PEPTIDE SYNTHESIS

Synthesis of proteins by automated flow chemistry

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[1,2], C. K. Schissel[1], A. J. Quartararo[1], X. Ye[1], A. J. Mijalis[1]§, M. D. Simon[1]¶, A. Loas[1], S. Liu[1]#, C. Jessen[3], T. E. Nielsen[3], B. L. Pentelute[1]**

Ribosomes can produce proteins in minutes and are largely constrained to proteinogenic amino acids. 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 long over 327 consecutive reactions. The machine is rapid: 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.

Mechanical pumps, valves, solid supports, and computers have transformed the way we perform chemical reactions. Continuous, multistep flow technology has enabled routine access to small molecules ranging from pharmaceutical ingredients to natural products and bulk commodities (*1*). Advantages of flow synthesis over batch methods are in-line spectroscopic monitoring, efficient mixing, and precise control over the reaction parameters (*2*). Translating these capabilities to the total chemical synthesis of peptides and proteins will provide rapid access to an expanded chemical space.

Protein production is an essential part of research in academia and industry and can be accomplished by biological methods or chemical synthesis. Most proteins are obtained by biological expression, a process that largely limits their chemical composition to the canonical proteinogenic amino acids (*3*). Advances in genetic code expansion have allowed for the incorporation of up to two unnatural amino acids in the structures of native proteins (*4*). By contrast, chemical synthesis offers unmatched flexibility when incorporation of multiple unnatural amino acids, posttranslational modifications, or artificial backbones is desired (*3*). 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 (*5*). Despite decades of optimization, peptides longer than 50 amino acids are difficult to synthesize with standard SPPS instrumentation, owing in large part to generation of by-products from deletion, truncation, and aggregation of the growing chains (*6*, *7*). It was not until the development of native chemical ligation (NCL) that chemical synthesis of protein chains became practical (*8*). Despite the efforts dedicated to improving NCL techniques (*9*), a major bottleneck resides in the absence of a routine, widely applicable protocol to access the requisite peptide fragments (*10*, *11*). 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—for example, control over physical parameters and greatly reduced formation of side products (*12*–*14*). Studies carried out as early as 1970 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 (*15*, *16*). Instead of complex systems for liquid handling to dispense reagents and wash the resin, high-performance liquid chromatography (HPLC) pumps were used to continuously deliver reagents, establishing the principles of peptide synthesis in flow. Inspired by this early work, over the past 5 years we developed rapid, automated fast-flow peptide synthesis (AFPS) instrumentation that incorporates amino acid residues in as little as 40 s at temperatures up to 90°C (*17*–*19*).

Even though prior work by us and others on flow-based SPPS considerably 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*–*23*). We set out to optimize our AFPS technology to meet this challenge (*19*). We report here a routine protocol that allows for stepwise chemical total synthesis of peptide chains exceeding 50 amino acids in length, with a cycle time of ~2.5 min per amino acid (Fig. 1A). The optimized protocol was built on a collection of analytical data acquired with an AFPS system and delivers products with high fidelity and of high chiral purity. Using this protocol, single-domain protein chains ranging from barstar (90 amino acids) to sortase $A_{59–206}$ (sortase A*, 164 amino acids) were synthesized in 3.5 to 6.5 hours. To demonstrate production of functional proteins, these sequences were folded, and their biophysical properties and enzymatic activities were determined. The time scale of chemical protein synthesis is on par with that of recombinant expression and therefore offers a practical alternative to biochemical methods while opening up the chemical space beyond canonical amino acids.

Rapid screening of reaction variables for AFPS protocol development

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 (*17*, *19*), reagents are mixed, heated, and delivered onto a pretempered 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 (Fig. 1B and tables S1 to S7). Modifications to our original AFPS protocol included increasing reagent concentrations to 0.4 M (*24*), the use of amine-free *N*,*N*-dimethylformamide (DMF), and an increase in temperature to 85° to 90°C for reagent activation and coupling. The performance of different activators for the coupling step was also investigated, identifying the azabenzotriazol-reagents PyAOP [(7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate] and HATU (hexafluorophosphate azabenzotriazole tetramethyl uranium) as optimal.

Automated collection and analysis of data combined with the synthesis parameters allowed for 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 (tables S8 and S9).

[1]Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [3]Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark.
*Present address: Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
†These authors contributed equally to this work.
‡Present address: p53Lab, Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Singapore 138648, Singapore.
§Present address: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA.
¶Present address: Ginkgo Bioworks, 27 Drydock Avenue, Boston, MA 02210, USA.
#Present address: Department of Chemistry, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China.
**Corresponding author. Email: blp@mit.edu

Downloaded from https://www.science.org at University of Bath on October 27, 2026

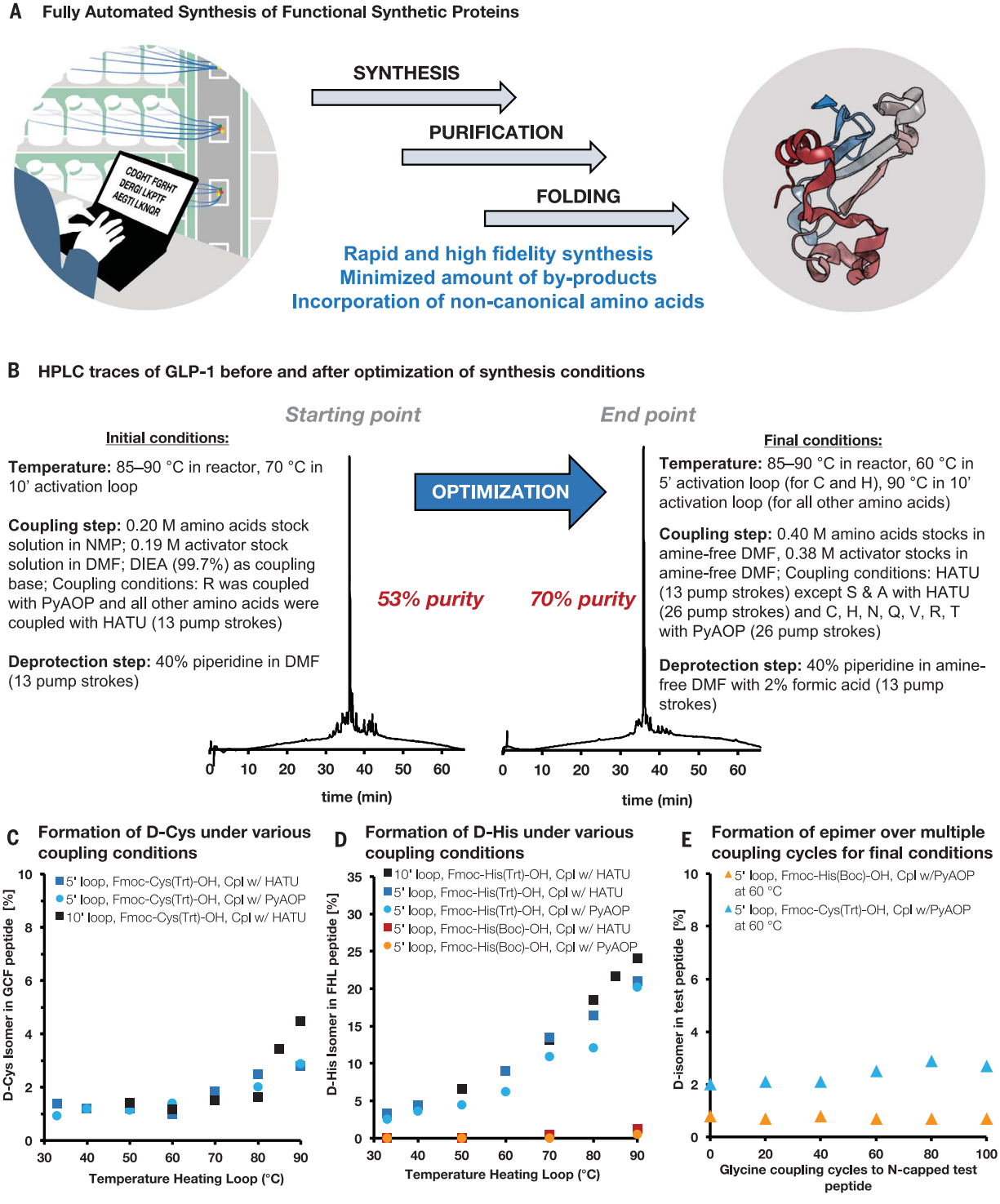


Fig. 1. Optimized conditions for automated fast-flow solid-phase peptide synthesis enable high-fidelity production of long amino acid sequences. (**A**) Fully automated chemical flow synthesis yields peptide chains, which, after purification and folding, give functional proteins. Protein Data Bank (PDB) 1BRS (barnase) (*57*) was used. (**B**) Synthesis of GLP-1 using starting conditions and optimized conditions. The concentrations listed refer to stock solutions. NMP, *N*-methylpyrrolidone; DIEA, diisopropylethylamine. (**C**) Quantification of cysteine epimerization as a function of activation temperature, heating time (5′ loop and 10′ loop), and activator in a GCF test peptide. Isomer was quantified from extracted ion chromatograms on LC-MS by comparison to reference peptides. (**D**) Quantification of histidine epimerization as a function of 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. (**E**) Quantification of epimerization over multiple coupling cycles. GCF and FHL were synthesized under optimized conditions, and the N terminus was manually capped with a Boc-protecting group. One-hundred glycine couplings were executed, and a sample was taken out for analysis every 20 amino acid couplings. Cpl w/, coupled with. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Downloaded from https://www.science.org at University of Bath on October 27, 2026

Analytical comparison of the products obtained for glucagon-like peptide-1 (GLP-1) is illustrative of the improvement in crude peptide quality achieved with the optimized synthesis conditions (Fig. 1B).

We aimed to suppress aspartimide formation, a major side reaction in SPPS and AFPS. Because increased temperature leads to more aspartimide formation, various deprotection bases, additives, and aspartic acid protecting groups were screened to minimize this unwanted side reaction (*25*, *26*). We found that milder deprotection bases [i.e., piperazine and 1-hydroxybenzotriazole (HOBt) with piperidine] and bulky aspartic acid protecting groups [i.e., *O*-3-methylpent-3-yl (OMpe)] decreased the level 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 (2% stock solution in 40:60 v/v piperidine: DMF) was therefore used as an additive for deprotection, and backbone protection was applied for collagen and fibroblast growth factor 1 (FGF1) syntheses.

We confirmed retention of chirality for amino acids at high risk of epimerization (i.e., cysteine and histidine) in a final optimization step (figs. S4 to S9) (*27*). The influence of temperature, time, and activating agent, as well as different side-chain protecting groups were screened (Fig. 1, C and D) (*17*). For both amino acids, epimerization 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 time at 60°C resulted in <2% D-epimer formation (Fmoc, fluorenylmethyloxycarbonyl; Trt, trityl; Boc, *tert*-butyloxycarbonyl). Next, we determined that the amount of epimerization under these optimized conditions does not increase over multiple coupling cycles (Fig. 1E). The amount of D-isomer did not change over 100 amino acid couplings executed after manual capping of the N-terminus, indicating that epimerization of cysteine and histidine only occurs during the activation step. Implementation of these conditions allowed us to solidify the general AFPS protocol, which was then applied to the production of sequences exceeding 50 amino acids [table S11 and supplementary materials (SM) section 3.10].

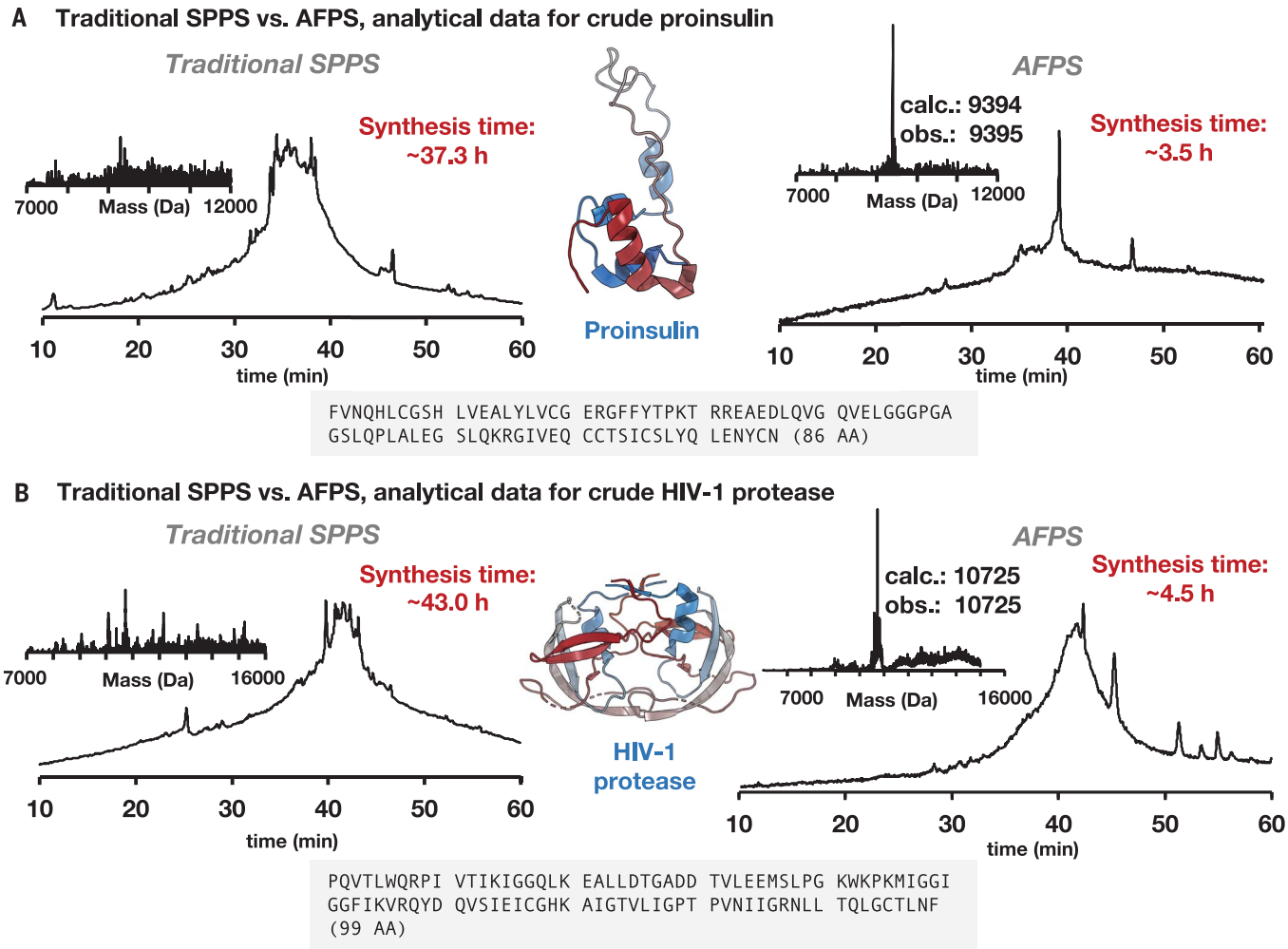


Fig. 2. Synthesis of proinsulin and HIV-1 protease demonstrates the advantage of AFPS over traditional SPPS methods. (**A** and **B**) Analytical HPLC data of the crude proinsulin (A) and HIV-1 protease (B) are presented as the main chromatographic traces with absorbance detection at 214 nm (additional details in the SM). Deconvoluted masses are displayed in the insets. Analytical data for the synthesis of crude protein chain using SPPS on a commercially available synthesizer at 70°C with total cycle times of 26 min per amino acid and 40 equivalents of amino acid for each coupling are displayed on the left; analytical data for the synthesis of crude protein chain using AFPS at 90°C with 60 equivalents of amino acid for each coupling are displayed on the right. PDB 2KQP (proinsulin) (*58*) and 2JE4 (HIV-1 protease dimer with inhibitor, not identical to the sequence synthesized) (*30*) were used. Crude material of higher quality was obtained in the case of HIV-1 protease by substitution of Cys and Met residues, as described in SM section 5. AA, amino acids.

Downloaded from https://www.science.org at University of Bath on October 27, 2026

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 human immunodeficiency virus-1 (HIV-1) protease (99 amino acids) as test sequences. The total synthesis of human proinsulin was previously reported using NCL of three peptide fragments individually prepared by SPPS (*28*). HIV-1 protease was previously prepared using stepwise and chemical ligation routes under Boc-SPPS conditions (*29*, *30*). 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 batch SPPS syntheses on commercially available synthesizers at room temperature, 70°C, and 90°C indicated substantially improved synthetic outcome for the optimized AFPS protocol (Fig. 2 and SM section 4). On each instrument, machine-specific, optimized conditions were used to achieve the best synthesis outcome. For HIV-1 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 substantial improvement when directly compared with traditional SPPS methods, both with respect to time and performance.

Optimized AFPS enables routine access to single-domain protein chains

To demonstrate general applicability of our AFPS protocol, the synthesis of additional protein chains ranging from ~70 to ~170 amino acids was performed (Fig. 3A and SM section 5). These sequences were chosen to enable comparison with literature data. We chose not only historically relevant targets for drug discovery, such as HIV-1 protease and murine double minute 2 (MDM2) (*31*, *32*), but also proteins that serve as therapeutics themselves, such as FGF1 and proinsulin (*33*, *34*). Barstar, barnase, lysozyme, MDM2, and sortase A* allowed for a direct comparison of recombinant and synthetic proteins. The ability of AFPS technology to rapidly and simultaneously incorporate noncanonical amino acids in greater number and of greater diversity than biological methods was tested by synthesizing derivatives of barnase and HIV-1 protease containing site-directed mutations. In the case of barnase, we incorporated *p*-bromophenylalanine at a site previously investigated for mutational tolerance (*35*). Then, we produced synthetic HIV-1 protease in which two methionine and one cysteine residues were replaced as previously described with norleucine and aminobutyric acid, respectively (Fig. 3B), to avoid potential oxidation side products and increase synthetic efficiency (*30*). All sequences were successfully synthesized in 3.5 to 6.5 hours.

The desired protein was the main product in every synthesis, and HPLC purification yielded milligram quantities of product. Isolated yields after HPLC purification ranged from 2.2 to 19.0 mg (1 to 5%), a sufficient amount of material for folding and evaluation of tertiary structure and biological function (Fig. 3B and SM section 5). In conclusion, optimized AFPS allows for the routine stepwise chemical synthesis of peptide chains of up to ~170 amino acids and therefore substantially decreases time and labor associated with the chemical production of single-domain proteins.

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 distinctive biological activity, is determined by its amino acid sequence (*36*). 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 (*37*).

We folded and purified selected 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 conditions.

Chemical denaturation is diagnostic for assessing structural integrity and stability of synthetic proteins. The globular protein barnase, a bacterial ribonuclease (RNase) isolated from *Bacillus amyloliquefaciens*, is a model system to investigate protein folding, denaturation, and binding to its inhibitor protein barstar (Fig. 4A) (*38*, *39*). The primary structures of synthetic and recombinant barnase were indistinguishable by LC-MS and HPLC methods (Fig. 4B). We used a chemical denaturation fluorometric assay as a readout for the integrity of the tertiary structure (Fig. 4C). In this assay, tryptophan 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, $[D]_{50\%}$) 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 (mean ± SE); $[D]_{50\%,\ literature}$ = 4.57 M} (*39*). 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} (mean ± SE); $m_{literature}$ = 2.06 kcal mol^{-1} M^{-1}] (*39*). If the synthetic protein were microheterogeneous (e.g., contained a distribution of isomers or deletion coproducts), then the apparent *m* value may be altered owing to the distribution of $[D]_{50\%}$ values represented within the mixture. Therefore, because the synthetic sample exhibited an *m* value within the error of the recombinant sample, we concluded that microheterogeneity was negligible.

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 (*40*, *41*). 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 fluorophore pair (*42*). The enzymatic efficiency of synthetic barnase was k_{cat}/K_M = (7.6 ± 0.2) × 10^6 M^{-1} s^{-1} (mean ± SE), which is comparable to that of recombinant barnase [k_{cat}/K_M = (9.0 ± 0.3) × 10^6 M^{-1} s^{-1} (mean ± SE)] determined using the same assay (Fig. 4D).

The primary structure of HIV-1 protease was confirmed by LC-MS and HPLC methods (Fig. 5B). HIV-1 protease hydrolyzes the peptides of HIV, and using a fluorogenic peptide allows for quantification of its proteolytic activity (*43*). Synthetic HIV-1 protease displays a Michaelis constant of K_M = 20.9 ± 1.0 μM (mean ± SE) and a turnover number of k_{cat} = 29.6 ± 4.1 s^{-1} (mean ± SE), close to literature values published for a similar synthetic sample obtained by SPPS (Fig. 5C) (*30*). 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. 5D) (*29*).

Sortase $A_{59–206}$ is a transpeptidase produced by Gram-positive bacteria that catalyzes a cell wall sorting reaction at a threonine-glycine

Downloaded from https://www.science.org at University of Bath on October 27, 2026

bond in the LPXTG motif (Leu-Pro-X-Thr-Gly, where X is any amino acid) (*44*). We synthesized the 164–amino acid–long sortase A* variant (P94S/D160N/K196T; P, Pro; S, Ser; D, Asp; N, Asn; K, Lys; T, Thr) to allow for direct comparison to a recombinant standard (*45*, *46*). At a concentration of 0.01 mg/ml, synthetic sortase A* led to 47% product formation by LC-MS within 24 hours (starting from 0.2 mg/ml GGGGGLY and AQALPETGEE as test substrates; G, Gly; L, Leu; Y, Tyr; A, Ala; Q, Gln; E, Glu) (fig. S23). This conversion value is comparable to that determined for the recombinant protein (50% product formation within 24 hours). Enzymatic activity assays of synthetic

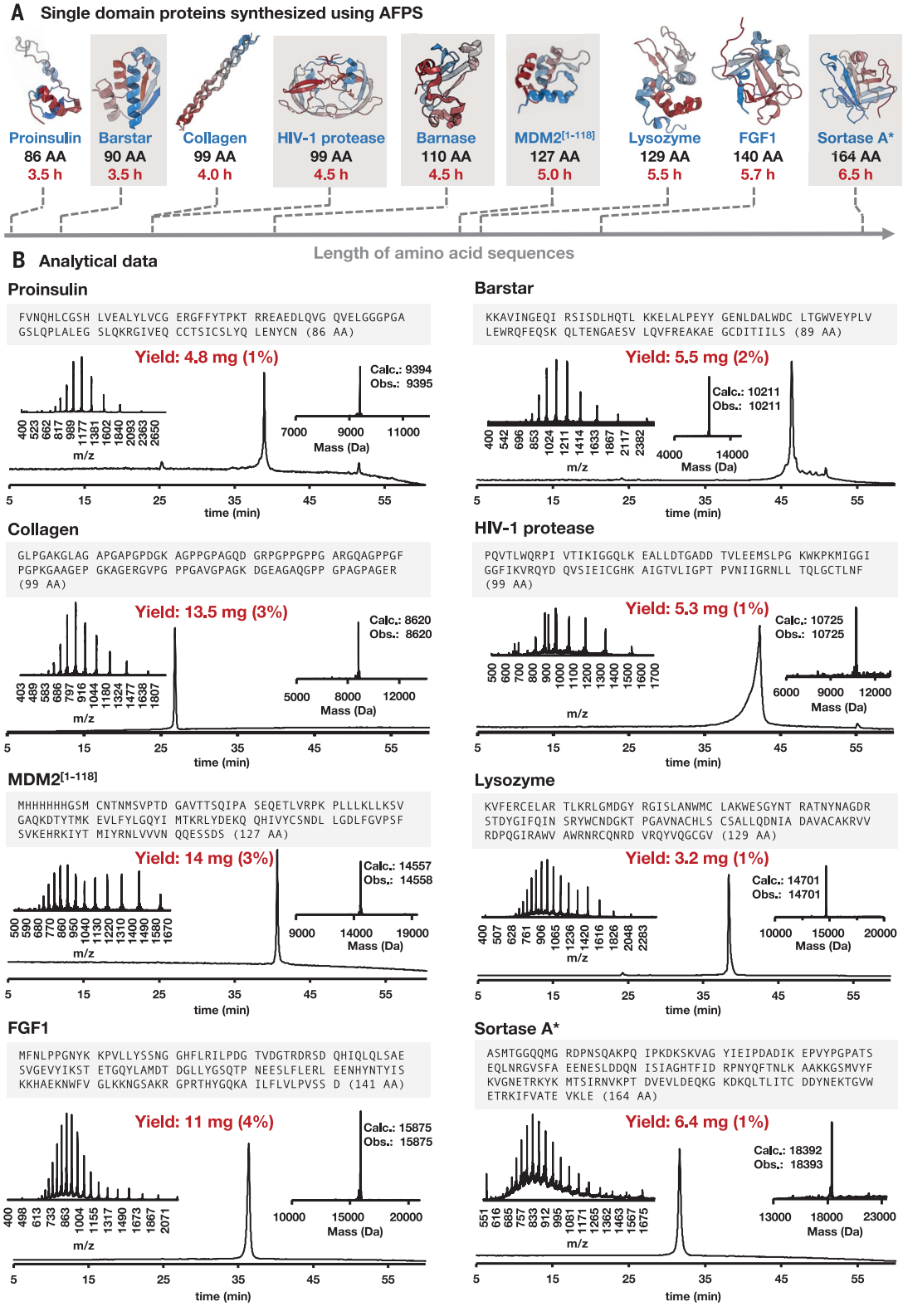


Fig. 3. AFPS enables high-fidelity production of long amino acid sequences in hours. (**A**) Sequences produced using an AFPS instrument. Sequences highlighted in gray were folded and purified, and their structure and biological activity were evaluated. All sequences were synthesized using the same standard recipe. PDB 1AY7 (barstar) (*59*), 2KQP (proinsulin) (*58*), 1CGD (collagen) (*60*), 2JE4 (HIV-1 protease dimer with inhibitor) (*30*), 1BRS (barnase) (*57*), 3G03 (MDM2) (*61*), 2NWD (lysozyme) (*62*), 4Q9G (FGF1) (*63*), and 2KID (sortase A) (*64*) were used. (**B**) Analytical data for the purified sequences of proinsulin, barstar, collagen, HIV-1 protease, MDM2[1–118], lysozyme, FGF1, and sortase A*. For all cases, analytical HPLC data of the purified protein chains are presented as the main chromatographic trace with absorbance detection at 214 nm. The gradient for analytical HPLC was 5 to 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. 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 LC peak; additional details on purification and analytical methods are in the SM section 5.

Downloaded from https://www.science.org at University of Bath on October 27, 2026

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. 4E) (*47*). 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] (*32*). The binding of MDM2 to p53 is a key

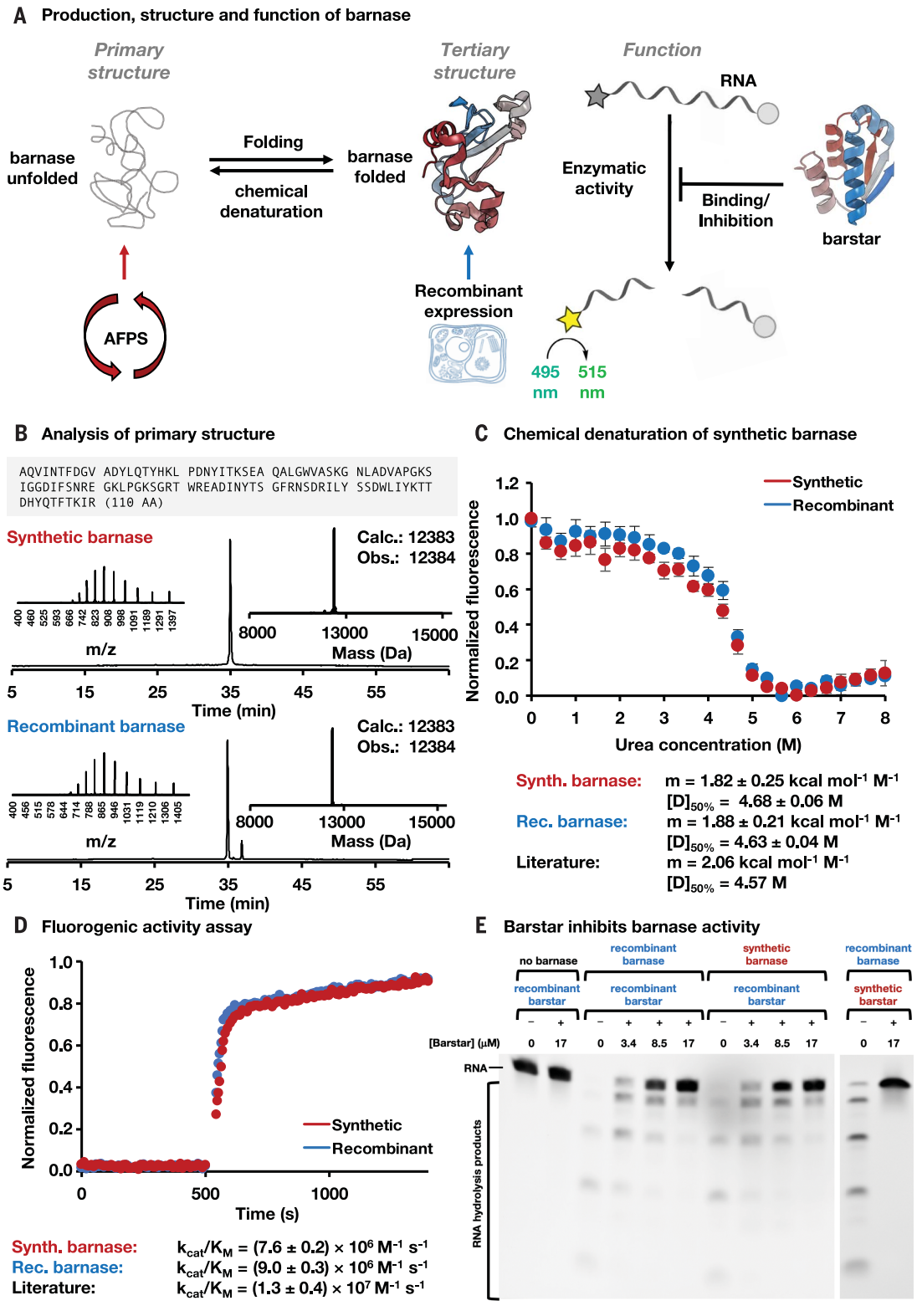


Fig. 4. Synthetic barnase and synthetic barstar fold into the native tertiary structure and display enzymatic activity comparable to recombinant samples. (**A**) Conceptual overview of production and analysis methods. (**B**) Comparison of primary structures obtained from AFPS and recombinant expression. 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 in the SM). 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 LC peak in the chromatogram. (**C**) Structural evaluation of barnase in a chemical denaturation assay using urea as denaturant performed in triplicate; results are reported as mean ± SE. Error bars on the graph indicate SE. (**D**) Quantitative enzymatic activity assay performed in triplicate; error bars are not displayed for clarity. Details are outlined in the SM. k_{cat}/K_M values are reported as mean ± SE. (**E**) Barnase inhibition and binding assay using recombinant and synthetic barstar. 3.4 nM barnase was used in all conditions. Details are outlined in the SM. PDB 1BRS (barnase) (*57*) and 1AY7 (barstar) (*59*) were used.

Downloaded from https://www.science.org at University of Bath on October 27, 2026

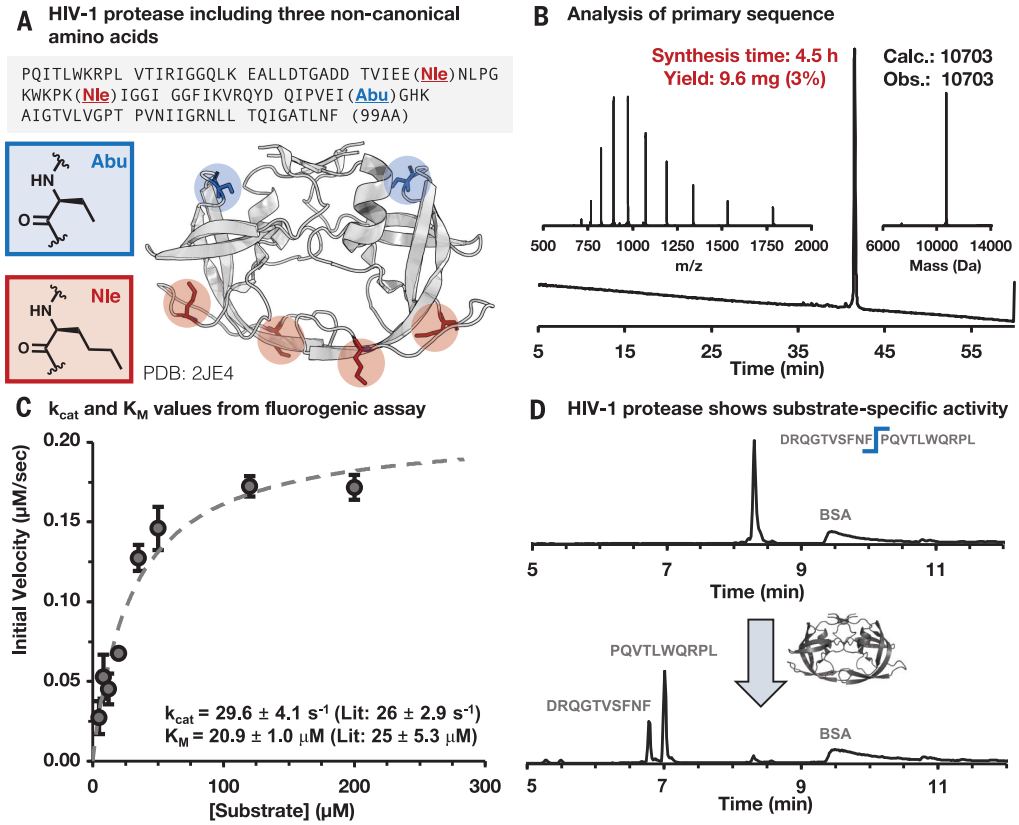


Fig. 5. Synthetic HIV-1 protease containing three noncanonical amino acids folds into the native dimer structure and displays enzymatic activity and substrate specificity comparable to literature samples. (**A**) Crystal structure of HIV-1 protease dimer with highlighted noncanonical amino acids aminobutyric acid (Abu, blue) and norleucine (Nle, red). PDB 2JE4 (HIV-1 protease dimer with inhibitor) (*30*) was used. (**B**) Primary structure obtained from AFPS. Analytical HPLC data of the purified HIV-1 protease is presented as the main chromatographic trace with absorbance detection at 214 nm (additional details in the SM). ESI mass spectrum and deconvoluted mass spectrum of the purified sample are displayed in the upper-left and the upper-right insets, respectively. Both spectra were obtained by summation over the entire LC peak in the chromatogram. (**C**) Quantitative enzymatic activity assay performed in triplicate for the determination of k_{cat} and K_M values. Results are reported as mean ± SE. Error bars on the graph indicate SE. Lit., literature. (**D**) Qualitative substrate specificity assay with model substrate p12nt, in which HIV-1 protease exclusively cleaves at a single Phe-Pro site whereas bovine serum albumin (BSA) stays intact; details are outlined in the SM.

interaction in multiple pathways up-regulated in cancer (*48*, *49*). We folded milligram quantities of synthetic MDM2$^{[1–118]}$ and characterized its binding to immobilized p53$^{[14–29]}$ using biolayer interferometry (figs. S24 and S25). Synthetic MDM2$^{[1–118]}$ displayed an affinity toward p53 [dissociation constant (K_d) = 6.25 μM] comparable to the literature value (K_d = 5.45 μM) obtained under the same folding conditions.

Discussion

The optimized AFPS protocol demonstrates advantages of flow chemistry over common batch methods, yielding peptide chains more than three times longer than previously accessible by routine standard SPPS (*6*). 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 SPPS 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. A potentially limiting feature of our setup is synthesis scale. The capacity of the reactor used in our study allows up to 200 mg of resin with a loading of 0.49 mmol/g. Increased production output can be achieved by incorporating a larger reactor in the current system, but such a modification will likely require specific optimization, toward which we performed preliminary investigations (*19*). Since we implemented AFPS, we have produced more than 5000 peptides and automatically collected in-line analysis data for all syntheses. Moving forward, this extensive, high-quality dataset 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 (*50*).

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. Our 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. 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 (those with a mass of up to ~30 kDa) (*10*, *51*). In this respect, we envisage adapting to our AFPS protocol the incorporation of peptide hydrazides for thioester-based ligation, an approach previously achieved with manual flow instrumentation (*52*). Additional research avenues opened by our method include rapid access to mirror-image proteins, posttranslationally modified proteins, and de novo–designed, abiotic proteins. Introduction of noncanonical amino acids as point mutations in native proteins will make accessible variants with considerably altered biological function, for example, catalytic activity (*53*, *54*). 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 (*55*, *56*).

REFERENCES AND NOTES

1. J. Britton, C. L. Raston, *Chem. Soc. Rev.* **46**, 1250–1271 (2017).
2. D. Webb, T. F. Jamison, *Chem. Sci.* **1**, 675–680 (2010).
3. S. Bondalapati, M. Jbara, A. Brik, *Nat. Chem.* **8**, 407–418 (2016).
4. J. W. Chin, *Nature* **550**, 53–60 (2017).
5. R. B. Merrifield, *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
6. S. B. H. Kent, *Chem. Soc. Rev.* **38**, 338–351 (2009).
7. A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, *Future Med. Chem.* **1**, 361–377 (2009).
8. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **266**, 776–779 (1994).
9. V. Agouridas *et al.*, *Chem. Rev.* **119**, 7328–7443 (2019).
10. P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **69**, 923–960 (2000).
11. A. El-Faham, F. Albericio, *Chem. Rev.* **111**, 6557–6602 (2011).

Downloaded from https://www.science.org at University of Bath on October 27, 2026

12. N. Ahmed, *Chem. Biol. Drug Des.* **91**, 647–650 (2018).
13. C. P. Gordon, *Org. Biomol. Chem.* **16**, 180–196 (2018).
14. A. J. Mijalis, A. Steinauer, A. Schepartz, B. L. Pentelute, in *Flow Chemistry in Organic Synthesis* (Georg Thieme Verlag, Stuttgart, 2018), pp. 381–398.
15. E. Bayer, G. Jun, I. Halász, I. Sebestian, *Tetrahedron Lett.* **11**, 4503–4505 (1970).
16. T. J. Lukas, M. B. Prystowsky, B. W. Erickson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2791–2795 (1981).
17. A. J. Mijalis *et al*., *Nat. Chem. Biol.* **13**, 464–466 (2017).
18. S. K. Mong, A. A. Vinogradov, M. D. Simon, B. L. Pentelute, *ChemBioChem* **15**, 721–733 (2014).
19. M. D. Simon, "Fast flow biopolymer synthesis," thesis, Massachusetts Institute of Technology, Cambridge, MA (2017).
20. S. Fuse, Y. Otake, H. Nakamura, *Chem. Asian J.* **13**, 3818–3832 (2018).
21. S. Fuse, Y. Mifune, T. Takahashi, *Angew. Chem. Int. Ed.* **53**, 851–855 (2014).
22. L. K. Spare, V. Laude, D. G. Harman, J. R. Aldrich-Wright, C. P. Gordon, *React. Chem. Eng.* **3**, 875–882 (2018).
23. E. T. Sletten, M. Nuño, D. Guthrie, P. H. Seeberger, *Chem. Commun.* **55**, 14598–14601 (2019).
24. M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Protein Res.* **40**, 180–193 (1992).
25. T. Michels, R. Dölling, U. Haberkorn, W. Mier, *Org. Lett.* **14**, 5218–5221 (2012).
26. M. Mergler, F. Dick, B. Sax, P. Weiler, T. Vorherr, *J. Pept. Sci.* **9**, 36–46 (2003).
27. S. A. Palasek, Z. J. Cox, J. M. Collins, *J. Pept. Sci.* **13**, 143–148 (2007).
28. S. Luisier, M. Avital-Shmilovici, M. A. Weiss, S. B. H. Kent, *Chem. Commun.* **46**, 8177–8179 (2010).
29. J. Schneider, S. B. H. Kent, *Cell* **54**, 363–368 (1988).
30. E. C. B. Johnson *et al*., *J. Am. Chem. Soc.* **129**, 11480–11490 (2007).
31. A. Brik, C.-H. Wong, *Org. Biomol. Chem.* **1**, 5–14 (2003).
32. J. D. Oliner *et al*., *Nature* **362**, 857–860 (1993).
33. D. Bresson *et al*., *J. Clin. Invest.* **116**, 1371–1381 (2006).
34. E. Gasser, C. P. Moutos, M. Downes, R. M. Evans, *Nat. Rev. Endocrinol.* **13**, 599–609 (2017).
35. L. Serrano, J. T. Kellis Jr., P. Cann, A. Matouschek, A. R. Fersht, *J. Mol. Biol.* **224**, 783–804 (1992).
36. C. B. Anfinsen, *Science* **181**, 223–230 (1973).
37. S. B. Kent, P. F. Alewood, *Curr. Opin. Chem. Biol.* **22**, viii–xi (2014).
38. R. W. Hartley, *Trends Biochem. Sci.* **14**, 450–454 (1989).
39. J. T. Kellis Jr., K. Nyberg, A. R. Fersht, *Biochemistry* **28**, 4914–4922 (1989).
40. J. R. Knowles, *Science* **236**, 1252–1258 (1987).
41. U. Arnold *et al*., *J. Am. Chem. Soc.* **125**, 7500–7501 (2003).
42. B. R. Kelemen *et al*., *Nucleic Acids Res.* **27**, 3696–3701 (1999).
43. M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **36**, 544–550 (1990).
44. H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12424–12429 (1999).
45. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao, B. L. Pentelute, *J. Am. Chem. Soc.* **134**, 10749–10752 (2012).
46. R. L. Policarpo *et al*., *Angew. Chem. Int. Ed.* **53**, 9203–9208 (2014).
47. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, *Chem. Sci.* **6**, 2997–3002 (2015).
48. F. Touti, Z. P. Gates, A. Bandyopadhyay, G. Lautrette, B. L. Pentelute, *Nat. Chem. Biol.* **15**, 410–418 (2019).
49. P. Chène, *Mol. Cancer Res.* **2**, 20–28 (2004).
50. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **341**, 1238149 (2013).
51. S. J. Wheelan, A. Marchler-Bauer, S. H. Bryant, *Bioinformatics* **16**, 613–618 (2000).
52. M. D. Simon *et al*., *ChemBioChem* **15**, 713–720 (2014).
53. J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, *J. Am. Chem. Soc.* **140**, 15997–16000 (2018).
54. H. Xiao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6961–6966 (2015).
55. R. J. Desnick, E. H. Schuchman, *Nat. Rev. Genet.* **3**, 954–966 (2002).
56. N. L. Truex *et al*., *Sci. Rep.* **10**, 723 (2020).
57. A. M. Buckle, G. Schreiber, A. R. Fersht, *Biochemistry* **33**, 8878–8889 (1994).
58. Y. Yang *et al*., *J. Biol. Chem.* **285**, 7847–7851 (2010).
59. J. Sevcík, L. Urbanikova, Z. Dauter, K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.* **54**, 954–963 (1998).
60. J. Bella, B. Brodsky, H. M. Berman, *Structure* **3**, 893–906 (1995).
61. A. Czarna *et al*., *Cell Cycle* **8**, 1176–1184 (2009).
62. T. Durek, V. Y. Torbeev, S. B. H. Kent, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4846–4851 (2007).
63. X. Xia, L. M. Longo, M. Blaber, *J. Pharm. Sci.* **104**, 566–576 (2015).
64. N. Suree *et al*., *J. Biol. Chem.* **284**, 24465–24477 (2009).
65. A. J. Mijalis, amijalis/afps-integration: AFPS-analysis: Code used to extract and analyze data from AFPS synthesis files. Zenodo (2020).

ACKNOWLEDGMENTS

We thank T. F. Jamison, H. U. Stilz, L. F. Iversen, K. Little, and D. Lundsgaard for productive discussions and administrative support. We acknowledge the participants of the 26th American Peptide Symposium and the 8th Chemical Protein Synthesis Meeting, especially P. E. Dawson, R. T. Raines, and S. B. H. Kent, for helpful discussions and for suggesting additional experiments. We also thank E. D. Evans, F. W. W. Hartrampf, and R. L. Holden for careful proofreading of the manuscript. We are grateful to N. L. 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., and 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., A.J.M., and S.L. optimized synthesis conditions; A.J.M. provided the software for UV data analysis, and M.D.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 coauthors. **Competing interests:** B.L.P. is a cofounder of Amide Technologies and Resolute Bio. Both companies focus on the development of protein and peptide therapeutics. A.J.M. and M.D.S. hold equity in Amide Technologies. The following authors are inventors on patents and patent applications related to the technology described: A.J.M., M.D.S., and B.L.P. are co-inventors on U.S. patent application 20170081358A1 (23 March 2017) describing methods and systems for solid-phase peptide synthesis; M.D.S. and B.L.P. are co-inventors on U.S. patents 9,868,759 (16 January 2018), 9,695,214 (4 July 2017), and 9,169,287 (27 October 2015) describing solid-phase peptide synthesis processes and associated systems. **Data and materials availability:** Code for analysis of the UV-vis traces obtained from the AFPS instrument is available in a GitHub repository (*65*). All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6494/980/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S25
Tables S1 to S14
References (*66*–*71*)
MDAR Reproducibility Checklist

10 February 2020; accepted 30 March 2020
10.1126/science.abb2491

Downloaded from https://www.science.org at University of Bath on October 27, 2026

Science
AAAS

Synthesis of proteins by automated flow chemistry

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

Science **368** (6494), . DOI: 10.1126/science.abb2491

Fully synthetic whole proteins in reach

Solid-phase peptide synthesis of homogeneous peptides longer than about 50 amino acids has been a long-standing challenge because of inefficient coupling and side reactions. Hartrampf *et al.* used an automated chemistry platform to optimize fast-flow peptide synthesis and were able to produce fully synthetic single-domain proteins (see the Perspective by Proulx). The targets included proinsulin and enzymes such as barnase and a version of HIV-1 protease containing multiple noncanonical amino acids. Refolded peptides were nearly indistinguishable from recombinant proteins, and the synthesized enzymes had activity close to that of their ribosomally synthesized counterparts. This method will enable fast, on-demand synthesis of small proteins with a vastly expanded pool of precursor amino acids.

Science, this issue p. 980; see also p. 941

View the article online
https://www.science.org/doi/10.1126/science.abb2491
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from https://www.science.org at University of Bath on October 27, 2026