Synthesis of Proteins by Automated Flow Chemistry

Authors: N. Hartrampf[1], A. Saebi[1†], M. Poskus[1†], Z. P. Gates[1], A. J. Callahan[1], A. E. Cowfer[1], S. Hanna[1], S. Antilla[1], C. K. Schissel[1], A. J. Quartararo[1], X. Ye[1], A. J. Mijalis[1,2], M. D. Simon[1], A. Loas[1], S. Liu[1,3], C. Jessen[4], T. E. Nielsen[4] and B. L. Pentelute[1*]

Affiliations:

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

[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

† authors contributed equally.

***Abstract**:*

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.

One Sentence Summary:

A benchtop automated machine synthesizes protein chains in hours.

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.

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 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.

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 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 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.

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 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)*

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 $A_{59\text{-}206}$ (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.

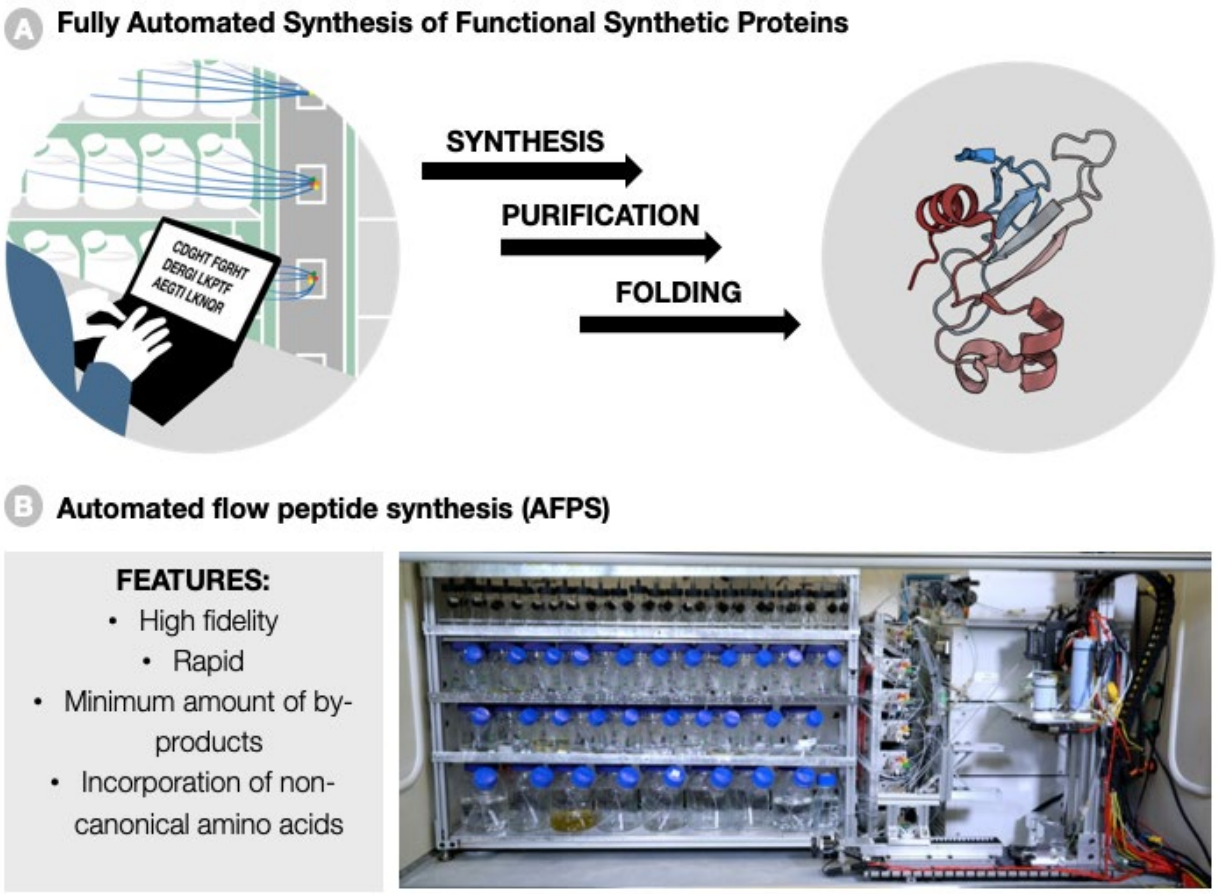

Fig. 1 Automated fast-flow solid-phase peptide synthesis enables high-fidelity production of long amino acid sequences. A) Fully automated chemical flow synthesis yields peptide chains, which —after purification and folding— give functional proteins; B) Automated fast-flow peptide synthesizer (AFPS)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

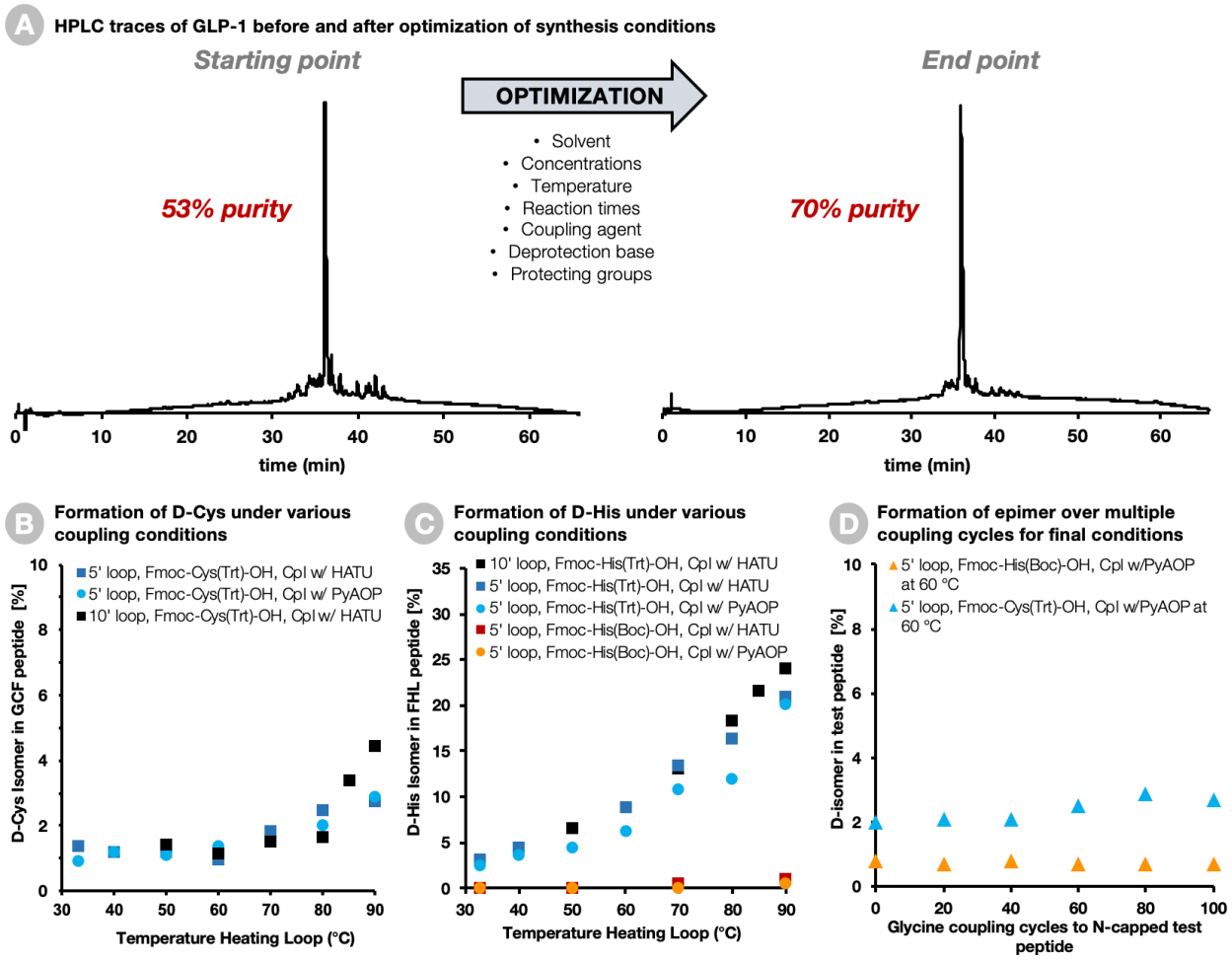

Fig. 2 Optimization of AFPS conditions improves synthesis outcome and reduces the amount of cysteine and histidine epimerization. A) Synthesis of GLP-1 using starting conditions and optimized conditions; B) Quantification of cysteine epimerization on activation temperature, heating time (5' loop and 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.

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

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.

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-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 methods, both with respect to time and performance.

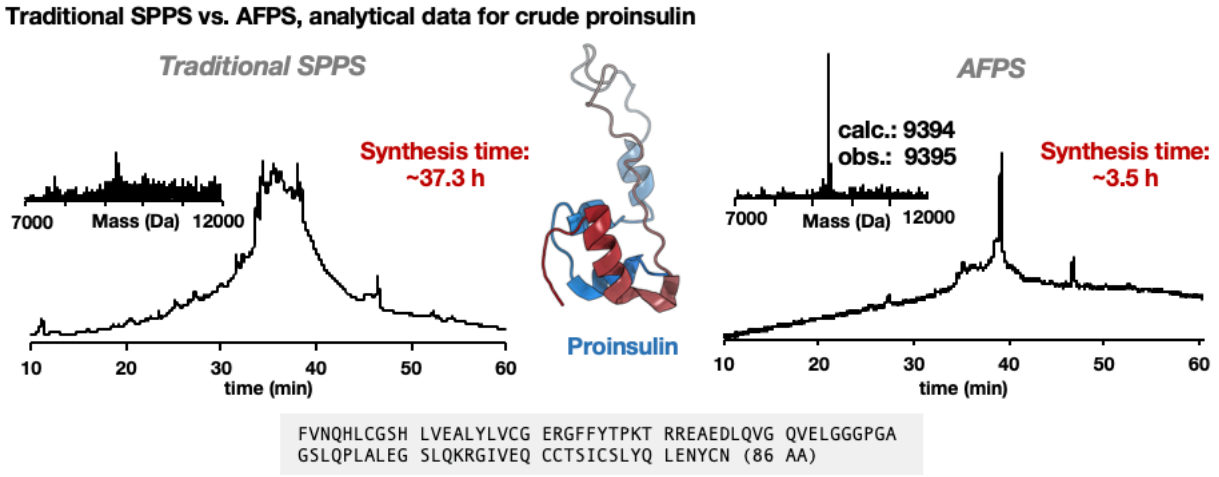

Fig. 3 Synthesis of proinsulin demonstrates the advantage of AFPS over traditional SPPS methods. Analytical HPLC data of the crude proinsulin are presented as the main chromatographic trace with absorbance detection at 214 nm (additional details can be found in the SI). Deconvoluted masses are displayed 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.

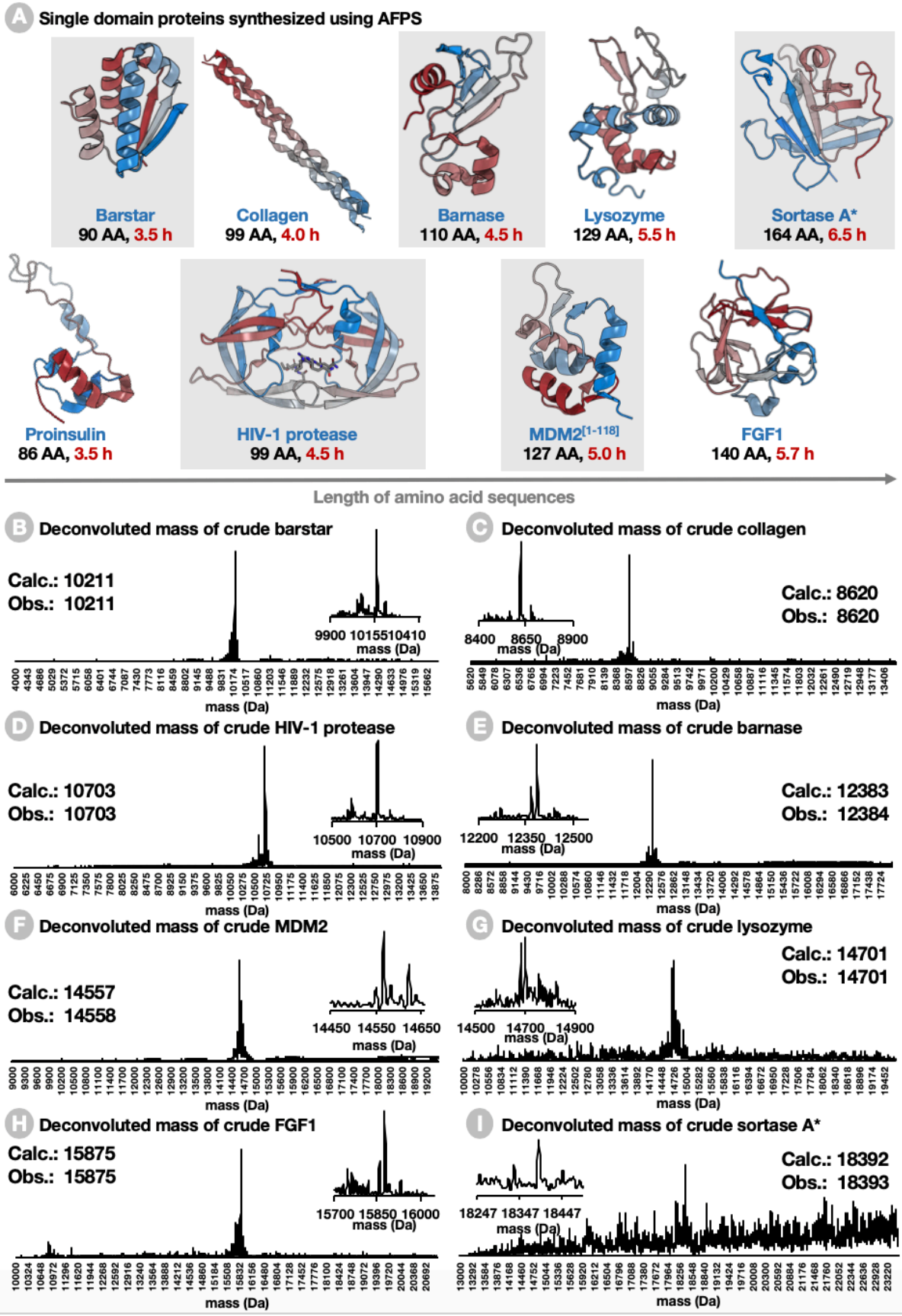
A Single domain proteins synthesized using AFPS
Barstar
90 AA, 3.5 h
Collagen
99 AA, 4.0 h
Barnase
110 AA, 4.5 h
Lysozyme
129 AA, 5.5 h
Sortase A*
164 AA, 6.5 h
Proinsulin
86 AA, 3.5 h
HIV-1 protease
99 AA, 4.5 h
MDM2[1-118]
127 AA, 5.0 h
FGF1
140 AA, 5.7 h
Length of amino acid sequences
B Deconvoluted 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
mass (Da)

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

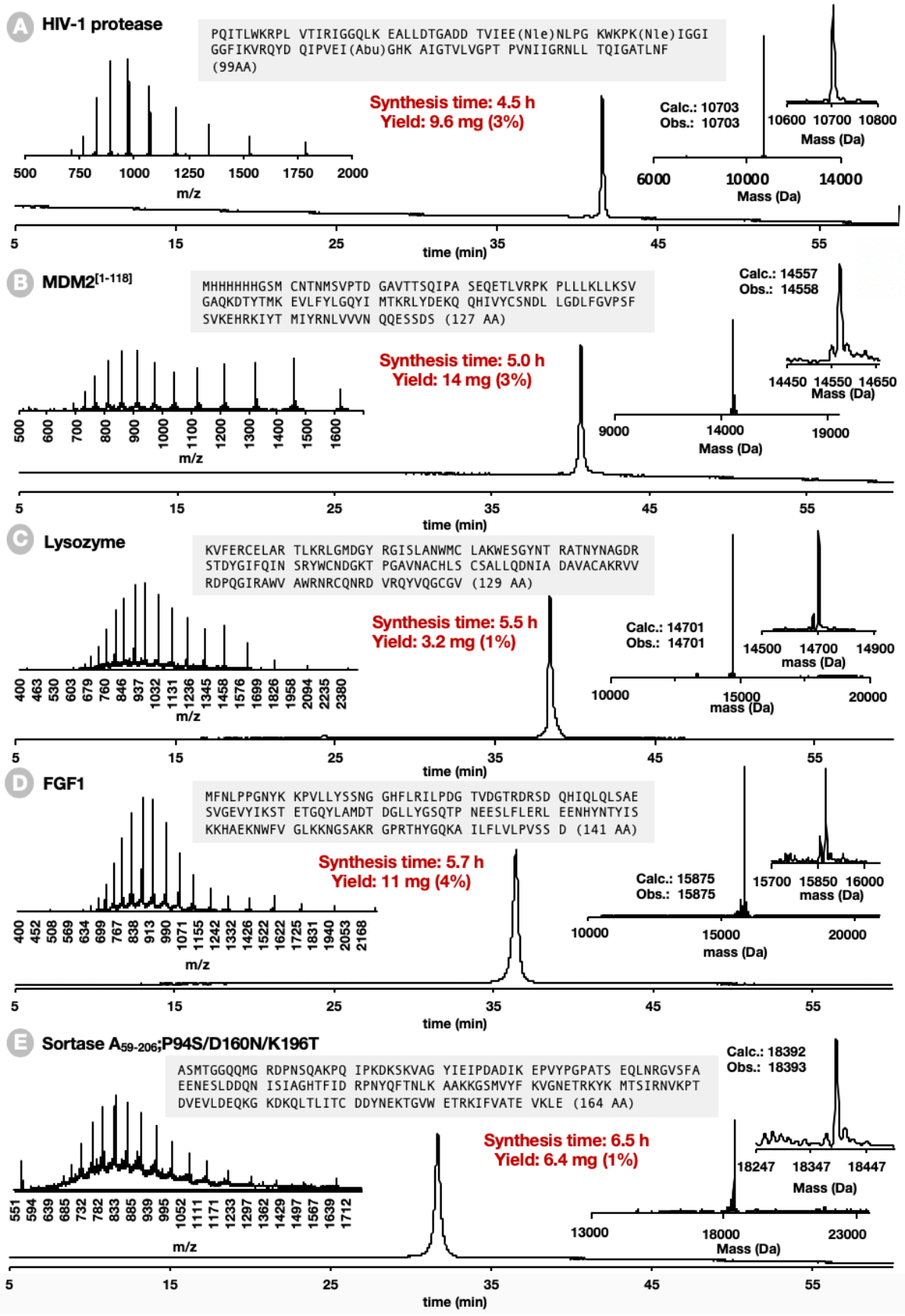
A HIV-1 protease
PQITLWKRPL VTIRIGGQLK EALLDTGADD TVIEE(Nle)NLPG KWKPK(Nle)IGGI
GGFIKVRQYD QIPVEI(Abu)GHK AIGTVLVGPT PVNIIGRNLL TQIGATLNF
(99AA)
Synthesis time: 4.5 h
Yield: 9.6 mg (3%)
Calc.: 10703
Obs.: 10703
B MDM2[1-118]
MHHHHHHGSM CNTNMSVPTD GAVTTSQIPA SEQETLVRPK PLLLKLLKSV
GAQKDTYTMK EVLFYLGQYI MTKRLYDEKQ QHIVYCSNDL LGDLFGVPSF
SVKEHRKIYT MIYRNLVVVN QQESSDS (127 AA)
Synthesis time: 5.0 h
Yield: 14 mg (3%)
Calc.: 14557
Obs.: 14558
C Lysozyme
KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR
STDYGIFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV
RDPQGIRAWV AWRNRCQNRD VRQYVQGCGV (129 AA)
Synthesis time: 5.5 h
Yield: 3.2 mg (1%)
Calc.: 14701
Obs.: 14701
D FGF1
MFNLPPGNYK KPVLLYSSNG GHFLRILPDG TVDGTRDRSD QHIQLQLSAE
SVGEVYIKST ETGQYLAMDT DGLLYGSQTP NEESLFLERL EENHYNTYIS
KKHAEKNWFV GLKKNGSAKR GPRTHYGQKA ILFLVLPVSS D (141 AA)
Synthesis time: 5.7 h
Yield: 11 mg (4%)
Calc.: 15875
Obs.: 15875
E Sortase A59-206;P94S/D160N/K196T
ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS EQLNRGVSFA
EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF KVGNETRKYK MTSIRNVKPT
DVEVLDEQKG KDKQLTLITC DDYNEKTGVW ETRKIFVATE VKLE (164 AA)
Synthesis time: 6.5 h
Yield: 6.4 mg (1%)
Calc.: 18392
Obs.: 18393
m/z
time (min)
Mass (Da)
mass (Da)

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$^{[1\text{-}118]}$ 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.

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. 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*)

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.

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 ($[D]_{50\%,\ synthetic}$ = 4.68 ± 0.06 M; $[D]_{50\%,\ recombinant}$ = 4.63 ± 0.04 M; $[D]_{50\%,\ literature}$ = 4.57 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 ($m_{synthetic}$ = 1.82 ± 0.25 kcal mol^{-1} M^{-1}, $m_{recombinant}$ = 1.88 ± 0.21 kcal mol^{-1} M^{-1} and $m_{literature}$ = 2.06 kcal mol^{-1} 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.

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 $k_{cat}/K_M = (7.6 \pm 0.2) \times 10^6\ M^{-1}\ s^{-1}$, which is comparable to that of recombinant barnase ($k_{cat}/K_M = (9.0 \pm 0.3) \times 10^6\ M^{-1}\ s^{-1}$) 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 $K_M = 20.9 \pm 1.0$ mM and a turnover number of $k_{cat} = 29.6 \pm 4.1\ s^{-1}$, 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 $A_{59\text{-}206}$ 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**). 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.

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 $MDM2^{[1\text{-}118]}$.(*31*) The binding of MDM2 to p53 is a key interaction in multiple pathways upregulated in cancer.(*47*, *48*) We folded milligram quantities of synthetic $MDM2^{[1\text{-}118]}$ and characterized its binding to immobilized $p53^{[14\text{-}29]}$ using biolayer interferometry (**Fig S24 and S25**). Synthetic $MDM2^{[1\text{-}118]}$ displayed an affinity toward p53 ($K_d = 6.25$ μM) comparable to the literature value ($K_d = 5.45$ μM) obtained under the same folding conditions.

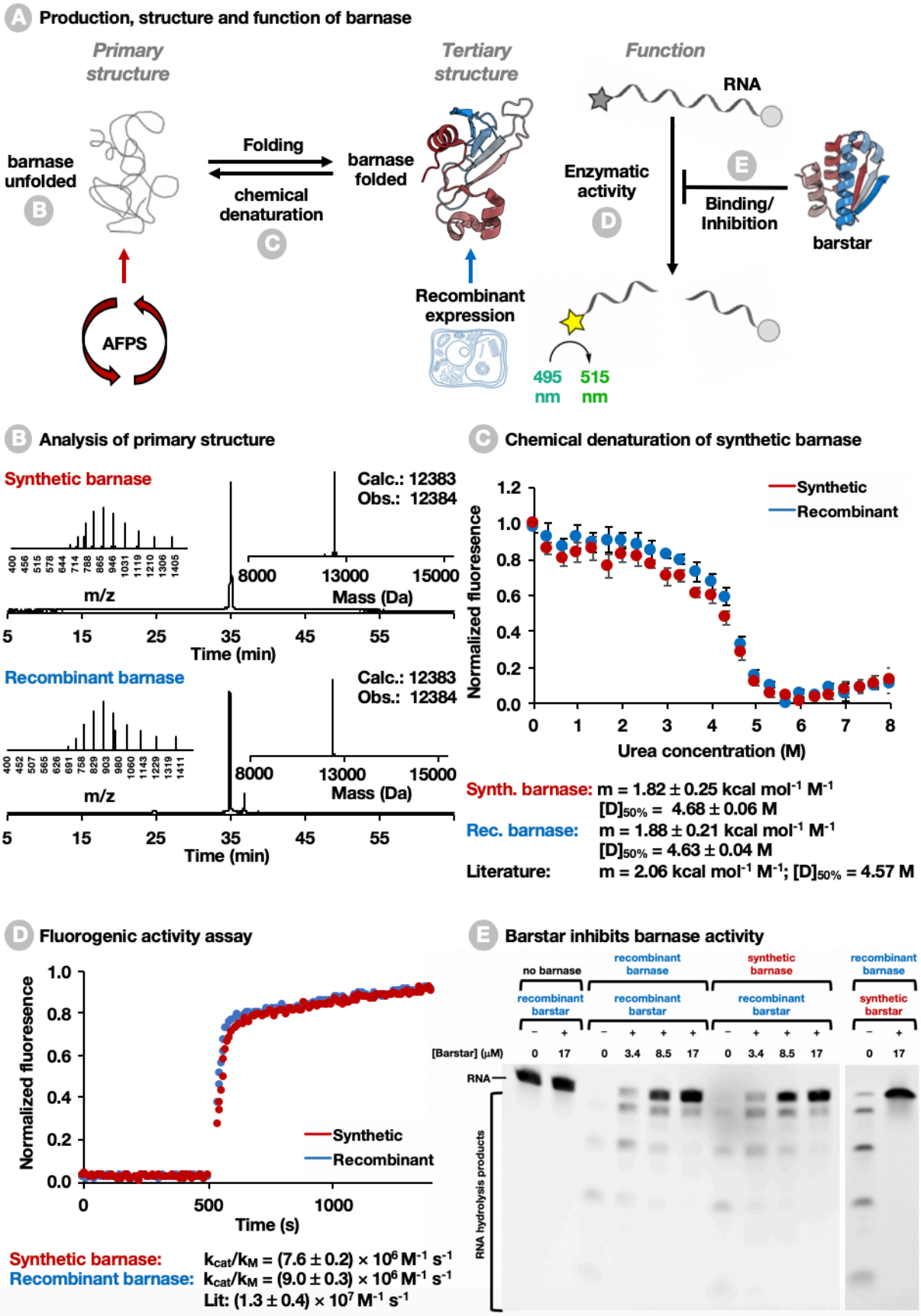
A Production, structure and function of barnase
Primary structure
Tertiary structure
Function
barnase unfolded
B
Folding
chemical denaturation
C
barnase folded
RNA
Enzymatic activity
D
E
Binding/ Inhibition
barstar
AFPS
Recombinant expression
495 nm
515 nm
B Analysis of primary structure
Synthetic barnase
Calc.: 12383
Obs.: 12384
m/z
8000
13000
15000
Mass (Da)
5
15
25
35
45
55
Time (min)
Recombinant barnase
Calc.: 12383
Obs.: 12384
C Chemical denaturation of synthetic barnase
Synthetic
Recombinant
Normalized fluoresence
Urea concentration (M)
Synth. barnase: m = 1.82 ± 0.25 kcal mol-1 M-1
[D]50% = 4.68 ± 0.06 M
Rec. barnase: m = 1.88 ± 0.21 kcal mol-1 M-1
[D]50% = 4.63 ± 0.04 M
Literature: m = 2.06 kcal mol-1 M-1; [D]50% = 4.57 M
D Fluorogenic activity assay
Normalized fluoresence
Time (s)
Synthetic
Recombinant
Synthetic barnase: kcat/kM = (7.6 ± 0.2) × 10^6 M-1 s-1
Recombinant barnase: kcat/kM = (9.0 ± 0.3) × 10^6 M-1 s-1
Lit: (1.3 ± 0.4) × 10^7 M-1 s-1
E Barstar inhibits barnase activity
no barnase
recombinant barnase
synthetic barnase
recombinant barstar
synthetic barstar
[Barstar] (μM)
0
17
3.4
8.5
RNA
RNA hydrolysis products

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). Electrospray 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*)

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:

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).

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. Phanumartwiwath, 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, 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).

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.

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).

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).

15. C. P. Gordon, The renascence 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*.

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).

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).

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), 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).

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**, 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).

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).

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).

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).

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.

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).

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).

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 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 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 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. *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).

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, 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).

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).

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 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 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 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 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 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-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

Supplementary Text

Figs. S1 to S26

Tables S1 to S16