

Characterization of a Ligand–Receptor Binding Event Using Receptor-Dependent Four-Dimensional Quantitative Structure–Activity Relationship Analysis

Dahua Pan, Jianzhong Liu, Craig Senese, Anton J. Hopfinger,* and Yufeng Tseng

Laboratory of Molecular Modeling and Design (M/C 781), College of Pharmacy, The University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231

Received November 21, 2003

Receptor-dependent four-dimensional quantitative structure–activity relationship (RD-4D-QSAR) analysis is used to map the ligand–receptor binding event characteristic of a set of 47 glucose analogue inhibitors of glycogen phosphorylase (GPb). Specifically, the geometric and energetic binding profiles are constructed, conformational changes are determined, and conformational couplings among structural units are identified for the composite set of ligand–receptor complexes. A pruned ligand–receptor model is used to estimate ligand–receptor thermodynamics. Rather than explicitly handling the large amount of structural data generated from each of the pruned ligand–receptor models, these complexes were divided into three subregions. The subregions consist of a “functional” region, the smallest region providing definitive information about inhibitor binding, and two “allosteric” regions that surround the “functional” region and are based on distances from the center of the catalytic site. Maximum information on inhibitor binding and/or inhibitor–receptor conformational changes is extracted from each of these subregions. The key sites for inhibitor binding and conformational changes in GPb are presented as grid cell occupancy descriptors (GCODs), which can be both numerically and graphically represented. An induced conformational change in both the inhibitor and the binding site of GPb occurs in a distinct manner for each complex. The inter-relationships (correlations) between GCODs from different regions are identified and probed. Such correlations validate the ligand–receptor interactions identified from the “functional” region. A long-range network of conformational associations involving ligands and the receptor is also found by exploring correlations among the GCODs for the set of inhibitors.

Introduction

Medicinal chemists have long sought the capability of mapping the evolution of interactions and conformational changes that take place within, and between, a ligand and a receptor over the event of their binding to each other. Of even greater practical value, with respect to lead optimization, would be the ability to concatenate each individual mapped binding profile from a set of analogue ligands to a common receptor and extract a set of rules for tight binding.

The purpose of this paper is to demonstrate that receptor-dependent four-dimensional quantitative structure–activity relationship (RD-4D-QSAR) analysis, a quantitative structure-based design tool, can be used to accomplish detailed mapping of a ligand–receptor binding event and to integrate the information from a set of such binding profiles across an analogue series of ligands. The composite information from the binding events of the ligands is packaged as a site-occupancy-weighted three-dimensional (3D) pharmacophore. The weights provide a relative ranking of the importance of each of the pharmacophore sites for ligand binding. Occupancies of the pharmacophore sites are determined by sampling the family of thermodynamically accessible conformational states of the ligand–receptor complex.

All of the weighted pharmacophore sites are coupled to one another such that the coupling constants reflect the degree of cooperativity occurring among sites of the ligand–receptor complex over the course of the binding event. Thus, the dynamic nature of the ligand binding process is captured.

Structural-based design (SBD) has evolved as the tool of choice for de novo drug design and is increasingly being used in virtual screening and lead optimization applications. Several comprehensive reviews regarding the progress and current status of SBD have been published.^{1–4} In an SBD study the starting point is usually a 3D structure of a target protein, preferably a ligand–protein complex, with a known function that is of pharmaceutical and/or medicinal interest. The 3D structure is obtained from X-ray crystallography, NMR spectrometry, and/or homology modeling. The protein data bank,⁵ PDB, is currently the most comprehensive source for high-quality 3D X-ray structures.

Once a 3D structure of a receptor is available, the locations of ligand binding are sought. If the 3D structure of a ligand–protein complex is available, the bound ligand provides valuable information regarding the location and corresponding alignment requirement of a binding site.

Many methods have been developed to analyze protein structures including grid-based hot-spot analysis,⁶ multiple copy simultaneous search (MCSS),⁷ and com-

* To whom correspondence should be addressed. Phone: 312-996-4816. Fax: 312-413-3479. E-mail: hopfingr@uic.edu.

posite crystal-field analysis.^{8,9} Information gained from analyzing the 3D structure of the target protein, especially from the binding site, is subsequently used in searching for novel ligands. Two ligand-search approaches are popular. A direct way to find new ligands is to search existing combinatorial and/or virtual libraries for ligands that are readily incorporated into the binding site based on complementary shape and/or electronic features compared to that of the binding site lining.^{10,11} The second method is *de novo* ligand design based on the findings from a ligand–protein binding interaction analysis. In most of the cases, a new ligand is built up from pieces of atoms and/or chemical fragments that fit or bind at various locations in the binding site.¹² The potency of the putative ligand, from virtual screening or *de novo* design, is then computed using a scoring function(s) or a ligand–protein binding energy estimation. Ligands ranked highest in predicted potencies are further investigated, which may include their synthesis and assaying.

SBD, however, currently has four classes of limitations that diminish the reliability and efficiency of this design strategy, especially in identifying tight binders.

(1) Ligand alignment and/or conformational flexibilities are often not taken into consideration. Moreover, the criteria used to select the single representative binding conformer are usually ambiguous and based on conformations of the free ligand, which can be different from those realized when the ligand binds to a receptor. In some studies multiple conformers have been evaluated in an effort to minimize the binding conformation problem, which has been reviewed by Lyne.¹³ Unfortunately, the limited conformational ensemble profile is again obtained for the free ligand without assessing the physicochemical environment at the binding site that may alter ligand conformation.

Normally, only a single ligand binding alignment is selected, and this alignment is arrived at by analogy to “similar” ligand–receptor bound complexes. In addition, the alignment is often “frozen” so that the ligand cannot move when bound to the receptor.

(2) The 3D structure of the receptor (a protein) is usually treated as rigid, and a common binding alignment is assumed for all ligands in an SBD study. Ligand–receptor induced conformational changes, which can play an important role in ligand recognition, are effectively omitted from most SBD methods for the purpose of reducing the complexity and computational cost. However, it has been demonstrated that the conformational change of a protein (receptor) can be very large upon ligand binding.^{14,15} Significant binding mode changes at a binding site¹⁶ would be expected to invalidate a fixed binding alignment profile for new ligand design. Moreover, ligand docking, an essential part of an SBD study, would not be meaningfully realized unless the protein (receptor) flexibility was taken into consideration.¹⁷

Some recently developed SBD methods utilize multiple protein conformations (MPS),¹⁷ which incorporate protein flexibility to varying extents. These methods usually investigate a relatively small number of conformations of the protein, or protein–ligand complexes, to generate a composite virtual model of the receptor binding site. This composite virtual model is then used

as a virtual screen of a large trial database, which may have hundreds of thousands of compounds, in a search for “hit compounds”. However, the results of an MPS study, owing to the small number of conformations considered, may not present a full spectrum of binding sites to many of the virtual compounds in the database. That is, a protein (receptor) may adopt a distinct conformation for the binding of each individual ligand and only some of these conformer states are sampled by MPS. Moreover, MPS does not estimate differences in energy among protein conformations that can contribute significantly to a scoring function.

Other current SBD methods permit limited receptor flexibility, most often side chain flexibility in the binding site.¹⁸ Of course, any constraint on flexibility begs the question if enough flexibility is allowed for a given system to yield reliable results.

(3) Scoring functions and/or force fields are the major means of predicting the binding affinity of a ligand to a receptor. No force field or scoring function appears sufficiently robust to permit accurate estimation of ligand binding, especially tight binding (potent) ligands. Moreover, most estimations of ligand–receptor binding leave out free energy contributions to binding (entropy, for example) and/or neglect the unbound state.

(4) The desolvation and resolution of both the ligand and its receptor that is inherent to their mutual binding to one another can have a direct and significant impact on the strength of binding and the corresponding measurement of the resultant biological activity. However, there are problems and limitations in both state sampling and force field representation when modeling aqueous solvation contributions in an SBD study. Accurately estimating both the geometrical and energetic effects of aqueous solvation upon ligand–receptor binding is quite difficult.

The ability to map the ligand–receptor binding event must necessarily overcome, to an appreciable extent, each of these four classes of limitations to current SBD. RD-4D-QSAR analysis was designed to minimize the impact of each of these limitations and, therefore, also offers an opportunity to carry out *quantitative* structure-based design. This capability is discussed in a previous paper detailing the RD-4D-QSAR methodology and an example of its application.¹⁹

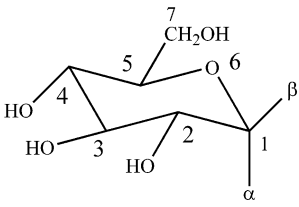
Materials and Methods

Training Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase-b (GPb). The structure–activity data of the 47 glucose analogue inhibitors forming the training set used to conduct RD-4D-QSAR analysis are given in Table 1. These structures, and the corresponding inhibitory binding constants (K_i), are reported in refs 20–23. The free energy of binding, ΔG , is estimated from K_i by

$$\Delta G = -RT \ln K_i \quad (1)$$

where T is the absolute temperature and R is the gas constant. ΔG is used as the measure of the biological response for constructing the RD-4D-QSAR models.

Crystal Structure of Glucose-Bound Receptor Complex. The crystal structure of the T state of glucose-bound glycogen phosphorylase-b was determined to a resolution of 2.3 Å by Martin and co-workers.²⁴ The corresponding PDB file (2GPB) was obtained from the Brookhaven Protein Data Bank.⁵ Water molecules identified in the crystal structure have been excluded for the RD-4D-QSAR analyses.

Table 1. Structure–Activity Data of the Training Set GPb Inhibitors Used in the RD-4D-QSAR Analyses


compd	α	β	K_i (mM)	ΔG_{303} (kcal/mol)
1	H	NHCOCH ₃	0.032	6.23
2	H	NHCOCH ₂ CH ₃	0.039	6.11
3	H	NHCOCH ₂ Br	0.044	6.04
4	H	NHCOCH ₂ Cl	0.045	6.03
5	H	NHCOC ₆ H ₅	0.081	5.67
6	H	NHCOCH ₂ CH ₂ CH ₃	0.094	5.58
7	H	NHCONH ₂	0.14	5.34
8	H	CONHCH ₃	0.16	5.26
9	H	NHCOCH ₂ NH ₃ ⁺	0.37	4.76
10	CONH ₂	H	0.37	4.76
11	H	CONH ₂	0.44	4.65
12	H	CONHNH ₂	0.4	4.17
13	H	SH	1	4.16
14	CH ₂ OH	H	1.5	3.92
15	OH	H	1.7	3.84
16	H	CONHC ₆ H ₅	5.4	3.14
17	H	OH	7.4	2.95
18	H	CH ₂ CN	9	2.84
19	OH	CH ₂ OH	15.8	2.5
20	H	OCH ₃	24.7	2.23
21	CH ₂ NH ₃ ⁺	H	34.5	2.03
22	CONHCH ₃	H	36.7	1.99
23	CH ₃	H	53.1	1.77
24	CONH ₂	NHCOOCH ₃	0.016	6.65
25	H	NHCOOCH ₂ Ph	0.35	4.79
26	H	NHCOCH ₂ NHCOCH ₃	0.99	4.17
27	H	CONHNHCH ₃	1.8	3.81
28 ^a	OH	H	2	3.74
29	H	CONHCH ₂ CH ₂ OH	2.6	3.58
30	H	COOCH ₃	2.8	3.54
31	CONHNH ₂	H	3	3.5
32	H	SCH ₂ CONHPh	3.6	3.39
33	H	CONH–4–OHPH	4.4	3.27
34	H	CH ₂ CH ₂ NH ₃ ⁺	4.5	3.25
35	CONH–4–OHPH	H	5.6	3.12
36	OH	CH ₂ N ₃	7.4	2.95
37	OH	CH ₂ CN	7.6	2.94
38	H	CONHCH ₂ CF ₃	8.1	2.9
39	CONHPh	H	12.6	2.63
40	COO [–]	H	15.2	2.52
41	H	CH ₂ NH ₃ ⁺	16.8	2.46
42	CONHCH ₂ CH ₂ OH	H	16.9	2.46
43	H	SCH ₂ CONH–2,4–F ₂ Ph	18.9	2.39
44	H	SCH ₂ CONH ₂	21.1	2.32
45	CH ₂ N ₃	H	22.4	2.29
46	COOCH ₃	H	24.2	2.24
47	CONHCH ₂ –2,4–F ₂ Ph	H	27.2	2.17

^a The O6 on glucose ring is replaced by S.

RD-4D-QSAR Method. The RD-4D-QSAR methodology has been reported in detail in a previous paper.¹⁹ A brief description of the method is given in this paper only to expedite a general understanding of the approach. The current commercial version of the RI-4D-QSAR software package,²⁵ version 3.0, was used to perform many RD-4D-QSAR steps in the study.

Step 1. Receptor Pruning and Atom Charge Assignment. Receptor pruning, using HyperChem 5.01,²⁶ is performed for the complete glycogen phosphorylase-b glucose bound complex to scale down the protein to a manageable size structure effectively containing only the “lining” of the binding site. Atom 1 of the bound glucose inhibitor in Figure 1 is chosen as the center of the pruning volume. In a previous study the total inhibitor–GPb interaction energy for the largest inhibitor in Table 1 shows no significant change after the size (radius)

of the pruned GPb receptor increases beyond 10 Å.²⁷ Hence, residues more than 12 Å from this center, are “cut off” subject to the constraint that GPb residues having at least one atom in the 12 Å region are completely included in the pruned receptor (the binding lining) model. The 2 Å outer layer (12–10 Å) of the pruned receptor model is treated as a “frozen shell” in order to maintain the conformational integrity of the composite binding site environment of the receptor. Isolated residues, separated by less than five intervening residues, are connected through the missing residues. The important 280 loop of GPb (residues 282–286) is maintained by including the residues cut off in the pruning process. The GPb cofactor, pyridoxaldehyde phosphate (PLP), is also retained in the pruned receptor model. The atoms are assigned AMBER partial charges.²⁸

Step 2. 3D Model Building of the Training Set. The

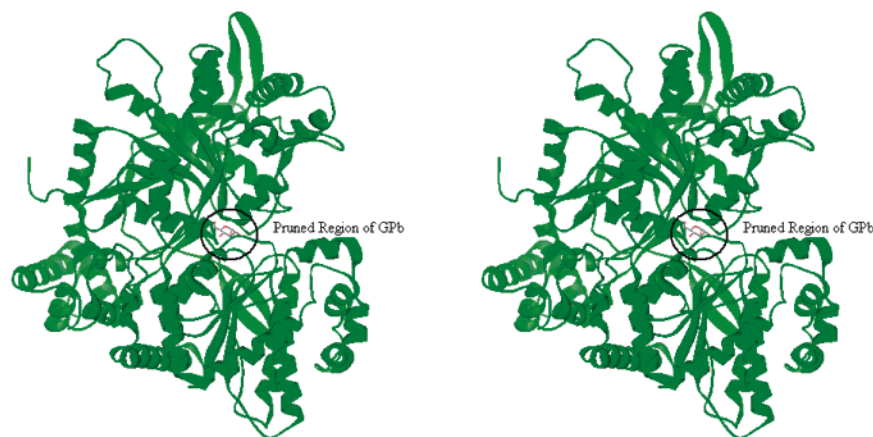


Figure 1. Graphical representation of the glucose-bound GPb complex. The (schematic) protein is in green. The bound glucose is colored by elements: oxygen atoms are in red and carbon atoms in gray. The pruned receptor model of the protein is defined by the structure within the 12 Å black circle.

geometry of glucose bound in the catalytic site of GPb, from the 2GPB file, was used as the template for the starting conformation and the binding alignment for each of the training set inhibitors. Partial atomic charges of the inhibitors were computed using the semiempirical AM1 method.²⁹ A minimum energy conformation of each inhibitor, before being docked into the binding site of GPb, was determined using the MM steepest descent method.²⁹

Step 3. Ligand Docking. All the inhibitors of the training set are glucose derivatives, which make it possible to align them to the binding topology of glucose to generate an initial binding geometry of each inhibitor–GPb complex. The inhibitors were aligned to the invariant three atoms (6, 2, 4) of the bound glucose (see the structures listed in Table 1). The bound glucose was then removed from the binding site, and the corresponding set of inhibitor-bound complexes were generated.

Step 4. Atom Type (Interaction Pharmacophore Element, IPE) Assignment. All the atoms of the pruned GPb–inhibitor complex are assigned interaction pharmacophore elements (IPEs), which are defined as follows: (a) any type of atom (any); (b) nonpolar atom (np); (c) polar atom of positive partial charge (p+); (d) polar atom of negative partial charge (p−); (e) hydrogen bond acceptor (hba); (f) hydrogen bond donor (hbd); (g) aromatic atoms (ar). The atom type (IPE) assignment permits the classification of ligand–receptor interactions.

Step 5. Constraining Selected Receptor and Inhibitor Atoms. Overall, the residues located at the ends of the fragments, and those also located at the outer edges of the pruned receptor model, have considerably more conformational flexibility than when they are part of the complete parent protein. To facilitate the pruned receptor model retaining the highest conformational similarity to the complete protein, a constraining approach called coordinate fixation has been applied. In the coordinate fixation method, the coordinates of only those atoms that are outside a 10 Å inner sphere of the 12 Å radius spherical pruned ligand–receptor complex are fixed. All of the atoms inside the “rigid outer layer” retain their intrinsic conformational freedom. The translational and rotational movements of the bound inhibitor in response to the pruned GPb model are constrained at atoms (6, 2, 4) of the glucose ring (see top of Table 1). This binding alignment constraint is necessary to permit the generation of consistent grid cell occupancy descriptors (GCODs) across the training set.

Step 6. Molecular Dynamic Simulation (MDS). MDS, using in this study the MOLSIM package,³⁰ generates the conformational ensemble profile (CEP) of each pruned ligand–receptor complex. Before the MDS was performed, geometry optimization was performed using both MM steepest descent and conjugate gradient methods to obtain the lowest potential energy for each individual complex. Each resulting low-energy complex was selected as the initial structure in the corre-

sponding MDS. An MDS of 20 ps sampling time with time-step intervals of 0.001 ps was performed for a total sampling of 20 000 conformations of each pruned GPb–inhibitor complex at a simulation temperature of 300 K and a molecular dielectric of 3.5. A total of 2000 frames were retained to construct the CEP and the corresponding GCOD trial descriptor set of each model ligand–receptor complex.

Step 7. Alignment in a Binding Site. In a RD-4D-QSAR study the lining of the binding site is included and inappropriate test alignments usually introduce sterically forbidden overlaps of parts of an inhibitor with parts of the binding site model. Moreover, it is known from previous work²³ that the glucose rings of most of the inhibitors in Table 1 bind to the receptor in very similar alignments. Atoms (6, 2, 4) of the glucose rings of the inhibitors (see Table 1) is the only allowed alignment found for this RD-4D-QSAR study.

Step 8. Grid Analysis. Each CEP conformation of a pruned receptor–ligand complex is placed in the reference grid cell lattice according to the alignment under consideration in step 7. The resolution of the grid cell lattice is 1 Å. The normalized absolute occupancy of each grid cell by each IPE atom type over the CEP for the alignment provides the trial pool of RD-4D-QSAR independent variables referred to as the grid cell occupancy descriptors.

Step 9. Local Lattice Generation and Subregion Delineating. The previous RI-4D-QSAR analysis of the training set in Table 1 indicates that the most important GCODs are located very close to the glucose ring.³¹ Thus, initially focusing this RD-4D-QSAR study to small regions (local lattices) centered at the binding (GPb catalytic) site, instead of the entire ligand–receptor complex, can provide an opportunity to explore such a “local” binding phenomenon.

Generation of a local lattice is an operation similar to the receptor pruning process. Atom 1 of the glucose ring (see Table 1) is chosen as the center of each of the local lattices having side lengths of n Å ($n = 2, 4, 6, 8, 10, 12$) “cut out” as cubes from the complete reference grid lattice (see Figure 2). Structural information contained in each of these cubes is evaluated individually to determine the appropriate ligand–receptor volume that possesses the definite information for ligand–receptor binding. Work done in a previous study¹⁹ demonstrates that a cube with side length 8 Å is essential for “explaining” inhibitor–GPb binding and is defined as the “functional” region. Two “allosteric” regions, which wrap about the “functional” region (8 Å cube), are then defined relative to their distances to the center of the GPb–ligand complex: “allosteric” region 1 is defined as the layer between the inner 8 Å (“functional” region) and the outer 10 Å cube, and “allosteric” region 2 is the layer between the inner 10 Å (“functional” and the first “allosteric” region) and outer 12 Å cube (see Figure 3). Structural information contained within each subregion is then defined from its local lattice and used to provide members to a trial pool of GCODs for constructing

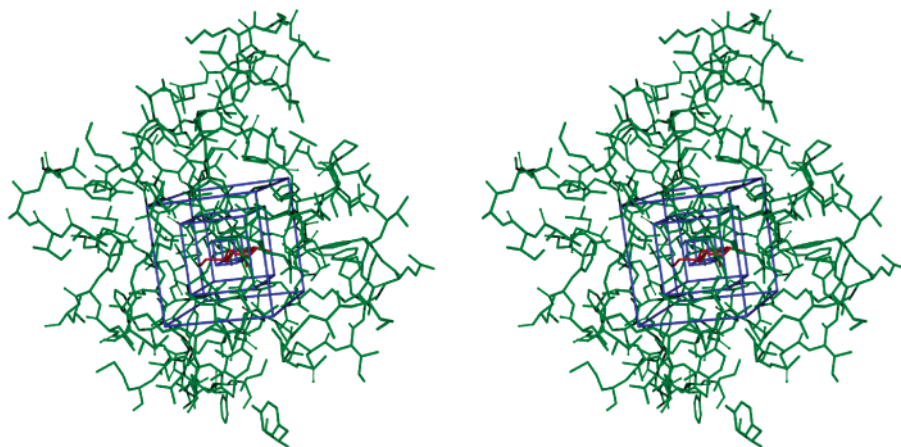


Figure 2. Conceptual representation of the local lattice generation process (step 9). The cubes, local grid lattices with side lengths of 2, 6, and 10 Å, are embedded in the pruned receptor model.

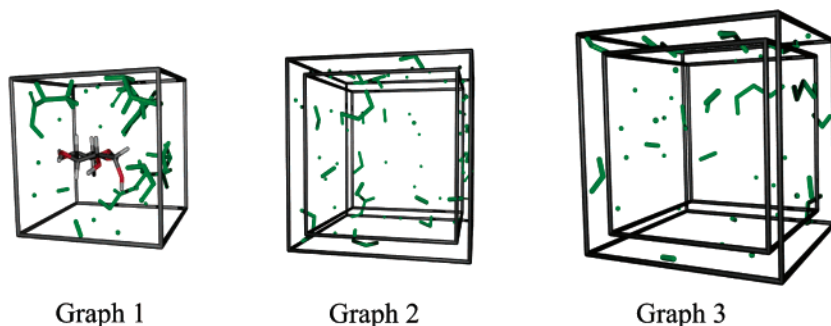


Figure 3. Conceptual representation of the subregion dividing process (step 9): (graph 1) 8 Å “functional” region; (graph 2) the first “allosteric” region between the 8 and 10 Å cubes; (graph 3) the second “allosteric” region between the 10 and 12 Å cubes. Glucose is colored by atom type. The protein fragment structures are in green.

corresponding RD-4D-QSAR models with the goal of extracting binding information.

Step 10. Partial Least-Squares (PLS) Regression Analysis. PLS³² is employed as a data reduction tool to identify the most highly weighted GCODs from the entire set generated for a given subregion (“functional” or “allosteric” region) in step 9.

Step 11. Identification of Key Sites by Genetic Function Approximation–Multiple Linear Regression (GFA–MLR). The N most highly ranked PLS GCOD descriptors determined in step 10, for a given subregion, are chosen to form the trial descriptor basis set for model optimization application using the genetic function approximation (GFA).³³ Diagnostic measures to analyze the resultant QSAR models are determined as part of the GFA optimization and include the correlation coefficient r^2 and the leave-one-out cross-validation correlation coefficient q^2 .

The top 10 models from GFA optimization, which have the highest q^2 without data overfitting, are determined and investigated for each subregion. Any GCOD used more than once for each subregion among the top 10 models is retained for the next model building operation employing multiple linear regression analysis (MLR). Since the GCODs are determined by GFA, all of them may be significant. Thus, an appropriate MLR strategy, backward elimination,³⁵ is performed to filter out the relatively less important GCOD descriptors from the composite set found by application of GFA optimization to each subregion. The MLR backward elimination procedure was performed using SAS 8.1.³⁶ Significant GCODs are defined as those that survive the MLR back-elimination process at the $p < 0.05$ level.

Step 12. Active Conformation. The active conformation of each ligand–GpB complex is defined as the conformation from the MDS trajectory that predicts the highest binding affinity but is within 2 kcal/mol of the global minimum energy conformation. The active conformations of all the complexes

from the training set are obtained from the quantitative representation of the model given by

$$\begin{aligned} \Delta G = & 2.4 \text{ GC1}(0, -3, 2, \text{any}) + 6.4 \text{ GC2}(1, -2, 2, \text{hbd}) + \\ & 8.7 \text{ GC3}(-3, -1, 3, \text{p}+) + 2.5 \text{ GC4}(-1, -2, -3, \text{any}) + \\ & 2.9 \text{ GC5}(1, 4, 3, \text{any}) - 1.8 \text{ GC6}(1, 4, 4, \text{p}+) + 3.03 \quad (2) \\ n = & 47r^2 = 0.85q^2 = 0.82 \end{aligned}$$

which has also been demonstrated as the best QSAR model found in the RD-4D-QSAR analysis.¹⁹

Results

The 11 most significant GCODs found in the “functional” region (8 Å cube), denoted as GCAn, $n = 1–11$, together with their corresponding atom and/or group occupants are presented in Table 2. Each GCOD captures one key pharmacophore site important for ligand–receptor binding. Two criteria are used to determine the possible ligand–receptor interacting groups associated with the GCODs:

(1) The IPE type of a GCOD is complementary to the atom type of its interacting group. For instance, if a GCOD has an IPE type hbd, its interacting group should be a hydrogen acceptor or partially negatively charged polar atom/group.

(2) The distance between a GCOD and its interacting group should be appropriate for the nature of the interaction proposed for them. For example, the distance between a GCOD with an IPE type hbd and its interacting group, as a hydrogen bond acceptor, should be less than 3.5 Å in order to realize a hydrogen bond.

Table 2. The Most Significant GCODs (GCAs) in the 8 Å "Functional" Region

	GCOD			
	GCA1(0, -3,2,any)	GCA2(1, -2,2,hbd)	GCA3(-3, -1,3,p+)	GCA4(-1, -2, -3,any)
complex used to demonstrate occupancy	24^a	24^a	10^a	10^a and 24^a
occupant	-CO- of the β substituent of the inhibitor	proton of -NHCO- of the β substituent of the inhibitor	proton of the amide of the α substituent of the inhibitor's	backbone of Leu 136
GPb or inhibitor interacting group	proton of the side chain of Asn 284	backbone -CO- of His 377	cofactor PLP and/or -O- of the side chain of Asn 284	<i>c</i>
distance between GCOD and interacting group (Å)	2.4	2.3	2.6, 3.6	<i>c</i>
nature of interaction	hydrogen bonding	hydrogen bonding	electrostatic or hydrogen bonding	conformational change of Leu 136
	GCOD			
	GCA5(1,4,3,any)	GCA6(1,4,4,p+) ^b	GCA7(-2, -2, -1,p+)	GCA8(-2,2,3,any) ^b
complex used to demonstrate occupancy	1^a	39^a	10^a	24^a
occupant	backbone -NH- of Ser 674	proton of backbone -NH- of Ser 674	proton of the backbone -NH- of Leu 136	C ₃ -OH of the inhibitor
GPb or inhibitor interacting group	C ₃ and C ₄ -OHs of the inhibitor	C ₃ and C ₄ -OHs of the inhibitor	-CO- of the inhibitor's α substituent	adapting best orientation for interaction between C ₃ -OH and the main chain -NH- of Ser 674
distance between GCOD and interacting group (Å)	around 2.5	around 3.0	1.1	<i>c</i>
nature of interaction	conformational change of GPb for better accommodating the inhibitor	conformational change of GPb for better accommodating the inhibitor	electrostatic	conformational change of the inhibitor's glucose ring
	GCOD			GCA11(0,2,3,np)
	GCA9(0, -2,1,p+)	GCA10(-2, -1, -1,p+)		
complex used to demonstrate occupancy	8^a	10^a		24^a
occupant	proton of -NH- of the β substituent of the inhibitor	proton of the backbone -NH- of Leu 136		C ₃ of the inhibitor
GPb or inhibitor interacting group	backbone -CO- of His 377	-CO- of the α substituent of the inhibitor		adapting best orientation for interaction between 3 -OH and main chain -NH- of Ser 674
distance between GCOD and interacting group (Å)	3.8	1.7		<i>c</i>
nature of interaction	electrostatic interaction	hydrogen bonding		conformational change of the inhibitor's glucose ring

^a Inhibitor number in Table 1. ^b GCOD occupancy decreases inhibition potency. ^c No interacting group can be found for the GCOD.

The interacting groups and their appropriate interactions for each of the significant GCODs identified in the "functional" region, using the two criteria, are also represented in Table 2. The significant GCOD descriptors obtained from the two "allosteric" regions (the layer between the 8 and 10 Å cubes, allosteric region 1, and the layer between the 10 and 12 Å cubes, allosteric region 2) and corresponding interacting groups and interactions are presented in Tables 3 and 4. These GCODs are denoted by GCB n (region 1) and GCC n (region 2).

Overall, Tables 2–4 summarizes the essential ligand–receptor interactions and the key binding points, as represented by the significant GCODs, of the ligand–receptor recognition process in each of three subregions. These tables provide an overview of the complexity of inhibitor binding and the corresponding diversity of the binding modes for individual ligand recognition.

An operational objective of a RD-4D-QSAR study is to identify and characterize the most essential ligand–receptor interactions that may be directly used in structure-based drug design. Thus, only GCODs involved directly with GPb–inhibitor binding and induced conformational changes in the receptor are investigated further. These GCODs are graphically illustrated in Figures 4–8. The complexes having high occupancy values for these GCODs have been used in Figures 4–8 to graphically demonstrate how these GCODs can reveal the interacting chemical groups. Complex **24** illustrating GCA1, GCA2, GCA4, GCA8, and GCA11 is shown in Figure 4; complex **10** with GCA3, GCA4, GCA7, GCA10, and GCB4 illustrated is displayed in Figure 5; complex **1** with GCA5 is presented in Figure 6; complex **39** with GCA6 highlighted is shown in Figure 7; complex **8** for GCA9 is given in Figure 8.

Two distinct sets of GCODs can be identified from the

Table 3. The Most Significant GCODs (GCBs) Identified from the First “Allosteric” Region (the 2 Å Layer between the 8 and 10 Å Cubes)

	GCOD			
	GCB1(−2,−5,5,hba)	GCB2(−2,5,3,np) ^a	GCB3(−5,0,−1,np)	GCB4(−5,0,2,hba)
complex used to demonstrate occupancy	7^b	24^b	1^b	10^b
occupant	−CO− of the side chain of Asn 284	Gly 675	Gly 135	oxygen of the phosphate of the cofactor PLP
GPb or inhibitor interacting group	<i>c</i>	<i>c</i>	<i>c</i>	NH ₂ of the β substituent of the ligand
distance between GCOD and interacting group (Å)	<i>c</i>	<i>c</i>	<i>c</i>	2.4
nature of interaction	conformational change of GPb	conformational change of GPb	conformational change of GPb	hydrogen bonding

	GCOD		
	GCB5(0,−4,5,any) ^a	GCB6(0,−5,5,hba)	GCB7(5,2,1,any) ^a
complex used to demonstrate occupancy	39^b	24^b	24^b
occupant	−CO− of the amide of Asn 284	−CO− of the amide of Asn 284	Val 445
GPb or inhibitor interacting group	<i>c</i>	<i>c</i>	<i>c</i>
distance between GCOD and interacting group (Å)	<i>c</i>	<i>c</i>	<i>c</i>
nature of interaction	conformational change of GPb	conformational change of GPb	conformational change of GPb

^a GCOD occupancy decreases inhibition potency. ^b Inhibitor number in Table 1. ^c No interacting group can be found for the GCOD.

Table 4. The Most Significant GCODs (GCCs) Identified from the Second “Allosteric” Region (the 2 Å Layer between the 10 and 12 Å Cubes)

	GCOD			
	GCC1(−2,6,−1,np) ^a	GCC2(2,6,−5,np)	GCC3(3,−3,6,hba) ^a	GCC4(3,6,−1,hba)
complex used to demonstrate occupancy	24^b	24^b	39^b	24^b
occupant	Thr 676	side chain of Leu 139	OH of side chain of Thr 378	nitrogen of side chain of Asn 484
GPb or inhibitor interacting group	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>
distance between GCOD and interacting group (Å)	<i>c</i>	<i>c</i>	<i>c</i>	<i>c</i>
nature of interaction	conformational change of GPB	conformational change of GPB	conformational change of GPB	conformational change of GPB

	GCOD	
	GCC5(4,1,−6,any)	GCC6(4,4,6,np)
complex used to demonstrate occupancy	7^b	1^b
occupant	side chain of Thr 375	side chain of Ala 673
GPb or inhibitor interacting group	<i>c</i>	<i>c</i>
distance between GCOD and interacting group (Å)	<i>c</i>	<i>c</i>
nature of interaction	conformational change of GPB	conformational change of GPB

^a GCOD occupancy decreases inhibition potency. ^b Inhibitor number in Table 1. ^c No interacting group can be found for the GCOD.

optimized RD-4D-QSAR models (see Tables 2–4). One set describes the binding features of the ligands, showing which functional groups help enhance ligand binding potency. The other set of GCODs identify the corresponding interacting groups of the receptor and/or the rearrangement of GPb residues to realize optimal contacts with a specific inhibitor. The ligand-based GCOD set includes GCA1, GCA2, GCA3, GCA8, GCA9,

and GCA11, while the second, receptor-based GCOD set consists of GCA4, GCA5, GCA6, GCA7, GCA10, and GCODs from the two “allosteric” regions. Information regarding GCOD occupancy and spatial correlation among the corresponding interacting groups can be discerned from Tables 2–4 and Figures 4–8. Examples of the occupancy and spatial correlations for three GCODs from each of the two sets (GCA1, GCA2, and

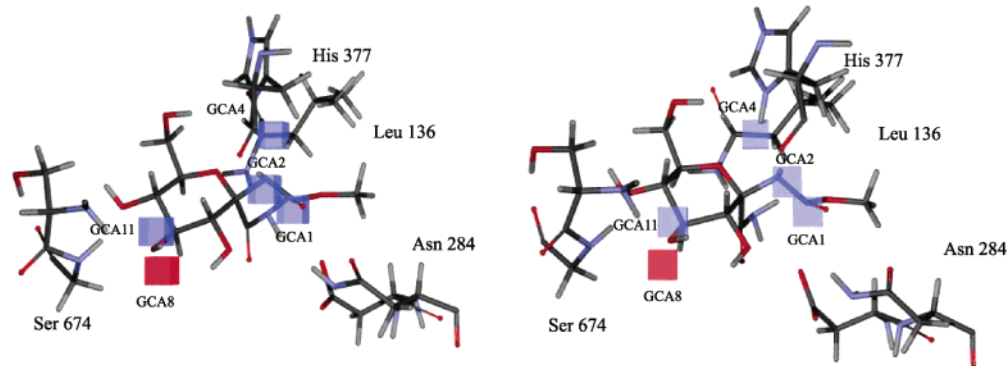


Figure 4. Simplified graphical representation of the spatial relationship between complex **24** and GCA1, GCA2, GCA4, GCA8, and GCA11.

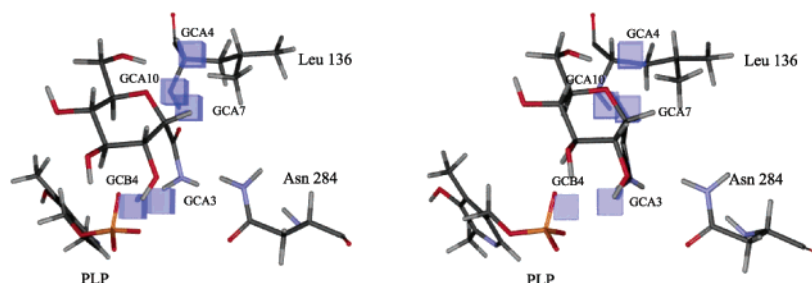


Figure 5. Simplified graphical representation of the spatial relationship between complex **10** and GCA3, GCA4, GCA7, GCA10, and GCB4.

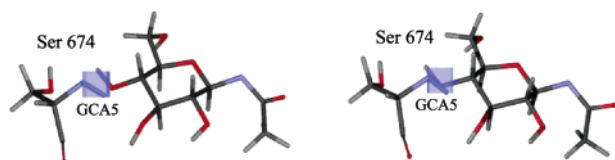


Figure 6. Simplified graphical representation of the spatial relationship between complex **1** and GCA5.

GCA3 from set 1; GCA4, GCA5, and GCA6 from set 2) are presented below to provide a perspective on how further understanding of the ligand–receptor binding event has been extracted from the 4D-RD-QSAR analysis.

The GCODs in the first set are directly related to the structures of the inhibitors and provide a direct view of some components (pharmacophore sites) of ligand–receptor binding. Thus, it is straightforward and best in a rational drug design study to first consider this set of GCODs. As shown in Figure 4, GCA1(0,−3,2,any) is situated at the −CO− group of compound **24**. GCA1 is located only 2.4 Å from an NH₂ hydrogen atom of the side chain of Asn 284, suggesting the formation of a hydrogen bond between the NH₂ of Asn 284 and the carbonyl oxygen of compound **24**. However, in a previous study¹⁹ it was established that GCA1 can be occupied by both a hydrogen bond donor and acceptor with respect to different ligand–GPb complexes. This distinct dual functionality of GCA1 is captured by the atom type “any”. The atom type “any” is used when *more than one* specific atom [IPE] type across the training set is found to satisfy the interaction being captured by a particular GCOD. This multiple IPE type occupancy of GCA1 has been discussed in the previous paper reporting the RD-4D-QSAR methodology.¹⁹

GCA2(1,−2,2,hbd) is situated at the “top” of the −NH− group of inhibitor **24** and is within hydrogen-

bonding distance (2.3 Å) to the backbone carbonyl oxygen of His 377. This observation suggests that a second ligand–receptor hydrogen bond of compound **24** with the −CO− of the backbone of His 377 contributes to the binding of this inhibitor.

GCA3 is close to the NH₂ group of the α substituent of compound **10** (see Figure 5). This GCOD is 3.6 Å from the carbonyl group of the side chain of Asn 284. An attractive electrostatic interaction, or weak hydrogen bond, is suggested by this GCOD. In addition, this GCOD is very close to the negatively charged cofactor PLP. Electrostatic interactions between the phosphate group of PLP and the NH₂ protons may also be captured by GCA3.

The second set of GCODs captures the interacting groups of the receptor and/or the conformational changes of GPb due to ligand binding. The realignment of residues at the inhibitor binding site gives the receptor enough flexibility to accommodate each inhibitor according to its structural uniqueness. The conformational changes are seemingly crucial in ligand recognition. Again, only GCODs indicative of conformational changes of the receptor that directly affects GPb–inhibitor binding are discussed in detail. As shown in Figure 4, GCA4 is occupied by the backbone of Leu 136. This observation may reflect that Leu 136 can play an important role in the binding of ligands and, in particular, a hydrogen bond between the −CO− of inhibitor **10** and the −NH− of Leu 136 contributes to the observed potency of this inhibitor. GCA3(−3,−1,3,p+) identifies an attractive electrostatic interaction between the NH₂ of the α substituent of inhibitor **10** and the phosphate group of the cofactor PLP. The conformational alignment of PLP is captured by the occupancy and role of GCB4(−5,0,2,hba) as shown in Figure 5. GCB4 is occupied by the oxygen of the phosphate group

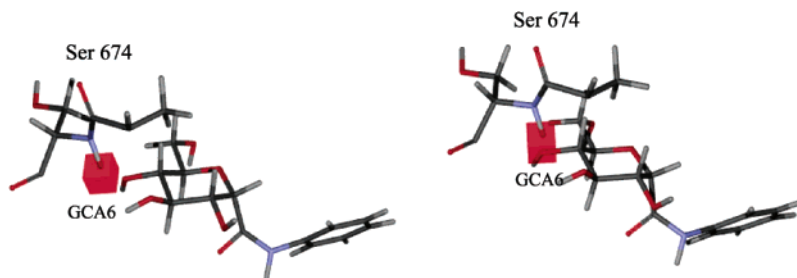


Figure 7. Simplified graphical representation of the spatial relationship between complex **39** and GCA6.

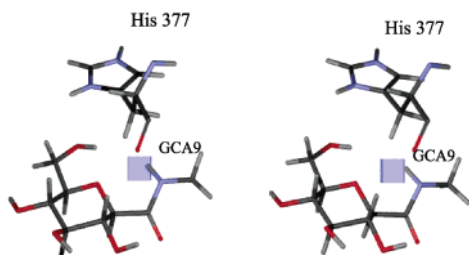


Figure 8. Simplified graphical representation of the spatial relationship between complex **8** and GCA9.

of PLP and is only 2.6 Å from the NH₂ of the α substituent of inhibitor **10**.

Rearrangement of Ser 674 upon ligand binding is captured by GCA5(1,4,3,any) as displayed in Figure 6. GCA5 is situated on the backbone –NH– of Ser 674, and it is only 2.5 Å from the hydroxyl groups of both C₃ and C₄ of the inhibitor glucose ring. Hydrogen bonding between Ser 674 and the C₃ and C₄ hydroxyl groups contribute to the relatively high inhibition potency of inhibitor **1**.

Another GCOD, GCA6(1,4,4,p+), also suggests the importance of the binding site region near Ser 674, which is shown in Figure 7. GCA6 is occupied by the proton of the backbone –NH– of Ser 674. However, unlike GCA5, its occupancy by a p+ IPE decreases the potency of the corresponding GPb inhibitor. This behavior may be due to steric hindrance and/or the electrostatic repulsion resulting from the unfavorable alignment of Ser 674 relative to an inhibitor.

The RD-4D-QSAR model explicitly demonstrates that receptor–ligand recognition, and subsequent binding, involves contributions from ligand–ligand, ligand–receptor, and receptor–receptor interactions. Moreover, induced conformational changes within the receptor are triggered by inhibitor binding and may have consequential effects on selective regions of the receptor binding site. Investigation of each individual GCOD gives a partial view of this multidimensional binding process. GCODs identified from each subregion can be correlated to one another to reveal a network of associated interactions comprising a dynamic 3D pharmacophore model. The GCOD association results are presented in Tables 5–8. GCOD pairs with correlation coefficients greater than 0.5 are considered to represent coupled interactions that can occur in a GPb–ligand complex.

Five pairs of correlated ($r \geq 0.5$) GCODs are observed in Table 5: GCA1 and GCA9; GCA4 and GCA10; GCA5 and GCA8; GCA7 and GCA10; GCA8 and GCA11. The high correlation coefficient between GCA1 and GCA9 may simply result from their spatial proximity (1.4 Å) with occupancy information actually shared by these

two GCODs. These two GCODs are occupied by members of the –NHCO– group present in many of the inhibitors. The GCODs GCA4, GCA7, and GCA10 each can be occupied by atoms of the same amino acid residue, Leu 136, as is shown in Figure 5. The correlation among these three GCODs for different ligands demonstrates the importance of the change in receptor conformation involving Leu 136 for ligand recognition. As can be seen in Figure 4, Leu 136 is located close to the β substituents of most of the inhibitors in the training set. The β substituents also possess the most structural diversity among the inhibitors. Because the β substituent of the inhibitor changes, realignment of Leu 136 is important to form favorable interactions between the receptor and the inhibitor.

The structural coupling between GCA8 and GCA11 can be readily seen in Figure 4 because they are occupied by the same part of the ligands, namely, glucose ring C₃ atom and its hydroxyl substituent. A special binding conformation in this part of the molecule is crucial to realizing high inhibitor potency due to the potential for forming multiple hydrogen bonds between the C₃ and C₄ hydroxyl groups and Ser 674.

Ligand–receptor interactions are reflected by the high correlation coefficient between GCA5 and GCA8 that is different from the structural correlations between other GCOD pairs that involve occupancy only from the same component (the same group from the inhibitor or same residue from the receptor). One of these two GCODs is occupied by the IPE atom of the inhibitor (GCA8) and the other from an atom of GPb (GCA5). Moreover, the correlation coefficient is negative, which suggests a possible steric ligand–receptor repulsion if both GCODs are simultaneously occupied. Realignment of both the glucose ring and Ser 674 is needed in order to avoid this unfavorable interaction upon complexing.

The correlation coefficients between GCOD sets from different subregions are also calculated and presented in Tables 6–8. The high correlation coefficients found in each of these three tables are, in turn, actually intercorrelated to one another. It is useful to discuss the composite set of packaged ligand–receptor interactions from all three tables. However, priority has been given to Table 6 to exploit information regarding direct ligand–receptor coupling. But possible implications of the high cross-correlation coefficients given in Tables 7 and 8 with respect to the binding event are also included below in related discussions. The most noticeable correlation coefficient in Table 6 is that between GCA3(–3,–1,3,hbd) and GCB4(–5,0,2,hba) with a value of 0.85. As has been discussed above for inhibitor **10**, GCA3 is occupied by the NH₂ group of the α substituent of inhibitor **10** and GCB4 is situated on one of the oxygens

Table 5. Cross-Correlation Coefficients for Pairs of Significant GCAs Identified from the "Functional" Region of GPb Complexes^a

	GCA1 (0,-3,2,any)	GCA2 (1,-2,2,hbd)	GCA3 (-3,-1,3,hbd)	GCA4 (-1,-2,-3,any)	GCA5 (1,4,3,any)	GCA6 (1,4,4,p ⁺)	GCA7 (-2,-1,-1,any)	GCA8 (-2,2,3,np)	GCA9 (0,-2,1,p ⁺)	GCA10 (-2,-2,-1,any)
GCA2 (1,-2,2,hbd)	0.28	1.00								
GCA3 (-3,-1,3,hbd)	-0.24	-0.12	1.00							
GCA4 (-1,-2,-3,any)	-0.06	-0.05	-0.19	1.00						
GCA5 (1,4,3,any)	0.14	0.01	-0.08	0.18	1.00					
GCA6 (1,4,4,p ⁺)	-0.17	-0.17	0.01	0.07	-0.29	1.00				
GCA7 (-2,-1,-1,any)	0.04	0.05	-0.05	0.42	-0.06	0.02	1.00			
GCA8 (-2,2,3,np)	-0.10	-0.23	0.12	-0.43	-0.62	0.15	-0.15	1.00		
GCA9 (0,-2,1,p ⁺)	0.56	0.02	-0.09	-0.19	-0.14	-0.25	-0.09	0.17	1.00	
GCA10 (-2,-2,-1,any)	0.03	0.21	-0.09	0.56	0.12	-0.30	0.57	-0.09	-0.11	1.00
GCA11 (0,2,3,np)	-0.13	-0.07	-0.04	-0.26	-0.21	-0.31	-0.07	0.56	0.06	0.13

^a Correlation coefficients ≥ 0.5 in absolute value are reported in bold print.

Table 6. Cross-Correlation Coefficients for Pairs of Significant GCODs from the "Functional" Region (GCAs) and the First "Allosteric" Region (GCBs)^a

	GCB1 (-2,-5,5,hba)	GCB2 (-2,5,3,any)	GCB3 (-5,0,-1,np)	GCB4 (-5,0,2,hba)	GCB5 (0,-4,5,any)	GCB6 (0,-5,5,any)	GCB7 (5,2,1,any)
GCA1(0,-3,2,any)	0.33	-0.10	0.33	-0.15	-0.20	0.34	-0.06
GCA2(1,-2,2,hbd)	0.12	0.10	0.01	0.03	-0.15	0.44	-0.07
GCA3(-3,-1,3,hbd)	-0.17	-0.09	-0.17	0.85	-0.12	-0.12	-0.06
GCA4(-1,-2,-3,any)	0.21	-0.33	0.27	-0.20	0.03	0.11	-0.24
GCA5(1,4,3,any)	0.64	-0.56	0.48	-0.05	-0.21	0.21	0.01
GCA6(1,4,4,p ⁺)	-0.26	0.28	-0.24	-0.07	0.48	-0.40	0.13
GCA7(-2,-1,-1,any)	0.09	-0.24	0.08	-0.10	0.08	0.29	-0.12
GCA8(-2,2,3,np)	-0.36	0.42	-0.34	0.02	0.10	-0.18	0.11
GCA9(0,-2,1,p ⁺)	0.20	0.21	0.02	-0.08	-0.17	0.30	0.09
GCA10(-2,-2,-1,any)	0.12	-0.38	0.13	-0.12	-0.14	0.45	-0.20
GCA11(0,2,3,np)	-0.05	-0.05	-0.04	0.04	0.00	-0.05	-0.19

^a Correlation coefficients ≥ 0.5 in absolute value are reported in bold print.

Table 7. Cross-Correlation Coefficients for Pairs of GCODs from the "Functional" Region (GCAs) and the Second "Allosteric" Region (GCCs)^a

	GCC1 (-2,6,-1,np)	GCC2 (2,6,-5,np)	GCC3 (3,-3,6,hba)	GCC4 (3,6,-1,hba)	GCC5 (4,1,-6,any)	GCC6 (4,4,6,np)
GCA1(0,-3,2,any)	-0.02	-0.09	-0.31	0.22	0.26	0.19
GCA2(1,-2,2,hbd)	0.19	0.19	0.00	0.42	0.07	-0.01
GCA3(-3,-1,3,hbd)	-0.18	-0.12	-0.06	-0.10	0.14	-0.10
GCA4(-1,-2,-3,any)	-0.34	0.01	0.00	-0.03	-0.01	0.03
GCA5(1,4,3,any)	-0.29	0.04	0.11	0.14	-0.07	0.81
GCA6(1,4,4,p ⁺)	-0.05	-0.27	-0.32	-0.45	-0.11	-0.25
GCA7(-2,-1,-1,any)	-0.33	-0.12	-0.25	-0.13	-0.01	0.20
GCA8(-2,2,3,np)	0.26	-0.09	-0.08	-0.38	0.16	-0.48
GCA9(0,-2,1,p ⁺)	0.18	-0.10	-0.20	0.13	0.50	-0.14
GCA10(-2,-2,-1,any)	-0.28	0.02	0.02	0.11	0.02	0.12
GCA11(0,2,3,np)	-0.01	-0.09	0.22	0.09	0.17	-0.14

^a Correlation coefficients ≥ 0.5 in absolute value are reported in bold print.

of the phosphate group of PLP, as displayed in Figure 5. The distance between these two GCODs, 2.6 Å, is ideal for hydrogen bonding, and the IPE types of these two GCODs, hydrogen bond donor and hydrogen bond acceptor, match for hydrogen bond formation. Thus, the cofactor PLP can facilitate ligand-receptor binding for glucose analogue ligands possessing specific substituents.

It is also interesting to probe the high correlation, 0.64, between GCA5(1,4,3,any) and GCB1(-2,-5,5,hba). These two GCODs are about 10 Å from each other, and their occupants belong to two different residues of GPb: Ser 674 and Asn 284. The long-range correlation between these two GCODs may be due to the coupled and cooperative conformational conversions involving

both the inhibitor and GPb. To be specific, a conformational change in the inhibitor, which is PARTIALLY reflected by GCA8, serves as the coupling agent to join conformational conversions in these two distantly separated residues of GPb. In addition to GCA5, GCB1, and GCA8, another two GCODs, GCB3(-5,0,-1,np) and GCC6(4,4,6,np), are also part of a network of highly correlated GCODs as can be deduced from the correlation coefficients among all pairs of the five GCODs (GCA5, GCA8, GCB1, GCB3, and GCC6) shown in Table 9. In summary, a network of coupled conformational realignments, shown in Figure 9, are characterized by these five GCODs and involve both the inhibitor and receptor (Ala 673-Ser 674, Gly 135, and Asn 284), with

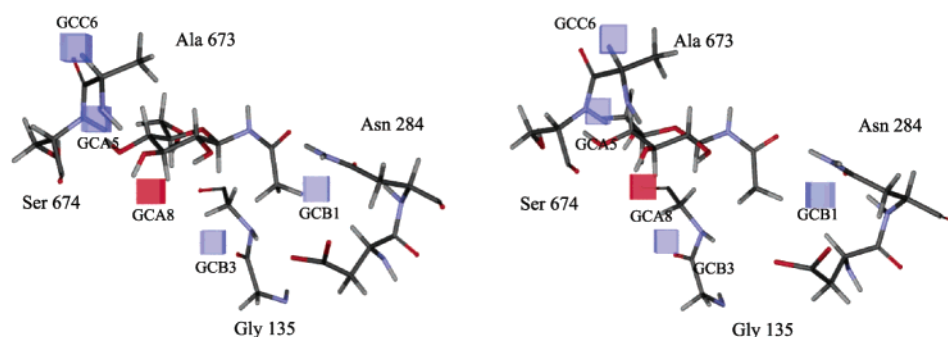
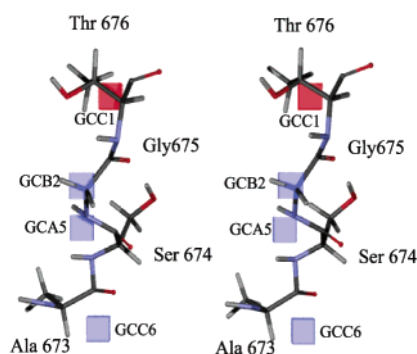
Table 8. Cross-Correlation Coefficients for Pairs of GCODs from the First “Allosteric” Region (GCBs) and the Second “Allosteric” Region (GCCs)^a

	GCC1 (-2,6,-1,np)	GCC2 (2,6,-5,np)	GCC3 (3,-3,6,hba)	GCC4 (3,6,-1,hba)	GCC5 (4,1,-6,any)	GCC6 (4,4,6,np)
GCB1(-2,-5,5,hba)	-0.04	0.07	0.05	0.04	0.35	0.59
GCB2(-2,5,3,any)	0.56	0.05	-0.22	-0.25	0.14	-0.52
GCB3(-5,0,-1,np)	-0.23	-0.14	-0.06	0.14	0.30	0.60
GCB4(-5,0,2,hba)	-0.14	-0.07	0.02	0.10	0.16	-0.08
GCB5(0,-4,5,any)	-0.08	-0.29	-0.27	-0.31	-0.13	-0.18
GCB6(0,-5,5,any)	-0.05	0.09	-0.15	0.26	-0.15	0.18
GCB7(5,2,1,any)	0.35	-0.19	0.00	-0.20	-0.10	0.08

^a Correlation coefficients ≥ 0.5 in absolute value are reported in bold print.

Table 9. Cross-Correlation Coefficient Matrix for GCA5, GCA8, GCB1, GCB3, and GCC6

	GCA5(1,4,3,any)	GCA8(-2,2,3,any)	GCB1(-2,-5,5hba)	GCB3(-5,0,-1,np)	GCC6(4,4,6,np)
GCA5(1,4,3,any)	1.00				
GCA8(-2,2,3,any)	-0.62	1.00			
GCB1(-2,-5,5hba)	0.64	-0.36	1.00		
GCB3(-5,0,-1,np)	0.48	-0.34	0.44	1.00	
GCC6(4,4,6,np)	0.81	-0.48	0.59	0.60	1.00

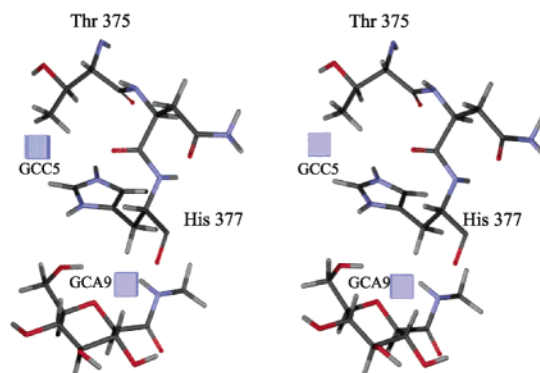
**Figure 9.** Simplified graphical representation of the spatial relationship between complex **1** and GCA5, GCA8, GCB1, GCB3, and GCC6, which illustrates a long-range structural coupling in an inhibitor–receptor complex.**Figure 10.** Spatial representation between loop 670 and the four GCODs that are occupied by this loop (GCA5, GCB2, GCC1, and GCC6). The high correlation coefficients between GCOD pairs reflect coupled conformational changes in loop 670.

the inhibitor serving as an intermediary agent to couple structurally unrelated residues of the receptor.

The final pair of correlated GCODs in Table 6, GCA5-(1,4,3,any) and GCB2(-2,5,3,any), are located at two adjacent amino residues, Ser 674 and Gly 675. Another two GCODs, GCC1(-2,6,-1,np) and GCC6(4,4,6,np), are also located in this region of the receptor, as shown in Figure 10. The high correlation coefficients among pairs of these four GCODs, reported in Table 10, provide a clear indication of conformational changes occurring in the tetrapeptide sequence (Ala 673)–(Ser 674)–(Gly 675)–(Thr 676) during inhibitor binding.

Table 10. Cross-Correlation Coefficient Matrix for GCA5, GCB2, GCC1, and GCC6

	GCA5 (1,4,3,any)	GCB2 (-2,5,3,np)	GCC1 (-2,6,-1,np)	GCC6 (4,4,6,np)
GCA5(1,4,3,any)	1.00			
GCB2(-2,5,3,np)	-0.56	1.00		
GCC1(-2,6,-1,np)	-0.29	0.56	1.00	
GCC6(4,4,6,np)	0.81	-0.52	-0.30	1.00

**Figure 11.** Simplified graphical representation of the spatial relationship between complex **8** and GCODs, GCA9, and GCC5.

GCA9(0,-2,1,p+), as shown in Figure 11, is occupied by the proton of the -NH- group of the β substituent of inhibitor **8**, and it may interact with the backbone -CO- of His 377. This interaction can be indirectly inferred from the correlation between GCA9 and GCC5-(4,1,-6,any), which is occupied by Thr 375. Thus, while

the conformational changes of His 377 are not specifically captured by any GCOD descriptor, GCC5 still provides the opportunity of probing the possible ligand–receptor interactions involving His 377 by indirectly relating the interacting group of the inhibitor to a residue (Thr 375) that is coupled to His 377.

Discussion and Conclusions

Conformational changes in the protein initiated by inhibitor binding can be investigated using the sub-region division method that permits correlating interaction sites across the defined regions of the receptor. “Functional” regions identified by subregion division provide information on the individual functional groups of the inhibitors that influence inhibition potency and also on the residues of the receptor binding site that define a unique binding mode, or signature, to each inhibitor in a particular analogue set.

The X-ray structures of several training set–GPb complexes have been determined in previous studies,^{20–22} all of which suggest that the glucose analogue inhibitor–GPb binding modes are unpredictable. Favorable ligand–receptor interactions designed for one pharmacophore site may induce loss of binding at one or more other sites. For this type of situation, where ligand–receptor binding is not “site-additive”, it may be advantageous to perform quantitative structure-based design VHTS using a RD-4D-QSAR model that evaluates the overall induced conformational changes of an inhibitor–receptor complex.

Information taken directly from the RD-4D-QSAR GCODs in principle can be used to streamline lead compound optimization in the rational drug design process. That is, the conceptual use of the information from an RD-4D-QSAR model is fairly straightforward: GCODs, whose occupancy increases inhibition potency, should be occupied with atoms of the correct IPE type, while those GCODs whose occupancy decreases potency should be avoided. However, the difficulty of applying the GCOD occupancy construct is that RD-4D-QSAR analysis also reveals, as shown in this study, that couplings can occur among the GCODs. *This coupling of GCODs may actually be a revelation of why lead optimization is quite difficult in most cases. Simply put, increasing favorable GCOD occupancy at one site may have an adverse effect at another coupled GCOD to overall binding potency.*

In the case of GPb inhibition by glucose analogues, GCODs directly related to the structure of the inhibitors, the *GCA_n*, suggest necessary functional groups for a potent inhibitor. Ideally, information contained in every GCOD can be employed. However, the nature of GCOD occupancy and the correlation between these key GCODs are not all unique or independent of each other. GCA1 has a high correlation, 0.56, with GCA9, which implies that using one of these two GCODs would be sufficient in order to avoid duplicate, competing interactions between the inhibitor and the receptor. Both GCA8 and GCA11 demonstrate the significance of conformational changes in the glucose ring on inhibitor potency. However, the conformational behavior of the glucose ring, and therefore, occupancy of GCA8 and GCA11 is difficult to control.

Functional groups already present in the training set inhibitors can be chosen to build new lead compounds.

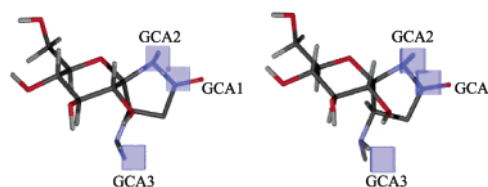


Figure 12. New lead compound arrived at by utilizing GCOD occupancy information from the “functional” region of the protein–ligand complex. The fit of the new lead compound to three key GCODs (GCA1, GCA2, and GCA3) is shown. Carbon atoms are in gray, oxygen atoms are in red, and nitrogens are in blue.

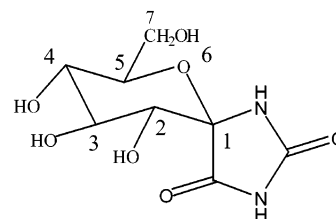


Figure 13. Structure of the known analogue of the virtual compound in Figure 12. This analogue has a measured K_i value of 3.2 μM and a predicted K_i of 5.5 μM (eq 2).

To be specific, a $-\text{CO}-$, which is suitable to occupy GCA1, a $-\text{NH}-$ appropriate for GCA2, and an amino group for GCA3 are convenient to select as a set of “correct” functional groups. These functional groups should be located at the positions specified by the corresponding GCODs. Structural constraints are thus necessary to achieve the correct positions of the functional groups. The structure of a “virtual” potent compound designed using these constraints is presented in Figure 12 together with the GCA1, GCA2, and GCA3 used to propose it. To keep all three functional groups at positions where favorable interactions form, a pseudo-spiro-ring structure is needed as shown in Figure 12. This virtual compound is predicted to have a K_i value of 0.7 μM , which is 200 times more potent than the most active compound in the training set based on eq 2. Even though no experimental data has been reported for the proposed compound, one of its analogues, shown in Figure 13, is found to have a K_i value of 3.2 μM and a predicted K_i value of 5.5 μM (using eq 2), which provides support for reliability in the predicted high inhibition potency of the virtual compound shown in Figure 12. Only three functional groups are used to achieve high inhibition potency based on three GCODs for spiro-ring inhibitors. Thus, it cannot be concluded that compounds with these three functional groups would bind tightly to GPb. The common structure of the training set compounds, the glucose ring, and its hydroxyl substituents also contributes to the binding of a glucose analogue inhibitor.

Unlike the “ligand-based” GCODs discussed above, a second set of GCODs identified from the “functional” region describes residues of the receptor involved in ligand recognition and are termed “receptor-based” GCODs. This set of GCODs includes GCA4, GCA5, GCA6, GCA7, and GCA10. GCA4, GCA7, and GCA10 indicate the significance of Leu 136, while GCA5 and GCA6 identify the contribution of Ser 674 in inhibitor recognition. In addition, possible interacting groups of the receptor, namely, Asn 284, His 377, Ser 674, and cofactor PLP, with the six ligand-based GCODs, provide

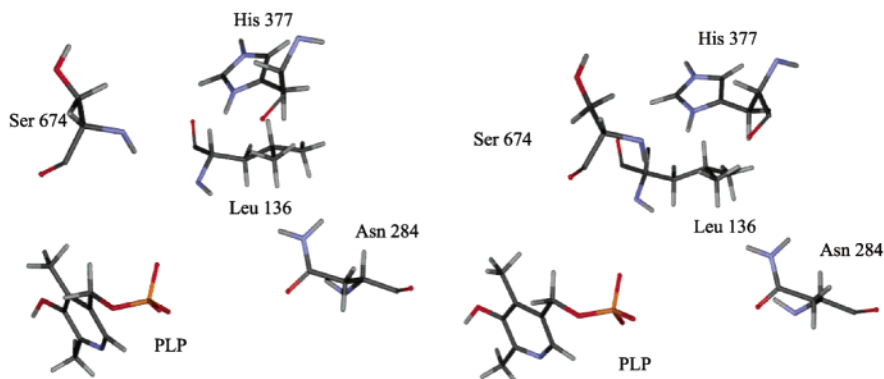


Figure 14. Geometry of the GPb binding site specific to glucose analogue inhibitors as deduced from an analysis of the GCODs. Four residues, Leu 136, Asn 284, His 377, and Ser 674, and the cofactor PLP are identified as significant for glucose analogue binding.

indirect information regarding the roles of these residues in ligand binding. All these binding residues and cofactor identified from both “ligand-based” and “receptor-based” GCODs are shown in Figure 14. These residues and cofactor are identified with differences in inhibition potency across the glucose analogues and, consequently, can be considered as target residues in structure-based drug design.

The two “allosteric” regions of the inhibitor–receptor complex do not provide direct information regarding inhibitor binding. However, these regions comprise both a supplementary and complementary component to both the inhibitor recognition process and the propagation of the binding event over the binding site. The two “allosteric” regions of the inhibitor–GPb complex embed conformational changes in the receptor that are driven by inhibitor binding. The key GCODs identified for each of these two allosteric regions are exclusively occupied by receptor atoms. Investigation of the correlations between significant GCODs in these two “allosteric” regions and the “functional” region reveals a comprehensive view of the onset and evolution of inhibitor–receptor interactions over the course of ligand–receptor complex formation.

The inhibitor–Gpb complex crystal structures are in agreement with respect to the major interactions and corresponding interacting groups (from both the inhibitor and receptor) found in this RD-4D-QSAR analysis. A comparison of structural information from crystal^{20–22,37} and RD-4D-QSAR analyses can be summarized as (1) hydrogen bonding between the amine group of Asn 284 and the $-\text{CO}-$ group of the β substituent of an inhibitor as indicated by GCA1, (2) hydrogen bonding between the main chain $-\text{CO}-$ of His 377 and the $-\text{NH}-$ of the β substituent of an inhibitor as captured by GCA2, (3) hydrogen bonding between main chain $-\text{NH}-$ of Leu 136 and the $-\text{CO}-$ of the α substituent of an inhibitor as indicated by GCA4, (4) hydrogen bonding between main chain $-\text{NH}-$ of Ser 674 and the glucose ring C5 and C6 hydroxyl groups as captured by GCA5 and GCA6, (5) conformational changes in loop 280 as signaled by GCB1, GCB5, and GCB6, and (6) conformational changes in loop 380 as represented by GCC3 and GCC5.

Crystal structures are, however, average static representations of the dynamic behavior of an inhibitor–receptor complex. Thus, X-ray structures contain limited information on possible conformational changes and

corresponding coupled motions occurring in the inhibitor–receptor binding event. Investigating regions of the ligand–receptor complex possessing highly correlated GCODs using RD-4D-QSAR analysis may provide a way to compute, process, and package dynamic information complementary to that provided by X-ray structures.

The high correlation, 0.85, between GCA3 and GCB4, seen in Table 6 and Figure 5, is evidence of a cooperative alignment of the two interacting groups, NH_2 of the inhibitor and an oxygen of cofactor PLP, to enhance ligand binding. The extent of correlation found between GCODs may serve as probes to reveal where major ligand–receptor binding interactions occur across a set of inhibitors.

The occupancy values of GCA9 and GCC5 are reasonably highly correlated, as shown in Table 7. This correlation implies the structural associations presented in Figure 11. GCA9 is occupied by the $-\text{NH}-$ of the β substituent of inhibitor **8** and indicates a hydrogen bond between this $-\text{NH}-$ and the backbone $-\text{CO}-$ of His 377 of the receptor. GCC5 is occupied by side chain atoms of Thr 375, which in turn are influenced by His 377. This long-range correlation thus reflects a ligand–receptor hydrogen bond.

If conformational changes in one receptor residue are significant for inhibitor recognition, it is reasonable to expect that neighboring residues may also be affected because of conformational coupling among residue neighbors. As shown in Figure 10, four GCODs (GCA5, GCB2, GCC1, and GCC6) are located at the same region, loop 670, and have relatively high correlation coefficients among themselves (see Table 10). Although some correlation coefficients are less than 0.50, the overall correlation among these GCODs maps the conformation coupling that takes place among the residues in this loop due to ligand binding.

Allosteric conformational changes can be identified from the five GCODs presented in Table 9 and Figure 9. These five GCODs form a correlation network starting with GCB1 and ending with GCA5. Moreover, the distance between GCB1 and GCA5 is about 10 Å, yet they have a correlation of 0.64. This high correlation between these two GCODs occurs because of a conformational coupling to each other through the inhibitor. *A quantitative structure–activity relationship can only be constructed within the RD-4D-QSAR paradigm if this long-range allosteric cooperativity is properly assessed as part of the modeling process.*

Explicit water molecules have been excluded in this study to simplify the analysis of the ligand binding process. The resolution of the typical training set inhibitor–receptor crystal complex is moderate, with an average resolution of around 2.5 Å, resulting in some uncertainty in both the locations and the orientations of the water molecules at the binding site. Moreover, crystal packing forces may exert additional constraints on the water molecules in the crystalline inhibitor–receptor complex. Conversely, excluding explicit waters in the MDS used to construct the RD-4D-QSAR model may expose both the inhibitor and the receptor to more direct interactions with one another and, thus, sample too large of conformational changes in both the inhibitor and receptor. The actual “in vivo” behavior of an inhibitor–receptor complex may be a “state” between that of the crystal structure and that found in the MDS excluding explicit water molecules. Explicit waters may be considered in a future RD-4D-QSAR study of glucose analogue–GpB complexes now that we have some understanding of the molecular motions and behavior of these complexes.

Because of limitations in computer capabilities, as well as our current understanding of “allosteric” effects in proteins, it remains difficult to generate and explore information discerned over a complete receptor protein. Such information transfer may involve long-range networks of subtle conformational changes over the entire protein. However, RD-4D-QSAR is a tool that holds the potential to investigate such complex events.

Acknowledgment. Partial funding for this study was provided by National Institutes of Health, Grant P01-GM 62195. Resources of the Laboratory of Molecular Modeling and Design at UIC and The Chem21 Group, Inc. were used in performing this work. An unrestricted financial gift from the Procter & Gamble Company is also gratefully acknowledged.

References

- Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design. *Med. Res. Rev.* **1996**, *16*, 3–50.
- Böhm, H. J. Computational tools for structure-based ligand design. *Prog. Biophys. Mol. Biol.* **1996**, *66* (3), 197–210.
- Sotriffer, C.; Klebe, G. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *Farmaco* **2002**, *57*, 243–251.
- Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78*, 269–281.
- Web site: <http://www.rcsb.org/pdb/>.
- Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29–34.
- Klebe, G. The use of composite crystal-field environments in molecular recognition and *de novo* design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
- Murray-Rust, P.; Glusker, J. P. Directional hydrogen bonding to sp²- and sp³-hybridized oxygen atoms and its relevance to ligand–macromolecule interactions. *J. Am. Chem. Soc.* **1984**, *106*, 1018–1025.
- Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078–1082.
- Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
- Böhm, H. J. Current computational tools for *de novo* ligand design. *Curr. Opin. Biotechnol.* **1996**, *7*, 433–436.
- Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7* (20), 1047–1055.
- Freire E. Statistical thermodynamic linkage between conformational and binding equilibria. *Adv. Protein Chem.* **1998**, *51*, 255–279.
- Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **1999**, *12*, 713–720.
- Brem, R.; Dill, K. A. The effect of multiple binding modes on empirical modeling of ligand docking to proteins. *Protein Sci.* **1999**, *8*, 1134–1143.
- Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
- Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235* (1), 345–356.
- Pan, D.; Tseng, Y.; Hopfinger, A. J. Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Com. Sci.* **2003**, *43* (5), 1591–607.
- 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. Glucose analogue inhibitors of glycogen phosphorylase: the design of potential drugs of diabetes. *Biochemistry* **1991**, *30*, 10101–10116.
- Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Bichard, C. J. F.; Orchard, M. G.; Fleet, G. W. J.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M.; Papageorgiou, A. C. Design of inhibitors of glycogen phosphorylase: a study of α - and β -C-glucosides and 1-thio- β -D-glucose compounds. *Biochemistry* **1994**, *33*, 5745–5758.
- Watson, K. A.; Mitchell, E. P.; Johnson, L. N. Glucose analogue inhibitors of glycogen phosphorylase: from crystallographic analysis to drug prediction using GRID force field and GOLPE variable selection. *Acta Crystallogr.* **1995**, *D51*, 458–472.
- Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure–activity relationship. *J. Med. Chem.* **1997**, *40*, 1455–1464.
- Martin, J. L.; Johnson, L. N.; Withers, S. G. Comparison of the binding of glucose and glucose 1-phosphate derivatives to T-state glycogen phosphorylase b. *Biochemistry* **1990**, *29*, 10745–10757.
- 4D-QSAR, version 3.0; The ChemBats21 Group Inc.: Lake Forest, IL, 2001.
- HyperChem, release 5.01 (for Windows); Hypercube, Inc.: Ontario, Canada, 1996.
- Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand–receptor binding thermodynamics by free energy force field three-dimensional quantitative structure–activity relationship analysis: applications to a set of glucose analog inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **1999**, *42*, 2169–2179.
- Weiner, S. J.; Kollman, P. A.; Nguyen, D. T. An all atom force field for simulation of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
- HyperChem Reference Manual; Hypercube, Inc.: Ontario, Canada, 1996.
- Doherty, D. C. *MOLSIM User's Guide*; The ChemBats21 Group, Inc.: Lake Forest, IL, 1997.
- Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand–receptor binding free energy by 4D-QSAR analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1141–1150.
- Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal components analysis and partial least squares. *Tetrahedron Comput. Methodol.* **1989**, *2*, 349–354.
- Rogers, D. G.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- Friedman, J. *Multivariate Adaptive Regression Splines*; Technical Report No. 102; Laboratory for Computational Statistics, Department of Statistics, Stanford University: Stanford, CA, 1988 (revision 1990).
- Kleinbaum, D. G.; Kupper, L. L.; Muller, K. E.; Nizam, A. *Applied Regression Analysis and Other Multivariable Methods*, 3rd ed.; Books/Cole Publishing Company: Pacific Grove, CA, 1998; pp 330–332.
- SAS; Version 8.1 for Windows; SAS Institute Inc.: Pittsburgh, PA, 2001.
- Bichard, C. J.; Mitchell, E. P.; Wormald, M. R.; Watson, K. A.; Johnson, L. N.; Zographos, S. E.; Koutra, D. D.; Oikonomakos, N. G.; Fleet, G. W. J. Potent inhibition of glycogen phosphorylase by a spirohydantoin of glucopyranose: first pyranose analogues of hydantocidin. *Tetrahedron Lett.* **1995**, *36* (12), 2145–2148.