



UNIVERSITY OF LEEDS

This is a repository copy of *Structural and functional insight into human O-GlcNAcase.*

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/117506/>

Version: Accepted Version

Article:

Roth, C, Chan, S, Offen, WA et al. (7 more authors) (2017) Structural and functional insight into human O-GlcNAcase. *Nature Chemical Biology*, 13 (6). pp. 610-612. ISSN 1552-4469

<https://doi.org/10.1038/nchembio.2358>

© 2017, Nature America, Inc. This is an author-produced version of a paper published in *Nature Chemical Biology*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

1

2

3

Structural and functional insight into human O-GlcNAcase

4

5 Christian Roth¹, Sherry Chan¹, Wendy A Offen¹, Glyn R Hemsworth¹, Lianne I Willems²,
6 Dustin T King², Vimal Varghese², Robert Britton², David J Vocadlo^{2,*} and Gideon J Davies^{1,*}

7 1. York Structural Biology Laboratory, Department of Chemistry University of York, York,
8 YO10 5DD UK

9 2. Department of Chemistry, Simon Fraser University, Burnaby, V5A 1S6 British Columbia,
10 Canada

11 * corresponding authors: gideon.davies@york.ac.uk, dvocadlo@sfu.ca

12 **O-GlcNAc hydrolase, OGA, removes O-linked N-acetylglucosamine (O-GlcNAc) from myriad**
13 **nucleocytoplasmic proteins. Through co-expression and assembly of OGA fragments we**
14 **determined the 3-D structure of human OGA, revealing an unusual helix exchanged dimer**
15 **that lays a structural foundation for an improved understanding of substrate recognition**
16 **and regulation of OGA. Structures of OGA in complex with a series of inhibitors define a**
17 **precise blueprint for the design of inhibitors having clinical value.**

18

19 The dynamic O-GlcNAc modification of hundreds of nuclear and cytoplasmic proteins plays
20 diverse roles in a range of cellular processes including, for example, transcriptional
21 regulation and stress response (reviewed in Ref^{1,2}). Dysregulation of O-GlcNAcylation has
22 been implicated in diseases including cancer³, obesity⁴, and neurodegenerative diseases.^{5,6}
23 Notably, therapeutic agents targeting the O-GlcNAc modification have entered phase I
24 clinical trials, stimulating interest in the molecular and chemical basis of O-GlcNAcylation
25 and its manipulation with small molecules.⁷

26 Within mammals, this modification of serine and threonine residues is installed by O-GlcNAc
27 transferase, OGT, for which extensive structural data are available.^{8,9} The O-GlcNAc
28 modification is removed by O-GlcNAc hydrolase, OGA.¹⁰ Structures of bacterial homologs of
29 OGA from CAZY family GH84¹¹ (originally, and most notably *Bacteroides thetaiotaomicron*
30 (*BtGH84*)¹² and *Clostridium perfringens* (*CpNGA*)¹³), having a conserved active site with OGA,
31 have aided glycomimetic inhibitor design. Indeed, compounds based on the neighboring-
32 group catalytic mechanism^{12,14} have been applied in cellular and animal studies. The absence
33 of structural data for mammalian OGA, however, has limited efforts toward inhibitor design
34 and curtailed insight into peptide substrate binding and association with binding partners.
35 Accordingly, as part of our long-standing effort to understand O-GlcNAcase, we set out to

36 dissect human OGA (hOGA) and study its structure as well as its binding to different inhibitor
37 classes.

38 We first sought to establish a functional construct of human OGA for structural analyses.
39 hOGA is a complex multi-domain protein, produced as two splice variants; long-form OGA-L
40 and a less-active short form, OGA-S (reviewed in ref ¹⁵). OGA-L consists of an N-terminal
41 catalytic domain (GH84), a helical domain, extensive regions predicted as disordered, and a
42 C-terminal domain of unknown function having similarity to histone acetyl transferase (HAT)
43 domains (**Supplementary Results, Supplementary Fig. 1a**). We generated various truncated
44 constructs and screened possible domain boundaries¹⁵, none of these (**Supplementary Table**
45 **1**), however, yielded protein amenable to structural analysis.

46 Central to our successful strategy was that hOGA can be cleaved by caspase-3 into two
47 fragments, which remain tightly associated and active in solution.¹⁶ Accordingly, co-
48 expression of two hOGA fragments should permit their systematic truncation and removal of
49 putative disordered regions, yet allow their assembly into active hOGA in a form suitable for
50 crystallization. We co-expressed an extensive series of N- and C-terminal constructs in which
51 the putative disordered regions were systematically truncated (**Supplementary Table 1**).
52 Constructs were screened for formation of stable complexes and screened for crystallization.
53 Ultimately, a construct comprising amino acids 11-396 (N-terminal fragment) and 535-715
54 (C-terminal fragment) (**Supplementary Fig. 1b**) yielded crystals suitable for structure
55 determination. Furthermore, this construct, which we term "Split1", has essentially wild-
56 type catalytic activity toward synthetic substrates (**Fig. 1a**) and processes O-GlcNAcylated
57 proteins (**Fig. 1b, Supplementary Fig. 2, 3**).

58 We solved the structure of Split1 (**Supplementary Table 2**) by molecular replacement using a
59 sculpted model of *BtGH84* having 31% identity in the catalytic domain. This initial model was
60 greatly improved by extensive rebuilding, guided by the positions of sulfur atoms from
61 methionine and cysteine residues, observed using a long wavelength dataset,
62 (**Supplementary Fig. 4**). The final model of Split1 (11-396 / 535-715; in which residues 11-58,
63 341-370, 535-536, 596-598, 674-675 and 696-706 are disordered) reveals a two-domain
64 structure, with an N-terminal catalytic domain followed by a C-terminal helical bundle (**Fig.**
65 **1c**). hOGA forms a dimer with a contact interface of 4390 Å², calculated using PISA.¹⁷ We
66 confirmed that the dimer also forms in solution by size exclusion chromatography with
67 multi-angle light scattering (SEC-MALS) (**Supplementary Fig. 5a**). Notably, a dimer "swap"
68 of the C-terminal helix from each of the helical domains (Q676-P694) completes the open 3-
69 helix bundle to yield a closed 4-helix coiled coil bundle (**Fig. 1c**). This swap is essential for
70 stable interactions between both the two domains and the two monomers. Deletion of this
71 helix compromised dimer formation as assessed by SEC-MALS, and led to disassembly into
72 its peptide components to yield a near inactive construct (**Fig. 1a, Supplementary Fig. 5b, 6**).
73 SEC-MALS also showed near full-length hOGA (His₍₆₎9-916) is primarily a dimer
74 (**Supplementary Fig. 5c**). Previous evidence on the native multimerization status of hOGA
75 was conflicting and depended on the method used.^{10,18,19} When superposed with the

76 bacterial homolog structures, *BtGH84*¹², *CpNGA*¹³ and *Oceanicola granulosus* OGA
77 (*OgOGA*)²⁰ there is structural conservation of the catalytic domain and in particular the “-1”
78 sugar binding subsite that we define as the active site pocket (**Supplementary Fig. 7**). The
79 majority of active site residues comprising this subsite are conserved and mutations of these
80 residues in these enzymes *BtGH84*^{12,13,20} and in hOGA²¹, coupled with detailed mechanistic
81 studies,^{14,21,22} provide clear support for a catalytic mechanism involving substrate-assisted
82 catalysis from the substrate acetamido group. The helical regions that contribute to the
83 putative peptide-binding cleft (see below) are, however, markedly different; both in
84 organization and residue identity (**Supplementary Fig. 7-9**). The most similar helical domain
85 is seen for *OgOGA*²⁰, though that model was assigned a monomeric organization lacking the
86 helix exchange observed in hOGA.

87 The active site pocket of hOGA is located at the base of a V-shaped cleft (~ 22 x 25 Å with an
88 angle of ~70°) formed between the catalytic domain of monomer 1 and the C-terminal
89 helical bundle of the other monomer of the dimer (**Fig. 1c, Supplementary Fig. 10**). The
90 unusual dimer topology thus plays an unforeseen functional –as well as structural – role by
91 contributing to formation of this groove, which likely binds the peptide component of
92 substrates. The helical bundle contributes a rigid structure (**Supplementary Fig. 10**), but also
93 a flexible loop connecting the swapped helix, which can adopt different conformations as
94 discussed below. This mobility may confer plasticity to this groove, perhaps allowing it to
95 accommodate different substrates and inhibitors. Notably, our analysis revealed unexpected
96 density in this groove (**Fig. 1d, Supplementary Fig. 11a**) that we assigned as the C-terminal
97 end (P707-Y715) of the helical bundle with Y715 adopting a position that is consistent with
98 the position of an O-GlcNAcylated serine/threonine residue (**Supplementary Fig. 12**).
99 Notably, this crystal-packing derived peptide runs in the opposite direction to that observed
100 in bacterial structures²⁰ and thus is only indicative of the peptide binding surface; further
101 study of peptide complexes will be needed to clarify binding of protein substrates.

102 Key to our analysis of hOGA is how it binds inhibitors. Extensive soaking of diverse chemical
103 inhibitors was performed to displace the C-terminal peptide. Structures of complexes with
104 the mechanism-derived inhibitor Thiamet-G ($K_i=0.9$ nM),²³ the “PUGNAc-imidazole” hybrid
105 ($K_i=3.9$ μM)²⁴ and a potent derivative (VV347) of recently described pyrrolidine inhibitors
106 ($K_i=8$ nM),²⁵ were determined (**Fig. 2, Supplementary Fig. 11b, 13, Supplementary Note**).
107 The sugar-like moieties of these inhibitors occupy the -1 GlcNAc binding site, making
108 hydrogen bonds with, for example G67, K98, N280, D285 and N313. The alkyl amino-group
109 of Thiamet-G fills the pocket, formed by C215, Y219 and W278. In all three cases the
110 catalytic residues D174 and D175 are engaged with the inhibitors, representing a “closed”
111 conformation. D174 interacts with both nitrogens of the N-aminothiazoline moiety as
112 expected for this transition state mimic. The acetamido groups of VV347, and PUGNAc-
113 imidazole, point into the same pocket with the acetamido nitrogen interacting with D174
114 (**Fig. 2, Supplementary Fig. 14**), consistent with its role as polarizing residue. D175 points
115 towards the anomeric carbon as expected for its role as general acid/base in the catalytic
116 cycle.²² Interactions of the sugar-like moiety of these inhibitors within the -1 sugar binding

117 subsite are conserved between hOGA and their bacterial counterparts, explaining the
118 success of inhibitors targeting this site. Not conserved, however, are residues outside the -1
119 sugar binding subsite. These residues participate in the recognition of the aglycon
120 component of inhibitors that project out of the active site pocket and can be observed here
121 for the first time (**Fig. 2, Supplementary Fig. 11b**).

122 The aglycon of the glucoimidazole inhibitor mainly interacts with residues of the catalytic
123 domain that are outside the -1 binding site pocket, notably F223 and V254. The phenyl group
124 is within 4.5 Å of the loop comprising residues 677-683 of the helical bundle of monomer 2
125 (680-loop) suggesting a direct contribution of the helical bundle domain to the binding of the
126 inhibitor and potential substrates. The complex with VV347 (**Fig. 2**) reveals interactions with
127 multiple residues of the helical bundle domain. The helix comprising residues 633-662 and
128 the 680-loop, are core structural elements of the peptide-binding groove. The
129 trifluoromethyl-phenyl group of VV347 binds in a pocket formed by the side chain of W645,
130 part of helix α 4 and W679, part of the 680-loop, which undergoes a major reorientation
131 (**Supplementary Fig. 15**). Previous mutagenesis studies suggest the 680-loop interacts with
132 protein substrates²⁰ though further study will illuminate the precise roles of this feature.

133 In summary, we report a functional construct of human OGA obtained by exploiting the abil-
134 ity of the N- and C-terminal fragments of hOGA to associate stably. The structure revealed an
135 unusual obligate dimer with intertwined helical bundle domains that leads to residues from
136 both domains contributing to formation of the substrate-binding site. Structures of OGA in
137 complex with rationally-designed high affinity inhibitors define both the active site pocket
138 and, crucially, the surrounding peptide-binding and aglycon regions in a manner that is
139 unique to mammalian OGA. Strikingly, part of this peptide binding site is a flexible loop con-
140 necting the swapped helix of the helical bundle domain with the opposite peptide binding
141 groove, which may open-up possible communication between active sites in response to
142 peptide binding. Exploitation of this peptide binding groove will offer new opportunities for
143 the design of OGA inhibitors as research tools and for potential clinical use in treating O-
144 GlcNAc related diseases.

145 **Accession codes:**

146 The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes
147 5M7R for the apo-structure, 5M7S for the Thiamet-G complex, 5M7T for the PugNAc-imidazole complex and 5M7U for the
148 complex with VV347.

149 **Acknowledgements:**

150 The authors thank Diamond Light Source for beamtime (proposals mx-1221, mx-7864 and mx-9948), and the staff of beam-
151 lines I02, I03 and I24 for assistance. The authors are grateful to Dr. Johan P Turkenburg and Sam Hart for their help in crys-
152 tal testing and data collection. The authors thank Juliet Borgia, Simon Grist, Andrew Leech and Louise Haigh for technical
153 support. This research was supported by funding from the Biotechnology and Biological Sciences Research Council
154 (BB/K003836/1) and the Canadian Institutes of Health Research (MOP-123341), Brain Canada, Genome British Columbia,
155 the Michael Smith Foundation for Health Research. LIW is supported by the Netherlands Organization for Scientific Re-
156 search (NWO) and the Banting Postdoctoral Fellowships program are also thanked for financial support. DJV is supported as
157 a Tier I Canada Research Chair in Chemical Glycobiology. GJD is supported by the Royal Society through a Ken Murray re-
158 search professorship.

159

160

161

162 **Author contributions:**

163 C.R. designed truncated constructs, cloned, expressed, crystallized and solved the structure. S. C. designed truncated
164 constructs, cloned, expressed, purified and crystallized protein. W.A.O. cloned purified and crystallized protein. G.R.H.
165 designed experiments, cloned and purified protein. L.I.W. performed cell culture and western blot assays. D.T.K. performed
166 the kinetic characterization. V.V. synthesized VV347. R.B. and D.J.V designed the pyrrolidine inhibitors. G.J.D designed
167 cloning and structural experiments. D.J.V designed biochemical and inhibition experiments. C.R., D.J.V. and G.J.D wrote the
168 manuscript with contributions from all authors.

169

170 **References:**

- 171 1. Hardiville, S. & Hart, G.W. Nutrient regulation of signaling, transcription, and cell physiology
172 by O-GlcNAcylation. *Cell Metab* **20**, 208-13 (2014).
- 173 2. Groves, J.A., Lee, A., Yildirim, G. & Zachara, N.E. Dynamic O-GlcNAcylation and its roles in the
174 cellular stress response and homeostasis. *Cell Stress Chaperones* **18**, 535-58 (2013).
- 175 3. Ferrer, C.M. et al. O-GlcNAcylation regulates cancer metabolism and survival stress signaling
176 via regulation of the HIF-1 pathway. *Mol Cell* **54**, 820-31 (2014).
- 177 4. Lagerlof, O. et al. The nutrient sensor OGT in PVN neurons regulates feeding. *Science* **351**,
178 1293-6 (2016).
- 179 5. Liu, F. et al. Reduced O-GlcNAcylation links lower brain glucose metabolism and tau
180 pathology in Alzheimer's disease. *Brain* **132**, 1820-1832 (2009).
- 181 6. Yuzwa, S.A. et al. Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against
182 aggregation. *Nat Chem Biol* **8**, 393-399 (2012).
- 183 7. Smith, S.M. et al. Eearly Clinical Results and Preclinical Validation of the O-GlcNAcase (OGA)
184 Inhibitor MK-8719 as a novel Therapeutic for the Treatment of Tauopathies. *Alzheimer's &*
185 *Dementia: The Journal of the Alzheimer's Association* **12**, P261 (2016).
- 186 8. Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase
187 and its complex with a peptide substrate. *Nature* **469**, 564-567 (2011).
- 188 9. Jinek, M. et al. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits
189 structural similarities to importin alpha. *Nat Struct Mol Biol* **11**, 1001-1007 (2004).
- 190 10. Gao, Y., Wells, L., Comer, F.I., Parker, G.J. & Hart, G.W. Dynamic O-glycosylation of nuclear
191 and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-
192 acetylglucosaminidase from human brain. *J Biol Chem* **276**, 9838-9845 (2001).
- 193 11. Cantarel, B.L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource
194 for Glycogenomics. *Nucleic Acids Res* **37**, D233-8 (2009).
- 195 12. Dennis, R.J. et al. Structure and mechanism of a bacterial beta-glucosaminidase having O-
196 GlcNAcase activity. *Nat Struct Mol Biol* **13**, 365-371 (2006).
- 197 13. Rao, F.V. et al. Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc
198 hydrolysis. *EMBO J* **25**, 1569-78 (2006).
- 199 14. Macauley, M.S., Whitworth, G.E., Debowski, A.W., Chin, D. & Vocadlo, D.J. O-GlcNAcase uses
200 substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-
201 inspired inhibitors. *J Biol Chem* **280**, 25313-25322 (2005).
- 202 15. Vocadlo, D.J. O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and
203 enzyme regulation. *Curr Opin Chem Biol* **16**, 488-97 (2013).
- 204 16. Butkinaree, C. et al. Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3
205 during apoptosis. *J Biol Chem* **283**, 23557-66 (2008).
- 206 17. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res* **43**,
207 W314-9 (2015).

- 208 18. Wells, L. et al. Dynamic O-glycosylation of nuclear and cytosolic proteins: further
209 characterization of the nucleocytoplasmic beta-N-acetylglucosaminidase, O-GlcNAcase. *J Biol*
210 *Chem* **277**, 1755-1761 (2002).
- 211 19. Izumi, T. & Suzuki, K. Neutral beta-N-acetylhexosaminidases of rat brain. Purification and
212 enzymatic and immunological characterization. *J Biol Chem* **258**, 6991-9 (1983).
- 213 20. Schimpl, M., Schuttelkopf, A.W., Borodkin, V.S. & van Aalten, D.M. Human OGA binds
214 substrates in a conserved peptide recognition groove. *Biochem J* **432**, 1-7 (2010).
- 215 21. Cetinbas, N., Macauley, M.S., Stubbs, K.A., Drapala, R. & Vocadlo, D.J. Identification of
216 Asp174 and Asp175 as the key catalytic residues of human O-GlcNAcase by functional
217 analysis of site-directed mutants. *Biochemistry* **45**, 3835-3844 (2006).
- 218 22. He, Y., Macauley, M.S., Stubbs, K.A., Vocadlo, D.J. & Davies, G.J. Visualizing the reaction
219 coordinate of an O-GlcNAc hydrolase. *J Am Chem Soc* **132**, 1807-1809 (2010).
- 220 23. Cekic, N. et al. Analysis of transition state mimicry by tight binding aminothiazoline inhibitors
221 provides insight into catalysis by human O-GlcNAcase. *Chemical Science* **7**, 3742-3750 (2016).
- 222 24. Shanmugasundaram, B. et al. Inhibition of O-GlcNAcase by a gluco-configured nagstatin and
223 a PUGNAc-imidazole hybrid inhibitor. *Chem Commun (Camb)*, 4372-4 (2006).
- 224 25. Bergeron-Brlek, M. et al. A Convenient Approach to Stereoisomeric Iminocyclitols:
225 Generation of Potent Brain-Permeable OGA Inhibitors. *Angew Chem Int Ed Engl* **54**, 15429-33
226 (2015).

227 **Figure legends:**

228 **Figure 1:** Function and structure of hOGA. a) Michaelis-Menten kinetics of crystallized construct Split 1, full-
229 length hOGA (OGA-L) and Split 2. The kinetic parameters for the OGA variants are, OGA-L: $V_{max} = 1.10 \pm 0.02 \mu\text{M}$
230 min^{-1} ; $K_M = 92.47 \pm 4.86 \mu\text{M}$, and OGA-Split 1: $V_{max} = 1.11 \pm 0.03 \mu\text{M} \text{min}^{-1}$; $K_M = 40.93 \pm 3.30 \mu\text{M}$. Therefore, the
231 catalytic efficiency (k_{cat}/K_M) of OGA-L is roughly similar (2.3 fold lower) to OGA-Split 1 when using pNP-GlcNAc
232 as the substrate. Data represent average of quadruplicate rate measurements \pm s. d. b) Immunoblot analysis
233 showing that Split1 can digest recombinant O-GlcNAcylated TAB1, evaluated using anti-O-GlcNAc antibody
234 CTD110.6 anti-Histidine antibody as a loading control. Thiamet-G inhibits digestion. c) Ribbon diagram of the
235 hOGA dimer, colored by chain. The helix (Q676-P694) swapped between the two helical bundles is marked with
236 a star. The position of the active site (-1 sugar binding site) is indicated by the Van der Waals' surface of the
237 inhibitor Thiamet-G in bluegreen. d) Binding of the C-terminal peptide (shown with its Van der Waals surface in
238 bluegreen) of the helical bundle fragment to the hOGA peptide-binding groove. Full blots are provided in the
239 **Supplementary Fig. 15.**

240 **Figure 2:** Ligand binding to hOGA. a) Binding of Thiamet-G (left), the PUGNAc-imidazole" hybrid (middle) and
241 the pyrrolidine derivative "VV347" in the active site of hOGA. The corresponding electron density is shown at
242 2.0σ r.m.s.d. ($0.28 \text{ e}/\text{\AA}^3$) for Thiamet-G. For the "PUGNAc-imidazole" hybrid (middle) at 1.5σ r.m.s.d. (0.17
243 $\text{e}/\text{\AA}^3$), and the pyrrolidine derivative "VV347" at 1.5σ r.m.s.d. ($0.21 \text{ e}/\text{\AA}^3$). The catalytic residues as well the
244 acetamido pocket forming residues are shown as sticks. c) Van der Waals' surface of the inhibitor binding sites
245 showing that the -1-subsite is deeply buried with the aglycon units extending outwards interacting with both
246 domains of hOGA.

247 **Online Methods:**

248 **Cloning, expression and purification**

249 The gene of the longest isoform of hOGA (hOGA-L (Uniprot Accession number: O60502)) was
250 synthesized in a codon optimized form for recombinant expression in *Escherichia coli*. The successful
251 expression construct encoded the N-terminal region of hOGA, comprising amino acids 11-396, and
252 the C-terminal region, comprising amino acids 535-715. The N-terminal construct was cloned in the
253 vector pACYC-Duet (Millipore) using the sequence and ligation independent cloning method ²⁶ in
254 frame with an N-terminal His₆-Tag. The C-terminal construct was cloned into the vector pET-YSBLIC3C

255 ²⁷ with an N-terminal His₆-Tag followed by a 3C-protease cleavage site using the same method. The
256 nucleotide sequences of all made constructs were confirmed by sequencing. Both vectors were
257 simultaneously transformed into *E. coli* BL21(DE3)-Gold (Agilent) for subsequent protein expression.
258 Cells were grown in 2L TB-medium to an OD₆₀₀ of ~1.0 and protein synthesis was then induced by
259 adding IPTG to a final concentration of 0.1 mM. Protein expression was carried out at 16 °C with an
260 induction time of 20 h. The cells were harvested by centrifugation at 4500xg for 20 min, flash frozen
261 and stored at -20°C until required.

262 For purification of the hOGA complex, cells were resuspended in 50 mM HEPES pH 7.0, 750 mM NaCl,
263 20 mM imidazole, and 0.5 mM DTT (resuspension buffer). Cells were lysed using a French Press at 25
264 kPsi. The lysate was cleared by centrifugation at 50,000 g for 1h and the supernatant was passed
265 through a 10 mL HisTrap FF column (GE Healthcare) pre-equilibrated with resuspension buffer. The
266 bound hOGA was purified by gradient elution over 10 column volumes using 0 to 50% of elution
267 buffer (50 mM HEPES pH 7.0, 750 mM NaCl, 500 mM imidazole and 0.5 mM DTT). hOGA containing
268 fractions were combined, concentrated by ultrafiltration using Vivaspin columns (Sartorius) with a
269 molecular weight cut-off (MWCO) of 30 kDa, and applied to a Superdex S200 column (GE Healthcare)
270 pre-equilibrated with size-exclusion buffer (10 mM HEPES pH 7.0, 250 mM NaCl, 1 mM DTT).
271 Fractions corresponding to the dimeric form of hOGA were combined and concentrated to 20 mg/ml
272 by ultrafiltration with a Vivaspin (MWCO: 30 kDa) column, flash frozen using liquid nitrogen, and
273 stored at -80°C until required.

274 **Crystallization and data collection**

275 Initial crystallization conditions were identified using commercially available screens from Hampton
276 and Molecular Dimension in a 96 well sitting drop screening format. Further optimization in a 48 well
277 sitting drop format provided suitable conditions for reliable crystallization (crystallization solution:
278 0.1-0.2 M (NH₄)₃-citrate pH 6.5-7.5; 16-24 % PEG 3350). Optimal crystals were reliably obtained by
279 micro seeding with previously obtained crystals. For data collection, protein crystals were transferred
280 into crystallization solution containing 25% PEG3350 (cryoprotectant solution), which enabled
281 cryoprotection of the crystals. Crystals were recovered using a Nylon microfibre loop (Hampton) and
282 flash frozen in liquid nitrogen. For soaking experiments the inhibitors were dissolved in 10 % (v/v)
283 DMSO to a concentration of 100 mM and added to a drop containing the cryoprotectant solution to a
284 final inhibitor concentration of 10 mM. Crystals were soaked with inhibitors for times ranging from
285 48 hours to 1 week. The resulting crystals were handled as described above. Data were collected at
286 the Diamond light source beamlines I02, I03 or I04 using a Pilatus 6M detector (Dectris) at a
287 wavelength of 0.979 Å. Data were collected over 180° with an oscillation angle of 0.1°. Data were
288 integrated with XDS²⁸, integrated in the XIA2 pipeline²⁹ and scaled using AIMLESS³⁰.

289 **Structure solution and refinement**

290 The structure was solved by molecular replacement using Phaser³¹ in conjunction with a sculpted
291 model of *BtGH84* (PDB-ID 2CHO)¹². The initial model was rebuilt and refined using Buccaneer³². The
292 final model was obtained by alternating rounds of manual model building in COOT³³, followed by
293 reciprocal refinement with Refmac³⁴ or Phenix³⁵. For inhibitor complexes the apo-structure was
294 refined against data from a crystal soaked with the respective inhibitor. If clear density for a bound
295 inhibitor could be identified a model of the ligand was built using Acedrg, part of the CCP4- software
296 package³⁶ and incorporated in the apo model which was then subsequently refined. The quality of
297 the final models were judged using MolProbity³⁷. The number of outliers is between 0.2 to 0.4 %.
298 Figures of the structural models were prepared using CCP4MG³⁸.

299 **Enzyme kinetics**

300 Initial rate experiments of OGA-L, OGA-Split 1, and OGA-Split 2 catalyzed *pNP*-GlcNAc hydrolysis were
301 carried out in PBS buffer (pH 7.4) and monitored continuously at 25°C at a wavelength of 405 nm

302 using a SpectraMax i3x multi-mode plate reader from Molecular Devices. Reactions were performed
303 in a 384 well clear assay plate from Corning (Product #3702) in a final reaction volume of 45 μ L.
304 Steady state kinetic values were attained from substrate dose response curves using 50 nM of the
305 OGA-L, OGA-Split 1, and OGA-Split 2 variants unless otherwise stated and varying concentrations of
306 *p*NP-GlcNAc. Reaction velocities were determined by linear regression of the progress curves over a
307 15 min period. The amount of product formed was assessed by creating a *p*NP standard curve in PBS
308 buffer. The substrate dose-response curves were then fit to the Michaelis-Menten equation using the
309 GraphPad Prism5 software package.

310

311 Inhibition assays were performed using a final concentration of 20 nM OGA-L and 200 μ M of 4-
312 methylumbelliferyl-2-acetamido-2-deoxy- β -D-glucopyranoside with various concentrations of VV-347
313 and Thiamet-G. The assay was performed at 37°C in a final volume of 45 μ L in inhibition buffer: PBS
314 pH7.4, 0.0025% BSA, 0.0025mM DTT. Reactions were initiated with the addition of enzyme, and re-
315 action progress was monitored continuously (excitation and emission wavelengths: 350, and 445nm)
316 over a 10-minute period. The amount of fluorophore liberated was assessed using a standard curve
317 for 4-methylumbelliferone in inhibition buffer. Inhibitor K_i values were determined using the Morri-
318 son equation for tight binding inhibition as described previously³⁹. All curve fitting for enzyme kinet-
319 ics and inhibition experiments was performed using GraphPad Prism, and error bars correspond to
320 S.D. from three technical replicates (triplicate reads). The experiments were all repeated at least
321 twice to ensure reproducibility of the data.

322

323 **SEC-MALS**

324

325 Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light
326 scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system
327 (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A
328 autosampler). Work was conducted at room temperature (20 \pm 2°C). Sample injection volume was
329 100 μ L at a protein concentration of 5 mg/ml. The samples were separated on a Superdex S200
330 10/300 (GE Healthcare) using 10 mM Tris pH 7, 250 mM NaCl as buffer. Shimadzu LC Solutions
331 software was used to control the HPLC and Astra V software for the HELEOS-II and rEX detectors.
332 Data were analyzed using the Astra V software. MWs were estimated using the Zimm fit method with
333 degree 1. A value of 0.174 was used for protein refractive index increment (dn/dc).

334

335

336

337

338 ***Digestion of HEK293 cell extracts with OGA constructs***

339 HEK293 cells were obtained from ATCC. They were not further authenticated. However, as we are
340 looking at total GlcNAc levels in cell extracts and the digestion of O-GlcNAc by OGA, rather than cell-
341 type specific physiology, the exact cell type is less relevant in our experiments. The cells were tested
342 for mycoplasma contamination in June 2016 (negative result). HEK293 cells were cultured in high
343 glucose (4.5 g/L) Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) Fetal
344 Bovine Serum (FBS), 100 units/mL Penicillin and 100 μ g/mL Streptomycin in a 5% CO₂ humidified
345 incubator at 37 °C. Cells were seeded into a 100 mm cell culture dish and grown to 95% confluency.
346 Cells were washed with ice-cold PBS, harvested by scraping on ice in ice-cold PBS, pooled and centri-
347 fuge-d at 700 g for 10 min at 4 °C. The cell pellet was resuspended in 150 μ L of lysis buffer (50 mM
348 NaH₂PO₄, pH 7.0, 100 mM NaCl, 1% (v/v) NP-40 substitute, 0.5% (w/v) sodium deoxycholate, and 1
349 mM PMSF) and incubated on ice for 30 min. The cell extracts were then centrifuged at 14,000 g for
350 10 min at 4 °C and the clear supernatant was collected. The protein concentration, determined by
351 DCTM protein assay (Bio-Rad), was 29 mg/mL. HEK293 cell lysates (5 μ L, 150 μ g) were mixed with 5 or

352 25 μ M of OGA-Split 1, OGA-L or BtGH84 (5.5 μ L 2x solution in 50 mM NaH₂PO₄, 100 mM NaCl, pH 7.0)
353 in the presence or absence of 250 μ M Thiamet-G (0.6 μ L 5 mM in PBS). The reactions were incubated
354 at 25 °C for 3 hrs. Under these conditions, enzyme activity, as determined by *p*NP-GlcNAc hydrolysis,
355 was shown to be stable (see Supplementary Figure 13). After the reaction, 4 μ L of binding buffer (500
356 mM NaH₂PO₄, 100 mM NaCl, pH 8.0) was added and the mixtures were added to 30 μ L of Nickel-NTA
357 Agarose beads (prewashed in binding buffer). The samples were incubated for 1 hr at rt with rotation
358 to remove the His-tagged enzyme. The beads were spun down at 5,000 g for 2 min and the superna-
359 tant was collected (15 μ L), mixed with 2x Laemmli's sample buffer containing β -mercaptoethanol and
360 boiled for 5 min at 100 °C. A third (~ 50 μ g of protein) of each sample was resolved on a 4-20% Mini-
361 PROTEAN TGX gradient gel (Bio-Rad). The proteins were transferred onto a nitrocellulose membrane
362 using a Bio-Rad wet western blotting system (1 hr at 100V). The membrane was blocked with 2% BSA
363 in PBSt for 1 hr at rt and hybridized with mouse CTD110.6 antibody (BioLegend, 1:3,000) and rabbit
364 anti- β -actin antibody (LI-COR, 1:5,000) in PBS with 2% BSA overnight at 4 °C, followed by IRDye 680LT
365 goat anti-mouse antibody (LI-COR, 1:10,000) and IRDye 800CW goat anti-rabbit antibody (LI-COR,
366 1:10,000) in PBSt with 2% BSA for 1 hr at rt. Proteins were visualized using a LI-COR Odyssey scanner.
367 The protein ladder used was a PageRuler Prestained Plus Protein Ladder (Thermo Fisher Scientific).
368

369 ***Stability of OGA-Split 1 under reaction conditions***

370 The various OGA constructs were dissolved at 25 μ M in 50 mM NaH₂PO₄, 100 mM NaCl, pH 7.0. The
371 mixtures were incubated at 25 °C for 3 hrs. Before (t = 0 hrs) and after incubation (t = 3 hrs), 2 ali-
372 quots of 0.5 μ L were taken out and each added to 125 μ L of PBS. The diluted enzymes (100 nM) were
373 then mixed 1:1 with of 400 μ M 4-Nitrophenyl N-acetyl-D-glucosaminide (*p*NP-GlcNAc) in PBS and two
374 45 μ L aliquots (for duplicate reads) were transferred to a 384 well Corning clear-bottom plate. *p*NP-
375 GlcNAc hydrolysis was monitored continuously at a wavelength of 405 nm at 25°C using a Spectra-
376 Max i3x multi-mode plate reader (Molecular Devices). Background hydrolysis (the average value in
377 samples without enzyme at the corresponding time point) was subtracted from all values, after
378 which reaction velocities were determined by linear regression of the progress curves. These values
379 were then normalized to 100% activity (the average value in samples with the corresponding OGA
380 construct at 0 hrs incubation). Data were analyzed using GraphPad Prism5 software and represent
381 mean values \pm standard deviation from two biological replicates with two technical replicates (dupli-
382 cate reads) each (**Supplementary Fig. 16**).
383

384

385 ***Treatment of O-GlcNAcylated TAB1 with OGA constructs***

386 TAB1 was coexpressed with OGT in *E.coli* and purified as previously described.¹⁷ O-GlcNAcylated and
387 HIS-tagged TAB1 (5 μ g per sample, 2.5 μ L 2 mg/mL in 50 mM NaH₂PO₄, 100 mM NaCl, pH 7.0) was
388 treated with 5 or 25 μ M of OGA-Split 1, OGA-L or BtGH84 (3 μ L 2x solution in 50 mM NaH₂PO₄, 100
389 mM NaCl, pH 7.0) in the presence or absence of 250 μ M Thiamet-G (0.5 μ L 3 mM in PBS) at 25 °C for
390 3 hrs. The reactions were quenched by addition of 4 μ L of 5x Laemmli's sample buffer containing β -
391 mercaptoethanol and boiled for 5 min at 100 °C. A fifth (~ 1 μ g of TAB1) of each sample was resolved
392 on a 12% SDS-PAGE gel and analyzed by western blot as described above. Instead of rabbit anti- β -
393 actin antibody, a rabbit anti-HIS antibody was used as loading control (Cedarlane, 1:5,000).

394 **Data Availability:**

395 The atomic coordinates and structure factors have been deposited in the Protein Data Bank under
396 the accession codes 5M7R for the apo-structure, 5M7S for the Thiamet-G complex, 5M7T for the
397 PugNAc-imidazole complex and 5M7U for the complex with VV347. Any other datasets generated

398 during and/or analyzed during the current study are available from the corresponding author on
399 reasonable request.

400

401 **Online Methods References**

402

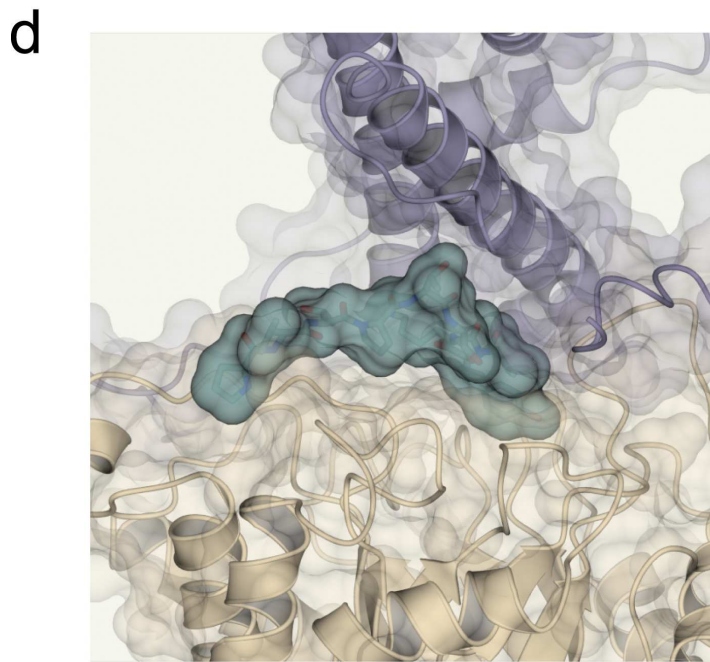
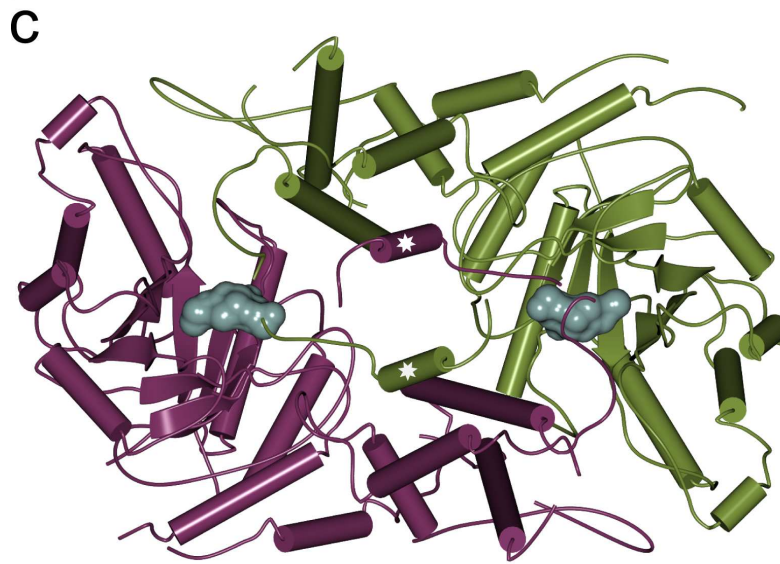
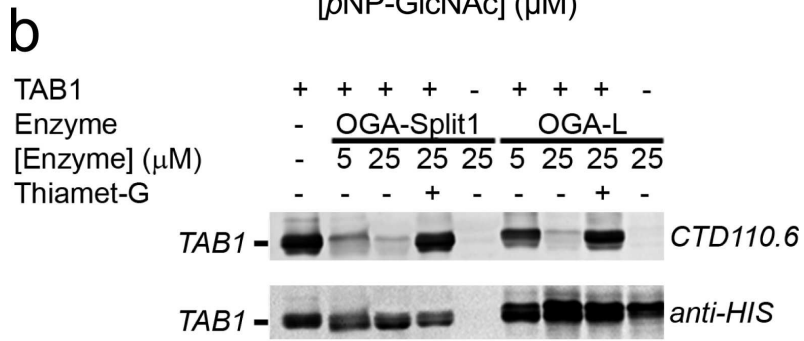
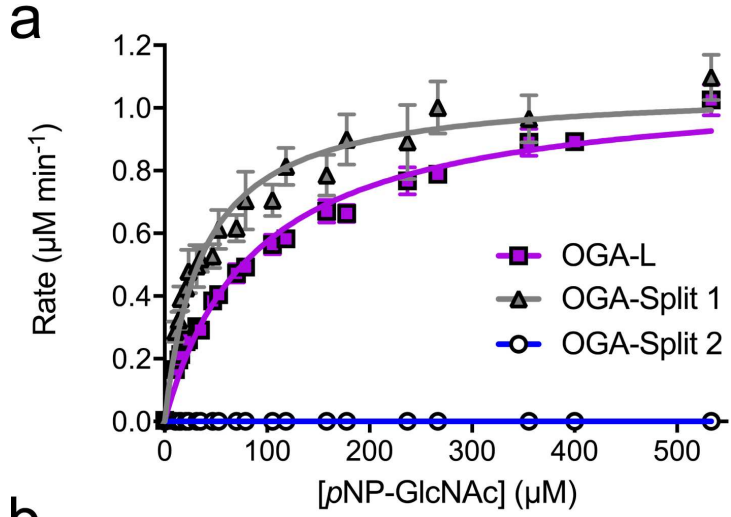
403

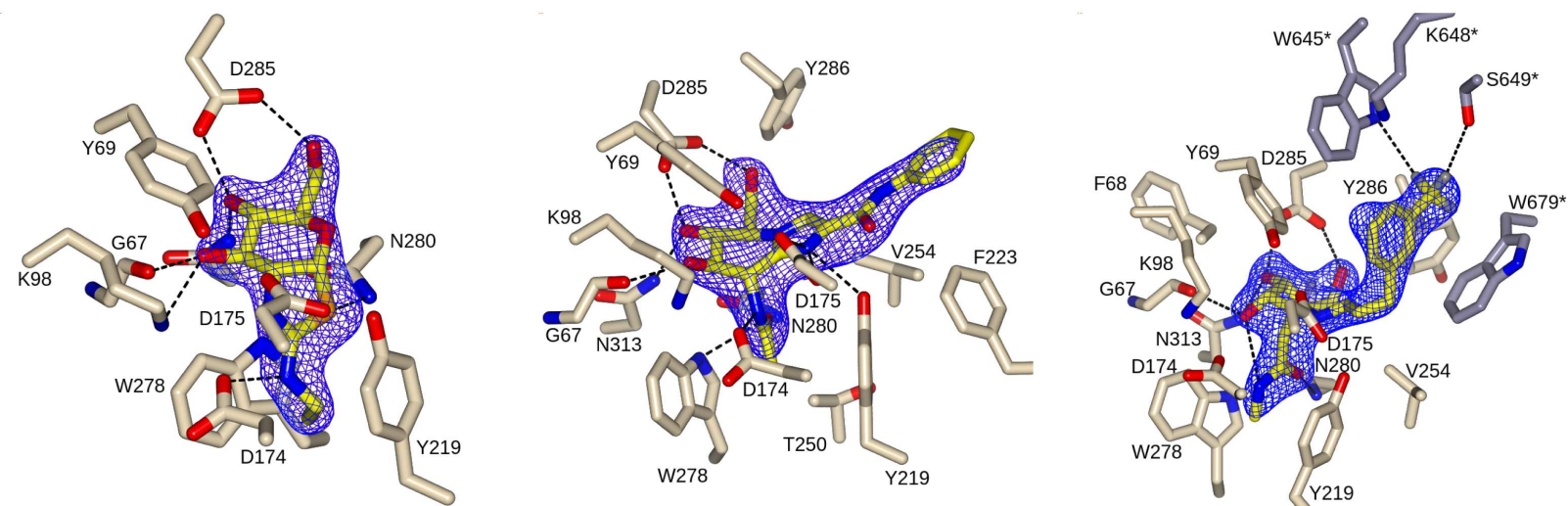
- 404 26. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning.
405 *Methods Mol Biol* **852**, 51-9 (2012).
- 406 27. Fogg, M.J. & Wilkinson, A.J. Higher-throughput approaches to crystallization and crystal
407 structure determination. *Biochem Soc Trans* **36**, 771-5 (2008).
- 408 28. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
- 409 29. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl*
410 *Crystallogr* **43**, 186-190 (2010).
- 411 30. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta*
412 *Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).
- 413 31. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
- 414 32. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains.
415 *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11 (2006).
- 416 33. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta*
417 *Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
- 418 34. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures.
419 *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
- 420 35. Adams, P.D. et al. The Phenix software for automated determination of macromolecular
421 structures. *Methods* **55**, 94-106 (2011).
- 422 36. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D*
423 *Biol Crystallogr* **67**, 235-242 (2011).
- 424 37. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular
425 crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
- 426 38. McNicholas, S., Potterton, E., Wilson, K.S. & Noble, M.E. Presenting your structures: the
427 CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**, 386-394 (2011).
- 428 39. Cekic, N. et al. Analysis of transition state mimicry by tight binding aminothiazoline inhibitors
429 provides insight into catalysis by human O-GlcNAcase. *Chem. Sci.* **7**, 3742-3750 (2016).

430

431 **Competing Financial Interests**

432 D.J.V. is a co-founder of and holds equity in the company Alectos Therapeutics. D.J.V. serves as CSO
433 and Chair of the Scientific Advisory Board of Alectos Therapeutics of which G.J.D is a member.



a**b**