

3D-QSAR CoMFA Study on Imidazolinergic I₂ Ligands: A Significant Model through a Combined Exploration of Structural Diversity and Methodology

Nicolas Baurin,[†] Eric Vangrevelinghe,[†] Luc Morin-Allory,[†] Jean-Yves Mérour,[†] Pierre Renard,[‡] Marc Payard,[§] Gérald Guillaumet,[†] and Christophe Marot*,[†]

Institut de Chimie Organique et Analytique, associé au CNRS, Université d'Orléans, BP-6759, 45067 Orléans Cedex 2, France, ADIR et Cie, 1 rue Carle Hébert, 92415 Courbevoie, France, and Laboratoire de Chimie Thérapeutique, Faculté de Pharmacie de Toulouse, Université Paul Sabatier, 35 chemin des Maraîchers, 31400 Toulouse, France

Received October 27, 1999

Displaying an unprecedented structural diversity, 119 I₂ ligands, and their pK_i values, were collected and submitted to a comparative molecular fields analysis (CoMFA) study. They were discerned into three structural subsets (A, B, C), to explore the I₂ 3D-QSARs from finite structural systems (A, B, C) to more complex ones (AB, AC, BC, ABC). In addition, various key steps of the CoMFA methodology were explored. The applied method used two pharmacophore templates and seven molecular field combinations (electrostatic, lipophilic, steric), as well as eight alignment methods (two point-by-point and six similarity-based variations). That way, 644 CoMFA models were obtained and further selected according to their predictive ability through two filters. The first filter was mainly based on the q^2 , which internally evaluates the predictive ability from the training set. For the second filter, the predictive ability was externally evaluated through the prediction of test sets. Finally, one model was extracted from the whole data as the best. Indeed, it combines three features of utmost importance for the further design of ligands endowed with high I₂ affinity: structural diversity ($n = 73$), robustness ($N = 9$, $r^2 = 0.96$, $s = 0.28$, $F = 148$), and a great fully assessed predictive ability ($q^2 = 0.50$, $r^2_{\text{test set}} = 0.81$, $n_{\text{test set}} = 46$). On the basis of structural data and CoMFA isocontours, some elements of the I₂ tridimensional pharmacophore are also suggested.

Introduction

Working on Clonidine (Figure 1), a centrally acting antihypertensive with an imidazoline structure, Bousquet and co-workers¹ first established the existence of receptors, mediating the lowering of blood pressure, which were distinct from the α_2 -adrenoceptors 15 years ago. Here was initiated the concept of the IBS (imidazoline binding sites) which actually reflects the fact that imidazolines are the most important identified ligands defining those non-adrenoceptor recognition sites. Subsequent radioligand binding studies and then functional studies,² while providing new evidence supporting the IBS concept such as the identification of endogenous ligands,^{2–4} also pointed out that the IBS constituted a heterogeneous family. Indeed, the IBS, in addition to different ligand recognition properties² and subcellular localizations,² present a wide tissue distribution² in different species including humans. They have been found to be part of both central and peripheral nervous systems and localized in various organs² such as the brain, kidney, stomach, prostate, lung, and heart. As for their implications in physiological functions, we can cite the decrease of blood^{2,5,6} and intraocular pressures,⁷ increase of neuron activity,⁸ insulin release,^{3,9,10} protein synthesis in astrocytes,¹¹ and also inhibitions in various systems such as the release of neurotransmitters¹² or

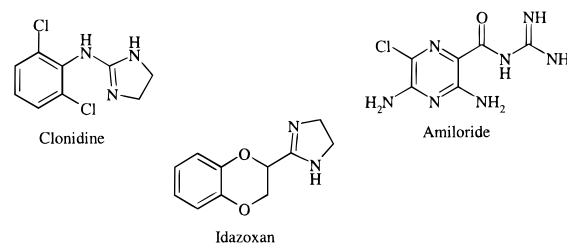


Figure 1. Molecular structures of Clonidine, Idazoxan, and Amiloride.

MAO (monoamine oxidase) activity.^{2,13} IBS have also been investigated for their role in food intake,¹⁴ gastric acid secretions,^{2,15} mood disorders,^{2,16} opioid tolerance,¹⁷ neuroprotection,^{2,17} renal function,^{2,18} and Alzheimer's¹¹ and Parkinson's¹⁹ diseases. Nevertheless, the heterogeneity of the IBS, if it emphasizes their biological importance, is also the major obstacle for the IBS concept to be fully validated. In this context, with respect to their selectivity profile in radioligand studies, two major subtypes have first been discerned:²⁰ I₁ and I₂ IBS subtypes respectively characterized by [³H]-Clonidine and [³H]Idazoxan (Figure 1). Second, I₂-IBS was further divided²¹ into I₂A-IBS and I₂B-IBS respectively characterized by a high and low affinity for Amiloride (Figure 1). Now, the further characterization of the IBS is crucial in order to identify new targets specifically linked to a physiopathology: this step implies the development of compounds with better affinity and selectivity. Computer-aided drug design can help us this way by rationalizing and optimizing the process consisting in using what is already established,

* To whom correspondence should be addressed. Tel: +33 (0)2 38 49 45 91. Fax: +33 (0)2 38 41 72 81. E-mail: Christophe.Marot@univ-orleans.fr.

[†] Université d'Orléans.

[‡] ADIR et Cie.

[§] Université Paul Sabatier.

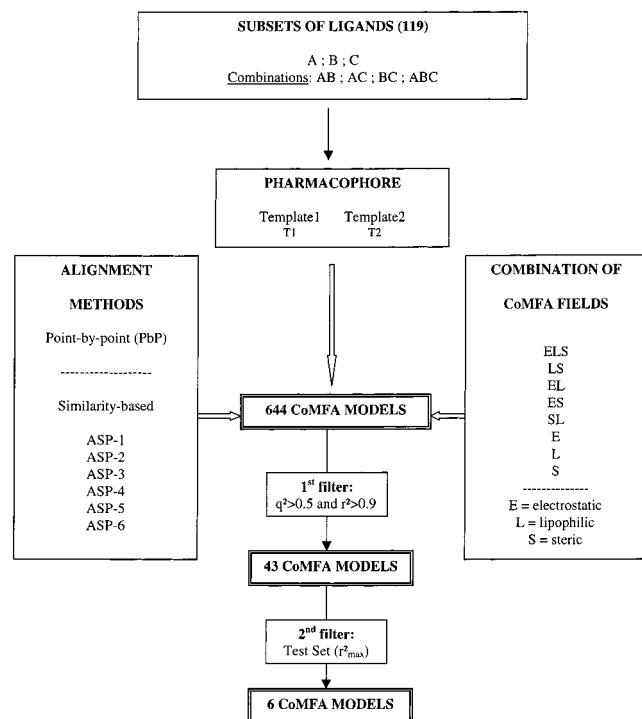


Figure 2. Methodology of the 3D-QSAR CoMFA approach.

i.e the ligands and their biological activities, to direct the design of new bioactive compounds. In the case of the IBS, as no tridimensional structure is available, the comparative molecular field analysis (CoMFA)²² is a valuable tool to study the tridimensional quantitative structure–activity relationships (QSAR) within a set of aligned ligands: this tridimensional grid-based QSAR technique tries to correlate variations of their biological activities with variations of their electrostatic, lipophilic, and steric fields. If a validated QSAR is obtained, it is first possible to visualize the results graphically, i.e the regions where a variation of the electrostatic, lipophilic, or steric potentials has been correlated with a variation of the biological activity. This qualitative information gives the key features on the ligand–receptor interaction, but much care has to be taken on their extrapolations on the topology of the receptor. Finally, central to the goals of such a validated CoMFA model is the ability to predict the activity of new chemical structures in an effort to guide the synthesis of new bioactive compounds. As soon as enough I₂ ligands were synthesized and tested, classical CoMFA technique gave the first validated models^{23,24} as well as their associated physico-chemical requirements for the I₂ affinity.

In this paper are reported further insights into the I₂ scope which result from a global approach (Figure 2) combining 3D-QSAR tools and methods previously used in our group.^{25,26} Indeed, aiming at optimizing the whole 3D-QSAR CoMFA process, we have explored not only the molecular diversity of the I₂ ligands but also the key steps of the CoMFA methodology such as the choice of the pharmacophore or the alignment methods. We therefore succeeded in obtaining new models which, in addition, were further evaluated using external sets of ligands: the comparison of their predicted and measured activities constitutes the first external test of predictivity ever applied to I₂ models. Closely related to the applied use of such models, these ultimate steps

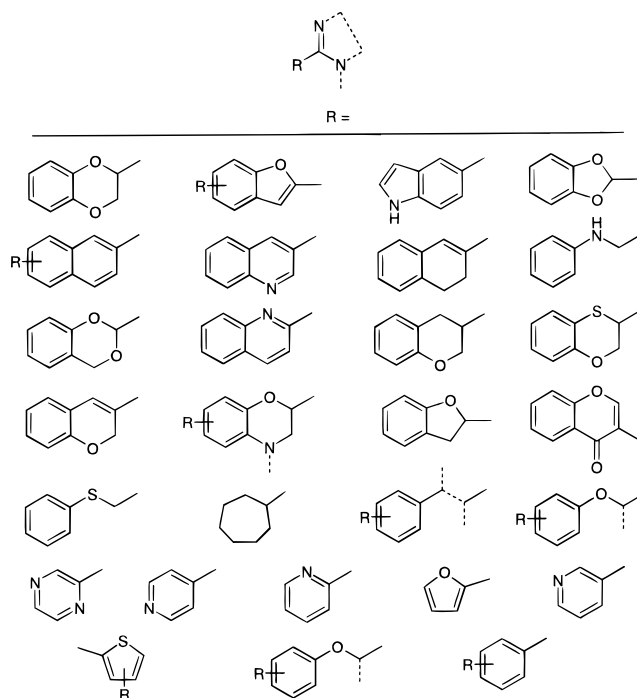


Figure 3. Definition of the 28 classes of R moiety describing the I₂ ligands investigated in this study.

point out the whole work reported here as a major breakthrough in the I₂ scope through the development of new I₂ ligands endowed with better affinity.

Materials and Methods

Selection of I₂ Ligands and Binding Affinities. In the present study, 119 compounds were finally selected from literature and our databases, and each had their I₂ binding affinity specifically expressed in terms of pK_i as a result of [³H]Idazoxan binding studies on rabbit kidney homogenates.²⁷ As reported in Figure 3, they belong to 28 families which differ in the moiety linked to the imidazoline cycle. To apply the external test of predictivity, the original pool was first divided into training and test sets; then, second, we have discerned three different groups with respect to their structure and the type of alignment pattern (A, B, or C). Thus, training set A (Table 1) and test set A (Table 4) have been obtained respectively constituted of 26 and 18 compounds, as well as training set B (Table 2) and test set B (Table 5) respectively constituted of 23 and 28 compounds. Both subsets A and B come from our groups,²⁸ whereas subset C, which presents a third different moiety, was found in the literature.²⁴ Due to the low availability of ligands of the C type, training set C contained only 24 compounds (Table 3). This subdivision of the global set of 119 I₂ ligands will enable us to explore the I₂ 3D-QSAR from finite structure systems (A, B) to more complex systems through combination of the subsets (A–B, A–C, B–C, A–B–C). In addition, it will make the external test of predictivity, which is the ultimate test set step, easier to interpret. Subset C was not used to get C type I₂ CoMFA models, since it has already been successfully done,^{23,24} but it was interesting to combine it with subsets A and B for structural diversity reasons. The building of each training set was handmade in order to optimize both structural diversity and homogeneous pK_i distribution, the remaining ligands constituting the test set. The pK_i distribution is reported in Figure 4 for the global set. The homogeneity of the pK_i with respect to both the experimental protocol and the distribution as well as the structural diversity are essential prerequisites of the training and test sets for the validity of potential results and their meaningful interpretation. Finally, special mention has to be made about 38 of the 119 ligands, which have one chiral center: whatever the sources, binding experiments were

Table 1. p*K_i* Values for the 26 Compounds of the Training Set: Alignment Pattern Type A

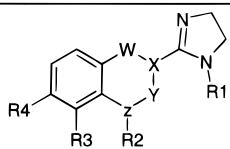
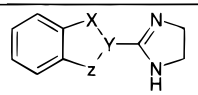
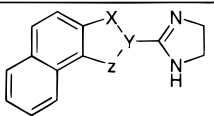
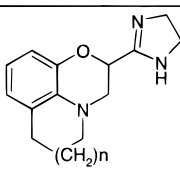
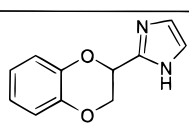
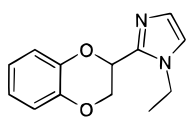
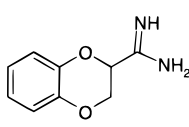
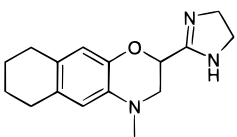
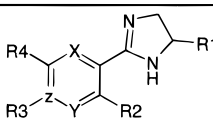
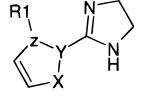
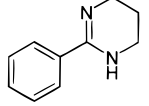
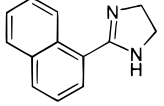
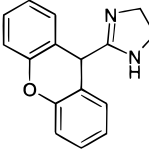
									
Cpds	W	X	Y	Z	R1	R2	R3	R4	p <i>K_i</i>
A1	O	CH	CH2	N	H	H	H	H	7.3
A2	O	CH	CH2	N	H	Me	H	H	8.1
A3	O	CH	CH2	N	H	Me	H	Me	8.7
A4	O	CH	CH2	N	H	Me	Me	H	6.3
A5	O	CH	CH2	N	H	Me	H	Cl	6.8
A6	O	CH	O	CH	H	H	H	H	8.0
A7	O	CH	CH2	N	H	benzyl	H	H	5.8
A8	O	C-(n-propyl)	CH2	N	H	Me	H	H	5.1
A9		CH=C		CH=C	H	H	H	H	9.2
A10		CH=C		CH=C	H	H	H	O-Me	8.0
A11		CH=C		CH=N	H	-	H	H	8.5
A12		N=C		CH=C	H	H	H	H	8.8
									
		X		Y		Z			
A13		O		CH		O			7.1
A14		O				C=CH			8.8
									
		X		Y		Z			
A15		O		CH		CH2			6.2
A16		O				C=CH			8.6
									
					n=0				
A17					n=0				7.5
A18					n=1				8.0
A19					n=2				6.2
									
A20					5.5	A24			6.2
									
A21					5.0	A25			6.1
									
A22					6.8	A26			8.6
									
A23					6.1				

Table 2. pK_i Values for the 23 Compounds of the Training Set: Alignment Pattern Type B

								
Cpds	X	Y	Z	R1	R2	R3	R4	pK_i
B1	CH	CH	C	H	H	H	H	7.4
B2	CH	CH	C	H	H	Me	H	8.1
B3	CH	CH	C	H	H	phenyl	H	8.0
B4	CH	CH	C	H	H	n-butyl	H	6.8
B5	CH	CH	C	H	H	O-Me	H	8.3
B6	CH	CH	C	H	H	O-CF ₃	H	7.1
B7	CH	CH	C	H	H	O-phenyl	H	7.4
B8	CH	CH	C	H	H	H	Me	8.3
B9	CH	CH	C	H	H	H	O-Me	7.8
B10	CH	CH	C	H	Me	H	H	8.0
B11	CH	CH	C	H	S-Me	H	H	5.0
B12	CH	CH	C	H	H	F	Cl	6.8
B13	CH	CH	C	H	F	H	Me	8.2
B14	CH	CH	C	Me	H	H	H	5.4
B15	CH	CH	N	H	H	-	H	5.2
B16	CH	N	C	H	H	H	H	5.9
B17	N	CH	C	H	H	H	H	6.2

				
X	Y	Z	R1	
B18	S	C=C	H	6.4
B19	S	C=C	Me	7.4
B20	CH=C	O	-	5.6
B21		6.2	B23	8.8
				
B22		5.9		
				

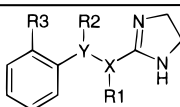
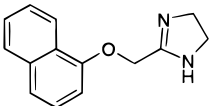
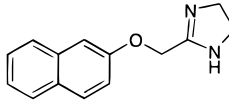
carried out on racemic ones, and hence, a unique pK_i is associated with the R/S couple. As will be seen in the Alignment Methods section, this problem was handled by selecting, for the model derivation, the enantiomer of the R/S couple which is the most likely to be bioactive. That means that, for each R/S couple, the two enantiomers were modeled and submitted to the conformational analysis, then the alignment. That is to say that 157 structures were finally handled.

Molecular Modeling. The similarity-based alignment was performed with ASP V3.2²⁹ implemented in TSAR V3.1²⁹ running on a R4600 Silicon Graphics workstation. All the other modeling techniques were applied on R4400 and R10000 Silicon Graphics workstations running Sybyl 6.5³⁰ in which were implemented the MOPAC 6.0 semiempirical package³¹ for the AM1 charge calculation, the CLIP module³² for the calculation of the molecular lipophilic potential (MLP),³³ and the Sybyl QSAR module for the CoMFA analysis.

Conformational Analysis. Since the imidazoline, imidazole, or amidine groups should be protonated at physiological pH, and former CoMFA studies on unprotonated I₂ ligands were unsuccessful,²⁴ the 3D models of the ligands were built protonated. To better take into account the mesomery of these protonated structures (Figure 5), each of them was then optimized with MOPAC using the AM1 Hamiltonian, a net charge of +1, and a "full" and "precise" optimization. Display-

ing Coulson charges, from the previous AM1 calculation, and a fixed planar conformation of their protonated group, each ligand was then subjected to a conformational analysis through a random search process taking into account all the torsion angles. This process involves many cycles of random torsion changes and minimization of the resulting structures using Tripos force field: a comparison test of the diverse minimized structure enables to finally obtain a sampling of unique minima. The main parameters of the random search process are the rms threshold (root mean square = 0.2 Å) which defines the maximum rms difference between two conformations before they are considered as different, the minimum hits ($n = 10$) needed before recording a conformation as unique, the energy cutoff ($E = +70$ kcal/mol) which defines the maximum allowable energy for a minimized conformation, and the maximum cycles ($c = 500$) of perturbation-minimizations if the minimum hits criterion has not been met for every minima. Once more, for a closest relevance to their mesomery, all the generated minimized conformations were fully reoptimized with the semiempirical package MOPAC using the AM1 Hamiltonian. As will be seen in the 3D Pharmacophore Investigation section, particular attention has been paid to the conformational analysis of the template. Indeed, both random search and grid search were used to screen its conformational space. Suitable, since the template only has one torsionable

Table 3. p*K_i* Values for the 24 Compounds of the Training Set: Alignment Pattern Type C

							
Cpds	R1	X	Y	R2	R3	p <i>K_i</i>	
C1	H	CH	O	-	allyl	8.9	
C2	H	CH	O	-	benzyl	6.7	
C3	H	CH	O	-	cyclopropyl	8.4	
C4	H	CH	O	-	i-propyl	8.7	
C5	H	CH	O	-	ethyl	8.6	
C6	H	CH	O	-	propyl	8.2	
C7	H	CH	O	-	Me	9.1	
C8	H	CH	O	-	O-(i-propyl)	7.1	
C9	H	CH	O	-	phenyl	6.5	
C10	H	CH	O	-	H	9.0	
C11	H	CH	O	-	s-butyl	7.3	
C12	H	CH	O	-	t-butyl	6.7	
C13	H	CH	CH	H	H	8.6	
C14	H	CH	CH	phenyl	H	5.5	
C15	H	CH	S	-	H	7.3	
C16	H	CH	N	H	H	7.5	
C17	H	C=C		H	H	8.7	
C18	Me		O	-	i-propyl	5.3	
C19	Me		O	-	Me	6.7	
C20	Me	CH	O	-	O-(i-propyl)	5.6	
C21	Me	CH	O	-	H	5.6	
C22	Me	C=C		H	H	8.0	
C23			8.7	C24			8.4

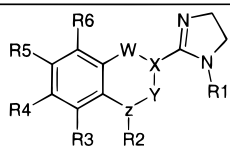
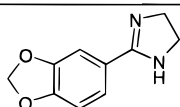
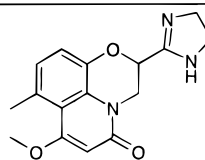
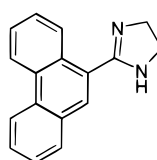
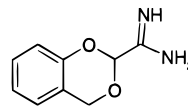
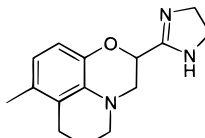
bond (Figure 6), the grid search was parametrized as follows: Due to its symmetry, the N3–C2–C6–C7 torsion bond was only explored from 0° to 180° with a 10° increment. The force field used was Tripos and the charges considered were of Coulson type. Finally, as for the random search process, all the grid search-generated conformers were finally reoptimized with the semiempirical package MOPAC using the AM1 Hamiltonian.

3D Pharmacophore Investigation. Measured when the receptor–ligand system is at equilibrium, the I₂ p*K_i* reflects the thermodynamic aspect of the binding process and, thus, is mainly connected to the bioactive conformation of the ligand within the I₂ site. Nevertheless, no experimental data concerning ligands bound to I₂ receptor exists, and the proposition of bioactive conformers is thus based on the pharmacophore concept. In this context, we dispose of one ligand endowed with great I₂ affinity and rigid moiety, namely the benzazoline (Figure 6), which was already used for the I₂ CoMFA process in previous works.²⁴ Benzazoline only has one torsion bond, and hence we decided to thoroughly explore its conformational space with both random search and grid search techniques. Two distinct minima (T1, T2) were conjointly found: T1 displays a dihedral angle of 38°, whereas T2 displays a dihedral angle of 142° (Figure 7). Both have an AM1 heat of formation (ΔH_f°) of 208.5 kcal·mol⁻¹. As these two conformers are not superposable and, as a result, would be distinct in an anisotropic environment, we decided to go on the CoMFA methodology with the two templates. In addition, no experimental ligand–I₂ receptor data being available, we could not exclude any of the two templates. Thus, T1 and T2 will each serve as a template for the selection of the assumed bioactive conformation of all the other I₂ ligands, through the further alignment process. The caveat of the above process is directly connected to the choice of the template which did not take the

receptor into account, since it was rationalized on basis of energy calculation in vacuum.

Alignment Methods. Both point-by-point and similarity-based methods were applied in this study. For each I₂ ligand considered, both methods finally result in a superposition of one assumed bioactive conformer to the template. The bioactive feature of this conformation directly refers to the template conformation, and thus, the superposition of the conformation to the fixed template one has to be optimized. This is the point where the methods differ. First, the point-by-point method superimposes the two structures on their common backbone and quantifies the superposition quality with the rms (root mean squares of the distances between the nuclei of the chosen atoms). The similarity-based method superimposes them, using various techniques, and quantifies the superposition with a similarity index. Second, the point-by-point method optimized the superposition through a rms-based selection among all the AM1-optimized conformers, which were generated by the Random Search conformational analysis. The similarity-based method directly optimized the superposition from an initial conformation of the ligand whose similarity index to the template was maximized. Another goal of the alignment concerned 38 *R/S* couples of our I₂ ligands from which one of the enantiomers had to be selected. We selected the enantiomer endowed with the best rms-fitting or similarity index. The details of the point-by-point method were as follows (Figure 6): As the templates have a common backbone with the three subsets, subset A fit was realized according to the 1–2–3–4–5–6–7–8–9–10–12–13 benzazoline profile, subset B according to the 1–2–3–4–5–6–7–8–13 one, and finally subset C according to the 1–2–3–4–5–6–7–8–9 one. The fitting process was automatized using an “in-house” developed method written in SPL (Sybyl programming language). Due to the mesomery of the imidazoline and its related cycles, both

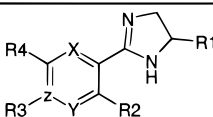
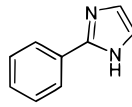
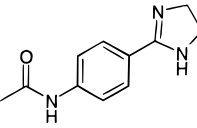
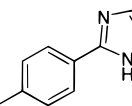
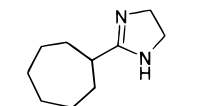
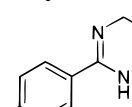
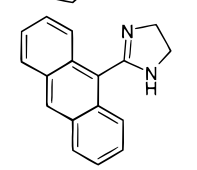
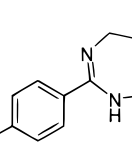
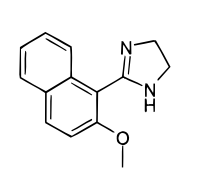
Table 4. pK_i Values for the 18 Compounds of the Test Set: Alignment Pattern Type A

											
Cpds	W	X	Y	Z	R1	R2	R3	R4	R5	R6	p <i>K</i> _i
A27	O	CH	CH2	N	H	Me	H	H	Me	H	7.3
A28	O	CH	CH2	N	H	Me	H	F	H	H	7.2
A29	O	CH	CH2	N	H	propyl	H	H	H	H	6.8
A30	O	CH	CH2	N	H	Me	H	H	H	Me	7.3
A31	O	CH	CH2	N	H	Me	H	O-Me	H	H	7.2
A32	O	CH	CH2	O	H	-	H	H	H	H	8.1
A33	O	C-Me	CH2	N	H	Me	H	H	H	H	5.0
A34	CH2	CH	CH2	O	H	-	H	H	H	H	7.8
A35		CH=C	CH2	O	H	-	H	H	H	H	8.4
A36		CH=C	CH2	CH	H	H	H	H	H	H	8.0
A37	C=O		C=CH	O	H	-	H	H	H	H	5.7
A38	O	CH	CH2	S	H	-	H	H	H	H	7.3
A39	O	C-Me	C=O	N	H	Me	H	H	H	H	5.1
A40					8.3	A43					4.7
A41					7.2	A44					6.4
A42					6.0						

directions of numbering were considered: for example, the conformational sample of an A type ligand was submitted to both 1-2-3-4-5-6-7-8-9-10-12-13 and 3-2-1-5-4-6-7-8-9-10-12-13 superposition rules before selection of the ligand endowed with the best fit. More over, when different directions of alignment were possible on the whole structure of the ligand, with respect to the naphthalenic nature of the template, they were all tried and compared. For isolated cases such as the **A25** and **A43** ligands, the best fit was selected. Nevertheless, subset B case was problematic since almost all these ligands can undergo the two directions of numbering without significant rms differences. We then decided to further divide B into subsets B α and B β with respect to the position of the substituent on the aligned ligand. Thus, B α ligands have an *ortho*-substituent and a *meta*-substituent respectively numbered as 7 and 14, whereas B β ligands have an *ortho*-substituent and a *meta*-substituent respectively numbered as 15 and 8. This arbitrary subclassification, where the *ortho*- and *meta*-substituents are each placed on a distinct side of the phenyl moiety, enabled us to better explore the *ortho*- and *meta*-associated regions by avoiding their potential overfitting. In addition, whatever B α or B β rules, the numbering directions of **B16**, **B17**, and **B20** were the same as their bicyclic active analogues **A11**, **A12**, and **A14**, whereas the five-membered ring ligands **B18** and **B19** were numbered to the best fit. Finally, all ligands C were aligned so that the substituent on the phenyl moiety was numbered as 9. All the above underground rules of alignment are valid since the main goal of the final

alignment is to describe the causes rather than the consequences of the variation of pK_i . As for the similarity-based method, it was applied as follows: For each I₂ ligand, we selected the conformation which was the most stable, on basis of the AM1 heat of formation (ΔH_f), among all the conformers AM1 optimized of the Random Search conformational analysis. This latter conformation was then submitted to the alignment onto the template, called the leader. The superposition was quantified with a combined similarity index, C_{AB} , defined as: $C_{AB} = (E_{AB} + L_{AB} + S_{AB})/3$, where E_{AB} , L_{AB} , and S_{AB} are respectively the electrostatic, lipophilic, and shape similarity indices. E_{AB} is a Carbo similarity index³⁴ where the electrostatic potential is the propriety being compared. The electrostatic potential $P(r)$ is expressed as: $P(r) = \sum_{i=1}^n q_i/|r - r_i|$, where n is the number of atoms in the molecule and r is the position of the point, the partial charge q_i coming from MOPAC calculation. The L_{AB} expression is similar to the E_{AB} one and compares the lipophilic potential, whose expression is the same as the electrostatic one, the partial log P_i ³⁵ being treated as the q_i . S_{AB} is also evaluated using the Carbo equation and compares the atomic electron densities. These three indices are calculated faster through an analytical method,³⁶ which fits the inverse distance dependence of the potentials, or the electron density functions, to an expansion of Gaussian functions. As for the optimization of the superposition, we combined four different techniques available in the software. The six different combinations applied are reported in Table 6. First,

Table 5. p*K_i* Values for the 28 Compounds of the Test Set: Alignment Pattern Type B

<div></div>								
Cpds	X	Y	Z	R1	R2	R3	R4	p <i>K_i</i>
B24	CH	CH	C	Me	H	Me	H	7.0
B25	CH	CH	C	H	H	C-F ₃	H	7.3
B26	CH	CH	C	H	H	ethyl	H	7.1
B27	CH	CH	C	H	H	i-propyl	H	7.1
B28	CH	CH	C	H	H	t-butyl	H	6.3
B29	CH	CH	C	H	H	propyl	H	7.2
B30	CH	CH	C	Me	H	H	Me	6.8
B31	CH	CH	C	H	H	H	F	7.6
B32	CH	CH	C	H	H	H	phenyl	7.6
B33	CH	CH	C	H	F	H	H	7.0
B34	CH	CH	C	H	phenyl	H	H	5.0
B35	CH	CH	C	H	Cl	F	H	5.5
B36	CH	CH	C	H	Me	H	F	5.3
B37	CH	CH	C	H	H	Me	F	8.5
B38	CH	CH	C	H	H	H	ethyl	7.5
B39	CH	CH	C	H	H	S-Me	H	8.4
B40	CH	CF	C	H	H	H	F	7.1
B41	CH	C-(t-butyl)	C	H	H	OH	t-butyl	6.6
B42	C-Me	CH	C	H	H	H	CN	8.3
B43	N	N	C	H	H	H	H	5.1
B44	<div></div>		5.9	B48	<div></div>		5.0	
B45	<div></div>		6.6	B49	<div></div>		7.4	
B46	<div></div>		5.0	B50	<div></div>		6.0	
B47	<div></div>		5.0	B51	<div></div>		5.0	

the *simple alignment* fits the molecules using a function based on their shapes and electrostatic charges. Second, the *full rigid search* considers all the orientations of the molecule onto the fixed leader. The molecule is exhaustively rotated around orthogonal axes centered on the centroid of the template. The orientation with the best-combined similarity index C_{AB} is stored. Third, the *rigid optimization* uses the Simplex algorithm³⁷ to direct rigid rotations and translations toward a fitting endowed with a high C_{AB} . Six variables were entered into the algorithm: the XYZ rotations and XYZ translations. Fourth, the *flexible optimization* uses the Simplex algorithm to direct random rigid rotations and translations as well as torsional rotations. The XYZ rotations and XYZ translations as well as all the torsional rotations were entered into the algorithm. We chose to consider all the rotatable single bonds as well as the double and amide bonds. The other Simplex parameters were set as follows: The criterion of convergence was 0.001 and the maximum iterations set to 500. The contraction, expansion, and reflection scales were respectively

set to 0.5, 2, and 1. The last parameters concerns the *flexible optimization* technique in order to avoid a conformation endowed with a great C_{AB} but energetically unrealizable. Indeed, the C_{AB} is then weighted with a Boltzmann penalty factor and is expressed as: $C_{AB} = C_{AB} \times e^{\Delta E/RT}$, where c is a weighting factor set to 0.01 and ΔE is the energy difference between the initial and new conformations. The energy was calculated using the COSMIC force field.³⁸

Technical Specifications of the CoMFA Study. Each alignment was submitted to the CoMFA analysis for each combination of the electrostatic, lipophilic, and steric molecular fields. The steric and electrostatic interaction energies, respectively calculated with the Lennard-Jones and Coulomb potential expressions, are sampled on grid points of a 3D-lattice with a sp³ carbon probe endowed with a charge of +1. The resolution of the lattice is 2.0 Å. The AM1 charges from the MOPAC calculation were used as well as a 1/ r dielectric function. The field's values were neither *smoothed* nor *dropped* but a 30 kcal/mol energy cutoff was applied. As for the

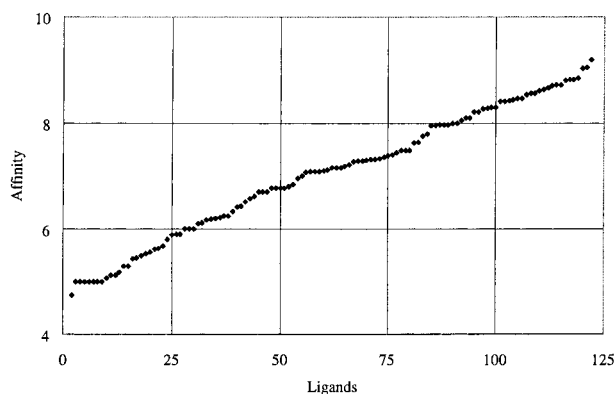


Figure 4. Distribution of affinity ($pK_i I_2$) of all ligands investigated.

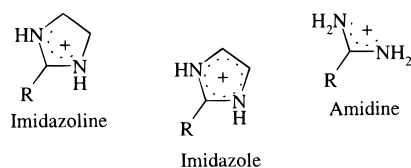


Figure 5. Imidazoline, imidazole, and amidine groups.

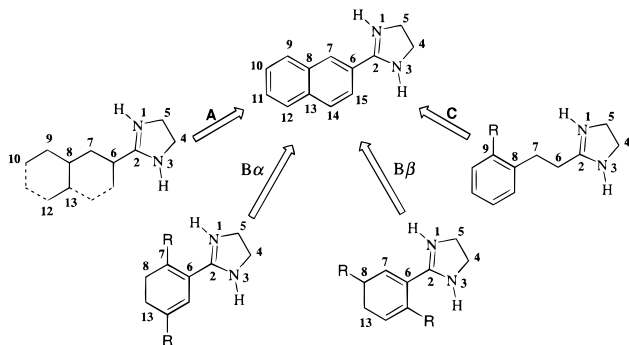


Figure 6. Benazoline ($pK_i I_2 = 9.2$) displaying the numbering on which the PbP alignment rules are based: A rule for A ligands, B α and B β rules for B ligands, C rule for C ligands.

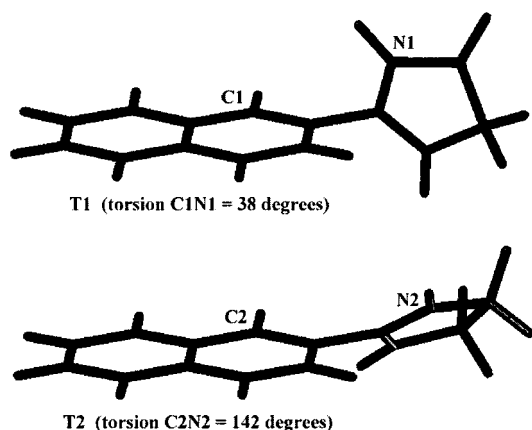


Figure 7. Pharmacophores: templates T1 and T2.

sampling of the lipophilic interaction energies, we used the molecular lipophilic field expression (MLP) through the CLIP module. The PLS (partial least-squares) analysis produces model equations which explain the variance of the pK_i data in terms of linear combinations of the sampled molecular fields, extracted into components (latent variables). First, to avoid overfitted 3D-QSARs, the optimum number of components N to use in the model derivation is chosen from the analysis with the highest q^2 . The cross-validated r^2 (q^2) quantifies the

Table 6. Parametrization of the Six ASP-Based Alignment Methods^a

ASP Methods	Simple alignment	Full rigid search	Simplex optimization
ASP-1	Yes	No	No
ASP-2	Yes	No	Rigid optimization
ASP-3	Yes	No	Flexible optimization
ASP-4	Yes	Angle increment = 10°	No
ASP-5	Yes	Angle increment = 10°	Rigid optimization
ASP-6	Yes	Angle increment = 10°	Flexible optimization

^a See the Materials and Methods section for the common parametrization.

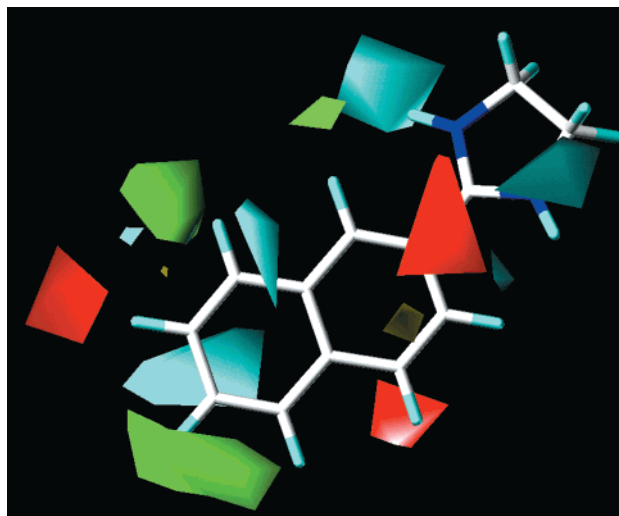


Figure 8. Steric and lipophilic graphical results of PLS model M43 with template T2 (red = steric-decreased affinity; green = steric-increased affinity; cyan = hydrophilicity-increased affinity; yellow = lipophilicity-increased affinity).

predictive ability of the model and was determined in this study by a "leave-one-out" (LOO) procedure of cross-validation in which each compound is excluded once, from the model derivation, so that its pK_i can be predicted. Once more, to avoid overfitted models, the number of components to be extracted from the LOO run was limited to 6, 5, and 10 for respectively training sets A, B, and others. In addition, we omitted from the analysis the lattice points whose energy variation was less than 2 kcal/mol. Second, the CoMFA model was derived through a PLS analysis using the optimal number of components (N), previously determined, as well as all the lattice points. The model can now be used for the determination of both the training and test sets' pK_i . From the training set, the robustness of this model is internally evaluated with the calculation of the r^2 , s , and F test values. It is externally evaluated with the calculation of the r^2 from the test set. Finally, from such a fully validated model, the CoMFA color contour maps are derived for the electrostatic, lipophilic, and steric fields.

Interpretation of the CoMFA Contour Maps. The graphic interpretation of a fully validated CoMFA model is reported in Figure 8, on which the most relevant regions of the space (isocontours), with the largest variations of a given statistical field, are plotted. The color code used to characterize the isocontours for each field signal is described in Table 7. Concerning the electrostatic field, a white zone can mean a favorable influence of some electron deficiency or an unfavorable influence of some high electron density. The interpretation of the lipophilicity field is similarly ambiguous. Indeed, lipophilicity encodes two major structural contributions,³⁹ namely a bulk term reflecting hydrophobic and dispersive forces and a polar term reflecting more directional electrostatic interac-

Table 7. Color Code of Graphical CoMFA Results

Field or component of field	Increase affinity	Decrease affinity
Steric (S)	Green	red
Electrostatic (E)		
Positive charge	White	magenta
negative charge	Magenta	white
Lipophilic (L)		
lipophilic component	Yellow	cyan
hydrophilic component	Cyan	yellow

tions and hydrogen bonds. The favorable influence of the hydrophobic and hydrophilic components is described by yellow or cyan colors, respectively. By examining the moieties and functionalities involved, a chemist should be able to rely on common sense to remove the ambiguity in interpreting the influence of the electrostatic and lipophilicity fields.

Results and Discussion

Selection and Validation of the CoMFA Model.

Central to the goal of this study was the design of new compounds endowed with a great I₂ affinity and, as a result, the design of the best CoMFA model in terms of predictivity. In this context, this global approach (Figure 2) was developed, where all the steps of the CoMFA methodology that could potentially improve the predictivity of the model were taken into account. The subdivision of the original pool of I₂ ligands into structural families was decided in order to explore the impact of the structural diversity onto the CoMFA results. For the same reasons, it was decided to apply the point-by-point (PbP) alignment and some variants of the similarity-based (ASP) alignment as well as, more classically, to combine the electrostatic, lipophilic, and steric fields. In addition, we also considered the elements that could not be eluded on sound evidence. That is why we decided, no data on the I₂ receptor structure being available yet, to consider the two conformers T1 and T2 as templates. So, 644 CoMFA models were produced as a basis to the rest of our work. Next comes the notion of filter to retain the best predictive model.

First Filter Analysis. The first filter selected models with respect to their q^2 and r^2 , which respectively quantify their predictive ability, internally evaluated by the LOO method, and their robustness. We set the filter to a q^2 and r^2 of at least 0.5 and 0.9, respectively; 43 models were obtained (Table 8). The first striking fact is that, for each training set type, several models were obtained. More precisely, for subset A, twelve models including seven ASP and five PbP models were found: three T1-based ASP models, including M3 which reaches a q^2 of 0.59 through the ASP-6 alignment method; four T2-based ASP models, including M5 which reaches a q^2 of 0.59 through ASP-4; and five T2-based PbP models, including M12 which reaches a q^2 of 0.69. M3, M5, and M12 are based on different combinations of molecular fields, respectively, the lipophilic, electric-steric, and steric ones. For subset B, sixteen models including fifteen ASP and one PbP models were found: seven T1-based models, including M14 which reaches a q^2 of 0.61 through the ASP-2 method and is based on an electric-lipophilic field combination; eight T1-based ASP models,

Table 8. Results of the First Filter*

Models	Training Set	Templates	Alignment	q^2 ^a	N ^b	r^2 ^c	S ^d	F ^e	E ^f	L ^f	S ^f
Set		Method									
M1	A	T1	ASP-2	0.52	6	0.98	0.23	130	34	66	
M2			ASP-5	0.52	5	0.97	0.24	139	40		60
M3			ASP-6	0.59	6	0.98	0.18	199		100	
M4		T2	ASP-4	0.57	6	0.99	0.09	777	20	42	38
M5				0.59	4	0.96	0.28	126	29		71
M6				0.55	2	0.91	0.40	116		52	48
M7			ASP-5	0.51	3	0.92	0.38	85		56	44
M8			PbP	0.61	6	0.99	0.12	433	27	40	33
M9				0.56	6	0.99	0.13	386	38		62
M10				0.62	6	0.99	0.18	217		54	46
M11				0.50	6	0.98	0.18	200	35	65	
M12				0.69	5	0.98	0.18	246			100
M13	B	T1	ASP-2	0.54	4	0.95	0.27	88	24	52	54
M14				0.61	5	0.95	0.29	62	30	70	
M15				0.57	3	0.94	0.35	85		100	
M16			ASP-3	0.56	3	0.94	0.31	89	14	64	22
M17				0.51	5	0.96	0.28	67	28		72
M18				0.60	3	0.92	0.36	66		75	25
M19		T2	ASP-6	0.50	5	0.94	0.32	50	30		70
M20			ASP-1	0.67	4	0.97	0.21	142	32	48	20
M21				0.71	4	0.96	0.24	115	38	62	
M22				0.65	4	0.94	0.31	67		67	33
M23				0.73	4	0.94	0.30	72		100	
M24			ASP-2	0.52	4	0.92	0.35	51	21	52	27
M25			ASP-3	0.55	5	0.99	0.15	246	23	49	28
M26				0.52	5	0.98	0.19	141	32	68	
M27		A+B	ASP-6	0.51	3	0.90	0.39	56		100	
M28			PbP- α	0.55	5	0.94	0.33	46	27		73
M29			ASP-2	0.53	7	0.95	0.29	104	36	64	
M30			ASP-5	0.56	5	0.94	0.31	131	30	40	30
M31			ASP-6	0.55	6	0.96	0.26	150	24	46	30
M32				0.54	6	0.93	0.32	100	29	71	
M33	A+C	T2		0.52	6	0.92	0.35	83		100	
M34			PbP	0.53	4	0.90	0.41	104	16	48	36
M35				0.53	6	0.92	0.37	89	26		74
M36				0.59	7	0.97	0.24	185		52	48
M37				0.58	7	0.96	0.27	144			100
M38			PbP- α	0.51	7	0.95	0.30	95			100
M39	B+C	T2	PbP- β	0.53	10	0.99	0.14	318	21	42	37
M40			PbP- β	0.50	10	0.98	0.21	138	30		70
M41			ASP-1	0.50	5	0.90	0.39	121	24	50	26
M42	A+B+C	T2	PbP- α	0.50	6	0.91	0.40	87			100
M43			PbP- β	0.50	9	0.96	0.28	148		51	49

* For each training set, the models were selected with regards to their $q^2 > 0.5$ and $r^2_{\text{training set}} > 0.9$. ^a Cross-validation correlation coefficient. ^b Number of components, used in final PLS analyses, corresponding to the first maximum of the function $q^2 = f(N)$ in cross-validation analyses. ^c Correlation coefficient of the final PLS analysis. ^d Standard error of estimate measuring the unexplained uncertainty. ^e F ratio; the higher the F ratio, the better the PLS analysis. ^f Relative contributions of the steric (S), electrostatic (E), and lipophilic (L) field in the final PLS analysis.

including M23 which reaches a q^2 of 0.73 through the ASP-1 method and is based on a lipophilic field; and one T2-based PbP model, M28, which reaches a q^2 of 0.55 through the α rule of alignment and is based on an electric-lipophilic field combination. For subset AB, five T1-based ASP models were found: M30 reaches a q^2 of 0.56 through the ASP-5 method and is based on an electric-lipophilic-steric field combination. For subset AC, four T2-based PbP models were found, including M36, which reaches a q^2 of 0.59 and is based on a lipophilic-steric field combination. For subset BC, three T2-based PbP models were found, including M39 which reaches a q^2 of 0.53 through the β rule of alignment and is based on an electric-lipophilic-steric field combination. For subset ABC, three T2-based PbP models were found, including M43, which reaches a q^2 of 0.50 through the β rule of alignment and is based on

Table 9. Results of the Second Filter*

Selected	Training	Template	q^2	r^2	r^2	r^2
model	set type	Alignment method		training set	A test set	B test set
M2	A	T1-ASP5	0.52	0.97	0.51	-
M37	AC	T2-PbP	0.58	0.96	0.62	-
M27	B	T2-ASP6	0.51	0.90	-	0.30
M38	BC	T2-PbP α	0.51	0.95	-	0.19
M32	AB	T1-ASP6	0.54	0.93	0.25	0.33
M43	ABC	T2-PbP β	0.50	0.96	0.80	0.81

* For each training set, the best model was selected with regards to its $r^2_{\text{test set}}$.

a lipophilic–steric field combination. If we consider the alignment methods, we can say that, applied to subsets A, B, and A+B, the ASP alignments are very effective to find correlations, whereas the PbP alignment is limited to subset A. This can be explained by the fundamental difference between ASP and PbP alignments with respect to the ligand–receptor binding. Indeed, whereas ASP aligns the ligands taking into account electrostatic, steric, and shape features, three elements clearly identified as critical in the binding interactions, the PbP only fits onto the backbones. As a result we can claim that the ASP alignment is by nature more relevant to the affinity values and, that way, a correlation is more easily obtained. Contrary to A and B ligands, C ligands are problematic for ASP alignments since, for every subset including the C ligands, the PbP supremacy is highlighted. We can suppose here that the flexibility of these ligands is poorly handled by the ASP method which, as was observed, fitted similar C molecules in a different way. That latter fact points out the reproducibility of an alignment method as a critical feature for the predictive ability of the model. Finally, neither variant of the ASP alignment nor B rules of alignment (α/β) were highlighted since each was effective for correlation. Concerning the templates, it is striking to note that whereas T1- and T2-based ASP models were obtained, all the selected PbP models are based on T2, the conformer of the benazoline which already gave I_2 CoMFA results in previous studies.^{23,24} On the whole, we note that, first, the q^2 value decreases when the molecular diversity increases. Second, the fact that all the field combinations are represented emphasizes the electrostatic, lipophilic, and steric fields as complementary binding descriptors for the description of the I_2 binding. This latter statement was already reported in I_2 related previous studies²³ as well as in other works.

Second Filter Analysis. Although this prolific production of models is interesting on the one hand, on the other hand it is problematic since we did not know what model to use in order to explore the I_2 3D-QSARs or predict the affinity of a new compound. We then used the test set procedure as the second filter. This last selection is based on an external test of predictivity, of more practical relevance, which retained, for each combination of subsets, the best model. This externally assessed predictive ability is expressed as the r^2 , correlation coefficient of the predicted pK_i versus experimental pK_i . The results, reported in Table 9, were

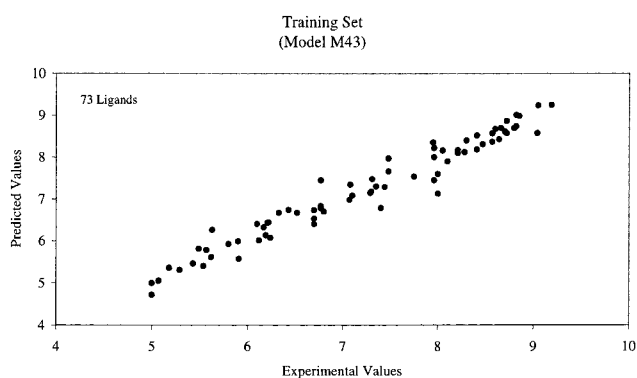
obtained by removing, from test set A, compound **A37**, which was generally very poorly predicted, as it was in previous CoMFA works.²³ We will discuss later compound **A37** in detail. The first point we can notice is that the ASP and PbP methods are equally distributed for the six subsets. For subset A, the T1-based ASP-5 model M2 displays a $r^2_{\text{test set}}$ (0.51) similar to the q^2 (0.52). For subset B, the T2-based ASP-6 model M27 displays a largely lower $r^2_{\text{test set}}$ (0.30) compared with its q^2 (0.51). For subset AB, the T1-based ASP-6 model M32 displays largely lower $r^2_{\text{test set}}$ (0.25 and 0.33 for respectively test sets A and B) compared with the q^2 (0.54). All these finally selected variants of the ASP alignment are based on a full combined optimization (*simple alignment + full rigid search + simple optimization*), but only one, M2 for subset A, is valid in terms of predictive ability. For subset AC, the T2-based PbP model M37 displays a $r^2_{\text{test set}}$ (0.62) similar to the q^2 (0.58). Interestingly, M37 and M2, both valid with respect to the prediction of subset A, are based on different templates, respectively T2 and T1. For subset BC, the T2-based PbP- α model M38 displays a $r^2_{\text{test set}}$ (0.19) largely lower than the q^2 (0.51). Finally, for subset ABC, the T2-based PbP- β model M43 displays $r^2_{\text{test set}}$ (0.80 and 0.81 for respectively test sets A and B) much larger than the q^2 (0.50). On the whole, the structural diversity is likely to reinforce the predictive ability, despite the parallel decrease of the q^2 . Indeed, the predictive ability, $r^2_{\text{test set}}$, for test set A is reinforced from 51% (M2, training set A) to 62% (M37, training set AC) and 80% (M43, training set ABC). But the result is spectacular for test set B since it reaches 80% (M43), whereas it was not previously predicted in valid terms (<35% for training sets B, BC, and AB). M43 was finally selected since it combines the structural diversity (73 ligands of three structural families), the robustness ($r^2 = 0.96$), and the best externally assessed predictive ability ($r^2_{\text{test set}} = 0.81$ for global test set AB). Indeed, these three latter features are of upmost importance for the practical application of a CoMFA model to direct synthesis of new compounds toward a better I_2 affinity. In connection to the M43 supremacy, the importance of both the T2 template and the β rule of alignment has to be noticed and, beyond this particular combination, the interest to apply a global methodology to the CoMFA process is confirmed. The pK_i values predicted by model M43, as well as the experimental pK_i and the associated residual values, are given for the training and test sets in respectively Tables 9 and 10. These results are graphically represented in Figures 9–11. With regard to Figure 10, we noticed that the best predicted products are the ones endowed with the best affinity. We suppose this is related to the fact there are many physicochemical features responsible for a lack of I_2 affinity. As a result, a validated CoMFA model is likely to better describe the binding features and, thus, to better predict the compounds endowed with good affinities. This latter point, which was already reported for CoMFA studies on melatoninergic ligands,²⁶ once more underlines the interest of M43 for practical applications. To come back to compound **A37**, a reason we may advance for its general very poor prediction is the absence of any physicochemical information in the training set concerning a carbonyl at that position. More precisely, we

Table 10. Experimental versus Predicted p*K*_i Values for the Training Set of Model M43

Cpds	Experimental Values	Predicted Values	Gap Values
A1	7.3	7.2	0.1
A2	8.1	7.9	0.2
A3	8.7	8.6	0.1
A4	6.3	6.7	0.4
A5	6.8	6.8	0.0
A6	8.0	8.0	0.0
A7	5.8	5.9	0.1
A8	5.1	5.1	0.0
A9	9.2	9.3	0.1
A10	8.0	8.4	0.4
A11	8.5	8.3	0.2
A12	8.8	9.0	0.2
A13	7.1	7.1	0.0
A14	8.8	8.7	0.1
A15	6.2	6.1	0.1
A16	8.6	8.4	0.2
A17	7.5	8.0	0.5
A18	8.0	7.1	0.9
A19	6.2	6.1	0.1
A20	5.5	5.8	0.3
A21	5.0	4.7	0.3
A22	6.8	6.8	0.0
A23	6.1	6.4	0.3
A24	6.2	6.4	0.2
A25	6.1	6.0	0.1
A26	8.6	8.4	0.2
B1	7.4	6.8	0.6
B2	8.1	8.2	0.1
B3	8.0	7.6	0.4
B4	6.8	6.7	0.1
B5	8.3	8.1	0.2
B6	7.1	7.4	0.3
B7	7.4	7.3	0.1
B8	8.3	8.4	0.1
B9	7.8	7.5	0.3
B10	8.0	7.5	0.5
B11	5.0	5.0	0.0
B12	6.8	7.5	0.7
B13	8.2	8.2	0.0
B14	5.4	5.5	0.1
B15	5.2	5.4	0.2
B16	5.9	5.6	0.3
B17	6.2	6.3	0.1
B18	6.4	6.7	0.3
B19	7.4	7.3	0.1
B20	5.6	6.3	0.7
B21	6.2	6.4	0.2
B22	5.9	6.0	0.1
B23	8.8	8.7	0.1
C1	8.9	9.0	0.1
C2	6.7	6.7	0.0
C3	8.4	8.2	0.2
C4	8.7	8.7	0.0
C5	8.6	8.6	0.0
C6	8.2	8.1	0.1
C7	9.1	9.2	0.1
C8	7.1	7.0	0.1
C9	6.5	6.7	0.2
C10	9.0	8.6	0.4
C11	7.3	7.2	0.1
C12	6.7	6.4	0.3
C13	8.6	8.7	0.1
C14	5.5	5.4	0.1
C15	7.3	7.5	0.2
C16	7.5	7.7	0.2
C17	8.7	8.6	0.1
C18	5.3	5.3	0.0
C19	6.7	6.5	0.2
C20	5.6	5.6	0.0
C21	5.6	5.8	0.2
C22	8.0	8.2	0.2
C23	8.7	8.9	0.2
C24	8.4	8.5	0.1

Table 11. Experimental versus Predicted p*K*_i Values for the Test Set of Model M43

Cpds	Experimental Values	Predicted Values	Gap Values
A27	7.3	6.8	0.5
A28	7.2	8.0	0.8
A29	6.8	6.5	0.3
A30	7.3	7.6	0.3
A31	7.2	6.8	0.4
A32	8.1	7.5	0.6
A33	5.0	6.0	1.0
A34	7.8	8.2	0.4
A35	8.4	8.6	0.2
A36	8.0	8.9	0.9
A37	5.7	9.2	3.5
A38	7.3	7.1	0.2
A39	5.1	6.5	1.4
A40	8.3	7.5	0.8
A41	7.2	7.4	0.2
A42	6.0	7.1	1.1
A43	4.7	5.6	0.9
A44	6.4	5.8	0.6
B24	7.0	7.6	0.6
B25	7.3	7.8	0.5
B26	7.1	8.0	0.9
B27	7.1	7.4	0.3
B28	6.3	7.0	0.7
B29	7.2	7.7	0.5
B30	6.8	6.9	0.1
B31	7.6	7.2	0.4
B32	7.6	7.4	0.2
B33	7.0	7.0	0.0
B34	5.0	6.4	1.4
B35	5.5	6.1	0.6
B36	5.3	4.1	1.2
B37	8.5	8.6	0.1
B38	7.5	7.9	0.4
B39	8.4	8.9	0.5
B40	7.1	7.6	0.5
B41	6.6	7.3	0.7
B42	8.3	8.2	0.1
B43	5.1	5.4	0.3
B44	5.9	5.6	0.3
B45	6.6	6.7	0.1
B46	7.2	7.6	0.4
B47	5.0	7.1	2.1
B48	5.0	5.7	0.7
B49	7.4	6.7	0.7
B50	6.0	6.9	0.9
B51	5.0	6.3	1.3

**Figure 9.** p*K*_i values predicted by model M43 versus experimental values for the 73 compounds of the training set.

were interested by the conformational induction of this carbonyl. Indeed, the interaction of the carbonyl and one NH of the imidazoline cycle constrains molecule **A37** in a planar conformation (dihedral angle of 170°). We may suppose that the distortion necessary to bring the

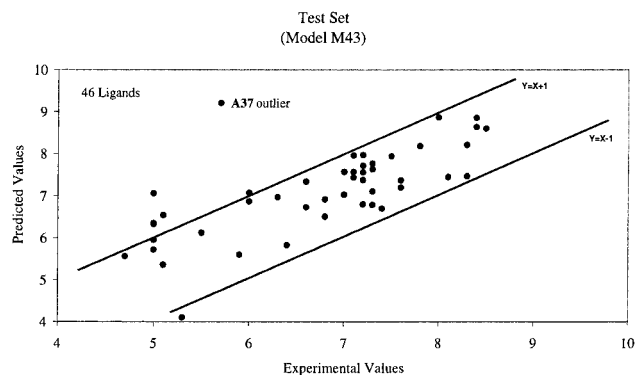


Figure 10. pK_i values predicted by model M43 versus experimental values for the 46 compounds of the test set.

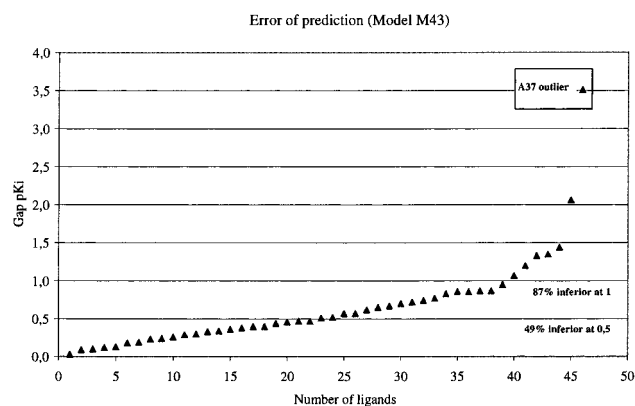


Figure 11. Gap between pK_i values predicted by model M43 versus experimental values for the 46 compounds of the test set.

dihedral angle from 170° to 142° is energetically unfavorable and, as a result, the I_2 binding will be low. In addition, if this feature is not described by a CoMFA model, the pK_i of compound **A37** will be overestimated. If these latter elements might look spurious to explain, first, why compound **A37** has a poor I_2 affinity and, second, why it is overpredicted by model M43, they nevertheless underline the limit of a fully validated CoMFA model to the design of ligands similar to those of the training set.

CoMFA Contour Maps. We graphically compared the CoMFA contour maps of the models selected by the first filter (not shown), and we noted that the ASP CoMFA contour maps were diverse and unworkable when the molecular diversity was increasing. On the opposite, we noted similar zones for the good PbP models obtained (e.g. models M37 and M43) which became more refined in the model M43. This latter fact correlates, for a series of models based on a specific alignment, the variability of the CoMFA contour maps with a poor general predictive ability and vice versa. Between two models based on the same training set and evaluated with the same test set, the model showing the best predictive ability will have the best explicative ability. As M43 displays these refined zones and additional ones, only its CoMFA contour map will be displayed and discussed.

M43 CoMFA Contour Map. Finally selected as the most powerful model, M43 displays an equal contribution of the lipophilic (49%) and steric (51%) fields. Compared to previous I_2 CoMFA models,^{23,24} our model does not include the electrostatic field as a descriptor.

However, it has to be reminded that the MLP encodes, in addition to hydrophobic and dispersive forces, a polar term, which reflects electrostatic interactions. An iso-contour map of CoMFA model M43 is reported in Figure 8, onto which are displayed the regions where variations in the lipophilic and steric features of the I_2 training set ligands are connected to variations in the affinity. On the whole, the most salient features to emerge from this CoMFA map are as follows: Two favorable hydrophilic regions (cyan) surround the two nitrogens of the imidazoline cycle. Two cyan regions are located on both faces of T2. The larger one can be interpreted as a hydrophilic favorable region since **B5** have an oxymethyl substituent in this area. One unfavorable steric region (red) is above the junction of the naphthalene and the imidazoline cycle. This region was already reported in CoMFA works on C ligands.^{23,24} For example, **C18** as well as **A8** which display a substituent in that position have very low I_2 affinities. A favorable steric region (green) is positioned around the 9-position of T2. Here again, this region was already reported and is mainly due to C ligands, which display substituents at that position. A favorable steric region surrounds the 11-position extremity of the benzazoline. For example, this region is reached by the 11-substituent of **A3**. Finally, an unfavorable steric region is located in front of the 10-position of T2. This latter is reached by **A23**.

Elements for a Tridimensional Pharmacophore.

Taken as a whole, the present report collected global structural diversity and QSARs dealing with the I_2 scope. Some of these points were used to advance the first elements of an I_2 pharmacophore. As common structural features, the three structural families A–C display one imidazoline as well as one phenyl moiety. More precisely, since every modification made to the imidazoline cycle is connected to a dramatic decrease of the I_2 affinity, the integrity of the imidazoline is stressed out. The importance of the two nitrogens is also underlined by two of the M43 CoMFA isocontours. Another pharmacophoric feature may be extracted from the alignment on which the global I_2 model M43 is based. Indeed, template T2 and its related 142° dihedral angle are the basis of the whole model. This latter point describes a pharmacophoric system defined by two planes relatively positioned by a 142° angle, each being respectively characterized by an imidazoline cycle and an aromatic moiety. As for the positions of the previously mentioned aromatic moiety, with respect to the imidazoline cycle, they are different for subsets A–C. These differently positioned aromatic moieties may be regarded as an associated I_2 receptor zone. To define this zone and its width, we used the centroids of the B and C aromatic moieties. That way, the maximal zone extent is defined at the position of the centroid for the aromatic moiety of subset C, whereas the minimal zone extent is defined at the position of the centroid for the aromatic moiety of subset B: the line between these two extreme authorized positions of the aromatic centroids is named the aromatic axis. Thus, the following elements are proposed as I_2 pharmacophoric elements and displayed in Figure 12: (1) a nonsubstituted imidazoline cycle and its associated plane p1, (2) an aromatic moiety and its associated plane p2, (3) a 142° dihedral angle between the two latter planes such as

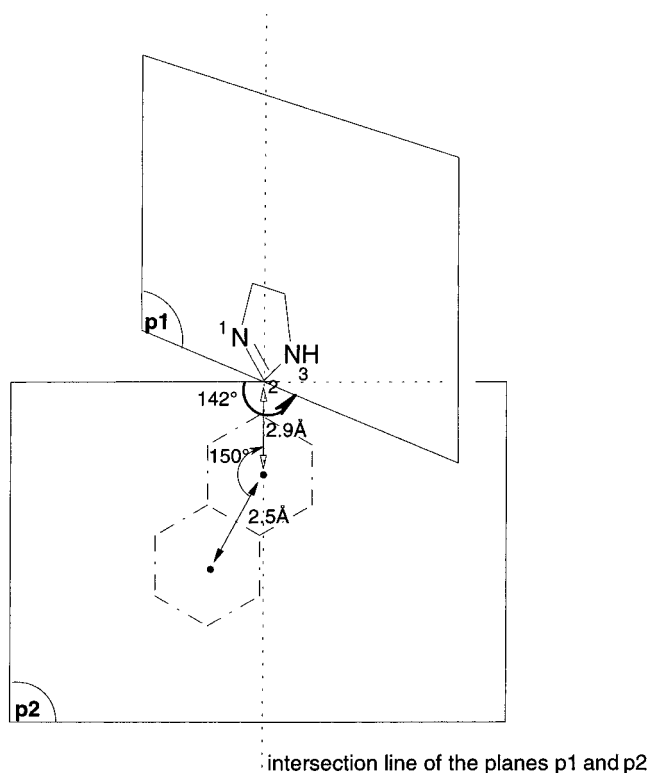


Figure 12. I₂ pharmacophore map displaying the imidazoline cycle (plane p1) and the 2.5 Å wide aromatic zone (plane p2): the dihedral angle is 142°, and the intersection of the two planes goes through the imidazoline C2 atom. The aromatic axis defines the authorized position for the centroid of the aromatic moiety: the angle between this axis and the interplane intersection line is 150°, an extremity of this axis belongs to the interplane intersection line at 2.9 Å from the imidazoline C2 atom, and the distance between the axis extremities is 2.5 Å; dashed line, the naphthalene moiety of the template molecule, benzazoline (see the Elements for a Tridimensional Pharmacophore section for details).

the imidazoline C2 atom belongs to both planes p1 and p2, (4) a 2.5 Å long aromatic axis, defining the authorized position of the centroid of an aromatic system, with an extremity at 2.9 Å from the imidazoline C2 atom following the interplane intersection line and the other extremity following a line at 150° with respect to the interplane intersection line, and (5) a set of tridimensional regions which are favorable and unfavorable to the I₂ affinity as displayed on the M43 CoMFA contour map (Figure 8).

Conclusion

In this paper, we have reported on new imidazoline-based structures which, combined to previously known ones, are displaying an unprecedented I₂ structural diversity. To exploit to the best such an amount of data, we have applied a combined CoMFA analysis where various key points of the CoMFA methodology were explored. Keeping in mind the potential use of a CoMFA model, the large amount of CoMFA models first derived was handled by applying filters in relation with the predictive ability. That way, we succeeded in extracting a unique CoMFA model, M43, as the best in terms of predictive ability. This outlines the interest of the combined CoMFA methodology (exploration of the template conformations, the alignment methods, and CoM-

FA fields; application of predictivity-based filters) here applied. CoMFA model M43 is highly interesting since it is representative of the three I₂ structural subsets and was fully validated for its predictive ability, both internally and externally. This latter test characterizes model M43 as a valuable tool for the upcoming process consisting in proposing I₂ high-affinity ligands to synthesize. Indeed, since the I₂ receptor has been neither cloned nor sequenced, our CoMFA model is at the moment the more advanced tool available for the pharmacochemist to direct I₂ drug design. That way, by combining our medicinal and computational chemistry means, this CoMFA model will be exploited for both its associated CoMFA constraints and its predicting power. Finally, the whole structural diversity and CoMFA results were used to advance the first elements of an I₂ tridimensional pharmacophore. In the near future, similar work applied to the I₁ and α₂ receptors should give access to features relating to selectivity, another fundamental element toward therapeutic applications. In this connection, the identification of the imidazoline receptor 3D structures, which could be integrated in a combined CoMFA process, is a promising event to further rationalize the I₂ drug design.

Acknowledgment. We are especially grateful to Céline Michelin for the preliminary work on the subject.

References

- (1) Bousquet, P.; Feldman, J.; Schwartz, J. Central cardiovascular effects of alpha adrenergic drugs: differences between catecholamines and imidazolines. *J. Pharmacol. Exp. Ther.* **1984**, *230*, 232–236.
- (2) Molderings, G. J. Imidazoline receptors: basic knowledge, recent advances and future prospects for therapy and diagnosis. *Drugs Future* **1997**, *22*, 757–772.
- (3) Chan, S. L. F. Clonidine-displacing substance and its putative role in control of insulin secretion: A minireview. *Gen. Pharmacol.* **1998**, *31*, 525–529.
- (4) Reis, D. J.; Regunathan, S. Agmatine: an endogenous ligand at imidazoline receptors may be a novel neurotransmitter in brain. *J. Auton. Nerv. Syst.* **1998**, *72*, 80–85.
- (5) Wenzel, R. R.; Spieker, L.; Qui, S.; Shaw, S.; Luscher, T. F.; Noll, G. I₁-imidazoline agonist moxonidine decreases sympathetic nerve activity and blood pressure in hypertensives. *Hypertension* **1998**, *32*, 1022–1027.
- (6) Molderings, G. J.; Gothert, M. Imidazoline binding sites and receptors in cardiovascular tissue. *Gen. Pharmacol.* **1999**, *32*, 17–22.
- (7) Chu, T. C.; Socci, R. R.; Ogidigben, M. J.; Potter, E. D. Potential mechanisms of moxonidine-induced ocular hypotension: role of norepinephrine. *J. Ocul. Pharmacol. Ther.* **1997**, *13*, 489–496.
- (8) Ugedo, L.; Pineda, J.; Ruiz-Ortega, J. A.; Martin-Ruiz, R. Stimulation of locus coeruleus neurons by non-I₁/I₂-type imidazoline receptors: an in vivo and in vitro electrophysiological study. *Br. J. Pharmacol.* **1998**, *125*, 1685–1694.
- (9) Ernsberger, P.; Ishizuka, T.; Liu, S.; Farrell, C. J.; Bedol, D.; Koletsky, R. J.; Friedman, J. E. Mechanisms of antihyperglycemic effects of moxonidine in the obese spontaneously hypertensive Koletsky rat (SHROB). *J. Pharmacol. Exp. Ther.* **1999**, *288*, 139–147.
- (10) Rondu, F.; Le Bihan, G.; Wang, X.; Lamouri, A.; Touboul, E.; Dive, G.; Bellahsene, T.; Pfeiffer, B.; Renard, P.; Guardiola-Lemaitre, B.; Manechez, D.; Penicaud, L.; Ktorza, A.; Godfroid, J. J. Design and Synthesis of Imidazoline Derivatives Active on Glucose Homeostasis in a Rat Model of Type II Diabetes. 1. Synthesis and Biological Activities of N-Benzyl-N'-(arylalkyl)-2-(4',5'-dihydro-1'H-imidazol-2'-yl)piperazines. *J. Med. Chem.* **1997**, *40*, 3793–3803.
- (11) Garcia-Sevilla, J. A.; Escriba, P. V.; Walzer, C.; Bouras, C.; Guimon, J. Imidazoline receptor proteins in brains of patients with Alzheimer's disease. *Neurosci. Lett.* **1998**, *247*, 95–98.
- (12) Heemskerck, F. M. J.; Dontenwill, M.; Grenay, H.; Vonthron, C.; Bousquet, P. Evidence for the existence of imidazoline-specific binding sites in synaptosomal plasma membranes of the bovine brainstem. *J. Neurochem.* **1998**, *71*, 2193–2202.
- (13) Remaury, A.; Missy, K.; Parini, A. Characterization of [H-3]-idazoxan binding proteins in solubilized membranes from rabbit and human liver. *J. Auton. Nerv. Syst.* **1998**, *72*, 111–117.

- (14) Nutt, D. J.; French, N.; Handley, S.; Hudson, A.; Husbands, S.; Jackson, H.; Jordan, S.; Lallies, M. D.; Lewis, J.; Lione, L.; Mallard, N.; Pratt, J. Functional studies of specific imidazoline-2 receptor ligands. *Ann. N. Y. Acad. Sci.* **1995**, *763*, 125–139.
- (15) Molderings, G. J.; Donecker, K.; Burian, M.; Simon, W. A.; Schroder, D. W.; Gothert, M. Characterization of I-2 imidazoline and a binding sites in the rat and human stomach. *J. Pharmacol. Exp. Ther.* **1998**, *285*, 170–177.
- (16) Ivanov, T. R.; Feng, Y.; Wang, H.; Regunathan, S.; Reis, D. J.; Chikkala, D. N.; Gupta, P.; Jones, J. C.; Piletz, J. E. Imidazoline receptor proteins are regulated in platelet-precursor MEG-01 cells by agonists and antagonists. *J. Psychiatr. Res.* **1998**, *32*, 65–79.
- (17) Boronat, M. A.; Olmos, G.; Garcia-Sevilla, J. A. Attenuation of tolerance to opioid-induced antinociception and protection against morphine-induced decrease of neurofilament proteins by idazoxan and other I-2-imidazoline ligands. *Br. J. Pharmacol.* **1998**, *125*, 175–185.
- (18) Smyth, D. D.; Penner, S. B. Imidazoline receptor mediated natriuresis: central and/or peripheral effect. *J. Auton. Nerv. Syst.* **1998**, *72*, 155–162.
- (19) Gargalidis-Moudanos, C.; Pizzinat, N.; Javoy-Agid, F.; Remaury, A.; Parini, A. I₂ Imidazoline binding sites and monoamine oxidase activity in human postmortem brain from patients with Parkinson's disease. *Neurochem. Int.* **1997**, *30*, 31–36.
- (20) Michel, M. C.; Ernsberger, P. Keeping an eye on the I site: imidazoline-preferring receptors. *Trends Pharmacol. Sci.* **1992**, *13*, 369–370.
- (21) Tesson, F.; Prib-Buus, C.; Lenoire, A.; Pegorier, J. P.; Parini, A. Subcellular distribution of imidazoline-guanidinium-receptive sites in human and rabbit liver. Major localization to the mitochondrial outer membrane. *J. Biol. Chem.* **1991**, *266*, 155–160.
- (22) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (23) Pignini, M.; Bousquet, P.; Brasili, L.; Carrieri, A.; Cavagna, R.; Dontenwill, M.; Gentili, F.; Giannella, M.; Leonetti, F.; Piergentili, A.; Quaglia, W.; Carotti, A. Ligand binding to I₂ imidazoline receptor: The role of lipophilicity in quantitative structure–activity relationship models. *Bioorg. Med. Chem.* **1998**, *6*, 2245–2260.
- (24) Carrieri, A.; Brasili, L.; Leonetti, F.; Pignini, M.; Giannella, M.; Bousquet, P.; Carotti, A. 2-D and 3-D modeling of imidazoline receptor ligands: insights into pharmacophore. *Bioorg. Med. Chem.* **1997**, *5*, 843–856.
- (25) Comoy, C.; Marot, C.; Podona, T.; Baudin, M. L.; Morin-Allory, L.; Guillaumet, G.; Pfeiffer, B.; Caignard, D. H.; Renard, P.; Rettori, M. C.; Adam, G.; Guardiola-Lemaitre, B. 3-Amino-3,4-dihydro-2H-1-benzopyran Derivatives as 5-HT_{1A} Receptor Ligands and Potential Anxiolytic Agents. 2. Synthesis and Quantitative Structure–Activity Relationship Studies of Spiro[pyrrolidine- and piperidine-2,3'(2'H)-benzopyrans]. *J. Med. Chem.* **1996**, *39*, 4285–4298.
- (26) Marot, C.; Chavatte, P.; Morin-Allory, L.; Viaud, M. C.; Guillaumet, G.; Renard, P.; Lesieur, D.; Michel, A. Pharmacophoric Search and 3D-QSAR Comparative Molecular Field Analysis Studies on Agonists of Melatonin Sheep Receptors. *J. Med. Chem.* **1998**, *41*, 4453–4465.
- (27) Pignini, M.; Bousquet, P.; Carotti, A.; Dontenwill, M.; Giannella, M.; Moriconi, R.; Piergentili, A.; Quaglia, W.; Tayebati, S. K.; Brasili, L. Imidazoline receptors: qualitative structure–activity relationships and discovery of trazoline and benazoline. Two Ligands with high affinity and unprecedented selectivity. *Bioorg. Med. Chem.* **1997**, *5*, 833–841.
- (28) Compounds from our laboratories; synthesis and structural data to be published elsewhere.
- (29) Oxford Molecular Ltd., The Medawar Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA, U.K.
- (30) Sybyl 6.5, Tripos Associates Inc., 1699 South Hanley Rd., St. Louis, MO 63144.
- (31) Stewart, J. P. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–103.
- (32) (a) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular lipophilicity potential, a toll in 3D-QSAR. Method and Applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83–96. (b) Institut of Medicinal Chemistry, University of Lausanne, BEP CH-1015 Lausanne, Switzerland.
- (33) Audry, E.; Dubost, J. P.; Dallet, P.; Langlois, M. H.; Colleter, J. C. The Molecular Lipophilic Potential: Application to a series of beta-adrenolytic amines. *Eur. J. Med. Chem.* **1989**, *24*, 155–161.
- (34) Carbo, R.; Domingo, L. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* **1987**, *32*, 517–545.
- (35) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure–activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (36) Good, A. C.; Richards, W. G. Rapid Evaluation of Shape Similarity Using Gaussian Functions. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112–116.
- (37) Nelder, J. A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308–313.
- (38) Vinter, J. G.; Davis, A.; Saunders, M. R. Strategic approaches to drug design. 1. An integrated software framework for molecular modeling. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31–51.
- (39) Van de Waterbeemd, H.; Testa, B. The parametrization of lipophilicity and other structural properties in drug design. In *Advances in Drug Research*; Testa, B., Ed.; Academic Press: London, 1987; Vol. 16, pp 87–227.

JM991124T