A glycal-based photoaffinity probe that enriches sialic acid binding proteins

Peter S. Thuy-Boun and Dennis W. Wolan*

Departments of Molecular Medicine and Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

*To whom correspondence should be addressed. Tel. 1.858.784.7936, email: wolan@scripps.edu

ABSTRACT

To identify sialic acid binding proteins from complex proteomes, three photocrosslinking affinity-based probes were constructed using Neu5Ac (**5** and **6**) and Neu5Ac2en (**7**) scaffolds. Kinetic inhibition assays and Western blotting revealed the Neu5Ac2en-based **7** to be an effective probe for the labeling of a purified gut microbial sialidase (BDI_2946) and a purified human sialic acid binding protein (hCD33). Additionally, LC-MS/MS affinity-based protein profiling verified the ability of **7** to enrich a low-abundance sialic acid binding protein (complement factor H) from human serum thus validating the utility of this probe in a complex context.

Mucus layer integrity is a key determinant of intestinal homeostasis in the context of host-microbiome interactions. During active ulcerative colitis (UC), microbes can be found in contact with, or translocated through the host epithelium where they can trigger host inflammatory responses. This is in contrast to healthy or remissive states when microbes are held at bay by a thick, impenetrable, and continuously shed mucus layer.[1-9] Though the aftermath of microbial infiltration is often apparent, the initiation and progression of this process is poorly understood. The early stages of infiltration likely involve microbial adherence to, and degradation of, the mucus layer and it is thought that specific microbes and microbial products are culpable.[10-14] The mucus layer is principally composed of Muc2 mucin, a heavily *O*-glycosylated gel-forming protein secreted from intestinal goblet cells.[15,16] The structures of mucin *O*-glycans are diverse but many glycans terminate with sialic acid (*N*-acetyl neuraminic acid in humans) making this residue one of the most exposed host components of the mucus layer, and very likely an early target of microbial adhesion. To understand the microbiome component of UC development, we aim to identify microbes and microbial proteins responsible for mucus layer degradation in UC patients. Our overall approach relies on LC-MS/MS shotgun proteomics to identify proteins of interest; however, the species- and protein-level complexity of the gut microbiome renders it highly impervious to deep proteome profiling efforts without a pre-enrichment step.[17-22] Toward the goal of identifying gut microbes and microbial proteins that degrade the mucus layer in UC patients, we focus our efforts herein on the design and evaluation of affinity-based protein profiling probes for the enrichment of sialic acid binding proteins (sialidases and lectins) from complex mixtures.

Several tools currently enable the identification of proteins that bind carbohydrates. With respect to sialic acid, the Wong group has reported a mechanism-based alkynyl difluoro-sialic acid probe that effectively targets sialidases but not sialic acid binding lectins.[23,24] Tools based on a metabolic labeling strategy are powerful standouts but we sought a culture-independent strategy to preserve the unique *in vivo* phylogenetic and transcriptional configuration of our clinical samples and because many gut microbes are anaerobic and/or "unculturable" outside of their native environments.[25-31] Moreover, the abundance of *O*-glycolytic activity originating from microbiome samples limits our ability to use *O*-linked carbohydrates as affinity groups.[32] We were inspired by early works of the Bednarski group where non-hydrolyzable sialic acid *C*-glycosides were used to visualize whole influenza virus attachment (via hemagglutinin), and sought to adapt this concept for modern mass spectrometry-based proteomics.[33] Thus, we designed a set of probes with three requirements: a non-hydrolyzable affinity tag, an enrichment tag, and a covalent crosslinker.

We first set out to identify affinity scaffolds applicable to the gut microbiome. Using the commercially available turn-on fluorescent substrate 4-methylumbelliferyl-α-*N*-acetyl neuraminic acid (4-MUNANA), we found an abundance of sialidase activity in several human patient microbiome samples and determined that neither commercially available **1**, Zanamivir (**3**),[34] or Oseltamivir (**4**)[35] were able to inhibit this sialidase activity despite being potent inhibitors of influenza sialidases (Fig. 1). We did however identify **2** as a potent inhibitor of microbiome sialidase activity (Table 1, entry 2).[36] Interestingly, **2** and **3** only differ by their C4 substituents, yet **2** has >100-fold more inhibitory aptitude than **3**. This observation informed our decision to refrain from excessively modifying the sialic acid-based affinity group at this time, except at the anomeric carbon for the sake of probe generality. Due to the limited supply of clinical microbiome samples, we also identified BDI_2946, a sialidase from the commensal gut microbe *Parabacteroides distasonis,* as a suitable protein surrogate with a similar inhibition profile for further testing (Fig. 1B, Table 1).
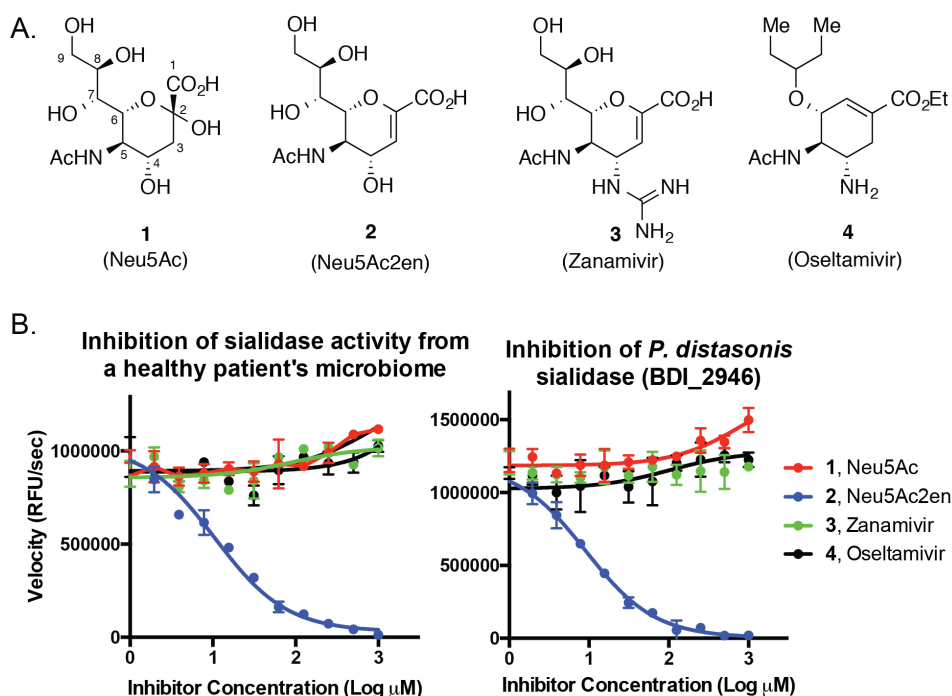

Fig. 1. A. Sialic acid and sialidase inhibitors. B. Inhibition of sialidase activity from a healthy patient's microbiome and from a purified gut microbial sialidase.

We first constructed a modular crosslinking enrichment tag utilizing a UV-active diazirine moiety as a crosslinker and a terminal-alkyne as a proxy for biotin via the copper catalyzed azide alkyne cycloaddition (CuAAC click reaction) (Fig. 2A, "DA").[37-41] We constructed three probes (**5**, **6**, **7**) based on **1** and **2** (Fig. 2A). **5** and **6** were elaborated from *C*-allylated intermediates **9a** and **9b** generated via tin-mediated

allylation of **9** (Fig. 2B, a).[42-45] *C*-allylation yielded a mixture of α/β glycosides which were chromatographically separated prior to further functionalization and conjugation to the crosslinking enrichment tag "DA." Na_2HPO_4 mediated HCl elimination of **8** yielded **10** which was fully deprotected to give **2**.[46,47] **7** was prepared from **2** after reacetylation of free hydroxyl groups followed by linker conjugation to the C1 carboxylic acid (see Supplementary data, Scheme 5).
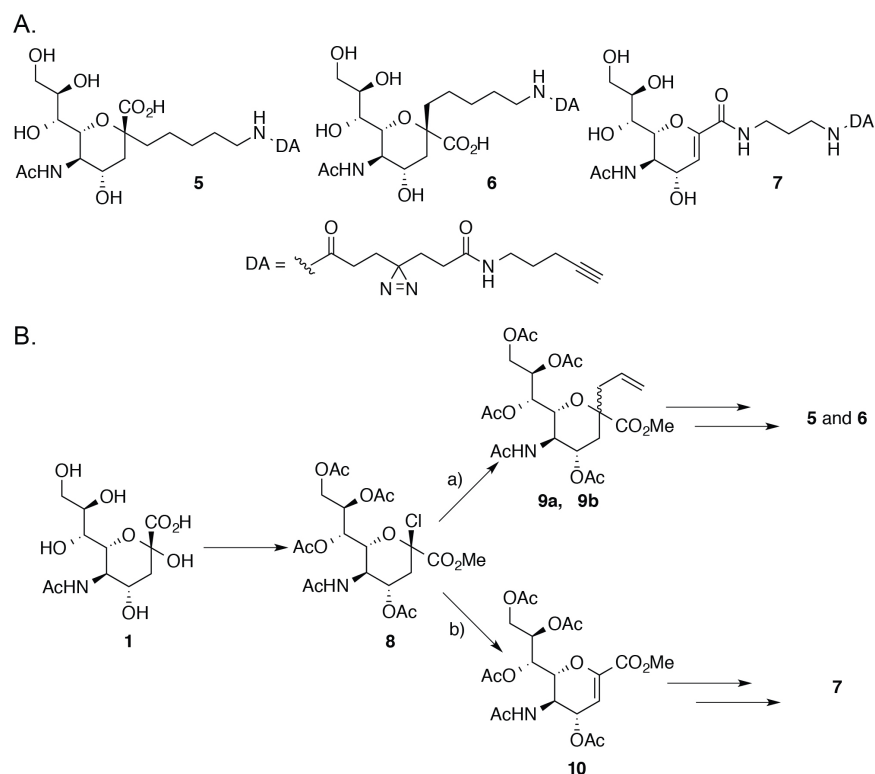

Fig. 2. A. Probes for sialic acid binding proteins examined in this study. B. Synthetic routes to **5**, **6**, and **7**; a) $AllylSnBu_3$, Bu_6Sn_2, THF, hv, r.t., N_2; b) Na_2HPO_4, MeCN, reflux.

Table 1. Inhibition of sialidase activity from (1) a healthy patient microbiome sample or (2) a purified bacterial sialidase by various compounds. N.D.= not determined.

Entry	Compound	IC$_{50}$(microbiome)	IC$_{50}$ (BDI_2946)	% Inhibition (BDI_2946)
1	**1**	> 1 mM	> 1 mM	N/A
2	**2**	11 ± 1 µM	9 ± 1 µM	>99%
3	**3**	> 1 mM	> 1 mM	N/A
4	**4**	> 1 mM	> 1 mM	N/A
5	**5**	N.D.	> 1 mM	N/A
6	**6**	N.D.	> 1 mM	N/A
7	**7**	N.D.	181 ± 2 µM	51%

Compounds **5**-**7** were evaluated as inhibitors of BDI_2946 and **7** was found to be the only candidate to appreciably inhibit this sialidase at a concentration < 1mM (Table 1, entries **5**-**7**). The importance of the C1-carboxylate on **2** for binding affinity is underscored by the 16-fold reduction in inhibitory aptitude observed with **7**.
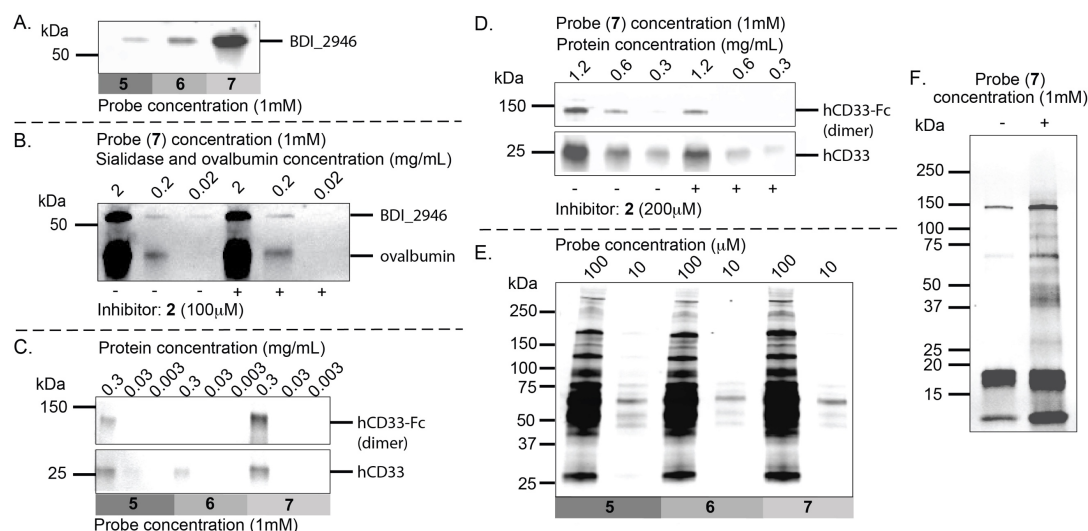

Fig. 3. A. Western blot of purified BDI_2946 sialidase (0.2 mg/mL) labeled by various photocrosslinking probes (1 mM), 1 h UV irradiation; B. Western blot of purified BDI_2946 labeled with **7** (1 mM) and the blocking of **7**-labeling using **2** (100 μM), 30 min UV irradiation. C. Western blot of recombinant human-CD33 labeled with probes **5**-**7** (1 mM), 1 h UV irradiation D. Western blot of CD33 labeled with **7** (1 mM) and the blocking of **7**-labeling using **2** (200 μM). E. Western blot of human serum labeled with probes **5**-**7** (100 μM and 10 μM), 30 min UV irradiation. F. Western blot of a cultured murine distal gut microbiome sample labeled with probe **7** (1 mM), 30 min UV irradiation.

We next evaluated compounds **5**-**7** as photoaffinity probes for BDI_2946. BDI_2946 was treated with 1 mM each of **5**-**7** then irradiated 1h with a 365 nm UV black light bulb. After irradiation, the mixture was treated with Biotin-Peg3-azide under CuAAC conditions (see Supporting Information).[40,41] Streptavidin bead enrichment and Western blotting revealed strong labeling with **7** but not **5** or **6** (Fig. 3A), an observation that is consistent with our kinetic inhibition data. Because **7** produced an intense labeling signal, we evaluated its selectivity further by labeling BDI_2946 in the presence of ovalbumin and **2** (a more potent inhibitor of BDI_2946). Signal for BDI_2946 appears to diminish with the application of **2** suggesting a degree of active site specificity (Fig. 3B**,** 0.02 mg/mL lanes). **7** also appears to non-specifically label ovalbumin alongside BDI_2946 at higher concentrations (Fig. 3B, 2 mg/mL and 0.2 mg/mL lanes), but **7** labels BDI_2946 exclusively at low concentrations (Fig. 3B, 0.02 mg/mL lanes).

We initially tested **5-7** using BDI_2946 because we were able to use sialidase activity as a straightforward readout for target-engagement. We wondered whether **5-7** would be effective probes for the enrichment of lectins which could be challenging targets because their preferred ligands are often multiple monosaccharides in length, and because binding affinities for their preferred ligands can be high (millimolar range) in the absence of multivalency.[48,49] Using relatively high concentrations of human siglec 3 (hCD33), we observed consistently strong labeling by **7** relative to **5** and **6** (Fig. 3C). This is surprising given that the preferred monovalent ligand of hCD33 Neu5Acα(2,3)Galβ(1-4)GlcNacβOR is three residues long while **7** is comprised of only one residue.[50]

Table 2. LC-MS/MS analysis of complement factor H (CFH) and serum albumin (SA) enrichment from a healthy patient's serum treated with probes **5-7** relative to unenriched serum (background). "n.s." = not significant.

Protein	Probe	mean Ln(NSAF)	Δ mean Ln(NSAF)	*p*-value	fold change
CFH (P08603)	5	-9.461	0.095	0.803	1.1
	6	-9.607	-0.050	0.913	0.9
	7	-8.401	1.156	0.026	3.2
SA (P02768)	5	-4.453	-1.970	<0.0001	0.1
	6	-4.754	-2.272	<0.0001	0.1
	7	-3.743	-1.261	<0.0001	0.3

Finally, we wondered whether **7** could enrich for sialic acid binding proteins from a complex mixture. We chose human serum as a test proteome with the intention of targeting complement factor H (CFAH_HUMAN or CFH, P08603), an important attenuator for the alternative pathway of the complement system. CFH is a 139 kDA protein that possesses multiple polyanion binding sites and is known to prefer Neu5Acα(2,3)Gal-containing glycoconjugates.[51-53] We chose serum because it is protein-dense (>50 mg/mL), though only a small fraction of this mixture is comprised of our protein of interest (CFH: 0.1-0.5 mg/mL).[54,55] Approximately half of the serum proteome is serum albumin (SA, P02768) by weight and it is expected that this protein would be a suitable specificity control.[56,57] Because of the complexity of serum, LC-MS/MS was used to both identify and quantify protein enrichment.

We treated undiluted human serum with 100 µM and 10 µM of probes **5-7** preconjugated to biotin to simplify labeling. Labeled proteins were then enriched with streptavidin beads and partitioned. One half of the partitioned sample was used for Western blotting (Fig. 3D) and the other half for LC-MS/MS

analysis. Triplicates of the 100 µM probe-treated samples were analyzed and protein abundances were quantified by normalized spectral abundance factor (NSAF).[58] No significant enrichments of CFH were observed in **5**- and **6**-treated serum compared to unenriched serum control (background). However, treatment with probe **7** was found to significantly enrich CFH 3.2-fold over background with a *p*-value of <0.05. In addition, all probes (**5-7**) were found to negatively enrich for SA 0.1-0.3-fold compared to background with *p*-values <0.0001 (Table 2).

The finding that probe **7** outperformed probes **5** and **6** in the labeling of sialidase BDI_2946 is not surprising given that **7** is based on transition-state like inhibitor **2**. However, it is surprising that probe **7** outperformed probes **5** and **6** for the labeling of purified hCD33, and again for CFH from human serum especially in light of the fact that both of these sialic acid binding proteins have preferred ligands containing α-linked sialic acid >1 monosaccharide in length. The positive results observed with probe **7** are encouraging and provide a lead for further structural optimization. In addition, multivalent display strategies are being pursued to enhance probe affinity.

In summary, we synthesized and evaluated **5-7** as photocrosslinking probes for sialic acid binding proteins. Of compounds **5-7**, Neu5Ac2en-based **7** was identified as a competent probe for both a gut microbial sialidase and a purified sialic acid binding protein. We also demonstrate using LC-MS/MS protein profiling that **7** is capable of enriching for a naturally occurring sialic acid binding protein (CFH) from human serum 3.2-fold over background. We anticipate that **7** will prove to be a useful tool to further our efforts in the area of microbiome proteomics, as the molecule labels distinct bacterial proteins from a cultured murine distal gut microbiome sample that has measurable sialidase activity (Fig. 3F). Protein profiling efforts employing **7** to identify bacterial sialidases and lectins will greatly assist in determining the role these proteins have in health and microbiome-related diseases.

Author Contributions

P.S.T.-B. and D.W.W. conceived of the project. P.S.T.-B. synthesized all probes, purified recombinant proteins, performed enzyme assays, and collected and analyzed all LC-MS/MS data. Both authors contributed to the preparation and editing of the manuscript and have given approval to the final version of the manuscript. The authors declare no competing financial interests.

Supplementary data

Supplementary data associated with this article can be found in the online version at http://xxxxxxx.

References

1. Danese, S.; Fiocchi, C. *N. Engl. J. Med.* **2011**, *365*, 1713–1725.
2. Baumgart, D. C.; Sandborn, W. J. *Lancet* **2007**, *369*, 1641–1657.
3. Johansson, M. E. V.; Sjövall, H.; Hansson, G. C. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10,* 352–361.
4. Johansson, M. E. V. *Inflamm. Bowel Dis.* **2014**, *20*, 2124–2131.
5. Van der Sluis, M.; De Koning, B. A. E.; De Bruijn, A. C. J. M.; Velcich, A.; Meijerink, J. P. P.; Van Goudoever, J. B.; Büller, H. A.; Dekker, J.; Van Seuningen, I.; Renes, I. B.; Einerhand, A. W. C. *Gastroenterol.* **2006**, *131*, 117–129.
6. Johansson, M. E. V.; Phillipson, M.; Petersson, J.; Velcich, A.; Holm, L.; Hansson, G. C. *Proc. Nat. Acad. Sci. USA* **2008**, *105*, 15064–15069.
7. Johansson, M. E.; Gustafsson, J. K.; Sjoberg, K. E.; Petersson, J.; Holm, L.; Sjovall, H.; Hansson, G. C. *PLoS ONE* **2010**, *5*, e12238.
8. Johansson, M. E. V.; Larsson, J. M. H.; Hansson, G. C. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 4659–4665.
9. Johansson, M. E. V.; Gustafsson, J. K.; Holmén-Larsson, J.; Jabbar, K. S.; Xia, L.; Xu, H.; Ghishan, F. K.; Carvalho, F. A.; Gewirtz, A. T.; Sjövall, H.; Hansson, G. C. *Gut* **2014**, *63*, 281–291.
10. Strober, W. *Trends Immunol.* **2013**, *34*, 423–430.
11. Honda, K.; Littman, D. R. *Annu. Rev. Immunol.* **2012**, *30*, 759–795.
12. Dalal, S. R.; Chang, E. B. *J. Clin. Invest.* **2014**, *124*, 4190–4196.
13. Schwechheimer, C.; Kuehn, M. J. *Nat. Rev. Microbiol.* **2015**, *13*, 605–619.
14. Hickey, C. A.; Kuhn, K. A.; Donermeyer, D. L.; Porter, N. T.; Jin, C.; Cameron, E. A.; Jung, H.; Kaiko, G. E.; Wegorzewska, M.; Malvin, N. P.; Glowacki, R. W. P.; Hansson, G. C.; Allen, P. M.; Martens, E. C.; Stappenbeck, T. S. *Cell Host Microbe* **2015**, *17*, 672–680.
15. Bergstrom, K. S. B.; Xia, L. *Glycobiology* **2013**, *23*, 1026–1037.
16. Marcobal, A.; Southwick, A. M.; Earle, K. A.; Sonnenburg, J. L. *Glycobiology* **2013**, *23*, 1038–1046.
17. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. *Chem. Rev.* **2013**, *113*, 2343–2394.
18. Xiong, W.; Abraham, P. E.; Li, Z.; Pan, C.; Hettich, R. L. *Proteomics* **2015**, *15*, 3424–3438.
19. Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. *Annu. Rev. Biochem.* **2008**, *77*, 383–414.
20. Verberkmoes, N. C.; Russell, A. L.; Shah, M.; Godzik, A.; Rosenquist, M.; Halfvarson, J.; Lefsrud, M. G.; Apajalahti, J.; Tysk, C.; Hettich, R. L.; Jansson, J. K. *ISME J.* **2009**, *3*, 179–189.
21. Chatterjee, S.; Stupp, G. S.; Park, S. K. R.; Ducom, J.-C.; Yates, J. R.; Su, A. I.; Wolan, D. W. *BMC Genomics* **2016**, *17*, 642.
22. Mayers, M. D.; Moon, C.; Stupp, G. S.; Su, A. I.; Wolan, D. W. *J. Proteome Res.* **2017**, *16*, 1014–1026.
23. Tsai, C.-S. C.; Yen, H.-Y. H.; Lin, M.-I. M.; Tsai, T.-I. T.; Wang, S.-Y. S.; Huang, W.-I. W.; Hsu, T.-L. T.; Cheng, Y.-S. E. Y.; Fang, J.-M. J.; Wong, C.-H. C. *Proc. Nat. Acad. Sci. USA* **2013**, *110*, 2466–2471.

24. Watts, A. G.; Damager, I.; Amaya, M. L.; Buschiazzo, A.; Alzari, P.; Frasch, A. C.; Withers, S. G. *J. Am. Chem. Soc.* **2003**, *125*, 7532–7533.
25. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.
26. Luchansky, S. J.; Goon, S.; Bertozzi, C. R. *Chembiochem* **2004**, *5*, 371–374.
27. Han, S.; Collins, B. E.; Bengtson, P.; Paulson, J. C. *Nat. Chem. Biol.* **2005**, *1*, 93–97.
28. Tanaka, Y.; Kohler, J. J. *J. Am. Chem. Soc.* **2008**, *130*, 3278–3279.
29. Feng, L.; Hong, S.; Rong, J.; You, Q.; Dai, P.; Huang, R.; Tan, Y.; Hong, W.; Xie, C.; Zhao, J.; Chen, X. *J. Am. Chem. Soc.* **2013**, *135*, 9244–9247.
30. Stewart, E. J. *J. Bacteriol.* **2012**, *194*, 4151–4160.
31. Frank, D. N.; Pace, N. R. *Curr. Opin. Gastroenterol.* **2008**, *24*, 4–10.
32. Chang, T.-C.; Lai, C.-H.; Chien, C.-W.; Liang, C.-F.; Adak, A. K.; Chuang, Y.-J.; Chen, Y.-J.; Lin, C.-C. *Bioconjug. Chem.* **2013**, *24*, 1895–1906.
33. Charych, D. H.; Nagy, J. O.; Spevak, W.; Bednarski, M. D. *Science* **1993**, *261*, 585–588.
34. Itzstein, Von, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W. *Nature* **1993**, *363*, 418–423.
35. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. *J. Am. Chem. Soc.* **1997**, *119*, 681–690.
36. Meindl, P.; Bodo, G.; Palese, P.; Schulman, J.; Tuppy, H. *Virology* **1974**, *58*, 457–463.
37. Li, Z.; Hao, P.; Li, L.; Tan, C. Y. J.; Cheng, X.; Chen, G. Y. J.; Sze, S. K.; Shen, H. M.; Yao, S. Q. *Angew. Chem. Int. Ed.* **2013**, *52*, 8551–8556.
38. Church, R.; Weiss, M. J. *J. Org. Chem.* **1970**, *35*, 2465–2471.
39. Rostovtsev, V. V. V.; Green, L. G. L.; Fokin, V. V. V.; Sharpless, K. B. K. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
40. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192–3193.
41. Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004**, *6*, 2853–2855.
42. Kuhn, R.; Lutz, P.; Macdonald, D. L. *Chemische Berichte* **1966**, *99*, 611–617.
43. Roy, R.; Laferriere, C. A. *Can. J. Chem.* **1990**, *68*, 2045–2054.
44. Nagy, J. O.; Bednarski, M. D. *Tetrahedron Lett.* **1991**, *32*, 3953–3956.
45. Paulsen, H.; Matschulat, P. *Eur. J. Org. Chem.* **1991**, *1991*, 487–495.
46. Kulikova, N. Y.; Shpirt, A. M.; Kononov, L. O. *Synthesis* **2006**, *24*, 4113–4114.
47. Okamoto, K.; Tadao, K.; Toshio, G. *Bull. Chem. Soc. Japan* **2006**, *60*, 631–636.
48. Mammen, M.; Choi, S.-K.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **1998**, *37*, 2754–2794.
49. Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E. W.; Haag, R. *Angew. Chem. Int. Ed.* **2012**, *51*, 10472–10498.
50. Blixt, O.; Collins, B. E.; van den Nieuwenhof, I. M.; Crocker, P. R.; Paulson, J. C. *J. Biol. Chem.* **2003**, *278*, 31007–31019.
51. Blaum, B. S.; Hannan, J. P.; Herbert, A. P.; Kavanagh, D.; Uhrín, D.; Stehle, T. *Nat. Chem. Biol.* **2015**, *11*, 77–82.
52. Ferreira, V. P.; Pangburn, M. K.; Cortés, C. *Mol. Immunol.* **2010**, *47*, 2187–2197.
53. Ram, S.; Sharma, A. K.; Simpson, S. D.; Gulati, S.; McQuillen, D. P.; Pangburn, M. K.; Rice, P. A. *J. Exp. Med.* **1998**, *187*, 743–752.
54. Esparza-Gordillo, J.; Soria, J. M.; Buil, A.; Almasy, L.; Blangero, J.; Fontcuberta, J.; Rodríguez de Córdoba, S. *Immunogenetics* **2004**, *56*, 77–82.
55. Sofat, R.; Mangione, P. P.; Gallimore, J. R.; Hakobyan, S.; Hughes, T. R.; Shah, T.; Goodship, T.; D'Aiuto, F.; Langenberg, C.; Wareham, N.; Morgan, B. P.; Pepys, M. B.; Hingorani, A. D. *J. Immunol. Methods* **2013**, *390*, 63–73.
56. Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.
57. Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. *Mol. Cell. Proteomics* **2002**, *1*, 947–955.

58. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. *J. Proteome Res.* **2006**, *5*, 2339–2347.