Communication

Israel Journal of Chemistry

DOI: 10.1002/ijch.201400182

A Triply Divergent Reagent for Glycoprotein Synthesis

Sonia De Munari,[a] Torben Schiffner,[b] and Benjamin G. Davis*[a]

Abstract: Chemical synthesis of glycoproteins through three different strategies, from the same reagent, is herein described. The flexibility of this system, which allows the comparison of different linkage motifs between the same glycan and protein, is shown by one-step diversification of a GlcNAc-ylating reagent and its application in both site-specific and non-site-specific glycoprotein synthesis to create different conjugates from common representative protein scaffolds.

Keywords: glycoconjugates · glycoproteins · glycosylation · protein modifications · tag-and-modify

Synthesis of glycoproteins through chemical modification is a powerful tool for the preparation of new glycoconjugates.[1] Here we report that this can be achieved from one single versatile chemical entity, through a divergent derivatization approach.

Protein modification plays a key role at the frontier between chemistry, chemical biology and biochemistry. New strategies have been developed to increase control over site-selective modifications and to enable the multifunctionalization of natural and artificially engineered proteins.[2] Expanding the post-translational modification toolbox available to scientists has become of primary importance in many fields, from basic research to therapeutic applications. The ability to efficiently and specifically modify complex systems such as proteins, under mild aqueous conditions, is a necessary requirement to explore the potential of such entities. Among all the post-translational modifications, glycosylation is the most diverse[3] and hence potentially intriguing. The associated kaleidoscopic variety of modifications that it creates, optimized by nature through evolution, imparts to similar basic biomolecular scaffolds a huge number of different and potentially highly specific functions. Synthetic glycoproteins can recapitulate and modulate such functions and their construction can be achieved both enzymatically and chemically, by reacting selected protein functionalities with appropriately pre-activated carbohydrates.[1]

While non-specific protein modification generates a heterogeneous mixture of conjugates, site-specific modification of proteins enables access to precisely defined structures. Both methods have their differing utilities in terms of practicality and outcome and ready flexible comparison may be of increasing use as e.g. synthetic biologics[4] find increasing therapeutic application. One flexible, convergent protein synthesis methodology (that has the potential for both non- and site-specific application) is the 'tag-and-modify' approach.[5] In this strategy, a chemical functionality (the 'tag') is positioned within a protein, often within the side chain of an amino acid residue. This moiety can then be selectively chemically targeted ('modified') to create a desired alteration in the protein. Using this method, a number of chemical modifications of proteins have been developed by our group, including several first applications in synthetic glycoprotein construction: from triazole formation,[6] thiol-ene,[7] and Suzuki–Miyaura couplings,[8] to traceless Staudinger ligation[9] and cross metathesis,[10] allowing incorporation of both natural and usefully unnatural sugars (e.g. [^{18}F]-fluorosugars)[11] and even the creation of synthetic glycoproteins in living cells.[12]

Such chemistry often requires strategic flexibility. The consequent linkage between protein and sugar can often be determined by the tag and reagent identities at an early stage when the appropriate tag is installed in the protein and/or when the corresponding sugar reagent has been assembled. This can preclude diversification at a later stage to, for example, explore the effect of linkage type, site or copy number. A flexible method that would allow more ready later-stage diversification would be useful. Here we describe such a strategy, which uses both

[a] S. De Munari, B. G. Davis
Department of Chemistry
University of Oxford
Mansfield Road
Oxford, OX1 3TA (UK)
Phone: (+44) 01865 275652
e-mail: ben.davis@chem.ox.ac.uk

[b] T. Schiffner
The Sir William Dunn School of Pathology
University of Oxford
South Parks Road
Oxford, OX1 3RE (UK)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/ijch.201400182.

Israel Journal of Chemistry

a common sugar reagent intermediate and common protein platform intermediates. In this work we thus demonstrate that diverse glycoprotein syntheses can be achieved using a single model sugar reagent intermediate (here containing the natural sugar GlcNAc) that can be readily activated in three different ways and then used in forming conjugates containing three linkage types from single protein partners. Such an approach that uses a 'triply divergent' reagent for glycoprotein synthesis should allow the use of multiple modification strategies to maximize the efficiency and diversity of chemical glycoprotein synthesis and potentially enable the introduction of different sugars on different glycosylation sites at different tags (Figure 1a).

In the design of the triply divergent sugar reagent (Figure 1b), we envisaged that the nitrile group could provide a single precursor functional group that could then be readily elaborated to act potentially as a nucleophile, an electrophile or a 1,3-dipole, possibly even in a one-pot manner. Thus, a single sugar bearing a nitrile attached at the reducing (anomeric) terminus of the glycan reagent would potentially allow creation of three different linkages that would mimic the typical natural presentation of glycans on proteins (where the reducing terminus is attached to amino acid side chains). We envisaged a cascade of reactions (Scheme 1b) from this nitrile with activation at each stage of the cascade to create imidate, isothiocyanate and azide, respectively – each one ready for direct glycoprotein synthesis selectively on natural (Lys) or unnatural (homopropargylglycine, Hpg) amino acids in a single protein scaffold to create amidine,[13] thiourea,[14] and triazole[6] linkages, respectively. In this method, we could envisage the modification of proteins either non-specifically or site-specifically depending upon chosen conditions, tag density and accessibility and desired copy number or linkage distribution.

To assess the readiness of this cascade of interconversions to create different reactive coupling groups, the peracetylated GlcNAc derivative **1** was used as the common triply divergent precursor and was readily prepared from the parent sugar (Scheme 1a see the Supporting Information for full details). In the first step of the cascade (Scheme 1b), the cyano group of **1** was reduced to the primary amine **2** by hydrogenation[15] over PtO_2; this hetero-

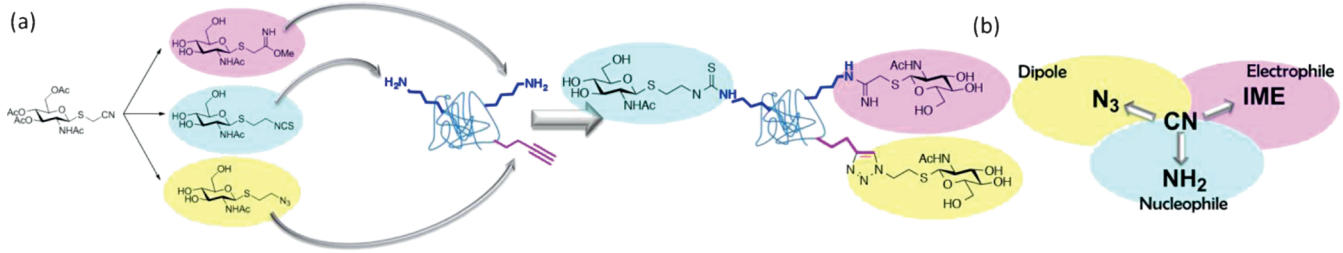
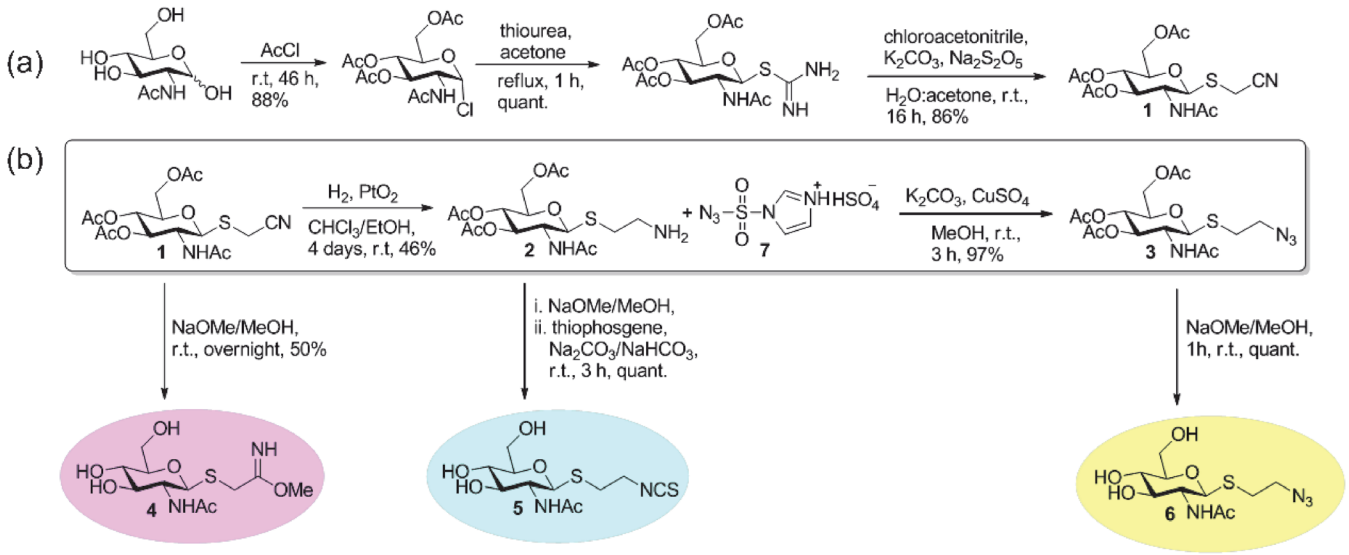


Figure 1. Triply divergent glycoprotein synthesis. From the triple diversification of the same reagent, three different types of glycoproteins can be constructed on the same scaffold, either on natural or unnatural amino acids.



Scheme 1. A cascade of reactions from a common sugar intermediate (here **1**) allows the synthesis of activated imidate, isothiocyanate and azido sugars, ready to create glycoproteins in a triply diverse manner. See the Supporting Information for full details.

geneous system has been shown to be less sensitive to poisoning by the thioether moiety in the sugar[16] than palladium or other catalysts and proved optimal here. A mixed chloroform–ethanol solvent system avoided the use of the relatively harsh acidic conditions frequently employed in such reductions and we anticipate that these mild conditions would also be appropriate for more complex and acid-sensitive sugars. Amine **2** was then transformed into the corresponding azido sugar **3** by the diazo transfer reagent **7**.[17]

Before modification of the proteins, each sugar was deprotected and activated as appropriate (Scheme 1b). GlcNAc-CN **1** was simultaneously deprotected and converted into the 2-imido-2-methoxyethyl (IME)[18] derivative **4** by stirring overnight with sodium methoxide in methanol. Whilst complete deprotection was rapid (5 min), the activation reaction is an equilibrium process and required more time to reach the maximum conversion; at a concentration of 0.1 M with 0.04 M NaOMe in MeOH this typically reaches 50 % equilibrium position. Following glycoprotein synthesis, the unreacted nitrile can be recovered, reactivated and reused. Similarly, GlcNAc-NH_2 **2** was also deprotected under Zemplén deacetylation conditions[19] and then activated with thiophosgene to form the isothiocyanate **5**.[14] The reaction was performed in a mixture of $NaHCO_3$ and Na_2CO_3 (0.3 M), and the product could be used crude or easily purified by size exclusion chromatography (Biogel P2). Lastly, peracetylated GlcNAc-N_3 **3** was deprotected with NaOMe to afford azido reagent **6**.

Two representative protein scaffolds were selected for modification: Np276 is a right-handed quadrilateral β-helix protein scaffold derived from the genome of *Nostoc punctiforme*[20] and Qβ is a coat protein scaffold from the genome of *Allolevivirus Q-beta*[21] that readily forms a 180-copy homomultimer that is an icosahedral protein particle. Reassignment of the Met sense codon to control the position of unnatural amino acid Hpg as an alkyne tag was achieved through heterologous expression in *E. coli* B834(DE3) auxotroph and allowed the creation of variants of the two proteins containing both Lys and Hpg residues: Np276-Hpg61[22] and Qβ-Hpg16[11b] (Figure 2). These protein scaffolds, ready for diverse glycoprotein synthesis with **4**–**6**, contain 3 × Lys + 1 × Hpg and 6 × Lys + 1 × Hpg, respectively, which together with the N-terminal amines were expected to allow varied modifications.

GlcNAc-imidate reagent **4** was reacted with proteins under mildly basic conditions (pH 9). Use of varying amounts of reagent allowed variation of the Lys modification levels; the reaction of Np276-Hpg61 with 1000 equivalents of **4** for 2 h, for example, consistently gave a mixture of di- and tri-GlcNAc-ylated protein (TAPS buffer 0.6 M, see Table 1, entry 1), whereas reaction of the more bulky $(Qβ\text{-}Hpg16)_{180}$ with 1000 equivalents of **4** for 1 h

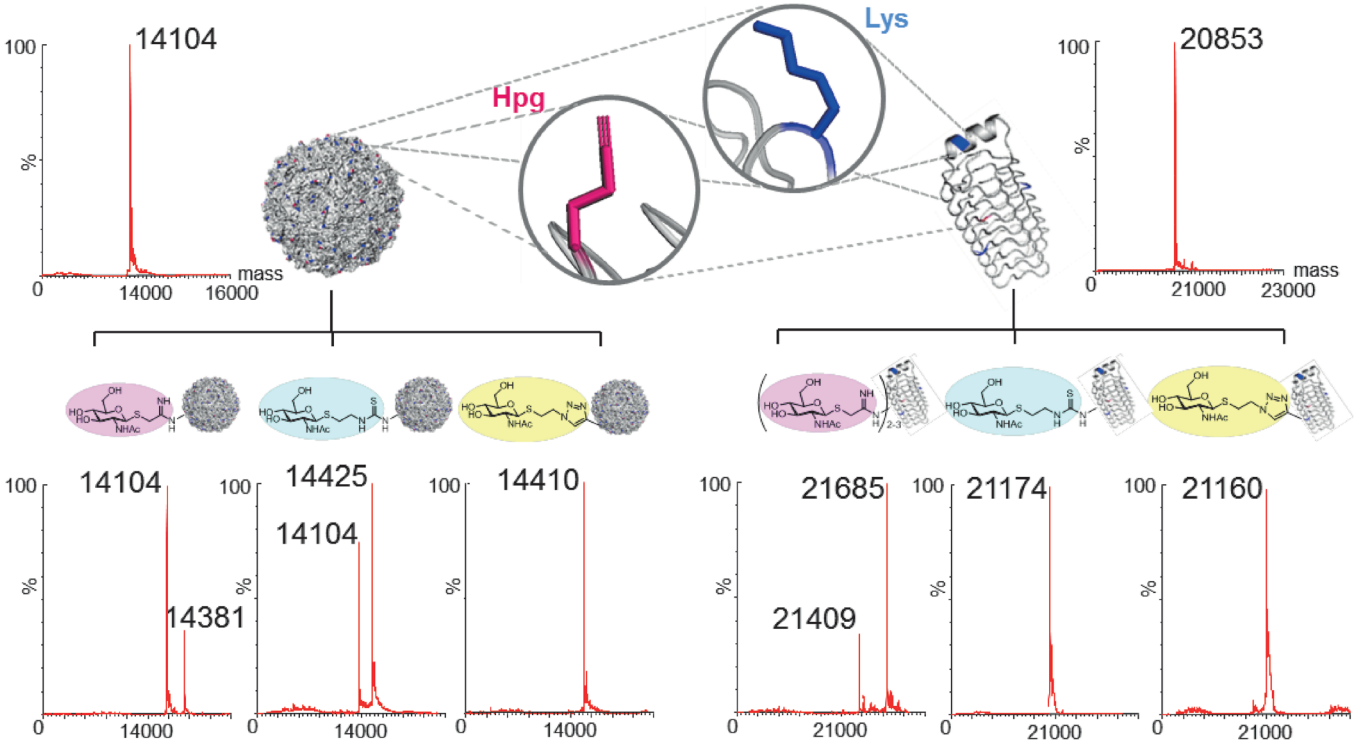
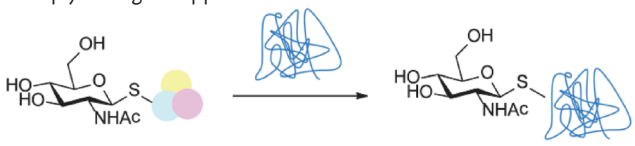


Figure 2. Construction of glycoproteins from Qβ capsid protein (left) and Np276 (right). The unnatural amino acid Hpg is shown in magenta and Lys is shown in blue (top). The reaction with the three different glycan reagents resulted in three different glycoconjugates with different linkages (bottom).

Israel Journal of Chemistry

Table 1. Glycoprotein construction and conversions obtained using the triply divergent approach



Entry	Protein	Sugar	Glycans added[a]	Conversion %[b]
1	Np276-Hpg61	**4**	2 : 3	28 : 72
2	Np276-Hpg61	**5**	1	> 95
3	Np276-Hpg61	**6**	1	> 95
4	Qβ-Hpg16	**4**	1	30
5	Qβ-Hpg16	**5**	1	58
6	Qβ-Hpg16	**6**	1	> 95

[a] Number of sugars added to the protein during each conjugation. [b] Conversion obtained (in %), determined by MS analysis.

gave mono- and GlcNAc-ylated protein along with unreacted protein (phosphate buffer 50 mM, see Table 1, entry 4). GlcNAc-isothiocyanate **5** reacted more slowly with both proteins and, even after 16 h at pH 8 (bicarbonate buffer 0.3 M), gave only mono-GlcNAc-ylated protein from Np276-Hpg61 (see Table 1, entry 2), while $(Qβ\text{-}Hpg16)_{180}$ gave mono-GlcNAc-ylated protein along with unreacted protein (see Table 1, entry 5). GlcNAc-azide **6** was reacted with Np276-Hpg61 for 3 h at pH 8 (phosphate buffer 50 mM) in the presence of Cu(I)·TTMA complex (preformed in acetonitrile from Cu(I)Br)[6a] to give only mono-GlcNAc-ylated product. Under the same conditions, $(Qβ\text{-}Hpg16)_{180}$ gave fully mono-GlcNAc-ylated protein in 16 h. The generally more sluggish reactivity of homomultimer $(Qβ\text{-}Hpg16)_{180}$ compared to monomeric Np276-Hpg16 may be logically attributed to slower reaction of the Qβ 'particle', since non-denaturing conditions were used here. Under these conditions, residues (including some occluded Lys residues) are predicted to be less accessible from molecular mechanics analysis.

In this way, we were able to access derivatized proteins with, in many cases, similar copy numbers (mono-GlcNAc-ylation), but each containing different linkage types (amidine, thiourea and triazole). The nature of the linkage has logically been suggested to sometimes affect presentation and function in such conjugates. Given the differing properties (e.g. length, distance from protein surface, pK_a), diverse functional comparison could be of utility. To compare the functional variance of these synthetic glycoproteins we evaluated their binding to the model carbohydrate-binding[23] lectin wheat germ agglutinin (WGA) by surface plasmon resonance (SPR).

WGA was immobilized on the sensor chip and the synthetic glycoproteins were used as the flow analytes (Figure 3). The level of binding observed for the synthetic Np276-based glycoproteins proved to be consistent with the level of glycosylation, with the more heavily GlcNAc-ylated protein derived from modification of Lys with GlcNAc-imidate **4** (2–3 glycans) binding more strongly than those derived from **5** or **6** and containing a single

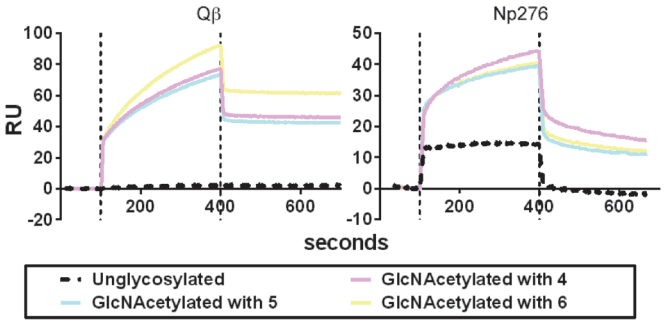


Figure 3. Binding of synthetic glycoproteins to the lectin WGA. Biotinylated WGA was captured onto a streptavidin-coated SPR biosensor chip and the binding of glycoproteins as analytes (20 μg/mL) was followed for 5 min followed by a 5 min dissociation period. Dashed vertical lines indicate start and stop of the sample injection period.

GlcNAc. All bound more strongly than non-glycosylated Np276, which also showed some non-specific interaction with WGA. Interestingly, although the site of glycosylation is different for the latter two, no consequent significant difference was observed in binding. All three Np276 glycoproteins dissociated relatively rapidly.

The three synthetic glycoproteins derived from multimeric Qβ showed strong binding, with no detectable non-specific binding for the non-glycosylated protein (Figure 3). Although the level of glycosylation per monomer was lower for Qβ, the total level of glycan display in the 180-mer homomultimeric particle, which Qβ assembles to form, provides a higher level of valency. The observed binding enhancement is consistent with our previous observations in highly valent protein-derived glycodendrinanoparticles.[24] Again, the level of this interaction was consistent with the overall level of glycan loading, with the binding of the synthetic glycoprotein particle derived from modification of Hpg with **6** giving the highest binding. Interactions that persisted throughout the dissociation period were observed for all.

In conclusion, we have demonstrated that a 'triply divergent' protein-glycosylating reagent can be iteratively generated from a single divergent intermediate by a cascade of reactions, allowing formation of glycoconjugates containing three different linkage types from single protein partners. It has allowed here diversified modulation of both the site of glycosylation and the copy number of glycosylation as well as the protein platform, which in turn generates diverse data to allow evaluation of the associated differences in biological function (here binding to a lectin). This glycoprotein construction approach will allow the use of multiple modification strategies, in order to maximize efficiency and diversity, via the formation of

different linkage types, allowing strategic flexibility. One example of the usefulness of this new methodology could be in the ready creation and then comparison of glycoconjugates as vaccines, where some linkages, but not others, may generate so-called 'anti-linker' responses.[25] The ability to take, for example, the same glycan immunogen from a pathogen and rapidly create similar copy numbers for protein loading but differing linkage types, as we have shown here, may be highly advantageous in the study of synthetic glycoconjugates as vaccines.[26]

Acknowledgements

We would like to thank the EU FP7-ITN Marie-Curie Network programme RADDEL (290023) and the EU FP7-Integrated Infrastructure Initiative–I3 programme ESTEEM2 (312483) for funding. We would also like to thank Dr. T. B. Parsons for useful discussions.

References

[1] a) B. G. Davis, *Chem. Rev.* **2002**, *102*, 579–602; b) D. P. Gamblin, E. M. Scanlan, B. G. Davis, *Chem. Rev.* **2009**, *109*, 131–163; c) L. X. Wang, M. N. Amin, *Chem. Biol.* **2014**, *21*, 51–66.

[2] a) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074; b) A. J. de Graaf, M. Kooijman, W. E. Hennink, E. Mastrobattista, *Bioconjugate Chem.* **2009**, *20*, 1281–1295; c) Y. Takaoka, A. Ojida, I. Hamachi, *Angew. Chem. Int. Ed.* **2013**, *52*, 4088–4106; d) C. D. Spicer, B. G. Davis, *Nat. Commun.* **2014**, *5*, 4740.

[3] A. Varki, *Glycobiology* **1993**, *3*, 97–130.

[4] B. G. Davis, in *Biotherapeutics: Recent Developments using Chemical and Molecular Biology* (Eds.: L. H. Jones, A. J. McKnight), Royal Society of Chemistry, Cambridge, **2013**, pp. 130–144.

[5] J. M. Chalker, G. J. L. Bernardes, B. G. Davis, *Acc. Chem. Res.* **2011**, *44*, 730–741.

[6] a) S. I. van Kasteren, H. B. Kramer, D. P. Gamblin, B. G. Davis, *Nat. Protoc.* **2007**, *2*, 3185–3194; b) S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature* **2007**, *446*, 1105–1109.

[7] a) N. Floyd, B. Vijayakrishnan, J. R. Koeppe, B. G. Davis, *Angew. Chem. Int. Ed.* **2009**, *48*, 7798–7802; b) M. Lo Conte, S. Staderini, A. Marra, M. Sanchez-Navarro, B. G. Davis, A. Dondoni, *Chem. Commun.* **2011**, *47*, 11086–11088.

[8] a) J. M. Chalker, C. S. C. Wood, B. G. Davis, *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347; b) C. D. Spicer, B. G. Davis, *Chem. Commun.* **2011**, *47*, 1698–1700; c) C. D. Spicer, T. Triemer, B. G. Davis, *J. Am. Chem. Soc.* **2012**, *134*, 800–803.

[9] G. J. L. Bernardes, L. Linderoth, K. J. Doores, O. Boutureira, B. G. Davis, *ChemBioChem* **2011**, *12*, 1383–1386.

[10] Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 9642–9643.

[11] a) O. Boutureira, G. J. Bernardes, F. D'Hooge, B. G. Davis, *Chem. Commun.* **2011**, *47*, 10010–10012; b) O. Boutureira, F. D'Hooge, M. Fernandez-Gonzalez, G. J. Bernardes, M. Sanchez-Navarro, J. R. Koeppe, B. G. Davis, *Chem. Commun.* **2010**, *46*, 8142–8144.

[12] a) C. D. Spicer, B. G. Davis, *Chem. Commun.* **2013**, *49*, 2747–2749; b) M. Yang, J. Li, P. R. Chen, *Chem. Soc. Rev.* **2014**, *43*, 6511–6526.

[13] M. A. Robinson, S. T. Charlton, P. Garnier, X. T. Wang, S. S. Davis, A. C. Perkins, M. Frier, R. Duncan, T. J. Savage, D. A. Wyatt, S. A. Watson, B. G. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14527–14532.

[14] M. K. Patel, B. Vijayakrishnan, J. R. Koeppe, J. M. Chalker, K. J. Doores, B. G. Davis, *Chem. Commun.* **2010**, *46*, 9119–9121.

[15] D. Srimani, M. Feller, Y. Ben-David, D. Milstein, *Chem. Commun.* **2012**, *48*, 11853–11855.

[16] M. Breysse, R. Frety, *J. Catal.* **1987**, *154*, 144–154.

[17] N. Fischer, E. D. Goddard-Borger, R. Greiner, T. M. Klapotke, B. W. Skelton, J. Stierstorfer, *J. Org. Chem.* **2012**, *77*, 1760–1764.

[18] Y. C. Lee, C. P. Stowell, M. J. Krantz, *Biochemistry* **1976**, *15*, 3956–3963.

[19] Z. Wang, in *Comprehensive Organic Name Reactions and Reagents*, John Wiley & Sons, Inc., **2010**, pp. 3123–3128.

[20] M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, *Protein Sci.* **2007**, *16*, 755–760.

[21] M. Kajitani, A. Kato, A. Wada, Y. Inokuchi, A. Ishihama, *J. Bacteriol.* **1994**, *176*, 531–534.

[22] M. Fernández-González, O. Boutureira, G. J. L. Bernardes, J. M. Chalker, M. A. Young, J. C. Errey, B. G. Davis, *Chem. Sci.* **2010**, *1*, 709–715.

[23] Y. Nagata, M. M. Burger, *J. Biol. Chem.* **1974**, *249*, 3116–3122.

[24] R. Ribeiro-Viana, M. Sánchez-Navarro, J. Luczkowiak, J. R. Koeppe, R. Delgado, J. Rojo, B. G. Davis, *Nat. Commun.* **2012**, *3*, 1303.

[25] T. Buskas, Y. Li, G. J. Boons, *Chem. Eur. J.* **2004**, *10*, 3517–3524.

[26] Y. L. Huang, J. T. Hung, S. K. Cheung, H. Y. Lee, K. C. Chu, S. T. Li, Y. C. Lin, C. T. Ren, T. J. Cheng, T. L. Hsu, A. L. Yu, C. Y. Wu, C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 2517–2522.

Received: November 15, 2014
Accepted: December 16, 2014
Published online: March 3, 2015