JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Silent Encoding of Chemical Post-Translational Modifications in Phage-Displayed Libraries

Katrina F. Tjhung, Pavel I. Kitov, Simon Ng, Elena N. Kitova, Lu Deng, John S. Klassen, and Ratmir Derda*

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

Supporting Information

ABSTRACT: *In vitro* selection of chemically modified peptide libraries presented on phage, while a powerful technology, is limited to one chemical post-translational modification (cPTM) per library. We use unique combinations of redundant codons to encode cPTMs with "silent barcodes" to trace multiple modifications within a mixed modified library. As a proof of concept, we produced phage-displayed peptide libraries Ser-$[X]_4$-Gly-Gly-Gly, with Gly and Ser encoded using unique combinations of codons (TCA-$[X]_4$-GGAGGAGGA, AGT-$[X]_4$-GGTGGTGGT, etc., where $[X]_4$ denotes a random NNK library). After separate chemical modification and pooling, mixed-modified libraries can be panned and deep-sequenced to identify the enriched peptide sequence and the accompanying cPTM simultaneously. We panned libraries bearing combinations of modifications (sulfonamide, biotin, mannose) against matched targets (carbonic anhydrase, streptavidin, concanavalin A) to identify desired ligands. Synthesis and validation of sequences identified by deep sequencing revealed that specific cPTMs are significantly enriched in panning against the specific targets. Panning on carbonic anhydrase yielded a potent ligand, sulfonamide−WIVP, with K_d = 6.7 ± 2.1 nM, a 20-fold improvement compared with the control ligand sulfonamide−GGGG. Silent encoding of multiple cPTMs can be readily incorporated into other *in vitro* display technologies such as bacteriophage T7 or mRNA display.

Since their development in the early 1990s, *in vitro* display of polypeptide chains—on phage,[1] bacteria,[2] yeast,[3] mRNA,[4] and ribosomes[5]—has rapidly become a dominant methodology for the discovery and optimization of biological (protein-based) drugs. Although peptides can serve as potent small-molecule inhibitors, in general linear peptides made of natural amino acids have low conformational stability and poor pharmacokinetic properties. Chemical modification of peptides,[6] their cyclization,[7] or introduction of stabilizing moieties inside or at the terminus of the chain can yield potent and specific ligands where unmodified peptides fail.[8] Nature increases the diversity of proteins made of the 20 canonical amino acids through a handful of post-translational modifications (PTMs).[9] Similarly, chemical PTMs (cPTMs), when applied to peptide libraries, convert genetically encoded libraries of linear peptides to genetically encoded glycosylated,[10] cyclic,[7a] or bicyclic variants[11] with advanced recognition elements (e.g., carbohydrates[12]) that are not present in the side chains of natural peptides. Display of cPTM-peptides accelerates the development of new drug leads that combine the advantages of both "small-molecule" and "biological" (or protein-based) classes of drugs and is a rapidly growing method for lead discovery in the biotechnology industry and large pharmaceutical companies.[13]

Because of their small size, peptide derivatives can have desired tissue permeability, while genetic selection allows for rapid discovery and optimization of these molecules.[6b,7b] To date, cPTM has been used in phage-displayed[11b,14] and mRNA-displayed[15] libraries, but examples of modifications have been limited to one cPTM per screen. While it is possible to introduce different cPTMs into parallel libraries to find an optimal combination of peptide and cPTM,[16] multiple modifications cannot be performed on the same library because the identities of the modifications cannot be traced. We aimed to develop a method that allows simple encoding of a relatively small number of modifications (2−100) that would permit rapid diversification of readily available phage-displayed peptide libraries.

Ideal encoding of cPTMs should not require any modification to existing panning or sequencing protocols to determine both the modification and synergistic peptide motif. We reasoned that multiple modifications can be encoded if (i) multiple chemically identical copies of libraries are available, (ii) the chemical modifications are robust and have reproducible yields, and (iii) each library contains noninterfering, unique features that can be traced genetically once the libraries are pooled (Figure 1). We recognized that requirements (i) and (iii) can be satisfied by encoding the same amino acid sequence by a different set of genetic codons (Figures 2A−D and S1). We refer to this approach as "silent barcoding", as it employs chemically identical peptides translated from genetically distinct DNA sequences. We implemented the silent barcoding approach on the example of phage-displayed libraries using the M13KE vector,[17] which is also the basis for the commercial Ph.D.-7 and Ph.D.-12 libraries (New England BioLabs). These libraries have been used by over 100 research groups to identify more than 3000 peptide ligands to date.[18] Other phage or phagemid vectors and display platforms such as mRNA display of chemically modified peptides[4,15a] can be used analogously.

Production of libraries with silent barcodes requires no modifications to the phage library cloning process[19] (Figure 1): KpnI and EagI restriction enzyme cloning in M13KE vector

Received: October 5, 2015
Published: December 18, 2015

DOI: 10.1021/jacs.5b10390
J. Am. Chem. Soc. 2016, 138, 32−35

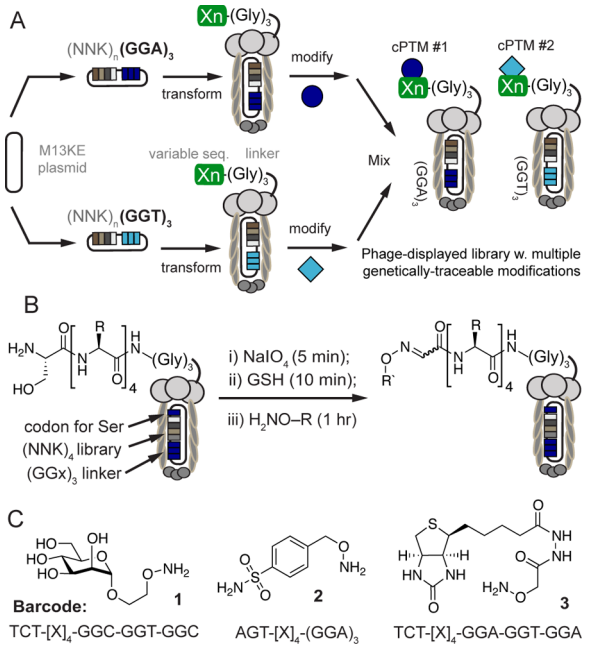

Figure 1. (A) Encoding of the cPTM is achieved by cloning an identifiable barcode sequence, e.g., $(GGA)_3$ or $(GGT)_3$ in the M13 phage DNA near the region that encodes the peptide library. As barcodes, we use translationally identical Gly_3 peptide sequences that produce identical libraries at the peptide level but are distinguishable at the genetic level. Libraries bearing different barcodes are chemically modified and then combined to yield a mixed library of chemically modified peptides. The redundancy of the genetic code allows for 4^3 = 64 different ways to express the Gly_3 linker. (B) Chemical modification of the phage-displayed peptide library through oxime ligation to an N-terminal glyoxal generated by $NaIO_4$ oxidation of N-terminal serine. (C) Hydroxylamine modifiers and silent barcodes used in this article.

produced pIII-displayed peptide libraries Ser-$[X]_4$-Gly-Gly-Gly, with Gly and Ser encoded using unique combinations of codons (TCA-$[X]_4$-GGAGGAGGA, AGT-$[X]_4$-GGTGGTGGT, etc., where $[X]_4$ denotes a random NNK library; see the Supporting Information for details). To avoid bias due to sequencing errors, we chose a set of barcodes that are separated from each other by a Hamming distance of 2 or 3 (Figure 2A–D). Silent barcodes can be introduced at any conserved location of the phage vector, including translationally active and silent regions, or sequences excised from phage proteins (e.g., leader sequence), but it is best to position silent barcodes in close proximity to the variable region to enable simultaneous characterization by deep sequencing of short amplicons.[20]

In a proof-of-principle experiment, we generated four barcoded libraries and used previously established aldehyde–oxime chemistry (Figure 1B)[10,12] to convert the N-terminal serine of each member of the Ser-$[X]_4$-$(Gly)_3$ library to an N-terminal glyoxal. Oxime ligation introduced three modifiers: hydroxylamine–sulfonamide (SA), hydroxylamine–biotin (Bio), and hydroxylamine–mannose (Man). We confirmed the presence of specific modifications in each library (Figure 2A–D) by biotin-capture assay[10] and panning (Figure S2). We then pooled the four barcoded libraries with N-terminal modifications, including the native N-terminal Ser (Figure 2B), and panned the mixed library against matched targets that were expected to recognize each of the chemical modifiers: bovine carbonic anhydrase (BCA) for SA, streptavidin (Str) for biotin, and concanavalin A (ConA) for Man. Panning of libraries with silent barcodes is limited to one round; however, analysis by deep sequencing makes it possible to identify productive ligands even from a single round of panning (Figure 2E).

Analysis of the relative barcode composition after selection showed that the same mixed library enriched distinct barcodes when panned on different matched targets. The magnitude of enrichment mirrored the order of the dissociation constants (K_d) of the ligands that anchor the library to the three different protein targets (Figure 2F). In panning on Str, 95% of the barcodes in the selected library corresponded to modification with biotin; in panning on BCA, 90% of the barcodes corresponded to modification with SA. An increase in the number of barcodes corresponding to Man from 25% to 50% was observed in panning on ConA, which was consistent with previously published panning experiments[12] (Figure 2F). There was no preference for specific barcodes emerging from selection on a mismatched or irrelevant target (here, a bovine serum albumin (BSA)-coated plate). Importantly, the ratio of barcodes remained relatively stable during amplification without any selection, demonstrating that codon-induced bias in amplification of the phage was small (Figure 2F, last set of bars). While silent barcoding can be hidden in sequences of arbitrary length, we caution that the minor bias observed after amplification (Figure 2F) might be exacerbated in silent barcodes of greater length. Barcoding in longer sequences should thus be designed to account for bias in codon usage.

Panning of the library with multiple modifications enriched not only specific modifiers but also specific peptide sequences. Using a previously published enrichment analysis[12] (Tables S1–S2 and Figure S3), we identified SA-modified peptides as sequences that were significantly enriched ($p < 0.05$) in panning on BCA compared with panning on an unrelated target (BSA, Str, or ConA; Figure 3A). As an additional control, we screened for and discarded fast-growing "parasite sequences"[21] that emerged from amplified library phage without panning. For each test–control pair, the analysis can be illustrated as a volcano plot (Figure 3A compares selection against BCA to selection against streptavidin; additional volcano plots are shown in Figure S3). Sequences significantly enriched in panning on BCA exhibited a conserved penultimate Trp residue (Figure 3B and Table S1), whereas sequences enriched in panning on Str yielded different peptide sequences: they were biotin-terminated and contained fragments of the known streptavidin-binding sequence HPQ (Figure 3A and Table S2).

Synthesis and testing of 12 peptides identified in this screen confirmed that the Trp-rich peptide motif acts synergistically with sulfonamide (Figure 4A). Electrospray ionization mass spectrometry (ESI-MS) yielded relative binding affinities of all SA-modified peptides and estimated that SA–WIVP binds to BCA with 20-fold higher affinity than the control SA–GGGG ligand (Figure 4A,B). Subsequent isothermal titration calorimetry (ITC) measurements of the K_d values for the peptides SA–WIVP, SA–FVVR, SA–TRPA, and SA–GGGG confirmed the relative affinities estimated by ESI-MS (Figure 4C for SA–WIVP and Figure S4 for the remaining modified peptides). The best ligand, SA–WIVP, had single-digit-nanomolar affinity (K_d = 6.7 ± 2.1 nM) and was ~20 times more potent than the control ligand SA–GGGG (K_d = 140 ± 10 nM; Figure S4). We note that our unbiased search identified a sulfonamide–indole motif that has been found in other potent ligands for carbonic anhydrase.[22] SA–WIVP shares structural features with the most potent carbonic anhydrase inhibitor reported to date (K_d = 0.03 nM; see Figure S5).[23]

DOI: 10.1021/jacs.5b10390
J. Am. Chem. Soc. 2016, 138, 32–35

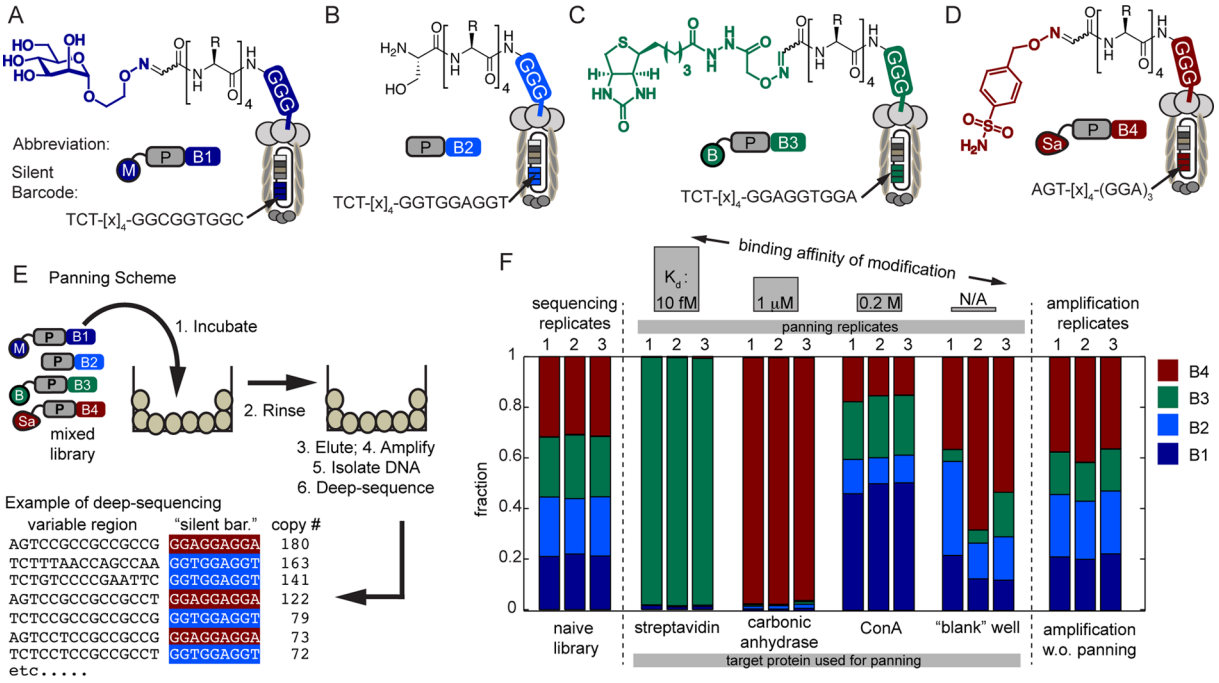

Figure 2. (A–D) Phage-displayed libraries of peptides with four distinct modifications and sequences of "silent barcodes" that encode these modifications. (E) In a model panning assay, a mixed library containing equal ratios of each modified library was incubated with immobilized streptavidin, bovine carbonic anhydrase, ConA, or BSA-blocked well ("blank well"), followed by a washing step and acid elution. Eluted phage was amplified and processed for DNA isolation and sequencing by Illumina. (F) Analysis of sequencing data. Phage eluted from streptavidin, carbonic anhydrase, ConA, or uncoated well showed a pronounced binding preference for the corresponding protein target. Sequencing data confirmed that each modification type was equally represented in the naïve mixed library and that changes in relative ratios were not due to amplification bias. Three bars describe the results of sequencing of three independent panning experiments.

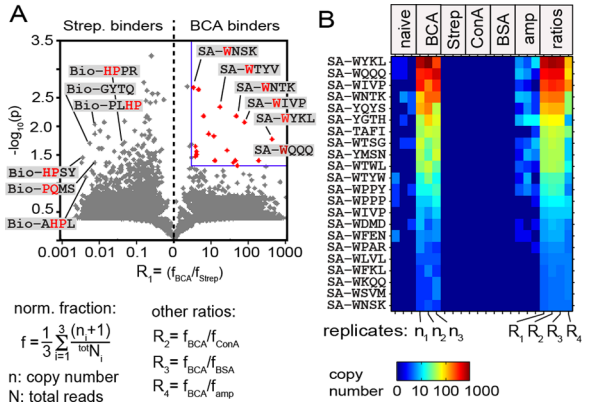

Figure 3. (A) Plot of p value vs ratio (R) ("volcano plot") in selection of the barcoded library against BCA and streptavidin (Str); ligands with $R > 3$ and $p < 0.05$ were enriched in panning on BCA; those with $R < 0.3$ and $p < 0.05$ were enriched in panning on Str. Two enriched sets have different chemical modifications—sulfonamide (SA) for BCA and biotin (Bio) for Str—and different sequence motifs (see Tables S1 and S2). (B) Sequences significantly enriched in panning on BCA but not in four control panning experiments on ConA, BSA, or Str or during amplification without selection (all in triplicate). The heat map describes copy numbers determined by deep sequencing.

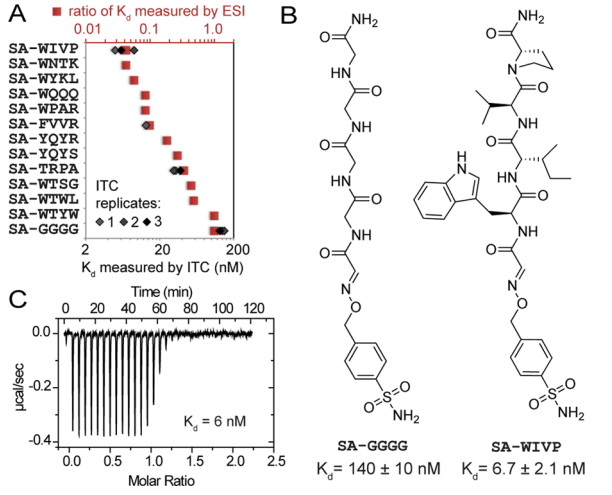

Figure 4. (A) Relative dissociation constants (K_d) for binding of 12 putative SA–peptide hits and one control (SA–GGGG) to BCA as determined by ESI-MS (for the raw data, see Figure S7); absolute values of K_d (black diamonds) were determined by ITC. The most active hit, SA–WIVP, had K_d = 6.7 ± 2.1 nM. (B) Structure of synthetic SA–peptide conjugates. (C) Representative ITC trace for SA–WIVP (for additional traces, see Figure S4).

Next, we evaluated the affinity of three modified peptides enriched in selection against streptavidin using ESI-MS. The biotinylated peptides Bio–AHPL, Bio–HFTN, and Bio–GYTQ exhibited a modest but measurable 2–3 fold increase in binding to Str compared with the control ligand Bio–GGGG. The K_d values were 114, 57, and 49 fM for the selected biotin–peptide conjugates, respectively, and 133 fM for Bio–GGGG (Figure S9).

To compare the performance of mixed-barcode libraries with a previously validated panning experiment,[12] we analyzed a selection of libraries containing N-terminal mannose, methoxy, or Ser in panning against ConA. Each modified library was panned either separately or as a pooled mixture (Figure S10A–

DOI: 10.1021/jacs.5b10390
J. Am. Chem. Soc. 2016, 138, 32–35

C). Similarly to the previously reported selection,[12] both panning experiments enriched the Man–WY motif (Figure S10E-H). Mixed-barcode libraries identified this motif from four replicates of panning, whereas the use of separately panned libraries mandated 12 independent panning experiments. The enrichment in panning that contained mixed libraries was different from the enrichment in conventional panning, potentially because of the competition of ligands with productive and unproductive modifications present in the same solution. We anticipate that the presence of control libraries in the same solution will permit the development of selection methods that are currently impossible with non-encoded modifications.

In conclusion, silent barcoding enables tracing of cPTMs in multiplexed phage-displayed peptide libraries. We envision that silent barcoding will be used as a powerful tool in selections that employ phage-displayed or mRNA-displayed peptide-derived macrocycles.[6b] Since the structure of the cross-linking moiety has a strong influence on the conformation and activity of macrocycles, silent encoding can be used to find an optimal sequence and optimal cross-linker in one screen rather than in separate screens.[11a] In the area of genetically encoded fragment-based design,[24] the challenge is identification of an optimal linker that connects a known promiscuous fragment to the genetically encoded synergistic peptide fragment; silent encoding can be used to encode multiple linkers and identify the optimal structure of the linker that connects two fragments.

Although the selection of "silently-encoded" cPTM libraries is limited to one round of panning, modern deep-sequencing tools and bioinformatics analyses can effectively deconvolute the results of panning even after a single round[12,20b] and yield potent binders. Additional rounds of panning may be achieved by modification of the selected population with a single modification to further enrich for synergistic peptide motifs. With only minor modifications, this technology can be applied to encode any chemical modifier, such as modifying linkers[12] or cross-linking moieties used for cyclization,[11a] and can be expanded to chemically modified libraries displayed on mRNA or T7-phage, providing a powerful strategy for multiplexing chemically modified protein or peptide libraries.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b10390.

MatLab scripts 400 × 400plots (ZIP)
MatLab scripts DeepSeqAnalysisScripts (ZIP)
MatLab scripts MakeFigureS10_Scripts (ZIP)
Procedures, ESI-MS and NMR data, Tables S1 and S2, and Figures S1–S10 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*ratmir.derda@ualberta.ca

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Dr. Dax Torti for helpful discussions regarding Illumina sequencing. This work was supported by NSERC Discovery Grant 402511, the Alberta Glycomics Centre, and the Canadian Foundation for Innovation. K.F.T. acknowledges Alberta Innovates (AI) Health Solutions for a summer fellowship; S.N. acknowledges AI Technology Futures for a graduate fellowship.

■ REFERENCES

(1) Scott, J.; Smith, G. *Science* **1990**, *249*, 386.
(2) Georgiou, G.; Stathopoulos, C.; Daugherty, P. S.; Nayak, A. R.; Iverson, B. L.; Curtiss, R., III *Nat. Biotechnol.* **1997**, *15*, 29.
(3) Boder, E. T.; Wittrup, K. D. *Nat. Biotechnol.* **1997**, *15*, 553.
(4) Li, S.; Roberts, R. W. *Chem. Biol.* **2003**, *10*, 233.
(5) Hanes, J.; Plückthun, A. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4937.
(6) (a) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chem. Biol. Drug Des.* **2013**, *81*, 136. (b) Heinis, C.; Winter, G. *Curr. Opin. Chem. Biol.* **2015**, *26*, 89.
(7) (a) Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2014**, *9*, 443. (b) Passioura, T.; Katoh, T.; Goto, Y.; Suga, H. *Annu. Rev. Biochem.* **2014**, *83*, 727.
(8) (a) Clackson, T.; Wells, J. A. *Trends Biotechnol.* **1994**, *12*, 173. (b) Patgiri, A.; Jochim, A. L.; Arora, P. S. *Acc. Chem. Res.* **2008**, *41*, 1289. (c) Freire, F.; Gellman, S. H. *J. Am. Chem. Soc.* **2009**, *131*, 7970.
(9) Jensen, O. N. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 391.
(10) Ng, S.; Jafari, M. R.; Matochko, W. L.; Derda, R. *ACS Chem. Biol.* **2012**, *7*, 1482.
(11) (a) Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. *Angew. Chem., Int. Ed.* **2014**, *53*, 1602. (b) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5*, 502.
(12) Ng, S.; Lin, E.; Kitov, P. I.; Tjhung, K. F.; Gerlits, O. O.; Deng, L.; Kasper, B.; Sood, A.; Paschal, B. M.; Zhang, P.; Ling, C. C.; Klassen, J. S.; Noren, C. J.; Mahal, L. K.; Woods, R. J.; Coates, L.; Derda, R. *J. Am. Chem. Soc.* **2015**, *137*, 5248.
(13) (a) Kawakami, T.; Ishizawa, T.; Fujino, T.; Reid, P. C.; Suga, H.; Murakami, H. *ACS Chem. Biol.* **2013**, *8*, 1205. (b) Josephson, K.; Ricardo, A.; Szostak, J. W. *Drug Discovery Today* **2014**, *19*, 388. (c) Baeriswyl, V.; Rapley, H.; Pollaro, L.; Stace, C.; Teufel, D.; Walker, E.; Chen, S.; Winter, G.; Tite, J.; Heinis, C. *ChemMedChem* **2012**, *7*, 1173. (d) Santoso, B.; Murray, B. W. *Methods Mol. Biol.* **2015**, *1248*, 267.
(14) Fukunaga, K.; Hatanaka, T.; Ito, Y.; Taki, M. *Mol. BioSyst.* **2013**, *9*, 2988.
(15) (a) Guillen Schlippe, Y. V.; Hartman, M. C.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10469. (b) Millward, S. W.; Takahashi, T. T.; Roberts, R. W. *J. Am. Chem. Soc.* **2005**, *127*, 14142. (c) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. *Chem. Biol.* **2011**, *18*, 1562.
(16) Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. *Angew. Chem., Int. Ed.* **2014**, *53*, 1602.
(17) Noren, K. A.; Noren, C. J. *Methods* **2001**, *23*, 169.
(18) Huang, J.; Ru, B.; Zhu, P.; Nie, F.; Yang, J.; Wang, X.; Dai, P.; Lin, H.; Guo, F. B.; Rao, N. *Nucleic Acids Res.* **2012**, *40*, D271.
(19) Ng, S.; Tjhung, K. F.; Paschal, B. M.; Noren, C. J.; Derda, R. *Methods Mol. Biol.* **2015**, *1248*, 155.
(20) (a) Matochko, W. L.; Chu, K.; Jin, B.; Lee, S. W.; Whitesides, G. M.; Derda, R. *Methods* **2012**, *58*, 47. (b) Miersch, S.; Sidhu, S. S. *Methods* **2012**, *57*, 486.
(21) Matochko, W. L.; Li, S. C.; Tang, S. K. Y.; Derda, R. *Nucleic Acids Res.* **2014**, *42*, 1784.
(22) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem. Rev.* **2008**, *108*, 946.
(23) Grzybowski, B. A.; Ishchenko, A. V.; Kim, C. Y.; Topalov, G.; Chapman, R.; Christianson, D. W.; Whitesides, G. M.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1270.
(24) (a) Meyer, S. C.; Shomin, C. D.; Gaj, T.; Ghosh, I. *J. Am. Chem. Soc.* **2007**, *129*, 13812. (b) Wang, W.; Hirano, Y.; Uzawa, T.; Liu, M.; Taiji, M.; Ito, Y. *MedChemComm* **2014**, *5*, 1400.

DOI: 10.1021/jacs.5b10390
J. Am. Chem. Soc. 2016, 138, 32–35