PNAS RESEARCH ARTICLE | BIOCHEMISTRY CHEMISTRY

Phage display uncovers a sequence motif that drives polypeptide binding to a conserved regulatory exosite of O-GlcNAc transferase

Matthew G. Alteen[a], Richard W. Meek[b,c], Subramania Kolappan[d], Jil A. Busmann[d], Jessica Cao[e], Zoe O'Gara[e], Ying Chou[e], Ratmir Derda[e,f], Gideon J. Davies[b], and David J. Vocadlo[a,d,1]

Edited by Robert S. Haltiwanger, The University of Georgia, Athens, GA; received March 5, 2023; accepted August 29, 2023 by Editorial Board Member Stephen J. Benkovic

The modification of nucleocytoplasmic proteins by *O*-linked N-acetylglucosamine (*O*-GlcNAc) is an important regulator of cell physiology. *O*-GlcNAc is installed on over a thousand proteins by just one enzyme, *O*-GlcNAc transferase (OGT). How OGT is regulated is therefore a topic of interest. To gain insight into these questions, we used OGT to perform phage display selection from an unbiased library of ~10^9 peptides of 15 amino acids in length. Following rounds of selection and deep mutational panning, we identified a high-fidelity peptide consensus sequence, [Y/F]-x-P-x-Y-x-[I/M/F], that drives peptide binding to OGT. Peptides containing this sequence bind to OGT in the high nanomolar to low micromolar range and inhibit OGT in a noncompetitive manner with low micromolar potencies. X-ray structural analyses of OGT in complex with a peptide containing this motif surprisingly revealed binding to an exosite proximal to the active site of OGT. This structure defines the detailed molecular basis driving peptide binding and explains the need for specific residues within the sequence motif. Analysis of the human proteome revealed this motif within 52 nuclear and cytoplasmic proteins. Collectively, these data suggest a mode of regulation of OGT by which polypeptides can bind to this exosite to cause allosteric inhibition of OGT through steric occlusion of its active site. We expect that these insights will drive improved understanding of the regulation of OGT within cells and enable the development of new chemical tools to exert fine control over OGT activity.

glycosyltransferase | phage display | allosteric inhibition | exosite | X-ray structure

Post-translational glycosylation of proteins has emerged as serving essential roles in regulating cellular functions (1). One form of glycosylation involves addition of *N*-acetylglucosamine (GlcNAc) to the hydroxyl groups of serine and threonine residues of nuclear and cytosolic proteins (2, 3). This modification, known as *O*-GlcNAc, modulates human physiology through its diverse molecular roles in regulating transcription (4, 5) protein stability (6) and stress response (7). Despite the widespread occurrence of this posttranslational modification on hundreds of protein targets (8), cycling of *O*-GlcNAc is controlled by just two enzymes. The glycosyltransferase (GT) *O*-GlcNAc transferase (OGT) from family 41 (GT41) catalyzes the installation of *O*-GlcNAc to acceptor polypeptides using uridine diphosphate GlcNAc (UDP-GlcNAc) as the glycosyl donor (9). Hydrolytic cleavage of *O*-GlcNAc is effected by the glycoside hydrolase (GH) from family 84 (GH84) known as *O*-GlcNAcase (OGA), which hydrolyzes the glycosidic linkage to liberate free serine or threonine residues (9, 10). Since dysregulation of *O*-GlcNAc cycling has been genetically linked to X-linked intellectual disability (XLID) (11, 12) and also implicated in numerous pathological conditions, including neurodegenerative diseases (13, 14) and cancers (15, 16) there is considerable interest in studying the regulation of these two enzymes and generating new tools and insights that can be exploited to enable their manipulation (17).

The structure of OGT consists of two principal parts: an N-terminal superhelical section comprising 13.5 tetratricopeptide repeat (TPR) domains and a C-terminal catalytic region (Fig. 1*A*) (18–20). This C-terminal catalytic region can be divided into three folded domains: two catalytic domains (N-Cat and C-Cat) that together comprise the active site and an intervening domain (Int-D) which separates the catalytic domains and has a yet undefined function (9, 21). Selection of protein substrates by OGT is thought to be mediated by its TPR domains (21, 22), which form a superhelical groove extending into the active site of OGT (Fig. 1 *B* and *C*), but molecular details as to how it does this remain ill defined. An asparagine ladder lining the TPR superhelical groove has been found to coordinate the peptide backbone of substrates (23), and aspartate residues within this groove enable selection of a protein subset, but how OGT achieves greater specificity for various sets of proteins remains unknown.

Significance

Thousands of proteins within humans are modified by the monosaccharide N-acetylglucosamine (O-GlcNAc). O-GlcNAc regulates cellular physiology and is being pursued to create therapeutics. Remarkably, only one enzyme, O-GlcNAc transferase (OGT), installs O-GlcNAc, and its regulation is poorly understood. By affinity selection using a vast peptide library, we uncover a sequence motif that drives binding of polypeptides to OGT. An OGT-peptide complex shows how this motif binds to an allosteric site proximal to the active site and inhibits OGT in an unprecedented manner. Given the distribution of this sequence motif within the human proteome, proteins containing this motif likely regulate the activity of OGT, outlining a potential mode by which OGT could be controlled and opening up opportunities for research.

Author contributions: M.G.A., R.W.M., R.D., and D.J.V. designed research; M.G.A., R.W.M., S.K., J.A.B., and Y.C. performed research; M.G.A., R.W.M., S.K., J.A.B., J.C., Z.O., Y.C., R.D., G.J.D., and D.J.V. analyzed data; and M.G.A. and D.J.V. wrote the paper.

Competing interest statement: R.D. is a consultant with 48 Hour Discovery. J.C., Z.O, and Y.C. are employees or consultants with 48 Hour Discovery. R.D. is a shareholder of 48 Hour Discovery.

This article is a PNAS Direct Submission. R.S.H. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: dvocadlo@sfu.ca.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303690120/-/DCSupplemental.

Published October 11, 2023.

Because OGT has hundreds of known protein targets, yet lacks a clear consensus sequence governing glycosylation, the activity of OGT is thought to be regulated in various ways by cellular factors (21, 24, 25). Among these regulatory mechanisms, OGT dimer status has been found to influence its activity (20) as has an interaction with an intrinsically disordered region of OGA (26). Interactions between OGT and other proteins, mediated through direct binding to the TPR region, have also been proposed to regulate OGT activity (26). This last idea has gained support through quantitative proteomics studies showing that different cellular conditions alter the set of its interacting partners (27, 28). Interestingly, direct binding of candidate proteins identified in these proteomic studies has not yet been observed, and how such proteins might modulate OGT activity remains to be elucidated. In brief, how OGT is regulated remains enigmatic and a topic of considerable interest.

Among various approaches to aiding understanding of the roles of OGT is the use of inhibitors of this enzyme (29, 30). Moreover, OGT inhibitors may hold promise for the treatment of conditions, including various cancers, which are linked to excessive *O*-GlcNAcylation (15, 16). However, despite several high-throughput screening campaigns to identify small-molecule inhibitors of OGT, most compounds arising from these efforts have limited selectivity and cell permeability (31, 32). Extensive medicinal chemistry efforts, however, have led to OSMI-4, a nanomolar inhibitor OGT that shows utility within cells (33). We noted that identifying these HTS-derived inhibitors relies on activity-based assays or competitive displacement assays that are biased toward those compounds binding to the active site. Accordingly, compounds that bind OGT at alternative sites could be useful complements to the existing set of competitive OGT inhibitors. Toward this goal of identifying ligands that bind to OGT, we recently turned to display methods as an alternative approach to uncover ligands for OGT (34). We used the RaPID mRNA display system (35) to identify cyclic peptide inhibitors of OGT that bind to the TPR domain and exert allosteric inhibition (34, 36). In parallel to those efforts, we turned to using standard phage display as an alternative approach for the identification of linear peptides that bind OGT. We expected that the advantages of DNA-encoded linear peptide libraries, including high structural diversity, synthetic accessibility, and general biocompatibility, could enable uncovering peptide sequences binding to the TPR domains or active site that might serve as interesting leads for new inhibitors (37, 38). Additionally, since phage display screening methods rely on affinity-based selection, we reasoned that we might identify peptide ligands of OGT found within native proteins that may be either inhibitors or simply noninhibitory ligands. In pursuit of this goal, we screened immobilized OGT against a deep library comprising ~10^8 sequences of linear 15-amino acid peptides expressed on the surface of the pIII major coat protein of filamentous M13 phage. After identification of sequence preferences and affinity maturation of the selected phage, a panel of peptides with submicromolar noncompetitive inhibition of OGT were found. Furthermore, we made the surprising observation that these peptides bind an exosite within the intervening domain of the C-terminal section of OGT. Finally, we showed that inhibition by these peptides is correlated with the length of their sequences, consistent with a steric occlusion model that antagonizes binding of polypeptide substrates to the active site of OGT.

Results

Phage Library Development and Screening Results. To identify potential peptide ligands of OGT, we screened full-length OGT against an unbiased phage-display library constructed using codon-corrected trinucleotide phosphoramidite cassettes (TriNuc), which produce a single codon for each amino acid and result in equal distribution of amino acids within translated sequences (39). This

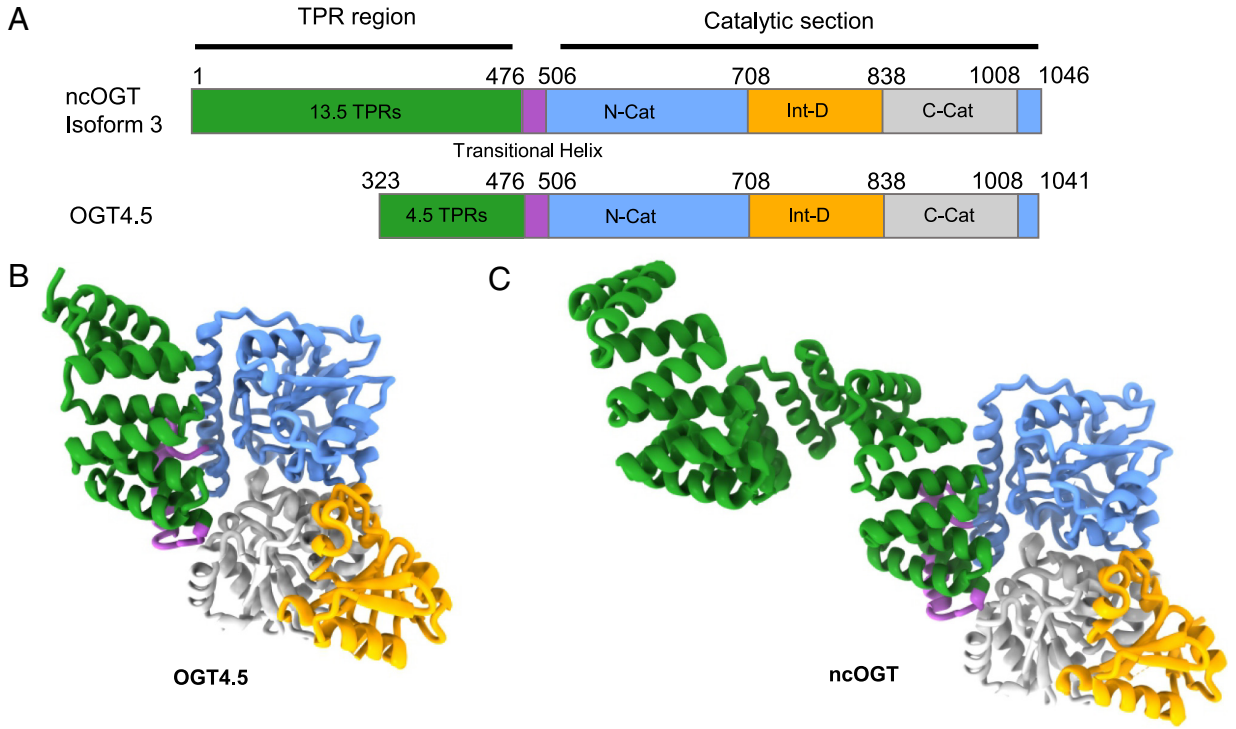

Fig. 1. Domain arrangement of OGT: (*A*) Schematic showing the distribution of folded domains across the primary sequence of full-length OGT (ncOGT) and the crystallized form of OGT (OGT4.5). Tetratricopeptide repeats (TPRs), N-terminal catalytic domain (N-Cat), intervening domain (Int-D), and C-terminal catalytic domain (C-Cat) are shown with the amino numbers defining the domain boundaries. (*B*) The domains, colored as in the primary sequence representation, mapped onto the OGT4.5 crystal structure and (*C*) onto the structure of full-length ncOGT (isoform 3).

library strategy eliminates amino acid sequence representation biases arising from variable codon redundancies within the standard genetic code. As a first step, we used a TriNuc-encoded library containing 19 codons for each amino acid, excluding cysteine, to produce a library of linear peptides comprising 15 randomized amino acids. These peptides were presented on M13 phage by fusion to the N terminus of the pIII coat protein. To do this, we generated a library we termed X15 that has a sequence diversity of ~10^9. We immobilized recombinant full-length OGT, bearing a hexahistidine-tag (His_6) fused to the N terminus, on magnetic nickel-NTA beads and then incubated the beads with the X15 phage library (Fig. 2*A*). We performed four rounds of phage panning to enrich phage binding to OGT from the larger phage population and analyzed enriched sequences from selected samples (Fig. 2 *B* and *C* and *SI Appendix*, Table S1) by lysis followed by DNA isolation and next-generation sequencing (NGS) (*SI Appendix*, Fig. S1) (40).

Phage populations exhibited a significant increase in recovery in panning on OGT-modified protein in rounds 3 and 4 when compared to rounds 1 and 2. This increased recovery was not

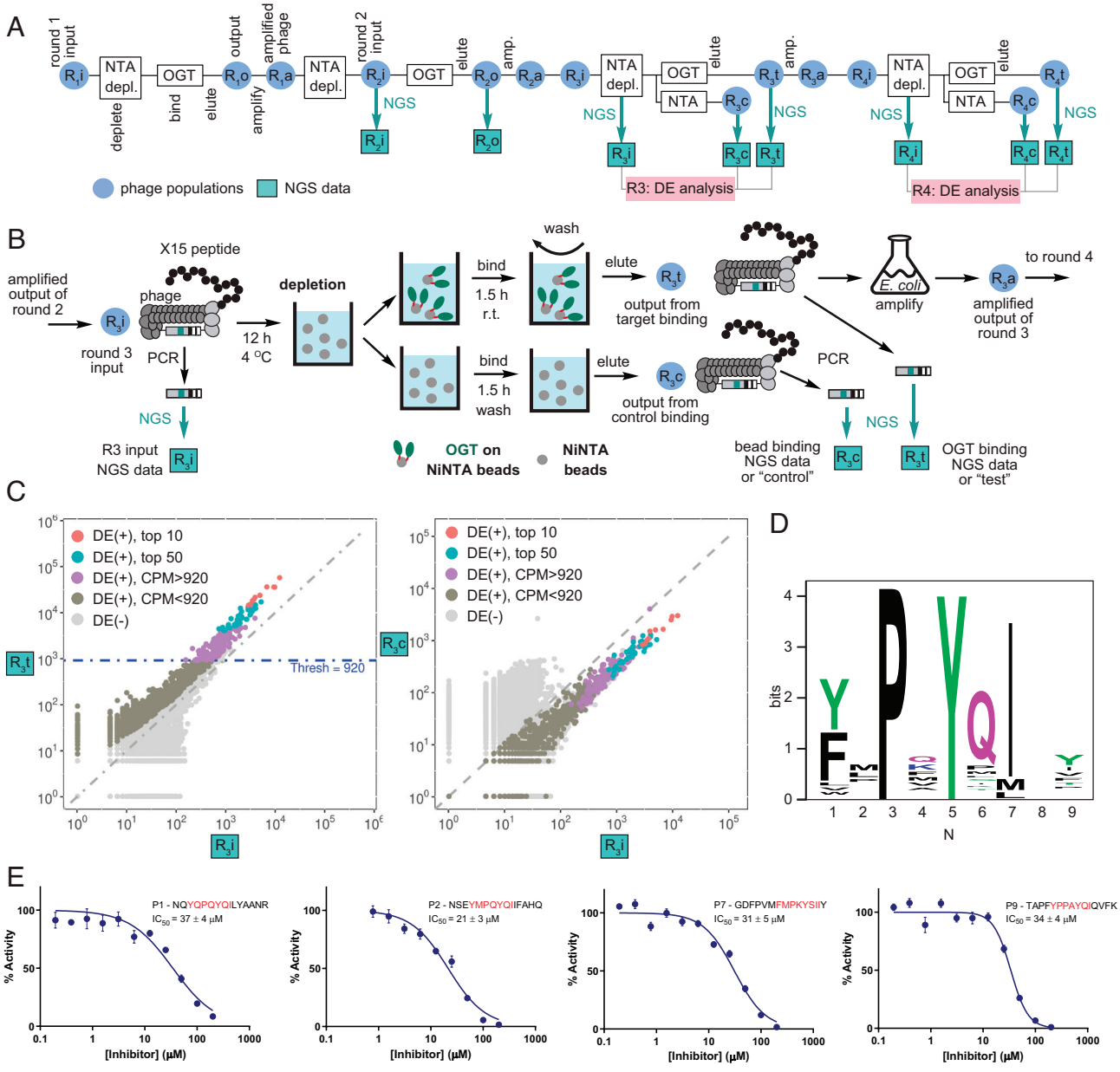

Fig. 2. Outline of phage display screening approaches and sequence conservation of the peptide binders of OGT: (*A*) Schematic of four rounds of the phage panning procedure using bead-immobilized OGT ("OGT") or target-free Ni-NTA beads ("NTA") as a control; bar chart representing the number of plaque forming units (PFU) of phage input and recovered from each selection round. (*B*) Detailed schematic of the third round of panning. In each round, the phage library was depleted against Ni-NTA beads to generate the input library, which was incubated with bead-immobilized OGT, followed by washing and elution. Samples from the resulting phage library were submitted for next generation sequencing (NGS) and amplified in *E. coli* for the next round. (*C*) Bioinformatic analysis of round 3 of screening NGS data. Each dot represents a sequence that was observed in the NGS data, colored dots represent 645 sequences that pass differential enrichment (DE) analysis and exhibit significant enrichment in OGT samples whereas sequences denoted by the gray dots have not passed the DE test. The left scatter plot (*C*) shows the counts per million (CPM) of each sequence in the bound to bead-immobilized OGT sample as compared to Input, whereas the right plot shows comparison of the same Input in binding to NiNTA-magnetic beads. (*D*) Sequence logo indicating the prevalence of amino acid residues in top 300 aligned peptide sequences selected that bind to OGT after four rounds of phage panning. (*E*) Concentration–response curves of four of the top peptide binders derived from the first round of phage panning that inhibit OGT. The consensus motif determined from analysis of phage panning results is shown in red for each peptide. % OGT activity was normalized relative to controls. Error bars represent SEM of quadruplicate measurements.

observed when panning on nickel-NTA beads (Fig. 2*A*), indicating binding to OGT enhanced recovery of phage. Analysis of the sequencing results was accomplished using multiple strategies. Initially, we ranked the sequences identified through NGS according to their normalized sequence count per million sequences (NCPM). Notably, out of 18,000 total sequences in the final enriched phage library, 81% of the overall enrichment as assessed using NCPM values was attributable to just the top 200 sequences (*SI Appendix*, Table S1), revealing a high degree of enrichment for this set of peptides. In the third round of selection, 645 sequences also passed a differential enrichment analysis (DE) in which the sequences were significantly enriched in binding to OGT when compared to input and but not in the experiments that used protein-free bead control (Fig. 2*C* and *SI Appendix*, Fig. S1*D*). By the fourth round of phage panning, the differential enrichment of peptide sequences with respect to input was modest (*SI Appendix*, Fig. S1*E*), and we therefore stopped selections at this point. We then binned the sequences into groups of 100 that we analyzed using the Kalign2 multiple sequence alignment tool. This analysis revealed that among these top 200 sequences, a strong consensus motif was present with 90% fidelity. Further analysis uncovered that the motif [Y/F]-x-P-x-Y-x-[I/M/L] dominated the top enriched sequences in rounds 3 and 4 (Fig. 2*D*), and it was detectable as early as round 2 of panning (*SI Appendix*, Fig. S1*C*). Revealingly, this seven-amino acid motif was present at different registers within the 15-amino acid sequence of the displayed peptide, indicating that variability at either end of the motif did not markedly impair binding to OGT. We observed some variability in positions between the conserved residues of this motif, suggesting oriented binding of these peptides to OGT. We also remarked on the presence of a conserved proline residue within this sequence that is consistent with observations from large-scale proteomic studies on the peptide substrate preferences of OGT (8, 41, 42). This last observation suggested to us that these peptides may bind within the active site of OGT.

In Vitro Inhibition of OGT by Prioritized Peptide Sequences. Having identified putative binders, we next proceeded to use synthetic peptides to examine whether these candidate ligands might bind to OGT. As noted above, the proline within the observed consensus motif suggested to us that these peptides were likely binding within the active site, which pointed to inhibition experiments as being potentially informative. We prioritized peptides selected from among the top 30 enriched sequences as assessed by NCPM. All these sequences exhibited at least a 1,000-fold enrichment over the level seen for background peptide sequences (*SI Appendix*, Table S1). The top 30 sequences were then grouped into five families based on amino acid composition within the consensus region and a total of 10 peptides, two from each

Table 1. Sequences and OGT inhibition of prioritized peptides from phage panning showing in bold the conserved motifs

First Generation Phage Peptides			
Peptide ID	Sequence	Motif family	OGT activity percent at 100 μM
P1	NQ**YQPQYQI**LYAANR	Y-x-P-Q-Y-Q	34
P2	NSE**YMPQYQI**IFAHQ	Y-x-P-Q-Y-Q	5
P3	YDK**VMPEYQI**KYTTS	[V/L]-x-P-x-Y	100
P4	GPRN**LIPKYQI**QYTY	[V/L]-x-P-x-Y	100
P7	GDFPVM**FMPKYSI**IY	F-x-P-x-Y	1
P8	TMPRD**FTPKYQI**MVK	F-x-P-x-Y	97
P9	TAPF**YPPAYQI**QVFK	P-A-Y-Q-I	2
P10	DRL**FLPAYQI**YPDVV	P-A-Y-Q-I	100
Second Generation Phage Peptides			
Peptide ID	Sequence	Motif family	IC_{50} (μM)
P7 (parent)	GDFPVM**FMPKYSI**IY	F-x-P-x-Y	31
P11	GDFPVM**FMPKYSI**I**F**		6.0
P12	GDFPVM**FMPKYS**W**I**Y		6.4
P1 (parent)	NQ**YQPQYQI**LYAANR	Y-x-P-Q-Y-Q	37
P13	NQ**YQP**R**YQI**LYAANR		>100
P14	NQ**P**Q**PQYQI**LYAANR		>100
P9 (parent)	TAPF**YPPAYQI**QVFK	P-A-Y-Q-I	34
P15	TAPF**YPPA**E**QI**QVFK		>100
P16	TAPF**YPPAYM**I**QVFK		39
P2 (parent)	NSE**YMPQYQI**IFAHQ	Y-x-P-Q-Y-Q	21
P17	N**W**E**YMPQYQI**IFAHQ		5.1
P18	NSE**YMPK**Y**QI**IFAHQ		28
P19	N**W**E**YMPK**Y**QI**IFAHQ		18

family, were synthesized for testing (Table 1). We then assessed whether these 10 peptides could inhibit the glycosyltransferase activity of OGT using a direct fluorescence activity-based OGT assay (43). Using this assay, we first screened these peptides for inhibition of OGT at a concentration of 100 μM and identified four peptides from four families that showed notable inhibition (Table 1 and *SI Appendix*, Fig. S2), while peptide P6 had inadequate solubility for testing. We then generated concentration–response curves for members from these four families and determined their IC_{50} values against OGT, which ranged from 21 to 37 μM (Fig. 2*E*). Interestingly, five peptides showed no significant inhibition toward OGT. Because these sequences were likely OGT ligands, this observation suggested that there are features within these peptides, outside of the motif, that are required for inhibition.

Phage Library Maturation and Evaluation of Second-Generation Peptides. After determining IC_{50} values of these peptides from the initial phage panning and thereby validating the importance of this seven-amino-acid consensus motif, we next sought to build upon these observations by generating a focused library for subsequent phage selection experiments. We therefore constructed a phage library consisting of 11 amino acids centered on the top four inhibitory sequences, with an additional 762 randomized sequences also included to ensure quality control of the library. We first used 11-mer peptides because this length allowed full positional maturation through saturation mutagenesis. Using this focused phage library (FL1), we carried out a round of phage panning against OGT to identify matured peptide sequences having greater affinity for OGT (Fig. 3*A* and *SI Appendix*, Fig. S3). While the enrichment of 11-mers appeared to be weaker than 15-mers overall, we were encouraged to

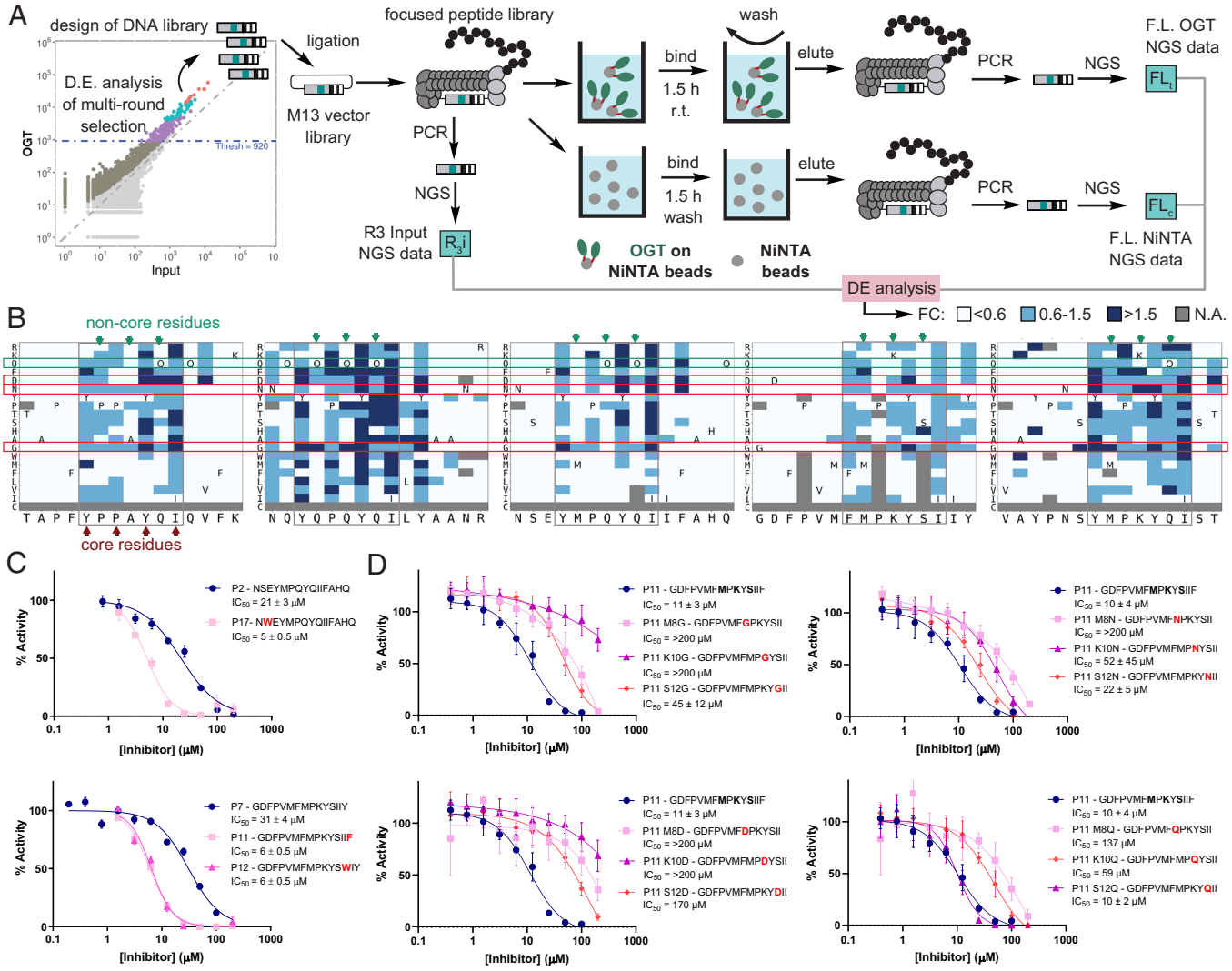

Fig. 3. Dose-dependent inhibition of OGT activity by phage display-derived peptides: (*A*) Focused phage displayed libraries (FL) of peptides were designed using peptide sequences that passed differential enrichment analysis in multiround selection as well as peptides within Hamming distance of 1 (i.e., each amino acid position was replaced by other 18 natural amino acids, excluding Cys). DNA pools encoding all sequences were synthesized, cloned into the phage vector, and transformed to yield two focused phage libraries. FL1 was encoded 11-mers and FL2 encoded 15-mers. The libraries were sequentially panned against OGT protein immobilized on Ni-NTA beads or Ni-NTA beads alone, and the output of the panning was sequenced. (*B*) Heat map of the binned fold change (FC) enrichment of sequences obtained from panning of FL2 for which FC > 1.5 corresponds to binding and FC < 0.6 corresponds to loss of function, as determined by the correlation between FC and IC_{50} values (*SI Appendix*, Fig. S3 *L* and *M*). (*C*) IC_{50} values for three inhibitory peptides derived from the first round (FL1) of saturation mutagenesis phage panning against OGT using 11-mers. Inhibitory activities of peptides having lower IC_{50} values are shown in comparison with the IC_{50} values obtained for the original sequences identified in the first rounds of selection. Mutations arising in the sequences from the FL1 phage selection are in red bolded font. (*D*) IC_{50} values for 12 inhibitory peptides derived from the second round (FL2) of saturation mutagenesis phage panning against OGT using 15-mers. We examined peptides containing G, N, D, or Q amino acid substitutions at each of the three intervening "noncore" positions as indicated in red bolded font. Error bars represent SD of quadruplicate measurements.

find, after sequencing analyses, that four families of peptides were enriched compared to the overall 11-mer population. We ranked these peptides based on their relative fold-enrichment over the parent sequences (*SI Appendix*, Table S1) and, on this basis, prepared 10 additional 15-mer peptides (Table 1), containing overrepresented point mutations. We found three proposed positional substitutions in 15-mer sequences gave rise to peptides with greater inhibitory activity toward OGT than parental 15-mer sequences (Fig. 3 *B* and *C*) with IC_{50} values ranging from 5 to 7 μM, representing a fivefold to sixfold improvement over the corresponding first generation of parental peptides (Table 1). Seven other substitutions had no effect or even weaker inhibitory potency than the corresponding parent 15-mer peptide. Finally, given the greater inhibitor potency of the 15-mer sequences, we generated a second focused phage library (FL2) that expressed each of the top 15-mer from the four inhibitory peptide families as well as corresponding peptides in which each position was fully mutagenized with each amino acid aside from cysteine. Again, an additional 762 randomized sequences were also included to ensure quality control of the library. In this round of screening (Fig. 3*A* and *SI Appendix*, Fig. S3), we focused on the intervening residues between the conserved side chains comprising the binding motif. In particular, we analyzed whether certain amino acid residues were underrepresented in these positions based on the sequencing showing lower levels of these variants. This analysis revealed certain residues appear to compromise peptide binding to OGT (Fig. 3*B* and *SI Appendix*, Fig. S3). To test whether these intervening residues influence binding to OGT, we generated peptides in which we incorporated residues that appear to be disfavored at some of these positions. We examined the effects of substituting G, N, D, and Q at each of the intervening residue sites within the P11 15-mer peptide sequence. OGT inhibition assays using these 12 peptides showed these disfavored residues do indeed impair binding, some in a deleterious manner (Fig. 3*D*). These collective observations, including the improvements in inhibitory potency, suggest that while the consensus motif is critical for binding to OGT, the flanking positions and positions between the conserved residues also play a role in the ability of these peptides to inhibit OGT, with hydrophobic residues enhancing binding and charged or hydrophilic residues within or immediately adjacent to the motif being unfavorable.

To better understand how these peptides were inhibiting OGT, we performed kinetic analyses on two of the most potent affinity matured peptides (P11 and P17) to identify their mode of inhibition (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S4). We accomplished this by varying either the concentration of the fluorescent UDP-GlcNAc donor or the peptide acceptor used in the OGT activity assay in the presence of several concentrations of peptide P17. Surprisingly, global fitting of the data revealed both P11 (Fig. 4 *A* and *B*) and P17 (*SI Appendix*, Fig. S4) inhibited OGT in a noncompetitive manner with respect to the acceptor peptide and donor substrate. The calculated K_I value for P11 inhibition with respect to UDP-GlcN-BODIPY was 2.7 ± 0.3 μM and with respect to Biotin-HCF was 3.8 ± 0.4 μM (R^2 = 0.87 and 0.85). The noncompetitive mode of inhibition, coupled with the concordance in K_I values for P11 inhibition with respect to each substrate, strongly suggested that binding of the peptides occurs outside of the active site of OGT. Given the peculiar mode of inhibition of P11, we decided to determine the K_d of P11 using biolayer interferometry (BLI). Immobilization of biotinylated P11

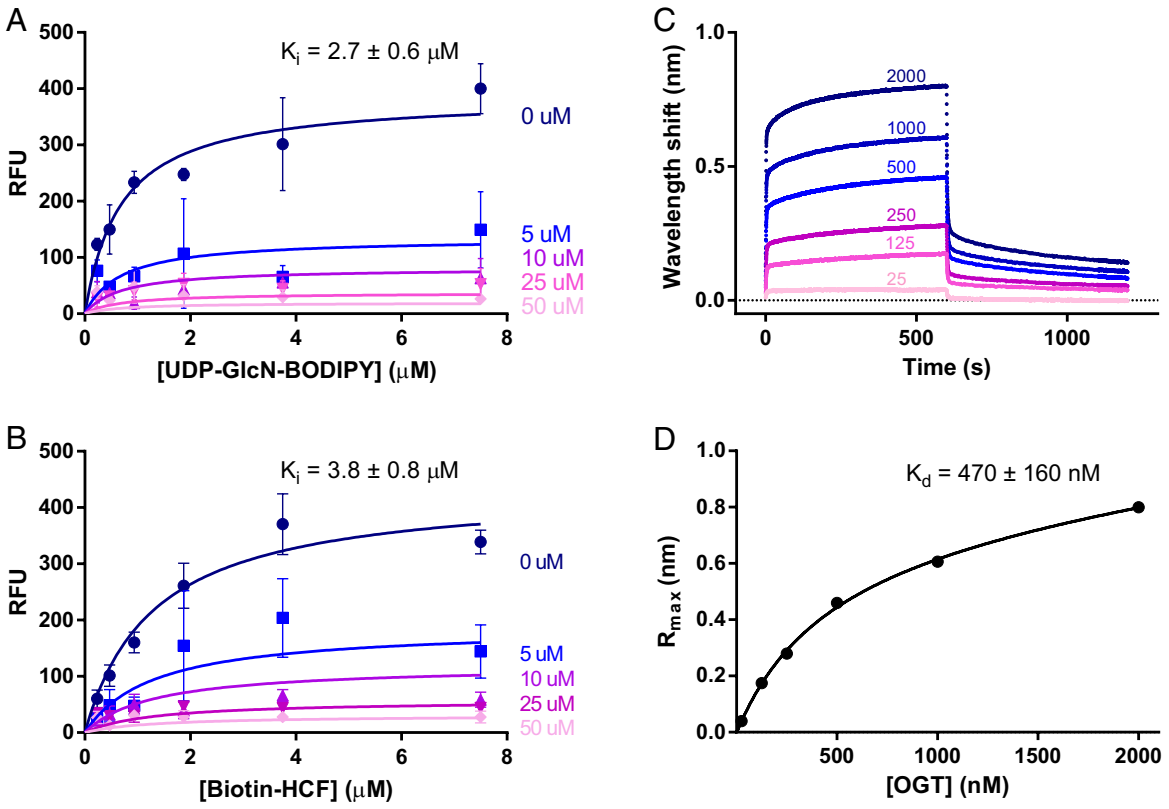

Fig. 4. Inhibition and binding of OGT by peptide P11: (*A*) Inhibition of OGT by P11 at varying concentrations of P11 and UDP-GlcN-BODIPY reveals a mode of noncompetitive inhibition. (*B*) Inhibition of OGT by P11 at varying concentrations of P11 and Biotin-HCF peptide reveals a mode of noncompetitive inhibition. (*C*) BLI data for binding of P11 to OGT determined by immobilization of P11 followed by binding of OGT at different concentrations. (*D*) Fitting to a single binding model to a plot of maximum response (R_{max}) as a function of [OGT] reveals a dissociation constant (K_d) consistent with the K_I value determined by global fitting of the inhibition data.

onto the chip surface followed by monitoring of binding by OGT showed binding reached equilibrium over time (Fig. 4*C*). Plotting these equilibrium positions reflected by the maximum response (R_{max}) and revealed an excellent fit to a single site—no background binding model that reflected high-affinity binding (K_d = 470 ± 160 nM, R^2 = 0.99) to OGT (Fig. 4*D*). This tighter binding, as reflected by the 10-fold lower K_d value, as compared to inhibition, reflected by the K_I value, was unexpected, but we reasoned that it could arise from an allosteric mode of inhibition in which a perturbation arising from binding is incompletely transmitted to the active site. These data collectively support that these peptides are binding to OGT and inducing inhibition through an allosteric mechanism, perhaps by inducing a conformational shift that causes inhibition in a manner akin to that observed for recently described macrocyclic peptide inhibitors of OGT (36).

Structure–Activity Relationships Using Targeted Peptide Truncations and Deletions. To gain a better understanding of the features of these peptides that mediate inhibition of OGT, we proceeded to assess the inhibitory potency of two truncated peptides (tP11 and tP17) that comprised only the 7-residue consensus motif ([Y/F]-x-P-x-Y-x-[I/M/F]) identified from our sequencing results. Strikingly, none of the short peptides that we tested inhibited OGT activity at concentrations of up to 100 μM (*SI Appendix*, Table S2 and Fig. S5). Such results were also in accordance with overall weak binding potency of shorter 11-mer peptides in a maturation screen. Nevertheless, because the phage screening results contained longer, 15-mer peptides with variable residues beyond the sequence motif, we reasoned that inhibition required the presence of amino acid residues pendent to the consensus motif. We therefore hypothesized that these extensions likely somehow drive inhibition. To investigate this idea, we prepared a series of peptide constructs based around the sequence of the P11 and P17 peptides and then tested these as inhibitors of OGT (*SI Appendix*, Table S2 and Fig. S5). As a first step, we enhanced the solubility of P11 to determine whether inhibition was influenced by solubility or aggregation of these peptides. We found that the addition of two lysine residues at either the N or C terminus of the peptides (P21 and P22) had only marginal effects on inhibition, indicating that inhibition was likely mediated by the soluble monomeric peptide. In addition, deletion of three N-terminal residues (P24) also did not significantly perturb inhibition while deletion of all residues N-terminal of the motif drove a reduction in inhibition but did not totally abolish the effect (P25). Deletion of the two C-terminal residues (P23) resulted in complete loss of inhibition at concentrations up to 200 μM. No inhibition from a control peptide (P26) was observed. These data collectively suggested that, in addition to the importance of the motif, residues outside of the motif at the C-terminal end are essential for inhibition of OGT whereas residues N-terminal of the motif regulate the potency of inhibition. Together, these collective data support these peptides authentically binding to OGT and inhibition being sensitive to small changes within the sequence motif, its intervening residues, and the presence of residues C and N-terminal to the motif.

X-ray Structure of OGT in Complex with Truncated Peptides. Based on the unique binding and inhibition properties of these peptides, coupled with their noncompetitive mode of inhibition, we judged that structural insights into their binding to OGT could be illuminating. We therefore set out to determine the X-ray crystallographic structure of OGT in complex with P11. To do so, we used a previously reported, readily crystallizable construct of OGT (OGT4.5) that consists of the catalytic region appended to 4.5 TPRs (19). To aid crystallization, we used a truncated version of the P11 peptide (tP11; FMPKYSI) bearing just the motif of interest. The residues of peptide tP11 are numbered to match P11. Due to limited peptide solubility, we found it necessary to concentrate OGT4.5 in the presence of the tP11 peptide to obtain a soluble OGT-peptide complex. Crystal screening yielded diffraction quality crystals that were processed in the *P*321 spacegroup to 2.8 Å with four copies of OGT comprising the asymmetric unit (*SI Appendix*, Table S3). Most surprisingly, clear F_o-F_c electron density, incompatible with the OGT main chain, was observed across a face of the intervening domains (Int-D) (*SI Appendix*, Fig. S6), suggesting that these linear peptides were not binding to the TPR domains as we had expected. Into this density, it was possible to unambiguously model the tP11 peptide. The tP11 peptide binds across the Int-D of the catalytic region, a region we hereafter refer to as an exosite, lying outside of the active site (Fig. 5*A*). Despite numerous structural insights into OGT, no function has yet been attributed to the Int-D and only OGT enzymes from metazoans possess this domain (9, 21). While the electron density of tP11 is of sufficient quality to model in the peptide, electron density is diffuse for the sidechains of F7 and I13 (Fig. 5*B*). The peptide is bound in place by hydrogen bonding between N791 and the I13 backbone and hydrogen bonding between Y11 and the carbonyl group of L837. In addition, the carbonyl of S833 hydrogen bonds to the backbone nitrogen of K10. (Fig. 5*C*). Further interactions occur along the length of the peptide (Fig. 5*D*). Notably, I13 in the tP11 peptide sequence faces and binds into a hydrophobic pocket formed by F723, I734, and I787 (Fig. 5*D*). Strikingly, the residues making up the exosite that cradles this peptide motif are conserved among mammalian OGT orthologs (Fig. 6 and *SI Appendix*, Fig. S7).

Analysis of the structure of this complex revealed that residues that are not conserved within this peptide motif are solvent exposed, clearly explaining the tolerances at these sites that are seen within the deep sequencing of the pool of selected phage (*SI Appendix*, Table S1). Conserved features of the motif can also be understood in the context of the structure. By constraining geometry, the proline residue (P9) of the peptide induces a kink that adjusts its orientation and permits its flanking residues to bind to the surface of OGT. The phenylalanine residue (F7) packs against OGT to strengthen the overall peptide interaction and may facilitate reorientation of the peptide toward the active site. These two contributions from the F7 residue provide a rationale for the loss of inhibition we observed when this residue was replaced with a proline residue in the P14 peptide. Since peptide tP11 is unable to suppress OGT activity and residues which would lie C-terminal of the consensus motif are not positioned close to the active site, we believe that the disordered residues N-terminal of the consensus motif are likely to be responsible for occluding the active site and promoting inhibition. Accordingly, the consensus motif likely acts as an anchoring point for the peptide, orienting these N-terminal residues, so they can interfere with binding of polypeptide substrates at the active site (Fig. 6*A*). This model is also consistent with the K_d value (470 ± 160 nM) for the P11 peptide being 10-fold lower than the K_I value (4.0 ± 0.9 mM) of P11 since inhibition in this model must be exerted by a subset of the peptide conformations that occlude the active site. We are unaware of other reports of such a steric occlusion model for inhibition of enzymes. However, though clearly different, this mode of inhibition is reminiscent of how binding of autoproteolytically processed peptides allosterically antagonize enzyme activity by binding to exosites present within certain pepsin-family aspartic acid proteases (45).

To gain a sense for the level of conservation of this exosite among OGT orthologs, we leveraged the AlphaFold2 structure

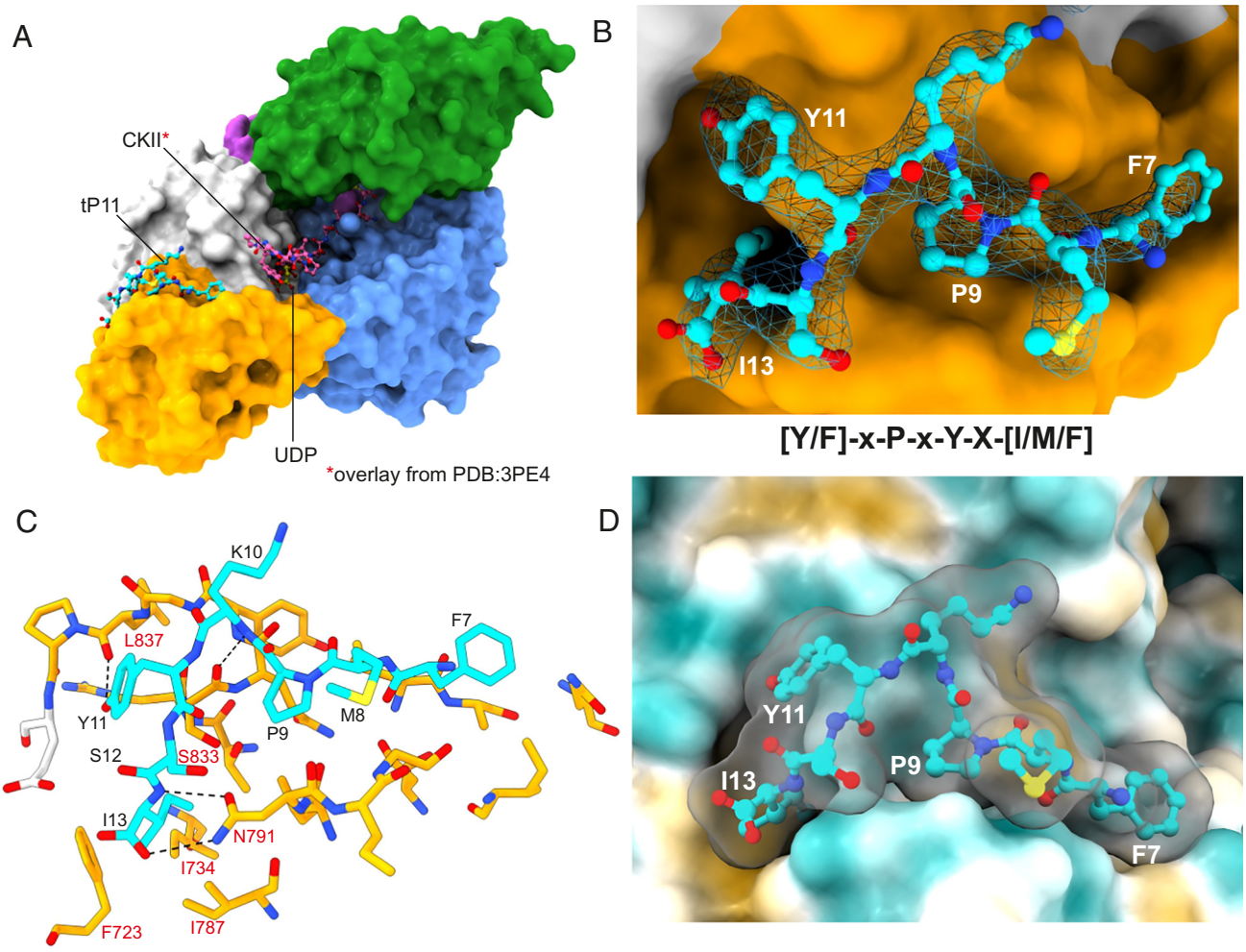

Fig. 5. Structure of OGT in complex with tP11: (*A*), OGT4.5 depicted as a surface with bound tP11 peptide and UDP depicted in ball and sticks. CKII from PDB: 3PE4 is also shown to demonstrate relative position between a peptide acceptor and the bound tP11 peptide. (*B*) Expanded view of the tP11 peptide binding to the surface of OGT, with $2F_o$-F_c electron density contoured to 1 σ localized to the tP11 peptide. The density is incompatible with the OGT main chain and lies across the face of the intervening domain. (*C*) Hydrogen bonding interactions between the tP11 peptide and OGT displayed as dashed lines. Residue numbering in black refers to P11 residues; red numbering highlights OGT residues involved in tP11 binding. (*D*) Surface displaying hydrophobic (brown) and hydrophilic (blue) regions of the surface of OGT. The semitransparent surface of the tP11 peptide is also shown to highlight the complementary shape of the tP11 peptide to the surface of OGT.

prediction database (46) to see whether OGT from model organisms that contain the Int-D domain are likely to have an exosite capable of accommodating a peptide bearing our motif. Unsurprisingly, given the high sequence similarity, the AlphaFold generated mouse OGT exosite is identical to the human, as was that of the AlphaFold OGT structure from zebrafish (*Danio rerio*) (Fig. 6*C*). In comparison the AlphaFold generated structure of *Caenorhabditis elegans* and the crystal structure of *Drosophila melanogaster* have distinct differences in the features of the exosite. An arginine residue in *D. melanogaster* which replaces S833, causes significant steric hindrance to binding of the tyrosine residue of the peptide motif. In addition, residue N791 of human OGT is absent from both models. Furthermore, the hydrophobic pocket used to accommodate the I13 residue of the peptide motif has been disrupted in both organisms. Notably, a growing number of mutations in OGT have been linked to XLID and a survey of these, as recently summarized (47), suggests that two of these mutations (I721V and R792P) localize near the OGT exosite. Structural considerations suggest that either mutation could disrupt peptide binding, with I721V potentially impacting the hydrophobic network which supports I13 (tP11s) and R792P potentially destabilize the C terminus of an α-helix likely causing reorientation of the peptide coordinating residue N791 (*SI Appendix*, Fig. S8). These results suggest that this exosite-mediated binding of protein ligands may be a recent evolutionary feature and impaired binding at this site may be associated with XLID. However, it is also possible that these organisms use their Int-D domain for other purposes yet to be defined or that other peptide sequence motifs are bound by this site in these other organisms.

Taken together, these results define an exosite on the surface of OGT that can bind peptides containing the sequence motif identified in our phage display studies. Given the strong consensus motif of peptide sequences that emerged through the selections, along with the conservation of this exosite within higher vertebrates, coupled with tolerance to considerable variation of the residues at intervening positions, we speculate that native proteins containing this exosite-binding motif are likely present within mammalian proteomes. Moreover, the low micromolar to nanomolar binding and inhibition constants of these exosite-binding peptides are consistent with proteins containing such sequences having biological activity within cellular settings. Given their potential for disruption of the glycosyltransferase activity of OGT

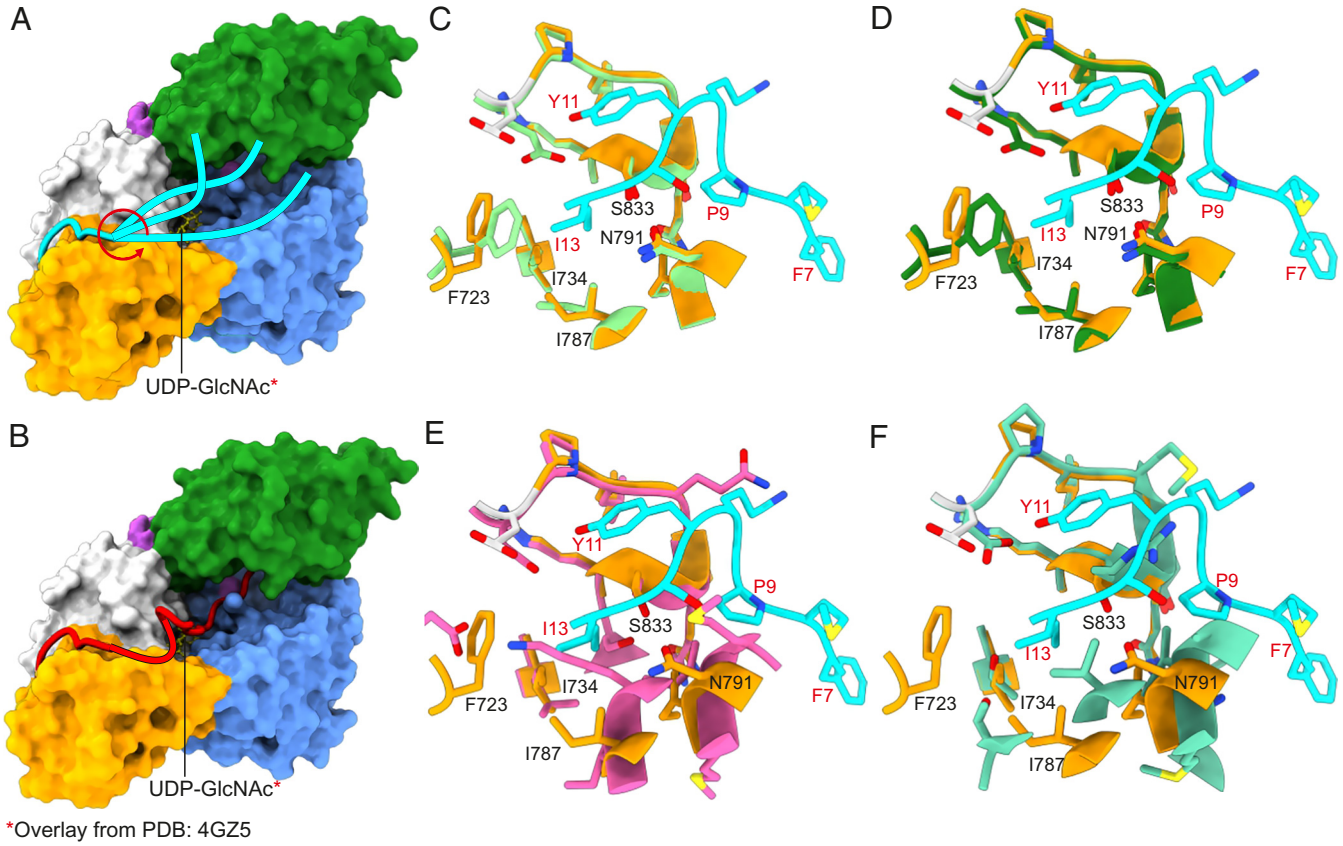

Fig. 6. Mechanism of inhibition of OGT by exosite binding polypeptides and conservation of the OGT exosite: (*A*) A proposed mechanism by which peptides bearing the conserved motif prevent the binding of OGT protein substrates. (*B*) A second proposed mechanism by which the conserved motif in OGT protein substrates enables tighter binding and enhanced O-GlcNAcylation. (*C–F*) Comparisons of the human OGT (*Hs*OGT) structure against structures and AlphaFold models of OGT homologs. The structure of *Hs*OGT superimposed onto AlphaFold generated *Mus Musculus* (*C*), *Danio rerio* (*D*), and *C. elegans* (*E*) OGT models. (*F*) *Hs*OGT superimposed onto the X-ray crystal structure of *D. melanogaster* OGT,PDB: 5A01 (44).

as we have seen here, binding of peptides and proteins to this exosite may have regulatory significance, either as we see here through binding and inhibiting of OGT but, alternatively, by serving to recruit specific protein substrates for modification (Fig. 6*B*), which may then either diffuse away if binding with modest affinity or remain to antagonize OGT activity.

Given these considerations, we examined the prevalence of this consensus motif within the human proteome we conducted a bioinformatic analysis of the human proteome using the MOTIF search tool (48). Using the consensus sequences as a search string ([Y/F]-x-P-x-Y-x-[I/M/F]), we found that 112 human genes encode proteins bearing some variation of this sequence (Dataset S1). Of these, we curated this set to include only those proteins localized to the nucleus or cytoplasm, which are the cellular compartments in which OGT acts, and identified 52 proteins (Dataset S1). Among these 52 proteins are represented protein families of various function, including transcription factors, kinases, deacetylases, and nucleic acid editing enzymes. Notably, only nine of the 52 proteins have been previously identified as substrates of OGT that can be modified with O-GlcNAc (8). We note that the sites of O-GlcNAc modification on these nine proteins are at least 50 amino acid residues away from the motif, suggesting that if binding of the motif within the protein to the OGT exosite serves to promote O-GlcNAcylation modification, it is unlikely to drive modification immediately adjacent to the motif. We expect that some members of this set of 52 proteins are likely to bind to OGT through the identified motif, and some of these may function as regulatory partners. Given the limited number of O-GlcNAcylated proteins observed in this set, we reason that binding of putative partners to this exosite likely mediates recruitment or inhibitory regulation of OGT. However, a number of factors will need to be considered in advancing on this question since the accessibility of the peptide motif within the protein and the secondary structure in which the motif is found are likely to influence binding to OGT. Notably, prior to performing the round of selections using the second focused library (FL2) to perform saturation mutagenesis of the selected 15-mers, we examined whether peptide sequences containing the motif selected from among these 52 proteins might antagonize OGT. We obtained and assayed four 15-mer peptides containing the motif with sequences from repeat 1 of SORB1 (PEENP**Y**F**P**T**Y**K**F**PEL), repeat 2 of SORB1 (DDSDL**YSPRYSF**SED), SOS1 (PTSKA**YSPRYSI**SDR), and SCLY (RGLGE**F**T**PLYPM**LFG). Surprisingly, however, none of these four peptides inhibited OGT, which prompted us to further consider the features influencing binding to the OGT exosite. Notably, phage panning using the second saturation mutagenesis library indicated that certain residues at intervening—and adjacent—residues around the motif are disfavored, but this depends, to some extent, on the parent sequence (Fig. 3*B*). We find that generally disfavored amino acid residues are present in all four peptides from the proteins we selected. Accordingly, these data collectively suggest that some proteins may be preferred based on the sequence containing the motif; however, given that binding at more distant sites may enhance binding, the data pointing to disfavored intervening residues may serve to guide future efforts but should likely not be used to definitively exclude possible candidate exosite

interacting proteins. In addition, the specific amino acid composition adjacent to the motif may contribute to inhibition. Accordingly, experiments to investigate in greater detail these putative interaction partners and the amino acid composition driving inhibition are warranted.

Discussion

The *O*-GlcNAc post-translational modification has been firmly established as a key regulator of several signaling pathways that govern important cellular processes. However, the molecular mechanisms that control the glycosyltransferase activity and substrate specificity of OGT have been slow to emerge. Indeed, the large number of *O*-GlcNAcylated targets in the human proteome raises questions about how OGT activity is regulated and how this enzyme identifies and discriminates between its great number of protein substrates. Attempts have been made to identify a generalized consensus motif surrounding serine or threonine O-GlcNAcylation sites (8, 41, 42). Although a clear consensus motif likely does not exist, certain preferences in primary peptide sequence have been identified and these are consistent with the structural requirements of the active site of OGT (42). The roles of the TPR region in recruiting substrates to OGT for glycosylation are gradually emerging (22, 23, 49). Other regulatory mechanisms are also being defined (20, 26) including the general concept that binding partners of OGT may influence substrate recruitment (21, 24, 50). The molecular details of these interactions, however, remain unknown. Binding to the TPR region is one logical potential means of controlling substrate specificity. However, it is notable that panning of our diverse phage library did not return sequences binding to the TPR region and instead converged to deliver a sequence motif that binds to a defined exosite on the surface of the Int-D of OGT that is proximal to the active site. Notably, the amino acid residues that define the topological features of this exosite are strikingly conserved among mammals (*SI Appendix*, Fig. S7), suggesting that it likely fulfills important evolutionarily conserved roles in modulating OGT function. While this exosite may serve to recruit OGT substrates, we find it is more likely that this site serves a regulatory role by recruiting OGT polypeptides, including proteins, that contain this motif such that their binding occludes the active site of OGT and prevents its subsequent binding of protein substrates. With respect to the linear peptides identified here, their low to submicromolar binding affinity suggests that these peptides could be purposed in useful ways. For example, by inserting or fusing this motif to proteins one may be able to control or direct OGT activity within cells, which has emerged as a topic of some interest (51, 52). Furthermore, chemical refinement of these structures may also offer up chemical tools to control OGT activity in an allosteric manner as well as to interrogate the biological effects of antagonizing this exosite within cells. Perhaps most immediately salient is that the exosite that we have uncovered is likely to have biological significance by functioning as a regulatory site for OGT at which binding of proteins and polypeptides containing this sequence motif act to modulate its activity. Although the short peptides identified in this study display fair but not exceptional binding affinity toward OGT, it is likely that larger proteins in cells that are capable of binding to this site may do so with significantly higher potency by virtue of their greater steric bulk, which would more effectively occlude the active site. Moreover, full-length proteins may have features enabling additional contacts within the active site or elsewhere on the surface of OGT. Notably, a comparative analysis of the 52 proteins containing this motif with OGT interacting proteins identified in proteomic screening experiments shows that six of these 52 proteins have been found to interact with OGT (Dataset S1) (28, 53, 54). Intriguingly, the conservation of tyrosine within this motif suggests a molecular mechanism for relief of inhibition of OGT. The sequence of this motif is compatible with tyrosine kinases being able to induce phosphorylation of this tyrosine residue (55) which would abrogate binding of potential exosite-binding protein antagonists of OGT as the phosphate would prohibit binding to the exosite of OGT. Analysis of the 52 candidate proteins strikingly revealed that five of candidate proteins have been found to be tyrosine phosphorylated within this conserved motif (Dataset S1) (56). Studies directed toward identifying protein regulatory factors and their potential control by tyrosine kinases are likely to prove a promising line of inquiry into an unusual mechanism for regulation of OGT activity within cells.

Notably, following initial identification of this motif by phage display (34) and during the course of preparation of this manuscript, a parallel independent report detailing related research appeared (57). That work from the Jiang group describes a closely related motif along with biophysical and structural data relating to its binding to the same site on the Int-D domain of OGT and is broadly consistent with our findings. Their data also support tyrosine phosphorylation of this motif-blocking peptide binding to OGT. While that work proposes this exosite is principally for recruitment of substrates, whereas our findings are interpreted as this site acting in an inhibitory capacity, the convergence of these two separate lines of research and similarity of core findings provides strong support for the conclusion that this exosite is most likely to play important roles in the regulation of OGT.

Materials and Methods

Selection of Phage Peptides Bound to OGT. Four rounds of panning were performed using the X15 library and histidine-tagged O-GlcNAc transferase (OGT) as a target. Manipulations of solutions and magnetic beads were performed manually in the first round; for subsequent rounds, a Kingfisher Duo Prime magnetic bead purification system was used (Thermo Fisher). The X15 library was depleted for bead-binding phage by incubation with Dynabeads (Thermo Fisher) before panning. All bead incubations were set up in 1.7-mL Eppendorf tubes on a rotator. One hundred microliters of beads were washed in HN buffer (50 mM HEPES and 500 mM NaCl, pH 7.4), resuspended in a total volume of 100 μL of HN containing X15 library (1.4^{11} PFU/mL), and incubated at 4 °C overnight. The depleted library supernatant was then separated from magnetically immobilized beads and saved for panning. OGT beads were prepared by incubating 10 μg of OGT protein with 20 μL of HN-washed Dynabeads in total volume of 100 μL HN at 4 °C overnight. OGT beads were blocked with 2% bovine serum albumin (BSA) in HN at room temperature for 1 h. Four replicate library panning tubes were set up, each containing depleted X15 library (10^{11} PFU/mL) with blocked OGT beads in 1 mL of 2% BSA/HN. The tubes were incubated at room temperature for 1.5 h. After incubation, the OGT beads were immobilized by magnet, and the supernatant was removed. The OGT beads with bound phage were washed three times with 0.1% Tween/HN. Phage were released by incubating beads in 200 μL of 0.2M glycine–HCl pH 2.2 on a rotator for 9 min. The supernatant containing released phage was removed from the beads and neutralized with 30 μL of 1M Tris–HCl pH 9.1. This pool of released phage was separated into three aliquots: i) 200 μL of the pool was added to 25 mL of early log phase *Escherichia coli* TG1 pRLA4 and amplified in a shaking incubator at 37 °C for 4.5 h; ii) 20 μL was used as a source of DNA for next-generation sequencing; and iii) the remaining phage was quantified by plaque assay. Phage amplification and subsequent polyethylene glycol (PEG) precipitation were performed as described (58). PEG precipitated phage was resuspended, titered, and used as the library for the next round of panning. For rounds 2, 3, and 4 of panning, beads were washed 1, 3, and 3 times, respectively, using a Kingfisher instrument. For rounds 3 and 4 of panning, parallel negative panning controls were performed using depleted X15 libraries and beads containing no target protein. Focus library panning was performed

twice, once under the same conditions as round 3 of regular panning without a bead depletion step, and the second time with an additional three washes, both with the addition of parallel negative panning controls.

BLI. BLI experiments were performed using an Octet RED96e instrument (FortéBio, Sartorius) and 96-well black microplate (part no. 655209, Greiner). The binding assay was carried out at 25°C in assay buffer 50 mM Hepes NaOH pH 7.4, 500 mM NaCl, 1 mM TCEP, 0.1 % (w/v) BSA, 0.02 % (v/v) Tween-20, and 5 % (v/v) DMSO with sample plate shaking speed 1000 rpm. The Octet® streptavidin (SA) dip and read biosensor (part no 18-5019, Sartorius) was equilibrated in the assay buffer for 120 s. Then, the ligand, biotinylated peptide P11 (sequence: GDFPVMFMPKYSIIF with N-terminal modification Biotin-Ahx), at a concentration 0.85 μg mL^{-1} (400 nM) was immobilized on the biosensor for 600 s. The biosensor was then washed in the assay buffer for 120 s and dipped in the sample solution, OGT4.5 dimer, at a concentration range of 25, 125, 250, 500, 1,000, and 2,000 nM. Samples at a concentration of 2,000 nM or buffer alone 0 nM were added in the rest of the wells to check for nonspecific binding and apply correction for baseline drift. Following the association step, dissociation was performed for 600 s. The data correction and analysis were carried out using Octet Data Analysis High Throughput (HT) software version 12. The rate constants were determined by global fit type using all the six sample concentrations.

X-ray Crystallography. A reaction of 52.5 μM OGT4.5 with 1050 μM UDP in size-exclusion buffer was incubated for 1 h at 4 °C. The combined OGT4.5/UDP mix was then diluted 28.3 times in size-exclusion buffer containing tP11 at 57 μM to yield a final reaction mix of 1.85 μM protein, 37 μM UDP, and 55 μM tP11. The OGT4.5/UDP/tP11 mixture was incubated together for 4 h before being concentrated to ~8 mg/mL using a 3000 MWCO Vivaspin concentrator (protein concentration determined via A_{280} using the extinction coefficient and molecular weight of OGT4.5 only). The best crystals obtained were in 1.5 M potassium phosphate dibasic, 5% xylitol, and 10 mM EDTA. Crystal drops were set up at a 2:1 ratio. Crystals were cryoprotected in mother liquor supplemented with 27% xylitol and flash-cooled. Diffraction data were collected at Diamond Light Source in Oxford, UK on beamline I03. Data reduction and processing was completed through *xia2* using *DIALS* and *AIMLESS* (59, 60) A structure solution was obtained through *Phaser* using chain A of PDB 3PE3 (19) as a search model. Simulated annealing was run on the structure solution through *Phenix* before cycles of refinement and interactive model building were undertaken using *REFMAC5* and *Coot* (61–64). Residues have been numbered relative to OGT isoform 3. Figures were prepared through UCSF ChimeraX (65).

Data, Materials, and Software Availability. X-ray structural data; raw sequencing data (17 files) have been deposited in PDB; 48HD.cloud (8CM9; https://48hd.cloud/file/2852; https://48hd.cloud/file/2883; https://cloud/file/2862; https://48hd.cloud/file/2818; https://48hd.cloud/file/2820; https://48hd.cloud/file/2819; https://48hd.cloud/file/2821; https://48hd.cloud/file/2825; https://48hd.cloud/file/2822; https://48hd.cloud/file/2823; https://48hd.cloud/file/2824; https://48hd.cloud/file/2826; https://48hd.cloud/file/2829; https://48hd.cloud/file/2827; https://48hd.cloud/file/2828; https://48hd.cloud/file/6074; https://48hd.cloud/file/6073). All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. We are grateful for support from GlycoNet, the Canadian Glycomics Network (CD-1), the Canadian Institutes of Health Research (PJT-156202), and the Natural Sciences and Engineering Council of Canada (RGPIN-05426). We thank Diamond Light Source for access to beamline I03 (proposal number mx24948). G.J.D. thanks the Royal Society for the Ken Murray Research Professorship and R.W.M. for the associated PDRA funding (RP\EA\180016). We also wish to acknowledge Dr. Johan Turkenburg and Sam Hart for helping coordinate data collection. D.J.V. thanks the Canada Research Chairs program for support as a Tier I Canada Research Chair in Chemical Biology. R.W.M. thanks the University of Southampton for support with an Anniversary Fellowship. We also thank the Centre for High-Throughput Chemical Biology (HTCB) for access to core facilities.

Author affiliations: [a]Department of Chemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; [b]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, United Kingdom; [c]School of Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom; [d]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; [e]48 Hour Discovery, Nanotechnology Research Centre, Edmonton, AB T6G 2M9, Canada; and [f]Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada

1. K. W. Moremen, M. Tiemeyer, V. A. Nairn, Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).
2. T. Carmen-Rosa, G. W. Hart, Topography and polypeptide distribution of terminal N-Acetylglucosamine residues on the surfaces of intact lymphocytes. *J. Biol. Chem.* **259**, 3308–3317 (1984).
3. X. Yang, K. Qian, Protein O-GlcNAcylation: Emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **18**, 452–465 (2017).
4. M. P. Parker, K. R. Peterson, C. Slawson, O-GlcNAcylation and O-GlcNAc cycling regulate gene transcription: Emerging roles in cancer. *Cancers* **13**, 1666 (2021).
5. S. Brimble, E. Wollaston-Hayden, C. Teo, A. Morris, L. Wells, The role of the O-GlcNAc modification in regulating eukaryotic gene expression. *Curr. Sig. Trans. Therapy* **5**, 12–24 (2010).
6. H. B. Ruan, Y. Nie, X. Yang, Regulation of protein degradation by O-GlcNAcylation: Crosstalk with ubiquitination. *Mol. Cell Proteom.* **12**, 3489–3497 (2013).
7. K. M. M. Fahie, K. N. Papanicolaou, N. E. Zachara, Integration of O-GlcNAc into stress response pathways. *Cells* **11**, 3509 (2022).
8. E. Wulff-Fuentes *et al.*, The human O-GlcNAcome database and meta-analysis. *Sci. Data* **8**, 25 (2021).
9. C. M. Joiner, H. Li, J. Jiang, S. Walker, Structural characterization of the O-GlcNAc cycling enzymes: Insights into substrate recognition and catalytic mechanisms. *Curr. Opin. Struct. Biol.* **56**, 97–106 (2019).
10. A. A. Elbatrawy, E. J. Kim, G. Nam, O-GlcNAcase: Emerging Mechanism, substrate recognition and small-molecule inhibitors. *ChemMedChem* **15**, 1244–1257 (2020).
11. A. P. Willems *et al.*, Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. *J. Biol. Chem.* **292**, 12621–12631 (2017).
12. K. Vaidyanathan *et al.*, Identification and characterization of a missense mutation in the O-linked beta-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J. Biol. Chem.* **292**, 8948–8963 (2017).
13. A. T. Balana, M. R. Pratt, Mechanistic roles for altered O-GlcNAcylation in neurodegenerative disorders. *Biochem. J.* **478**, 2733–2758 (2021).
14. S. A. Yuzwa *et al.*, Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **8**, 393–399 (2012).
15. L. Ciraku, E. M. Esquea, M. J. Reginato, O-GlcNAcylation regulation of cellular signaling in cancer. *Cell Signal* **90**, 110201 (2022).
16. C. Slawson, G. W. Hart, O-GlcNAc signalling: Implications for cancer cell biology. *Nat. Rev. Cancer* **11**, 678–684 (2011).
17. C. Fehl, J. A. Hanover, Tools, tactics and objectives to interrogate cellular roles of O-GlcNAc in disease. *Nat. Chem. Biol.* **18**, 8–17 (2022).
18. M. Jínek *et al.*, The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004).
19. M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz, S. Walker, Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–569 (2011).
20. R. W. Meek *et al.*, Cryo-EM structure provides insights into the dimer arrangement of the O-linked beta-N-acetylglucosamine transferase OGT. *Nat. Commun.* **12**, 6508 (2021).
21. D. T. King, A. Males, G. J. Davies, D. J. Vocadlo, Molecular mechanisms regulating O-linked N-acetylglucosamine (O-GlcNAc)–processing enzymes. *Curr. Opin. Chem. Biol.* **53**, 131–144 (2019).
22. C. M. Joiner, Z. G. Levine, C. Aonbangkhen, C. M. Woo, S. Walker, Aspartate residues far from the active site drive O-GlcNAc transferase substrate selection. *J. Am. Chem. Soc.* **141**, 12974–12978 (2019).
23. Z. G. Levine *et al.*, O -GlcNAc transferase recognizes protein substrates using an asparagine ladder in the tetratricopeptide repeat (TPR) superhelix. *J. Am. Chem. Soc.* **140**, 3510–3513 (2018).
24. H. M. Stephen, T. M. Adams, L. Wells, Regulating the regulators: Mechanisms of substrate selection of the O-GlcNAc cycling enzymes OGT and OGA. *Glycobiology* **31**, 724–733 (2021), 10.1093/glycob/cwab005.
25. J. Biwi, C. Biot, Y. Guerardel, A. S. Vercoutter-Edouart, T. Lefebvre, The many ways by which O-GlcNAcylation may orchestrate the diversity of complex glycosylations. *Molecules* **23**, 2858 (2018).
26. H. Gao *et al.*, Mutual regulation mechanism of the O-GlcNAcylation enzyme pair revealed by Cryo-EM structure of human OGT–OGA complex. Research Square [Preprint] (2023). https://doi.org/10.21203/rs.3.rs-2275302/v1 (Accessed 20 January 2023).
27. M. Martinez *et al.*, Quantitative proteomics reveals that the OGT interactome is remodeled in response to oxidative stress. *Mol. Cell Proteom.* **20**, 100069 (2021).
28. H. M. Stephen, J. L. Praissman, L. Wells, Generation of an interactome for the tetratricopeptide repeat domain of O-GlcNAc transferase indicates a role for the enzyme in intellectual disability. *J. Proteome Res.* **20**, 1229–1242 (2021).
29. M. G. Alteen, H. Y. Tan, D. J. Vocadlo, Monitoring and modulating O-GlcNAcylation: Assays and inhibitors of O-GlcNAc processing enzymes. *Curr. Opin. Struct. Biol.* **68**, 157–165 (2021).
30. R. Trapannone, K. Rafie, D. M. F. Van Aalten, O-GlcNAc transferase inhibitors: Current tools and future challenges. *Biochem. Soc. Trans.* **44**, 88–93 (2016).
31. B. J. Gross, J. G. Swoboda, S. Walker, A strategy to discover inhibitors of O-linked glycosylation. *J. Am. Chem. Soc.* **130**, 440–441 (2008).
32. B. J. Gross, B. C. Kraybill, S. Walker, Discovery of O-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **127**, 14588–14589 (2005).
33. S. E. S. Martin *et al.*, Structure-based evolution of low nanomolar O-GlcNAc transferase inhibitors. *J. Am. Chem. Soc.* **140**, 13542–13545 (2018).
34. M. G. Alteen, "Development of chemical tools for the study of intracellular glycosylation," PhD thesis, Simon Fraser University, Burnaby, BC, Canada (2020), p. 258. etd21031, https://summit.sfu.ca/_flysystem/fedora/2022-08/input_data/21444/etd21031.pdf.

35. T. Passioura, H. Suga, A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* **53**, 1931–1940 (2017).
36. M. G. Alteen *et al.*, Potent de novo macrocyclic peptides that inhibit O-GlcNAc transferase through an allosteric mechanism. *Angew. Chem. Int. Ed. Engl.* **62**, e202215671 (2023).
37. L. Wang *et al.*, Therapeutic peptides: Current applications and future directions. *Signal Trans.* **7**, 48 (2022).
38. M. Muttenthaler, G. F. King, D. J. Adams, P. F. Alewood, Trends in peptide drug discovery. *Nat. Rev. Drug. Disc.* **20**, 309–325 (2021).
39. L. R. H. Krumpe, K. M. Schumacher, J. B. McMahon, L. Makowski, T. Mori, Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library. *BMC Biotechnol.* **7**, 65 (2007).
40. W. L. Matochko, R. Derda, Next-generation sequencing of phage-displayed peptide libraries. *Meth Mol. Biol.* **1248**, 249–266 (2015).
41. R. J. Chalkley, A. Thalhammer, R. Schoepfer, A. L. Burlingame, Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8894–8899 (2009).
42. S. Pathak *et al.*, The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat. Struct. Mol. Biol.* **22**, 744–749 (2015).
43. M. G. Alteen *et al.*, A direct fluorescent activity assay for glycosyltransferases enables convenient high-throughput screening: Application to O -GlcNAc transferase. *Angew. Chem. Int. Ed. Engl.* **132**, 9688–9696 (2020).
44. D. Mariappa *et al.*, Dual functionality of O-GlcNAc transferase is required for Drosophila development. *Open Biol.* **5**, 150234 (2015).
45. I. Hanova *et al.*, Novel structural mechanism of allosteric regulation of aspartic peptidases via an evolutionarily conserved exosite. *Cell. Chem. Biol.* **25**, 318–329.e314 (2018).
46. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. M. Omelková *et al.*, An O-GlcNAc transferase pathogenic variant linked to intellectual disability affects pluripotent stem cell self-renewal. *Dis. Model Mech.* **16**, dmm049132 (2023).
48. M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, The KEGG databases at GenomeNet. *Nucleic Acids Res.* **30**, 42–46 (2002).
49. C. M. Joiner, F. A. Hammel, J. Janetzko, S. Walker, Protein substrates engage the lumen of O-GlcNAc transferase's tetratricopeptide repeat domain in different ways. *Biochemistry* **60**, 847–853 (2021).
50. A. Berthier *et al.*, Combinatorial regulation of hepatic cytoplasmic signaling and nuclear transcriptional events by the OGT/REV-ERBalpha complex. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11033–E11042 (2018).
51. D. H. Ramirez *et al.*, Engineering a proximity-directed O-GlcNAc transferase for selective protein O-GlcNAcylation in cells. *ACS Chem. Biol.* **15**, 1059–1066 (2020).
52. Y. Zhu, G. W. Hart, Dual-specificity RNA aptamers enable manipulation of target-specific O-GlcNAcylation and unveil functions of O-GlcNAc on beta-catenin. *Cell* **186**, 428–445.e427 (2023).
53. J. Gao *et al.*, Proteomic analysis of the OGT interactome: Novel links to epithelial–mesenchymal transition and metastasis of cervical cancer. *Carcinogenesis* **39**, 1222–1234 (2018).
54. M. E. Griffen *et al.*, Functional glycoproteomics by integrated network assembly and partitioning. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.06.13.541482 (Accessed 26 June 2023).
55. A. Li, R. Voleti, M. Lee, D. Gagoski, N. H. Shah, High-throughput profiling of sequence recognition by tyrosine kinases and SH2 domains using bacterial peptide display. *eLife* **12**, e82345 (2023).
56. P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham, Skrzypek E PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
57. C. Blankenship *et al.*, Motif-dependent binding on the intervening domain regulates O-GlcNAc transferase. *Nat. Chem. Biol.*, 10.1038/s41589-023-01422-2 (2023).
58. S. Ng, K. F. Tjhung, B. M. Paschal, C. J. Noren, R. Derda, Chemical posttranslational modification of phage- displayed peptides. *Meth. Mol. Biol.* **1248**, 155–172 (2015).
59. G. Winter *et al.*, DIALS: Implementation and evaluation of a new integration package. *Acta Cryst. D.* **74**, 85–97 (2018).
60. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Cryst. D.* **69**, 1204–1214 (2013).
61. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D.* **67**, 355–367 (2011).
62. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Cryst. D.* **66**, 486–501 (2010).
63. A. J. McCoy *et al.*, Phaser crystallographic software. *J. App. Crystall* **40**, 658–674 (2007).
64. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Cryst. D* **75**, 861–877 (2019).
65. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Prot. Sci.* **30**, 70–82 (2021).