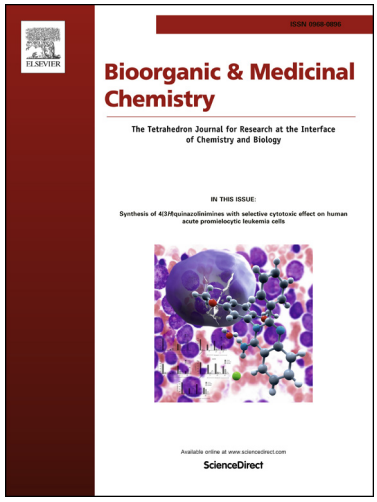
Accepted Manuscript



Covalent inhibitors of LgtC: a blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases

Yong Xu, Ruth Smith, Mirella Vivoli, Masaki Ema, Niina Goos, Sebastian Gehrke, Nicholas J. Harmer, Gerd K. Wagner

PII: S0968-0896(17)30179-7
DOI: http://dx.doi.org/10.1016/j.bmc.2017.04.006
Reference: BMC 13675

To appear in: *Bioorganic & Medicinal Chemistry*

Received Date: 2 February 2017
Revised Date: 1 April 2017
Accepted Date: 4 April 2017

Please cite this article as: Xu, Y., Smith, R., Vivoli, M., Ema, M., Goos, N., Gehrke, S., Harmer, N.J., Wagner, G.K., Covalent inhibitors of LgtC: a blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases, *Bioorganic & Medicinal Chemistry* (2017), doi: http://dx.doi.org/10.1016/j.bmc.2017.04.006

Covalent inhibitors of LgtC: a blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases[♯]

Yong Xu [a], Ruth Smith [b], Mirella Vivoli [c], Masaki Ema [a], Niina Goos [b], Sebastian Gehrke [b,d], Nicholas J. Harmer [c] & Gerd K. Wagner [a]*

[a] King's College London, Department of Chemistry, Faculty of Natural & Mathematical Sciences, Britannia House, 7 Trinity Street, London, SE1 1DB, UK. Phone: +44 (0)20 7848 1926; e-mail: gerd.wagner@kcl.ac.uk

[b] King's College London, Institute of Pharmaceutical Science, 150 Stamford Street, London, SE1 9NH, UK

[c] University of Exeter, Henry Wellcome Building for Biocatalysis, Stocker Road, Exeter, EX4 4QD, UK

[d] University of East Anglia, School of Pharmacy, Earlham Road, Norwich, NR4 7TJ, UK

*Corresponding author

1 April 2017

Keywords: covalent inhibitor, enzyme, glycosyltransferase, chemical tool, virulence factor

[♯]GW dedicates this work to his mother, on the occasion of her 65th birthday.

Abstract

Non-substrate-like inhibitors of glycosyltransferases are sought after as chemical tools and potential lead compounds for medicinal chemistry, chemical biology and drug discovery. Here, we describe the discovery of a novel small molecular inhibitor chemotype for LgtC, a retaining α-1,4-galactosyltransferase involved in bacterial lipooligosaccharide biosynthesis. The new inhibitors, which are structurally unrelated to both the donor and acceptor of LgtC, have low micromolar inhibitory activity, comparable to the best substrate-based inhibitors. We provide experimental evidence that these inhibitors react covalently with LgtC. Results from detailed enzymological experiments with wild-type and mutant LgtC suggest the non-catalytic active site residue Cys246 as a likely target residue for these inhibitors. Analysis of available sequence and structural data reveals that non-catalytic cysteines are a common motif in the active site of many bacterial glycosyltransferases. Our results can therefore serve as a blueprint for the rational design of non-substrate-like, covalent inhibitors against a broad range of other bacterial glycosyltransferases.

1. Introduction

Glycosyltransferases (GTs) are key enzymes for the biosynthesis of complex glycans and glycoconjugates in all domains of life [1]. In bacteria, GT activity is required for virulence and viability, and individual GTs are emerging as novel targets for anti-microbial and anti-virulence drug discovery [2,3]. Small molecular inhibitors of GTs are therefore sought after as chemical tools for the interrogation, manipulation and disruption of cellular glycosylation pathways [4], and as potential lead compounds for drug discovery. Most existing GT inhibitors are substrate analogues that are structurally derived from the respective GT donor and/or acceptor [5]. Such substrate-based GT inhibitors are usually not drug-like and often suffer from modest bioactivity and/or intrinsic physicochemical liabilities, such as limited stability and poor cell penetration. While some of these drawbacks can be circumnavigated by using metabolic precursors [6-8], the identification of alternative, non-substrate-like GT inhibitors remains an important goal in medicinal chemistry and chemical biology [9]. Despite recent progress [10,11], very few such non-substrate-like GT inhibitors are currently available, in stark contrast to other enzyme classes of similar size and biological importance (e.g. kinases, proteases).

Covalent enzyme inhibitors are currently undergoing a renaissance in chemical biology [12] and drug discovery [13]. Covalent inhibitors generally display a range of attractive features, including high potency, prolonged duration of action, and amenability to rational design. A systematic study has recently shown that even reactive electrophiles, such as the Michael acceptor acrylamide, do not react indiscriminately with any biological nucleophile [14]. This differential reactivity can be harnessed for the development of covalent inhibitors, especially those that target non-catalytic residues [15]. This approach has been used successfully for inhibitor development against several challenging enzyme targets, including drug-resistant kinases [16] and proteases

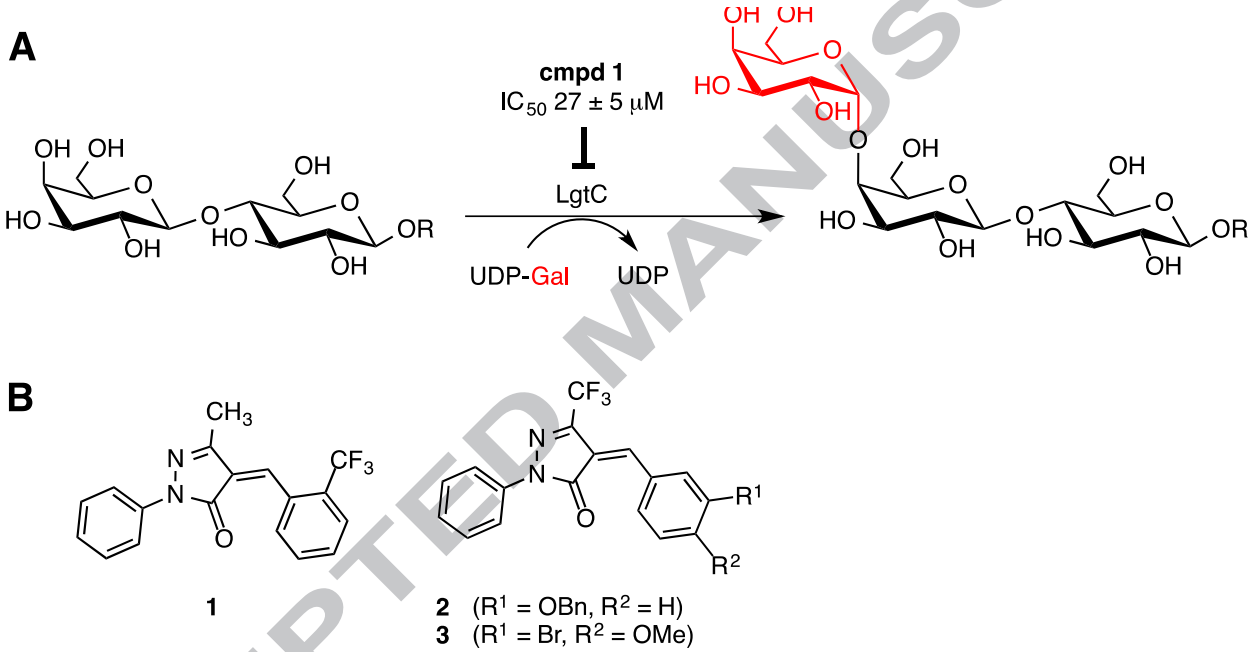
[17]. The covalent targeting of non-catalytic residues, in particular cysteines, therefore represents a promising strategy for inhibitor development against intractable targets such as GTs. To date, no inhibitors with this mode of action have been reported for any bacterial GT. This is particularly striking, as non-catalytic cysteines are a common motif in the active site of bacterial GTs. Even beyond bacterial enzymes, there is only a single mechanistically related example in the entire GT family [11]. Inhibitors of the mammalian O-linked N-acetylglucosamine transferase (OGT) act via an unusual double-displacement mechanism, which draws on the unique architecture of this enzyme – not found in other GTs – and requires a very specific arrangement of two separate target residues [11].

Herein, we describe a new class of non-substrate-like, covalent inhibitors of the retaining α-1,4-galactosyltransferase LgtC from *Neisseria meningitidis*. LgtC catalyzes the transfer of D-galactose (D-Gal) from a UDP-Gal donor to a lactose acceptor (Fig. 1A) [18]. In pathogenic Gram-negative bacteria such as *Neisseria* and *Haemophilus*, LgtC is required for the biosynthesis of the terminal digalactoside epitope α-D-Gal*p*-(1,4)-β-D-Gal*p* in the lipooligosaccharide (LOS) structures of the outer core of the cell envelope [19]. LgtC expression has been associated with high-level serum resistance in non-typeable *Haemophilus influenzae* (NTHi) [19], and inhibition of LgtC and related GTs that are involved in LOS biosynthesis has been suggested as a strategy for anti-microbial drug discovery [18]. LgtC is an excellent starting point for such an approach not only because of its role in bacterial virulence, but also because of its structural and mechanistic communalities with other bacterial GTs [18], which have made it a widely used model system for mechanistic and structural studies in this enzyme family.

We have identified small molecular inhibitors of LgtC that are structurally unrelated to both the LgtC donor and acceptor (Fig. 1B). We demonstrate that these inhibitors act via a covalent mechanism of inhibition, targeting a non-catalytic cysteine

residue in the LgtC active site. Analysis of sequence and structural data shows that active-site cysteines are a common feature in bacterial GTs. Our results may therefore serve as a blueprint for the rational development of non-substrate-like, covalent inhibitors of other bacterial enzymes in this family.

Figure 1 (**A**) The LgtC reaction, and its inhibition by pyrazol-3-one **1**.[a] (**B**) Chemical structures of pyrazol-3-ones **1**-**3**.



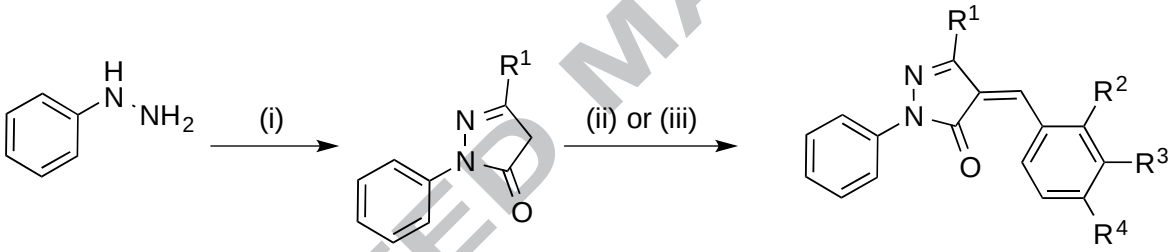
[a]*Assay conditions*: LgtC (activated with 5 mM DTT), lactose (2 mM), UDP-Gal (28 μM), $MnCl_2$ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and inhibitor (0.1-100 μM) were incubated in HEPES buffer (pH 7.0) for 20 min at 30 °C. The progress of the reaction was determined with malachite green as previously described [20].

2. Results and discussion

2.1 Rationale and chemical synthesis of inhibitors. We have recently adapted a biochemical GT assay for inhibitor studies with non-substrate-like chemotypes [20]. We have subsequently used this assay to evaluate a collection of 130 small molecules as potential non-substrate-like inhibitors of LgtC. The collection was designed around

structural scaffolds that are known from other GT inhibitors, including nucleosides, steroids, and pyrazol-3-ones (Fig. 1B). Pyrazol-3-ones have previously been reported as screening hits against other GTs [21,22], but their molecular mode of action has not been investigated. From our own screen against LgtC, pyrazol-3-one **1** emerged as the most promising hit from this series and is the primary target molecule of the present study. **1** was synthesized in two steps from phenylhydrazine via a ring-closing condensation with ethyl acetoacetate, followed by a Knoevenagel condensation with 2-trifluoromethyl benzaldehyde (Scheme 1). This synthesis was readily adapted to also provide access to the congeners **2** and **3** (Scheme 1).

Scheme 1 Synthesis of target molecules **1**-**3**.[a]



6 (R = CH_3)
7 (R = CF_3)

1 (R^1 = CH_3, R^2 = CF_3, R^3 = R^4 = H)
2 (R^1 = CF_3, R^2 = R^4 = H, R^3 = OBn)
3 (R^1 = CF_3, R^2 = H, R^3 = Br, R^4 = OMe)

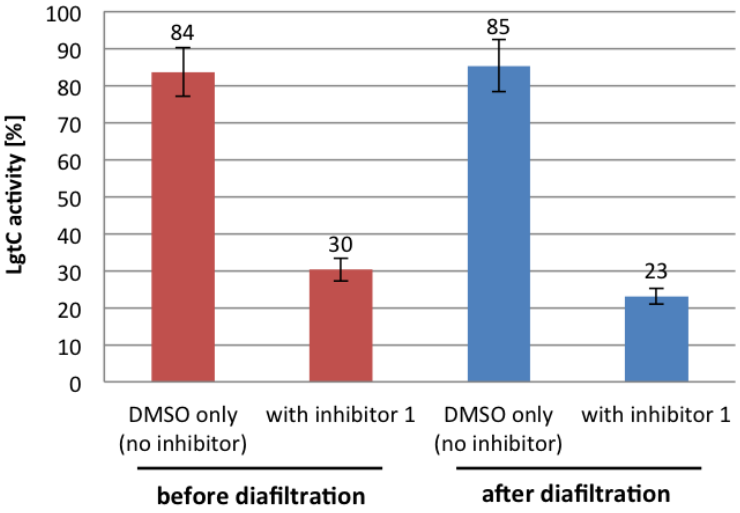
[a]*Reagents & conditions*: (i) ethyl 3-oxobutanoate, acetic acid, 110 ºC, 97% (**6**) or ethyl 4,4,4-trifluoroacetoacetate, acetic acid, 110 ºC, 72% (**7**); (ii) 2-(trifluoromethyl) benzaldehyde, acetic acid, reflux, 48% (for **1**); (iii) 3-(benzyloxy)benzaldehyde (for **2**) or 3-bromo-4-methoxybenzaldehyde (for **3**), 160 ^{o}C/microwave, 15 min, 53% (**2**) or 75% (**3**)

2.2 Biochemical evaluation of inhibitors. In initial experiments, pyrazol-3-one **1** inhibited LgtC with an IC_{50} value of 27 μM (Fig. 1A). All inhibition assays were carried out in the presence of surfactant to suppress non-specific aggregation and avoid assay artefacts [23]. This assay protocol provided confidence that the inhibitory activity of **1** was genuine. Equally importantly, **1** showed no inhibitory activity in control experiments

against the calf intestinal phosphatase (CIP) used as a secondary enzyme in the biochemical assay (SI).

We hypothesised that **1** may act as a covalent inhibitor towards LgtC, due to the presence of the Michael acceptor system, which is a known cysteine-reactive electrophile in covalent inhibitors and probes [24]. As no reports on pyrazol-3-ones as covalent enzyme inhibitors have previously been published, we set out to experimentally test this hypothesis. First, we compared enzyme activity in the assay mixture of LgtC and **1** before and after diafiltration. The inhibitory activity of **1** was preserved after diafiltration, which strongly suggested a covalent mode of inhibition (Fig. 2).

Figure 2 LgtC activity in the presence of inhibitor **1**, before and after diafiltration.[a]

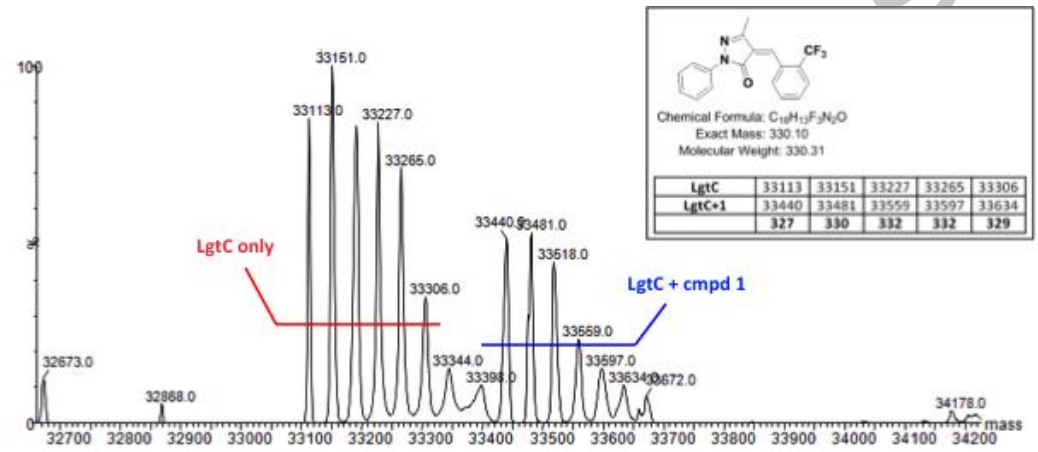


[a]*Conditions*: LgtC (activated with 5 mM DTT) was pre-incubated with inhibitor (100 µM) or DMSO, in the presence of UDP-Gal (28 µM), for 10 min at 30 °C. Lactose (2 mM) was added, and the reactions were incubated for 20 min at 30 °C. Samples were drawn, and enzyme activity before diafiltration was determined under standard assay conditions. Enzyme reactions (3 mL) were subjected to diafiltration in Vivaspin concentrator tubes (centrifugation at 4 °C, 4000 rpm). The initial residual volume (300 µL) was washed once with HEPES buffer (pH 7.0, total volume 3 mL). The residual volume (300 µL) from the wash step was diluted to 450 µL with HEPES buffer (pH 7.0), and enzyme activity after diafiltration was determined under standard assay conditions (2 mM lactose, 28 µM UDP-Gal, 20 min incubation at 30 °C). Each experiment was carried out in triplicate (100% activity = complete conversion of UDP-Gal). Error bars represent standard deviation.

Next, we directly analysed the reaction of LgtC with **1** by mass spectrometry. These experiments showed the formation of a covalent adduct, providing direct evidence for the covalent reaction of **1** and LgtC (Fig. 3A). Importantly, only a single adduct was observed in these experiments, at a 1:1 ratio of inhibitor and enzyme. This suggests that **1** reacts only with a single residue in LgtC.

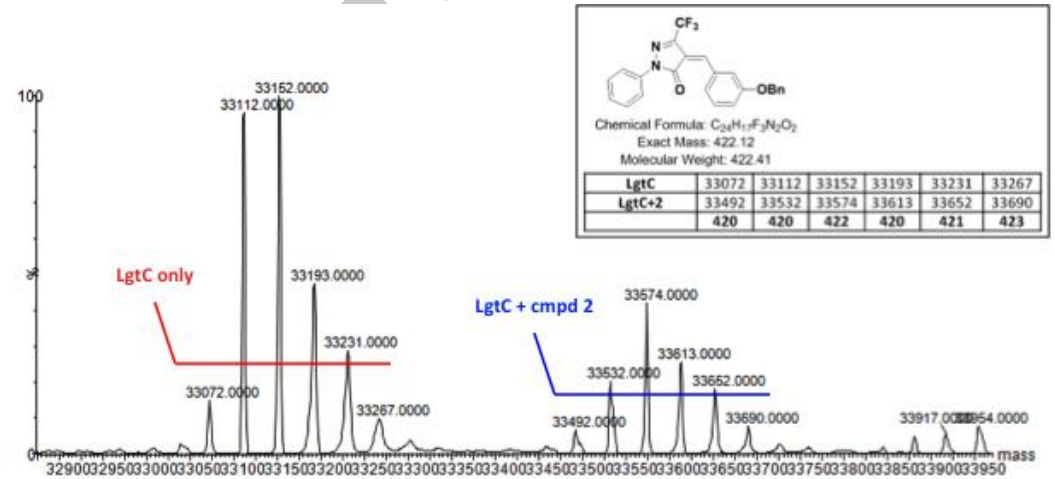
Figure 3 Mass spectrometry reveals the formation of covalent adducts of LgtC with pyrazol-3-ones **1** (panel **A**), **2** (panel **B**) and **3** (panel **C**).[a]

A



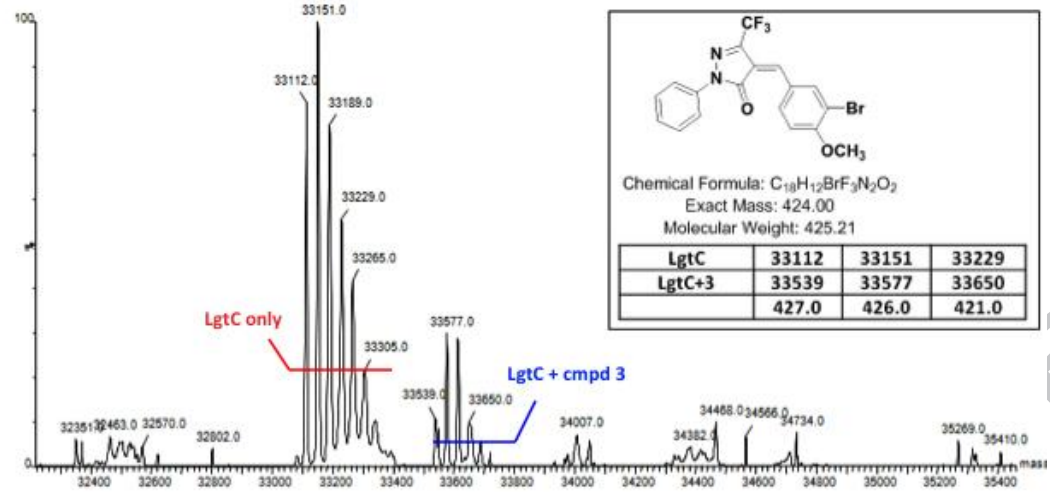
LgtC	33113	33151	33227	33265	33306
LgtC+1	33440	33481	33559	33597	33634
	327	330	332	332	329

B



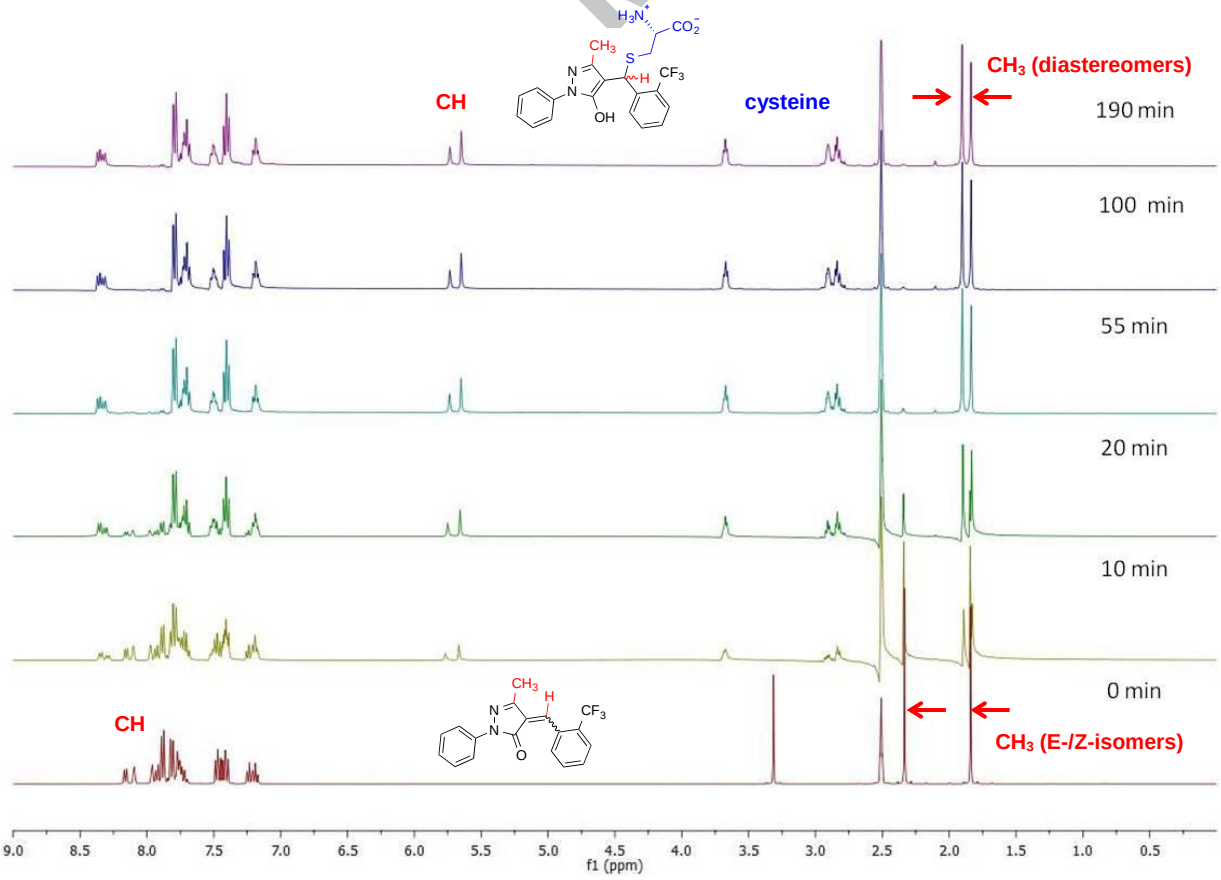
LgtC	33072	33112	33152	33193	33231	33267
LgtC+2	33492	33532	33574	33613	33652	33690
	420	420	422	420	421	423

C



[a]*Conditions*: LgtC in HEPES buffer (pH 7.0) was incubated with inhibitor (**1** and **2:** 100 µM in DMSO; **3**: 50 µM in DMSO) for 20 min at 30 °C (10% final DMSO concentration). Samples were directly analysed by electrospray mass spectrometry.

Figure 4 ^{1}H-NMR spectra of the time-dependent, covalent reaction of **1** with cysteine.[a]



[a]*Conditions*: Solutions of **1** (19.8 mg, 0.06 mmol, in *d6*-DMSO) and L-cysteine (7.26 mg, 0.06 mmol, in *d6*-DMSO) were mixed in a glass NMR tube. ^{1}H-NMR spectra were recorded at various time points from 0 min to 190 min (δ 3.3 ppm: H_2O).

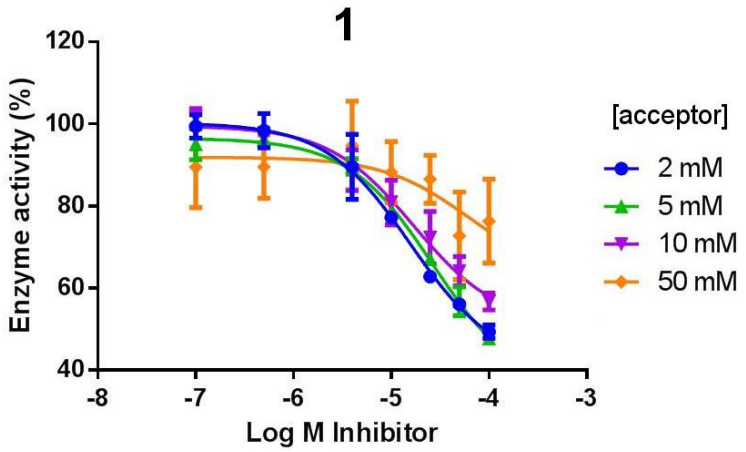
In order to identify potential target residues for the covalent reaction, we incubated **1** separately with three amino acids containing different side chain nucleophiles (L-cysteine, L-serine and L-lysine), and followed the individual reactions by ^{1}H-NMR. The time-dependent formation of a covalent adduct at the exocyclic double bond was observed upon incubation of **1** with L-cysteine (Fig. 4), but not with L-serine or L-lysine (SI). These results show that under otherwise identical conditions, the Michael acceptor system in **1** reacts preferentially with a sulfur nucleophile.

This preference suggested that the target residue in LgtC may be a cysteine, although the reactivity of a specific residue in the environment of the protein will necessarily depend on neighboring groups. LgtC contains five cysteines, of which three are solvent-accessible [18]. One of these is Cys246, which sits at the interface of the LgtC acceptor and donor binding sites, in close proximity to the lactose acceptor (SI, Fig. S1). Molecular simulations using a covalent docking protocol [25] suggested that **1** is binding at the substrate binding site of LgtC in an ideal orientation for the covalent reaction of the Michael acceptor system with Cys246 (SI, Fig. S1). In order to experimentally test the hypothesis that **1** may indeed bind at the substrate binding site, we investigated the effect the concentration of the LgtC acceptor, lactose, may have on inhibition. Under standard assay conditions (no pre-incubation, turnover 20-50%), inhibitory activity decreased markedly at higher acceptor concentrations (Fig. 5A). Qualitatively, this is consistent with an acceptor-competitive binding mode for **1**. In order to obtain quantitative data and enable detailed enzyme kinetics analysis, we repeated this experiment at turnover rates of <10%, at a fixed concentration of UDP-Gal donor (100 μM), and variable concentrations of lactose acceptor (5-25 mM) and inhibitor (1-5 μM). Dixon and Cornish-Bowden analyses of the resulting data show that under these conditions, **1** acts as a competitive inhibitor of LgtC relative to the lactose acceptor (Fig.

5B). These results show that **1** does indeed bind at the substrate binding site, as required for a potential covalent reaction with Cys246.

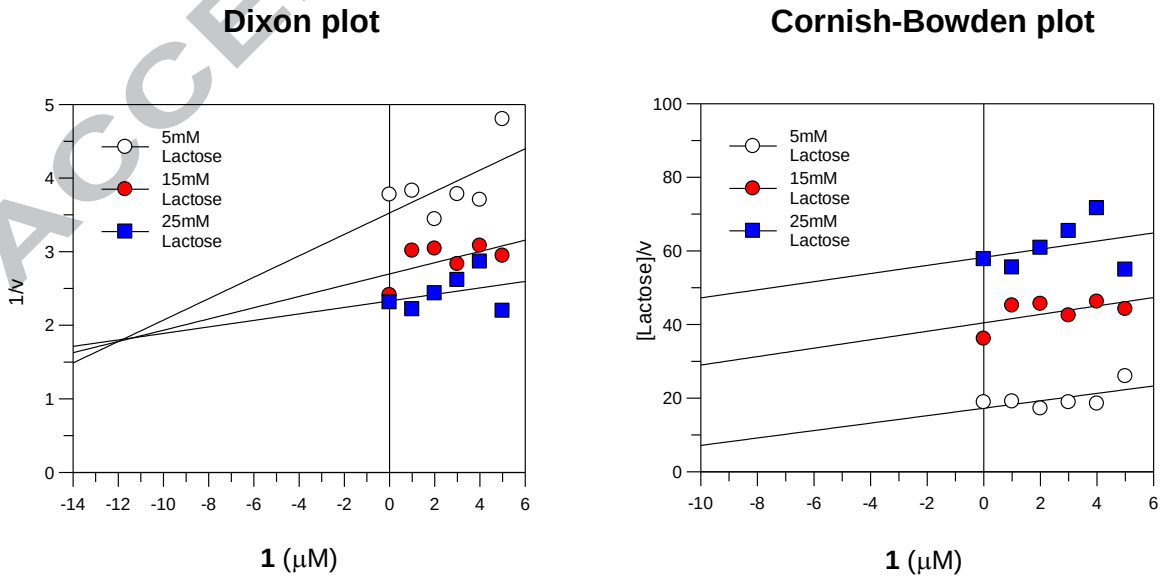
Figure 5 Effect of different acceptor concentrations on the inhibitory activity of pyrazol-3-one **1** towards LgtC under standard assay conditions (**A**) and enzyme kinetics conditions (**B**).

A Standard assay conditions (donor turnover 20-50%)[a]



[a]*Conditions*: LgtC was incubated with **1** (0.1-100 µM), UDP-Gal (28 µM), $MnCl_2$ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and lactose acceptor (2-50 mM) for 20 min at 30 °C in 13 mM HEPES buffer (pH 7.0). Each experiment was carried out in triplicate.

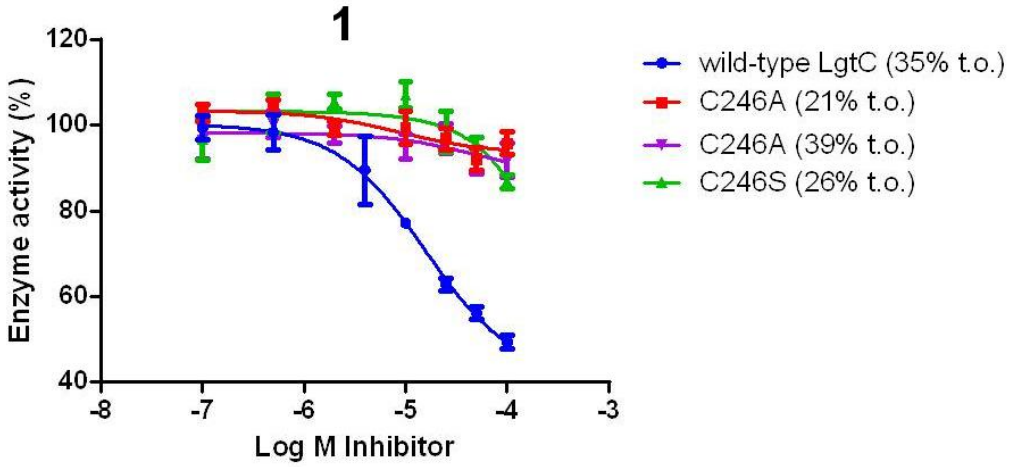
B Enzyme kinetics conditions (donor turnover <10%)[a]



[a]*Conditions*: LgtC was incubated with **1** (1-5 µM), UDP-Gal (100 µM), $MnCl_2$ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and lactose acceptor (5-25 mM) for 20 min at 30 °C in 13 mM HEPES buffer (pH 7.0). Data points were fitted to a Dixon or Cornish-Bowden plot with GraFit 7.

In order to investigate the role of Cys246 for the inhibitory activity of **1**, we created two LgtC mutants, in which Cys246 is replaced with a non-nucleophilic (C246A) or only weakly nucleophilic (C246S) residue. Both mutants were catalytically active under our standard conditions, albeit less so than wild-type LgtC. At the highest enzyme concentration tested, the observed turnover of donor was approximately 36% (C246A) or 22% (C246S) of that of the wild-type enzyme (SI). This level of enzymatic activity was sufficient for inhibition experiments. Under the standard assay conditions, pyrazol-3-one **1** at the highest concentration (100 µM) showed only weak inhibition against C246S LgtC, and almost none against C246A LgtC (Fig. 6). These results provide direct evidence that Cys246 is a critical residue for the inhibition of LgtC by pyrazol-3-one **1**, possibly as the target for a covalent interaction with the inhibitor. They also suggest that **1** is only weakly effective as a non-covalent inhibitor of LgtC, at least at the concentrations tested.

Figure 6 Compound **1** shows only weak inhibition of the LgtC mutants C246A and C246S.[a]



[a]*Conditions*: LgtC was incubated with **1** (0.1-100 µM), UDP-Gal (28 µM), $MnCl_2$ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and lactose acceptor (2 mM) for 20 min at 30 °C in 13 mM HEPES buffer (pH 7.0). Control experiments without inhibitor were carried out in parallel, showing a donor turnover of, respectively, 23% (wild-type), 21% or 39% (C246A), and 26% (C246S). Each experiment was carried out in triplicate.

To further dissect the structural requirements for the interaction between **1** and LgtC, we next carried out inhibition experiments with structural analogues of **1** (SI). As expected, compound **4**, a hydrogenated variant of **1** lacking the exocyclic double bond and hence the capacity for covalent interaction, showed no inhibitory activity against LgtC (SI, Table S1). Compound **5**, which contains a different heterocyclic scaffold, but the same Michael acceptor system as pyrazol-3-one **1**, was also inactive (SI, Table S1). The lack of activity of compounds **4** and **5** shows that the Michael acceptor system in **1** is necessary, but not sufficient, for the inhibition of LgtC. It suggests that non-covalent interactions also contribute to the LgtC inhibitory activity of **1**, probably by orienting the inhibitor correctly in the substrate binding site prior to a covalent interaction with Cys246.

The substrate binding sites of GTs with the same donor and/or acceptor are highly conserved. We therefore decided to explore the behaviour of **1** against three other GTs (Table 1), which use either the same UDP-Gal donor (bovine β-1,4-GalT) or the same lactose acceptor (*Pasteurella multocida* α-2,3-SiaT, human FUT-2) as LgtC, but lack an active-site cysteine. Pyrazol-3-one **1** was inactive against all three of these GTs (Table 1 & SI). These results demonstrate that **1** does not act simply as a non-covalent GT substrate mimic. Taken together, our results suggest strongly that it is the capacity to bind non-covalently at the substrate binding site of LgtC, in combination with a covalent interaction, most likely with Cys246, which forms the basis for the inhibitory activity of **1** towards LgtC.

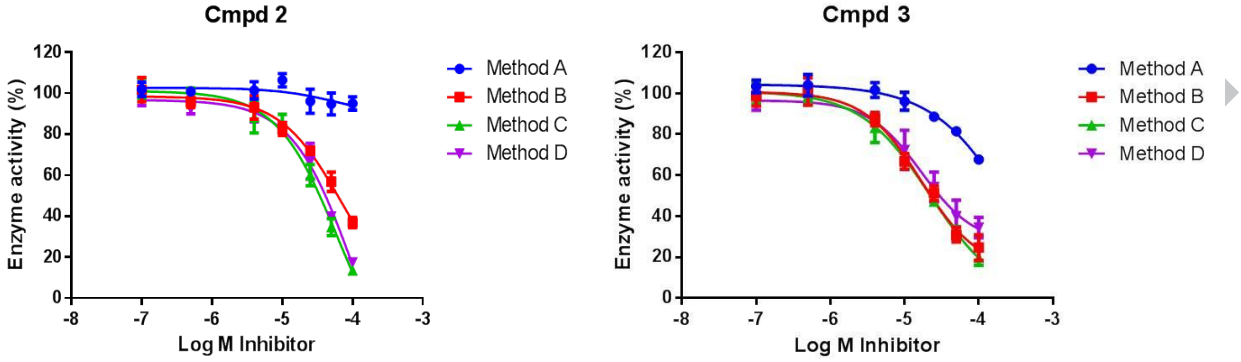
Table 1 Substrate specificities and inhibition of four glycosyltransferases[1]

GT	Organism	Donor substrate	Acceptor substrate	Inhibition (cmpd **1**)
LgtC	*Neisseria meningitidis*	UDP-Gal	β-lactose	27 ± 5 μM[2]
β-1,4-GalT	*Bos taurus*	UDP-Gal	GlcNAc	no inhibition[3]
α-2,3-SiaT	*Pasteurella multocida*	CMP-Neu5Ac	β-lactose	no inhibition[4]
FUT-2	*H. sapiens*	GDP-Fuc	α-lactose	no inhibition[4]

[1]For experimental details of the inhibition experiments see SI; [2]IC_{50}; [3]at highest concentration of inhibitor (100 μM); [4]at highest concentration of inhibitor (150 μM)

Having established that prototype pyrazol-3-one **1** does indeed act via a covalent mode of action, we set out to improve the potency of this new LgtC inhibitor class. To increase the reactivity of the Michael acceptor system, we replaced the 5-CH_3 group in **1** with a strongly electron-withdrawing 5-CF_3 group (Fig. 1B). The exact 5-CF_3 congener of **1**, bearing a (2-trifluoromethyl)benzylidene substituent in position 4, was not synthetically accessible. We therefore introduced alternative benzylidene substituents in position 4 (Fig. 1B). While the new inhibitors **2** and **3** showed no, or only modest, activity under our standard assay conditions, their inhibitory activity increased markedly upon pre-incubation of LgtC with inhibitor, prior to starting the enzyme reaction (Fig. 7 & 8). Such a time-dependent inhibition is characteristic for a covalent mode of action. Interestingly, the pronounced effect of pre-incubation on inhibitory activity was largely independent of the presence of donor or acceptor substrate (Fig. 7, methods B-D).

Figure 7 Pre-incubation with LgtC significantly increases the inhibitory activity of **2** and **3**, both in the presence and absence of LgtC substrates.[a]



[a]*Conditions*: Following pre-incubation (Methods A-D, see Table 2), LgtC was incubated with inhibitor (0.1-100 µM), UDP-Gal (28 µM), $MnCl_2$ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and lactose acceptor (2 mM) for 20 min at 30 °C in 13 mM HEPES buffer (pH 7.0). Each experiment was carried out in triplicate.

Table 2 Pre-incubation protocols and enzyme turnovers.

	Pre-incubation protocol	Turnover (%)[1]	
		2	**3**
Method A	No pre-incubation	29	30
Method B	Pre-incubation (10 min) of LgtC, inhibitor and donor, prior to addition of acceptor	29	30
Method C	Pre-incubation (10 min) of LgtC and inhibitor, prior to addition of donor and acceptor	25	25
Method D	Pre-incubation (10 min) of LgtC, inhibitor and acceptor, prior to addition of donor	23	23

[1]Turnover is defined as the percentage consumption of donor UDP-Gal in the biochemical assay, relative to the total amount of UDP-Gal used.

For the two inhibitors **2** and **3**, inhibitory activity increased consistently with increasing pre-incubation time, albeit at different rates, before reaching a plateau after 60 mins (**2**) and 20 mins (**3**), respectively (Fig. 8 & Table 3). The covalent mode of action for **2** and **3** was further confirmed with several key experiments, as for the parent compound **1**. Thus, for both **2** and **3**, covalent adducts with LgtC were detected by mass spectrometry (Fig. 3B & C), and the time-dependent formation of a covalent adduct with

cysteine was observed by ^{1}H NMR (SI). In control experiments, neither **2** nor **3** showed activity against CIP or β-1,4-GalT (SI).

Figure 8 LgtC inhibition by pyrazolones **2** and **3** at different pre-incubation times. For conditions and IC_{50} values see Table 3.

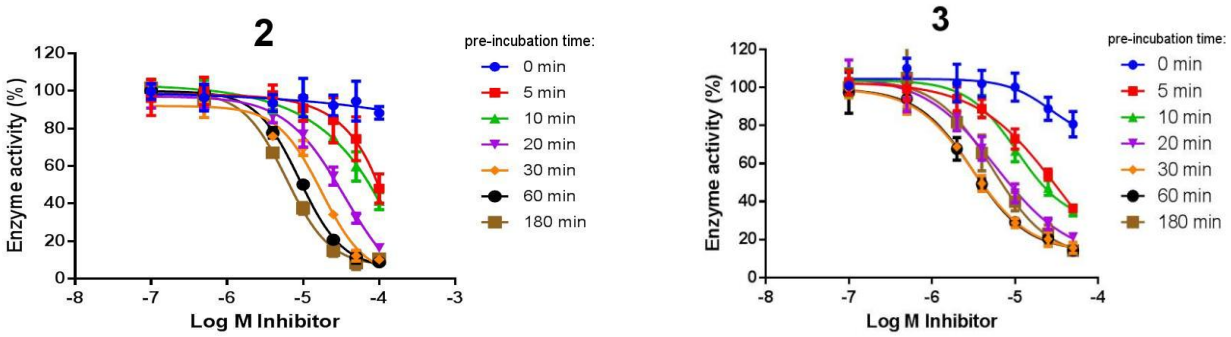


Table 3 LgtC inhibition by pyrazolones **2** and **3** at different pre-incubation times. IC_{50} values extracted from inhibition curves in Figure 8.[a]

	Pre-incubation time (min)						
	0	5	10	20	30	60	180
2 IC_{50} (μM) ± SD	>100	>50	>50	38 ± 13.0	17.6 ± 1.1	9.0 ± 0.2	6.0 ± 0.2
LgtC turnover (%)	29	30	41	27	24	29	19
3 IC_{50} (μM) ± SD	>50	24 ± 7.1	15.8 ± 0.8	6.5 ± 1.6	3.1 ± 0.5	3.1 ± 0.3	5.5 ± 1.2
LgtC turnover (%)	30	32	46	32	32	39	26

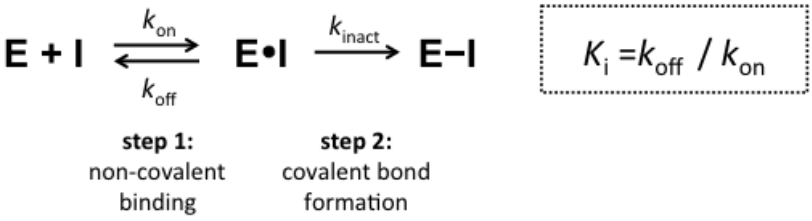
[a]*Conditions*: LgtC was pre-incubated with inhibitor (**2**: 0.1-100 μM; **3**: 0.1-50 μM) or DMSO for the given time (0-180 min) at 30 °C. Lactose (2 mM) and UDP-Gal (28 μM) were added, and the reactions were incubated for 20 min at 30 °C. Enzyme activity was determined under standard assay conditions. Each experiment was carried out in triplicate.

The activity of a covalent inhibitor is governed by two parameters: the strength of the initial, non-covalent binding interaction with the target enzyme, and the reactivity of the electrophilic warhead (Fig. 9). In order to dissect the contribution of these two steps towards the overall activity of **2** and **3**, we determined k_{obs} at different inhibitor concentrations. By plotting k_{obs} over inhibitor concentration, the kinetic parameters for

the first (K_i) and second (k_{inact}) interaction step can be extracted [13]. This analysis shows that k_{inact} for both **2** and **3** is of a similar order of magnitude, whereas K_i is about 10-times greater for compound **2** than for **3**. These results suggest that due to the different substitution pattern of the 4-substituent, **2** and **3** differ significantly in their non-covalent binding affinity for LgtC, but not so much in the reactivity of the Michael acceptor system. This profile corresponds to the results from experiments with different pre-incubation times, which show that **2** requires a longer pre-incubation period for maximal activity than **3** (Table 3). The different non-covalent affinities of **2** and **3** are also consistent with the slightly different binding modes that were obtained for both inhibitors from covalent docking simulations. These docking results suggest that both **2** and **3** can bind at the substrate binding site of LgtC in an orientation that allows a covalent reaction with residue Cys246, but that the smaller 3-bromo-4-methoxyphenyl substituent in **3** can be more readily accommodated than the bulky 3-benzyloxyphenyl substituent in **2** (SI, Fig. S1).

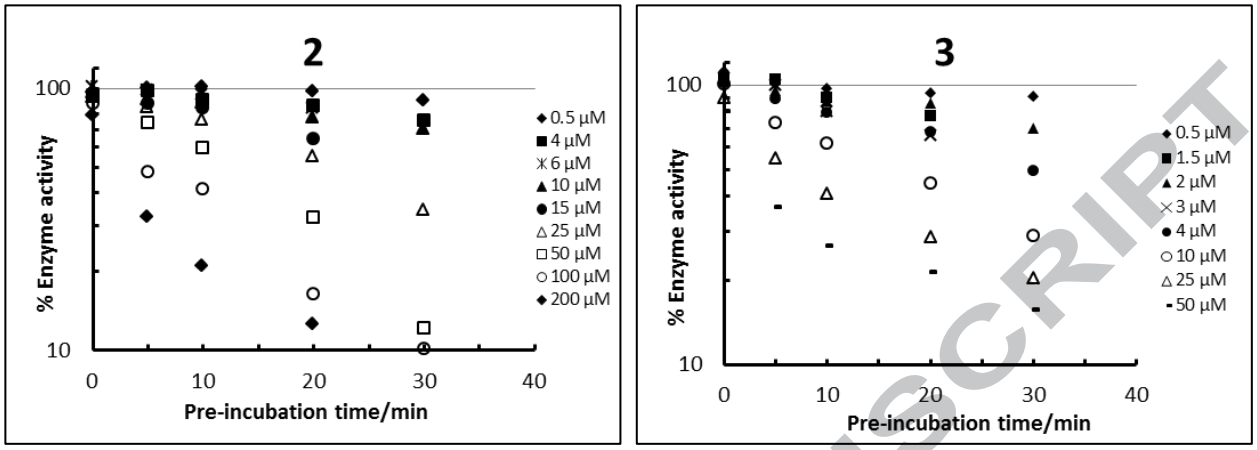
Figure 9 Covalent inhibition kinetics for **2** and **3**.

(a) Two-step model of covalent inhibition.

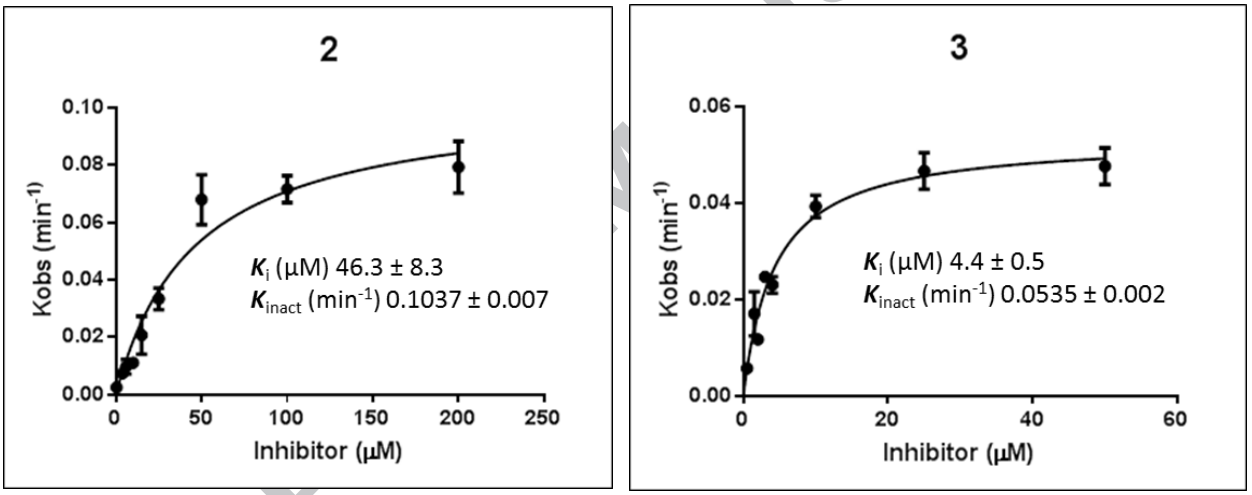


(b) Determination of kinetic parameters k_{obs}, K_i and k_{inact}.[a]

A Semi-logarithmic plots for the time-dependent inactivation of LgtC at different inhibitor concentrations



B Observed rate constants k_{obs} at different inhibitor concentrations, extracted from the plots in **A**



[a]*Conditions*: **A** LgtC activity was determined at various concentrations of **2** (0.5-200 µM) or **3** (0.5-50 µM) and in function of different pre-incubation times (0, 5, 10, 20 and 30 min). Enzyme activity is expressed as percentage of control (DMSO only) and plotted on a semi-logarithmic scale over pre-incubation time. Values for k_{obs} were extracted from exponential regression using the equation *enzyme activity* [%] = $Ae^{-kobs \times t}$. **B** Observed rate constants k_{obs} were extracted from the plots in **A**, re-plotted over inhibitor concentrations, and fitted to the hyperbolic equation $k_{obs} = k_{inact} \times [I]/(K_i+[I])$. From these plots, K_i and k_{inact} were obtained as previously described [13]. All experiments were performed in triplicate.

3. Conclusions

Existing inhibitors of the bacterial α-1,4-galactosyltransferase LgtC are substrate analogues based on a sugar-nucleotide scaffold [18,26]. Our results establish suitably substituted pyrazol-3-ones as the first non-substrate-like, covalent inhibitor chemotype

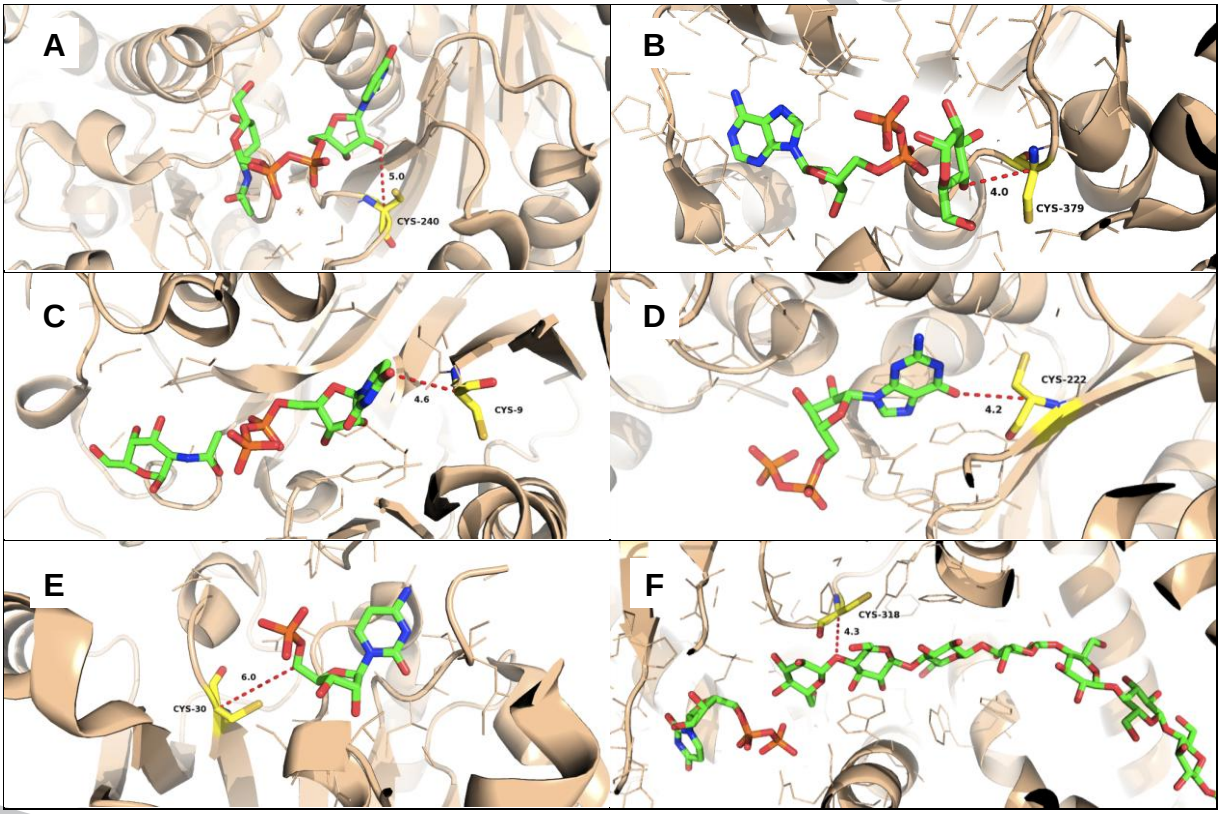
for this enzyme. The activity of inhibitors **2** and **3** upon preincubation is comparable to the most potent substrate-based LgtC inhibitors reported to date [26]. Our data suggest that pyrazol-3-ones **1**-**3** bind at the substrate binding site of LgtC prior to the covalent reaction, which most likely occurs at the non-catalytic residue Cys246 in the enzyme active site. Importantly, the new inhibitors are inactive against three related GTs that use either the same donor or acceptor substrate as LgtC, but lack a non-catalytic cysteine in the active site. This behaviour is notable, as other pyrazol-3-ones have previously been classified as non-selective, pan-assay interfering compounds (PAINS) [27]. It suggests that, while it is likely that the unoptimised inhibitors **1**-**3** may also react with other proteins that contain a carbohydrate- or nucleotide-binding site with a non-catalytic cysteine, target selectivity could be engineered into these inhibitors through further structural modification.

Non-catalytic cysteine residues are a common feature of bacterial GTs. In principle, the direct targeting of non-catalytic cysteines can therefore serve as a general inhibitor discovery strategy against many other bacterial GTs. In the sequence-based CAZy database of carbohydrate-active enzymes [28], LgtC is a member of family GT-8. *Neisseria meningitidis* LgtC is the only bacterial member of this CAZy family for which structural information is currently available. Sequence alignment with other members of CAZy family GT-8 shows that Cys246 in LgtC is part of a highly conserved 6-aa motif, which is present in several other GTs from different bacteria (SI, Fig. S2). Interestingly, these enzymes are not limited to galactosyltransferases, but also include GTs with other substrate specificities as well as GTs with unknown function.

Beyond family GT-8, structural information is currently available for a total of 48 GTs from bacterial organisms that have a human host. 23 of these structures, including *Neisseria meningitidis* LgtC, have been solved at a resolution of <3Å and contain a ligand in the active site (SI, Table S3). Analysis of these 23 structures reveals that all but

three, across 14 different CAZy families, have a non-catalytic cysteine at or near the active site. In 11 of these structures, this cysteine is within 8Å of the active site ligand (see Fig. 10 for selected examples, and SI, Table S3). This analysis demonstrates that the presence of non-catalytic, active site cysteines is widespread in bacterial GTs from both Gram-negative and Gram-positive organisms, independent of enzyme substrate specificity, CAZy family, overall fold type and biological function (Fig. 10).

Figure 10 Examples for bacterial GTs with non-catalytic active-site cysteines.[a]



[a]Proteins are shown in cartoon representation (wheat), active site residues are shown as lines (wheat), and the active-site cysteine as sticks (yellow). Substrates or substrate analogues are shown as sticks (green). Broken red lines indicate the distance from the α-carbon of the non-catalytic cysteine to the nearest atom of the respective ligand. **(A)** *E. coli* chondroitin polymerase (PDB 2Z87), donor: UDP-GalNAc, CAZy family: GT-2, general fold type: A; **(B)** *E. coli* glycogen synthase (PDB 2QZS) [40], donor: ADP-Glc, CAZy family: GT-5, general fold type: B; **(C)** *Bacteroides ovatus* BoGT6a (PDB 4CJC) [41], donor: UDP-GalNAc, CAZy family: GT-6, general fold type: A; **(D)** *Bradyrhizobium sp. WM9* fucosyltransferase NodZ (PDB 3SIX) [42], donor: GDP-Fuc, CAZy family: GT-23, general fold type: B; **(E)** *Campylobacter jejuni* sialyltransferase-II (PDB 2X61) [29], donor: CMP-Neu5Ac, CAZy family: GT-42, general fold type: A variant; **(F)** *Rhodobacter sphaeroides* cellulose synthase (PDB 4HG6) [43], donor: UDP-Glc, CAZy family: GT-2, general fold type: A.

A representative example is the bifunctional CMP-Neu5Ac:α-2,3/-2,8-sialyltransferase Cst-II from *Campylobacter jejuni* [29]. In Cst-II, the non-catalytic cysteine residue Cys30 is located to the rear of active site, at a distance of 6.0Å from the CMP ligand (SI, Fig. S3). Cst-II is a member of CAZy family GT-42. Sequence alignments within this family suggest that Cys30 in Cst-II is part of a conserved 10-aa sequence, which is also present in several other bacterial sialyltransferases in this family (SI, Fig. S4).

This analysis, in conjunction with our experimental results, suggests that the presence of non-catalytic active site cysteines can be systematically exploited for the discovery of new inhibitor classes against a broad range of bacterial GTs, including enzymes with very different, or indeed unknown, substrate specificities. In the latter case, such inhibitors would be of considerable use to annotate the biochemical and biological function of these GTs. Our results provide a blueprint for such an inhibitor discovery strategy. Such a strategy will be facilitated further by the recent publication of freely available, web-based protocols for covalent docking [25] and virtual screening [30].

Many of the bacterial GTs that are amenable to this general inhibitor discovery strategy are involved in bacterial virulence processes and complex pathogen/host interactions [31]. Covalent inhibitors would therefore be eminently useful as tool compounds for anti-virulence drug discovery [32], not only for inhibition studies, but also as affinity probes for the labelling and imaging of bacterial GTs *in vivo*. In preliminary experiments, we have already successfully used a fluorescent derivative of **1** for the labelling of LgtC (unpublished results). Results from this ongoing work will be reported in due course.

4. Experimental section

4.1 Chemistry

All chemical reagents were obtained commercially and used as received. Target compounds and synthetic intermediates were purified by flash chromatography column and characterized by TLC, ^{1}H-NMR, ^{13}C-NMR, and ESI-MS. Flash chromatography columns were packed wet. Thin layer chromatography (TLC) was performed on precoated aluminium plates (Silica Gel 60 F_{254}, Merck). Compounds were visualized by exposure to UV light (254/365 nm). NMR spectra were recorded on a Bruker BioSpin at 400 MHz (^{1}H) or 100 MHz (^{13}C). Mass spectra were recorded at the EPSRC National Mass Spectrometry Service Centre, Swansea. Analytical HPLC was carried out on Agilent 1260 Infinity machine equipped with an Agilent Poroshell 120 EC-C18 column (2.7 µm, 4.6 x 50 mm) under the following conditions: 0.1% formic acid in water/methanol, flow rate: 0.5 mL/min, detection wavelength: 254 nm. The purity of all target compounds was >95%.

4.1.1 5-Methyl-2-phenyl-2,4-dihydro-3H-pyrazol-3-one (**6**). Phenylhydrazine (219 mg, 2.03 mmol) and ethyl 3-oxobutanoate (264 mg, 2.03 mmol) were dissolved in glacial acetic acid. The reaction mixture was stirred at 110 ºC until TLC (hexane/EtOAc 1:1) showed complete consumption of the starting material. Upon cooling a white solid precipitated from the solution and was filtered and washed with ice-cold ethanol. Purification by flash column chromatography afforded the title compound [33] as a light yellow solid (343 mg, 1.97 mmol, 97 %). ^{1}H-NMR (400 MHz, $CDCl_3$) δ = 2.26 (s, 3H), 3.49 (s, 2H), 7.27-7.32 (m, 1H), 7.48-7.54 (m, 2H), 7.96-8.01 (m, 2H) ppm; ^{13}C-NMR (100 MHz, $CDCl_3$) δ = 16.9, 43.0, 118.8, 125.0, 128.8, 138.0, 156.4, 170.6 ppm.

4.1.2 1-Phenyl-3-(trifluoromethyl)-1H-pyrazol-5-ol (**7**). The title compound [34] was obtained as a white solid (652 mg, 2.86 mmol, 72 %) from phenylhydrazine (433 mg, 4 mmol) and ethyl 4,4,4-trifluoroacetoacetate (736 mg, 4 mmol) via the method described for **6**. ^{1}H-NMR (400 MHz, DMSO-*d6*) δ = 5.94 (s, 1H), 7.39 (t, 1H, *J* = 8.0 Hz), 7.52 (t, 2H, *J* = 8.0 Hz), 7.71 (d, 2H, *J* = 8.0 Hz), 12.49 (s, 1H) ppm; ^{13}C-NMR (100 MHz, DMSO-*d6*) δ = 85.6, 121.3 (q, $^{1}J_{CF}$ = 267 Hz), 122.3, 127.2, 129.1, 137.7, 140.4 (d, $^{2}J_{CF}$ = 37 Hz), 153.7 ppm.

4.1.3 (Z)-5-Methyl-2-phenyl-4-(2-(trifluoromethyl)benzylidene)-2,4-dihydro-3H-pyrazol-3-one (**1**). **6** (220 mg, 2.03 mmol) and 2-(trifluoromethyl)benzaldehyde (706 mg, 4.06 mmol) were dissolved in glacial acetic acid. The reaction was heated to reflux until TLC (hexane/EA 3:1) showed complete consumption of starting material. The reaction mixture was cooled and the product was isolated by column chromatography. The title compound was obtained as an orange solid (320 mg, 0.97 mmol, 48%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ = 2.36 (s, 3H), 7.18 (t, 1H, *J* = 8.0 Hz), 7.39 (t, 2H, *J* = 8.0 Hz), 7.59 (t, 1H, *J* = 8.0 Hz), 7.66 (t, 1H, *J* = 8.0 Hz), 7.77-7.78 (m, 2H), 7.90 (d, 2H, *J* = 8.0 Hz), 8.49 (d, 1H, *J* = 8.0 Hz) ppm; ^{13}C-NMR (100 MHz, $CDCl_3$) δ = 13.1, 118.9, 123.9 (q, $^{1}J_{CF}$ = 272 Hz), 125.1, 126.1 (q, $^{3}J_{CF}$ = 6 Hz), 128.8, 129.4 (d, $^{2}J_{CF}$ = 30 Hz), 129.7 (d, $^{3}J_{CF}$ = 2 Hz), 129.8, 131.3, 131.5, 132.7, 138.0, 141.3, 150.4, 161.0 ppm. ESI-MS: m/z 331.1 (100 %) $[M+H]^+$, 348.1 (8 %) $[M+NH_4]^+$; HR-MS: m/z 331.1058 $[M+H]^+$, $[C_{18}H_{14}F_3N_2O]^+$ calcd for 331.1053. HPLC: 98%.

4.1.4 (Z)-4-(3-(Benzyloxy)benzylidene)-2-phenyl-5-(trifluoromethyl)-2,4-dihydro-3H-pyrazol-3-one (**2**). **7** (57 mg, 0.25 mmol) and 3-(benzyloxy)benzaldehyde (64 mg, 0.30 mmol) were placed in a microwave-proof glass tube and heated for 15 min at 160 ^{o}C in a commercial microwave apparatus. The reaction was cooled to room temperature. The

reaction product was precipitated by addition of ethylacetate and hexane, collected by filtration, and recrystallized from hexane and ethylacetate. The title compound was obtained as orange solid (56 mg, 0.13 mmol, 53%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ = 5.23 (s, 2H), 7.27-7.33 (m, 2H), 7.35-7.39 (m, 1H), 7.41-7.45 (m, 2H), 7.47-7.53 (m, 5H), 7.76 (s, 1H), 7.84 (d, 1H, *J* = 8.0 Hz), 7.92-7.95 (m, 2H), 8.68 (t, 1H, *J* = 2.0 Hz) ppm; ^{13}C-NMR (100 MHz, $CDCl_3$) δ = 70.3, 118.1, 119.8 (q, $^{1}J_{CF}$ = 270 Hz), 120.1, 121.6, 122.9, 126.3, 127.9, 128.2, 128.5, 128.6, 129.0, 129.8, 133.8, 136.4, 137.5, 140.7 (d, $^{2}J_{CF}$ = 37 Hz), 150.6, 159.0, 161.1 ppm. ESI-MS: m/z 423.1 (91 %) $[M+H]^+$, 455.2 (100 %) $[M+MeOH+H]^+$, 477.1 (65 %) $[M+MeOH+Na]^+$; HR-MS: m/z 423.1314 $[M+H]^+$, $[C_{24}H_{18}F_3N_2O_2]$ calcd for 423.1315. HPLC: 97%.

4.1.5 (Z)-4-(3-Bromo-4-methoxybenzylidene)-2-phenyl-5-(trifluoromethyl)-2,4-dihydro-3H-pyrazol-3-one (**3**). The title compound was obtained as a white solid (80 mg, 0.19 mmol, 75%) from **7** (54 mg, 0.25 mmol) and 3-bromo-4-methoxybenzaldehyde (65 mg, 0.30 mmol) under the conditions described for **2**. ^{1}H-NMR (400 MHz, $CDCl_3$) δ = 4.05 (s, 3H), 7.06 (d, 1H, *J* = 12.0 Hz), 7.30 (t, 1H, *J* = 8.0 Hz), 7.48 (t, 2H, *J* = 8.0 Hz), 7.63 (s, 1H), 7.93 (d, 2H, *J* = 8.0 Hz), 8.71-8.74 (m, 1H), 8.88 (d, 1H, *J* = 4.0 Hz); ^{13}C-NMR (100 MHz, $CDCl_3$) δ = 57.8, 111.6, 112.3, 119.8, 119.9 (q, $^{1}J_{CF}$ = 270 Hz), 120.1, 126.3, 127.0, 129.0, 136.7, 137.6, 140.0, 140.6 (d, $^{2}J_{CF}$ = 37 Hz), 148.2, 160.9, 161.4 ppm. ESI-MS: m/z 425.0 (100 %) $[M+H]^+$, 479.0 (60 %) $[M+MeOH+Na]^+$; HR-MS: m/z 425.0107 $[M+H]^+$, $[C_{18}H_{13}BrF_3N_2O_2]^+$ calcd for 425.0107. HPLC: 98%.

*4.2 Biochemistry***.** LgtC was expressed and purified as previously described [26]. Recombinant LgtC was activated with DTT (10 mM, in HEPES buffer) in a 1:1 ratio for 30 min at 30 °C prior to each experiment, unless otherwise stated. For enzyme activity and inhibition experiments, we used a recently reported modification [20] of a

colorimetric glycosyltransferase assay originally developed by Wu and co-workers [35]. All assays were carried out in Nunc clear, flat-bottom 96-well microplates, unless otherwise stated.

4.3 Inhibition assays. For IC_{50} experiments, LgtC activity was adjusted to 20-50% turnover of UDP-Gal donor. We have previously shown that within this turnover range IC_{50} values are obtained reproducibly [20]. Assay components were added in the order required for the respective experiment (Methods A-D). All concentrations for the assay components are final concentrations.

4.3.1 Method A (standard assay protocol): no pre-incubation with inhibitor. Aliquots (15 µL each) of activated LgtC, $MnCl_2$ (5 mM), chicken egg-white lysozyme (CEL, 1 mg/mL), calf-intestinal alkaline phosphatase (CIP, 10 U/mL), Triton (0.01%) and HEPES buffer (13 mM, pH 7.0) were combined with lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control for non-specific hydrolysis) in the requisite microplate wells. Inhibitor at various concentrations in DMSO (15 µL, 10% final DMSO concentration) or DMSO only (15 µL, control) was added. Reactions were started by addition of UDP-Gal donor (15 µL, 28 µM) and incubated for 20 min at 30 °C. Reactions were stopped by addition of Malachite Green reagent A (30 µL) [20]. The microplate was shaken carefully, and Malachite Green Reagent B (30 µL) [20] was added. The colour was allowed to develop over 20 min, and the absorbance in each well was recorded at 620 nm on a Polarstar Optima plate reader (BMG Labtech). The absorbance measurements were used to calculate enzyme activity.

4.3.2 Method B: pre-incubation of LgtC with inhibitor, in the presence of donor. Aliquots (15 µL each) of activated LgtC, $MnCl_2$ (5 mM), CEL (1 mg/mL), CIP (10 U/mL), Triton

(0.01%) and HEPES buffer (13 mM, pH 7.0) were combined with inhibitor at various concentrations in DMSO (15 µL, 10% final DMSO concentration) or DMSO only (15 µL, control) in the requisite microplate wells. UDP-Gal donor (15 µL, 28 µM) was added, and the mixtures were pre-incubated for the requisite time at 30 °C. Lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control) were added, and the reactions were incubated for 20 min at 30 °C. The reactions were stopped and analysed as described.

4.3.3 Method C: pre-incubation of LgtC with inhibitor. Aliquots (15 µL each) of activated LgtC, $MnCl_2$ (5 mM), CEL (1 mg/mL), CIP (10 U/mL), Triton (0.01%) and HEPES buffer (13 mM, pH 7.0) were combined with inhibitor at various concentrations in DMSO (15 µL, 10% final DMSO concentration) or DMSO only (15 µL, control) in the requisite microplate wells. The mixtures were pre-incubated for the requisite time at 30 °C. UDP-Gal donor (15 µL, 28 µM) and lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control) were added, and the reactions were incubated for 20 min at 30 °C. The reactions were stopped and analysed as described.

4.3.4 Method D: pre-incubation of LgtC with inhibitor, in the presence of acceptor. Aliquots (15 µL each) of activated LgtC, $MnCl_2$ (5 mM), CEL (1 mg/mL), CIP (10 U/mL), Triton (0.01%) and HEPES buffer (13 mM, pH 7.0) were combined with inhibitor at various concentrations in DMSO (15 µL, 10% final DMSO concentration) or DMSO only (15 µL, control) in the requisite microplate wells. Lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control) were added, and the mixtures were pre-incubated for the requisite time at 30 °C. UDP-Gal donor (15 µL, 28 µM) was added, and the reactions were incubated for 20 min at 30 °C. The reactions were stopped and analysed as described.

4.4. Data analysis. On each microplate, a UDP calibration curve (0-12.5 µM) was included. Wells for the calibration curve comprised of all components of the standard reaction except for acceptor (replaced with buffer), inhibitor (replaced with DMSO), and UDP-Gal (replaced with UDP). For each experiment, absorbance at 620 nm (AU) was converted to [UDP] (µM) using linear regression of the respective calibration curve. A negative control (0 µM inhibitor) and a blank value (0 µM inhibitor, 0 µM acceptor) were included on each plate in triplicate. After linear regression, subtracting the blank from the negative control afforded the assay window. The background value for each inhibitor concentration (no acceptor, but otherwise identical components) was subtracted from each inhibitor concentration data point. Once corrected for the background, the absorbance in the presence of each inhibitor concentration was divided by the assay window and represented as a percentage. These percentage values were plotted against log [inhibitor] and analysed using GraphPad Prism (version 6) to afford relative IC_{50} values, if the data represented a sigmoidal curve. Each experiment was carried out at least in triplicate.

4.5. Inhibition kinetics. The kinetic parameters k_{obs}, K_i and k_{inact} for covalent inhibition by **2** and **3** were determined with the assay protocols described above, following the general approach described by Singh and co-workers [13]. LgtC activity was determined at various concentrations of **2** (0.5-100 µM) or **3** (0.5-50 µM) and in function of different pre-incubation times (0, 5, 10, 20 and 30 min). Enzyme activity was expressed as percentage of control (with DMSO) and plotted on a semi-logarithmic scale over pre-incubation time. From these plots, the observed rate constant k_{obs} at each inhibitor concentration was extracted. k_{obs} values were re-plotted over inhibitor concentrations and fitted to a hyperbolic equation ($k_{obs} = k_{inact} \times [I]/(K_i+[I])$) with GraphPad Prism (version

6) to obtain the values for K_i and k_{inact} as previously described [13]. All experiments were performed in triplicate.

4.6. Acceptor kinetics. The inhibition kinetics of **1** relative to acceptor were determined with the standard assay protocol (Method A) at a fixed concentration of UDP-Gal (100 μM) and variable concentrations of **1** (1, 2, 3, 4 and 5 μM) and lactose acceptor (5, 15 and 25 mM). LgtC activity was adjusted to <10% turnover of UDP-Gal donor. The type of enzyme inhibition was determined graphically by fitting the relevant set of data points to Dixon or Cornish-Bowden plots with GraFit 7 software.

4.7. LgtC mutants. C246A and C246S mutants of LgtC were prepared in the pCWori+ vector. Primers (Life Technologies) were designed using the Agilent primer design tool available online [36], and mutagenesis performed using the Quikchange Lightning II site-directed mutagenesis kit (Agilent). Mutations were confirmed by sequencing. The mutant plasmids were transformed into the BL21 (DE3) strain of *E. coli*, and incubated in 1 L of LB broth supplemented with 100 μg/mL ampicillin. Cells were incubated at 37 °C until the optical density at 600 nm reached 0.6, induced with 0.5 mM IPTG, and incubated overnight at 20 °C. Cells were harvested, resuspended in 25 mL of buffer A (20 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazole pH 8.0) supplemented with one tablet of Complete EDTA free protease inhibitor cocktail (Roche), and lysed using a sonicator. The sample was clarified by centrifugation at 20,000*g*, and the protein purified using an ÄKTAxpress automated purification system. The sample was first purified over a 1 mL HisTrap crude FF column (equilibrated with buffer A) and eluted with buffer A supplemented with 250 mM imidazole. It was then purified over a Superdex 200 16/600 hr column and eluted isocratically with 10 mM HEPES pH 7.0, 0.5 M NaCl. Glycerol was added to 20% (v/v) and the samples stored at -80 °C until use. The activity of both mutants and inhibition by

compound **1** (method A) were determined in our colorimetric glycosyltransferase assay [20].

4.8. Diafiltration. Aliquots (15 µL each) of activated LgtC, $MnCl_2$ (5 mM), CEL (1 mg/mL), CIP (10 U/mL), Triton (0.01%), HEPES buffer (13 mM, pH 7.0), and inhibitor in DMSO (100 µM) or DMSO only (control) were combined in individual wells of a 96-well microplate. UDP-Gal donor (15 µL, 28 µM) was added, and the reactions were incubated for 10 min at 30 °C. Lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control for non-specific hydrolysis) was added to each well, and the reactions were incubated for another 20 min at 30 °C. Samples for pre-diafiltration activity measurements were diluted with HEPES buffer at a ratio of 1:1 and analysed colorimetrically as described. Samples for diafiltration were collected in Vivaspin concentrators (MW cut-off: 10 kDa, total volume: 2-6 mL) and kept on ice prior to centrifugation. Samples (3 mL) were centrifuged at 4 °C (4000 rpm) for 50 min. The residual volume (300 µL) was diluted to 3 mL with HEPES buffer (pH 7.0) and centrifuged again for 50 min. The residual volume (300 µL) from the wash step was diluted to 450 µL with HEPES buffer (pH 7.0) and combined with aliquots (450 µL each) of $MnCl_2$ (5 mM), CEL (1 mg/mL), CIP (10 U/mL), Triton (0.01%) and HEPES buffer (13 mM, pH 7.0). To this master mix (90 µL), lactose acceptor (30 µL, 2 mM) or HEPES buffer (30 µL, control for non-specific hydrolysis) were added, followed by DMSO (15 µL). Reactions were started by addition of UDP-Gal donor (15 µL, 28 µM), incubated for 20 min at 30 °C, and analysed colorimetrically as described, to determine enzyme activity after diafiltration.

4.9. Time-dependent NMR experiments. A solution of the requisite amino acid (0.06 mmol; cysteine in d_6-DMSO, lysine and serine in D_2O) was added to a solution of the

requisite inhibitor (**1** or **2**, 0.06 mmol) in d_6-DMSO in a glass NMR tube. The solution was mixed thoroughly by shaking, and ^{1}H-NMR spectra were recorded at 298 K on a Bruker BioSpin machine at 400 MHz at various time points from 0-190 min.

4.10 Mass spectrometry. A LgtC stock solution (no DTT treatment) was diluted with HEPES buffer (13 mM, pH 7.0) and incubated with inhibitor (**1** and **2**: 100 µM in DMSO; **3**: 50 µM in DMSO) for 20 min at 30 °C (10% final DMSO concentration). Samples were directly analysed by electrospray mass spectrometry on a Waters LCT Premier machine (injection volume: 10 µL; desolvation temperature: 350 °C; source temperature: 120 °C).

4.11 Molecular docking protocol. Starting files for inhibitors **1**-**3** in .mol2 format were prepared in MarvinSketch [37]. Structures were energy minimized in UCSF Chimera version 1.8 [38] with the following settings: Steepest Descent Steps: 100, Steepest Descent Step size (A): 0.02, Conjugate Gradient Steps: 100, Conjugate Gradient Step size (A): 0.02, Update interval: 10, Fixed atoms: none. Covalent docking was carried out on the CovalentDock Cloud server [25] with the following settings: receptor PDB file 1GA8 [18], binding site Cys246, ligand: minimized starting files for inhibitors **1**-**3**. Docking results were analysed with MacPyMOL [39].

Acknowledgements

This study was supported by King's College London (King's China Award to Y.X., PhD studentship to N.G), the Norwich Research Park (PhD studentship to S.G.), and the Japanese Patent Office (visiting fellowship to M.E.). The plasmids for LgtC and β-1,4-GalT were generous gifts from Dr Warren Wakarchuk (Toronto) and Dr Christelle Breton (Grenoble). Mass spectrometry analysis of the LgtC/inhibitor adducts was carried out by Dr Lisa Haigh at Imperial College London. Mass spectrometry data for compounds **1**-**5** was acquired at the EPSRC UK National Mass Spectrometry Facility at Swansea University. We thank Dr Sarah Barry (King's College London) for many valuable discussions of the enzymological results, and Evie Dineva (University of Exeter) for assistance in cloning LgtC mutants.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/xxx

References

[1] Gloster, T.M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28**, 131-141 (2014).

[2] Petrou, V.I.; Herrera, C.M.; Schultz, K.M.; Clarke, O.B.; Vendome, J.; Tomasek, D.; Banerjee, S.; Rajashankar, K.R.; Dufrisne, M.B.; Kloss, B.; Kloppmann, E.; Rost, B.; Klug, C.S.; Trent, M.S.; Shapiro, L.; Mancia, F. Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **2016**, *351*, 608-612.

[3] Chen, Y.; Seepersaud, R.; Bensing, B.A.; Sullam, P.M.; Rapoport, T.A. Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E1190-E1199.

[4] Gloster, T.M.; Vocadlo, D.J. Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nat. Chem. Biol.* **2012**, *8*, 683-694.

[5] for a recent review of substrate-based glycosyltransferase inhibitors see e.g. Wang, S.; Vidal, S. Recent design of glycosyltransferase inhibitors. *Carbohydr. Chem.* **2013**, *39*, 78-101.

[6] Rillahan, C.D.; Antonopoulos, A.; Lefort, C.T.; Sonon, R.; Azadi, P.; Ley, K.; Dell, A.; Haslam, S.M.; Paulson, J.C. Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat. Chem. Biol.* **2012**, *8*, 661-668.

[7] Zandberg, W.F.; Kumarasamy, J.; Pinto, B.M.; Vocadlo, D.J. Metabolic Inhibition of Sialyl-Lewis X Biosynthesis by 5-Thiofucose Remodels the Cell Surface and Impairs Selectin-Mediated Cell Adhesion. *J. Biol. Chem.* **2012**, *287*, 40021-40030.

[8] Gloster, T.M., Zandberg, W.F.; Heinonen, J.E.; Shen, D.L.; Deng, L.H.; Vocadlo, D.J. Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nat. Chem. Biol.* **2011**, *7*, 174-181.

[9] Tedaldi, L.; Wagner, G.K. Beyond substrate analogues: new inhibitor chemotypes for glycosyltransferases. *MedChemComm* **2014**, *5*, 1106-1125.

[10] Ortiz-Meoz, R.F.; Jiang, J.; Lazarus, M.B.; Orman, M.; Janetzko, J.; Fan, C.; Duveau, D.Y.; Tan, Z.-W.; Thomas, C.J.; Walker, S. A Small Molecule That Inhibits OGT Activity in Cells. *ACS Chem. Biol.* **2015**, *10*, 1392-1397.

[11] Jiang, J.; Lazarus, M.B.; Pasquina, L.; Sliz, P.; Walker, S. A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nat. Chem. Biol.* **2012**, *8*, 72-77.

[12] Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S.J.; Jones, L.H.; Gray, N.S. Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Bio.* **2013***, 20*, 146-159.

[13] Singh, J.; Petter, R.C.; Baillie, T.A.; Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug. Discov.* **2011**, *10*, 307-317.

[14] Jost, C.; Nitsche, C.; Scholz, T.; Roux, L.; Klein, C.D. Promiscuity and Selectivity in Covalent Enzyme Inhibition: A Systematic Study of Electrophilic Fragments. *J. Med. Chem.* **2014**, *57*, 7590-7599.

[15] Serafimova, I.M.; Pufall, M.A.; Krishnan, S.; Duda, K.; Cohen, M.S.; Maglathlin, R.L.; McFarland, J.M.; Miller, R.M.; Frödin, M.; Taunton, J. Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **2012**, *8*, 471-476.

[16] Kwak, E.L.; Sordella, R.; Bell, D.W.; Godin-Heymann, N.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Driscoll, D.R.; Fidias, P.; Lynch, T.J.; Rabindran, S.K.; McGinnis, J.P.; Wissner, A.; Sharma, S.V.; Isselbacher, K.J.; Settleman, J.; Haber, D.A. Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7665-7670.

[17] Hagel, M.; Niu, D.; St Martin, T.; Sheets, M.P.; Qiao, L.; Bernard, H.; Karp, R.M.; Zhu, Z.; Labenski, M.T.; Chaturvedi, P.; Nacht, M.; Westlin, W.F.; Petter, R.C.; Singh, J. Selective irreversible inhibition of a protease by targeting a non-catalytic cysteine. *Nat. Chem. Biol.* **2011**, *7*, 22-24.

[18] Persson, K.; Ly, H.D.; Dieckelmann, M.; Wakarchuk, W.W.; Withers, S.G.; Strynadka, N.C.J. Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.* **2001**, *8*, 166-175.

[19] Erwin, A.L.; Allen, S.; Ho, D.K.; Bonthius, P.J.; Jarisch, J.; Nelson, K.L.; Tsao, D.L.; Unrath, W.C.T.; Watson, M.E.; Gibson, B.W.; Apicella, M.A.; Smith, A.L. Role of l*gtC* in resistance of nontypeable *Haemophilus influenzae* strain R2866 to human serum. *Infect. Immun.* **2006**, *74*, 6226-6235.

[20] Tedaldi, L.; Evitt, A.; Göös, N.; Jiang, J.; Wagner, G.K. A practical glycosyltransferase assay for the identification of new inhibitor chemotypes. *MedChemComm* **2014***, 5,* 1193-1201.

[21] Moreau, F.; Desroy, N.; Genevard, J.M.; Vongsouthi, V.; Gerusz, V.; Le Fralliec, G.; Oliveira, C.; Floquet, S.; Denis, A.; Escaich, S.; Wolf, K.; Busemann, M.; Aschenbrenner, A. Discovery of new Gram-negative antivirulence drugs: Structure and properties of novel E. coli WaaC inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4022-4026.

[22] Hu, Y.N.; Helm, J.S.; Chen, L.; Ginsberg, C.; Gross, B.; Kraybill, B.; Tiyanont, K.; Fang, X.; Wu, T.; Walker, S. Identification of Selective Inhibitors for the Glycosyltransferase MurG via High-Throughput Screening. *Chem. Biol.* **2004***, 11*, 703-711.

[23] Jadhav, A.; Ferreira, R.S.; Klumpp, C.; Mott, B.T.; Austin, C.P.; Inglese, J.; Thomas, C.J.; Maloney, D.J.; Shoichet, B.K.; Simeonov, A. Quantitative Analyses of

Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of a Thiol Protease. *J. Med. Chem.* **2010**, *53*, 37-51.

[24] Shannon, D.A.; Weerapana, E. Covalent protein modification: the current landscape of residue-specific electrophiles. *Curr. Opin. Chem. Biol.* **2015**, *24*, 18-26.

[25] Ouyang, X.; Zhou, S.; Su, C.T.T.; Ge, Z.; Li, R.; Kwoh, C.K. CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constrains. *J. Comput. Chem.* **2013**, *34*, 326-336.

[26] Descroix, K.; Pesnot, T.; Yoshimura, Y.; Gehrke, S.; Wakarchuk, W.; Palcic, M.M.; Wagner, G.K. Inhibition of galactosyltransferases by a novel class of donor analogues. *J. Med. Chem.* **2012**, *55*, 2015-2024.

[27] Baell, J. B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) From Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719-2740.

[28] Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2013**, *42*, D490-D495.

[29] Lee, H.J.; Lairson, L.L.; Rich, J.R.; Lameignere, E.; Wakarchuk, W.W.; Withers, S.G.; Strynadka, N.C.J. Structural and kinetic analysis of substrate binding to the sialyltransferase Cst-II from Campylobacter jejuni. *J. Biol. Chem.* **2011**, *286*, 35922-35932.

[30] London, N.; Miller, R.M.; Krishnan, S.; Uchida, K.; Irwin, J.J.; Eidam, O.; Gibold, L.; Cimermancic, P.; Bonnet, R.; Shoichet, B.K.; Taunton, J. Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **2014**, *10*, 1066-1072.

[31] Tytgat, H.L.P.; Lebeer, S. The Sweet Tooth of Bacteria: Common Themes in Bacterial Glycoconjugates. *Mol. Microbiol. Rev.* **2014**, *78*, 372-417.

[32] Allen, R.C.; Popat, R.; Diggle, S.P.; Brown, S.P. Targeting virulence: can we make evolution-proof drugs? *Nat. Rev. Microbiol.* **2014**, *12*, 300-308.

[33] Panda, M., Ramachandran, S., Ramachandran, V., Shirude, P.S., Humnabadkar, V., Nagalapur, K., Sharma, S., Kaur, P., Guptha, S., Narayan, A., Mahadevaswamy, J., Ambady, A., Hegde, N., Rudrapatna, S.S., Hosagrahara, V.P., Sambandamurthy, V.K. & Raichurkar, A. Discovery of pyrazolopyridones as a novel class of noncovalent DprE1 inhibitor with potent anti-mycobacterial activity. *J. Med. Chem.* **57**, 4761-4771 (2014).

[34] Gilbert, A. M., Bursavich, M. G., Lombardi, S., Georgiadis, K.E., Reifenberg, E., Flannery, C.R. & Morris, E.A. 5-((1*H*-pyrazol-4-yl)methylene)-2-thioxothiazolidin-4-one inhibitors of ADAMTS-5. *Bioorg. Med. Chem. Lett.* **17**, 1189-1192 (2007).

[35] Wu, Z.L.; Ethen, C.M.; Prather, B.; Machacek, M.; Jiang, W. Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology* **2011**, *21*, 727-733.

[36] http://www.genomics.agilent.com/primerDesignProgram.jsp

[37] MarvinSketch (version 5.12.4) for Mac OS X 10.7.5; www.chemaxon.com

[38] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

[39] The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

[40] Sheng, F.; Jia, X.; Yep, A.; Preiss, J.; Geiger, J.H. The crystal structures of the open and catalytically competent closed conformation of Escherichia coli glycogen synthase. *J. Biol. Chem.* **2009**, *284*, 17796-17807.

[41] Pham, T.T.K.; Stinson, B.; Thiyagarajan, N.; Lizotte-Waniewski, M.; Brew, K.; Acharya, K.R. Structures of complexes of a metal-independent glycosyltransferase GT6 from Bacteroides ovatus with UDP-N-acetylgalactosamine (UDP-GalNAc) and its hydrolysis products. *J. Biol. Chem.* **2014**, *289*, 8041-8050.

[42] Brzezinski, K.; Dauter, Z.; Jaskolski, M. Structures of NodZ α-1,6-fucosyltransferase in complex with GDP and GDP-fucose. *Acta Crystallogr. Sect. D* **2012**, *68*, 160-168.

[43] Morgan, J.L.; Strumillo, J.; Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **2012**, *493*, 181-186.

Author contributions

YX and SG synthesised the inhibitors; YX, ME and NG carried out the biochemical experiments; RS and GKW carried out the analysis of sequence and structural data and the docking experiments; MV and NH designed, generated and prepared the LgtC mutants; GKW designed the study, supervised the research and wrote the manuscript, with contributions from YX, RS and NH.

The bacterial galactosyltransferase LgtC... **.... and its covalent inhibition.**