up

Entry	peptide	Coupling base	HPLC purity	Crude yield
1	GLP-1	DIEA	66%	24%
2	GLP-1	NMM	24%	9%
3	GLP-1	Collidine + HOBT (1.0eq)	5%	2%
4	GLP-1	Anisole (7%, v/v)	56%	21%

Table S5. DIEA is superior to other bases for the synthesis of GLP-1. Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, coupling base as indicated in the table; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = amine-free DMF (40 pump strokes). “Amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: DIEA is the most effective coupling base on AFPS.

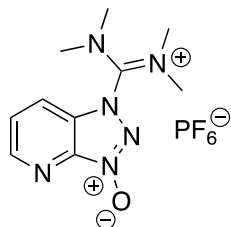
3.5 Deprotection solutions

Entry	peptide	base	solvent	Coupling agent	HPLC purity	Crude yield
1	GLP-1	Piperidine (40%, v/v)	DMF	HATU	69%	33%
2	GLP-1	Piperidine (40%, v/v)	DMF + TFE (1%)	HATU	75%	31%
3	GLP-1	piperazine (6%, w/v)	DMF	HATU	41%	14%
4	GLP-1	piperazine (6%, w/v)**	DMF	HATU	59%	22%
5	GLP-1	piperazine (8%, w/v)	DMF	HATU	43%	12%
6	GLP-1	DBU (2%)**	DMF	HATU	78%	33%
7	GLP-1	DBU (4%)	DMF	HATU	66%	31%
8	GLP-1	DBU (4%)**	DMF	HATU	66%	31%
9	GLP-1	DBU (8%)	DMF	HATU	61%	21%
10	GLP-1	DBU (2%) + piperazine (6%)	DMF	HATU	75%	34%
11	GLP-1	DBU (2%) + piperazine (6%)	DMF	PyAOP	74%	35%
12	GLP-1	DBU (2%) + piperazine (6%)	DMF + TFE (1%)	PyAOP	76%	41%
13	GLP-1	Piperidine (40%) + FA (2%)	DMF	HATU	77%	30%
14	GLP-1	Piperidine (40%) + FA (2%)	DMF	PyAOP	78%	36%
15	GLP-1	Piperidine (40%) + FA (2%)	DMF + TFE (1%)	PyAOP	74%	34%
16	GLP-1	Piperidine (40%) + 0.1M HOBt	DMF	HATU	68%	29%

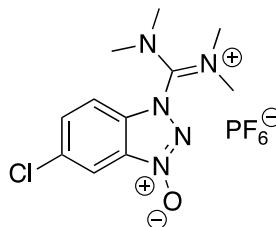
Table S6. Multiple deprotection solutions show improved synthesis outcome. (NOTE: DBU is significantly more basic than piperidine and all solutions will be tested in epimerization studies again.) Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = as indicated in the table (40 pump strokes); washing steps = amine-free DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage. ** = 80 pump strokes for deprotection.

Conclusion: Piperidine with formic acid additive improves synthesis outcome.

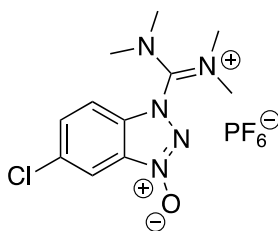
3.6 Activators



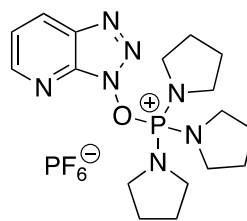
HATU



HCTU



HBTU



PyAOP

Entry	Peptide	Coupling agent	Purity by HPLC	Crude Yield
1	GLP-1	HATU ^{a)}	71%	24%
2	GLP-1	HBTU ^{a)}	66%	20%
3	GLP-1	HCTU ^{a)}	69%	26%
4	GLP-1	PyAOP	76%	22%

Table S7. PyAOP was identified as the most effective activator for the synthesis of GLP-1. Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, all other amino acids were coupled as described on this table (13 pump strokes); deprotection = 40% piperidine + 2% formic acid in DMF (40 pump strokes); washing steps = amine-free DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: PyAOP performed the best in our screening and was therefore used moving forward.

3.7 Treatment of individual amino acids

For an amino acid-based optimization we used statistical data on amide bond formation that was indirectly obtained from the deprotection integrals on the AFPS system (real time UV data at 310 nm). Data from the synthesis of multiple peptides (excluding GLP-1 synthesis traces) was used to improve synthesis conditions as demonstrated in **Table S8**.

	A	D	E	F	G	H	I	K	L	M
initial coupling conditions	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
before individual amino acid optimization	97,9%	101,2%	99,9%	100,8%	103,2%	99,6%	99,6%	99,9%	99,7%	100,4%
changes to the protocol	PyAOP, 26	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
after individual amino acid optimization	101,4%	98,1%	99,0%	100,6%	101,0%	98,7%	98,9%	98,4%	98,3%	99,8%

	N	P	Q	R	S	T	V	W	Y
initial coupling conditions	HATU, 13	HATU, 13	HATU, 13	PyAOP, 26	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
before individual amino acid optimization	94,1%	99,8%	97,8%	96,2%	98,2%	98,3%	97,9%	98,7%	100,5%
changes to the protocol	PyAOP, 26	HATU, 13	PyAOP, 26	PyAOP, 26	HATU, 21	PyAOP, 26	PyAOP, 26	HATU, 13	HATU, 13
after individual amino acid optimization	97,4%	100,2%	100,5%	98,3%	101,3%	100,1%	99,7%	100,3%	99,4%

Table S8. UV integrals for deprotection signals before and after amino acid-based optimization of coupling protocol shows increased integral values for optimized residues. NOTE Un-optimized amino acid coupling resulted in different integral values because coupling is residue and sequence-specific. This table is intended to display deprotection (and thereby coupling) trends rather than absolute values.

Finally, GLP-1 synthesis under various coupling conditions was examined as demonstrated in **Table S9**. Since Valine and Histidine deletions were the most abundant by-products of GLP1 syntheses, these residues were coupled with various conditions. After analyzing statistical peptide data for the various set of conditions (**Table S8**), GLP-1 was resynthesized with the optimal coupling conditions.

Entry	Peptide	Coupling conditions	Purity by HPLC	Crude Yield
1	GLP-1	H, V double coupling time	67%	30%
2	GLP-1	H, V double coupling time 1% TFE in Q,A,K solution	64%	27%
3	GLP-1	H, V double coupling time (deprotected by 2% DBU+6%Piperizine)	66%	30%
4	GLP-1	S,A = HATU (21 strokes) Q, N, V, T, R = PyAOP (21 strokes)	70%	24%

Table S9. Set of developed coupling conditions from coupling statistics proves to increase synthesis outcome more than optimization based on by-product formation (by LCMS). Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, all other amino acids were coupled as described on this table (13 pump strokes); deprotection = 40% piperidine + 2% formic acid in DMF (40 pump strokes); washing steps = amine-free. DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage.

3.8 Aspartimide formation

Optimization effort was put into reducing aspartimide formation in sequences containing Asp. Peptide NN92 was used as a test sequence since it showed significant aspartimide formation in initial experiments. To determine the ratio of aspartimide to peptide, the integrals of the extracted ion count [EIC] of both compounds was integrated and compared as shown in **Fig. S3**.

Sequence: RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)

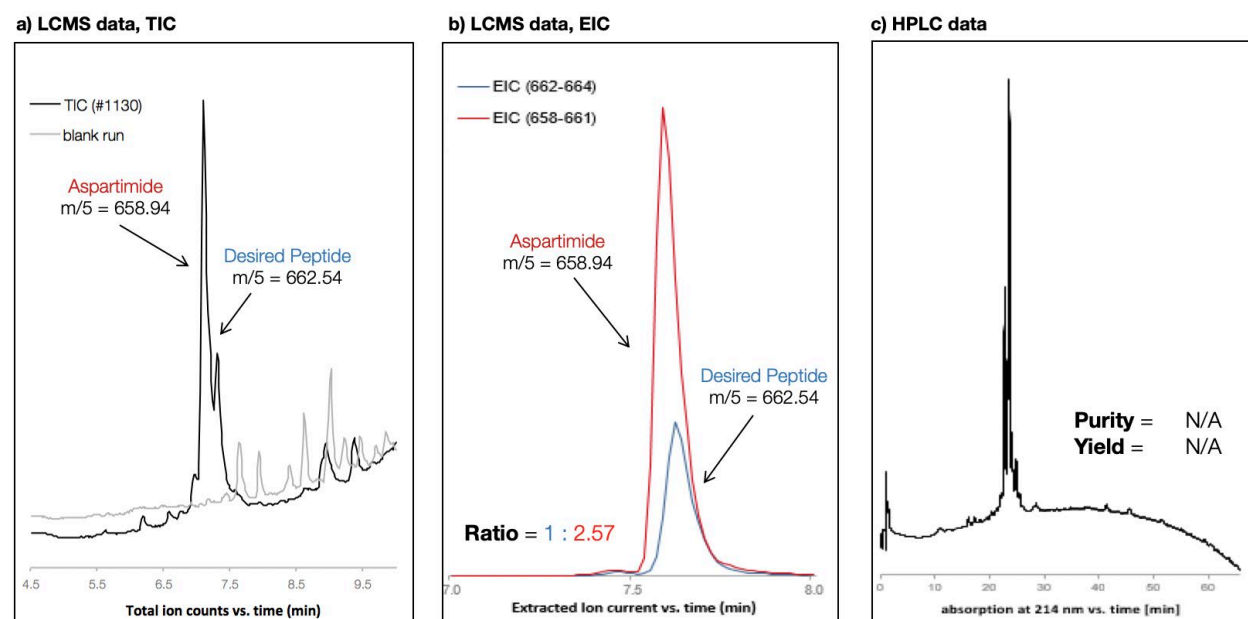
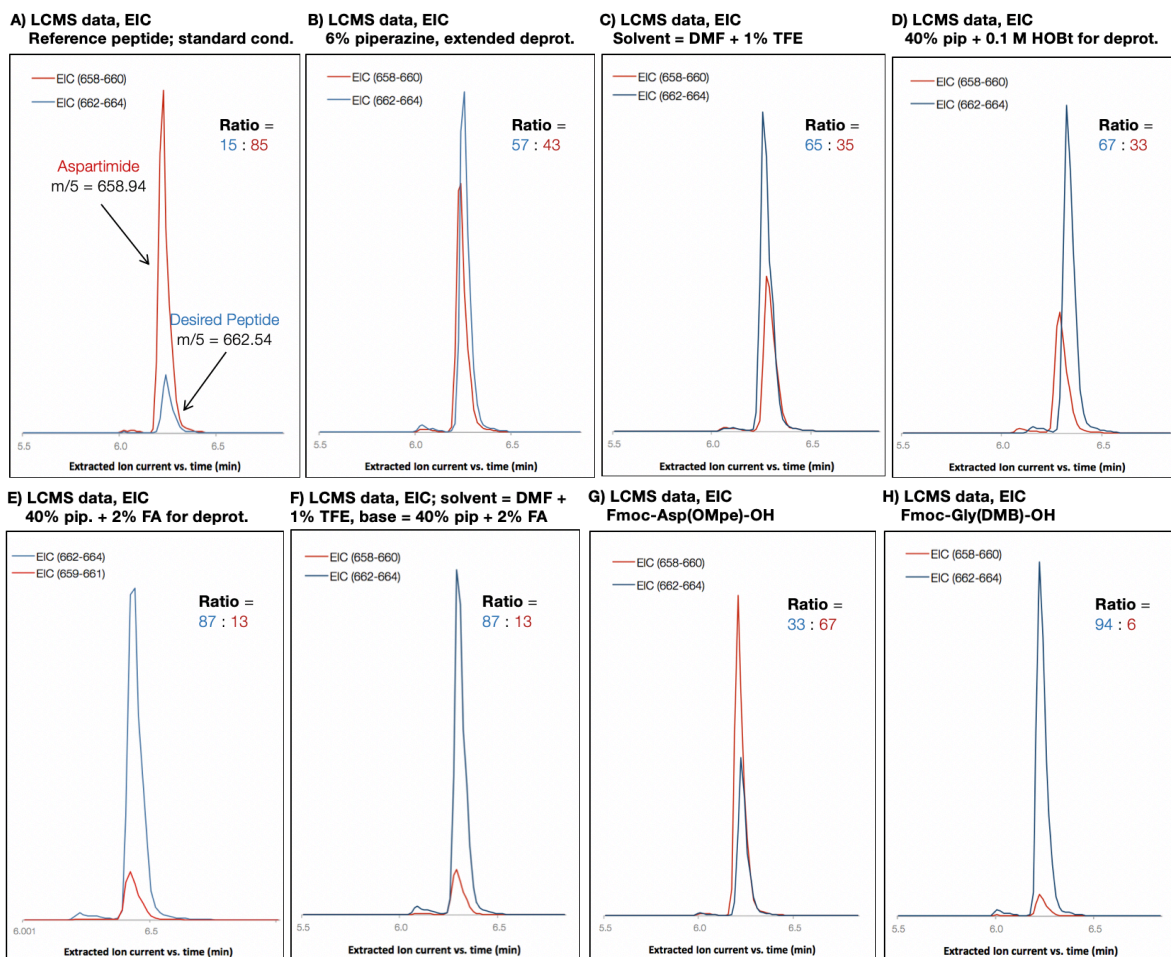


Fig. S3. Synthesis data for aspartimide-forming peptide NN92. A) LC-MS analysis of by-products; b) extracted ion counts of peptide m/z (662-664 Da) and aspartimide m/z (658-661); c) analytical HPLC trace.



Entry	base	solvent	Coupling agent	HPLC purity	Crude yield
1	Reference peptide; standard cond.	DMF	HATU ^{a)}	61%	17%
2	6% piperazine, extended deprot.	DMF	HATU ^{a)}	67%	31%
3	Solvent = DMF + 1% TFE	DMF	HATU ^{a)}	63%	27%
4	40% pip + 0.1 M HOBt for deprot.	DMF	HATU ^{a)}	63%	24%
5	40% pip + 2% formic acid for deprot.	DMF	HATU ^{a)}	65%	24%
6	solvent = DMF + 1% TFE, base = 40% pip + 2% FA	DMF	HATU ^{a)}	74%	34%
7	Fmoc-Asp(OMpe)-OH	DMF	HATU ^{a)}	70%	15%
8	Fmoc-Gly(DMB)-OH	DMF	HATU ^{a)}	72%	20%

Table S10. Formic acid additive and Fmoc-(DMB)Gly-OH were the most effective strategies to minimize aspartimide formation in the test peptide NN92. Synthesis conditions: 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading), flowrate: 40mL/min, temperature = 90 °C, 0.4 M amino acid stock solutions (diluted to 0.2 M solution during synthesis on the machine), 0.38 M activator (diluted to 0.19 M during synthesis on the machine), 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine) unless an alternative deprotection base was indicated in the figure.

Conclusion: In all cases the ratios of desired mass to aspartimide and HPLC purity (combined signal of product and aspartimide) improved; best results were achieved with formic acid additive and Fmoc-Gly(DMB)-OH.

3.9 Epimerization experiments

Epimerization experiments were carried out for Fmoc-Cys(Trt)-OH, Fmoc-His(Trt)-OH and Fmoc-His(Boc)-OH. Initial experiments on epimerization and the correlation between flow rate and epimerization were already reported.⁽¹⁷⁾ In additional experiments, we determined the correlation between epimerization and protecting group, temperature, heating time, reactor temperature, coupling strokes and additional coupling cycles. Four reference peptides (GCF, GcF, FHL and FhL) were prepared using standard batch solid-phase peptide synthesis. All peptides in this study were cleaved using a modified version of the protocol for short peptides (see section 1.3). Peptides were cleaved as described, the cleavage mixture was removed under a stream of nitrogen and the peptides were dissolved in 50% acetonitrile in water with 0.1% TFA. The peptide solution was filtrated by a Nylon syringe filter with a 0.22 μm pore size and directly submitted for analysis. For Fmoc-Cys(Trt)-OH epimerization was quantified by integrating EIC [$m = 325\text{--}327$] traces for both diastereomers. For Fmoc-His(Trt)-OH and Fmoc-His(Boc)-OH, epimerization was quantified by integrating analytical HPLC traces for both diastereomers (instrument and analysis method: analytical HPLC with Zorbax C18 column, 1–41 gradient over 20 minutes). Variability of approx. $\pm 0.5\%$ can be attributed to the inaccuracy of integration.

3.9.1 Temperatures for activation in 10' heating loop

First, we examined the dependence of epimerization on the temperature during activation in the 10' heating loop. We synthesized the two test peptides GCF and FHL using our standard coupling protocol on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g). All amino acids were coupled with different temperatures (90, 85, 80, 70, 60, and 50 $^{\circ}\text{C}$) in the 10' heating loop. The D-Cys values obtained from this analysis as a function of temperature as well as the respective EIC traces are plotted in **Fig. S4 A and B**. Approximately 4.5–5.0% D-isomer was observed under standard coupling conditions at 90 $^{\circ}\text{C}$, which matches the previously reported value.⁽¹⁷⁾ The D-His values obtained from this analysis as a function of temperature as well as the respective HPLC traces are plotted in **Fig. S4 C and D**.

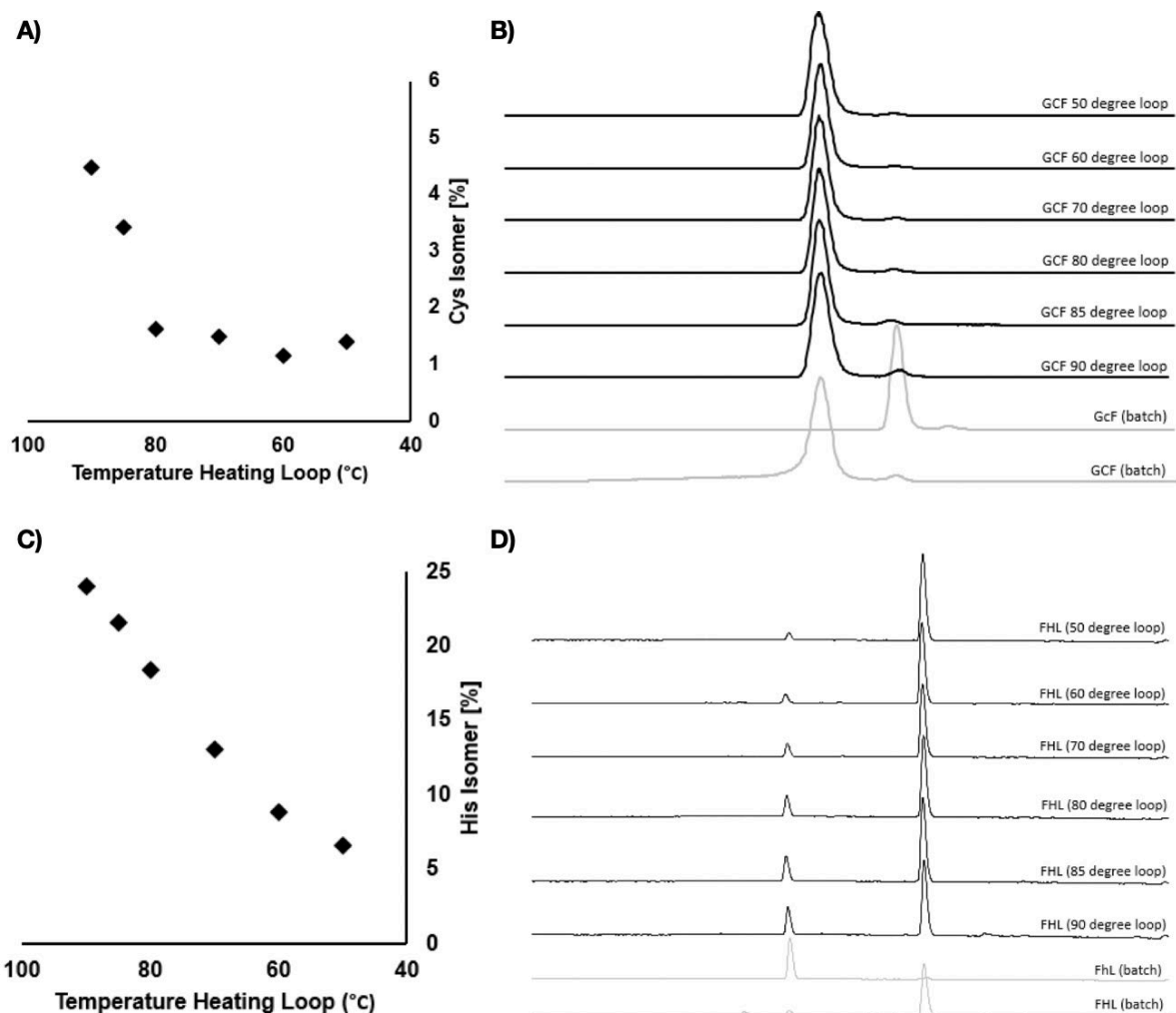


Fig. S4. Correlation between temperature and epimerization for cysteine (Fmoc-Cys(Trt)-OH) and histidine (Fmoc-His(Trt)-OH) coupling via a 10' activation loop at different temperatures. A) Temperature in the heating loop vs. formation of D-Cys isomer in % resulting from Fmoc-Cys(Trt)-OH coupling; B) EIC [$m = 325-327$] plotted for different temperatures; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Temperature in the heating loop vs. formation of D-His isomer in % resulting from Fmoc-His(Trt)-OH coupling; D) analytical HPLC traces plotted for syntheses at different temperatures; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

3.9.2 Temperatures for activation in 5' heating loop

Next, we optimized conditions for His and Cys coupling by investigating a shorter heating loop, different protecting groups, different activators and different temperatures. GCF and FHL were synthesized on the amidator using our standard coupling protocol (10' heating loop at 90 °C) on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g), except for His and Cys which were diverted to a 5' heating loop for activation (under the indicated conditions being studied).

Fmoc-Cys(Trt)-OH was coupled using either HATU or PyAOP as an activator with different temperatures (90, 80, 70, 60, 50, 40, 33 °C) in the 5' heating loop. For Fmoc-Cys(Trt)-OH

epimerization was quantified by integrating EIC [$m = 325\text{--}327$] traces for both diastereomers (instrument and analysis method: LCMS 6550-1 with LunaC18 column, 1–61 gradient over 18 minutes). The isomer values obtained from this analysis as a function of temperature as well as the respective EIC traces are plotted in **Fig. S5**. We found that both activators, HATU and PyAOP, led to an equal amount of epimerization, and epimerization was not further reduced below approx. 60 °C, where D-isomer remained approximately 1.5%. This is in accordance with experiments in the 10' heating loop.

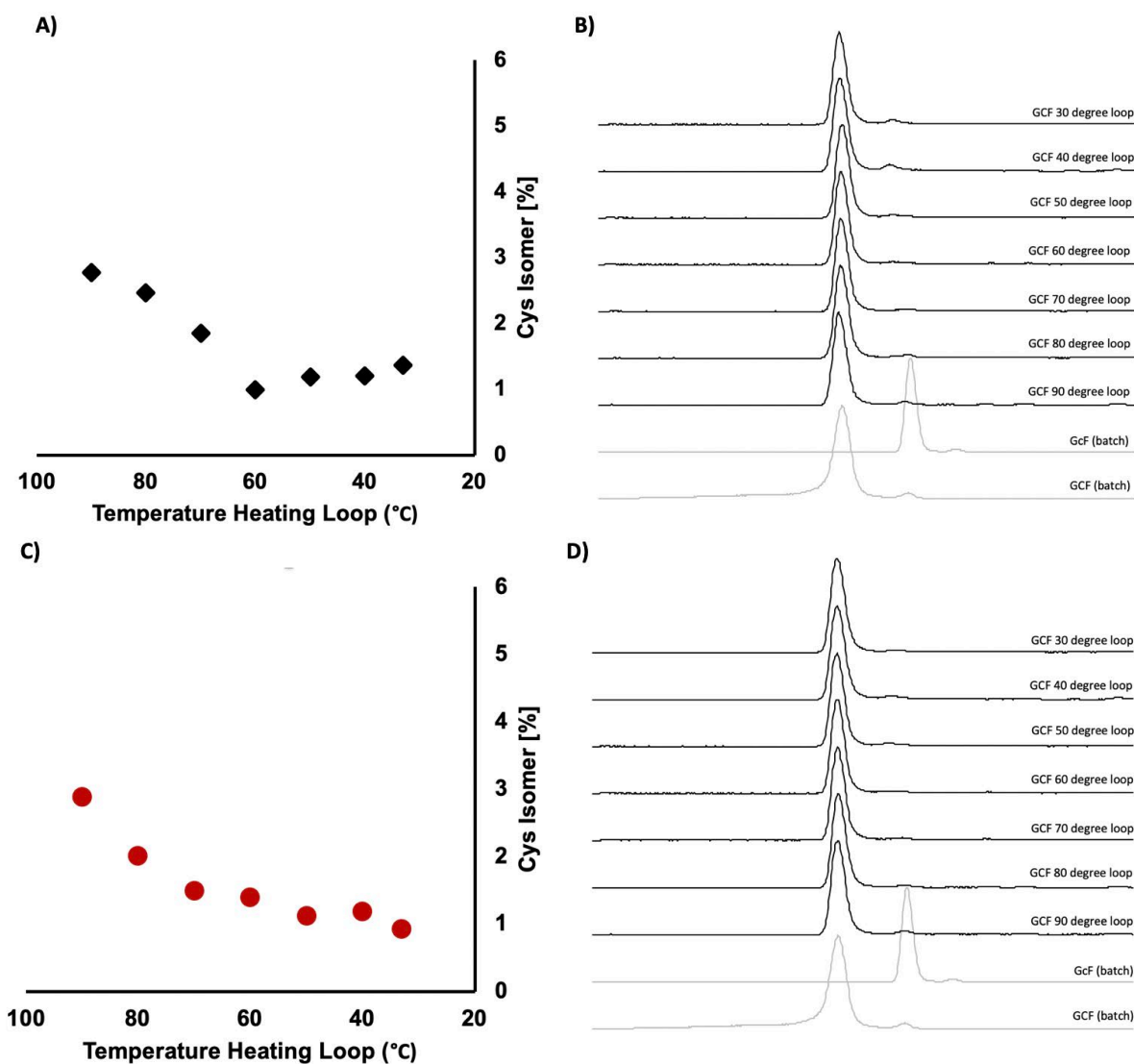


Fig. S5. Correlation between temperature and epimerization for cysteine coupling via a 5' activation loop with different activators and at different temperatures. A) Temperature in the heating loop vs. formation of Cys isomer in %, all Fmoc-Cys(Trt)-OH couplings with HATU; B) EIC [$m = 325\text{--}327$] plotted for different temperatures; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS; C) Temperature in the heating loop vs. formation of Cys isomer in %, all Fmoc-Cys(Trt)-OH couplings with PyAOP; D) EIC [$m = 325\text{--}327$] plotted for different temperatures; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

Ideal coupling conditions for histidine-coupling on the AFPS systems were investigated next. In the first set of experiments, Fmoc-His(Trt)-OH was coupled using HATU and PyAOP as an activator with different temperatures (90, 80, 70, 60, 50, 40, 30 °C) in the 5' heating loop. The epimerization values obtained from this analysis as a function of temperature are plotted in SI Fig. S6.

As histidine-epimerization could not be suppressed further using Fmoc-His(Trt)-OH, we investigated Fmoc-His(Boc)-OH as an alternative building block. Fmoc-His(Boc)-OH was coupled using HATU and PyAOP as an activator with different temperatures (90, 70, 50, 30 °C) in the 5' heating loop. The isomer values obtained from this analysis as a function of temperature are plotted in SI Fig. S6.

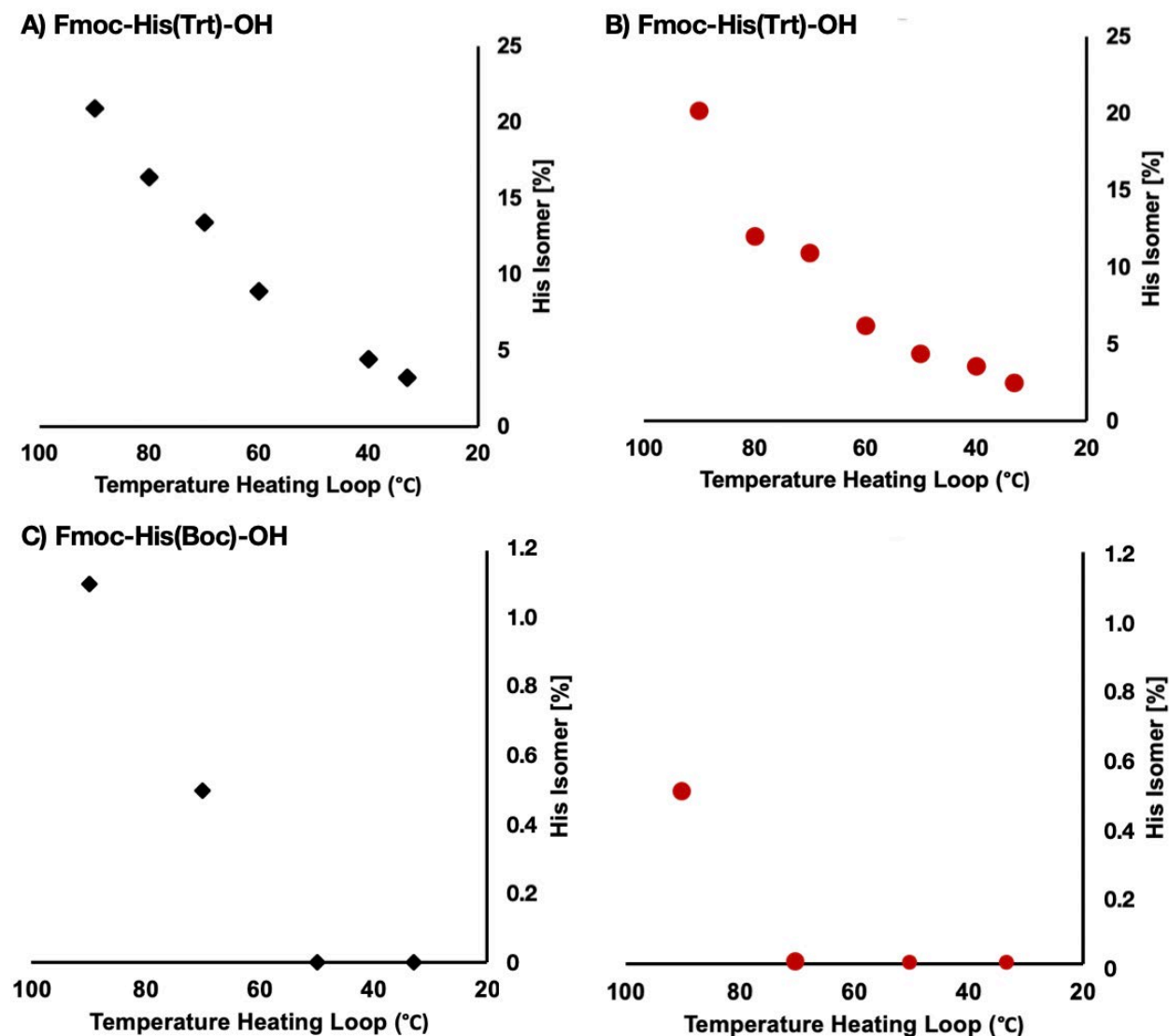


Fig. S6. Correlation between temperature and epimerization coupling of histidine via a 5' activation loop at with different activators and at different temperatures. A) Coupling of Fmoc-His(Trt)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Trt)-OH couplings with HATU; B) Coupling of Fmoc-His(Trt)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Trt)-OH couplings with PyAOP; C) Coupling of Fmoc-His(Boc)-OH; temperature in the heating loop vs. formation of His isomer in %, all

Fmoc-His(Boc)-OH couplings with HATU; D) Coupling of Fmoc-His(Boc)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Boc)-OH couplings with PyAOP.

3.9.3 Extended coupling time

Then we determined if Fmoc-protected resin-bound Cys(Trt) is prone to epimerization under extended coupling times. GCF was synthesized on the amidator using our standard coupling protocol (heating loop to 90 °C in 10') on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g) with various coupling strokes. The D-Cys values obtained from this analysis as a function of coupling cycles as well as the respective EIC traces are plotted in **Fig. S7**.

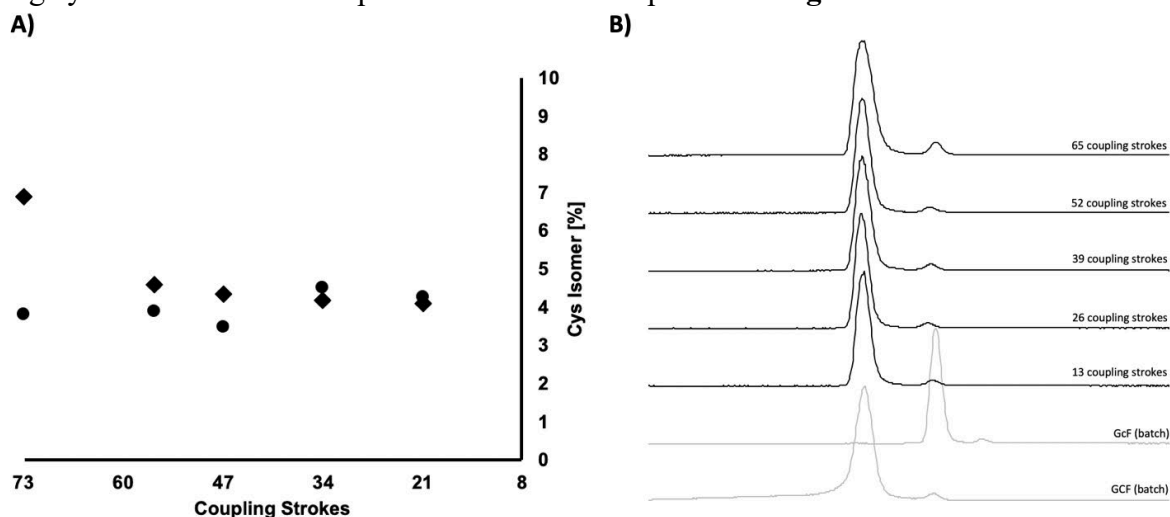


Fig. S7. Correlation between epimerization and coupling strokes for Fmoc-Cys(Trt)-OH. A) Fmoc-Cys(Trt)-OH coupling strokes vs. formation of Cys isomer in %, the study was performed in duplicate and both data sets are presented in the graph; B) EIC [m = 325-327] plotted for number of Fmoc-Cys(Trt)-OH coupling strokes; only first set of experiment is plotted; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

3.9.4 Multiple coupling cycles

GCF and FHL were synthesized on the amidator using our standard coupling protocol (10' heating loop to 90 °C) on 200 mg ChemMatrix Rink amide resin (0.49 mmol/g). The *N*-terminus was capped with a Boc-protecting group. To simulate the synthesis of long sequences and the effect on cysteine- and histidine-isomerization the resin was exposed to 100 glycine couplings carried out under standard conditions. After every 20 amino acid couplings approximately 20 mg of resin was removed for cleavage and analysis. The D-Cys values obtained from this analysis as a function of coupling cycles as well as the respective EIC traces are plotted in **Fig. S8 A and B**. The initial epimerization did not change over multiple coupling cycles. Variability of $\pm 0.5\%$ can be attributed to the inaccuracy of integration.

For Fmoc-His(Trt)-OH, epimerization was quantified by integrating analytical HPLC traces for both diastereomers. The D-His values obtained from this analysis as a function of coupling cycles as well as the respective HPLC traces are plotted in **Fig. S8 C and D**. The initial epimerization did not change over multiple coupling cycles.

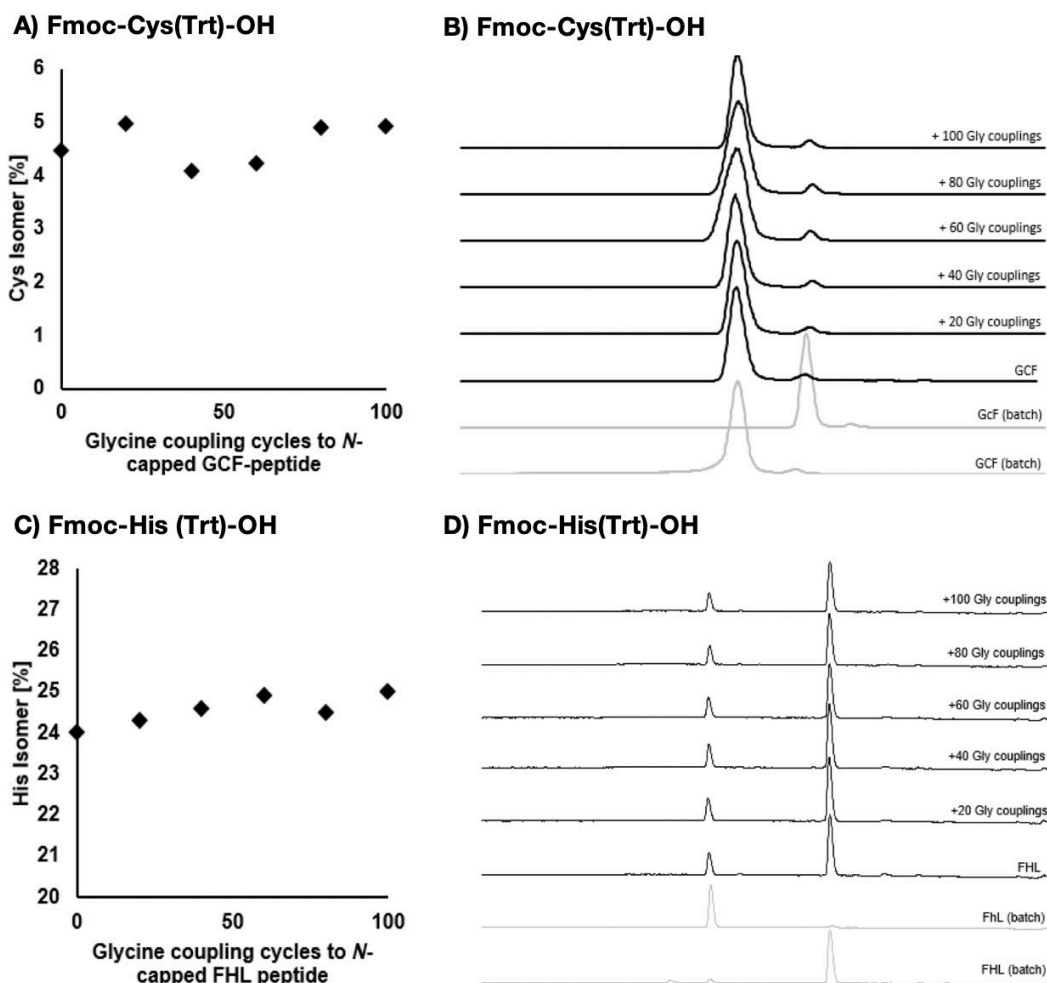


Fig. S8. Correlation between epimerization and number of amino acid coupling cycles for Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH using coupling conditions resulting in highest epimer formation. Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH were both coupled at 90 °C with HATU through the 10' heating loop. A) Glycine coupling cycles vs. formation of Cys isomer in %; B) EIC [m = 325-327] plotted for number of Glycine coupling cycles; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Glycine coupling cycles vs. formation of His isomer in %; D) analytical HPLC traces plotted for syntheses at different temperatures; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

In the final coupling protocol, Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH were both coupled with PyAOP through the 5' heating loop at 60 °C to achieve a balance between epimerization and coupling efficiency. The iterative coupling experiment was repeated with the final set of synthesis conditions as demonstrated in **Fig. S9**.

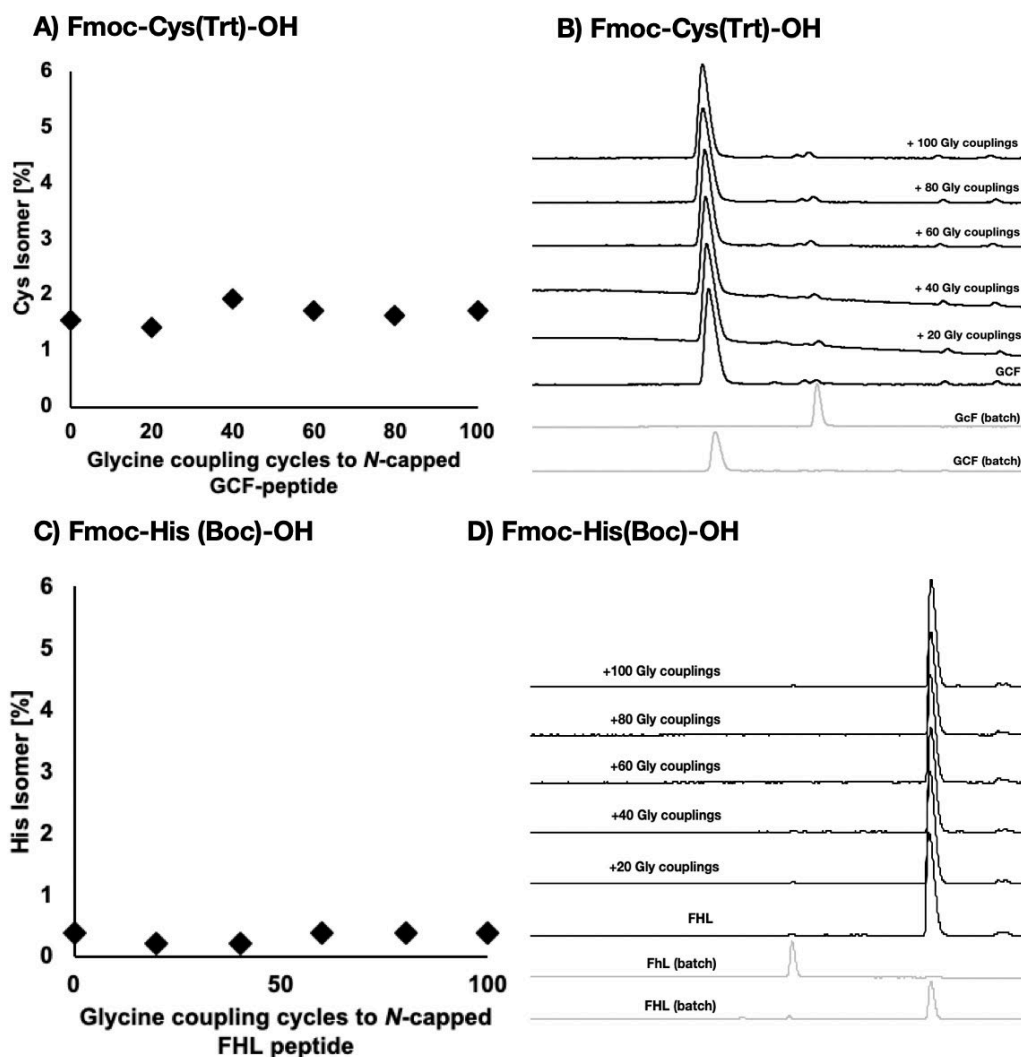


Fig. S9. Correlation between epimerization and number of amino acid coupling cycles for Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH using final coupling conditions. Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH were both coupled with PyAOP at 60 °C through a 5' heating loop. A) Glycine coupling cycles vs. formation of Cys isomer in %; B) analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Glycine coupling cycles vs. formation of His isomer in %; D) analytical HPLC traces plotted for number of glycine coupling cycles; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

3.10 Final synthesis protocol

Moving forward all peptide sequences were synthesized using the general synthesis protocol developed in **sections 2.1 to 2.9**:

Parameter	Conditions
Temperature	85–90 °C in reactor, 60 °C in 5' activation loop (for C and H), 90 °C in 10' activation loop (for all other amino acids)
Flow Rate	40 mL/min
Coupling step	0.40 M amino acids stocks in amine-free DMF 0.38 M activator stocks in amine-free DMF Coupling conditions: HATU (13 pump strokes) except S&A w/ HATU (26 pump strokes) and C, H, N, Q, V, R, T w/ PyAOP (26 pump strokes)
Deprotection step	40% pip in amine-free DMF with 2% formic acid (13 pump strokes)
Washing steps	Amine-free DMF (40 pump strokes)

Table S11. Set of optimized synthesis conditions on the AFPS. Pump strokes refer to volumes described in the general synthesis protocol (section 1.2).

3.10.1 GLP-1

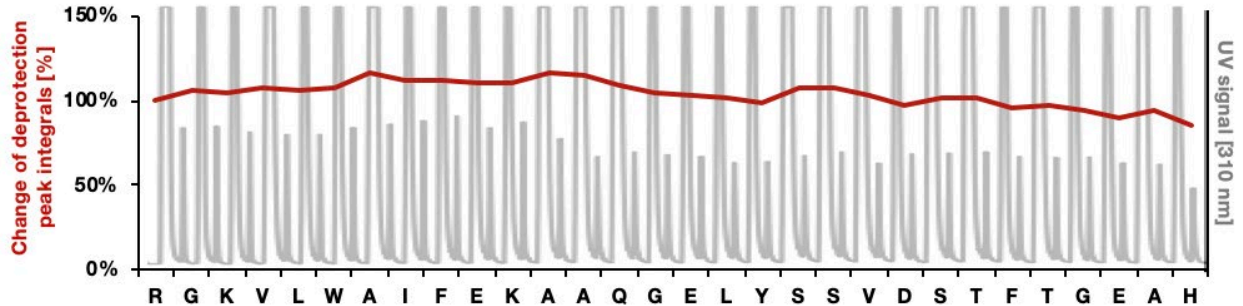
A) Synthesis Data for GLP-1

Sequence: HAEGTFTSDV SSYLEGQAAK EFLAWLVKGR (30 AA)

Resin: 100 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 1.1 h

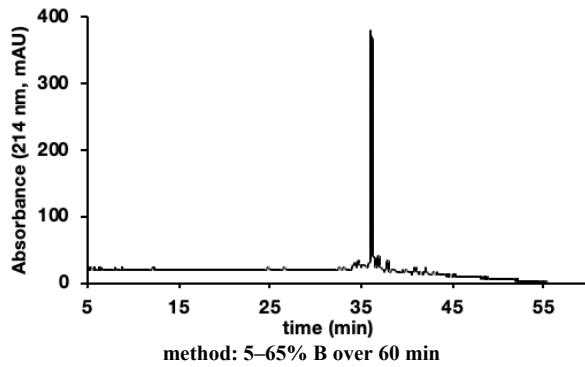
Synthesis UV trace from AFPS:



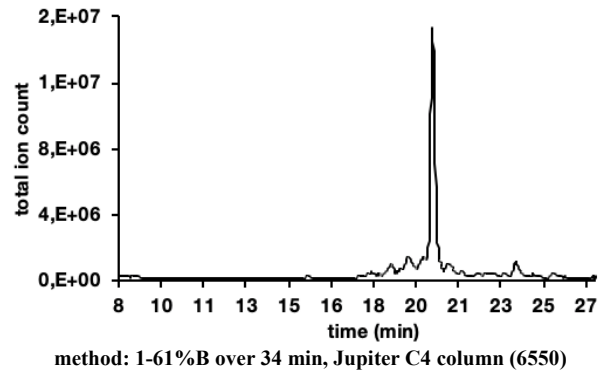
B) Cleavage and analytical Data for crude GLP-1

Cleavage: Cleavage protocol A for peptides

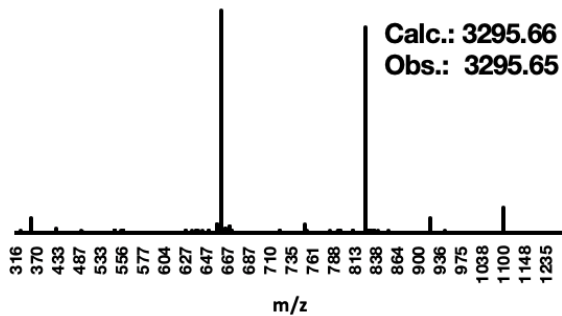
Analytical HPLC:



LC-MS:



MS data from LC-MS:



3.10.2 NN92

A) Synthesis Data for NN92

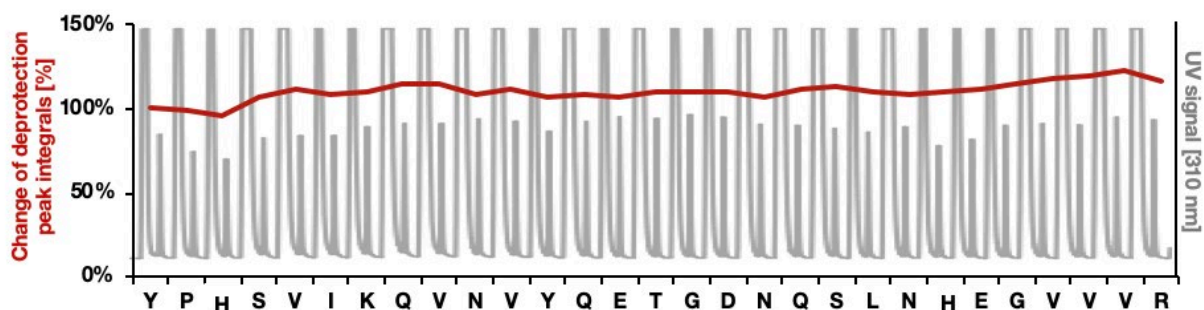
Sequence: RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)

Resin: 100 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 1.1 h

Synthesis UV trace from AFPS:

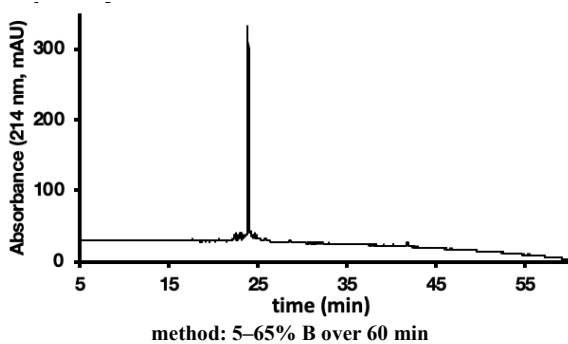
a) UV trace from AFPS



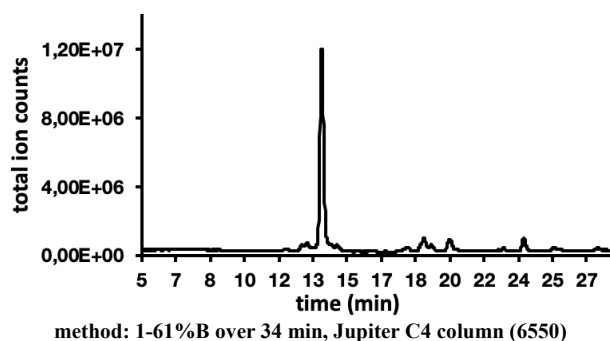
B) Cleavage and analytical Data for crude NN92

Cleavage: Cleavage protocol A for peptides

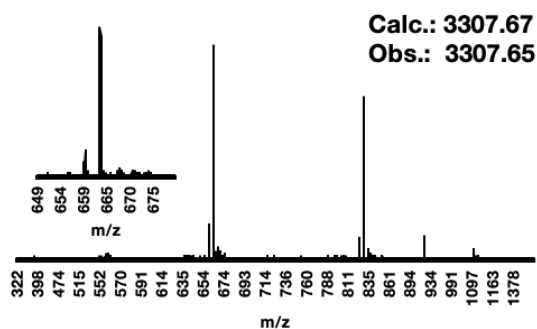
Analytical HPLC:



LC-MS:



MS data from LC-MS:



3.10.3 Amyloid beta [1-42]

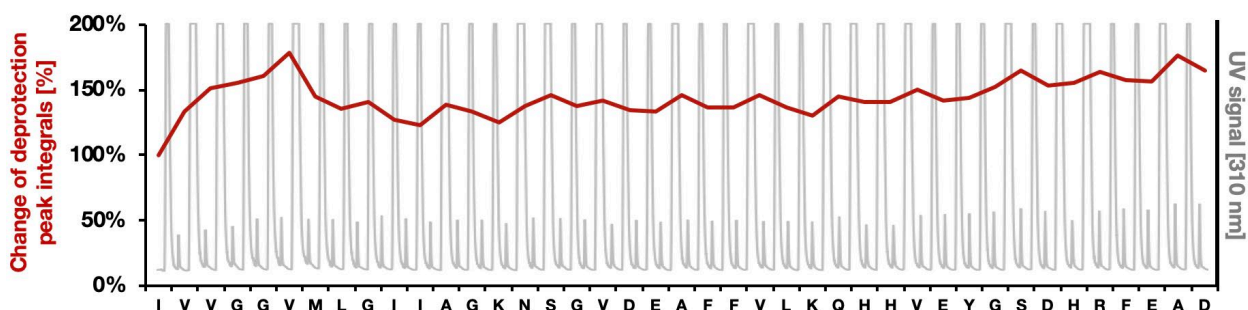
A) Synthesis Data for Abeta

Sequence: DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVV IA (42 AA)

Resin: 90 mg of HMPB ChemMatrix® (0.09 mmol/g) pre-coupled to alanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 1.7 h

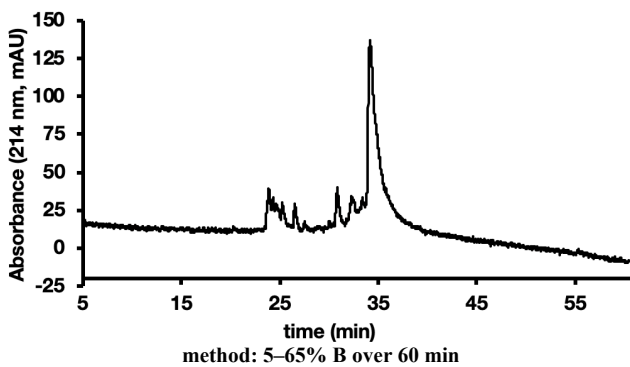
Synthesis UV trace from AFPS:



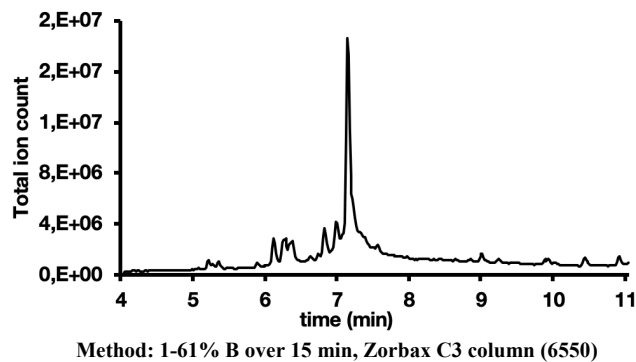
B) Cleavage and analytical Data for crude Abeta

Cleavage: Cleavage protocol A for peptides

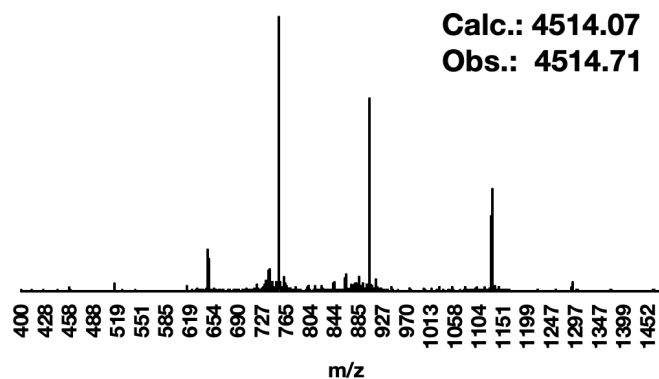
Analytical HPLC:



LC-MS:



MS data from LC-MS:



4 Synthesis of longer peptides - comparison between AFPS and traditional synthesis methods

4.1 Synthesis on AFPS

All proteins were synthesized using the general synthesis protocol developed in section 2 unless otherwise noted.

4.1.1 GLP-1

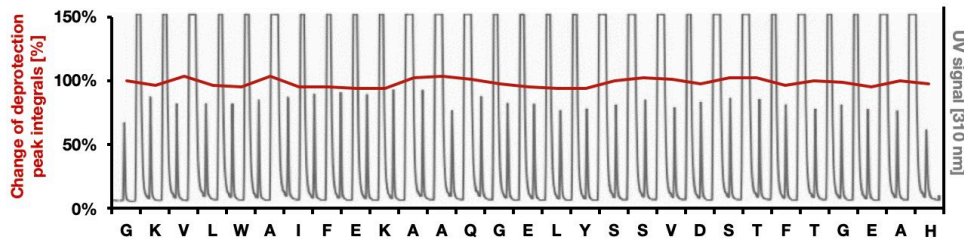
A) Synthesis Data for GLP-1

Sequence: HAEGFTSDV SSYLEGQAAK EFLAWLVKGR (30 AA)

Resin: 140 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage.

Synthesis time: 1.1 h

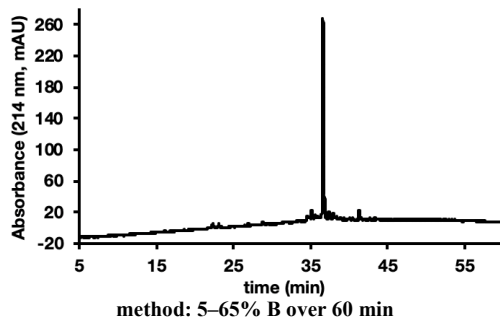
Synthesis UV trace from AFPS:



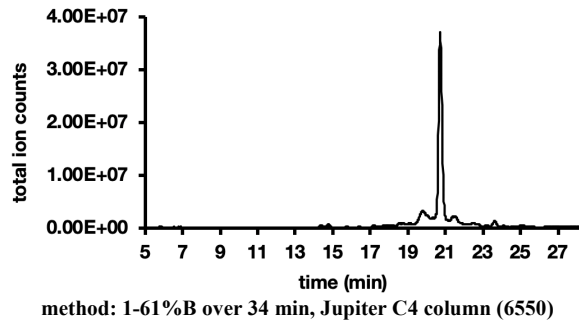
B) Cleavage and analytical Data for crude GLP-1

Cleavage: Cleavage protocol A for peptides

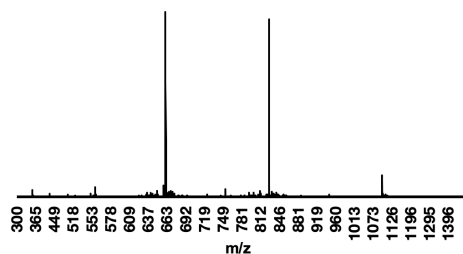
Analytical HPLC:



LC-MS:



MS data from LC-MS:

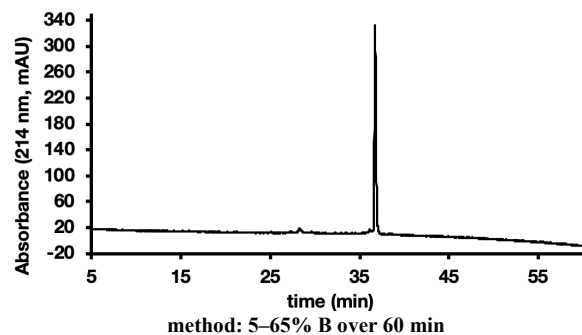


C) HPLC purification and analytical data for GLP-1

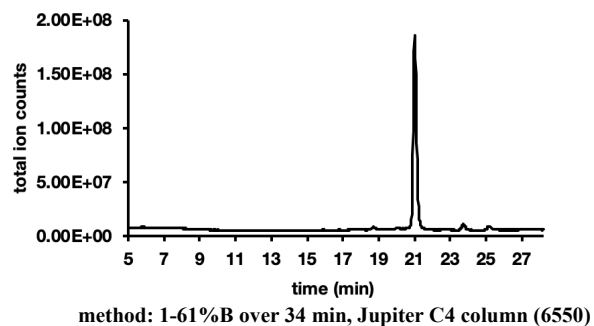
HPLC purification: Purified on semipreparative Zorbax 300SB-C3 column at 60 °C and 4 mL/min (gradient: 5–31% B at 1%/min, 32–52% B at 0.2%/B min, 53–65% B at 1%B/min).

Yield past HPLC: 8 mg (8%)

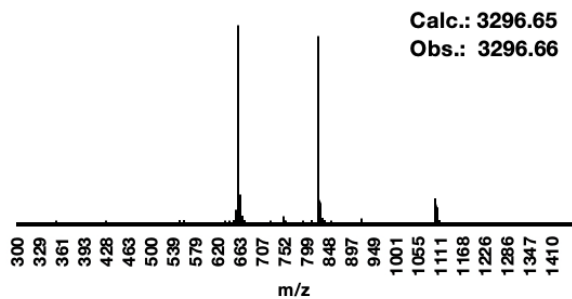
Analytical HPLC:



LC-MS:



MS data from LC-MS:



4.1.2 Proinsulin

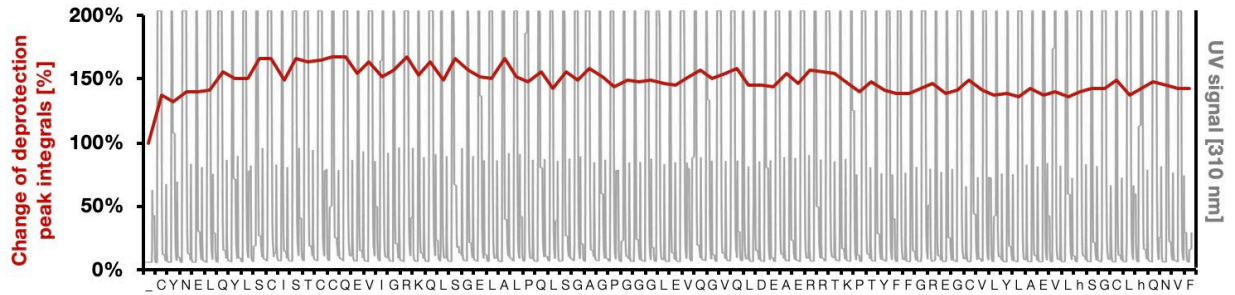
A) Synthesis Data for proinsulin

Sequence: FVNQHLGGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA
GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 3.5 h

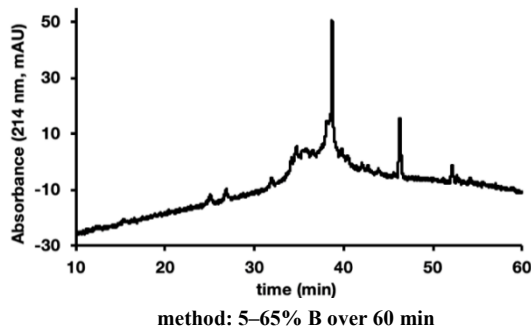
Synthesis UV trace from AFPS:



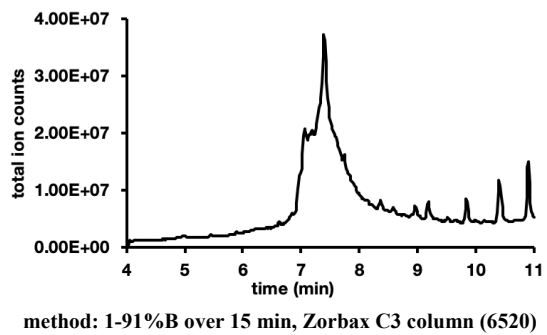
B) Cleavage and analytical Data for crude proinsulin

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

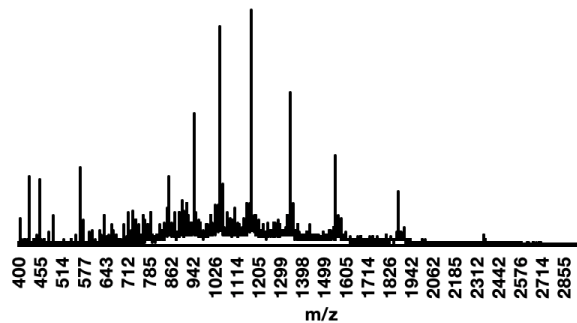
Analytical HPLC:*)



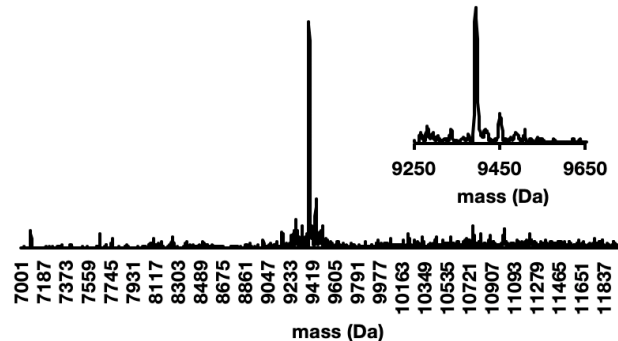
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

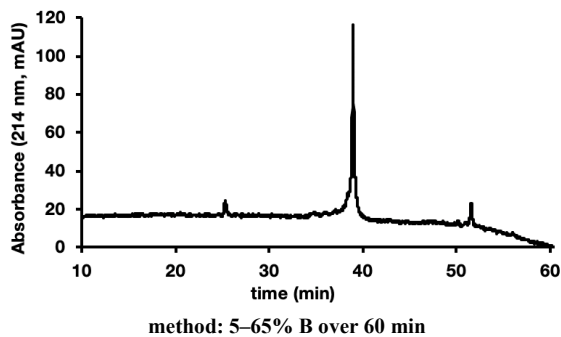


C) HPLC purification and analytical data for proinsulin

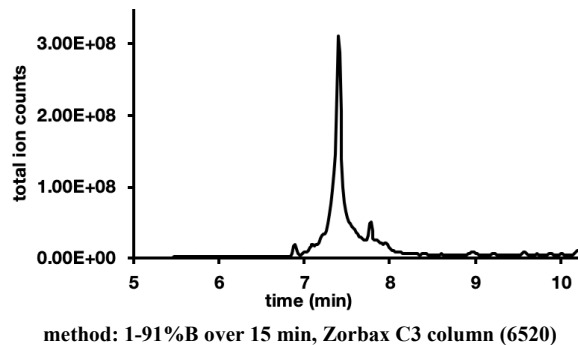
HPLC purification: Purified on semipreparative Zorbax 300SB-C3 column at 60 °C and 4 mL/min (gradient: 5–29% B at 1%/min, 29–49% B at 0.2%/B min, 49–65% B at 1%B/min).

Yield past HPLC: 4.8 mg (1%)

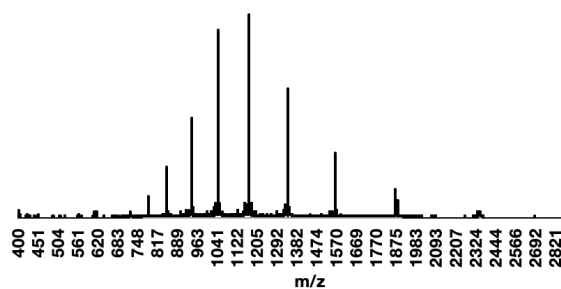
Analytical HPLC:*)



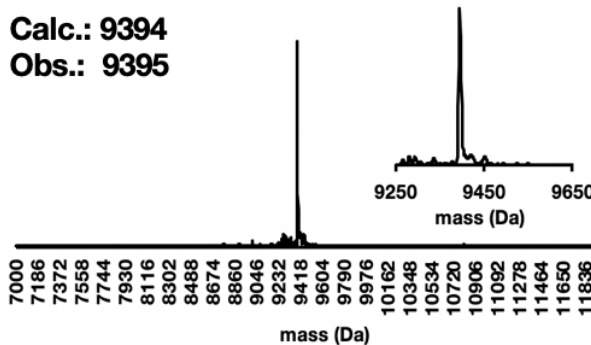
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



*) peptide is not very soluble, the absorbance of this sample is therefore low

4.1.3 HIV protease (wild type sequence)

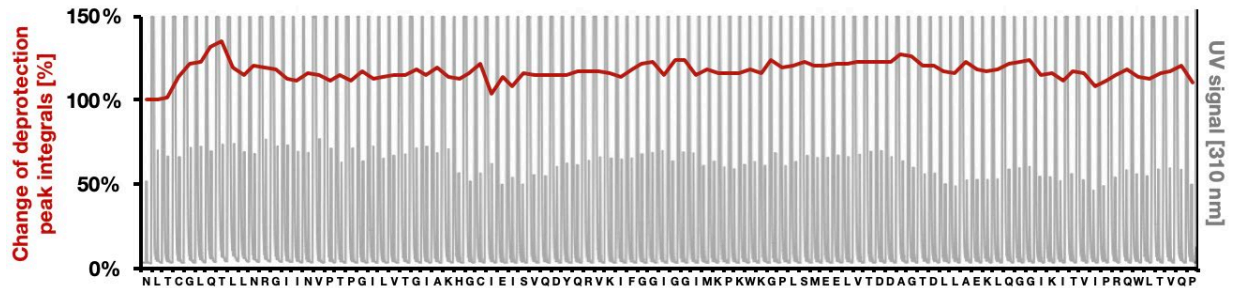
A) Synthesis Data

Sequence: PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI GGFIVRQYD
QVSIEICGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF
(99 AA)

Resin: 100 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

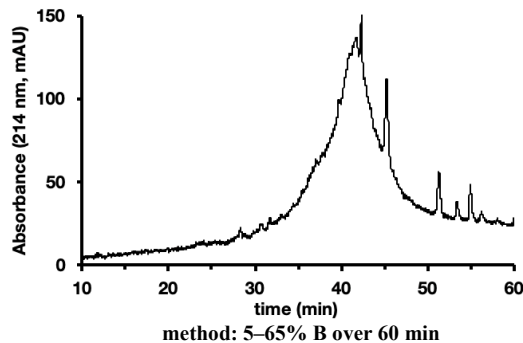
Synthesis UV trace from AFPS:



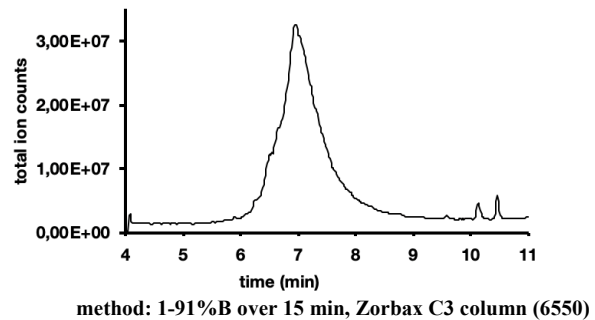
B) Cleavage and analytical Data for crude HIV-1 protease

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

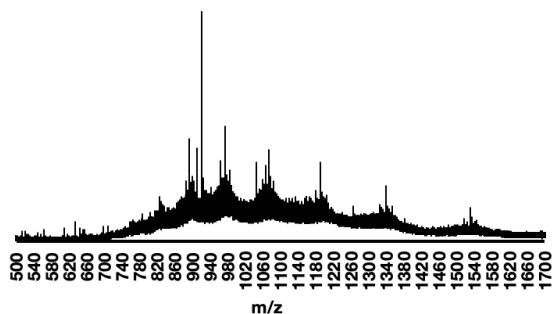
Analytical HPLC:



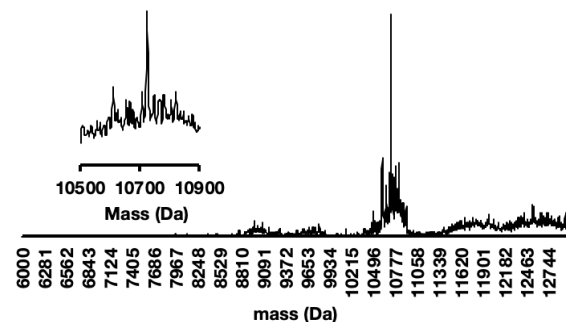
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

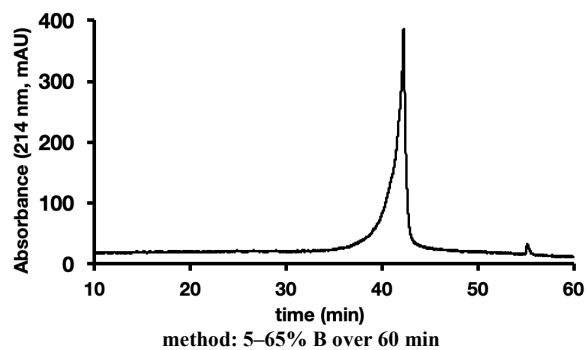


C) HPLC purification and analytical data for HIV-1 protease

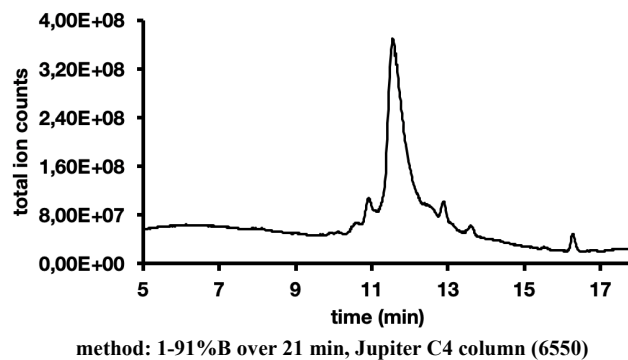
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-29%B with 1% B/min, 29-49%B with 0.2% B/min, 49-65% B with 1% B/min)

Yield past HPLC: 5.3 mg (1%)

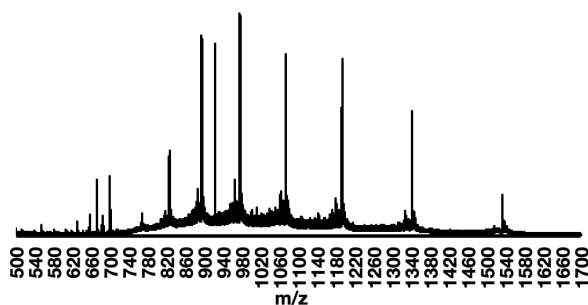
Analytical HPLC:



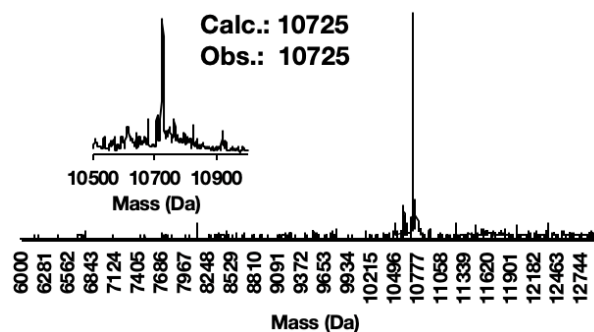
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



4.2 Synthesis on commercially available synthesizers at r.t., 70 °C and 90 °C

All proteins were synthesized at Novo Nordisk using PurePep™ amino acids (Gyros Protein Technologies) on a commercially available peptide synthesizer at room temperature and at elevated temperatures. A standard peptide synthesis protocol used at Novo Nordisk was applied for all syntheses. For coupling, Oxyma (0.3 M) and amino acid (0.3 M) were premixed, before DIC (3.0 M) was added. For the room temperature synthesis Collidine (3.0M) was added as activator. 0.1M Oxyma in 20% Piperidine in DMF was used for deprotection.

For room temperature synthesis (including capping), the general protocol included the following steps:

step	Volume (uL)	Mixing time (min)	Heated?	Repeat
Deprotection	6000	10	-	2
DCM wash	1500	0	-	1
DMF wash	6000	0	-	6
Amino acid/Oxyma	5000	0	-	1
DIC	1000	0	-	1
Collidine	1000	0	-	1
Mix and coupling	-	120	-	1
Drain/Dry	-	0	-	1
DCM wash	300	0	-	1
Mix	-	9	-	1
Cap	5000	0	-	1
Collidine	1000	0	-	1
Mix	-	20	-	1
Drain/Dry	-	0	-	1
DMF	6000	0	-	3
DCM	1500	0	-	1
DMF	6000	0	-	1

Total time per coupling and deprotection cycle: **3h 37min**

For 70 and 90 °C syntheses, the general protocol included the following steps:

step	Volume (uL)	Mixing time (min)	Heated?	Repeat
Deprotection	3000	1	-	1
DMF wash	1000	0	-	1
Deprotection	3500	2	x	1

DMF wash	1000	0	-	4
DCM wash	3000	0	-	1
DMF wash	3000	0.3	-	3
Amino acid/Oxyma	3000	0	-	1
DIC	300	0	-	1
DCM wash	300	0	-	1
Mix	-	9	x	1
Drain/Dry	-	0	-	1
DCM wash	3000	0	-	1
DMF wash	1000	0	-	4
DMF wash	3000	0.3	-	2

Total time per coupling and deprotection cycle: **26 min**

4.2.1 GLP-1

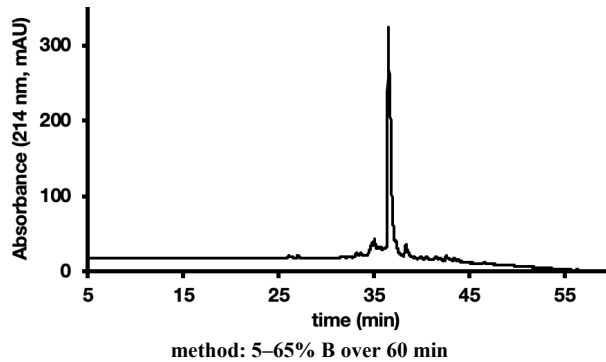
A) Synthesis Data for GLP-1

Sequence:	HAEGTFTSDV SSYLEGQAAR EFLAWLVRGR GF (32 AA)
Resin:	50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage
Synthesis protocol:	Commercially available peptide synthesizer at 70 °C

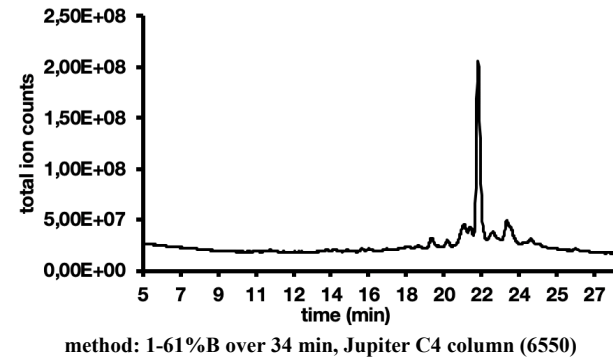
B) Cleavage and analytical Data for crude GLP-1

Cleavage: Cleavage protocol A for peptides

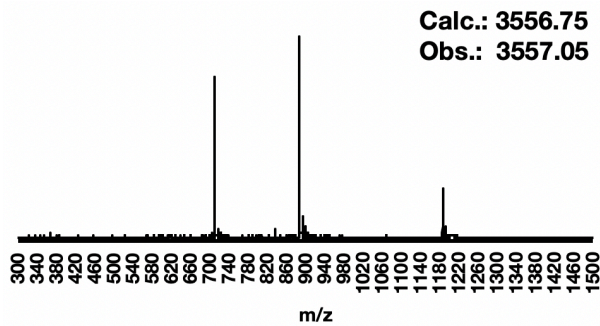
Analytical HPLC:



LC-MS:



MS data from LC-MS:



4.2.2 Proinsulin

A) Synthesis Data for proinsulin

Sequence: FVNQHLGGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA
GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

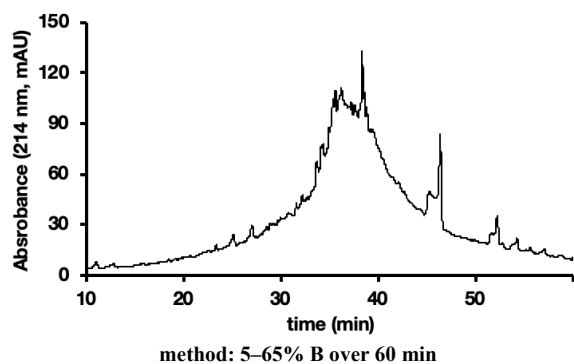
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at room temperature**

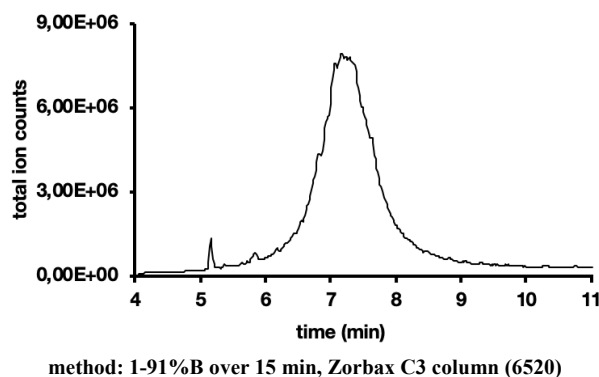
B) Cleavage and analytical Data for crude proinsulin

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

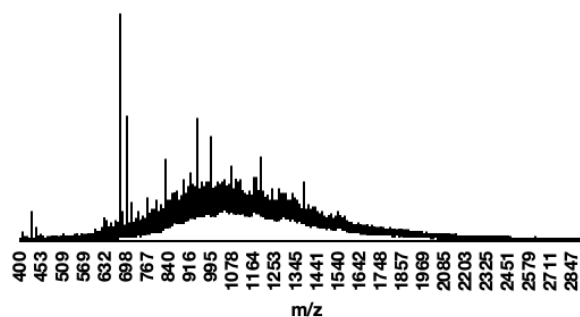
Analytical HPLC:*)



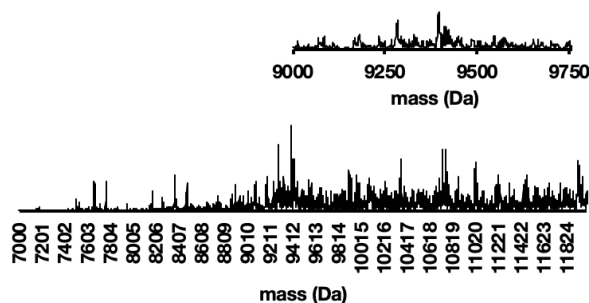
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for proinsulin

Sequence: FVNQHLGGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA
GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

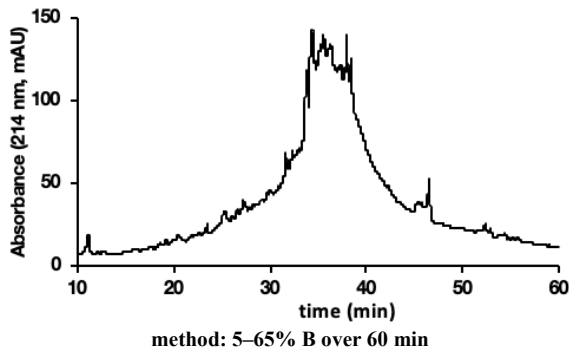
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **PRELUDE® X peptide synthesizer at 70 °C**

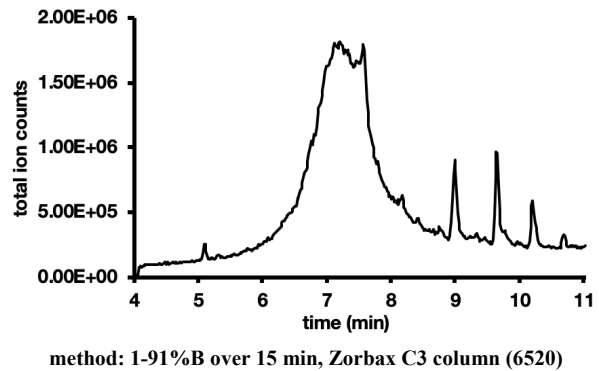
B) Cleavage and analytical Data for crude proinsulin

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

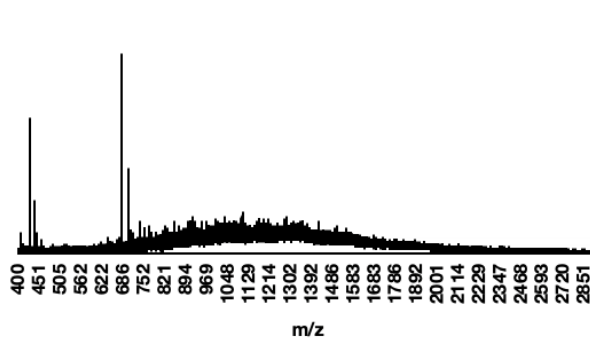
Analytical HPLC:*)



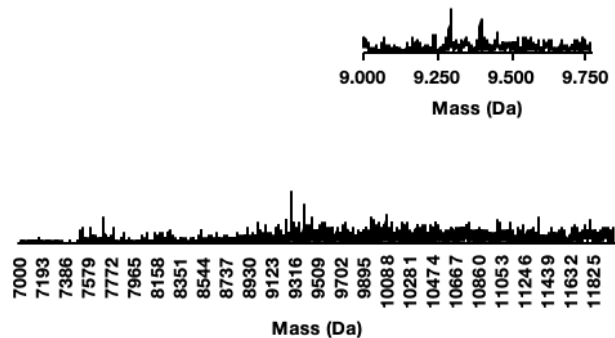
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for proinsulin

Sequence: FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA
GSLQLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

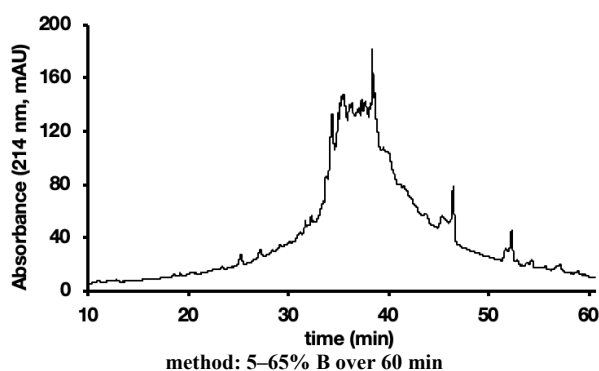
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at 90 °C**

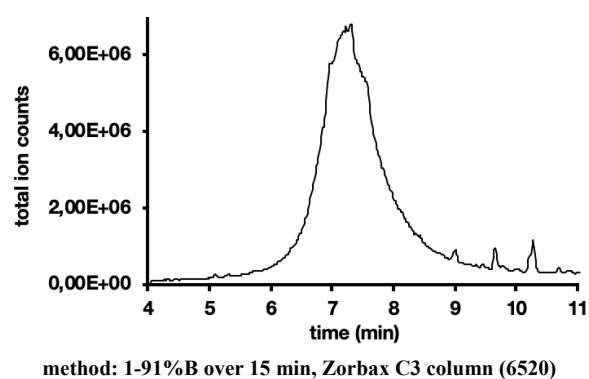
B) Cleavage and analytical Data for crude proinsulin

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

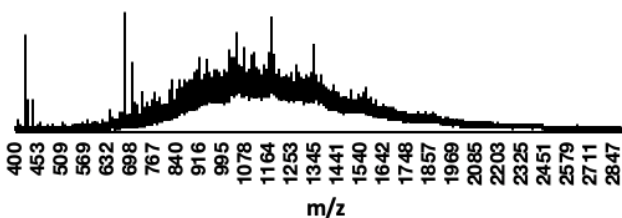
Analytical HPLC:*)



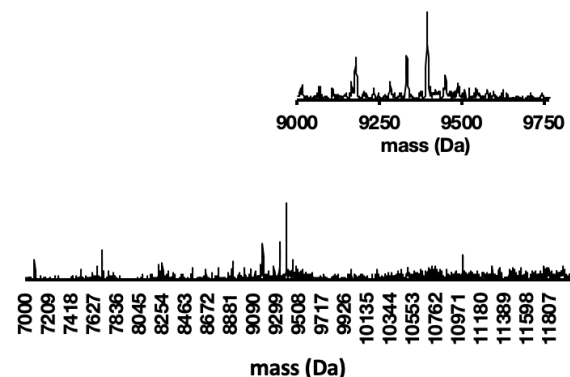
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



4.2.3 HIV protease (wild type sequence)

A) Synthesis Data

Sequence: PQVTWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI
GGFIKVRQYD QVSIEICGHK AIGTVLIGPT PVNIIGNRLL TQLGCTLNF
(99 AA)

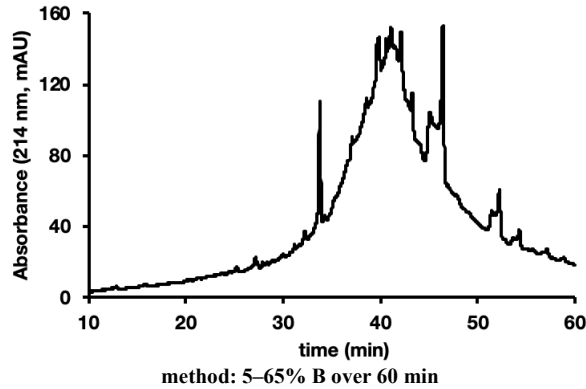
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at room temperature**

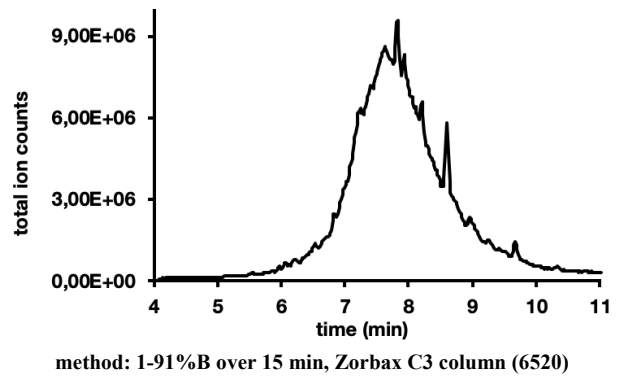
B) Cleavage and analytical Data for crude HIV-1 protease

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

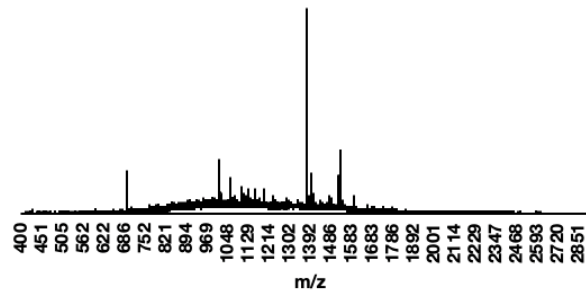
Analytical HPLC:



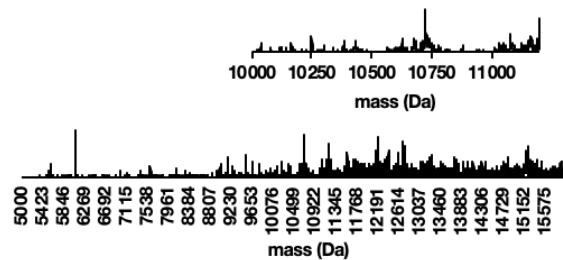
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for crude HIV-1 protease

Sequence: PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI
GGFIKVRQYD QVSIEIGGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF
(99 AA)

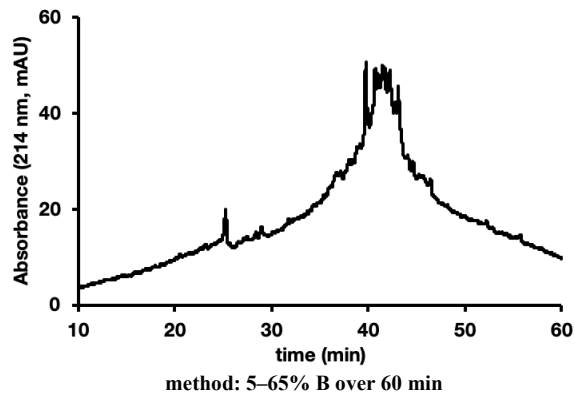
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at 70 °C**

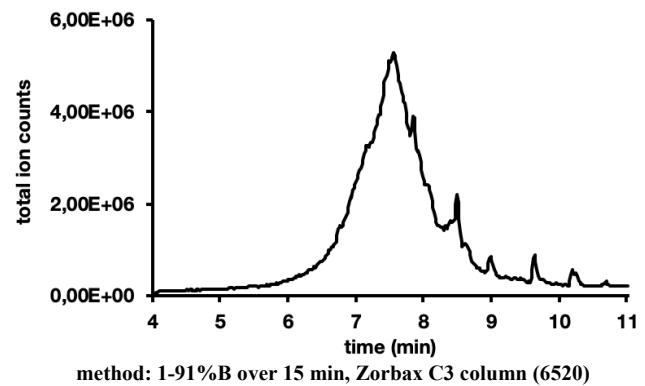
B) Cleavage and analytical Data for crude HIV-1 protease

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

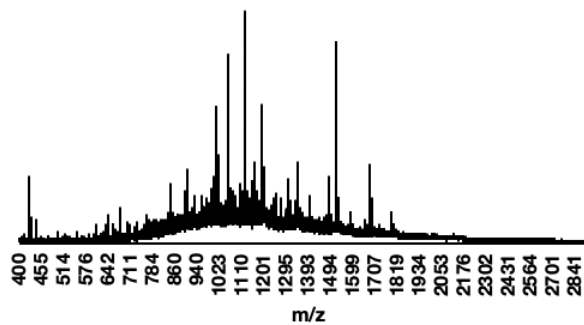
Analytical HPLC:



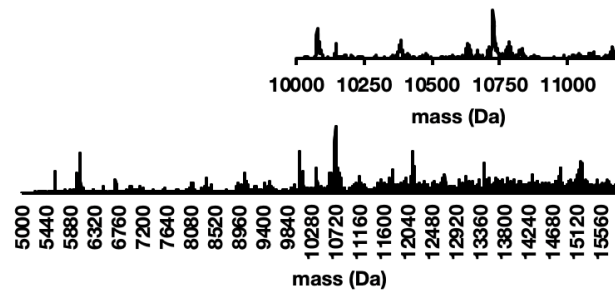
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



5 AFPS Synthesis of peptides in the length range of single domain proteins

5.1 MDM2^[25-109]

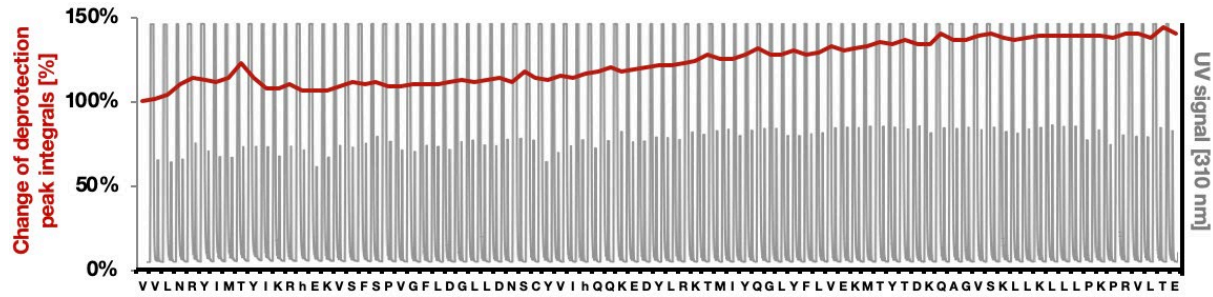
A) Synthesis Data for MDM2^[25-109]

Sequence: ETLVRPKLL LKLLKSVGAQ KDTYTMKEVL FYLGQYIMTK RLYDEKQQHI VYCSNDLLGD LFGVPSFSVK EHRKIYTMIIY RNLVV (85 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 3.5 h

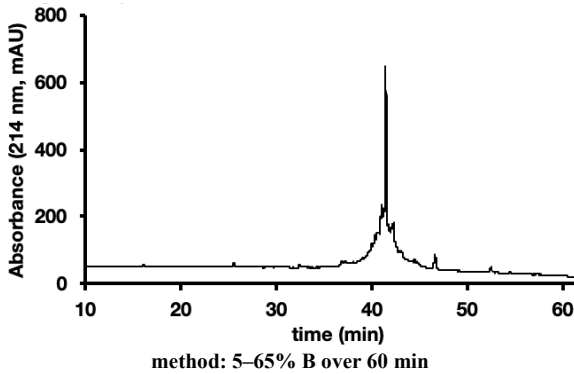
Synthesis UV trace from AFPS:



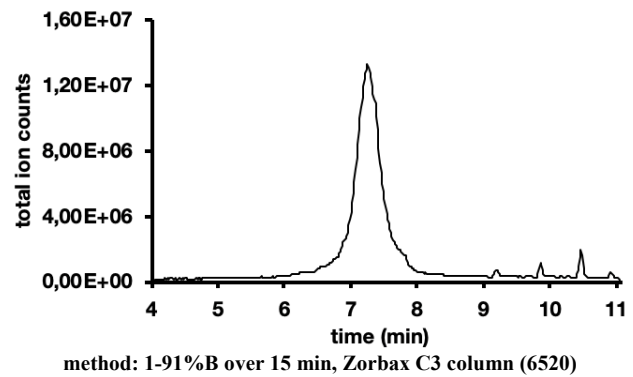
B) Cleavage and analytical Data for crude MDM2^[25-109]

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

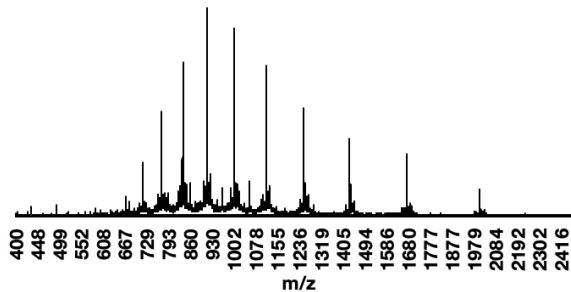
Analytical HPLC:



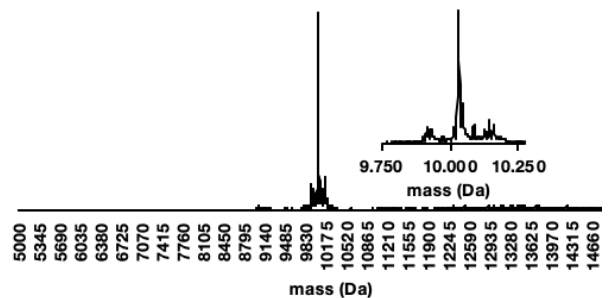
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

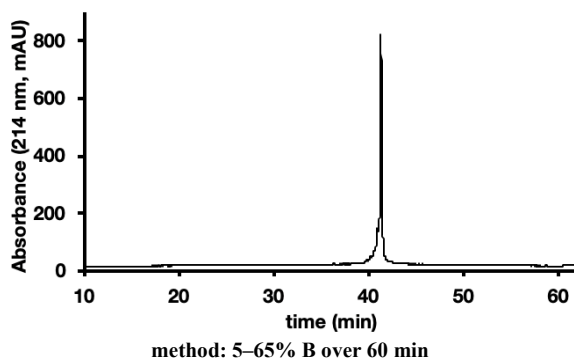


C) HPLC purification and analytical data for MDM2^[25-109]

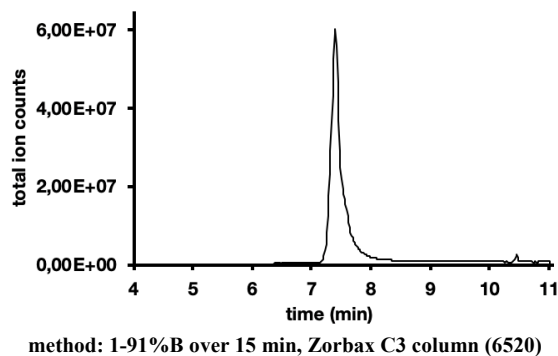
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C
(gradient: 5-29% B with 1% B/min, 29-49%B with 0.2% B/min, 49-65%
B with 1%B/min)

Yield past HPLC: 19.1 mg (6%)

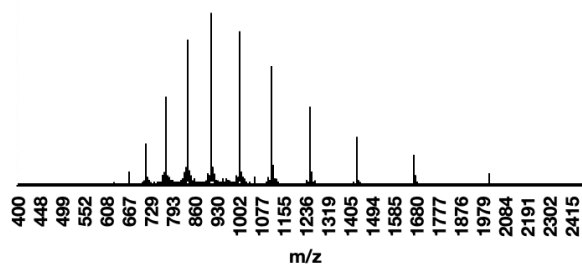
Analytical HPLC:



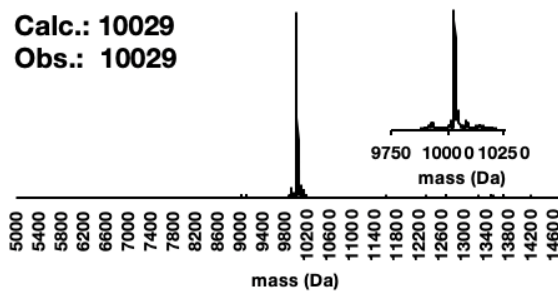
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



5.2 Barstar

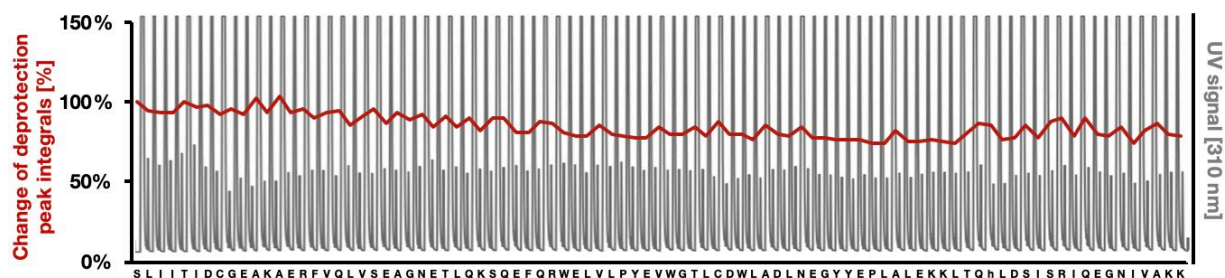
A) Synthesis Data of barstar

Sequence: KKAVINGEQI RSISDLHQTL KKEALPEYY GENLDALWDC LTGWVEYPLV
LEWRQFEQSK QLTENGAESV LQVFREAKAE GCDITIILS (89 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 3.5 h

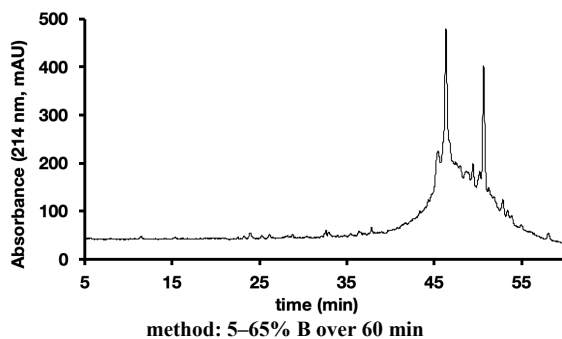
Synthesis UV trace from AFPS:



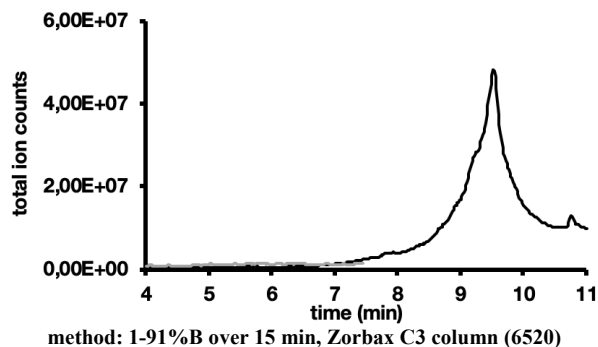
B) Cleavage and analytical Data for barstar

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

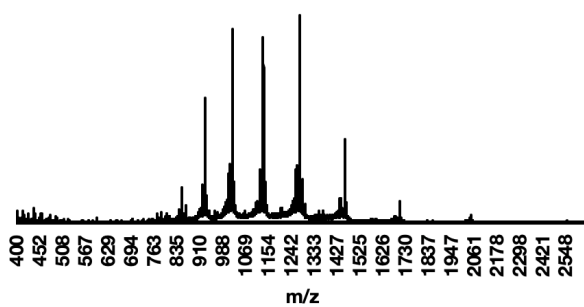
Analytical HPLC:



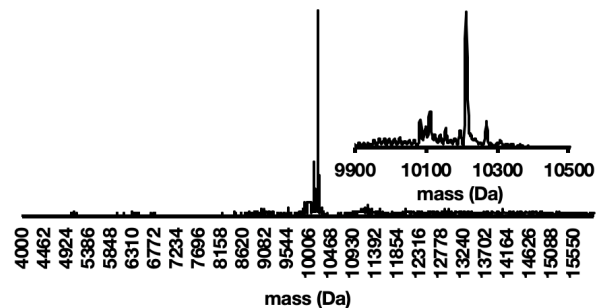
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

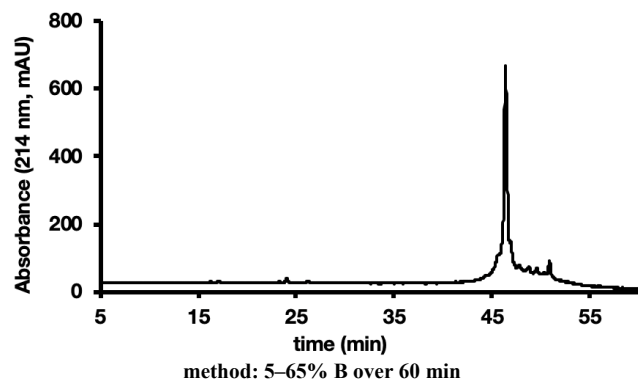


C) HPLC purification and analytical data for barstar

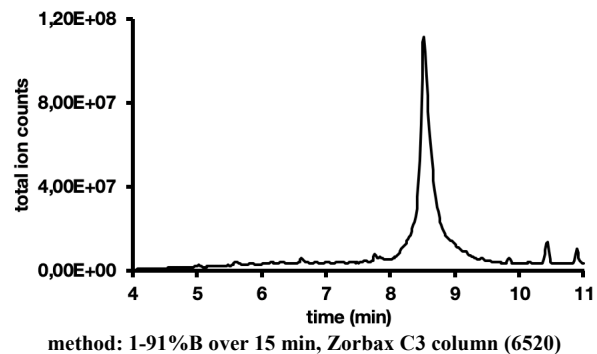
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-43%B with 1.5% B/min, 43-63%B with 0.2% B/min, 63-73% B with 1% B/min)

Yield: 5.5 mg (2%)

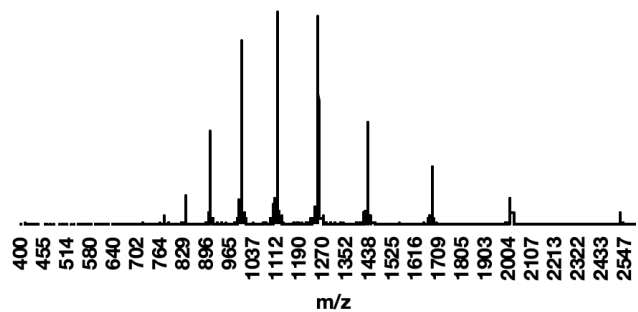
Analytical HPLC:



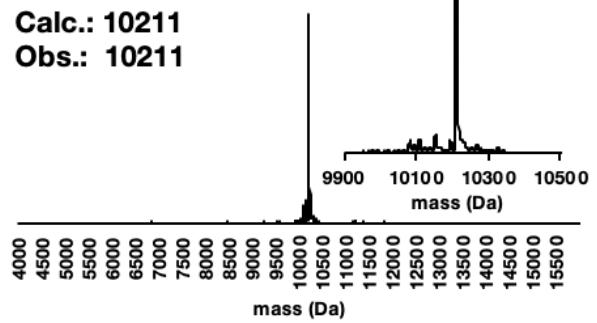
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



Comments:

*) Desired deconvoluted mass is the primary peak in both crude and purified samples.

5.3 Collagen

A) Synthesis Data

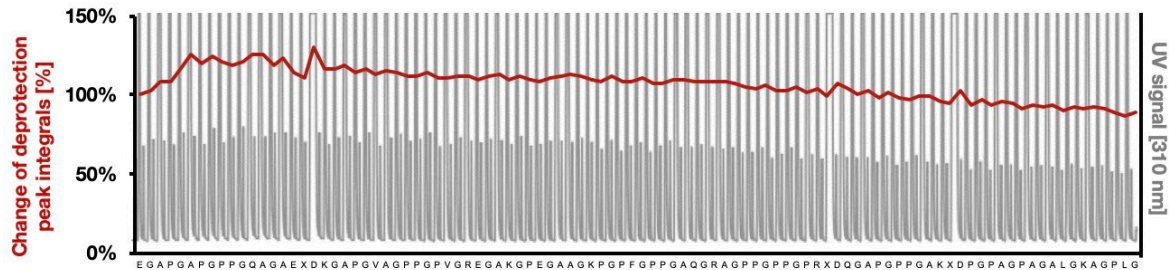
Sequence: GLPGAKGLAG APGAPGPDGK AGPPGPAGQD GRPGPPGPPG ARGQAGPPGF. PGPKGAAGEP
GKAGERGVPG PPGAVGPAGK DGEAGAQQPP GPAGPAGER

Note: the three glycines (G) highlighted in blue were introduced as Fmoc-(Dmb)Gly-OH building blocks

Resin: 100 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.0 h

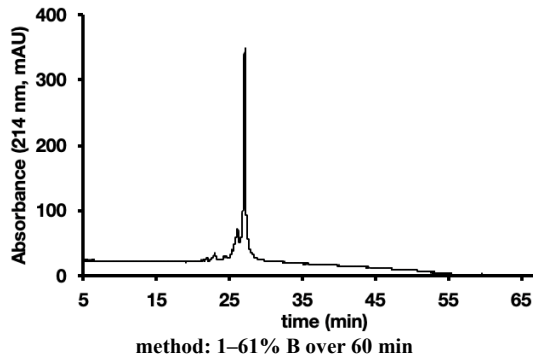
Synthesis UV trace from AFPS:



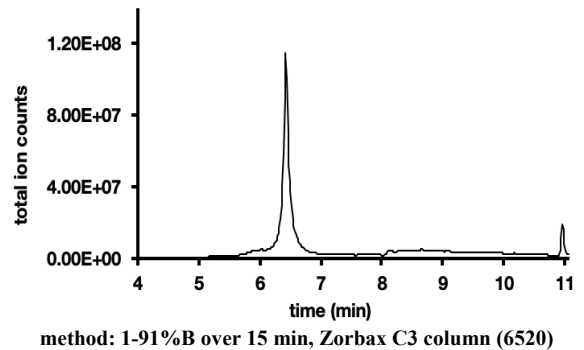
B) Cleavage and analytical Data for crude collagen

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:

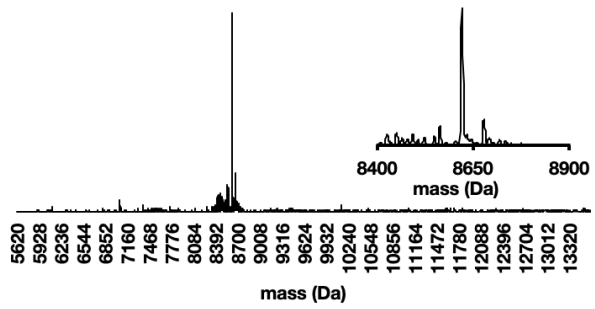
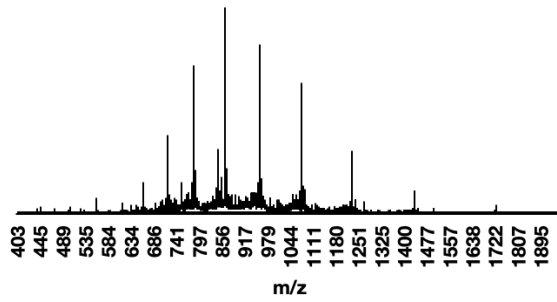


LC-MS:



MS data from LC-MS:

Deconvolution of LC-MS data:

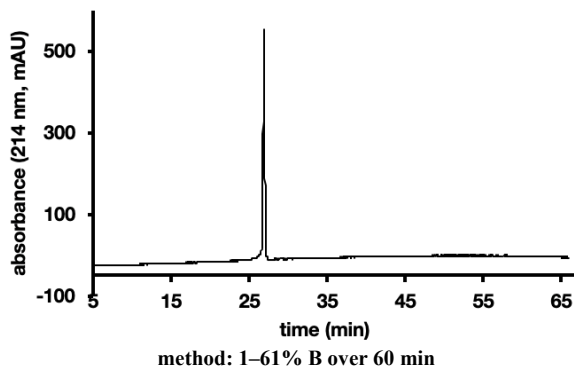


C) HPLC purification and analytical data for collagen

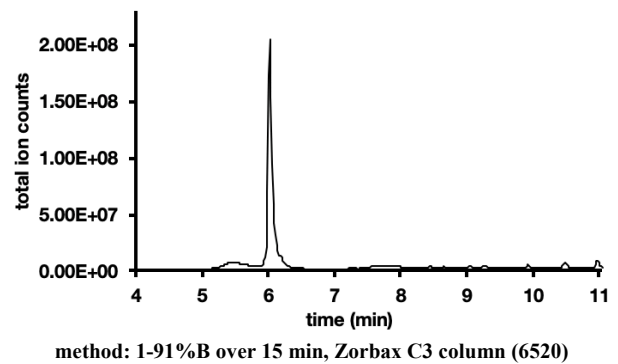
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 1-31%B with 0.5 B/min)

Yield past HPLC: 13.5 mg (3%)

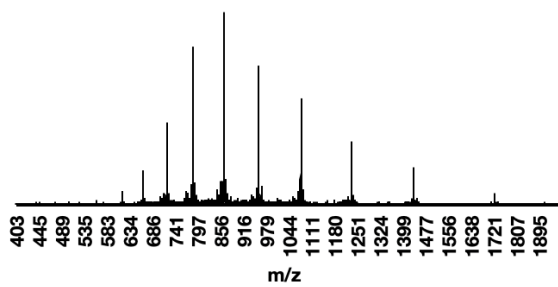
Analytical HPLC:



LC-MS:

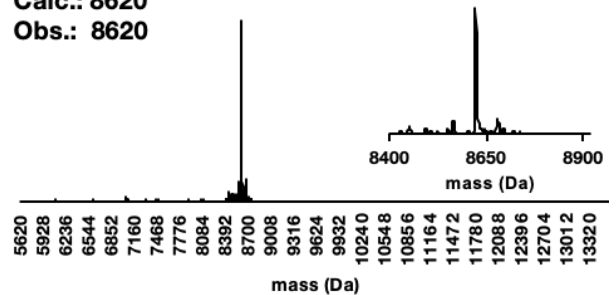


MS data from LC-MS:



Deconvolution of LC-MS data:

Calc.: 8620
Obs.: 8620



5.4 HIV-1 protease (Kent sequence)

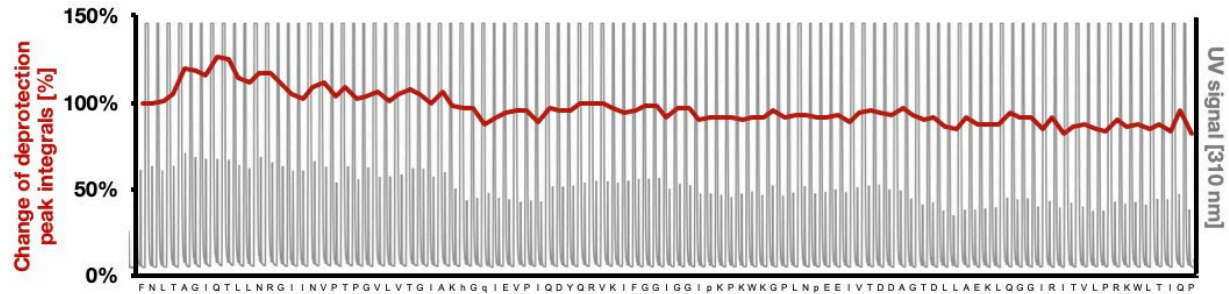
A) Synthesis Data

Sequence: PQITLWKRPL VTIRIGGQLK EALLDTGADD TVIEE (N1e)NLPG KWKPK (N1e) IGGI
GGFIKVRQYD QIPVEI (Abu)GHK AIGTVLVGPT PVNIIGRNLL TQIGATLNF (99AA)

Resin: 130 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 4.5 h

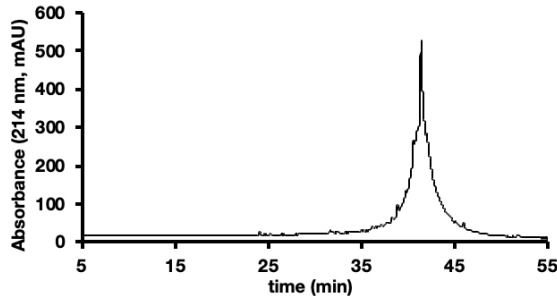
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude HIV-1 protease

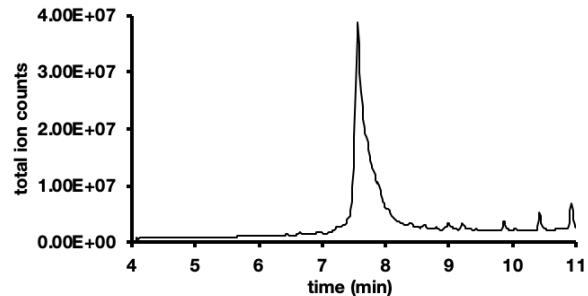
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



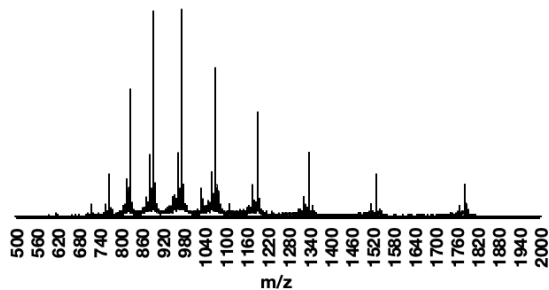
method: 5–65% B over 60 min

LC-MS:

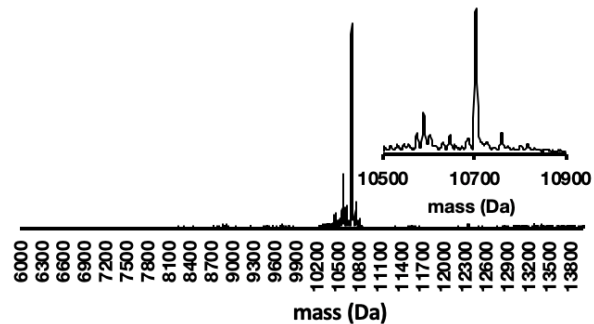


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

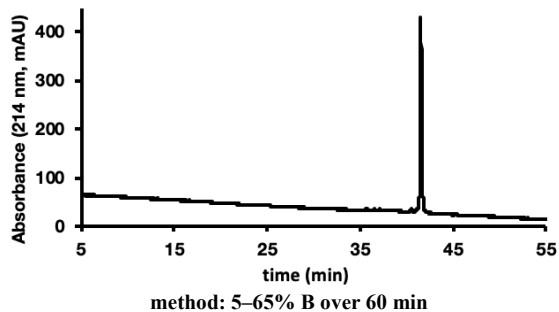


C) HPLC purification and analytical data for HIV-1 protease

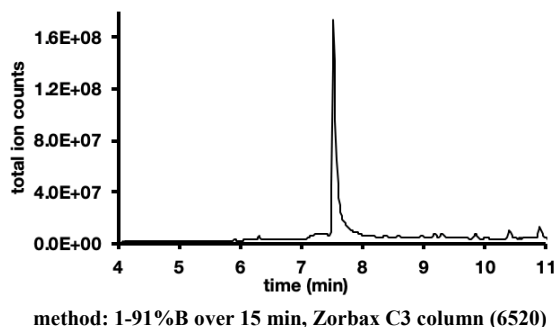
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C
(gradient: 5-29%B with 1% B/min, 29-49%B with 0.2% B/min, 49-65%
B with 1% B/min)

Yield past HPLC: 9.6 mg (3%)

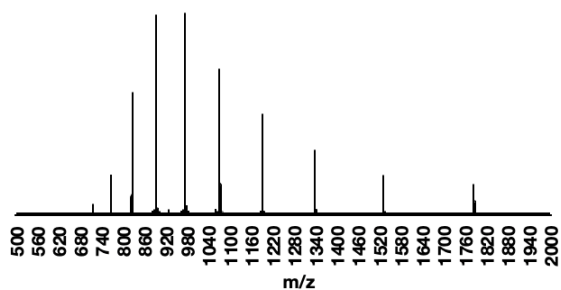
Analytical HPLC:



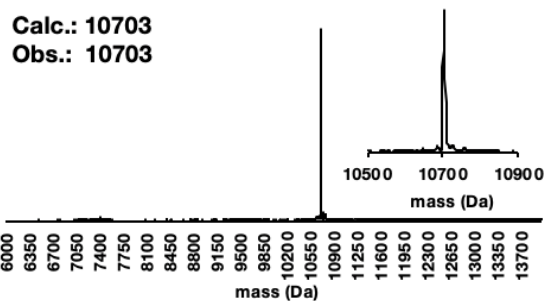
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



5.5 Barnase

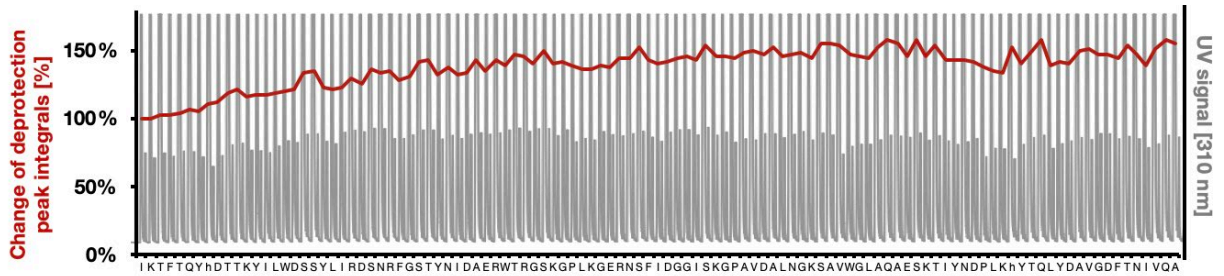
A) Synthesis Data

Sequence: AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS
IGGDIFSNRE GKLPKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT
DHYQTFTKIR (110 AA)

Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

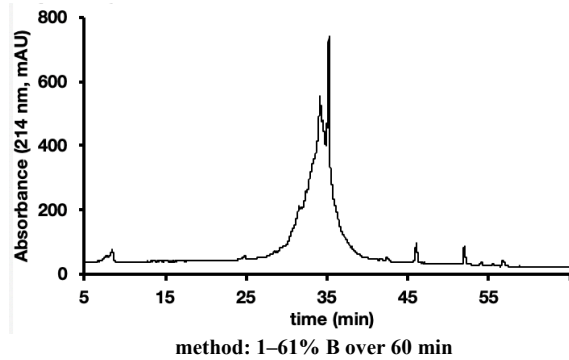
Synthesis UV trace from AFPS:



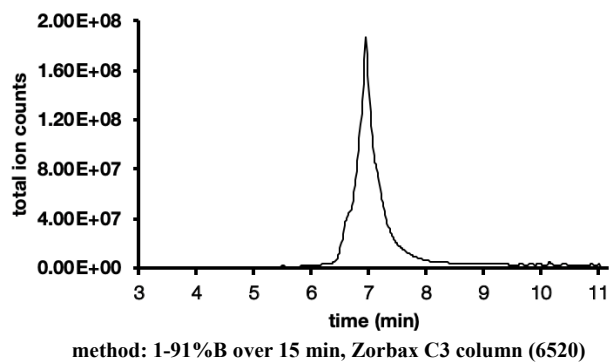
B) Cleavage and analytical Data for crude barnase

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

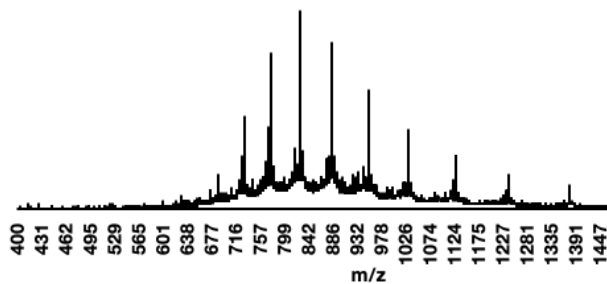
Analytical HPLC:



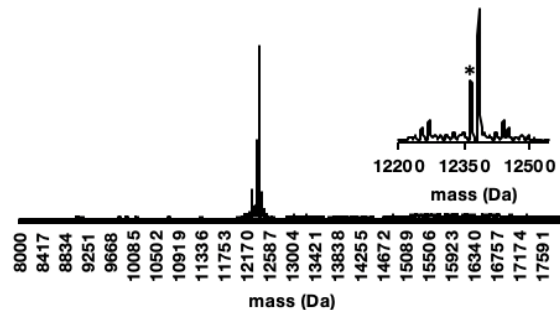
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

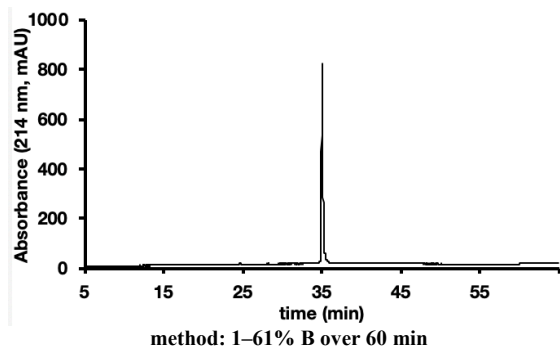


C) HPLC purification and analytical data for barnase

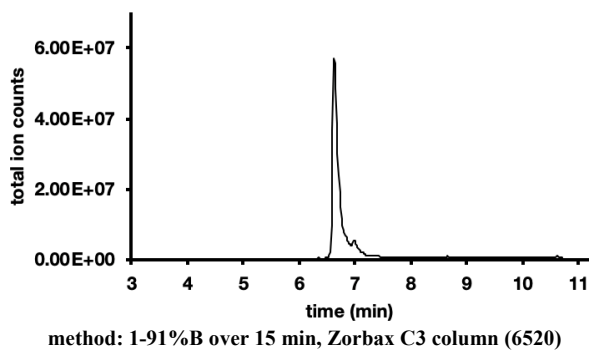
HPLC purification: Zorbax SB300 C3 column, 60 °C
Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C
(gradient: 5-25% B with 1% B/min, 25-45%B with 0.2% B/min, 45-65% B with 1%B/min)

Yield past HPLC: 8.4 mg (2%)

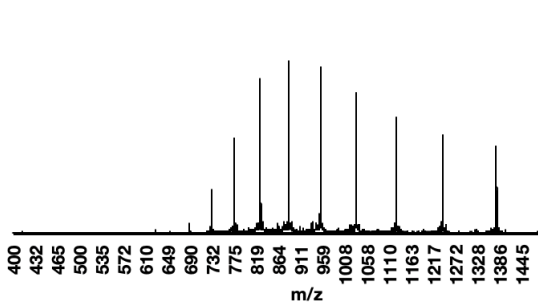
Analytical HPLC:



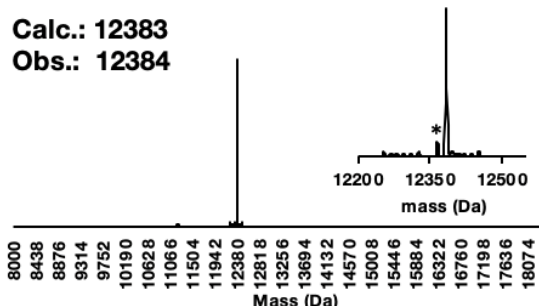
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) aspartimide by-product

5.6 Barnase (R110F_{Br})

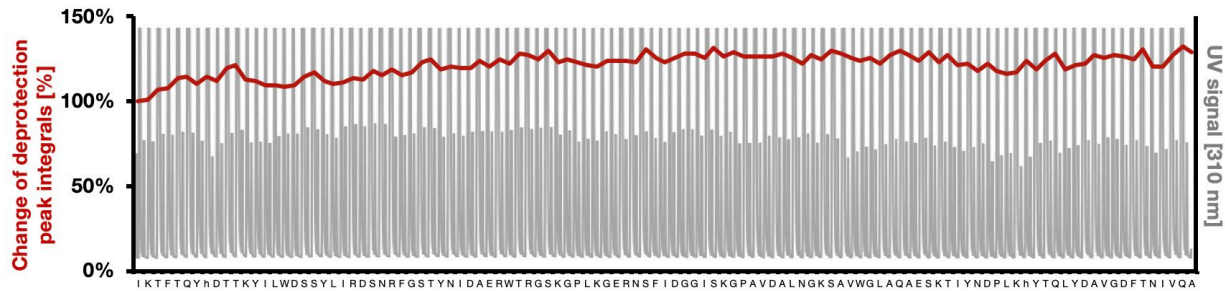
A) Synthesis Data

Sequence: AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS
IGGDIFSRE GKLPKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT
DHYQTFKIF_{Br} (110 AA)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to 4-bromophenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

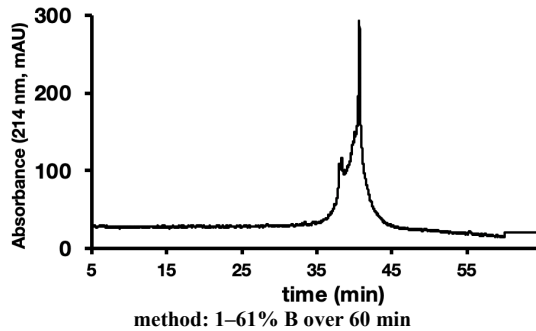
Synthesis UV trace from AFPS:



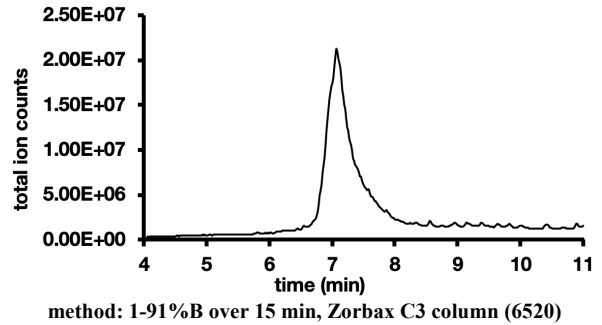
B) Cleavage and analytical Data for crude barnase(R110F_{Br})

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:

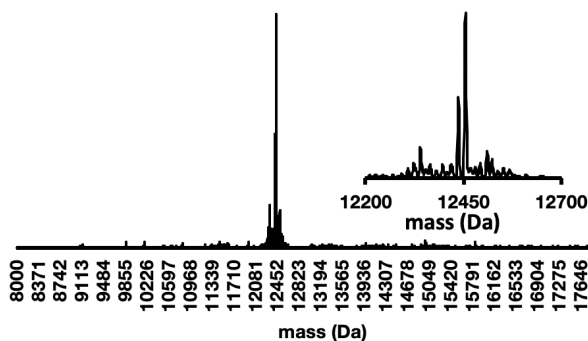
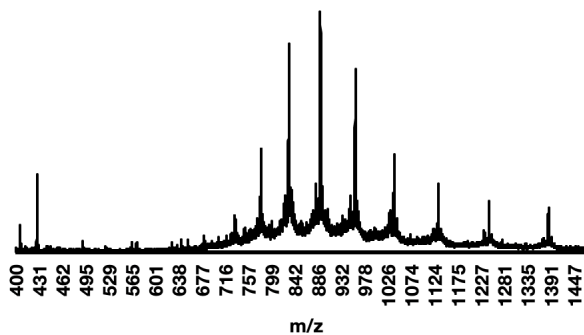


LC-MS:



MS data from LC-MS:

Deconvolution of LC-MS data:

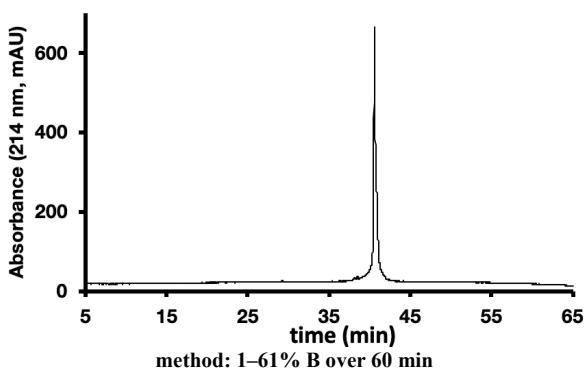


C) HPLC purification and analytical data for barnase(R110F_{Br})

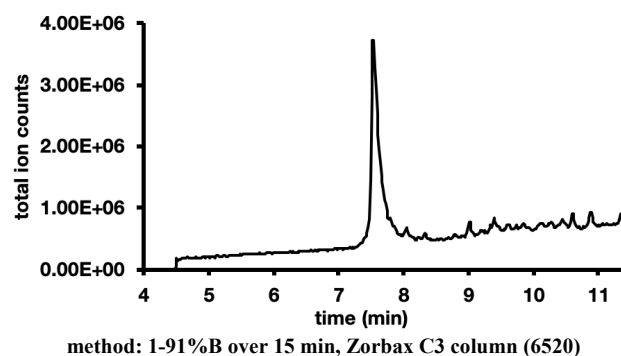
HPLC purification: Purified on a preparative Zorbax 300SB-C3 column at 25 °C at 20 mL/min. (gradient: 15-55%B with 0.33% B/min).

Yield past HPLC: 15 mg (3%)

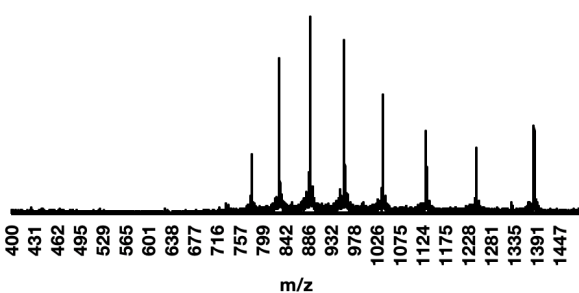
Analytical HPLC:



LC-MS:

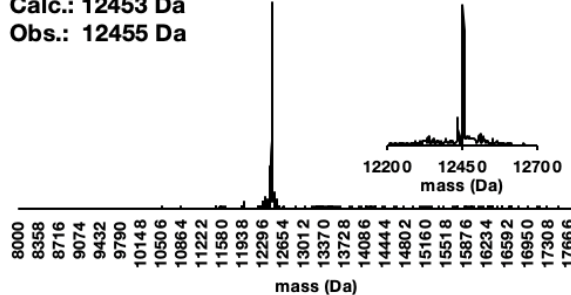


MS data from LC-MS:



Deconvolution of LC-MS data:

Calc.: 12453 Da
Obs.: 12455 Da



5.7 MDM2^[1-118]

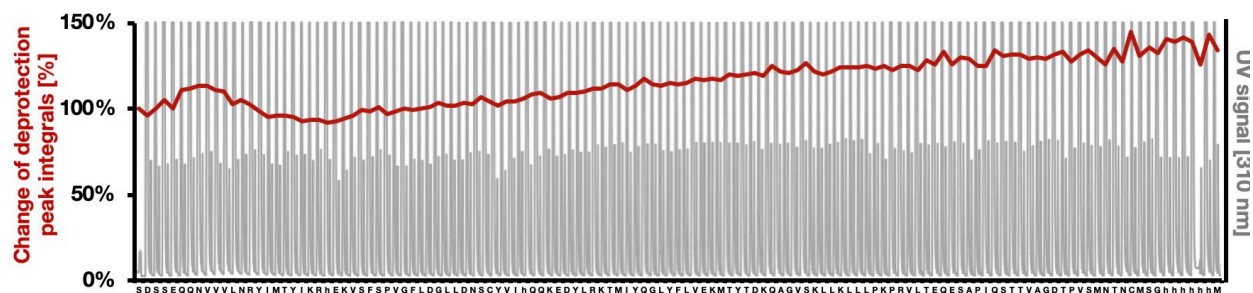
A) Synthesis Data

Sequence: MHHHHHGSM CNTNMSVPTD GAVTTSQIPA SEQETLVRPK PLLKLLKSV GAQKDTYTMK
EVLFYLGQYI MTKRLYDEKQ QHIVYGSNDL LGDLFGVPSF SVKEHRKIYT MIYRNLVVVN
QQESSDS (127 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 5 h

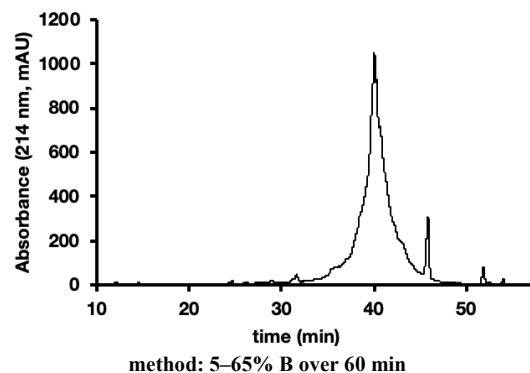
Synthesis UV trace from AFPS:



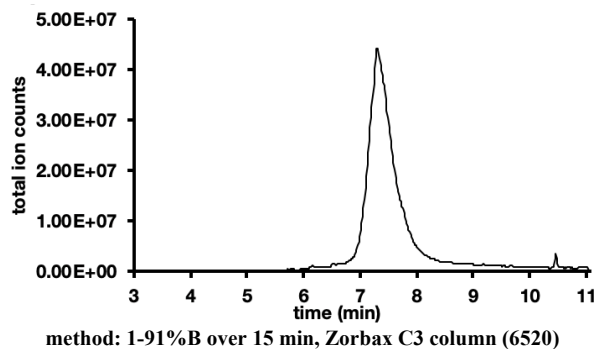
B) Cleavage and analytical Data for crude MDM2^[1-118]

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

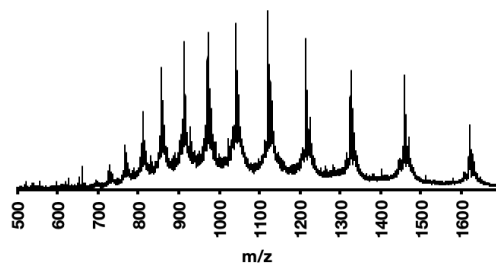
Analytical HPLC:



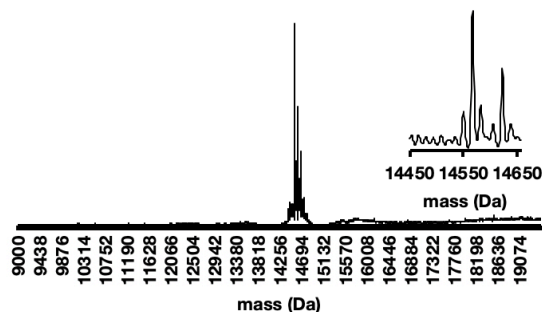
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:

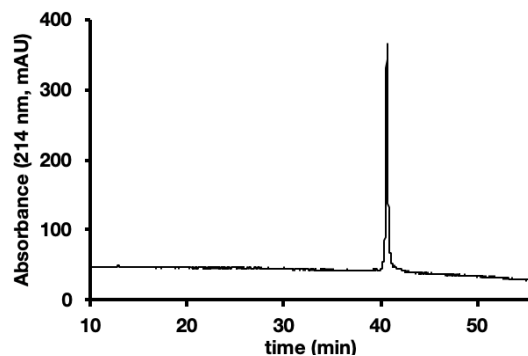


C) HPLC purification and analytical data for MDM2^[1-118]

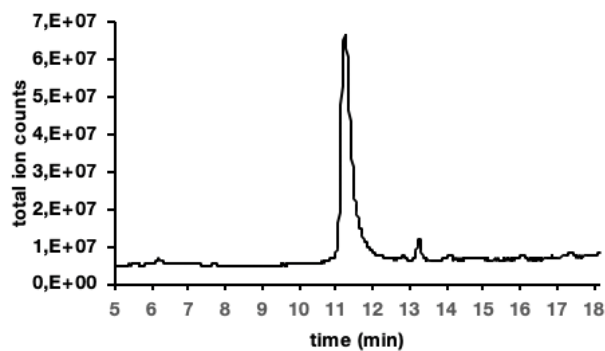
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-29% B with 1% B/min, 29-49%B with 0.2% B/min, 49-65% B with 1%B/min)

Yield past HPLC: 14 mg (3%)

Analytical HPLC:

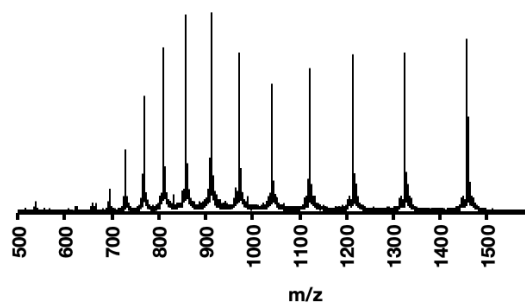


method: 5-65% B over 60 min



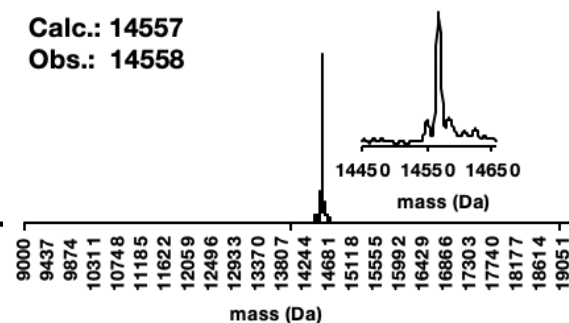
method: 1-91%B over 20 min, Zorbax C3 column (6550)

MS data from LC-MS:



Deconvolution of LC-MS data:

Calc.: 14557
Obs.: 14558



Comments:

*) oxidation by-products detected

5.8 Lysozyme

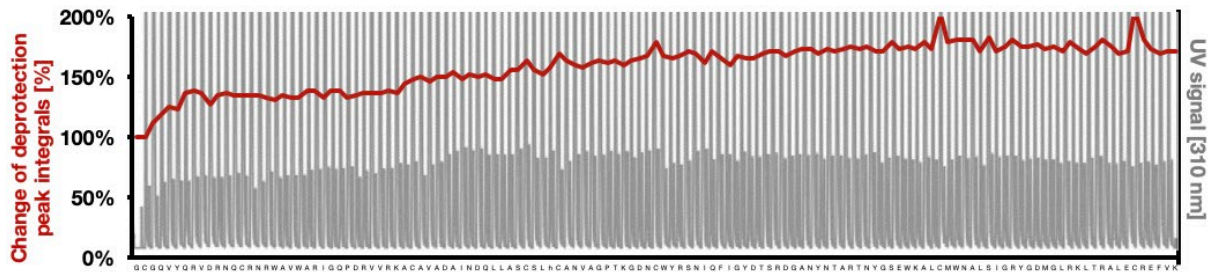
A) Synthesis Data for lysozyme

Sequence: KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR
STDYGFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV
RDPQGIRAW AWRNRCQNRD VRQYVQGGCV (129 AA)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to valine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 5.5 h

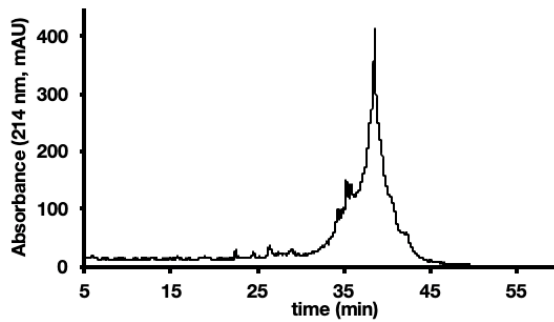
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude lysozyme

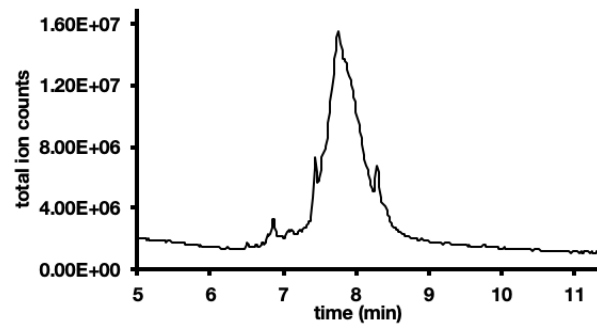
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



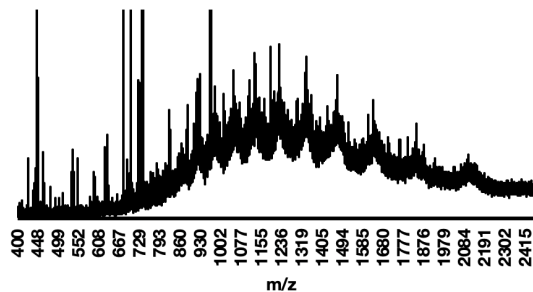
method: 5–65% B over 60 min

LC-MS:

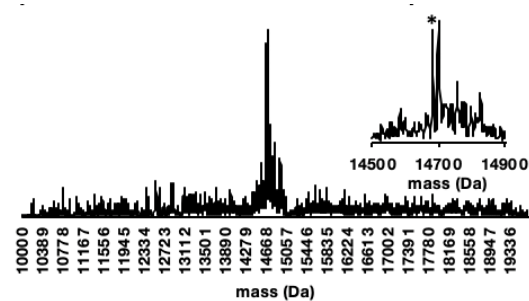


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

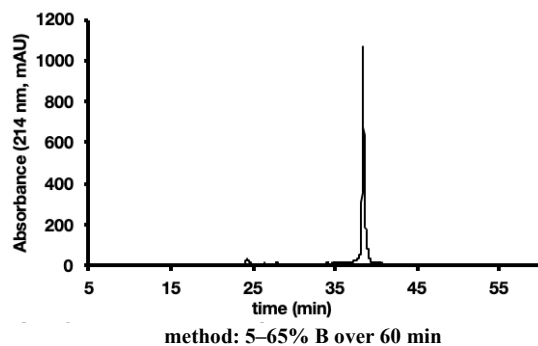


C) HPLC purification and analytical data for lysozyme

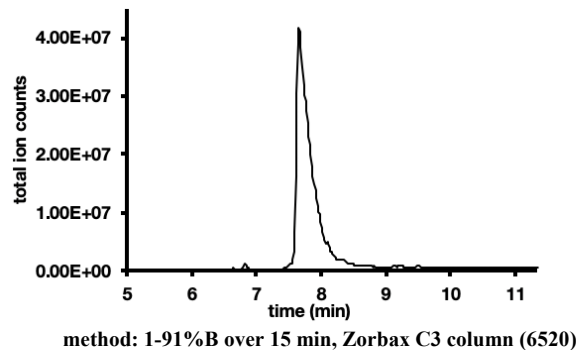
HPLC purification: Purified on Zorbax 300SB-C3 semipreparative column; Gradient: 5-28%B at 1%B/min; 29 to 49%B at 0.2% B/min; 50-65%B at 1%B/min.

Yield past HPLC: 3.2 mg (1%)

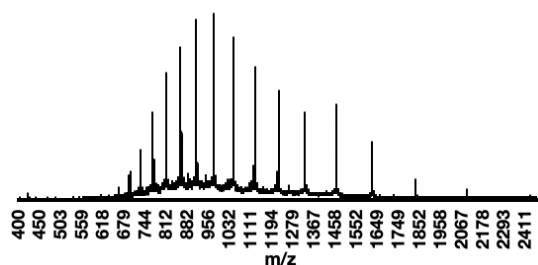
Analytical HPLC:



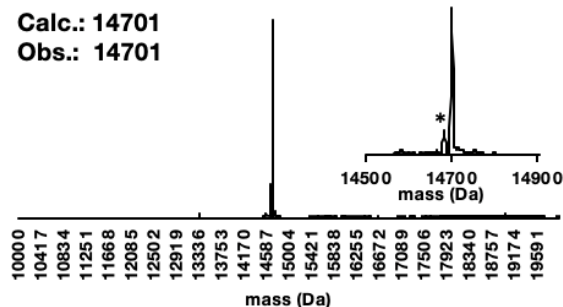
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) aspartimide by-product

5.9 FGF1

A) Synthesis Data for FGF1

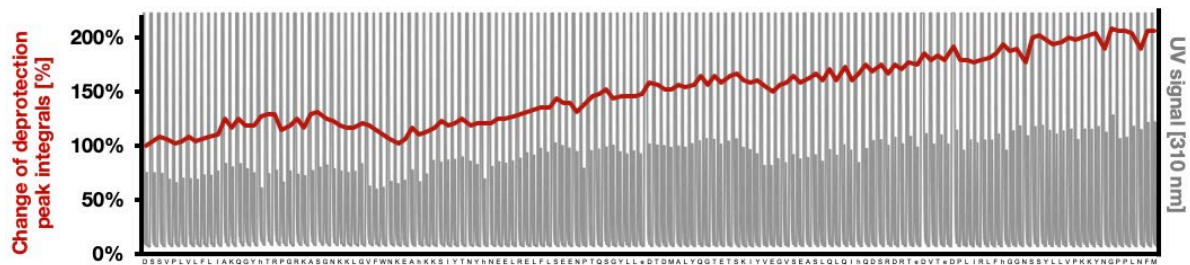
Sequence: MFNLPPEGNYK KPVLLYSSNG GHFLRILPDG TVDGTDRSD QHIQLQLSAE
SVGEVYIKST ETGQYLAMDT DGLLYGSQTP NEESLFLERL EENHYNTYIS
KKHAEKNWFV GLKKNQSAKR GPRTHYGQKA ILFLVLPVSS D (141 AA)

Note: the three glycines (G) highlighted in blue were introduced as Fmoc-(Dmb)Gly-OH building blocks

Resin: 100 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 5.7 h

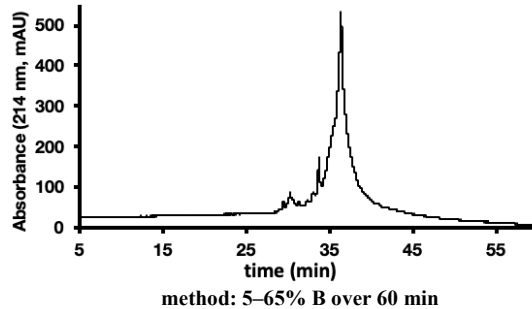
Synthesis UV trace from AFPS:



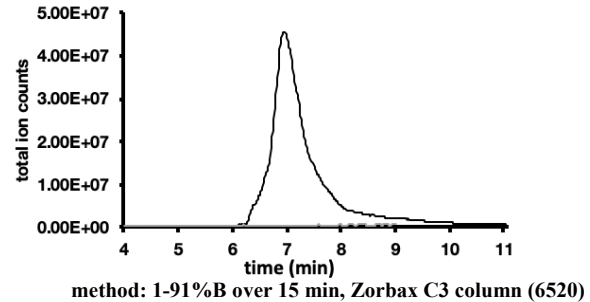
B) Cleavage and analytical Data for crude FGF1

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:

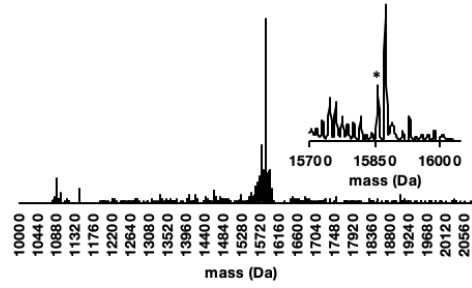
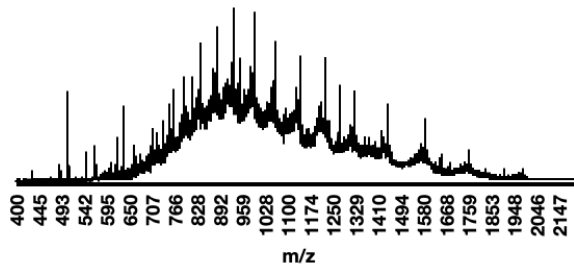


LC-MS:



MS data from LC-MS:

Deconvolution of LC-MS data:

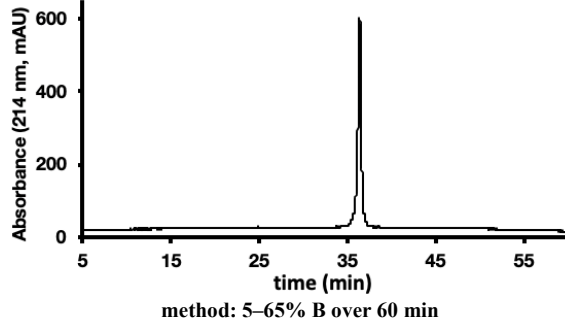


C) HPLC purification and analytical data for FGF1

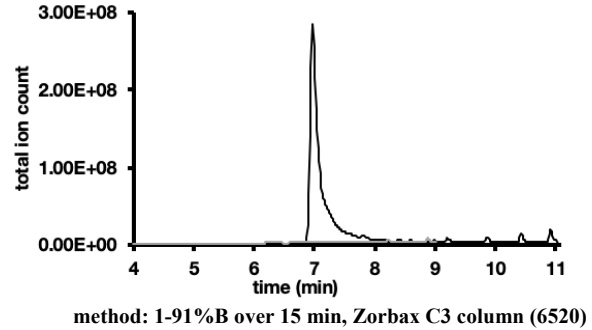
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-29%B with 1% B/min, 29-49%B with 0.2% B/min, 49-65%B with 1%B/min)

Yield past HPLC: 11 mg (4%)

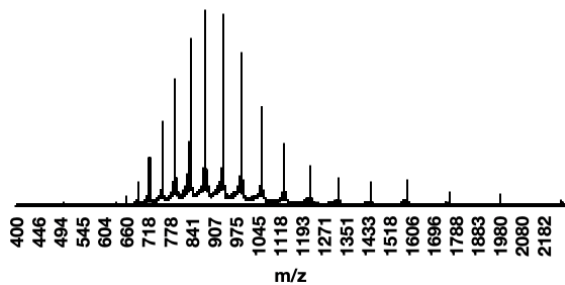
Analytical HPLC:



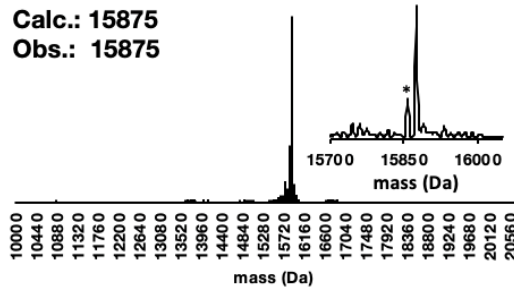
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *)aspartimide by-product is present

5.10 Sortase A₅₉₋₂₀₆;P94S/D160N/K196T

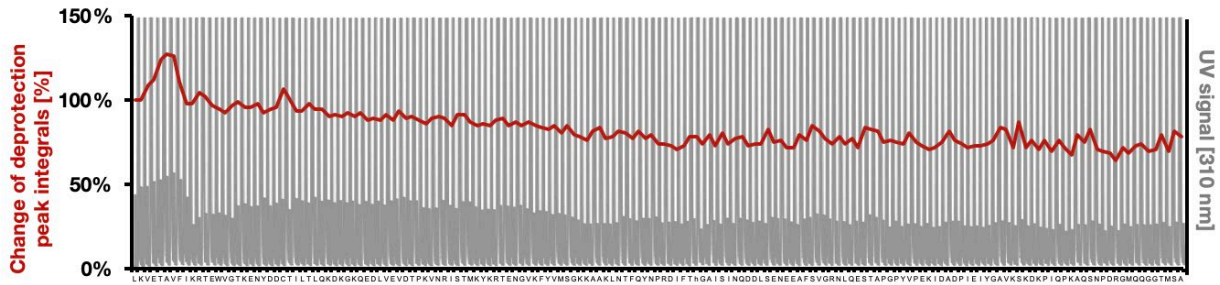
A) Synthesis Data

Sequence: ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS
EQLNRGVSFA EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF
KVGNETRKYK MTSIRNVKPT DVEVLDEQKG KDKQLTLITC DDYNEKTGVW
ETRKIFVATE VKLE (164 AA)

Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to glutamic acid, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 6.5 h

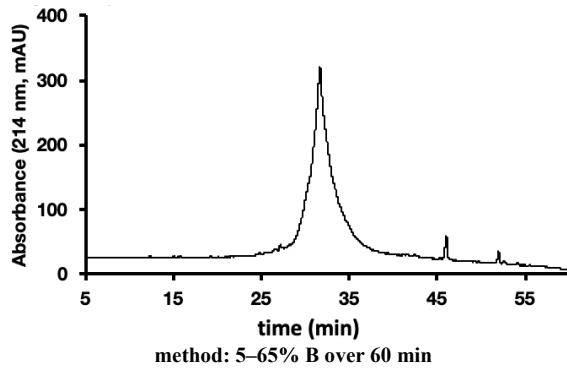
Synthesis UV trace from AFPS:



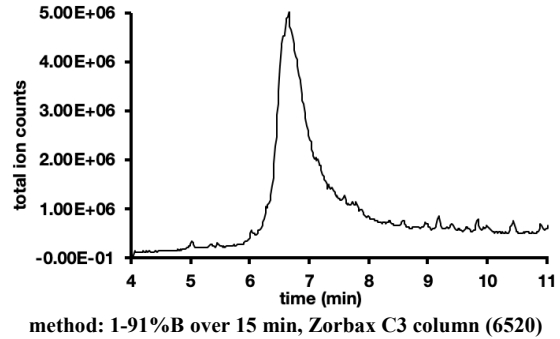
B) Cleavage and analytical Data for crude sortase A₅₉₋₂₀₆;P94S/D160N/K196T

Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:

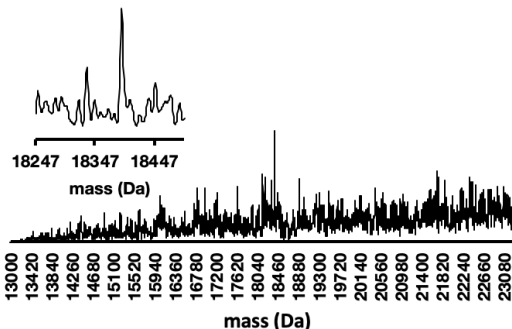
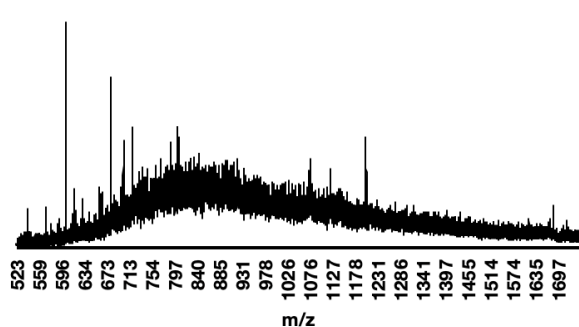


LC-MS:



MS data from LC-MS:

Deconvolution of LC-MS data:

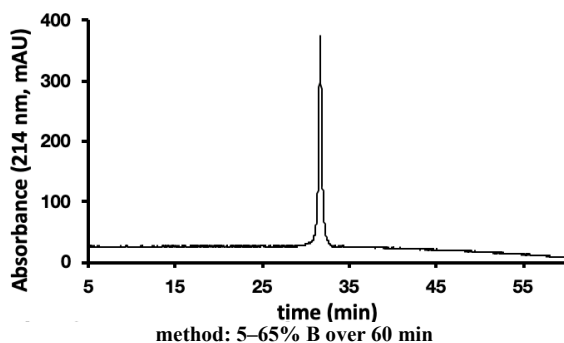


C) HPLC purification and analytical data for sortase A₅₉₋₂₀₆;P94S/D160N/K196T

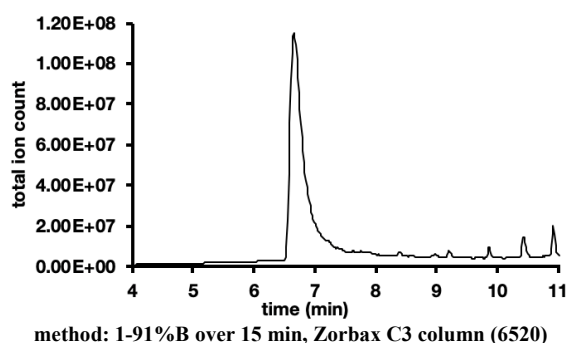
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-24%B with 1% B/min, 24-44%B with 0.2% B/min, 44-65%B with 1%B/min)

Yield past HPLC: 6.4 mg (1%)

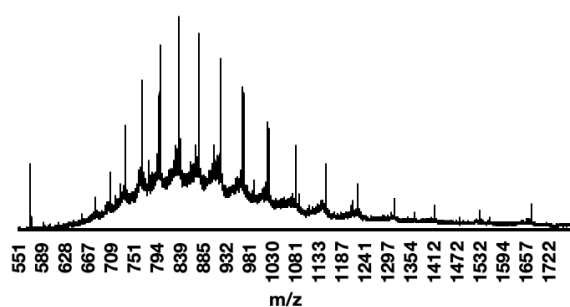
Analytical HPLC:



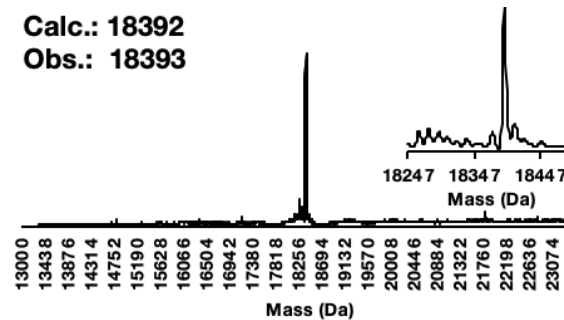
LC-MS:



MS data from LC-MS:



Deconvolution of LC-MS data:



6 Folding, Purification and Characterization of synthetic proteins

6.1 Barnase and barstar

6.1.1 Expression and SEC purification of recombinant barnase

Barnase was expressed based on previous report from Okorokov et al.⁽⁵⁵⁾ In short, *E. Coli* strain XL-1 Blue was transformed with plasmid pMT1002 (Addgene plasmid 8621). Next, 1 mL of 8-h culture of XL-1 Blue cells carrying the plasmid grown in LB medium with carbenicillin (100 µg/mL), was diluted 1:1000 into the same medium (1L) and grown at 28 °C, shaking at 200 rpm. The culture had a density of ~OD600 = 0.6 in approximately 24 h, at which point a further 500 mL of preheated LB medium (85 °C) was poured into the cell culture, and the shaker was incubated at 42 °C for 30 minutes. Next, the temperature was adjusted to 37 °C and cells were cultured for another 18 hours. Acetic acid was added to the culture until a final pH of ~4.3 as determined by pH paper. The cells were centrifuged for 30 minutes at 8000 rpm. The pellet was discarded while the supernatant was divided into 4000 mL portions, flash frozen with liquid nitrogen and kept at -80 °C until purification.

For purification, 400 mL of the supernatant was thawed on ice and buffer-exchanged into buffer SPA (50 mM MES, pH 6.3) using Amicon 3K concentrator (15 mL, EMD Millipore). The resulting solution was loaded onto two Capto S columns attached in series (5 mL, GE Healthcare) and the desired protein was eluted with a gradient of SPB (50 mM MES, pH 6.3, 1 M NaCl) in SPA. The fractions containing the desired protein were combined and concentrated using Amicon 3K concentrator (15 mL, EMD Millipore). The protein was then filtered using a 0.22 µm filter and purified via size exclusion chromatography with buffer P (20 mM Tris, pH 7.5, 150 mM NaCl). The fractions containing pure barnase were combined and concentrated using Amicon 3K concentrators and the molecular weight of the final protein was determined by LC-MS (**Fig. S10**). Overall, 3 mg of recombinant barnase was isolated from 400 mL of the culture supernatant.

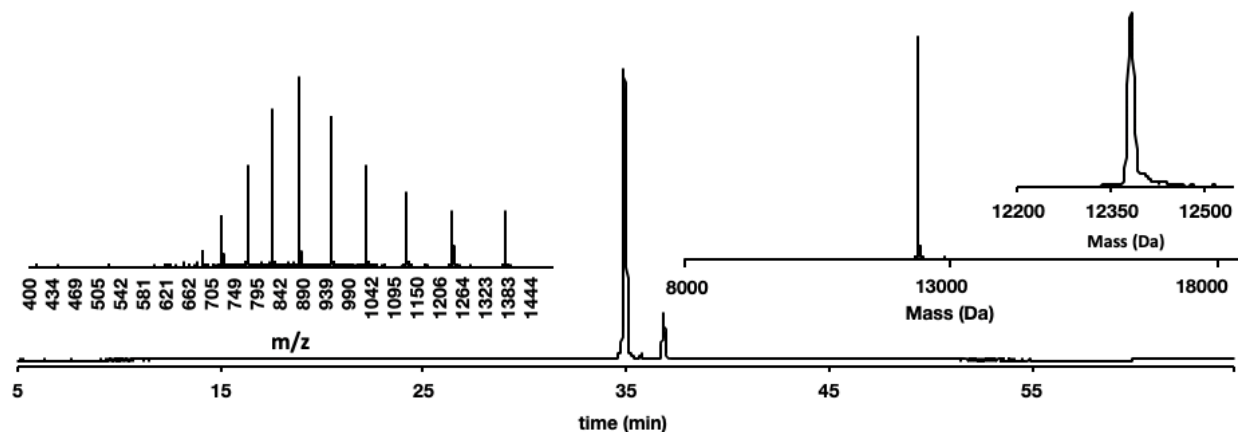


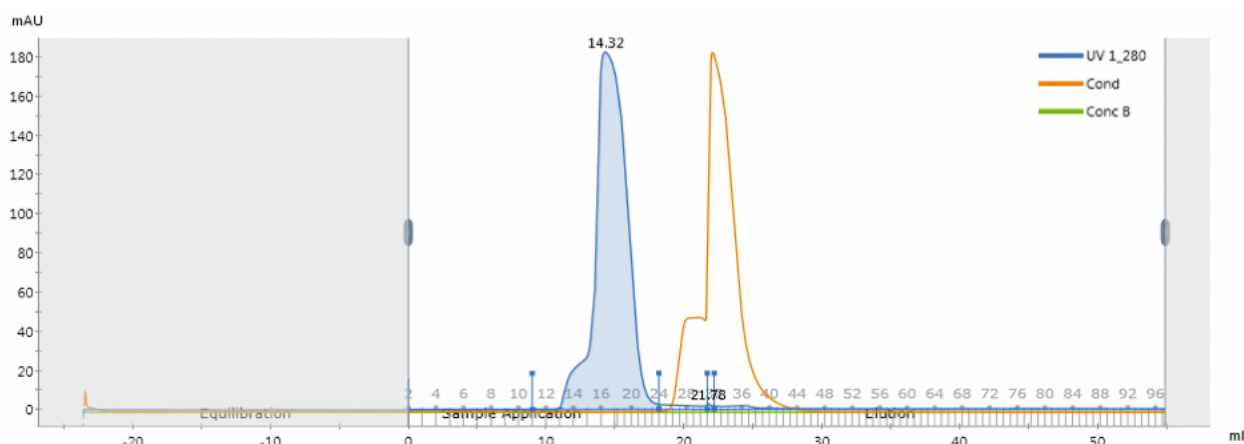
Fig. S10. Expression and purification yields recombinant barnase for comparison with synthetic enzyme. Analytical HPLC traces of purified recombinant barnase (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12383 Da, Obs. 12384 Da.

6.1.2 Folding and SEC purification of synthetic WT barnase

The folding procedure of WT barnase was adapted from Mong *et al.*(18) 2.5 mg of WT barnase was dissolved in 150 μ L of 6 M Gn·HCl solution in 50 mM NaH₂PO₄ RNase free¹ buffer, pH = 7.8. The mixture was serially diluted in four steps to 0.3 M Gn·HCl using a buffered RNase free solution (150 mM Tris buffer pH 7.5, 150 mM NaCl). The final solution was then filtered and subjected to size exclusion chromatography (SuperdexTM 75 Increase 10/300 GL, 0.25 mL/min), using isocratic eluent 150 mM Tris buffer pH 7.5, 150 mM NaCl, RNase free. The elution profile is depicted in **Fig. S11**.

Fractions containing the protein were concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at -80 °C. A total of 0.8 mg of WT barnase was isolated, corresponding to a 33% isolated yield. The purity of the final product was assessed by LCMS and HPLC.

A) SEC purification of synthetic barnase (wild type sequence)



B) Analytical data for purified synthetic barnase past SEC (wild type sequence)

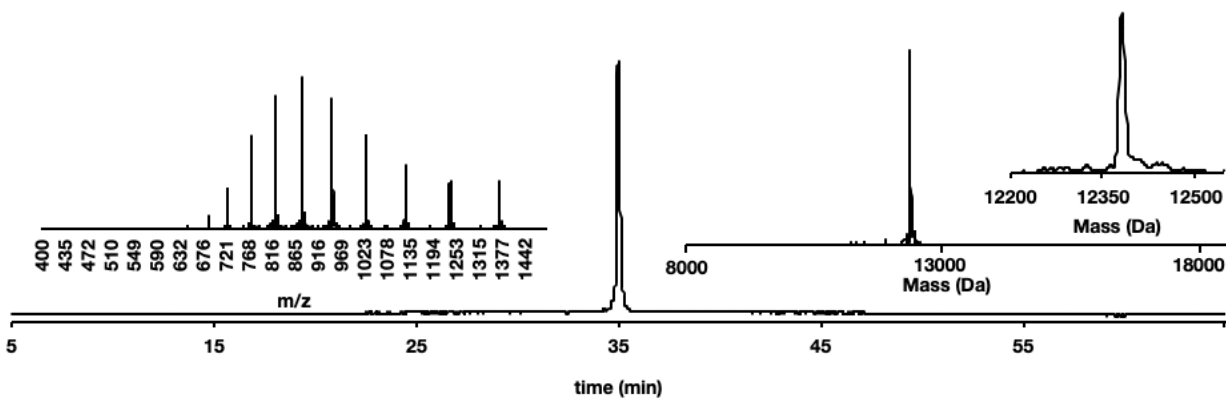
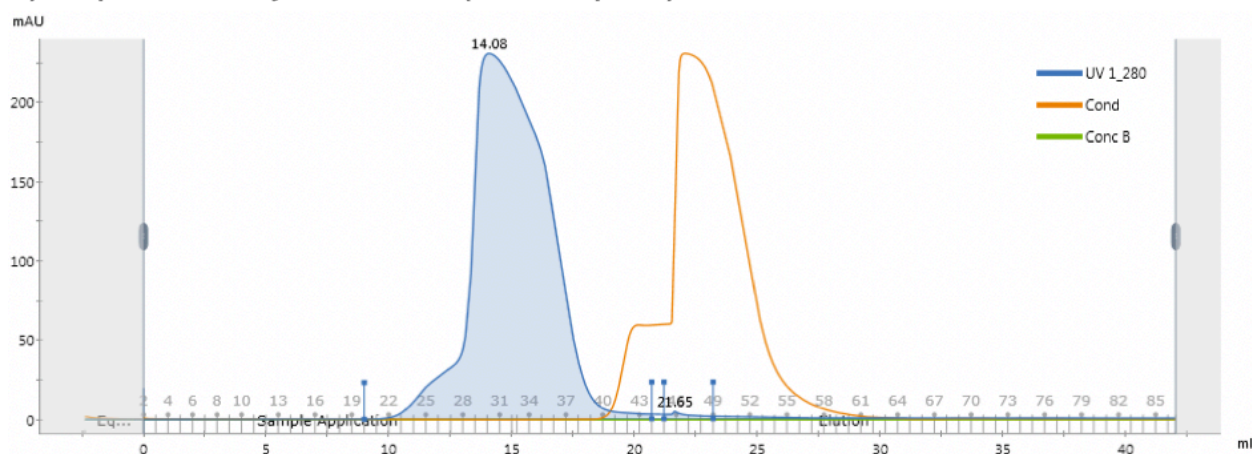


Fig. S11. Folding and SEC purification of synthetic barnase yields folded protein indistinguishable from recombinant barnase. A) HPLC-purified barnase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) analytical HPLC traces of purified synthetic barnase (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12383 Da, Obs.: 12384 Da.

6.1.3 Folding and SEC purification of synthetic barnase R110F_{Br}

The folding procedure of barnase R110F_{Br} protein was adapted from Mong *et al.*(18) 5.9 mg of barnase R110F_{Br} variant was dissolved in 150 μ L of 6 M Gn·HCl solution in 50 mM NaH₂PO₄ RNase free⁴ buffer, pH = 7.8. The mixture was then serially diluted in four steps to 0.3 M Gn·HCl using a buffered RNase free solution (150 mM Tris buffer pH 7.5, 150 mM NaCl). The final solution was then filtered and subjected to size exclusion chromatography (Superdex™ 75 Increase 10/300 GL, 0.25 mL/min), using isocratic eluent 150 mM Tris buffer pH 7.5, 150 mM NaCl, RNase free. The elution profile is depicted in **Fig. S12**. Fractions containing the protein were concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at -80 °C. A total of 1.3 mg of barnase R110F_{Br} was isolated, corresponding to 22% yield. The purity of the final product was assessed by LCMS and HPLC.

A) SEC purification of synthetic barnase (R110F_{Br} sequence)



B) Analytical data for purified synthetic barnase past SEC (R110F_{Br} sequence)

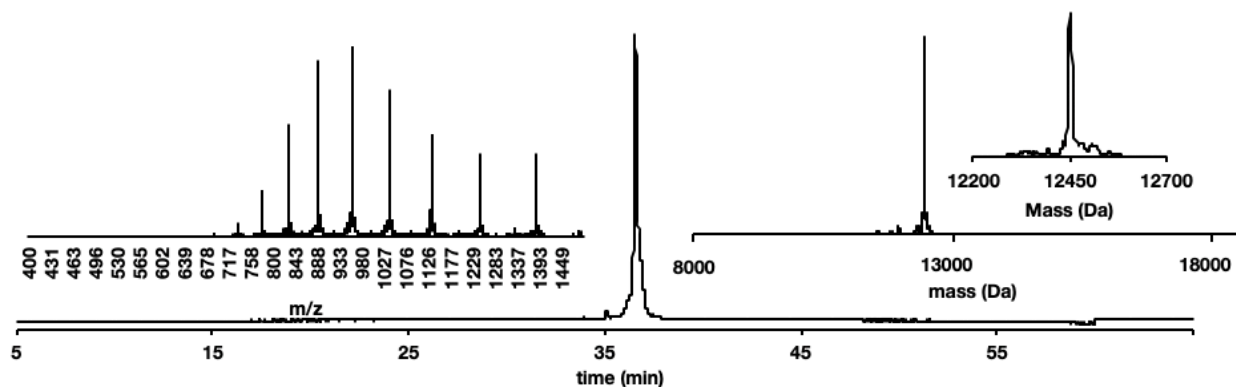


Fig. S12. Folding and SEC purification of synthetic barnase R110F_{Br} yields clean protein sample. A) HPLC-purified barnase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) analytical HPLC traces of purified synthetic barnase R110F_{Br} (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12453 Da, Obs.: 12452 Da.

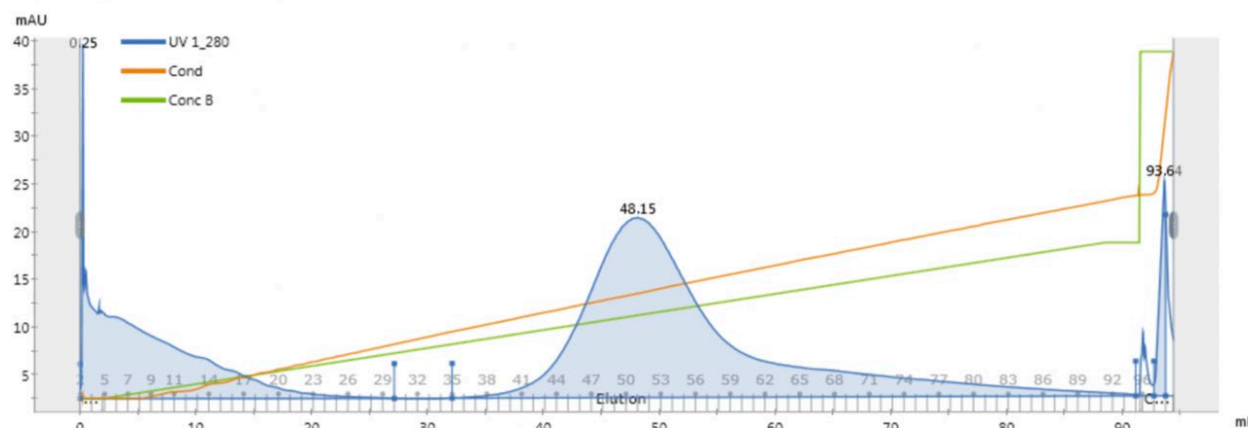
⁴ RNase free water was prepared by treatment of Milli-Q® water with DEPC at 1000:1 (v/v) for 2 hours at 37°C, followed by autoclaving the mixture for 20 minutes

6.1.4 Folding and IEC purification of synthetic WT Barstar

The folding procedure of WT Barstar was adapted from Udgaonkar *et al.* (1-3), (56) 3.0 mg of HPLC-purified WT barstar was dissolved in 1050 μL of a 6 M Gn·HCl solution in 5 mM NaH_2PO_4 , 250 μM EDTA, 250 μM DTT RNase free buffer, pH 8. The mixture was serially diluted in ten steps to 0.6 M Gn·HCl using a buffered RNase free solution (5mM NaH_2PO_4 , 250 μM EDTA, 250 μM DTT, pH 8). The resulting solution then desalted using a 7K Zeba™ Spin Desalting Column and filtered, and subjected to anion exchange chromatography (GE Healthcare HiTrap™ Q Sepharose Fast Flow IEX Column), using gradient elution with RNase free mobile phases 20 mM Tris buffer pH 7.4 and 20 mM Tris, 1 M NaCl pH 7.4 (gradient: 0-45% at 0.5% B/min). The elution profile is depicted in **Fig. S13**.

Fractions containing the protein were concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at -80°C . A total of 0.14 mg of WT barstar was isolated corresponding to 5% isolated yield. The purity of the final product was assessed by LCMS and HPLC.

A) SEC purification of synthetic barstar



B) Analytical data for purified synthetic barstar

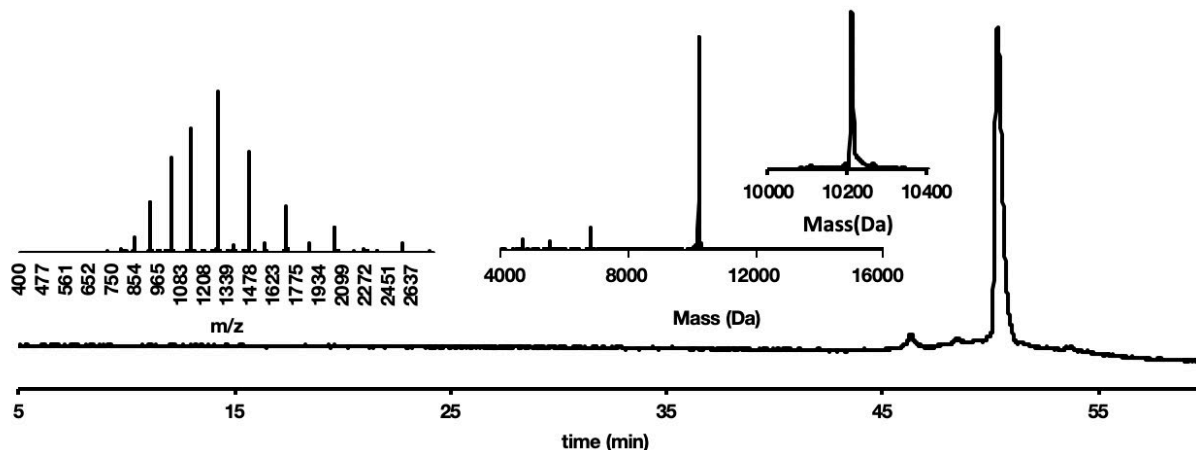


Fig. S13. Folding and IEC purification of synthetic barstar. A) HPLC-purified barstar was subjected to folding conditions followed by anion exchange chromatography. The elution profile from the anion exchange chromatography is displayed in this figure. B) analytical HPLC traces of purified synthetic barstar (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:10211 Da, Obs.: 10211 Da.

6.1.5 RNA hydrolysis assay

Hydrolysis of RNA by the synthesized barnase proteins was used to assess the activity of these proteins. Substrate RNA* was purchased from Millipore-Sigma as PAGE-purified dry solid. The RNA was dissolved in a solution of 10 mM Tris, 1 mM EDTA, pH 7.7 according to the manufacturer's recommendation and stored in 0.5 µg/µL aliquots at -80 °C.

RNA hydrolysis was initiated by addition of the RNA to a solution of barnase variants (recombinant WT, synthetic WT, synthetic R110F_{Br}) in 10 mM Tris, 1 mM EDTA, pH 7.7 buffer, and left at room temperature for 20 minutes. The total reaction volume was 10 µL, with 7 µM final concentration of RNA and 3 nM concentration of barnase. The reactions were quenched at the time points by addition of 10 µl of TBE-Urea gel loading dye (LC6876, Thermofisher Scientific) and rapidly frozen on liquid nitrogen.

Barnase activity was inhibited by its known inhibitor barstar to ensure any RNase activity observed in the assay was barnase-specific. Prior to the introduction of RNA substrate, the barnase variants (recombinant WT, synthetic WT, synthetic R110F_{Br}) were incubated with either recombinant barstar (**Fig. S14**), or synthetic barstar (**Fig. S15**), for 10 minutes on ice. The final concentration of RNA in the reaction was 7 µM and the concentration of barnase was 3 nM. The concentration of recombinant barstar incubated in each reaction was either 0 µM, 4 µM, 9 µM or 18 µM. The concentration of synthetic barstar incubated in each reaction was either 0 nM, 3.5 nM, 35 nM or 350 nM.

Immediately prior to gel analysis, the quenched reaction mixtures were heated at 95 °C for 5 minutes. Analysis was done on 15% denaturing polyacrylamide gel (Novex™ TBE-Urea Gels, EC68855BOX, Thermofisher Scientific) using 180 V for 60 minutes. The gel was stained with ethidium bromide at 1 µg/mL for 30 minutes and then washed with water three times. The gel was visualized on the ChemiDoc gel imager (Bio-Rad).

(*) sequence of substrate RNA – bolded are the potential sites of cleavage by barnase.

CAACAUCUUG**CU**AUACAAUG**CCAAUCCAU**GCUACACUAC**GU**UACA

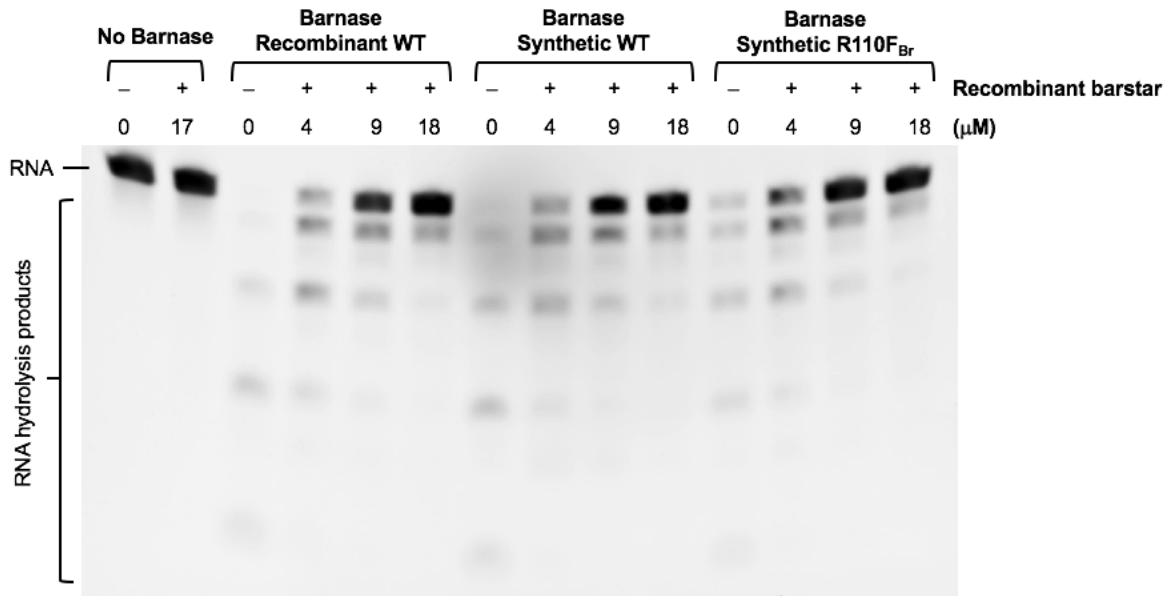


Fig. S14. RNA hydrolysis assay shows comparable RNase activity for recombinant WT barnase, synthetic WT barnase, and synthetic R110F_{Br} barnase. The synthetic barnase proteins are able to digest native RNA within 20 minutes. The dose-dependent inhibition of barnase activity upon pre-incubation with recombinant barstar protein points to association of these proteins.

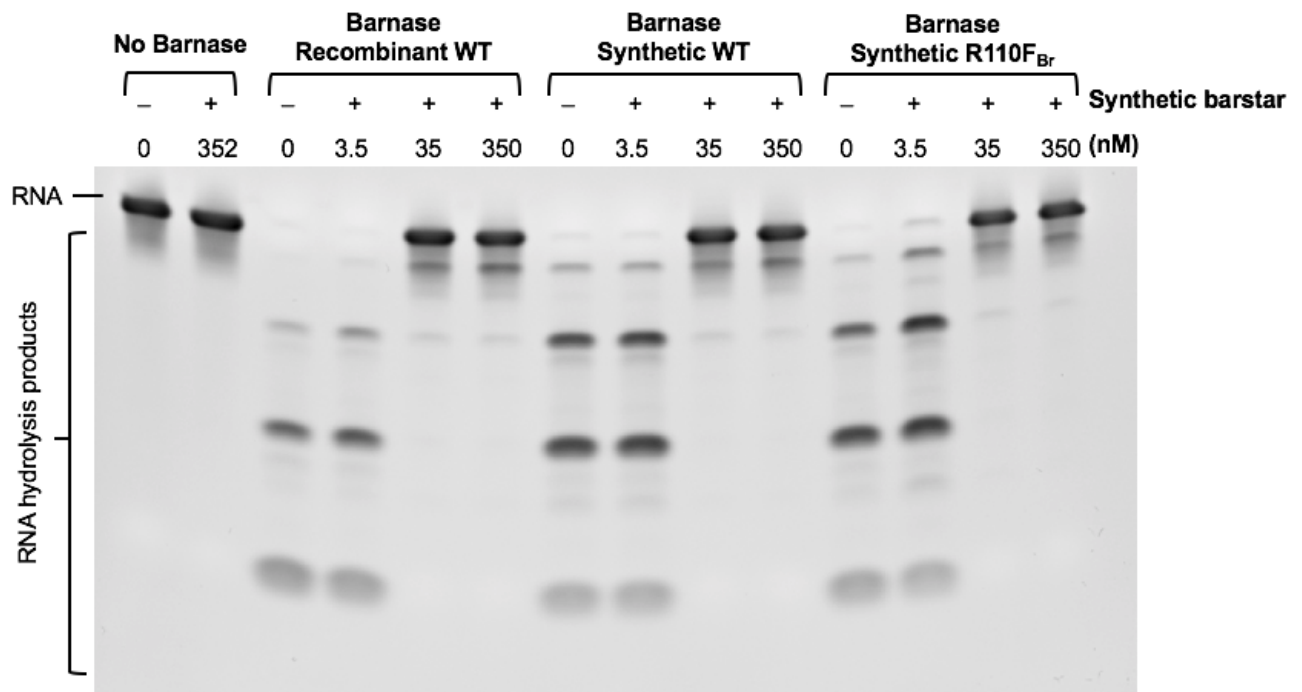


Fig. S15. Synthetic barstar inhibits the RNase activity of recombinant and synthetic barnase variants. Pre-incubation of barnase variants with synthetic barstar protein leads to dose-dependent inhibition of RNase activity of barnases.

6.1.6 Fluorogenic RNase activity assay

Hydrolysis of a fluorogenic substrate (6-FAM-dA^DrG^DdA^DdA^D-6-TAMRA) by the recombinant and synthesized barnases can be used to calculate the k_{cat}/K_M of the proteins based on the first order rate equation (1).

$$I = I_f - (I_f - I_0)e^{-\frac{k_{cat}}{K_M} [E](t-t_0)} \quad (1)$$

In this equation, I is the fluorescence at time t , I_0 is fluorescence of the intact substrate, I_f is the fluorescence of the hydrolyzed substrate, $[E]$ is the total enzyme concentration, k_{cat} , K_M are steady state enzyme kinetic parameters.(41, 46)

The fluorogenic substrate for the assay was purchased from ChemGenes as HPLC-purified solid. The substrate was dissolved in assay buffer solution (100 mM MES, 100 mM NaCl, pH 6.0) and kept in 200 μ M aliquots at -80 °C until use.

In each assay, the substrate was diluted in the assay buffer and the fluorescence of the substrate solution was monitored at 515 nm upon excitation at 495 nm using Tecan plate reader M1000 (**Fig. S16, Table S12**). The fluorescence was monitored every 10 seconds, after 3 second shake, for at least 500 seconds to measure the starting fluorescence, I_0 , and to ensure no background cleavage took place prior to the addition of enzymes. To start the cleavage reaction, an aliquot of enzyme was added to the substrate and the solution was rapidly mixed. The increase in fluorescence was then monitored for at least 1000 seconds. The concentration of the substrate in the reaction mixture at the time of addition of barnase was 200 nM and barnase was approximately at 2 nM. Total reaction volume was 250 μ L.

The experiment was performed in triplicates. Blank runs were performed in parallel with the experimental runs. For the blank runs, assay buffer was added to wells instead of barnase. Blank runs were otherwise set up identical to the experimental runs.

The fluorescence readings from experimental runs were first subtracted from the fluorescence readings of their blank. The data was normalized and then analyzed using the nonlinear regression, exponential, plateau followed by one phase association analysis using Prism 8 software from Graphpad. The K generated from this analysis corresponds to $k_{cat} [E]/K_M$ based on equation 1 above. The exact concentration of protein in each case was determined via absorbance at 280 nm using reported molar extinction coefficient values for barnase, 27411 $M^{-1} cm^{-1}$. (57)

Accordingly, $k_{cat}/K_M \pm SE$ of the synthetic and recombinant wild-type barnase were determined to be $(7.6 \pm 0.2) \times 10^6 M^{-1} s^{-1}$ and $(9.0 \pm 0.3) \times 10^6 M^{-1} s^{-1}$, respectively. The values match the literature reported $k_{cat}/K_M \pm SE$ of wild-type recombinant barnase $(1.3 \pm 0.4) \times 10^7$. (46) Furthermore, the $k_{cat}/K_M \pm SE$ of the R110F_{Br} synthetic variant barnase was calculated to be $(2.3 \pm 0.8) \times 10^6 M^{-1} s^{-1}$.

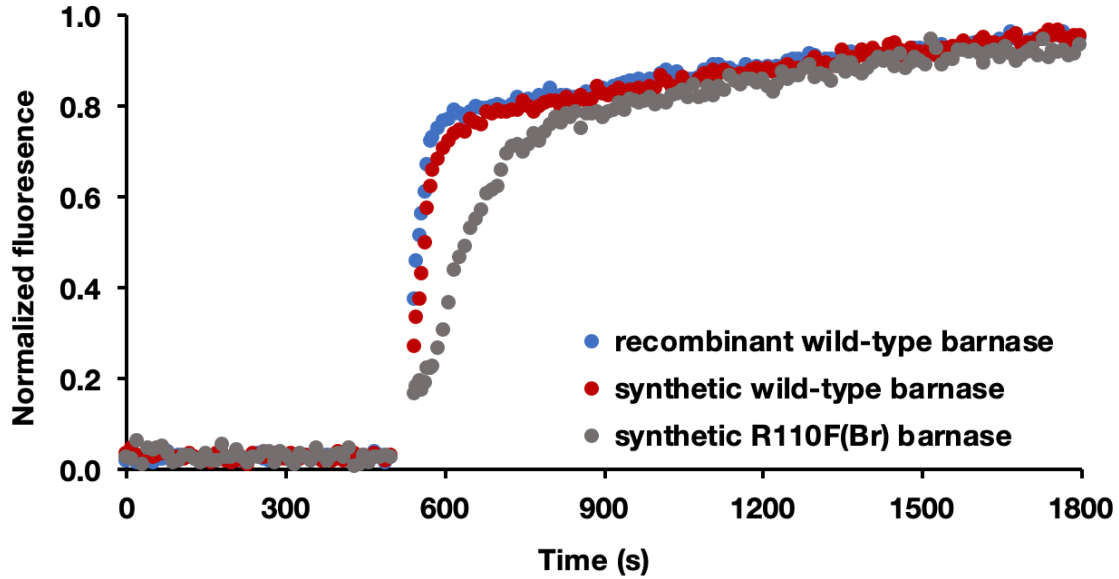


Fig. S16. Fluorogenic RNase activity determines the k_{cat}/K_M value of the barnase variants. A sample of the data produced from cleavage of substrate 6-FAM-dA^DrG^DdA^DdA^D-6-TAMRA with synthetic and recombinant barnase. At 500 s, barnase was added to the substrate. The increase in fluorescence indicates cleavage of substrate by the barnase protein.

	recombinant wild-type barnase			synthetic wild-type barnase			synthetic R110F(Br) barnase		
	Normalized Fluorescence			Normalized Fluorescence			Normalized Fluorescence		
time	mean	SD	N	mean	SD	N	mean	SD	N
0	0.020037	0.017308	3	0.034435	0.007823	3	0.028378	0.023046	3
10	0.02132	0.015889	3	0.046331	0.016813	3	0.022022	0.007184	3
20	0.015086	0.015035	3	0.030857	0.019189	3	0.06372	0.012992	3
30	0.021942	0.017726	3	0.027624	0.02024	3	0.012052	0.008125	3
40	0.026442	0.018517	3	0.032546	0.004795	3	0.04555	0.002984	3
50	0.016073	0.00999	3	0.025117	0.008412	3	0.044298	0.018378	3
60	0.036291	0.016581	3	0.031619	0.006571	3	0.045795	0.004736	3
70	0.023665	0.007032	3	0.035765	0.003841	3	0.049083	0.009578	3
80	0.036835	0.003653	3	0.03402	0.005061	3	0.034325	0.001505	3
90	0.020951	0.004466	3	0.027982	0.006634	3	0.01559	0.013845	3
100	0.030705	0.009477	3	0.027342	0.01022	3	0.017765	0.005798	3
110	0.030156	0.009942	3	0.024142	0.012263	3	0.031525	0.002104	3
120	0.024709	0.008243	3	0.032683	0.003773	3	0.028929	0.012342	3
130	0.024256	0.005014	3	0.02834	0.015835	3	0.023533	0.005085	3
140	0.017455	0.017315	3	0.018469	0.004924	3	0.015224	0.014591	3
150	0.029419	0.004726	3	0.026854	0.008953	3	0.033465	0.004178	3
160	0.017336	0.006205	3	0.0151	0.0181	3	0.018083	0.003085	3
170	0.029225	0.008823	3	0.032867	0.0199	3	0.018952	0.010341	3
180	0.033916	0.011287	3	0.03369	0.0175	3	0.052823	0.011273	3

190	0.02599	0.007404	3	0.017313	0.017574	3	0.026656	0.011297	3
200	0.016351	0.002159	3	0.016372	0.015168	3	0.025757	0.012297	3
210	0.020221	0.015403	3	0.020087	0.006475	3	0.044108	0.009565	3
220	0.018462	0.016127	3	0.023417	0.020989	3	0.028391	0.012195	3
230	0.018707	0.00495	3	0.010505	0.003587	3	0.013807	0.005961	3
240	0.031657	0.01581	3	0.033561	0.006329	3	0.027224	0.017443	3
250	0.035476	0.01616	3	0.032288	0.016513	3	0.025987	0.012745	3
260	0.022506	0.009581	3	0.025767	0.011322	3	0.038104	0.007806	3
270	0.020404	0.013912	3	0.026544	0.020409	3	0.038312	0.010112	3
280	0.028206	0.011171	3	0.033368	0.010215	3	0.016367	0.014619	3
290	0.029076	0.013204	3	0.022812	0.004462	3	0.040011	0.012231	3
300	0.025502	0.021369	3	0.021907	0.018448	3	0.029034	0.006337	3
310	0.018464	0.016006	3	0.032555	0.01751	3	0.018069	0.004663	3
320	0.030029	0.026137	3	0.026226	0.008982	3	0.008987	0.008256	3
330	0.025448	0.004314	3	0.023299	0.015746	3	0.033125	0.006954	3
340	0.028496	0.007847	3	0.028521	0.011935	3	0.026245	0.012744	3
350	0.021273	0.00765	3	0.026386	0.020324	3	0.019992	0.006786	3
360	0.0253	0.010263	3	0.035107	0.021094	3	0.045855	0.006217	3
370	0.022551	0.011895	3	0.020279	0.021479	3	0.031424	0.006564	3
380	0.029261	0.004228	3	0.022874	0.023375	3	0.018642	0.004327	3
390	0.028017	0.002704	3	0.018335	0.0139	3	0.031547	0.011161	3
400	0.036284	0.009215	3	0.037668	0.014133	3	0.034496	0.005877	3
410	0.023297	0.012709	3	0.017666	0.017925	3	0.025033	0.017906	3
420	0.03321	0.014274	3	0.030426	0.003552	3	0.045924	0.013302	3
430	0.023832	0.009899	3	0.034993	0.022073	3	0.006082	0.005307	3
440	0.020559	0.005707	3	0.018712	0.007472	3	0.00967	0.009478	3
450	0.029683	0.009379	3	0.027875	0.013458	3	0.029794	0.002359	3
460	0.021359	0.024315	3	0.024734	0.022602	3	0.014599	0.009345	3
470	0.03843	0.022088	3	0.035663	0.018396	3	0.024636	0.020606	3
480	0.031492	0.015798	3	0.02752	0.015067	3	0.028337	0.003348	3
490	0.014854	0.00971	3	0.019215	0.015841	3	0.029691	0.005326	3
500	0.02735	0.013916	3	0.031689	0.025211	3	0.025918	0.005519	3
543	0.376124	0.036526	3	0.271574	0.032096	3	0.167876	0.059504	3
548,1	0.459421	0.047275	3	0.336779	0.039623	3	0.181806	0.053791	3
553,3	0.515691	0.046399	3	0.374202	0.04335	3	0.194694	0.068961	3
558,5	0.561773	0.04446	3	0.429502	0.037342	3	0.176133	0.064137	3
563,6	0.612995	0.036094	3	0.500059	0.060141	3	0.188731	0.083854	3
568,8	0.670432	0.053508	3	0.577149	0.056928	3	0.221164	0.088519	3
574	0.723481	0.068816	3	0.621854	0.052468	3	0.222829	0.071291	3
580	0.730962	0.074672	3	0.659043	0.058191	3	0.225588	0.061999	3
590	0.751674	0.056278	3	0.682233	0.044692	3	0.268627	0.047726	3
600	0.765715	0.063007	3	0.70632	0.066043	3	0.30802	0.032704	3
610	0.772352	0.054325	3	0.723838	0.063124	3	0.365948	0.020063	3

620	0.792288	0.042424	3	0.739504	0.061735	3	0.439824	0.014831	3
630	0.78369	0.058984	3	0.7489	0.066118	3	0.467312	0.047987	3
640	0.774608	0.043515	3	0.745366	0.064396	3	0.491913	0.039488	3
650	0.787995	0.070144	3	0.76974	0.060571	3	0.531004	0.054388	3
660	0.798371	0.060211	3	0.763413	0.051354	3	0.550379	0.038997	3
670	0.797081	0.072206	3	0.759797	0.047309	3	0.5723	0.035127	3
680	0.796562	0.072284	3	0.786097	0.052311	3	0.606657	0.057093	3
690	0.79919	0.067915	3	0.785046	0.065701	3	0.614952	0.05746	3
700	0.802912	0.059498	3	0.791514	0.073707	3	0.622057	0.075278	3
710	0.801541	0.055021	3	0.789526	0.054387	3	0.657701	0.059946	3
720	0.79664	0.070561	3	0.785719	0.067301	3	0.693549	0.061446	3
730	0.802072	0.052877	3	0.790921	0.046656	3	0.709792	0.044731	3
740	0.81815	0.048111	3	0.791962	0.064269	3	0.71556	0.043365	3
750	0.807192	0.056963	3	0.810828	0.060985	3	0.700964	0.043515	3
760	0.817287	0.052582	3	0.797686	0.059471	3	0.716954	0.031734	3
770	0.811981	0.062775	3	0.785714	0.061371	3	0.740284	0.027179	3
780	0.82347	0.048289	3	0.799313	0.077963	3	0.724472	0.026804	3
790	0.807626	0.042271	3	0.809195	0.053371	3	0.741614	0.020228	3
800	0.840929	0.051624	3	0.810611	0.078902	3	0.759457	0.028248	3
810	0.817837	0.054528	3	0.809661	0.061736	3	0.776538	0.007411	3
820	0.823256	0.055015	3	0.79439	0.061698	3	0.761966	0.010145	3
830	0.822055	0.071487	3	0.820107	0.06036	3	0.782278	0.018724	3
840	0.820815	0.052934	3	0.812052	0.062859	3	0.776735	0.002586	3
850	0.818606	0.052253	3	0.808896	0.058742	3	0.786466	0.013331	3
860	0.825116	0.061536	3	0.823394	0.067653	3	0.752564	0.006443	3
870	0.832907	0.063302	3	0.811726	0.059798	3	0.782646	0.012596	3
880	0.829368	0.062037	3	0.814136	0.053184	3	0.78504	0.02902	3
890	0.844227	0.053276	3	0.844583	0.045882	3	0.78659	0.009901	3
900	0.838947	0.055164	3	0.828261	0.037764	3	0.775951	0.029465	3
910	0.833839	0.040875	3	0.824427	0.044241	3	0.786218	0.025418	3
920	0.841965	0.043857	3	0.831476	0.05839	3	0.792628	0.017983	3
930	0.848654	0.056661	3	0.841001	0.045014	3	0.806894	0.012068	3
940	0.851123	0.064403	3	0.829143	0.057962	3	0.790717	0.012342	3
950	0.856127	0.059306	3	0.835491	0.046216	3	0.814523	0.022747	3
960	0.855458	0.055894	3	0.836658	0.044652	3	0.813464	0.002953	3
970	0.858759	0.043555	3	0.838659	0.048534	3	0.806437	0.007446	3
980	0.846229	0.039644	3	0.836323	0.052457	3	0.81123	0.010721	3
990	0.861021	0.051743	3	0.844542	0.047788	3	0.818862	0.017697	3
1000	0.860124	0.037773	3	0.838449	0.032969	3	0.802104	0.010954	3
1010	0.871405	0.047664	3	0.867588	0.030417	3	0.812119	0.010048	3
1020	0.879705	0.045614	3	0.857412	0.040826	3	0.807209	0.008592	3
1030	0.863406	0.032734	3	0.845084	0.027521	3	0.825026	0.015542	3
1040	0.876683	0.045794	3	0.843043	0.034408	3	0.825936	0.014285	3

1050	0.862598	0.048212	3	0.86368	0.025266	3	0.835548	0.010757	3
1060	0.860973	0.038292	3	0.849547	0.043789	3	0.819451	0.011808	3
1070	0.860911	0.051288	3	0.848751	0.041862	3	0.845981	0.009484	3
1080	0.863709	0.054628	3	0.859445	0.0489	3	0.821491	0.010403	3
1090	0.881335	0.042886	3	0.8713	0.033364	3	0.824353	0.007542	3
1100	0.859804	0.043836	3	0.855567	0.03273	3	0.845307	0.025729	3
1110	0.891695	0.037454	3	0.880997	0.052833	3	0.821864	0.024502	3
1120	0.890083	0.024555	3	0.872138	0.023785	3	0.842958	0.016154	3
1130	0.88306	0.056312	3	0.879636	0.029181	3	0.835432	0.005226	3
1140	0.882613	0.057051	3	0.870926	0.040086	3	0.86883	0.013364	3
1150	0.87425	0.05003	3	0.86957	0.038553	3	0.854641	0.007618	3
1160	0.87671	0.049984	3	0.87283	0.031383	3	0.847458	0.033206	3
1170	0.892104	0.052538	3	0.868942	0.038176	3	0.861389	0.022785	3
1180	0.87411	0.060786	3	0.880557	0.041124	3	0.858163	0.021748	3
1190	0.887225	0.051185	3	0.883772	0.00694	3	0.850603	0.010208	3
1200	0.879748	0.026432	3	0.88162	0.02692	3	0.859979	0.00223	3
1210	0.887951	0.036666	3	0.884597	0.02469	3	0.845525	0.009499	3
1220	0.881522	0.028912	3	0.874719	0.020365	3	0.830608	0.014017	3
1230	0.885875	0.045187	3	0.881455	0.037266	3	0.849051	0.016682	3
1240	0.888335	0.043274	3	0.883107	0.028542	3	0.874278	0.013654	3
1250	0.899468	0.018936	3	0.889966	0.014284	3	0.860377	0.00852	3
1260	0.902428	0.055806	3	0.884093	0.019233	3	0.868625	0.010584	3
1270	0.903998	0.037564	3	0.884141	0.010158	3	0.869546	0.009208	3
1280	0.898531	0.045361	3	0.883313	0.039547	3	0.885399	0.013973	3
1290	0.91715	0.035162	3	0.897579	0.037825	3	0.892564	0.014781	3
1300	0.912233	0.030149	3	0.905962	0.02742	3	0.865452	0.013866	3
1310	0.900096	0.042752	3	0.88913	0.042974	3	0.872594	0.009218	3
1320	0.901899	0.040921	3	0.897186	0.037983	3	0.86845	0.013984	3
1330	0.904747	0.030157	3	0.889655	0.026229	3	0.855536	0.009329	3
1340	0.905374	0.046754	3	0.892457	0.047966	3	0.889587	0.018802	3
1350	0.905217	0.042014	3	0.924449	0.051979	3	0.900231	0.010541	3
1360	0.903975	0.03494	3	0.898476	0.033105	3	0.876535	0.008822	3
1370	0.921486	0.039052	3	0.909217	0.030589	3	0.897478	0.012951	3
1380	0.917068	0.033473	3	0.908642	0.030608	3	0.881089	0.010173	3
1390	0.922752	0.029892	3	0.923588	0.011692	3	0.869858	0.002327	3
1400	0.917129	0.042799	3	0.905918	0.051671	3	0.882578	0.015365	3
1410	0.91162	0.031577	3	0.925903	0.018351	3	0.894472	0.018157	3
1420	0.913361	0.044968	3	0.913765	0.040699	3	0.908207	0.030894	3
1430	0.903894	0.04421	3	0.913699	0.036355	3	0.899196	0.01454	3
1440	0.92812	0.037626	3	0.934564	0.032934	3	0.909512	0.016236	3
1450	0.9104	0.035023	3	0.937841	0.046852	3	0.887865	0.011087	3
1460	0.91388	0.03966	3	0.922646	0.045705	3	0.916184	0.013709	3
1470	0.917405	0.038273	3	0.908025	0.046071	3	0.883533	0.004676	3

1480	0.926008	0.032702	3	0.926801	0.050733	3	0.895135	0.022068	3
1490	0.929518	0.052096	3	0.906598	0.049841	3	0.907352	0.015465	3
1500	0.927991	0.037908	3	0.91788	0.02429	3	0.900141	0.015125	3
1510	0.91493	0.03213	3	0.913883	0.041736	3	0.890612	0.009286	3
1520	0.927749	0.057806	3	0.929176	0.037133	3	0.949479	0.0082	3
1530	0.913	0.047798	3	0.913628	0.036506	3	0.928299	0.004685	3
1540	0.935032	0.034354	3	0.926339	0.038336	3	0.891125	0.013185	3
1550	0.928485	0.045773	3	0.927216	0.032867	3	0.903895	0.006207	3
1560	0.919521	0.049643	3	0.930735	0.039302	3	0.904905	0.017096	3
1570	0.931742	0.034419	3	0.931093	0.033827	3	0.92241	0.003846	3
1580	0.934999	0.039336	3	0.943812	0.043961	3	0.924075	0.016872	3
1590	0.935931	0.046387	3	0.930803	0.03432	3	0.920086	0.008909	3
1600	0.9401	0.030024	3	0.922722	0.029411	3	0.923504	0.024047	3
1610	0.93869	0.047519	3	0.943155	0.038442	3	0.91726	0.036838	3
1620	0.945422	0.034078	3	0.950145	0.02908	3	0.894994	0.028412	3
1630	0.931694	0.01677	3	0.940597	0.053523	3	0.921357	0.02071	3
1640	0.940147	0.025434	3	0.941288	0.039321	3	0.922655	0.015768	3
1650	0.946097	0.040666	3	0.937356	0.025513	3	0.907572	0.010849	3
1660	0.935475	0.031454	3	0.938526	0.042052	3	0.933259	0.02133	3
1670	0.963623	0.025302	3	0.950404	0.036139	3	0.913833	0.013341	3
1680	0.947625	0.042977	3	0.960243	0.033238	3	0.899055	0.022951	3
1690	0.921471	0.025743	3	0.936323	0.025505	3	0.913698	0.015741	3
1700	0.94573	0.03974	3	0.938312	0.028674	3	0.921568	0.005354	3
1710	0.948126	0.033203	3	0.941317	0.036537	3	0.926841	0.014384	3
1720	0.949213	0.029979	3	0.949526	0.044255	3	0.912354	0.011957	3
1730	0.948841	0.039448	3	0.955409	0.05122	3	0.94784	0.025639	3
1740	0.961175	0.019217	3	0.968252	0.047841	3	0.906015	0.019192	3
1750	0.963186	0.045429	3	0.95592	0.048029	3	0.922572	0.011168	3
1760	0.956187	0.026524	3	0.965879	0.013982	3	0.920943	0.020585	3
1770	0.963125	0.024549	3	0.952226	0.053125	3	0.924948	0.010928	3
1780	0.944291	0.035203	3	0.940995	0.040823	3	0.911972	0.006214	3
1790	0.95027	0.025093	3	0.954923	0.056191	3	0.91409	0.014675	3
1800	0.952615	0.036223	3	0.957022	0.030378	3	0.935238	0.009781	3
1810	0.95206	0.026894	3	0.959288	0.044304	3	0.946181	0.010277	3
1820	0.955887	0.027048	3	0.955439	0.020433	3	0.928095	0.018649	3
1830	0.961337	0.030839	3	0.950123	0.0436	3	0.932606	0.019155	3
1840	0.956388	0.035006	3	0.943586	0.0386	3	0.900708	0.017476	3
1850	0.948485	0.040128	3	0.958944	0.039501	3	0.9268	0.013184	3
1860	0.944026	0.032726	3	0.953577	0.024781	3	0.934471	0.007455	3
1870	0.969306	0.048112	3	0.975616	0.042234	3	0.94489	0.024365	3

Table S12. Averaged data from fluorogenic RNase activity assay was used for the determination of k_{cat}/K_M values.

6.1.7 Chemical denaturation assay

The chemical denaturation profile of barnase was obtained by fluorescence spectroscopy. In this assay, a solution of each barnase variant (recombinant WT, synthetic WT, synthetic R110F_{Br}) was prepared at 1 μM in either buffer A (50 mM MES, pH 6.3) or buffer B (8 M urea, 50 mM MES, pH 6.3). In a 384-well plate, using the two stock solutions, mixtures of each barnase variant was prepared at 25 urea concentrations ranging 0 – 8 M (at 0.33 M steps). The final concentration of barnase in each case was 1 μM while the buffer was 50 mM MES at pH 6.3. The mixtures were prepared in triplicates. The solutions were equilibrated at room temperature for two hours. Then the emission at 315 nm upon excitation at 290 nm was recorded (Fig. S17, Table S13).

The transition midpoint ($[D]_{50\%}$ - the concentration of urea at which half of the sample is unfolded) and m -values (the slope of the unfolding transition) were determined by fitting of the normalized fluorescence readings as described (58), using Prism 8 software from Graphpad. The values for these parameters were:

	m in [kcal mol ⁻¹ M ⁻¹]	$[D]_{50\%}$ in [M]
Literature	2.06	4.57
recombinant wild-type barnase	1.88 ± 0.21	4.63 ± 0.04
synthetic wild-type barnase	1.82 ± 0.25	4.68 ± 0.06
synthetic R110F(Br) barnase	1.62 ± 0.21	4.46 ± 0.06

Table S13. $[D]_{50\%}$ and m values for recombinant wild-type, synthetic wild-type and synthetic R110F(Br) barnase determined from a chemical denaturation assay shows comparable melting behavior for recombinant and synthetic samples.

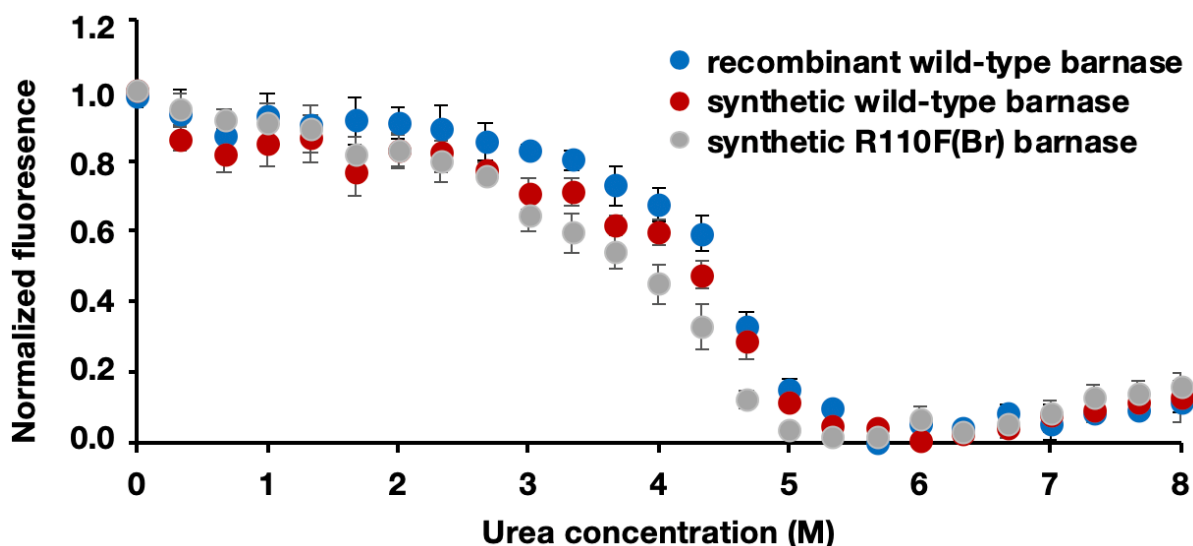


Fig. S17. Urea denaturation of the barnase variants demonstrates the structural similarity of the synthetic and recombinant samples. The fitted data was used to calculate the m -values as well as $[D]_{50\%}$ of recombinant WT barnase, synthetic WT barnase, and synthetic R110F_{Br} of barnase.

	recombinant wild-type barnase	synthetic wild-type barnase	synthetic R110F(Br) barnase
--	-------------------------------	-----------------------------	-----------------------------

Urea conc. [M]	Normalized Fluorescence			Normalized Fluorescence			Normalized Fluorescence		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
8.00	0.112	0.023	3	0.127	0.071	3	0.155	0.023	3
7.67	0.093	0.013	3	0.113	0.038	3	0.142	0.033	3
7.33	0.085	0.019	3	0.090	0.035	3	0.129	0.035	3
7.00	0.056	0.050	3	0.076	0.046	3	0.082	0.013	3
6.67	0.082	0.029	3	0.043	0.028	3	0.055	0.023	3
6.33	0.040	0.014	3	0.026	0.023	3	0.027	0.025	3
6.00	0.052	0.024	3	0.003	0.006	3	0.069	0.036	3
5.67	0.000	0.000	3	0.040	0.009	3	0.016	0.020	3
5.33	0.097	0.017	3	0.049	0.020	3	0.018	0.031	3
5.00	0.150	0.028	3	0.114	0.021	3	0.034	0.020	3
4.67	0.330	0.042	3	0.284	0.049	3	0.121	0.025	3
4.33	0.593	0.051	3	0.476	0.040	3	0.330	0.063	3
4.00	0.676	0.048	3	0.597	0.036	3	0.448	0.055	3
3.67	0.730	0.056	3	0.617	0.028	3	0.542	0.050	3
3.33	0.801	0.027	3	0.711	0.037	3	0.595	0.056	3
3.00	0.828	0.018	3	0.705	0.047	3	0.645	0.045	3
2.67	0.853	0.055	3	0.774	0.023	3	0.754	0.016	3
2.33	0.889	0.067	3	0.821	0.057	3	0.800	0.059	3
2.00	0.907	0.047	3	0.831	0.047	3	0.827	0.046	3
1.67	0.911	0.068	3	0.767	0.064	3	0.817	0.051	3
1.33	0.900	0.055	3	0.863	0.066	3	0.888	0.067	3
1.00	0.926	0.066	3	0.844	0.057	3	0.907	0.057	3
0.67	0.869	0.047	3	0.813	0.045	3	0.916	0.028	3
0.33	0.934	0.069	3	0.860	0.034	3	0.945	0.047	3
0.00	0.982	0.031	3	1.000	0.000	3	1.000	0.000	3

Table S14. Normalized fluorescence and standard deviation from chemical denaturation assay with urea.

6.2 HIV-1 protease

6.2.1 Folding of synthetic HIV-1 protease (Kent sequence)

The procedure for folding HIV Protease was taken from Johnson *et al.*(29) Lyophilized HIV-1 protease (2.5 mg) was dissolved in 2 mL of denaturing buffer (6 M guanidine hydrochloride, 200mM NaPi, pH 7.4). A dialysis cassette (Slide-A-Lyzer™ 3.5MWCO, Sigma Aldrich), was pretreated for 2 min. with dialysis buffer A (50 mM NaOAc pH 5.6), before the denaturated HIV-1 protease sample was injected. The dialysis cassette was placed in a vessel containing 350 mL dialysis buffer A and left to stir at room temperature for 3 h. The cassette was then transferred to dialysis buffer B (350 ml of 10 mM NaOAc pH 5.6 prechilled to 4 °C) and left at 4 °C for 14 h. The protein containing buffer was removed from the cassette, filtered by a Pall™ 0.22 µm PTFE syringe filter, and stored in 100 µL aliquots at -80 °C. For all further assays, frozen fractions were used within 1 hour after defrosting and excess from the aliquot was discarded after use.

Concentration of the resulting sample was determined using a Pierce™ Rapid Gold BCA Protein Assay Kit (supplier) of the undiluted stock HIV-1 Protease in triplicate compared to the BSA standard. Stock was found to contain 0.41 mg/ml protein which corresponds to a refolding yield of 50%. The concentration of active enzyme within the stock was determined by active site titration in the method of Windsor *et al.*(59) using darunavir and was found to be 76%.

6.2.2 Fluorogenic protease activity assay

Assay conditions were taken from Johnson *et al* with slight adaptations taken from the original report for the assay from Toth *et al.*(29, 42) The fluorogenic substrate Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH₂ (Abz = 2, aminobenzoic acid, Nle= norleucine, p-nitro-Phe = p-nitro phenylalanine) was purchased from Bachem (product no. 4030748.0005) and was used as a 4 mM stock in DMSO (concentration determined gravimetrically). Fluorescence time-courses were measured at 37 °C on a Tecan™ M1000 Pro plate reader (ex: 355nm, em: 430nm, bandwidths 5m) with a gain setting of 124 in a Costar™ black chimney-well 96-well plate. Assays were conducted in a final volume of 200 µL of 50 mM NaOAc pH 5.6 maximum 4.5% DMSO, substrate (5-200 µM), and protease (10 nM active dimer) with 3 replicates. Calibration of fluorophore was performed as described in Toth *et al*(42), and quantitation of the initial rates were made within the linear range of the detector (40 µM cleaved substrate at a gain setting of 124) (3 replicates). Initial velocity data was used with less than 20% of total substrate cleavage (**Fig. S18**). Kinetic constants were derived from fitting the initial velocities to the Michaelis-Menten equation using Graphpad™ Prism's nonlinear least-squares regression (**Fig. S19**).

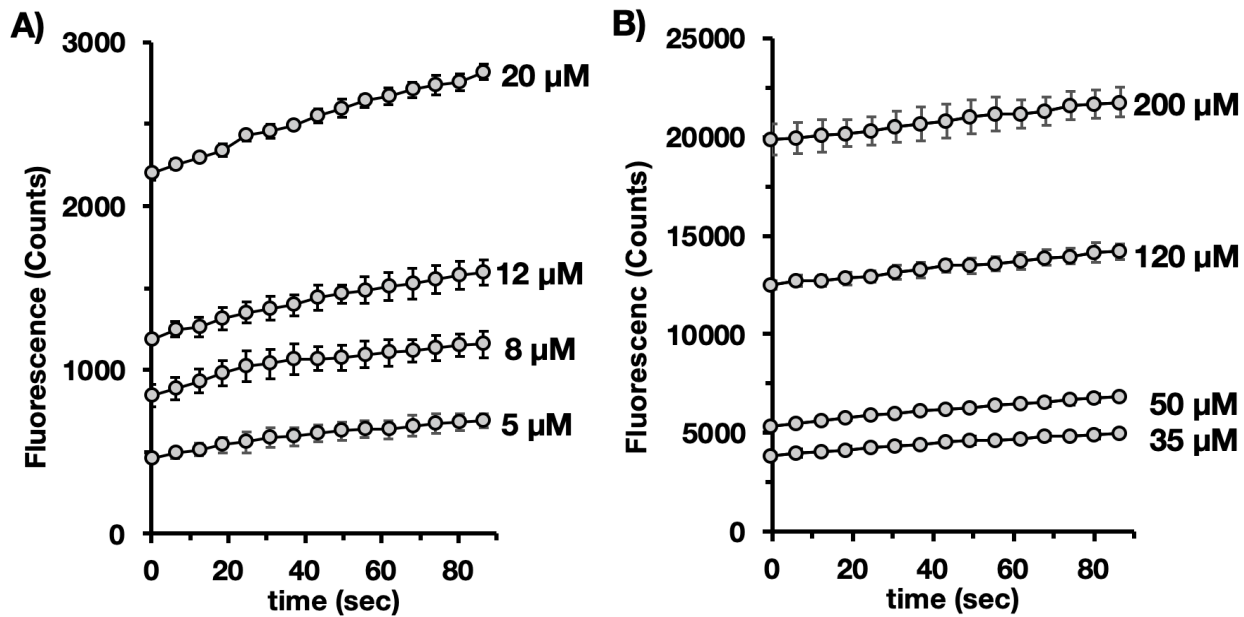


Fig. S18. Raw time course measurement data for fluorogenic protease activity assay. Traces shown represent increases in fluorescence of each sample due to cleavage of the peptide substrate by the protease. All data represent averages over 3 replicates, with error bars plotted as standard deviations.

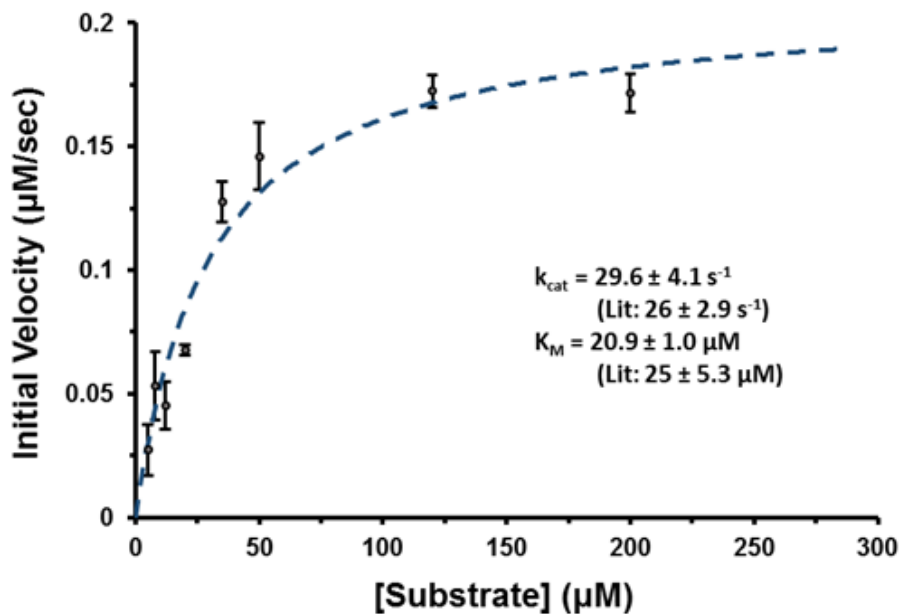


Fig. S19. k_{cat} and K_M values of HIV-1 determined by fluorogenic assay.

6.2.3 Substrate specificity assay

The qualitative assay with model substrate p12nt was previously reported by Schneider *et al.*(28) The 20mer substrate was synthesized on the Amidator using standard optimized peptide synthesis conditions described in section 2. 21 mg of the peptide was purified on a C3 Zorbax semi-prep column (9.4 x 250mm) with a gradient from 15% to 75% ACN over 62 minutes with a flow-rate of 4 ml/min. The cleanest fraction affording 1 mg was used for all further assays, and analytical LCMS data is shown in **Fig. S20**. p12NT (10 μ g) was incubated with HIV-1 Protease (300 nM) in a total volume of 30 μ L of 50 mM NaOAc pH 5.6 with 0.5 mg/ml BSA. Reactions were left at 37 $^{\circ}$ C for 14 h and quenched with snap-freezing in liquid N₂ followed by storage at -80° C. The crude reaction mixture was analyzed by LCMS. Fragments of the substrate peptide corresponding to cleavage at the conserved site were observed with loss of full-length peptide. No peptides corresponding to non-specific cleavage of the substrate peptide nor the native BSA in solution were observed even after prolonged incubation of 16 hours (**Fig. S21**).

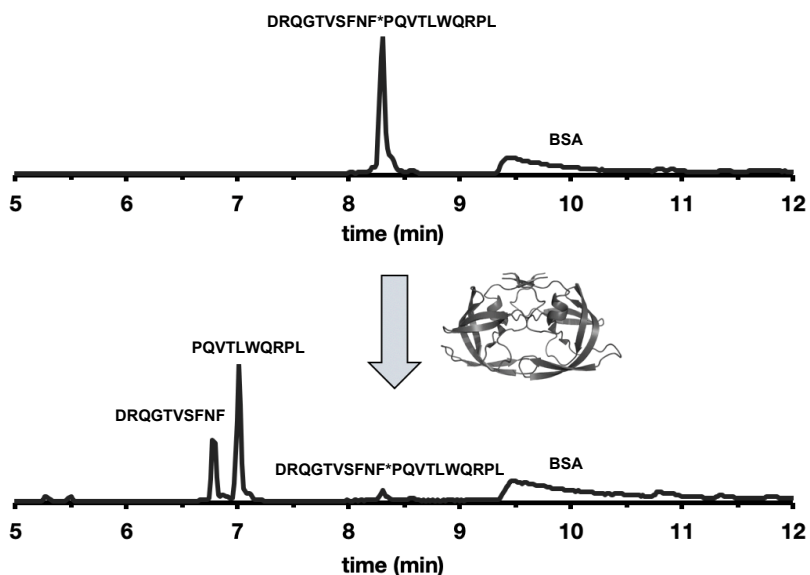


Fig. S20. Synthetic HIV-1 protease shows substrate specific activity. Incubation of the synthetic protease with a peptide containing a conserved HIV-Pr consensus cleavage site leads to specific hydrolysis. No observation of off-target hydrolysis at separate sites, nor of native BSA in solution was observed. Traces shown represent separation of the reaction mixture onto LCMS using standard method 1-61% B over 15 min, Zorbax C3 column (6550).

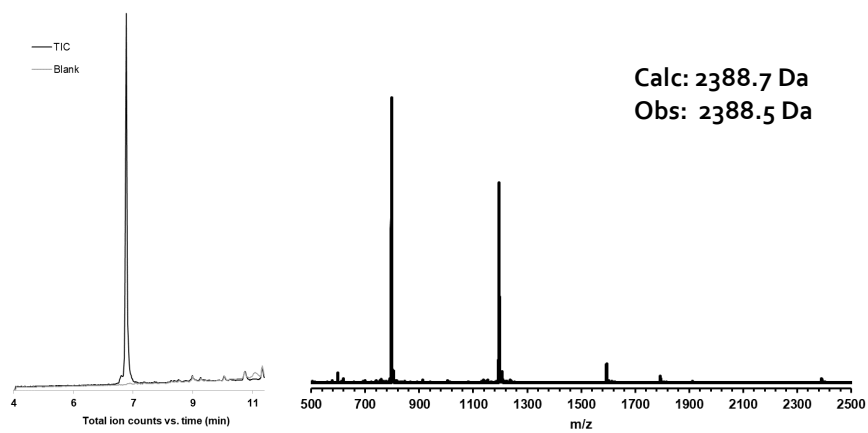


Fig. S21. Characterization of purified qualitative substrate P12NT. LCMS trace recorded with standard method 1-91% B over 9 min, Zorbax C3 column (6520). Left figure shows the TIC trace, and the m/z extraction is shown for the major peak on the right. Calculated mass and observed mass are in good agreement.

6.3 Sortase A^[59-206];P94S/D160N/K196T

6.3.1 Expression and purification of recombinant sortase A^[59-206];P94S/D160N/K196T
Recombinant sortase A^[59-206] was recombinantly expressed following a protocol described in the literature.⁽⁴⁵⁾

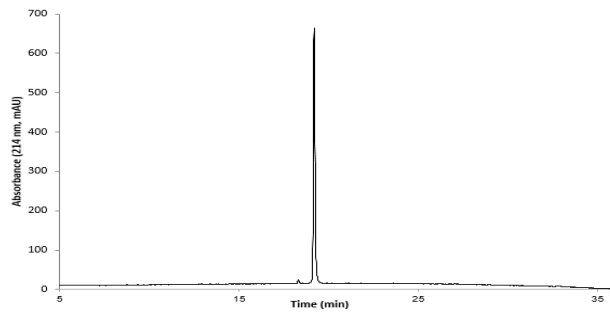
A) Recombinant Sortase A^[59-206];P94S/D160N/K196T

Sequence:

ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS EQLNRGVSFA
EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSVMVYF KVGNETRKYK MTSIRNVKPT
DVEVLDEQKG KDKQLTLITC DDYNEKTGVW ETRKIFVATE VKLEHHHHHH (170 AA)

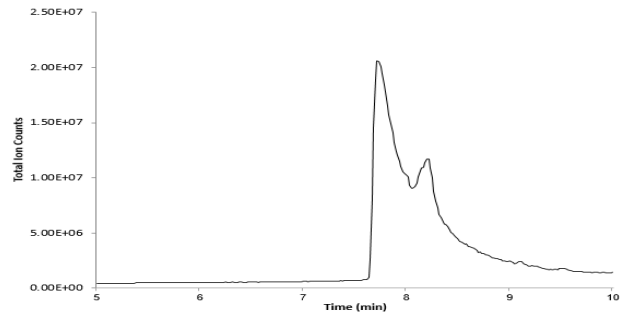
B) Analytical data for recombinant Sortase A^[59-206];P94S/D160N/K196T

Analytical HPLC:



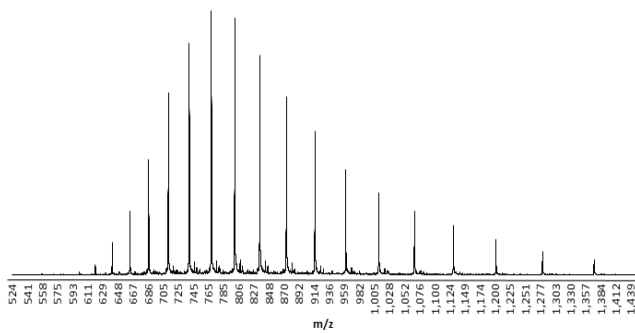
method: 5–65% B over 30 min

LC-MS:



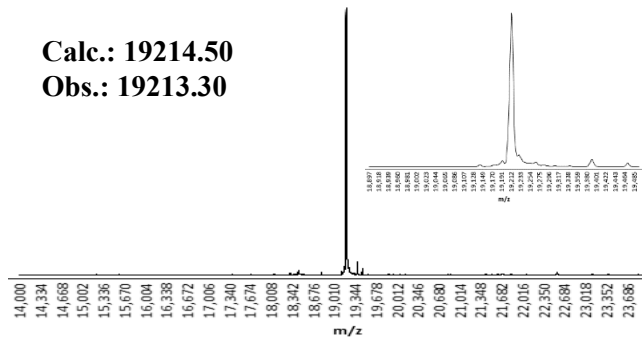
method: 1-61%B over 15 min, Zorbax C3 column (6520)

MS Data from LC-MS:



Deconvolution of LC-MS data:

Calc.: 19214.50
Obs.: 19213.30

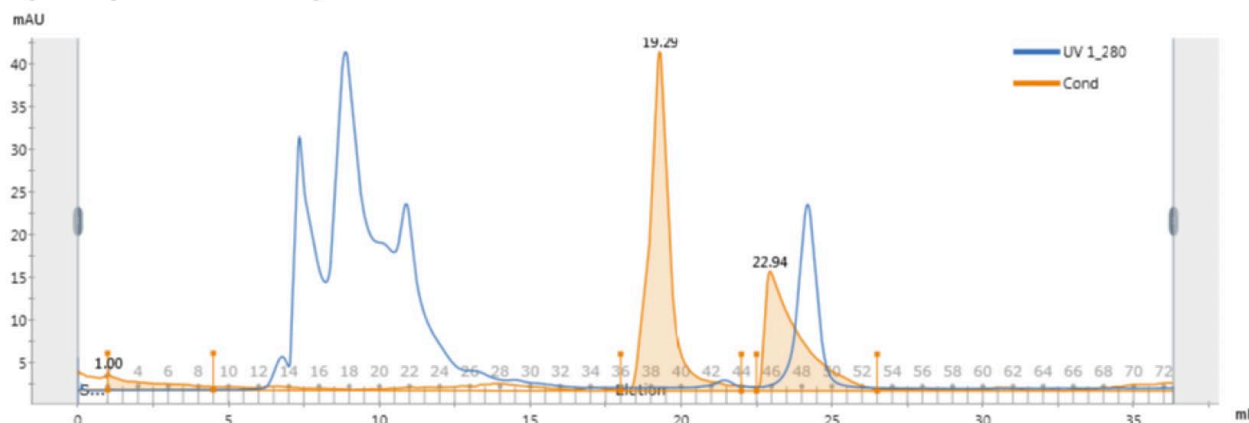


6.3.2 Folding and SEC purification of synthetic sortase A[59-206];P94S/D160N/K196T

Synthetic sortase A[59-206];P94S/D160N/K196T (1.8 mg) was dissolved in 150 μL of 6 M Gn·HCl denaturing solution in 50 mM Tris, 20 mM DTT, pH 7.5. Concentration was determined by A_{280} reading on plate reader and adjusted by the addition of Gn·HCl solution to a concentration of 8 mg/ml sortase A. (extinction coefficient of sortase A is $14440 \text{ M}^{-1} \text{ cm}^{-1}$).

The mixture was then serially diluted twenty fold in nine steps utilizing 50 mM Tris, 150 mM NaCl, 20 mM DTT, pH 7.5 resulting in 0.3 M Gn·HCl and 0.4 mg/ml sortase A. The solution was left overnight (18 hrs) at room temperature on the lab bench. The solution was then concentrated and filtered and subjected to size exclusion chromatography (SuperdexTM 75 Increase 10/300 GL, 0.4 mL/min), using isocratic eluent 20 mM Tris, 150 mM NaCl, pH 7.5. The elution profile is depicted in Fig. S22A. Fractions were analyzed by LCMS and the cleanest fractions were pooled and concentrated utilizing a 3K molecular weight cut off spin filter and stored at -80°C . A total of 0.11 mg of sortase A was isolated corresponding to a 6% isolated yield. The purity of the final product was assessed by LCMS and HPLC (Fig. S22B).

A) SEC purification of synthetic sortase A



B) Analytical Data for purified synthetic sortase A post SEC

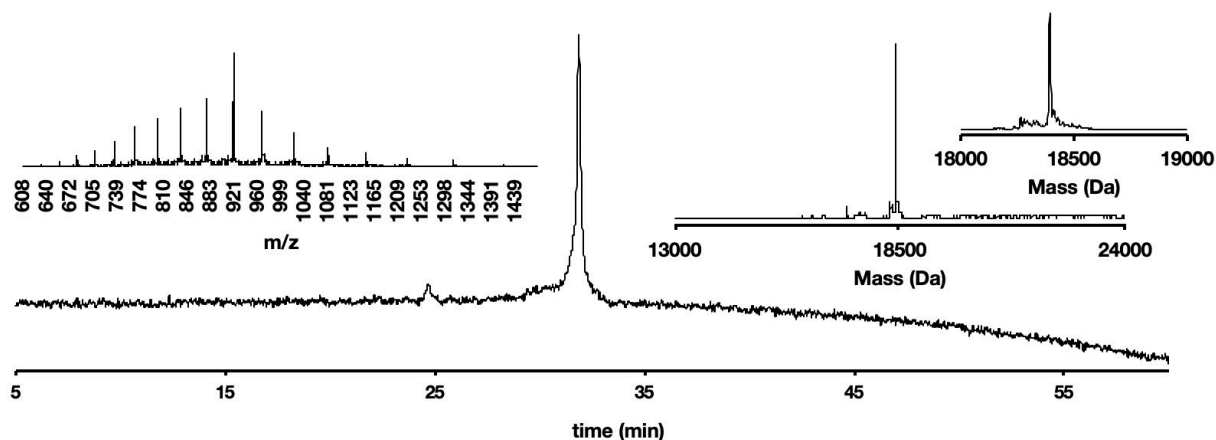


Fig. S22. Folding and SEC purification of synthetic sortase A yields clean protein sample. A) HPLC-purified sortase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) Analytical HPLC trace of purified synthetic sortase A post size exclusion chromatography (absorbance at 214 nm), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals).

6.3.3 Semiquantitative activity assay

Sortase A performs a transpeptidation reaction where the active enzyme cleaves the threonine-glycine bond in the LPXTG motif and ligates to a polyglycine. Therefore, an active sample of sortase A in the presence of peptides AQALPETGEE and GGGGGLY should generate the ligation product AQALPETGGGGGLY. Samples of folded recombinant sortase A, folded synthetic sortase A, and folded synthetic sortase A post-SEC were diluted to a concentration of 0.02 mg/ml in 50 mM Tris, 150 mM NaCl, 20 mM DTT, pH 7.5 as determined by A_{280} reading on a plate reader. The sortase A samples (15 μ l) were then added to a reaction mixture containing 6 μ l AQALPETGEE_(am) (1 mg/ml), 6 μ l GGGGGLY_(am) (1 mg/ml) and 3 μ l calcium chloride (50 mM). Mixtures were left at room temperature for 20 hours and quenched with 30 μ l of a 2% trifluoroacetic acid solution with 50:50 (v/v) water: acetonitrile. Samples were analyzed by LCMS (Fig. S23).

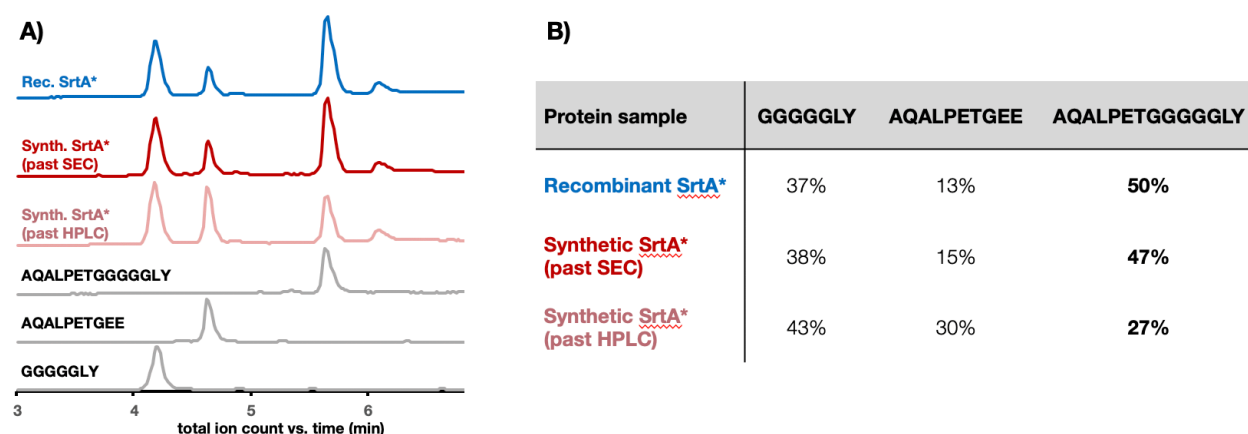


Fig. S23. Synthetic sortase A shows enzymatic activity in semiquantitative assay. A) LCMS traces (TIC) of the three assay reactions and the three peptide standards. Ligation product AQALPETGGGGGLY was formed for all three assay conditions indicating active sortase A in each reaction. B) Area under each peak in the TIC, manually integrated in MassHunter software. The SEC purified synthetic sortase A has similar values to the recombinant sample, with improved activity over synthetic folded sample which was not purified by SEC.

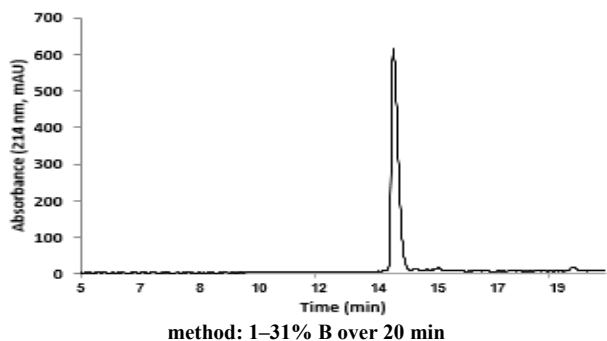
A) Synthesis data for substrate 1 (Sortase A semiquantitative assay)

Sequence: GGGGGLY (7 AA)
Resin: 200 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

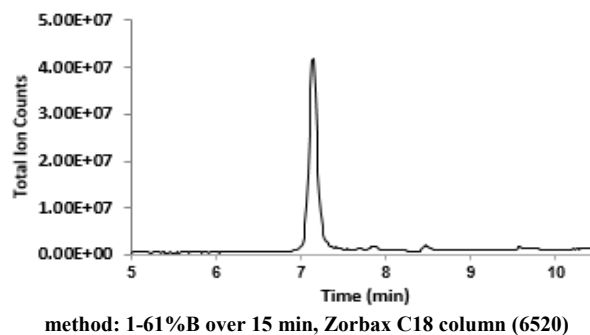
B) Cleavage and analytical data for crude substrate 1 (Sortase A semiquantitative assay)

Cleavage: Cleavage protocol A for peptides, no ether wash, evaporated under N₂

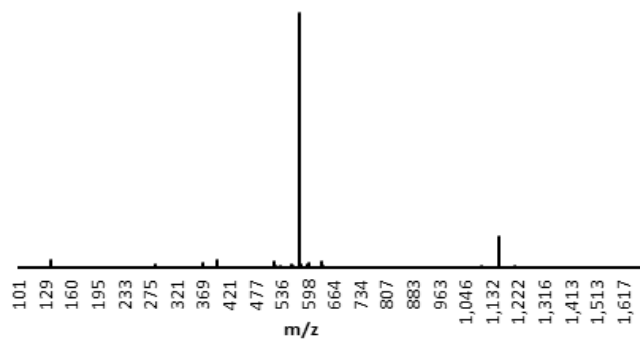
Analytical HPLC:



LC-MS:



MS Data from LC-MS:



A) Synthesis data for substrate 2 (Sortase A semiquantitative assay)

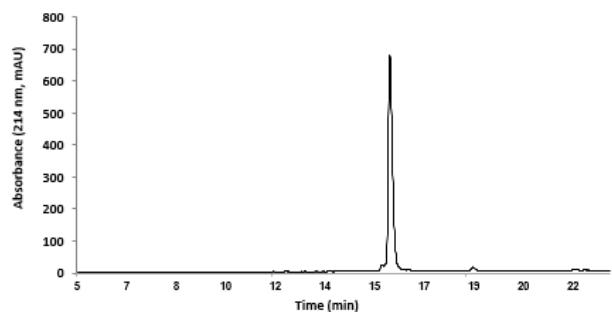
Sequence: AQALPETGEE (10 AA)

Resin: 150 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

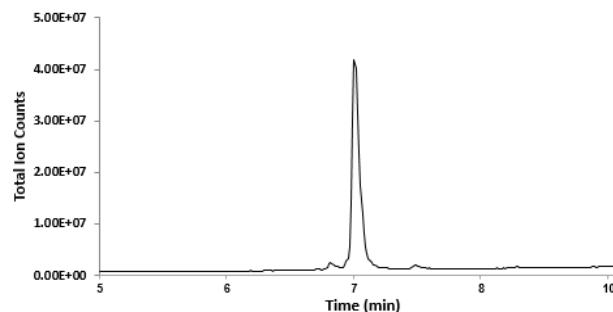
B) Cleavage and analytical data for crude substrate 2 (Sortase A semiquantitative assay)

Cleavage: Cleavage protocol A for peptides

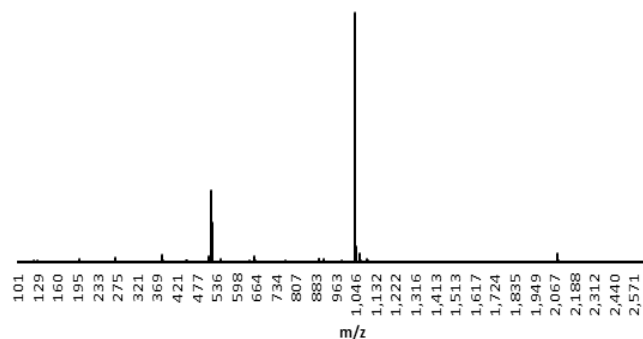
Analytical HPLC:



LC-MS:



MS Data from LC-MS:



A) Synthesis Data for ligation standard (SrtA semiquantitative assay)

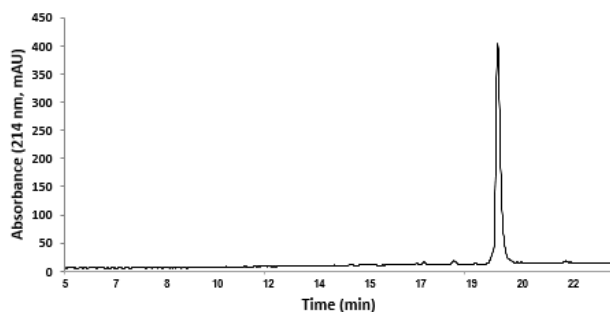
Sequence: AQALPETGGG GGLY (14 AA)

Resin: 150 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

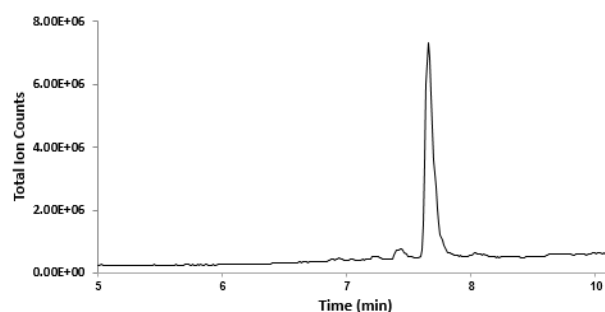
B) Cleavage and analytical data for crude ligation standard (SrtA semiquantitative assay)

Cleavage: Cleavage protocol A for peptides

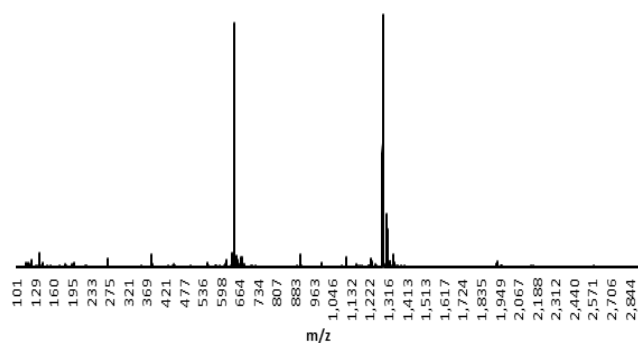
Analytical HPLC:



LC-MS:



MS Data from LC-MS:



6.4 MDM2^[1-118]

6.4.1 Folding of synthetic MDM2^[1-118]

Synthetic MDM2^[1-118] was refolded according to the procedure from Zhan *et al.*(60) Lyophilized MDM2^[1-118] (0.6 mg; 35 nmol) was dissolved in phosphate buffered saline (PBS) containing 6 M Guanidine hydrochloride (70 μ L) and 20 mM DTT at pH 7.11. MDM2^[1-118] concentration was determined by UV²⁸⁰ and adjusted to 150 μ M (ext. coeff. Of MDM2: 10430 m⁻¹ cm⁻¹). 50 μ L of the resulting solution was serial diluted into 250 μ L of folding buffer containing PBS and 20 mM DTT at pH 7.31, to give final conditions of 25 μ M MDM2 in PBS containing 1 M guanidine hydrochloride and 20 mM. DTT. The resulting solution was kept at room temperature for 1 h before performing biolayer interferometry.

6.4.2 Preparation of biotinylated p53(15-29)

15-29P53-like peptide (SQETFSDLWKLLPEN) was synthesized by Fmoc-based SPPS, with H-Rink Amide-ChemMatrix resin (130mg/synthesis; 0.06 mmol). After the Fmoc deprotection, peptide labeling with N-terminal D-biotin was performed on the resin bound protected peptides by treating the resin with a solution of Biotin-PEG₄-propionic acid (ChemPep Inc., 0.75 mmol), HATU (0.38M in DMF; 1.87 mL; 0.71 mmol) and DIEA (1.5 mmol) for 1.5 h at room temperature. Upon completion, the resin was washed with DMF (5x) and DCM (5x) and dried under reduced pressure. The peptide was cleaved and processed using the standard protocol for peptide cleavage described in **section 1.3** (method A).

6.4.3 Assay of MDM2 binding activity, by biolayer interferometry

Synthetic MDM2 in refolding buffer (see **section 5.4.1**) and commercial MDM2 (Abcam 167941) were brought to 1 mg/mL bovine serum albumin (BSA), 0.02% Tween 20 by addition of 10 mg/mL BSA, 0.2% Tween 20 in PBS. The resulting solutions were diluted serially into 1 mg/mL BSA, 0.02% Tween 20 in PBS for BLI assay.

Biolayer interferometry was performed using an Octet Red96 system (ForteBio; Menlo Park, CA) and black, polypropylene, chimney well, flat-bottom 96 well plates (Greiner Bio-One, Kremsmünster, Austria). Wells were filled with 200 μ L of the appropriate solution. Streptavidin biosensors (ForteBio) were equilibrated in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer for at least 10 min prior to use. Sample plates were equilibrated at 30 °C for 5 minutes before the start of an experiment, and kept at 30 °C throughout. Sample plates were agitated at either 1000 or 1500 rpm throughout the assay.

The assay protocol was as follows: 1) 60 sec 'baseline' in 1 mg/mL BSA, 0.02% Tween 20, PBS buffer; 2) 120 sec 'p53 immobilization' in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer containing ~400 nM biotin-p53(15-29); 3) 120 sec 'baseline' in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer; 4) 300 sec 'association' in MDM2 (**Fig. S24**, indicated concentrations). BLI assay was run in triplicates. Equilibrium response (nm) was plotted against MDM2 concentration to determine K_d of synthetic and recombinant MDM2 (**Fig. S25**).

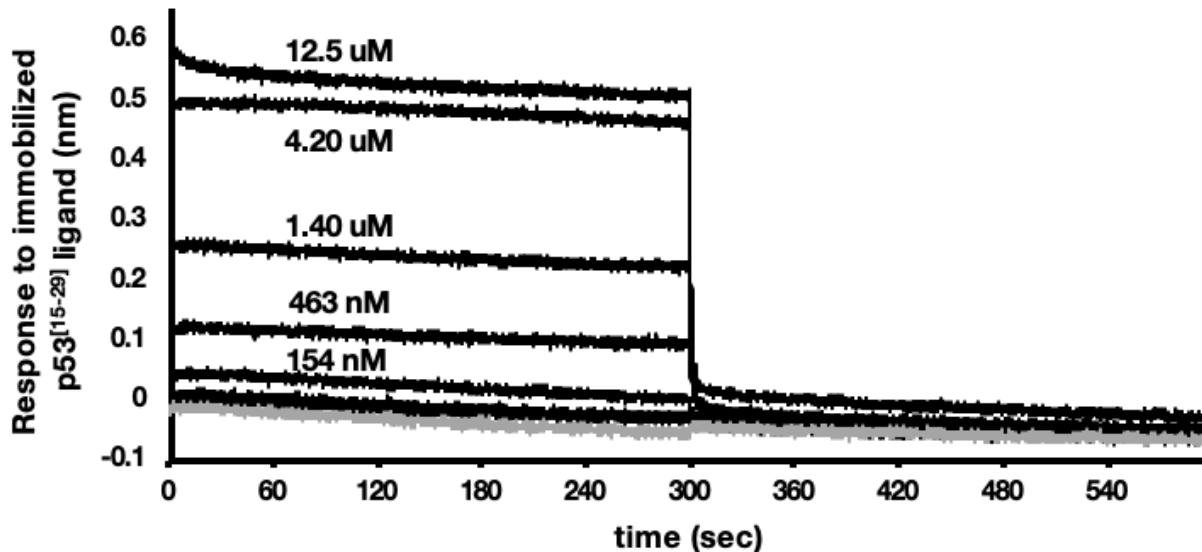


Fig. S24. Sensogram of MDM2^[1-118]. Experiment was conducted in triplicates, for clarity only a single data set is displayed here.

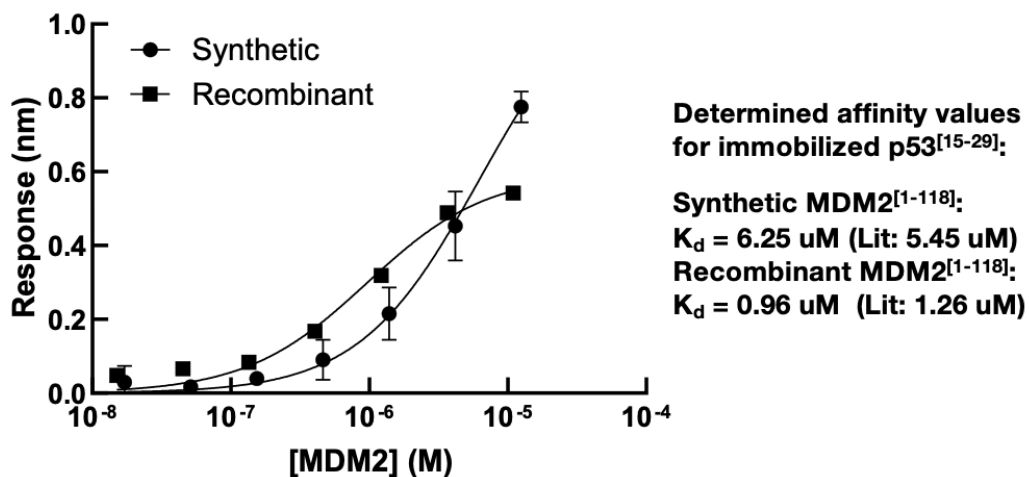


Fig. S25. Determination of K_d from concentration-dependent equilibrium binding responses. All experiments were carried out in triplicates and K_d values were determined from equilibrium binding responses of immobilized p53 to MDM2 at various concentrations.(58)

200210 Supplementary Material_ChemRxiv.pdf (9.58 MiB)

[view on ChemRxiv](#) • [download file](#)
