Central Lancashire Online Knowledge (CLoK)

Title	Structure based inhibitor design targeting glycogen phosphorylase b. Virtual screening, synthesis, biochemical and biological assessment of novel N-acyl-β-d-glucopyranosylamines
Type	Article
URL	https://clok.uclan.ac.uk/13291/
DOI	https://doi.org/10.1016/j.bmc.2014.06.058
Date	2014
Citation	Parmenopoulou, Vanessa, Kantsadi, Anastassia L., Tsirkone, Vicky G., Chatzileontiadou, Demetra S.M., Manta, Stella, Zographos, Spyros E., Molfeta, Christina, Archontis, Georgios, Agius, Loranne et al (2014) Structure based inhibitor design targeting glycogen phosphorylase b. Virtual screening, synthesis, biochemical and biological assessment of novel N-acyl-β-d-glucopyranosylamines. Bioorganic & Medicinal Chemistry, 22 (17). pp. 4810-4825. ISSN 0968-0896
Creators	Parmenopoulou, Vanessa, Kantsadi, Anastassia L., Tsirkone, Vicky G., Chatzileontiadou, Demetra S.M., Manta, Stella, Zographos, Spyros E., Molfeta, Christina, Archontis, Georgios, Agius, Loranne, Hayes, Joseph M., Leonidas, Demetres D. and Komiotis, Dimitri

It is advisable to refer to the publisher's version if you intend to cite from the work.
https://doi.org/10.1016/j.bmc.2014.06.058

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

Structure based inhibitor design targeting glycogen phosphorylase b. Virtual screening, synthesis, biochemical and biological assessment of novel *N*–(*β*-D-glucopyranosyl) amides

Vanessa Parmenopoulou[a#] Anastassia L. Kantsadi[a#], Vicky G. Tsirkone[a#§], Demetra S.M. Chatzileontiadou[a], Stella Manta[a], Spyros E. Zographos[b], Nikolaos Kollatos[a], Georgios Archontis[c], Loranne Agius[d], Joseph M. Hayes[e], Demetres D. Leonidas[a*] and Dimitri Komiotis[a,*]*

[a]Department of Biochemistry and Biotechnology, University of Thessaly, 26 Ploutonos Str. 41221 Larissa, Greece

[b]Institute of Biology, Medicinal Chemistry and Biotechnology, National Hellenic Research Foundation, Athens, Greece

[c]Department of Physics, University of Cyprus, Nicosia, Cyprus.

[d]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, U.K.

[e]Centre for Materials Science, Division of Chemistry, University of Central Lancashire, Preston PR1 2HE, U.K.

[§]Current work address: Laboratory for Biocrystallography, Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Belgium.
[#]Authors contributed equally to this work

* To whom correspondence should be addressed:

Assoc. Prof. Demetres D. Leonidas, Department of Biochemistry and Biotechnology, University of Thessaly, 26 Ploutonos Str. 41221 Larissa, Greece. Phone: +30 2410 565278; Fax: +30 2410 565290; e-mail: ddleonidas@bio.uth.gr.

Prof. Dimitri Komiotis, Laboratory of Bio-organic Chemistry, Department of Biochemistry and Biotechnology, University of Thessaly, 26 Ploutonos Str., 41221 Larissa, Greece. Tel.+302410 565285; Fax. +302410 565290; email: dkom@bio.uth.gr

Dr Joseph M. Hayes, Centre for Materials Science, Division of Chemistry, University of Central Lancashire, Preston PR1 2HE, United Kingdom. Tel. +441772894334; Fax: +44 (0) 1772 894981; email: jhayes@uclan.ac.uk

Abstract

Glycogen phosphorylase (GP) is a validated target for the development of new type 2 diabetes treatments. Exploiting the Zinc docking database, we report the *in silico* screening of 1888 β-D-glucopyranose-NH-CO-R putative GP inhibitors differing only in their R groups. CombiGlide and GOLD docking programs with different scoring functions were employed with the best performing methods combined in a "consensus scoring" approach to ranking of ligand binding affinities for the active site. Six selected candidates from the screening were then synthesized and their inhibitory potency was assessed both *in vitro* and *ex vivo*. Their inhibition constants' values, *in vitro*, ranged from 5 to 377 µM while two of them were effective at causing inactivation of GP in rat hepatocytes at low µM concentrations. The crystal structures of GP in complex with the inhibitors were defined and provided the structural basis for their inhibitory potency and data for further structure based design of more potent inhibitors.

1. Introduction

Glycogen phosphorylase (GP), an enzyme that catalyzes the first step of intracellular degradation of glycogen to Glc-1-P [1], has been an important target for the discovery of novel hypoglycaemic agents [1-4]. The efficacy of such inhibitors on blood glucose control and hepatic glycogen balance has been confirmed from biological studies [5-9]. The main physiological inhibitor of GP is glucose, albeit a weak inhibitor (Ki= 2-3 mM). Inhibition of the catalytic site of the enzyme has therefore been probed with glucose based analogues[10], designed on the basis of information derived from the crystal structure of the T-state GPb-α-D-glucose complex [11]. These inhibitors mainly have α- and β-substitutions at the anomeric C1 atom of the glucose moiety [2, 10] with the β-substitutions aiming to exploit the catalytic subsite called β-cavity, an empty space at the β-1-C configuration lined by both polar and non-polar groups. A key structural feature of glucose based ligands is the so called linker which connects the glucose moiety with an R group (Scheme 1). One such linker is the -NH-CO- group and has led to the identification of ligands in the μM range for GP inhibition when following the general reaction scheme shown in Scheme 2 [2, 12]. Using this linker, however, only selected R groups have been previously studied towards exploiting the β-cavity features. Hence in the current work we have adopted a more systematic approach and used a virtual screening methodology to discover new –R groups. Docking using both the Glide (CombiGlide) [13, 14] and GOLD[15] programs with different scoring functions was employed for this purpose, with the final ranking of ligand potencies based on a consensus scoring approach [16] combining the best performing scoring functions based on calculations for a training set of ligands. The diverse set of R groups for virtual screening was derived from acid chloride (RCOCl) molecules downloaded from the Zinc docking database [17]. Selected candidates from the virtual screening were then experimentally evaluated. Absorption, distribution, metabolism and excretion (ADME) property predictions from the QikProp program [14]

allowing us to select a final set of beta-D-glucopyranose-NH-CO-R ligands with drug-like properties as a criterion for further evaluation [18, 19]. Thus, we also report here the synthesis of novel *N*-acyl-*β*-D-glucopyranosylamine type inhibitor series (**S**) shown in Table 1 and their biochemical and *ex vivo* assessment. We have determined the crystal structures of their complexes with GPb to analyze the molecular elements that govern molecular recognition by the catalytic site of the enzyme with emphasis in the *β*-pocket environment. The common kinetic and structural features of these compounds are that they are highly selective for GPb and they are competitive inhibitors with respect to the substrate Glc-1-P. They bind at the catalytic site and stabilise the T-state conformation (closed) of the 280s loop by making several favourable contacts with residues of this loop. Furthermore, compounds were designed to fill an empty pocket at the *β*-1-*C* configuration of glucose [20], denoted as the *β*-channel, with bulky non-polar groups.

Three of the best N-acyl-*β*-D-glucopyranosylamines inhibitors of GP have been identified in this work (< 10 µM activity), which also have revealed promising activity in the cell in agreement with ADME predictions. The present work constitutes the first integrated structure-based drug design with high-throughput screening of commercially available small molecule libraries for glycogen phosphorylase inhibitors (GPIs). It also highlights the modeling perspective on GP and inhibitor and how calculations can support experiment in a more rational and time-efficient approach to design of potent, drug-like GPIs.

2. Materials and methods

2.1. Virtual Screening

2.1.1 Protein Preparation. The GPb receptor structure was prepared for modeling using the 2.15 Å resolution GPb-*N*-4-phenylbenzoyl-2-*N*'-*β*-D-glycopyranosyl urea crystal complex (PDB ID: 2QLN). This receptor was chosen based on its previous performance in docking studies, its co-crystallized ligand having a bulkier R substituent which opens up the β-cavity

in regions complementary to binding of the other ligands [4]. Schrödinger's 'Protein Preparation Wizard'[14] was used for the preparation with water molecules within 5 Å of the ligand initially retained, but deleted for subsequent docking. Bond orders were assigned and hydrogen atoms added, with protonation states for basic and acidic residues based on residue pKa's at normal pH (7.0). However, subsequent optimization of hydroxyl groups, histidine C/N atom "flips" and protonation states, and side chain O/N atom "flips" of Asn and Gln residues was based on optimizing hydrogen bonding patterns, so that the final assignments were checked on visual inspection of the protein. In particular, all final His residues were assigned as neutral, either in a HIE or HID state. The PLP co-factor was used in its mono-anion form. Finally, the system was softly minimized using Impref and the OPLS-AA(2005) forcefield [21] with the root-mean-square deviation (RMSD) of heavy atoms kept within 0.3 Å of the crystallographic positions.

2.1.2 Ligand Preparation The training set of 40 ligands (Table S1*)* was selected from previous studies [2, 22] with K_i values ranged from 0.35 µM to no inhibition. For consistency, a similar protocol for preparation of both the training and screening set ligands was used. In effect, the training set –R groups were first built as acid chlorides using Maestro and LigPrep. The acid chlorides were then "reacted" at a β-D-glucopyranose-NH₂ attachment point (-NH₂) using modules from the combinatorial screening CombiGlide program of Schrodinger [14] in a manner that reflected the actual synthetic routine (Scheme 2). For the virtual screening set, the acid chloride molecules were downloaded in pre-determined low energy conformations (any relevant tautomers/ionization states included) from the Zinc database. To ensure that the final ligands were 'drug-like', the following criteria were used as filters: the structural pattern - *C(Cl)=O, 0 ≤ MW ≤ 250 amu and 0 ≤rotatable bonds ≤ 4*. A 1888 ligand screening set resulted. In effect, the CombiGlide workflow provided a 3D minimized structure of a molecule that contains a "core" (in this case D-glucose) and a defined attachment point for a

set of reagents. CombiGlide exploits the program LigPrep for this purpose with the final structures ready to be employed in the docking studies.

2.1.3 CombiGlide Docking In the docking calculations with CombiGlide 2.6 [14], the shape and properties of the catalytic binding site were mapped onto grids with dimensions of 28.6 Å x 28.6 Å x 28.6 Å centred on the native 2QLN ligand. Standards parameters were applied including van der Waals scaling for non-polar atoms (by 0.8) to include modest ''induced-fit'' effects. Positional constraints were applied on the hydroxyl (-OH) groups of glucose [23]. Glide standard (SP) and extra-precision (XP) modes were used.

2.1.4 GOLD Docking In the docking calculations with GOLD 5.0 [15], all atoms and their associated residues within 10Å of the ligand were used to define the active site. The information about the ligand hydrogen-bonding interactions and conformation was encoded into the corresponding genetics algorithms (GA) of GOLD GoldScore, ChemScore, ASP and ChemPLP using the default GA parameters. Scaffold constraints were applied for the glucose ring atoms with weight 1.0 Å.

2.1.5 Consensus Scoring In combining the results from individual docking methods, the ranks from the best performing scoring functions were added to give a rank-sum for each ligand, otherwise known as a Simple Sum Rank [16].

2.1.6 Statistical Analysis of Docking Performance For the training set, performance with respect to recovery of the actives (top 10 inhibitors) was analyzed in a statistical manner [24-27]. Since only a small fraction of a database is tested experimentally in practical VS studies, it is important to recognize actives or leads as "early" as possible. We therefore examined several metrics ranging from the commonly used hit rate and enrichment factor, to those that better highlight "early recognition". Hit-rate is the simplest metric and represented the % of actives in the predicted top 10 ligands for a particular docking method. The EF may be defined as the ratio of the number of actives retrieved relative to the number of database molecules tested.

An EF is often described with respect to a given percentage of the database screened. In general, if n represents the total number of actives and N represents the total number of molecules in the database, then if there are n_a actives among the $N_{x\%}$ molecules in the first $x\%$ of the database screened, the EF is given by Mackey and Melville [28]

$$EF_{x\%} = \frac{n_a/N_{x\%}}{n/N} \quad Eq.(1)$$

In our case we are looking at simply recovery of $n_a = 10$ actives in a set of N=40 ligands, where x%=25% A ROC plot is conceptually similar to an enrichment plot and shows the relationship between the true positive rate (*TPR, or sensitivity*) and the false positive rate (*FPR, or specificity*) [29]. The Area Under the ROC curve (AU-ROC) is a common way to summarize the overall quality of a ROC plot.

$$AUC = \frac{1}{n}\sum_{i=1}^{n}(1 - f_i) \quad Eq.(2)$$

where f_i is the fraction of decoys or non-actives ranked higher than the ith active. The value of the AU-ROC varies between 0 and 1, where 1 represents a perfect ranking (all actives ranked above the decoys) while 0.5 corresponds to a random ranking. However, EF and AU-ROC values do not distinguish early and late performance.[24] Accordingly, we also used the metric "sum of logarithms of ranks" (**SLR**) which emphasizes on the "early recognition" of actives. The SLR metric is calculated as:

$$SLR = -\sum_{i=1}^{n} log\left(\frac{r_i}{N}\right) \quad Eq.(3)$$

where r_i is the rank of the ith active. The negative logarithm emphasizes early recognition. Noting that for an ideal case, a VS method would rank all actives within the first n positions, a theoretical maximum SLR may be calculated as:

$$SLR_{max} = -\sum_{i=1}^{n} log\left(\frac{i}{N}\right) \quad Eq.(4)$$

This allows a normalized SLR (NSLR) to be calculated as:

$$NSLR = \frac{SLR}{SLR_{max}} \quad Eq.(5)$$

This metric ranges from 0 to 1 (best achievable ranking).

2.1.7 Screening Set Analysis/Candidate Selection

a) Consensus Scoring: Based on the training set results and a statistical evaluation of the performance of the different docking methods and their combination in consensus scoring, the protocol employed for final re-ranking of the screening set ligands was a Simple Sum Rank [16, 30, 31] of Glide-SP and GOLD-ChemPLP individual rankings. A Simple Sum Rank value of <150 was chosen as the cut off for the 'final set' of ligands. Votes (coarse quantite voting) [16] received by the final set of ligands from other scoring functions/algorithms if ranked in the top ~ 1.5% and top ~4% with these methods were considered as a criteria in selecting ligands for synthesis; a minimum–maximum of 0-3 votes was possible, one from each of GOLD-ChemScore, -GoldScore, ASP.

b) ADME property predictions: QikProp from Schrödinger [14] was used for ADME properties predictions allowing us to select ligands with drug-like properties as an extra criteria.[18, 19] Specifically, Lipinski's rules of five [32] [MW<500, logP <5, HBA<10, and HBD<5] and Jorgensen's rule of three [33] [QPlogS > -5.7, QP PCaco > 22 nm/s, number of Primary Metabolites < 7] were applied to predict oral bioavailability. Default settings were employed for these calculations.

*c) Ligand 2D Fingerprint Calculations***:** For the "final set" of ligands based on Simple Sum Rank (<150), 2D fingerprints for the unbound ligands (type: linear with atoms distinguished

by ring size, aromaticity, hydrogen bond acceptor/donor (HBA/HBD), ionization potential, whether terminal, whether halogen and bonds distinguished by bond order) were calculated using the related Schrodinger cheminformatics python script. Canvas similarity (Tanimoto coefficient) and clustering of the fingerprints (average linkage method) was then computed.[14] The result was grouping of the final 29 compounds into 15 families (Table S3).

2.2 Chemistry

Thin layer chromatography (TLC) was performed on Merck precoated 60F254 plates. Reactions were monitored by TLC on silica gel, with detection by UV light (254nm) or by charring with sulfuric acid. Flash column chromatography was performed using silica gel (240-400 mesh, Merck). ^{1}H and ^{13}C NMR spectra were obtained at room temperature with a Bruker 400 spectrometer at 400 and 100 MHz, respectively using chloroform-*d* (CDCl$_3$) and dimethylsulfoxide-*d$_6$* (DMSO-*d$_6$*) with internal tetramethylsilane (TMS). Chemical shifts (*δ*) were given in ppm measured downfield from TMS, and spin–spin coupling constants are in Hz. Mass spectra were obtained on a ThermoQuest Finnigan AQA Mass Spectrometer (electrospray ionization). Optical rotations were measured using an Autopol I polarimeter. The IR spectra were obtained with a Thermo Scientific Nicolet IR100 FTIR spectrometer.

All coupling reactions were performed under an atmosphere of oxygen-free nitrogen, and the solvent (or solution) was vigorously deoxygenated with N$_2$ prior to addition of catalysts. DMF was distilled from calcium hydride and stored over 3E molecular sieves. Azide **AcGlc-N$_3$** was prepared according to the procedures described in literature and its chemical and physical properties were in agreement with previous data [34].

*2.2.1 General procedure for preparation of the glucopyranosyl amine (**AcGlc-NH$_2$**).* After two vacuum/H$_2$ cycles to remove air from the reaction tube, the stirred mixture of azide **AcGlc-N$_3$** (5.00 g, 13.4 mmol), 10% Pd/C (0.95 g) in ethyl acetate (EtOAc, 184.6 mL) was hydrogenated at ambient pressure (balloon) and temperature (ca. 20 $^\circ$C) for 24 h. The reaction

mixture was filtered and the filtrate was concentrated in vacuum. The residue was purified by flash column chromatography using EtOAc. The spectroscopic properties of amine **AcGlc-NH₂** were in agreement with previous data [35].

2.2.2 General procedure for preparation of the protected amides (AcS). To a stirred mixture of β-D-glucopyranosyl amine (**AcGlc-NH₂**) (1.44 mmol) in 3 ml anhydrous DMF and Et₃N (0.96 mmol) the corresponding acylating agents RCOCl (**1, 2, 3, 6, 20, 21**) (0.96 mmol) were added in one portion. The reaction mixture was left for 1 hour at room temperature, then concentrated and the residue was purified by flash chromatography to give the desired analogues (**AcS**) as foams.

2.2.2.1.   N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-4(prop-1-enyl)biphenyl   amide (AcS1). White foam, 45% yield; Rf = 0.30 (*n*-hexane/EtOAc 1:1); $[\alpha]_D^{22}$ = +6 (*c* = 0.20, CHCl₃); ¹H NMR (400 MHz, CDCl₃): δ 7.69 (d, 1H, *J* = 15.7 Hz, CH=C̲H̲), 7.64-7.37 (m, 9H, 2Ph), 6.42 (d, 1H, $J_{1,2}$ = 9.4 Hz, H-1), 6.36 (d, 1H, C̲H̲=CH), 5.44-5.33 (m, 2H, H-2, NH), 5.11 (t, 1H, *J* = 9.6 Hz, H-3), 5.01 (t, 1H, *J* = 9.6 Hz, H-4), 4.35 (dd, 1H, $J_{5,6a}$ = 4.3 Hz, $J_{6a,6b}$ = 12.6 Hz, H-6a), 4.11 (dd, 1H, $J_{5,6b}$ = 1.9 Hz, H-6b), 3.93-3.86 (m, 1H, H-5), 2.09, 2.06, 2.05, 2.04 (4s, 12H, 4OAc); ¹³C NMR (100 MHz, CDCl₃) δ 171.3, 170.6, 169.9, 169.6, 165.8, 143.1, 142.9, 140.1, 133.1, 128.9, 128.6, 127.8, 127.5, 127.0, 119.0, 78.5, 73.5, 72.6, 70.7, 68.1, 61.6, 20.8, 20,7, 20.6; MS (ESI) m/z: 554.21 [M+H]⁺. Anal. (C₂₉H₃₁NO₁₀): C, 62.92; H, 5.64; N, 2.53. Found: C, 63.10; H, 5.52; N, 2.74.

2.2.2.2. N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-6-butyl-1,2,3,4-tetrahydronapthalene amide (AcS2). Yellow foam, 42% yield; Rf = 0.6 (*n*-hexane/EtOAc 1:1); $[\alpha]_D^{22}$ = -8 (*c* = 0.20, CHCl₃); ¹H NMR (400 MHz, CDCl₃): δ 6.99-6.86 (m, 3H, Ph), 6.16 (d, 1H, $J_{1,2}$ = 9.3 Hz, H-1), 5.34-5.24 (m, 2H, H-2, NH), 5.06 (t, 1H, *J* = 9.6 Hz, H-3), 4.91 (t, 1H, *J* = 9.6 Hz, H-4), 4.32 (dd, 1H, $J_{5,6a}$ = 4.3 Hz, $J_{6a,6b}$ = 12.6 Hz, H-6a), 4.07 (dd, 1H, $J_{5,6b}$ = 1.9 Hz, H-6b), 3.84-3.79 (m, 1H, H-5), 2.73 (s, 4H, cyclohexane), 2.53 (t, 2H, *J* = 7.4 Hz, CH₂), 2.22-2.15 (m, 2H,

CH$_2$), 2.07, 2.03, 2.02, 2.01 (4s, 12H, 4OAc), 1.94-1.87 (m, 2H, CH$_2$), 1.78 (m, 4H, cyclohexane); ^{13}C NMR (100 MHz, CDCl$_3$) δ 173.1, 171.0, 170.6, 169.9, 169.6, 138.1, 137.1, 134.8, 129.1, 129.0, 125.5, 78.0, 73.5, 72.6, 70.5, 68.0, 61.6, 35.8, 34.5, 29.3, 29.0, 26.6, 23.2, 23.1, 20.7, 20.6, 20.5; MS (ESI) m/z: 548.63 [M+H]$^+$. Anal. (C$_{28}$H$_{37}$NO$_{10}$): C, 61.41; H, 6.81; N, 2.56. Found: C, 61.75; H, 7.14; N, 2.30.

2.2.2.3. N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-4-ethoxybiphenyl amide (AcS3). Yellow foam, 55% yield; Rf = 0.33 (*n*-hexane/EtOAc 1:1); [α]$_D^{22}$ = -6 (*c* = 0.20, CHCl$_3$); ^{1}H NMR (400 MHz, DMSO-*d$_6$*): δ 8.93 (d, 1H, $J_{1,NH}$ = 9.4 Hz, NH), 7.62-7.58 (m, 4H, Ph), 7.42 (t, 2H, J = 7.3 Hz, Ph), 7.32 (t, 1H, Ph), 6.99 (d, 2H, J = 8.8 Hz, Ph), 5.48 (t, 1H, $J_{1,2}$ = 9.3 Hz, H-1), 5.37 (t, 1H, J = 9.5 Hz, H-3), 5.01-4.88 (m, 2H, H-2, H-4), 4.59 (s, 2H, CH$_2$), 4.19-4.09 (m, 2H, H-6), 4.00-3.95 (m, 1H, H-5), 1.98, 1.97, 1.91, 1.84 (4s, 12H, 4OAc); ^{13}C NMR (100 MHz, CDCl$_3$) δ 170.6, 169.9, 169.5, 169.2, 156.4, 140.2, 135.3, 128.8, 128.4, 127.0, 126.7, 114.8, 77.9, 73.6, 72.5, 70.0, 67.9, 67.1, 61.5, 20.7, 20.5, 20.3; MS (ESI) m/z: 558.43 [M+H]$^+$. Anal. (C$_{28}$H$_{31}$NO$_{11}$): C, 60.32; H, 5.60; N, 2.51. Found: C, 60.72; H, 5.26; N, 2.40.

2.2.2.4. N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-1-isopropyl-4(prop-1enyl)benzene amide (AcS6). White foam, 44% yield; Rf = 0.53 (*n*-hexane/EtOAc 1:1); [α]$_D^{22}$ = +10 (*c* = 0.20, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.63 (d, 1H, J = 15.7 Hz, CH=<u>CH</u>), 7.44 (d, 2H, J = 8.2 Hz, o-Ph), 7.25 (d, 2H, J = 8.4 Hz, m-Ph), 6.33 (m, 2H, $J_{1,2}$ = 8.5 Hz, H-1, <u>CH</u>=CH), 5.43-5.33 (m, 2H, H-2, NH), 5.10 (t, 1H, J = 9.4 Hz, H-3), 5.00 (t, 1H, J = 9.6 Hz, H-4), 4.34 (dd, 1H, $J_{5,6a}$ = 4.4 Hz, $J_{6a,6b}$ = 12.5 Hz, H-6a), 4.10 (dd, 1H, $J_{5,6b}$ = 1.7 Hz, H-6b), 3.91-3.86 (m, 1H, H-5), 2.98-2.88 (m, 1H, CH), 2.08, 2.06, 2.04, 2.03 (4s, 12H, 4OAc), 1.29, 1.27 (2s, 6H, 2CH$_3$); ^{13}C NMR (100 MHz, CDCl$_3$) δ 171.3, 170.6, 169.9, 169.6, 166.0, 151.6, 143.3, 131.8, 128.2, 127.0, 118.2, 78.5, 73.5, 72.7, 70.7, 68.1, 61.6, 34.1, 23.8, 20.7, 20.6, 20.5; MS (ESI) m/z: 520.58 [M+H]$^+$. Anal. (C$_{26}$H$_{33}$NO$_{10}$): C, 60.11; H, 6.40; N, 2.70. Found: C, 60.51; H, 6.15; N, 2.31.

2.2.2.5. N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-1-sec-butyl-4-ethylbenzene amide (AcS20). Yellow foam, 41% yield; Rf = 0.42 (*n*-hexane/EtOAc 3:2); $[\alpha]_D^{22}$ = -2 (*c* = 0.20, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.14-7.11 (m, 4H, Ph), 6.23 (dd, 1H, $J_{1,2}$ = 9.2 Hz, H-1), 5.32-5.17 (m, 2H, H-2, NH), 5.08-5.01 (m, 1H, H-3), 4.92-4.82 (m, 1H, H-4), 4.32 (dd, 1H, $J_{5,6a}$ = 4.3 Hz, $J_{6a,6b}$ = 12.5 Hz, H-6a), 4.08-4.02 (m, 1H, H-6b), 3.83-3.76 (m, 1H, H-5), 3.29-3.20 (m, 1H, CH), 2.65-2.41 (m, 4H, 2CH$_2$), 2.08-1.85 (m, 12H, 4OAc), 1.31-1.19 (m, 6H, 2CH$_3$); ^{13}C NMR (100 MHz, CDCl$_3$) δ 176.4, 172.1, 170.9, 170.5, 169.5, 142.6, 142.1, 127.9, 126.5, 77.9, 73.3, 72.5, 70.3, 67.9, 61.5, 45.1, 42.3, 35.7, 28.2, 21.7, 20.5, 15.4; MS (ESI) m/z: 522.62 [M+H]$^+$. Anal. (C$_{26}$H$_{35}$NO$_{10}$): C, 59.87; H, 6.76; N, 2.69. Found: C, 60.24; H, 6.48; N, 2.98.

2.2.2.6. N-(2,3,4,6-Tetra-O-acetyl-β-D-glucopyranosyl)-1-isobutyl-4-isopropylbenzene (AcS21). Yellow foam, 42% yield; Rf = 0.47 (*n*-hexane/EtOAc 1:1); $[\alpha]_D^{22}$ = +2 (*c* = 0.20, CHCl$_3$); ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.13-7.04 (m, 4H, Ph), 6.25 (dd, 1H, $J_{1,2}$ = 9.3 Hz, $J_{1,NH}$ = 9.4 Hz, H-1), 5.33-5.20 (m, 2H, H-2, NH), 5.09-5.01 (m, 1H, H-3), 4.91-4.88 (t, 1H, J = 9.6 Hz, H-4), 4.36-4.30 (m, 1H, H-6a), 4.09-4.05 (m, 1H, H-6b), 3.83-3.80 (m, 1H, H-5), 3.04-2.80 (m, 2H, CH$_2$), 2.57-2.40 (m, 2H, 2CH), 2.08, 2.02, 2.01, 1.99 (4s, 12H, 4OAc), 1.23-1.05 (m, 9H, 3CH$_3$); ^{13}C NMR (100 MHz, CDCl$_3$) δ 176.3, 171.0, 170.6, 169.9, 169.6, 146.8, 136.3, 128.9, 126.4, 78.1, 73.5, 72.5, 70.4, 68.1, 61.6, 43.1, 38.8, 33.6, 24.0, 20.7, 20.6, 17.1, 17.0; MS (ESI) m/z: 536.38 [M+H]$^+$. Anal. (C$_{27}$H$_{37}$NO$_{10}$): C, 60.55; H, 6.96; N, 2.62. Found: C, 60.67; H, 6.72; N, 2.84.

2.3. General procedure for the preparation of unprotected nucleosides(S). The protected nucleosides (**AcS**) (1 mmol) were treated with ammonia/MeOH (saturated at 0 °C, 41.8 mL) overnight at room temperature. The solvent was evaporated under reduced pressure to afford pure **S**, in 66-81% yields as syrups.

*2.3.1. N-(β-D-Glucopyranosyl)-4(prop-1-enyl)biphenyl amide (**S1**).* White syrup, 80% yield; Rf = 0.27 (EtOAc/MeOH 4:1); $[\alpha]_D^{22}$ = -26 (*c* = 0.20, MeOH); IR (KBr) 3423, 1655 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.68 (d, 1H, $J_{1,NH}$ = 9.2 Hz, NH), 7.76-7.30 (m, 10H, 2Ph, CH=<u>CH</u>), 6.72 (d, 1H, *J* = 15.7 Hz, <u>CH</u>=CH), 5.15-4.89 (m, 2H, 2-OH, 3-OH), 4.86 (t, 1H, $J_{1,2}$ = 8.9 Hz, H-1), 4.57-4.43 (m, 1H, 4-OH), 3.68-3.53 (m, 2H, H-3, H-4), 3.23-3.04 (m, 4H, H-2, H-5, H-6); ^{13}C NMR (100 MHz, DMSO-d_6) δ 171.5, 165.2, 141.2, 139.3, 133.9, 129.0, 128.2, 127.2, 126.6, 122.0, 79.6, 78.6, 77.5, 72.6, 69.9, 60.9; MS (ESI) m/z: 386.67 [M+H]$^+$. Anal. ($C_{21}H_{23}NO_6$): C, 65.44; H, 6.02; N, 3.63. Found: C, 65.62; H, 5.74; N, 3.99.

*2.3.2. N-(β-D-Glucopyranosyl)-6-butyl-1,2,3,4-tetrahydronapthalene amide (**S2**).* Yellow syrup, 76% yield; Rf = 0.28 (EtOAc/MeOH 4:1); $[\alpha]_D^{22}$ = -36 (*c* = 0.20, MeOH); IR (KBr) 3453, 1654 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.29 (d, 1H, $J_{1,NH}$ = 9.2 Hz, NH), 6.95-6.85 (m, 3H, Ph), 5.01-4.79 (m, 3H, 2-OH, 3-OH , 4-OH), 4.70 (t, 1H, $J_{1,2}$ = 9.0 Hz, H-1), 4.47 (t, 1H, $J_{6-OH,6}$ = 4.8 Hz, 6-OH), 3.65-3.03 (m, 6H, H-2, H-3, H-4, H-5, H-6), 2.66 (m, 4H, 2CH$_2$), 2.12-2.07 (m, 2H, CH$_2$), 1.75-1.71 (m, 8H, cyclohexane); ^{13}C NMR (100 MHz, DMSO-d_6) δ 172.3, 138.6, 136.2, 133.7, 128.7, 125.5, 79.4, 78.4, 77.5, 72.4, 69.9, 60.9, 34.8, 34.2, 28.7, 28.4, 26.7, 22.8, 22.5; MS (ESI) m/z: 380.71 [M+H]$^+$. Anal. ($C_{20}H_{29}NO_6$): C, 63.31; H, 7.70; N, 3.69. Found: C, 63.61; H, 8.01; N, 3.31.

*2.3.3. N-(β-D-Glucopyranosyl)-4-ethoxybiphenyl amide (**S3**).* Yellow syrup, 68% yield; Rf = 0.25 (EtOAc/MeOH 4:1); $[\alpha]_D^{22}$ = +2 (*c* = 0.20, MeOH); IR (KBr) 3399, 1655 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.67 (d, 1H, $J_{1,NH}$ = 9.1 Hz, NH), 7.67-6.98 (m, 9H, 2Ph), 5.15-4.90 (m, 3H, 2-OH, 3-OH, 4-OH), 4.81 (t, 1H, $J_{1,2}$ = 8.8 Hz, H-1), 4.56 (s, 2H, CH$_2$), 4.52 (d, 1H, $J_{6-OH,6}$ = 5.3 Hz, 6-OH), 3.65 (dd, 1H, $J_{6a,6b}$ = 11.4, $J_{5,6a}$ = 4.3 Hz, H-6a), 3.50-3.39 (m, 1H, H-6b), 3.23-3.07 (m, 4H, H-2, H-3, H-4, H-5); ^{13}C NMR (100 MHz, DMSO-d_6) δ 168.0, 157.3, 139.7, 133.1, 128.8, 127.7, 126.7, 126.2, 115.1, 79.4, 78.7, 77.3, 72.1, 69.9, 66.7, 60.9; MS

(ESI) m/z: 390.38 [M+H]$^+$. Anal. (C$_{20}$H$_{23}$NO$_7$): C, 61.69; H, 5.95; N, 3.60. Found: C, 61.33; H, 6.09; N, 3.77.

*2.3.4. N-(β-D-Glucopyranosyl)-1-isopropyl-4(prop-1enyl)benzene amide (**S6**).* White syrup, 78% yield; Rf = 0.32 (EtOAc/MeOH 4:1); [α]$_D^{22}$ = +1 (*c* = 0.20, MeOH); IR (KBr) 3343, 1663 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-*d$_6$*): δ 8.56 (d, 1H, *J*$_{1,NH}$ = 9.2 Hz, NH), 7.49-7.26 (m, 5H, Ph, CH=<u>CH</u>), 6.70-6.60 (m, 2H, H-1, <u>CH</u>=CH), 5.05-4.79 (m, 4H, H-2, H-3, H-4, 2-OH), 4.52 (t, 1H, *J*$_{6\text{-OH},6}$ = 5.9 Hz, 6-OH), 3.70-3.37 (m, 2H, H-6, 4-OH), 3.27-3.14 (m, 2H, CH, H-5), 1.46 (d, 3H, *J* = 15.7 Hz, CH$_3$), 1.23 (d, 3H, *J* = 12.5 Hz, CH$_3$); ^{13}C NMR (100 MHz, DMSO-*d$_6$*) δ 171.4, 165.3, 150.2, 132.4, 127.6, 126.9, 121.0, 78.5, 77.5, 72.6, 69.9, 60.9, 33.3, 23.6, 22.5; MS (ESI) m/z: 352.51 [M+H]$^+$. Anal. (C$_{18}$H$_{25}$NO$_6$): C, 61.52; H, 7.17; N, 3.99. Found: C, 61.69; H, 6.77; N, 4.19.

*2.3.5. N-(β-D-Glucopyranosyl)-1-sec-butyl-4-ethylbenzene amide (**S20**).* Yellow syrup, 66% yield; Rf = 0.5 (EtOAc/MeOH 4:1); [α]$_D^{22}$ = -2 (*c* = 0.20, MeOH); IR (KBr) 3436, 1655 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-*d$_6$*): δ 9.35 (d, 2H, *J* = 8.2 Hz, o-Ph), 7.41 (d, 2H, *J* = 8.2 Hz, m-Ph), 7.34-7.29 (m, 2H, 2-OH, 3-OH), 6.17-5.98 (m, 4H, H-1, H-2, H-3, H-6a), 5.88-5.76 (m, 2H, H-4, H-6b), 5.64-5.47 (m, 5H, 2CH$_2$, H-5), 3.00 (m, 7H, 2CH$_3$, CH); ^{13}C NMR (100 MHz, DMSO-*d$_6$*) δ 171.2, 143.8, 141.0, 127.4, 126.4, 79.4, 78.3, 77.4, 72.3, 70.0, 60.9, 43.7, 35.0, 27.5, 21.4, 15.4; MS (ESI) m/z: 354.36 [M+H]$^+$. Anal. (C$_{18}$H$_{27}$NO$_6$): C, 61.17; H, 7.70; N, 3.96. Found: C, 61.57; H, 7.34; N, 3.58.

*2.3.6. N-(β-D-Glucopyranosyl)-1-isobutyl-4-isopropylbenzene (**S21**).* Yellow syrup, 81% yield; Rf = 0.41 (EtOAc/MeOH 4:1); [α]$_D^{22}$ = +2 (*c* = 0.20, MeOH); IR (KBr) 3349, 1654 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-*d$_6$*): δ 8.40-8.33 (m, 1H, NH), 7.13-7.04 (m, 4H, Ph), 4.74-4.63 (m, 1H, H-1), 3.68-3.54 (m, 1H, H-2), 3.25-3.00 (m, 5H, H-3, H-4, H-5, H-6), 2.94-2.70 (m, 3H, CH$_2$, CH), 2.44-2.34 (m, 1H, CH), 1.72, 1.15, 1.13 (3s, 9H, 3CH$_3$); ^{13}C NMR (100 MHz, DMSO-*d$_6$*) δ 175.7, 171.5, 145.8, 137.3, 128.8, 126.0, 79.6, 78.5, 77.5, 72.5, 70.0, 60.9,

33.0, 24.0, 22.5, 16.9; MS (ESI) m/z: 368.62 [M+H]$^+$. Anal. ($C_{19}H_{29}NO_6$): C, 62.11; H, 7.96; N, 3.81. Found: C, 62.25; H, 7.67; N, 3.44.

2.4 Protein purification and kinetic measurements

Rabbit skeletal muscle glycogen phosphorylase b was isolated, purified recrystallized and assayed as described previously [36]. Kinetic studies were performed at 30° C in the direction of glycogen synthesis by measuring the inorganic phosphate released in the reaction as reported previously [36]. The enzyme (5 µg/ml) was assayed in a buffer of 30 mM imidazole/HCl (pH 6.8) containing 60 mM KCl, 0.6 mM EDTA, and 0.6 mM dithiothreitol using constant concentrations of glycogen (0.2% w/v), AMP (1 mM), and various concentrations of Glc-1-P (2, 4, 6, 10 and 20 mM) and inhibitors. Enzyme and glycogen were preincubated for 15 min at 30°C before initiating the reaction with Glc-l-P. Initial velocities were calculated from the pseudo-first order rate constants. For the calculation and statistical evaluation of the kinetic and binding parameters, the non-linear regression program GRAFIT was used [37].

2.5 Protein crystallography

Crystallographic binding studies were performed by diffusion of a solution of the *C*-5 halogen substituted glucopyranosyl nucleosides in the crystallization media of preformed GPb crystals. GPb crystals were grown in the tetragonal lattice, space group $P4_32_12$, under conditions similar to those described previously [38]. These crystals were soaked in a 5 mM inhibitor solution in a buffer of 10 mM Sodium N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic (pH 6.7) containing 0.1 mM EDTA, and 0.02 % w/v sodium azide at room temperature for 2 hours prior to data collection. X-ray diffraction data were collected using synchrotron radiation on station ID911-2 at MAX-Lab Synchrotron Radiation Source in Lund, Sweden or for the GPb-S**6** complex on an Oxford Diffraction SuperNova X-ray system using a 135 mm CCD detector at room temperature. Crystal orientation, integration of reflections, inter-frame scaling, partial reflection summation, and data reduction was

performed by the program Mosflm [39]. For scaling and merging of intensities the program SCALA of CCP4 Suite [40] was used. Crystallographic refinement of the complexes was performed by maximum-likelihood methods using REFMAC [40]. The starting model employed for the refinement of the complexes was the structure of the native T state GPb complex determined at 1.9 Å resolution (Leonidas et al., unpublished results). Ligand models were constructed using the PRODRG server [41] and they were fitted to the electron density maps after adjustment of their torsion angles. Alternate cycles of manual rebuilding with the molecular graphic program COOT [40] and refinement with REFMAC [40] improved the quality of the models. A summary of the data processing and refinement statistics for the inhibitor complex structures is given in Table 6. The stereochemistry of the protein residues was validated by MolProbity [42]. Hydrogen bonds and van der Waals interactions were calculated with the program CONTACT as implemented in CCP4 [40] applying a distance cut off 3.3 Å and 4.0 Å, respectively. Figures were prepared with PyMol [43]. The coordinates of the new structures have been deposited with the RCSB Protein Data Bank (http://www.rcsb.org/pdb) with codes presented in Table 6.

2.6 Ex vivo experiments

Hepatocytes were isolated from male Wistar rats, 200-300g body wt (obtained from Harlan UK, Ltd) by collagenase perfusion [44]. Procedures conformed to Home Office Regulations and were approved by Newcastle University Ethical Committee. Hepatocytes were suspended in Minimum Essential Medium (MEM) with 5% calf serum and seeded in multi-well plates at a density of 6×10^4 cells / cm^2. After cell attachment the medium was replaced by serum-free MEM containing 5 mM glucose, 10 nM dexamethasone, 1 nM insulin and the hepatocytes were cultured for 24 h. For determination of the effects of the GP inhibitors on conversion of GPa to GPb, the compounds or glucose (to a final concentration of 25 mM) were added to the medium 1 h after the change of medium and the incubations were continued for a further 30

min. Incubations with 5 mM or 25 mM glucose were also done in the additional presence of 5-thioglucose (9 mM), a hexokinase inhibitor, that prevents the elevation in glucose 6-P at 25 mM glucose [45]. The monolayers were snap frozen in liquid nitrogen and the activity of GPa was determined by coupling glycogen phosphorylase to the glucose 6-phosphate dehydrogenase reaction as described previously [45].   GPa activity in cells incubated with the GP inhibitors or with 25 mM glucose was expressed as percentage of the corresponding controls containing 5mM glucose (without or with 5-thioglucose). Results are Means ± SEM for 5 hepatocyte preparations. Statistical analysis was by the paired t-test.

3. Results

3.1. Virtual Screening

3.1.1 Training Set Results To evaluate the performance of CombiGlide and GOLD with different scoring functions for the task at hand, a training set of ligands was employed with known inhibitory constants, each of these ligands containing a –NH-CO- group at the C1 anomeric position, but included ligands with extended linkers (e.g. –NHCONHCO-) so as to increase training set size as well as the spread of experimental binding affinities (Table S1). The performances were evaluated with respect to the recovery of active molecules, which we defined as those with K_i's < 10 µM (10 ligands within a set of 40) and is reported in Table 2. Statistical performance metrics used were hit-rate, efficiency factor (EF), receiver operator characteristic (ROC) plots and (normalized) sum of the log ranks ((N)SLR), as described in the Experimental Section. Docking algorithms/scoring functions tested comprised Glide-SP and –XP; GOLD with -ChemScore, -ChemPLP, and -GoldScore and -ASP (two knowledge-based scoring functions). The best scoring functions were combined in a consensus scoring approach to ranking of ligand binding affinities [16]. Each of the six scoring functions has strengths and weaknesses. To increase the success rate in virtual screening, it may therefore

be beneficial to use several scoring functions and apply "consensus scoring" for which a considerable reduction of false positives has previously been reported [16, 30].

Although all scoring methods performed reasonably well, Glide-SP and GOLD-ChemPLP scores were found to be the most robust (Table 2). Hit rates of 70% (Gold-ChemPLP) and 80% (Glide-SP) for recovery of actives were obtained and enrichment factor (EF) values of 2.8 (Gold-ChemPLP) and 3.2 (Glide-SP). For both of these methods, AU-ROC values > 0.9 were obtained, while (N)SLR "early enrichment" values of ~ 0.9 reflect the excellent early recovery of actives. Combining Glide-SP and COLD-ChemPLP in consensus scoring using a Simple Sum Rank [16] reproduces the best hit rate and EF values (from Glide-SP) of 80% and 3.2, respectively; the best AU-ROC value of 0.92 (from Gold-ChemPLP); while the (N)SLR value (0.91) is similar to the best value of 0.92 obtained from Gold-ChemPLP. The method also produces the best correlation (R^2 = 0.46) between the predicted and actual ligand ranking of binding affinities.

3.1.2 Screening Set Results The five docking algorithms/scoring functions Glide-SP, GOLD-ChemPLP, -ChemScore, -ASP and –GoldScore were used. Glide-XP was the weakest performer for the training set, and given its computational expense for screening of 1888 ligands, its inclusion was not warranted. Re-ranking of predicted ligand potencies was via consensus scoring involving a Simple Sum Rank of results from Glide-SP and Gold-ChemPLP. Ligands with a Simple Sum Rank <150 (29 ligands) were chosen for further evaluation with their structures shown in Table S2. The docking/consensus scoring results are shown in Tables 3 (ligands synthesized) and S3 (all 29 ligands). Ligands are listed as **S***N*, where the **S** represents Screening Set ligand and the *N* is the overall rank of the ligand with **S1** having the best (lowest) Simple Sum Rank and **S29** the worst (highest).

3.1.3 Prioritization of Ligands for Synthesis & Biological Evaluation The final set of 29 ligands were clustered into groups based on 2D fingerprints (c.f. Experimental Section; Table

S3). Criteria used to select ligands from the clusters were the Simple Sum Rank, coarse quantile votes [16] from the other scoring functions, and ADME properties (Tables 3 and S3) considering Lipinksi's 'rule of five' [32] and Jorgensen's 'rule of three' [33]. Ultimately, the six ligands shown in Table 1 were chosen due to availability and therefore synthetic viability. In terms of recovery of actives, 50% of the six molecules synthesized were found to be actives (K_i's $< 10\,\mu$M as defined for by our initial training set).

3.2. Chemistry

The approach towards the preparation of the target *N*-(β-D-glucopyranosyl) amides **S** with various aralkyl and aralkenyl moieties, involved the cross-coupling reaction of the primary glucopyranosylamine **AcGlc-NH₂** with a diverse set of commercially available acyl chlorides RCOCl (Scheme 2). As starting material was selected the per-*O*-acetylated β-D-glucopyranosyl azide (**AcGlc-N₃**)[34], which upon catalytic hydrogenation in the presence of palladium-on-carbon (Pd/C) led to the desired β-D-glucosylamine (**AcGlc-NH₂**). A direct acylation of D-glucopyranosylamine (**AcGlc-NH₂**) with the appropriate acid chlorides (**1,2,3,6,20,21**) using triethylamine (Et₃N) as base and *N,N*-dimethylformamide (DMF) as solvent, resulted in the protected *N*-acyl-β-D-glycosylamines (**AcS**), in reasonable to moderate yields (41-55%). This acylation reaction was stereospecific producing the β-anomers only, within 1 hour at room temperature. In the final step, removal of all the protecting groups of the amido-derivatives (**AcS**) with saturated methanolic ammonia, furnished the target analogues **S1,S2,S3,S6,S20** and **S21**, in good yields (66-81%).

All compounds gave satisfactory analytical and spectroscopic data, which were in full accordance with their depicted structure. In the ¹H NMR spectra of compounds (**AcS**) and (**S),** the H-1 protons appeared as doublets ($J_{1,2} \geq 8.5$ Hz), indicating a β configuration of the anomeric linkage. In the IR spectrum of glucosylamine (**AcGlc-NH₂**), the presence of the amino absorption band at 3467-3267 cm^{-1} and the simultaneous absence of the azido group

absorption band at 2119 cm^{-1} confirmed the proposed structure, while the ester carbonyl functions featured at 1753 cm^{-1}. Derivatives (**S**) gave in their IR spectra the characteristic bands of both the amide bond and the C=O stretch between 3343-3453 cm^{-1} and 1654-1663 cm^{-1}, respectively.

3.3 Biochemical assessment

To validate our synthesized compounds we tested their inhibitory potency on kinetics experiments with rabbit muscle glycogen phosphorylase b (GPb) in the direction of glycogen synthesis. Human liver GPb and rabbit muscle GPb active sites are identical in amino-acid sequence and structural architecture and hence active site ligands tested against rabbit muscle GPb are expected to be potent against human liver glycogen phosphorylase as well. The inhibition constants are shown in Table 1. All compounds displayed competitive inhibition with respect to the substrate Glc-1-P, at constant concentrations of glycogen and AMP. The most potent are compounds **S1** (K_i=5.14 µM) and **S6** (K_i=4.98 µM) followed closely by **S3** (K_i=8.25 µM). From the inhibition constant values (Table 1) it seems that compounds **S2**, **S21** and **S20** that have a certain degree of freedom in the linker that connects the Glc-NHCO group to the aromatic rings are worse inhibitors than those that have a linker with a more rigid geometry which directs the position of the phenyl rings.

3.3 Cellular evaluation

GP ligands that stabilise the T-state promote conversion of GPa to GPb in hepatocytes because the T-state is a better substrate for dephosphorylation by phosphorylase phosphatase, whereas compounds that stabilise the R-state have the converse effect and promote conversion of GPb to GPa [46]. We compared the effects of **S1**, **S2**, **S3** and **S6** with 25 mM glucose on depletion of GPa during a 30 min incubation of the hepatocytes (Table 4). The effects of 25 mM glucose were also tested in the presence of the hexokinase inhibitor, 5-thioglucose, to prevent the elevation in glucose 6-P caused by 25mM glucose.[45] Glucose (25

mM) lowered GPa by 36% in the absence of inhibitor and by 21% in the presence of the hexokinase inhibitor, in agreement with previous findings that glucose inactivates GPa through the combined effects of glucose and the elevated glucose 6-P [45]. **S1** and **S3** caused significant lowering of GPa at concentrations of 2 μM (11-15%) and at 20 μM (17-20%), showing that these compounds mimic the effect of elevated glucose at low micromolar concentrations, and also that their effect is comparable to the inhibition caused by glucose in the presence of the hexokinase inhibitor. **S2** and **S6** had no significant effect (Table 4).

3.4 Structural studies

Towards elucidating the structural basis of inhibition and most importantly the differences in the inhibition constants, we have determined the crystal structures of the six selected inhibitors in complex with GPb. All inhibitors were found bound at the active site with their atoms well defined within the electron density maps (Fig. 1). The anchor point for each of the inhibitors appears to be the glucose moiety which is engaged in hydrogen bond (Table 5) and van der Waals interactions almost identical to those that have observed for glucose and in other glucose containing inhibitors [1, 2, 10, 47]. Inhibitors **S1** and **S3** (the most potent) bind very similar with the R group in the β-pocket participating in van der Waals interactions with residues Glu88, Asn133, Tyr280, Asn282, Asp283, Phe286, Arg292, His341, and Glu385 (Fig 2). The introduction of ether oxygen in the linker (**S3**) instead of methyl did not seem to affect the binding as this oxygen is not involved in any hydrogen bond interactions with protein residues in the vicinity (Fig 2). Meanwhile the van der Waals interactions of this oxygen in the **S3** complex are similar to those of the corresponding carbon in the **S1** complex. Inhibitors **S6, S20** and **S21** bind similar to each other, within the β-pocket, but remarkably different from **S1** and **S3**. Their R groups are closer to the location of the 280s loop in the free GPb structure than **S1** and **S3**. The distance between the benzene rings of **S6, S20,** and **S21** are those of **S1** and **S3** is approximately 1.6, 1.8 and 2.6 Å in the respective complexes.

However, the methyl groups of **S6**, and **S21,** and the ethyl group of **S20** on the benzyl group are all in the same location (within a radius of 0.5 Å) involved in similar van der Walls interactions with residues Tyr280, Asn282, Phe285, His341, and Ala383 (Fig 2). The methyl group in the linkers of **S20** and **S21** point towards the same direction and participates in van der Waals interactions with side chain atoms of Leu136 and Asp339. Inhibitor **S2** binds similarly to **S1** but the linker follows a different route to place the tetrahydro-naphtalene group in the place where the two benzyl rings of S1 and **S3** bind, participating in similar van der Waals interactions with protein residues. The differences in the conformation of the linker observed are probably imposed by the architecture of the β-pocket in the protein and its hydrophobic characteristics that attract the R group of all six inhibitors. Thus, since the R group is attracted at this location the linker is rearranged (according to its stereochemistry) to allow the binding.

All six inhibitors upon binding induce a significant conformational change of the 280s loop (residues 280-288) is observed when comparing the complex structures to the unliganded (Fig. 3). Thus, it seems that the presence of the R group of the two inhibitors induces the movement of this loop away from the active site making space for the inhibitor. Thus the atoms from residues 282-287 upon superposition onto the corresponding ones of the unliganded GPb structure show an rms displacement of 1.5, 1.2, 2.0, 2.2, 1.2 and 2.2 Å for the **S1, S2, S3, S6, S20** and **S21** complexes, respectively. From the above values it is evident that **S2** and **S20** cause the smaller disturbance to the 280s loop. Comparing the conformational change of the 280s loop induced by the binding of the inhibitors it seems that **S1**, **S2** and **S20** trigger a mere displacement of the 280s loop while the rest of the inhibitors cause also more significant conformational changes. Briefly, in the **S3, S6,** and **S21** complexes Asp283 and Asn284 shift away from the ligand to avoid clashes with the R group, with concomitant shifts of the other residues of the 280s loop. More specifically, Asn284 flips 180° and is positioned

between the phenyl rings of Phe285 and Tyr613, two residues that constitute the inhibitor site at the entrance of the catalytic site in GPb, resulting in its disruption. This rearrangement of the 280s loop is in the direction of the R to T allosteric conversion and has been also observed in other GPb ligand complexes with bulky benzoyl moieties [48-50].

4. Discussion and Conclusions

The structural studies offer an explanation for the differences in the inhibitory potency of the six inhibitors. Thus, since all six inhibitors are involved in the same hydrogen bond interactions (direct and water mediated) our attention shifted to the van der Waals interactions of the ligands. The virtual screening was performed with the goal to maximize the non-polar interactions of the R group within the β-pocket of the active site. The two most potent inhibitors **S1** and **S6** that display similar potency form the same number of van der Waals interactions with the protein (100) while the third best (**S3**) participates in slightly fewer (91) and thus has a slight increased K_i value. This pattern is repeated with the other two inhibitors, **S2** and **S21** that engage in 78 and 71 van der Waals interactions with the protein and have K_i values of 48.7 and 75.3 μM, respectively. The only exception to this pattern is **S20** that forms 71 non-polar interactions like **S21** but has a K_i value five times bigger than **S21** (Table 1). The main difference in the binding of these two inhibitors is the conformational change they induce on the 280s loop. Thus, the binding of **S20** causes a more severe rearrangement of the loop than the one caused by the binding of **S21**. This is pronounced by the shifts of Asp283 and Asn284 described above, and this probably has energy cost that increases the K_i value of **S20** with respect to that of **S21**.

A structural comparison with the GPb - *N*-4-phenylbenzoyl-2-*N'*-*β*-D-glycopyranosyl urea complex [49] (PDB id: 2qln), the inhibitor (K_i=3.7 μM) used as a scaffold for modelling reveals that **S1** and **S3** bind very similar with this compound. However, both **S1** and **S3** lack the second carbonyl oxygen in the linker of the scaffold compound and thus are not involved

in any of the additional water-mediated of the scaffold compound and Gly134, Gly135, Gly137, and Asp283. These interactions probably are responsible for the slightly better inhibitory potency of the scaffold compound with respect to **S1** and **S3.**

S1 and **S3** are effective at causing inactivation of GPa in hepatocytes at low micromolar concentrations. The magnitude of the inhibition is smaller than for 25 mM glucose which is mediated in part through the raised concentration of glucose 6-P[45], but it is comparable to the effect of 25 mM glucose in the presence of the hexokinase inhibitor which blocks glucose phosphorylation and the elevation in glucose 6-P.

The current synthetic scheme/combinatorial protocols were for glucose-NHCO-R ligands, where by default the –R group was connected to the –NHCO- via a carbon. We now plan to extend our screening approach to glucose analogues with a more extended linker such as – NHCONH- since compounds with this linked have shown high potency [49]. In conclusion, the results of the present work, which is the first high-throughput structure-guided screening of commercially available small chemical entities databases for GPIs reported, show that this approach can generate potent drug-like GPIs that are effective both *in vitro* and *ex vivo*.

Supporting Information Available: Tables of the virtual screening results.
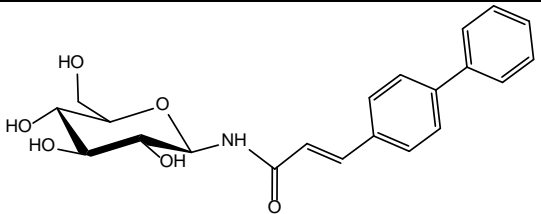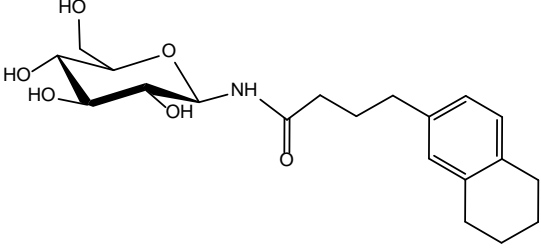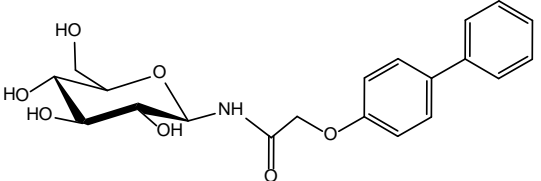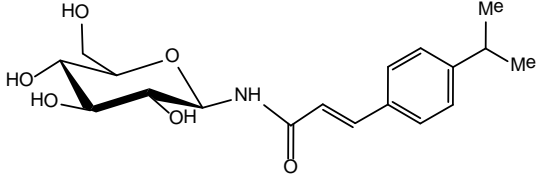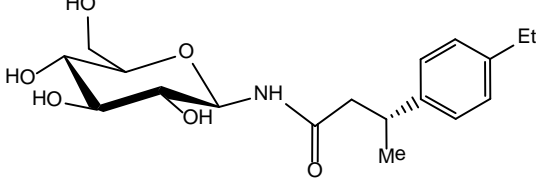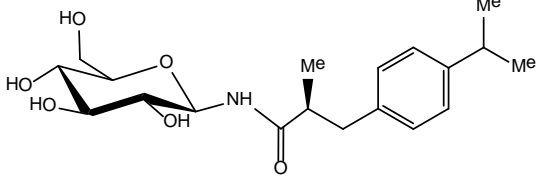
References

1.   Oikonomakos, N. G. *Curr. Protein. Pept. Sci.* **2002,** *3*, 561.
2.   Somsak, L.; Czifrak, K.; Toth, M.; Bokor, E.; Chrysina, E. D.; Alexacou, K. M.; Hayes, J. M.; Tiraidis, C.; Lazoura, E.; Leonidas, D. D.; Zographos, S. E.; Oikonomakos, N. G. *Curr Med Chem* **2008,** *15*, 2933.
3.   Oikonomakos, N. G.; Somsak, L. *Curr. Opin. Investig. Drugs* **2008,** *9*, 379.
4.   Hayes, J. M.; Leonidas, D. D. *Mini Rev. Med. Chem.* **2010,** *10*, 1156.
5.   Agius, L. *Mini Rev. Med. Chem.* **2010,** *10*, 1175.
6.   Treadway, J. L.; Mendys, P.; Hoover, D. J. *Expert Opin. Investig. Drugs* **2001,** *10*, 439.
7.   Agius, L. *Best Pract. Res. Clin. Endocrinol. Metab.* **2007,** *21*, 587.
8.   Kurukulasuriya, R.; Link, J. T.; Madar, D. J.; Pei, Z.; Richards, S. J.; Rohde, J. J.; Souers, A. J.; Szczepankiewicz, B. G. *Curr. Med. Chem.* **2003,** *10*, 123.
9.   Kurukulasuriya, R.; Link, J. T.; Madar, D. J.; Pei, Z.; Rohde, J. J.; Richards, S. J.; Souers, A. J.; Szczepankiewicz, B. G. *Curr. Med. Chem.* **2003,** *10*, 99.
10.   Somsak, L. *Comptes Rendus Chimie* **2011,** *14*, 211.
11.   Martin, J. L.; Veluraja, K.; Ross, K.; Johnson, L. N.; Fleet, G. W. J.; Ramsden, N. G.; Bruce, I.; Orchard, M. G.; Oikonomakos, N. G.; Papageorgiou, A. C.; Leonidas, D. D.; Tsitoura, H. S. *Biochemistry (USA)* **1991,** *30*, 10101.
12.   Oikonomakos, N. G.; Somsak, L. *Curr Opin Investig Drugs* **2008,** *9,* 379.
13.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J Med Chem* **2004,** *47*, 1739.
14.   .
15.   Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995,** *245*, 43.
16.   Feher, M. *Drug Discov Today* **2006,** *11*, 421.
17.   Irwin, J. J.; Shoichet, B. K. *J Chem Inf Model* **2005,** *45*, 177.
18.   Van de Waterbeemd, H. *Expert Opin Drug Metab Toxicol* **2005,** *1*, 1.
19.   Thompson, T. N. *Curr Drug Metab* **2000,** *1*, 215.
20.   Papageorgiou, A. C.; Oikonomakos, N. G.; Leonidas, D. D.; Bernet, B.; Beer, D.; Vasella, A. *Biochem. J.* **1991,** *274*, 329.
21.   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J Phys Chem B* **2001,** *105*, 6474.
22.   Somsak, L.; Kovacs, L.; Toth, M.; Osz, E.; Szilagyi, L.; Gyorgydeak, Z.; Dinya, Z.; Docsa, T.; Toth, B.; Gergely, P. *J Med Chem* **2001,** *44*, 2843.
23.   Benltifa, M.; Hayes, J. M.; Vidal, S.; Gueyrard, D.; Goekjian, P. G.; Praly, J.-P.; Kizilis, G.; Tiraidis, C.; Alexacou, K. M.; Chrysina, E. D.; Zographos, S. E.; Leonidas, D. D.; Archontis, G.; Oikonomakos, N. G. *Bioorg. Med. Chem.* **2009,** *17*, 7368.
24.   Truchon, J. F.; Bayly, C. I. *J. Chem. Inf. Model.* **2007,** *47*, 488.
25.   Zhao, W.; Hevener, K. E.; White, S. W.; Lee, R. E.; Boyett, J. M. *BMC Bioinformatics* **2009,** *10*.
26.   Sheridan, R. P. *J. Chem. Inf. Model.* **2008,** *48*, 426.
27.   Swamidass, S. J.; Azencott, C. A.; Daily, K.; Baldi, P. *Bioinformatics* **2010,** *26*, 1348.
28.   Mackey, M. D.; Melville, J. L. *J. Chem. Inf. Model.* **2009,** *49*, 1154.
29.   Fawcett, T. *Pattern Recogn. Lett.* **2006,** *27*, 861.

30.   Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999,** *42*, 5100.

31.   Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. *J. Mol. Graph. Model.* **2002,** *20*, 281.

32.   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced Drug Delivery Reviews* **2001,** *46*, 3.

33.   Jorgensen, W. L.; Duffy, E. M. *Advanced Drug Delivery Reviews* **2002,** *54*, 355.

34.   Tropper, F. D.; Andersson, F. O.; Braun, S.; Roy, R. *Synthesis-Stuttgart* **1992**, 618.

35.   Jarrahpour, A.; Alvand, P. *Iran.J. Sci. Tech. Trans. a-Sci.* **2007,** *31*, 17.

36.   Tsirkone, V. G.; Tsoukala, E.; Lamprakis, C.; Manta, S.; Hayes, J. M.; Skamnaki, V. T.; Drakou, C.; Zographos, S. E.; Komiotis, D.; Leonidas, D. D. *Bioorg. Med. Chem.* **2010,** *18*, 3413.

37.   Leatherbarrow, R. J. *Erithacus Software Ltd. Staines, U.K.* **1998**.

38.   Alexacou, K. M.; Tenchiu Deleanu, A. C.; Chrysina, E. D.; Charavgi, M. D.; Kostas, I. D.; Zographos, S. E.; Oikonomakos, N. G.; Leonidas, D. D. *Bioorg. Med. Chem.* **2010,** *18*, 7911.

39.   Leslie, A. G. W.; Powell, H. R. In *Evolving Methods for Macromolecular Crystallography;* Read, R. J., Sussman, J. I., Eds.; Springer, 2007; Vol. 245, 41.

40.   CCP4. *Acta Crystallogr.* **1994,** *D 50*, 760.

41.   Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. D. Biol. Crystallogr.* **2004,** *60*, 1355.

42.   Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D. Biol. Crystallogr.* **2010,** *66*, 12.

43.   DeLano, W. L.: Sa Carlos, CA, USA, 2002.

44.   Tudhope, S. J.; Wang, C. C.; Petrie, J. L.; Potts, L.; Malcomson, F.; Kieswich, J.; Yaqoob, M. M.; Arden, C.; Hampson, L. J.; Agius, L. *Diabetes* **2012,** *61*, 49.

45.   Aiston, S.; Green, A.; Mukhtar, M.; Agius, L. *Biochem. J.* **2004,** *377*, 195.

46.   Aiston, S.; Hampson, L.; Gomez-Foix, A. M.; Guinovart, J. J.; Agius, L. *J Biol Chem* **2001,** *276*, 23858.

47.   Kantsadi, A. L.; Hayes, J. M.; Manta, S.; Skamnaki, V. T.; Kiritsis, C.; Psarra, A. M.; Koutsogiannis, Z.; Dimopoulou, A.; Theofanous, S.; Nikoleousakos, N.; Zoumpoulakis, P.; Kontou, M.; Papadopoulos, G.; Zographos, S. E.; Komiotis, D.; Leonidas, D. D. *ChemMedChem* **2012,** *7*, 722.

48.   Oikonomakos, N. G.; Kosmopoulou, M.; Zographos, S. E.; Leonidas, D. D.; Chrysina, E. D.; Somsak, L.; Nagy, V.; Praly, J. P.; Docsa, T.; Toth, B.; Gergely, P. *Eur. J. Biochem.* **2002,** *269*, 1684.

49.   Nagy, V.; Felfoldi, N.; Konya, B.; Praly, J. P.; Docsa, T.; Gergely, P.; Chrysina, E. D.; Tiraidis, C.; Kosmopoulou, M. N.; Alexacou, K. M.; Konstantakaki, M.; Leonidas, D. D.; Zographos, S. E.; Oikonomakos, N. G.; Kozmon, S.; Tvaroska, I.; Somsak, L. *Bioorg. Med. Chem.* **2012,** *20*, 1801.

50.   Kantsadi, A. L.; Manta, S.; Psarra, A. M.; Dimopoulou, A.; Kiritsis, C.; Parmenopoulou, V.; Skamnaki, V. T.; Zoumpoulakis, P.; Zographos, S. E.; Leonidas, D. D.; Komiotis, D. *Eur. J. Med. Chem.* **2012,** *54*, 740.

TABLES

Table1. The chemical structures of the inhibitors studied with the K_i values for each inhibitor.

Compound	Chemical Structure	K_i (μM)
S1		5.14 ± 0.77
S2		48.68 ± 4.49
S3		8.25 ± 0.77
S6		4.98. ± 0.58
S20		376.87 ± 17.04
S21		75.29 ± 6.96

Table 2. Statistical results for ranking of training set ligand binding affinities from individual scoring functions and after consensus scoring with a Simple Sum Rank of Glide-SP and GOLD-ChemPLP.[a]

Statistical Metric	XP	Chemscore	Goldscore	ASP	ChemPLP	SP	Simple Sum Rank
							SP+ChemPLP
Hit Rate (%)	50	50	60	50	70	80	80
E.F.	2.0	2.0	2.4	2.0	2.8	3.2	3.2
AU-ROC	0.67	0.71	0.86	0.84	0.92	0.91	0.92
(N)SLR	0.60	0.67	0.82	0.81	0.92	0.89	0.91
R^2	0.02	0.13	0.20	0.14	0.31	0.28	0.46
Ranking of actives	3,4,5,6,9,14, 17,26,29,37	1,3,6,7,8,13,2 1,22,26,29	1,2,4,6,8,9,12 ,16,17,18	1,2,4,5,6,11,13 ,16,19,22	1,2,3,4,5,7,9,1 1,14,17	1,2,3,5,7,8,9, 10,15,16	1,2,3,5,6,7,8,10, 14,17

[a] Statistical metrics calculated as described in text. R^2 represents the correlation between the predicted and actual ranking of training set ligands using the different methods.

Table 3. Docking/consensus scoring and ADME results for the ligands shown in Table 1.[a]

Ligand	Zinc code (Acid Chloride)	Rank Glide-SP	Rank GOLD-ChemPLP	Simple Sum Rank	Votes 1.5% (4%)[b]	RMSD[c]	Lipinski's Rule of Five & Violations (V)					Jorgensen's Rule of Three & Violations (V)			
							MW (<500)	Num. H-bond donors (≤ 5)	Num. H-bond acceptors (≤ 10)	log P (o/w) (< 5)	V	Caco-2 perm. (>22 nm/s)	log S (>-5.7)	Num. Prim. Metab (< 7)	V
S1	33434881	1	1	2	3(3)	0.364	385.4	5	7*	0.851	0	80.2	-3.704	5	0
S2	11631072	10	8	18	1(2)	0.934	379.5	5	7*	0.208	0	95.1	-2.926	*9**	1
S3	26470275	18	5	23	3(3)	1.505	389.4	5	8*	-0.082	0	50.9	-2.385	7	1
S6	26470293	17	54	71	1(2)	1.707	351.4	4	7*	0.296	0	98.5	-3.08	6	0
S20	11629160	106	16	122	1(1)	0.910	353.4	5	7*	-0.162	0	89.0	-2.343	*8*	1
S21	11629207	78	65	143	0(1)	0.654	367.4	5	7*	0.01	0	76.7	-2.168	*9**	1
Range 95% Known Drugs[c]	-	-	-	-			130 – 725	2.0-20.0	0.0-6.0	-2.0 to 6.5	-	< 25 poor; >500 great	-6.5 to 0.5	1-8	-

[a] See text for full details. Violations of rules highlighted in italics, while those values outside the range of 95% of know drugs are highlighted with an asterisk (*). [b] Coarse quantile votes (maximum of 3) received by ligands if ranked in the top ~1.5% and 4%(in parentheses) of screened database using the other docking methods/scoring functions (GOLD-ChemScore, -GoldScore and –ASP). [c]RMSD is the difference between Glide-SP and GOLD-ChemPLP ligand poses (heavy atoms) in-place. [c] QikProp 3.3 User Manual.

Table 4. Effects of glucose analogues on GPa depletion in hepatocytes.

Compound	Concentration used	GPa (% control)
Glucose	25 mM	64.1 ± 4.4 **
5-thioglucose	9 mM	106.2 ± 4.0
Glucose - 5-thioglucose mixture	25 mM - 9 mM	78.7 ± 2.9 **
S1	2 µM	89.4 ± 2.1 **
S1	20 µM	83.3 ± 2.0 **
S2	2 µM	93.0 ± 4.8
S2	20 µM	100 ± 6.6
S3	2 µM	84.5 ± 6.3 *
S3	20 µM	80.3 ± 3.1 **
S6	2 µM	105 ± 11.5
S6	20 µM	106 ± 7.8

Hepatocytes were incubated in MEM containing 5mM and the analogues indicated (at 2 or 20 µM) or with 25 mM glucose without or with hexokinase inhibitor (5-thioglucose). After 30 min, hepatocytes were snap frozen for determination of GPa in the cell extracts. GPa is expressed as percentage of the respective controls with 5 mM glucose without or with 9 mM 5-thioglucose.   Means ± SEM (n=5).

* $P < 0.05$; ** $P < 0.005$, effect of analogue or 25 mM glucose relative to the respective control

Table 5. Potential hydrogen bond interactions of inhibitors with GPb residues in the crystal. Numbers shown are distances in Å.

Inhibitor atom	Protein atom	Distance (Å)					
		S1	S2	S3	S6	S20	S21
O2'	Tyr573 OH	3.2	3.1	3.1	-	3.1	3.2
	Glu672 OE1	3.2	3.2	3.1	3.1	3.0	3.1
	Water	2.6	2.6	2.7	2.7	3.2	2.8
	Water	-	2.8	2.9	-	2.8	2.7
O3'	Glu672 OE1	2.8	2.8	2.8	2.6	2.8	2.8
	Ala673 N	3.3	3.3	3.3	-	3.3	-
	Ser674 N	3.0	3.0	3.1	3.1	3.0	2.9
	Gly675 N	3.2	3.1	3.3	3.0	3.1	3.0
O4'	Gly675 N	2.8	2.9	2.8	2.8	2.9	2.9
	Water	2.7	2.7	2.6	2.6	2.6	2.7
O6'	His377 ND1	2.7	2.7	2.6	2.6	2.7	2.8
	Asn484 OD1	2.7	2.8	2.9	2.9	2.8	2.7
O3	Leu136 N	3.0	-	3.0	3.0	3.2	-
	Water	2.8	2.8	3.0	3.1	2.6	-
	Water	-	-	3.2	2.9	-	2.7
N1	His377 O	-	2.9	-	-	-	3.3
	Water	-	-	2.8	-	-	-
Total		**13**	**14**	**16**	**12**	**14**	**13**

Table 6: Summary of the diffraction data processing and refinement statistics for the GPb
complexes

GPb complex	S1	S2	S3	S6	S20	S21
Data collection and processing statistics						
Resolution (Å)	28.83 – 1.99	23.94 – 1.90	23.94 – 2.00	18.23 – 2.45	23.82 – 1.85	23.89 – 2.15
Outermost shell (Å)	2.10 – 1.99	2.00 – 1.90	2.11 – 2.00	2.58 – 2.45	1.95 – 1.85	2.27 – 2.15
Reflections	169449	254758	104316	72136	216184	242135
measured	(17599)	(36440)	(15051)	(10932)	(30152)	(33396)
Unique reflections[a]	63559 (6924)	75897 (10836)	52566 (7918)	33056 (4997)	82325 (11899)	52824 (7533)
Multiplicity[a]	2.7 (2.5)	3.4 (3.4)	2.0 (1.9)	2.2 (2.2)	2.6 (2.5)	4.6 (4.4)
R_{symm}[a]	0.094 (0.451)	0.055 (0.365)	0.073 (0.347)	0.099 (0.531)	0.056 (0.414)	0.103 (0.452)
Completeness (%)[a]	94.2 (71.3)	98.1 (97.1)	79.7 (83.1)	91.4 (95.6)	98.5 (98.8)	98.4 (97.3)
$<I/\sigma I>$ [a]	6.2 (2.2)	13.2 (3.2)	7.6 (2.7)	6.3 (1.8)	10.2 (2.3)	8.8 (3.3)
Refinement statistics						
R_{cryst}[a]	0.165 (0.264)	0.161 (0.242)	0.162 (0.241)	0.187 (0.256)	0.161 (0.271)	0.163 (0.237)
R_{free} [a]	0.189 (0.307)	0.188 (0.254)	0.200 (0.289)	0.234 (0.271)	0.190 (0.286)	0.203 (0.298)
No of solvent	223	265	295	113	301	208
molecules						
r.m.s. deviation from ideality						
in bond lengths	0.008	0.007	0.009	0.005	0.009	0.009
(Å)						
in angles (°)	1.2	1.2	1.2	1.0	1.3	1.3
Average B factor (Å²)						
Protein atoms	36.3	31.3	31.0	36.7	32.8	38.9
Solvent molecules	37.4	34.8	36.2	28.7	37.4	37.7
Ligand atoms	38.3	23.6	23.1	27.1	29.6	32.6
PDB entry	4MHS	4MI6	4MHO	4MIC	4MI9	4MI3

[a]Values in parentheses are for the outermost shell.

LEGENDS FOR FIGURES AND SCHEMES

Scheme 1: (a) glucose based analogues with α- and β-substitutions at the anomeric C1 atom. (b) glucose based analogues with linker which connects the glucose moiety with an R group. The –NHCO- linker was investigated in this work.
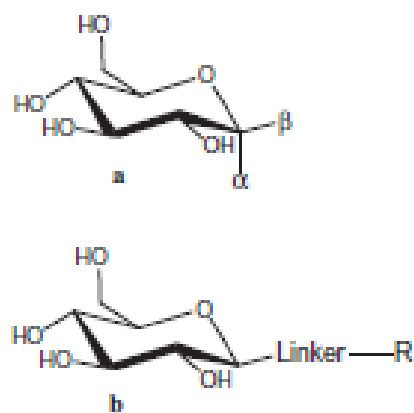
Scheme 2: (i) H_2, 10% Pd/C, EtOAc, 20 ^{o}C, 24 hours; **(ii)** DMF, RCOCl, Et_3N; **(iii)** ammonia/MeOH.

Figure 1. The REFMAC $2F_o$-F_c electron density maps of bound ligands contoured at 1.0 σ before the incorporation of the ligand models in the refinement process. The final refinemend models of **S1**, **S2**, **S3**, **S6**, **S20**, and **S21** are shown.

Figure 2. Stereo diagrams of the binding of **S1** (a), **S2** (b), **S3** (c), **S6** (d), **S20** (e), and **S21** (f) at the active site of GPb. Hydrogen bonds are shown as dashed lines and water molecules as black spheres labeled W.

Figure 3. Stereo diagram of the superposition of the six GPb ligand complexes (white) at the active site of the free GPb (black). **S1** (a), **S2** (b), **S3** (c), **S6** (d), **S20** (e), and **S21** (f).

Scheme 1

R	K_i (μM)
(phenyl group)	81^{20}, 144^{32}
(naphthalen-1-yl group)	444^{21}
(naphthalen-2-yl group)	10^{21}

Figure 1

Figure 2

Figure 3