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

Selective Metal-Site-Guided Arylation of Proteins

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

Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K.

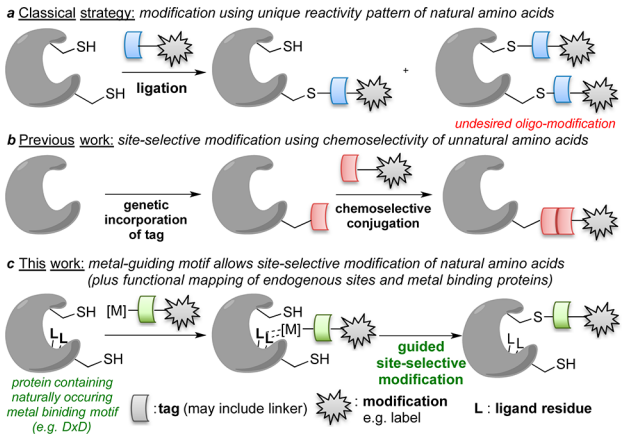
Supporting Information

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 also 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] While 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 directs regioselection and enables selectivity for certain residues, often related to or guided by function. While regioselectivity directed simply by accessibility has proved to be a potentially successful approach in chemical protein modification,[3] the ability to direct a protein-modifying *chemical* catalytic center 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 toward 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 the 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] Elegant prior metal-guided covalent protein modifications have typically[11] exploited unnatural metal-binding peptide sequences at the termini of proteins, such as oligoaspartate,[12,13] -histidine[14−16] or -cysteine[17] motifs. However, covalent protein modifications based on naturally occurring motifs to achieve site selectivity are rare.[18] 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 enable not only labeling of metal-binding proteins in vitro but also identification in cell lysates (e.g., via "pull-down handles").

Scheme 1. Established Strategies Compared with Metal-Site-Guided Protein Modification

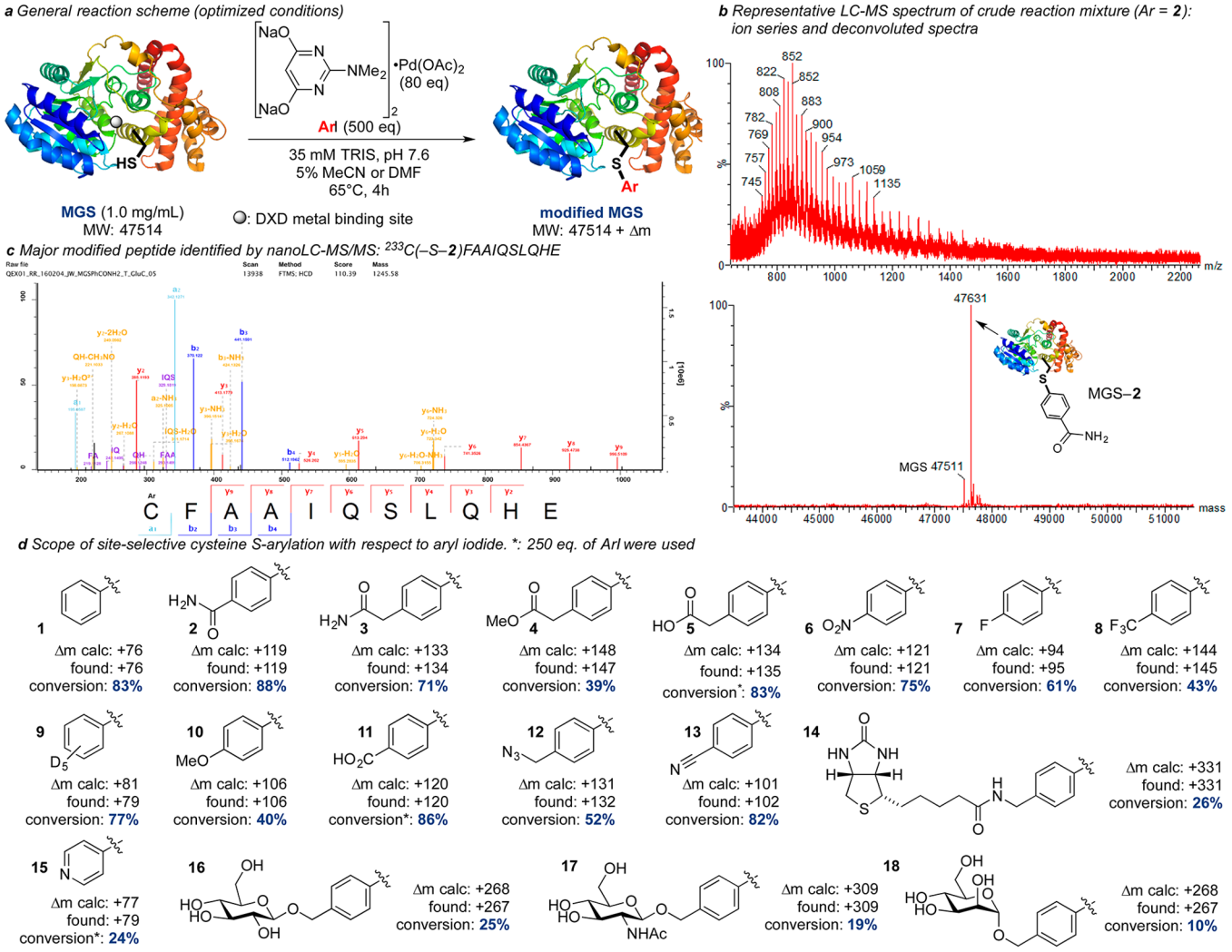


One form of a common, naturally occurring metal-binding motif is that found in enzyme active sites that enable metal-dependent catalysis. We reasoned that these could direct a reactive metal complex into a given protein, thereby guiding selective reaction. Our proposed protein modification thus relied 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 A[19] glycosyltransferases (GTs) are archetypal metal-dependent enzymes, and we chose mannosyl-glycerate synthase (MGS), originally isolated from the thermophilic bacteria *Rhodothermus marinus*,[20] as a model system. MGS was the first mannosyltransferase to be fully structurally characterized[21] and displays a well-examined[22] metal binding profile. GT-A folds contain a common DxD sequence as

Received: April 20, 2016
Published: June 23, 2016

DOI: 10.1021/jacs.6b04043
J. Am. Chem. Soc. 2016, 138, 8678−8681

Scheme 2. Development, Analysis, and Scope of Guided Site-Selective, Pd-Mediated Cysteine Arylation



a metal-binding motif;[23] in MGS it is D100A101D102 (PDB entry 2BO6).[22] We speculated that this could steer an M(II)–aryl species generated in situ into the active site, allowing it to react with a range of potential residues, depending on reactivity and mechanism, including proximal Cys233 thiolate (Scheme 1c).[24]

We first investigated several transition metal precatalysts[25,26,27] in combination with iodobenzene for modifying MGS-His_6. Major emphasis was placed on palladium[28] as a suitable metal mediator, as previous studies have shown its versatility in aqueous Suzuki[25,29] and Sonogashira coupling reactions.[30] 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 to be the most reactive at 65 °C and was used as a precatalyst (Scheme 2a). The unique performance (see SI) of DM-ADHP as a 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 and co-workers have also elegantly shown that Pd-mediated cysteine S-arylation can be rendered general by tuning of the surrounding ligand.[31] 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 Pd–aryl species.

The site of the modification was examined by LC–MS/MS analysis after in-gel digestion with trypsin/endoproteinase GluC. This confirmed C233 as the major reaction site (Scheme 2c). Modified versus unmodified Cys sites were also established semiquantitatively through a carbamidomethylation strategy (see the 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 nondirected 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 C34A or C209A (Table S.8, entries 2 and 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 (Figure S.9) only from combined mutagenesis of D100 and D102 that constitute the DxD and T139 (found in its accessory ion-engagement site[21]) (Table S.8, entries 10, 11,

DOI: 10.1021/jacs.6b04043
J. Am. Chem. Soc. 2016, 138, 8678–8681

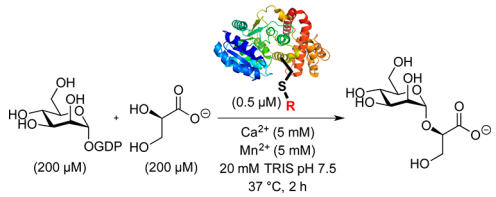
and 13). Although other backbone interactions that cannot be "scanned" in this way likely also contribute, the partial role of this combined site was confirmed by triple mutation (entries 14 and 15), which reduced modification to less than ~30%; other putative binding/basic residues (e.g., H217, R131) had essentially no effect (Table S.8). It should be noted that detailed analysis of the structure of MGS (see the SI) shows that D100, D102, and T139 contribute to *part* of the metal/ion-binding site; remaining backbone interactions and protein shell may explain the retained residual directing activity of even triple DDT → AAA mutants.

As poly-His sites have been used to guide other protein chemistries,[14−16] we tested the endogenous site versus such unnatural sites. The outcompeting directing effect of the identified metal/ion-binding D100-D102-T139 motif over the artificial (i.e., His_6) motif was suggested by the essentially negligible influence of His_6 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 the 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 toward 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 labeled with fluoro- (**7**, **8**) or deutero-labeled (**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. The observed conversions varied in a manner that may reflect solubility in buffered aqueous media and potentially differing abilities to enter the regions proximal to the metal-binding motif (although no clear trends supporting, e.g., hydrophobic interactions, 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 years,[32] and several enzyme target families[33] contain conserved reactive residues, such as the modified Cys233 found here. 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[21,34] of representative C233-S-arylated MGS variants revealed modification-dependent inhibition of MGS's endogenous enzymatic metabolic role (generation of stress protectant mannosylglycerate; Table 1). This vitally confirmed two key aspects: (i) the fact that 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 and functional group and the inclusion of biomimetic motifs (e.g., glycans that might additionally engage the active site) in the aryl moieties allowed the development of quite potent inhibition. A good correlation between the level of inhibition and modification was observed in most cases, consistent with effective blocking of the active site (Table 1). Interestingly, Glc-bearing **16** showed relative levels of inhibition higher than others; Glc is a recognized

Table 1. 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	$C_6H_4CONH_2$ (**2**)	62.3 ± 0.1	77
4	$C_6H_4CH_2CO_2H$ (**5**)	58.4 ± 1.4	75
5	C_7H_6−*O*-Glc (**16**)	72.1 ± 0.4	21
6	C_7H_6−NH−biotin (**14**)	15.6 ± 0.5	24

[a]Average of two following arylation with ~2 mM reagent. [b]From GDP release[34] verified by MS[21] (see the SI). [c]Determined by LC−MS. [d]Unmodified MGS subjected to identical conditions without ArI.

sugar substrate for MGS,[21] and we tentatively assign this to potential site-to-site engagement of the Glc moiety by MGS.

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,

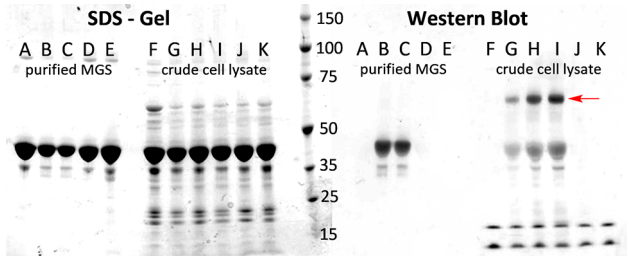


Figure 1. Identification of metal-binding proteins in crude cell lysate: SDS-PAGE (left) and Western blot stained with anti-Biotin antibody (right). Reaction conditions: 35 mM TRIS (pH 7.6), 67 °C, 3 h. Lanes: (A) MGS; (B) 2 mM [Pd], 10 mM **14**-I; (C) 2 mM [Pd], 5 mM **14**-I; (D) no [Pd], 5 mM **14**-I; (E) 2 mM [Pd], 10 mM PhI; (F) lysate; (G) 1 mM [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.

purified MGS as well as the heat-shocked crude cell lysates of MGS overexpression were treated with different $Pd(\mathbf{L2})_2(OAc)_2$ and biotin−aryl (**14**) iodide concentrations (along with associated controls) and visualized using antibiotin antibody. These enabled the detection of an additional metal-binding protein with a molecular weight of ~60 kDa (Figure 1, lanes G−I). In-gel tryptic digestion allowed this band to be identified as the 60 kDa chaperonin GroEL,[35] a thermophilic, ATP-dependent protein that assists folding during the last steps of protein biosynthesis. In line with our designed method, GroEL binds Mg^{2+} through an aspartate-based motif.[36,37] Mapping of modified GroEL by MS/MS (see the SI) revealed Cys138 as the primary site for arylation; notably, direct interaction of Cys138 with the Mg·ATP binding site has been previously demonstrated.[38] The method was further verified via controls and also allowed biotin-mediated affinity extraction (see the SI).

In conclusion, we have developed a Pd-mediated site-selective S-arylation that exploits endogenous metal-binding motifs. In contrast to previous Pd-mediated bioconjugation reac-

DOI: 10.1021/jacs.6b04043
J. Am. Chem. Soc. 2016, 138, 8678−8681

tions,[31,39,40] the isolation of a preformed Pd−aryl complex was not necessary, rendering 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: First, not only can endogenous metal binding sites be used to guide site-selective, metal-mediated modification (Scheme 1c), but second, since the 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 have shown examples of covalent inhibition of GTs,[41] which inspire future investigations. Third, 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 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 here reactions of quite different directing motifs in proteins, suggesting generality based on metal or metal/ion binding; of course, we cannot 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 compared with, e.g., Suzuki reactions[25] on proteins are likely a consequence of the more demanding nature of the Pd-mediated C−S bond-forming process;[42] higher reaction temperature may also enable critical protein flexibility to allow interaction between the 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 to incorporate either UAAs or unnatural motifs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b04043.

Procedures and additional data (PDF)
Supplementary Proteomic Data Tables 1 and 2 (XLSX)
Assay data (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*ben.davis@chem.ox.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

■ REFERENCES

(1) Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*; Roberts and Company: Englewood, CO, 2006.
(2) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem., Int. 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*, 4740.
(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) Uchinomiya, S.; Ojida, A.; Hamachi, I. *Inorg. Chem.* **2014**, *53*, 1816.
(12) Nonaka, H.; Fujishima, S.-H.; Uchinomiya, S.-H.; Ojida, A.; Hamachi, I. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6696.
(13) Ojida, A.; Honda, K.; Shinmi, D.; Kiyonaka, S.; Mori, Y.; Hamachi, I. *J. Am. Chem. Soc.* **2006**, *128*, 10452.
(14) Uchinomiya, S.; Nonaka, H.; Fujishima, S.; Tsukiji, S.; Ojida, A.; Hamachi, I. *Chem. Commun.* **2009**, 5880.
(15) Hauser, C. T.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 3693.
(16) Rosen, C. B.; Kodal, A. 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.
(17) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269.
(18) Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I. *J. Am. Chem. Soc.* **2015**, *137*, 5372.
(19) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annu. Rev. Biochem.* **2008**, *77*, 521.
(20) Martins, L. g. O.; Empadinhas, N.; Marugg, J. D.; Miguel, C.; Ferreira, C.; da Costa, M. S.; Santos, H. *J. Biol. Chem.* **1999**, *274*, 35407.
(21) 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.
(22) 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.
(23) Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. *Glycobiology* **2006**, *16*, 29R.
(24) Lee, C. F.; Liu, Y. C.; Badsara, S. S. *Chem. - Asian J.* **2014**, *9*, 706.
(25) Chalker, J.; Wood, C.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131*, 16346.
(26) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. *Angew. Chem., Int. Ed.* **2013**, *52*, 3916.
(27) Gao, Z.; Gouverneur, V.; Davis, B. *J. Am. Chem. Soc.* **2013**, *135*, 13612.
(28) Yang, M.; Li, J.; Chen, P. R. *Chem. Soc. Rev.* **2014**, *43*, 6511.
(29) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2011**, *47*, 1698.
(30) Li, N.; Lim, R.; Edwardraja, S.; Lin, Q. *J. Am. Chem. Soc.* **2011**, *133*, 15316.
(31) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. *Nature* **2015**, *526*, 687.
(32) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. *Nat. Rev. Drug Discovery* **2011**, *10*, 307.
(33) Zhang, J.; Yang, P. L.; Gray, N. S. *Nat. Rev. Cancer* **2009**, *9*, 28.
(34) Wu, Z. L.; Ethen, C. M.; Prather, B.; Machacek, M.; Jiang, W. *Glycobiology* **2011**, *21*, 727.
(35) Braig, K.; Otwinowski, Z.; Hegde, R.; Boisvert, D. C.; Joachimiak, A.; Horwich, A. L.; Sigler, P. B. *Nature* **1994**, *371*, 578.
(36) Wang, J.; Boisvert, D. C. *J. Mol. Biol.* **2003**, *327*, 843.
(37) Boisvert, D. C.; Wang, J.; Otwinowski, Z.; Norwich, A. L.; Sigler, P. B. *Nat. Struct. Biol.* **1996**, *3*, 170.
(38) Bochkareva, E.; Horovitz, A.; Girshovich, A. *J. Biol. Chem.* **1994**, *269*, 44.
(39) Simmons, R.; Yu, R.; Myers, A. G. *J. Am. Chem. Soc.* **2011**, *133*, 15870.
(40) Cheng, G.; Lim, R.; Li, N.; Lin, Q. *Chem. Commun.* **2013**, *49*, 6809.
(41) Tedaldi, L.; Wagner, G. K. *MedChemComm* **2014**, *5*, 1106.
(42) Kondo, T.; Mitsudo, T. *Chem. Rev.* **2000**, *100*, 3205.

DOI: 10.1021/jacs.6b04043
J. Am. Chem. Soc. 2016, 138, 8678−8681