Activity based probes for functional interrogation of retaining β-glucuronidases

Liang Wu[§#], Jianbing Jiang[‡#], Yi Jin[§#], Wouter W. Kallemeijn[†], Chi-Lin Kuo[†], Marta Artola[‡], Wei Dai[‡], Cas van Elk[‡], Marco van Eijk[†], Gijsbert A. van der Marel[‡], Jeroen D. C. Codée[‡], Bogdan I. Florea[‡], Johannes M. F. G. Aerts[†], Herman S. Overkleeft[‡], Gideon J. Davies[§*]*

[‡]Department of Bioorganic Synthesis, Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, the Netherlands

[†]Department of Medical Biochemistry, Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, the Netherlands

[§]York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, YO10, 5DD, UK

[#] L.W., J.J. and Y.J. contributed equally to this work

Corresponding authors: gideon.davies@york.ac.uk; h.s.overkleeft@chem.leidenuniv.nl

Abstract

Humans express at least two distinct β-glucuronidase enzymes involved in disease: *exo*-acting β-glucuronidase (GUSB), whose deficiency gives rise to mucopolysaccharidosis type VII, and *endo*-acting heparanase (HPSE), implicated in inflammation and cancers. The medical importance of these enzymes necessitates reliable methods to assay their activities in tissues. Herein, we present a set of β-glucuronidase specific activity based probes (ABPs) which allow for rapid and quantitative visualization of GUSB and HPSE in biological samples, providing a powerful tool for dissecting their activities in normal and disease states. Unexpectedly, we find that the supposedly inactive HPSE proenzyme proHPSE is also labeled by our ABPs, leading to surprising insights regarding structural relationships between proHPSE, mature HPSE, and their bacterial homologs. Our results demonstrate the application of β-glucuronidase ABPs in tracking pathologically relevant enzymes, and provide a case study of how ABP driven approaches can lead to discovery of unanticipated structural and biochemical functionality.

Introduction

Retaining β-glucuronidases are enzymes responsible for hydrolytic cleavage of β-linked glucuronides from polysaccharide and glycoconjugate molecules, with net retention of anomeric stereochemistry at the released glucuronide. Humans express at least two major retaining β-glucuronidases: *exo*-acting GUSB, responsible for cleaving β-linked glucuronides from the non-reducing end of diverse glycosaminoglycans (GAGs) in the lysosome, and *endo*-acting heparanase (HPSE), specifically responsible for breakdown of heparan sulfate (HS) in lysosomes and the extracellular matrix (ECM). Both enzymes are strongly implicated in disease processes: deficiency of GUSB is the basis of the autosomal recessive disease mucopolysaccharidosis type VII (MPSVII), also known as Sly syndrome[1-3], whilst HPSE overexpression is linked to a variety of pathologies including inflammation and cancer metastasis[4,5].

Although both GUSB and HPSE possess β-glucuronidase activity, these enzymes are dissimilar at the sequence level and fall under different families of the Carbohydrate Active enZymes (CAZy) classification scheme[6]: GH2 for GUSB and GH79 for HPSE. Structurally, human GUSB is a large homotetrameric assembly, with each protomer comprising a $(\beta/\alpha)_8$ barrel domain, a jelly roll domain and an Ig constant chain like domain[7]. In contrast, HPSE is a heterodimer comprised of 8 kDa and 50 kDa subunit chains, which fold to produce a $(\beta/\alpha)_8$ barrel flanked by a smaller β-sandwich domain[8]. The mature HPSE heterodimer is formed by proteolytic removal of a 6 kDa linker peptide from a single chain proenzyme - proHPSE.

Given the importance of β-glucuronidases in human health and disease, a facile method to visualize and quantitate their activity would be of great utility. We have previously reported the development of activity based probes (ABPs) based upon the cyclophellitol aziridine scaffold, which can be used to specifically detect enzymatic activity for a range of glycosidases[9-14]. These probes provide valuable tools to rapidly determine enzyme activities within their native physiological contexts.

Herein, we unveil the synthesis of ABPs designed to selectively target and label retaining β-glucuronidases. We demonstrate the utility of these probes in quantitating β-glucuronidase activity

in a range of cell and tissue samples, via both fluorescence and chemical proteomics approaches. Unexpectedly, we find that a monosugar β-glucuronidase probe is sufficient to label not only *exo*-acting GUSB, but also *endo*-acting HPSE, despite binding just one of multiple subsites within the HPSE active site cleft. Furthermore, the supposedly inactive proHPSE proenzyme is also labeled by ABPs, prompting us to investigate the nature of proHPSE 'inactivation' by its 6 kDa linker, and how this structure relates to other GH79 enzymes. Our results demonstrate a wide ranging potential of β-glucuronidase ABPs as biological and biomedical tool compounds, and highlight the general power of ABPs for driving the discovery of novel biological insights[15].

Results

Glucuronidase specific inhibitor and probe design

We have previously demonstrated cyclophellitol derived epoxides and aziridines to be powerful mechanism based inhibitors for retaining β-glucosidases[16], due to their ability to specifically label the enzyme catalytic nucleophile, in a conformation resembling the covalent intermediate of glycoside hydrolase reactions. (**Fig. 1a, b**)[17]. Inhibition is typically tolerant to functionalization at the ring nitrogen of cyclophellitol aziridines, allowing fluorophore or biotin tagging to create inhibitor probes which can label specific glycosidases within complex biological mixtures[18].

Conceptually, we envisioned that β-glucuronidase specific ABPs could be accessed from cyclophellitol by oxidation at the C6 equivalent position, to emulate the carboxylate of glucuronic acid (GlcUA). ABPs **1–4** are composed of such a β-glucuronide configured cyclophellitol aziridine, bearing a spacer from the aziridine nitrogen terminating in BODIPY-FL, a BODIPY-TMR analog, Cy5, or biotin respectively. Alongside these functionalized ABPs we also prepared azide substituted ABP **5** (the precursor of **1–4**), unsubstituted aziridine **6** and cyclophellitol-6-carboxylate **7**. (**Fig. 1c**; structures of additional compounds **8–16** used in this study are shown in **Supplementary Results**, **Supplementary Fig. 1**).

GlcUA ABPs target β-glucuronidases *in vitro* and *in situ*

To assess the potency of our β-glucuronidase ABPs *in vitro*, we first turned to the *exo*-acting GH79 β-glucuronidase AcGH79 from *Acidobacterium capsulatum*, whose activity is readily followed using the fluorogenic substrate 4-methylumbelliferyl-glucuronic acid (4MU-GlcUA)[19]. All compounds tested were effective inhibitors of AcGH79, with apparent IC50s in the low to sub nM range (**Table 1** left panel). Core ABP 'warhead' **6** inhibited AcGH79 with apparent IC50 of ~5 nM. This was potentiated by further functionalization: apparent IC50 of Cy5 substituted ABP **3** was ~1 nM, whilst **1**, **2**, **4** and **5** were all sub-nanomolar inhibitors of AcGH79. Apparent IC50 for epoxide **7** was ~34 nM, consistent with lower reactivity of the epoxide moiety compared to aziridines.

Kinetic parameters for inhibition of AcGH79 were determined using a continuous assay, whereby substrate and inhibitor react with enzyme simultaneously (**Supplementary Note 1**)[20], allowing us to derive a combined inhibition parameter k_i/K_I for all ABPs tested. k_i/K_I values largely reflected the trend seen with IC50s, with the activity of core aziridine **6** potentiated by further functionalization, and epoxide **7** substantially less active than aziridines (**Table 1** middle panel, **Supplementary Fig. 2**).

Finally, we tested the ability of our probes to inhibit β-glucuronidases in live fibroblast cells. *In situ* apparent IC50s were determined for ABPs **2** and **3** to be in the low μM range (~1.7 and ~1.8 μM respectively). We were unable to determine *in situ* apparent IC50s for **1** or **4–7**, likely reflecting a limited ability of these compounds to permeate the cell membrane (**Table 1** right panel).

Fluorescent labeling of AcGH79 by ABP **1** was readily visualized after running on SDS-PAGE, and could be blocked by competition with **2–7**, 4MU-GlcUA, or iminosugar **8**. Labeling also was abolished by SDS denaturation of protein, in line with a mechanism-based mode of action requiring catalytically competent enzyme (**Supplementary Fig. 3a**).

To dissect the mechanistic mode of action of our probes, we obtained crystal structures of **5** in complex with wild type AcGH79, and an inactive AcGH79(E287Q) nucleophile mutant (**Supplementary Fig. 3b, c**). Both complexes showed a single molecule of **5** bound within the active site of AcGH79, with no labeling of off-target residues. In wild type AcGH79, reacted **5** was observed bound via C1 to the enzyme nucleophile (Glu287) in a $^{4}C_{1}$ conformation, making identical non-covalent contacts as previously observed for GlcUA or 2F-GlcUA[19]. In the AcGH79(E287Q) mutant, **5** occupied the same active site position, but was instead found to adopt a $^{4}H_{3}$ conformation, due to restricted rotation across the C1-C7 bond imposed by the aziridine. Notably, the $^{4}H_{3}$ conformation observed for unreacted **5** is the same as that postulated for oxocarbenium-like transition states of retaining β-glycosidase substrates during hydrolysis (**Supplementary Fig. 3d**)[21]. The high affinities of cyclophellitol derived ABPs for their target enzymes may thus be in part due to their conformational mimicry of this transition state[22].

ABP profiling reveals GUSB and HPSE as probe targets

To determine the targets of retaining β-glucuronidase ABPs in complex biological samples, human splenic lysates (which we have previously shown to express a range of glycosidases[10-12]) were treated with one or more ABPs, resolved by SDS-PAGE, and labeled proteins visualized by fluorescent scanning (the typical ABP workflow is shown in **Supplementary Fig. 4**).

Several fluorescent bands were observed in samples treated with Cy5 ABP **3** which were absent in a mock (DMSO) control, and which could be competed for by biotin ABP **4** (**Fig. 2a**). Based on literature reports, we tentatively assigned the prominent double bands at ~78–80 kDa as full length and C-terminal truncated isoforms of GUSB[23], and the lowest molecular weight band as the ~64 kDa isoform of GUSB[24] ; these bands were also identified by an anti-GUSB western blot (**Supplementary Fig. 5a**). A band at ~60 kDa did not correspond with any known glucuronidases but could be abrogated by pretreatment with β-glucosidase ABP **9**, suggesting this was the lysosomal acid β-glucosidase GBA, which is specifically labeled by **9**[12]. Correspondingly, this ~60 kDa band was also absent in splenic lysates from patients with Gaucher disease, which is characterized by lack of GBA activity.

To unambiguously establish the targets of our ABPs, we carried out a set of chemical proteomics experiments using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Lysates from normal or Gaucher spleens were incubated with biotin ABP **4**, with or without pretreatment using **9**. Labeled proteins were then pulled down using streptavidin beads, and processed using a standard proteomics workflow, utilizing both 'in-gel' and 'on-bead' digest

protocols (**Supplementary Fig. 4**). The 'on bead' digest protocol was also applied to human fibroblast lysates, in order to assess the efficacy of ABP pulldown in another tissue type.

Proteomic profiling using the 'in-gel' digest protocol identified GUSB as the predominant splenic protein labeled by **4**, with particular enrichment in bands corresponding to those previously identified by **3** (**Fig. 2b**). As with fluorescent labeling, we detected GBA at ~60 kDa, which was abrogated by pretreatment with **9** and reduced in Gaucher spleen. 'In-gel' proteomic results were largely mirrored by those from the 'on bead' protocol, which showed GUSB to be the most abundant glycosidase after pulldown from spleen, and also highly enriched after pulldown from fibroblast (**Supplementary Data Set 1**). Covalent modification of the GUSB nucleophile (Glu540) by **4** was directly characterized by MS/MS fragmentation of the 13 amino acid peptide containing this residue (**Supplementary Fig. 5b**).

The high sensitivity of proteomic profiling enabled detection of two glycosidase enzymes in splenic lysates not observed by fluorescent labeling with **3**: β-galactosidase GLB1, and *endo* β-glucuronidase HPSE. Both enzymes were substantially less abundant in pull-down fractions compared to GUSB, as estimated by their exponentially modified protein abundance index (emPAI) scores (**Supplementary Fig. 5c**). We theorized that GLB1 was likely a weak non-specific target of **4**, as it showed the lowest emPAI of all detected glycosidases despite robust transcriptional expression reported in spleen (**Supplementary Fig. 5d**)[25], and strong histochemical staining of β-galactosidase activity in mammalian splenic tissues[26]. In contrast, HPSE is predicted to be poorly expressed in spleen by transcriptomics, yet showed a higher emPAI score than GLB1 after pull-down by **4**. These observations suggested HPSE was a *bona fide* target of our ABPs, despite the inability of the monosugar probes to make a full complement of interactions within the HPSE *endo*-acting substrate cleft, which normally accommodates at least a trisaccharide[27].

β-glucuronidase ABPs label endogenous HPSE and proHPSE

Following the surprising discovery that splenic HPSE was labeled by **4**, we re-examined fluorescent labeling of HPSE in cells and tissues expressing higher levels of HPSE. In the first instance, we induced HPSE overexpression in HEK293T cells, and probed harvested lysates at set intervals. Two plasmids were tested: pGEn1-HPSE, encoding for N-His-Strep-TEV tagged proHPSE, and pGEn2-HPSE, encoding for N-His-Avitag-eGFP-TEV tagged proHPSE; HEK293T cells subsequently process these proHPSE precursors to mature HPSE.

Using **3**, we tracked increasing expression of a band at ~50 kDa, corresponding to the mass of the large HPSE subunit, which contains the nucleophile Glu343 (**Fig. 3a**)[28]. Unexpectedly, we also detected bands at ~75 kDa for pGEn1-HPSE and ~100 kDa for pGEn2-HPSE, corresponding to the masses of proHPSE chains expressed by these plasmids. Western blotting using an anti-HPSE antibody confirmed these bands to be proHPSE, indicating that **3** also labeled the supposedly inactive proenzyme. Initial comparison of band intensities from ABP labeling using **3** *vs.* western blotting suggested that **3** labeled mature HPSE with greater efficiency than it labeled proHPSE. However, the opposite trend was observed when using purified recombinant proHPSE (**Supplementary Fig. 6a**), suggesting that the majority of proHPSE overexpressed by HEK293T cells was likely inactive.

HPSE maturation is thought to be mediated by the cysteine protease cathepsin L (CTSL), through multiple proteolytic cleavages of the 6 kDa linker peptide[29]. We tested the effect of CTSL inhibitors CAA0225[30] (**10**), Z-FY(tBu)-DMK[31] (**11**) and leupeptin on proHPSE maturation in HEK293T cells by incubating cells transfected pGEn1-HPSE with CTSL inhibitors for 2 d and labeling lysates with **3**. In inhibitor treated cells, we observed modest but dose dependent accumulation of a band ~5 kDa below the principal proHPSE band, corresponding to loss of the unstructured N-His-Strep-TEV tag followed by blockade of further proteolysis, in line with the role of CTSL in HPSE maturation (**Supplementary Fig. 6b**). However, accumulation of mature HPSE was still detected in the presence of all inhibitors tested, suggesting either incomplete CTSL inhibition, or the presence of non-cathepsin mediated HPSE maturation pathways also utilized by HEK293T cells.

We also tracked internalization and processing of proHPSE to mature HPSE in fibroblasts, a process which may be utilized by cancer cells to increase their own levels of HPSE: by capturing and internalizing extracellular proenzyme[32]. Within 90 minutes of introducing proHPSE into culture medium, internalization and processing of HPSE by fibroblasts was detectable using **3** (**Fig. 3b**). These experiments demonstrate the ability of β-glucuronidase ABPs to detect and track key biological processes such as HPSE internalization and maturation.

Lastly, we reattempted fluorescent labeling of endogenous HPSE in human tissues. As splenic HPSE expression was below the fluorescent detection limit, we turned to platelets, which are known to contain high levels of mature HPSE[33]. Using **3**, we observed labeling of a band at ~50 kDa in platelet lysates corresponding to HPSE, in addition to the same GUSB bands previously detected in spleen (**Fig. 3c**). Comparison of fluorescence and western blotting intensities between platelet HPSE and a 200 fmol recombinant standard suggested fluorescent sensitivity for recombinant HPSE to be in the fmol range, somewhat more sensitive than western blotting in our hands. Fluorescent HPSE detection in platelet lysates was slightly less sensitive, possibly due to the presence of competing protein targets or inactive HPSE *in situ*. 10 nM of **3** was sufficient to produce a detectable HPSE signal in platelets after 30 minutes (**Supplementary Fig. 6c**). Labeling of HPSE, but not GUSB, was also improved by the addition of NaCl (**Supplementary Fig. 6d**). Optimum labeling of HPSE by **3** was achieved at pH 4.5–5.0, consistent with literature reports of its optimum pH for enzymatic activity[34]. In contrast, optimum pH for labeling GUSB was higher than expected at pH ~5.5–6.0 (**Fig. 3d**), compared to its reported optimum for activity at the lysosomal pH ~4.5[35]. This unexpected pH of GUSB ABP labeling may be due to facile aziridine ring opening occurring independently of a protonated acid/base residue[36]. However, optimum labeling of GUSB at a non-lysosomal pH presents its own serendipitous advantages, allowing both GUSB and HPSE to be analyzed either jointly or independently of each other through modulation of labeling pH.

Competitive ABP labeling identifies HPSE specific inhibitors

Because ABPs can detect a complete complement of enzymes in a cellular/environmental sample, competitive ABP labeling provides a powerful tool to assess inhibitor efficacy and specificity within a single experiment[37]. We sought to establish whether GlcUA ABPs could be used for the assessment of enzyme specific inhibitors, by testing platelet labeling at pH 5.0 (where both GUSB and HPSE react) in the presence of a set of known inhibitors.

In the presence of the monosaccharide-like β-glucuronidase inhibitor siastatin B[38], both GUSB and HPSE labeling were abrogated in a dose-dependent manner, demonstrating the ability of this

molecule to outcompete ABP binding in both *endo*- and *exo*- acting enzymes. Using quantitated band intensities, IC50s for GUSB and HPSE labeling inhibition were measured to be ~3.3 μM and ~6.7 μM respectively, indicating slightly greater affinity for GUSB by siastatin B (**Fig. 4a**).

In contrast, HPSE labeling in platelets was selectively inhibited by competition with heparin (**12**, IC50 ~0.17 mg/mL, **Fig. 4b**), a large polysaccharide which cannot be accommodated by the *exo*- acting active site of GUSB. Selective inhibition was also observed upon competition with HS (**13**), the substrate of HPSE, albeit with slightly lower potency (**Fig. 4c**, IC50 ~0.50 mg/mL). Negligible inhibition was observed for N-Acetyl-O-desulfated heparin (**14**) (**Supplementary Fig. 7a**), highlighting the importance of sulfation for interactions between heparin/HS and HPSE. A lower degree of sulfation in HS *vs.* heparin may partly account for its slightly weaker abrogation of ABP labeling[39].

We next tested labeling inhibition by GAGs with different linkages and sulfation patterns to heparin and HS. Hyaluronic acid (**15**) and chondroitin sulfate (**16**) both showed no inhibition of either HPSE or GUSB labeling at concentrations sufficient for inhibition by heparin/HS (**Supplementary Fig. 7b, c**), highlighting the critical role of sugar linkage and sulfation in interactions between GAGs and HPSE. Taken together, these assays provide proof of principle that GlcUA ABPs are amenable for use in a competitive format, to assess inhibition of specific β-glucuronidases within a mixture of related activities.

Structural basis of HPSE and proHPSE ABP labeling

To investigate how efficient labeling of *endo*-acting HSPE was achieved by a monosugar ABP, we obtained crystal structures of both wild type and nucleophile mutant (E343Q) HPSE in complex with ABP **5**. Complexes of HPSE with **5** were similar to those obtained with AcGH79, showing a single molecule of probe occupying the –1 subsite of the HPSE substrate binding cleft (nomenclature according to Ref. 40) in reacted $^4C_1$ (wild type) or unreacted $^4H_3$ (mutant) conformations (**Supplementary Fig. 8a**). The network of interactions made to the probe was highly similar between AcGH79 and HPSE, with the primary difference being a lack of interaction by HPSE to O4 of the probe, due to extension of its natural HS substrate towards this position (**Supplementary Fig. 8b**). A –1 subsite C6 carboxylate recognition motif, comprising 3 H-bonds from a tyrosine and two consecutive backbone amides, is highly conserved in GH79 β-glucuronidases (Tyr 334, Gln293-Gly294 in AcGH79; Tyr391, Gly349-Gly350 in HPSE; Tyr 302, Gly261-Gly262 in the recently characterized heparanase from *Burkholderia pseudomallei*)[41]. This strong network of H-bonds to C6 carboxylate likely offsets the absence of only a single H-bond to O4 of the ABP in HPSE compared to AcGH79, thus rationalizing robust labeling of HPSE by a monosaccharide probe that only occupies a single subsite within its extensive binding cleft. Additional binding affinity may also derive from the transition state like $^4H_3$ conformation adopted by unreacted ABPs.

We next sought to solve the structure of proHPSE, in order to characterize the basis of its 'inactivation' by the 6 kDa linker peptide, and to determine how β-glucuronidase ABPs are able to circumvent this. Herein, we report the first crystal structure of proHPSE in both apo and ABP complexed forms, which together with previously reported HPSE structures completes a structural characterization of the HPSE maturation process.

The proHPSE structure was similar to that of mature HPSE (RMSD: 0.52 Å over 451 Cα), with the same $(\beta/\alpha)_8$ and β-sheet domains clearly discernible. The 6 kDa linker (110-157) forms a large helical

domain which sits directly 'above' the active site cleft, blocking access to the bulky HS substrates of HPSE. The final loop of the linker leading into the 50 kDa subunit (His155-Lys159) is substantially more disordered than the rest of the protein, as evidenced by higher B-factors for these residues in the crystallographic model (**Supplementary Fig. 9a, b**). Mutation studies have established Tyr156 of the proHPSE linker to be critical for recognition by CTSL in the first step of HPSE maturation[42]. Disorder of the His155-Lys159 loop allows for unencumbered CTSL access to Tyr156 without disrupting preexisting secondary structures, consistent with the important role of Tyr156 in HPSE maturation.

Unexpectedly, steric blockage by the linker peptide was found to be incomplete in proHPSE, leaving a 'binding pocket' on the protein surface containing exposed catalytic nucleophile and acid/base residues, similar to the *exo*-acting active site of AcGH79 (**Fig. 5a**). When compared in a sequence alignment, the proHPSE linker corresponds to a loop in AcGH79, which forms the 'rear' face of its active site pocket, suggesting that whilst this sequence has expanded in the human enzyme, proHPSE still retains some structural characteristics reminiscent of a GH79 *exo*-glycosidase (**Supplementary Fig. 8c**). ABP **5** was found to bind to this proHPSE 'active-site pocket' in a configuration identical to that observed for HPSE. The O4 proximal position, vacant in HPSE, was occupied by His155 in proHPSE, contributed by the linker, which blocks off extension towards this position by HS substrates (**Fig. 5b, c**). The disordered proHPSE His155-Lys159 loop was slightly displaced upon binding **5** (~1.54 Å for Tyr156 Cα), due to steric clashes with the bound ABP (**Supplementary Fig. 9c**).

As with mature HPSE, proHPSE was inactive against the artificial fluorogenic substrate 4MU-GlcUA, indicating it does not possess any additional *exo*-glucuronidase activity against this substrate which is lost upon maturation (**Supplementary Fig. 8d**). To assess the accessibility of the proHPSE 'pocket' compared to mature HPSE, we conducted competitive ABP experiments against recombinant proHPSE and HPSE using **3**. As with platelets, siastatin B inhibited ABP labeling of both pro- and mature HPSE (**Supplementary Fig. 10a**), indicating it could efficiently occupy the 'binding pocket' of proHPSE as well as HPSE. In contrast, heparin only inhibited labeling of HPSE, albeit with lower efficacy than seen in platelets, due to more facile labeling of the recombinant enzyme. Unexpectedly, proHPSE labeling was slightly increased at moderate heparin concentrations (**Supplementary Fig. 10b**). Finally, we tested the ability of GlcUA to inhibit labeling of HPSE and proHPSE. No substantial inhibition of either proHPSE or HPSE labeling was observed at up to 20 mM GlcUA (**Supplementary Fig. 10c**), suggesting that GlcUA cannot occupy the active site of (pro)HPSE with sufficient affinity to prevent binding and reactivity of an ABP. Further subsite interactions may be required for binding of simple glucuronides to (pro)HPSE.

It has previously been demonstrated that proHPSE uptake by cells is a HS dependent process, and can be disrupted by addition of exogenous heparin[43]. To investigate possible roles for the proHPSE 'binding pocket' in proHPSE uptake and maturation, we prelabeled recombinant proHPSE with either untagged ABP **6**, fluorescent ABPs **1** or **3**, or a mock DMSO control, and examined its uptake by fibroblasts at 90 or 180 min. In all cases, prelabeled proHPSEs were taken up and processed to mature HPSE, as evidenced by western blot and fluorescence of internalized **1** or **3** (**Supplementary Fig. 11**). These data indicate that the proHPSE 'pocket' does not participate in the HS interactions involved in cellular uptake, and that occupation of the proHPSE 'pocket' does not inhibit HPSE maturation.

Discussion

The important role of β-glucuronidases in human biology is highlighted by the pathologies associated with aberrant expression of these enzymes. Lack of GUSB activity leads to accumulation of glucuronide-containing GAGs within lysosomes in MPSVII (Sly Syndrome). Conversely, overexpression of HPSE leads to aberrant breakdown of HS in the ECM, causing increased cancer growth and metastasis. Accurate tracking of β-glucuronidase activities is an essential prerequisite for fully understanding their role in both physiological and disease states.

Here we have reported the design and application of novel β-glucuronidase configured ABPs, and demonstrated their broad utility for interrogating activities of these enzymes. We show that ABP profiling is a viable method to assay β-glucuronidase activity in a variety of samples, ranging from recombinant proteins, to complex cell, tissue and organ lysates. Fluorescent labeling provides a facile method for probing β-glucuronidases in tissues with sufficient expression, allowing for tracking of processes such as proenzyme uptake and processing, and how these are affected by biological or pharmacological perturbation. In tissues with lower enzyme abundance, we have demonstrated detection of β-glucuronidases using a proteomic approach, which is also applicable for the discovery of previously uncharacterized β-glucuronidase activities in biological samples.

Use of ABPs provides several advantages over more traditional methods to quantitate glycoside hydrolase activities. Compared to techniques such as western blotting, ABPs specifically detect active enzymes, rather than an entire protein complement which may include misfolded or inactive isoforms. Whilst fluorometric or colorimetric assays also provide assessments of enzyme activity, they cannot distinguish between overlapping activities in complex mixtures, which arise from several enzymes or enzyme isoforms active on the same substrate. Indeed, many carbohydrate processing enzymes are processed from precursors into one or more isoforms with differing activities[23,28,44,45]. ABP profiling allows for multiple activities to be visualized and their responses to perturbation or inhibition to be individually assessed *in situ*.

Many *endo*-glycosidases such as HPSE are inactive in traditional activity assays, necessitating the use of expensive specialized substrates and/or cumbersome assay procedures to follow their activities. The discovery that aziridine ABPs label HPSE paves the way for more rapid and practicable methods to assess the activity of this enzyme, and may inspire development of probes to assay other *endo*-glycosidases. Whilst this current generation of β-glucuronidase ABPs shows some off-target effects against GBA and GLB1, limiting their use in diagnostic applications, further optimization based upon the crystal structures of HPSE (and proHPSE) may lead to improved probes with increased potency and specificity. Optimization efforts will be aided by the use of competitive ABP techniques, which we have demonstrated to be a viable method for assessing selective inhibitors of individual β-glucuronidases.

ABPs also provide powerful tools for characterization of novel enzyme activities, which may escape detection in traditional biochemical experiments. The use of an ABP driven approach in this study lead us to the surprising observation that the HPSE precursor proHPSE is in principle catalytically competent, an entirely unanticipated outcome based on previous studies. We have reported the first structural views of proHPSE, illustrating how its 6 kDa linker restricts access to the active site cleft for HS substrates. This linker does not entirely block access to the catalytic residues of

proHPSE, but instead contributes to the formation of an *exo*-glycosidase like 'binding pocket', which can accommodate smaller molecules. It remains to be determined whether this proHPSE 'pocket' is simply a structural relic from evolutionary expansion of an ancestral GH79 active site loop, or if there are *bona fide* endogenous substrates which are hydrolyzed by proHPSE.

In conclusion, we have presented a set of ABPs for functional interrogation of β-glucuronidases in their native contexts. The application of ABP methodology to carbohydrate processing enzymes provides a powerful set of tools to study the activity of these key enzymes, and will contribute towards our understanding of fundamental processes in glycobiology.

Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5G0Q (AcGH79(wt)-**5** complex), 5L77 (AcGH79(E287Q)-**5** complex), 5L9Y (HPSE(wt)-**5** complex), 5L9Z (HPSE(E343Q)-**5** complex), 5LA4 (*apo* proHPSE), 5LA7 (proHPSE-**5** complex).

Acknowledgements

We thank Diamond Light Source for access to beamlines i02, i03 and i04 (proposals mx-9948 and mx-13587), which contributed to the results presented here. We acknowledge the Netherlands Organization for Scientific Research (NWO, ChemThem Grant to J.M.F.G.A. and H.S.O.), the China Scholarship Council (CSC, PhD Grant to J.J.), the European Research Council (ErC-2011-AdG-290836 to H.S.O.; ErC-2012-AdG-322942 to G.J.D.), and the Royal Society (Ken Murray Research Professorship to G.J.D) for financial support.

Author contributions

L.W., J.M.F.G.A., H.S.O. and G.J.D. conceived and designed the experiments. J.J., M.A., W.D. and C.v.E. carried out synthesis of probes, with guidance from G.A.v.d.M. and J.D.C.C.. L.W. and Y.J. carried out protein expression and structural studies on enzyme-probe complexes. J.J., L.W., W.W.K. and C-L.K. carried out gel labeling experiments. J.J. and B.I.F. carried out proteomics experiments. C-L.K. and W.W.K. determined IC50 and kinetic parameters for ABPP inhibition. M.v.E. obtained tissue samples. L.W., J.J., H.S.O., and G.J.D. wrote the manuscript with input from all authors.

Competing financial interests

The authors declare no competing financial interests.

Corresponding authors

Correspondence to Gideon Davies or Herman Overkleeft.

Methods and Supplementary Information

Supplementary results containing Supplementary Tables 1 and 2, Supplementary Figures 1–12, Supplementary Notes 1 and 2 and Supplementary Data Set 1 are available in the online version of this paper.

Main text references

1. Khan, F.I. et al. Large scale analysis of the mutational landscape in beta-glucuronidase: A major player of mucopolysaccharidosis type VII. *Gene* **576**, 36-44 (2016).
2. Sly, W.S., Quinton, B.A., McAlister, W.H. & Rimoin, D.L. Beta glucuronidase deficiency: report of clinical, radiologic, and biochemical features of a new mucopolysaccharidosis. *J. Pediatr.* **82**, 249-257 (1973).
3. Naz, H. et al. Human beta-glucuronidase: structure, function, and application in enzyme replacement therapy. *Rejuvenation Res.* **16**, 352-363 (2013).
4. Rivara, S., Milazzo, F.M. & Giannini, G. Heparanase: a rainbow pharmacological target associated to multiple pathologies including rare diseases. *Future Med. Chem.* **8**, 647-680 (2016).
5. Vlodavsky, I. et al. Significance of heparanase in cancer and inflammation. *Cancer Microenviron.* **5**, 115-132 (2012).
6. Cantarel, B.L. et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic. Acids. Res.* **37**, D233-238 (2009).
7. Jain, S. et al. Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat. Struct. Biol.* **3**, 375-381 (1996).
8. Wu, L., Viola, C.M., Brzozowski, A.M. & Davies, G.J. Structural characterization of human heparanase reveals insights into substrate recognition. *Nat. Struct. Mol. Biol.* **22**, 1016-1022 (2015).
9. Jiang, J. et al. Detection of Active Mammalian GH31 alpha-Glucosidases in Health and Disease Using In-Class, Broad-Spectrum Activity-Based Probes. *ACS Cent. Sci.* **2**, 351-358 (2016).
10. Kallemeijn, W.W. et al. Novel activity-based probes for broad-spectrum profiling of retaining beta-exoglucosidases in situ and in vivo. *Angew. Chem. Int. Ed.* **51**, 12529-12533 (2012).
11. Willems, L.I. et al. Potent and selective activity-based probes for GH27 human retaining alpha-galactosidases. *J. Am. Chem. Soc.* **136**, 11622-11625 (2014).
12. Witte, M.D. et al. Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat. Chem. Biol.* **6**, 907-913 (2010).
13. Jiang, J. et al. In vitro and in vivo comparative and competitive activity-based protein profiling of GH29 alpha-l-fucosidases. *Chem. Sci.* **6**, 2782-2789 (2015).
14. Kwan, D.H. et al. Chemoenzymatic synthesis of 6-phospho-cyclophellitol as a novel probe of 6-phospho-beta-glucosidases. *FEBS Lett.* **590**, 461-468 (2016).
15. Barglow, K.T. & Cravatt, B.F. Activity-based protein profiling for the functional annotation of enzymes. *Nat. Methods* **4**, 822-827 (2007).
16. Li, K.Y. et al. Exploring functional cyclophellitol analogues as human retaining beta-glucosidase inhibitors. *Org. Biomol. Chem.* **12**, 7786-7791 (2014).
17. Koshland, D.E. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* **28**, 416-436 (1953).
18. Willems, L.I. et al. From covalent glycosidase inhibitors to activity-based glycosidase probes. *Chemistry* **20**, 10864-10872 (2014).
19. Michikawa, M. et al. Structural and biochemical characterization of glycoside hydrolase family 79 beta-glucuronidase from Acidobacterium Capsulatum. *J. Biol. Chem.* **287**, 14069-14077 (2012).
20. Baici, A., Schenker, P., Wachter, M. & Ruedi, P. 3-Fluoro-2,4-dioxa-3-phosphadecalins as inhibitors of acetylcholinesterase. A reappraisal of kinetic mechanisms and diagnostic methods. *Chem. Biodivers.* **6**, 261-282 (2009).
21. Speciale, G., Thompson, A.J., Davies, G.J. & Williams, S.J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1-13 (2014).

22. Gloster, T.M. & Davies, G.J. Glycosidase inhibition: assessing mimicry of the transition state. *Org. Biomol. Chem.* **8**, 305-320 (2010).
23. Islam, M.R., Grubb, J.H. & Sly, W.S. C-terminal processing of human beta-glucuronidase. The propeptide is required for full expression of catalytic activity, intracellular retention, and proper phosphorylation. *J. Biol. Chem.* **268**, 22627-22633 (1993).
24. Ono, M. et al. Phosphorylation of beta-glucuronidases from human normal liver and hepatoma by cAMP-dependent protein kinase. *J. Biol. Chem.* **263**, 5884-5889 (1988).
25. Forrest, A.R. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462-470 (2014).
26. Weiss, D.J., Liggitt, D. & Clark, J.G. Histochemical Discrimination of Endogenous Mammalian beta-galactosidase Activity from that Resulting from lac-Z Gene Expression. *Histochem. J.* **31**, 231-236 (1999).
27. Peterson, S.B. & Liu, J. Multi-faceted substrate specificity of heparanase. *Matrix Biol.* **32**, 223-227 (2013).
28. Fairbanks, M.B. et al. Processing of the human heparanase precursor and evidence that the active enzyme is a heterodimer. *J. Biol. Chem.* **274**, 29587-29590 (1999).
29. Abboud-Jarrous, G. et al. Cathepsin L is responsible for processing and activation of proheparanase through multiple cleavages of a linker segment. *J. Biol. Chem.* **283**, 18167-18176 (2008).
30. Takahashi, K. et al. Characterization of CAA0225, a novel inhibitor specific for cathepsin L, as a probe for autophagic proteolysis. *Biol. Pharm. Bull.* **32**, 475-479 (2009).
31. Shaw, E., Mohanty, S., Colic, A., Stoka, V. & Turk, V. The affinity-labelling of cathepsin S with peptidyl diazomethyl ketones. Comparison with the inhibition of cathepsin L and calpain. *FEBS Lett.* **334**, 340-342 (1993).
32. Nadav, L. et al. Activation, processing and trafficking of extracellular heparanase by primary human fibroblasts. *J. Cell Sci.* **115**, 2179-2187 (2002).
33. Vlodavsky, I. et al. Expression of heparanase by platelets and circulating cells of the immune system: possible involvement in diapedesis and extravasation. *Invas. Metast.* **12**, 112-127 (1992).
34. McKenzie, E. et al. Biochemical characterization of the active heterodimer form of human heparanase (Hpa1) protein expressed in insect cells. *Biochem. J.* **373**, 423-435 (2003).
35. Islam, M.R. et al. Active site residues of human beta-glucuronidase. Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. *J. Biol. Chem.* **274**, 23451-23455 (1999).
36. Kallemeijn, W.W. et al. A sensitive gel-based method combining distinct cyclophellitol-based probes for the identification of acid/base residues in human retaining beta-glucosidases. *J. Biol. Chem.* **289**, 35351-35362 (2014).
37. Niphakis, M.J. & Cravatt, B.F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341-377 (2014).
38. Kawase, Y. et al. A-72363 A-1, A-2, and C, novel heparanase inhibitors from Streptomyces nobilis SANK 60192, II. Biological activities. *J. Antibiot. (Tokyo)* **49**, 61-64 (1996).
39. Rabenstein, D.L. Heparin and heparan sulfate: structure and function. *Nat. Prod. Rep.* **19**, 312-331 (2002).
40. Davies, G.J., Wilson, K.S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321 (Pt 2)**, 557-559 (1997).
41. Bohlmann, L. et al. Functional and structural characterization of a heparanase. *Nat. Chem. Biol.* **11**, 955-957 (2015).
42. Abboud-Jarrous, G. et al. Site-directed mutagenesis, proteolytic cleavage, and activation of human proheparanase. *J. Biol. Chem.* **280**, 13568-13575 (2005).
43. Gingis-Velitski, S. et al. Heparanase uptake is mediated by cell membrane heparan sulfate proteoglycans. *J. Biol. Chem.* **279**, 44084-44092 (2004).

44. Moreland, R.J. et al. Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor. *J. Biol. Chem.* **280**, 6780-6791 (2005).
45. Tollersrud, O.K. et al. Purification of bovine lysosomal alpha-mannosidase, characterization of its gene and determination of two mutations that cause alpha-mannosidosis. *Eur. J. Biochem.* **246**, 410-419 (1997).

Figure 1 Concept and design of β-glucuronidase targeting ABPs. **a** Generalized schematic of double-displacement mechanism employed by retaining β-glycosidases. **b** Mechanism based inhibition by cyclophellitol derived ABPs. **c** Structures of cyclophellitol, cyclophellitol aziridine, and β-glucuronidase specific ABPs used in this study. Numbering of atomic positions is shown on cyclophellitol.

Figure 2 ABP labeling of retaining β-glucuronidases in human spleen lysates. **a** Three isoforms of GUSB are fluorescently labeled by Cy5 ABP **3** in human wild type spleen, along with off target labeling of the β-glucosidase GBA. Labeling of these proteins by **3** can be competed for by biotin ABP **4**. GBA labeling in wild type spleen is also specifically competed for by **9**, and is absent in lysates from Gaucher spleen. **b** Silver stained SDS-PAGE gel of proteins captured from human wild type spleen by labeling with **4** (with or without competition by **9**) followed by streptavidin pulldown. Glycosidase enzymes identified in each gel band by proteomic profiling are listed. Full proteomics datasets for proteins identified by ABP pulldowns are available in **Supplementary Data Set 1**.

Figure 3 Human HPSE is readily visualized by fluorescent β-glucuronidase ABPs. **a** Induced overexpression of HPSE and proHPSE in HEK293T cells can be tracked by ABP **3**. Fluorescent labeling by **3** correlates with bands from western blotting using an anti-HPSE antibody. **b** ABP tracking of uptake and processing of proHPSE to HPSE by fibroblast cells. **c** Endogenous HPSE in human platelets can be labeled by **3**, along with the same GUSB bands as observed in spleen. **d** pH dependence of HPSE and GUSB labeling in platelet lysates, demonstrating how general or specific enzyme labeling can be achieved by modulating pH.

Figure 4 General and endo-specific inhibition of β-glucuronidases assessed by competitive ABP profiling. **a** Monosugar like β-glucuronidase inhibitor siastatin B can be accommodated *exo*- and *endo*- acting β-glucuronidase active sites, and competes out ABP **3** labeling of both GUSB and HPSE. **b** Polysaccharide heparin (**12**) only inhibits ABP labeling of HPSE, due to its inability to interact with the *exo*- configured active site of GUSB. **c** Selective HPSE inhibition is also achieved by heparan sulfate (**13**). Competitive ABP gels shown are representative of three technical replicates. Plots are mean values ± s.d. (N=3) for quantitated HPSE and GUSB fluorescent band intensities, normalized to band intensities in the no inhibitor control lane. For all plots, quantitated GUSB fluorescence is a sum of the three assigned bands. n.d.: not determined.

Figure 5 3-dimensional structure of proHPSE, and its active site interactions with ABP **5**. **a** Ribbon and surface diagram of proHPSE, demonstrating steric blockage of the HPSE binding cleft by the 6 kDa linker. An exposed 'pocket' in proHPSE can still interact with small molecules such as **5** (highlighted pink for clarity). **b** ABP **5** in complex with proHPSE within its 'binding pocket'. The O4 position, where HS substrates would extend in mature HPSE, is blocked by His155, contributed by the linker (colored in green). Density is REFMAC maximum-likelihood/σ_A weighted 2Fo−Fc contoured to 0.38 electrons/Å^3. **c** Schematic of H-bonding interactions between reacted **5** and proHPSE active site residues. Interactions are identical to those observed for the mature enzyme (**Supplementary Fig. 8a, b**), except for His155 proximal to O4 of the probe. (nuc.: nucleophile; a/b: acid base).

Table 1 Apparent IC50 values for *in vitro* and *in situ* inhibition of β-glucuronidase activity by ABPs, and kinetic parameters for inhibition of AcGH79 by ABPs. Data are mean values ± s.d. (N=3) from three biological replicates.

Compound	*In vitro* AcGH79 apparent IC50 (nM)	Kinetic Parameters (AcGH79) k_i/K_I ($\mu M^{-1}min^{-1}$)	*In situ* fibroblast apparent IC50 (μM)
1	0.6±0.2	25.0±0.7	>15
2	0.8±0.2	18.2±0.9	1.7±0.6
3	1.1±0.1	14.0±0.8	1.8±0.4
4	0.4±0.02	5.5±0.2	>15
5	0.1±0.01	18.8±0.7	>15
6	4.6±0.03	3.5±0.2	>15
7	33.4±3.1	0.49±0.05	>15

Methods

Chemical probes and inhibitors

4MU-GlcUA, Leupeptin, Siastatin B, Hyaluronic Acid (**15**) and Chondroitin Sulfate (**16**) were obtained from Sigma Aldrich. CTSL inhibitors CAA0225 (**10**) and Z-FY(tBu)-DMK (**11**) were obtained from Merck. Heparin (**12**) and Heparan Sulfate (**13**) were obtained from Iduron. *N*-Acetyl-*O*-desulfated Heparin (**14**) was obtained from Dextra. Cyclophellitol, cyclophellitol aziridine[46] and ABP **9**[10] were synthesized according to described procedures. Syntheses of compounds **1–8** are described in **Supplementary Note 2**.

Tissue and cell samples

Gaucher patients were diagnosed on the basis of reduced GBA activity and demonstration of an abnormal genotype. Spleens from a normal subject and a patient suffering from type 1 Gaucher disease were collected after splenectomy and frozen at –80 °C until use. Platelets were collected from healthy donors, using EDTA as the anti-coagulant. Platelet rich plasma (PRP) was prepared by centrifugation at 100 g for 20 min at 22 °C to remove red and white blood cells. Platelets were isolated from PRP by centrifugation at 220 g for 10 min at 22 °C, and frozen at –80 °C until use. Approval for tissue collection was obtained from the Academisch Medisch Centrum (AMC) and University of York medical ethics committees. Informed consent was obtained from all donors.

Primary human fibroblasts (CC-2511) were obtained from Lonza. HEK293T cells (ATCC-CRL-3216) were obtained from the American Type Culture Collection (ATCC). Sf21 and High Five cells for protein production were obtained from Invitrogen. Cells were used as obtained from the supplier without further authentication. All cells used tested negative for mycoplasma contamination.

All tissue lysates were prepared in KPI buffer (25 mM potassium phosphate [pH 6.5], supplemented with 1 × cOmplete protease inhibitor cocktail (Roche)). Cells/tissues were homogenized with a silent crusher S equipped with a type 7 F/S head (30,000 rpm, 3 × 7 sec) on ice. Lysate protein concentrations were determined with a Qubit 2.0 Fluorometer (Invitrogen) or Bradford assay using BSA as a standard. Lysates were stored in aliquots at –80 °C until use.

Recombinant protein cloning, expression and purification

AcGH79

The coding sequence of AcGH79 with an N-terminal 6×His tag was cloned into pET28a (Novagen), which was used to transform *E. coli* BL21-Gold(DE3) (Agilent). Transformants were grown at 37 °C in LB media containing 50 μg/mL kanamycin to an OD_{600} of 0.8, induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside, and protein production carried out out at 25 °C for 12 h. Harvested cells were resuspended in 50 mL AcGH79 HisTrap buffer A (20 mM HEPES [pH 7.0], 200 mM NaCl, 5 mM imidazole), lysed by sonication, and lysate clarified by centrifugation at 12000 g. Supernatant containing AcGH79 was filtered before loading onto a HisTrap 5 mL FF crude column (GE Healthcare) pre-equilibrated with AcGH79 HisTrap buffer A. The loaded HisTrap column was washed with 10 column volumes (CV) of AcGH79 HisTrap buffer A, before eluting with AcGH79 HisTrap buffer B (20 mM HEPES [pH 7.0], 200 mM NaCl, 400 mM imidazole) over a 20 CV linear gradient.

Fractions containing AcGH79 were pooled, concentrated using a 30 kDa cutoff Vivaspin concentrator (GE Healthcare) and further purified by size exclusion chromatography (SEC) using a Superdex 75 16/600 column (GE Healthcare) in AcGH79 SEC buffer (20 mM HEPES [pH 7.0], 200 mM NaCl). Fractions containing AcGH79 were pooled and concentrated using a 30 kDa Vivaspin concentrator to a final concentration of 14.5 mg/mL, and flash frozen for use in further experiments.

E287Q mutagenesis was carried out using a PCR based method[47]. Mutant protein was purified using the same protocol as for wild type protein. Mutagenesis primers are listed in **Supplementary Table 2**.

Mature HPSE

Mature HPSE cloning, expression, purification was carried out as previously described[8].

E343Q mutagenesis was carried out using a PCR based method. Mutant protein was purified using the same protocol as for wild type protein. Mutagenesis primers are listed in **Supplementary Table 2**.

proHPSE

Insect cells are unable to process proHPSE to mature HPSE, allowing the former to be isolated following expression. cDNA encoding for proHPSE, minus the first 35 amino acid codons comprising the native signal sequence, was cloned behind a 5' honeybee mellitin signal sequence, 6×His tag, and TEV cleavage site, into the pOMNIBac plasmid (Geneva Biotech) using SLIC[48]. pOMNIBac-proHPSE was used to generate recombinant bacmid using the Tn7 transposition method in DH10EMBacY cells[49] (Geneva Biotech). Baculovirus preparation and protein expression was carried out as previously described for mature HPSE.

For purification, 3 L of conditioned media was cleared of cells by centrifugation at 400 g for 15 min at 4 °C, followed by further clearing of debris by centrifugation at 4000 g for 60 min at 4 °C. DTT (1 mM) and AEBSF (0.1 mM) were added to cleared media, which was loaded onto a HiTrap Sepharose SP FF 5 mL column (GE healthcare) pre-equilibrated in IEX buffer A (20 mM HEPES [pH 7.4], 100 mM NaCl, 1 mM DTT). The loaded SP FF column was washed with 10 CV of IEX buffer A, and eluted with a linear gradient over 30 CV using IEX buffer B (20 mM HEPES [pH 7.4], 1.5 mM NaCl, 1 mM DTT). proHPSE containing fractions were pooled and diluted 10 fold into proHPSE HisTrap buffer A (20 mM HEPES [pH 7.4], 500 mM NaCl, 20 mM Imidazole, 1 mM DTT), before loading onto a HisTrap 5 mL FF crude column pre-equilibrated in proHPSE HisTrap buffer A. The loaded HisTrap column was washed with 10 CV HisTrap buffer A, and eluted with a linear gradient over 20 CV using proHPSE HisTrap buffer B (20 mM HEPES [pH 7.4], 500 mM NaCl, 1 M Imidazole, 1 mM DTT). proHPSE containing fractions were pooled and concentrated to ~2 mL using a 30 kDa cutoff Vivaspin concentrator, and treated with 5 μL EndoH (NEB) and 5 μL AcTEV protease (Invitrogen) for >72 h. Digested protein was purified by SEC using a Superdex S75 16/600 column in proHPSE SEC buffer (20 mM HEPES [pH 7.4], 200 mM NaCl, 1 mM DTT). proHPSE containing fractions were concentrated to 10 mg/mL using a 30 kDa Vivaspin concentrator, exchanged into IEX buffer A via at least 3 rounds of dilution/reconcentration, and flash frozen for use in further experiments.

Overexpression of HPSE in HEK293T cells

pGEn1-HPSE and pGEn2-HPSE plasmids were obtained from the DNASU repository[50]. HEK293T cells were grown in DMEM media supplemented with 10% newborn calf serum (NBCS; Sigma) and 1% penicillin/streptomycin (Sigma). 20 μg DNA was transfected into HEK293T cells at ~80% confluence, using linear PEI at a ratio of 3:1 (PEI:DNA). At relevant timepoints cells were washed with PBS, harvested into KPI buffer using a cell scraper, and pelleted by centrifugation at 200 g for 5 min at 4 °C. Cell pellets were frozen at –80 °C prior to use.

For CTSL inhibition experiments, transfections were carried out as above, except media was exchanged 7 h post transfection for DMEM supplemented with CTSL inhibitor or vehicle only control (0.05% v/v EtOH).

proHPSE fibroblast uptake experiment

Primary human fibroblasts were grown in DMEM/F12 media supplemented with 10% NBCS and 1% penicillin/streptomycin. Cells at 80% confluence were washed with PBS, before exchanging into DMEM/F12 media supplemented with proHPSE to 10 μg/mL final concentration. At relevant timepoints, cells were washed with ice cold PBS twice, harvested into KPI buffer using a cell scraper, and pelleted by centrifugation at 200 g for 5 min at 4 °C. Cell pellets were stored at –80 °C prior to use.

For uptake experiments with prelabeled proHPSEs, prelabeling was carried out in McIlvaine citrate/phosphate buffer [pH 5.0], 300 mM NaCl in 200 μL volume, using 50 μM proHPSE and 200 μM ABP. Reactions were incubated for 1 h for 37 °C, then excess ABP removed by desalting using a 40 kDa MWCO Zeba spin column (Thermo). Extent of prelabeling by **1** and **3** was quantified by comparison of protein and fluorophore UV/Vis absorption values. Extent of prelabeling by **6** was estimated by testing residual reactivity to **3** (**Supplementary Fig. 11b**).

Enzyme activity and inhibition assays

Recombinant AcGH79 enzyme activity was assayed using 1.67 ng protein in 150 mM McIlvaine buffer [pH 5.0]. To determine apparent IC50 values, 25 μL AcGH79 was preincubated with a range of inhibitor dilutions for 30 min at 37 °C, followed by addition of 100 μL 4MU-GlcUA solution to give final concentrations of 260 pM AcGH79 and 2.5 mM 4MUGlcUA. Reactions were carried out for 30 min at 37 °C, quenched with 200 μL of 1 M NaOH-glycine [pH 10.3], and 4-MU fluorescence measured using a LS55 Fluorometer (Perkin Elmer) at λ_{ex} 366 nm and λ_{em} 445 nm. Apparent IC50 values were determined in Prism (GraphPad) using a one phase decay function.

Kinetic parameters for inhibition of AcGH79 were determined using a continuous method (**Supplementary Note 1**; also Ref. 20). AcGH79 was added to pre-warmed mixtures of 4MU-GlcUA and ABP, to give final concentrations of 260 pM AcGH79 and 2.5 mM 4MU-GlcUA in a final reaction volume of 125 μL. Reactions were incubated at 37°C. At set timepoints, aliquots of reaction mixture were transferred to 96-well microplates (Greiner), quenched with 1 M NaOH-glycine [pH 10.3], and 4MU fluorescence measured immediately using a LS-55 Fluorometer. The apparent rate of inactivation (k_{obs}) was calculated for each ABP concentration by fitting with the exponential function [4MU]=A*(1-e^(–k_{obs}*t)). The resulting plot of k_{obs} *vs.* [ABP] was fitted using a linear function, which

gives the combined apparent inhibition parameter ki/K'_I as the gradient. k_i/K_I was derived from k_i/K'_I by correcting for the presence of competing 4MU-GlcUA substrate, using the relationship $K'_I=K_I(1+[S]/K_M)$, where [S] = 2.5 mM and K_M = 18.2 μM. All fittings were carried out using Prism.

In situ fibroblast IC50s were determined by incubating human fibroblast cells with a range of inhibitor dilutions for 2 hours, followed by 3 × washing with PBS and harvesting into KPI buffer supplemented with 0.1% Triton X-100. Harvested cells were pelleted by centrifugation at 200 g for 5 min at 4 °C, and pellets stored at –80 °C prior to use. Enzymatic reactions and apparent IC50 calculations were performed as described for the *in vitro* IC50 determination experiments, but with 5 μg of total lysate protein per reaction.

Fluorescent labeling

Initial abeling reactions were carried out in McIlvaine buffer [pH 5.0], except for pH range experiments, which were carried out in McIlvaine buffer at the stated pHs. Typically 200 fmol recombinant protein was used for labeling AcGH79, HPSE and proHPSE (20 nM in 10 μL final reaction volume). 20 μg total protein was used for labeling cell/tissue lysates, except for HEK293T overexpression experiments, where 10 μg total protein was used. Unless otherwise specified, labeling reactions were carried out by incubation with 100 nM fluorescent ABP in a reaction volume of 10 μL for 1 h at 37 °C. Gels were scanned for ABP-emitted fluorescence using a Bio-Rad ChemiDoc MP imager with the settings: Cy2 (λ_{EX} 470 nm, bandpass 30 nm; λ_{EM} 530 nm, bandpass 28 nm) for **1**, Cy3 (λ_{EX} 530 nm, bandpass 28 nm; λ_{EM} 605 nm, bandpass 50 nm) for **2**, and Cy5 (λ_{EX} 625 nm, bandpass 30 nm; λ_{EM} 695 nm, bandpass 55 nm) for **3**.

For labeling rate experiments with pro- and mature HPSE, reactions were carried out as above, except 2 pmol recombinant protein was incubated with an equimolar amount of **3** at 37 °C. At specified timepoints, aliquots were removed from the reaction and denatured by boiling in Laemmli buffer. Denatured samples were stored on ice until all timepoints were collected, and run together on SDS-PAGE.

For competition experiments, optimized labeling reactions were carried out in McIlvaine buffer [pH 5.0], 300 mM NaCl. Protein samples were preincubated with inhibitor for 60 min at 37 °C prior to addition of 100 nM **3** for labeling. Platelet lysates were labeled at 37 °C for 1 h, recombinant proteins were labeled at 37 °C for 30 min. Following labeling, samples were denatured by boiling with Laemmli buffer for 5 min, and resolved by SDS-PAGE. Gels used for quantitation were scanned using a laser based Bio-Rad FX molecular imager, using the λ_{EX} 635 nm external laser and 690BP emission filter. Images were analyzed using Quantity One (Bio-Rad). Full-length images of all fluorescent gels used in this study can be found in **Supplementary Fig. 12**.

Chemical proteomics

3 mg total protein from human wild type spleen, Gaucher spleen lysate, or human fibroblast lysate was incubated with either 10 μM **4**, 10 μM **9** for 30 min followed by 10 μM **4**, or a vehicle only control (0.1% DMSO). All labeling reactions were carried out for 30 min at 37 °C in 500 μL McIlvaine buffer [pH 5.0], before denaturation by addition of 125 μL 10% SDS and boiling for 5 min. Samples were prepared for pull-down with streptavidin coupled DynaBeads (Invitrogen) as described previously[51]. Following pull-down the samples were divided: 1/3 for in-gel digest and 2/3 for on-

bead digest. In-gel digest samples were eluted by boiling beads at 100 °C in 30 µL Laemmli buffer. Eluted proteins were separated by SDS-PAGE, and visualized by silver staining using the SilverQuest kit (Invitrogen). Bands were excised by scalpel and treated with gel digestion buffer (10 mM NH_4HCO_3, 5% ACN, 1mM $CaCl_2$, 10 ng/µL trypsin) at 37 °C overnight. The resulting trypsin-digested peptides were desalted using stage tips, followed by evaporation of ACN and resuspension into 70 µL sample solution (95:3:0.1 H_2O:ACN:TFA) for LC-MS analysis. Samples were analyzed with a 2h gradient of 5–25% ACN on a nano-LC, hyphenated to an LTQ-Orbitrap. Peptides were identified via the Mascot protein search engine.

On bead digest samples were treated with trypsin digestion buffer (100 mM Tris [pH 7.8], 100 mM NaCl, 1 mM $CaCl_2$, 2% ACN and 10 ng/µL trypsin) at 37 °C overnight with shaking. Trypsin digested peptides were desalted using stage tips, and analyzed as above. For identification of GUSB active-site peptides, trypsin digested beads were further treated with endoproteinase Glu-C digestion buffer (100 ng/µL Glu-C in PBS) at 37 °C overnight with shaking. Digested peptides were desalted and analyzed as above.

Peptide identification data were stringently filtered at a false discovery rate (FDR) of 1% and an MS/MS assignment Mascot score >40. Full proteomics data are available in **Supplementary Data Set 1.** The data presented in **Supplementary Data Set 1** show: the ranking of the proteins by the protein score, the UniProt accession number, trivial name of the protein, protein score, protein mass predicted from the RNA sequence, the amino acid coverage of the protein achieved by MS/MS sequencing, the query number of the peptide in the LC/MS run, the experimentally determined m/z of the peptide, the measured molecular weight, the charge state z, the predicted molecular weight, the delta accuracy between predicted and experimentally determined mass, the MS/MS assignment Mascot score, the expectancy value, peptide sequence and emPAI value (which gives an approximate relative estimation of peptide abundance). ~~Full proteomics data are available in **Supplementary Data Set 1**.~~

Western blotting

Proteins resolved by SDS-PAGE were transferred to a PVDF membrane using a Trans-Blot Turbo system (BioRad). Membranes were blocked in 5% BSA for 1 h at rt, then incubated with anti-HPSE (AbCam ab59787) or anti-GUSB (AbCam ab103112) at 1:1000 dilution in 5% BSA at 4 °C overnight. Membranes were washed 3 × with TBST (50 mM Tris [pH 7.5], 150 mM NaCl, 0.05% Tween 20), then incubated with HRP conjugated goat-anti-rabbit (Sigma A0545) at 1:5000 dilution in 5% BSA for 1 h at room temperature. Membranes were washed again 3 × with TBST, blots visualized using Amersham prime ECL reagent (GE Healthcare), and recorded using a Bio-Rad ChemiDoc XRS. Full-length images of all blots used in this study can be found in **Supplementary Fig. 12**.

Protein crystallization

AcGH79

AcGH79 was tested against a range of commercial crystallization screens. Well diffracting crystals of wild type AcGH79 were obtained by the sitting drop vapor diffusion method at 20 °C using 0.8–1.2 M 0.5:9.5 NaH_2PO_4:K_2HPO_4 (v/v) at a protein:well ratio of 700:500 nL. Crystals of AcGH79 E287Q mutant were obtained using 1.2–1.5 M 1.0:9.0 NaH_2PO_4:K_2HPO_4 (v/v) at a protein:well ratio of

Comment [L1]: This seemed like a reasonable place to move this statement (from the Data Availability Statement below). However, if you think there is a better place for it in the Figure legend or in the running online methods, please move it.

Comment [l2]: Moved to slightly earlier in paragraph, just before the description of the columns in the data set.

Suppl data set 1 is also called out in relevant parts of the main text (e.g. L142) and the caption for Fig. 2.

Comment [L3]: This seemed like a reasonable place to move this statement (from the Data Availability Statement below). However, if you think there is a better place for it in the Figure legend or in the running online methods, please move it.

Comment [l4]: Removed the word 'gels', as this is the section about western blotting.

A similar statement about gels has been added to the previous page (L672), in the 'fluorescent labeling' subsection of the methods.

500:500 nL. Crystals typically appeared after 1 day. Crystals were cryoprotected using 2 M lithium sulfate prior to flash freezing in liquid N_2 for data collection.

Mature HPSE

Mature HPSE was crystallized and data collected as previously described[8].

proHPSE

proHPSE at 10 mg/mL was tested against a range of commercial crystallization screens. Thin plate crystals were found in the JCSG screen, which were used to microseed subsequent rounds of crystal screening[52]. Well diffracting single crystals were obtained by the sitting drop vapor diffusion method at 20 °C using 100 mM succinate [pH 7.0], 17% PEG3350 and 1:250 diluted seed stock at a protein:seed:well ratio of 700:100:400 nL. Crystals typically appeared after 1 day and grew to maximum size within 10 days. proHPSE crystals were cryoprotected using mother liquor solution supplemented with 25% ethylene glycol prior to flash freezing in liquid N_2 for data collection.

Xray data collection and structure solution

Xray diffraction data were collected at 100 K at beamlines i02 (5LA4, 5LA7), i03 (5G0Q, 5L77, 5L9Y) and i04 (5L9Z) of the Diamond Light Source UK. Reflections were autoprocessed with the xia2 pipeline[53] of the CCP4 software suite, or manually processed using XDS[54] and Aimless[55]. Apo proHPSE was solved by molecular replacement with the mature HPSE model (5E8M) using MolRep[56], followed by alternating rounds of manual model building and refinement using Coot and REFMAC5 respectively[57,58].

For all complexes with ABP **5**, crystals were soaked in their respective mother liquors supplemented with 2–5 mM **5** for 1–3 h at 20 °C. Soaked crystals were cryoprotected using their respective cryoprotectant solutions prior to flash freezing in liquid N_2 for data collection.

Complexes were solved by molecular replacement with their respective apo structures, followed by rounds of manual model building and refinement using Coot and REFMAC5. Ligand coordinates were built using jLigand[59]. Active site diagrams were generated using ccp4mg[60]. Ribbon and protein surface diagrams were generated using PyMOL.

Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5G0Q (AcGH79(wt)-**5** complex), 5L77 (AcGH79(E287Q)-**5** complex), 5L9Y (HPSE(wt)-**5** complex), 5L9Z (HPSE(E343Q)-**5** complex), 5LA4 (*apo* proHPSE), 5LA7 (proHPSE-**5** complex).

Comment [I5]: This is a verbatim repeat of L350-352. I originally added this information here as part of the data availability statement.

Is it still necessary?

Data availability

All ~~other~~ data generated or analyzed during this study are included in this published article (and its Supplementary Information files), or are available from the corresponding authors on reasonable request.

Methods references

46. Li, K.-Y. et al. Synthesis of Cyclophellitol, Cyclophellitol Aziridine, and Their Tagged Derivatives. *Eur. J. Org. Chem.*, 6030–6043 (2014).
47. Zheng, L., Baumann, U. & Reymond, J.L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**, e115 (2004).
48. Li, M.Z. & Elledge, S.J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51-59 (2012).
49. Berger, I., Fitzgerald, D.J. & Richmond, T.J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583-1587 (2004).
50. Seiler, C.Y. et al. DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. *Nucleic Acids Res.* **42**, D1253-1260 (2014).
51. Li, N. et al. Relative quantification of proteasome activity by activity-based protein profiling and LC-MS/MS. *Nat. Protoc.* **8**, 1155-1168 (2013).
52. D'Arcy, A., Bergfors, T., Cowan-Jacob, S.W. & Marsh, M. Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallogr. F* **70**, 1117-1126 (2014).
53. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186-190 (2010).
54. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125-132 (2010).
55. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204-1214 (2013).
56. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22-25 (2010).
57. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).
58. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355-367 (2011).
59. Lebedev, A.A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D* **68**, 431-440 (2012).
60. McNicholas, S., Potterton, E., Wilson, K.S. & Noble, M.E.M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **67**, 386-394 (2011).