Genetically Encoded Boronolectin as a Specific Red Fluorescent UDP-GlcNAc Biosensor

Jing Zhang,[1,2] Zefan Li,[1,2] Yu Pang,[2,3] Yichong Fan,[1,2†] and Hui-wang Ai[1,2,3,4*]

1 Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, 1340 Jefferson Park Ave, Charlottesville, Virginia, 22908, USA

2 Center for Membrane and Cell Physiology, University of Virginia School of Medicine, Charlottesville, Virginia, 22908, USA

3 Department of Chemistry, University of Virginia, Charlottesville, Virginia, 22904, USA

4 UVA Comprehensive Cancer Center, University of Virginia, Charlottesville, Virginia, 22903, USA

ABSTRACT: There is great interest in developing boronolectins, which are synthetic lectin mimics containing a boronic acid functional group for reversible recognition of diol-containing molecules, such as glycans and ribonucleotides. However, it remains a significant challenge to gain specificity. Here, we present a genetically encoded boronolectin, which is a hybrid protein consisting of a noncanonical amino acid (ncAA) p-boronophenylalanine (pBoF), natural-lectin-derived peptide sequences, and a circularly permuted red fluorescent protein (cpRFP). The genetic encodability permitted a straightforward protein engineering process to derive a red fluorescent biosensor that can specifically bind uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), an important nucleotide sugar involved in metabolic sensing and cell signaling. We further characterized the resultant boronic acid- and peptide-assisted UDP-GlcNAc sensor (bapaUGAc) both in vitro and in live mammalian cells. Because UDP-GlcNAc in the endoplasmic reticulum (ER) and Golgi apparatus plays essential roles in glycosylating biomolecules in the secretory pathway, we genetically expressed bapaUGAc in the ER and Golgi and validated the sensor for its responses to metabolic disruption and pharmacological inhibition. In addition, we combined bapaUGAc with UGAcS, a recently reported green fluorescent UDP-GlcNAc sensor based on an alternative sensing mechanism, to monitor UDP-GlcNAc level changes in the ER and cytosol simultaneously. We expect our work to facilitate the future development of specific boronolectins for carbohydrates. In addition, this newly developed genetically encoded bapaUGAc sensor will be a valuable tool for studying UDP-GlcNAc and glycobiology.

Carbohydrates constitute one of the four major classes of biomolecules, and they play a crucial role in many physiological and pathological processes. As such, specific carbohydrate binders are highly sought after for their potential applications in research, diagnostics, therapy, and drug delivery.[1-3] In recent decades, researchers have focused on developing carbohydrate-binding antibodies, lectins, aptamers, and synthetic lectin mimics.[4-6] Despite these efforts, the development of specific binders for biologically significant carbohydrates remains a significant challenge in the field.

Boronic acids can undergo reversible esterification reactions with molecules that contain cis-diol functional groups (Fig. 1a).[7] This property allows boronic acids to recognize and bind to carbohydrates, leading to the development of synthetic lectin mimics known as "boronolectins". However, these interactions are often promiscuous and lack specificity. Researchers have attempted to improve the specificity of boronic acid-

carbohydrate interactions by using multiple coordination sites, steric effects, and molecular geometry, but these approaches have only been successful for a few examples.[6,8-10] A generalizable approach is still lacking.

Genetic code expansion (GCE) is a technology that allows for the convenient expression of proteins containing noncanonical amino acids (ncAAs) in live cells and organisms.[11,12] GCE uses engineered tRNA and aminoacyl-tRNA synthetase to introduce ncAAs into proteins in a site-specific manner. Previous studies have enabled the genetic encoding of p-boronophenylalanine (pBoF, Fig. 1b) in E. coli and mammalian cells.[13,14] In addition, pBoF has been introduced into a phage display library, leading to the identification of pBoF-containing single-chain variable antibody fragments (scFvs) that bind to glucosamine (GlcN).[15] However, the specificity of these pBoF-containing scFvs to glucosamine, compared to other structurally similar carbohydrates, was not sufficiently addressed in this previous study.[15]
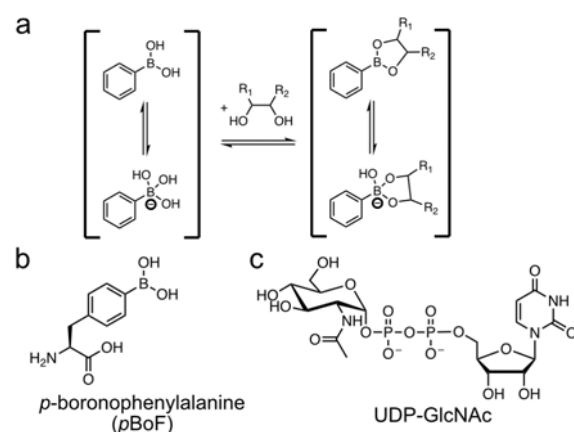

Figure 1. (a) Chemical reaction between phenylboronic acid and diol. (b, c) Chemical structures of the indicated compounds.

Fluorescent protein (FP)-based biosensors have become popular tools for biological and medical research because they can provide real-time and specific detection of a wide range of biomolecules and bioactivities.[16] Introducing ncAAs into FPs is an attractive strategy for developing new FP-based biosensors.[17,18] Previously, we and others incorporated pBoF into FPs and developed reaction-based biosensors for hydrogen peroxide and peroxynitrite.[14,19-21] Despite the progress,

pBoF-modified FPs have not yet been used to sense glycans.

Based on the knowledge and results described above, we hypothesize that pBoF-modified FPs may be engineered into specific biosensors for carbohydrates. Because GCE provides a practical method for the genetic encoding of pBoF-modified FPs, specificity may be achieved via genetic protein engineering. To test our hypothesis, we selected uridine diphosphate N-acetylglucosamine (UDP-GlcNAc, Fig. 1c), an important cell metabolite, as our first target.

UDP-GlcNAc is considered an integrator of nutritional and metabolic signals.[22] In mammalian cells, UDP-GlcNAc can be synthesized de novo from glucose, glutamine, acetyl-coenzyme A (Ac-CoA), adenosine triphosphate (ATP), and uridine triphosphate (UTP) via the hexosamine biosynthetic pathway (HBP).[23] Subsequently, UDP-GlcNAc, along with other nucleotide sugars, is used as glycosyl donors in the process of glycosylation that covalently attach carbohydrate moieties to biomolecules. In the cytosol and nucleus, UDP-GlcNAc is used by O-GlcNAc transferase (OGT) to add O-GlcNAc modifications to proteins, resulting in reversible and regulated cell signaling.[24] In the endoplasmic reticulum (ER) and Golgi apparatus, UDP-GlcNAc and other nucleotide sugars are used by glycosyltransferases to form complex glycosylation products, which may be further displayed on the cell surface or secreted to the extracellular space.[25]

Because of the importance of UDP-GlcNAc in metabolism and signaling, we have devoted efforts to developing fluorescent UDP-GlcNAc biosensors. We recently reported a green fluorescent biosensor, namely UGAcS, which is an insertion of a circularly permuted green FP (cpGFP) into an inactive mutant of an E. coli UDP-GlcNAc transferase.[26] UGAcS has the right affinity for monitoring the levels of UDP-GlcNAc in the cytosol. In this manuscript, we describe a parallel effort in our lab, which explores a pBoF-modified circularly permuted red FP (cpRFP) for the specific detection of UDP-GlcNAc (Fig. 2a). The work has resulted in a novel genetically encoded red fluorescent boronolectin, denoted as bapaUGAc, with a suit-

able affinity for detecting the levels of UDP-GlcNAc in the ER and Golgi.

Our biosensor engineering journey started with cpmApple, a cpRFP mutant derived from a previously reported calcium indicator, R-GECO1 (Supporting Information, Fig. S1).[27,28] Since cpmApple showed relatively low folding efficiency, we randomized the gene using error-prone polymerase chain reactions (EP-PCRs) and screened the library for increased fluorescence in E. coli colonies at 37 °C. We repeated this process for three rounds, resulting in an enhanced cpmApple mutant (ecpApple) with three mutations from cpmApple (Fig. S2). Next, we sought to introduce pBoF into ecpApple. Because we aimed at using carbohydrate binding to modulate fluorescence, we selected two residues (S14 and K30, Figs. S2 & S3) in spatial proximity to the chromophore of ecpApple and site-specifically introduced pBoF to each of the two residues. Among the two mutants, ecpApple with pBoF at residue 30 (ecpApple-K30B) showed a reproducible fluorescence turn-off response to 5 mM UDP-GlcNAc, but the response was not specific since several other carbohydrate analogs could also cause fluorescence changes.
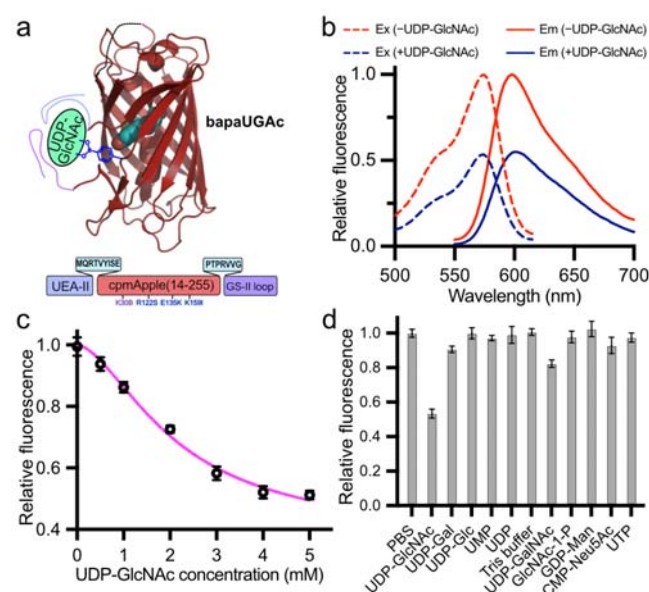

Figure 2. (a) Schematic illustration of bapaUGAc and its interaction with UDP-GlcNAc. The cpRFP chromophore is displayed as cyan spheres. The bottom shows the domain arrangement with the linkers and mutations highlighted, and B represents the pBoF residue. (b) Fluorescence excitation (dashed line) and emission (solid line) spectra of bapaUGAc in the presence (blue) and absence (red) of UDP-GlcNAc. (c) Dose-dependent fluorescence response. (d) Specificity characterization with 1 µM bapaUGAc protein and 5 mM of each compound. Data are presented as mean and s.d. of three technical replicates.

To increase the specificity of the response, we appended a 13-amino-acid peptide sequence derived from a natural N-acetylglucosamine (GlcNAc)-binding lectin, UEA-II (see bapaUGAc0.1 in Fig. S1).[29] Next, we added several randomized residues to the C-terminus, followed by examining the addition of another peptide sequence derived from natural GlcNAc binding lectins. We tested several options, and a 14-amino-acid loop sequence derived from the GS-II lectin[30,31] led to the best responsiveness and specificity (bapaUGAc0.3 in Fig. S1). Furthermore, we carried out three additional rounds of error-prone PCRs, and in addition, optimized the linker between the UEA-II-derived peptide and the N-terminus of ecpApple. Our effort resulted in the final bapaUGAc mutant (Figs. 2a, S1, and S2).

The fluorescence excitation and emission peaks of bapaUGAc as a purified protein were at 575 and 600 nm, respectively (Fig. 2b). The fluorescence decreased by ~50% in response to 5 mM UDP-GlcNAc. We further examined the response using several UDP-GlcNAc concentrations, and the apparent dissociate constant (Kd) was determined to be 2.14 ± 0.29 mM (Fig. 2c). The response displayed some cooperative effect, as the Hill coefficient was 1.7. In addition, we examined the specificity of bapaUGAc among common nucleotide sugars and nucleotides (Fig. 2d). UDP-GlcNAc caused robust fluorescence turn-off, while other tested compounds induced no or only small fluorescence changes. As a negative control, we prepared the ecpApple protein and confirmed its unresponsiveness to UDP-GlcNAc (Fig. S4).

We next explored the use of bapaUGAc to image UDP-GlcNAc level changes in mammalian cells. The Kd of bapaUGAc is too high compared to the concentrations of UDP-GlcNAc in the cytosol. On the other hand, nucleotide sugars are more concentrated in the ER and Golgi,[32] so we

tested whether bapaUGAc could sense UDP-GlcNAc in the ER and Golgi. We used signal peptides to genetically express bapaUGAc in the corresponding subcellular compartments of human embryonic kidney (HEK) 293T cells (Fig. 3a,b). We first treated the cells with GlcN, a potent precursor that can boost the biosynthesis of UDP-GlcNAc (Fig. 3c). As expected, the fluorescence of both ER- and Golgi-localized bapaUGAc decreased in response to GlcN (Fig. 3d,e). Next, we treated the cells with 2-deoxy-d-

glucose (2-DG), a glucose analog inhibiting hexokinase (HK) and phosphoglucose isomerase (PGI) (Fig. 3c). The addition of 2-DG caused a robust rise of bapaUGAc signals in the ER or Golgi (Fig. 3 f,g). In contrast, cells expressing ecpApple in the ER or Golgi were unresponsive to GlcN and 2-DG (Fig. 3 d,e,f,g). Collectively, the results support that bapaUGAc can detect UDP-GlcNAc level changes in the ER and Golgi in living mammalian cells in response to metabolic disruption and pharmacological inhibition.
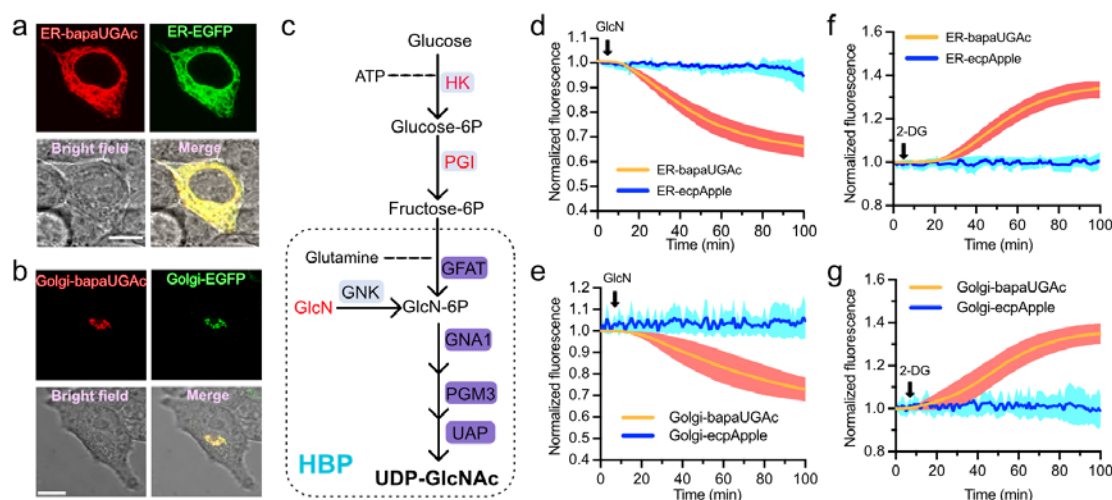

Figure 3. Monitoring UDP-GlcNAc level changes in the ER and Golgi in live mammalian cells using bapaUGAc. (a, b) Co-expression of ER-bapaUGAc and ER-EGFP (a), or Golgi-bapaUGAc and Golgi-EGFP (b) in HEK 293T cells. Scale bar, 10 µm. (c) Schematic illustration of the hexosamine biosynthetic pathway (HBP). Glucosamine (GlcN), as well as hexokinase (HK) and phosphoglucose isomerase (PGI) that are enzymes inhibited by 2-deoxy-d-glucose (2-DG), are highlighted in magenta. (d, e) Time-lapse responses of ER (d) or Golgi (e) localized bapaUGAc in HEK 293T to extracellular addition of GlcN (10 mM) (f, g) Time-lapse responses of ER (f) or Golgi (g) localized bapaUGAc in HEK 293T to extracellular addition of 2-DG (10 mM). Data are presented as mean and s.d. of eight individual cells from three technical replicates.
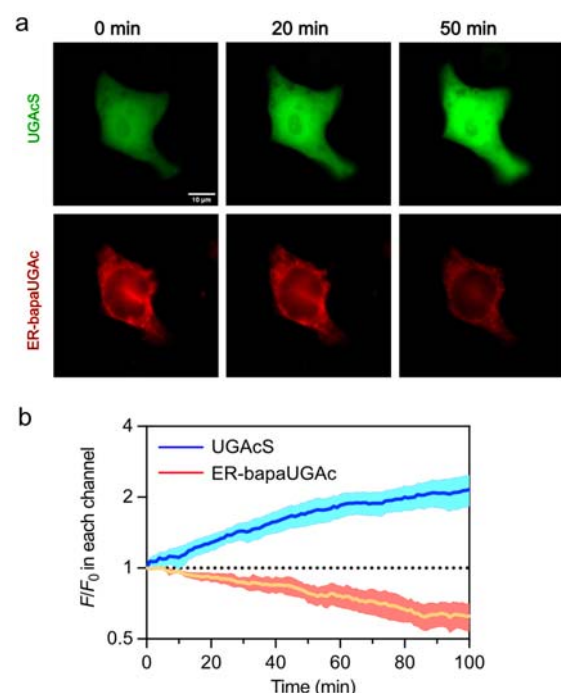
Figure 4. Dual-color imaging of UDP-GlcNAc level changes in the cytosol and ER using green fluorescent UGAcS and red fluorescent bapaUGAc. (a) Representative fluorescence images. Scale bar, 10 µm. (b) Quantitation of the responses of UGAcS and bapaUGAc to extracellular addition of GlcN (10 mM). F/F_0 represents the fluorescence intensities of individual cells normalized to the value at starting time. Data are presented as mean and s.d. of 12 individual cells from three technical replicates.

To visualize UDP-GlcNAc in different subcellular compartments, we paired ER-localized bapaUGAc with our cytosolic UGAcS biosensor. We co-expressed the two biosensors in HEK 293T cells and monitored fluorescence changes in the green and red channels after adding GlcN, a precursor for UDP-GlcNAc synthesis (Fig. 4). Since UGAcS is a fluorescence turn-on sensor, the green fluorescence of UGAcS increased in response to GlcN, indicating an increase in UDP-GlcNAc in the cytosol. Meanwhile, the red fluorescence of bapaUGAc decreased over time, suggesting an accompanied increase in UDP-GlcNAc in the ER. These results support the compatibility of bapaUGAc and UGAcS for dual-color, dual-compartment imaging of UDP-GlcNAc.

In summary, we have integrated GCE and protein engineering to create a new FP-based boronolectin called bapaUGAc, which can serve as a specific sensor for UDP-GlcNAc. We characterized the functionality of bapaUGAc both in vitro and in live cells. We further achieved dual-color monitoring of UDP-GlcNAc levels in two different subcellular compartments. We expect bapaUGAc to be a valuable research tool for studying UDP-GlcNAc and glycobiology. In addition, our results will catalyze the future development of specific boronolectins for other carbohydrate molecules.

AUTHOR INFORMATION

Corresponding Author

* Corresponding author. E-mail: huiwang.ai@virginia.edu

Present Addresses

† Present Address: AbbVie Inc., Irvine, CA 92612

Author Contributions

J.Z., Y.F., and Y.P. engineered the sensors. J.Z. performed characterization experiments with the help of Z.L.. J.Z. analyzed data, prepared figures, and wrote the manuscript. H.-w.A. supervised research, revised figures, and wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

REFERENCES

(1) Jin, S.; Cheng, Y.; Reid, S.; Li, M.; Wang, B. Carbohydrate recognition by boronolectins, small molecules, and lectins. *Med. Res. Rev.* **2010,** *30* (2), 171-257.

(2) Bies, C.; Lehr, C. M.; Woodley, J. F. Lectin-mediated drug targeting: history and applications. *Adv Drug Deliv Rev* **2004,** *56* (4), 425-35.

(3) Ghazarian, H.; Idoni, B.; Oppenheimer, S. B. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **2011,** *113* (3), 236-47.

(4) Heimburg-Molinaro, J.; Rittenhouse-Olson, K. Development and characterization of antibodies to carbohydrate antigens. *Methods Mol. Biol.* **2009,** *534*, 341-57.

(5) Li, M.; Lin, N.; Huang, Z.; Du, L.; Altier, C.; Fang, H.; Wang, B. Selecting aptamers for a glycoprotein through the incorporation of the boronic acid moiety. *J. Am. Chem. Soc.* **2008,** *130* (38), 12636-8.

(6) Edwards, N. Y.; Sager, T. W.; McDevitt, J. T.; Anslyn, E. V. Boronic acid based peptidic receptors for pattern-based saccharide sensing in neutral aqueous media, an application in real-life samples. *J. Am. Chem. Soc.* **2007,** *129* (44), 13575-83.

(7) Springsteen, G.; Wang, B. A detailed examination of boronic acid-diol complexation. *Tetrahedron* **2002,** *26* (58), 5291-5300.

(8) James, T. D.; Shinkai, S. Artificial receptors as chemosensors for carbohydrates. *Host-guest chemistry* **2002**, 159-200.

(9) Williams, G. T.; Kedge, J. L.; Fossey, J. S. Molecular Boronic Acid-Based Saccharide Sensors. *ACS Sens* **2021,** *6* (4), 1508-1528.

(10) Zhao, J.; Davidson, M. G.; Mahon, M. F.; Kociok-Köhn, G.; James, T. D. An enantioselective fluorescent sensor for sugar acids. *J. Am. Chem. Soc.* **2004,** *126* (49), 16179-86.

(11) Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010,** *79*, 413-44.

(12) Ai, H. W. Biochemical analysis with the expanded genetic lexicon. *Anal Bioanal Chem* **2012,** *403* (8), 2089-2102.

(13) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G. A genetically encoded boronate-containing amino acid. *Angew. Chem. Int. Ed. Engl.* **2008,** *47* (43), 8220-3.

(14) Chen, Z. J.; Ren, W.; Wright, Q. E.; Ai, H. W. Genetically encoded fluorescent probe for the selective detection of peroxynitrite. *J. Am. Chem. Soc.* **2013,** *135* (40), 14940-3.

(15) Liu, C. C.; Mack, A. V.; Brustad, E. M.; Mills, J. H.; Groff, D.; Smider, V. V.; Schultz, P. G. Evolution of proteins with genetically encoded "chemical warheads". *J. Am. Chem. Soc.* **2009,** *131* (28), 9616-7.

(16) Greenwald, E. C.; Mehta, S.; Zhang, J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chem Rev* **2018,** *118* (24), 11707-11794.

(17) Niu, W.; Guo, J. Expanding the chemistry of fluorescent protein biosensors through genetic incorporation of unnatural amino acids. *Mol. Biosyst.* **2013,** *9* (12), 2961-70.

(18) Chen, S.; Chen, Z. J.; Ren, W.; Ai, H. W. Reaction-based genetically encoded fluorescent hydrogen sulfide sensors. *J. Am. Chem. Soc.* **2012,** *134* (23), 9589-92.

(19) Chen, Z. J.; Tian, Z.; Kallio, K.; Oleson, A. L.; Ji, A.; Borchardt, D.; Jiang, D. E.; Remington, S. J.; Ai, H. W. The N-B Interaction through a Water Bridge: Understanding the Chemoselectivity of a Fluorescent Protein Based Probe for Peroxynitrite. *J. Am. Chem. Soc.* **2016,** *138* (14), 4900-7.

(20) Chen, Z.; Zhang, S.; Li, X.; Ai, H. W. A high-performance genetically encoded fluorescent biosensor for imaging physiological peroxynitrite. *Cell Chem*

Biol **2021**, DOI:10.1016/j.chembiol.2021.01.013.

(21) Wang, F.; Niu, W.; Guo, J. T.; Schultz, P. G. Unnatural Amino Acid Mutagenesis of Fluorescent Proteins. *Angewandte Chemie-International Edition* **2012,** *51* (40), 10132-10135.

(22) Bond, M. R.; Hanover, J. A. A little sugar goes a long way: the cell biology of O-GlcNAc. *J. Cell Biol.* **2015,** *208* (7), 869-80.

(23) Akella, N. M.; Ciraku, L.; Reginato, M. J. Fueling the fire: emerging role of the hexosamine biosynthetic pathway in cancer. *BMC Biol.* **2019,** *17* (1), 52.

(24) Schwein, P. A.; Woo, C. M. The O-GlcNAc Modification on Kinases. *ACS Chem. Biol.* **2020,** *15* (3), 602-617.

(25) Khoder-Agha, F.; Sosicka, P.; Escriva Conde, M.; Hassinen, A.; Glumoff, T.; Olczak, M.; Kellokumpu, S. N-acetylglucosaminyltransferases and nucleotide sugar transporters form multi-enzyme-multi-transporter assemblies in golgi membranes in vivo. *Cell Mol Life Sci* **2019,** *76* (9), 1821-1832.

(26) Li, Z.; Zhang, J.; Ai, H. W. Genetically Encoded Green Fluorescent Biosensors for Monitoring UDP-GlcNAc in Live Cells. *ACS Cent Sci* **2021,** *7* (10), 1763-1770.

(27) Zhao, Y.; Araki, S.; Wu, J.; Teramoto, T.; Chang, Y. F.; Nakano, M.; Abdelfattah, A. S.; Fujiwara, M.; Ishihara, T.; Nagai, T.et al. An expanded palette of genetically encoded Ca^{2+} indicators. *Science* **2011,** *333* (6051), 1888-91.

(28) Fan, Y.; Chen, Z.; Ai, H. W. Monitoring redox dynamics in living cells with a redox-sensitive red fluorescent protein. *Anal. Chem.* **2015,** *87*, 2802-2810.

(29) Loris, R.; De Greve, H.; Dao-Thi, M. H.; Messens, J.; Imberty, A.; Wyns, L. Structural basis of carbohydrate recognition by lectin II from Ulex europaeus, a protein with a promiscuous carbohydrate-binding site. *J. Mol. Biol.* **2000,** *301* (4), 987-1002.

(30) Zhu-Salzman, K.; Shade, R. E.; Koiwa, H.; Salzman, R. A.; Narasimhan, M.; Bressan, R. A.; Hasegawa, P. M.; Murdock, L. L. Carbohydrate binding and resistance to proteolysis control insecticidal activity of Griffonia simplicifolia lectin II. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (25), 15123-8.

(31) Zhu, K.; Bressan, R. A.; Hasegawa, P. M.; Murdock, L. L. Identification of N-acetylglucosamine binding residues in Griffonia simplicifolia lectin II. *FEBS Lett.* **1996,** *390* (3), 271-4.

(32) Parker, J. L.; Newstead, S. Gateway to the Golgi: molecular mechanisms of nucleotide sugar transporters. *Curr. Opin. Struct. Biol.* **2019,** *57*, 127-134.