Three-Dimensional Quantitative Structure−Activity Relationships of Cyclo-oxygenase-2 (COX-2) Inhibitors: A Comparative Molecular Field Analysis

Philippe Chavatte,*,§ Saïd Yous,§ Christophe Marot,† Nicolas Baurin,† and Daniel Lesieur§

Institut de Chimie Pharmaceutique Albert Lespagnol, Université de Lille 2, BP 83, F-59006 Lille Cédex, France, and Institut de Chimie Organique et Analytique, UMR 6005, Université d'Orléans, BP 6759, F-45067 Orléans Cédex 2, France

Received March 26, 2001

The three-dimensional quantitative structure−activity relationship (3D-QSAR) approach using comparative molecular field analysis (CoMFA) was applied to an extensive series of 305 varied diarylheterocyclic derivatives known as COX-2 selective inhibitors. X-ray crystal structure of COX-2 bound with SC-558, a selective COX-2 inhibitor, was used to derive the putative bioactive conformation of these inhibitors. Five statistically significant models were obtained from the randomly constituted training sets (229 compounds) and subsequently validated with the corresponding test sets (76 compounds). The best predictive model ($n = 229$, $q^2 = 0.714$, $N = 8$, $r^2 = 0.905$, $s = 0.291$, $F = 261.545$) was selected for further comparison of the CoMFA contour maps obtained for steric, electrostatic, and lipophilic fields with the enzyme structure. The high level of compatibility with the COX-2 enzyme topology shows the great accuracy of this model that can predict inhibitory activities for a wide range of compounds and offers important structural insight into designing novel antiinflammatory drugs prior to their synthesis.

Introduction

The classical nonsteroidal antiinflammatory drugs (NSAIDs) such as aspirin, ibuprofen, or indomethacin are therapeutic agents widely used in the treatment of inflammation, pain, and fever.[1] Their principal pharmacological effect is their ability to inhibit prostaglandin synthesis. They act via inhibition of the enzyme prostaglandin H_2 synthase, also referred to as cyclooxygenase (COX), which catalyses the conversion of arachidonic acid to prostaglandin H_2 (PGH_2). In 1990, Fu et al. discovered the existence of two isoforms of this enzyme: COX-1 and COX-2.[2] COX-1 is constitutively expressed in most tissues and, particularly, in the gastrointestinal tract and kidneys where it is mainly responsible for the synthesis of cytoprotective prostaglandins. COX-2 is selectively induced by proinflammatory cytokines (IL-1) and growth factors (TNFα) and facilitates the release of prostaglandins involved in the inflammatory process. This discovery led to the hypothesis that side effects such as ulcers and renal failure associated with the clinically useful NSAIDs are caused by the inhibition of COX-1, whereas the antiinflammatory properties result from the inhibition of the inducible COX-2. Selective inhibition of COX-2 provided a new class of antiinflammatory, analgesic, and antipyretic drugs with significantly reduced side effects. Recent works suggest that inhibiting COX-2 could also be an important strategy for preventing or treating a number of cancers[3] and could be used to delay or slow the clinical expression of Alzheimer's disease.[4]

The discovery of two lead compounds arising from distinct chemical classes, NS-398[5] and DuP-697[6] (Figure 1), has led to two general classes of selective COX-2 inhibitors, the diarylheterocycles and the methanesulfonanilides. When tested on animal models at high doses, they have shown antiinflammatory, analgesic, and antipyretic activities without causing gastrointestinal lesions. Recently, the structures of some nonselective NSAIDs such as indomethacin, zomepirac, aspirin, and flurbiprofen have been successfully converted into selective COX-2 inhibitors.[7]

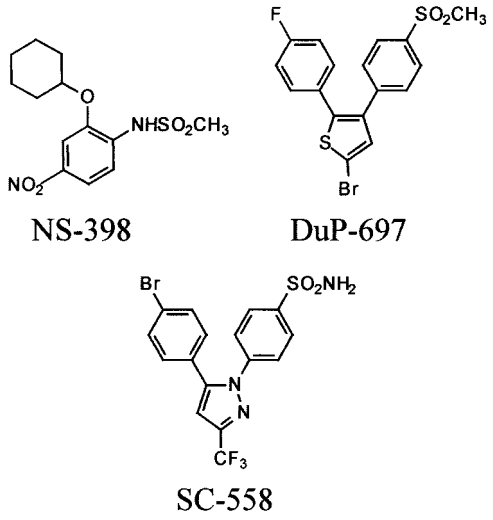


Figure 1. Molecular structures of NS-398, DuP-697, and SC-558.

In this paper, we report a 3D-QSAR study which applied comparative molecular field analysis (CoMFA)[8] methodology, including the steric, electrostatic, and lipophilic fields,[9,10] to derive predictive models from a wide series of 305 varied diarylheterocyclic derivatives selected from other papers as COX-2 selective inhibitors. The determination of the active conformation which constitutes the critical step in 3D-QSAR CoMFA studies, was facilitated by the availability of the X-ray crystal structure of murine COX-2 bound with SC-558

* To whom correspondence should be addressed. Tel: +33 (0)320 964 040. Fax: +33 (0)320 964 375. E-mail: pchavatt@phare.univ-lille2.fr.
§ Université de Lille 2.
† Université d'Orléans.

10.1021/jm0101343 CCC: $20.00
Published on Web 09/05/2001

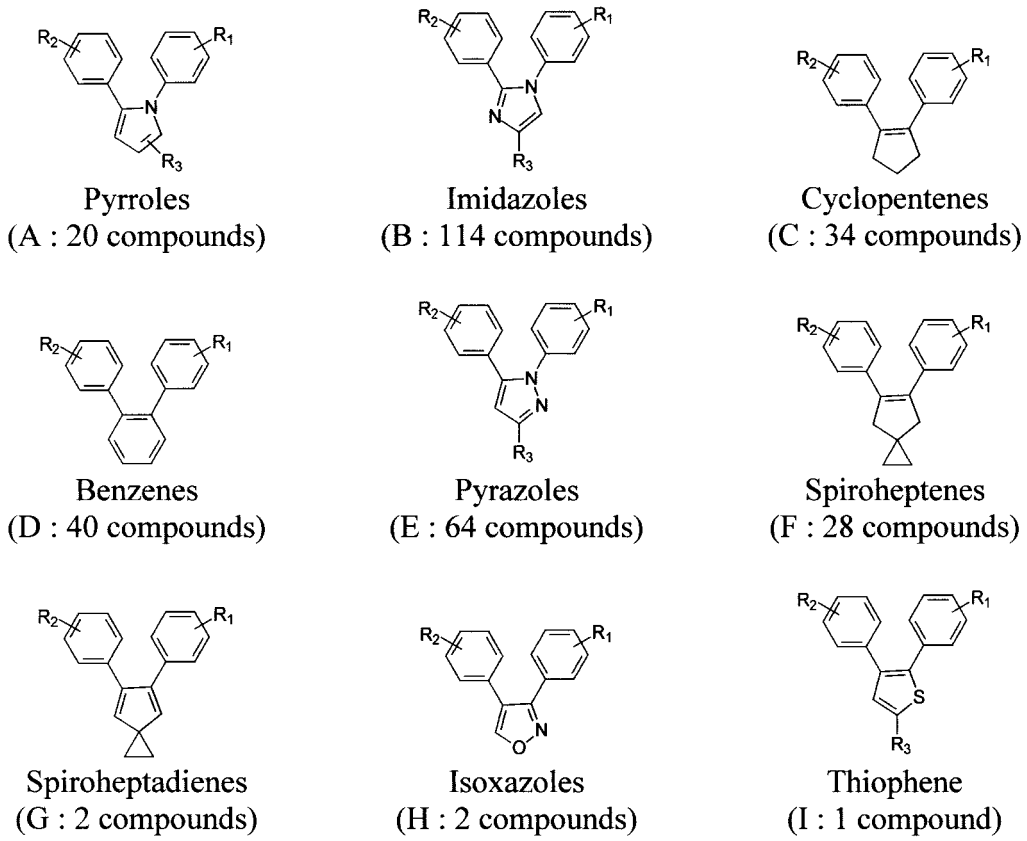


Figure 2. Definition of the nine families of inhibitors studied.

(Figure 1), a selective COX-2 inhibitor.[11] Our recent success in deriving a 3D-QSAR predictive model on a set of selective COX-2 inhibitors[12] prompted us to extend the study to a larger and more diverse set of inhibitors of the diarylheterocycle class. This should facilitate the design and development of new selective COX-2 inhibitors.

Materials and Methods

Selection of Ligands. A compilation was made of the inhibitory activity values of 305 compounds against the COX-2 inducible isoform. All the values retained had been obtained using the same biological method on the human recombinant enzyme[13] and are expressed in terms of pIC_{50} or $\log(1/IC_{50})$ where IC_{50} represents the drug concentration that inhibits 50% of activity. The inhibitory activities of reference compounds were checked to ensure that no difference occurred between the different series. The compounds retained belong to nine structurally different families depending on the central cyclic tensor (Figure 2). These tensors were pyrrole[14] (family A: 20 compounds in Table 1, Supporting Information), imidazole[15,16] (family B: 114 compounds in Table 2, Supporting Information), cyclopentene[17,18] (family C: 34 compounds in Table 3, Supporting Information), benzene[19] (family D: 40 compounds in Table 4, Supporting Information), pyrazole[20] (family E: 64 compounds in Table 5, Supporting Information), spiroheptene[21] (family F: 28 compounds in Table 6, Supporting Information), spiroheptadiene[21] (family G: 2 compounds in Table 7, Supporting Information), isoxazole[22] (family H: 2 compounds in Table 8, Supporting Information), and thiophene (family I: 1 compound in Table 9, Supporting Information). Their enzyme inhibitory activities are widespread and homogeneous (Figure 3). Thirty-eight compounds display pIC_{50} values between 5 and 6 (very low activity), 107 display pIC_{50} values between 6 and 7 (low activity), 104 exhibit pIC_{50} values between 7 and 8 (moderate activity), and 56 show pIC_{50} values higher than 8 (good activity). This is a prerequisite if meaningful results are to be

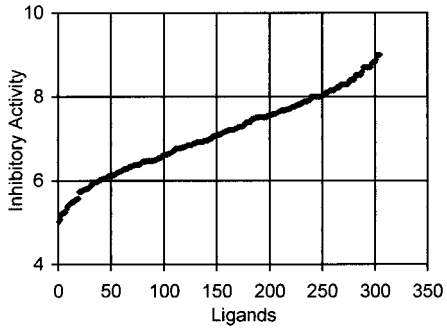
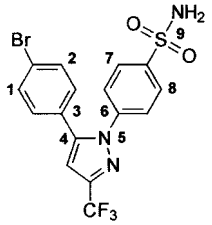


Figure 3. Distribution of inhibitory activities (pIC_{50}) of all compounds investigated.

obtained from a three-dimensional quantitative structure–activity relationship study (3D-QSAR) using the comparative molecular field analysis (CoMFA) method. Approximately 75% of the 305 compounds were divided into five individual training sets of 229 compounds each, and the remaining 25% were used as test sets with 76 compounds each. The compounds were split between training and test sets at random. Random numbers were generated and assigned to each compound before they were sorted in increasing order. The test set elaboration follows suggestions by Oprea et al.:[23] (1) the biological assay methods for both training and test set should be compatible; (2) for the test set, the biological activity values should span several orders of magnitude but should not exceed activity values in the training set by more than 10%; (3) the test set should represent a balanced number of both active and inactive compounds for uniform sampling of the data. This multimodel approach was used to assess the predictive power of the final model.

Molecular Modeling. Molecular modeling studies were performed using SYBYL software version 6.6[24] running on Silicon Graphics workstations. The X-ray crystal structure of SC-558 (Figure 1), a selective COX-2 inhibitor bound to the enzyme active site and available in the RCSB Protein Data



SC-558

Figure 4. Superposition modes.

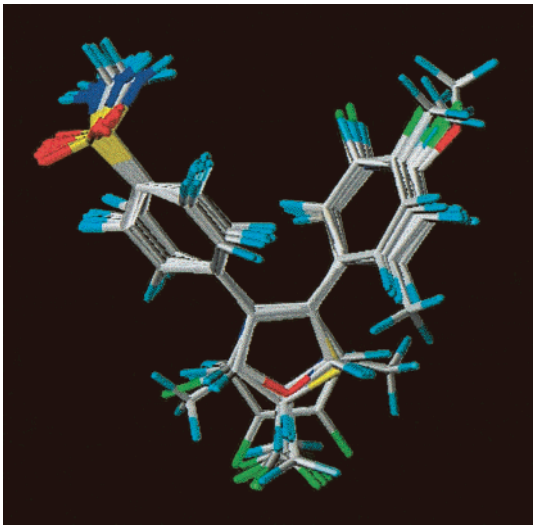


Figure 5. Alignment of the compounds.

Bank (1CX2),[11] was used as template to construct the three-dimensional models of all the compounds by replacing phenyl substituents or modifying the heterocycle. The geometry of these putative bioactive conformations was subsequently optimized using the Tripos force field[25] including the electrostatic term calculated from Gasteiger and Hückel atomic charges. The method of Powell available in Maximin2 procedure was used for energy minimization until the gradient value was smaller than 0.001 kcal/mol Å. To determine the best orientation of the phenyl substituents, the torsional angle was manually rotated by 180° and the two rotamers were rigidly docked into the enzyme active site using the Dock module. The inhibitor−enzyme complexes were energy-minimized, using a dielectric constant of 4.0 and a 10 Å nonbonded cutoff, until the gradient value was smaller than 0.05 kcal/mol Å. The inhibitor structure in the most stable docking model was extracted from the complex for subsequent energy minimization. Each compound geometry was then optimized with the semiempirical MOPAC package version 6.0[26] using the Hamiltonian AM1 (keywords: PRECISE, NOMM, PARASOK), and Coulson partial atomic charges were calculated using the same method.

Alignment of the Compounds. One of the most important adjustable parameters in CoMFA is the relative alignment of all the compounds to one another so that they have a comparable conformation and a similar orientation in space. The SC-558 conformation extracted from the X-ray crystallographic inhibitor−enzyme complex was used as template for superimposition, assuming that this conformation represents the most probable bioactive conformation of the diarylheterocycle derivatives at the enzyme active site level. Nine features were selected for the alignment of all compounds. These atoms are numbered 1 to 9 in Figure 4. An example of this alignment is presented in Figure 5. The fitting process was performed using an "in-house" developed method written in SPL (SYBYL Programming Language). This SPL program makes it possible to fit all conformations of a compound on a template conformation (SC-558) according to selected pairs of atoms.

Table 1. Color Code of Graphic CoMFA Results

molecular fields	increased activity	decreased activity
steric (Ste)	green	yellow
electrostatic (Ele)		
positive charge	blue	red
negative charge	red	blue
lipophilic (Lip)		
lipophilicity	magenta	white
hydrophilicity	white	magenta

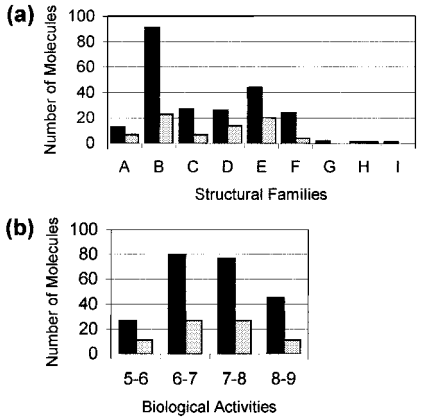


Figure 6. Distributions of structural families (a) and biological activities (pIC_{50}) (b) versus number of molecules for the training (black) and test (gray) sets of model 3.

CoMFA Studies. The CoMFA[8] studies were performed with the QSAR module of SYBYL for each combination of the three molecular fields—(Ste) steric, (Ele) electrostatic, and (Lip) lipophilic—which were sampled at each point of a regularly spaced grid of 1.5 Å within an automatically defined region. The steric and electrostatic fields were calculated using a sp^3-carbon with a +1 charge as probe. The lipophilic field was calculated by the molecular lipophilic potential (MLP) implemented in the CLIP[9] module of SYBYL. The method of partial least squares (PLS) implemented in the QSAR module of SYBYL was used to construct and validate the models. Cross-validation was performed with the leave-one-out procedure. The optimal number of components N retained for final PLS analyses was defined as the one that yielded the highest cross-validated q^2 value and which normally had the smallest standard error of prediction s_{cv}. The robustness of the models was internally evaluated by calculating the r^2, s, and F test values from the training set and was externally validated by calculating the r^2_{pred} from the test set. To obtain the statistical confidence limits on the analyses, bootstrapping was carried out with 100 groups.

Presentation of the Results. From fully validated CoMFA models, contour maps are presented displaying the most relevant regions of the space where variations in the statistical steric, electrostatic, and lipophilic fields are the largest. The color code used to characterize these isocontours for each field signal is described in Table 1.

Results and Discussion

A multimodel approach was used to obtain the best predictive CoMFA model.

Training and Test Sets. Great attention was paid to the distribution of biological activities and structural classification of compounds in both the training and test sets. All models show homogeneity in the distribution of biological activities and structural characteristics of their compounds split between training and test sets randomly as presented for model 3 in Figure 6.

CoMFA Models. Knowledge of the SC-558 bioactive conformation obtained from the X-ray crystallographic inhibitor−enzyme complex was of great help in deter-

Table 2. Statistical Results for the Five CoMFA Models

	model 1	model 2	model 3	model 4	model 5
q^2 [a]	0.741	0.722	0.714	0.708	0.707
N [b]	8	8	8	8	8
s_{cv} [c]	0.490	0.487	0.504	0.503	0.522
r^2 [d]	0.907	0.899	0.905	0.904	0.902
s [e]	0.294	0.292	0.291	0.289	0.301
F [f]	267.284	246.028	261.545	257.996	254.529
r^2_{bs} [g]	0.937	0.934	0.933	0.935	0.933
s_{bs} [h]	0.008	0.008	0.008	0.007	0.008
r^2_{pred} [i]	0.654	0.708	0.740	0.710	0.664

[a] Cross-validation correlation coefficient. [b] Number of components. [c] Standard error of prediction. [d] Correlation coefficient. [e] Standard error of estimate. [f] *F*-ratio. [g] Bootstrapped correlation coefficient. [h] Bootstrapped standard deviation. [i] Predicted correlation coefficient.

mining an experimentally derived alignment rule for the five constituted training sets and their corresponding test sets. The five different training sets were then used to derive five separate CoMFA models taking into account the combination of the three molecular fields (i.e., steric, electrostatic, and lipophilic). The results obtained are presented in Table 2. Although cross-validation reflects the predictive power of the models, the results indicate the good predictive capacity of the models which yield high cross-validated correlation coefficients q^2 (from 0.707 to 0.741) with reasonable respective standard errors of prediction s_{cv} (from 0.490 to 0.522). A q^2 value of 0.3 corresponds to a confidence limit greater than 95%, which minimizes the risk of finding correlation just by mere chance.[27] The optimal number of components in each case was 8. This value is reasonable considering the high number of compounds used to derive the models. The leave-one-out procedure might produce high q^2 values which do not necessarily give a suitable representation of the real predictive power of the models.[28] We therefore performed leave-half-out cross-validation using half of the training set compounds to predict the activity of the remaining half. As the random constitution of the two groups may have an effect on the results, cross-validation was repeated 100 times. The results show that the q^2 mean values obtained for the five models with low corresponding standard deviations are slightly lower than those obtained with the leave-one-out method. We never observed any negative q^2 mean values and can confirm the great internal consistency of our training sets. We checked this again by scrambling the biological data and repeating the model derivation process.[27] Chance correlation can be detected in this way. The biological activities were randomly assigned to molecular structures before leave-one-out cross-validation was performed 50 times. In all cases, negative q^2 values were noted, and no model was obtained with more than three components. We can therefore conclude that our models are significantly better than random models.

Using eight principal components, our five models yielded high conventional r^2 (from 0.899 to 0.907) with relatively low standard errors of estimate s (from 0.289 to 0.301) as shown in Table 2. These data suggest a good correlation between the three molecular fields and the activities registered for the different compounds of the different models. The high bootstrapped correlation coefficients r^2_{bs} (from 0.933 to 0.937) and the small standard deviations s_{bs} (from 0.007 to 0.008) reflect a high degree of confidence in the analyses performed.

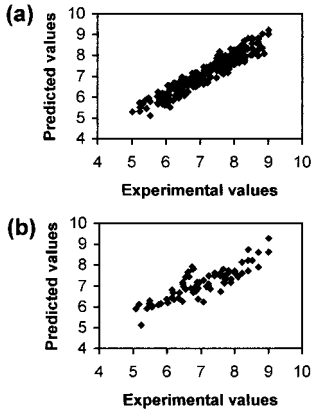


Figure 7. Predicted values versus experimental values for the training (a) and test (b) sets of model 3.

Finally, to validate our models, we attempted to predict activities for the 76 compounds of the test sets. The calculated predictive correlation coefficients r^2_{pred} are given in Table 2. Multiple models were constructed using overlapping test sets. Each compound may appear several times in the test sets. We looked at the molecular redundancy impact on prediction variations by calculating the standard error of prediction s_{pred} for each compound compiled over the number of times it appeared. A total of 115 compounds were selected once, 72 compounds were selected twice ($s_{pred} = 0.121 \pm 0.111$), 27 compounds were selected three times ($s_{pred} = 0.164 \pm 0.090$), 10 compounds were selected four times ($s_{pred} = 0.237 \pm 0.099$), and none was selected five times. We first noticed that there was molecular redundancy among the five test sets but it was not exclusive since 115 compounds were selected only once. Moreover, for compounds that were selected more than once, s_{pred} values were low showing that, for different combinations in the training set, the predictions were similar. This is further proof of the robustness of the various models.

Among the five models, model 3 yielding a good r^2_{pred} (0.740) concording with the q^2 value (0.714), appears to be the best predictive one. The determination coefficient of this linear regression has a value of 0.750. The estimated pIC_{50} values versus the experimental values for both the training and test sets are graphically represented in Figure 7. A total of 68% of compounds were predicted with an error lower than 0.5, and 96% of compounds with an error lower than 1. This model did not predict the pIC_{50} values accurately for only three compounds, **168**, **206**, and **303**, which were overestimated with error values, respectively, of 1.03, 1.05, and 1.18 log units. With regard to the size of our test set (76 compounds), these outliers do not call the statistical validity of our model into question. The discrepancy observed for compound **168** may be explained by the bad orientation of its phenyl ring substituents since when the torsional angle was rotated by 180°, the activity of the corresponding rotamer was correctly predicted with an error value of 0.69. Regarding compound **206**, our model cannot completely explain the sharp decrease in activity observed when replacing the two fluorine atoms in the central ring with chlorine atoms.[19] The activity of the difluorinated analogue of compound **206**, compound **177**, was correctly predicted in the corresponding training set, with an error value of 0.23. Compound **303** was identified as an active metabolite of valdecoxib,

Table 3. Statistical Results for the Seven Possible Combinations of the Three Molecular Fields: Steric, Electrostatic, and Lipophilic Fields Considered in CoMFA Models

	Ste[g]	Ele[g]	Lip[g]	Ste/Ele[g]	Ste/Lip[g]	Ele/Lip[g]	Ste/Ele/Lip[g]
q^2 [a]	0.553	0.554	0.572	0.601	0.622	0.652	0.714
N [b]	5	5	9	6	10	8	8
s_{cv} [c]	0.626	0.625	0.618	0.593	0.582	0.556	0.504
r^2 [d]	0.729	0.703	0.825	0.802	0.893	0.871	0.905
s [e]	0.488	0.511	0.395	0.418	0.509	0.338	0.291
F [f]	119.697	105.351	115.001	149.401	182.396	186.198	261.545

[a] Cross-validation correlation coefficient. [b] Number of components. [c] Standard error of prediction. [d] Correlation coefficient. [e] Standard error of estimate. [f] *F*-ratio. [g] Molecular field(s) used in CoMFA (Ste: steric field, Ele: electrostatic field, Lip: lipophilic field).

Table 4. Statistical Results for the Six Constituted Subsets

	A[h]	B[h]	C[h]	D[h]	E[h]	F[h]
n [a]	13	91	27	26	44	24
q^2 [b]	−0.043	0.703	0.658	0.368	0.156	0.451
N [c]	3	6	18	4	1	5
s_{cv} [d]	0.884	0.412	0.882	0.551	0.682	0.534
r^2 [e]		0.941	1.000	0.892		0.961
s [f]		0.183	0.022	0.228		0.141
F [g]		224.853	2027.002	43.282		89.897

[a] Number of compounds. [b] Cross-validation correlation coefficient. [c] Number of components. [d] Standard error of prediction. [e] Correlation coefficient. [f] Standard error of estimate. [g] *F*-ratio. [h] Subset constituted according to their structural belonging from the training set of model 3 (A: pyrrole, B: imidazole, C: cyclopentene, D: benzene, E: pyrazole, F: spiroheptene).

which is at present being clinically evaluated to control pain and inflammation.[22] Its surprising biological profile might explain why it is difficult to predict its pIC_{50} value accurately. The activity of valdecoxib, included in the training set, was very well predicted with an error value of 0.16. In the same way, celecoxib, which has recently been marketed, also displayed an excellent predicted activity with an error value of 0.19.

Model 3 was then selected for further investigations. We paid particular attention to the combination of the three molecular fields to determine their degree of interdependence. CoMFA models were calculated for each molecular field considered alone or in combination (Table 3). All the q^2 values were always higher than 0.5. These data suggest a good level of interdependency among the three molecular fields. The best CoMFA model was obtained with the combination of the three molecular fields: steric, electrostatic, and lipophilic.

Then, to determine the influence of structural diversity on the CoMFA model, we divided the training set of model 3 into six subsets according to structural characteristics. These subsets were pyrrole (subset A: 13 compounds), imidazole (subset B: 91 compounds), cyclopentene (subset C: 27 compounds), benzene (subset D: 26 compounds), pyrazole (subset E: 44 compounds), and spiroheptene (subset F: 24 compounds) derivatives. CoMFA models were derived for each subset (Table 4). No model with $q^2 > 0.3$ was found for the pyrrole (subset A) and pyrazole derivatives (subset E). For subset A, the number of compounds (13) was certainly too small to obtain a meaningful CoMFA model and, for subset E, structural variability seemed too poor to obtain a suitable model. On the other hand, statistically significant models were obtained for imidazole (subset B), benzene (subset D), and spiroheptene (subset F) derivatives. The cyclopentene derivatives (subset C) provide a statistically poor model since the latter needs 10 principal components for only 27 compounds. This leads to a very high unrealistic correlation coefficient r^2 (0.998).

Graphic representations of the CoMFA model 3 are displayed in Figure 8. They show regions where variations of steric, electrostatic, or lipophilic nature in the structural features of the different compounds of the training set lead to increases or decreases in activity. They do not take the common structural features of the molecules into account insofar as the CoMFA method relates the differences in biological properties to differences in the shapes of the noncovalent fields (electrostatic, steric, and lipophilic) surrounding the tested molecules. The CoMFA contour plot shows green-colored regions (>85% contribution) where increased steric bulk is associated with enhanced activity and yellow-colored regions (<15% contribution) where increased steric bulk is associated with diminished activity. The favorable steric region is close to the tensor between the two aromatic rings and the nonsulfonamide-containing phenyl ring. The unfavorable steric region may be mainly related to the critical size of the substituent at position-4 of this phenyl ring as discussed in a previous standard SAR study.[20] Regions where increased positive charge is favorable for activity (>85% contribution) are indicated in blue, while those where increased negative charge is favorable for activity (<15% contribution) are in red. Two blue zones can be observed close to the bottom of the heterocycle and the substituent at 4-position of the nonsulfonamide-containing phenyl ring. Another blue zone in the vicinity of the 2-position indicates that electron-withdrawing substituents in that area will decrease COX-2 inhibitory activity. For example, compounds **295** and **296** were less active than respective compounds **273** and **275** due to the presence of a chlorine or fluorine atom in 2-position. Regions colored in magenta (>85% contribution) correspond to a favorable influence of lipophilicity, and regions colored in white (<15% contribution) correspond to a favorable influence of hydrophilicity. The position of a magenta zone near the substituent at 4-position emphasizes the importance of its lipophilic nature. The presence of a white zone close to the sulfonamide group confirms this importance in COX-2 side-pocket binding. The structure of COX-2 inhibitors belonging to the 1,2-diarylheterocycle class exploits binding within the COX-2 side-pocket (often via sulfonyl, sulfone, or sulfonamide groups) to achieve selectivity.[29] These moieties are not structural requisites in the new class of 1,3-diaryltetrahydro-2*H*-indoles.[30] As CoMFA models were obtained from structural information provided only by the enzyme inhibitors, it seems interesting to describe the contour maps in relation to the enzyme site structure (Figure 9). Such an approach may provide valuable insight into binding requirements.[28,31−34] However, interpretation of results must be carried out carefully since contour maps cannot really be compared to receptor maps.[8] Sterically favorable green contours are located along the enzyme channel which is largely hydrophobic and contribute to enhancing activity by interacting with Val349, Leu352, Tyr355, Trp387, Phe518, Ala527, and Leu531, while the sterically un-

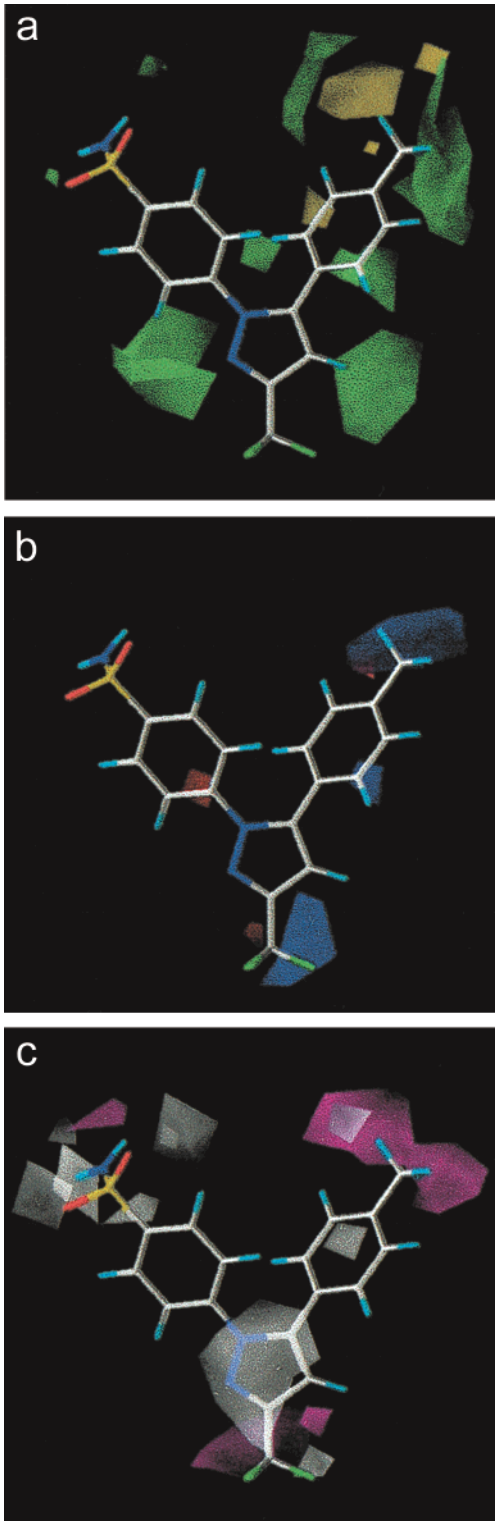


Figure 8. Contour plots of steric (a), electrostatic (b), and lipophilic (c) fields from the model 3 CoMFA in combination with celecoxib. (a) Steric contour plots: green contours (>85% contribution) indicate regions where an increase in steric bulk will enhance activity, and yellow contours (<15% contribution) indicate regions where an increase in steric bulk will reduce activity. (b) Electrostatic contour plots: blue contours (>85% contribution) and red contours (<15% contribution) correspond to regions where an increase in positive or negative charge, respectively, will enhance activity. (c) Lipophilic contour plots: magenta contours (>85% contribution) and white contours (<15% contribution) show regions where an increase in lipophilicity or hydrophilicity, respectively, will enhance activity.

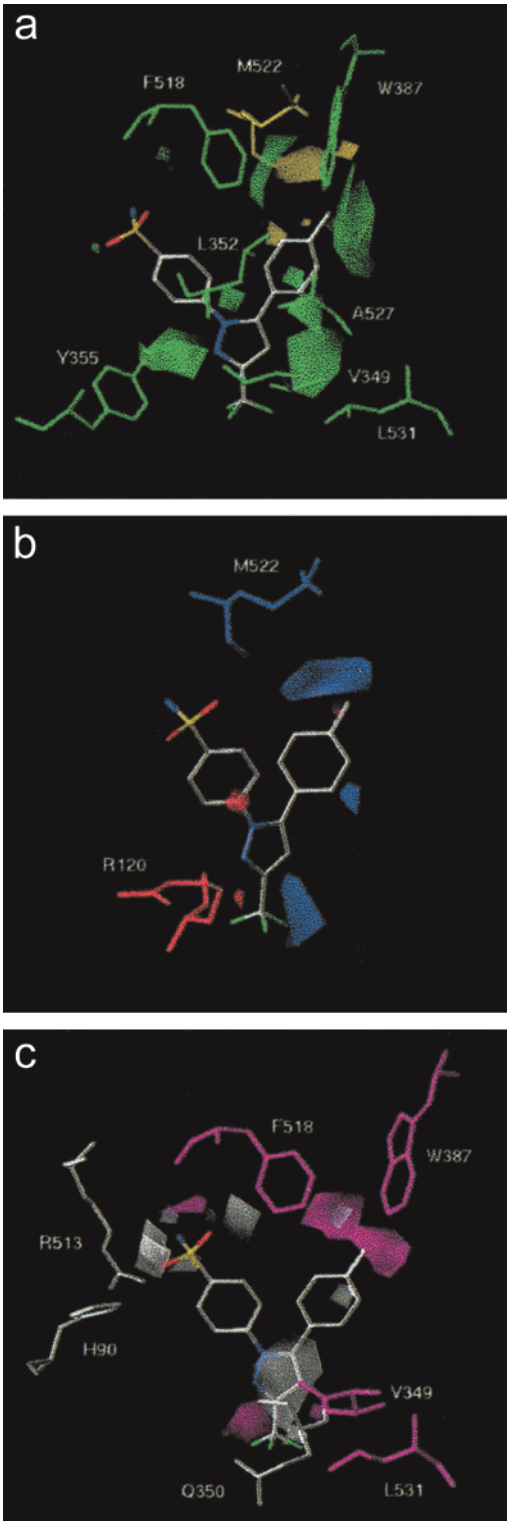


Figure 9. Superimposition of the steric (a), electrostatic (b), and lipophilic (c) model 3 CoMFA contour plots and some active site residues of COX-2.

favorable yellow region overlaps with atoms of Met522. The negative charge favorable red contour observed at the lower part of the heterocycle may be related to the ability of the compounds to bind to the positive-charged Arg120 in the COX-2 nonselective binding-site.[29] A positive charge favorable blue contour was also observed in the vicinity of the Met522 sulfide atom. The position of magenta contours suggests that this favorable hy-

drophobic influence may be related to the favorable influence of bulky nonpolar substituents in the vicinity of Val349, Trp387, Phe518, and Leu531. The hydrophilicfavorable white contour that occupies the COX-2 side-pocket may be explained by the possibility of sulfone or sulfonamide groups of the inhibitor to interact with the hydrophilic residues His90 and Arg513. Another white contour was also observed near Gln350. These observations indicate that our 3D-QSAR model concords with enzyme site topology and is able to take into account the interactions shown by X-ray data.

Conclusion

In this paper, a large series (305 compounds) of structurally different selective COX-2 inhibitors belonging to the diarylheterocycle class was used to generate and validate multiple 3D-QSAR models using CoMFA methodology and considering steric, electrostatic, and lipophilic parameters. These models were built using the X-ray crystal structure based alignment rule. The highly consistent and predictive selected CoMFA model was confronted with the three-dimensional structure of the enzyme. The CoMFA contour maps show total compatibility with protein topology. This confirms the great accuracy of the model that can predict inhibitory activities for a wide range of compounds of the COX-2 enzyme and offers important structural insight into designing novel selective COX-2 inhibitors prior to their synthesis.

Acknowledgment. We are especially grateful to Professor G. Vergoten for advice given on the subject and helpful discussions. We also thank Mrs. A. Tavernier for her assistance in the preparation of the manuscript.

Supporting Information Available: Tables of structures and COX-2 inhibitory activities (pIC_{50}) of pyrrole (Table 1), imidazole (Table 2), cyclopentene (Table 3), benzene (Table 4), pyrazole (Table 5), spiroheptene (Table 6), spiroheptadiene (Table 7), isoxazole (Table 8), and thiophene (Table 9) derivatives. Table of predicted values versus experimental values for the training and test sets of model 3 (Table 10). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Reitz, D. B.; Isakson, P. C. Cyclooxygenase-2 Inhibitors. *Curr. Pharm. Des.* **1995**, *1*, 211−220.
(2) Fu, J. Y.; Masferrer, J. L.; Seibert, K.; Raz, A.; Needleman, P. The Induction and Suppression of Prostaglandin H_2 Synthase (Cyclooxygenase) in Human Monocytes. *J. Biol. Chem.* **1990**, *265*, 16737−16740.
(3) Subbaramaiah, K.; Zakim, D.; Weksler, B. B.; Dannenberg A. J. Inhibition of Cyclooxygenase: A Novel Approach to Cancer Prevention. *Proc. Soc. Exp. Biol. Med.* **1997**, *216*, 201−210.
(4) Pasinetti, G. M. Cyclooxygenase and Inflammation in Alzheimer's Disease: Experimental Approaches and Clinical Intervention. *J. Neurosci. Res.* **1998**, *54*, 1−6.
(5) Futaki, N.; Yoshikawa, K.; Hamasaka, Y.; Arai, I.; Higuchi, S.; Iizuka, H.; Otomo, S. NS398, a Novel Nonsteroidal Antiinflammatory Drug with Potent Analgesic and Antipyretic Effects, which Causes Minimal Stomach Lesions. *Gen. Pharmacol.* **1993**, *24*, 105−110.
(6) Gans, K.; Galbraith, W.; Roman, R.; Haber, S.; Kerr, J.; Schmidt, W.; Smith, C.; Hewes, W.; Ackerman, N. Antiinflammatory and Safety Profile of DuP697, a Novel Orally Prostaglandin Synthesis Inhibitor. *J. Pharm. Exp. Ther.* **1990**, *254*, 180−187.
(7) Kalgutkar, A. S.; Marnett, A. B.; Crews, B. C.; Remmel, R. P.; Marnett, L. J. Ester and Amide Derivatives of the Nonsteroidal Antiinflammatory Drug, Indomethacin, as Selective Cyclooxygenase-2 Inhibitors. *J. Med. Chem.* **2000**, *43*, 2860−2870.
(8) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Fields Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(9) (a) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular Lipophilicity Potential, a Tool in 3D-QSAR. Methods and Applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83−96. (b) CLIP 1.0, Institute of Medicinal Chemistry, University of Lausanne, BEP-Dorigny, CH-1015 Lausanne, Switzerland.
(10) Carrupt, P. A.; Gaillard, P.; Billois, F.; Weber, P.; Testa, B.; Meyer, C.; Pérez, S. The Molecular Lipophilicity Potential (MLP) a New Tool for log P Calculation and in Comparative Molecular Field Analysis (CoMFA). In *Lipophilicity in Drug Research*; Pliska, V., Testa, B., Van de Waterbeemd, H., Eds.; VCH Publishers: Weinheim, 1996; pp 195−217.
(11) Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C. Structural basis for selective inhibition of cyclooxygenase-2 by antiinflammatory agents. *Nature* **1996**, *384*, 644−648.
(12) Marot, C.; Chavatte, P.; Lesieur, D. Comparative Molecular Field Analysis of Selective Cyclooxygenase-2 (COX-2) Inhibitors. *Quant. Struct.-Act. Relat.* **2000**, *19*, 127−134.
(13) Gierse, J. K.; Hauser, S. D.; Creely, D. P.; Koboldt, C.; Rangwala, S. H.; Isakson, P. C.; Seibert, K. Expression and Selective Inhibition of the Constitutive and Inducible Forms of Human Cyclo-oxygenase. *Biochem. J.* **1995**, *305*, 479−484.
(14) Khanna, I. K.; Weier, R. M.; Yu, Y.; Collins, P. W.; Miyashiro, J. M.; Koboldt, C. M.; Veenhuizen, A. W.; Currie, J. L.; Seibert, K.; Isakson, P. C. 1,2-Diarylpyrroles as Potent and Selective Inhibitors of Cyclooxygenase-2. *J. Med. Chem.* **1997**, *40*, 1619−1633.
(15) Khanna, I. K.; Weier, R. M.; Yu, Y.; Xu, X. D.; Koszyk, F. J.; Collins, P. W.; Koboldt, C. M.; Veenhuizen, A. W.; Perkins, W. E.; Casler, J. J.; Masferrer, J. L.; Zhang, Y. Y.; Gregory, S. A.; Seibert, K.; Isakson, P. C. 1,2-Diarylimidazoles as Potent, Cyclooxygenase-2 Selective, and Orally Active Antiinflammatory Agents. *J. Med. Chem.* **1997**, *40*, 1634−1647.
(16) Khanna, I. K.; Yu, Y.; Huff, R. M.; Weier, R. M.; Xu, X.; Koszyk, F. J.; Collins, P. W.; Cogburn, J. N.; Isakson, P. C.; Koboldt, C. M.; Masferrer, J. L.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Yuan, J.; Yang, D. C.; Zhang, Y. Y. Selective cyclooxygenase-2 inhibitors: heteroaryl modified 1,2-diarylimidazoles are potent, orally active antiinflammatory agents. *J. Med. Chem.* **2000**, *43*, 3168−3185.
(17) Reitz, D. B.; Li, J. J.; Norton, M. B.; Reinhart, E. J.; Collins, J. T.; Anderson, G. D.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Isakson, P. C. Selective Cyclooxygenase Inhibitors: Novel 1,2-Diarylcyclopentenes are Potent and Orally Active COX-2 Inhibitors. *J. Med. Chem.* **1994**, *38*, 3878−3881.
(18) Li, J. J.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Collins, J. T.; Garland, D. J.; Gregory, S. A.; Huang, H. C., Isakson, P. C.; Koboldt, C. M.; Logusch, E. W.; Norton, M. B.; Perkins, W. E.; Reinhart, E. J.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y.; Reitz, D. B. 1,2-Diarylcyclopentenes as Selective Cyclooxygenase-2 Inhibitors and Orally Active Antiinflammatory Agents. *J. Med. Chem.* **1995**, *38*, 4570−4578.
(19) Li, J. J.; Norton, M. B.; Reinhart, E. J.; Anderson, G. D.; Gregory, S. A.; Isakson, P. C.; Koboldt, C. M.; Masferrer, J. L.; Perkins, W. E.; Seibert, K.; Zhang, Y.; Zweifel, B. S.; Reitz, D. B. Novel Terphenyls as Selective Cyclooxygenase-2 Inhibitors and Orally Active Antiinflammatory Agents. *J. Med. Chem.* **1996**, *39*, 1846−1856.
(20) Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y. Y.; Isakson, P. C. Synthesis and Biological Evaluation of the 1,5-Diarylpyrazole Class of Cyclooxygenase-2 Inhibitors: Identification of 4-[5-(4-Methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-1-yl]benzenesulfonamide (SC-58635, Celecoxib). *J. Med. Chem.* **1997**, *40*, 1347−1365.
(21) Huang, H. C.; Li, J. J.; Garland, D. J.; Chamberlain, T. S.; Reinhart, E. J.; Manning, R. E.; Seibert, K.; Koboldt, C. M.; Gregory, S. A.; Anderson, G. D.; Veenhuizen, A. W.; Zhang, Y.; Perkins, W. E.; Burton, E. G.; Cogburn, J. N.; Isakson, P. C.; Reitz, D. B. Diarylspiro[2.4]heptenes as Orally Active, Highly Selective Cyclooxygenase-2 inhibitors: Synthesis and Structure−Activity Relationships. *J. Med. Chem.* **1996**, *39*, 253−266.
(22) Talley J. J.; Brown, D. L.; Carter, J. S.; Graneto, M. J.; Koboldt, C. M.; Masferrer, J. L.; Perkins, W. E.; Rogers, R. S.; Shaffer, A. F.; Zhang, Y. Y.; Zweifel, B. S.; Seibert, K. 4-[5-Methyl-3-phenylisoxazol-4-yl]-benzenesulfonamide, Valdecoxib: a Potent and Selective Inhibitor of COX-2. *J. Med. Chem.* **2000**, *43*, 775−777.

(23) Oprea, T. I.; Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure−Activity Relationship of Human Immunodeficiency Virus (I) Protease Inhibitors. 2. Predictive Power Using Limited Exploration of Alternate Binding Modes. *J. Med. Chem.* **1994**, *37*, 2206−2215.
(24) SYBYL 6.6, Tripos Associates, Inc., 1699 South Hanley Road, St. Louis, MO 63144.
(25) Clark, M.; Cramer, R. D. III; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(26) Stewart, J. J. P. MOPAC: a Semiempirical Molecular Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−103.
(27) Clark, M.; Cramer III, R. D. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137−145.
(28) Kulkarni, S. S.; Kulkarni, V. M. Three-Dimensional Quantitative Structure−Activity Relationship of Interleukin 1-β Converting Enzyme Inhibitors: A Comparative Molecular Field Analysis Study. *J. Med. Chem.* **1999**, *42*, 373−380.
(29) Hawkey, C. J. COX-2 inhibitors. *Lancet* **1999**, *353*, 307−314.
(30) Portevin, B.; Tordjman, C.; Pastoureau, P.; Bonnet, J.; De Nanteuil, G. 1,3-Diaryl-4,5,6,7-tetrahydro-2*H*-isoindole Derivatives: A New Series of potent and Selective COX-2 Inhibitors in which a Sulfonyl Group Is Not a Structural Requisite. *J. Med. Chem.* **2000**, *43*, 4582−4593.
(31) Debnath, A. K. Three-Dimensional Quantitative Structure−Activity Relationship Study on Cyclic Urea Derivatives as HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis. *J. Med. Chem.* **1999**, *42*, 249−259.
(32) Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; Lonze, P.; Weithmann, K. U. Quantitative Structure−Activity Relationship of Human Neutrophil Collagenase (MMP-8) Inhibitors Using Comparative Molecular Field Analysis and X-ray Structure Analysis. *J. Med. Chem.* **1999**, *42*, 1908−1920.
(33) Recanatini, M.; Cavalli, A.; Belluti, F.; Piazzi, L.; Rampa, A.; Bisi, A.; Gobbi, S.; Valenti, P.; Andrisano, V.; Bartolini, M.; Cavrini, V. SAR of 9-Amino-1,2,3,4-tetrahydroacridine-Based Acetylcholinesterase Inhibitors: Synthesis, Enzyme Inhibitory Activity, QSAR, and Structure-Based CoMFA of Tacrine Analogues. *J. Med. Chem.* **2000**, *43*, 2007−2018.
(34) Ducrot, P.; Legraverend, M.; Grierson, D. S. 3D-QSAR CoMFA on Cyclin-Dependent Kinase Inhibitors. *J. Med. Chem.* **2000**, *43*, 4098−4108.

JM0101343