Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Genetically encoded fragment-based discovery

Ratmir Derda and Simon Ng[1]

This opinion describes recent advances of molecular discovery technology dubbed Genetically Encoded Fragment-Based Discovery (GE-FBD). GE-FBD starts from a known ligand or 'fragment' that binds to a desired target weakly and often with low specificity. Covalent incorporation of fragment into a diverse, genetically encoded library of peptides yields a library of peptide–fragment combinations. Selection from such a library has a high likelihood to identify ligands, in which the peptides bind to distinct adjacent pockets of the target in synergy with the fragment and exhibits enhanced affinity and specificity when compared to the fragment itself. GE-FBD could employ fragments that bind non-covalently as well as reversible covalent warheads. The key advances in GE-FBD include (i) synthetic chemistry that enables incorporation of diverse fragments into both linear and cyclic peptide libraries; (ii) quantification of multi-step modifications in million-to-billion library members, (iii) and chemical transformations that permit incorporation of fragments with concurrent topological change from linear to macrocyclic topologies.

Address
Department of Chemistry, University of Alberta, 11227 Saskatchewan Dr., Edmonton, AB T6G 2G2, Canada

Corresponding author: Derda, Ratmir (ratmir@ualberta.ca)
[1] Current address: MSD, Singapore.

Current Opinion in Chemical Biology 2019, **50**:128–137

This review comes from a themed issue on **Next generation therapeutics**

Edited by **Yimon Aye** and **Paul J Hergenrother**

For a complete overview see the Issue and the Editorial

Available online 30th April 2019

https://doi.org/10.1016/j.cbpa.2019.03.014

Introduction

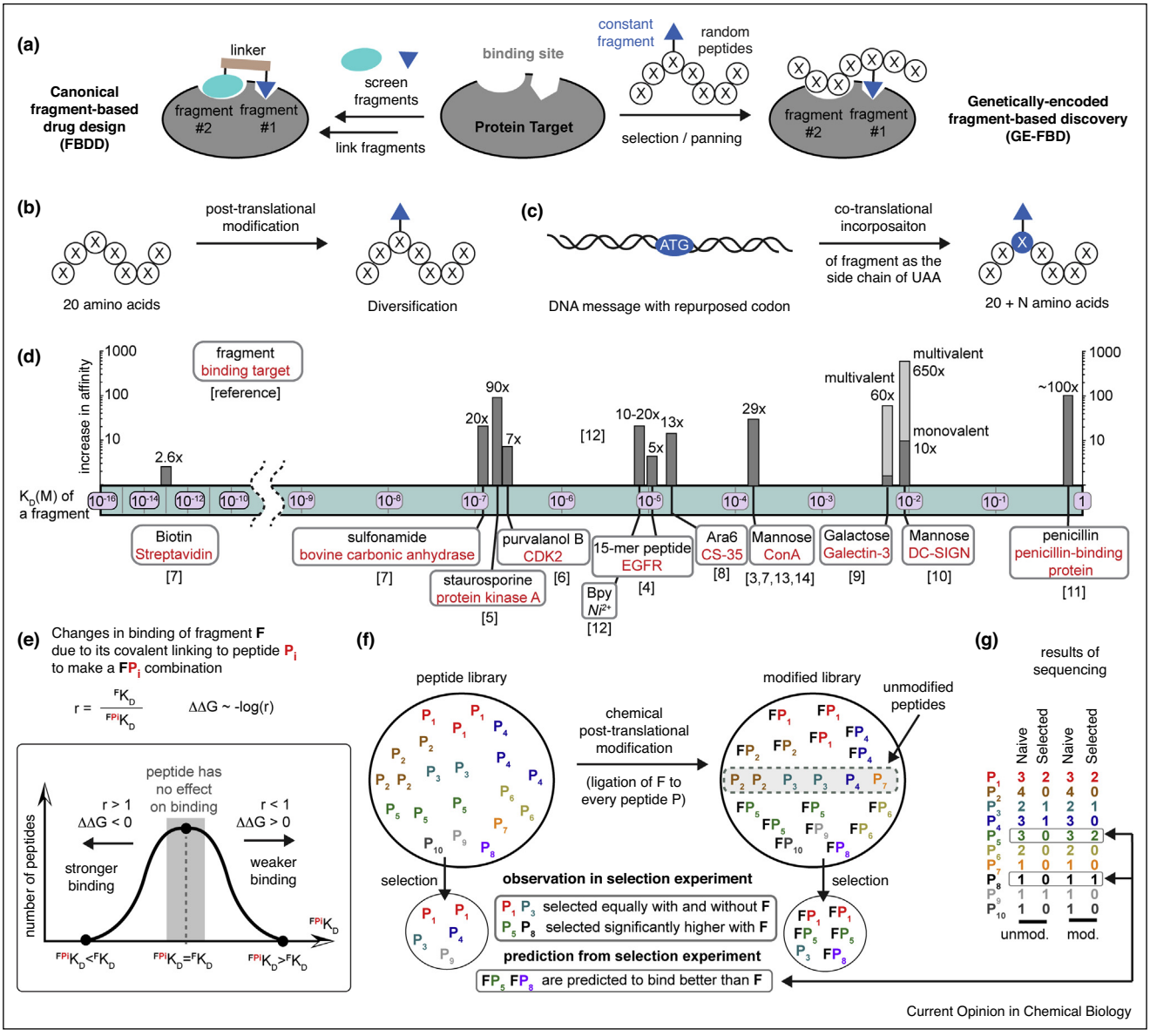
Fragment-based drug design (FBDD) is an important process in medicinal chemistry that gives rise to specific, and selective binding ligands to a protein of interest by linking weak and often promiscuous synthetic molecules termed 'molecular fragments' (Figure 1a) [1,2]. This review focuses on technology that combines the basic principles of FBDD with the power of genetically encoded (GE) discovery of peptides or macrocyclic peptide ligands [3••]. Covalent incorporation of a fragment **F** into a GE-library of peptides, denoted as {**P**} yields a new library, denoted as {**FP**} in which *most* members contain the fragment of interest. This opinion focuses on chemical post-translational incorporation of fragments (Figure 1b). Alternatively, {**FP**} libraries can be built from unnatural amino acids (UAA) that contain fragments as part of their side-chain (Figure 1c). Selection from {**FP**} libraries has been shown to identify ligands in which the fragment covalently linked to peptide segment bind the target with improved affinity and specificity when compared to the original fragment (Figure 1d). Improvement can be defined as $r = {}^{\mathbf{F}}K_D/{}^{\mathbf{FP}}K_D > 1$ or $\Delta\Delta G_{bind} \sim -\log(r) < 0$, where ${}^{\mathbf{F}}K_D$ is the affinity of the fragment towards the target of interest (Figure 1e). Figure 1d summarizes up-to-date outcomes from published GE-FBD reports and shows that $r = 10$–100 can be found across diverse fragment with ${}^{\mathbf{F}}K_D$ ranging from 100 nM to 1 mM [3••,4–6,7••,8,9•,10,11•,12,13•,14]. GE-FBD was also reported to be successful with reversible covalent fragments.

Historically, peptides are considered as the suboptimal drug modality due to their poor stability and pharmacological properties. However, peptides are increasingly recognized as potential starting point for lead optimization through systematic medicinal chemistry studies. Chemical modification of peptides—cyclization, *N*-methylation, capping of N-terminus and C-terminus and substitution of L-amino acid with unnatural or D-amino acid—can yield peptide derivatives with excellent stability and cell permeability that are suitable for pre-clinical and 'investigational new drug' (IND) studies [15,16]. These efforts have translated to a stapled peptide (ALRN-6924) that has entered the clinical trial (NCT02264613). The encouraging developments show the potential of the peptide modality, especially for targeting protein–protein interactions that are historically intractable to small-molecule approach [17,18].

Biophysical considerations of GE-FDB

GE-FBD offers a unique opportunity to study genetically encoded ligand discovery. Unlike traditional selection from a random library of molecules where the vast majority of ligands exhibit non-detectable binding for the target, members of the {**FP**} library contain the fragment that has a measurable affinity ${}^{\mathbf{F}}K_D$ towards the target. Specifically, {**FP**} library can be demarcated into members in which the affinity of i^{th} peptide–fragment combination (${}^{\mathbf{FPi}}K_D$) are stronger ($\Delta\Delta G_{bind} < 0$), similar ($\Delta\Delta G_{bind} \sim 0$) or weaker ($\Delta\Delta G_{bind} > 0$) when compared to the fragment **F** (Figure 1e). Historically, GE-FBD employed multiple rounds of selection and Sanger sequencing to discover singular ligands with favourable $\Delta\Delta G_{bind}$. Use of deep sequencing permitted not only reducing number of the rounds of selection but also

Figure 1



General concepts of GE-FBD. **(a)** canonical FBD versus GE-FBD. **(b)** Nature relies on PTM to increase functional diversity using 20 amino acids. **(c)** Incorporation of unnatural amino acids. **(d)** Affinities of fragments used in GE-FBD and the reported increase in affinity after GE-FBD. **(e)** Peptide–fragment combination (PF) that have similar, stronger or weaker affinity when compared to original fragment F. Ligands in a {**PF**} library of modified peptides, thus, exhibit a distribution of binding affinities; GE-FBD strives to identify the lowest $^{FPi}K_D$, but understanding the entire distribution can further illuminate structure–activity relationship (S.A.R). **(f–g)** Multiset representation of GE-FBD: chemical modification of a peptide library (multiset {**P**}) creates a {**FP**} library, in which *most* peptides bear the fragment and some remain unmodified. Differential enrichment analysis of selections from {**P**} and {**FP**} libraries can identify peptide P_i that yield FP_i combination that bind with higher affinity than parental fragment [3••].

changing the analysis approach. Instead of consecutive multi-round selection, the decisions can be made by the analysis of multiple parallel instances of selection of both modified {**FP**} and unmodified {**P**} libraries (Figure 1f) [3••,8,10]. The differential enrichment (DE) analysis identify the peptides that are reproducibly enriched in {**FP**} but not in {**P**} libraries. DE analysis calculates the ratio $R_i = \mathbf{FP_i/P_i}$ and p-value between replicates p_i = ttest ($\mathbf{FP_i}$,$\mathbf{P_i}$) [3••]. Including additional controls, such as panning of {**P-F**} library on irrelevant target or panning a library modified with irrelevant fragment {**P-f**} improved identification of the consensus motifs of peptide fragments that bind productively [3••]. For example, in parallel experiments, we selected {**FP**} library against the desired receptor concanavalin A (ConA) and promiscuous receptor BSA and defined $R_i = {}^{\mathbf{ConA}}\mathbf{FP_i}/{}^{\mathbf{BSA}}\mathbf{FP_i}$ and

p_i = ttest($^{ConA}\mathbf{FP_i}$, $^{BSA}\mathbf{FP_i}$) [3••]. DE analysis flagged putative 'non-specific binders' ligands that exhibited low R_i and high p_i exemplifying a statistically insignificant binding to ConA and BSA. Testing the same **{FP}** library against other controls and applying the same DE analysis further decreases the likelihood of identifying poly-specific binders. Searching databases of previous phage display and other selection experiments could serve as additional 'control'. Of course, regardless how many controls are performed, it is unrealistic to expect that a ligand identified by the screen will bind to only one protein. This specificity should be validated further, for example, through binding of synthesized ligands to large protein arrays [3].

Understanding how $\Delta\Delta G_{bind}$ is distributed in the context of the entire library (Figure 1e) can yield structure–activity relationships (SAR) and provide guidance for further optimization of ligands. Recent reports teach that learning the shape of the $\Delta\Delta G_{bind}$-distribution (Figure 1e) is possible [19••]. These reports systematically perturb protein–ligand interactions by methods similar to 'shotgun alanine scanning' [20] or 'deep mutational scanning' [21,22]. Starting from 10^3 to 10^4 DNA variants that comprehensively cover point mutants of one polypeptide, quantification of the enrichment of the selected variants by deep sequencing can relate changes in strength of protein–ligand interaction and systematic changes in amino acid composition [23•]. These methods have several important differences from traditional screens:

(1) Unlike traditional selections that focus only on the increase in recovery after the selection, 'deep mutational' approaches quantify both the increase and decrease in recovery. (2) Decrease the library size and one round of selection maximize the correlation between binding affinity and recovery in selection as measured by DNA-sequencing [23•,24••], and make it possible to measure Z-factors to estimate robustness and reproducibility of discovery [24••]. Smaller libraries of $<10^4$ members capture a broader dynamic range of enrichment and depletion from the sequencing of $\sim 10^7$ reads of DNA. (3) Regression models can be trained using a high-quality set of 10^3–10^4 peptides to predict the sequences of other peptides that minimize $\Delta\Delta G_{bind}$ for the target of interest while simultaneously increasing $\Delta\Delta G_{bind}$ for closely related targets [19••].

Chemical synthesis of libraries for GE-FBD

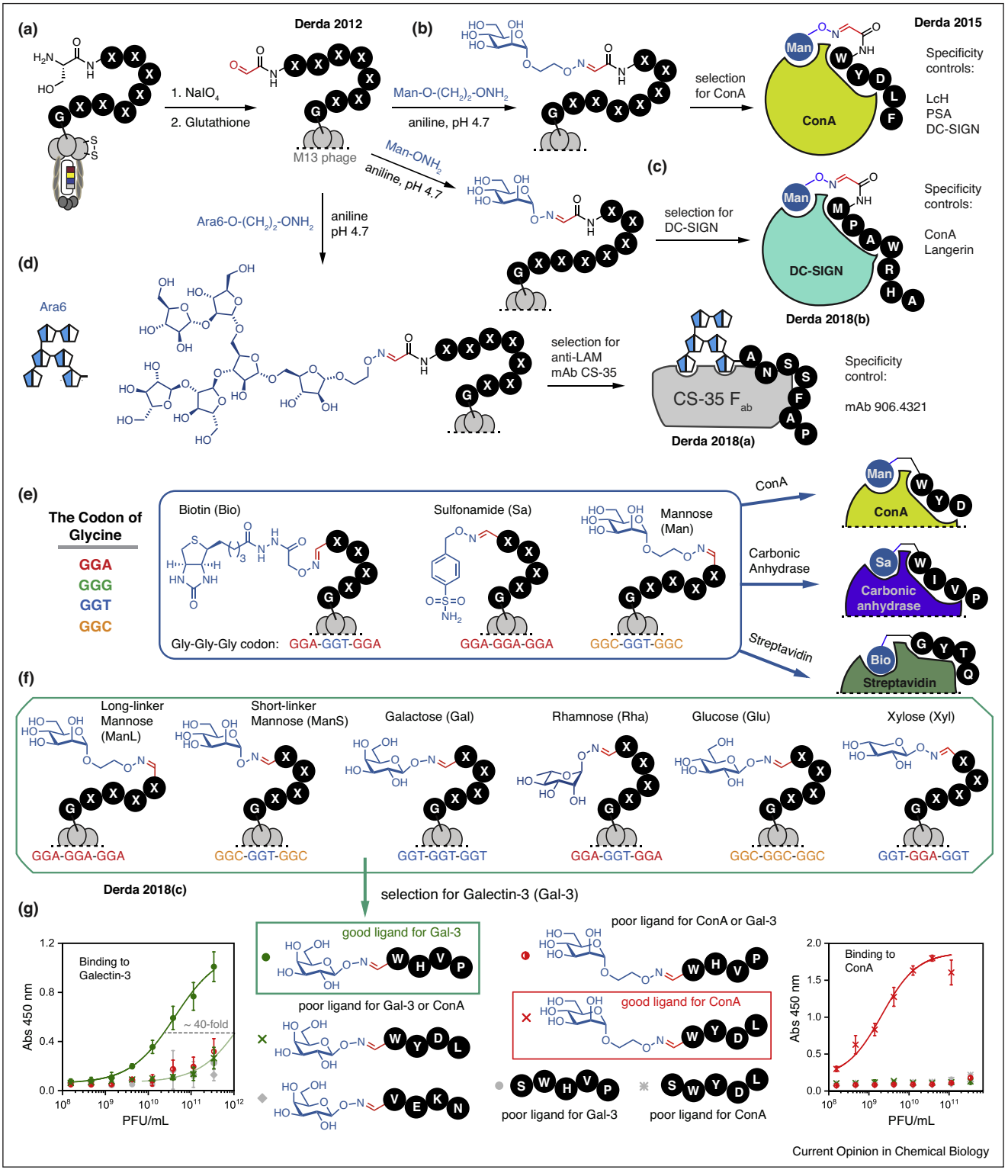
Predicting *a priori* which molecular topology provides the best selection outcome for a particular target is challenging. Solving this question practically necessitates building libraries of diverse architectures. Nature builds diverse proteins from the canonical 20 amino acids and further diversifies them via hundreds of different PTMs (Figure 1b) [25,26]. The same approach can build diverse libraries for GE-FBD. Chemical post-translational modifications (cPTM) are attractive because one ligation strategy can diversify the *same* library with many different fragments. Drawback of cPTM is the uncertainty in the incorporation of the fragment: as the fragment is not encoded genetically, the fidelity of fragment incorporation cannot be assessed via sequencing. It is thus crucial to develop methods for quantifying the yield of chemical modifications of libraries to ensure the complete conjugation of the fragments to all members of the library.

cPTM of N-terminus

Oxidation of native N-terminal Ser/Thr residue yields a bio-orthogonal aldehyde handle for conjugation of fragments (Figure 2a) [27]. We employed oxime ligation to attach a mannose (Man) onto the N-terminus of the random heptapeptide library SX_7 (Figure 2b). Selection from Man-X_7 library against protein ConA identified a conserved peptide segment (WYD) that binds in synergy with the Man [3••]. X-ray structure (PDB code: 4CZS) confirmed interaction of Man-WYDLF with ConA; K_D of this binding was measured 4.6 μM by isothermal titration calorimetry. Furthermore, the compound competitively inhibit the ConA:Man interaction with an IC_{50} of 4.0 μM. This outcome represented a 30-fold to 50-fold enhancement when compared to K_D = 137 μM and IC_{50} = 190 μM of CH_3O-Man. The selectivity of the Man-WYK(ε-Cy3)-OH was assessed using an array of 85 lectins: only three lectins, ConA and related Man-binding lectins LcH and PSA, demonstrated statistically significant binding to the compound. Binding of Man-WYDLF to LcH and PSA was validated by ITC with K_D of 1 mM and 0.7 mM respectively. These affinities were similar to that of CH_3O-Man fragment (K_D = 2 mM for LcH and 1 mM for PSA) indicative of no enhancement of binding affinity with the peptide. The affinity enhancement by WYDLF, thus, was specific to ConA.

GE-FBD with similar Man-X_7 library identified Man-MPAWRHA glycopeptide that bound to cells expressing dendritic cell receptor DC-SIGN but not cells expressing closely related mannose-binding receptor Langerin (Figure 2c) [10]. In two campaigns starting from the Man-X_7 library, GE-FBD reproducibly boosted the specificity of the promiscuous Man fragment. The targets—lectins ConA and DC-SIGN—are examples of proteins that evaded traditional lead discovery due to the presence of multiple binding sites in these proteins. Published selections that employed peptide phage display reproducibly converged to a peptide that binds to pocket distal from the carbohydrate-binding site [28]. Similar outcome was observed in screening of non-glycan fragments against DC-SIGN: many fragment targeted the binding pockets that are distal from the carbohydrate-binding site [29•]. For proteins with multiple potential binding sites, approaches like GE-FBD can steer ligand discovery process towards the desired pocket.

Figure 2



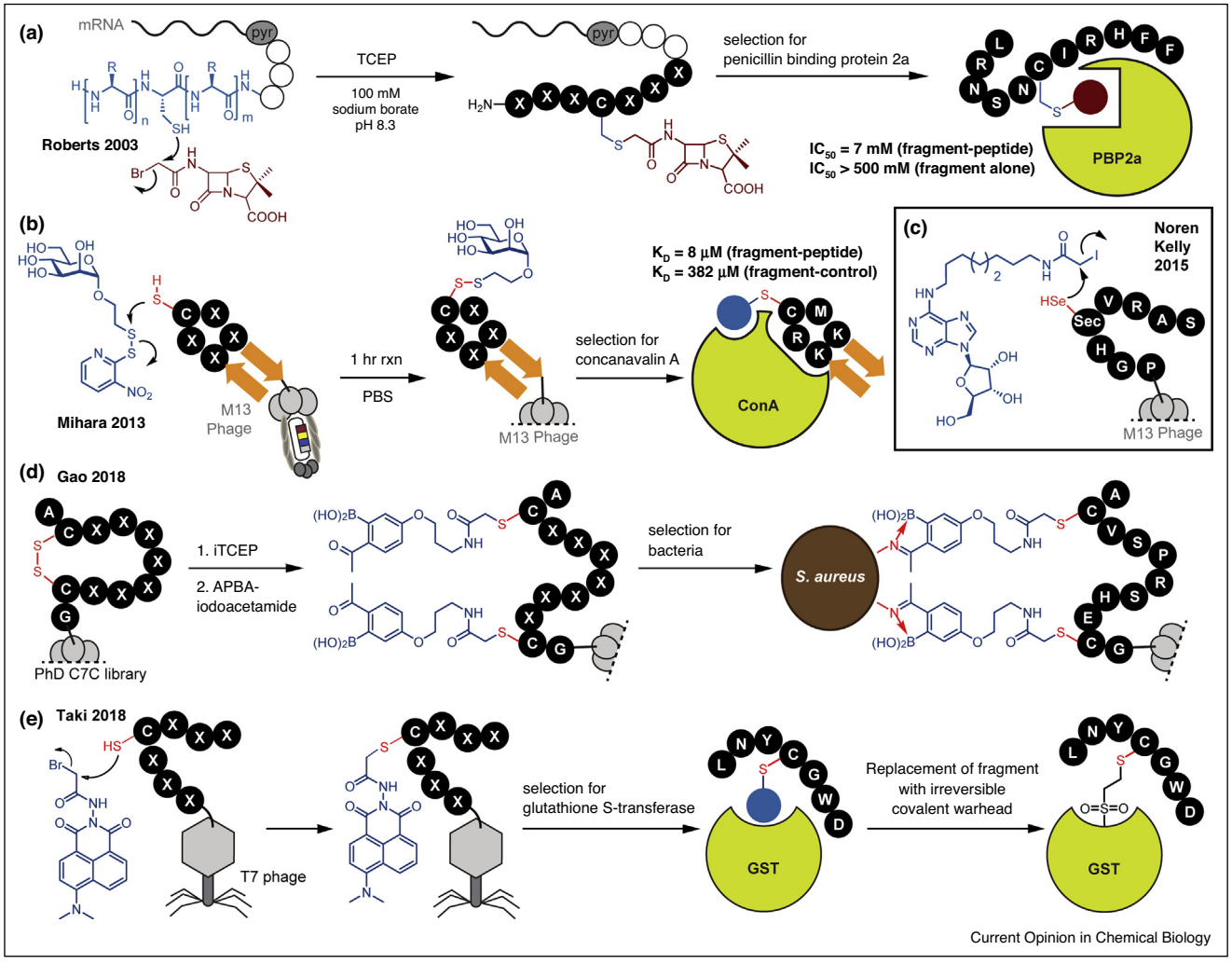
GE-FBD through cPTM of N-terminus. N-terminal modification via **(a)** oxidation of Ser/Thr followed by oxime ligation of **(b–c)** monosaccharides or **(d)** complex glycans. **(e–f)** 'silent encoding' permits modification of libraries with different fragments, mixing these modified libraries, and decoding both peptide sequence and identity of the fragment after selection. **(g)** Selection of peptide library modified with different monosaccharides against galectin-3 identified galactose fragment as the best binder when conjugated to a specific peptide (WHVP). Phage that displays Gal-WHVP demonstrated binding to immobilized galectin-3 using an ELISA assay. Substituting the galactose with a mannose or the WHVP with a WYDL

(Figure 2 Legend Continued) significantly attenuated the phage binding. Similarly, only phage that displays a mannose pairing up with the WYDL demonstrated binding to ConA.

To show that a complex oligosaccharide could serve as 'fragment', we modified SX_7 phage libraries with hexasaccharide Ara_6 derived from lipoarabinomannan (LAM) glycolipid from the cell wall of *Mycobacterium Tuberculosis* (Figure 2d). Selection of Ara_6-X_7 library identified Ara_6-ANSSFAP that binds to anti-LAM mAb CS-35 with a $K_D = 1.4\,\mu M$ [8], whereas the fragment alone (Ara_6-$-OCH_3$) exhibited $K_D = 19\,\mu M$. A related anti-LAM mAb 906.4321 bound Ara_6-OCH_3 and Ara_6-ANSSFAP with comparable affinity. Increased specificity of the ligands identified by GE-FBD was preserved when they were immobilized on a glass surface. In a prototype serological assay, Ara_6-ANSSFAP discriminated closely related anti-LAM mAbs more effectively than the parental Ara_6 epitope: The EC_{50} of $(Ara_6\text{-ANSSFAP})_2$-BSA for CS-35 was $0.2\,\mu g\,mL^{-1}$ (i.e., 1.3 nM) and for 906.4321 was $3.8\,\mu g\,mL^{-1}$ (i.e., 25 nM), whereas the EC_{50} of $(Ara6)_4$–BSA is indistinguishable for both anti-LAM mAb ($0.5–0.8\,\mu g\,mL^{-1}$) [8].

In libraries produced by cPTM, fragments are not genetically encoded. To overcome this problem, chemically identical libraries can be constructed using different DNA-codons ('silent barcodes'). A published example

Figure 3



GE-FBD through cPTM of Cys residue. **(a)** S_N2 alkylation of Cys residue in mRNA libraries with 6-bromoacetyl penicillin and its selection. **(b)** Disulfide exchange introduces mannose into phage libraries for selection against ConA. **(c)** Introduction of synthetic molecule through alkylation with selenocysteine displayed on phage. **(d)** Alkylation of Cys residues with reversible covalent warhead to facilitate selection of specific binders for bacterial surface. **(e)** T7 phage library was first modified with a fragment and selected for glutathione *S*-transferase (GST). The fragment was then replaced with an irreversible covalent warhead that reacts selectively with GST in the presence of serum proteins.

placed 'silent barcodes' in the Gly-Gly-Gly linker that connected the random peptide to phage (Figure 2e) [7••]. Modifying barcoded peptide libraries separately with different fragments (biotin, sulfonamide, and mannose), pooling these libraries and selection against ConA, carbonic anhydrase, and streptavidin enriched the matched fragment for each target. Moreover, it identified distinct peptides that synergize with each fragment (Figure 2e) [7••]. 'Silent encoding' can encode closely related modifications (e.g., glycan diastereomers) (Figure 2f). Panning of these glycopeptides against Galectin-3 protein identified Gal-WHVP conjugate that exhibited a 40-fold increase in binding potency for Galectin-3 as compared to a control conjugate (Figure 2g) [9•]. Seemingly similar WYDL and WHVP peptides synergized specifically with Man or Gal fragments. Fragment swapping confirmed their specificity (Figure 2g) [9•].

GE-FBD through cPTM of cys residue

Historically important examples of modification of GE libraries used Cys as handle: (i) S_N2 alkylation of a Cys-containing antibody libraries with a fluorophore [30]; (ii) native chemical ligation of a 32-amino-acid eglin 'fragment' to the N-terminal Cys of a peptide library [31]; and (iii) S_N2 alkylation of Cys-containing mRNA library with 6-bromoacetyl penicillin [11•]. Screening the penicillin-modified library successfully identified LRNSNC(penicillin)IRHFF that exhibited 100-fold stronger inhibition for penicillin-binding protein 2a relative to the penicillin itself (Figure 3a).

In recent examples, Mihara *et al.* employed disulfide exchange to introduce mannose into phage-displayed peptide libraries [13•,14]. Selection against ConA protein identified loop-constrained and helix-constrained Man-peptide conjugates that exhibited a 46-fold and >100-fold higher binding affinity to ConA when compared to the controls (Figure 3b). Other classical reactions, such as maleimide-Cys ligation can graft fragments into phage-displayed libraries [4]. As an alternative to Cys, selenocysteine residues can be employed [32] (Figure 3c). In contrast to GE-FBD focused on fragments that bind non-covalently, Gao *et al.* alkylated the cysteines in ACX_7C phage-displayed peptide library with reversible covalent warhead 2-acetylphenylboronic acid (APBA) (Figure 3d) [33••]. Screening of this bis-APBA-peptide library against bacteria yielded hits that bind to specific bacteria strain at low nanomolar affinity. Chemical modification [33••] allowed upgrading commercial library and bypass the need for UAA mutagenesis to incorporate boronic acids [34]. A less direct approach can start from a library with an unreactive fragment, selection against the target, and replacement of the fragment in the resulting ligands with a covalent warhead (Figure 3e) [35].

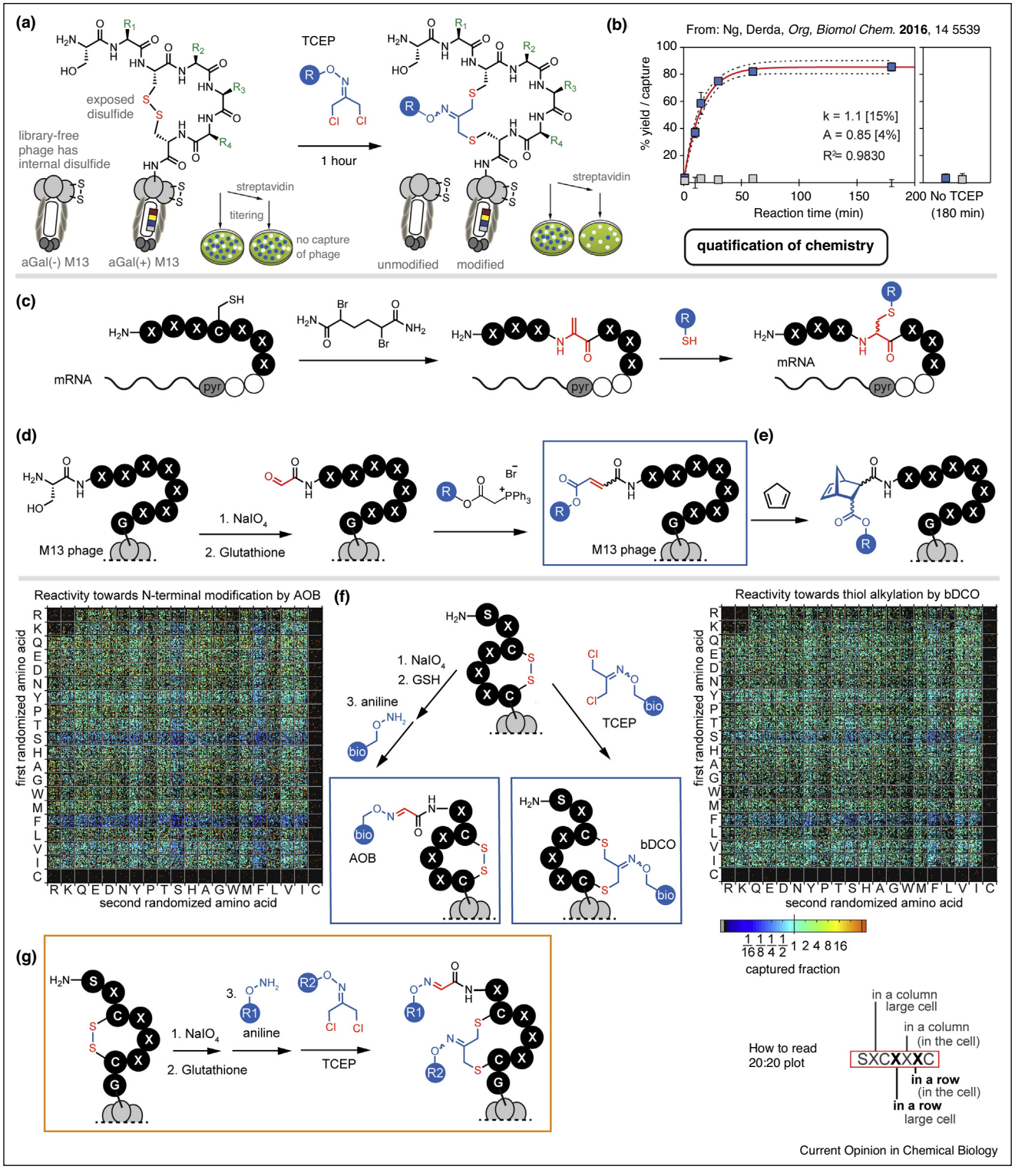
Enabling chemistry for cPTM

Growing interest in macrocyclic peptides makes synthetic strategies that can introduce a fragment and cyclize the peptide library highly desirable. Cys-alkylation with dichloroacetone-derived oxime cyclizes the peptide library and introduces the fragments concomitantly [36•] (Figure 4a). An alternative approach cyclizes a peptide and introduces an orthogonal reactive group first, followed by ligation of the fragments in the second step. Phage decorated with aldehyde or ketone are stable in storage [27] but low reactivity of ketone-macrocycle necessitated long reaction time at acidic pH [36•]. Tetradentate halo-electrophiles linkers form bicyclic peptide with an exocyclic thiol reactive group [37]; but high reactivity of halo-electrophiles may present a problem. For example, endocyclic bromobenzyl groups in peptide macrocycles react with side chains of Lys in an intramolecular fashion [38]. Suga *et al.* employed method of Davis [39] to convert Cys in mRNA-displayed macrocycles to endocyclic dehydroalanine. Michael addition of anomeric thioglycosides furnished a 1:1 mixture of diastereomeric glycopeptides (Figure 4c) [40•]. This report bridged GE-FBD with a dehydroalanine intermediate that acts not only as electrophile for *S* and *N*-nucleophiles but also a dipolarophile electronically matched to react with 2-diazoacetamide [41] and a SOMO-phile accepting $C(sp^3)$ radicals [39].

A C–C bond forming Wittig reaction with ester-stabilized ylide (Figure 4d) [42•] can diversify aldehyde-peptide phage-displayed libraries. The resulting electrophilic handle can be used to introduce the fragments via Michael addition or Diels–Alder reactions (Figure 4e). Aldehyde-containing libraries can, theoretically, be diversified by reactions that have been validated in aldehyde-containing proteins: Pictet-Splenger [43], aminobenzamidoxime ligation [44], Mukaiyama aldol [45] and Ugi reactions [46]. The barrier for adaptation of these reactions is effective determination of their conversion in the context of the entire library. It is tempting to extrapolate the yields of modification of 10^9 peptides from the modification in a few model peptides [38]. However, the rate of modifications of peptides unpredictably varies with sequence. Inequality of Cys-reactivity in different peptides is well-documented giving rise to peptides with 'fast reactive' Cys residues. The same inequality is problematic when uniform modification of all library members is desired.

To monitor the extent of modification in GE-libraries of peptides, we developed biotin-containing reactive probes that allow capture of phage that has been modified (biotinylated) [27,36•,48] (Figure 4b). Conducting this reaction in a mixture of library (Lib) phage and wild-type (WT) phage that displays no peptide tests the specificity of both reaction and capture. Conveniently, Lib and WT form plaques of two colors (blue and white respectively)

Figure 4



Enabling chemistry for the production of cPTM libraries. **(a)** Fragment incorporation with concomitant cyclization. When R group is biotin, the conversion can be quantified by biotin-capture assay. **(b)** Quantification of the kinetics of the on-phage reaction shown in panel (a). **(c)** Two step elimination and Michael addition; **(d)** Incorporation of reactive electrophiles via Wittig reaction; **(e)** Incorporation of fragment via tandem Wittig/Diels–Alder reaction; **(f)** Quantification of the library-wide incorporation of biotin handles either via cysteine alkylation with bDCO or N-terminal oxidation followed by oxime ligation with AOB. **(g)** Site-specific incorporation of two different fragments into one library.

and can be quantified in the same 'pot'. Ideal outcome—a progressive increase in biotinylation of Lib particles and no observable biotinylation of WT—quantify the chemistry, rate constant (k), and regioselectivity in libraries with single-molecule resolution. In six reports to date, the k of reaction measured on phage using biotin capture, or a 'pulse-chase' capture, was similar to k measured in purified peptides by methods such as HPLC [27,36•,42•,44,47,48]. Sequencing of the pulled-down biotinylated phage can further illuminate a library-wide conversion. We used this approach to compare the modification of the same library SXCXXXC by N-terminal ligation and Cys alkylation (Figure 4f) [49•]. Modification of N-termini by aminooxy-biotin exhibited bias against reaction on peptides with penultimate Ser and Phe whereas Cys alkylation by biotin dichloro-oxime was more uniform.

Other methods for quantification exist: MALDI mass-spectrometry can characterize reaction of peptides cleaved from a large number ($>10^{10}$ copies) of monoclonal phage [50]; unfortunately, this method is not straightforward for mixtures of peptides with different masses. The western blots [32] and fluorescent SDS-PAGE [33••] monitor modification of phage-displayed libraries and simultaneously assess their regioselectivity by juxtaposing the modification of pIII protein and other phage coat proteins in one gel. Throughout the 20th century, growth of synthetic chemistry was enabled by development of tools for analysis and characterization (X-ray, IR, NMR, ESI-MS). Similarly, development of new methods for library-wide quantification of chemical reactions will facilitate the merger of synthetic chemistry with genetically encoded technologies.

Conclusions and outlook

Innovations in molecular discovery are driven by unsolved demands. One of such unsolved challenges is the discovery of ligands that elicit a defined mechanism of action such as antagonism or agonism. *De novo* screening with libraries of small molecules or peptides has a likelihood of finding binders occupying a binding site that does not elicit any biological effect. In contrast, GE-FBD incorporates a well-characterized and functional binding fragment into the library and guides the selection towards the active site. The likelihood of finding peptide fragments in the vicinity of the active site, therefore, increases. GE-FBD thus holds the promise to be a robust approach for discovery of ligands with well-defined agonism/antagonism response.

New synthetic methodologies that are bio-orthogonal or bio-compatible will expand the versatility of GE-FBD. Merger of small molecule and biological therapeutics will drive innovation in reactions that ligate 'unnatural' molecules to native peptide residues. Future advances in chemistry will broaden the scope of GE-FBD by allowing incorporation of two or more distinct fragments into the same peptide (Figure 4g). Molecular biology, in turn, can upgrade GE-FBD by providing new methods for encoding of fragments. 'Silent encoding' introduces the DNA tag before modification [7••]; introducing the DNA tag after the ligation of the fragment is a complementary approach. Such strategies build DNA-encoded chemical libraries [24••]. A hybrid approach may start from a *translated* DNA-displayed or RNA-displayed library of peptides, followed by library-wide modification of peptides and ligation of a nucleotide strand to encode this modification.

Unlike synthetic chemistry that converts one substrate **S** to one product **P**, cPTM of GE-libraries transforms a vast multiset of substrates simultaneously: $\{\mathbf{P}\} = \mathbf{cPTM}\{\mathbf{S}\}$ [51]. Modifications of libraries of peptides, thus, should be cautiously distinguished from modification of singular polypeptides. Library-wide deep sequencing should make it possible to quantify manipulations of the library, such as selection or chemical modification, in form of defined 'operators' that act on multisets of sequences (Figure 1f) [52,53]. Modern engineering builds on robust numerical models; similarly, reproducible molecular discoveries from GE-FBD may become possible by improving quantitative treatments of modifications and selection procedures.

Conflict of interest statement

R.D. is the Founder and Chief Executive Officer of 48Hour Discovery Inc., the company that commercializes genetically-encoded, chemically-modified peptide library technologies.

Acknowledgement

R.D. gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (#402511), NSERC Discovery Accelerator Supplements Program, GlycoNet Canadian Glycomics Network and Alberta Glycomics Centre.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Erlanson DA: **Introduction to fragment-based drug discovery**. *Fragment-based Drug Discovery and X-ray Crystallography*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011, 1-32.

2. Murray CW, Rees DC: **The rise of fragment-based drug discovery**. *Nat Chem* 2009, **1**:187.

3. •• Ng S, Lin E, Kitov PI, Tjhung KF, Gerlits OO, Deng L, Kasper B, Sood A, Paschal BM, Zhang P, Ling C-C *et al.*: **Genetically encoded fragment-based discovery of glycopeptide ligands for carbohydrate-binding proteins**. *J Am Chem Soc* 2015, **137**:5248-5251.

Biochemical and X-ray evidence, for synergistic interaction between a mannose fragment and peptide binding to the desired binding site of glycan-binding protein ConA.

4. Santoso B, Lam S, Murray BW, Chen G: **A simple and efficient maleimide-based approach for peptide extension with a**

cysteine-containing peptide phage library. *Bioorg Med Chem Lett* 2013, **23**:5680-5683.

5. Meyer SC, Shomin CD, Gaj T, Ghosh I: **Tethering small molecules to a phage display library: Discovery of a selective bivalent inhibitor of protein kinase a**. *J Am Chem Soc* 2007, **129**:13812-13813.

6. Wang W, Hirano Y, Uzawa T, Liu M, Taiji M, Ito Y: ***In vitro* selection of a peptide aptamer that potentiates inhibition of cyclin-dependent kinase 2 by purvalanol**. *MedChemComm* 2014, **5**:1400-1403.

7. •• Tjhung KF, Kitov PI, Ng S, Kitova EN, Deng L, Klassen JS, Derda R: **Silent encoding of chemical post-translational modifications in phage-displayed libraries**. *J Am Chem Soc* 2016, **138**:32-35.
Encoding of multiple unnatural fragments in the canonical phage-displayed library was achieved through modification of libraries constructed using different DNA codons.

8. Chou Y, Kitova EN, Joe M, Brunton R, Lowary TL, Klassen JS, Derda R: **Genetically-encoded fragment-based discovery (ge-fbd) of glycopeptide ligands with differential selectivity for antibodies related to mycobacterial infections**. *Org Biomol Chem* 2018, **16**:223-227.

9. • Vinals DF, Kitov PI, Tu Z, Zou C, Cairo CW, Lin HC-H, Derda R: **Selection of galectin-3 ligands derived from genetically encoded glycopeptide libraries**. *Peptide Sci* 2018, **0**:e24097.
Example of encoding of six closely related glycan structures in one glycopeptide library and its use in selection against specific glycan binding protein.

10. Ng S, Bennett NJ, Schulze J, Gao N, Rademacher C, Derda R: **Genetically-encoded fragment-based discovery of glycopeptide ligands for DC-SIGN**. *Biorg Med Chem* 2018, **26**:5368-5377.

11. • Li S, Roberts RW: **A novel strategy for in vitro selection of peptide-drug conjugates**. *Chem Biol* 2003, **10**:233-239.
The first well-documented genetically encoded fragment-based discovery from an 'unnatural fragment'.

12. Day JW, Kim CH, Smider VV, Schultz PG: **Identification of metal ion binding peptides containing unnatural amino acids by phage display**. *Biorg Med Chem Lett* 2013, **23**:2598-2600.

13. • Ven Chang I, Tsutsumi H, Mihara H: **Screening for concanavalin a binders from a mannose-modified alpha-helix peptide phage library**. *Mol Biosyst* 2017, **13**:2222-2225.
The authors combined mannose fragment and helical peptide libraries to identify a Man-peptide ligand for ConA with 1.2 μM K_D: >100-fold enhancement in affinity over original Man fragment.

14. Arai K, Tsutsumi H, Mihara H: **A monosaccharide-modified peptide phage library for screening of ligands to carbohydrate-binding proteins**. *Bioorg Med Chem Lett* 2013, **23**:4940-4943.

15. Walport LJ, Obexer R, Suga H: **Strategies for transitioning macrocyclic peptides to cell-permeable drug leads**. *Curr Opin Biotechnol* 2017, **48**:242-250.

16. Henninot A, Collins JC, Nuss JM: **The current state of peptide drug discovery: Back to the future?** *J Med Chem* 2018, **61**:1382-1414.

17. Morrison C: **Constrained peptides' time to shine?** *Nat Rev Drug Discov* 2018, **17**:531.

18. Sawyer TK, Partridge AW, Kaan HYK, Juang Y-C, Lim S, Johannes C, Yuen TY, Verma C, Kannan S, Aronica P, Tan YS *et al.*: **Macrocyclic α helical peptide therapeutic modality: a perspective of learnings and challenges**. *Biorg Med Chem* 2018, **26**:2807-2815.

19. •• Jenson JM, Xue V, Stretz L, Mandal T, Reich L, Keating AE: **Peptide design by optimization on a data-parameterized protein interaction landscape**. *Proc Natl Acad Sci U S A* 2018, **115**:E10342-E10351.
A simple regression model in a well-calibrated set of a few thousands of selected helical peptides to reliably predicted affinities and specificities in the library of billions of peptides.

20. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS: **Rapid mapping of protein functional epitopes by combinatorial alanine scanning**. *Proc Natl Acad Sci U S A* 2000, **97**:8950-8954.

21. Araya CL, Fowler DM: **Deep mutational scanning: assessing protein function on a massive scale**. *Trends Biotechnol* 2011, **29**:435-442.

22. Araya CL, Fowler DM, Chen WT, Muniez I, Kelly JW, Fields S: **A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function**. *Proc Natl Acad Sci U S A* 2012, **109**:16858-16863.

23. • Rogers JM, Passioura T, Suga H: **Nonproteinogenic deep mutational scanning of linear and cyclic peptides**. *Proc Natl Acad Sci U S A* 2018, **115**:10959-10964.
The authors expanded the concept of deep-mutational scanning to assess the role of both natural and unnatural amino acids in macrocycles in one experiment.

24. •• Denton KE, Wang SJ, Gignac MC, Milosevich N, Hof F, Dykhuizen EC, Krusemark CJ: **Robustness of in vitro selection assays of DNA-encoded peptidomimetic ligands to cbx7 and cbx8**. *Slas Discov* 2018, **23**:417-428.
Authors encoded multiple instances of the positive and negative control ligands in DNA-encoded library to yield the first example of robust calculation of Z-factor in genetically encoded discovery.

25. Tan M, Luo H, Lee S, Jin F, Yang Jeong S, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, Lu Z *et al.*: **Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification**. *Cell* 2011, **146**:1016-1028.

26. Li Q, Shortreed MR, Wenger CD, Frey BL, Schaffer LV, Scalf M, Smith LM: **Global post-translational modification discovery**. *J Proteome Res* 2017, **16**:1383-1390.

27. Ng S, Jafari MR, Matochko WL, Derda R: **Quantitative synthesis of genetically encoded glycopeptide libraries displayed on m13 phage**. *ACS Chem Biol* 2012, **7**:1482-1487.

28. Jain D, Kaur K, Sundaravadivel B, Salunke DM: **Structural and functional consequences of peptide-carbohydrate mimicry: crystal structure of a carbohydrate-mimicking peptide bound to concanavalin a**. *J Biol Chem* 2000, **275**:16098-16102.

29. • Aretz J, Baukmann H, Shanina E, Hanske J, Wawrzinek R, Zapol'skii VA, Seeberger PH, Kaufmann DE, Rademacher C: **Identification of multiple druggable secondary sites by fragment screening against dc-sign**. *Angew Chem Int Ed* 2017, **56**:7292-7296.
Biophysical and structural evidence for multiple binding sites in therapeutically relevant carbohydrate binding proteins and steering of fragment discovery away from desired carbohydrate-binding site.

30. Jespers LSAT, Winter GP, Bonnert TP, Simon TM: Bacteriophage library displaying immunoglobulin repertoires with a chemical moiety covalently bound within the binding site: Production and selection thereof. In: *PCT International Patent*. (1995).

31. Dwyer MA, Lu WY, Dwyer JJ, Kossiakoff A: **Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity**. *Chem Biol* 2000, **7**:263-274.

32. Beech J, Saleh L, Frentzel J, Figler H, Corrêa IR, Baker B, Ramspacher C, Marshall M, Dasa S, Linden J, Noren CJ *et al.*: **Multivalent site-specific phage modification enhances the binding affinity of receptor ligands**. *Bioconj Chem* 2015, **26**:529-536.

33. •• McCarthy KA, Kelly MA, Li K, Cambray S, Hosseini AS, van Opijnen T, Gao J: **Phage display of dynamic covalent binding motifs enables facile development of targeted antibiotics**. *J Am Chem Soc* 2018, **140**:6137-6145.
GE-FBD converted promiscuous glycan-binding phenylboronic acid derivative to a potent and specific peptide that recognizes specific glycans on the surface of bacteria.

34. Liu CC, Mack AV, Brustad EM, Mills JH, Groff D, Smider VV, Schultz PG: **Evolution of proteins with genetically encoded 'chemical warheads'**. *J Am Chem Soc* 2009, **131**:9616-9617.

35. Uematsu S, Tabuchi Y, Ito Y, Taki M: **Combinatorially screened peptide as targeted covalent binder: alteration of bait-conjugated peptide to reactive modifier**. *Bioconj Chem* 2018, **29**:1866-1871.

36. • Ng S, Derda R: **Phage-displayed macrocyclic glycopeptide libraries**. *Org Biomol Chem* 2016, **14**:5539-5545.
Concurrent macrocyclization and incorporation of diverse fragments into phage-displayed peptide libraries.

37. Ernst C, Sindlinger J, Schwarzer D, Koch P, Boeckler FM: **The symmetric tetravalent sulfhydryl-specific linker natba facilitates a combinatorial "tool kit" strategy for phage display-based selection of functionalized bicyclic peptides**. *ACS Omega* 2018, **3**:12361-12368.

38. Heinis C, Rutherford T, Freund S, Winter G: **Phage-encoded combinatorial chemical libraries based on bicyclic peptides**. *Nat Chem Biol* 2009, **5**:502-507.

39. Dadova J, Galan SRG, Davis BG: **Synthesis of modified proteins dehydroalanine**. *Curr Opin Chem Biol* 2018, **46**:71-81.

40. • Jongkees SAK, Umemoto S, Suga H: **Linker-free incorporation of carbohydrates into in vitro displayed macrocyclic peptides**. *Chem Sci* 2017, **8**:1474-1481.
The authors converted natural Cys residued in mRNA-displayed macrocycles into valuable dehydroalanine intermediates and employed them for installation of thiglycosides.

41. Aronoff MR, Gold B, Raines RT: **1,3-dipolar cycloadditions of diazo compounds in the presence of azides**. *Org Lett* 2016, **18**:1538-1541.

42. • Triana V, Derda R: **Tandem wittig/Diels-Alder diversification of genetically encoded peptide libraries**. *Org Biomol Chem* 2017, **15**:7869-7877.
Late-stage functionalization of phage-displayed peptide libraries via multi-step carbon–carbon bond forming reactions.

43. Agarwal P, van der Weijden J, Sletten EM, Rabuka D, Bertozzi CR: **A pictet-spengler ligation for protein chemical modification**. *Proc Natl Acad Sci U S A* 2013, **110**:46.

44. Kitov PI, Vinals DF, Ng S, Tjhung KF, Derda R: **Rapid, hydrolytically stable modification of aldehyde-terminated proteins and phage libraries**. *J Am Chem Soc* 2014, **136**:8149-8152.

45. Alam J, Keller TH, Loh T-P: **Functionalization of peptides and proteins by mukaiyama aldol reaction**. *J Am Chem Soc* 2010, **132**:9546-9548.

46. Morejón MC, Laub A, Kaluđerović GN, Puentes AR, Hmedat AN, Otero-González AJ, Rivera DG, Wessjohann LA: **A multicomponent macrocyclization strategy to natural product-like cyclic lipopeptides: synthesis and anticancer evaluation of surfactin and mycosubtilin analogues**. *Org Biomol Chem* 2017, **15**:3628-3637.

47. Jafari MR, Deng L, Kitov PI, Ng S, Matochko WL, Tjhung KF, Zeberoff A, Elias A, Klassen JS, Derda R: **Discovery of light-responsive ligands through screening of a light-responsive genetically encoded library**. *ACS Chem Biol* 2014, **9**:443-450.

48. Kalhor-Monfared S, Jafari MR, Patterson JT, Kitov PI, Dwyer JJ, Nuss JM, Derda R: **Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone**. *Chem Sci* 2016, **7**:3785-3790.

49. • He B, Tjhung KF, Bennett NJ, Chou Y, Rau A, Huang J, Derda R: **Compositional bias in naïve and chemically-modified phage-displayed libraries uncovered by paired-end deep sequencing**. *Sci Rep* 2018, **8**:1214.
Deep-sequencing of the peptide libraries modified by biotin-containing reagents uncovered reaction-specific, library-wide bias in chemical modification of peptide libraries.

50. Chen S, Touati J, Heinis C: **Tracking chemical reactions on the surface of filamentous phage using mass spectrometry**. *Chem Commun* 2014, **50**:5267-5269.

51. Abbas A, Xing B: Loh**Allenamides as orthogonal handles for selective modification of cysteine in peptides and proteins**. *Angew Chem Int Ed* 2014, **53**:7491-7494.

52. Matochko WL, Derda R: **Error analysis of deep sequencing of phage libraries: Peptides censored in sequencing**. *Comput Math Methods Med* 2013, **2013**:491612.1-491612.13 http://dx.doi.org/10.1155/2013/491612.

53. Matochko WL, Cory Li S, Tang SKY, Derda R: **Prospective identification of parasitic sequences in phage display screens**. *Nucleic Acids Res* 2014, **42**:1784-1798.