

Peptide synthesis

Peptides were purchased from Tufts University Core Facility (Boston, Ma) or from Synthetic Biomolecules (San Diego, Ca) and synthesized using standard Fmoc chemistry. Fluorescein was incorporated at the time of synthesis either at the N-terminus using a 6-aminohexanoic acid-linked FITC or at the C-terminus using Fmoc-Lys(5/6 - FAM) Wang resin (AnaSpec™; catalog # 20146). The second label was either 7-diethylaminocoumarin-3-carboxylic acid, succinimidyl ester (DEAC, SE; AnaSpec™ catalog # 81211) or QXL™ -520, succinimidyl ester (QXL™ -520, SE; AnaSpec™ catalog # 81831) and was incorporated after synthesis to the crude peptide under the following conditions: Peptides were dissolved to ~5 mM in 250 mM NaHCO₃ buffer, pH 8.2; 0.5 eq of the NHS-ester was dissolved in DMSO (1/10 volume of the aqueous buffer) and quickly added to the peptide. After 30 minutes an additional 0.5 eq of the NHS-ester in a minimal amount of DMSO at room temperature. This was repeated one or two times and completion of the reaction was confirmed by LC/MS. Acidic and basic LC/MS conditions were developed using an Eclipse 5μm XD8-C8 column (Agilent) and could be used for all labeled peptides. Acidic conditions were used for all peptides with a net positive charge after labeling (typically DEAC labeled peptides) and basic conditions used for all net negatively charged peptides. Acidic LC/MS conditions were: 0 → 100% CH₃CN + 0.1% formic acid gradient over 20 minutes and basic conditions were: 0 → 100% CH₃CN + 0.1% NH₄OH gradient over 20 minutes. Peptides were purified in preparative scale using a Phenomenex Luna C18 5μm column, using the following conditions: 0 → 100% CH₃CN + 0.1% TFA gradient over 40 minutes for peptides with a net positive charge, and 0 → 100% CH₃CN + 0.1% NH₄OH gradient over 40 minutes for peptides with a net negative charge.

Preparative glycosylation of labeled peptides

For OGT: Purified FRET labeled peptides were added (5 μM final concentration) to OGT assay buffer (50 mM Tris, 20 mM CaCl₂, 0.05% Tween-20, 0.05% NP-40, pH=7.8), 200 μM UDP-GlcNAc, 0.05 units of alkaline phosphatase, and 500 nM OGT. For ppGalNAc T1 and T2: Purified FRET labeled peptides were added (5 μM final concentration) to ppGalNAc-T assay buffer (20mM imidazole, 0.05% Tween-20, 0.05% NP-40, 10mM MnCl₂, 1 mM β-mercaptoethanol, pH=7.7), 15 μM UDP-GalNAc, 0.05 units of alkaline phosphatase, and 20 nM ppGalNAc-T. Reactions were monitored by LC/MS using the column and conditions previously described. When complete, small amounts of glycosylated peptide were purified using the same column and conditions. Peptides with DEAC incorporated were quantitated by absorption at 492 nm using an extinction coefficient of 86,000 cm⁻¹M⁻¹; peptides with QXL™ -520 quencher were quantitated by absorption at 530 nm using an extinction coefficient of 26,000 cm⁻¹M⁻¹ x Q, where Q is the number of QXL™ -520 molecules incorporated into the peptide.

peptide	enzyme	Sequence
S1	OGT	FITC(Aha)STPV S SANMK _ε (DEAC)
S2	OGT	FITC(Aha)STPV S RANMK _ε (DEAC)
S3	OGT	FITC(Aha)STPV S RANMK _ε (QXL™ 520)
S4	OGT	FITC(Aha)STPV S RANMK _ε (QXL™ 520)R K _ε (QXL™ 520)
S5	ppGalNAc-T1	(DEAC)EPGPTEAPK _ε (FAM)
S6	ppGalNAc-T1	(QXL™ 520)EPGPTEAPK _ε (FAM)
S7	ppGalNAc-T2	(DEAC) EDAVTPGPK _ε (FAM)
S8	ppGalNAc-T2	(QXL™ 520) EDAVTPGPK _ε (FAM)

Table S1. Peptides used in this study. The glycosylated residue is shown in bold.

Determination of Z' and assay linearity

From the assay window (Figure 2a) a concentration of protease that completely cut the unglycosylated while not affecting the glycosylated peptide can be determined. Using this concentration of protease,

several hundred positive (BSA) and negative (OGT, ppGalNAc-T1, or ppGalNAc-T2) control wells were analyzed and Z' was calculated using the following equation:

$$Z' = 1 - [(3SD_+ + 3SD_-) / |Ave_+ - Ave_-|]$$

Assay linearity was tested for all peptides by combining purified glycosylated and unglycosylated peptide. Peptides that contained quencher were linear over the entire glycosylation range (Figure S1), while peptides that contained coumarin deviated from linearity at low and high levels of glycosylation. Linearity allows use of this assay for lower throughput applications that require greater sensitivity, such as kinetics. It also makes a quantitative data workup possible, which was exploited after the high-throughput screen (see below).

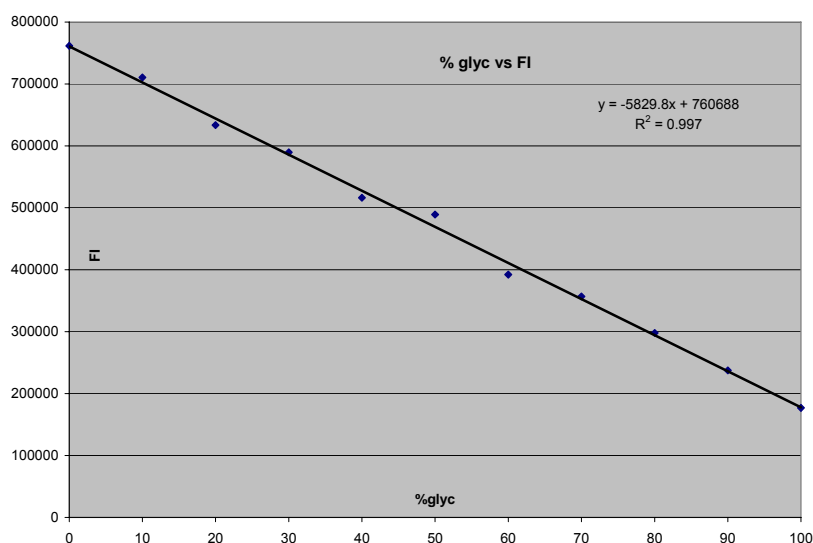


Figure S1: 535 nm emission (RFU) of mixtures of glycosylated and unglycosylated **S3**. RFU is shown on the y-axis, and percent glycopeptide is shown on the x-axis. The slope and R^2 values are shown.

Determining the effect of compound autofluorescence on assay readout

To test for autofluorescence, 100 nL of ICCB plate number 1364 (a mixture of commercially available known bioactive compounds) was pin transferred in duplicate into 20 μ L of a 5 μ M solution **S2** in buffer and the plates were read at 460 nm and 535 nm. These compounds are structurally diverse and represent the spectral characteristics of the larger libraries. A scatter plot of the duplicate 535nm/460nm ratio (Figure S2) reveals that roughly 5% of these compounds reproducibly autofluoresce near 460nm, and would be registered as positives in a high-throughput assay (since

increased 460 nm emission will result in a decrease in the 535nm/460nm ratio). This

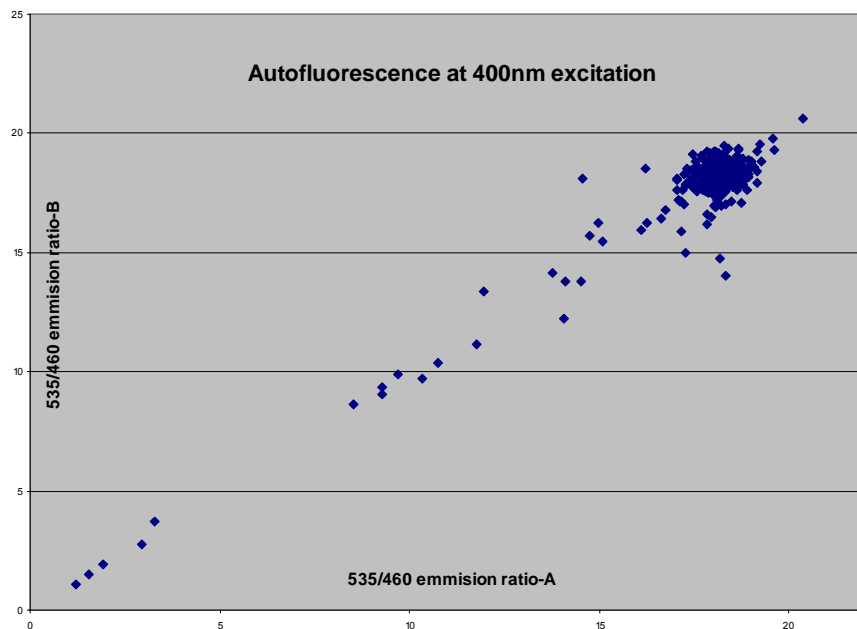


Figure S2. Scatter plot of 535/460 emission ratio of **S2** in the presence of known bioactive small molecules. The drop in ratio for ~5% of these compounds is dues to compound autofluorescence.

corresponds to a false positive frequency that is much higher than the ICCB cherry pick limit (0.2% of compounds screened) and would swamp out true hits. We considered adding an additional read step in between compound transfer and proteolysis which would successfully identify autofluorescent compounds, but this would also eliminate data for ~ 5% of the library members, including entire structural classes.

To solve this problem we switched to the donor quencher FRET approach. Compound autofluorescence with **S3** was assessed using the same pilot experiment described above. No compound autofluorescence was observed with 485 nm excitation. Note that since the readout in this case is total fluorescence, autofluorescence would result in a reproducible increase in signal.

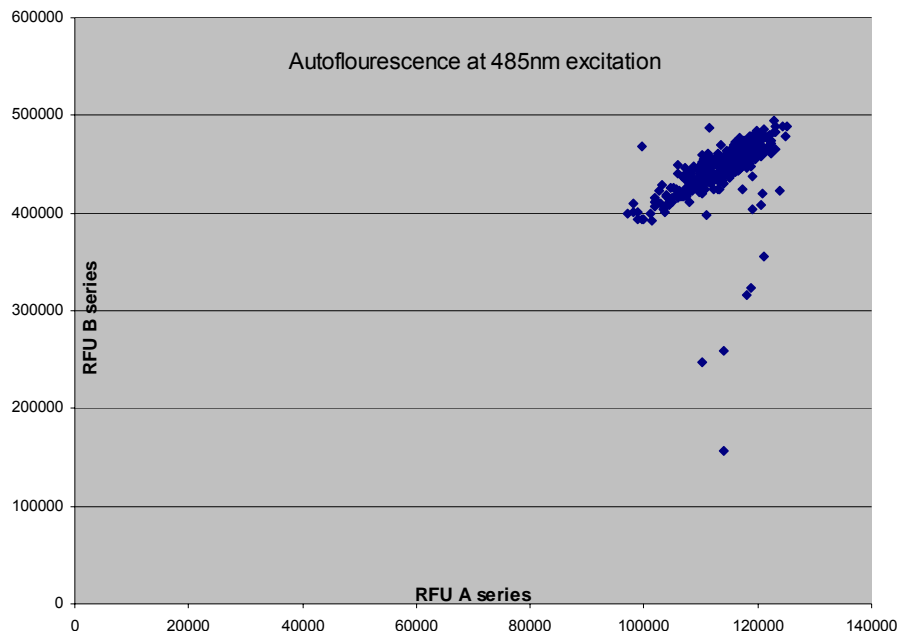


Figure S3. Scatter plot of 535 emission (RFU) of **S3** in the presence of known bioactive small molecules. Note that in this case, autofluorescence would lead to a reproducible increase in signal.

HTS assay development

It appeared that **S3** began to precipitate over time at stock concentrations (~ 1 mM), which prompted the synthesis of a slightly modified variant, **S4**. In addition to an extra arginine, **S4** has a second modified QXL™ 520 lysine residue, which led to a modest increase in the change in fluorescence between cut and uncut peptide and a five-fold lower optimal proteinase K concentration (~ 10 ng/mL).

We next sought to establish appropriate conversion targets for negative controls (in this case, the extent of glycosylation in the absence of inhibitors) and determine the concentration of reagents (*i.e.* OGT and both substrates) required to achieve this level of conversion. A target of 80% conversion was chosen, since this will result in a large signal but will still ensure that the reaction is not run past completion, potentially missing weak inhibitors. Using 4 μ M **S4**, 20 μ M UDP-GlcNAc, and 50-100 nM sOGT (depending on the enzyme batch), 70-80% conversion could be achieved in 2-4 hours. Because of the small but inherent variability, an aliquot of assay mix for each round of high throughput screening was analyzed by LC/MS to determine the actual conversion for that run.

OGT HTS Protocol. For high throughput screening, 384-well plates (Costar # 3710) were filled using a liquid handling robot with 10 μ L of a mixture of 8 μ M **S4**, 40 μ M UDP-GlcNAc, and buffer (50 mM Tris, 20 mM CaCl_2 , 0.05% Tween-20, 0.05% NP-40, pH=7.8, 500 μ M tris(hydroxypropyl)phosphine). Compound libraries were then transferred to the assay plates using a 100 nL pin array and 10 μ L of 500 nM sOGT in buffer (50 mM Tris, 20 mM CaCl_2 , 0.05% Tween-20, 0.05% NP-40, pH=7.8, 500 μ M tris(hydroxypropyl)phosphine) was added using a liquid handling robot. The compounds were assayed at a concentration of 25 μ g/mL or ~ 70 μ M, assuming an average compound MW of 350. After 2-3 hours, 20 μ L of a 20 ng/mL proteinase K solution in buffer (50 mM Tris, 20 mM CaCl_2 , 0.05% Tween-20, 0.05% NP-40, pH=7.8) was added and allowed to react for 30-90 minutes. The plates were read using a Perkin-Elmer Envision® microplate reader with a 480 nm excitation filter and 530 nm emission filter.

HTS data workup

Data workup began by calculating the average positive and negative control values independently for both duplicates of each plate. Using the linear relationship between percent glycosylation and fluorescence, the percent glycosylation was calculated for each well, with 0% corresponding to the average value of the positive controls and X% corresponding to the average value of the negative controls, where X is the empirically determined (by LC/MS) percent glycosylation for that reaction batch. Batches were normalized and compounds with < 70% activity in one duplicate and < 80% activity in the second duplicate were scored as hits.

Hit Validation

Hits were examined the secondary assay previously described with minor modifications. One microliter aliquots (5 mg/mL) of the hits were obtained from the ICCB, and these were normalized to 5.0 mM with DMSO. These compounds were screened at 8 two-fold dilutions ranging from 125 μ M to 900 nM. The reaction mixture containing 250 μ M peptide, 6.25 μ M UDP-¹⁴C-GlcNAc, ~20-40 nM sOGT, and buffer (50 mM Tris, 20 mM CaCl₂, 0.05% Tween-20, 0.05% NP-40, pH=7.8, 500 μ M tris(hydroxypropyl)phosphine). Reactions were run for 20-30 minutes, spotted on Whatman P81 phosphocellulose discs, washed three times for five minutes in 1% phosphoric acid, and counted by liquid scintillation counting. The IC₅₀ curve was fit to the following equation using Prism 4 software (Graphpad).

$$Y = Y_{\min} + (Y_{\max} - Y_{\min}) / (1 + 10^{((\log(X) - \log(IC_{50})) * h)})$$

where: X is the inhibitor concentration, Y is the reaction rate, and h is the hill slope.