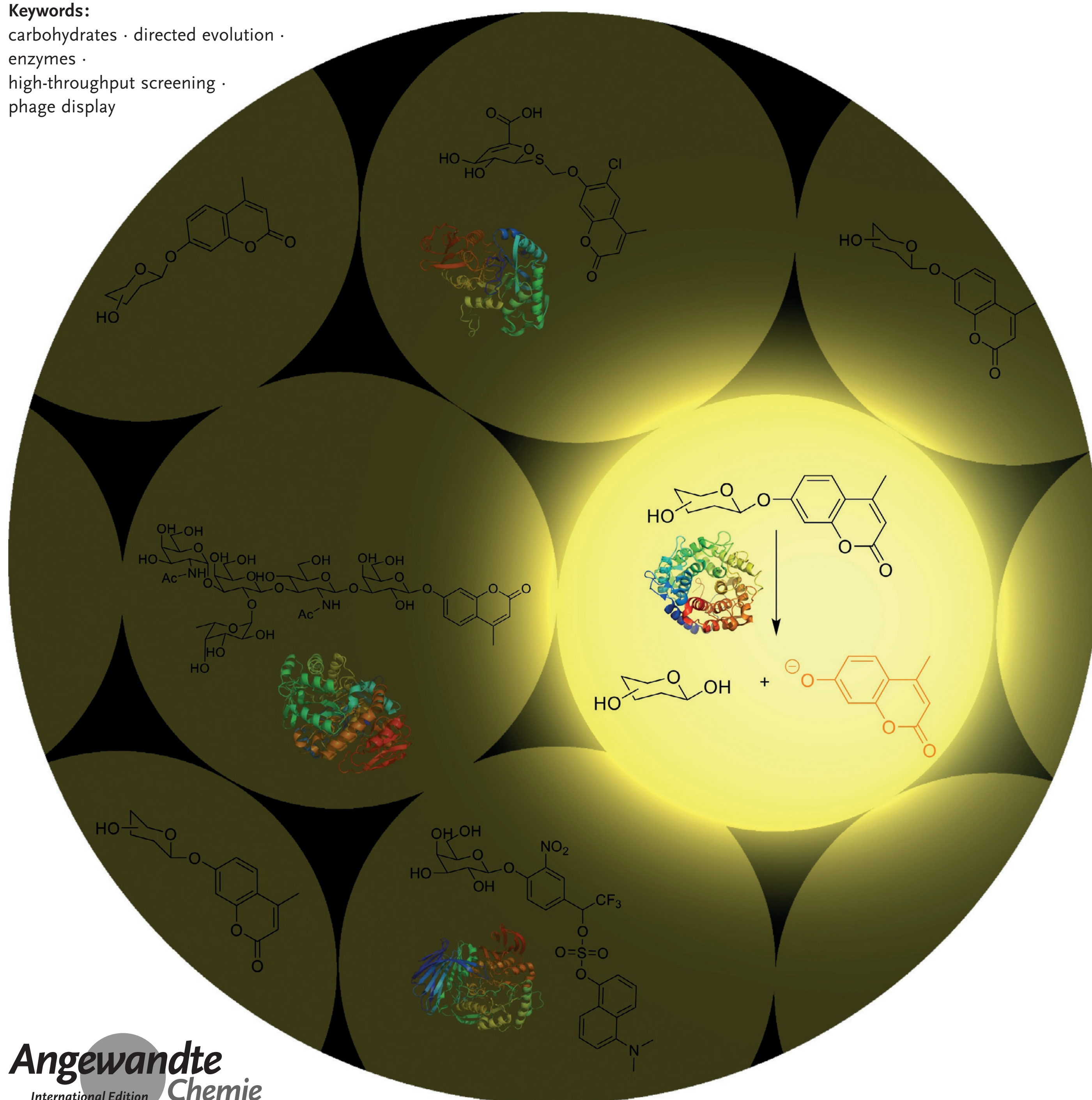
Carbohydrate-Active Enzymes

International Edition: DOI: 10.1002/anie.201900055
German Edition: DOI: 10.1002/ange.201900055

High-Throughput Approaches in Carbohydrate-Active Enzymology: Glycosidase and Glycosyl Transferase Inhibitors, Evolution, and Discovery

*Lemeng Chao and Seino Jongkees**

Keywords:
carbohydrates · directed evolution · enzymes · high-throughput screening · phage display



Angewandte Chemie International Edition

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Carbohydrates are attached and removed in living systems through the action of carbohydrate-active enzymes such as glycosyl transferases and glycoside hydrolases. The molecules resulting from these enzymes have many important roles in organisms, such as cellular communication, structural support, and energy metabolism. In general, each carbohydrate transformation requires a separate catalyst, and so these enzyme families are extremely diverse. To make this diversity manageable, high-throughput approaches look at many enzymes at once. Similarly, high-throughput approaches can be a powerful way of finding inhibitors that can be used to tune the reactivity of these enzymes, either in an industrial, a laboratory, or a medicinal setting. In this review, we provide an overview of how these enzymes and inhibitors can be sought using techniques such as high-throughput natural product and combinatorial library screening, phage and mRNA display of (glyco)peptides, fluorescence-activated cell sorting, and metagenomics.

1. Introduction

Carbohydrate-active enzymology has undergone a revolution in the last few decades, driven at least in part by advances in high-throughput technologies. Beyond just allowing "more of the same," high-throughput approaches open up entire new approaches and fields of work that could not be attempted using traditional means. The inhibitors discovered from high-throughput technologies are not just optimised versions of the same types of molecules arising from rational approaches but rather represent entire new classes and modes of action. The enzymes discovered and engineered through high-throughput approaches are not just slightly altered versions of existing enzymes but can have dramatically different properties or reactions catalysed. From an outside perspective these discoveries risk appearing as isolated lucky breaks from nebulous "fishing expeditions", but they provide fuel to the furnace of more hypothesis-driven and detail-oriented science, and innovative approaches with an element of rational design can dramatically improve the chances of success. High-throughput approaches represent a way of accessing the unknown unknowns that can truly push science forward.

Carbohydrate-active enzymology in particular stands to benefit from the power of high-throughput approaches, as there still remains much that is unknown. Inhibition of glycosidases is typically approached through transition-state mimics,[1–3] and this approach has had many successes, but development of these remains a slow process that requires a deep understanding of the mechanism of an enzyme and its structure[4] and often struggles in selectivity. For glycosyl transferases the situation is worse.[5,6] Some progress has been made through rational approaches, particularly in substrate mimics that are converted by intracellular kinases to the active inhibitor,[7] but as yet no general paradigm exists for inhibition of this important class of enzyme. Despite this, several inhibitors of glycosyl transferases are in clinical use, illustrating their importance as a target.[8] Alternative approaches that do not require extensive background knowledge, and that allows very precise targeting of only one specific enzyme, would open up new avenues of research on the biology of the relevant glycans.

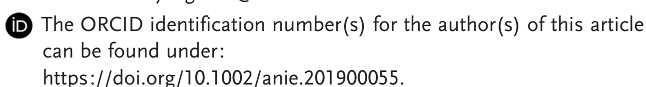
Many of the same approaches that are used for testing inhibition can also be applied to testing for improved activity, either in evolving existing enzymes for better activity, tuning optimal reaction conditions, or finding completely new enzymes. As the role of enzymes in synthesis grows,[9] discovery of new catalysts to expand the range of amenable reactions becomes ever more important. For example, given the abundance of glycans in nature, carbohydrate-active enzymes for biomass conversion are particularly important to stimulate a transition to a post-hydrocarbon world.[10] For pharmaceuticals, biologicals are a very quickly growing class of drug, but it is becoming apparent that their glycosylation state can have a large influence on their activity,[11] and enzymes that can remodel protein- or cell-linked glycans to an optimal state are thus in high demand.

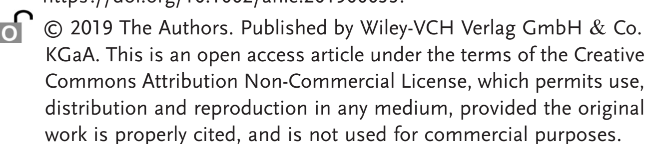
In this review, we outline and discuss new developments that push the envelope on what can be achieved using high-throughput approaches with this challenging class of biomolecule, the carbohydrates. Applications of such technologies (summarised in Figure 1) are discussed, presenting highlights in the discovery of inhibitors, tuning the activity of existing enzymes, and the search for new enzymes.

2. Searching for Inhibitors

Inhibitors are powerful tools to adjust the level of enzymatic activity, and the biological effects that result. Discovery of new inhibitors is particularly appealing as a source of treatments for the many diseases involving malfunctioning enzymes.[12] Already at the turn of the millennium, high-throughput screening was the primary engine driving lead discovery in many pharmaceutical companies,[13] but also non-druglike molecules arising from high-

[*] L. Chao, Dr. S. A. K. Jongkees
Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University
Universiteitsweg 99, 3581AG Utrecht (The Netherlands)
E-mail: s.a.k.jongkees@uu.nl


The ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201900055.


© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, and is not used for commercial purposes.

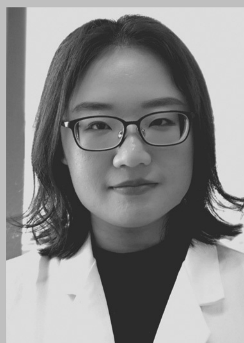
throughput approaches can be of great use as tools in the research community.[14] High-throughput approaches have recently yielded new classes of molecules that can circumvent some of the limitations of more traditional approaches.

Glycosidase activity can often be linked to a colour or fluorescence signal through the use of chromo- or fluorogenic substrates (see below), meaning they are often easily amenable to high-throughput screening approaches. High-throughput screening of commercial and in-house natural product libraries using such approaches has recently revealed several alternative strategies for glycosidase inhibition, as illustrated by two different inhibitors of human pancreatic α-amylase. One of these, a small protein of marine origin called helianthamide,[15] showed a low-picomolar K_i value derived in large part through very close shape complementarity to the target enzyme's active site,[16] while the other, a glycosylated flavonol of plant origin called montbretin A[17] (Figure 2 A), demonstrated a new inhibitory motif in its interactions with the conserved catalytic residues of retaining glycosidases.[18] These illustrate that the paradigm of transition-state mimicry need not be the only approach to glycosidase inhibition, with potential advantages of these new approaches in selectivity between highly related enzymes.

For glycosyl transferases, the situation is more challenging, as it is difficult to directly couple sugar transfer to a convenient readout that can be run in high throughput. However, through the use of indirect or coupled assays, the generation of NDP or NMP (from the NDP/NMP-sugar donor) can be detected. In the commercialised UDP-Glo assay (Promega), UDP release is coupled to generation of ATP and subsequently a luminescent signal. Another approach uses the diphosphate moiety to relieve quenching of a fluorophore.[19] A further alternative uses release of one or two equivalents of inorganic phosphate from the nucleotides by a phosphatase to allow its quantification using malachite-based reagents, although this precludes the use of phosphate buffer.[20] However, these can struggle under some conditions, such as testing of crude lysates. A substrate-affinity-based approach offers another alternative, such as by using a fluorescence polarisation change from the labelled-substrate displacement by the inhibitor. In an example application of this approach from the Walker group, initial hits against O-GlcNAc transferase from an expanded version of the the commercial ChemDiv library[21] were diversified through combinatorial chemistry based on a conserved quinolinone-6-sulfonamide core,[22] before the elucidation of an X-ray crystal structure allowed structure-based optimisation to a molecule with low-nanomolar potency in vitro and a low-micromolar EC_{50} in cellular assays (Figure 2 B).[23] The quinolinone-6-sulfonamide core in these applications acts as a mimic of the sugar donor's nucleotide base, suggesting much broader applicability. Coupled assays can also be useful, such as using glycosyl transfer to block the activity of one or more exo-acting glycosidases on a fluorescent substrate.[24]

Peptide display on phage or mRNA[25] can be used to select peptides with a strong affinity for in principle any carbohydrate-active enzyme or carbohydrate-binding protein, thereby allowing discovery of peptide-based inhibitors. However, a limitation is that the biopanning used to find hits is an affinity-based approach, only selecting for peptides that bind to the targeted enzyme. Whether they increase or inhibit, or indeed have no effect on, the enzymatic activity needs further study in each case. This is illustrated by work in which phage display was applied to maltase-glucoamylase,[26] revealing two cyclic peptides that had a weakly inhibitory effect at millimolar concentrations, while two linear peptides from the same enriched library were found to increase enzymatic activity at the same concentration. While these molecules do not display high affinity, this ability of peptide-based ligands to increase as well as decrease enzymatic activity is promising for modulating an enzyme's activity with more nuance, but at present it has not been demonstrated how such an effect can be deliberately selected for.

Messenger RNA-based display[27] has several advantages over phage display for finding inhibitors because it does not require any in vivo steps to limit throughput or sequence space and it is more amenable to chemical modifications. In practice, this obviates the need for many subsequent optimisation or refinement steps, as is often required for phage display, and reduces (but does not remove) the chance of a selection yielding no useable sequences. An example selection carried out using this approach to find macrocyclic peptide amylase inhibitors revealed dramatically more potent inhibitors than those found by phage display methods.[28] The inhibitors found show exemplary selectivity for the target over a panel of other glycosidases, including other amylases.



Lemeng Chao received her bachelor degree in Science from Nankai University (China) in 2015. She then studied drug innovation in Utrecht University and obtained her MS degree in pharmaceutical sciences in 2017. She is currently a PhD student in the same university, in the group of Chemical Biology and Drug Discovery. Her research topic is synthetic neuraminidase inactivators.

Seino Jongkees is Assistant Professor in the department of Chemical Biology and Drug Discovery at the Utrecht Institute of Pharmaceutical Science, Utrecht University. He received his BSc and BA from Otago University (New Zealand), then pursued a PhD from the University of British Columbia working on the mechanistic dissection of an unusual glycoside hydrolase. Subsequently, he took up a JSPS post-doctoral fellowship at the University of Tokyo to receive training on peptide selection technologies and genetic code reprogramming. Research in the group is focused on the discovery of ligands for carbohydrate-active enzymes and carbohydrate-binding proteins, as well as adapting peptide display technology to interrogate glycobiology.

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

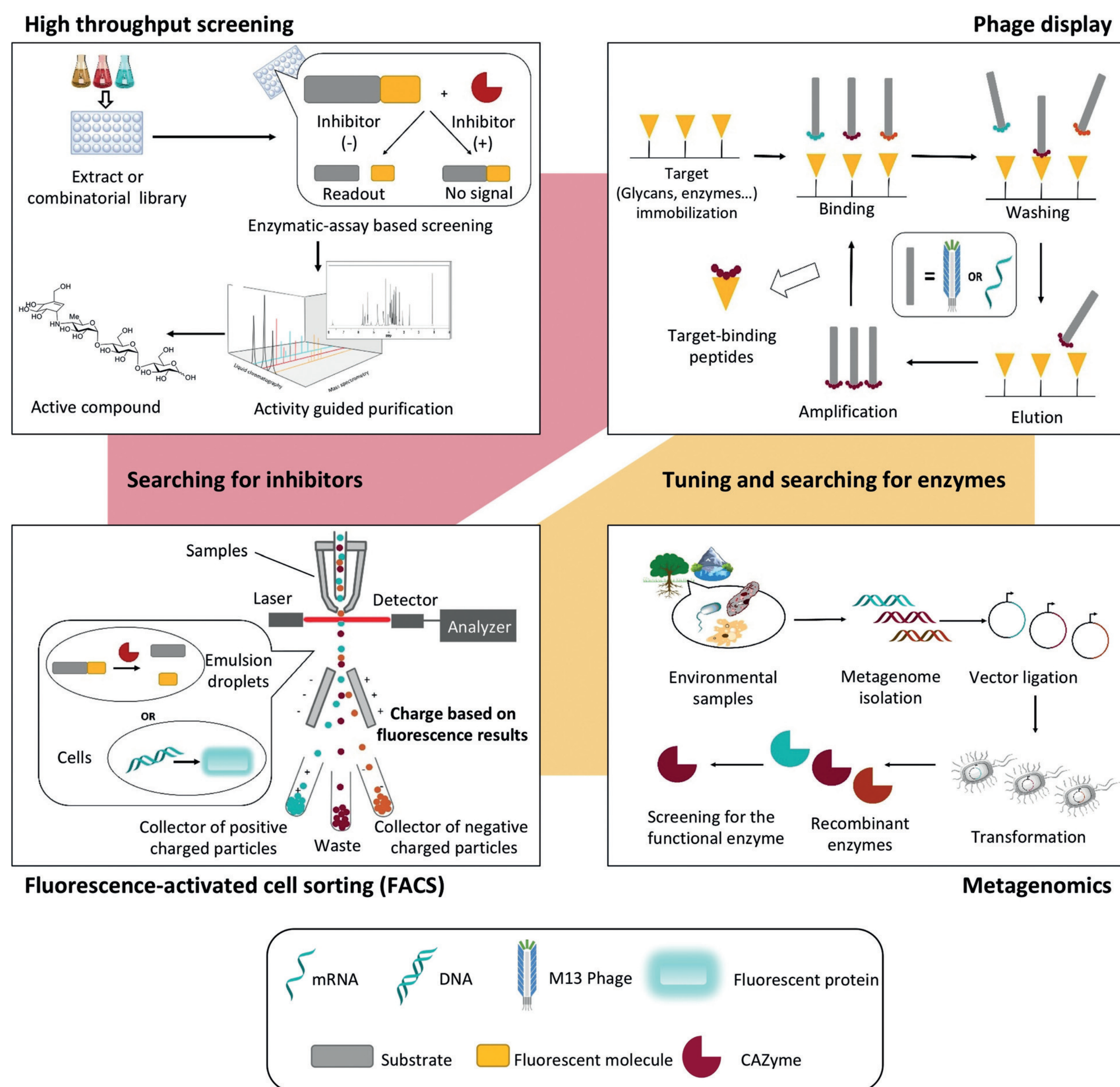


Figure 1. Overview of common high-throughput technologies and applications in the context of carbohydrate-active enzymes. In classical high-throughput screening with 96 well plates, activity of an enzyme is typically read out by a chromogenic or fluorogenic assay, and this activity is typically used to guide subsequent purification. In FACS, cells are sorted based on fluorescence of either a reporter assay or gene. In display technologies, a glycan or protein can be immobilised and phage or mRNA displaying a (poly)peptide can be panned to enrich for sequences that bind to this target. In metagenomics, DNA from the environment is isolated and ligated into a vector then transformed to bring about expression, which can be assayed for activity.

This approach also lends itself to facile genetic code reprogramming and post-translational modification because of its completely in vitro nature, expanding the scope of chemical space that can be explored and allowing incorporation of elements such as stable macrocyclization or *N*-methylation directly in the discovery phase.[29] In the amylase inhibitor selection mentioned, the most potent hit (high-picomolar K_i values) contained a non-canonical L-dopa residue that was incorporated into the library during the initial selection (Figure 2C).[30]

These affinity-based approaches are also applicable to deriving inhibitors of carbohydrate-binding proteins. A phage-display-based approach looking for hemagglutinin-binding peptides found surprisingly short sequences that were able to bind with low-nanomolar affinity to both H1 and H3 (Figure 2D),[31] while stearoyl derivatives of these sequences also proved effective in preventing influenza infection of MDCK cells with low-micromolar IC_{50} values. This lipid tail induces self-assembly, with the resultant multivalent presentation of the peptide to virus particles increasing efficacy.[32] It is interesting to note that the key peptide residues involved in haemagglutinin binding appear to be two arginine residues and a leucine residue, with these positively charged amino acids counter-intuitively binding to a receptor site for

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

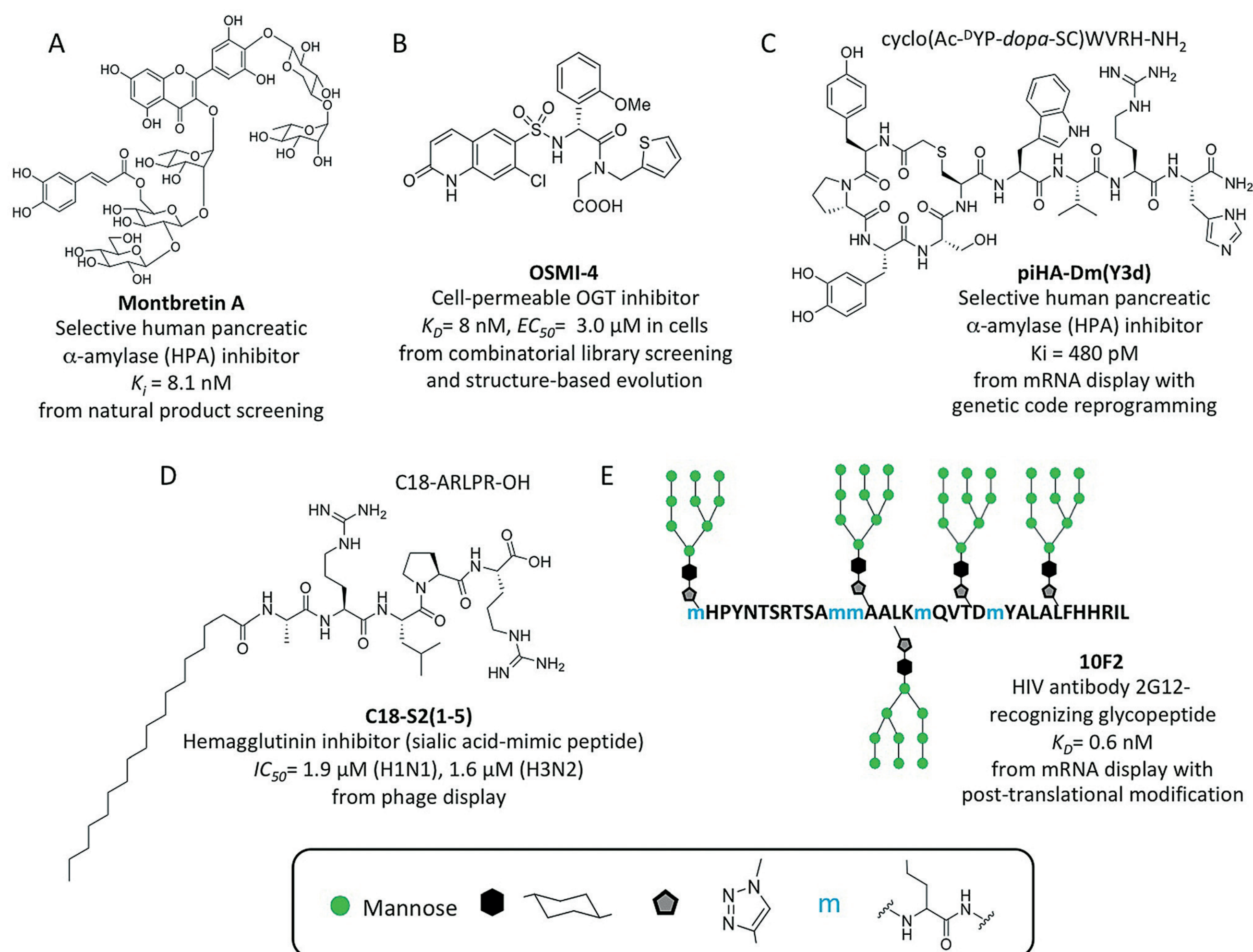


Figure 2. Example inhibitors of carbohydrate-active enzymes and carbohydrate-binding proteins derived from recent innovations in high-throughput approaches. See Refs. [17], (A); [23], (B); [30], (C); [31], (D); [40], (E).

a negatively charged sugar. In this case, the sialic acid is not making charge contacts with the protein, the carboxylate only making hydrogen bonds, and so this does not give a charge mismatch. Moreover, the binding pocket contains a glutamate residue available for interaction with these positively charged residues, but docking suggested it is not engaged in this way. Similarly, phage display also revealed sequences with mid-nanomolar IC_{50} binding to cholera toxin B, in this case using predominantly aromatic amino acids. Notably, these peptides were also able to prevent both toxin adsorption on cells and intracellular signal transduction as measured by cAMP levels.[33] These studies illustrate the diversity of binding modes available to glycan-binding sites and serve to reinforce the difficulty in predicting the results of such high-throughput approaches.

A promising recent development is the use of display technologies for the discovery of glycopeptides. In these methods, a carbohydrate attached to the peptide can be used to direct binding to the correct site, while the peptide portion can provide additional affinity and selectivity. Several approaches for attaching glycans to peptide libraries have been reported, including by disulfide linkage to cysteines,[34] 1-thiosugar addition to dehydroalanine,[35] sugar hydroxylamine addition to N-terminal aldehydes from periodate oxidation of serine,[36] and cysteine cross-linking with dichloroacetone pre-reacted with sugar hydroxylamines.[37] These reactions need to be extremely selective to be suitable for phage or mRNA displayed libraries without damaging the requisite phage particle or nucleic acid tag. In recent examples from the Derda group, phage-derived short glycopeptides were found to increase both affinity and selectivity relative to the sugar alone.[36,38] Of particular note is that even a low-affinity glycan can be used to target the site of binding, thus increasing the chance of discovering a glycopeptide with the desired biological activity.[39] An mRNA display selection by the Krauss group has also revealed an optimised peptide scaffold on which high-mannose glycans could be attached by a click reaction (Figure 2E), thereby presenting them in a way that binds to a broadly neutralising antibody with high potency, similar to the native interaction.[40] While in this instance the goal was not an inhibitor, the example nonetheless illustrates the power of such a glycopeptide selection approach in providing a 3D scaffold for glycan presentation. Such an approach has not yet been applied to carbohydrate-active

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

enzymes, only glycan-binding proteins, but the ability to target the sugar-binding site through attachment of a non-hydrolysable substrate analogue such as a thioglycoside should help to ensure competitive inhibition.

3. Improving or Altering Existing Enzymes

Carbohydrate-active enzymes that efficiently construct, remodel, or decompose glycan structures are in high demand in both academic research and industrial production. Apart from the type of reaction catalysed, enzymes that have other properties such as tolerance of high temperatures, acidic or basic environments, and organic co-solvents are also in demand. Enzymes with these characteristics can be created by directed evolution, as long as there exists a starting point with a low level of initial activity on which to build. At its core, directed evolution involves mutating or recombining a gene to create a library of mutants, followed by screening or selection methods to find improved versions.[41] Directed evolution allows the engineering of enzymes without the need for a complete understanding of the enzyme structure or mechanism, although it can be rationally guided, particularly in the library generation stage. However, for this to be successful, an assay is required that accurately reflects the desired properties in a readout with appropriately high throughput. By far the most convenient and sensitive such readout for many hydrolytic enzymes is through cleavage of a fluorescent substrate, as illustrated in Figure 3A, but label-free approaches that allow measurement of rate for a native substrate can have advantages such as in screening for turnover of the appropriate aglycone. These can be based on a reducing sugar assay[42] or mass spectrometry,[43] for example.

Directed evolution using plate-based screening approaches is a tried-and-true approach, but it is limited in its throughput by its requirement for one-by-one testing of mutants. Nonetheless, in cases in which the library size can be limited, such as through the use of rational guidance, it can prove very effective. In one example, error prone PCR and saturation mutagenesis of targeted residues in the $+1$ and $+2$ binding subsites of an enzyme capable of cleaving both type A and type B blood group antigens resulted in 170-fold improved processing of the minor type 1 A linkage, with only a minor cost in cleavage efficiency of other related linkages.[44] Because this enzyme cleaves several different substrates, a single coupled assay was employed that liberated a fluorophore only when an exo-acting β-galactosidase further cleaved the target enzyme's product (Figure 3B).

Fluorescent measurements of activity can also be achieved through a change in FRET signal, giving a ratiometric response that is more precise than a simple single-wavelength measurement. Such assays are most commonly applied to endo-acting enzymes, with a fluorophore attached to each end of the molecule. Examples include non-specific modification of heparin with DABCYL and EDANS fluorophores using NHS chemistry,[45] an endo-β-N-acetylglucosaminidase assay with two simple dyes at the reducing and non-reducing ends,[46] and a ganglioside derivative with a lipophilic second fluorophore in the lipid portion,[47] which can be used to assay several different enzymes in a cellular setting (Figure 3C). For exo-acting glucosidases, a FRET-based substrate can offer an advantage in allowing sterically limited active site pockets to still generate a fluorescent signal, through the use of a bis-acetal at the anomeric position. Fragmentation on enzymatic cleavage separates the donor and acceptor fluorophores.[48]

Fluorescence-activated cell sorting (FACS) offers a much higher throughput screening method for using these fluorescent signals, on the order of 10^7 per hour. In this, a difference in fluorescent signal is the basis for the sorting of cells.[49,50] Aside from the direct assays mentioned, the change in fluorescence signal can be achieved by a fluorescent reporter protein such as GFP, with expression under the control of a product-sugar inducible promoter.[51] To avoid biases from cell-based systems, enzymes can also be generated in situ from the encoding DNA by a cell-free transcription and translation system inside droplet emulsions (typically water/oil or water/oil/water), then reacted with substrate and sorted by flow cytometry[52] or in a microfluidic system.[53,54] In vitro compartmentalization based fluorescence-activated cell sorting (IVC-FACS),[55] encapsulating a cell within such an emulsion droplet, allows the evolution of surface-displayed[56] and secreted enzymes,[57] otherwise not possible using FACS.

With microfluidics-based systems, care must be taken to minimise cross-talk between droplets, which typically occurs by diffusion of lipophilic fluorescent products. By modifying the substrate with a charged group such as a sulfate on the fluorophore,[58] or by generating a sulfate on fragmentation after cleavage,[59] encapsulation of the fluorescent signal is ensured and discrimination of small differences is improved. In an alternative approach, a quinone methide reactive group can be generated upon enzyme-catalysed hydrolysis, which attaches to nearby nucleophiles for covalent immobilisation of the attached fluorophore, while fragmentation to generate this quinone methide simultaneously alleviates fluorophore quenching (Figure 3D).[60]

Evolution based on activated synthetic substrates risks optimising the enzyme for activity on the wrong substrate. In a label-free approach intended to overcome this,[61] cellulase activity on the natural carboxymethyl cellulose substrate yields glucose monomers, which hexose oxidase then converts to lactones while also producing hydrogen peroxide. This peroxide is subsequently used in oxidation of 3-carboxy-7-(4′-aminophenoxy) coumarin by vanadium bromoperoxidase to produce a fluorescent coumarin derivative (Figure 3E). The strength of the fluorescent signal is thus correlated to the cellulase activity but is applicable to many different hydrolases that generate free reducing sugars. Using this approach in fluorescence-activated cell sorting (FACS) with double emulsion technology, a 12-fold enrichment of cellulase-expressing cells was achieved after one sorting round. While broadly applicable and relevant to the true biological reaction to be catalysed, such multistep assays do carry the risk of enriching for changes that influence the final readout directly.

As with inhibitor screening, a particular challenge in the application of FACS for the evolution of glycosyl transferases is the lack of efficient direct screening systems for sugar

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

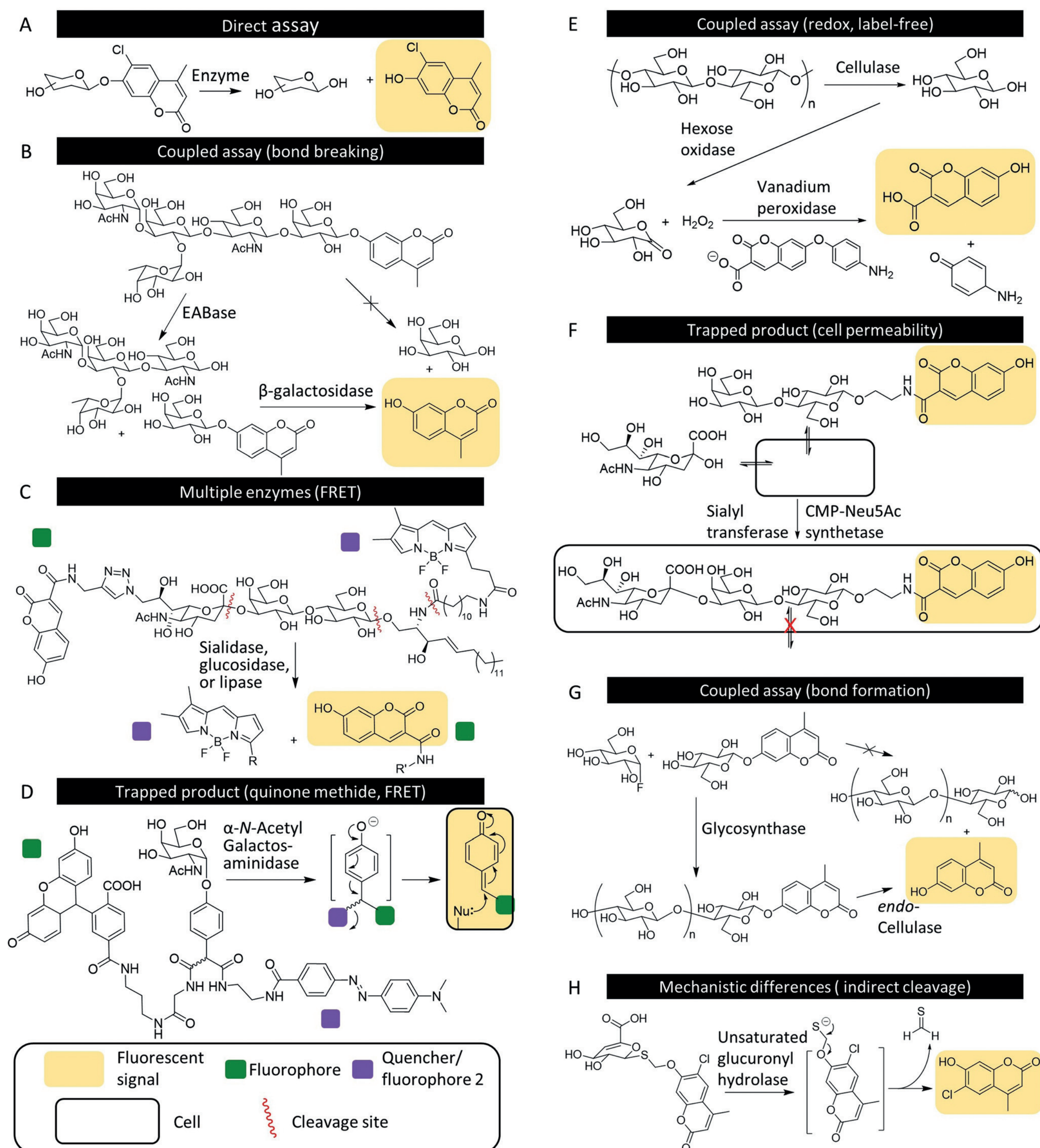


Figure 3. Innovative strategies to screen for carbohydrate-active enzyme activity with generation of a fluorescent signal A) by direct hydrolysis, B) by coupled exo-glycosidase activity (Ref. [44]), C) by change in FRET signal from one or more cleavage sites (Ref. [47]), D) by fragmentation to relieve quenching and generate a reactive fluorophore electrophile (Ref. [60]), E) by coupled assay with redox enzymes (Ref. [61]), F) by generation of a product not recognised by transporters (Ref. [62]), G) by coupling to an endo-glycosidase activity (Ref. [68]), and (H) by requiring initial cleavage of a thioglycosidic linkage (Ref. [83]).

transfer activity. Aside from the methods listed in the previous section, alternatives exist for FACS that take advantage of product entrapment. In one example, the formation of sialosides inside *Escherichia coli* led to selective trapping of the fluorescently labelled transfer products, allowing analysis and sorting of the resulting cell population using FACS (Figure 3 F).[62] Evolution of improved CstII sialyl transferase activity was carried out using a fluorescent bodipy-labelled derivative of lactose as an acceptor substrate. The bodipy-lactose and sialic acid are efficiently transported

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

into the cytoplasm of *E. coli* cells, while a plasmid-encoded CMP-sialic acid synthetase gene was introduced to allow its in situ generation. The sialylated fluorescent product is no longer recognised by the transporters, leaving the cells fluorescent to levels proportional to the CstII activity. A library of 4106 ST mutants was screened using this method and a variant with up to 400-fold higher catalytic efficiency was found for transfer to a variety of fluorescently labelled acceptor sugars.

A further step in directed evolution complexity is enabled by a novel application of phage display, wherein a phage protein displays a desired glycan. In this approach, a glycosyl acceptor peptide is made as a fusion with an M13 coat protein, while enzymes encoded on another plasmid synthesise and transfer an oligosaccharide onto this fusion, ultimately giving phage display of *N*-linked glycans.[63] In this way a genotype-phenotype link can be established between the phage-displayed glyco-epitope and the phagemid-encoded genes for any of the three essential components of the glycosylation process (the glycan synthesis pathway, an oligosaccharyl transferase, and an acceptor protein). Such a system shows great promise in allowing the engineering of entire pathways based on the display of the final glycan product, enriched through pull-down by lectins or antibodies. Phages also offer a unique opportunity in the powerful phage-based continuous evolution (PACE) technique, which continuously feeds host bacteria to a phage pool under selective pressure.[64] While this does not appear to have been applied to carbohydrate-active enzymes to date, a combination of such phage-based approaches would offer a truly exceptional method for the synergistic engineering of multiple glycan synthesis enzymes in a pathway.

Generation of efficient enzymes with completely new activities can also be driven by enzyme evolution processes.[65] The glycosynthases, for example, are mutant retaining glycosidases in which the catalytic nucleophile has been mutated to a smaller residue, allowing the catalysis of glycosidic bond formation between an acceptor sugar alcohol and a synthetic activated donor such as a glycosyl fluoride.[66,67] While this initial mutation can be introduced rationally, bond formation can be further optimised through directed evolution using high-throughput approaches. An early plate-based approach that detected bond formation through a coupled assay with an exclusively endo-acting cellulase (Figure 3G) was used to screen a library generated by saturation mutagenesis of the catalytic nucleophile, identifying glycine and cysteine mutants as alternative viable substitutions.[68] A broader, but still targeted, library generated by saturation mutagenesis of residues around the substrate binding pocket revealed mutants permissive of 3-*O*-methylation of the glycosyl fluoride donor sugar.[69] In a more general plate-based assay, colour change from a pH indicator detects generation of HF from the turnover of any glycosyl fluoride donor. Using this method, a library generated by error-prone PCR revealed mutants with increased k_{cat} values, but this came at a cost of increased K_m values.[70] Such approaches hold further promise when combining the power of higher throughput approaches such as cell sorting with different library generation approaches, such as random mutagenesis together with rational library design guided by an understanding of carbohydrate-active enzymology.

4. Searching for Novel Enzymes

Directed evolution can in some cases be used to guide known enzymatic activities in new directions, but in other cases it may be more efficient to find an already optimised catalyst from nature. Metagenomics is one particularly successful high-throughput approach used to discover novel enzymes from micro-organisms.[71] Over the two decades leading up to 2016, 861 novel glycosidases were found from metagenomic libraries by function-based screening while 96 were found by sequence-based screening.[72]

At the genetic level, transcriptomics of bacteria from the gut microbiome is one very successful recent example, revealing carbohydrate-active enzymes in known families as well as others with new substrates and mechanisms. Using such an approach, the genes activated during growth on yeast α-mannans revealed the pathway by which gut bacteria, especially *Bacteriodes* species, are able to degrade this food source in a way that controls the extent of extracellular depolymerisation, and thus maximises the benefit for the expressing microorganism.[73] A similar approach applied to the more complex plant glycan rhamnogalacturonan-II revealed a more diverse array of enzymes,[74] including founding members of seven new families in the carbohydrate-active enzymes database (http://www.cazy.org/).[75,76]

A function-based screen can be a relatively simple set-up, such as glycosidase activity selected for on agar plates supplemented with natural or synthetic substrates (e.g., the absence of starch colouration by iodine around colonies with active amylase). Studies using this method have been applied to discover enzymes with cellulose,[77] amylase,[78] and xylanase[79] activity, for example. Although low-tech functional screening requires no special devices and can be performed at high throughput, signals are often faint and this has been suggested as a reason for the common low hit rates.[80] Aiming to improve the sensitivity of the assays, function-based screening has been performed by fluorogenic enzymatic assays similar to those described for inhibitor screening and enzyme engineering. For example, the Withers group synthesized a set of nine 6-chloro-4-methylumbelliferyl glycosides as fluorogenic reagents to screen for different glycosidase activities.[81] An illustrative screen of a mining bioreactor fosmid library, which was known to contain glycosidase activity, was carried out using four different substrates in one reaction. Hits were subsequently deconvoluted to identify which specific activity was present, which allowed throughput to further be increased as all of the first-pass screens to identify clones with any activity could be combined in a single run.

Such an approach based on fluorogenic substrates is typically able to find enzymes with the desired activity but for common reactions often will find the same or similar enzymes each time. In order to guide a screen to find more esoteric enzymes, especially those that operate through a different mechanism,[82] a set of substrates were developed that contain

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

an additional linker between the sugar and the fluorophore (Figure 3H).[83] Important in this design is that fluorescence is generated only on cleavage of a thioglycosidic bond, which is typically intractable to glycoside hydrolases that operate through the more common retaining or inverting mechanisms. Application of these substrates to a small test metagenomic library derived from beaver fecal matter showed clear discrimination between desired and undesired activities, while simple substrates without the thioglycosidic immolative linker showed false positives from side activities of more common enzymes. It is worth noting, however, that while this substrate is cleverly designed to find certain types of new reactivities, it has its own limitations in which types of new mechanisms it could find. Surprises can also still be found using more conventional approaches not tailored specifically for finding new classes of enzymes, such as a new type of inverting glycosidase that appears to not be dependent on the usual two catalytic carboxylate residues, as revealed by the human gut microbiome transcriptomics approach of the Gilbert group.[84]

To find enzymes with improved tolerances, rather than a different substrate scope or mechanism, the source of the metagenomic library is perhaps more important than the assay used. For example, to discover enzymes with thermally stable amylase activity, a fosmid library from the microbial population of a deep-sea hydrothermal vent[85] was screened at the DNA level by random sequencing and homology searching.[86] Primers derived from this screen allowed identification of a new member of CaZY family GH57 that showed maximal amylase activity at 90 °C, and stability at this temperature for several hours.

In order to target a screen at discovery of only secreted enzymes, such as for incorporation into a bioreactor, an altered phage display system can be used. By deleting the gene for the pIII protein from the helper phage, as well as the pIII signal sequence in the library phagemid, the only properly assembled phage particles will derive from library members that provide their own signal sequence, which are typically secreted or cell surface proteins.[87] This system was applied to a rumen microbial community library of with a diversity of circa 5×10^6, with the only selection criterion being properly assembled phage particles as assessed by stability to ionic detergent treatment. Analysing enriched sequences by high-throughput sequencing revealed a diverse set of extracellular glycoside hydrolases, polysaccharide lyases, and carbohydrate esterases.

Where mechanism-based inactivators provide a rational approach to the design of inhibitors, their derivative activity-based probes offer a more rationally targeted approach to finding enzymes, through modification of covalent inactivators with tags such as biotin for recovery and sequencing or fluorophores for localisation and visualisation in complex mixtures.[88] Approaches include 2-fluoro[89] or 5-fluoro sugars,[90] cyclophellitol epoxide[91,92] and aziridine derivatives,[93] or more recently described compounds that generate allylic and cyclopropyl cations[94] (Figure 4). While extremely powerful tools, these approaches are exclusively suited for use with retaining glycosidases, as they form a covalent intermediate with the substrate during hydrolysis. An alternative approach, which is applicable to any hydrolase, is through release of an

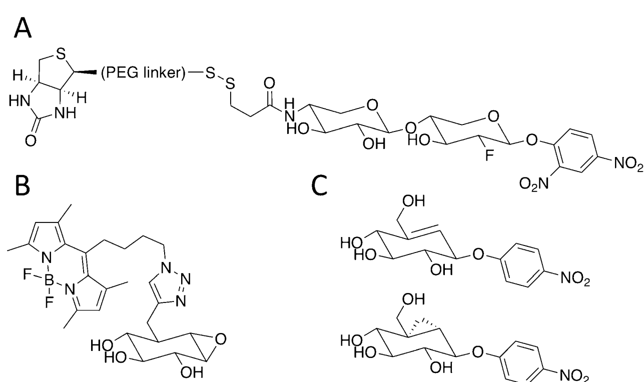


Figure 4. Activity-based probes and mechanism-based inhibitors based on A) a 2-fluoro sugar targeting cellulase activity for pull-down and sequencing (Ref. [89]), B) a cyclophellitol derivative for glucocerebrosidase activity visualisation (Refs. [91],[92]), and C) allylic and cyclopropyl cations for yeast α-glucosidase inactivation (Ref. [94]). Carbocyclic sugar analogues containing sp^2 centers or fused rings shown as chair structures for simplicity.

electrophile from the aglycone that becomes activated on hydrolysis, such as the quinone methide shown in Figure 3D. The labelling that arises in this case is less specific to the enzyme responsible for the activity, as diffusion can occur before reaction, making enzyme identification more challenging.

5. Summary and Outlook

High-throughput approaches are driving new discoveries in many different aspects of the glycosciences. In searching for inhibitors, mature technologies such as plate-based screening of combinatorial and natural product libraries or phage display of peptides and small proteins have proven their worth, but innovative developments in how these techniques are applied will increase this further still. In particular, increasing throughput and expanding the chemical space accessible drives initial hits to ever-higher affinities and selectivities. For screening approaches, whether applied to discovery of inhibitors, evolution of existing enzymes, or discovery of new enzymes, creative approaches such as coupled assays and product trapping continue to be reported for reliably coupling only the desired activity to a convenient read-out, in ever-higher sensitivity, throughput, and discrimination. Progression to assays amenable to cell sorting has opened up a new realm of possibilities in the scale of libraries that can be meaningfully investigated, also for secreted enzymes or reactions such as glycan synthesis. New sources of enzymes, particularly the unculturable microbiota represented in metagenomic libraries, present new starting points for evolutionary trajectories, as well as providing useful enzymes in their own right. The future of high-throughput approaches lies in ever-higher throughput and also in clever chemistry and molecular biology to generate meaningful questions to answer and useful collections to investigate, often yielding answers that would never be considered in a rational approach.

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Acknowledgements

We are grateful for funding support provided by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 746631. Professor Geert-Jan Boons is thanked for critical reading and feedback.

Conflict of interest

The authors declare no conflict of interest.

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 12750–12760
Angew. Chem. **2019**, *131*, 12880–12890

[1] T. M. Gloster, G. J. Davies, *Org. Biomol. Chem.* **2010**, *8*, 305–320.
[2] N. Asano, *Cell. Mol. Life Sci.* **2009**, *66*, 1479–1492.
[3] N. Asano, *Glycobiology* **2003**, *13*, 93R–104R.
[4] V. L. Schramm, *Annu. Rev. Biochem.* **2011**, *80*, 703–732.
[5] P. Compain, O. R. Martin, *Bioorg. Med. Chem.* **2001**, *9*, 3077–3092.
[6] L. Tedaldi, G. K. Wagner, *MedChemComm* **2014**, *5*, 1106–1125.
[7] T. M. Gloster, W. F. Zandberg, J. E. Heinonen, D. L. Shen, L. Deng, D. J. Vocadlo, *Nat. Chem. Biol.* **2011**, *7*, 174–181.
[8] P. A. Videira, F. Marcelo, R. K. Grewal, in *Carbohydr. Chem. Chem. Biol. Approaches*, *Vol. 43*, The Royal Society of Chemistry, London, **2017**.
[9] R. M. Schmaltz, S. R. Hanson, C. H. Wong, *Chem. Rev.* **2011**, *111*, 4259–4307.
[10] L. T. Mika, E. Cséfalvay, Á. Németh, *Chem. Rev.* **2018**, *118*, 505–613.
[11] F. Higel, A. Seidl, F. Sörgel, W. Friess, *Eur. J. Pharm. Biopharm.* **2016**, *100*, 94–100.
[12] N. F. Brás, N. M. Cerqueira, M. J. Ramos, P. A. Fernandes, *Expert Opin. Ther. Pat.* **2014**, *24*, 857–874.
[13] R. P. Hertzberg, A. J. Pope, *Curr. Opin. Chem. Biol.* **2000**, *4*, 445–451.
[14] T. M. Gloster, D. J. Vocadlo, *Nat. Chem. Biol.* **2012**, *8*, 683–694.
[15] C. Tysoe, L. K. Williams, R. Keyzers, N. T. Nguyen, C. Tarling, J. Wicki, E. D. Goddard-Borger, A. H. Aguda, S. Perry, L. J. Foster, et al., *ACS Cent. Sci.* **2016**, *2*, 154–161.
[16] C. Tysoe, S. G. Withers, *Biochemistry* **2018**, *57*, 5384—5387.
[17] C. A. Tarling, K. Woods, R. Zhang, H. C. Brastianos, G. D. Brayer, R. J. Andersen, S. G. Withers, *ChemBioChem* **2008**, *9*, 433–438.
[18] L. K. Williams, X. Zhang, S. Caner, C. Tysoe, N. T. Nguyen, J. Wicki, D. E. Williams, J. Coleman, J. H. McNeill, V. Yuen, et al., *Nat. Chem. Biol.* **2015**, *11*, 691–696.
[19] J. Ryu, M. S. Eom, W. Ko, M. S. Han, H. S. Lee, *Bioorg. Med. Chem.* **2014**, *22*, 2571–2575.
[20] Z. L. Wu, C. M. Ethen, B. Prather, M. MacHacek, W. Jiang, *Glycobiology* **2011**, *21*, 727–733.
[21] B. J. Gross, B. C. Kraybill, S. Walker, *J. Am. Chem. Soc.* **2005**, *127*, 14588–14589.
[22] R. F. Ortiz-Meoz, J. Jiang, M. B. Lazarus, M. Orman, J. Janetzko, C. Fan, D. Y. Duveau, Z. W. Tan, C. J. Thomas, S. Walker, *ACS Chem. Biol.* **2015**, *10*, 1392–1397.
[23] S. E. S. Martin, Z.-W. Tan, H. M. Itkonen, D. Y. Duveau, J. A. Paulo, J. Janetzko, P. L. Boutz, L. Törk, F. A. Moss, C. J. Thomas, et al., *J. Am. Chem. Soc.* **2018**, *140*, 13542–13545.
[24] X. Zhang, F. Chen, A. Petrella, F. Chacón-Huete, J. Covone, T.-W. Tsai, C.-C. Yu, P. Forgione, D. H. Kwan, *ACS Chem. Biol.* **2019**, *14*, 715–724.
[25] C. Heinis, G. Winter, *Curr. Opin. Chem. Biol.* **2015**, *26*, 89–98.
[26] I. Roskar, P. Molek, M. Vodnik, M. Stempelj, B. Strukelj, M. Lunder, *J. Diabetes Invest.* **2015**, *6*, 625–631.
[27] R. Obexer, L. J. Walport, H. Suga, *Curr. Opin. Chem. Biol.* **2017**, *38*, 52–61.
[28] S. A. K. Jongkees, S. Caner, C. Tysoe, G. D. Brayer, S. G. Withers, H. Suga, *Cell Chem. Biol.* **2017**, *24*, 381–390.
[29] T. Passioura, H. Suga, *Chem. Commun.* **2017**, *53*, 1931–1940.
[30] R. Yoshisada, L. van Gijzel, S. A. K. Jongkees, *ChemBioChem* **2017**, *18*, 2333–2339.
[31] T. Matsubara, A. Onishi, T. Saito, A. Shimada, H. Inoue, T. Taki, K. Nagata, Y. Okahata, T. Sato, *J. Med. Chem.* **2010**, *53*, 4441–4449.
[32] T. Matsubara, M. Sumi, H. Kubota, T. Taki, Y. Okahata, T. Sato, *J. Med. Chem.* **2009**, *52*, 4247–4256.
[33] R. K. Yu, S. Usuki, Y. Itokazu, H. C. Wu, *Glycobiology* **2016**, *26*, 63–73.
[34] K. Arai, H. Tsutsumi, H. Mihara, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4940–4943.
[35] S. A. K. Jongkees, S. Umemoto, H. Suga, *Chem. Sci.* **2017**, *8*, 1474–1481.
[36] S. Ng, E. Lin, P. I. Kitov, K. F. Tjhung, O. O. Gerlits, L. Deng, B. Kasper, A. Sood, B. M. Paschal, P. Zhang, et al., *J. Am. Chem. Soc.* **2015**, *137*, 5248–5251.
[37] S. Ng, R. Derda, *Org. Biomol. Chem.* **2016**, *14*, 5539–5545.
[38] Y. Chou, E. N. Kitova, M. Joe, R. Brunton, T. L. Lowary, J. S. Klassen, R. Derda, *Org. Biomol. Chem.* **2018**, *16*, 223–227.
[39] D. F. Vinals, P. I. Kitov, Z. Tu, C. Zou, C. W. Cairo, H. C.-H. Lin, R. Derda, *Pept. Sci.* **2019**, *111*, e24097.
[40] S. Horiya, J. K. Bailey, J. S. Temme, Y. V. Guillen Schlippe, I. J. Krauss, *J. Am. Chem. Soc.* **2014**, *136*, 5407–5415.
[41] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chem. Int. Ed.* **2015**, *54*, 3351–3367; *Angew. Chem.* **2015**, *127*, 3408–3426.
[42] E. Hammond, C. P. Li, V. Ferro, *Anal. Biochem.* **2010**, *396*, 112–116.
[43] G. Nagy, T. Peng, N. L. B. Pohl, *Anal. Chem.* **2016**, *88*, 7183–7190.
[44] D. H. Kwan, I. Constantinescu, R. Chapanian, M. A. Higgins, M. P. Kötzler, E. Samain, A. B. Boraston, J. N. Kizhakkedathu, S. G. Withers, *J. Am. Chem. Soc.* **2015**, *137*, 5695–5705.
[45] J. C. Sistla, S. Morla, A. H. B. Alabbas, R. C. Kalathur, C. Sharon, B. B. Patel, U. R. Desai, *Carbohydr. Polym.* **2019**, *205*, 385–391.
[46] N. Ishii, C. Sunaga, K. Sano, C. Huang, K. Iino, Y. Matsuzaki, T. Suzuki, I. Matsuo, *ChemBioChem* **2018**, *19*, 660–663.
[47] G. Y. Yang, C. Li, M. Fischer, C. W. Cairo, Y. Feng, S. G. Withers, *Angew. Chem. Int. Ed.* **2015**, *54*, 5389–5393; *Angew. Chem.* **2015**, *127*, 5479–5483.
[48] S. Cecioni, D. J. Vocadlo, *J. Am. Chem. Soc.* **2017**, *139*, 8392–8395.
[49] M. Wójcik, A. Telzerow, W. J. Quax, Y. L. Boersma, *Int. J. Mol. Sci.* **2015**, *16*, 24918–24945.
[50] G. Yang, S. G. Withers, *ChemBioChem* **2009**, *10*, 2704–2715.
[51] L. F. Ribeiro, J. Tullman, N. Nicholes, S. R. B. Silva, D. S. Vieira, M. Ostermeier, R. J. Ward, *Biotechnol. Biofuels* **2016**, *9*, 119.
[52] G. Körfer, C. Pitzler, L. Vojcic, R. Martinez, U. Schwaneberg, *Sci. Rep.* **2016**, *6*, 26128.
[53] E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, A. D. Griffiths, *Chem. Biol.* **2005**, *12*, 1291–1300.
[54] A. Autour, M. Ryckelynck, *Micromachines* **2017**, *8*, 128.
[55] O. J. Miller, K. Bernath, J. J. Agresti, G. Amitai, B. T. Kelly, E. Mastrobattista, V. Taly, S. Magdassi, D. S. Tawfik, A. D. Griffiths, *Nat. Methods* **2006**, *3*, 561–570.
[56] F. Ma, Y. Xie, C. Huang, Y. Feng, G. Yang, *PLoS One* **2014**, *9*, e89785.
[57] R. Tu, R. Martinez, R. Prodanovic, M. Klein, U. Schwaneberg, *J. Biomol. Screening* **2011**, *16*, 285–294.

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[58] M. Najah, E. Mayot, I. P. Mahendra-Wijaya, A. D. Griffiths, S. Ladame, A. Drevelle, *Anal. Chem.* **2013**, *85*, 9807 – 9814.
[59] F. Ma, M. Fischer, Y. Han, S. G. Withers, Y. Feng, G. Y. Yang, *Anal. Chem.* **2016**, *88*, 8587 – 8595.
[60] K. Kalidasan, Y. Su, X. Wu, S. Q. Yao, M. Uttamchandani, *Chem. Commun.* **2013**, *49*, 7237 – 7239.
[61] R. Ostafe, R. Prodanovic, U. Commandeur, R. Fischer, *Anal. Biochem.* **2013**, *435*, 93 – 98.
[62] A. Aharoni, K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. J. Strynadka, W. W. Wakarchuk, S. G. Withers, *Nat. Methods* **2006**, *3*, 609 – 614.
[63] E. Çelik, A. C. Fisher, C. Guarino, T. J. Mansell, M. P. DeLisa, *Protein Sci.* **2010**, *19*, 2006 – 2013.
[64] K. M. Esvelt, J. C. Carlson, D. R. Liu, *Nature* **2011**, *472*, 499 – 503.
[65] D. H. Kwan, S. G. Withers, *J. Carbohydr. Chem.* **2011**, *30*, 181 – 205.
[66] L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583 – 5584.
[67] M. R. Hayes, J. Pietruszka, *Molecules* **2017**, *22*, 1434.
[68] C. Mayer, D. L. Jakeman, M. Mah, G. Karjala, L. Gal, R. A. J. Warren, S. G. Withers, *Chem. Biol.* **2001**, *8*, 437 – 443.
[69] J. H. Shim, H. M. Chen, J. R. Rich, E. D. Goddard-Borger, S. G. Withers, *Protein Eng. Des. Sel.* **2012**, *25*, 465 – 472.
[70] A. Ben-David, G. Shoham, Y. Shoham, *Chem. Biol.* **2008**, *15*, 546 – 551.
[71] T. Davids, M. Schmidt, D. Böttcher, U. T. Bornscheuer, *Curr. Opin. Chem. Biol.* **2013**, *17*, 215 – 220.
[72] M. Ferrer, M. Martínez-Martínez, R. Bargiela, W. R. Streit, O. V. Golyshina, P. N. Golyshin, *Microb. Biotechnol.* **2016**, *9*, 22 – 34.
[73] F. Cuskin, E. C. Lowe R. McLean, M. J. Temple, Y. Zhu, E. A. Cameron, N. A. Pudlo, N. T. Porter, K. Urs, A. J. Thompson, A. Cartmell, et al., *Nature* **2015**, *517*, 165 – 169.
[74] D. Ndeh, A. Rogowski, A. Cartmell, A. S. Luis, A. Baslé, J. Gray, I. Venditto, J. Briggs, X. Zhang, A. Labourel, et al., *Nature* **2017**, *544*, 65 – 70.
[75] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490 – D495.
[76] The CAZypedia Consortium, *Glycobiology* **2018**, *28*, 3 – 8.
[77] Y. Feng, C. J. Duan, H. Pang, X. C. Mo, C. F. Wu, Y. Yu, Y. L. Hu, J. Wei, J. L. Tang, J. X. Feng, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 319 – 328.
[78] K. Lämmle, H. Zipper, M. Breuer, B. Hauer, C. Buta, H. Brunner, S. Rupp, *J. Biotechnol.* **2007**, *127*, 575 – 592.
[79] Y. Hu, G. Zhang, A. Li, J. Chen, L. Ma, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 823 – 830.
[80] T. Uchiyama, K. Miyazaki, *Curr. Opin. Biotechnol.* **2009**, *20*, 616 – 622.
[81] H. M. Chen, Z. Armstrong, S. J. Hallam, S. G. Withers, *Carbohydr. Res.* **2016**, *421*, 33 – 39.
[82] S. A. K. Jongkees, S. G. Withers, *Acc. Chem. Res.* **2014**, *47*, 226 – 235.
[83] S. A. Nasseri, L. Betschart, D. Opaleva, P. Rahfeld, S. G. Withers, *Angew. Chem. Int. Ed.* **2018**, *57*, 11359 – 11364; *Angew. Chem.* **2018**, *130*, 11529 – 11534.
[84] J. Munoz-Munoz, A. Cartmell, N. Terrapon, B. Henrissat, H. J. Gilbert, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4936 – 4941.
[85] F. Wang, H. Zhou, J. Meng, X. Peng, L. Jiang, P. Sun, C. Zhang, J. D. Van Nostrand, Y. Deng, Z. He, et al., *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4840 – 4845.
[86] H. Wang, Y. Gong, W. Xie, W. Xiao, J. Wang, Y. Zheng, J. Hu, Z. Liu, *Appl. Biochem. Biotechnol.* **2011**, *164*, 1323 – 1338.
[87] D. Jankovic, M. A. Collett, M. W. Lubbers, J. Rakonjac, *Genome Biol.* **2007**, *8*, R266.
[88] K. T. Barglow, B. F. Cravatt, *Nat. Methods* **2007**, *4*, 822 – 827.
[89] O. Hekmat, Y.-W. Kim, S. J. Williams, S. He, S. G. Withers, *J. Biol. Chem.* **2005**, *280*, 35126 – 35135.
[90] J. D. McCarter, S. G. Withers, *J. Am. Chem. Soc.* **1996**, *118*, 241 – 242.
[91] M. D. Witte, W. W. Kallemeijn, J. Aten, K.-Y. Li, A. Strijland, W. E. Donker-Koopman, A. M. C. H. van den Nieuwendijk, B. Bleijlevens, G. Kramer, B. I. Florea, et al., *Nat. Chem. Biol.* **2010**, *6*, 907 – 913.
[92] L. Wu, J. Jiang, Y. Jin, W. W. Kallemeijn, C. Kuo, M. Artola, W. Dai, C. Van Elk, M. Van Eijk, G. A. Van Der Marel, et al., *Nat. Chem. Biol.* **2017**, *13*, 867 – 873.
[93] L. I. Willems, T. J. M. Beenakker, B. Murray, S. Scheij, W. W. Kallemeijn, R. G. Boot, M. Verhoek, W. E. Donker-Koopman, M. J. Ferraz, E. R. van Rijssel, et al., *J. Am. Chem. Soc.* **2014**, *136*, 11622 – 11625.
[94] S. Shamsi Kazem Abadi, M. Tran, A. K. Yadav, P. J. P. Adabala, S. Chakladar, A. J. Bennet, *J. Am. Chem. Soc.* **2017**, *139*, 10625 – 10628.

Manuscript received: January 2, 2019
Revised manuscript received: March 5, 2019
Accepted manuscript online: March 26, 2019
Version of record online: July 17, 2019

 © 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim