

Phosphoinositide-3-kinases (PI3Ks): Combined Comparative Modeling and 3D-QSAR To Rationalize the Inhibition of p110 α

Raphaël Frédérick* and William A. Denny

Auckland Cancer Society Research Centre (ACSRC), School of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland 1020, New Zealand

Received September 18, 2007

The p110 α isoform of the class IA PI3Ks was recently genetically validated as a promising target for anticancer therapy. However, up to now, only one compound (**PIK75** = **1**) has been reported as a very potent and selective inhibitor of this isoform. The lack of a 3D structure for this enzyme has clearly hindered the discovery of new p110 α selective compounds. In view of this, we combined target-based (homology modeling) and ligand-based (3D-QSAR) approaches in an attempt to define an integrated interaction model for p110 α inhibition. Twenty-five analogues of **1** were docked within the putative p110 α binding site, and the molecular alignment generated was subsequently used to derive QSAR models based on scoring function, free energy of binding, CoMFA, and CoMSIA. The predictive power of these models was then analyzed using a challenging test set of 5 compounds. CoMSIA, and particularly CoMFA, models were found to outperform the other methods, predicting accurately the potency of 100% of the compounds in the test set, thereby validating our p110 α homology model for use in further drug development.

INTRODUCTION

Phosphoinositide 3-kinases (PI3Ks) constitute a class of enzymes that catalyze phosphorylation of the 3-hydroxyl position of phosphoinositides (PIs). The resulting second messengers, phosphatidylinositol 3,4-bisphosphate (PIP2) and phosphatidylinositol 3,4,5-trisphosphate (PIP3), can regulate a remarkably diverse array of physiological processes, including glucose homeostasis, cell growth, differentiation, and motility.^{1,2} Eight related PI3Ks, possessing unique substrate specificity, localization, and mode of regulation, have been identified in vertebrates. The class IA PI3Ks, comprised of a regulatory subunit (p85) and three different catalytic subunits (p110 α , p110 β , p110 δ), are activated by receptor tyrosine kinases.³

The recent discoveries that the p110 α isoform undergoes gene amplification⁴ and is frequently mutated in primary tumors,^{5–7} together with evidence that PTEN, a lipid phosphatase which acts as a negative regulator, is a commonly inactivated tumor suppressor,⁸ have genetically validated p110 α as an attractive target for cancer therapy. This has fueled drug development, with intensive efforts underway to develop selective inhibitors of the ATP-binding site of p110 α as anticancer agents.^{9–14} The selective inhibition of p110 γ/δ and p110 β has also been proven valuable in the fields of inflammation,¹⁵ allergy,¹⁶ and thrombosis,¹⁷ respectively.

The p110 γ isoform is the only PI3K whose 3D-coordinates have been experimentally determined (X-ray diffraction), either alone¹⁸ or in complex with small molecules (Figure 1) such as pan-PI3K inhibitors (wortmannin and LY294002),¹⁹

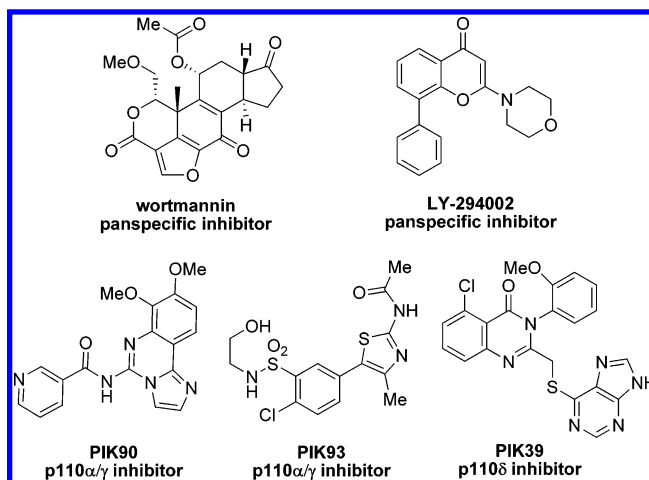


Figure 1. Structures of PI3K inhibitors that have been cocrystallized in the active site of p110 γ .

p110 α/γ inhibitors (PIK90 and PIK93), and selective p110 δ inhibitors (PIK39; IC87114).²⁰ The structural analysis of these complexes revealed essential features for PI3K inhibition. Notably, an H-bond between the Val882 backbone NH located in the p110 γ hinge region (ATP-binding site) and one H-bond acceptor atom on the ligand is recognized as the most critical interaction.

In the present work, we have combined target-based (homology modeling and docking) and ligand-based (3D-QSAR) approaches in an attempt to generate an integrated interaction model capable of correlating the chemical structures of p110 α selective inhibitors with their biological activities. Such a model would be of great assistance in a drug development program, as the activity of new analogs could be quantitatively predicted before their synthesis and testing.

* Corresponding author phone: +64-9-373 7599 ext. 86155; fax: +64-9-373 7502; e-mail: raphael.frederick@gmail.com. Present address: Belgian National Foundation for Scientific Research (FNRS), University of Namur, CBS Laboratory, 61 rue de Bruxelles, B-5000 Belgium.

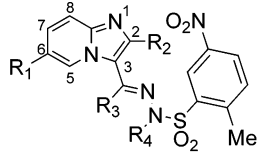
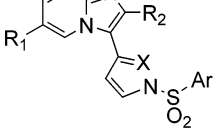
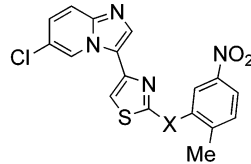
															
Cmpd	R ₁	R ₂	R ₃	R ₄	P ₁₁₀ α IC ₅₀ (μM)	Cmpd	R ₁	R ₂	X	Ar	P ₁₁₀ α IC ₅₀ (μM)	Cmpd	X	R	P ₁₁₀ α IC ₅₀ (μM)
1	Br	H	H	Me	0.0003	14	H	Me	N	4-fluorophenyl	0.67	22	S	Me	0.082
2	Cl	H	H	Me	0.0008	15	Cl	Me	N	4-fluorophenyl	0.76	23	SO ₂	Me	0.0028
3	F	H	H	Me	0.0053	16	Cl	Me	N	4-nitrophenyl	>30	24	SO	Me	0.031
4	CF ₃	H	H	Me	0.014	17	Cl	Me	N	3-nitrophenyl	0.28	25	SO	CH ₂ OH	0.020
5	Me	H	H	Me	0.006	18	Cl	Me	N	2-methyl-5-nitrophenyl	0.0028				
6	CN	H	H	Me	0.00026	19	Br	Me	N	2-methyl-5-nitrophenyl	0.0031				
7	CO ₂ Et	H	H	Me	0.34	20	Br	H	N	2-methyl-5-nitrophenyl	0.0018				
8	CONH ₂	H	H	Me	0.78	21	Cl	H	CH	2-methyl-5-nitrophenyl	0.10				
9	Br	Me	Me	Me	0.081										
10	Br	H	H	H	0.17										
11	Br	H	Me	H	0.021										
12	Br	Me	Me	H	0.017										
13	Br				0.40										

Figure 2. Structure and p110 α inhibitory potency of imidazo[1,2-*a*]pyridine derivatives **1**–**25** studied in the present work. The IC₅₀s have all been determined using the same experimental conditions (scintillation proximity assay) as described in refs 10 and 13.

RESULTS

The Inhibitor Set. To date, very few compound types have been reported as potent and selective p110 α inhibitors. Among these, the imidazo[1,2-*a*]pyridine derivative PIK75 (**1**) (Figure 2) appears to be the most potent and selective inhibitor of this isoform described to date, with IC₅₀ values of 0.0058, 1.3, 0.076, and 0.51 μ M against p110 α , β , γ , and δ , respectively.²⁰ Additional structure–activity relationships (SAR) around the imidazo[1,2-*a*]pyridine scaffold have recently been disclosed,^{10,13} including compounds with potencies ranging from sub-nM to low- μ M (Figure 2) against p110 α . These represent an excellent set to rationalize the activity and selectivity in this series using a computational approach.

Comparative Modeling. Since no experimentally derived structural data for the class IA PI3Ks (p110 α , p110 β , p110 δ) have been reported to date, molecular models of these three isoforms were first developed using homology modeling. This technique comprises 4 principal steps: (1) identification of a template, i.e., a protein of known 3D structure that shares sequence homology with the target protein, (2) alignment of the sequences of the target and template, (3) building and optimization of the 3D-model, and finally (4) quality assessment of the resulting structure. Human p110 γ , the only PI3K isoform crystallized up to now, was the obvious template. This protein shares 35% sequence identity and up to 54% sequence homology (similar residues) with each of p110 α , β , and δ . As the target/template alignment step is known to be critical to the quality of models, we used the automated homology program *ESyPred3D*.²¹ This performs a consensus alignment between the sequences of the target, the template, and other homologous proteins with the help of several different alignment algorithms and then uses *MODELLER* to generate the 3D coordinates.²² The overall quality of the resulting models was finally evaluated using several methods. Ramachandran plots (available as Supporting Information) proved to be very satisfactory with only 0.6, 0.7, and 1.2% of the residues in disallowed regions (Phi and Psi torsion angles larger than usual) for p110 α , β , and δ , respectively. It should be noted that all the residues lying in those areas are located far from the hinge region.

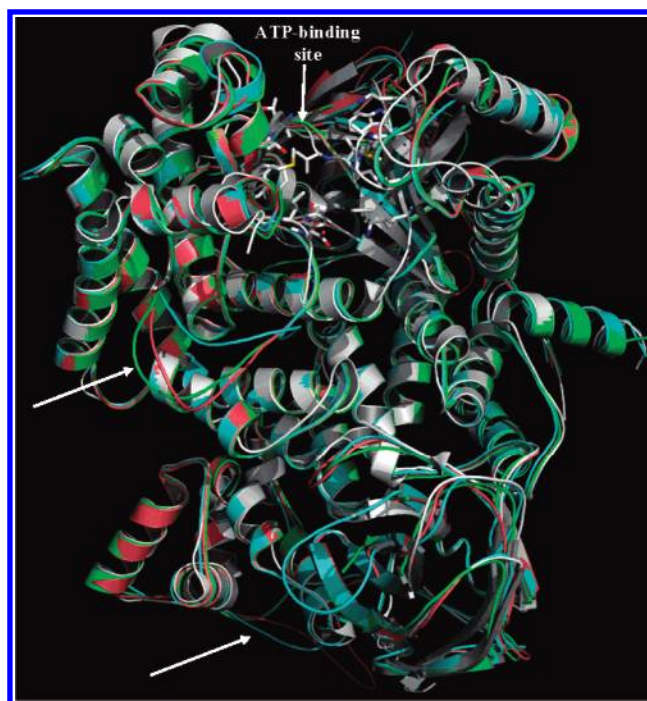


Figure 3. Superimposition of the p110 α (cyan), β (pink), and δ (green) homology models with the experimentally determined p110 γ structure (white). Arrows indicate loops difficult to model. Picture made using *PYMOL*.²³

A superimposition of p110 γ (PDB code: 2CHX) with the p110 α , β , and δ models (Figure 3) shows that the overall folding pattern (β -sheets, helices, and main loops) is well conserved among the different isoforms. Only loops, indicated by arrows in Figure 3, were difficult to model properly, since these regions were missing in the original template.

The ATP-binding site region is also very well conserved. In each isoform, the active site consists of a deep cavity bordered by conserved lipophilic residues (Figure 4), notably Met804, Trp812, Ile831, Ile879, Tyr867, Val882, and Met953 (using the p110 γ numbering). However, significant differences in amino acid constitution are also observed, depending on the isoform. For instance, Lys802 in p110 γ , is an arginine in p110 α (Arg770), a lysine in p110 β (Lys777), and a threonine in p110 δ (Thr750). More importantly, Thr886 in p110 γ , located in a key position close to

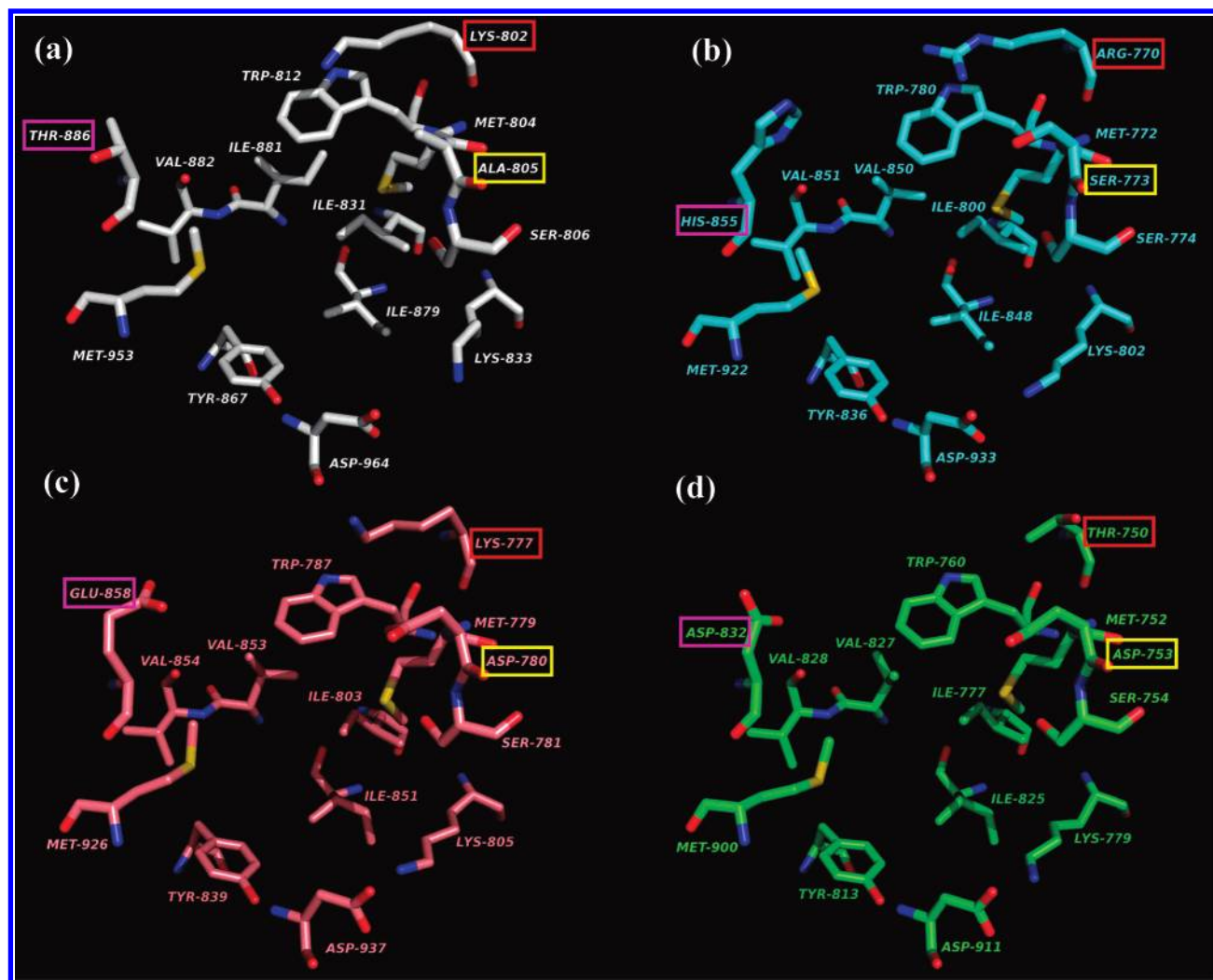


Figure 4. Comparison of the PI3K active sites of (a) p110 γ , (b) p110 α , (c) p110 β , and (d) p110 δ . Picture made using *PYMO*L.²³

the hinge region, is a histidine in p110 α (His855), a glutamic acid in p110 β (Gln858), and an aspartic acid in p110 δ (Asp832). Finally, Ala805 in p110 γ is a serine in p110 α (Ser773) and an aspartic acid in both p110 β (Asp780) and δ (Asp753). These mutations, located in crucial positions, could influence the binding of inhibitors by forming specific interactions and therefore account for their selectivity between the PI3K isoforms.

It should be noted that in the crystal structure of p110 γ in complex with PIK39 (Figure 1), the most selective p110 δ inhibitor reported until now, an important conformational rearrangement of Met804 occurred, leading to a shift in the peptide backbone that propagates through a loop which possesses low sequence identity among PI3Ks. This sequence variation was suggested to account for its selectivity, despite it being at some distances from the active site.²⁰ While no evidence has been reported so far, a similar reorganization was also proposed to rationalize p110 β inhibition. Although it is very likely that pan-inhibitors like LY294002 or wortmannin bind these isoforms in a “normal” conformation, without this rearrangement, these findings emphasize the fact that the p110 β and δ homology models should be used with care when rationalizing SAR. In contrast, such a structural rearrangement has not been suggested in p110 α . Moreover, the structural analysis of dual p110 α / γ inhibitors bound in the p110 γ active site also clearly suggests that both p110 α

and γ isoforms should adopt a similar “normal” conformation (without rearrangement) upon binding.²⁰

Therefore, to appraise the reliability of the p110 α homology model, we have investigated the binding of p110 α selective inhibitors (described above), within the modeled active site.

Molecular Docking. Compounds **1–25** were docked using the automated *GOLD* program (more details are given in the Experimental Section).²⁴ For compounds **24** and **25**, both enantiomers, i.e., with the chiral sulfoxides having the *R* and *S* configuration were evaluated.

When compounds **1–25** were docked within the p110 α active site, two different binding modes were observed. In the first conformation, illustrated with compound **1** (Figure 5a), the imidazo[1,2-*a*]pyridine scaffold is inserted deeply in the cavity, forming a hydrogen bond with the backbone NH of Asp933. One oxygen atom of the sulfonyl group interacts with the side chain of both Ser774 and Lys833, further stabilizing the ligand in the cleft. However, in that orientation, the generally highly conserved interaction with the Val 851 backbone NH is nonexistent. In addition, the substituent at the 6-position of the imidazo[1,2-*a*]pyridine ring, whose replacement has been shown to dramatically influence the p110 α inhibitory potency, is located in the core of the active site and does not form any particular interaction with active site residues. Similarly, the 2-methyl-5-nitrophe-

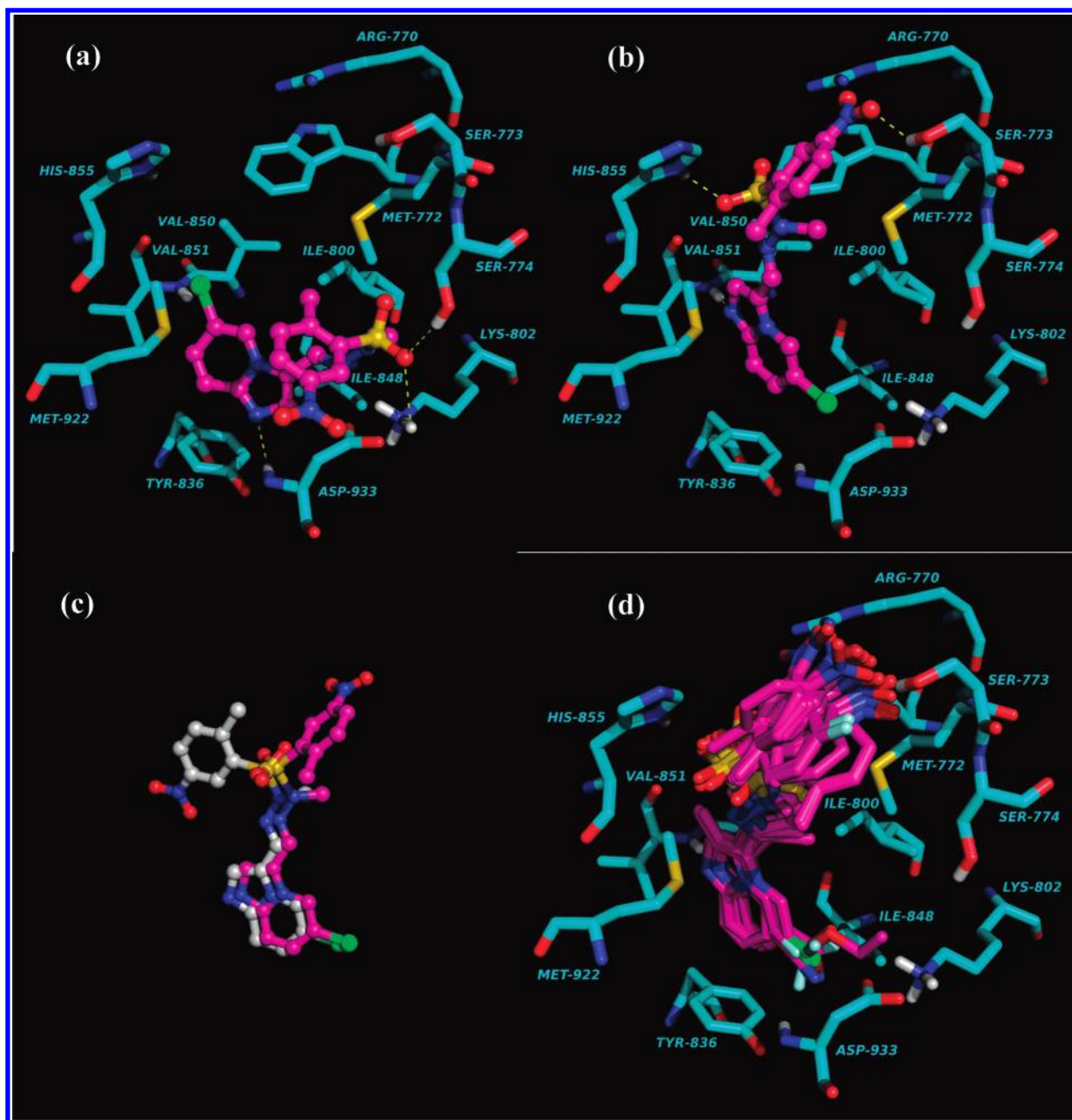


Figure 5. Docking of **1** into the p110 α cavity in (a) the first and (b) the second mode of binding; (c) superimposition of **1** docked in p110 α (magenta) and p110 γ (white); and (d) superimposition of compounds **1**–**25** docked in p110 α . Picture made with PYMOL.²³

nyl side chain, whose modification also strongly affects p110 α inhibition, does not provide any additional stabilization. Hence, this binding mode does not adequately describe the known SAR.

In contrast, the second predicted orientation (Figure 5b), similar to the one observed in p110 γ (see the Supporting Information), is consistent with the reported biological data. In this binding mode, the imidazo[1,2-*a*]pyridine scaffold was observed to be inserted deeply in the cavity, interacting with Tyr836 and Val851 through *T-Shape* and H-bond contacts, respectively. The bromine atom in the 6-position of the imidazo[1,2-*a*]pyridine scaffold is located in a small lipophilic cavity bordered by residues Ile848, Lys802, Asp933, Ile932 (not shown), and Tyr 836. An H-bond between one oxygen atom of the sulfonyl group and the imidazole NH of His855 further stabilizes this conformation. This corroborates our initial hypothesis that this His855

residue, which is a threonine (Thr886), a glutamate (Gln858), and an aspartate (Asp832) in p110 γ , β , and δ respectively, is particularly important for inhibitor selectivity within PI3K isoforms. However, in contrast to what we observed in p110 γ , the 2-methyl-5-nitrophenyl group is rotated when bound to p110 α (Figure 5c). In that orientation, the nitro moiety is poised to make an H-bond with the OH of Ser773 (Figure 5b). Interestingly, p110 α is again the only isoform which possesses a residue (Ser773) in that position, characterized by an H-bond donor group on its side chain (compared to Ala805, Asp780, and Asp753 in the p110 γ , β , and δ isoforms, respectively).

Surprisingly, for the sulfoxides **24** and **25**, only the *R*-enantiomers were predicted to bind in the p110 α cavity in this second binding mode, with the *S*-enantiomers predicted to bind in either the first mode or in conformations not observed previously. This suggests that the *R*-enantiomers

of **24** and **25** should be more potent compared to their *S*-enantiomers and that they would therefore primarily account for the activity observed in the racemic mixture.

The superimposition of compounds **1–25** (Figure 4d) reveals a well conserved overall conformation, particularly for the imidazo[1,2-*a*]pyridine ring, with slight variations in the orientation of the aromatic side chain. The two main interactions stabilizing the more active compounds within the cavity involve both residues His855 and Ser773 which are specific to the p110 α isoform and therefore might account for the selectivity profile observed in this series.

To support the proposed p110 α interaction pattern and help in the design of active compounds, different QSAR models were developed, compared, and eventually combined with the docking analysis.

QSAR Analysis. The quantitative analysis of structure–activity relationships (QSAR) is a computational technique that establishes models relating biological activities in a series of ligands (generally K_i or IC_{50}) to various descriptors (shape, size, chemical property, etc.). A QSAR analysis is therefore typically a ligand-based approach.

The Training Set. From the 25 inhibitors used in the docking study, five (**5**, **9**, **19**, **24**, and **25**) were retained as a test set of compounds to measure the predictive power of our models. Moreover, derivative **16** was discarded from the analysis as no accurate biological data was given for it. This left 19 structurally diverse p110 α inhibitors in the training set.

Docking-Based QSAR. 3D-QSAR techniques first require the molecular alignment of the studied compounds in Cartesian space. In the present study, our aim was to derive quantitative models in order to support the p110 α interaction model we have proposed. In this context, one interesting approach is to use the alignment generated during the docking study as a starting point for the QSAR analysis. This approach should be very useful as it ultimately allows a combination of both target- and ligand-based approaches in one general integrated model, confirming or disproving the interaction model initially generated.

During this work, (1) the statistical models linking the biological activities to the different descriptors were built by means of partial least-square (PLS) regression, (2) the degree of correlation of experimental vs predicted values was expressed in terms of the square of the correlation coefficient (r^2), indicating the fraction of explained variance, and (3) the internal predictability was measured in terms of cross-validated r^2 , hereafter referred to as q^2 , after cross-validation using the leave-one-out method.

The scoring function used by the docking software to rank the solutions according to their fit in the active site is the simplest descriptor related to the binding that can be used to derive a QSAR model. In the present case, when all the training set compounds are taken into account in the regression, no correlation between the GoldScore and the pIC_{50} is found ($r^2 = 0.017$; $q^2 = -0.206$) (Table 1, Figure 6a). If compounds possessing an error on the prediction larger than 1 log unit are considered as outliers, 12 derivatives appear to be mispredicted. Indeed, many papers have reported that docking programs reliably provide the bioactive conformation of ligands in the binding pocket but that the scoring function alone cannot be efficiently used to predict the

Table 1. PLS Statistics for the Prediction of the p110 α Inhibitory Potency of the Training Set Compounds

	method	r^2 ^a	q^2 ^b	RMSE ^c	outliers
docking-based	GoldScore	0.017	−0.206	1.207	12
	$\Delta E_{\text{binding}}$	0.484	0.347	0.875	2
	CoMFA	0.959	0.556	0.270	0
	CoMSIA	0.910	0.534	0.419	0
ligand-based	CoMFA	0.743	0.032	0.657	2
	CoMSIA	0.696	0.118	0.714	3

^a Square of the correlation coefficient. ^b Cross-validated square of the correlation coefficient (leave-one-out method). ^c Root-mean-square error (RMSE) in the non-cross-validated analysis.

ligands' affinity, particularly when using homology models rather than experimentally determined protein structures.^{25,26}

Alternatively, the free energy of binding ($\Delta E_{\text{binding}}$) can be considered as an improved descriptor compared with the docking score and is calculated using the following equation

$$\Delta E_{\text{binding}} = E_{\text{enzyme–ligand complex}} - E_{\text{free enzyme}} - E_{\text{free ligand}}$$

where $E_{\text{enzyme–ligand complex}}$ (kcal·mol^{−1}), $E_{\text{free enzyme}}$ (kcal·mol^{−1}), and $E_{\text{free ligand}}$ (kcal·mol^{−1}) are the total energy of the enzyme–ligand complexes, the energy of the enzyme without the ligand (free enzyme), and the energy of the free ligand as extracted from the complex, respectively. Although a significant improvement is observed using this descriptor when considering all derivatives ($r^2 = 0.484$; $q^2 = 0.347$) (Table 1, Figure 6b) compared to the GoldScore prediction, this model still does not provide the predictability necessary to estimate the potency of novel compounds.

Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) are 3D-QSAR methods intended to correlate the molecular features of a series of compounds with their biological activities. These two methods do not take into account receptor–ligand interactions but rely only on the calculation of molecular fields of the ligands and their subsequent correlations, by partial-least-square (PLS) regression, to their biological activities. In particular, CoMFA is based on the calculation of steric and electrostatic fields, while CoMSIA considers also hydrophobic and H-bond donor and acceptor fields. To provide meaningful models, CoMFA and CoMSIA rely on a knowledge of the ligands' bioactive conformation and their relative alignment. In this study, we derived this information directly from the docking studies, thus generating docking-based 3D-QSAR models. In order to assess the importance of this alignment to the quality of the model, more classical ligand-based 3D-QSAR models were also developed. Here, in contrast to docking-based alignment, ligands from the training set were built, minimized (Tripos force field), and aligned to each other on the basis of a common template, i.e., the heavy (hetero)atoms from the imidazo[1,2-*a*]pyridine, by means of the *ALIGN* command implemented in *SYBYL*.

When compared to docking score and energy-related calculations, the PLS analysis of the docking-based CoMFA ($r^2 = 0.959$; $q^2 = 0.556$) (Table 1, Figure 6c) and CoMSIA ($r^2 = 0.910$; $q^2 = 0.534$) (Table 1, Figure 6d) appeared statistically more robust and showed a higher internal predictivity. This is also true when these models are compared to the ones derived from the ligand-based approach (CoMFA, Table 1 Figure 6e, $r^2 = 0.743$; $q^2 = 0.032$;

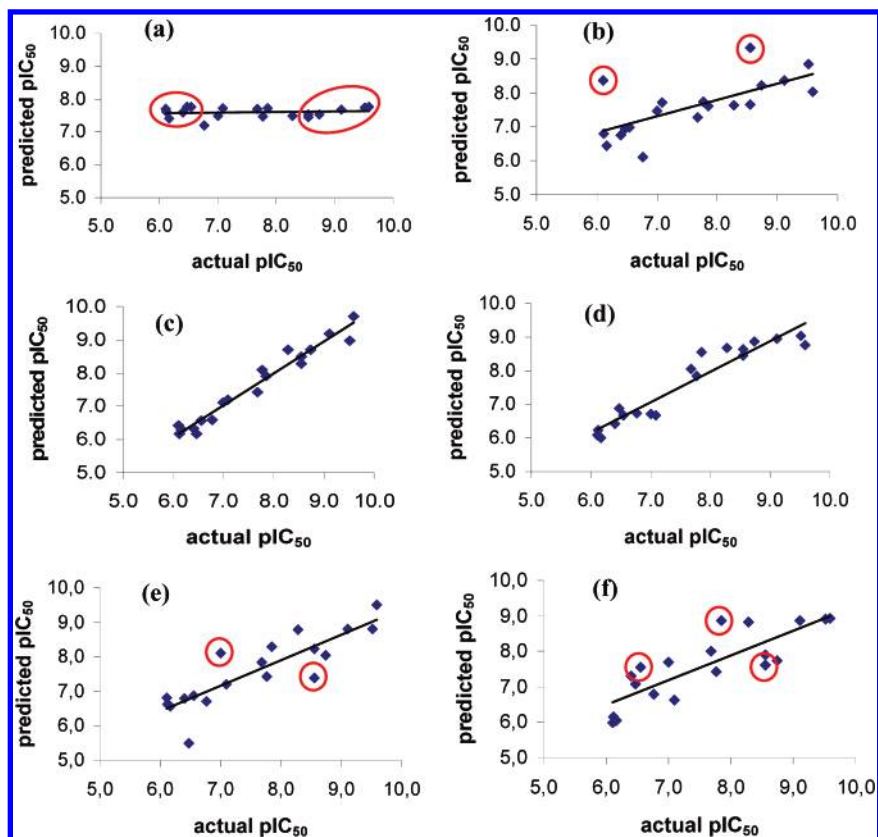


Figure 6. Plots of actual vs predicted p110 α inhibitory potency (pIC₅₀) for the training set compounds using (a) GoldScore; (b) the free energy of binding ($\Delta E_{\text{binding}}$); (c) docking-based CoMFA and (d) CoMSIA; (e) ligand-based CoMFA; and (f) CoMSIA. Red circles correspond to outliers (error of prediction > 1.0 log unit)

CoMSIA, Table 1 Figure 6f, $r^2 = 0.696$; $q^2 = 0.118$). This result clearly emphasizes the importance of considering the ligands' conformation the closest to the bioactive one to develop reliable 3D-QSAR models.

The docking-based CoMFA and CoMSIA 3D molecular fields, since they are derived from docking results, can be positioned into the p110 α binding cavity. This highlights the good complementarities between the 3D-QSAR and the proposed protein–ligand interaction model. First, the steric fields produced by both CoMFA and CoMSIA (Figure 7a,c) match precisely the topology of the receptor. Steric clashes are suggested for large substituents in the 2- and 6-positions of the imidazo[1,2-*a*]pyridine ring, corresponding respectively to interactions with the backbone carbonyl of Val851 and the side-chain carboxylate of Asp933. On the contrary, increased potency is expected for large groups in the proximity of the methyl-hydrazone moiety of **1**, which would occupy the central core of the p110 α cavity.

Second, the electrostatic fields mainly emphasize the importance of the imidazo[1,2-*a*]pyridine scaffold (Figure 7b,d), which forms a *T-shape* interaction (typically electronic) with the aromatic ring of Tyr836. A favorable hydrophobic area is found close to the 6-position of the imidazo[1,2-*a*]pyridine ring (Figure 7e), supporting the critical benefit of a small, lipophilic substituent in that position. More interestingly, H-bond donor fields are detected by CoMSIA in the proximity of the nitro and sulfonyl groups of **1** (Figure 7f) and are complemented by the Ser773 and His855 residues in the p110 α model. As hypothesized above, these two residues, specific to p110 α , probably contribute to selectivity within this series. Finally, an additional H-bond donor field,

matching the backbone NH of Asp933 in the p110 α model, is found around the 6-position of the aromatic scaffold.

In conclusion, the good complementarities found between the fields generated by both docking-based CoMFA and CoMSIA and the homology modeled structure strongly support our p110 α model. This encouraged us to analyze the predictivity of our models using a test set of inhibitors.

Test Set: Inhibitors. To measure the external predictivity of our models, a challenging test set of 5 compounds (**5**, **9**, **19**, **24**, **25**) was chosen. While belonging to the imidazo[1,2-*a*]pyridine family, each contains at least one substituent or a combination of substituents not featured in the training set and thus has interesting structural diversity. Consistent with the results from the docking study, the p110 α inhibitory potency prediction for compounds **24** and **25** was made using the *R*-configuration of their sulfoxide moieties.

Test Set: Prediction and Validation. The predicted p110 α inhibitory potencies of the compounds in the test set, expressed as pIC₅₀, are reported in Table 2. Plots of experimental versus predicted potencies are given in Figure 8.

The root-mean-square error of prediction (RMSEP) proved to be low for all methodologies. Therefore all compounds were correctly recognized as p110 α inhibitors. However, models based on predictions using the GoldScore and $\Delta E_{\text{binding}}$ exhibited bad ranking capabilities over the whole test set, measured in terms of r^2 . No correlation was found using the GoldScore predictions (Table 2, Figure 8a), and 2 out of the 5 compounds in the test set were found to be outliers using $\Delta E_{\text{binding}}$ (Table 2, Figure 8b). In contrast, both docking-based 3D-QSAR models (Table 2, Figure 8c,d, for

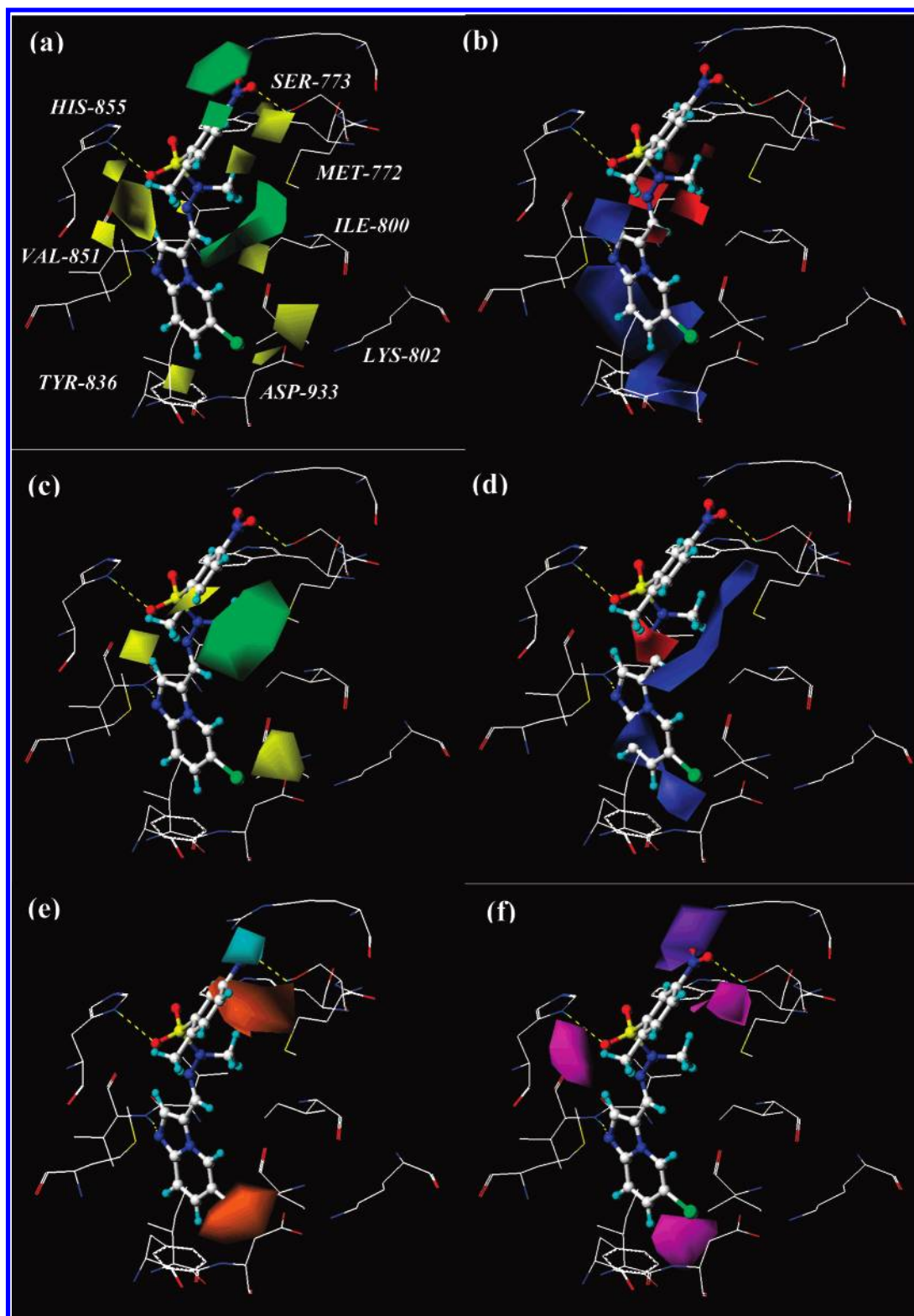


Figure 7. Contour maps of the docking-based CoMFA and CoMSIA models ($\text{StDev} \times \text{Coeff}$) docked into the p110 α binding site. Compound **1** is shown as a representative inhibitor. (a) CoMFA and (c) CoMSIA steric fields: bulky substituents increase the potency in the green areas and decrease it in the yellow areas (contour levels are 80% and 35% for CoMFA and 85% and 20% for CoMSIA); (b) (CoMFA) and (d) CoMSIA electrostatic fields: positive charges increase the potency in the blue areas and decrease it in the red areas (contour levels are 90% and 25% for CoMFA and 80% and 20% for CoMSIA); (e) CoMSIA hydrophobic fields: hydrophobic substituents increase the potency in the orange areas and decrease it in the cyan areas (contour levels are 80% and 25%); and (f) CoMSIA H-bond fields: H-bond donors on the protein increase the potency in the magenta areas and decrease in the violet areas (contour levels are 80% and 20%).

CoMFA and CoMSIA, respectively) performed significantly better, both being characterized by a concomitant low RMSEP (0.23 and 0.17, respectively) and a good predictive r^2 (0.51 and 0.52, respectively). Indeed, these models, also

performing better than classical ligand-based CoMFA (Table 2, Figure 8e, poor r^2 and 1 outlier) and CoMSIA (Table 2, Figure 8f, high RMSEP), accurately predicted the potency of 100% of the test set compounds.

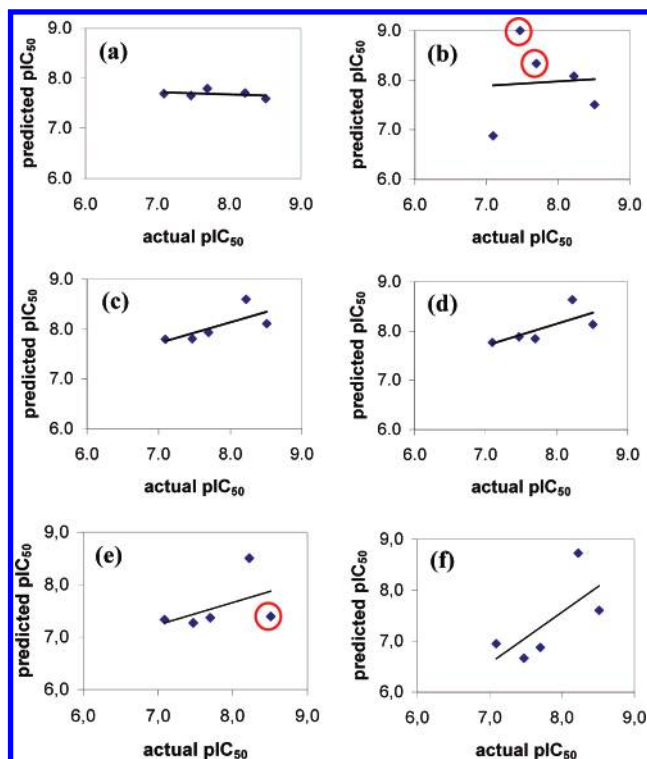


Figure 8. Plots of actual vs predicted p110 α inhibitory potency (pIC₅₀) for the test set compounds using (a) GoldScore; (b) the free energy of binding ($\Delta E_{\text{binding}}$); (c) docking-based CoMFA and (d) CoMSIA; (e) ligand-based CoMFA; and (f) CoMSIA. Red circles correspond to outliers (error of prediction >1.0 log unit).

DISCUSSION AND CONCLUSION

The PI3K p110 α isoform is a very attractive target for anticancer drug therapy. However, until now, only a few compounds have been reported as very potent and selective inhibitors of this isoform.^{10,13} The development of new p110 α selective compounds is therefore now the focus of intense research efforts. However, the absence of any experimental p110 α 3D structure is hindering the rational development of new potent and selective drugs.

Homology modeling still remains one of the most powerful techniques to predict the 3D structure of macromolecules when no experimental structural information is available. Applying this method to class IA PI3Ks (p110 α , β , δ) revealed that the 3 isoforms adopt very similar overall 3D conformations. However, comparison of their putative active sites highlighted significant differences, notably residues His855 and Ser773, in the hinge region, which are specific to the p110 α isoform and which provide opportunities for selective van der Waals and H-bonding. The docking of compounds **1–25** within the p110 α cavity further suggested that these amino acids could account for selectivity within this series.

One way to support an interaction model based on a protein inferred by homology modeling is to establish QSAR models using the molecular alignment generated by docking. By doing so, it is possible to combine both approaches into an integrated 3D-QSAR model intended to confirm or disprove the initial target-based approach. The use of GoldScore or the calculated free energy of binding ($\Delta E_{\text{binding}}$), as a unique descriptor relating the activity, unsurprisingly, did not provide satisfactory results as these are both highly

dependent on the precise topology of the active site. As a result, both descriptors performed poorly in predicting the activity of the test set compounds. In contrast, both docking-based 3D-QSAR studies (CoMFA and CoMSIA) yielded very good correlation coefficients and good internal predictivity. The good estimations of p110 α activity of compounds within the test set support the validity of these models, which predicted remarkably accurately the activity of 100% of the test set. Consequently, the CoMFA and CoMSIA maps were merged into the p110 α homology model, where they displayed excellent complementarities between the ligand-based 3D contour maps and the molecular features of the protein. This highlights the structural features required to design new p110 α selective inhibitors. A central (hetero)-aromatic scaffold bearing an H-bond acceptor that could interact with Val851, conserved among the different isoforms, is required to obtain PI3K inhibition. To selectively target p110 α , this core should be substituted on one side with a small lipophilic group that could interact deeply in its ATP-binding site and on the other side with two H-bond acceptors to form strong H-bonds with Ser773 and His855, two residues highly specific to p110 α .

This study has therefore yielded a useful interaction model for p110 α inhibition, providing structural requirements that could serve to design new p110 α selective inhibitors.

EXPERIMENTAL SECTION

Molecular modeling studies were carried out on a Linux workstation.

Homology Modeling. The human p110 α , β , and δ primary sequences were obtained from the Swiss-Prot database (accession number P42336, P42338, and O00329, respectively). Sequence analyses were performed using BLAST²⁷ through the Protein Data Bank (BLOSUM62 matrix).²⁸ The human p110 γ isoform (PDB code 1E7U)¹⁹ was selected as the most appropriate template in each case. Amino acid sequences were then aligned by means of the *ESyPred3D* program (available at www.fundp.ac.be/urbm/bioinfo/esypred/).²¹ This automated homology modeling program compares results from various multiple alignment algorithms (such as PSI-BLAST, ClustalW, Dialign2, etc.) to derive a “consensus” alignment between the target sequence and the templates. Furthermore, 3D models (built with Modeller²²) were also provided with *ESyPred3D*. The quality of each model was finally analyzed by means of the PDBsum server.²⁹ Ramachandran plots are available in the Supporting Information.

Docking Simulations. Compounds **1–25** were built using the *SKETCH* module, as implemented in *SYBYL* (version 7.3),³⁰ and their geometry was optimized using the *MINIMIZE* module. The minimization process uses the Powell method with the Tripos force field (dielectric constant 1 r) to reach a final convergence of 0.01 kcal·mol⁻¹. Docking simulations were performed in the homology model of the human p110 α isoform with the automated *GOLD* program (active site definition: residues within 5 Å around **1** as docked in p110 γ).²⁴ In each case, 50 conformations were produced and clustered by similarity. In order to take into account protein flexibility, the conformations with the best score (GoldScore) in both orientations were further refined using the minimization process described above. The free

Table 2. Experimental and Predicted p110 α Inhibitory Potency for the Test Set Compounds^a

compd	exp	p110 α inhibition expressed as pIC ₅₀ predicted using					
		GoldScore	$\Delta E_{\text{binding}}$	docking-based		ligand-based	
				CoMFA	CoMSIA	CoMFA	CoMSIA
5	8.22	7.71 (−0.51)	8.08 (−0.14)	8.60 (0.42)	8.63 (0.41)	8.51 (0.29)	8.73 (0.51)
9	7.09	7.69 (0.60)	6.88 (−0.21)	7.80 (0.79)	7.77 (0.69)	7.33 (0.24)	6.95 (−0.14)
19	8.51	7.59 (−0.92)	7.51 (−1.00)	8.11 (−0.40)	8.68 (0.17)	7.39 (−1.12)	7.61 (−0.90)
24	7.47	7.66 (0.19)	9.00 (1.53)	7.81 (0.34)	7.88 (0.41)	7.27 (−0.20)	6.67 (−0.80)
25	7.70	7.80 (0.10)	8.33 (0.63)	7.93 (0.23)	7.85 (0.15)	7.37 (−0.33)	6.88 (−0.88)
<i>r</i> ²		no correlation	0.01	0.51	0.52	0.22	0.47
RMSEP		0.30	0.76	0.23	0.17	0.31	0.50
outliers		0	2	0	0	1	0

^a Errors of predictions are shown in parentheses.

energy of binding ($\Delta E_{\text{binding}}$), generated by compounds **1–25** inside the p110 α active site, was calculated using the following equation

$$\Delta E_{\text{binding}} = E_{\text{enzyme-ligand complex}} - E_{\text{free enzyme}} - E_{\text{free ligand}}$$

where $E_{\text{enzyme-ligand complex}}$ (kcal·mol^{−1}), $E_{\text{free enzyme}}$ (kcal·mol^{−1}), and $E_{\text{free ligand}}$ (kcal·mol^{−1}) are the total energy of the enzyme–ligand complexes, the energy of the enzyme without the ligand (free enzyme), and the energy of the free ligand as extracted from the complex, respectively. Predictions of the pIC₅₀ values were obtained by PLS regression of the GoldScore and $\Delta E_{\text{binding}}$ descriptors using the QSAR module as implemented in SYBYL (version 7.3) with the default parameters. Experimental pIC₅₀ values were used as dependent variables.

CoMFA and CoMSIA. A 3D cubic lattice with grid spacing of 1 Å and extending 4 Å around the aligned molecules in all directions was used. The molecular fields (steric and electrostatic for CoMFA and steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor for CoMSIA) were calculated with a probe atom with the VdW properties of an sp³ carbon and with a charge of +1.0. For the generation of the CoMFA fields, a distance dependent dielectric constant was selected for the Coulombic electrostatic energy calculation, and a cutoff of 30 kcal·mol^{−1} with smooth transition was used. The standard parameters were used for the generation of the CoMSIA fields. The PLS regression, using the experimental pIC₅₀ values as dependent variables, and the leave-one-out cross-validation were performed using the QSAR module as implemented in SYBYL.³⁰ For both CoMFA and CoMSIA, the CoMFA standard scaling and a column filtering of 2 kcal·mol^{−1} were used. Gasteiger–Hückel charges were used for all ligands.

ACKNOWLEDGMENT

R.F. thanks Drs. C. Charlier, B. D. Palmer, and J. D. Kendall for their useful advice and critical reading of the manuscript. This work was funded in part by the Health Research Council of New Zealand and by the Maurice Wilkins Centre for Molecular Biodiscovery at the University of Auckland.

Supporting Information Available: Ramachandran plots obtained for p110 α (Figure S1), β (Figure S2), and δ (Figure S3); detailed description of the binding of **1** within the p110 γ active site (Figure S4); and enlargements of Figure 5a (Figure S5) and Figure 5b (Figure S6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- (1) Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **2002**, *296*, 1655–1657.
- (2) Leever, S. J.; Vanhaesebroeck, B.; Waterfield, M. D. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr. Opin. Cell. Biol.* **1999**, *11*, 219–225.
- (3) Katso, R.; Okkenhaug, K.; Ahmadi, K.; White, S.; Timms, J.; Waterfield, M. D. Cellular function of phosphoinositide 3-kinases: implications for development, homeostasis, and cancer. *Annu. Rev. Cell. Dev. Biol.* **2001**, *17*, 615–675.
- (4) Stephens, L.; Williams, R.; Hawkins, P. Phosphoinositide 3-kinases as drug targets in cancer. *Curr. Opin. Pharmacol.* **2005**, *5*, 357–365.
- (5) Gymnopoulos, M.; Elsliger, M. A.; Vogt, P. K. Rare cancer-specific mutations in PIK3CA show gain of function. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5569–5574.
- (6) Samuels, Y.; Wang, Z.; Bardelli, A.; Silliman, N.; Ptak, J.; Szabo, S.; Yan, H.; Gazdar, A.; Powell, S. M.; Riggins, G. J.; Willson, J. K.; Markowitz, S.; Kinzler, K. W.; Vogelstein, B.; Velculescu, V. E. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **2004**, *304*, 554.
- (7) Vogt, P. K.; Kang, S.; Elsliger, M. A.; Gymnopoulos, M. Cancer-specific mutations in phosphatidylinositol 3-kinase. *Trends Biochem. Sci.* **2007**, *32*, 342–349.
- (8) Cantley, L. C.; Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4240–4245.
- (9) Foukas, L. C.; Claret, M.; Pearce, W.; Okkenhaug, K.; Meek, S.; Peskett, E.; Sancho, S.; Smith, A. J.; Withers, D. J.; Vanhaesebroeck, B. Critical role for the p110 α phosphoinositide-3-OH kinase in growth and metabolic regulation. *Nature* **2006**, *441*, 366–370.
- (10) Hayakawa, M.; Kaizawa, H.; Kawaguchi, K.; Ishikawa, N.; Koizumi, T.; Ohishi, T.; Yamano, M.; Okada, M.; Ohta, M.; Tsukamoto, S.; Raynaud, F. I.; Waterfield, M. D.; Parker, P.; Workman, P. Synthesis and biological evaluation of imidazo[1,2-*a*]pyridine derivatives as novel PI3 kinase p110 α inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 403–412.
- (11) Hayakawa, M.; Kaizawa, H.; Moritomo, H.; Koizumi, T.; Ohishi, T.; Okada, M.; Ohta, M.; Tsukamoto, S.; Parker, P.; Workman, P.; Waterfield, M. Synthesis and biological evaluation of 4-morpholino-2-phenylquinazolines and related derivatives as novel PI3 kinase p110 α inhibitors. *Bioorg. Med. Chem.* **2006**, *14*, 6847–6858.
- (12) Hayakawa, M.; Kaizawa, H.; Moritomo, H.; Koizumi, T.; Ohishi, T.; Yamano, M.; Okada, M.; Ohta, M.; Tsukamoto, S.; Raynaud, F. I.; Workman, P.; Waterfield, M. D.; Parker, P. Synthesis and biological evaluation of pyrido[3',2':4,5]furo[3,2-*d*]pyrimidine derivatives as novel PI3 kinase p110 α inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2438–2442.
- (13) Hayakawa, M.; Kawaguchi, K. I.; Kaizawa, H.; Koizumi, T.; Ohishi, T.; Yamano, M.; Okada, M.; Ohta, M.; Tsukamoto, S. I.; Raynaud, F. I.; Parker, P.; Workman, P.; Waterfield, M. D. Synthesis and biological evaluation of sulfonylethylhydrazone-substituted imidazo[1,2-*a*]pyridines as novel PI3 kinase p110 α inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 5837–5844.
- (14) Zhao, J. J.; Cheng, H.; Jia, S.; Wang, L.; Gjoerup, O. V.; Mikami, A.; Roberts, T. M. The p110 α isoform of PI3K is essential for proper growth factor signaling and oncogenic transformation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16296–16300.
- (15) Ruckle, T.; Schwarz, M. K.; Rommel, C. PI3Kgamma inhibition: towards an 'aspirin of the 21st century'. *Nat. Rev. Drug Discovery* **2006**, *5*, 903–918.
- (16) Ali, K.; Bilancio, A.; Thomas, M.; Pearce, W.; Gilfillan, A. M.; Tkaczyk, C.; Kuehn, N.; Gray, A.; Giddings, J.; Peskett, E.; Fox, R.

- Bruce, I.; Walker, C.; Sawyer, C.; Okkenhaug, K.; Finan, P.; Vanhaesebroeck, B. Essential role for the p110delta phosphoinositide 3-kinase in the allergic response. *Nature* **2004**, *431*, 1007–1011.
- (17) Jackson, S. P.; Schoenwaelder, S. M.; Goncalves, I.; Nesbitt, W. S.; Yap, C. L.; Wright, C. E.; Kenche, V.; Anderson, K. E.; Dopheide, S. M.; Yuan, Y.; Sturgeon, S. A.; Prabakaran, H.; Thompson, P. E.; Smith, G. D.; Shepherd, P. R.; Daniele, N.; Kulkarni, S.; Abbott, B.; Saylik, D.; Jones, C.; Lu, L.; Giuliano, S.; Hughan, S. C.; Angus, J. A.; Robertson, A. D.; Salem, H. H. PI 3-kinase p110beta: a new target for antithrombotic therapy. *Nat. Med.* **2005**, *11*, 507–514.
- (18) Walker, E. H.; Perisic, O.; Ried, C.; Stephens, L.; Williams, R. L. Structural insights into phosphoinositide 3-kinase catalysis and signaling. *Nature* **1999**, *402*, 313–320.
- (19) Walker, E. H.; Pacold, M. E.; Perisic, O.; Stephens, L.; Hawkins, P. T.; Wymann, M. P.; Williams, R. L. Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol. Cell.* **2000**, *6*, 909–919.
- (20) Knight, Z. A.; Gonzalez, B.; Feldman, M. E.; Zunder, E. R.; Goldenberg, D. D.; Williams, O.; Loewith, R.; Stokoe, D.; Balla, A.; Toth, B.; Balla, T.; Weiss, W. A.; Williams, R. L.; Shokat, K. M. A pharmacological map of the PI3-K family defines a role for p110alpha in insulin signaling. *Cell* **2006**, *125*, 733–747.
- (21) Lambert, C.; Leonard, N.; De Bolle, X.; Depiereux, E. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* **2002**, *18*, 1250–1256.
- (22) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
- (23) Delano, W. L. The PyMOL Molecular Graphics System on World Wide Web. <http://pymol.sourceforge.net/> (accessed Dec 10, 2007).
- (24) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (25) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.
- (26) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (27) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
- (28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (29) Laskowski, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **2001**, *29*, 221–222.
- (30) SYBYL Version 7.3; Tripos International: St. Louis, MO, 2004.

CI700348M