Rapid Generation of Hyperdiverse Chemical Libraries

John S. Albin[1,2,3], Gha Young Lee[1,3], Corey Johnson[2,3], Dimuthu Ashcharya Vithanage[2,3], Wayne Vuong[1], and Bradley L. Pentelute[1,4-6*]

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

[2]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, 02114, USA

[3]Harvard Medical School, Boston, MA, 02115, USA

[4]The Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[5]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

*Correspondence: blp@mit.edu

Summary: Combinatorial peptidomimetic libraries facilitate the economical identification and refinement of lead compounds directed at diverse therapeutic targets. Further development of selection-based approaches to drug discovery utilizing such libraries is impeded, however, both by the slow pace of library generation and by the physical limitations to library diversity inherent to current methods. To overcome these barriers, we describe here the adaptation of peptide flow synthesis technology to the generation of combinatorial libraries. Using a simple and accessible semi-automated flow platform, we demonstrate methods for library synthesis including both canonical and noncanonical amino acid building blocks in a format that quickens the pace of library synthesis from days to < 1 hour per library while permitting individual library diversity orders of magnitude beyond current approaches up to a physical maximum of 10^{19} members. Flow synthesis is thus a powerful approach for the rapid generation of hyperdiverse libraries for selection-based drug discovery.

Keywords: Peptide, peptidomimetic, flow synthesis, combinatorial synthesis, chemical library, affinity selection-mass spectrometry

Introduction

Polyamide libraries are valuable tools for drug discovery. A wide variety of technologies may be used to generate such libraries, including molecular biology-based methods such as phage display and mRNA display(1,2). Although the latter of these, in particular, is capable of generating very large libraries of up to 10^{13} members, these methods are generally not equipped to readily incorporate extensive building block diversity outside of the 20 canonical amino acids except in selected, highly-engineered adaptations(3). They also suffer from the inclusion of large tags that may impact binding or other biological activities and typically have turnaround times on the order of weeks for the generation of custom libraries.

Polyamide libraries may also be made through combinatorial synthesis for use in selection systems such as affinity selection-mass spectrometry(4) (AS-MS). Although the maximum library diversity accessible under these approaches yields several orders of magnitude fewer members per library than would be accessible with molecular biology methods, synthetic libraries have the advantages of facile and complete chemical control over building block diversity and downstream library modifications. Library deconvolution via mass spectrometry further obviates the need for tagging systems, which in turn allows selection on lead compounds themselves without additional confounding moieties.

The methods used to make synthetic combinatorial polyamide libraries have historically fallen into one of two groups, both based on solid-phase synthesis. In split-and-pool synthesis, the addition of individual building blocks to a polyamide library at a given monomer site occurs on resin aliquots physically separated into different reaction vessels, which over multiple rounds of splitting and pooling of resin reliably generates diverse combinatorial libraries(5). Advantages of this approach include the ability to use large excesses of individual building blocks during

coupling to drive reactions to completion as well as the ability to approximate equimolar synthesis of each library member.

Despite its utility, there are several limitations to split-and-pool synthesis, the chief among these being the slow pace of synthesis. In the absence of automated robotic systems, the generation of any one library of modest length requires the full attention of one individual for days at a time, a bottleneck that limits both the ability to adapt such libraries to a given target with a practical turnaround time and to generate greater numbers of increasingly diverse libraries for applications such as the training of artificial intelligence tools(6). This synthetic speed bump further limits the ability to directly incorporate diversity regions into the synthetic protein domains accessible through automated fast flow synthesizers(7).

Aside from issues of speed and practicality, a further drawback to split-and-pool synthesis is that it results in one-bead, one-compound libraries, in which each bead is occupied by a single compound at a fixed molar quantity. This fixed molar quantity per compound, in turn, imposes major limitations to total library diversity, as increasing diversity requires increasing the number of individual beads to numbers occupying impractical amounts of physical space. Moreover, in the absence of the ability to reduce the molar quantity of each compound below the atomistic level established by bead size, library solubility past a certain level of diversity in volumes practical for a given selection is unlikely. In practice, the synthetic diversity limit of such libraries is around 10^9 members(8).

A second method for the generation of synthetic combinatorial polyamide libraries is the use of building block mixtures to randomize the incorporation of individual building blocks at each monomer site, a methodology developed extensively in the 1980-1990s(9). Synthesis under these methods proceeds in a fashion identical to that of a standard peptide synthesis, with one

caveat: the different reaction rates of individual building blocks at each monomer site necessitate molar compensation to normalize incorporation of building blocks to a targeted level – *e.g.*, equimolar quantities. The ability to make these compensatory changes, in turn, requires the empiric determination of the relative reaction rates of different monomer units in the system under use(10–15).

In principle, mixture-based library synthesis is compatible with the rapid flow-based peptide and protein synthesis systems developed in our laboratory over the past decade. These systems shorten the time required for synthesis by an order of magnitude relative to batch methods, allowing full coupling cycles in 1-3 minutes(16,17), while optimized implementations have the efficiency to make fully synthetic protein domains in a matter of hours(18–20).

In addition to speed and synthetic quality, the adaptation of our flow-based systems to library generation would have the added benefit of bypassing bead-based physical limitations on synthetic diversity. That is, in a mixture-based system in which all resin reaction sites are available for reaction at all times, the total mass of library synthesized per resin mass remains fixed, while the molar quantity of each individual library member varies inversely with library diversity. Thus, in mixture-based synthesis, the maximum achievable library diversity approaches the number of reaction sites in a given resin mass – *e.g.*, in 0.1 g of a 0.25 mmol/g loading resin, there are 2.5×10^{-5} mol * 6.02×10^{23} mol^{-1} reaction sites, allowing a theoretical physical maximum diversity of 10^{19} individual library members. Moreover, by holding library mass constant, libraries many orders of magnitude greater than can currently be made are more likely to remain soluble in practical volumes of solvent during selection experiments despite exponential expansions in library diversity. These and other advantages of flow synthesis over

current technology for library generation are summarized in **Figure 1** and in **Supplementary Table 1**.

Here we describe the adaptation of fast flow peptide synthesis technology to the generation of peptide libraries using building block mixtures. Starting from an empiric determination of the molar adjustments necessary to normalize building block incorporation in our flow-based system, we find that such adjustments can be approximated *in silico* through the calculated building block gyration radius, thus obviating the need for an empiric determination of incorporation efficiency for each new building block. Finally, we show that this system is able to generate libraries of expected content to a diversity approaching the theoretical maximum of 10^{19} distinct peptides in a single shot. Flow synthesis is thus a powerful method for the rapid generation of hyperdiverse polyamide libraries for drug discovery and development.

Results

Modest variability of amino acid reaction rates in semi-automated flow peptide synthesis

Multiple prior studies have indicated that the molar ratios of amino acid mixtures can be adjusted to achieve approximately equimolar building block incorporation at monomer sites in polyamide libraries made with different synthetic approaches(10,14,15,21). To determine the adjustments necessary in our semi-automated flow synthesis system, we made a series of 19 model 9-mer peptides of sequence VQRI**x**DFLR in which the 5th monomer position **x** was varied by the incorporation of equal volumes of Gly versus any of the other 19 canonical amino acids. These 1:1 competition peptides were then cleaved and analyzed by high performance liquid chromatography (HPLC) to quantify the intrinsic incorporation efficiency of each amino acid relative to Gly as evidenced by the integration of peaks observed at 205 nm, which permits for downstream correction for side-chain contributions to absorbance via a previously published method(22) when converting ratios for detection by mass spectrometry.

Figure 2 shows the relative incorporation efficiencies of the 19 canonical amino acids relative to Gly as detected at 214 nm (A-B), 205 nm (C-D), or 205 nm with adjustment for sidechain contributions to backbone absorbance via a previously published method (E-F). With adjustment for sidechain contributions, most amino acid incorporation efficiencies in our semi-automated flow system fall within a factor of two on either side of Gly artificially set to 1, for a dynamic range of approximately four. Of note, the amino acid proline is the lone canonical building block that substantially varies above 1 relative to Gly once absorbance is corrected for sidechain contributions. Setting this exception aside, then, most of the remaining canonical building blocks demonstrate incorporation of roughly 0.5-1 relative to Gly. Intrinsic variability of amino acid reaction rates in semi-automated flow synthesizers is thus limited and likely

amenable to the adjustments necessary to target amino acid incorporation to desired levels, equimolar or otherwise. Raw ratios of amino acid incorporation relative to Gly for each of 3-5 syntheses per test peptide as shown in **Figure 2** are summarized in **Supplementary Table 2**. Supporting analytical data from each synthesis are shown in **Supplementary Figures 1-19** (**Synthesis 1**), **20-38** (**Synthesis 2**), **39-57** (**Synthesis 3**), and **58-65** (**Syntheses 4-5** of selected test peptides). Conversion factors accounting for sidechain contributions to absorbance at 205 nm are provided in **Supplementary Table 4**.

Minimal effect of the preceding amino acid on downstream building block incorporation

In order for mixture-based synthesis to generate compound libraries with contiguous diversity sites, the effect of the peptide C-terminal amino acid on subsequent building block incorporation efficiencies at the monomer extension site must be either equal or at least equally felt by all downstream building blocks. That is, if a sterically hindered preceding amino acid causes a reduction in the Gly incorporation efficiency to half of that observed with a less hindered preceding amino acid, the reaction rate of another amino acid such as Arg must also be halved. While proportional effects have been observed in some systems(10,15), others have reported differences so great as to result in the loss of a large proportion of library members even in simple dipeptides(23).

To assess the effects of the peptide C-terminal amino acid on downstream building block incorporation in semi-automated flow synthesis, we prepared a series of 21 peptides of the design VQRI**xy**FLR, in which we utilized equal volumes of the sterically hindered and unhindered amino acids Arg and Gly as a coupling mixture at the **x** site and varied the **y** site to any of the 20 canonical amino acids plus the disubstituted amino acid aminoisobutyric acid

(Aib). As shown in **Figure 3**, the ratio of Arg to Gly remained relatively constant regardless of the upstream amino acid, with most ratios falling within a standard deviation of 0.05 about a mean of 0.5 for **y** site amino acids when excluding the sterically hindered α,α-disubstituted amino acid Aib (range 0.37-0.60) – or a standard deviation of 0.08 about a mean of 0.5 when including Aib. In either case, it is evident that Aib and to a lesser extent Pro caused disproportionately greater decreases in coupling of Arg relative to Gly at the **x** monomer extension site. There may thus be potential to bias library content where diversified monomer extension sites follow hindered α,α-disubstituted amino acids such as Aib or, to a lesser extent, a secondary amino acid such as proline. Outside of these special cases, however, our data suggest that most amino acids do not substantially bias relative downstream monomer reaction rates regardless of sidechain content. Moreover, use of Aib, Pro, or similar amino acids may still be feasible, so long as the downstream extension site is not a diversity site – or so long as one is willing to accept a degree of library content bias. Supporting analytical data from each synthesis are shown in **Supplementary Figures 66-86**, and the numerical Arg:Gly ratio for each preceding **y** site amino acid is provided in **Supplementary Table 3**.

Relative incorporation correlates with Fmoc building block gyration radius

One advantage of combinatorial synthesis for library generation over molecular biology methods is the ability to easily incorporate noncanonical amino acid building blocks to access chemical space beyond that accessible to canonical amino acids. Although it may be possible to empirically determine the reaction rate of every noncanonical amino acid that one might wish to use in a given library, it would be simpler to use calculated parameters to predict the adjustments

necessary to incorporate a given noncanonical amino acid – particularly where some degree of stochasticity makes it possible only to approximately control target incorporation.

Among the parameters important for reactivity in solid phase synthesis, the temperature, flow rate, solvent, and total reactant concentrations are constant in our semi-automated flow system. Other variables that may influence amino acid coupling are the diffusion rates and intrinsic reactivities of individual amino acid(24). Diffusion and intrinsic reactivity, in turn, may each be impacted by amino acid size and associated steric favorability or lack thereof. We therefore plotted the calculated gyration radii of the 20 canonical amino acids against the empirically observed reaction rates adjusted for sidechain contributions to absorbance at 205 nm as shown in **Figure 1E-F**. As shown in **Figure 4A**, this plot produces a curve that may be modeled as a hyperbola with an r-squared value of 0.70, consistent with the inverse relationship with between the size and diffusion coefficient of a given building block. Although additional factors may influence the reaction rate of any given building block, gyration radius likely accounts for the majority of the differences in reaction rates observed among Fmoc building blocks.

To determine whether noncanonical amino acids follow the same trend, we then characterized the relative incorporation of a series of singly synthesized noncanonical competition peptides of the design utilized in **Figure 1**, VQRI**x**DFLR where **x** is equal volumes of Gly versus a given noncanonical amino acid. When plotted alongside canonical amino acids in **Figure 4B**, these noncanonical incorporation rates generally followed the same curve, in this case yielding an r-squared value of 0.64 when pooled with canonical amino acid syntheses and 0.74 when analyzed as a separate series of single noncanonical syntheses. It may therefore be possible to utilize calculated gyration radii to predict the molar adjustments necessarily to control

incorporation of noncanonical building blocks where an empiric determination is impractical. Raw ratios of amino acid incorporation relative to Gly for each noncanonical amino acid tested are summarized in **Supplementary Table 5**; supporting analytical data are shown in **Supplementary Figures 87-99**.

To further explore the potential relationship between gyration radii and empirically observed incorporation ratios in semi-automated flow synthesis, we next proceeded to make a series of test peptides of the design **x**QRIKDFLR, where **x** is equal volumes of equimolar solutions of Gly versus any of 8 carboxylic acids covering a broader range of sizes than is readily achievable with Fmoc amino acids, where the primary limitation is the basal degree of steric hindrance imposed by the Fmoc group itself. As shown in **Figure 4C**, relative incorporation generally follows the hyperbolic model specified in **Figure 4B** for octanoic and elaidic acid. The higher than predicted incorporation ratio observed for oleic acid may be explained by the ability of this cis isomer of elaidic acid to take on more compact conformations than predicted by the calculated gyration radius. For carboxylic acids with gyration radii below approximately 3.3 Å, however, there was no apparent relationship between gyration radius and incorporation relative to Gly. Rather, the data suggest an incorporation plateau among these smaller building blocks, the exception being acrylic acid, thus further demonstrating the potential for intrinsic reactivity rather than steric factors to drive incorporation rates. Although building blocks of a size typical of Fmoc amino acids may be modeled by gyration radii, our data would suggest that smaller building block reaction rates are likely to be predominantly governed by other factors. Raw ratios of amino acid incorporation relative to Gly for each N-terminal carboxylic acid tested are summarized in **Supplementary Table 6**; supporting analytical data are shown in **Supplementary Figures 100-108**. Calculated gyration radii for all species tested in **Figure 4** are

summarized in **Supplementary Table 7**; gyration radii throughout were calculated with Vega-ZZ(25).

Molar compensation in building block mixtures controls amino acid incorporation

To determine whether the molar ratios of amino acids in a given mixture could be adjusted to compensate for the differential reaction efficiencies observed in **Figure 1**, we made a series of model libraries of the design VQRI**x**DFLR, as in **Figure 1**. In this case, however, the **x** site was varied to consist of any four amino acids in molar quantities adjusted as in **Methods** to compensate for the differential amino acid incorporation rates derived from **Figure 1**. Groupings were based solely on the propensity of the peptides within each mixture to separate on the HPLC column used for these studies – *i.e.*, the 1st, 6th, 11th, and 16th peptides by order of elution are in one grouping, the 2nd, 6th, 10th, and 14th in another, etc., across five test four-member libraries.

As shown in **Figure 5A** and **Supplementary Figures 109-113**, molar adjustment of different amino acids to compensate for their differential reaction rates resulted in comparable levels of incorporation for all 4-member libraries as assessed by HPLC absorbance at 205 nm in Synthesis Method 2. Similar results were observed with a second series of libraries using ratios as determined in Synthesis Method 1 (**Supplementary Figures 114-119**). To determine whether the levels of amino acid incorporation could also be adjusted up or down by the same methods, we further synthesized a series of three peptides using Method 1 and targeting **x** site incorporation of Arg to Gly ratios of 1:1, 1:2, or 2:1. As shown in **Supplementary Figures 120-123**, incorporation was comparable to the targeted ratio. Adjustment of molar ratios of amino acids may thus approximate targeted levels of amino acid incorporation at a given position.

Flow synthesis produces libraries of expected content up to at least 10^3 to 10^6 members

To determine whether mixture approaches such as those described in **Figure 5A** would scale to more complex libraries of up to 10^3 members, we made split-pool and flow libraries of the form VQR**xxx**FLR on ChemMatrix Rink Amide, where **x** indicates any canonical amino acid except Cys, Ile, or Pro, resulting in a theoretical diversity of 4,913 members. Of note, the choice of resin in this situation was guided by the respective requirements of flow and split-pool systems. That is, flow synthesis requires resins of relatively large particle sizes to avoid clogging the reactor outflow frits under high pressure, while this larger particle size limits the theoretical capacity of an equal amount of resin when synthesizing by split-pool. Under these conditions, the hypothetical maximum number of beads in 100 mg of resin, the amount typically used in our semi-automated flow reactors, with a bead size of 100-200 mesh is approximately 27,000, well above the number of members made in our 3-site (4,913 member) libraries. We then subjected the resultant 3-site libraries to comprehensive characterization by data-independent acquisition (DIA) mass spectrometry. As shown in **Figure 5B**, total identifications were equivalent across both split-pool and flow syntheses as analyzed in Spectronaut 18(26). Consistent with the similar amino acid ratios observed in test peptides synthesized on ChemMatrix versus Tentagel XV resins in **Figure 1**, identifications made from 3-site libraries synthesized on any of five resins from three manufacturers resulted in similar numbers of identifications (**Supplementary Figure 124**). To determine whether it might be possible to scale down the total equivalents used during synthesis, we further made a series of 3-site libraries using neat, 50%, or 25% dilutions of the amino acid mixtures used at diversity sites. As shown in **Supplementary Figure 125**, this resulted in similar identifications under all conditions. It may therefore be possible to conserve

total amino acid usage, particularly for costlier noncanonical amino acids, by cutting down on the total equivalents used without sacrificing substantial library quality.

As the total number of library members approaches 10^5, comprehensive characterization of library content by mass spectrometry becomes increasingly difficult, though independent mass spectrometric assessments outside of our laboratory have suggested that more than 70,000 of 83,521 members in a library of the design VQR**xxxx**LR and all 4,913 members of the smaller VQR**xxx**FLR library are readily detectable on later generation mass spectrometers than those used in **Figure 5B** (data not shown). To further assess the ability of flow synthesis to generate libraries of $\geq 10^6$ members for use in single-pass selections of the type commonly performed in our laboratory, we made a series of four libraries of the design shown in **Figure 5C**, in which a canonical haemagglutinin (HA) 12ca5 binding motif DxxDYA with a fixed theoretical frequency 0.1816% and total size 1.6 million members was frame-shifted within a 10-mer sequence. Such a synthesis can be completed with flow libraries in < 4 hours, while synthesis of these four libraries by a single user with standard split-pool methods in our laboratory would take days to weeks. We then subjected these to selection with immobilized biotinylated 12ca5 antibodies on magnetic beads and characterized the bound members by *de novo* sequencing using PEAKS as previously reported(27). As shown in **Figure 5C**, this procedure resulted in the identification of DxxDYA-containing peptides at a frequency typically >100-fold beyond that expected by chance. Flow synthesis is thus suitable for the generation of libraries of quality sufficient for downstream single-pass selection.

Flow library synthesis extends the practical range of library size by orders of magnitude

In a one-bead, one-compound library, total library size and solubility are limited by the requirement that each bead may consist of only one compound. This results in an upper limit to library diversity enforced by the physical size of the beads. This also results in an upper limit to library solubility enforced by the inability to decrease the molar amount of each library member below an atomistic quantity determined by the size of the beads in use (**Figure 6A**). In a library made by flow synthesis using amino acid mixtures, however, any one bead can hold multiple compounds, where the total number of compounds per bead is limited only by the number of reaction sites available on that bead (**Figure 6B**). Thus, while the library size achievable with a one-bead, one-compound library is around 10^9 members, the library size achievable with an optimal flow-synthesized library is around 10^{19} members.

The comprehensive characterization of 10^{19} sequences is not feasible with current technology, nor have mass spectrometry and other emergent technologies converged on a reliable way to sample single molecule peptide sequences to date. To determine whether flow synthesis is capable of generating library content of a quality sufficient to approximate such extreme diversity, we therefore synthesized a series of 16-mer libraries of the design shown in **Figure 6C**, where each position along the horizontal x-axis may be composed of Trp (yellow) or any canonical amino acid except Cys, Ile, Pro, Tyr or Trp (gray). We then utilized placement of Trp within each library as indicated by absorbance at 280 nm or fluorescence (Ex 280 / Em 360 nm) as a surrogate for expected library diversity. This functions in much the same way that amino acid analysis has been used historically to characterize vertical bulk library content, but adds information about horizontal library content by varying only Trp during C-to-N synthesis. As shown in **Figures 6D-E**, total absorbance or fluorescence of each library is approximately equal to target when varied horizontally at 4, 8, or 16 sites. This suggests that both horizontal and

vertical Trp library content is likely approximately equal to that intended, which in turn implies both horizontal and vertical amino acid content and resultant total library diversity approximating target levels. Of note, despite having a target diversity many orders of magnitude beyond what is possible with current one-bead, one-compound methods, libraries were largely soluble at 10 mg/mL in water with some residual clouding. Dilution to 5 mg/mL, 2 mg/mL, and 1 mg/mL demonstrated a gradual decrement in the small amount of insoluble material remaining after centrifugation, suggestive of a modest subset of poorly soluble members in these otherwise highly soluble library populations (**Supplementary Videos 1-4**).

Discussion

We describe here the development of methodologies for the rapid flow synthesis of peptide and peptidomimetic libraries, which decrease the time required for library synthesis by an order of magnitude while expanding the limits of achievable synthetic library diversity by orders of magnitude. In adapting library generation to a simple flow synthesis platform that has already been implemented in dozens of labs around the world for little startup cost, this work significantly lowers the barrier to the generation of synthetic combinatorial libraries for a variety of applications. In particular, the improved turnaround time offers enhanced flexibility for the iterative adaptation of library content to test specific hypotheses or to refine lead templates for a given purpose. Moreover, the ability to quickly and cheaply produce large numbers of libraries brings with it the opportunity to synergize with emergent artificial intelligence approaches to lead discovery and protein engineering by easily generating ever greater quantities of training data from which machines may learn. Adaptation of similar approaches to our fully automated flow peptide synthesizers(7) is expected to make possible the direct incorporation of diversity regions into small synthetic proteins for activity discovery and refinement.

Despite the advantages of flow library synthesis, barriers remain for the full realization of its potential. Among these, it is largely beyond current technology to precisely define the true upper limits of library diversity under real world conditions in which minor deviations in intended incorporation of a given amino acid at one site will propagate through the remainder of library synthesis to diminish overall diversity. Similarly, variation from the intended representation of any one building block at any one site will inevitably introduce bias into total sequence content in unpredictable ways, and it is impossible to know how local sequence context in an elongating polyamide chain might ultimately impact library content.

In considering these limitations, it is useful to recall that the same limitations might reasonably be assigned to one-bead, one-compound libraries, or to library synthesis generally. By sheer analogy to the synthesis of individual peptides, batch-synthesized split-pool libraries are unlikely to result in greater linear quality than flow libraries, and batch compatibility with larger scaffolds is limited. Even equimolar content, the most conceptually appealing feature of one-bead, one-compound library synthesis, is in part an assumption. Small deviations from an ideal monosized bead radius would be required to propagate to large variations in the molar production of individual compounds from individual beads when that radius deviation is squared (for surface-functionalized beads) or cubed (for porous beads). Thus, while some synthetic challenges are common to both methods, flow library synthesis brings clear advantages in speed, economy, and overall diversity.

Although the majority of the variability between building block incorporation ratios appears explicable by their gyration radii, other factors are likely to affect relative reaction rates. Further optimization of the methods described here may be achievable through the application of machine learning to empirical datasets such as ours in order to fine tune correction factors based on known physicochemical properties of the building blocks under use. An alternative approach that may also be of impactful would be to take advantage of the apparent breakdown in the relationship between gyration radius and relative incorporation at smaller radii by using smaller building blocks. While the gyration radii of Fmoc amino acids tend to be larger by virtue of the Fmoc group itself, for example, other building blocks such as the primary amines utilized in submonomer peptoid synthesis are likely to be smaller and perhaps better suited for incorporation at targeted levels in the absence of specific properties affecting intrinsic reactivity.

In the initial descriptions of mixture-based library synthesis, a major goal was simply to quickly and easily generate many libraries for use in positional scanning approaches to library deconvolution, which require a number of sub-libraries at each diversity position equal to the number of individual building blocks used at that position, easily totaling many dozens of libraries even for a single experimental application. Advances in peptide detection over recent decades, however, have substantially extended the technologies available for more efficient library deconvolution, which in turn opens up a new role for mixture-based methods in harnessing the power of extreme library diversity. With a theoretical synthetic limit on the order of 10^{19} library members, the effective limit to useful library size in flow synthesis is largely a function of the sensitivity of the downstream detection method. Under mass spectrometric conditions in routine use, the limit of detection is estimated at approximately 100 amol(28), yielding $2.5x10^{-5}$ mol / $100x10^{-18}$ mol, or a feasible diversity of 10^{11} members. To make a library of such diversity using one-bead, one-compound approaches similar to those previously used in our laboratory would require several kilograms of beads and similarly prohibitive quantities of individual reagents(27). To make such a library in flow, however, is logistically similar to making a single peptide on 100 mg of resin, neither more time-consuming nor more costly.

Approaches such as those used in **Figure 6** suggest that total library content is likely to approximate theoretical diversity; even missing this target by several orders of magnitude would still yield library diversity several orders of magnitude beyond that currently accessible. Precise quantification of this diversity is challenging with detection technologies currently available, though it may be possible to do so indirectly with mass spectrometry approaches analogous to those used in **Figure 5B** by detecting subsets of barcoded library members synthesized in linear series. In principle, this may provide a means by which to estimate the interval decrement in

diversity as a polyamide chain extends in much the same way that the ability to synthesize entire proteins is, in part, a function of the efficiency of each individual coupling(7). When combined with developing approaches to single-molecule detection such as fluorosequencing and further advances in mass spectrometry, one can envision a wide variety of approaches through which nimble, hyperdiverse chemical libraries might be deployed to navigate chemical space in search of solutions to otherwise intractable challenges in drug development and protein engineering (29–35). Even the expansion in single-pass library diversity may prove less important over time than the ability to rapidly generate overlapping libraries that can probe a given target from multiple angles and then be reassembled into consensus sequences of interest, ultimately covering more combinatorial space than could ever be accessed in any one library.

At the extreme, the combination of speed, economy, and diversity characteristic of flow peptide and peptidomimetic libraries brings us closer to the broader goal of developing synthetic methods that can fully harness the power of selection systems. By facilitating the rapid and iterative adaptation of chemical library content to a given purpose, flow library synthesis opens up new applications for AS-MS as a system for abiotic directed evolution with expansive underlying chemical diversity.

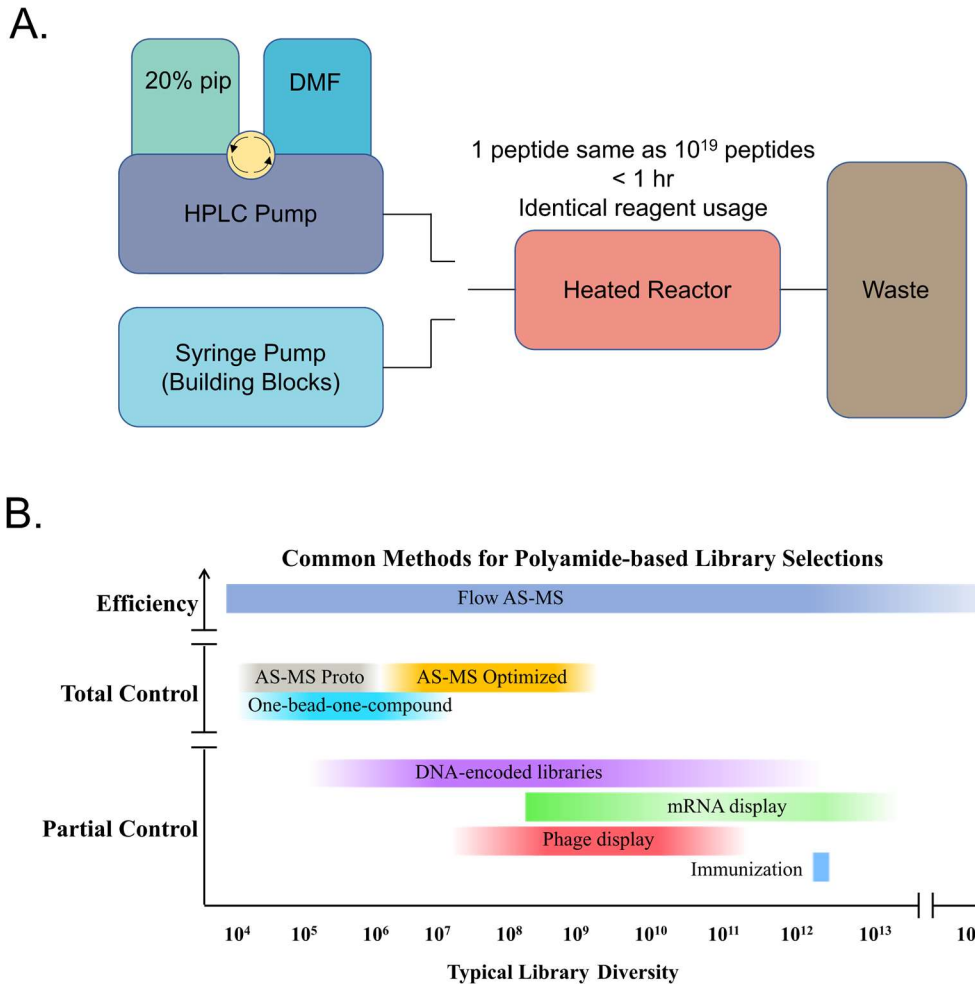


Figure 1 – Flow synthesis combines efficiency and extreme diversity with complete synthetic control for the generation chemical libraries: Total synthetic control over library content in earlier generations of affinity selection mass spectrometry (AS-MS) comes at the cost of speed, reagent use, and library diversity. Flow synthesis addresses each of these limitations, improving speed and reagent usage by an order of magnitude while dramatically expanding the upper limits of achievable library diversity beyond those previously reported. **A.** Schematic of the semi-automated flow synthesis platform utilized, in which conditions and reagent usage for a single peptide are identical to those used for any one library up to a total theoretical diversity of 10^{19} members. **B.** Flow synthesis of polyamide libraries in historical context. Methods such as phage and mRNA display are limited by tagging systems and difficulty accessing noncanonical building blocks, but have higher total achievable diversity within these limitations. AS-MS as previously described eliminates the need for tagging systems and provides complete control over builing block content, but is limited in total diversity relative to molecular biology methods and requires slow, labor-intensive synthesis. Flow synthesis combines the synthetic control of classic AS-MS with diversity at or above that achievable with molecular biology methods in a practical format wherein the synthesis of a hyperdiverse library is largely identical to the synthesis of a single peptide. **B** is adapted from Quartararo *et al. Nature Communications* 2020.

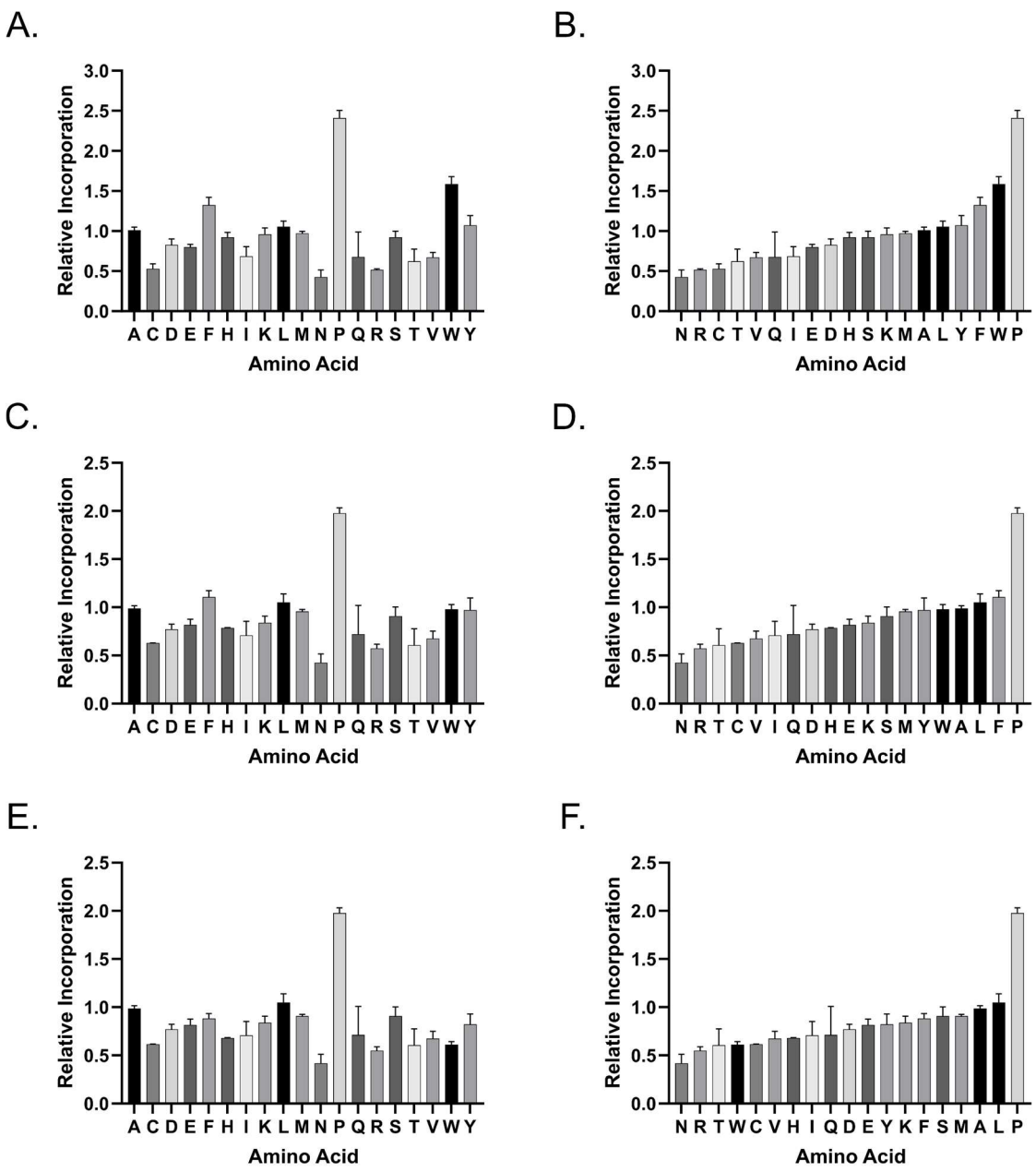


Figure 2 – Relative amino acid incorporation rates in semi-automated flow synthesis vary over a factor of four. Incorporation rates of 19 canonical amino acids relative to Gly, which is set to 1. **A.** Mean incorporation relative to Gly as measured at 214 nm across 3-5 syntheses ordered alphabetically; error bars show the standard deviation. **B.** Data in **A** reordered according to relative incorporation. **C.** Mean incorporation relative to Gly as measured at 205 nm across 2-4 syntheses (all overlapping with 214 nm data) ordered alphabetically. **D.** Data in **C** reordered according to relative incorporation. **E.** Data as in **C** corrected for sidechain contributions to backbone absorbance at 205 nm via a previously reported method ordered alphabetically. **F.** Data in **E** reordered according to relative incorporation.

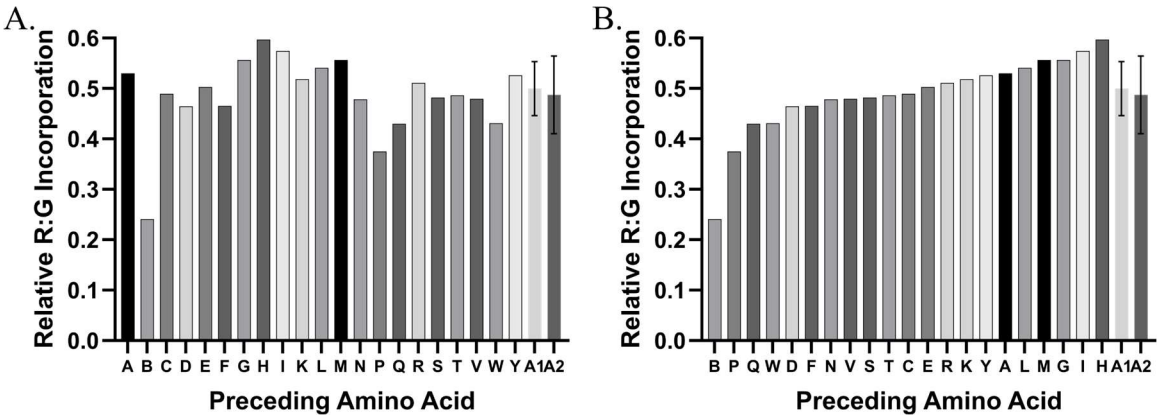
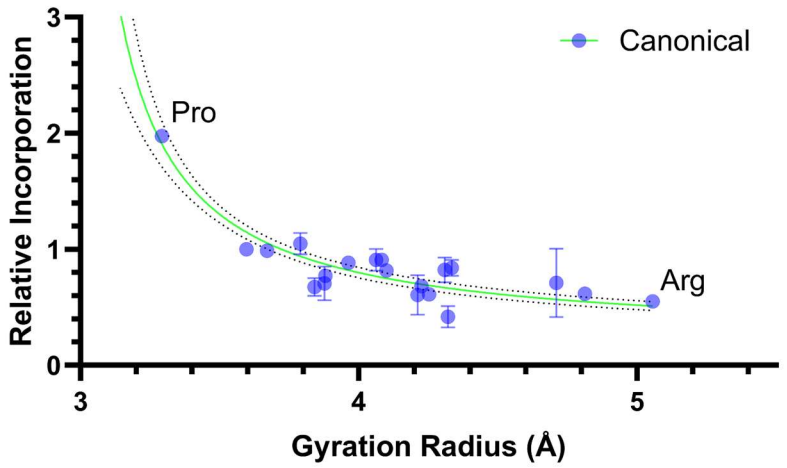
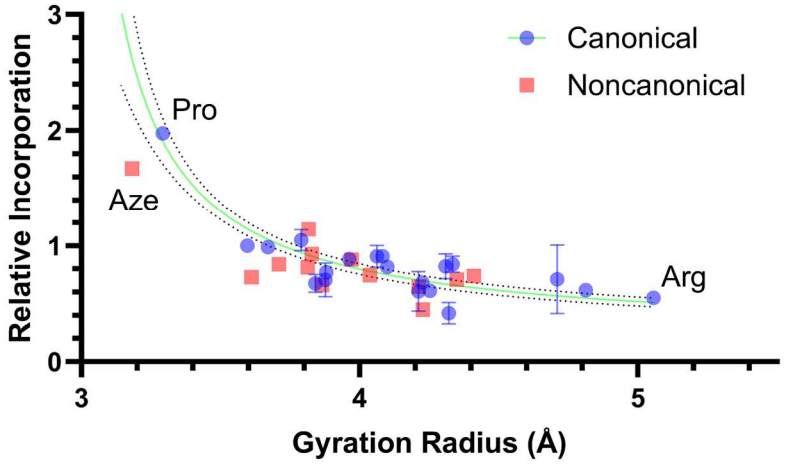


Figure 3 – Most amino acids do not substantially affect downstream relative amino acid coupling rates: Competition peptides of the design VQRI**xy**FLR were made to quantify the effect of placing any of the 20 canonical amino acids or Aib (B) at the **y** site ahead of coupling equal volumes of Arg and Gly at the **x** site. Mixtures of equal volumes of Arg and Gly at a given monomer extension site typically result in incorporation ratios of around 1:2 R:G as quantified by HPLC. **A.** Observed ratio of R:G when preceded by the the indicated amino acid. A1 is shown to indicate the average ratio of R:G and associated standard deviation for all preceding amino acids except the outlier Aib. A2 adds Aib to this mean and standard deviation. **B.** Data as in **A** ordered by relative reaction rate.

A.



B.



C.

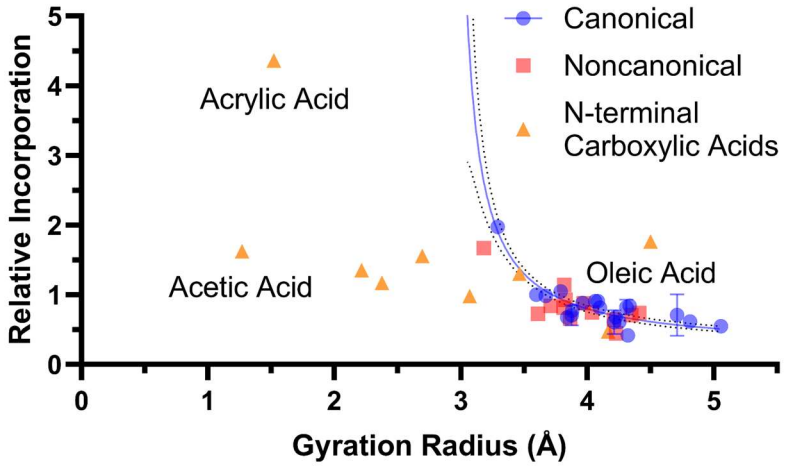
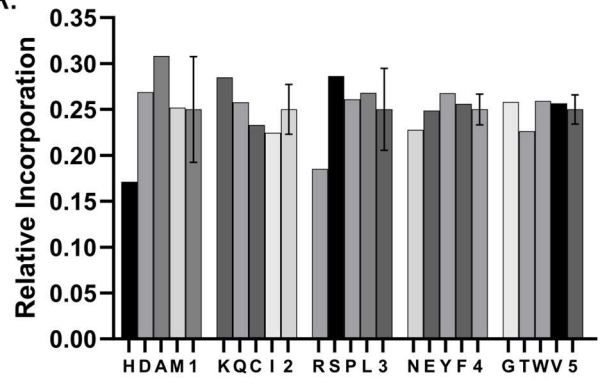
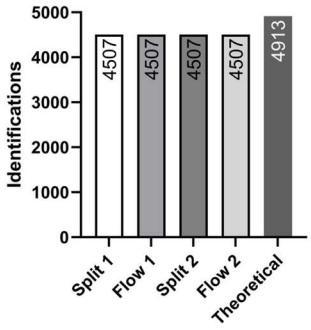


Figure 4 – Reaction rates in flow synthesis correlate with Fmoc building block gyration radius. A. The empirically determined relative incorporation of protected canonical amino acids correlates inversely with calculated Fmoc amino acid gyration radii. Canonical amino acid incorporation ratios are the mean of independent syntheses as in **Figure 1E-F**; error bars not shown are too small to be graphed. Relationship modeled as a hyperbola in Prism with r-squared value 0.70. **B.** Competition peptides in the style of **Figure 1** were synthesized utilizing equal volumes of Gly versus any of 13 noncanonical amino acids at the **x** site in test peptide VQRI**x**DFLR. Canonical signals are quantified by HPLC at 205 nm with correction as per a previously published method for the contributions of distinct sidechains to absorbance. Correction factors from the most closely related canonical amino acid were used when correcting noncanonical amino acids for sidechain contributions to absorbance at 205 nm. **C.** Relative incorporation ratios of carboxylic acids at the peptide N-terminus do not correlate with gyration radii below ~3.3 Å. Canonical and noncanonical amino acids are graphed as in **A-B** with the addition of incorporation ratios for each of 9 carboxylic acids of varying gyration radii relative to Gly. Competition peptides are of the form **x**QRIKDFLR, where the **x** site indicates competition between equal volumes of equimolar Gly versus any of 9 carboxylic acids.

A.



B.



C.

	Design	Identified	Enrichment
Library 1	xyyxxATSNK	0	0-fold
Library 2	NxyyxxATSK	3	384-fold
Library 3	SNxyyxxATK	4	120-fold
Library 4	TSNxyyxxAK	1	120-fold

Figure 5 – Correction to equalize amino acid incorporation results in targeted library content: Four-member libraries of the design VQRI**x**DFLR were made in which the **x** site was varied with any of five combinations of four amino acids per library. Subsequent analysis by HPLC demonstrated targeted amino acid incorporation levels, where the target for each amino acid is 0.25 of total and error bars indicate the standard deviation from this target among the four peptides in any one library. **B.** 4,913-member libraries of the design VQR**xxx**FLR were synthesized by flow or split-pool synthesis with downstream comprehensive analysis of library content by data-independent acquisition mass spectrometry. All conditions resulted in near comprehensive identification to this level of diversity. **C.** Four libraries of shifting design as shown, each with a diversity of 106 members, were synthesized and used for downstream selection for the enrichment of HA-binding peptides. Most resultant libraries were suitable for downstream enrichment of canonical HA-binding moieties by >100-fold over expected identification by chance.

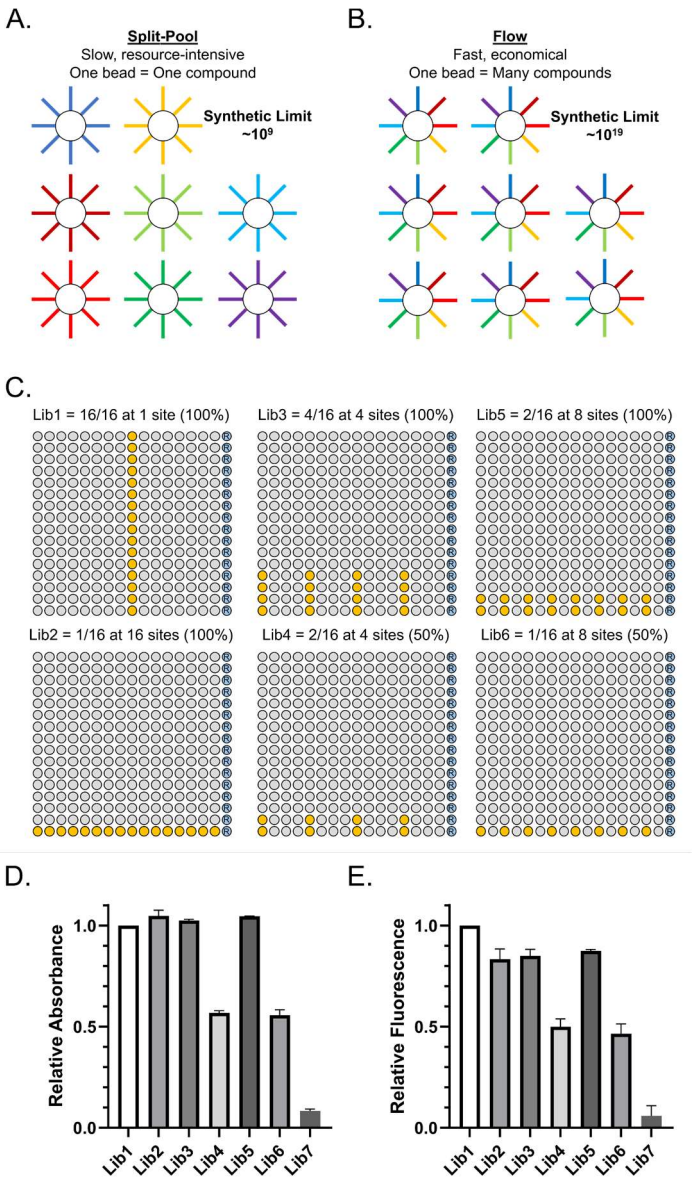


Figure 6 – Flow synthesis produces expected bulk library content up to a target diversity of 10^{19} library members. A. One-bead, one-compound libraries are limited in diversity and solubility by the fact that exactly one compound can be synthesized on any one bead. Physical size of beads and solvent required to dissolve a fixed molar quantity of each compounds as diversity expands makes libraries beyond ~10^9 members impractical. **B.** Flow synthesis can place any number of compounds on a given bead up to the total number of available reaction sites – in our system, a theoretical maximum of 10^{19} members – while keeping this diversity within a low, fixed library mass likely to remain soluble at volumes practical for downstream experimentation. **C.** Libraries of theoretical content $6.6x10^{18}$ (Lib1) to $1.8x10^{19}$ (Lib2) members were synthesized. Library design is shown schematically, where the vertical axis represents the targeted content of each of 16 amino acids (gray = canonical amino acids except C, I, P, Y, or W; yellow = W) and each circle along the horizontal axis represents a distinct library position, including a C-terminal R (blue). **D.** Distribution of W along each of 16 diversity sites in Lib2 results in bulk absorbance at 280 nm similar to that observed when synthesizing a library of fixed W content at a single position in Lib1. Similar results are observed with targeted W distribution across each of Lib3-Lib6. Error bars indicate the SEM of two, independent dissolutions and triplicate measurements of each library at 1 mg/mL in water. **E.** As in **D**, where W fluorescence (ex 280 / em 360) was measured for each library dissolved at 0.1 mg/mL in water. Lib7 is a library in which all 16 positions are gray (no C, I, P, Y, or W).

Acknowledgements

This work has been funded by the National Institute of Allergy and Infectious Diseases (T32 AI007061 and K08 AI166345 to JSA; U19 AI142780 to BLP) and by the Cystic Fibrosis Foundation (ALBIN19F0, ALBIN21Q0, ALBIN22A0-KB to JSA).

Author Contributions

JSA, GYL, CJ, DAV, and WV designed and conducted the experiments, analyzed data, prepared the figures, and wrote the paper. BLP designed experiments, supervised research, and wrote the paper.

Competing Interests Statement

JSA declares no competing interests. BLP is a co-founder and/or member of the scientific advisory board of several companies focusing on the development of protein and peptide therapeutics.

Methods

Common Reagents –

Common reagents were acquired from diverse manufacturers as follows: Dimethylformamide (DMF; VWR, Fisher, MilliporeSigma); hexafluorophosphate azabenzotriazole tetramethyluronium (HATU; P3 BioSystems); *N,N*-diisopropylethylamine (DIEA; Sigma); acetic acid (Sigma, VWR); piperidine (Sigma); dichloromethane (DCM; Sigma); trifluoracetic acid (Sigma); thioanisole (Sigma); water (Milli-Q source); phenol (Sigma); ethane-1,2-dithiol (EDT, Sigma); acetonitrile (Sigma); diethyl ether (Sigma). Canonical amino acids were purchased from Novabiochem; manufacturers for noncanonical amino acids and carboxylic acids described are available upon request. Resin sources included PCAS BioMatrix ChemMatrix Rink Amide with both high (0.49 mmol/g) and low (0.17 mmol/g) loading, ChemMatrix 4-(4-hydroxymethyl-3-methoxyphenoxy)butyric acid (HMPB, 0.44 mmol/g), AAPPTec OctaGel Fmoc-Rink Amide (0.44 mmol/g), and Rapp Polymere Tentagel XV Fmoc-Rink Amide (0.24-0.26 mmol/g).

Library Flow Synthesis –

Library synthesis is based on the methods initially described for rapid semi-automated flow library synthesis(16) with the addition of a capping step after each amino acid coupling. A summary of this approach is given in **Supplementary Figure 126**, and a schematic of the equipment set is provided in **Figure 1A**. The choice of semi-automated flow reflects the desire to develop the initial methods in a simple, accessible flow synthesizer that has already been exported to dozens of labs worldwide and that can be set up for an initial cost of a few thousand dollars, thus maximizing accessibility to the broader research community.

Under the definitive Method 2 used, Fmoc amino acids (AA) were dissolved at 0.4 M AA in a final concentration of 0.38 M HATU in DMF as stock solutions on the day of use. 0.4 M acetic acid (Sigma) in 0.38 M HATU in DMF was used as the capping solution. Under an earlier Method 1, AA were dissolved in 0.4 M HATU in DMF to a final concentration of 0.4 M AA on the day of use, with similar preparation of the capping solution. The definitive Method 2 is susceptible to AA / HATU falling out of solution, which can be remedied with brief heating at 60 °C or by lowering the equivalents of AA / HATU as in **Supplementary Figure 125**. Method 1 does not account for errors associated with the volume of the amino acids themselves but is easier to prepare and generally does not display major deviations from Method 2 in practice (**Supplementary Table 1**). Method 1 was utilized in Synthesis 1 (**Supplementary Figures 1-19**) and for test peptides analyzing the effects of different C-terminal amino acids on relative incorporation at the downstream extension site (**Supplementary Figures 66-86**). All remaining syntheses were completed using Method 2.

A schematic of the synthesis cycle is shown in **Supplementary Figure 126**. 2.5 mL of these AA / HATU or AcOH / HATU stocks were aliquoted to scintillation vials for individual couplings. Approximately 3-4 minutes prior to a given coupling (at the beginning of the prior cycle, or in the case of the first coupling, during a mock cycle to start each synthesis), 0.5 mL DIEA was added to each AA / HATU and AcOH / HATU mix to be coupled. Prior to the start of the cycle associated with each AA to be coupled, the AA / HATU / DIEA and AcOH / HATU / DIEA mixtures were aspirated into 5 mL syringes, residual air and bubbles were removed, and the AA / HATU / DIEA syringe was placed onto a syringe pump.

At the beginning of each cycle, the AA / HATU / DIEA mix was delivered over 30 seconds on a Harvard Apparatus syringe pump. This coupling was then immediately chased with

a second syringe containing AcOH / HATU / DIEA delivered over 30 seconds. Lines were then manually switched over to the HPLC pump (Varian) to wash the resin with DMF at 20 mL / minutes over 1 minute. After washing, the HPLC pump was switched over to a 20% piperidine in DMF deprotection solution for 20 seconds followed by an additional DMF wash for the final 1 minute of each cycle. All reactions were carried out at 70 °C. Steps in 3.5 minute cycles are further summarized as follows:

Coupling Cycle 0

- Add DIEA to AA_1 / HATU and $AcOH_1$ / HATU vials
- 3:30-0:00 – As in Coupling Cycle 1, but replace syringe pump steps with DMF washing
- Final 20-30 seconds of Coupling Cycle 0 – Add DIEA to AA_2 and $AcOH_2$ / HATU / DIEA vials

Coupling Cycle 1

- 3:30-3:00 – Couple AA_1 / HATU / DIEA via syringe pump
- 3:00-2:30 – Couple $AcOH_1$ / HATU / DIEA via syringe pump
- 2:30-2:20 – Manually switch lines from syringe pump to HPLC pump
- 2:20-1:20 – Wash with DMF at 20 mL / minute; syringes for the next Coupling Cycle are generally prepared during this step if not already assembled
- 1:20-1:00 – Deprotect with 20% piperidine in DMF (requires flipping a selector between DMF and deprotection solution on the HPLC apparatus)
- 1:00-0:00 – Wash with DMF at 20 mL / minute; add DIEA to AA_3 and $AcOH_3$ / HATU / DIEA vials in the final 20-30 seconds of this step

When defining incorporation rates of a given AA relative to that of Gly in single comparisons as in **Figure 1**, equal volumes of Gly / HATU versus AA_1 / HATU were added to the scintillation vial to be used at the central monomer site in the test peptide H-VQRI**x**DFLR-NH_2 in the amount of 1.25 mL apiece for a final volume of 2.5 mL. Subsequent coupling steps were identical to those used when coupling a single, defined AA.

When adjusting molar ratios of amino acids in a complex monomer site mixture to achieve a desired molar incorporation of a given amino acid – *e.g.*, equimolar for each building block as in **Figures 5-6** – the reciprocal of each of the defined reaction rates relative to Gly was taken. For example, if Arg has an incorporation rate of 0.5 relative to Gly, Pro has an incorporation rate of 2 relative to Gly, and Gly has an incorporation rate defined as 1, then the reciprocals are: Arg = 2, Pro = 0.5, and Gly = 1. The sum of these reciprocals (2 + 1 + 0.5 = 3.5) is then used as the denominator to define the percentage of each amino acid to be included in the mixture, and this percentage is multiplied by 2.5 to achieve a final AA_{mix} / HATU volume of 2.5 mL. For example:

Arg: (2 / 3.5) * 2.5 = 1.429 mL

Pro: (0.5 / 3.5) * 2.5 = 0.357 mL

Gly: (1 / 3.5) * 2.5 = 0.714 mL

Total = 2.5 mL

Where multiple randomized monomer sites are to be used, a master mix of greater volume is made from which 2.5 mL aliquots are distributed to each scintillation vial. In all cases, once a given mixture is in a scintillation vial, all subsequent steps are identical to those used to couple a single, defined AA.

After flow synthesis, each resin bed was transferred from the semi-automated flow reactor to an individual 5 mL fritted disposable reactor. Resin was then swollen and washed with DCM prior to drying under vacuum.

Resins used throughout were either ChemMatrix Rink Amide (Synthesis 1 and 2 in **Supplementary Table 1**) or Tentagel XV Rink Amide (Syntheses 3-5 in **Supplementary Table 1**). 3-site libraries in **Figure 5B** were synthesized in ChemMatrix Rink Amide; all remaining libraries were synthesized on Tentagel XV Rink Amide unless otherwise specified. 3-site libraries in **Figure 5** and the hyperdiverse libraries in **Figure 6** were synthesized using the adjustments empirically determined from Synthesis 2.

Library Split-Pool Synthesis –

Split-pool library synthesis for libraries in **Figure 5** of the design VQR**xxx**FLR where **x** is any of 17 canonical amino acids (no Cys, Ile, or Pro) was completed using ChemMatrix Rink Amide. For the split-pool synthesis of diversity sites, 200 mg of resin was resuspended with extensive pipetting to discourage bead aggregation in 40-50 mL DMF, with subsequent distribution of equal volumes of bead solution to each of 17 separate 3 mL fritted reaction vessels dedicated to each of the above 17 canonical amino acids. Residual bead solution was then resuspended in additional DMF and aliquoted to the 17 reaction vessels twice more prior to discarding the small volume of remaining solution. To each of these 17 reaction vessels was then added 1.2 mL of individual amino acid coupling solution as typically prepared for semi-automated flow synthesis. Following incubation at room temperature for 30 minutes, amino acid was drained by vacuum filtration followed by the addition of 1.2 mL of capping solution as typically prepared for semi-automated flow synthesis. After an additional 5 minutes incubation,

each resin was washed thrice with DMF prior to repooling of resin back into a single 50 mL reaction vessel and three additional washes. Resin was then deprotected with 15 mL 20% piperidine twice for 5 minutes each, followed by three additional washes with DMF. Resin was then split back to each of the 17 dedicated reaction vessels as above before coupling of a second diversity site. Prior to the coupling of the third diversity site, resuspended resin was split in two, with half carried forward for the addition of a third diversity site using halved reaction volumes.

Cleavage –

Cleavage throughout these studies was carried out with Reagent K (82.5% TFA, 5% water, 5% phenol, 5% thioanisole, 2.5% EDT) for 2 hours at room temperature with a handful of exceptions among the N-terminal carboxylic acid competition peptides. Subsequent workup differed according to the species under evaluation. For competition peptides like those in **Figure 2**, cleavage mixtures were precipitated with a 10:1 volume of chilled diethyl ether followed by spinning at 4000 rpm at 4 °C for at least 2 minutes and two subsequent diethyl ether washes under the same conditions. After evaporation of any residual ether, peptide mixtures were resuspended in 50% water : 50% acetonitrile : 0.1% TFA with brief heating at 60 °C where needed to promote complete dissolution. Mixtures were then flash frozen in liquid nitrogen and lyophilized. These crude competition peptides were analyzed directly by HPLC and LCMS.

Libraries consisting of thousands or more members were cleaved and precipitated as above. Following precipitation, rather than proceeding directly to lyophilization, dried products proceeded through an additional cleanup step prior to downstream mass spectrometry, selections, or other characterization. In brief, pellets were dissolved in 0.5 mL DMF per 50 mL tube, which was then resuspended in at least 50 mL total water + 1% DMF + 0.1% TFA per library.

Resuspended libraries were then subjected to solid phase extraction (SPE) under vacuum using Supelclean LC-18 SPE 1 g bed 6 mL tubes (Millipore) fitted with large volume SPE extension reservoirs (Sigma). In brief, the SPE beds were activated with 5 mL acetonitrile + 0.1% TFA and then equilibrated with 5 mL water + 0.1% TFA. Dissolved libraries were then bound, followed by three washes with 5 mL each 95% water + 5% acetonitrile + 0.1% TFA. Final elution of libraries was carried out with two applications of 3 mL apiece of 50% water + 50% acetonitrile + 0.1% TFA followed by two application of 3 mL apiece of 30% water + 70% acetonitrile + 0.1% TFA. The final eluted volume of 12 mL was then flash frozen in liquid nitrogen and lyophilized for downstream characterization.

High Performance Liquid Chromatography (HPLC) –

HPLC Method A: Column – Kinetex Evo C18, 5 µm, 100 Å, 4.6 x 250 mm; flow rate 1 mL min^{-1}; solvent system – A = Water + 0.1% TFA, B = Acetonitrile + 0.1% TFA; gradient – 3 min hold 1% B, 1-61% B gradient over 60 minutes, 3 min hold 61% B, 10-min post-run 1% B; instrument – Agilent 1200 series system with UV detection including 205 and 214 nm.

HPLC Method B: Column – Kinetex Evo C18, 5 µm, 100 Å, 4.6 x 250 mm; flow rate 1 mL min^{-1}; solvent system – A = Water with 10 mM NH_4OAc, B = Acetonitrile; gradient – 3 min hold 1% B, 1-61% B gradient over 60 minutes, 3 min hold 61% B, 10-min post-run 1% B; instrument – Agilent 1200 series system with UV detection including 205 and 214 nm.

Kinetex Evo C18 1-61_30 min Water + 0.1% TFA, B = Acetonitrile + 0.1% TFA

Kinetex Evo C18 1-61_30 min Water with 10 mM NH_4OAc, B = Acetonitrile

Liquid Chromatography – Mass Spectrometry (LCMS) –

LCMS Method A: Column – Aeris Widepore C4, 3.6 µm, 200 Å, 150 x 2.1 mm; flow rate 0.3 mL min^{-1}; solvent system – A = Water + 0.1% FA, B = Acetonitrile + 0.1% FA; gradient – 1-61% B over 2-12 minutes, MS on from 4-12 min; instrument – Agilent 6550-1, 1290 Infinity HPLC system with iFunnel QTOF MS run in position ionization mode with m/z range 100-1700.

LCMS Method B: Column – Zorbax 300SB-C3, 5 µm, 300 Å, 150 x 2.1 mm; flow rate 0.5 mL min^{-1}; solvent system – A = Water + 0.1% FA, B = Acetonitrile + 0.1% FA; gradient – 1-61% B over 2-12 minutes, MS on from 4-12 min; instrument – Agilent 6550-2, 1290 Infinity HPLC system with iFunnel QTOF MS run in position ionization mode with m/z range 100-3000.

Zorbax C18_1-61_8 min 6550#2

Zorbax C18_1-61_40 min 6550#2

Zorbax C18_1-61_15 min 6550#2

High resolution mass spectrometry characterization of library content –

For comparison of split-pool and flow-synthesized libraries of the design VQR**xxx**FLR, instrumentation included an EASY-nLC 1200 (Thermo) nano-liquid chromatography system coupled to an Orbitrap Eclipse mass spectrometer (Thermo). Libraries were injected at approximately 380 ng per injection including triplicate DIA acquisitions of each library and one DDA acquisition of each library. The gradient in both cases was as follows with a 300 nl/minute flow rate with Solvent A = Water / 0.1% formic acid and Solvent B = 80% Acetonitrile / 20%

Water / 0.1% formic acid on a PepMap C18 2 µm, 100 Å, 50 µm x 15 cm column preceded by a PepMap C18 3 µm, 100 Å, 75 µm x 2 cm nanoViper trap column:

0-5 minutes 1-6% B

5-35 minutes 6-21% B

35-55 minutes 21-41% B

55-60 minutes 41-61% B

DIA data were acquired with a method based on the template Xcalibur DIA method with MS1 resolution set to 60,000 and MS2 resolution set to 15,000. This was modified to cover a 300-700 m/z range using 12 m/z overlapping windows. HCD collision energy was set to 35%. DDA data were acquired with method based on the template Xcalibur DDA method for < 500 ng injection with MS1 resolution set to 120,000 and MS2 resolution set to 30,000 on a 3 second cycle time. This was modified to cover a 300-700 m/z range. HCD collision energy was set to 35%.

Analysis of library content was completed with Spectronaut 18 using directDIA on triplicate injections of each library. These triplicate injections as well as single injections of each library acquired in DDA mode were also used as Library Extension runs. Search settings were set to factory defaults with the following modifications:

Unspecific search, no enzyme, peptide length set to 9 amino acids

Fixed modification C-terminal amidation, variable modification Met oxidation.

Similar methods were used on an Orbitrap Fusion Lumos for the identification of members isolated from 3-site libraries synthesized with fewer amino acid equivalents or on different resin types as in **Supplementary Figures 124-125**, the primary difference being the use of a PepMap 2 µm, 100 Å, 75 µm x 25 cm column and 8 m/z overlapping windows. Amino acid equivalent comparison libraries are duplicate injections of syntheses using neat, 50%, or 25% amino acid

equivalents, where the file for one of two injections of the 25% equivalents synthesis was later found to be corrupt. Primary analysis was completed with Spectronaut 18 directDIA as above; an additional two injections of a mixture of all libraries acquired in DDA mode as well as the Eclipse injections characterizing split-pool versus flow synthesis as shown in **Figure 5B** were used as Library Extension runs. Resin comparison searches were separated into resins producing a C-terminal amide and ChemMatrix HMPB, which produces a C-terminal carboxylic acid to avoid erroneous assignment of the peptide C-terminus in each category if the C-terminus were to be assigned as a variable modification. An additional mixture of all libraries in each group was injected twice with DDA acquisition for use as a Library Extension run along with the **Figure 5B** injections as described above.

Library Selections –

Magnetic bead preparation

100 µL per replicate of MyOne Streptavidin T1 Dynabeads (10 mg/mL, Thermo) were transferred to centrifuge tubes applied to a magnetic separation rack (New England Biolabs) and washed three times with blocking buffer (phosphate-buffered saline (PBS), 10% fetal bovine serum (FBS), and 0.02% Tween). Washed beads were then resuspended in blocking buffer to be the same starting volume (100 µL per replicate).

Library preparation

Synthesized libraries were dissolved in PBS and stored as 4 nM stock solutions (average molecular weight of 1131.9 g/mol). Prior selections, libraries were diluted to 100 pM/member and applied to washed magnetic beads to negatively select against nonspecific binders. This

mixtures consisted of 100 µL beads, 25 µL library stock, and 875 µL blocking buffer. After mixing, the suspension was spun down at max speed for 10 minutes at 4°C. Supernatant was isolated as the final library sample to go through affinity selection with the target protein.

Magnetic bead capture of control binders

Automated affinity selection of peptide binders was carried out on a ThermoFisher KingFisher Duo Prime with methods for selection modified using the ThermoFisher BindIt 4.1 Software. In brief, 10 mg/mL magnetic beads in blocking buffer were washed three times followed by capture and transfer to wells containing biotinylated 12ca5 antibody for immobilization onto the streptavidin-functionalized bead surface. Excess unbound protein was then washed away prior to three additional washes in blocking buffer. Beads were then transferred to triplicate wells containing libraries dissolved at 0.1 nM and incubated in the presence of library for 1 hour with slow mixing at 10 °C. Beads were then washed a total of six times in PBS for 2 minutes per wash to remove low affinity binders and excess Tween. Beads were then transferred to a denaturing solution consisting of 6 M Guanidinium•HCl followed by repeat immobilization of beads and transfer to a second denaturing solution of the same composition. Eluted peptides were then pooled for desalting prior to mass spectrometry.

De-salting of eluates

Eluates were desalted via solid-phase extraction. CDS Empore™ C18 Extraction Disks were punched out using 18-gauge blunt tip needles (VWR Sterile Blunt Tip Needles NB18212) and placed into 200 µL pipette tips. The packed tips were inserted through centrifuge tube lids. The tips were first wetted with 60 µL of 80% acetonitrile/20% water (0.1% TFA) and

centrifuged at 500 rcf in 2-minute intervals until the level of the solvent was flush with the solid-phase material. This was repeated with 60 µL of 2% acetonitrile/98% water (0.1% TFA) for equilibration. Pooled eluates from the selection protocol (200 µL total per sample) were loaded into the tips and centrifuged in 4-minute intervals until spun down akin to the washing steps. This was followed by a wash with 60 µL of 3% acetonitrile/97% water (0.1% TFA), with centrifugation in 2-minute intervals until complete. The remaining peptides were eluted with 75 µL of 70% acetonitrile/30% water (0.1% TFA) and centrifuged at 500 rcf for 5-10 minutes. The final eluate was lyophilized, resuspended in 12 µL of water (0.1% formic acid), and 4 µL was submitted for nLC-MS/MS analysis.

nLC-MS/MS analysis

Analysis was performed on an EASY-nLC 1200 (Thermo) nano-liquid chromatography handling system connected to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo). Samples were run on a PepMap RSLC C18 column (2 µm particle size, 75 cm × 50 µm ID; Thermo Fisher Scientific). A nanoViper Trap Column (C18, 3 µm particle size, 100 Å pore size, 20 mm × 75 µm ID; Thermo Fisher Scientific) was used for desalting. The standard nano-LC method was run at 40 °C and a flow rate of 300 nL/min with Solvents A / B as defined above and a gradient of 1-45% B over 90 minutes. MS acquisition was DDA with MS1 set to Orbitrap detection with a resolution of 120,000 covering a range of 200-1400 m/z. Two fragmentation modes—higher-energy collisional dissociation (HCD), and electron-transfer/higher-energy collisional dissociation (EThcD)—were used for acquisition of secondary MS spectra, in all cases utilizing Orbitrap detection at a resolution of 30,000. Precursors with charge states 2-5 were subjected to HCD fragmentation while 3 sets of EThcD parameters were used for 3 charge state categories: 4-6, 3, and 2. For both fragmentation modes, detection was performed in the

Orbitrap. HCD collision energy was set to 25%. For charge states 4-6, EThcD collision energy was set to 25%; for charge states 2 or 3, collision energy was set to 30%.

Binders were identified using PEAKS 8.5 software for *de novo* sequencing with fixed C-terminal amidation and variable Met oxidation. Searches were completed with two subsets, where Subset 1 included {A, R, N, D, Q, E, G, H, M, L, K, F, S, T, W, Y, V} and Subset 2 included {A, R, N, Q, E, G, H, M, L, K, F, S, T, W, Y, V}. For library 1, for example, the search query was set to: '12211ATSNKU' where U is the C-terminal amidation variable.

Tryptophan Monitoring of Library Content –

Libraries were synthesized as described above based on the ratio adjustments empirically determined in Synthesis 2 (**Supplementary Table 1**). For each library, baseline content at any given monomer site consisted of any canonical amino acid except Cys, Ile, Pro, Tyr, or Trp. Trp was then introduced at the differing levels and positions indicated in **Figure 5** to act as a marker for targeted vertical and horizontal amino acid incorporation in hyperdiverse libraries with a theoretical diversity of 10^{18}-10^{19} members. Control libraries included one in which a single monomer position was set to Trp (100%) and another in which no Trp was utilized. After cleavage, SPE, and lyophilization, library powder was dissolved in LCMS grade water to a concentration of 1 mg/mL (see also **Supplementary Videos 1-4**). Absorbance at 280 nm was then measured for each library in triplicate in black, transparent bottom 384-well plates in a Tecan Spark multimodal plate reader. 0.1 mg/mL dilutions of the 1 mg/mL stock were similarly used for the evaluation of Tryptophan filter-based fluorescence with excitation at 280 nm and emission at 360 nm in Greiner black opaque bottom plates. Data shown in **Figure 5** are the mean and SEM of two independent library dissolutions and plate reader measurements of triplicate

wells separated by five months with intervening storage of lyophilized powder at -20 ºC. Background water signal was subtracted from all library signals, with subsequent use of the absolute values to account for negative fluorescence signals associated with the control library lacking Trp after subtraction of water background.

References

1. Newton MS, Cabezas-Perusse Y, Tong CL, Seelig B. Advantages of mRNA display. ACS Synth Biol. 2020 Feb 2;9(2):181.

2. Jaroszewicz W, Morcinek-Orłowska J, Pierzynowska K, Gaffke L, Węgrzyn G. Phage display and other peptide display technologies. FEMS Microbiol Rev. 2022 Mar 3;46(2):1–25.

3. Goto Y, Suga H. The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. Acc Chem Res. 2021 Sep 21;54(18):3604–17.

4. Koh LQ, Lim YW, Gates ZP. Affinity Selection from Synthetic Peptide Libraries Enabled by De Novo MS/MS Sequencing. Int J Pept Res Ther. 2022 Mar 1;28(2):1–14.

5. Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierskit WM, Knappt RJ. A new type of synthetic peptide library for identifying ligand-binding activity. Nature 1991 354:6348. 1991 Nov 7;354(6348):82–4.

6. For chemists, the AI revolution has yet to happen. Nature. 2023 May 18;617(7961):438.

7. Hartrampf N, Saebi A, Poskus M, Gates ZP, Callahan AJ, Cowfer AE, et al. Synthesis of proteins by automated flow chemistry. Science (1979) [Internet]. 2020 May 29 [cited 2020 May 31];368(6494):980–7. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.abb2491

8. Quartararo AJ, Gates ZP, Somsen BA, Hartrampf N, Ye X, Shimada A, et al. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. Nat Commun. 2020 Dec 1;11(1):1–11.

9. Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT, Cuervo JH. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. Nature 1991 354:6348. 1991;354(6348):84–6.

10. Ostresh JM, Winkle JH, Hamashin VT, Houghten RA. Peptide libraries: determination of relative reaction rates of protected amino acids in competitive couplings. Biopolymers. 1994;34(12):1681–9.

11. Acharya AN, Ostresh JM, Houghten RA. Determination of isokinetic ratios necessary for equimolar incorporation of carboxylic acids in the solid-phase synthesis of mixture-based combinatorial libraries. Biopolymers. 2002 Oct 5;65(1):32–9.

12. Ostresh JM, Schoner CC, Giulianotti MA, Kurth MJ, Dörner B, Houghten RA. Combinatorial libraries: Equimolar incorporation of benzaldehyde mixtures in reductive alkylation reactions.

13. Warrass R, Jung G, Wiesmüller KH. Combinatorial Oligocarbamate Collections: Synthesis by the Premix Method and Quality Control by HPLC-MS. QSAR Comb Sci. 2003 Nov 1;22(8):873–81.

14. Lee Herman W, Tarr G, Kates SA. Optimization of the synthesis of peptide combinatorial libraries using a one-pot method. Mol Divers. 1997;2(3):147–55.

15. Ivanetich KM, Santi D V. [15] Preparation of equimolar mixtures of peptides by adjustment of activated amino acid concentrations. Methods Enzymol. 1996 Jan 1;267:247–60.

16. Simon MD, Heider PL, Adamo A, Vinogradov AA, Mong SK, Li X, et al. Rapid Flow-Based Peptide Synthesis. ChemBioChem. 2014 Mar 21;15(5):713–20.

17. Mijalis AJ, Thomas DA, Simon MD, Adamo A, Beaumont R, Jensen KF, et al. A fully automated flow-based approach for accelerated peptide synthesis. Nat Chem Biol. 2017 Feb 28;13(5):464–6.

18. Hartrampf N, Saebi A, Poskus M, Gates ZP, Callahan AJ, Cowfer AE, et al. Synthesis of proteins by automated flow chemistry. Science (1979). 2020 May 29;368(6494):980–7.

19. Callahan AJ, Gandhesiri S, Travaline TL, Salazar LL, Hanna S, Lee YC, et al. Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Phage Display. 2023 Feb 3;

20. Saebi A, Brown JS, Marando VM, Hartrampf N, Chumbler NM, Hanna S, et al. Rapid Single-Shot Synthesis of the 214 Amino Acid-Long N-Terminal Domain of Pyocin S2. ACS Chem Biol. 2023 Mar 17;18(3):518–27.

21. Houghten RA, Pinilla C, Giulianotti MA, Appela JR, Dooley CT, Nefzi A, et al. Strategies for the use of mixture-based synthetic combinatorial libraries: Scaffold ranking, direct testing in vivo, and enhanced deconvolution by computational methods. J Comb Chem. 2008 Jan;10(1):3–19.

22. Anthis NJ, Clore GM. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. Protein Sci. 2013 Jun;22(6):851.

23. Boutin JA, Gesson I, Henlin JM, Bertin S, Lambert PH, Volland JP, et al. Limitations of the coupling of amino acid mixtures for the preparation of equimolar peptide libraries. Mol Divers. 1997;3(1):43–60.

24. Pinzón-López S, Kraume M, Danglad-Flores J, Seeberger PH. Reaction Chemistry & Engineering REVIEW Transport phenomena in solid phase synthesis supported by cross-linked polymer beads. Cite this: React Chem Eng. 2023;8:2951.

25. Pedretti A, Mazzolari A, Gervasoni S, Fumagalli L, Vistoli G. The VEGA suite of programs: an versatile platform for cheminformatics and drug design projects. Bioinformatics [Internet]. 2021 May 23 [cited 2024 Aug 26];37(8):1174–5. Available from: https://dx.doi.org/10.1093/bioinformatics/btaa774

26. Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng LY, Messner S, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics [Internet]. 2015 May 1 [cited 2024 Aug 26];14(5):1400–10. Available from: https://pubmed.ncbi.nlm.nih.gov/25724911/

27. Quartararo AJ, Gates ZP, Somsen BA, Hartrampf N, Ye X, Shimada A, et al. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. Nat Commun [Internet]. 2020 Dec 1 [cited 2020 Jul 11];11(1):1–11. Available from: https://doi.org/10.1038/s41467-020-16920-3

28. Krasny L, Huang PH. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. Mol Omics. 2021 Feb 22;17(1):29–42.

29. Swaminathan J, Boulgakov AA, Hernandez ET, Bardo AM, Bachman JL, Marotta J, et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nature Biotechnology 2018 36:11 [Internet]. 2018 Oct 22 [cited 2024 Aug 26];36(11):1076–82. Available from: https://www.nature.com/articles/nbt.4278

30. Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, et al. Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. J Proteome Res [Internet]. 2015 Dec 4 [cited 2024 Aug 26];14(12):5378–87. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jproteome.5b00932

31. Stewart HI, Grinfeld D, Giannakopulos A, Petzoldt J, Shanley T, Garland M, et al. Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. Anal Chem [Internet]. 2023 Oct 24 [cited 2024 Aug 26];95(42):15656–64. Available from: https://pubs.acs.org/doi/full/10.1021/acs.analchem.3c02856

32. Marx V. Proteomics sets up single-cell and single-molecule solutions. Nature Methods 2023 20:3 [Internet]. 2023 Mar 10 [cited 2023 Aug 15];20(3):350–4. Available from: https://www.nature.com/articles/s41592-023-01781-7

33. Beattie M, Jones OAH. Rate of Advancement of Detection Limits in Mass Spectrometry: Is there a Moore's Law of Mass Spec? Mass Spectrom (Tokyo) [Internet]. 2023 [cited 2023 Nov 16];12(1). Available from: https://pubmed.ncbi.nlm.nih.gov/37250598/

34. MacCoss MJ, Alfaro JA, Faivre DA, Wu CC, Wanunu M, Slavov N. Sampling the proteome by emerging single-molecule and mass spectrometry methods. Nature Methods

2023 20:3 [Internet]. 2023 Mar 10 [cited 2023 Nov 16];20(3):339–46. Available from: https://www.nature.com/articles/s41592-023-01802-5

35. Alfaro JA, Bohländer P, Dai M, Filius M, Howard CJ, van Kooten XF, et al. The emerging landscape of single-molecule protein sequencing technologies. Nature Methods 2021 18:6 [Internet]. 2021 Jun 7 [cited 2023 Nov 16];18(6):604–17. Available from: https://www.nature.com/articles/s41592-021-01143-1