

Computational Prediction of Structure–Activity Relationships for the Binding of Aminocyclitols to β -Glucocerebrosidase

Lucía Díaz,^{†,‡} Jordi Bujons,[§] Antonio Delgado,^{†,‡} Hugo Gutiérrez-de-Terán,^{*||} and Johan Åqvist^{*,‡}

[†]Research Unit on Bioactive Molecules (RUBAM), Departamento de Química Biomédica

[§]Departamento de Química Biológica y Modelización Molecular

Consejo Superior de Investigaciones Científicas, (CSIC), Instituto de Química Avanzada de Catalunya (IQAC), Jordi Girona 18-26, 08034 Barcelona, Spain

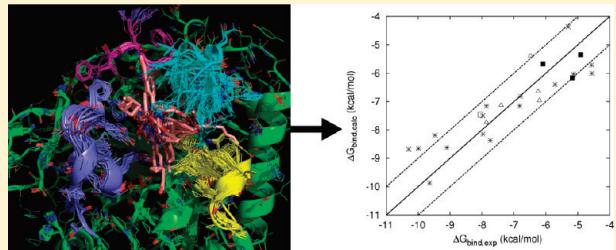
[‡]Facultad de Farmacia, Unidad de Química Farmacéutica (Asociada al CSIC), Universidad de Barcelona (UB), Avda. Joan XXIII s/n, 08028 Barcelona, Spain

^{||}Fundación Pública Galega de Medicina Xenómica- SERGAS, Complejo Hospitalario Universitario de Santiago, E-15706, Santiago de Compostela, Spain

^{*}Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, 751 24 Uppsala, Sweden

 Supporting Information

ABSTRACT: Glucocerebrosidase (GCCase, acid β -Glucosidase) hydrolyzes the sphingolipid glucosylceramide into glucose and ceramide. Mutations in this enzyme lead to a lipid metabolism disorder known as Gaucher disease. The design of competitive inhibitors of GCCase is a promising field of research for the design of pharmacological chaperones as new therapeutic agents. Using a series of recently reported molecules with experimental binding affinities for GCCase in the nanomolar to micromolar range, we here report an extensive theoretical analysis of their binding mode. On the basis of molecular docking, molecular dynamics, and binding free energy calculations using the linear interaction energy method (LIE), we provide details on the molecular interactions supporting ligand binding in the different families of compounds. The applicability of other computational approaches, such as the COMBINE methodology, is also investigated. The results show the robustness of the standard parametrization of the LIE method, which reproduces the experimental affinities with a mean unsigned error of 0.7 kcal/mol. Several structure–activity relationships are established using the computational models here provided, including the identification of hot spot residues in the binding site. The models derived are envisaged as important tools in ligand-design programs for GCCase inhibitors.



INTRODUCTION

The binding of small ligands to proteins is a well recognized mechanism to modulate the activity of a target macromolecule.¹ Structural information on the target macromolecule becomes essential in order to rationalize ligand–macromolecule interactions as well as to predict the binding mode for a particular type of compounds. In addition to the specific “on–off” stimulation or inhibition of the target protein by interaction with a small molecule, more sophisticated modes of action have also been recognized. This is the case with so-called pharmacological chaperones, small molecules designed to specifically bind a target protein in order to prevent its misfolding and eventual removal by the cellular proteasome in the course of the maturation process.² The development of pharmacological chaperones as potential drugs is a very attractive field of research for the biomedical community, since several diseases have been associated with abnormally low levels of certain proteins that show key mutations giving rise to misfolding and subsequent removal.³

Over the last years, we have been interested in the development of new aminocyclitols as pharmacological chaperones of the enzyme β -glucocerebrosidase (GCCase). A deficiency of this enzyme is associated with a lysosomal storage disorder known as Gaucher disease, due to the accumulation of glucosylceramide, the natural enzyme substrate.⁴ Several enzyme mutations have been characterized and considered to be responsible for a conformational destabilization that makes the protein more susceptible to misfolding, mistrafficking, and premature degradation.^{4,5} As a result, abnormally low levels of GCCase are found in lysosomes, where substrate accumulation takes place to trigger the observed symptoms of the disease. From a structural standpoint, pharmacological chaperones are expected to be found among reversible inhibitors of the target enzyme,² due to the ability of such inhibitors to stabilize the enzyme against thermal denaturation

Received: November 19, 2010

Published: March 08, 2011

upon binding.⁶ A collection of aminocyclitols, differing in the stereochemistry of the cyclitol scaffold⁷ and also in the functionalization of the amino group,⁸ has been reported from our laboratory. Results from this “first generation” of compounds showed that the best GCase inhibitors were found among the meso aminoinositol derivatives showing a scyllo stereochemistry and alkylamino or arylalkylamino functional groups. Interestingly, some of them were able to protect the enzyme against thermal denaturation,⁹ an indication of the compounds’ ability to behave as pharmacological chaperones *in vitro*.¹⁰ Moreover, these results were nicely corroborated in cellular assays carried out in human wild-type fibroblasts from Gaucher disease patients.¹¹ More recently, a “second generation” of GCase aminocyclitol inhibitors has been developed by exploring the chemical diversity of *N*-alkylaminocyclitols making use of the click chemistry approach based on the CuAAC Huisgen reaction. Interestingly, the presence of the triazole system thus unraveled the possibility of additional binding interactions with the target enzyme.¹²

Several of these “first and second generation” aminocyclitol derivatives showed an activity range as GCase inhibitors (IC_{50} values) spanning from 10^{-8} to 10^{-4} M.^{8,9,11–13} In addition, some of them were able to stabilize GCase against thermal denaturation as a model experiment to assess their potential as pharmacological chaperones.^{9,12} Moreover, promising results of cellular studies on mutant GCases have been observed for some of our aminocyclitols.¹¹ Preliminary docking studies were carried out to gain insight into the way these aminocyclitols interact with the GCase target.¹² In this paper we extend those studies to rationalize the experimental results with a computational approach aimed at disclosing the key ligand binding modes and structural requirements involved in the interaction of the aminocyclitols with the GCase active site. This structure-based analysis includes results from ligand docking, subsequent molecular dynamics (MD) simulations, and an estimation of binding affinities utilizing the linear interaction energy (LIE) method. The molecular mechanism of ligand binding for this class of compounds is revealed and presented as an invaluable tool to assist the design of GCase inhibitors.

METHODS

Molecular Docking. Docking calculations against two GCase–inhibitor complex structures (PDB codes 2NSX and 2V3E)^{14,15} were run as previously reported,¹² but in this case, a preliminary manual inspection and adjustment of the ionization state of the protein titratable residues were carried out. The program PROPKA 2.0^{16,17} was used to estimate the pK_a of these residues. The Schrödinger Suite 2008 package (Schrödinger, LLC, New York) through its graphical interface Maestro¹⁸ was used for all docking simulations. The structures of the proteins were prepared using the protein preparation wizard included in Maestro to remove the solvent molecules and ligands, adding hydrogens, setting protonation states, and minimizing the energy using the OPLS force field.¹⁹ Compounds of families 1–4 were built within Maestro and preoptimized before docking using the LigPrep application²⁰ included in the software. The program Glide^{21,22} was used for the docking calculations using the default XP precision settings²³ except for the following: (i) given the high flexibility of most of the compounds, a setting of 5 000 000 poses per ligand for the initial phase of docking and a scoring window of 500 for keeping initial poses was chosen; (ii) constraints were applied to maintain the cyclitol moiety of the ligands inside of the catalytic cavity of the GCase structures; (iii) up to 50 poses per ligand

were kept for the postdocking minimization. Glide XP scores were used to rank the resulting docked poses. Pose selection for inhibitors of families 1–3 was based on the possibility of a hydrogen bond between the triazole ring and the side chain of Gln 284 in the 2NSX structure or between Tyr 244 and the triazole ring in the 2V3E structure, as a consequence of the previously reported binding mode hypotheses.^{12,24} The top docking solutions in compliance with this requirement were used as starting ligand conformations for the subsequent MD simulations. For inhibitors of family 4, the best scored poses according to the Glide scoring function were selected instead. In addition to these docked compounds, the experimental pose of the cocrystallized ligand in the GCase structure with PDB code 2V3E, *N*-nonyldeoxyinosinic acid (NNNDNJ), was also considered for MD simulations.

Protocol for MD Simulations. MD simulations were done using the program Q²⁵ and the OPLS force field implemented therein.¹⁹ The parameters needed for the ligands that were not present in the original version of the force field were retrieved from the automatic parametrization performed with Macromodel²⁶ and translated into the syntax required by the program Q using a set of ad hoc scripts. The system was solvated with a 24 Å radius simulation sphere of TIP3P water²⁷ centered on the center of mass of the ligand. The water surface of this sphere was subjected to radial and polarization restraints²⁸ in order to mimic bulk water at the sphere boundary. Nonbonded interaction energies were calculated up to a 10 Å cutoff, except for the ligand atoms for which no cutoff was used. Beyond the cutoff, long-range electrostatics were treated with the local reaction field (LRF) multipole expansion method.²⁹ Protein atoms outside the simulation sphere were restrained to their initial positions and only interacted with the system through bonds, angles, and torsions. The ionization states of titratable residues inside the simulation sphere were manually assessed, taking into account that His 311 should be charged since it forms a salt bridge with Glu 235¹⁴ and the pK_a calculations mentioned in the previous section. This yielded an overall neutral system with the following residues charged: Asp127, Asp282, Asp283, Asp315, Asp399, Glu235, Glu340, Glu349, Glu388, Arg120, Arg285, Arg353, Arg359, Arg395, Arg463, Lys186, Lys346, and His311. On the contrary, Asp380 and Glu233 inside the simulation sphere have been considered in their neutral form as well as any other titratable residues closer than 3–5 Å to the boundary, which together with those outside the solvent sphere should be modeled as neutral because of the lack of dielectric screening.

For the ligand–protein simulations, a heating and equilibration procedure was applied before the data collection phase. The first phase, similar to energy minimization, consisted of 10 000 steps MD using a very short time step (0.1 fs) at 1 K temperature, coupled to a strong heat bath (1 fs bath coupling) with positional restraints of 25 kcal/mol·Å² on all heavy atoms. The system was then gradually heated up to 300 K during 50 ps, in which the bath coupling was relaxed until the final value of 100 fs, the time step was increased to 1 fs, and the positional restraints were gradually released. Unrestrained MD then followed for 2 ns, with energies collected at regular intervals of 25 fs. Energy averaging was performed on the energetically stable phase of this data collection period (average time for this collection period was 1.1 ns and the shortest collection period was 600 ps). Stability was addressed by an estimation of the convergence errors of the potential energies of the ligand with its surroundings (see Binding Free Energy Calculations Section).

MD trajectories were postprocessed and further analyzed with the available modules in Q (Qprep and Qcalc), including the generation of average energy values (ligand–surrounding energies, $\langle U_{l-s} \rangle$ and the ligand–residue interaction energies, $\langle U_{l-res} \rangle$) and geometrical analyses [root-mean-square deviation (rmsd), average structures, average interaction distances]. PyMOL (<http://www.pymol.org>) was used for visualization and images generation.

The MD sampling of the free ligand was done with an equivalent 24 Å TIP3P water sphere. The system was gradually heated with a 50 ps MD trajectory to 300 K, in which the heavy atoms of the ligand were restrained through a 10 kcal/mol·Å² force constant to their original position. MD followed for at least 2 ns under the same conditions as for the bound state but keeping the center of mass of the ligand restrained to the center of the sphere with a force constant of 10 kcal/mol·Å².

Binding Free Energy Calculations. Binding affinities were calculated using the LIE method, described in detail elsewhere.^{30,31} Basically, this approach estimates the ligand free energy of binding from the difference in the ligand–surrounding interaction energies in both its bound and free state. The relationship between the ligand intermolecular interaction energies and the free energy of binding is given by the equation:

$$\Delta G_{\text{bind}} = \alpha \Delta \langle U_{l-s}^{\text{vdW}} \rangle + \beta \Delta \langle U_{l-s}^{\text{el}} \rangle + \gamma \quad (1)$$

where $\langle U_{l-s}^{\text{vdW}} \rangle$ and $\langle U_{l-s}^{\text{el}} \rangle$ are, respectively, the Lennard-Jones and electrostatic interactions between the ligand and its surroundings (l-s). These interactions are evaluated as energy averages (denoted by the brackets) from separate MD simulations of the free (*f*) and bound (*b*) states of the ligand (solvated in water and bound to the solvated protein, respectively). The difference (Δ) between such averages for each type of potential is scaled by different coefficients (see ref 30), giving the polar and nonpolar contributions to the binding free energy. For the nonpolar contribution, this coefficient has been empirically set to $\alpha = 0.181$. The scaling factor for the polar contribution was initially derived from the linear response approximation ($\beta = 0.5$) but has subsequently been found, from free energy perturbation (FEP) calculations, to depend on the chemical nature of the ligand.³¹ According to that classification, ligands with at least two hydroxyl groups, like the ones handled in this study, are assigned a value of $\beta = 0.33$, from now on termed model A. More recently, Almlöf et al.³² proposed a refined model based on solvation free energies of chemical moieties calculated with the FEP method. According to this model, our ligands have β values ranking from 0.37 to 0.41 (termed model B). Two additional parametrizations of the LIE scaling parameters were tested in the present work, in order to gauge the robustness of the model, where model C is a free parametrization of the β parameter, while model D is a completely free parametrization of both α and β . In all the LIE models, γ is a constant term obtained by regression fitting that fixes the scale for absolute binding free energies. The nature of this parameter has been related to several descriptors of the binding site, such as the hydrophobic nature of the binding site.³³ Finally, the associated error to the LIE calculated free energies is estimated by combining the convergence errors associated to each simulation into a LIE-like equation but adding all the values since the error is additive:

$$\text{Error}_{\text{bind}} = \alpha[(\langle E_{l-s}^{\text{vdW}} \rangle_b + \langle E_{l-s}^{\text{vdW}} \rangle_f)] + \beta[(\langle E_{l-s}^{\text{el}} \rangle_b + \langle E_{l-s}^{\text{el}} \rangle_f)] \quad (2)$$

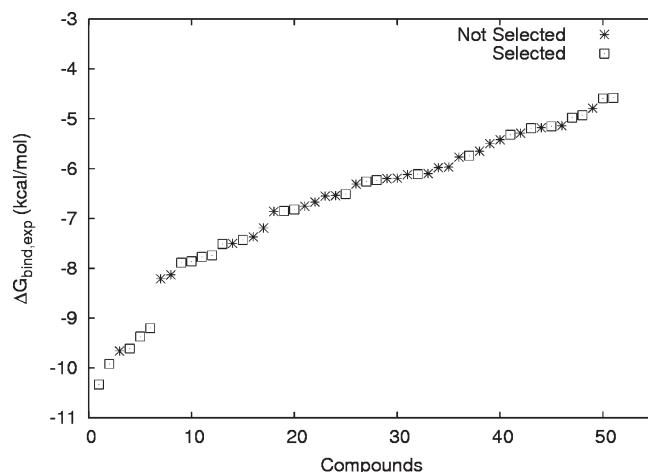


Figure 1. Activity range of the 51 compounds initially considered for this study. Squares indicate selected compounds for the computational exploration described in this work.

Where each term $\langle U_{l-s}^{\text{type}} \rangle$ (being type = vdW or el) is estimated as the difference between the total average of the given l-s potential energy and a subaverage of such energy term, calculated along any of the two halves of the simulation time.

Binding free energy estimates were compared to experimentally measured IC₅₀ values, which were converted into experimental free energy values through the equation:

$$\Delta G_{\text{bind, exp}}^0 = RT \ln IC_{50} + c \quad (3)$$

Where $c = -RT \ln(1 + [S]/K_M)$, i.e., a constant term.³⁴ Such a term has been shown to be close to zero in the assays performed in these set of compounds,^{12,13,24} but in any case, it does not affect the relative free energies and should be included in the optimized value of γ in eq 1.

COMBINE Analysis. The contribution of each residue to the interaction energy with the ligand was treated in a multivariate analysis, in a similar way as the COMparative BINding Energy method (COMBINE) of Ortiz et al.³⁵ The input matrix consisted of the interaction energies of the ligand with each residue, $\langle U_{l-res} \rangle_b$ plus the values of the interaction energies of the ligand with the solvent in the two states, bound $\langle U_{l-water} \rangle_b$ and free $\langle U_{l-water} \rangle_f$. The electrostatic and nonelectrostatic energy contributions were treated independently, thus resulting in an initial matrix of 998 descriptors [(497 residues + 2 solvent states) × 2 types of energy]. The statistical analysis was done with the Simca package.³⁶ A filtering process was applied in which all descriptors with only one value different from the mean value were removed, thus automatically eliminating any residue far from the ligand (mean value = 0). The scaling of the variables was based on the central values, and block scaling was performed in order to mix van der Waals and electrostatic terms, which have different average values.

■ RESULTS

Selection of Compounds. Some of the most significant aminocyclitols from our laboratory were grouped according to their activity range as GCase inhibitors. This generated a first set of 51 compounds, with IC₅₀ values spanning from 10⁻⁸ to 10⁻⁴ M (Figure 1). Among them, a subset representing around 50% of the total population (25 compounds), which covered the whole activity range, was selected for this work (see also Table 1).

Table 1. Experimental IC₅₀ (μ M) and Corresponding ΔG_{bind} (kcal/mol) Values of the Compounds Represented in Figure 2

entry	compound	IC ₅₀ ^a	ΔG_{bind}^b
1	1a	0.05 (0.06)	-9.99
2	1b	1.50 (1.80)	-7.97
3	1c	466 (nd)	-4.56
4	1d	10.3 (nd)	-6.82
5	1e	1.50 (0.30)	-7.97
6	1f	453 (nd)	-4.57
7	1g	1.80 (1.30)	-7.86
8	1h	134 (nd)	-5.30
9	1i	66.5 (nd)	-5.71
10	1j	0.12 (0.10)	-9.47
11	1k	0.09 (0.10)	-9.64
12	1l	2.20 (1.90)	-7.74
13	1m	10.8 (25.0)	-6.79
14	1n	178 (nd)	-5.13
15	1o	0.20 (0.20)	-9.10
16	1p	0.03 (0.06)	-10.29
17	1q	236.9 (nd)	-4.96
18	2a	167 (nd)	-5.17
19	3a	35.4 (18.2)	-6.09
20	3b	257 (44.6)	-4.91
21	4a	29.1 (nd)	-6.20
22	4b	18.2 (9.8)	-6.48
23	4c	3.9 (0.9)	-7.40
24	4d	1.80 (0.2)	-7.86
25	4e	27.6 (nd)	-6.24
26	NNDNJ	1.30 (0.30)	-8.05

^a Determined at pH = 5.2 or 7.4 (in parentheses); (nd) not determined.

^b The values have been calculated according to eq 3 using the IC₅₀ values determined at pH = 5.2.

Compounds 4a–e were selected as representative examples of the “first generation” approach, whereas compounds 1a–q, 2a, 3a and b are triazole containing “second generation” aminocyclitols (see Figure 2 for structures). Additionally, NNDNJ, a well recognized GCcase pharmacological chaperone,¹⁰ was also selected for this study since its binding mode has been revealed by X-ray crystallography.¹⁴

Inhibitors 1a–q only differ in the nature of the N-substitution at the triazolylalkyl side chain. Their activity as GCcase inhibitors and also as GCcase thermal stabilizers is higher than that shown by inhibitors 2a and 3a and b, with a longer carbon linker between the aminocyclitol core and the triazole moiety (Figure 2). Since variations in the placement of this ring along the N-alkyl side chain lead to noticeable differences in activity (compare 1o with 2a and 1k with 3a), a key role of the triazole moiety in the interaction with the target protein can be envisaged. Moreover, compounds 4a–e, retaining the aminocyclitol core but lacking the triazole ring, are weaker inhibitors. The experimental IC₅₀ values of the studied inhibitors, together with their corresponding binding free energies (ΔG_{bind}) are shown in Table 1.

Docking and Scoring. We recently reported the docking of the compound series 1–3.^{12,24} Two potential binding modes were identified, which differed in the arrangement of the substituted triazole while sharing a similar arrangement of the aminocyclitol moiety, deep inside the catalytic center of GCcase. In this work, we repeated such docking experiments, using the

same two structures for the protein (PDB codes 2NSX and 2V3E)^{14,15} but adjusting this time the ionization state of titratable residues in the binding site according to the MD simulations setup (see Methods). Concerning the docking poses, the results were mostly similar to the previous report.^{12,24} However, as shown in Figure 3, the correlation of the docking scores with experimental affinities was poor (2NSX target) or completely absent (2V3E target). Consequently, the use of MD and binding free energy calculations with LIE methodology was envisaged as a better approach to explain the binding affinity and structure–activity relationships of these GCcase inhibitors.

Concerning the “first-generation” GCcase inhibitors (compounds 4a–e), not previously studied by docking methods, Figure 4 shows how the best docking pose is in agreement with the experimental binding mode of the cocrystallized ligand NNDNJ, showing similar ligand arrangements to those previously found for the triazole containing analogs.

MD Simulations and Computational Analysis of Binding Affinities. All the compounds in the present series bear a secondary alkylamine which, depending on the environment and pH, could be considered in its charged or its neutral form. Additionally, two different conformations of the protein^{14,15} were initially considered, as already stated. Thus, our first concern was to elucidate which protonation state should be considered for the inhibitors in the binding site and which of the two protein crystal structures was more suitable for the MD simulations. The four possible combinations (i.e., the protonated or neutral state of inhibitors and the two available protein structures) were evaluated by running preliminary MD simulations on a subset of 14 compounds. The estimated binding affinities, according to the standard LIE model, are collected in Table S1 of the Supporting Information, which clearly shows that the best correlation was obtained using the 2V3E protein structure, considering the neutral form of the ligands. This last aspect was additionally assessed from different perspectives: First, we carried out standard empirical pK_a predictions in solution for the whole compound series, the results not being conclusive. While compounds with only one carbon linker between the amine and the triazole were predicted to be in their neutral form at physiological pH, the compounds where the linker is a longer alkyl chain were indicated to be in the charged form (data not shown). Second, since the reported experimental IC₅₀ values were measured at pH 5.2, we report in Table 1 additional IC₅₀ values at pH 7.4 for a subset of compounds, in order to get a more comprehensive view of the pK_a of the ligands in the binding site. This analysis shows that, whereas many compounds are unaffected, a slight increase in the affinity at neutral pH for some compounds (i.e., 1e, 3, and 4 subseries and NNDNJ) is observed. Finally, a closer look at the NNDNJ–GCcase crystal structure reveals the possibility of a neutral alkyl amine in the ligand not interacting with any acidic side chain in the protein. All these analyses support the results collected in Table S1, Supporting Information, suggesting that the ligands should be considered in their neutral form for the MD simulations, using 2V3E as the protein structure.

The 25 ligand–GCcase complexes, defined by the best molecular docking pose with 2V3E in each case (with the only exception of NNDNJ for which the experimental binding mode was used as starting point), were simulated for 2 ns of unrestrained MD. An additional 2 ns MD run in water was performed, in order to estimate binding affinities with the LIE methodology, thus resulting in a total of 100 ns MD simulation time for the

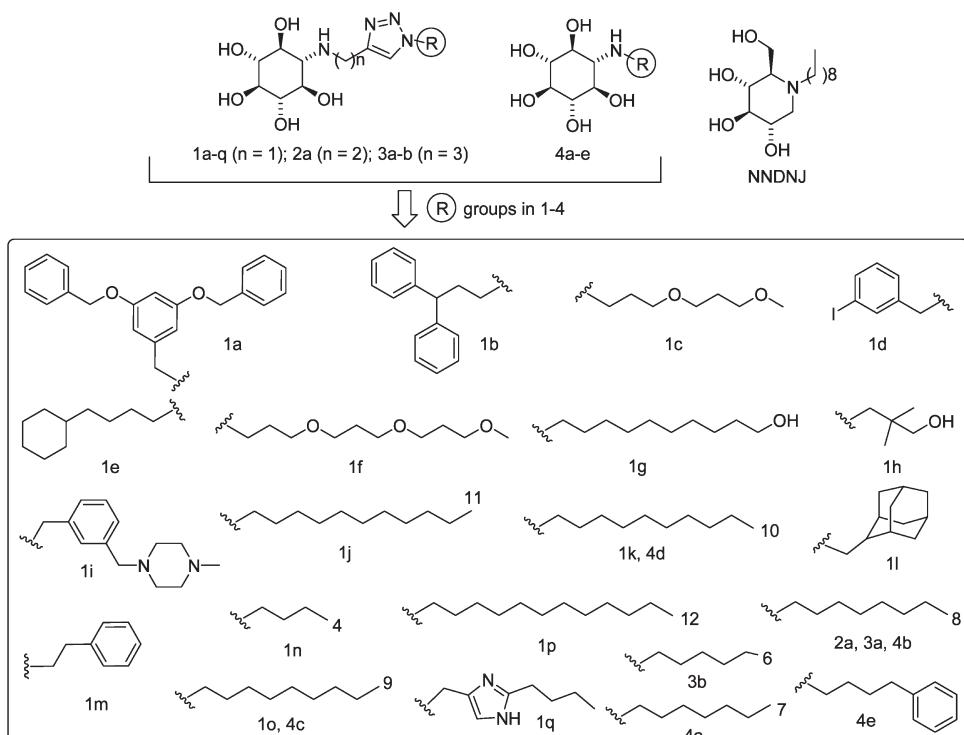


Figure 2. Structures of the GCase inhibitors used in this study.

whole series. The first remarkable observation from the dynamic exploration of the docking complexes is that some rearrangement of the binding site happens during the first part of the simulation. In fact, this is not surprising, since MD equilibration of rigid-protein docking has been largely recognized to be a useful tool in drug design.³⁷ However, the magnitude of these movements as well as the identification of the residues that show higher fluctuations can be quite informative. In the superposition of the 25 average structures, the highest variance among the different complexes is observed in the loop regions (Figure 5).

In particular, loop 1 is clearly the most flexible region of the binding site. This observation is in good agreement with the high B factors reported in the parent crystal structure.¹⁴ This region contains a flexible lysine at position 346, which acts as a gatekeeper to enable the access of the ligand, as observed by comparison of the crystal structures 2NSX and 2V3E.^{14,15} Moreover, it has already been noted that this loop might suffer from crystal packing effects, so a higher mobility during MD simulations might be expected.¹⁴ A detailed inspection of the ligands shows that the position of the cyclitol ring is very stable in the 25 complexes. This is due to a number of tight hydrogen bonds with the binding site residues Asp 127, Trp 179, Asn 234, Glu 340, and Trp 381. On the contrary, high fluctuations are observed for the flexible substituents in the secondary amine, which bear up to 14 torsional degrees of freedom. Although there is not a clear correlation between the mobility of the ligand and the activity, the highest fluctuations along the MD simulation are associated with compounds having the lowest affinity values (Figure S2, Supporting Information). The compounds identified in this group mainly belong to two particular chemical types: compounds with a polyether alkyl chain (**1c** and **1f**) and compounds with the longest linker between the triazole and the cyclitol rings (**2a**, **3a**, and **3b**). The exception is compound **1p**, the most active in the series, which registers an overaverage high rmsd due to the mobility

of its particularly long alkyl chain (12 carbons length). Furthermore, compound **1q**, one of the least active in the series, had to be discarded for further free energy estimations because the simulations did not converge even after extending them up to 10 ns.

The binding affinities calculated from the MD trajectories with the LIE methodology are in excellent agreement with the experimental data. Table 2 shows the average ligand-surrounding energies for the data collection period of the two simulations (bound and free). In order to gauge the robustness of the LIE model, we also investigated the possible dependence of the estimated binding affinities on the parameters of the LIE equation. The correlations between experimental and calculated values for different versions of the LIE equation, together with the statistical figures of merit, are shown in Table 3.

The standard LIE parametrization (model A) assigns values of $\alpha = 0.18$ and $\beta = 0.33$, since all the ligands bear more than two hydroxyl groups, thus expecting highest deviations from the linear response approximation.³¹ This model shows an excellent performance with a mean unassigned error of 0.69 kcal/mol and a good fit between experimental and calculated values ($R^2 = 0.77$). We then checked the possible improvement of the last version of the LIE method (derived for solvation free energies) in the present system (model B).³² In that version of the LIE method, a partial β value is assigned to every chemical group present in the ligand, and the global β value, which is initially assumed to be $\beta_0 = 0.43$, is then modulated by a weighted sum of the independent contributions of all the chemical groups present in the given ligand.³² Even if the results show the validity of these parameters, with statistical figures comparable to the standard LIE model, the additional complexity of this model does not improve the performance for the present system. We should note that the two above models have no free parameters affecting the relative free energies, which makes the agreement with experimental values quite remarkable. When the value of the electrostatic

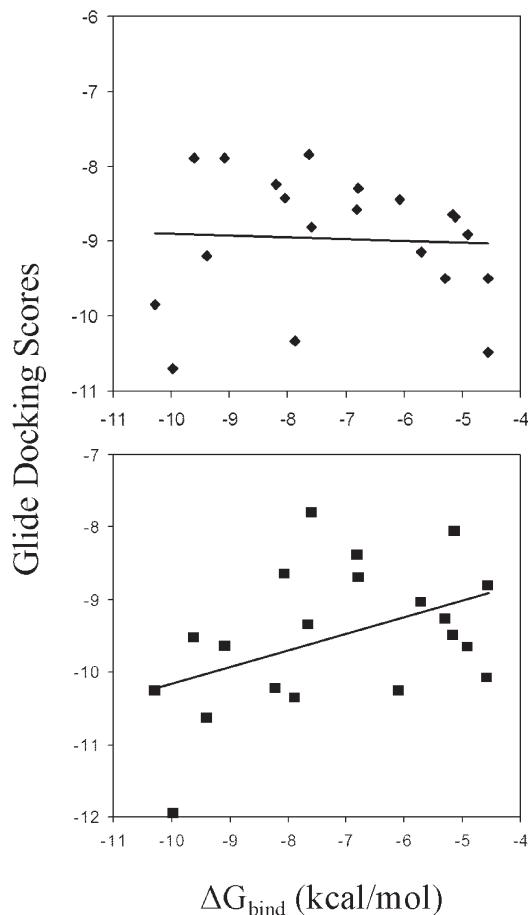


Figure 3. Glide scores for selected compounds obtained from the docking experiments against GCase targets 2V3E (♦) and 2NSX (■) vs experimental ΔG_{bind} values from Table 1.

contribution to the binding affinity is split into separate weighting factors for the water (β_w) and protein (β_p) states (model C), the optimized values remarkably both become $\beta_w = \beta_p = 0.37$, which is quite close to models A and B and indicates that protein prereorganization effects on ligand binding are minor.³⁸ In models A–C, the offset parameter γ shows a positive value close to 4 kcal/mol. The value of this offset parameter has been shown to be associated to the hydrophobicity of the binding site,³² in the sense that the more hydrophobic is the binding site, the more negative is the value of γ . The positive value obtained in our parametrizations is in agreement with these observations, given the high polarity observed for the binding site of GCase, with several charged amino acids present. A completely free parametrization of the LIE equation is also reported (model D). The slight increase for the weighting factors corresponding to the nonpolar (α) and polar (β) interaction energies is counterbalanced by a concomitant increase in the value of the offset γ parameter. This model primarily demonstrates the robustness of the earlier determined parameter values, while the slight improvement in statistical figures of merit is mainly an effect of adding more parameters. We can conclude that the standard LIE model not only performs extremely well in the present case, but most importantly, the model is robust to a free parametrization. Thus, model A will be the one selected for further discussion. The good agreement between calculated and observed affinities using this model can be seen in the scatter plot shown in Figure 6.

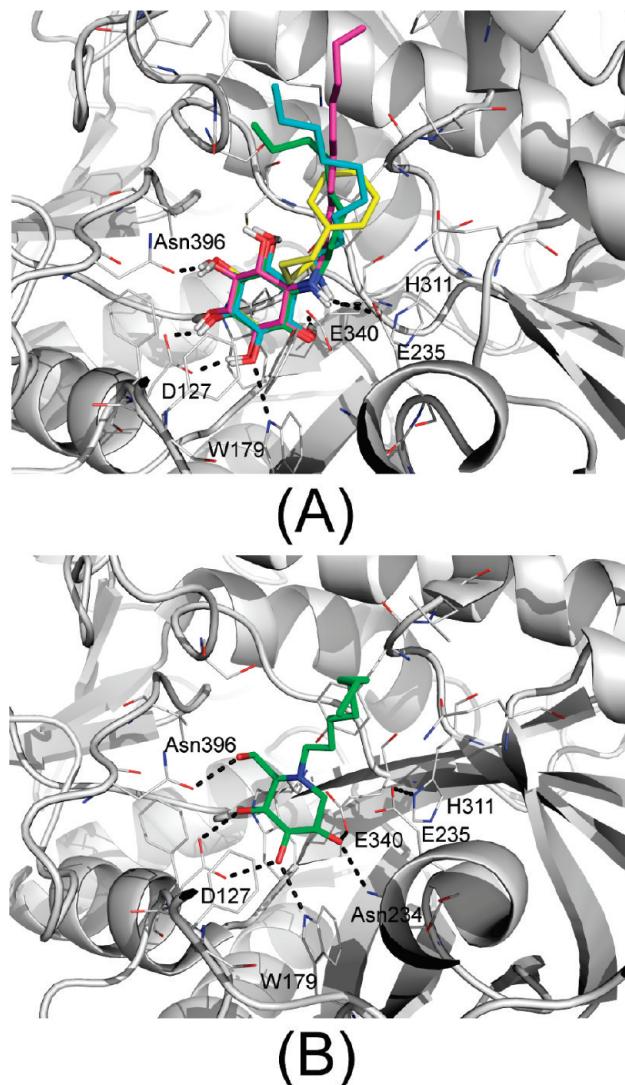


Figure 4. (A) Best poses obtained for compounds 4a–e docked against the GCase structure 2V3E. (B) The crystallographically determined structure of GCase-bound NNDNJ is also shown for comparison. Hydrogen bonds are depicted with dashed lines.

An advantage of the LIE method is not only its better performance for the calculation of ligand affinities, as compared to faster methods like empirical scoring functions, but also that it allows for a physical interpretation of binding contributions. This is important in order to establish structure–activity relationships and assist in further ligand design. As a first approximation, we were interested in the identification of the major forces governing the binding process, i.e., electrostatic or nonelectrostatic interactions. Not surprisingly, in the present system the magnitude of the electrostatic contribution to binding is approximately three times stronger than the nonpolar contribution (see Table 2). This is in agreement with the high polarity of the binding site, which compensates for the desolvation of the amino cyclitol moiety, as has been previously observed in other systems like in malarial aspartic proteases.³⁹ What is more surprising is the fact that the electrostatic energies also play a main role in the modulation of the binding affinities within the series. In fact, there is an important degree of correlation between the electrostatic interaction energy differences ($\Delta U_{\text{I-S}}^{\text{el}}$) and the experimental free energy values

Table 2. Average Ligand-Surrounding Energies (kcal/mol) for the Studied Inhibitors

inhibitor	$\langle U_{l-s}^{vdW} \rangle_b$	$\langle U_{l-s}^{vdW} \rangle_f$	$\langle U_{l-s}^{el} \rangle_b$	$\langle U_{l-s}^{el} \rangle_f$	$\Delta G_{bind,exp}$	$\Delta G_{bind,calc}^a$
1a	-63.89 ± 0.31	-40.60 ± 0.81	-107.13 ± 0.42	-83.29 ± 0.97	-9.99	-8.66 ± 0.66
1b	-53.75 ± 0.58	-31.57 ± 0.27	-101.91 ± 0.42	-80.99 ± 1.19	-7.97	-7.50 ± 0.68
1c	-44.73 ± 0.59	-26.22 ± 0.58	-99.14 ± 1.16	-80.73 ± 0.38	-4.56	-6.01 ± 0.71
1d	-43.85 ± 0.17	-26.29 ± 0.21	-95.72 ± 0.01	-73.62 ± 0.82	-6.82	-7.05 ± 0.34
1e	-41.91 ± 0.13	-28.93 ± 0.34	-102.06 ± 0.39	-74.14 ± 0.40	-7.97	-8.15 ± 0.34
1f	-50.02 ± 0.27	-31.82 ± 0.39	-106.89 ± 1.26	-89.20 ± 0.02	-4.57	-5.71 ± 0.54
1g	-42.37 ± 0.55	-28.74 ± 0.47	-115.33 ± 1.51	-90.78 ± 0.25	-7.86	-7.15 ± 0.76
1h	-36.68 ± 0.89	-19.97 ± 0.10	-99.68 ± 0.49	-85.28 ± 0.92	-5.30	-4.36 ± 0.64
1i	-50.93 ± 1.05	-31.68 ± 1.89	-113.29 ± 1.24	-94.10 ± 0.26	-5.71	-6.40 ± 0.84
1j	-48.01 ± 0.35	-32.73 ± 0.20	-97.44 ± 1.20	-70.65 ± 0.54	-9.47	-8.19 ± 0.67
1k	-49.44 ± 0.46	-30.78 ± 0.06	-102.72 ± 1.45	-72.67 ± 0.34	-9.64	-9.87 ± 0.68
1l	-38.92 ± 0.59	-28.49 ± 0.18	-101.7 ± 0.97	-71.71 ± 0.25	-7.74	-8.37 ± 0.54
1m	-43.01 ± 0.31	-26.36 ± 0.03	-91.88 ± 0.06	-70.01 ± 0.05	-6.79	-6.81 ± 0.22
1n	-37.66 ± 0.73	-20.74 ± 0.04	-92.37 ± 0.19	-73.03 ± 0.46	-5.13	-6.03 ± 0.10
1o	-47.6 ± 0.45	-28.83 ± 0.06	-101.03 ± 0.90	-74.83 ± 0.58	-9.09	-8.62 ± 0.35
1p	-53.47 ± 0.60	-34.71 ± 0.06	-100.59 ± 0.50	-74.20 ± 0.07	-10.29	-8.68 ± 0.58
2a	-45.07 ± 0.50	-28.67 ± 0.43	-93.16 ± 0.85	-73.12 ± 0.63	-5.17	-6.16 ± 0.13
3a	-50.94 ± 0.01	-29.75 ± 0.07	-90.09 ± 0.33	-74.15 ± 0.03	-6.09	-5.67 ± 0.31
3b	-0.15 ± 0.10	-27.02 ± 0.13	-94.39 ± 0.22	-75.05 ± 0.33	-4.91	-5.34 ± 0.65
4a	-34.14 ± 0.24	-17.64 ± 0.05	-85.89 ± 0.08	-63.49 ± 0.03	-6.20	-6.96 ± 0.71
4b	-34.96 ± 0.17	-19.28 ± 0.07	-81.65 ± 0.33	-63.56 ± 0.23	-6.48	-5.39 ± 0.09
4c	-38.09 ± 0.99	-21.41 ± 0.25	-84.93 ± 0.89	-61.14 ± 0.59	-7.40	-7.13 ± 0.22
4d	-41.2 ± 0.00	-22.82 ± 0.03	-86.81 ± 0.62	-63.10 ± 0.46	-7.86	-7.73 ± 0.37
4e	-37.5 ± 0.87	-20.52 ± 0.09	-88.29 ± 0.06	-67.11 ± 0.55	-6.24	-6.64 ± 0.36
NNDNJ	-36.13 ± 0.48	-21.4 ± 0.06	-77.69 ± 0.42	-52.80 ± 0.34	-8.05	-7.46 ± 0.35

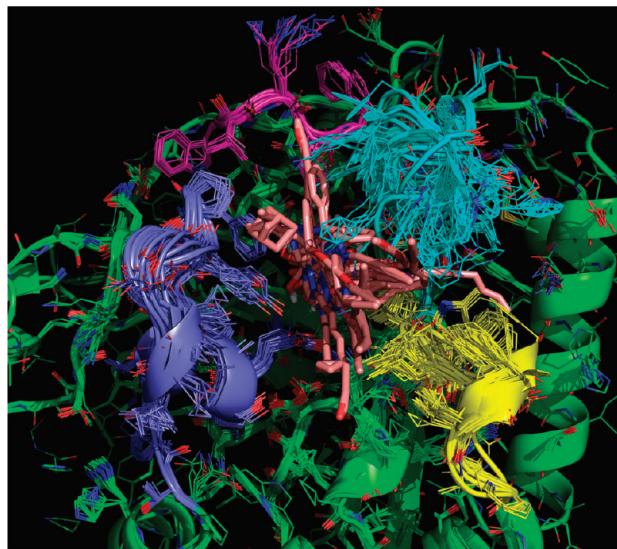
^a Model A (see Table 3).

Figure 5. Superposition of the 25 average structures over the MD trajectories. Ligands are colored in brown, while the protein carbon atoms are represented in green. This color has been changed in order to highlight the four loops of the binding site, according to the following scheme: cyan for loop 1; magenta for loop 2; yellow for loop 3; and violet for loop 4. It can be appreciated that the highest fluctuations among the complexes occur in loop 1.

which is lost if only the nonpolar contribution (ΔU_{l-s}^{vdW}) is considered (Figure 7). This behavior usually indicates a high degree of specificity, mainly due to the directionality of polar interactions in

the predicted binding mode of the aminocyclitol derivatives. In order to test the significance of this observation, a fifth parametrization of the LIE equation (model E in Table 3) is presented, where the nonelectrostatic contribution is treated as constant (i.e., $\alpha = 0$). Notably, this model does not deteriorate much compared to the standard model, with the total correlation (R^2) only decreasing by 14% and the mean unassigned error still below the 1 kcal/mol threshold. The offset parameter γ turns to approximately zero, while the value of the β parameter is very close to the theoretical value depicted in model A. This shows a high degree of correlation between the offset parameter γ and the nonpolar contribution to the ligand binding, as previously suggested by us³¹ and other authors.⁴⁰

To further identify the interactions responsible for both the overall strength and the modulation of ligand binding, we present in Figure 8 two plots of the average interaction energies (polar and nonpolar) of the ligands as a function of the protein (pseudo)sequence. Only those residues that register significant average interaction energies (i.e., $|\langle U_{l-s} \rangle| > 1$ kcal/mol) are considered. The “error bars” indicate the variance of the interaction energy for the given residue among the different ligands. The positions showing highest average interaction energies identify common key residues to the ligand binding within the series, while residues with the highest variance are expected to be responsible of the modulation of ligand affinities. The validity of this approach has been recently shown in the study of the binding process of a series of 43 non-nucleosidic HIV reverse transcriptase inhibitors,⁴¹ where a good agreement was observed between the predictions and the experimental data from mutation studies. In the present case of GCase inhibition, the residues forming strong hydrogen bonds with the hydroxyl groups of the cyclitol

Table 3. Statistical Figures of Merit for the Five LIE Models Discussed in the Text

model ^a	α	β	γ	$< \text{error} >$	rms	R^2
A ^b	0.18	0.33	+3.4	0.69	0.82	0.77
B ^c	0.18	0.37/0.38/0.39/0.40	+4.5	0.73	0.84	0.76
C ^d	0.18	0.37	+4.3	0.69	0.80	0.78
D ^e	0.26	0.39	+6.1	0.63	0.77	0.79
E ^f	0	0.33	0.3	0.84	1.04	0.63

^a The values of the optimized scaling factors are shown in bold.

^b Standard LIE model described in ref 31. ^c β values calculated as described in ref 32. ^d β and γ parameters are optimized. ^e Complete optimization of the three parameters (α , β , γ) in the LIE equation. ^f Optimized values of the β and γ parameters when the effect of the nonelectrostatic contribution is treated as constant ($\alpha = 0$).

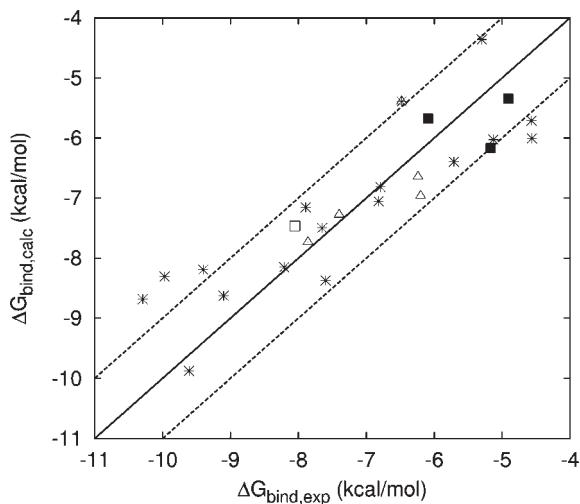


Figure 6. Scatter plot for calculated versus experimental free energies of binding for the 25 inhibitors, using the standard LIE parametrization (model A in Table 3). The solid line represents perfect agreement between $\Delta G_{\text{bind},\text{calc}}$ and $\Delta G_{\text{bind},\text{exp}}$. Dotted lines denote an interval of ± 1 kcal/mol. The different chemical families of compounds are represented using stars (triazole containing compounds 1a–p, where number of carbon linkers $n = 1$), closed squares (triazole-containing compounds where $n \geq 2$, 2a and 3a and 3b), triangles (compounds lacking triazole ring, 4a–e), and open square (compound NNDNJ, from the crystal structure 2V3E).

moiety are responsible for most of the electrostatic interaction energy (residues Asp 127, Trp 179, Asn 234, Glu 340, and Trp 381), with the strongest interactions associated with Asp 127 and Glu 340, both being charged in our simulations (Figure 8). All these residues were already highlighted in the structural analysis of the complex with NNDNJ, including Asn 234 which apparently can replace the role of Asn 396, as observed in the binding of glycerol.¹⁵ Two additional residues show high variance in their electrostatic contributions to ΔG_{bind} , namely Glu 235, which is close to the exocyclic amino group of the cyclitol moiety and Lys 346 interacting with the triazole moiety in the most active compounds (series 1, $n = 1$, see Figure 2). It is worth mentioning that the interactions established by the cyclitol moiety with the residues of the active site were already detected in the initial docking poses, which demonstrates the low mobility observed in this region during the MD phase. Conversely, in the regions outside of the GCase active site induced fit effects are observed, and while some electrostatic interactions seem to lose relevance

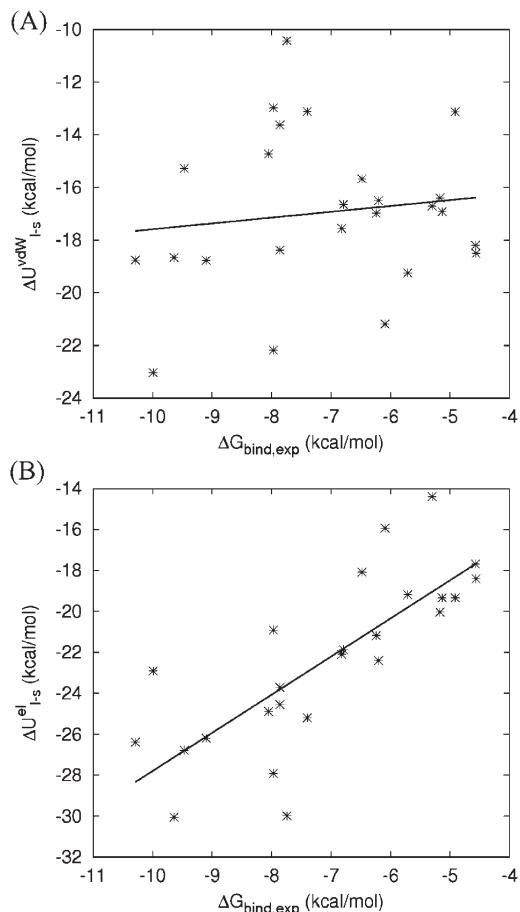


Figure 7. Scatter plot for the nonpolar (A) and electrostatic (B) components of binding energy versus experimental free energies of binding for the 25 inhibitors. The solid lines represent the best fit between $\Delta U_{\text{vdW}}^{\text{el}}$ and $\Delta G_{\text{bind},\text{exp}}$ ($R^2 = 0.02$) and $\Delta U_{\text{el}}^{\text{el}}$ and $\Delta G_{\text{bind},\text{exp}}$ ($R^2 = 0.63$).

(i.e., Tyr244, now just interacting through nonpolar forces), new interactions not detected in the previous docking runs (ie. with residue Lys 346) are now observed. All of the above suggest that the aforementioned positions showing strong average interactions with the ligand series should be considered in the ligand design process. A further look into the role of these residues in the ligand binding is provided in Figure 9.

The identification of the residues that most contribute to the modulation of the activities is, however, somewhat ambiguous. Thus, we were interested in testing the identification of the key residues by means of a completely automated and unbiased procedure. Inspired by the ideas behind the COMBINE approach,³⁵ we performed a partial least-squares (PLS) analysis on the electrostatic and van der Waals ligand interaction energies with each residue $\langle U_{\text{l-res}} \rangle$ and with the solvent $\langle U_{\text{l-water}} \rangle$. The results indicate that, although the statistical figures of merit using this kind of COMBINE approach are not as good as the results obtained with the LIE method ($R^2 = 0.67$, $Q^2 = 0.41$), the structural interpretation of the results is quite comparable. It is particularly interesting that completely neglecting the nonpolar interaction energies does not have a negative effect in the multivariate analysis; on the contrary, the correlation is even increased to a value of $R^2 = 0.72$, while predictivity only decreases slightly ($Q^2 = 0.36$). The identification of the variables that contribute the most to the PLS coefficients is a very convenient

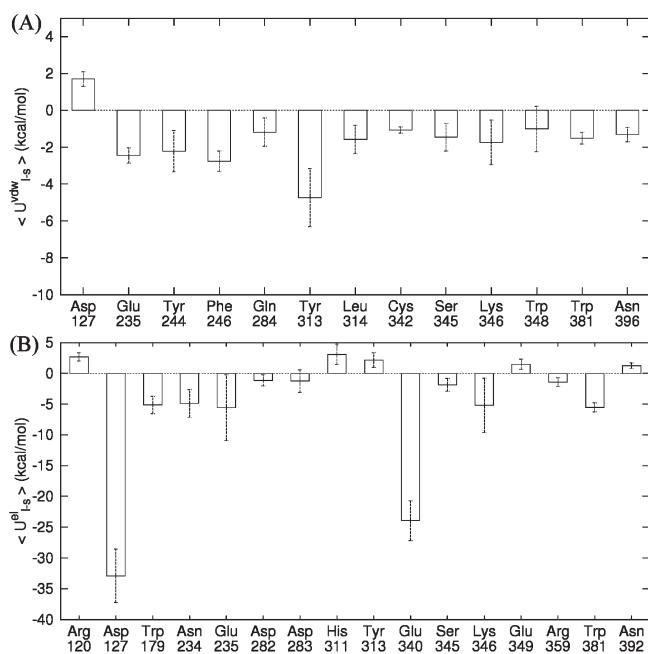


Figure 8. Average van der Waals (A) and electrostatic (B) ligand–residue interaction energies, $\langle U_{1-\text{res}} \rangle$, of the total 25 inhibitors. Only those residues that contribute significantly are shown (cutoff value ± 1 kcal/mol). The bars represent the average interaction energies, and the error bars represent the standard deviation.

way to identify the residues important for the modulation of ligand affinities within the series. Moreover, the statistical significance of this selection can be assessed with the confidence interval associated to each variable, provided by the PLS coefficients plot in Simca (data not shown). The list of the residues selected with this procedure is Asp 127, Glu 235, Asp 283, and Lys 346, in reasonable agreement with the above observations extracted from Figure 8, although with more reductionism in the model interpretation; some residues contributing importantly to the electrostatic interaction, such as Glu340, are missing, together with all hydrophobic interactions. Finally, it is worth mentioning that the inclusion of the desolvation term in the multivariate analysis, by means of the variables $\langle U_{1-\text{water}}^{\text{el}} \rangle_p$ and $\langle U_{1-\text{water}}^{\text{el}} \rangle_w$, is critical to obtain a decent correlation with the experimental values, since removing these descriptors reduces dramatically the model performance ($R^2 = 0.31$, $Q^2 = 0.22$). This justifies the need of considering the reference state of the ligand in the calculation of binding affinities, in contrast to the original COMBINE approach and other statistical methods like scoring functions.

■ DISCUSSION

This work follows up a series of recent reports from our laboratory concerning the development of a new structural class of aminocyclitol GCase inhibitors^{8,9,12,13} as promising pharmaceutical chaperones.¹¹ In this report, a detailed comprehensive analysis of the structural requirements for ligand binding is undertaken through the use of systematic MD sampling of the molecular complexes and calculations of ligand affinities with the LIE method. In particular, we were interested in understanding the role of a triazole linker between the aminocyclitol and the hydrophobic substituent (e.g., compound 4 vs compounds 1–3) and the importance of the spacer between the amino group and

the triazole ring, a key question that remained unsolved in previous work.¹² As for compounds lacking a triazole linker, our model reproduces reasonably well the observed dependence between the hydrophobicity of the alkyl chain and the GCase affinity of the compounds,¹² suggesting that the alkyl chain is the major modulator of affinity in this particular class of compounds. However, for compounds bearing a triazole ring, the length of the linker between this ring and the aminocyclitol core plays a significant role for ligand affinity. Thus, compounds with a two or three-carbon spacer (2a, 3a, and 3b, squares in Figure 6) have low activity as GCase binders. The results obtained here provide a rationale for these experimental data. The longer spacers in these compounds preclude those interactions of the triazole with the protein which are consistently observed for the compounds with only one carbon spacer, in particular with Lys 346 in loop 1 and the groove defined by Tyr 313 in loop 3. These two residues have been identified among those that most contribute to the modulation of affinities in the whole series (see Figure 8). Additionally, the compounds bearing two or three carbon spacers display higher rmsd fluctuations along the MD trajectory. In contrast, a one carbon spacer seems necessary, although not sufficient, for an optimal fitting of the triazole linker. As noted above, ligands with a polyether substitution (1c and 1f) also show high fluctuations along the MD trajectory, which result in weaker interactions of the ligand with the binding site. In these cases, the additional polarity and conformational constraints provided by the ether oxygen atoms are probably responsible for this loss of interactions.

This study has revealed an important role for the triazole linker, mainly interacting with residue Lys 346. Nevertheless, the nature of the alkyl substituent modulates this interaction, since it indirectly affects the attachment of the ligand in the binding site. In particular, the mobility of the ligand and the estimated strength of the electrostatic interactions of the triazole and aminocyclitol moieties, with the polar residues in the binding site, are dependent on the nature of the alkyl substitution in the ligand.

An exploration by MD simulations of 25 ligand–GCase complexes was performed in a systematic, semiautomated way that demonstrated the applicability of LIE binding free energy calculations based on MD sampling in the drug design process.^{37,42} The present series, consisting of 25 very flexible compounds, constitutes a realistic example of the application of this technique in a real drug design project, for which an impressive correlation with experimental results is found. This allows for establishing structure–activity relationships in terms of interactions with the protein, which complements the conclusions derived from our separate quantitative structure–activity relationship (QSAR) (L. Díaz, personal communication) and docking studies.¹² Correlation of docking scores in our previous work (compounds 1a–h in ref 12) could have been strongly influenced by correlation between cLog *P* and $-\text{Log IC}_{50}$. Following this argumentation, the only effect that was really predicted by the scoring function was the size dependence of the alkyl chain with the activity, thus the role of the triazole linker could not be established. The strong dependence of the scoring functions on the size of the inhibitors was previously observed in a detailed computational analysis of the non-nucleoside reverse transcriptase inhibitors family of HIV-RT.⁴³ In that study the limited applicability of scoring functions in the ranking of compounds was pointed out, while the LIE method showed an improved ability to predict good scoring poses and obtained good

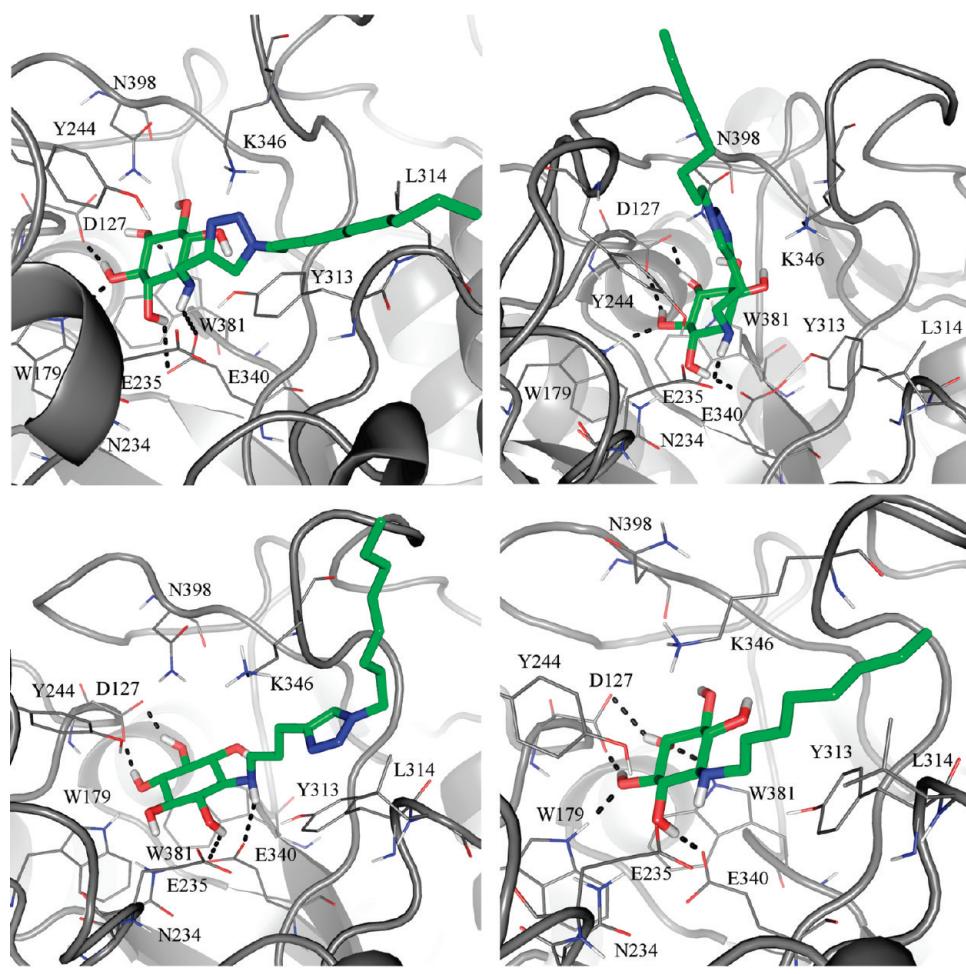


Figure 9. Average structures obtained for a representative ligand of each subfamily, showing the specific interactions with the residues in the binding site. Upper left: compound 1k; upper right: compound 2a; lower left: compound 3a; and lower right: compound 4d.

agreement with experimental results.⁴³ Similarly, the computational analysis of the binding of GCase inhibitors with the LIE methodology provides an impressive correlation with experimental results, while the scoring function failed for this purpose (see Figure 3). The current MD/LIE protocol and the results derived allow for a deeper knowledge of the molecular determinants of ligand binding, which could be further used in the ligand-design process of GCase.

■ ASSOCIATED CONTENT

S Supporting Information. The calculated affinities for a small set of compounds, using different MD conditions, is provided together with a figure showing the degree of correlation between the ligand mobility in the MD simulations and the experimental affinity values. We also provide the necessary parameter and library files for each ligand, according to the OPLS-AA force field, in the format required for the program Q. This information is available free of charge via the Internet at <http://pubs.acs.org>

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hugo.teran@usc.es, aqvist@xray.bmc.uu.se.

■ ACKNOWLEDGMENT

Financial support from the “Ministerio de Ciencia e Innovación”, Spain (Project CTQ2008-01426/BQU) and “Generalitat de Catalunya” (Grant 2009SGR-1072) is acknowledged. J.Å. acknowledges support from the Swedish Research Council (VR). L.D. is grateful to CSIC for predoctoral research training support within the JAE-Predoc program. H.G.T. is a researcher of the Isidro Parga Pondal program (Xunta de Galicia, Spain). Mr. Lars Boukharta is gratefully acknowledged for technical assistance and helpful discussions. Finally, the authors acknowledge the “Centre de Supercomputació de Catalunya” (CESCA) for allowing the use of its software and hardware resources.

■ REFERENCES

- (1) De Benedetti, P. G.; Fanelli, F. Ligand-receptor communication and drug design. *Curr. Protein Pept. Sci.* **2009**, *10*, 186–193.
- (2) Ringe, D.; Petsko, G. A. What are pharmacological chaperones and why are they interesting? *J. Biol.* **2009**, *8*, 80.
- (3) Grabowski, G. A. Treatment perspectives for the lysosomal storage diseases. *Expert Opin. Emerging Drugs* **2008**, *13*, 197–211.
- (4) Butters, T. D. Gaucher disease. *Curr. Opin. Chem. Biol.* **2007**, *11*, 412–418.
- (5) Parenti, G. Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics. *EMBO Mol. Med.* **2009**, *1*, 268–279.

- (6) Sanchez-Ruiz, J. M. Ligand effects on protein thermodynamic stability. *Biophys. Chem.* **2007**, *126*, 43–9.
- (7) Serrano, P.; Llebaria, A.; Delgado, A. Regio- and stereoselective synthesis of aminoinositol and 1,2-diaminoinsolos from conduritol B epoxide. *J. Org. Chem.* **2005**, *70*, 7829–7840.
- (8) Serrano, P.; Casas, J.; Zucco, M.; Emeric, G.; Egido-Gabas, M.; Llebaria, A.; Delgado, A. Combinatorial Approach to N-Substituted Aminocyclitol Libraries by Solution-Phase Parallel Synthesis and Preliminary Evaluation as Glucocerebrosidase Inhibitors. *J. Comb. Chem.* **2007**, *9*, 43–52.
- (9) Egido-Gabas, M.; Canals, D.; Casas, J.; Llebaria, A.; Delgado, A. Aminocyclitols as Pharmacological Chaperones for Glucocerebrosidase, a Defective Enzyme in Gaucher Disease. *ChemMedChem* **2007**, *2*, 992–994.
- (10) Sawkar, A. R.; Cheng, W. C.; Beutler, E.; Wong, C. H.; Balch, W. E.; Kelly, J. W. Chemical chaperones increase the cellular activity of N370S beta -glucosidase: a therapeutic strategy for Gaucher disease. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15428–15433.
- (11) Sanchez-Olle, G.; Duque, J.; Egido-Gabas, M.; Casas, J.; Lluch, M.; Chabas, A.; Grinberg, D.; Vilageliu, L. Promising results of the chaperone effect caused by iminosugars and aminocyclitol derivatives on mutant glucocerebrosidases causing Gaucher disease. *Blood Cells, Mol. Dis.* **2009**, *42*, 159–166.
- (12) Diaz, L.; Bujons, J.; Casas, J.; Llebaria, A.; Delgado, A. Click Chemistry Approach to New N-Substituted Aminocyclitols as Potential Pharmacological Chaperones for Gaucher Disease. *J. Med. Chem.* **2010**, *53*, 5248–5255.
- (13) Egido-Gabas, M.; Serrano, P.; Casas, J.; Llebaria, A.; Delgado, A. New aminocyclitols as modulators of glucosylceramide metabolism. *Org. Biomol. Chem.* **2005**, *3*, 1195–201.
- (14) Brumshtein, B.; Greenblatt, H. M.; Butters, T. D.; Shaaltiel, Y.; Aviezer, D.; Silman, I.; Futerman, A. H.; Sussman, J. L. Crystal structures of complexes of N-butyl- and N-nonyl-deoxynojirimycin bound to acid beta-glucosidase: insights into the mechanism of chemical chaperone action in Gaucher disease. *J. Biol. Chem.* **2007**, *282*, 29052–29058.
- (15) Lieberman, R. L.; Wustman, B. A.; Huertas, P.; Powe, A. C., Jr.; Pine, C. W.; Khanna, R.; Schlossmacher, M. G.; Ringe, D.; Petsko, G. A. Structure of acid beta-glucosidase with pharmacological chaperone provides insight into Gaucher disease. *Nat. Chem. Biol.* **2007**, *3*, 101–107.
- (16) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **2008**, *73*, 765–783.
- (17) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **2005**, *61*, 704–721.
- (18) Maestro, version 8.5; Schrödinger, LLC: New York, NY, 2008.
- (19) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (20) LigPrep, version 2.2; Schrödinger, LLC: New York, NY, 2005.
- (21) Glide, version 5.0; Schrödinger, LLC: New York, NY, 2008.
- (22) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (23) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (24) Diaz, L.; Bujons, J.; Casas, J.; Llebaria, A.; Delgado, A. New glucocerebrosidase inhibitors by exploration of chemical diversity of N substituted aminocyclitols using click chemistry and in situ screening. *J. Med. Chem.* **2011**, *S4*, accepted.
- (25) Marelus, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q. An MD program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Model.* **1999**, *16*, 213–225.
- (26) MacroModel, version 9.6; Schrödinger, LLC: New York, NY, 2005.
- (27) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Rw, I.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (28) King, G.; Warshel, A. A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. *J. Chem. Phys.* **1989**, *91*, 3647–3661.
- (29) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. Calculations of Antibody-Antigen Interactions: Microscopic and Semi-Microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to McPC603. *Protein Eng.* **1992**, *5*, 215–228.
- (30) Åqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
- (31) Hansson, T.; Marelus, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.
- (32) Almlöf, M.; Carlsson, J.; Åqvist, J. Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162–2175.
- (33) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. Binding Affinity Prediction with Different Force Fields: Examination of the Linear Interaction Energy Method. *J. Comput. Chem.* **2004**, *25*, 1242–1254.
- (34) Cheng, Y.; Prusoff, W. H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50% inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- (35) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.
- (36) SIMCA-P+, version 12.0; Umetrics: Umeå, Sweden.
- (37) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–568.
- (38) Sham, Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining Methods for Calculations of Binding Free Energies: LRA, LIE, PDLD-LRA and PDLD/S-LRA Calculations of Ligands Binding to an HIV Protease. *Proteins* **2000**, *39*, 393–407.
- (39) Gutiérrez-de-Terán, H.; Nervall, M.; Ersmark, K.; Dunn, B. M.; Hallberg, A.; Åqvist, J. Inhibitor Binding to the Plasmepsin IV Aspartic Protease from *Plasmodium Falciparum*. *Biochemistry* **2006**, *45*, 10529–10541.
- (40) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins* **1999**, *34*, 395–402.
- (41) Carlsson, J.; Boukharta, L.; Åqvist, J. Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. *J. Med. Chem.* **2008**, *51*, 2648–2656.
- (42) Bjelic, S.; Nervall, M.; Gutierrez-de-Teran, H.; Ersmark, K.; Hallberg, A.; Åqvist, J. Computational inhibitor design against malaria plasmepsins. *Cell. Mol. Life Sci.* **2007**, *64*, 2285–2305.
- (43) Nervall, M.; Hanspers, P.; Carlsson, J.; Boukharta, L.; Åqvist, J. Predicting binding modes from free energy calculations. *J. Med. Chem.* **2008**, *51*, 2657–2667.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on March 8, 2011 with errors in the last column of Table 2. The corrected version was published ASAP on March 11, 2011.