

Genetically encoded fragment-based discovery (GE-FBD) from phage-displayed macrocyclic libraries with genetically-encoded unnatural pharmacophores

Arunika Ekanayake¹, Lena Sobze¹, Jihea Youk¹, Nicholas J. Bennett¹, Raja Mukherjee¹,

Atul Bhardwaj,¹ Frank Wuest^{1,2}, Ratmir Derda^{1*}

¹ Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada;

² Cross Cancer Institute, University of Alberta, Edmonton, AB T6G 1Z2, Canada;

Corresponding author: *ratmir@ualberta.ca

Keywords (Macrocycles, Phage Display)

Abstract

Genetically-encoded macrocyclic peptide libraries with unnatural pharmacophores are valuable source for discovery of ligands for many targets of interest. Traditionally, generation of such libraries employs “early stage” incorporation of unnatural building blocks into the chemically or translationally-produced macrocycles. Here, we describe a divergent late-stage approach to such libraries starting from readily available starting material: genetically-encoded libraries of peptides. A diketone linchpin 1,5-dichloropentane-2,4-dione converts peptide libraries displayed on phage to 1,3-diketone bearing macrocyclic peptides (DKMP): shelf-stable precursors for Knorr pyrazole synthesis. Ligation of diverse hydrazine derivatives onto DKMP libraries displayed on phage that carries silent DNA-barcodes yields macrocycle libraries in which the amino acids sequence and the pharmacophore are encoded by DNA. Selection of this library against carbonic anhydrase enriched macrocycles with benzenesulfonamide pharmacophore and nanomolar K_d . The methodology described in this manuscript can graft diverse pharmacophores into many existing genetically encoded phage libraries and significantly increase the value of such libraries in molecular discoveries.

Introduction

Late-stage functionalization of unprotected peptides composed of natural amino acids in aqueous media provides a powerful approach to modify readily available million-to-billion scale genetically encoded peptide libraries, displayed on phage, mRNA or DNA.¹⁻⁵ Such functionalization expands existing genetically encoded chemical space to incorporate unnatural pharmacophores not present in the original peptide libraries, allowing the discovery of value-added molecules with properties not offered by peptides alone.⁶⁻⁹ Numerous reports demonstrated the power of discovery of potent ligands from phage- and mRNA-displayed libraries in which unnatural pharmacophores were grafted onto the peptides in million-to-trillion-scale genetically-encoded library¹⁰⁻¹². Genetically encoded fragment-based discovery (GE-FBD)¹⁰ from such libraries is conceptually similar to canonical fragment-based design (FBD), which is a powerful method for the development of ligands, drug leads and three FDA-approved drugs to date.¹³⁻¹⁹ Methods for production of GE-macrocyclic libraries with unnatural pharmacophores are bottom-up organic synthesis of DNA-encoded libraries (DEL),^{20,21} *in vitro* translation of mRNA-displayed macrocycles using modified protein translation systems^{11,12,22}, and engineering of biochemical pathways that produce ribosomally-made and postranslationally-modified peptides²³. Among these approaches late-stage chemical modification of existing GE-peptide libraries offers potentially the simplest path to generating chemical diversity. It combines a robust expression of million-to-billion scale peptides libraries made of 20 natural amino acids with a simple diversification of peptides by site-specific chemical conjugation. This report advances both the late stage functionalization of GE-libraries and GE-FBD approaches by introducing three important concepts: (i) ligation of unnatural fragments onto pre-formed GE-macrocyclic libraries; (ii) production of shelf-stable GE-macrocyclic libraries with a handle for biorthogonal reaction that forms an irreversible covalent bond, (iii) encoding and decoding of ligated unnatural fragments by DNA sequencing.

There exist several examples of late-stage functionalization of GE-peptide libraries to yield macrocycles with unnatural chemotypes. Oximes derived from dichloroacetone (DCA) linchpin²⁴

convert linear peptide libraries displayed on phage to a library of macrocycles and simultaneously introduces a diverse range of glycans²⁵ or reactive covalent warheads²⁶ into peptide macrocycles. Alkylation with other bi- and tri-dentate electrophile linchpins were employed to introduce non-covalent and covalent warheads into T7²⁷ and M13-displayed²⁸ phage libraries. These approaches use macrocyclization as a divergent step and require optimization of non-trivial ring closing reactions for every linchpin structure. More robust approaches graft the desired unnatural fragments onto the preformed macrocycles²¹. For example, Suga and co-workers converted free cysteines in mRNA displayed peptide macrocycles to dehydroalanines (DhA)²⁹, followed by Michael addition of thioglycosides to DhA³⁰. Ketone functionality in DCA-modified phage-displayed libraries can be used to introduce glycans and other pharmacophores via oxime ligation²⁴. Minor limitations of these approaches are the reversibility of formed bonds, slow reactivity of ketone and reactivity of DhA functionality to biological nucleophiles. To overcome these limitations, we describe synthesis and late-stage diversification of GE macrocyclic libraries with a 1,3-diketone reactive handle. We demonstrate that this approach offers advantages not present in prior reports such, as stability to storage, superior reactivity and formation of irreversible bond.

Aliphatic 1,3-diketones are bona fide bio-orthogonal moieties with long term stability *in vivo*. The evidence for the stability of 1,3-diketones *in vivo* comes from work of Barbas and Lerner who immunized mice with 1,3-diketone haptens and isolated antibody reactive to 1,3-diketones^{31,32}. They demonstrated that 1,3-diketones injected into blood circulation reacted with the circulating anti-1,3-diketone antibody selectively³² and only a rare unique peptide sequences³³ had any detectable reactivity with this group. Kate Carrol and co-workers discovered that 1,3-diketones react with biomolecules that contain sulfenic acid—a transient species formed from endogenous cysteines due to oxidative stress³⁴⁻³⁶. Sulfenic acid rapidly attacked nucleophilic carbon in derivatives of dimedone—1,3 diketone in a six-membered ring—but this reaction was significantly slower with open chain aliphatic 1,3-diketones³⁶. Unlike ketones that form reversible adducts with hydrazine-derivatives, 1,3-diketones undergo an irreversible

cyclocondensation with hydrazines to generate 1,2-diazoles (pyrazoles). Based on these observations, we rationalized that the Knorr pyrazole reaction³⁷ can be employed as bio-orthogonal late-stage functionalization of GE-libraries in aqueous, mild, biocompatible conditions.

Results

Our design started from aqueous modification of genetically encoded peptide libraries displayed on phage with 1,5-dichloropentane-2,4-dione (DPD) to yield 1,3-diketone modified macrocyclic peptides (DKMP) (Fig. 1a). We employed previously reported synthesis of DPD³⁸ and confirmed its structure by X-ray crystallography (Supplementary Fig. S1). In model studies, DPD converted synthetic peptides of structure, X_nCX_mC to DKMP's within 30 minutes in pH 8.5 aqueous bicarbonate buffer (Fig. 1a-c). We demonstrated that these DKMP's are ideally poised for late-stage functionalization using diverse alkyl and aryl hydrazine functionalities as precursors. Reaction in aqueous ammonium acetate buffer (pH 5.0) irreversibly grafted these functionalities onto the macrocycle via a hydrolytically stable endocyclic 1,2-diazole (Fig. 1a, e). Incubation of diazole peptide product and hydrazine in ammonium acetate buffer for 7 days produced no detectable cross-over product confirming the irreversibility of the formed bond (Supplementary Fig. S2). Substituted phenyl hydrazines formed N-aryl 1,2-diazole macrocyclic peptides with rate constants ranging from 0.01 to 1.27 $M^{-1}s^{-1}$ (Fig. 1 d, f). Reaction between DKMP and diverse N-alkyl hydrazines occurred with $k=1-2 M^{-1}s^{-1}$ in the same conditions (Fig. 1g). Hammett series plot for substituted phenyl hydrazines gave a $\rho = -0.44$ value, indicating a significant buildup of positive charge in the transition state of the rate determining step (Fig. 1c). Reaction between DKMP and the most electron poor perfluorophenyl hydrazine was completed after overnight incubation whereas other reactions were completed in <1 hour incubation (Fig. 1e, Supplementary Fig. S3-S13 and Table S1). We employed these reactions to demonstrate grafting of fluorophores, imaging chelators, *n*-alkyls, and polyethylene glycol moieties onto macrocycles (Fig. 1h-k).

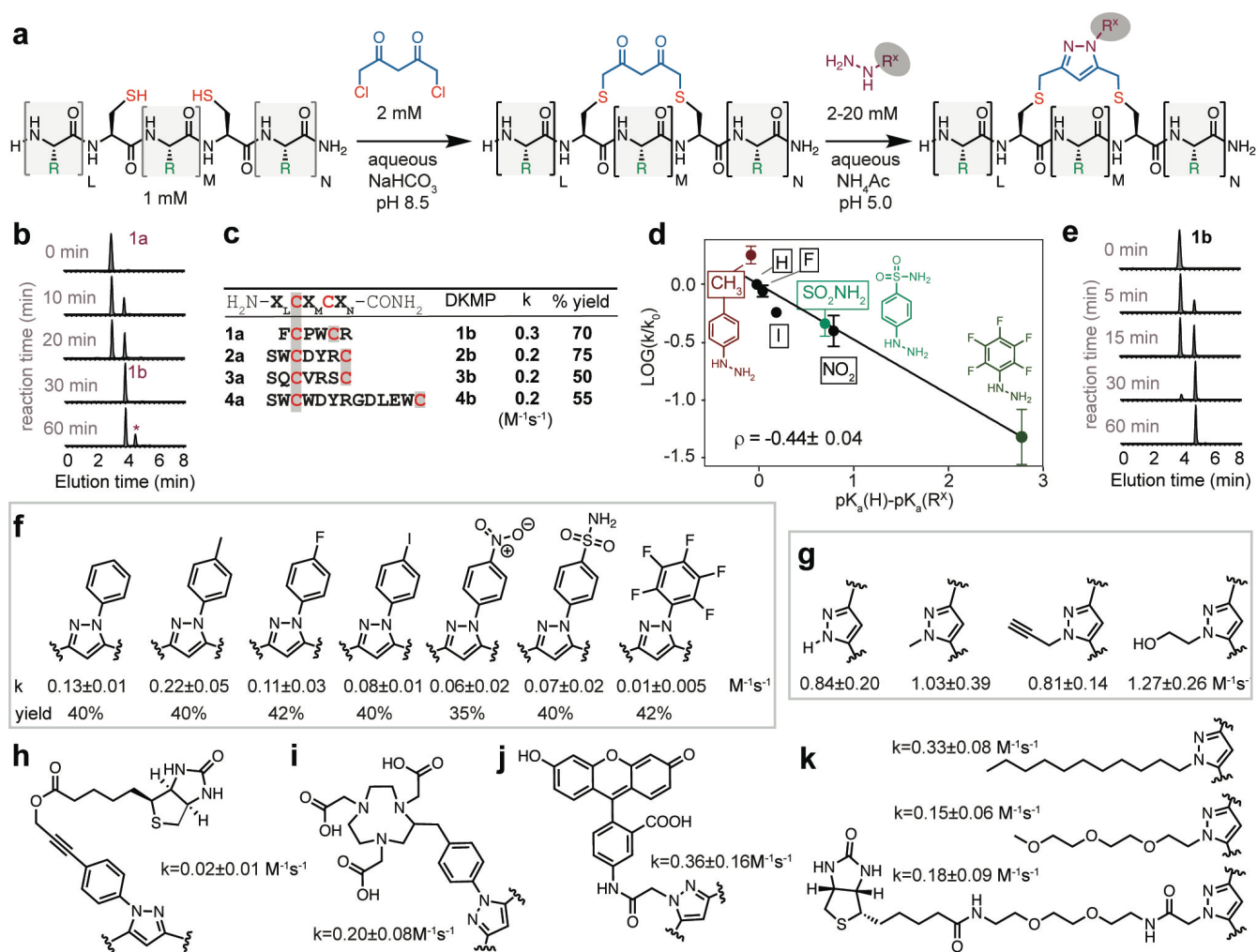


Figure 1. Model reactions on peptides. (a) Conversion of unprotected linear peptides to macrocycles with a diketone linchpin using dichloropentadione (DPD) at pH 8.5 and further functionalization of macrocycles using hydrazines at pH 5.0 (b) Liquid chromatography (LC) traces at 214 nm for the reaction between unprotected peptide **1a** and DPD. The reaction is complete at 30 minutes, prolonged reaction leads to an emergence of a minor by-product indicated by *. (c) Reactions between different peptides and DPD show similar reaction rates. (d) Ligation of aromatic hydrazines onto 1,3-diketone peptide displays positive Hammett correlation to the substituents on the phenyl ring indicating negative charge build up during the transition state. (e) LC traces for reaction between 1,3-diketone peptide and phenyl hydrazine (2 mM). The reaction is completed within 60 minutes. (f) Reaction rates for 1,3-diketone peptides with aryl hydrazines and yields for their HPLC purified products. (g) Reaction rates for 1,3-diketone peptide and alkyl hydrazines. (h-k) Reaction rates for 1,3-diketone peptide and hydrazine groups displaying affinity handles, imaging chelators, long alkyl chains and oligoethyleneglycol units.

Grafting of long chain *n*-alkyl chains (“lipidation”) or poly-ethylene glycol moieties (“pegylation”) are known to enhance pharmacokinetic properties of peptides and diketone ligation can be used for prospective “lipidation” and “pegylation” of macrocyclic peptide libraries.

When adapting the reaction conditions to phage-displayed peptide libraries, we employed well-established biotin-capture coupled to plaque forming units (PFU) assay approach^{25,39-42} to measure the conversion, regioselectivity⁴² and kinetics⁴⁰ of chemical modification of phage displayed libraries (Fig. 2). Hydrazine is known to damage of DNA in bacteria, viruses/phages and even in isolated DNA⁴³. Indeed, PFU assay confirmed that most hydrazine derivatives killed >99.999% of infective phage particles in less than 5 minutes (Supplementary Fig. S14). Importantly, we observed that addition of metal chelator such as EDTA rescued this toxicity⁴³. For example, in the presence of 5 mM EDTA, incubation of phage with 20 mM phenyl hydrazine for several hours did not show significant decrease in the number of infective particles. Importantly, EDTA did not influence the rate of reaction between diketone and hydrazine (Supplementary Fig. S14) and modification of DKPM-libraris on phage was possible in the presence of EDTA (Fig. 2a-c).

Specifically, we modified phage-displayed SXCX₃C library of 130,000 heptamer peptides⁴⁴ with 1 mM DPD for 30 minutes (Fig. 2a-b) and then 5 mM biotin hydrazine probe + 5 mM EDTA for 1 hour and used biotin-capture assay to monitor the reaction efficiency (Fig. 2c). We measured PFU before capture the phage particles with streptavidin beads (PFU^{before}) and after capture (PFU^{after}) and used the capture ratio $CapR = (PFU^{before} - PFU^{after}) / PFU^{before}$ to estimate the kinetics of the emergence of biotinylated phage particles at various reaction times (Fig. 2d). As in previous reports, blue-white PFU assay in agar overlay supplemented with colorimetric substrate X-gal distinguished phage particles that displayed a library of peptides and transduced LacZa reporter (i.e., produced PFU of blue color) in the presence of wild type (wt) phage that displayed no Cys-containing peptides and transduced no reporter (i.e., produced PFU of white color).⁴¹ The biotin-capture assay showed reaction between hydrazine and a library of 1,3-diketone macrocycles plateaued at ~70% modification yield after 30-60 min. Reaction was regioselective

because the control wild type phage present in the same solution was not biotinylated and not captured (Fig. 2d-e). The intermediate DKMP library was stable in storage, as evidenced from reproducible yield of biotinylation of DKMP library after 1 month of storage (Fig. 2f). Reactivity of DKMP handle was unchanged after 10 days of incubation in protein and metabolite rich media (yeast extract, Fig. 2f).

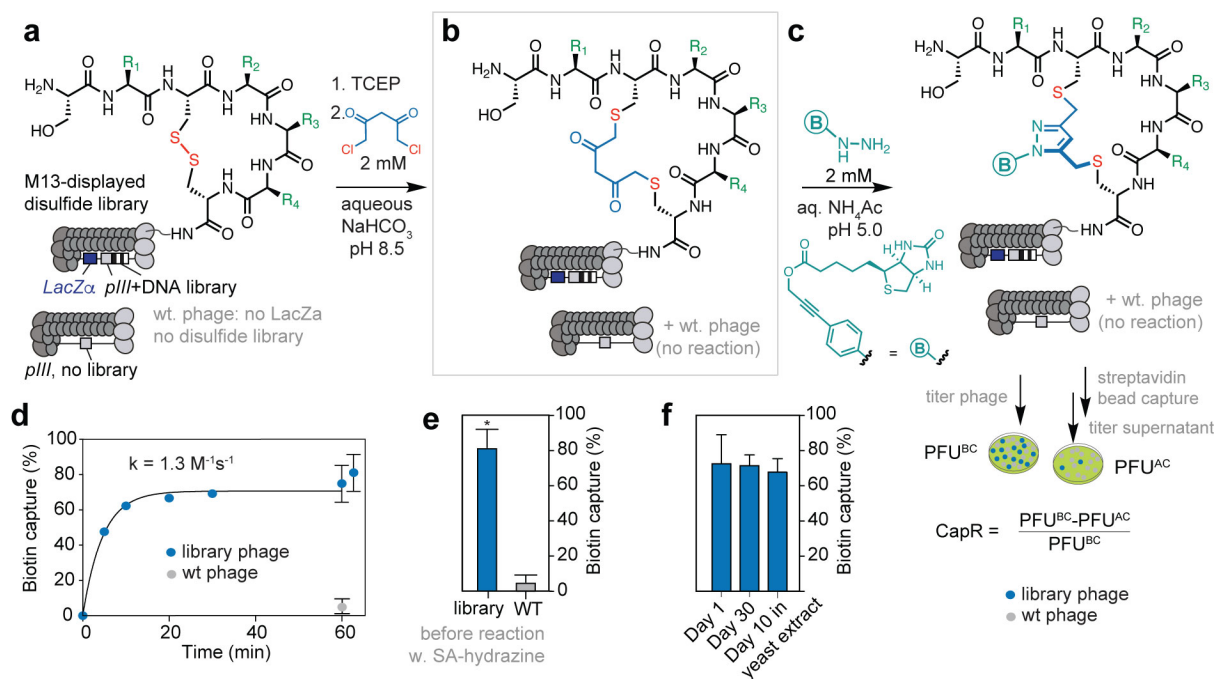


Figure 2. Macrocyclization of genetically-encoded phage libraries (a-b) M13-phage displayed disulfide library is reduced by TCEP to yield reactive thiols and then DPD to form 1,3-diketone bearing macrocyclic library. (c) The 1,3-diketone linchpin bearing macrocycles were further functionalized with hydrazines e.g. biotin hydrazine. A biotin capture experiment was used to measure the conversion of the 1,3-diketone. (d) The rate of conversion of the M13 library bearing 1,3-diketone was calculated using capture of biotinylated phage clones. (e) >80% of the library contains the 1,3-diketone modification. (f) Reactivity of DKMP handle was unchanged after storage in a buffer or in a protein and metabolite rich media (yeast extract).

To allow screening of the mixture of libraries modified by different unnatural fragment, we expressed chemically-identical SXCX₃C phage libraries, each containing a unique silent DNA barcode⁴⁵ and modified each library by a different pharmacophore (Fig 3, Fig. 4a). Silent-barcoded ^{SB2}SXCX₃C-DKPM library was modified by a biotin hydrazine forming biotinylated pyrazole-macrocyclic library

(Fig. 4a) in a 75% yield (Supplementary Fig S15). $^{SB3}SXCX_3C$ -DKPM library was modified by a *n*-decyl-hydrazine giving rise to “lipidated” pyrazole-macrocycles, $^{SB4}SXCX_3C$ -DKPM library was converted to a library of macrocycles with a sulfonamide warhead (Fig. 4a). Both reactions occurred in 75-80% modification yield and to confirm this yield, we employed previously developed “pulse-chase” biotin capture (Fig. 3, Supplementary Fig. S15).

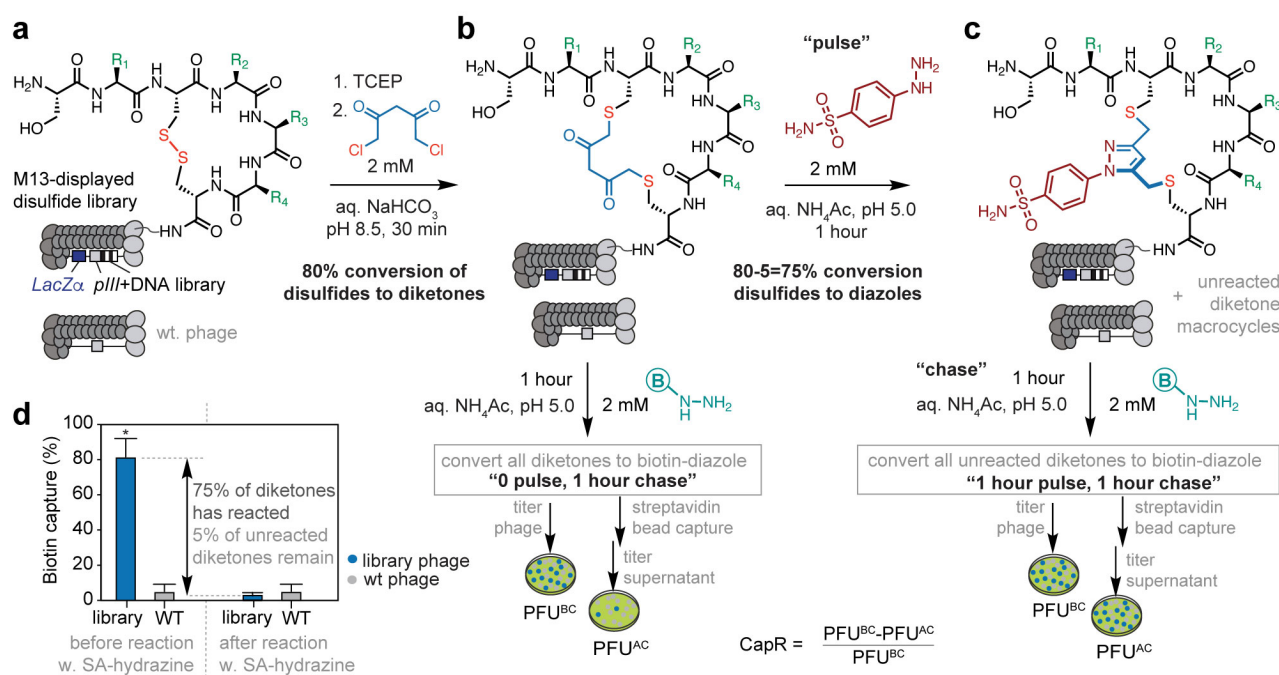


Figure 3. Pulse-chase monitoring of reactions in DKMP libraries. (a) M13-phage displayed disulfide library is reacted with TCEP and then DPD to form DKMP library. (b) Reaction of DKMP library with biotin hydrazine for 1 hour detects that 80% of clones contain reactive 1,3-diketones groups. (c) Reaction of DKMP library with 4-hydrazine benzenesulfonamide (4HBS) for 1 hour (“pulse”) followed by biotin hydrazine for 1 hour (“chase”) allows detection of residual 1,3 diketone groups that were not consumed by 4HBS. (b) From 80% library with reactive 1,3 diketones, <5% were not consumed after 1 hour reaction with 4HBS and can be biotinylated; these results indicate that 75% of the library were modified by 4HBS. Note that wild type phage present in the same solution is not biotinylated in any of the experiments indicating that only the disulfide library is modified with diketones and/or hydrazine derivatives

In short, reaction between DKMP library and biotin hydrazine confirmed that 80% of the clones in $^{SB4}SXCX_3C$ -DKPM library contained 1,3-diketone group (Fig. 3a-b). Reaction between $^{SB4}SXCX_3C$ -

DKPM and 4-hydrazine benzenesulfonamide for an hour (“pulse”) followed by addition of biotin hydrazine (“chase”) produced <5% of the biotinylated clones (Fig 3c). This observation confirmed that in 75% of the library the 1,3 diketone groups has been converted to benzenesulfonamide-pyrazole. Analogous pulse-chase confirmed 80% lipidation of ^{SB3}SXCX₃C-DKPM library (Supplementary Fig. S15).

Next we demonstrated the utility of late-stage modified libraries in discovery of macrocyclic ligands for protein receptors. We mixed the four libraries with DNA-encoded modifications (Fig. 4a) in a 1:1:1:1 ratio to produce a library of 4x160,000=640,000 macrocycles in which the identity of both peptide sequence and the unnatural chemotype can be decoded by simple DNA sequencing. We performed 4-6 parallel instances of one round of panning of this mixed library against bovine carbonic anhydrase (BCA) (Fig. 4b), streptavidin (Supplementary Fig S16) and bovine serum albumin (BSA) and analyzed each output by deep sequencing (Fig. 4c). The sub-library with the specific modification was enriched in a screen against a cognate target: ^{SB4}SXCX₃C-sulfonamide was enriched in panning against BCA and ^{SB2}SXCX₃C-biotin was enriched in panning against streptavidin when compared to the input (Fig. 4c). Interestingly, we also observed a modest enrichment of ^{SB4}SXCX₃C-sulfonamide and ^{SB3}SXCX₃C-n-decyl sub-libraries in panning on BSA-coated wells in line with known affinity of both chemotypes to albumin.

To analyze the deep-sequencing data, we employed previously published⁴⁶ Bioconductor EdgeR differential enrichment (DE) analysis^{47,48} with negative binomial model, Trimmed Mean of M-values (TMM) normalization⁴⁹, and Benjamini–Hochberg (BH)⁵⁰ correction to control the false discovery rate (FDR) at $\alpha = 0.05$. DE analysis of the output from BCA panning identified the 55 families of 688 peptide sequences that were significantly ($p < 0.05$) enriched in panning to BCA when compared to input and panning against BSA control (Supplementary Fig. S17; Supplementary file S1).

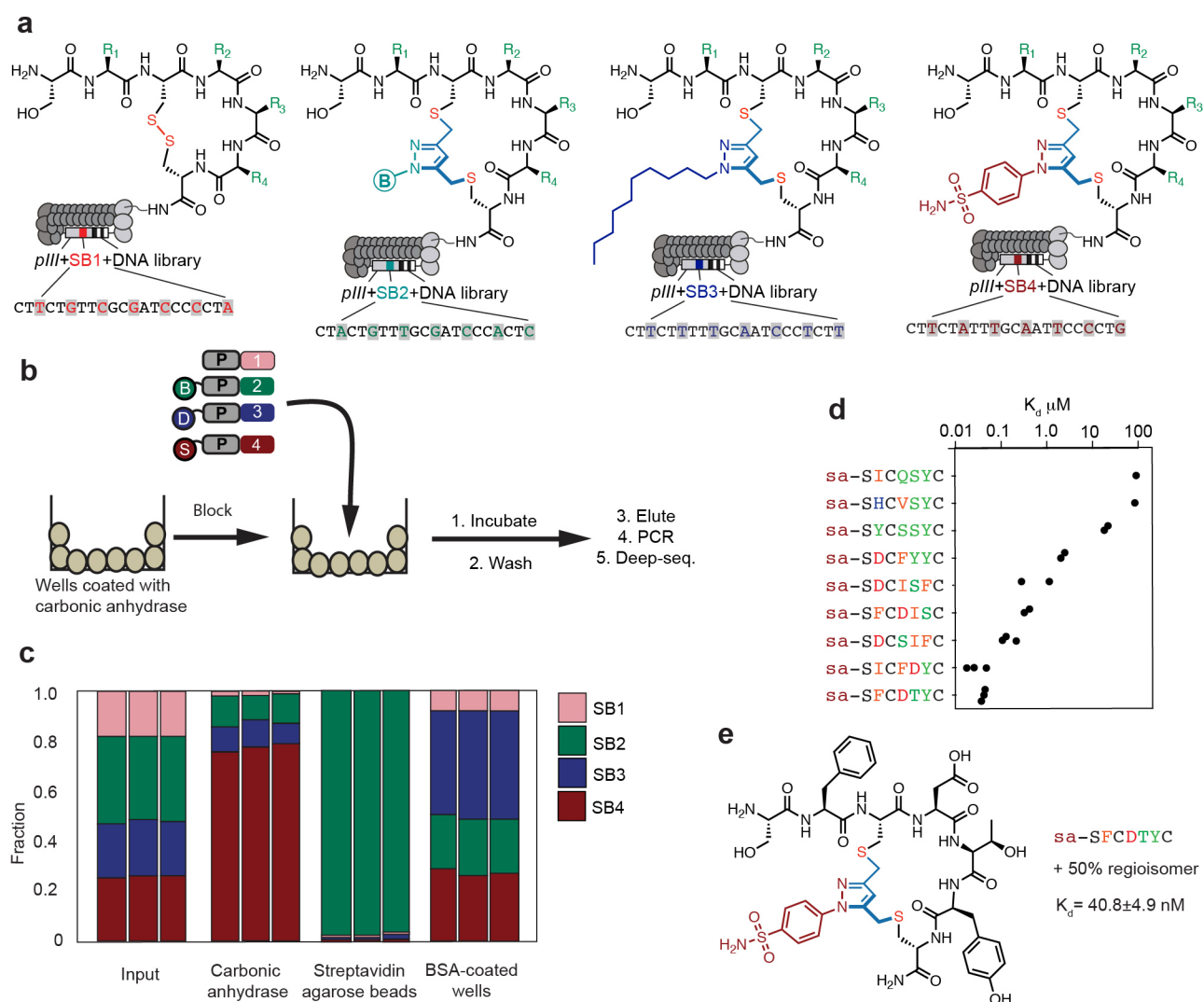


Fig. 4: Selection of ligands from functionalized macrocyclic libraries. **a**, Phage-displayed libraries of macrocycles with four distinct modifications and sequences of “silent barcodes” that encode these modifications. **b**, A mixture of modified libraries was incubated with immobilized protein, followed by washing steps and acid elution. Eluted phage was PCR-amplified. The amplicons were sequenced by Illumina. **c**, Abundance of silent barcodes in sequencing of libraries eluted from streptavidin, carbonic anhydrase, or BSA-coated wells showed an enrichment of specific barcodes that encode the fragments (**a**) that bind to the corresponding protein target. Three bars describe the results of sequencing of three independent panning experiment. **d**, K_d of binding to BCA for nine peptide macrocycles selected against BCA, determined with ITC. **e**, Structure of macrocyclic peptide displaying the lowest K_d value (40.8 ± 4.9 nM). Each macrocycle was synthesized as a 1:1 mixture of two regioisomers according to LC trace of LCMS.

To confirm that the sequences identified by DE-analysis are the binders for targets of interest, we synthesized nine out of 688 predicted sequences and modified peptides with DPD and sulfonamide hydrazines to form peptide macrocycles with grafted sulfonamides (Fig. 4d-e). Isothermal titration calorimetry (ITC) determined the binding constant K_d for the sulfonamide hydrazine to be $>50 \mu\text{M}$. K_d of sulfonamide macrocycles produced from a random SICQSYC sequence was $>100 \mu\text{M}$. From eight peptides predicted by DE analysis, six macrocycles exhibited 10-1000 fold enhancement in affinity (Fig. 4d); two most potent macrocycles interacted with BCA with $K_d = 40.8 \pm 4.9 \text{ nM}$ (SFCDTYC) and $41.1 \pm 26.1 \text{ nM}$ (SICFDYC). These results confirm that late-stage modified macrocycle libraries with multiple diverse unnatural fragments can be used for successful discovery of fragment-containing macrocycles with low-nanomolar potency.

Discussion

Broad substrate scope of Knorr pyrazole synthesis makes it an attractive strategy for diversification of macrocycles with built in 1,3-diketones using a large range of commercially available alkyl and aryl-hydrazine warheads. So far, we identified only two substrates with suboptimal reactivity: Reaction between 1,3-diketone-macrocycle and N-acyl hydrazines and benzenesulfonohydrazide in water was slow or incomplete. Even when formed, N-acyl 1,2-pyrazoles can be readily cleaved by thiols⁵¹ and other biological nucleophiles; such cleavage makes N-acyl 1,2-pyrazoles not suitable for stable grafting of functionalities onto peptides. Further substrate scope profiling may uncover further limitations; however, we foresee few problems in reactions that employ simple alkyl and aryl hydrazine fragments.

A limitation of Knorr-pyrazole ligation is formation of two regioisomers of 1,2-diazole: we observed formation of 1:1 mixtures of two isomers in many LC-traces; interestingly, this ratio was skewed towards one isomer in reaction with perfluorophenyl hydrazine (Supplementary Fig. S10). The reason for this regioinduction is presently not clear. Formation of isomers is a trait of many contemporary

modifications of polypeptides: For example, reactions between Cys and maleimide yields mixtures of stereoisomers. However, this reaction is used in manufacturing of FDA-approved antibody-drug conjugates (ADC) such as Ketruda™, Trodelvy™, Enhertu™, Polivy™, Adcetris™, and Padcev™. A ligation to dehydroalanine on polypeptides^{30,52,53} yields two diastereomers and has been successfully translated to manufacturing of ADCs²⁹. Pictet-Spengler⁵⁴ and Hydrazino-iso-Pictet-Spengler (HIPS) that yield diastereomeric linkages are employed in manufacturing of TRPH-222 ADC, which is currently in a Phase 1 clinical trial. Many reagents for modification of protein via strain-promoted cycloadditions and inverse demand Diels-Alder reactions form isomeric linkages. Other bioorthogonal ligations of aldehydes⁵⁵ to hydrazines, oximes⁴¹, 2-amino benzamidoximes⁴⁰ and Wittig ylides⁴² form E and Z products. In mRNA- and phage-displayed libraries, reactions that yield a mixture of stereo^{7,30} or regioisomers^{7,56,57} have been employed as well. In such mixed-isomer libraries, activity can be attributed to one synthetic isomer post-discovery^{7,56,57}. Formation of two regioisomers, thus, is not an impediment to a GE-discovery process: In this report, the macrocyclic peptides with phenyl sulfonamide fragments were discovered and synthesized post-discovery as mixtures of two regioisomers. It is likely that one isomer has higher activity than the other, but we did not attempt to measure the activity of separated isomers.

Late stage modification of GE-macrocycles described in this report makes important advances to the pioneering report of GE-FBD by Roberts⁵⁸ and Dwyer et al.⁵⁹ Simplicity and robustness of chemical modification makes it simple to adapt it to many existing phage and mRNA-displayed cysteine-containing phage libraries^{30,60}. These libraries when modified with DPD should yield shelf-stable, divergent macrocyclic precursors to GE-macrocyclic libraries with unnatural fragments. It is possible to produce GE-macrocyclic libraries with unnatural fragment by direct translation *via* unnatural amino acid mutagenesis⁶¹⁻⁶³, metabolic suppression^{64,65} and flexigyme technology^{11,22}. However, the unique advantage of late stage chemical modifications is the ability to introduce large functionalities such as fluorophores, metal chelators (Fig. 1i, j), or complex glycans^{30,66,67} (not shown in this report): such groups

might not be possible to introduce *via* translational machinery. Selection of peptides pre-modified with fluorescent probes or metal-chelating probes (Fig. 1i, j) offers an interesting opportunity to discover peptides for imaging while minimizing the commonly observed decrease in potency or specificity of peptide due to conjugation of imaging probes. Similarly, GE-screening of prospectively pegylated or lipidated macrocycles (Fig. 1k) can provide the pegylated and lipidated ligands and reduce the number of steps for optimization of leads. The “late” nature of 1,3-diketone ligation makes it an interesting candidate for ligation of reactive warheads for genetically-encoded discovery of covalent²⁶ or reversible covalent inhibitors⁶. Combination of this modification with silent encoding opens new opportunities in encoding of diverse number of macrocycles with pharmacophores, warheads and functionalities that would be difficult to introduce by other methods.

Materials:

Detailed **Synthetic Methods 1.1-1.4**, **Biochemical Methods 2.0-2.5**, **Data Processing Methods 3.1-3.2**, LC-MS traces of reaction kinetics (**Supplementary Fig. S3-S13**), high resolution mass spectrometry (**Table S3**) and LC-MS characterization of the DKPM and pyrazole macrocycles (**Supplementary Fig. S18-S37**), NMR spectra for synthetic compounds and representative 1D and 2D-NMR spectra of DKPM (compound 3b) and perfluorophenyl pyrazole macrocycle are available as part of the Supplementary Information document.

Supplementary information

Supplemental Figures S1-S37, Supplemental Tables S1-S3; detailed synthetic methods, biochemical methods describing the synthesis and selection of phage libraries, isothermal titration calorimetry (ITC) assay or protein-ligand binding, and data processing methods describing the analysis of the DNA sequencing data, statistical methods.

Source data: submitted as “data.zip” contain files describing (i) "Kinetics Matlab" directory with raw data used to monitor the kinetics of reactions and MatLab scripts for curve fit, (ii) "Sequence files" directory with *.txt files describing the raw deep-sequencing data; *.xlsx tables describing the silent barcoding; *.xlsx tables describing the differential enrichment (DE) analysis; and Supplementary Files S1 and S2 describing the output of differential enrichment analysis and clustering. (iii) Titters.xlsx describing the phage titers (PFU) for all experiments described in this manuscript

Data Availability: All raw deep-sequencing data is publicly available on <http://48hdcloud.ca/> with data-specific URL listed in Supplementary Table S2. MatLab, Python and R scripts used for analysis of deep-seq data have been deposited to <https://github.com/derdalab/diketone>.

Competing Interests:

R.D. is the C.E.O. and a shareholder of 48Hour Discovery Inc., the company that licensed the patent application describing silent encoding and chemical modification technologies.

Contributions:

A.E. performed synthesis of chemical and biochemical reagents, modification, selection and analysis of the phage display libraries, validation of the hits. N.J.B. expressed silently barcoded phage libraries. L.S. and J.Y. performed synthesis of DKMP macrocyclic peptides and pyrazole macorycles. R.M. performed synthesis of selected hydrazine reagents and macrocyclic peptides; A.B and F.W. contributed the critical reagents. R.D., and A.E. wrote the manuscript, edited the final manuscript and contributed intellectual and strategic input. All authors approved the final manuscript.

Acknowledgments:

We thank Dr. Ryan T. McKay at the University of Alberta NMR spectrometry facility, Dr. Randy Whittal and Béla Reiz at the mass spectrometry facility, and Dr. Michael J. Ferguson at the X-ray

crystallography laboratory (Chemistry Department). We thank Dr. Jessica Cao at 48Hour Discovery Inc for assistance with analysis of the sequencing data. The authors acknowledge funding from NSERC (RGPIN-2016-402511 to R.D.) and NSERC Accelerator Supplement (to R.D.). Infrastructure support was provided by CFI New Leader Opportunity (to R.D.). LS acknowledges summer research fellowship from GlycoNet.

References.

- 1 Cohen, D. T. *et al.* A chemoselective strategy for late-stage functionalization of complex small molecules with polypeptides and proteins. *Nat Chem* **11**, 78-85, doi:10.1038/s41557-018-0154-0 (2019).
- 2 Leroux, M. *et al.* Late-Stage Functionalization of Peptides and Cyclopeptides Using Organozinc Reagents. *Angew. Chem. Int. Ed. Engl.* **58**, 8231-8234, doi:10.1002/anie.201902454 (2019).
- 3 Noisier, A. F. M. *et al.* Late-Stage Functionalization of Histidine in Unprotected Peptides. *Angew. Chem. Int. Ed. Engl.* **58**, 19096-19102, doi:10.1002/anie.201910888 (2019).
- 4 Tang, J. *et al.* Peptide-guided functionalization and macrocyclization of bioactive peptidosulfonamides by Pd(II)-catalyzed late-stage C-H activation. *Nat Commun* **9**, 3383, doi:10.1038/s41467-018-05440-w (2018).
- 5 Heinis, C., Rutherford, T., Freund, S. & Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **5**, 502-507, doi:10.1038/nchembio.184 (2009).
- 6 McCarthy, K. A. *et al.* Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **140**, 6137-6145, doi:10.1021/jacs.8b02461 (2018).
- 7 Ng, S. *et al.* Genetically encoded fragment-based discovery of glycopeptide ligands for carbohydrate-binding proteins. *J. Am. Chem. Soc.* **137**, 5248-5251, doi:10.1021/ja511237n (2015).
- 8 Ven Chang, I., Tsutsumi, H. & Mihara, H. Screening for concanavalin A binders from a mannose-modified alpha-helix peptide phage library. *Mol Biosyst* **13**, 2222-2225, doi:10.1039/c7mb00495h (2017).
- 9 Vinogradov, A. A., Yin, Y. & Suga, H. Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* **141**, 4167-4181, doi:10.1021/jacs.8b13178 (2019).
- 10 Derda, R. & Ng, S. Genetically encoded fragment-based discovery. *Curr. Opin. Chem. Biol.* **50**, 128-137, doi:10.1016/j.cbpa.2019.03.014 (2019).
- 11 Huang, Y., Wiedmann, M. M. & Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem Rev* **119**, 10360-10391, doi:10.1021/acs.chemrev.8b00430 (2019).
- 12 Richardson, S. L., Dods, K. K., Abrigo, N. A., Iqbal, E. S. & Hartman, M. C. In vitro genetic code reprogramming and expansion to study protein function and discover macrocyclic peptide ligands. *Curr. Opin. Chem. Biol.* **46**, 172-179, doi:10.1016/j.cbpa.2018.07.013 (2018).
- 13 Erlanson, D. A. Introduction to fragment-based drug discovery. *Top Curr Chem* **317**, 1-32, doi:10.1007/128_2011_180 (2012).
- 14 Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat Chem* **1**, 187-192, doi:10.1038/nchem.217 (2009).
- 15 Scott, D. E., Coyne, A. G., Hudson, S. A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**, 4990-5003, doi:10.1021/bi3005126 (2012).

- 16 Bollag, G. *et al.* Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **467**, 596-599, doi:10.1038/nature09454 (2010).
- 17 Perera, T. P. S. *et al.* Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol Cancer Ther* **16**, 1010-1020, doi:10.1158/1535-7163.MCT-16-0589 (2017).
- 18 Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat Med* **19**, 202-208, doi:10.1038/nm.3048 (2013).
- 19 Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc Natl Acad Sci U S A* **105**, 3041-3046, doi:10.1073/pnas.0711741105 (2008).
- 20 Franzini, R. M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries. *Acc. Chem. Res.* **47**, 1247-1255, doi:10.1021/ar400284t (2014).
- 21 Li, Y. *et al.* Versatile protein recognition by the encoded display of multiple chemical elements on a constant macrocyclic scaffold. *Nat Chem* **10**, 441-448, doi:10.1038/s41557-018-0017-8 (2018).
- 22 Passioura, T., Katoh, T., Goto, Y. & Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu Rev Biochem* **83**, 727-752, doi:10.1146/annurev-biochem-060713-035456 (2014).
- 23 Hetrick, K. J., Walker, M. C. & van der Donk, W. A. Development and Application of Yeast and Phage Display of Diverse Lanthipeptides. *ACS Cent Sci* **4**, 458-467, doi:10.1021/acscentsci.7b00581 (2018).
- 24 Assem, N., Ferreira, D. J., Wolan, D. W. & Dawson, P. E. Acetone-Linked Peptides: A Convergent Approach for Peptide Macrocyclization and Labeling. *Angew. Chem. Int. Ed. Engl.* **54**, 8665-8668, doi:10.1002/anie.201502607 (2015).
- 25 Ng, S. & Derda, R. Phage-displayed macrocyclic glycopeptide libraries. *Org Biomol Chem* **14**, 5539-5545, doi:10.1039/c5ob02646f (2016).
- 26 Chen, S. & Bogyo, M. A phage display approach to identify highly selective covalent binders. *bioRxiv*, 791533, doi:10.1101/791533 (2019).
- 27 Uematsu, S., Tabuchi, Y., Ito, Y. & Taki, M. Combinatorially Screened Peptide as Targeted Covalent Binder: Alteration of Bait-Conjugated Peptide to Reactive Modifier. *Bioconjug Chem* **29**, 1866-1871, doi:10.1021/acs.bioconjchem.8b00301 (2018).
- 28 Ernst, C., Sindlinger, J., Schwarzer, D., Koch, P. & Boeckler, F. M. The Symmetric Tetravalent Sulfhydryl-Specific Linker NATBA Facilitates a Combinatorial "Tool Kit" Strategy for Phage Display-Based Selection of Functionalized Bicyclic Peptides. *ACS omega* **3**, 12361-12368, doi:10.1021/acsomega.8b01814 (2018).
- 29 Dadova, J., Galan, S. R. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr Opin Chem Biol* **46**, 71-81, doi:10.1016/j.cbpa.2018.05.022 (2018).
- 30 Jongkees, S. A. K., Umemoto, S. & Suga, H. Linker-free incorporation of carbohydrates into in vitro displayed macrocyclic peptides. *Chem Sci* **8**, 1474-1481, doi:10.1039/c6sc04381j (2017).
- 31 Barbas, C. F., 3rd *et al.* Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science* **278**, 2085-2092, doi:10.1126/science.278.5346.2085 (1997).

- 32 Rader, C. *et al.* A humanized aldolase antibody for selective chemotherapy and adaptor immunotherapy. *J. Mol. Biol.* **332**, 889-899, doi:10.1016/s0022-2836(03)00992-6 (2003).
- 33 Tanaka, F. *et al.* Development of a small peptide tag for covalent labeling of proteins. *Bioconjug Chem* **18**, 1318-1324, doi:10.1021/bc070080x (2007).
- 34 Gupta, V. & Carroll, K. S. Profiling the Reactivity of Cyclic C-Nucleophiles towards Electrophilic Sulfur in Cysteine Sulfenic Acid. *Chem Sci* **7**, 400-415, doi:10.1039/C5SC02569A (2016).
- 35 Gupta, V. & Carroll, K. S. Rational design of reversible and irreversible cysteine sulfenic acid-targeted linear C-nucleophiles. *Chem Commun (Camb)* **52**, 3414-3417, doi:10.1039/c6cc00228e (2016).
- 36 Gupta, V., Paritala, H. & Carroll, K. S. Reactivity, Selectivity, and Stability in Sulfenic Acid Detection: A Comparative Study of Nucleophilic and Electrophilic Probes. *Bioconjugate Chemistry* **27**, 1411-1418, doi:10.1021/acs.bioconjchem.6b00181 (2016).
- 37 Knorr, L. Einwirkung von Acetessigester auf Phenylhydrazin. *Berichte der deutschen chemischen Gesellschaft* **16**, 2597-2599, doi:10.1002/cber.188301602194 (1883).
- 38 Rychnovsky, S. D., Griesgraber, G., Zeller, S. & Skalitzky, D. J. Optically pure 1,3-diols from (2R,4R)- and (2S,4S)-1,2:4,5-diepoxy pentane. *The Journal of Organic Chemistry* **56**, 5161-5169, doi:10.1021/jo00017a032 (1991).
- 39 Kalhor-Monfared, S. *et al.* Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chem Sci* **7**, 3785-3790, doi:10.1039/c5sc03856a (2016).
- 40 Kitov, P. I., Vinals, D. F., Ng, S., Tjhung, K. F. & Derda, R. Rapid, hydrolytically stable modification of aldehyde-terminated proteins and phage libraries. *J. Am. Chem. Soc.* **136**, 8149-8152, doi:10.1021/ja5023909 (2014).
- 41 Ng, S., Jafari, M. R., Matochko, W. L. & Derda, R. Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage. *ACS Chem Biol* **7**, 1482-1487, doi:10.1021/cb300187t (2012).
- 42 Triana, V. & Derda, R. Tandem Wittig/Diels-Alder diversification of genetically encoded peptide libraries. *Org Biomol Chem* **15**, 7869-7877, doi:10.1039/c7ob01635b (2017).
- 43 Kimball, R. F. The mutagenicity of hydrazine and some of its derivatives. *Mutat Res* **39**, 111-126, doi:10.1016/0165-1110(77)90018-5 (1977).
- 44 He, B. *et al.* Compositional Bias in Naive and Chemically-modified Phage-Displayed Libraries uncovered by Paired-end Deep Sequencing. *Sci. Rep.* **8**, 1214, doi:10.1038/s41598-018-19439-2 (2018).
- 45 Tjhung, K. F. *et al.* Silent Encoding of Chemical Post-Translational Modifications in Phage-Displayed Libraries. *J Am Chem Soc* **138**, 32-35, doi:10.1021/jacs.5b10390 (2016).
- 46 Matochko, W. L. & Derda, R. Next-generation sequencing of phage-displayed peptide libraries. *Methods Mol Biol* **1248**, 249-266, doi:10.1007/978-1-4939-2020-4_17 (2015).
- 47 Robinson, M. D. & Smyth, G. K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881-2887, doi:10.1093/bioinformatics/btm453 (2007).
- 48 Robinson, M. D. & Smyth, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321-332, doi:10.1093/biostatistics/kxm030 (2008).
- 49 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25, doi:ARTN R25

- 10.1186/gb-2010-11-3-r25 (2010).
- 50 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300 (1995).
- 51 Flood, D. T. *et al.* Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for use in Native Chemical Ligation. *Angewandte Chemie International Edition* **57**, 11634-11639, doi:10.1002/anie.201805191 (2018).
- 52 Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nat Commun* **5**, 4388, doi:10.1038/ncomms5388 (2014).
- 53 Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354**, doi:10.1126/science.aag1465 (2016).
- 54 Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc Natl Acad Sci U S A* **110**, 46-51, doi:10.1073/pnas.1213186110 (2013).
- 55 Spears, R. J. & Fascione, M. A. Site-selective incorporation and ligation of protein aldehydes. *Org Biomol Chem* **14**, 7622-7638, doi:10.1039/c6ob00778c (2016).
- 56 Kale, S. S. *et al.* Cyclization of peptides with two chemical bridges affords large scaffold diversities. *Nat Chem* **10**, 715-723, doi:10.1038/s41557-018-0042-7 (2018).
- 57 Kong, X. D. *et al.* De novo development of proteolytically resistant therapeutic peptides for oral administration. *Nat Biomed Eng* **4**, 560-571, doi:10.1038/s41551-020-0556-3 (2020).
- 58 Li, S. & Roberts, R. W. A novel strategy for in vitro selection of peptide-drug conjugates. *Chem. Biol.* **10**, 233-239, doi:10.1016/s1074-5521(03)00047-4 (2003).
- 59 Dwyer, M. A., Lu, W., Dwyer, J. J. & Kossiakoff, A. A. Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity. *Chem. Biol.* **7**, 263-274, doi:10.1016/s1074-5521(00)00102-2 (2000).
- 60 Hacker, D. E. *et al.* Direct, Competitive Comparison of Linear, Monocyclic, and Bicyclic Libraries Using mRNA Display. *ACS Comb Sci* **22**, 306-310, doi:10.1021/acscombsci.0c00016 (2020).
- 61 Oller-Salvia, B. & Chin, J. W. Efficient Phage Display with Multiple Distinct Non-Canonical Amino Acids Using Orthogonal Ribosome-Mediated Genetic Code Expansion. *Angew. Chem. Int. Ed. Engl.* **58**, 10844-10848, doi:10.1002/anie.201902658 (2019).
- 62 Owens, A. E., Iannuzzelli, J. A., Gu, Y. & Fasan, R. MOrPH-PhD: An Integrated Phage Display Platform for the Discovery of Functional Genetically Encoded Peptide Macrocycles. *ACS Cent Sci* **6**, 368-381, doi:10.1021/acscentsci.9b00927 (2020).
- 63 Wang, X. S. *et al.* A Genetically Encoded, Phage-Displayed Cyclic-Peptide Library. *Angew. Chem. Int. Ed. Engl.* **58**, 15904-15909, doi:10.1002/anie.201908713 (2019).
- 64 Van Deventer, J. A., Yuet, K. P., Yoo, T. H. & Tirrell, D. A. Cell surface display yields evolvable, clickable antibody fragments. *ChemBioChem* **15**, 1777-1781, doi:10.1002/cbic.201402184 (2014).
- 65 Stieglitz, J. T., Kehoe, H. P., Lei, M. & Van Deventer, J. A. A Robust and Quantitative Reporter System To Evaluate Noncanonical Amino Acid Incorporation in Yeast. *ACS Synth Biol* **7**, 2256-2269, doi:10.1021/acssynbio.8b00260 (2018).
- 66 Chou, Y. *et al.* Genetically-encoded fragment-based discovery (GE-FBD) of glycopeptide ligands with differential selectivity for antibodies related to mycobacterial infections. *Org Biomol Chem* **16**, 223-227, doi:10.1039/c7ob02783d (2018).

67 Sojitra, M. *et al.* Genetically Encoded, Multivalent Liquid Glycan Array (LiGA).
bioRxiv, 2020.2003.2024.997536, doi:10.1101/2020.03.24.997536 (2020).