

Selective Metal-Site-Guided Arylation of Proteins

Jens Willwacher, Ritu Raj, Shabaz Mohammed, Benjamin G. Davis*

Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK

Supporting Information Placeholder

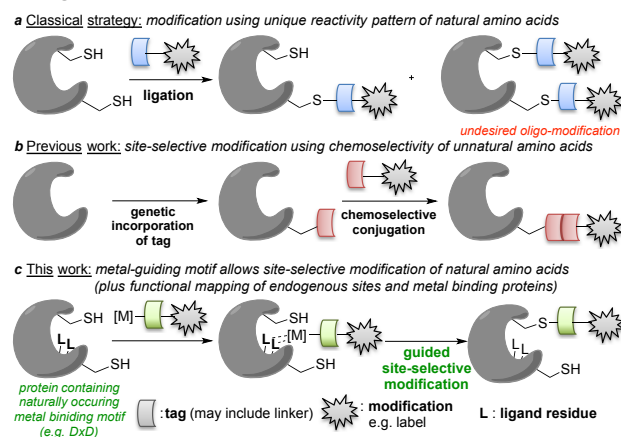
ABSTRACT: We describe palladium-mediated *S*-arylation that exploits natural metal-binding motifs to ensure high site-selectivity for a proximal reactive residue. This allows the chemical identification not only of proteins that bind metals but the environment of the metal-binding site itself through proteomic analysis of arylation sites. The transformation is easy to perform under standard conditions, does not require the isolation of a reactive Ar-Pd complex, is broad in scope and is applicable in cell lysates as well as to covalent inhibition/modulation of metal-dependent enzymatic activity.

Post-translational modification is nature's method to decorate proteins (typically enzymatically) with structurally diverse functional 'switches' and recognition sites.¹ Its advantage over the currently available chemical 'toolbox' for protein chemistry is its ability to carry out such reactions in a highly site-selective manner.² Whilst this is, in part, achieved via chemoselectivity, it is often also guided by mutual recognition of secondary structure in either the protein substrate or the modifying enzyme catalyst. This guides regioselection and enables selectivity for certain residues, often related to or guided by function. Whilst regioselectivity directed simply by accessibility has proved a potentially successful approach in chemical protein modification,³ the ability to direct a protein-modifying *chemical* catalytic centre by virtue of functionally inherent motifs might allow regioselection guided by endogenous features in a manner that partially mimics nature's approach. Here we show, as a proof-of-principle, a designed method for site-selective protein modification that appears to rely, at least in part, upon natural metal-binding.

Several strategies for chemical protein modification have been developed;^{4,5} a classic variant relies on (sometimes partial) chemoselectivity of certain natural side-chains (e.g. lysine (Lys) / cysteine (Cys), Scheme 1a).⁶ For example, the amine/thiol groups of these residues can exhibit higher nucleophilicity towards some electrophiles. Most of these methods rely on solvent accessibility/exposure of particular Lys- or Cys-residues for regioselectivity, and often generate heterogeneous product mixtures. To reduce such heterogeneity in classical protein-conjugates, conceptually different site-selective modification strategies have emerged. These often rely on the incorporation of functional groups with beneficial reactivity ('tags') for improved chemoselection and hence regioselection, if reaction is complete. These include genetically-, enzymatically- and chemically- installed unnatural amino acids (UAAs)^{7,8} with 'bioorthogonal' reactivity and enhanced chemoselectivity with an appropriate reagent (Scheme 1b).^{9,10} Metal-mediated examples include azide-alkyne 'cycloadditions',^{11,12} cross-couplings,¹³ and olefin cross-metathesis.¹⁴ Elegant, prior, metal-guided covalent protein

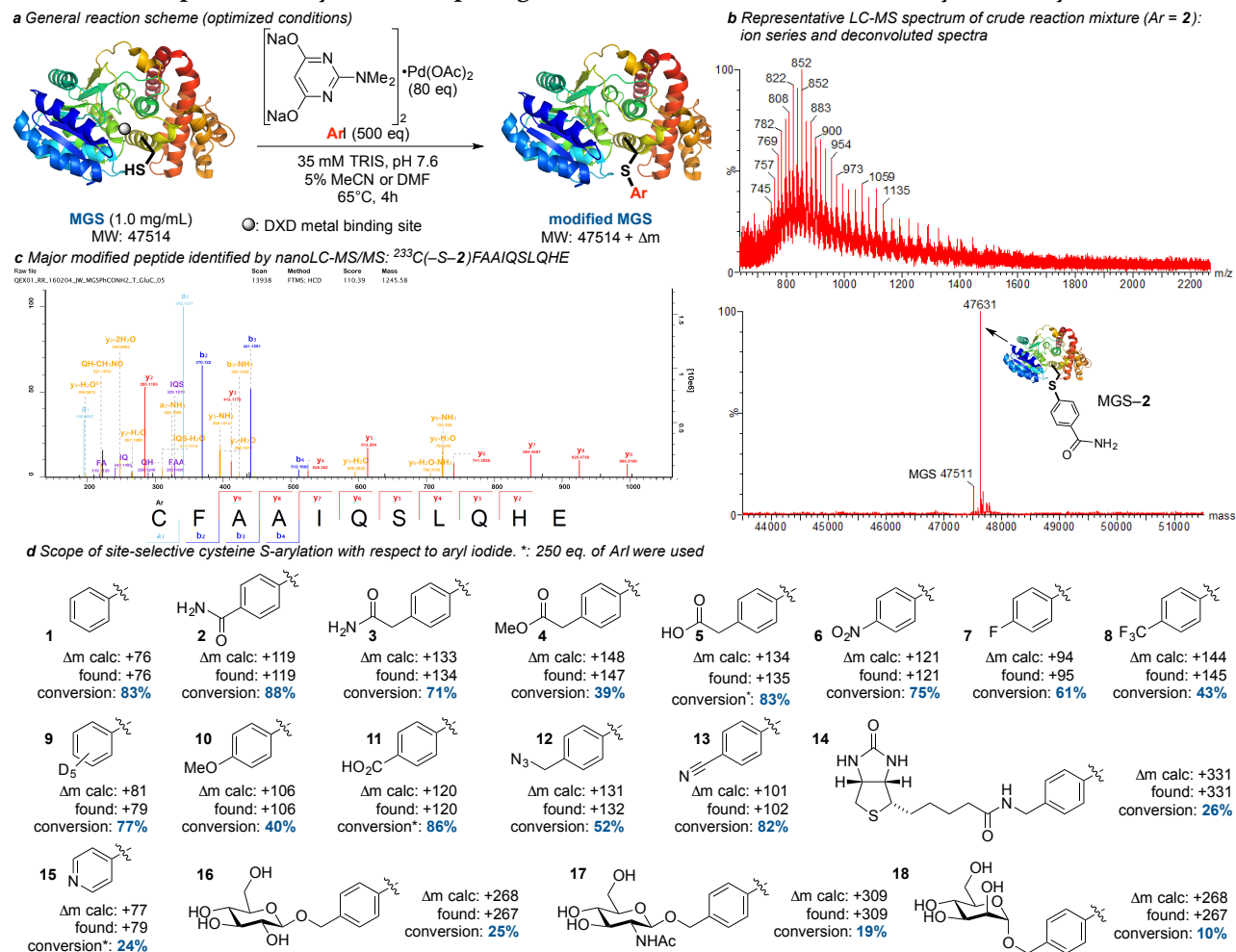
modifications have typically¹⁵ exploited unnatural metal-binding peptide sequences at the termini of proteins, such as oligo-aspartate,^{16,17} -histidine¹⁸⁻²⁰ or -cysteine²¹ motifs. However, covalent protein modifications based on naturally occurring motifs to achieve site-selectivity are rare.²² The need for programmed genetic incorporation of unnatural sequences at restricted locations prohibits their use as a tool for *a priori* identification of metal-binding proteins. We therefore set out to develop a conceptually different approach for site-selective protein modification by relying on endogenous metal-binding motifs (Scheme 1c). If successful, this would not only enable labeling of metal-binding proteins *in vitro* but also identification in cell lysates (e.g. through direct attachment of 'pull-down handles').

Scheme 1. Established strategies compared to metal-site guided protein modification.



One form of common, naturally-occurring metal-binding motif is that found in enzyme active sites that enable metal-dependent catalysis. We reasoned that these could direct a reactive metal complex into a given protein, thereby guiding selective reaction. Our proposed protein modification thus relied on two characteristic structural patterns: a) a metal-binding site that would steer the reactive complex; and b) a proximal, reactive amino-acid residue to react with the metal complex resulting in covalent modification. Class A²³ glycosyltransferases (GTs) are archetypal metal-dependent enzymes and we chose mannosylglycerate synthase (MGS), originally isolated from the thermophilic bacteria *Rhodothermus marinus*,²⁴ as a model system. MGS was the first mannosyltransferase to be fully structurally characterized²⁵ and displays a well-examined²⁶ metal binding profile. GT-A folds contain a common DxD sequence as a metal-binding motif²⁷; in MGS it is D100A101D102 (PDB 2BO6).²⁶ We speculated that this could steer an M(II)-Aryl species generated *in situ* into the active site, allowing it to react with

Scheme 2. Development, analysis and scope of guided site-selective, Pd-mediated, cysteine-arylation.



a range of potential residues, depending on reactivity & mechanism, including proximal Cys233 thiolate (Sch. 1c).²⁸

We first investigated several transition metal pre-catalysts^{13,29,30} in combination with iodobenzene for modifying MGS-His₆. Major emphasis was placed on palladium³¹ as a suitable metal mediator, as previous studies have shown its versatility in aqueous Suzuki-^{13,32} and Sonogashira-coupling reactions.³³ After a screen of a variety of systems, Pd(OAc)₂ ligated with the disodium salt of *N,N*-dimethyl-2-amino-4,6-dihydroxypyrimidine (DM-ADHP) proved most reactive at 65°C and was used as a precatalyst (Scheme 2a). Other pre-catalysts such as Pd(OC(O)CF₃)₂, Pd(OAc)₂, or other water-soluble palladium pre-catalysts were unreactive and using these only unmodified MGS was obtained (see SI). At ambient temperature or 37°C, only unmodified protein was detected by LC-MS analysis (see SI). This unique performance of DM-ADHP as ligand suggested that at least one molecule of ligand was still attached to the Pd(II)-Aryl species during the reaction or was necessary to deliver the Pd-complex to the guiding site. Independently, Buchwald *et al.* have also elegantly shown that Pd-mediated cysteine S-arylation can be rendered general by tuning the surrounding ligand.³⁴ Under our optimized conditions, only a single product was observed in >85% conversion after 4 h (Scheme 2b). Consistent with our selectivity hypothesis, potential di- or oligo-modified species were not detected by LC-MS despite other potential modifiable sites (e.g. Cys34, Cys209, Cys305); notably C305 is the most solvent exposed but does not arylate,

further suggesting regioselectivity directed by protein structure and aryl-Pd species. Longer reaction times led only to loss of protein through degradation / precipitation pathways.

The site of the modification was examined by LC-MS/MS analysis after in-gel digest with trypsin/endoproteinase GluC. This confirmed C233 as the major reaction site (Scheme 2c). Modified vs unmodified Cys sites were also established semi-quantitatively through a carbamidomethylation strategy (see SI); this not only confirmed C233 as the primary reaction site but also highlighted the retained high reactivity of the Cys residues (which were unmodified by our metal-guided process) to non-directed Cys alkylation chemistry. To further test these sites of reactivity, we prepared mutants in which potentially reactive Cys were exchanged for Ala (C34A, C209A, C233A). The reaction outcome remained unchanged with the C34A or C209A (Table 1, entry 2-3), whereas no product (<5%) was detected for the C233A variant, confirming the high site-selectivity of the S-arylation (entry 4).

Mutagenesis was also used to probe the dependence of this site-selective cysteine arylation on the potentially important sites within the protein. Ala-scanning mutational analysis revealed significant effects (SI Figure S9) only from combined mutagenesis of D100, D102 that comprise the DxD and T139 (found in its accessory ion-engagement site²⁵, entries 10,11,13). Although other backbone interactions that cannot be 'scanned' in this way likely also contribute, the partial role of this combined site was confirmed by triple mutation (entries 14,15) that reduced modification to <~30%; other puta-

tive binding/basic residues, e.g. H217, R131, had essentially no effect (Table 1). It should be noted that detailed analysis (see SI) of the structure MGS shows that D100, D102, T139 contribute to *part* of the metal/ion-binding site; remaining backbone interactions and protein shell may explain retained residual directing activity of even triple DDT→AAA mutants.

Table 1. Effect of host mutations on S-arylation

entry	protein	% S-arylation ^a		
		Ar = 1	Ar = 7	Ar = 3
1	His ₆ MGS-WT	75	58	49
2	C34A	71	49	53
3	C209A	75	55	46
4	C233A	<5	<5	<5
5	D100A	72	57	51
6	D102A	66	49	45
7	H217A	75	60	51
8	D100A H217A	58	35	58
9	D102A H217A	59	43	51
10	D100A D102A	42	29	36
11	D100A D102A H217A	48	33	39
12	R131A	66	43	55
13	T139A	48	44	31
14	D100A D102A T139A	30	34	27
15	D100A D102A T139A H217A	35	40	42
16	“His ₆ ” MGS-WT	65	38	52

^aLC-MS conversion after 3h under conditions of Scheme 2.

As poly-His sites have been used to guide other protein chemistries,¹⁸⁻²⁰ we tested endogenous vs such unnatural sites. The out-competing, directing effect of the identified metal/ion-binding D100-D102-T139 motif over artificial (i.e. His₆) motif was suggested by the essentially negligible influence of His₆ in MGS (entry 16). The role of the DDT motif was further supported by lower arylation rates for the AAA mutant and by the attempted modification of several different proteins that bear accessible Cys but no nearby metal-binding sites; from these only unmodified proteins were recovered (see SI). Together, these data suggested an important role for the endogenous motif (Scheme 1c).

Next, we tested the scope of the reaction beyond Ar = Ph towards functionalized aryls (**1** – **18**, Scheme 2d), including those containing useful labels and handles. The reaction is tolerant of a variety of different aryl iodides bearing reactive as well as biochemically-useful functional groups (Scheme 2d). Accordingly, proteins can be labelled with fluoro- (**7**, **8**) or deuterio-labelled (**9**) groups and sugar (e.g. Glc- **16**, GlcNAc- **17**, Man-**18**), biotin- (**14**) or azide- (**12**) bearing moieties enabling potential subsequent detection, further modification or purification by affinity chromatography. Observed conversions varied in a manner that may reflect solubility in buffered aqueous media & potentially differing abilities to enter the regions proximal to the metal-binding motif (although no clear trends supporting e.g. hydrophobic interaction were observed).

We also examined the capacity of this method to modulate protein (e.g. enzymatic) activity. The use of directed, covalent enzyme inhibition strategies has seen a resurgence in recent years³⁶ and several enzyme target families³⁷ contain conserved reactive residues, such as the Cys that we find here. Our identified site-selective S-arylation modified residue Cys233 in close proximity to the active site of MGS. Since metal binding sites also often coincide or abut active sites, we considered that this could enable active-site-directed,

metal-mediated modulation/inhibition. Strikingly, evaluation of the transferase activity^{25,35} of representative C233-S-arylated MGS variants revealed modification-dependent inhibition of MGS's endogenous enzymatic metabolic role (generation of stress protectant mannosylglycerate, Table 2). This vitally confirmed two key aspects: i) the arylation conditions are apparently benign with respect to global protein structure (here giving inhibited but measurable activity consistent with folded protein); and ii) proof-of-principle, structure-varied, covalent inhibition of a model GT. Indeed, variation of both size, functional group and the inclusion of bio-mimetic motifs (e.g. glycans that might additionally engage the active site) in the aryl moieties allowed the development of quite potent (~70-80% at ~2mM, 200 μM substrate) inhibition. A good correlation between level of inhibition and modification was observed in most cases, consistent with effective blocking of the active site (Table 2). Interestingly, Glc-bearing **16** showed relative levels of inhibition higher than others; Glc is a recognized sugar substrate for MGS²⁵ and we tentatively assign this to potential site-to-site engagement of the Glc moiety by MGS.

Table 2. Covalent inhibition of MGS^a

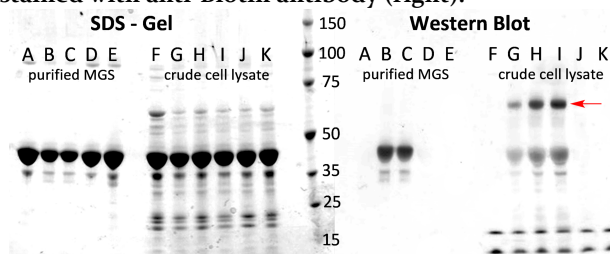
entry	R	% inhibition ^b	% modification ^c
1	-H (wild-type) ^d	-	-
2	-Ph (1)	77.8 ± 0.6	82
3	-C ₆ H ₄ -CONH ₂ (2)	62.3 ± 0.1	77
4	-C ₆ H ₄ -CH ₂ CO ₂ H (5)	58.4 ± 1.4	75
5	-C ₇ H ₆ -OGlc (16)	72.1 ± 0.4	21
6	-C ₇ H ₆ -NH-biotin (14)	15.6 ± 0.5	24

^aaverage of two following arylation with ~2mM reagent. ^bfrom GDP release³⁵ verified by MS²⁵ (see SI). ^cestimated by LC-MS. ^dunmodified MGS subjected to identical conditions without ArI.

Finally, we were further able to show that this method is suitable for the *a priori* identification of other metal-dependent proteins, even in complex mixtures (Figure 1). In this context, purified MGS as well as the heat-shocked crude cell lysates of MGS over-expression were treated with different Pd(L₂)₂(OAc)₂ and biotin-aryl (**14**) iodide concentrations (along with associated controls) and visualized using anti-biotin antibody. These enabled the detection of a metal-binding protein with a molecular weight of ~60 kDa (Figure 1, lanes G-I). In-gel tryptic digest allowed this band to be identified as the 60 kDa chaperonin GroEL,³⁸ a thermophilic, ATP-dependent protein that assists folding during the last steps of protein biosynthesis. In line with our designed method, GroEL binds Mg²⁺ through an aspartate-based motif.^{39,40} Mapping of modified GroEL by MS/MS (see SI) revealed Cys138 as the primary site for arylation; notably, direct interaction of Cys138 with the Mg•ATP binding site has been previously demonstrated.⁴¹ The method was further verified under differing conditions, such as no overexpression or overexpression of a different protein, and gave essentially similar results (see SI). As well as being able to visualize identified proteins in this way, we were also able to demon-

strate biotin-mediated affinity extraction using the attached biotin moiety found in **14** (see SI).

Figure 1. Identification of metal-binding proteins in crude cell lysate. SDS-PAGE (left) and Western blot stained with anti-Biotin antibody (right).



Reacn. 35 mM TRIS pH7.6, 67°C, 3h. Lanes: **A** MGS; **B** 2 mM [Pd], 10 mM **14-I**; **C** 2 mM [Pd], 5 mM **14-I**; **D** no [Pd], 5mM **14-I**; **E** 2 mM [Pd], 10 mM PhI; **F** lysate; **G** 1mM [Pd], 5 mM **14-I**; **H** 2 mM [Pd], 10 mM **14-I**; **I** 4 mM [Pd], 10 mM **14-I**; **J** no [Pd], 10 mM **14-I**; **K** 2 mM Pd, 10 mM PhI. Red arrow: excised bands.

In conclusion, we have developed Pd-mediated site-selective S-arylation that exploits endogenous metal-binding motifs. In contrast to previous Pd-mediated bioconjugation reactions,^{34,42,43} the isolation of a preformed Pd-aryl complex was not necessary and renders this protocol easy to implement. It also enables regioselective differentiation to enhance product homogeneity in proteins bearing multiple Cys. We suggest that this work has established three proofs-of-principle: firstly, not only that endogenous metal binding sites may be used to guide site-selective, metal-mediated modification (Scheme 1c) but secondly, since active sites of several enzymes (e.g. GTs) contain such binding motifs, it provides the opportunity for active-site-directed covalent modulation/inhibition of activity. Here, we show examples of covalent inhibition of GTs,⁴⁴ which inspires future investigations. Thirdly, this selectivity was demonstrated to be sufficient to allow protein-selective modification based on the presence of such an endogenous metal-binding motif (a selectivity that we tested here within complex cellular milieu). Thus, the method can also be used for *a priori* labeling, detection and identification of metal-dependent proteins by employing e.g. biotin-labeling in cell lysates. We have shown reaction here of quite different directing motifs in proteins suggesting generality based on metal or metal/ion binding; we cannot, of course, exclude other directing factors. We should note too that the conditions (e.g. temperature, reagents) may well be incompatible with certain proteins and future studies will focus on exploring breadth and altered reagent conditions to delineate such limitations further. The harsher reaction conditions for S-arylation as compared to e.g. Suzuki reactions¹³ on proteins are likely a consequence of the more demanding nature of the Pd-mediated C-S bond forming process;⁴⁵ higher reaction temperature may also enable critical protein flexibility to allow interaction between targeted Cys residue and the metal binding site. We believe that the concept presented herein might represent a useful additional way to achieve site-selective protein modifications (and associated uses) without the need for incorporating either UAAs or unnatural motifs.

Supporting Information

Additional information on optimization, detailed experimental procedures, LC-MS and MS/MS spectra are included in the SI. The authors declare no competing financial interests.

Corresponding Author ben.davis@chem.ox.ac.uk

ACKNOWLEDGMENT

We thank the German Academic Exchange Service (JW), & the Felix Foundation (RR) for financial support and Drs C. Fehl, A. Fernandez-Tejada, Y. Geng, Mr S. Galan for helpful discussions.

- (1) Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*; Roberts and Company: Englewood, Colorado, 2006.
- (2) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem. Intl Ed.* **2009**, *48*, 6974.
- (3) van Kasteren, S.I.; Kramer, H.; Jensen, H.; Campbell, S.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446*, 1105.
- (4) Boutureira, O.; Bernardes, G. J. L. *Chem. Rev.* **2015**, *115*, 2174.
- (5) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*.
- (6) Chalker, J.; Bernardes, G.; Lin, Y.; Davis, B. *Chem Asian J* **2009**, *4*, 630.
- (7) Young, T. S.; Schultz, P. G. *J. Biol. Chem.* **2010**, *285*, 11039.
- (8) Wals, K.; Ovaa, H. *Front. Chem.* **2014**, *2*, 15.
- (9) Kim, C.; Axup, J.; Schultz, P. G. *Curr Opin Chem Biol* **2013**, *17*, 412.
- (10) Lang, K.; Chin, J. W. *Chem. Rev.* **2014**, *114*, 4764.
- (11) Tornøe, C.; Christensen, C.; Meldal, M. *J Org Chem* **2002**, *67*, 3057.
- (12) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Intl Ed.* **2002**, *41*, 2596.
- (13) Chalker, J.; Wood, C.; Davis, B. G. *J Am Chem Soc* **2009**, *131*, 16346.
- (14) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G. *J. Am. Chem. Soc.* **2008**, *130*, 9642.
- (15) Uchinomiya, S.; Ojida, A.; Hamachi, I. *Inorg. Chem.* **2014**, *53*, 1816.
- (16) Nonaka, H.; Fujishima, S.-H.; Uchinomiya, S.-H.; Ojida, A.; Hamachi, I. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6696.
- (17) Ojida, A.; Honda, K.; Shinmi, D.; Kiyonaka, S.; Mori, Y.; Hamachi, I. *J. Am. Chem. Soc.* **2006**, *128*, 10452.
- (18) Uchinomiya, S.; Nonaka, H.; Fujishima, S.; Tsukiji, S.; Ojida, A.; Hamachi, I. *Chem. Commun.* **2009**, 5880.
- (19) Hauser, C. T.; Tsien, R. Y. *Proc Natl Acad Sci USA* **2007**, *104*, 3693.
- (20) Rosen, C. B.; KodanAnne, L. B.; Nielsen, J. S.; Schaffert, D. H.; Scavenius, C.; Okholm, A. H.; Voigt, N. V.; Enghild, J. J.; Kjems, J.; Tørring, T.; Gothelf, K. V. *Nat Chem* **2014**, *6*, 804.
- (21) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269.
- (22) Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I. *J. Am. Chem. Soc.* **2015**, *137*, 5372.
- (23) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annu Rev Biochem* **2008**, *77*, 521.
- (24) Martins, L. g. O.; Empadinhas, N.; Marugg, J. D.; Miguel, C.; Ferreira, C.; Costa, M. S. d.; Santos, H. *J. Biol. Chem.* **1999**, *274*, 35407.
- (25) Flint, J.; Taylor, E.; Yang, M.; Bolam, D. N.; Tailford, L. E.; Martinez-Fleites, C.; Dodson, E. J.; Davis, B. G.; Gilbert, H. J.; Davies, G. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 608.
- (26) Nielsen, M.; Suits, M. D. L.; Yang, M.; Barry, C. S.; Martinez-Fleites, C.; Tailford, L. E.; Flint, J. E.; Dumon, C.; Davis, B. G.; Gilbert, H. J.; Davies, G. J. *J. Biol. Chem.* **2011**, *286*, 15155.
- (27) Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. *Glycobiology* **2006**, *16*, 29R.
- (28) Lee, C. F.; Liu, Y. C.; Badsara, S. S. *Chem. Asian J.* **2014**, *9*, 706.
- (29) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. *Angew. Chem. Intl Ed.* **2013**, *52*, 3916.
- (30) Gao, Z.; Gouverneur, V.; Davis, B. *J Am Chem Soc* **2013**, *135*, 13612.
- (31) Yang, M.; Li, J.; Chen, P. R. *Chem. Soc. Rev.* **2014**, *43*, 6511.
- (32) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2011**, *47*, 1698.
- (33) Li, N.; Lim, R.; Edwardraja, S.; Lin, Q. *J Am Chem Soc* **2011**, *133*, 15316.
- (34) Vinogradova, E. V.; Zhang, C.; Spokoyniy, A. M.; Pentelute, B. L.; Buchwald, S. L. *Nature* **2015**, *526*, 687.
- (35) Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W. *Glycobiology* **2011**, *21*, 727.
- (36) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. *Nat Rev Drug Discov* **2011**, *10*, 307.
- (37) Zhang, J.; Yang, P. L.; Gray, N. S. *Nat Rev Cancer* **2009**, *9*, 28.
- (38) Braig, K.; Otwinowski, Z.; Hegde, R.; Boisvert, D. C.; Joachimiak, A.; Horwich, A. L.; Sigler, P. B. *Nature* **1994**, *371*, 578.
- (39) Wang, J.; Boisvert, D. C. *J. Mol. Biol.* **2003**, *327*, 843.
- (40) Boisvert, D. C.; Wang, J.; Otwinowski, Z.; Norwich, A. L.; Sigler, P. B. *Nat Struct Mol Biol* **1996**, *3*, 170.
- (41) Bochkareva, E.; Horovitz, A.; Girshovich, A. *J Biol Chem* **1994**, *269*, 44.
- (42) Simmons, R.; Yu, R.; Myers, A. G. *J Am Chem Soc* **2011**, *133*, 15870.
- (43) Cheng, G.; Lim, R.; Li, N.; Lin, Q. *Chem. Commun.* **2013**, *49*, 6809.
- (44) Tedaldi, L.; Wagner, G. K. *MedChemComm* **2014**, *5*, 1106.
- (45) Kondo, T.; Mitsudo, T. *Chem. Rev.* **2000**, *100*, 3205.

