pubs.acs.org/acschemicalbiology

Articles

Generation of a Large Peptide Phage Display Library by Self-Ligation of Whole-Plasmid PCR Product

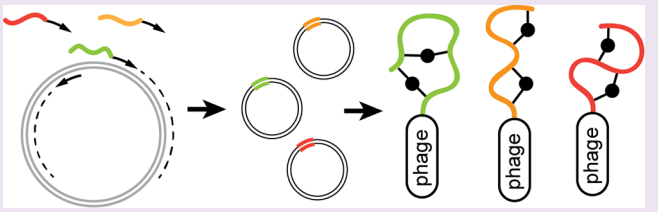
Xu-Dong Kong, Vanessa Carle, Cristina Díaz-Perlas, Kaycie Butler, and Christian Heinis*

Cite This: https://dx.doi.org/10.1021/acschembio.0c00497

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The success of phage display, used for developing target-specific binders based on peptides and proteins, depends on the size and diversity of the library screened, but generating large libraries of phage-encoded polypeptides remains challenging. New peptide phage display libraries developed in recent years rarely contained more than 1 billion clones, which appears to have become the upper size limit for libraries generated with reasonable effort. Here, we established a strategy based on whole-plasmid PCR and self-ligation to clone a library with more than 2×10^{10} members. The enormous library size could be obtained through amplifying the entire vector DNA by PCR, which omitted the step of vector isolation from bacterial cells, and through appending DNA coding for the peptide library via a PCR primer, which enabled efficient DNA circularization by end-ligation to facilitate the difficult step of vector-insertion of DNA fragments. Panning the peptide repertoires against a target yielded high-affinity ligands and validated the quality of the library and thus the new library cloning strategy. This simple and efficient strategy places larger libraries within reach for nonspecialist researchers to hopefully expand the possible targets of phage display applications.



INTRODUCTION

Peptides and proteins have become important therapeutics, with many developed with the help of in vitro evolution methods such as phage display.[1] Phage display is a powerful technique that allows for screening large libraries of these biomolecules for binding to targets of interest.[1,2] Although the chances of isolating good binders increase with a greater number and diversity of library molecules that can be sampled,[3] phage display library sizes and designs have stagnated in the past decade. Most phage display libraries used today contain between 1 million and 1 billion different polypeptides, and libraries over this size are generally prohibitively difficult to make due to technical hurdles, including the generation of large quantities of phage vector encoding the polypeptides and the transformation of the DNA into individual bacterial cells. Even for specialized laboratories and companies with longstanding experience in DNA library construction, cloning a phage display library comprising more than 1 billion different peptides is a major task that typically takes multiple weeks.

Alternative display methods such as ribosome display and mRNA display omit the step of transforming library DNA into cells and allow for screening larger libraries, but important applications such as chemical modification of peptides in organic-aqueous solutions,[4] or performing selections under stringent conditions such as high temperature or proteolytic pressure,[5,6] are not possible. The demand for new peptide phage display libraries is thus unbroken. Alone in the past year, several large libraries were reported including the most recent ones in which phages display linear X_{12} peptides (X = any random amino acid; library size: 1.2×10^8 clones),[7] Y*X_6C cyclic peptides (library size: 10^8 clones; Y* = unnatural amino acid),[8] and CX_6K* cyclic peptides (library size: 10^9 clones; K* = unnatural amino acid).[9] From these libraries with sizes ranging from 100 to 1000 million clones, peptides with desired qualities were isolated, but it is likely that larger libraries would yield peptides with even better properties.

Current techniques used for cloning phage display libraries are diverse and can be broadly divided into two groups. The first one includes strategies based on cassette mutagenesis in which double-stranded library DNA is ligated into linearized plasmid.[10] The second group are Kunkel mutagenesis-type approaches in which degenerated oligonucleotides are annealed to single-stranded DNA followed by the synthesis of a complementary strand.[11−14] Strategies based on cassette mutagenesis are frequently used for cloning phage display libraries, though they have multiple technical challenges that need to be overcome to be efficient. A first difficulty is the production of sufficient vector that needs to be isolated from bacterial cells and linearized by restriction endonucleases. This process is often plagued by the presence of variable forms of plasmid DNA, such as supercoiled DNA, which is occasionally

Received: June 19, 2020
Accepted: October 19, 2020

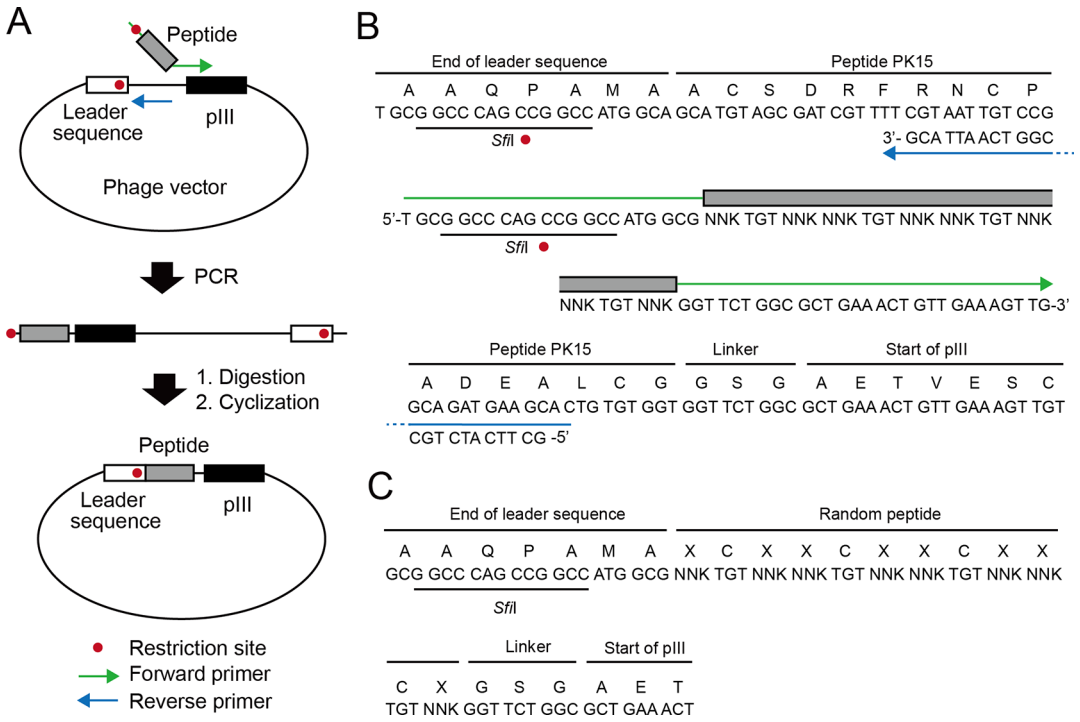


Figure 1. Whole-plasmid PCR phage display peptide library cloning strategy. (A) Experimental steps of cloning strategy. The forward primer anneals at the N-terminus of the pIII gene and appends a DNA sequence coding for a library peptide and an *Sfi*I restriction site. The reverse primer anneals downstream of the leader sequence that contains an *Sfi*I restriction site. The whole-plasmid PCR product contains an *Sfi*I restriction site at each end, allowing for self-ligation after *Sfi*I digestion. (B) Region of phage vector fd-tet-PK15 that encodes the N-terminus of pIII and the leader sequence. The sequences of the primers XD12_16 and XD_reverse and their annealing regions are shown. (C) Sequence of phage vector from a library subpool coding for 12-amino-acid peptides (library XD12). The regions showing the leader sequence, library peptide, and N-terminus of pIII are shown.

more difficult to linearize. This makes electrophoretically separating linearized phage vector DNA from some forms of noncleaved circular plasmid difficult, leading to the transformation of a vector without a library insert. A second difficulty is DNA insertion into a linearized vector, which can yield various side-products, such as self-ligated vector or multimeric products in which several copies of insert, vector, or both are ligated. The third major challenge is inserting the circular DNA into bacterial cells, which is a naturally inefficient step due to the low permeability of membranes for DNA. For transforming DNA into cells, the membrane needs to be disrupted transiently to allow DNA entry and then closed again to ensure cell survival, which typically works for only a small fraction of cells subjected to this process. Our laboratory has been cloning phage display libraries for the past 10 years, and most comprise between 10 million and 1 billion different peptides. Only one contains more than 1 billion members (4×10^9),[15] which was obtained by applying a large effort to isolate more than 1 mg of vector DNA from large *E. coli* cultures, transforming the ligated vector into electrocompetent cells grown in multiple liters of culture and performing a large number of electroporations.

We hypothesized that large peptide phage display libraries could be more easily generated through whole-plasmid PCR, single restriction enzyme digestion, and intramolecular self-ligation. Whole-plasmid PCR was used by Parikh and Guengerich for cloning a library of randomly mutated cytochrome P450[16] and by Matsumura and Rowe for generating a library of randomly mutated beta-galactosidase.[17] To our knowledge, whole-plasmid PCR has neither been applied for the generation of peptide libraries nor for cloning phage display libraries. Intramolecular self-ligation of the linearized vector was used by Scott and co-workers for generating the largest peptide phage display library reported so far (LX-6 library, 10^{10} transformants), but in their approach, single-stranded DNA was isolated from cells and filled by a Klenow fragment rather than amplified by whole-plasmid PCR.[14] The approach that we now propose would facilitate the procedure of phage vector preparation and yield circular DNA containing exactly one gene insert, thus a good quality library.

Our strategy for generating large phage display libraries is based on the amplification of the entire DNA of the phage vector by PCR, wherein the DNA coding for the peptide library is appended to the vector DNA through one of the two DNA primers (Figure 1A). The primer introducing the peptide library (forward primer) would anneal with its constant region to the N-terminal region of the pIII protein gene to display the peptides as a pIII fusion. The second primer (reverse primer) would anneal downstream of the pIII leader sequence and thus just before the region where the forward primer anneals. The two primers would additionally append *Sfi*I restriction sites for ligating the two ends of the PCR product in an intramolecular reaction after *Sfi*I digestion (red dots in Figure 1A). The strategy differs from approaches such as Kunkel mutagenesis and QuikChange cloning in which the DNA polymerase is synthesizing the complementary strand of a vector only but is not amplifying the entire plasmid DNA in repetitive PCR cycles to synthesize micrograms of library DNA. Overall, our new approach was surprisingly efficient and allowed us to generate a large library comprising more than 20 billion

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

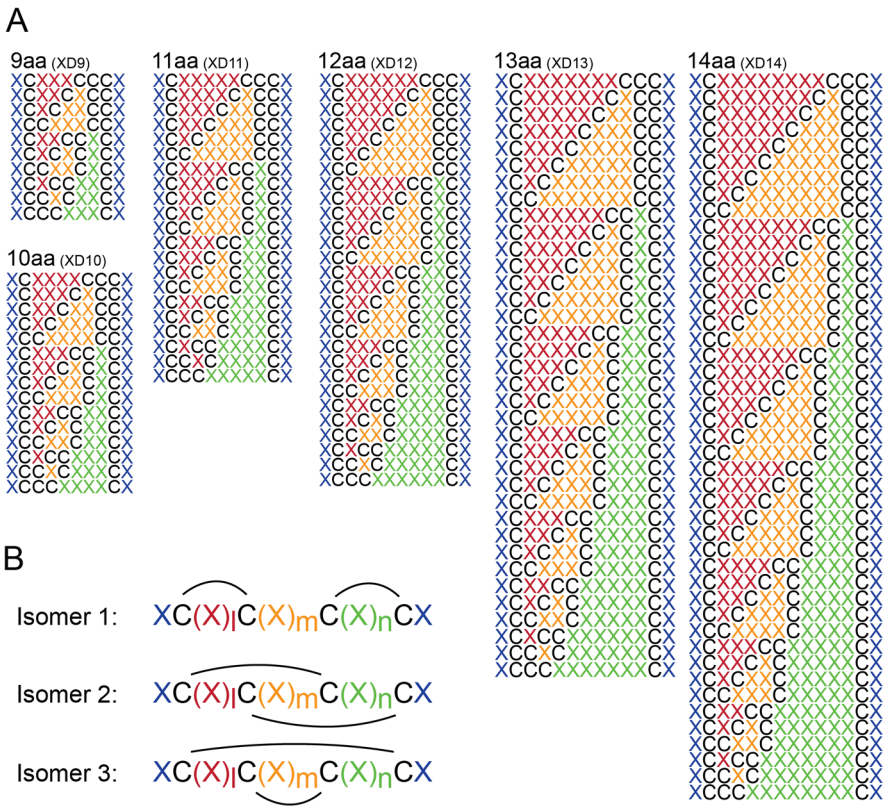


Figure 2. Peptide formats of the phage display library. (A) Positions of cysteines (C) and random amino acids (X) are shown for the six phage display sublibraries XD9 to XD14. The peptides have lengths between 9 and 14 amino acids and contain cysteines in four fixed positions. The libraries were cloned using 155 degenerate DNA primers (Supplementary Table S1) in which the random amino acids were encoded by NNK codons. The red, orange, and green colors show the three randomized segments. (B) The three peptide isomers that are formed upon oxidation of a four-cysteine peptide.

different peptides, all by applying a comparatively moderate effort. Panning the libraries against several targets revealed convergent peptide evolution and yielded nanomolar binders, both indications for a good library quality.

RESULTS

Design of Large Libraries of Cyclic Peptides. We applied our herein proposed cloning approach (Figure 1) to the generation of large and structurally diverse cyclic peptide libraries, chosen for their ability to bind targets with high affinity and selectivity and their relatively high stability.[18] Normally, cyclic peptide phage display libraries are obtained by displaying linear peptides on the phage tip that are subsequently cyclized by either forming a disulfide bridge between cysteines in the peptides or by bridging amino acids through a reaction with chemical linkers.[19] In a recent work, we generated bicyclic peptide libraries by chemically bridging two pairs of cysteines in phage-displayed peptides to increase the overall skeletal diversity (Figure 2), and so we designed a library based on this technology.[20] The connection of four cysteines in peptides allows for generating cyclic peptides with an enormously large skeletal diversity because the four cysteines can be placed in different positions, which gives $3(n-4)!/2(n-6)!$ different backbone structures in peptides with a length of n amino acids (Figure 2). We applied the whole-plasmid PCR cloning approach for generating six sublibraries of peptides with lengths of 9, 10, 11, 12, 13, and 14 amino acids. The peptides of the form $XCX_lCX_mCX_nCX$ all contained a random amino acid at both ends and four cysteines spaced by different numbers of random amino acids wherein $l + m + n = 3, 4, 5, 6, 7$, and 8. Connecting two pairs of cysteines in the six sublibraries yields 30, 45, 63, 84, 108, and 135 different backbone formats (Figure 2). For cloning the libraries, we used the phage vector fd-tet-PK15 (9255 bp),[21] 155 forward primers that appended the peptide DNA to the N-terminal end of the pIII gene as shown in Figure 1, and a reverse primer that annealed downstream of the leader sequence of pIII, leading to PCR amplification of the whole vector. In addition to the peptide sequences, the forward primers append an *Sfi*I restriction site to the vector DNA, which is required for DNA circularization. The reverse primer anneals downstream of the *Sfi*I site present in the leader sequence, yielding a PCR product with *Sfi*I sites at both ends (Figure 1).

Whole-Plasmid PCR and Self-Ligation. The PCR amplification of the large phage vector fd-tet-PK15, around 9.2 kb, required an efficient DNA polymerase with a low mutation rate. We used Phusion DNA polymerase, which can amplify a kb in around 15−30 s and has a lower error rate than, for example, Taq polymerase (around 50-fold lower) or Pfu polymerase (around 8-fold lower). PCR amplification of vector fd-tet-PK15 with the primer pair XD9_1 and XD_reverse at an elongation time of 7 min yielded a product with the desired size (around 30 ng/μL reaction; Figure 3A). A slightly better yield was obtained when the amount of vector template was increased from 5 to 40 ng, and it could be further increased to around 50 ng/μL when the amount of Phusion DNA polymerase was doubled to 0.08 units/μL (Figure 3A). We

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

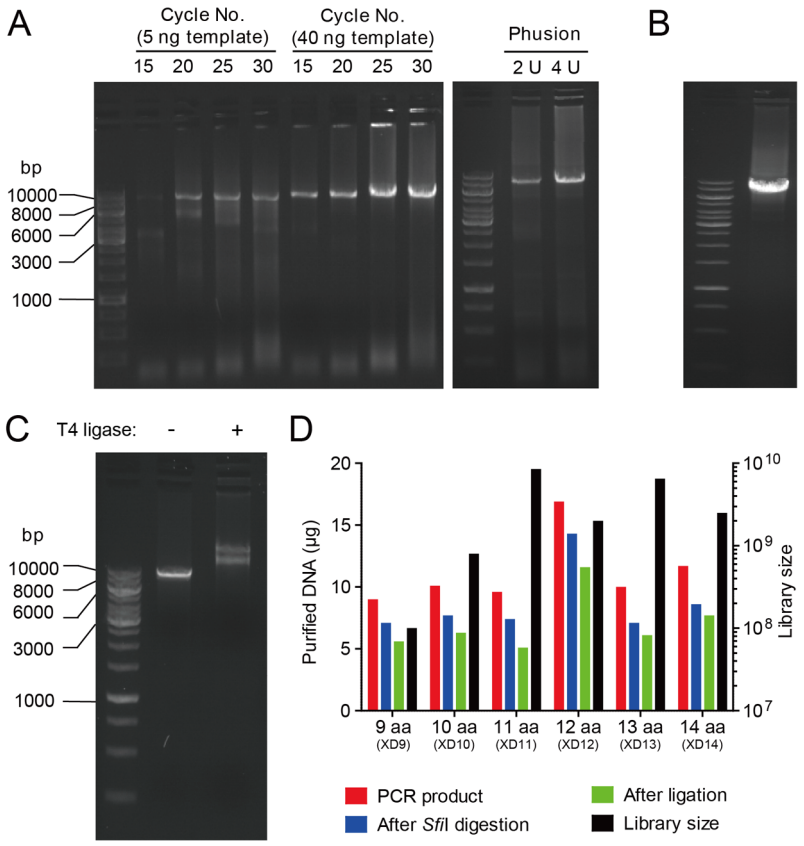


Figure 3. Peptide library cloning. (A) Optimization of whole-plasmid PCR by varying the quantity of template DNA, PCR cycle number, and quantity of Phusion polymerase, indicated for 50 μL reactions. (B) Whole-plasmid PCR product using optimized conditions and a reaction volume of 50 μL. (C) Self-circularization of linear DNA by T4 ligase. (D) Quantity of DNA at each of three experimental steps (scale on left side, quantity of each library is indicated in μg of DNA) and the library size determined based on the number of colonies formed from diluted transformed *E. coli* cells (scale on right side). The quantity of DNA was determined after purification (gel extraction after PCR, PCR purification kit after *Sfi*I digestion, PCR purification kit after ligation).

applied the optimized conditions to PCR-amplify the phage vector in separate 50 μL reactions using the 155 degenerated primers (Table S1), which yielded around 1 μg of linear vector DNA per reaction. We separated the DNA of the reactions by agarose gel electrophoresis (Figure 3B) and extracted DNA of the desired size from the gel. The DNA encoding peptides of identical length were pooled to generate the six sublibraries XD9 to XD14. For each sublibrary, between 9 and 17 μg of purified DNA was obtained.

We next cyclized the linear DNA of the six sublibraries by first generating complementary sticky ends using *Sfi*I, DNA purification, and incubation at a concentration of around 50 ng/μL with T4 ligase for 2 h at 20 °C. Gel electrophoretic analysis of a sample taken from a ligation reaction showed two bands, both running slower than the linear PCR product (Figure 3C). It was likely that one of the two bands corresponded to the desired cyclic product. After inactivation of the ligase and purification, between 5 and 12 μg of DNA remained of each library (Figure 3D). Electroporation of the six DNA pools into electrocompetent *E. coli* cells yielded peptide sublibraries with sizes between 1×10^8 and 6.5×10^9 as determined based on the number of colonies formed after electrotransformation (Figure 3D). For comparison, the largest one of the six sublibraries, XD11 (11-amino acid peptides), was slightly larger than the largest peptide phage display library that we had developed previously (6×6 library, 4×10^9 peptides). After electroporation of DNA into *E. coli* cells, we allowed the cells to recover by incubation in media for 1 h at 37 °C before plating them on selective agar plates. During the incubation time, cells could potentially duplicate or produce new phages that infect additional bacteria, which could lead to an overestimation of the library size. We assessed such potential effects and found that the number of colonies formed after 1 h of incubation is around 2-fold higher than if cells are plated immediately (1 min after electroporation; Figure S1). The moderate increase in the colony number is likely due to multiple effects such as the recovery of electroporated (and damaged) cells and cell division. Importantly, the relatively small increase ensured that we did not overestimate the library size, or at maximum by a factor of 2.

Quality of Peptide Libraries Analyzed by NGS. To test the quality of the library and determine the efficiency of the PCR, we analyzed phage DNA of the sublibrary XD12, encoding peptides of 12 amino acids, by sequencing around 1.7 million random clones, and we analyzed the data using previously described MatLab scripts.[22] Reads that showed low sequencing quality (5.2%) or were lacking constant regions of the vector (8.9%) were eliminated as described in the Methods section. Of the remaining 1.5 million sequences, 93.2% encoded peptides of exactly 12 amino acids and thus the desired length, 1.5% encoded peptides shorter than 12 amino acids, and 5.3% were longer. Many of the shorter or longer

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

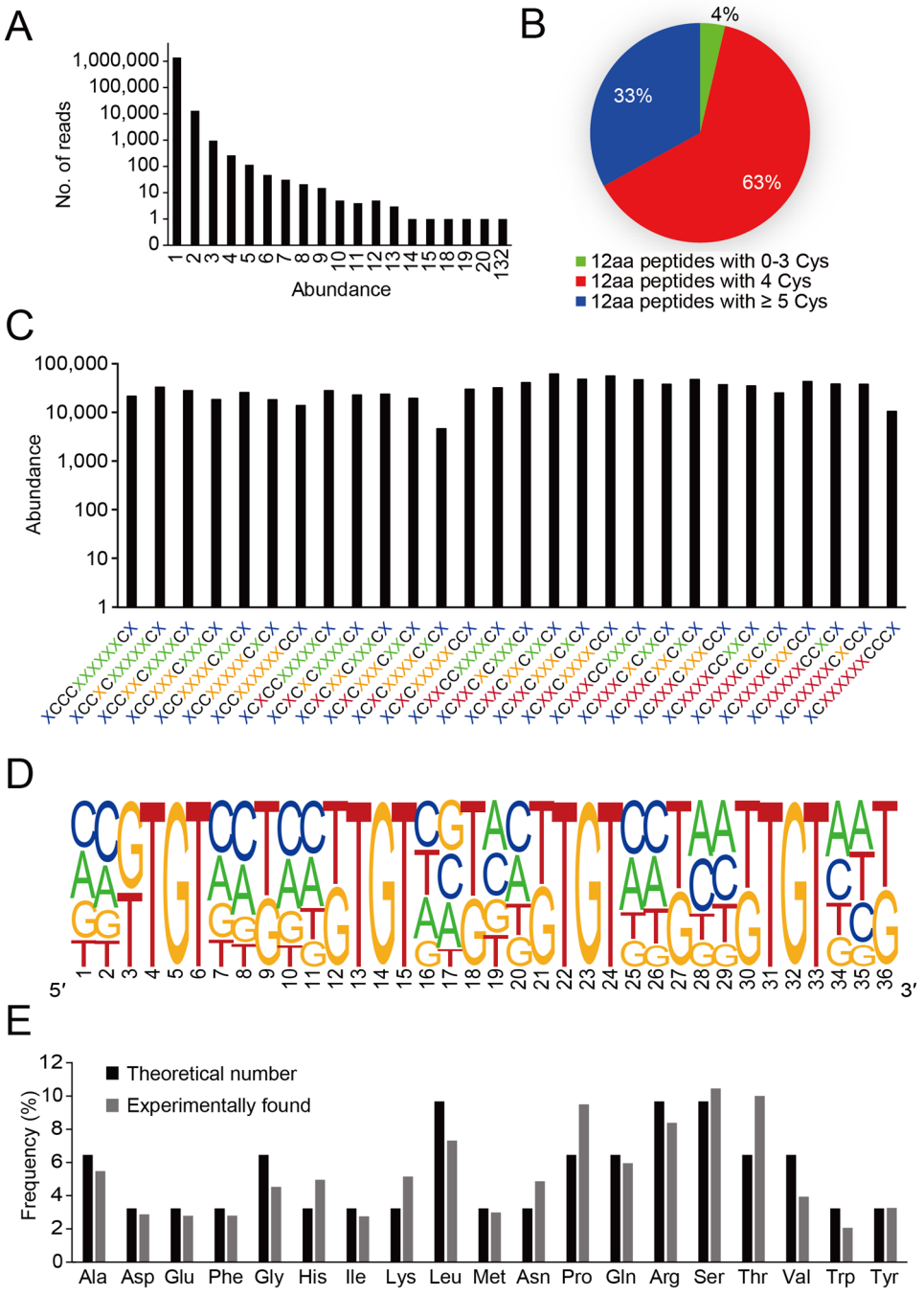


Figure 4. Quality of 12-amino-acid phage display peptide library (library XD12) analyzed by NGS. (A) Number of NGS reads (y) for the indicated copy numbers (x), analyzed for the 12-amino acid peptide library. The vast majority of peptides were sequenced once. (B) Number of cysteines in peptides shown for the 12-amino-acid peptide sublibrary. (C) Representation of the 28 pools of the 12-amino-acid peptide phage display sublibrary. The number of NGS reads for each pool is indicated. (D) Frequency of nucleotides at each position of the pool coding peptides of the format XCXXCXXCXXCX. (E) Theoretical and experimental frequency of all amino acids (except cysteine) in the pool coding peptides of the format XCXXCXXCXXCX.

sequences contained frame shifts and were expected to not yield pIII protein and phage. Of the group with longer sequences, nearly all encoded the peptide PK15, which was part of the phage vector fd-tet-PK15 that was used as the PCR template. The relatively low percentage of PK15 sequences further confirmed that the processes of plasmid PCR amplification and ligation were efficient. We did not make an effort to reduce the percentage of PK15 sequences as we did not expect a negative impact on the phage selections. However, in case this is desired, strategies may be applied to remove the template vector, for example, by digestion with *Dpn*I.

Of the around 1.4 million sequences coding for 12-amino-acid peptides, 97.7% were found only once, suggesting that the vast majority of the sequences in the library were different, as expected (Figure 4A). Of these sequences, 63% coded for peptides that contained exactly four cysteines, which was lower than the statistically expected number $[(31/32)^{(12-4)} = 78\%]$, and most of the remaining ones contained one or several

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

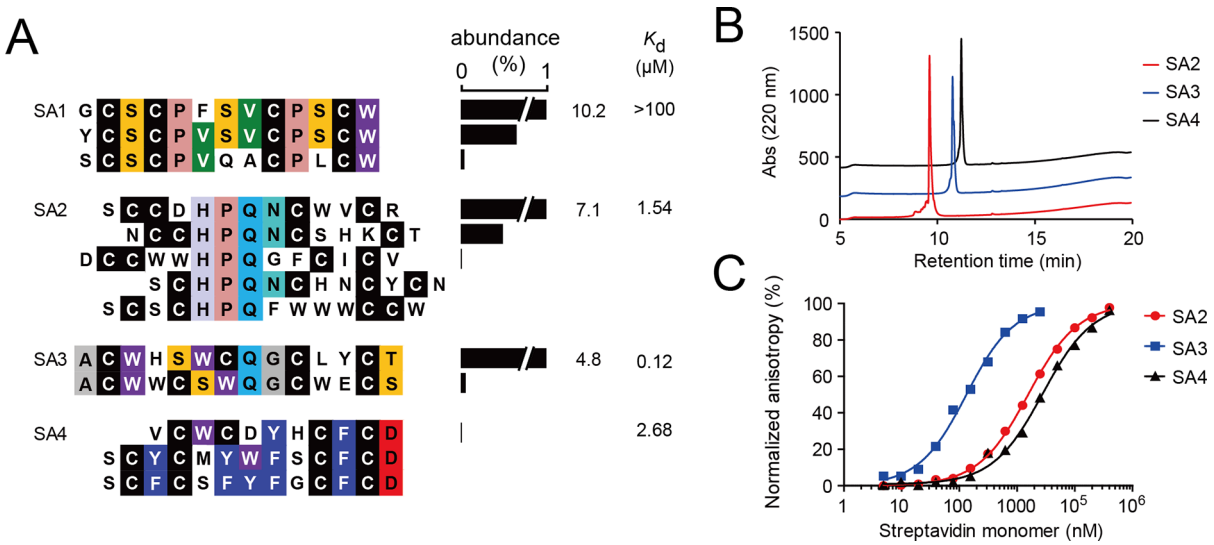


Figure 5. Phage display selections performed against streptavidin. (A) Alignment of peptides in groups based on sequence similarities. The abundance of the peptides is indicated. The binding affinity (K_d) of four peptides, labeled at the N-terminus with fluorescein, was determined by fluorescence polarization. Mean values of two measurements are shown. (B) Analytical HPLC chromatograms of the active isomers of peptides SA2, SA3, and SA4. (C) Fluorescence polarization of peptides at increasing streptavidin concentrations. Mean values of two measurements are shown.

additional cysteines (Figure 4B). The introduction of additional cysteines may be prevented in future libraries by using primers synthesized with mixtures of trinucleotides that lack cysteine codons.[23,24] An analysis of the 28 subpools of the 12-amino-acid peptide sublibrary revealed that all of them were present and similarly represented (Figure 4C).

To analyze the nucleotide and amino acid distributions, we looked at one sample subpool in more detail, the one with the format "XCXXCXXCXXCX." To start, we analyzed the frequency of nucleotides A, T, C, and G at each position in 10,000 randomly picked sequences (Figure 4D). The four cysteines and the random amino acid positions were confirmed to be encoded by TGT and NNK codons, respectively. The nucleotides at the "N" positions coded for A, T, C, and G with frequencies of 30.3%, 19.2%, 31.7%, and 18.8%. In position "K," T and G were found with equal frequency (52.7%, 47.3%). Translation of the 10,000 sequences into peptides and analysis of the amino acids in the randomized positions showed that the frequencies of most amino acids fit well with the theoretical number expected for NNK encoding, with His, Lys, Pro, and Thr showing a slightly higher frequency and Gly, Leu, Val, and Trp showing a slightly lower frequency than statistically expected (Figure 4E). A comparison of the codons of these two groups of amino acids revealed that amino acids with a higher frequency were encoded by codons with A or C in the first two positions and the amino acids with a lower frequency with T or G, suggesting that the small imbalance of amino acid representation resulted from the small bias in the nucleotide ratio described above.

Isolation of Low- to Submicromolar Binders. Finally, we tested whether target-specific cyclic peptides could be isolated from the six peptide phage display sublibraries in affinity selections against the model target streptavidin. We pooled phages of the six sublibraries, bicyclized the displayed peptides by disulfide bridge formation, and panned them in two consecutive rounds against magnetic streptavidin beads. The isolated peptides converged to four different consensus groups, one of them containing the known streptavidin-binding peptide motif His-Pro-Gln in one of the three cysteine-spaced segments, which indicated the isolation of target-specific peptides (Figure 5A). In the three other consensus groups, the sequence similarities were distributed over multiple cysteine-spaced segments. We synthesized four of the peptides with an N-terminal fluorescein moiety, cyclized the peptides by disulfide formation, separated the three regioisomers by reversed phase HPLC (Figure 5B), and tested the binding by fluorescence polarization. As expected, only one of the three regioisomers of each peptide sequence bound to the target. For three of the four peptide sequences, binding affinities in the single-digit micromolar or even submicromolar affinity range were found (K_d = 120 nM; Figure 5C).

Recently, we also applied the large peptide phage display library for the *in vitro* evolution of double-bridged peptides, a format in which two pairs of cysteines in the peptides are bridged by chemical linkers.[20] The linkers were introduced postdisplay by reacting the cysteines with bis-electrophile reagents. Parallel application of multiple structurally diverse chemical linkers increased the library size by approximately another factor of 10. Selections against various targets including coagulation factor XI and interleukin-23 receptor yielded nanomolar binders, which further confirmed the quality of the library.[21]

DISCUSSION

Technical improvements for manipulating and handling DNA have enabled the generation of ever larger phage display libraries, but because reported library sizes have not increased in the past 10 years, it seemed that an upper library size limit had been reached. The largest phage display peptide library that we had cloned with a large effort contains 4 billion different peptides,[19] and the largest phage display peptide library that we found in the literature, the LX-6 library, contains 10 billion different peptides.[14] Herein, we present a cloning strategy that shifts the upper limit for phage library generation and enabled us to clone a cyclic peptide library of more than 20 billion peptides. This new approach, based on

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

robust experimental procedures, enables the construction of larger libraries with a smaller effort than typical for even much smaller libraries.

Key for the large improvement in library generation was the technically simple generation of vector DNA by PCR and the efficient way of inserting library DNA into the vector. Despite the relatively large size of our vector of around 9 kb, large quantities of DNA could be easily generated, and the product was homogeneous, which is in contrast to vector DNA prepared from cells. The vector DNA generation by PCR was clearly facilitated by the tremendous improvements in DNA polymerase engineering that allowed for the amplification of long DNA strands with high efficiency and fidelity. An efficient insertion of peptide-coding DNA into the vector was achieved by appending the DNA to the vector DNA in a whole-plasmid PCR through an overhang PCR. The circularization of this DNA is then highly robust, because the ends are simply joined "intramolecularly." We have further facilitated the ligation by appending sticky ends via an *Sfi*I site.

Compared to classical cassette mutagenesis in which the vector and DNA need to be incubated at an optimal ratio and in which two ligation events are required, this approach provided three key advantages. First, the vector DNA is efficiently produced by PCR and thus does not need to be prepared from bacterial cells, and it does not need to be linearized by restriction enzymes—both steps that are not trivial as described in the Introduction. Second, the self-ligation of a PCR-amplified vector is an intramolecular reaction that is more efficient than the intermolecular ligation of an insert (digested PCR product) and a linearized vector, as also described above. Third, through the strong amplification of the vector DNA by PCR, the template vector represents a minor species, and most of the generated plasmid contains library DNA. This contrasts with cassette mutagenesis approaches in which vector DNA is often not linearized or self-ligated, which can lead to a substantial percentage of plasmid without insertion into the library.

For comparison with previous library-generation techniques, the largest of the six sublibraries we produced in this paper, XD11 (11-amino acid peptides), was slightly larger than the largest peptide phage display library that we had developed previously (6×6 library, 4×10^9 peptides). This previously developed 6×6 library was generated by transforming larger quantities of ligated DNA (33 μg vector, 9 μg insert) and by performing more electroporations (>50 electroporations)[19] and thus applying a much larger effort. In other words, the new strategy yielded larger libraries in a shorter time, despite the use of less material and a smaller effort. The six sublibraries together contain more than 2×10^{10} peptides, which is 5 times more than the 6×6 library. The effort for cloning the six sublibraries together was less than half compared to the work we had invested for creating the above-described 6×6 library, showing that it would be significantly less work to use this strategy to test greater numbers of library variations in the future, such as variations in backbones or ring sizes.

A risk of the whole-plasmid PCR cloning strategy was that mutations would be introduced that could lead to unforeseeable biases in the iteratively performed affinity selections. For example, it was reported that phage clones with mutations in the 5′-untranslated region of gene II were particularly enriched in selections with the NEB library Ph.D.-7,[25,26] and such mutations could potentially be introduced in the whole-plasmid PCR. We were pleased to see that in control selections performed without target protein, the vast majority of sequencing runs showed different peptide sequences (97.7%) and no dominant clones, suggesting that no phage clones with strong propagation advantages were created. Importantly, in affinity selections with the model target streptavidin, we enriched strong binders, which further confirms that the library generated by whole-plasmid PCR has a high diversity and is of good quality. More specifically, we found that families of peptides with similar sequences were enriched, including a peptide family with the His-Pro-Gln motif that is characteristic for streptavidin binders.[27] The convergent evolution is indicative of a successful selection. Additionally, sampling the large combinatorial libraries identified two new consensus sequences for streptavidin. Finally, the successful isolation of high affinity double-bridged peptides against several other targets (nanomolar binding constants), including coagulation factor XI and interleukin-23 receptor, reported in a recent work[21] confirmed the good quality of the library.

By generating a library comprising more than 20 billion clones, we have come closer to the theoretically largest possible phage display library size. The applied commercial TG1 electrocompetent cells have a transformation efficiency of 4×10^{10} cfu/μg of pUC19 vector if a small ratio of DNA/cells is used (10 pg pUC19/25 μL cells). Herein, we reached 1.3×10^9 cfu/μg of ligated vector for the largest one of the six sublibraries (XD11, 11-amino acid peptides). This value is rather close to the pUC19 control, in particular when considering that we used a higher ratio of DNA/cells to economize on expenses for cells (1,250,000 pg/25 μL cells) and that we used a much larger vector (9.2 kb versus 2.7 kb). It is tempting to speculate that even larger phage display libraries can be generated by using a smaller vector, such as a 4 kb phagemid, that would have higher yields in the whole-plasmid PCR and a higher transformation yield.

The whole-plasmid PCR library cloning method developed herein is particularly suited for generating peptide libraries, as the DNA coding for peptides can easily be appended to the vector DNA by overhang PCR. The method may be applied to generate any type of peptide format, ranging from linear peptides over cyclic peptides to bicyclic structures. In principle, the whole-plasmid PCR cloning strategy could be applied for the generation of antibody phage display libraries as well. In the case of antibodies, the DNA coding for the antibody repertoires may be generated in PCRs and the PCR products used as "long primers" for the whole-plasmid PCR. Overall, the ability to make larger phage display libraries with a smaller effort will allow for the expansion of biobased therapeutics, as more compounds and more formats can be tested with an increasingly shorter time.

METHODS

Phage Vector. The phage vector fd-tet-PK15[21] was used as a template for the whole-plasmid PCR. This vector is based on the vector fd-tet[28] that contains the genome of wild-type fd and a segment of transposon Tn10 coding for tetracycline resistance. The vector fd-tet-PK15 additionally contains DNA coding for the 17-amino-acid peptide PK15[19] and a Gly-Ser-Gly linker inserted between the modified leader sequence (changed from ···PFYSHS in wild-type fd genome to ···PFYAAQPAMA; altered sequence is underlined) and the pIII. The region of the DNA coding for the leader sequence, the peptide PK15, and the start of pIII is shown below together with the sequence of the encoded protein. The leader sequence is shown in italics, the peptide in bold. The annealing regions of the primers are underlined.

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX

DNA sequence:

*GTGAAAAAATTATTATTCGCAATTCCTTTAGTTGTTCCTT-TCTATGCGGCCCAGCCGGCCATGGCA***GCATGTAGCG-ATCGTTTTCGTAATTGTCCGGCAGATGAAGCA-CTGTGTGGT**GGTTCTGGCGCTGAAACTGTTGAAAGTTGT

Protein sequence:

*VKKLLFAIPLVVPFYAAQPAMA***ACSDRFRNCPADEALCG-**GSGAETVESC

Degenerated DNA Primers. The 155 primers introducing the DNA sequences coding for the peptides in the whole-plasmid PCR were ordered from Marcrogen who synthesized them at a 0.05 μmol scale and purified them by MOPC, which provided a purity >85%. The primers are shown in Supplementary Table S1. The reverse primer XD_reverse (5′-GCTTCATCTGCCGGACAATTACG-3′) was used for all PCRs.

Whole-Plasmid PCR. The whole-plasmid PCRs were performed in 0.2-mL thin-wall tubes in a total volume of 50 μL, wherein the following reagents were added in the indicated sequence while the tube and reagents were kept on ice: degenerated forward primer (200 nM, final conc.), XD_reverse primer (200 nM, final conc.), dNTP mix (200 μM each, final conc.), 40 ng of phage vector fd-tet-PK15 as a template, and 10 μL of 5 × HF Phusion buffer. The solution was filled up to 48 μL with ddH_2O. Four units (2 μL) of Phusion polymerase (CAT# M0530L, New England Biolabs) were added, and the reaction was mixed by pipetting and then immediately incubated in the thermocycler with the following program. An overview of the PCR reagents and their final concentrations is shown in Supplementary Table S2. The following program was used: initial denaturation of 2 min at 95 °C; 25 cycles of 30 s at 95 °C, 45 s at 55 °C, and 7 min at 72 °C; and final elongation for 7 min at 72 °C.

The PCR product was purified by electrophoresis on a 1% agarose gel (containing 0.005% v/v of ethidium bromide, 1 mM guanosine) in Tris-Acetic acid-EDTA (TAE) buffer. The guanosine was added to protect DNA against damage by UV light.[29] The DNA of each PCR (50 μL) was electrophoresed in separate agarose-gel slots 2 mm thick, 10 mm wide, and 8 mm deep. The bands running at around 10 kb were quickly excised under UV light, and the gel slices from the same sublibrary were mixed before extracting the DNA using a kit (QIAquick Gel Extraction Kit, Qiagen). The DNA was eluted from the kit columns using 10 mM Tris-HCl buffer, at pH 8.5 (three times 20 μL). The following numbers of columns were used for each 50-μL PCR: libraries XD9, XD10, XD11, XD12, one column; libraries XD13, XD14, one column for two or three 50-μL PCRs.

DNA-End Digestion by *Sfi*I. The two ends of the whole-plasmid PCR product were digested with *Sfi*I (CAT# R0123S, New England Biolabs) as follows. In a 0.5-mL tube, 10 μg of the PCR product and 20 μL of 10 × Cutsmart buffer were pipetted and diluted by filling up to 198 μL with ddH_2O. *Sfi*I (2 μL, 40 units) was added; the reaction was mixed well by pipetting and was incubated for 4 h at 50 °C. The digested PCR product was purified with a DNA purification kit (QIAquick Gel Extraction Kit, Qiagen). The DNA was eluted from the columns (two columns for 10 μg of DNA) with 10 mM Tris-HCl buffer, at pH 8.5 (three times 20 μL).

PCR Product Cyclization by Self-Ligation. For each library, around 10 μg of *Sfi*I-digested PCR product was mixed with 20 μL of 10 × ligation buffer containing ATP, and the reaction was diluted by filling up to 180 μL with ddH_2O. T4 ligase (20 μL, 100 Weiss units, Thermo Fisher) was added; the reaction was mixed by pipetting and was incubated at 20 °C for 2 h. The ligation reaction was analyzed by electrophoresis to confirm completion, which was assessed by the appearance of the band of linear PCR product. The ligation mixture was incubated at 70 °C for 10 min to inactivate the ligase, and the ligated DNA was purified with the DNA purification kit. Before the elution of DNA, the column was washed two additional times with wash buffer (Buffer EB, QIAquick Gel Extraction Kit, Qiagen) to thoroughly remove the salt. The DNA was eluted from the column with ddH_2O (three times 20 μL).

Electroporation of DNA into *E. coli* cells. Purified DNA (around 5 to 12 μg in 60 μL) was electroporated into 100 μL of commercial electrocompetent *E. coli* TG1 cells (Lucigen) in four cuvettes (around 40 μL DNA/cell mixture per cuvette) under the bacteria (*E. coli*) mode of the electroporator (Bio-Rad MicroPulser). After each electroporation, cells were resuspended in a prewarmed (37 °C) recovery medium (Lucigen) and were incubated for 1 h at 37 °C with shaking (200 rpm). The size of the library was determined by measuring the total volume of electroporated cells, taking an aliquot of 20 μL and plating a series of 10-fold dilutions on small 2 × YT/tetracycline (10 μg/mL) agar plates. The rest of the cells were plated on 10 large 2 × YT/tetracycline agar plates (14 cm diameter) and incubated at 37 °C overnight. The *E. coli* TG1 cells on the plates were recovered in around 50 mL of 2 × YT medium containing 10% glycerol. Aliquots of 1 mL were stored at −80 °C as glycerol stocks.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.0c00497.

Influence of postelectroporation incubation time on number of transformed bacteria; next-generation sequencing of peptide library; analysis of next-generation sequencing data; phage selection against streptavidin; synthesis of peptides; determination of binding affinity by fluorescence polarization; Supplementary Table 1: sequences of degenerated DNA primers used for the whole-plasmid PCR; Supplementary Table 2: whole-plasmid PCR reagents; Supplementary Table 3: primers used for the second PCR in NGS (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Christian Heinis − *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland;* orcid.org/0000-0001-9982-9457; Phone: +41 (0)21 693 93 50; Email: christian.heinis@epfl.ch; Fax: +41 (0)21 693 98 95

Authors

Xu-Dong Kong − *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland*

Vanessa Carle − *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland;* orcid.org/0000-0001-6372-5664

Cristina Díaz-Perlas − *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland;* orcid.org/0000-0002-9552-3896

Kaycie Butler − *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland;* orcid.org/0000-0002-8365-5867

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.0c00497

Funding

This work was supported by the Swiss National Science Foundation (grants 157842 and 192368).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully thank B. Mangeat from Gene Expression Core Facility of EPFL for help and useful suggestions on next generation sequencing of phage display libraries.

REFERENCES

(1) Nixon, A. E., Sexton, D. J., and Ladner, R. C. (2014) Drugs Derived from Phage Display from Candidate Identification to Clinical Practice. *MAbs 6* (1), 73−85.

(2) Smith, G. (1985) Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (Washington, DC, U. S.) 228* (4705), 1315−1317.

(3) Lindner, T., Kolmar, H., Haberkorn, U., and Mier, W. (2011) DNA Libraries for the Construction of Phage Libraries: Statistical and Structural Requirements and Synthetic Methods. *Molecules 16* (2), 1625−1641.

(4) Deyle, K., Kong, X.-D., and Heinis, C. (2017) Phage Selection of Cyclic Peptides for Application in Research and Drug Development. *Acc. Chem. Res. 50* (8), 1866−1874.

(5) Kristensen, P., and Winter, G. (1998) Proteolytic Selection for Protein Folding Using Filamentous Bacteriophages. *Folding Des. 3* (5), 321−328.

(6) Sieber, V., Plückthun, A., and Schmid, F. X. (1998) Selecting Proteins with Improved Stability by a Phage-Based Method. *Nat. Biotechnol. 16* (10), 955−960.

(7) Kara, E., Nielsen, N. V., Eggertsdottir, B., Thiede, B., Kanse, S. M., and Løset, G. Å. (2020) Design and Characterization of a New PVII Combinatorial Phage Display Peptide Library for Protease Substrate Mining Using Factor VII Activating Protease (FSAP) as Model. *ChemBioChem 21* (13), 1875−1884.

(8) Owens, A. E., Iannuzzelli, J. A., Gu, Y., and Fasan, R. (2020) MOrPH-PhD: An Integrated Phage Display Platform for the Discovery of Functional Genetically Encoded Peptide Macrocycles. *ACS Cent. Sci. 6* (3), 368−381.

(9) Wang, X. S., Chen, P. C., Hampton, J. T., Tharp, J. M., Reed, C. A., Das, S. K., Wang, D., Hayatshahi, H. S., Shen, Y., Liu, J., et al. (2019) A Genetically Encoded, Phage-Displayed Cyclic-Peptide Library. *Angew. Chem., Int. Ed. 58* (44), 15904−15909.

(10) Noren, K. A., and Noren, C. J. (2001) Construction of High-Complexity Combinatorial Phage Display Peptide Libraries. *Methods 23* (2), 169−178.

(11) Kunkel, T. A. (1985) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Proc. Natl. Acad. Sci. U. S. A. 82* (2), 488−492.

(12) Papworth, C., Bauer, J., and Braman, J. W. D. (1996) Site-Directed Mutagenesis in One Day with > 80% Efficiency. *Strategies 9* (8), 3−3.

(13) Huang, R., Fang, P., and Kay, B. K. (2012) Improvements to the Kunkel Mutagenesis Protocol for Constructing Primary and Secondary Phage-Display Libraries. *Methods 58* (1), 10−17.

(14) Bonnycastle, L. L. C., Mehroke, J. S., Rashed, M., Gong, X., and Scott, J. K. (1996) Probing the Basis of Antibody Reactivity with a Panel of Constrained Peptide Libraries Displayed by Filamentous Phage. *J. Mol. Biol. 258* (5), 747−762.

(15) Rentero Rebollo, I., and Heinis, C. (2013) Phage Selection of Bicyclic Peptides. *Methods 60* (1), 46−54.

(16) Parikh, A., and Peter Guengerich, F. (1998) Random Mutagenesis by Whole-Plasmid PCR Amplification. *BioTechniques 24* (3), 428−431.

(17) Matsumura, I., and Rowe, L. A. (2005) Whole Plasmid Mutagenic PCR for Directed Protein Evolution. *Biomol. Eng. 22* (1−3), 73−79.

(18) Zorzi, A., Deyle, K., and Heinis, C. (2017) Cyclic Peptide Therapeutics: Past, Present and Future. *Curr. Opin. Chem. Biol. 38*, 24−29.

(19) Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009) Phage-Encoded Combinatorial Chemical Libraries Based on Bicyclic Peptides. *Nat. Chem. Biol. 5* (7), 502−507.

(20) Kale, S. S., Villequey, C., Kong, X. D., Zorzi, A., Deyle, K., and Heinis, C. (2018) Cyclization of Peptides with Two Chemical Bridges Affords Large Scaffold Diversities. *Nat. Chem. 10* (7), 715−723.

(21) Kong, X.-D., Moriya, J., Carle, V., Pojer, F., Abriata, L. A., Deyle, K., and Heinis, C. (2020) De Novo Development of Proteolytically Resistant Therapeutic Peptides for Oral Administration. *Nat. Biomed. Eng. 4* (5), 560−571.

(22) Rentero Rebollo, I., Sabisz, M., Baeriswyl, V., and Heinis, C. (2014) Identification of Target-Binding Peptide Motifs by High-Throughput Sequencing of Phage-Selected Peptides. *Nucleic Acids Res. 42* (22), e169−e169.

(23) Krumpe, L. R., Schumacher, K. M., McMahon, J. B., Makowski, L., and Mori, T. (2007) Trinucleotide Cassettes Increase Diversity of T7 Phage-Displayed Peptide Library. *BMC Biotechnol. 7* (1), 65.

(24) He, B., Tjhung, K. F., Bennett, N. J., Chou, Y., Rau, A., Huang, J., and Derda, R. (2018) Compositional Bias in Naïve and Chemically-Modified Phage-Displayed Libraries Uncovered by Paired-End Deep Sequencing. *Sci. Rep. 8* (1), 1214.

(25) Brammer, L. A., Bolduc, B., Kass, J. L., Felice, K. M., Noren, C. J., and Hall, M. F. (2008) A Target-Unrelated Peptide in an M13 Phage Display Library Traced to an Advantageous Mutation in the Gene II Ribosome-Binding Site. *Anal. Biochem. 373* (1), 88−98.

(26) Zygiel, E. M., Noren, K. A., Adamkiewicz, M. A., Aprile, R. J., Bowditch, H. K., Carroll, C. L., Cerezo, M. A. S., Dagher, A. M., Hebert, C. R., Hebert, L. E., et al. (2017) Various Mutations Compensate for a Deleterious LacZα Insert in the Replication Enhancer of M13 Bacteriophage. *PLoS One 12* (4), e0176421.

(27) Katz, B. A. (1995) Binding to Protein Targets of Peptidic Leads Discovered by Phage Display: Crystal Structures of Streptavidin-Bound Linear and Cyclic Peptide Ligands Containing the HPQ Sequence. *Biochemistry 34* (47), 15421−15429.

(28) Zacher, A. N., Stock, C. A., Golden, J. W., and Smith, G. P. (1980) A New Filamentous Phage Cloning Vector: Fd-Tet. *Gene 9* (1−2), 127−140.

(29) Gründemann, D., and Schömig, E. (1996) Protection of DNA during Preparative Agarose Gel Electrophoresis against Damage Induced by Ultraviolet Light. *BioTechniques 21* (5), 898−903.

https://dx.doi.org/10.1021/acschembio.0c00497
ACS Chem. Biol. XXXX, XXX, XXX−XXX