nature structural & molecular biology

Structure of an *O*-GlcNAc transferase homolog provides insight into intracellular glycosylation

Carlos Martinez-Fleites[1], Matthew S Macauley[2], Yuan He[1], David L Shen[2], David J Vocadlo[2] & Gideon J Davies[1]

***N*-Acetylglucosamine (*O*-GlcNAc) modification of proteins provides a mechanism for the control of diverse cellular processes through a dynamic interplay with phosphorylation. UDP-GlcNAc:polypeptidyl transferase (OGT) catalyzes *O*-GlcNAc addition. The structure of an intact OGT homolog and kinetic analysis of human OGT variants reveal a contiguous superhelical groove that directs substrates to the active site.**

O-GlcNAc modification of the serine and threonine residues of nucleocytoplasmic proteins is believed to be a crucial intracellular signaling event. Unlike glycosylation within the secretory pathway, *O*-GlcNAc is a dynamic modification that is processed more rapidly than the polypeptides it decorates[1,2]. Thus, *O*-GlcNAc is akin to phosphorylation, and there is an intriguing reciprocal relationship between these modifications (**Supplementary Fig. 1** online). Accordingly, *O*-GlcNAc has been implicated in various signaling pathways[1,3]. Two enzymes are responsible for *O*-GlcNAc cycling in *Homo sapiens*, but the molecular mechanisms regulating these *O*-GlcNAc–processing enzymes are unclear.

The glycosyltransferase OGT catalyzes *O*-GlcNAc addition to proteins[4]. The absence of structural insights into OGT catalysis has slowed efforts both to engineer the enzyme[5,6] and to develop inhibitors[7]. There is also a confounding lack of an apparent consensus sequence directing which proteins are glycosylated[1]. OGT seems to recognize protein substrates through interactions with its N-terminal region, which comprises a series of tetratricopeptide repeat (TPR) domains[5,8]. Highlighting the role of the TPR domain, these previous studies have found that OGT constructs with different numbers of TPRs have different specificities for protein substrates, and constructs lacking the majority of TPRs are inactive against protein substrates but do modify peptide substrates[8]. The TPR region of human OGT (HsOGT) forms a superhelical structure[9] that is common to other TPR-containing proteins. How the topographical arrangement of the C-terminal glycosyltransferase domain and its N-terminal TPR region enables protein modification is unclear. Here we present the full-length structure of a homolog of OGT from the plant pathogen *Xanthomonas campestris* that comprises the catalytic and TPR domains (**Fig. 1**; see **Supplementary Methods** for experimental details). The catalytic center of XcOGT shows high similarity to eukaryotic OGTs (**Supplementary Fig. 2** online).

XcOGT is one of more than 130 TPR-containing bacterial OGT GT41 homologs, all of whose substrates are undefined. Like OGT, XcOGT catalyzes the transfer of UDP-GlcNAc to water *in vitro* (**Supplementary Fig. 3** online), a reaction that is inhibited by excess UDP, suggesting that XcOGT uses UDP-GlcNAc or a structurally similar donor sugar substrate. The three-dimensional structure of XcOGT, solved using single-wavelength anomalous dispersion methods and refined at 2.75 Å using data from the UDP complex of the enzyme (**Supplementary Table 1** online), enabled us to generate a model of HsOGT. The structure of XcOGT in complex with the UDP product reveals that residues highly conserved among eukaryotes are involved in recognition of the UDP-GlcNAc donor (**Fig. 1**). The organization of XcOGT (**Fig. 1** and **Supplementary Movie 1** online)

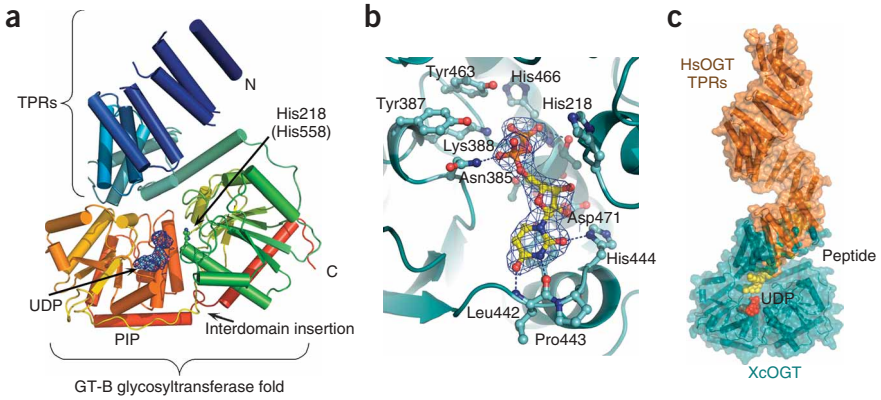

Figure 1 The three-dimensional structure of XcOGT. (**a**) Schematic cartoon of XcOGT. UDP is shown in ball-and-stick representation with $2F_o - F_c$ electron density at 1σ. The catalytic base, His218 (HsOGT His558) is shown, along with the predicted location of the putative HsOGT phosphatidylinositol 3,4,5-trisphosphate (PIP) interaction site and the interdomain insertion. (**b**) The XcOGT active center and interactions of UDP. (**c**) Composite of XcOGT (cyan, with UDP in red) overlaid with the HsOGT TPR domains[9] (gold) with a peptide ligand from the yeast importin complex[9,12]

[1]Structural Biology Laboratory, Department of Chemistry, The University of York, Heslington, York YO10 5YW, UK. [2]Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. Correspondence should be addressed to G.J.D. (davies@ysbl.york.ac.uk) or D.J.V. (dvocadlo@sfu.ca).

Received 29 February; accepted 9 May; published online 8 June 2008; doi:10.1038/nsmb.1443

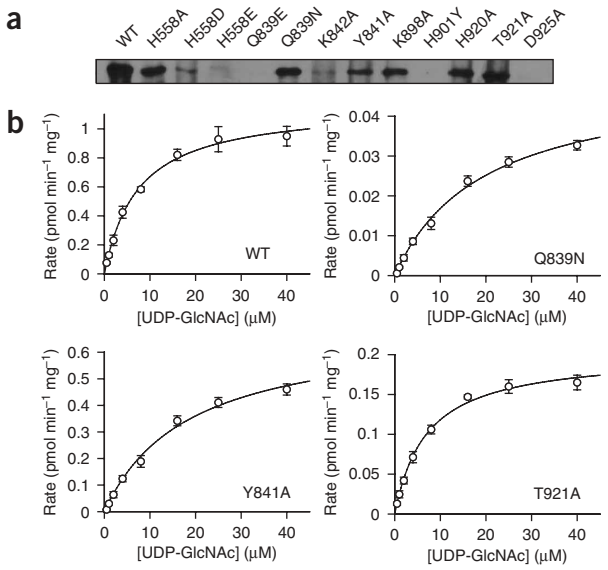

Figure 2 Kinetics of human OGT and its active center variants. (**a**) Western blot showing expression levels of OGT variants compared to wild type (WT). (**b**) Michaelis-Menten kinetics (s.d. shown) for WT and selected mutant OGTs (**Supplementary Table 2** and **Supplementary Methods** online).

parallels that of HsOGT, containing 5.5 TPRs (residues 1–202) and a C-terminal glycosyltransferase moiety (residues 203–568) that has the GT-B topology (one of two folds found in nucleotide-sugar glycosyltransferases[10]). Notably, XcOGT lacks a 120-residue insertion between the N- and C-terminal domains of the GT-B fold in mammalian OGTs; this insertion may thus mediate HsOGT-specific interactions. A putative phosphatidylinositol trisphosphate binding site on the human enzyme[11] maps to the C terminus of XcOGT (**Supplementary Fig. 2**).

On the basis of the XcOGT UDP complex (**Supplementary Fig. 4**) and the HsOGT model, we predicted active-site mutants of HsOGT and assayed the enzyme activity of soluble variants (**Fig. 2**; **Supplementary Table 2**). Mutants of Lys842 (HsOGT numbering) or Gln839, which interact with the pyrophosphate, showed low specific activities of ≈1%, whereas the Thr921 mutant showed 18% of the specific activity. Mutation of His920 was deleterious, whereas mutation of Tyr841 (both in the putative donor binding site) lowered specific activity to 24%. Mutation of Lys898, which is involved in uracil binding, resulted in a protein with no apparent activity. The XcOGT structure also suggests that His558 is likely to be the catalytic general base that facilitates nucleophilic attack of the serine/threonine hydroxyl group on UDP-GlcNAc, on the basis of its orientation and three-dimensional position within the HsOGT model relative to histidine general bases in other GT-B enzymes[10]. His558 mutations showed no apparent activity.

In addition to these insights into the catalytic center of OGT, the XcOGT structure reveals its spatial relationship to the substrate-binding TPRs. The XcOGT TPRs are arranged around a central axis, making a single turn of approximately 260° that is approximately 32 Å wide and 45 Å long, forming a near complete superhelical turn with similar dimensions to the equivalent region of HsOGT[9]. Notably, the last TPR domain of XcOGT is atypical; helix A is shorter and helix B longer than canonical TPR helices. These two helices, together with helix A of the third TPR domain, form an intimate interface (∼1,500 Å^2) with the GT domain. Sequence and structural comparisons reveal no conserved patterns of surface-exposed residues between XcOGT TPR domains and other OGT homologs, although, similarly to the HsOGT TPR complex, several asparagine, arginine and glutamine residues line the inner surface of the superhelix. Overlays of the XcOGT and the human TPR domains (**Fig. 1c**) indicate that the overall length of the composite superhelix is >130 Å, which may explain how OGT accommodates myriad protein partners. The dimensions of the groove, with a width of ∼17 Å and a depth of ∼25 Å, suggest that HsOGT may recognize both linear peptide substrates and secondary-structural elements and explain the lack of an apparent consensus sequence governing which proteins are *O*-GlcNAc modified[1,5,8]. Furthermore, the remarkable length and broad, concave inner face of the superhelical groove of HsOGT provides a large surface area to which different protein partners may bind. Substrates need not bind along the entire length of the groove; indeed, the specific binding of protein substrates could occur anywhere along the superhelix. Binding of the substrate to the TPR domain may simply serve to direct the region bearing the target residue to the active site and orient the acceptor hydroxyl group within the active site. The efficiency with which substrates are modified with *O*-GlcNAc may therefore be governed by their affinity for the TPR domains, their affinity to other protein partners, as well as their propensity to be bound within the active site in a productive orientation. Notably, the atypical TPR domains mediating the tight interface orient the catalytic domain and enable the superhelical groove of the TPR domains to continue smoothly into the catalytic machinery. Thus, binding in the TPR twist may allow polypeptide substrates to extend on into the active-site cleft where catalysis occurs. The three-dimensional structure also shows how OGT may be regulated by other proteins having affinity for the TPR domain. Binding of partners along the TPR provides a mechanism to prevent, or enhance, the binding of cognate proteins for *O*-GlcNAc modification.

Accession codes. Protein Data Bank: Coordinates for XcOGT have been deposited with accession code 2VSN.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work was funded by the UK Biotechnology and Biological Sciences Research Council and the the Canadian Institutes of Health Research. G.J.D. is a Royal Society/Wolfson research Merit Award recipient. D.J.V. is a Canada Research Chair in chemical glycobiology and a fellow of the Michael Smith Foundation for Health Research (MSFHR). M.S.M. is a scholar of the MSFHR.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hart, G.W., Housley, M.P. & Slawson, C. *Nature* **446**, 1017–1022 (2007).
2. Wells, L., Vosseller, K. & Hart, G.W. *Science* **291**, 2376–2378 (2001).
3. Vosseller, K., Wells, L., Lane, M.D. & Hart, G.W. *Proc. Natl. Acad. Sci. USA* **99**, 5313–5318 (2002).
4. Kreppel, L.K., Blomberg, M.A. & Hart, G.W. *J. Biol. Chem.* **272**, 9308–9315 (1997).
5. Lubas, W.A. & Hanover, J.A. *J. Biol. Chem.* **275**, 10983–10988 (2000).
6. Lazarus, B.D., Roos, M.D. & Hanover, J.A. *J. Biol. Chem.* **280**, 35537–35544 (2005).
7. Gross, B.J., Kraybill, B.C. & Walker, S. *J. Am. Chem. Soc.* **127**, 14588–14589 (2005).
8. Lazarus, B.D., Love, D.C. & Hanover, J.A. *Glycobiology* **16**, 415–421 (2006).
9. Jinek, M. *et al. Nat. Struct. Mol. Biol.* **11**, 1001–1007 (2004).
10. Lairson, L., Henrissat, B., Davies, G.J. & Withers, S.G. *Annu. Rev. Biochem.* published online, doi:10.1146/annurev.biochem.76.061005.092322 (14 April 2008).
11. Yang, X.Y. *et al. Nature* **451**, 964–969 (2008).
12. Conti, E. & Kuriyan, J. *Structure* **8**, 329–338 (2000).