

Glucose Analogue Inhibitors of Glycogen Phosphorylase: from Crystallographic Analysis to Drug Prediction using *GRID* Force-Field and *GOLPE* Variable Selection*

BY KIMBERLY A. WATSON,† EDWARD P. MITCHELL AND LOUISE N. JOHNSON

Oxford Centre for Molecular Sciences and Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

GABRIELE CRUCIANI

University of Perugia, Department of Chemistry, Via Elce di Sotto, 8, 06100 Perugia, Italy

JONG CHAN SON, CLAIRE J. F. BICHARD AND GEORGE W. J. FLEET

Dyson Perrins Laboratory, South Parks Road, Oxford OX1 3QY, England

AND NIKOS G. OIKONOMAKOS, MARIA KONTOU AND S. E. ZOGRAPHOS

The National Hellenic Foundation, 48, Vas. Constantinou Avenue, Athens 11635, Greece

(Received 24 April 1994; accepted 22 November 1994)

Abstract

Several inhibitors of the large regulatory enzyme glycogen phosphorylase (GP) have been studied in crystallographic and kinetic experiments. GP catalyses the first step in the phosphorylysis of glycogen to glucose-1-phosphate, which is utilized *via* glycolysis to provide energy to sustain muscle contraction and in the liver is converted to glucose. α -D-Glucose is a weak inhibitor of glycogen phosphorylase form b (GPb, $K_i = 1.7$ mM) and acts as a physiological regulator of hepatic glycogen metabolism. Glucose binds to phosphorylase at the catalytic site and results in a conformational change that stabilizes the inactive T state of the enzyme, promoting the action of protein phosphatase 1 and stimulating glycogen synthase. It has been suggested that in the liver, glucose analogues with greater affinity for glycogen phosphorylase may result in a more effective regulatory agent. Several *N*-acetyl glucopyranosylamine derivatives have been synthesized and tested in a series of crystallographic and kinetic binding studies with GPb. The structural results of the bound enzyme–ligand complexes have been analysed together with the resulting affinities in an effort to understand and exploit the molecular interactions that might give rise to a better inhibitor. Comparison of the *N*-methylacetyl glucopyranosylamine (*N*-methylamide, $K_i = 0.032$ mM) with the analogous β -methylamide derivative (*C*-methylamide, $K_i = 0.16$ mM) illustrate the importance of

forming good hydrogen bonds and obtaining complementarity of van der Waals interactions. These studies also have shown that the binding modes can be unpredictable but may be rationalized with the benefit of structural data and that a buried and mixed polar/non-polar catalytic site poses problems for the systematic addition of functional groups. Together with previous studies of glucose analogue inhibitors of GPb, this work forms the basis of a training set suitable for three-dimensional quantitative structure–activity relationship studies. The molecules in the training set are void of problems and potential errors arising from the alignment and bound conformations of each of the ligands since the coordinates were those determined experimentally from the X-ray crystallographic refined ligand–enzyme complexes. The computational procedure described in this work involves the use of the program *GRID* to describe the molecular structures and the program *GOLPE* to obtain the partial least squares regression model with the highest prediction ability. The *GRID/GOLPE* procedure performed using 51 glucose analogue inhibitors of GPb has good overall predictivity [standard deviation of error predictions (SDEP) = 0.98 and $Q^2 = 0.76$] and has shown good agreement with the crystallographic and kinetic results by reliably selecting regions that are known to affect the binding affinity.

Introduction

The knowledge of the three-dimensional structures of biological molecules has the potential to aid in the design of therapeutic agents. With increasing reports of macromolecular crystallographic structures, structure-based

* This work has been supported by the MRC (KAW and EPM), the Italian Government MURST (GC), the Oxford Centre for Molecular Sciences (JCS and CJFB) and an EEC Twinning SCIENCE programme (SC1*-CT91-0623) (NGO and MK).

† Author to whom correspondence should be addressed.

drug design has become a widely used strategy for modification and improvement of known inhibitors as well as for *de novo* inhibitor design. The strategy has been to systematically modify the structure of a lead compound using the three-dimensional macromolecular structure as a template for optimizing interactions. Analysis of the interactions between the macromolecular and its ligand target molecules provides insight into those regions of importance that influence binding. Furthermore, studies of ligand-macromolecule complexes offer the possibility of quantitative structure-activity relationships (QSAR) and may provide a rational explanation for the behaviour of certain ligands to aid prediction and subsequent testing of new drug molecules.

This work focuses on inhibitors of glycogen phosphorylase (GP) that may be of clinical interest for the treatment of diabetes. Diabetes mellitus is characterized by elevated blood glucose levels and affects 1–2% of the population of the Western world. Type II or non-insulin dependent diabetes mellitus (NIDDM) accounts for 75% of the cases. Treatment of NIDDM patients by diet, exercise, hypoglycaemic drugs or insulin therapy is not always satisfactory since it is difficult to control the concentrations of glucose in the blood as effectively as in normal patients. Consequently, there is continued interest in drugs that more closely resemble the physiological control of blood glucose. In NIDDM, high blood glucose levels are a result of diminished insulin release and/or insulin resistance with the consequence of impaired glucose uptake and impaired suppression of hepatic glucose production. In response to insulin release after eating, muscle glycogen synthesis is the primary method of glucose disposal. In diabetic patients, insulin does not have this effect and consequently fails to stimulate glycogen synthesis.

In mammals glycogen is the carbohydrate reserve of most metabolically active cells. Cellular demands to produce glucose-1-phosphate (Glc-1-P) from glycogen are met by the large regulatory enzyme GP which catalyses the first step in the phosphorylysis of glycogen to Glc-1-P. In muscle, the product is fed into glycolysis to provide energy to sustain muscle contraction and in the liver it is converted to glucose to supply energy for other tissues. GP is a classic control enzyme and exists in two (or at least two) interconvertible states, a T state which has low affinity for substrate and low activity, and an R state that has high affinity and high activity (Graves & Wang, 1972; Johnson *et al.*, 1989; Johnson, 1992). In the liver the major control mechanism is through hormonal stimulated phosphorylase kinase action that converts the inactive phosphorylase b (GPb) to active phosphorylase a (GPa) through phosphorylation at serine 14. In muscle activation of phosphorylase is exerted both through phosphorylation and by non-covalent binding of allosteric effectors. Glucose itself is a physiological regulator of hepatic glycogen metabolism with a K_i for GP = 1.7 mM (Stalmans, DeWulf & Hers, 1974). It

promotes inactivation of phosphorylase by competitive inhibition with the glucosyl substrate and by stabilization of the T state of the enzyme. The T state is a better substrate for phosphatase, the enzyme that converts active phosphorylase a to inactive phosphorylase b. Glucose acts synergistically with insulin and together these two agents lead to diminished hepatic glycogen degradation and enhanced glycogen synthesis. Muscle phosphorylase a is also regulated by glucose *in vitro*, but such regulation is unlikely to be significant *in vivo* since concentrations of glucose in muscle are negligible. There is the expectation that if a sufficiently powerful and specific inhibitor of phosphorylase could be achieved then such an agent should shift the balance between glycogen breakdown and glycogen synthesis in favour of the latter and may be of therapeutic benefit in the treatment of diabetes.

We have designed and synthesized a number of glucose-analogue inhibitors and studied their binding to GP in kinetic and crystallographic experiments (Martin *et al.*, 1991; Watson *et al.*, 1994). These studies have shown that binding modes can be unpredictable but may be rationalized with the benefit of structural data and that a buried and mixed polar/non-polar catalytic site poses problems for the systematic addition of functional groups. The knowledge of the three-dimensional structures of T-state rabbit muscle GPb (Acharya, Stuart, Varvil & Johnson, 1991) and the glucose-GPb complex (Martin, Johnson & Withers, 1990) have been used as models for the design of glucose analogues (Martin *et al.*, 1991). Glucose binds at the active site of GP buried 15 Å from the surface. Phosphorylase provides an extensive network of hydrogen bonds such that each hydroxyl group on the sugar is involved both as a donor and as an acceptor. There are 56 van der Waals interactions between glucose and the enzyme, however, only 15 are between complementary groups and there are no pyranose aromatic ring interactions that are characteristic of the tight-binding periplasmic monosaccharide-binding proteins (Quioco, 1989). Examination of the van der Waals packing surfaces for glucose has shown an empty pocket in the β -configuration at the anomeric carbon (C1) of glucose and a smaller, water-filled pocket in the α -configuration. These sites at the α and β positions of C1 have been the target for substituent groups. It has been postulated that, in the liver, glucose analogues with greater inhibition of GP may result in more effective regulatory agents than glucose. Although the three-dimensional structure of liver phosphorylase is not known, the amino-acid sequence has overall 80% identity to the rabbit muscle sequence and has 100% identity between residues involved in the catalytic site (Newgard, Nakano, Hwang & Fletterick, 1986).

A primary goal in any drug-design strategy is to predict the activity of new compounds. Comparative molecular field analysis (CoMFA) and three-dimensional quantitative structure-activity relationship (3D-QSAR)

have been used in drug-design strategies. The CoMFA approach permits analysis of a large number of quantitative descriptors and uses partial least squares (PLS) to correlate changes in biological activity with changes in chemical structure. One of the characteristics of the 3D-QSAR method is the large number of variables which are generated in order to describe the non-bonded interaction energies between one or more probes and each drug molecule. Since it is difficult to know *a priori* which variables affect the biological activity of the compounds, much effort has been expended to develop methods that optimize the selection of only those variables of importance. We have previously used a selected series of glucose analogue inhibitors of GPb to optimize the variable selection using (i) the program *GRID* to describe the molecular structures and (ii) a method of generating optimal partial least squares estimations (program *GOLPE*) as the chemometric tool (Cruciani & Watson, 1994).

The *GOLPE* procedure is an advanced variable selection method. The procedure evaluates the effects of individual variables on the model predictivity. This effectively means that only those variables which improve the model predictivity will be extracted from the large amount of redundant information produced from the 3D-QSAR method. The application of the *GOLPE* procedure is in three stages. The first step is the application of a normal linear PLS model using all the variables, followed by a variable preselection using a D-optimal design (Mitchell, 1974; Steinberg & Hunter, 1984). In this way, redundancy is reduced and sufficient collinearity among the remaining variables is maintained to satisfy the PLS algorithm. The second step is to build a design matrix which contains variable combinations according to a fractional factorial design (FFD). At this stage, dummy variables are added to the design matrix which allow comparison between the effect of a true variable and the average effect of the dummy variables on the model predictivity. This then leads to the third and final stage in which the variables are either fixed or excluded from the variable combinations. This allows only the significant variables which improve the predictivity of the model to be selected. Specifically, variables exhibiting a positive effect on the model predictivity are those which show an effect statistically higher than the reference value obtained by the dummies; consequently, these variables are fixed within the variable combinations and always used in the estimation of the prediction. Conversely, variables exhibiting a negative effect are always excluded from the variable combinations. This is an effective method of eliminating a number of variables and increasing the model stability. This is an iterative process which continues until all the variables have been assigned. In this way, a final model is derived which has the highest predictive power which may then be used in the design and prediction of new drug molecules.

In this work, we examine the three-dimensional X-ray structures and kinetics of a new range of glucose analogue inhibitors of glycogen phosphorylase. The present series of *N*-acetyl glucopyranosylamine derivatives shows an overall tenfold improvement in inhibition constant compared to the previous series of analogous *C*-glucosides (Watson *et al.*, 1994). Structural comparison of *N*-methylacetyl glucopyranosylamine (*N*-methylamide, $K_i=0.032\text{ mM}$), the best inhibitor to date, with the analogous β -methylamide derivative (*C*-methylamide, $K_i=0.16\text{ mM}$) indicates the importance of forming good hydrogen bonds and obtaining complementarity of van der Waals interactions. In this work, the *GOLPE* analysis suggests that reversal of the amide moiety (from the *C*-amide to the *N*-amide) causes a favourable increase in the electronegativity in the region of the carbonyl substituent in the case of the *N*-methylamide derivative. In addition, the results indicate that placing a more polar substituent in the region of the methyl substituent of these derivatives is unfavourable. Thus, the structural information, when combined with the quantitative results of the *GRID/GOLPE* study, gives rise to the possibility of observing features which may have been overlooked by the crystallographic analysis alone. Furthermore, using a QSAR method such as *GOLPE* not only provides additional information (which may be increased by the number and variety of *GRID* probes) but also allows the possibility of predicting new compounds in a rational manner and is useful for drug-design strategies in the absence of crystallographic binding studies. Although the application of a *GRID/GOLPE* procedure to the GPb training set is still in the early stages, this work (with the new series of compounds added to the training set) provides insight into the rational design and prediction of improved inhibitors of glycogen phosphorylase.

Materials

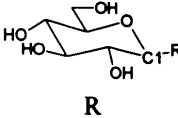



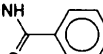
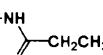
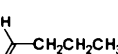
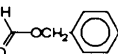

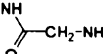
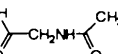
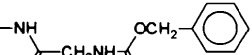
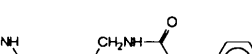
The chloro compound (1) (Table 1a) was prepared and kindly provided by R. J. Woods (Glycobiology Unit, University of Oxford). The syntheses of compounds (2) to (14) (Table 1a) were performed by G. W. J. Fleet and coworkers (Dyson Perrins Laboratory, University of Oxford) and are reported elsewhere (Bichard, Son & Fleet, 1995).

Experimental

Kinetic studies

The kinetic studies were performed as described previously (Martin *et al.*, 1991; Watson *et al.*, 1994). The glucose analogues were freshly dissolved immediately prior to use. Phosphorylase activity in the direction of glycogen synthesis was determined at pH 6.8 and 303 K by measuring the release of inorganic phosphate.

Table 1. *Glucose analogues and glucose analogue inhibitors of glycogen phosphorylase b*(a) Glucose analogues and their kinetic constants for glycogen phosphorylase *b*

Compound	Substituents at C1 position	K_i (mM)	<i>n</i>
			
(1)		0.045±0.001	1.4
(2)		0.044±0.005	1.5
(3)		0.032±0.001	1.5
(4)		0.081±0.007	1.5
(5)		0.039±0.002	1.4
(6)		0.094±0.005	1.4
(7)		0.35±0.02	1.2
(8)		0.14±0.01	1.4
(9)		0.37±0.05	1.2
(10)		0.99±0.11	1.2
(11)		0.021±0.001	1.5
(12)		0.10 ± 0.003	1.0

The enzyme (5–20 $\mu\text{g ml}^{-1}$) was assayed in 47 mM triethanolamine hydrochloride buffer (pH 6.8), 100 mM KCl, 1 mM dithiothreitol, 1 mM EDTA, 1% (w/v) glycogen and 1 mM AMP with concentrations of Glc-1-P varying between 3 and 20 mM and with concentrations of glucose analogues as indicated in the results (Table 1). Initial velocities were calculated from pseudo-first-order reaction constants (Engers, Shechosky & Madsen, 1970). Enzyme preparations exhibited V_{max} values of 70–80 $\mu\text{mol mg}^{-1} \text{min}^{-1}$ (with respect to Glc-1-P) and K_m values for Glc-1-P of 1.2–2.2 mM at saturating AMP (1 mM) and glycogen (1%) concentrations. Kinetic data were analysed by Michaelis–Menten kinetics using non-linear regression analysis (Leatherbarrow, 1987) in

which values of K_m and V_{max} and the standard error for these values were determined. Hill coefficients were generated by linear regression analysis of Hill plots. Determination of K_i values was made by plotting $K_m(\text{app})$ (slopes or intercepts) versus inhibitor concentration with an explicit value for the standard deviation of each point.

Preparation of crystals and data collection

The T-state crystals of GPb used in this study were grown as previously reported (Oikonomakos, Melpidou & Johnson, 1985; Johnson, Madsen, Mosely & Wilson, 1974). Prior to data collection, a crystal was mounted in a

Table 1 (cont.).

(b) Database of glucose analogue inhibitors of glycogen phosphorylase b

No.	Substituent at C1		K_i (mM)	No.	Substituent at C1		K_i (mM)	No.	Substituent at C1		K_i (mM)	No.	Substituent at C1		K_i (mM)
	α	β			α	β			α	β			α	β	
(1)	—OH		1.7	(14)	—CH ₂ CN		9.0	(27)			3.0	(40)			0.4
(2)		—OCH ₂ CH ₂ OH	25.3	(15)	—OH	—CH ₂ CN	7.6	(28)			21.1	(41)			0.16
(3)	—CH ₂ N ₃		22.4	(16)	-O-(1-6)-D-glucose		16.3	(29)			0.44	(42)			0.044
(4)		—CH ₂ N ₃	15.2	(17)			1.8	(30)			12.6	(43)			0.045
(5)	—OH	—CH ₂ N ₃	7.4	(18)			0.37	(31)			5.4	(44)			0.032
(6)	—CH ₂ OH		1.5	(19)	5-thio- α -D-glucose		2.0	(32)			36.7	(45)			0.039
(7)	—CH ₂ OH		21.9	(20)		—SH	1.0	(33)			5.6	(46)			0.094
(8)	—OH	—CH ₂ OH	15.8	(21)			8.6	(34)			4.4	(47)			0.35
(9)		—CH ₂ OSO ₂ CH ₃	4.8	(22)			8.6	(35)			24.2	(48)			0.14
(10)	—OH	—CH ₂ OSO ₂ CH ₃	3.7	(23)			2.6	(36)			3.6	(49)			0.99
(11)	—CH ₂ NH ₃ ⁺		34.5	(24)			8.1	(37)	1,2-dideoxy-1,2-difluoro- α -D-glucopyranose		0.2	(50)			0.081
(12)		—CH ₂ NH ₃ ⁺	16.8	(25)			2.6	(38)			1.3	(51)			0.37
(13)		—CH ₂ CH ₂ NH ₃ ⁺	4.5	(26)			16.9	(39)			2.8				

thin-walled glass capillary and soaked for 30 min in a buffered solution [10 mM *N,N*-bis(2-hydroxyethyl)-2-aminoethanesulphonic acid (BES), 0.1 mM EDTA pH 6.7] containing approximately 100 mM of the glucose analogue compound. A single crystal was used for each complete data collection. The X-ray diffraction data were collected on a Nicolet IPC multiwire area detector using a Rigaku RU-200 rotating-anode X-ray source, with a graphite monochromator, operating between 50 kV, 50 mA and 60 kV, 60 mA. The detector was placed 16 cm away from the crystal at a swing angle of 22° and data to 2.3 Å resolution were collected for each experiment. Data frames 0.2° (ω) oscillation were collected with exposure times ranging between 90 and 120 s. The data were processed using the *XDS* (Kabsch, 1988*a,b*) package to produce a scaled set of structure factors, which were then used to generate difference Fourier electron-density maps (with respect to native GPb) using the SERC Collaborative Computational Project, Number 4 (1994) suite of programs for protein crystallography

(Daresbury, England). In all but two cases [compounds (11) and (12)] the $F_o - F_c$ maps clearly showed the location and orientation of the ligand molecules ($> 3\sigma$). In all cases, the overall difference electron-density map was checked for any other changes, especially to solvent molecules. The only significant changes observed in each ligand-enzyme complex were at or near the catalytic site, with the largest changes in atomic positions of the enzyme occurring when compounds (11) and (12) were bound. A summary of the crystallographic data collection and refinement for compounds (1) to (10) is presented in Table 2.

Refinement of the ligand-enzyme complexes

Refinement of the enzyme complexes was performed on either a Convex C210 or a DECA Alpha Workstation using *X-PLOR* energy and crystallographic least-squares minimization (Brünger, 1988, 1989; Brünger, Karplus & Petsko, 1989) to *R* values of less than 0.19. The final *R* factors to 2.3 Å resolution for the glucose-enzyme

Table 2. *Crystallographic data collection and refinement statistics*

Compound	No. of reflections measured	No. of unique reflections*	No. of reflections $I/\sigma > 12$	$R_m(I)^\dagger$	R_{iso}^\ddagger	No. of reflections used in refinement	Final R^\S
(1)	75083	30222	13422	0.080	0.116	29330	0.170
(2)	87127	33625	23919	0.064	0.109	32577	0.182
(3)	86715	32746	30274	0.057	0.107	31772	0.177
(4)	76787	31649	14891	0.080	0.139	30739	0.180
(5)	76562	35293	24971	0.041	0.130	34416	0.185
(6)	76938	32042	23710	0.063	0.116	31151	0.181
(7)	77006	33912	25699	0.044	0.124	32985	0.181
(8)	76843	32078	27573	0.053	0.113	31194	0.178
(9)	74802	30038	14676	0.084	0.131	29188	0.170
(10)	75970	29472	21731	0.077	0.116	28775	0.173

* Data sets were between 77 and 95% complete to 2.3 Å resolution.

† Merging R factor for all reflections above $I/\sigma(I) = 0$.

‡ Mean fractional isomorphous difference in structure-factor amplitudes from native T state GPb.

§ Reflections between 8 and 2.3 Å resolution.

¶ Crystallographic R factor $\sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes, respectively.

complex (Martin *et al.*, 1990) and the native enzyme (Acharya *et al.*, 1991) are 0.181 and 0.191, respectively. After several cycles of positional refinement, individual atomic B -factor refinement was performed. The starting protein structure was the refined structure of the glucose complex (Martin *et al.*, 1990). The starting structure for each of the ligands was generated using ideal bond lengths, bond angles and torsion angles with the program SYBYL (Tripos Associates Inc., 1992). The glucosyl moiety was maintained in the standard 4C_1 chair conformation as observed in the glucose-GP complex. Additional groups were added in their standard minimum-energy conformation and, if necessary, adjustments were made to torsion angles to fit the groups to the observed difference density. Water molecules in the glucose-GPb complex were examined and those displaced by the ligand were removed. Final Fourier ($2F_o - F_c$) electron-density maps were calculated (with the CCP4 programs GENSFC, FFT and EXTEND) using the X-PLOR-refined coordinates for each of the ligand-enzyme complexes. A portion of a final electron-density map (r.m.s. = 0.3) in the region of the catalytic site is shown (Fig. 1) for

compound (3) bound to GPb. Estimates of the precision of coordinates indicate errors of the order of 0.2 Å in positional coordinates and are consistent with previous studies.

In each case, the final X-PLOR-refined ligand-phosphorylase complex was examined in detail and potential hydrogen bonds were assigned if the distance between two electronegative atoms was less than 3.3 Å and if the angles formed between these two atoms and the preceding atom was greater than 90°. Van der Waals interactions were assigned where the separation between non-H atoms was less than 4 Å.

GRID/GOLPE studies

Training set. Over 50 compounds bound to GPb have been studied to date (Martin *et al.*, 1991; Watson *et al.*, 1994). Most of these compounds display the same mechanism of action (by binding in a similar fashion to the catalytic site and promoting the T state of the enzyme) and cover a range in the biological activity (K_i mM to μ M). The training set consisted of 51 molecules

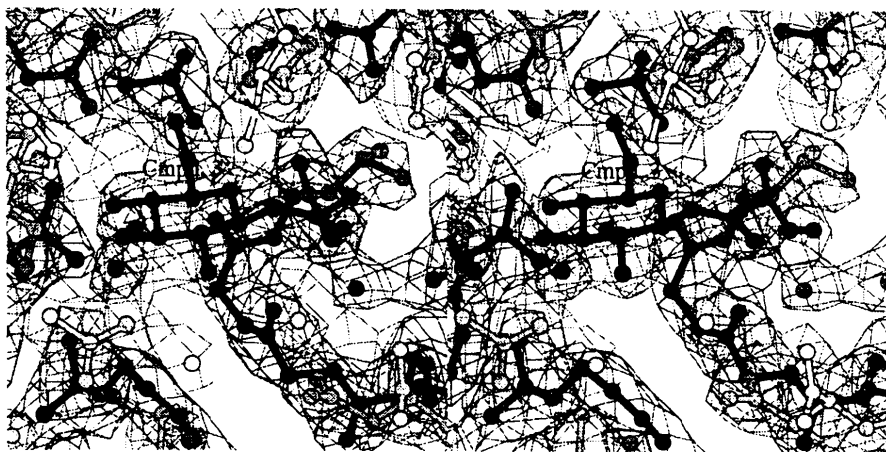


Fig. 1. A stereo diagram of a final $2F_o - F_c$ electron-density map in the region of the glucose analogue (3) bound at the catalytic site of GPb. The contour level corresponds to 3 r.m.s. deviations of the map. This figure and others was produced interactively using the program XOBJECTS (M. E. M. Noble, unpublished work).

(Table 1b) whose coordinates and bound conformations were determined from the X-ray crystallographic refined ligand–enzyme complexes. In most QSAR studies, the bound conformations and alignments of the ligands, with the enzyme and with respect to one another, are not known and this gives rise to potential errors. In the present work the conformations of the bound molecules have been established experimentally and these difficulties do not arise.

GRID force-field. The program *GRID* (Version 10.0) (Goodford, 1985; Boobbyer, Goodford, McWhinnie & Wade, 1989; Wade, Clerk & Goodford, 1993) was used to calculate the interactions between a small chemical group (the probe) and each of the ligand molecules in the training set (the targets). *GRID* is a computational procedure for detecting energetically favourable binding sites on molecules of known three-dimensional structure. The energies are calculated as Lennard–Jones, electrostatic and hydrogen-bond interactions between the probe and each target, using a position-dependent dielectric function to modulate the strong electrostatic interaction between charged centres.

Several probes have been used [such as a phenolic hydroxyl group (OH), a methyl group (C3), an sp^2 carbonyl O atom (O), a carboxy O atom (O:), an amide NH group (N1) and an O atom of either a sulfone or sulfoxide group (OS)]. For simplicity only the results of the phenolic hydroxyl group (OH) are presented in this work. Details of analysis still in progress will be discussed in future work. The energy calculations were performed using a 1.0 Å grid spacing between grid points in a rectangular box measuring $20 \times 20 \times 21$ Å. The *GRID* origin and axes were chosen such that all the atoms of the targets were included when maintained in the conformation when bound to the protein and, an additional region which extends over several active-site residues of GPb (within approximately 4 Å of the atoms of the targets) although the protein residues were not explicitly included in the calculations. A cut-off of $+5 \text{ kcal mol}^{-1}$ ($+20.9 \text{ kJ mol}^{-1}$) in positive variables was used in order to make the data more symmetrically distributed about zero, since high values for the interaction energy may be obtained when the probe is very close to the target.

One *GRID* calculation produced 7221 interaction energies for each probe with each target. Each set of calculated interaction energies contained in the resulting three-dimensional matrix from *GRID* is rearranged as a one-dimensional vector of variables (Cruciani & Goodford, 1993; Cruciani & Watson, 1994). This one-dimensional vector representation of the *GRID* data is conventionally used as input to the program *GOLPE* and the computational analysis performed as described below.

Generating optimal linear PLS estimation (GOLPE). *GOLPE* (Version 2.0) (Baroni, Constantina, Cruciani, Riganelli, Valigi & Clementi, 1993) is defined as an

advanced variable selection procedure aimed at obtaining PLS regression models with the highest prediction ability. By comparing the biological activity with the local descriptors for the ligand molecules (in this case *GRID* energies) the program tries to map those regions where particular groups of atoms (depending on the type of probe) may be added to improve the biological activity. In more technical terms, the PLS model is a two-block projection method that relates a matrix **X** (containing the chemical descriptors) to a matrix **Y** (containing the biological activities) with the aim of predicting the values **Y** from the information contained in **X** (Hoskulsso, 1988). The *GOLPE* procedure was applied in three stages as detailed previously (Baroni, Clementi, Cruciani, Kettaneh & Wold, 1993; Cruciani & Watson, 1994): (i) preselection of variables using a D-optimal criterion, (ii) building of a design matrix based on a FFD strategy, and (iii) evaluation of the effects of individual variables on the predictivity using a fixing/excluding procedure. However, the power of a *GOLPE* procedure depends on the method (and sequence) of pretreatment of the data such as, criteria to discriminate against short-range unfavourable interactions and long-range insignificant values in order to achieve a symmetrical distribution of energy values, and to establish a suitable weighting scheme. Various methods have been previously studied using a subset of the present data set (Cruciani & Watson, 1994) and optimized procedures have been established.

The D-optimal variable selection was carried out on the full data without scaling of variables, with an energy cut-off of $>+5 \text{ kcal mol}^{-1}$ (20.9 kJ mol^{-1}) on positive interactions. The D-optimal procedure was applied in three runs, obtaining a reduction in variables from 7221 to 2000, from 2000 to 1000 and, from 1000 to 871 for each run, respectively. The FFD strategy to build the design matrix used a 2:1 ratio of combinations/variables and a 4:1 ratio of true/dummy variables. The fixing/excluding procedure was used for the final variable selection from 871 variables. The final model selects only 460 out of the original 7221 variables. The *GOLPE* procedure produces a final model for each type of probe which can be evaluated by either the standard deviation error of calculations and standard deviation error of predictions (SDEC and SDEP, respectively) or R^2 , Q^2 (Baroni, Constantino *et al.*, 1993). Both R^2 and Q^2 parameters are dimensionless with values between 0 (no correlation in the data) and +1 (maximum correlation in the data). On the other hand, SDEC and SDEP have units of the actual **Y** values (in this case K_i in mM) and since they represent errors in fitting (SDEC) and errors in prediction (SDEP) these should be minimized. R^2 and SDEC represent the accuracy of the model in fitting, whereas Q^2 and SDEP represent the accuracy of the model in prediction. The Q^2 parameter is equivalent to the r^2 parameter used by CoMFA. Contour plots in loadings space represent the important three-dimensional

regions as characterized by the model. All contour maps were produced using the molecular graphics routine in *GOLPE* implemented on a Silicon Graphics CRIMSON workstation.

Results and discussion

X-ray crystallographic analysis

The new compounds examined in this study and their kinetic data (K_i) and Hill coefficients (n) are summarized in Table 1(a). The crystallographic results of ten of the new complexes are given in Table 2. X-ray analysis of the other complexes used in the training set (Table 1b) have been presented elsewhere (Martin *et al.*, 1991; Watson *et al.*, 1994).

Previous structural and kinetic studies on glucose analogues bound to GP led to a β -methyl glucoheptonic acid derivative with a K_i of 0.16 mM. This compound (the *C*-methylamide derivative of glucose) bound nearly 50 times better than the parent β -D-glucose ($K_i = 7.4$ mM). Analysis of the β -methyl glucoheptonic acid complex with GP showed only one hydrogen bond from the amide atom to O His377 (3.10 Å) with the O7 too far from ND2 Asn284 to form a hydrogen bond. A water molecule (OH4 Wat847) is displaced by the methyl substituent which makes six van der Waals contacts (three polar, three non-polar). The transfer of the methyl group from bulk solvent to the protein and displacement of a water molecule upon binding gives rise to a favourable entropic energy change. Comparison of this compound with several other glucoheptonic acid derivatives indicated that the hydrogen bond formed between the amide N atom and O His377 might play a significant role in lowering the K_i value. Comparison of binding constants for similar molecules indicated that this hydrogen bond contributed approximately 1.7 kcal mol⁻¹ (7.1 kJ mol⁻¹) to the binding energy (Watson *et al.*, 1994). This work suggested that an improvement in hydrogen bonding to O His377 and more favourable van der Waals contacts from substituent groups might lead to a better inhibitor. Using the β -methyl glucoheptonic acid derivative (the *C*-methylamide) as the lead molecule, the present series of β -*N*-acetyl glucopyranosylamines were synthesized and studied in crystallographic and kinetic binding experiments with GP with the aim of producing a more potent inhibitor. The small structural difference resulting from the reversal of the amide functionality made a significant difference in the K_i values.

β -*N*-chloroacetyl glucopyranosylamine (1) was the first compound tested in this series. The kinetic result showed three times greater binding affinity than the previous β -methyl glucoheptonic acid derivative. Two water molecules (OH4 Wat847 and OH2 Wat890, Table 3) are displaced from the catalytic site upon binding the ligand and there is evidence from the $2F_o - F_c$ electron-

Table 3. Summary of the polar contacts and the van der Waals interactions between GPb and the C1 substituent atoms of the glucose analogues

Compound (1)	No. of hydrogen bonds	No. of van der Waals contacts	Waters displaced by C1 substituents
(1)	1	23 15 non-polar 6 polar	OH4 Wat847 OH2 Wat890
(2)	1	16 10 non-polar 6 polar	OH4 Wat847 OH2 Wat890
(3)	1	18 15 non-polar 3 polar	OH4 Wat847 OH8 Wat872
(4)	1	28 22 non-polar 6 polar	OH4 Wat847
(5)	1	19 16 non-polar 3 polar	OH4 Wat847
(6)	1	21 18 non-polar 3 polar	OH4 Wat847 OH2 Wat890
(7)	1	38 30 non-polar 8 polar	OH4 Wat847 OH2 Wat890
(8)	2	17 8 non-polar 9 polar	None
(9)	1	22 13 non-polar 9 polar	OH4 Wat847 OH2 Wat890
(10)	4	29 24 non-polar 5 polar	OH4 Wat847

density map of a hydrogen bond (2.9 Å) between the chlorine substituent of the ligand and OD1 Asp339 of the protein (assuming the aspartic acid to be protonated) (Table 4). The analogous brominated derivative (2) exhibited the same binding mode as (1). The torsion angles observed for the C1 substituents of these two compounds (Table 5) place the two analogues in the same orientation. Similarly, the Br atom is within hydrogen-bonding distance of OD1 Asp339 and the same two water molecules are displaced. Not surprisingly, the kinetic results are similar for these two compounds [K_i compound (1) = 0.045 mM; K_i compound (2) = 0.044 mM].

The synthesis of the *N*-ethylacetyl and *N*-glycylglycyl glucopyranosylamines, compounds (5) and (9) respectively, followed with the aim of maintaining the substituent chain length and the possibility of a similar hydrogen bond with OD1 Asp339, particularly in the case of compound (9). Indeed both (5) (Fig. 2) and (9) (Fig. 3) adopt approximately the same orientation as (1) and (2). In the complex with (9), the N atom (N2) is just too far away from Asp339 to make a hydrogen bond (3.5 Å) but makes a hydrogen bond through O7 to OH0 Wat 887 (2.8 Å, Table 4). Conversely, compound (5) has no donor or acceptor group for a hydrogen bond with Asp339 but makes a favourable hydrogen bond through the amide N atom (N) to O His377 (2.8 Å). The van der

Table 4. Polar contacts between GPb and substituent atoms at C1 of the glucose analogue complexes

Potential hydrogen bonds are assigned if the distance between two electronegative atoms is less than 3.3 Å and if the angles formed between these two atoms and the preceding atom is greater than 90°.

Compound	Atom numbering -R	Ligand atom	Protein atom	Distance (Å)
α -D-glucose	-O1	O1	OH8 Wat872	3.0
(1)	O7 -N-C7-C8-Cl	Cl	OD1 Asp339	2.9
(2)	O7 -N-C7-C8-Br	Br	OD1 Asp339	3.1
(3)	O7 -N-C7-C8	N	O His377	2.9
(4)	O7 -N-C7-Ph	N	O His377	3.2
(5)	O7 -N-C7-C8-C9	N	O His377	2.9
(6)	O7 -N-C7-C8-C9-C10	N	O His377	2.9
(7)	O7 -N-C7-O8-C8-Ph	N	O His377	3.0
(8)	O7 -N1-C7-N2	N2 O7	OH4 Wat847 OH0 Wat887	2.8 3.2
(9)	O7 -N1-C7-C8-N2	O7	OH0 Wat887	2.8
(10)	O7 -N1-C7-C8-N2-C9-C10	O9 N1 N2 O9	O His377 OD1 Asp339 OH2 Wat890 N Asn284	3.1 2.7 3.0 2.9

Waals contacts to the methyl substituent of (5) (four non-polar, two polar) are slightly more favourable than those to the terminal N atom (N2) of compound (9) (three polar, two non-polar). There is a tenfold increase in K_i observed for (9) versus the K_i observed for (5). The

Table 5. Summary of the torsion angles of the substituent atoms at C1 of the glucose analogue complexes

Compound	Atoms at C1	Torsion angle (°)
(1)	O5-C1-N-C7 N-C7-C8-C1	-98 -51
(2)	O5-C1-N-C7 N-C7-C8-Br	-84 -46
(3)	O5-C1-N-C7	-98
(4)	O5-C1-N-C7 N-C7-C8-C9	-99 -9
(5)	O5-C1-N-C7 N-C7-C8-C9	-92 -72
(6)	O5-C1-N-C7 N-C7-C8-C9 C7-C8-C9-C10	-93 -176 -178
(7)	O5-C1-N-C7 N-C7-O8-C8 C7-O8-C8-C9 O8-C8-C9-C10	-77 -139 -106 -97
(8)	O5-C1-N1-C7	-89
(9)	O5-C1-N1-C7 N1-C7-C8-N2	-87 -16
(10)	O5-C1-N1-C7 N1-C7-C8-N2 C7-C8-N2-C9	-89 -157 -89

decrease in binding affinity for (9) may be a result of two factors; the gain in entropy from the displacement of a water molecule and the addition of a hydrogen bond is not sufficient to overcome the energy required to dehydrate the amine functionality in order to transfer it to the buried active site, and the absence of the hydrogen bond between the amide N atom and O His377 lowers the binding affinity. Conversely, with the methyl group of compound (5) there is less energy required for dehydration and the possibility of a gain from shielding the non-polar group from bulk water. This has been previously observed in the analogous C-amide series of compounds with GP (Watson *et al.*, 1994). Additional energy is gained through a favourable hydrogen bond from the amide N atom to O His377.

The result of the ligand-enzyme complex with the *N*-ethylacetyl derivative (5) led to the syntheses of

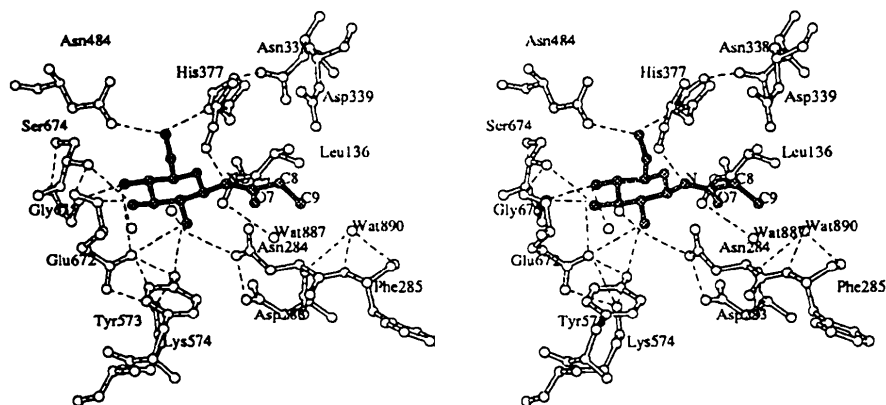


Fig. 2. Contacts between the β -*N*-ethylacetyl glucopyranosylamine, (5) ($K_i = 0.039$ mM), and GPb show potential hydrogen bonds between the amide N atom N and the main-chain O atom of His377. Analysis of the chloro- (1) and bromo-*N*-acetyl derivative (2) show similar binding to that observed for (5).

compounds (3), (4), (6), (7) and (8) in order to study the effects of various chain lengths and polar *versus* non-polar substituents. Analysis of the bound complexes with the *N*-methylacetyl (3), the *N*-ethylacetyl (5) and the *N*-propylacetyl glucopyranosylamine (6) show that all three ligands bind in a very similar orientation up to the C8 substituent (see Table 4 for labelling). The torsion angles around the C1—N bond are all very similar (average -94°) however, the torsion angles around C7—C8 in compounds (5) and (6) indicate a different orientation for

C9 in each case. In the complexes with (3) (Fig. 4) and (5) the final substituent does not come into contact with side chains of the protein, namely Asp339. The terminal methyl group in (5) (Fig. 2) does come close to the side chain of Asp339 but the side chain flips 90° about the CB—CG bond in order to avoid contact with the inhibitor. This has been observed previously (Watson *et al.*, 1994) and was generally the result of the necessity to accommodate additional substituents in the binding pocket. In contrast, compound (6) (Fig. 5) is too long

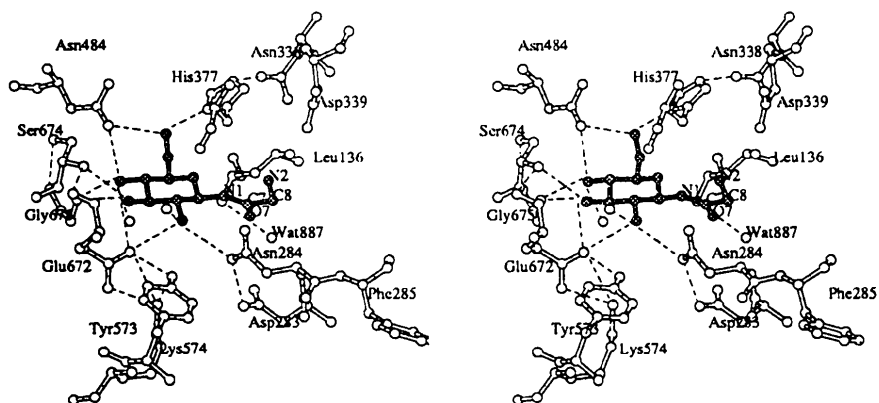


Fig. 3. Binding of *N*-glycylglycopyranosylamine (9) to the active site of GPb. The tenfold increase in K_i observed for (9) ($K_i = 0.37 \text{ mM}$) compared with (5) is the result of desolvation.

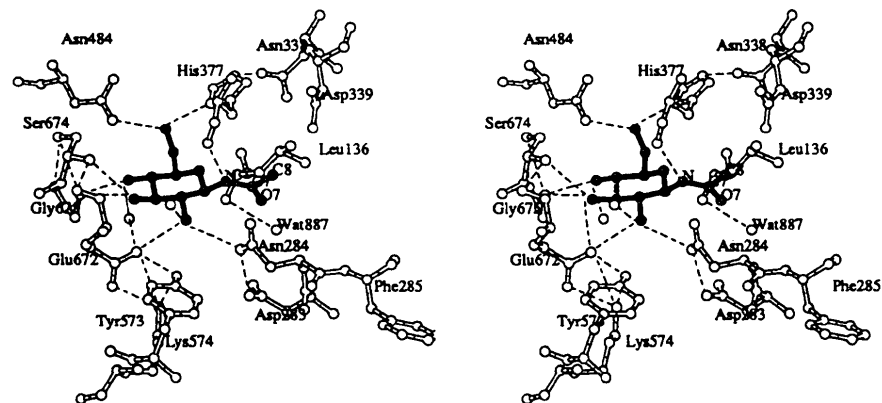


Fig. 4. Binding of the β -methylacetyl glucopyranosylamine, (3) ($K_i = 0.032 \text{ mM}$), to GPb gives rise to the best inhibitor in the glucopyranosylamine series. The decrease in K_i may be accounted for by a single good hydrogen bond from the amide N atom directly to the protein OHis377, increase in entropy through the displacement of OH4 Wat847 and OH8 Wat872 and possible gain in energy by shielding the methyl group from bulk solvent.

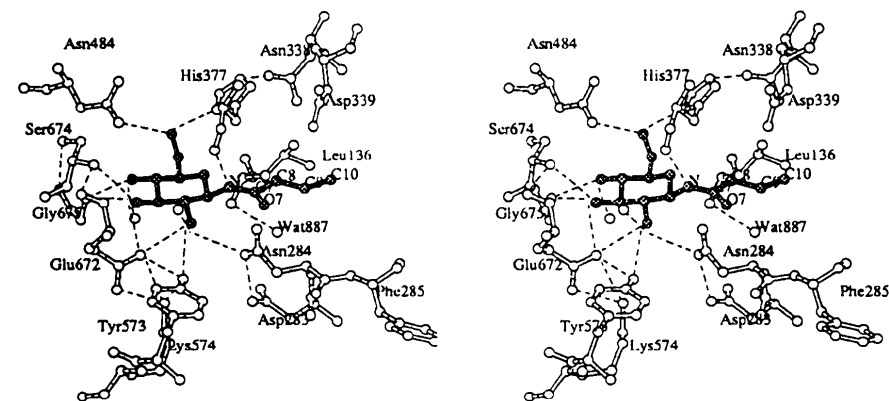


Fig. 5. Ligand-enzyme complex formed with *N*-propylacetyl glucopyranosylamine (6) ($K_i = 0.094 \text{ mM}$). The increase in K_i compared with (3) and (5) may be due to the loss in conformational entropy, since the substituents are rotated away from the protein side chain in order to avoid a clash with Asp339.

to follow the same orientation as (5) since the terminal methyl group in (6) would clash with the side chain of Asp339. Therefore, the substituents are rotated away from the protein side chain. The increase in K_i observed for the *N*-propylacetyl derivative (6) compared with compounds (3) and (5) may be due to the loss in conformational entropy. This has also been observed (Martin *et al.*, 1991) with related glucose analogue inhibitors of GP.

The *N*-methylacetyl glucopyranosylamine (3) is the best inhibitor of GP with a $K_i = 0.032$ mM, two orders of magnitude tighter than the parent glucose compound and one order of magnitude tighter than the corresponding glucoheptonic acid (*C*-amide) derivative. The structure analysis shows a better geometry of the hydrogen bond between the NH on the sugar and the main-chain O atom of His377 (Fig. 4). In the β -glucoheptonamide complex the hydrogen bond is 3.1 Å, there is a small distortion in the planarity of the amide (4°) and the hydrogen-bond angle deviates from linearity by 15° whereas in the *N*-methylacetyl glucopyranosylamine complex the hydrogen bond is 2.9 Å, the amide is exactly planar and the deviation from linearity of the hydrogen bond is 11° . In addition, OH8 Wat872 is displaced in the glucopyranosylamine complex which appears to correspond with an overall tightening of the binding pocket which can be

observed by an overall shortening of the hydrogen bonds to the peripheral hydroxyl groups of the sugar as compared to the slightly longer contacts observed in the glucoheptonamide complex. The differences in the hydrogen-bond geometry are at the limit of the precision of the data but they provide a partial explanation for the differences in affinity.

Following the analysis of these ligand-enzyme complexes it was of interest to study the effect of other more rigid non-polar substituents which led to the syntheses of compounds (4) and (7). In the ligand-enzyme complex with the *N*-benzylacetyl glucopyranosylamine (4) the phenyl ring was found to occupy the same region as the CH₂-CH₃ moiety of compound (5). The amide portion of (4) (Fig. 6) sits below those observed previously and makes a longer contact (3.2 Å) with the protein through the amide functionality. There is a long contact through O7 to OH0 Wat887 (3.4 Å). The N-C7-C8-C9 torsion angle (9°) brings the phenyl ring back up toward Asp339. In order to avoid a clash, this time the Asp339 side chain moves away from the ring but does not flip. The nearest contact is through C10 to OD1 Asp339 (3.3 Å) which may represent a weak aromatic/polar hydrogen bond (Burley & Petsko, 1988). By comparison, the amide portion of the *N*-benzyloxycarbonyl derivative (7) (Fig. 7) is found in the same region as in compounds

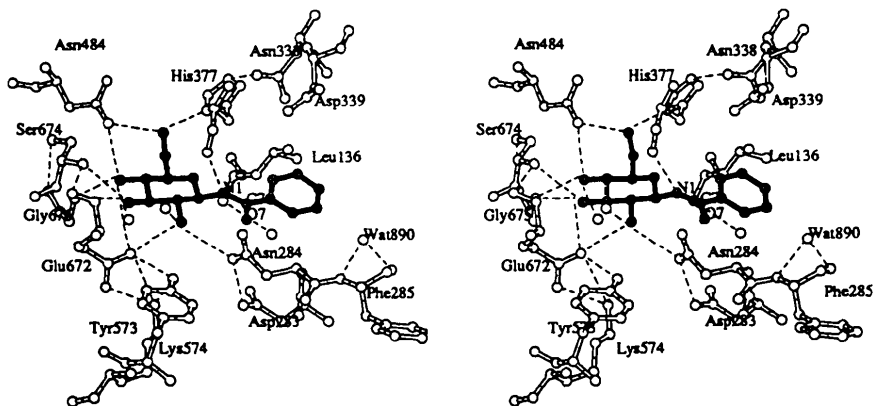


Fig. 6. Binding of the *N*-benzylacetyl derivative (4) ($K_i = 0.081$ mM) to GPb. The amine moiety sits well below that observed for the other derivatives and consequently cannot make contact to the main chain of His377. No potential hydrogen bonds are observed with this compound.

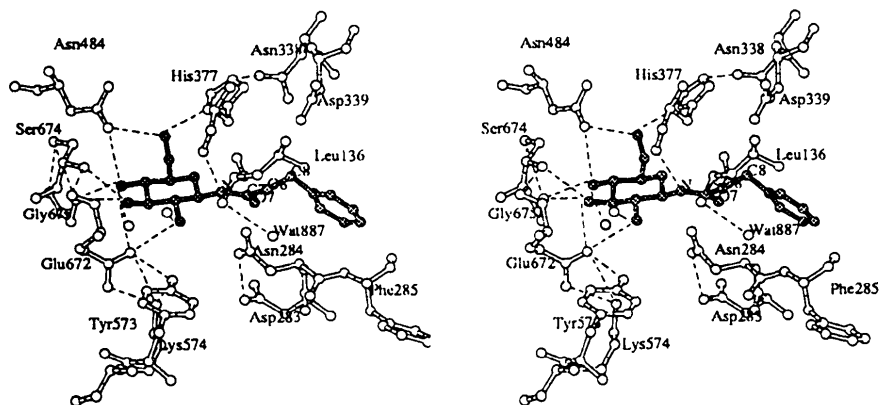


Fig. 7. Contacts between the *N*-benzyloxycarbonyl derivative (7) ($K_i = 0.35$ mM). Binding is similar to compounds (3), (5) and (6) whereby the amide N atom is within hydrogen-bonding distance to the main-chain O atom of His377.

(3), (5) and (6) gives rise to a hydrogen bond through the amide N to O His377 (3.0 Å, Table 4). The conformational freedom about the C7—O8 and O8—C8 bonds brings the phenyl ring to a similar position as that observed in (4). However, the additional substituents place the phenyl ring much further along the binding pocket such that the nearest contact with Asp339 is through C9 (3.9 Å) and there is no evidence of movement of the protein side chain away from the ligand. There is an additional water molecule displaced by the phenyl ring which sits in a relatively polar environment. Atoms of the phenyl ring make 14 van der Waals contacts to polar groups and three to non-polar groups. The ring sits perpendicular to the ring of His341. There are three separate van der Waals contacts to CE1 His341 through C9, C13 and C14 and, three to NE2 His341 through C10, C11 and C12. Although these interactions are favourable, the additional substituents (particularly C8) are placed in unfavourable proximity to Asp339.

Comparison of the complexes formed with the urea-*ido*-glucopyranose (8) (Fig. 8) and the *N*-glycinylyl glucopyranosylamine (9) (Fig. 3) show that the amide moiety sits in a similar orientation to that of (4). There is no hydrogen bond between the amide N atom and O His377 in either (8) or (9) but, the carbonyl O atom (O7) in each case forms a hydrogen bond with OH0 Wat887

(Table 4). There is an additional hydrogen bond in compound (8) from N2 of the amine functionality to OH4 Wat847 (2.8 Å). The *N*-glycinylyl derivative (9) has already been discussed in comparison with the *N*-ethylacetyl derivative (5). A similar comparison may be made between compounds (3) and (8). By analogy, there is a similar increase in binding energy due to the effects of dehydration.

Analysis of the ligand-enzyme complex with (10) (Fig. 9) shows that it superimposes well with compound (7) (Table 5). It makes a hydrogen bond through N1 to O His377 (3.1 Å) and three additional hydrogen bonds; N2 to OD1 Asp339, O9 to OH2 Wat890 and N Asn284 (Table 4). Further analysis shows substituents C8, C9 and C10 are placed in relatively polar environments (eight polar *versus* four non-polar van der Waals contacts) and N2 and O9 are placed in virtually non-polar regions (five non-polar *versus* one polar van der Waals contacts). Poor complementarity of van der Waals contacts was also observed for compound (7). Both compounds (7) and (10) are weak inhibitors of GP.

GRID/GOLPE analysis

Each of the *GRID* interactions energies (from each calculation with various probes) are rearranged as a one-

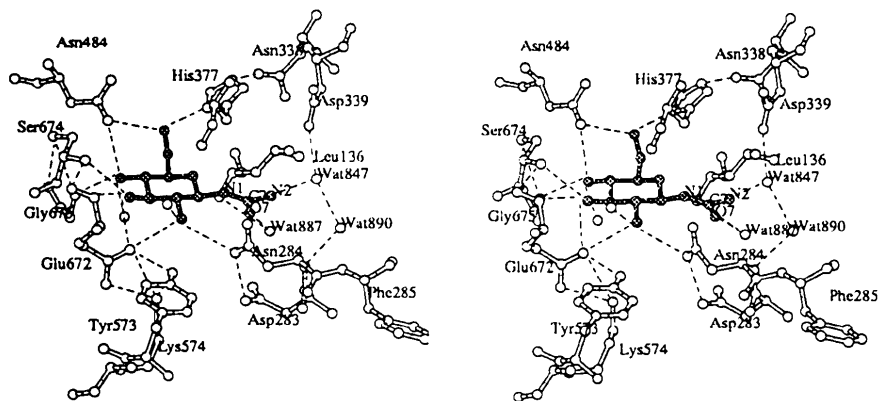


Fig. 8. Ligand-enzyme complex formed with the urea-*ido*-glucopyranose (8) ($K_i = 0.14$ mM). The compound forms two hydrogen bonds with the protein (N2 to OH4 Wat847, 2.8 Å and O7 to OH0 Wat887, 3.2 Å) however, the increase in binding energy arise from the effects of dehydration.

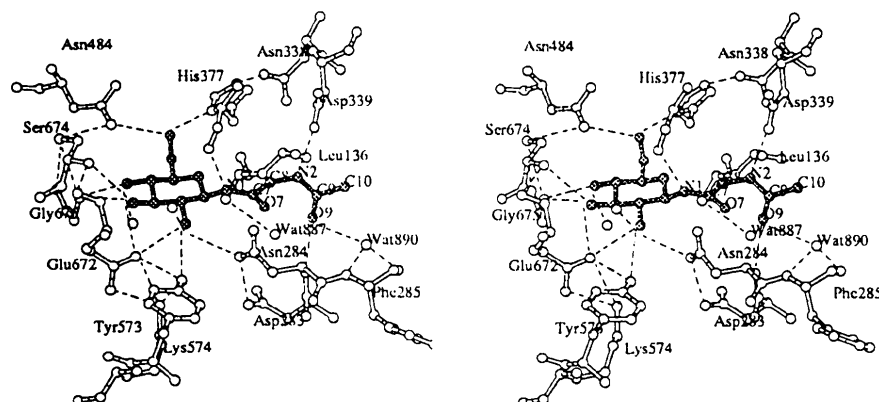


Fig. 9. Binding of compound (10) with GPb shows similarity with binding of compound (7). It makes four hydrogen bonds (three directly to the protein and one to water). Poor complementarity of van der Waals interactions is observed for both compounds (7) and (10). Both (7) and (10) ($K_i = 0.99$ mM) are weak inhibitors of GPb.

Table 6. Summary of the results of the application of the *GOLPE* procedure for 51 molecules in the *GPb* training set

The fitting (SDEC and R^2) and the predictivity (SDEP and Q^2) of the model using an OH probe and *GOLPE* variable selection

Variables*	Optimal dimension†	SDEC	R^2	SDEP	Q^2
7221	4	0.72	0.87	1.45	0.47
871	4	0.73	0.87	1.49	0.45
460	4	0.54	0.92	0.98	0.76

* Number of active variables at the three levels of variable selection. The final model with 460 variables was obtained using the fractional factorial selection procedure with 25% dummy variables and a variable combination ratio equal to 2.

† Optimal dimension is the number of principle components (PC) for a model for which there is the minimum estimated prediction error.

dimensional vector of variables (Cruciani & Goodford, 1993; Cruciani & Watson 1994) and this matrix of variables (similar in size to the *GRID* description matrix) was then used as standard input to the program *GOLPE*. The reduction of the *GOLPE* variables from 7221 to 2000, from 2000 to 1000 and, from 1000 to 871 was performed using the D-optimal procedure and from 871 to 460 using the FFD strategy and fixing/excluding procedure as described by Cruciani & Watson. Thus, the final model selects 460 variables from the initial 7221 possible. Examination of the results presented in Table 6 show that the accuracy of the model in fitting (SDEC, R^2) and in prediction (SDEP, Q^2) improves with *GOLPE* variable selection. The value of Q^2 increases from 0.47 for 7221 variables to 0.76 for 460 variables. Similarly,

the value of SDEP decreases from 1.45 for all the data to 0.98 for the final model. Comparison of the current values of SDEP (0.98) and Q^2 (0.76) with those reported earlier for an OH probe using a smaller training set (36 molecules, SDEP=1.21, Q^2 =0.49 in Cruciani & Watson, 1994) illustrates the importance of having as many different structures with a range in biological activity as possible in order to increase the reliability of the predictions. Addition of this series of *N*-acetyl glucopyranosylamine derivatives to the training set has made a significant difference in the prediction ability of the model. This increased precision of the model in prediction may then be reflected in the ability of the model to isolate only those regions of biological significance.

Figs. 10(a) and 10(b) show the contour maps of the positive and negative PLS coefficient values respectively for the final model of the interactions between an OH *GRID* probe and the 51 target molecules. There are essentially two regions identified in the loading plot of the positive PLS coefficients (Fig. 10a) which have been identified by *GOLPE*. The chemical interpretation of these regions, represented by the positive coefficients, is that a more electronegative substituent placed in these regions should give a corresponding decrease in the K_i value. The region near the carbonyl O atom of the amide moiety of the small molecule is of interest since this region was selected both by *GOLPE* and from our crystallographic analysis to be of importance to the biological activity. For example, the *C*-amide which places the NH in this region, is a less effective inhibitor than the *N*-amide which places the C=O group in this

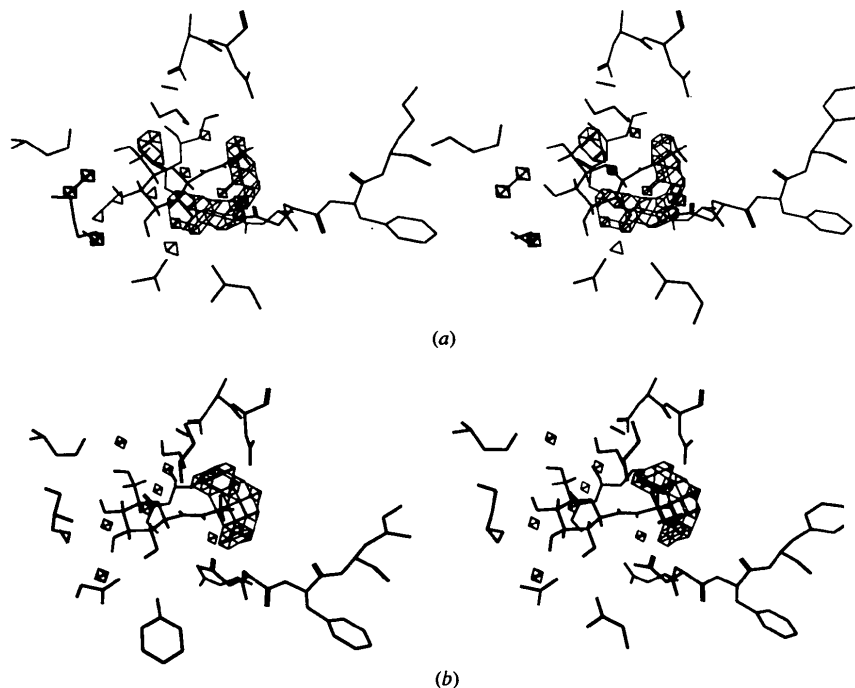


Fig. 10. Contour maps of the PLS coefficient values for the model (SDEP=0.98, Q^2 =0.76) of the interactions between an OH *GRID* probe and 51 target molecules in the *GPb* training set. (a) Loading plot of the positive PLS coefficients showing those regions in which modifications using more electronegative substituents should lead to a decrease in the K_i value. (b) Loading plot of the negative PLS coefficients showing those regions in which placing substituents that would result in more positive interactions should lead to a decrease in K_i value.

region. The loading plot of the negative PLS coefficients (Fig. 10b) shows regions where it is unfavourable to place an OH group but which it may be favourable to place a non-polar group. There is one major region located near the position of the methyl substituent in the *N*-methylacetyl glucopyranosylamine compound (3). Again, this region has been identified both by *GOLPE* and by crystallographic binding studies. For example, compound (3) ($K_i = 0.032 \text{ mM}$) which places a CH_3 group in this position is a better inhibitor than compound (8) ($K_i = 0.14 \text{ mM}$) which places an NH_2 substituent.

The results obtained by the *GRID/GOLPE* method for the design of inhibitors of glycogen phosphorylase are still in the early stages of interpretation. Preliminary results with the OH probe indicate that there is nothing to be gained by filling the large β -pocket. However, different results may be obtained with other probes. This work and computations using other probes is still in progress and will be addressed in more detail in future work.

Concluding remarks

A salutary message from the structure-based design of inhibitors of GPb is the difficulty in designing a ligand to interact with a mixed polar/non-polar active site. This has led to the systematic synthesis, crystallographic and kinetic analysis of several glucose analogues with a slow but progressive increase in binding affinity. The results of this continued study have provided some guidelines for the design of tighter inhibitors to GPb (Martin *et al.*, 1991; Watson *et al.*, 1994).

The best inhibitor to date is the β -methylacetyl glucopyranosylamine [(3)], ($K_i = 0.032 \text{ mM}$). The 250-fold improvement over the parent β -D-glucose compound is achieved through favourable entropic changes arising from the transfer of a methyl group from bulk solvent to the buried catalytic site of the protein and displacement of two water molecules leading to an overall tightening of the active site. In addition, comparison of this *N*-methylamide to the analogous *C*-methylamide ($K_i = 0.16 \text{ mM}$) (Watson *et al.*, 1994) illustrates the importance of forming good hydrogen bonds. The structure analysis showed that there was a better geometry of the hydrogen bond between the NH substituent of the sugar to the main-chain O of His377 in the *N*-methylamide complex with GPb. Although the differences noted in the hydrogen-bond geometries are at the limit of the precision of the structural data, they do provide an explanation for the differences in affinity. The van der Waals contacts made in each of these two complexes are similar and it is assumed that the difference in binding constants is a result of the hydrogen bond. This work has further emphasized the importance of forming good hydrogen bonds.

Although there is a temptation to try and fill the large pocket adjacent to the β -position of C1, addition of bulky

groups in this region often led to a diminished inhibitory effect. Substituting the small methyl substituent of (3) with a phenyl group as in compound (4) led to a difference in binding energy of $0.5 \text{ kcal mol}^{-1}$. The increase in K_i for (4) ($K_i = 0.081 \text{ mM}$) seems to be the result of the conformation geometry necessary to accommodate the large substituent and of placing a large hydrophobic group in a largely polar environment. Comparison of *N*-propylacetyl glucopyranosylamine (6) with the *N*-benzyloxycarbonyl derivative (7) shows a difference in binding energy of $0.8 \text{ kcal mol}^{-1}$, although both compounds form a hydrogen bond through the amide N atom to the carbonyl O atom of His377 in the protein. Comparison of the crystallographic and kinetic data for phenyl-substituted compounds *versus* small less bulky substituents shows that small groups can make as effective (or more effective) contributions to the K_i values.

The energy required to dehydrate polar molecules on transfer from bulk solvent to a buried active site in a protein is not compensated by the hydrogen bonds made to the protein as shown in complexes (3) *versus* (8) and (5) *versus* (9). The difference in binding energy between (3) and (8) is $0.9 \text{ kcal mol}^{-1}$, and $1.3 \text{ kcal mol}^{-1}$ between (5) and (9). The increases in K_i of compounds (8) and (9) despite the additional hydrogen bonds can be attributed to desolvation effects. This has been observed for other glucose analogue inhibitors of GPb (Martin *et al.*, 1991; Watson *et al.*, 1994).

In general, the structural studies have been successful in providing rational explanations for the differences observed in the binding affinities of the various compounds studied. However, this has not necessarily led to the discovery of a tighter inhibitor. It is recognised that the additional information from the *GRID/GOLPE* studies provide an alternative explanation for the observed differences in binding affinity and provide a quantitative method for the design and prediction of new potential drug molecules. The results so far have been encouraging. The regions predicted by *GOLPE* to have a significant effect on the biological activity, in particular the region near the amide moiety of the compounds and the region near the methyl substituent as in the *N*-methylacetyl derivative, have been identified and deduced as being important to the activity through our work on the structural studies. The *GOLPE* computation has produced a model with good overall predictivity (SDEP = 0.98 and $Q^2 = 0.76$). Further work is in progress using this model to design and predict novel compounds that might lead to a more potent inhibitor of GPb and be of therapeutic benefit in the treatment of NIDDM.

Clearly there is some way to go in our present inhibitor design studies to achieve a lowering in K_i by a further two orders of magnitude, the estimated potency necessary to modulate glycogen metabolism based on a nominal dose of 300 mg for an average adult (Watson *et al.*, 1994). There have been promising results from

biochemical and biophysical studies with the *N*-methyl-acetyl glucopyranosylamine (3). The compound has shown the anticipated effects on glycogen synthesis and degradation (Board, 1994).

We are grateful to Dr M. E. M. Noble for help with his program *XOBJECTS* for production of the figures for the crystallographic data.

References

- ACHARYA, K. R., STUART, D. I., VARVIL, K. M. & JOHNSON, L. N. (1991). *Glycogen Phosphorylase b*, 1st ed. Singapore: World Scientific Press.
- BARONI, M., CLEMENTI, S., CRUCIANI, G., KETTANEH, N. & WOLD, S. (1993). *Quant. Struct. Act. Relat.* **12**, 225–231.
- BARONI, M., COSTANTINO, G., CRUCIANI, G., RIGANELLI, D., VALIGI, R. & CLEMENTI, S. (1993). *Quant. Struct. Act. Relat.* **12**, 9–20.
- BICHARD, C. J. F., SON, J. C. & FLEET, G. W. J. (1995). In preparation.
- BOARD, M. (1995). In preparation.
- BOOBYER, D. N. A., GOODFORD, P. J., MCWHINNIE, P. M. & WADE, R. C. (1989). *J. Med. Chem.* **32**, 1083–1094.
- BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.
- BRÜNGER, A. T. (1989). *Acta Cryst.* **A45**, 42–50.
- BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* **A45**, 50–61.
- BURLEY, S. K. & PETSKO, G. A. (1988). *Adv. Protein Chem.* **39**, 125–189.
- COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* **D50**, 760–763.
- CRUCIANI, G. & GOODFORD, P. J. (1993). *J. Mol. Graphics*, **12**, 116–129.
- CRUCIANI, G. & WATSON, K. A. (1994). *J. Med. Chem.* **37**, 2589–2601.
- ENGERS, H. D., SHECHOSKY, S. & MADSEN, N. B. (1970). *Can. J. Biochem.* **48**, 746–754.
- GOODFORD, P. J. (1985). *J. Med. Chem.* **28**, 849–857.
- GRAVES, D. J. & WANG, J. H. (1972). *The Enzymes*, 3rd ed., Vol. 7, edited by P. D. BOYER, pp. 435–482. New York: Academic Press.
- HOSKULSSON, A. (1988). *J. Chemometrics*, **2**, 211–218.
- JOHNSON, L. N. (1992). *FASEB J.* **6**, 2274–2282.
- JOHNSON, L. N., HADJU, J., ACHARYA, K. R., STUART, D. I., MCLAUCHLIN, P. J., OIKONOMAKOS, N. G. & BARFORD, D. (1989). *Allosteric Enzymes*, edited by G. HERVE, pp. 81–127. Boca Raton, Florida: CRC Press.
- JOHNSON, L. N., MADSEN, N. B., MOSELY, J. & WILSON, K. S. (1974). *J. Mol. Biol.* **90**, 703–717.
- KABSCH, W. (1988a). *J. Appl. Cryst.* **21**, 67–71.
- KABSCH, W. (1988b). *J. Appl. Cryst.* **21**, 916–924.
- LEATHERBARROW, R. J. (1987). *Enzfitter. A Non-Linear Regression Data Analysis Program for the IBM PC*, Elsevier Biosoft, Cambridge, England.
- MARTIN, J. L., JOHNSON, L. N. & WITHERS, S. G. (1990). *Biochemistry*, **29**, 10745–10757.
- 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. (1991). *Biochemistry*, **30**, 10101–10116.
- MITCHELL, T. J. (1974). *Technometrics*, **16**, 203–210.
- NEWGARD, C. B., NAKANO, K., HWANG, P. K. & FLETTERICK, R. J. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 8132–8136.
- OIKONOMAKOS, N. G., MELPIDOU, A. E. & JOHNSON, L. N. (1985). *Biochim. Biophys. Acta*, **832**, 248–256.
- QUIOCHO, F. A. (1989). *Pure Appl. Chem.* **61**, 1293–1306.
- STALMANS, W., DE WULF, H. & HERS, H. G. (1974). *Eur. J. Biochem.* **41**, 127–134.
- STEINBERG, D. M. & HUNTER, W. G. (1984). *Technometrics*, **26**, 71–76.
- Tripos Associates Inc. (1992). *SYBYL Molecular Modeling Software*, St. Louis, Missouri, USA.
- WADE, R., CLERK, K. J. & GOODFORD, P. J. (1993). *J. Med. Chem.* **36**, 140–147.
- WATSON, K. A., MITCHELL, E. P., JOHNSON, L. N., SON, J. C., BICHARD, C. J. F., ORCHARD, M. G., FLEET, G. W. J., OIKONOMAKOS, N. G., LEONIDAS, D. D., KONTOU, M. & PAPAGEORGIU, A. C. (1994). *Biochemistry*, **33**, 5745–5758.