On the Use of Chemical Function-Based Alignments as Input for 3D-QSAR

Thierry Langer*,† and Rémy D. Hoffmann‡

University of Innsbruck, Institute of Pharmaceutical Chemistry, Innrain 52a, A-6020 Innsbruck, Austria, and Molecular Simulations SARL, Parc Club Orsay Université, 20 rue Jean Rostand, F-91838 Orsay, France

Received August 24, 1997

A set of 15 highly flexible competitive inhibitors of rat liver squalene epoxidase (EC.1.14.99.7) covering a wide activity range (IC_{50} = 2 nM−10 μM) has been investigated by three-dimensional quantitative structure−activity relationships (3D-QSAR). Conformational analysis of the ligands was done by a quasirandom sampling approach with sequential poling. The alignment rule has been defined by a chemical function mapping based method. A comparative molecular field analysis (CoMFA) was performed using interaction energy matrices generated within the GRID program. This approach was shown to yield predictive QSAR models.

INTRODUCTION

Progress in drug design depends on the ability to understand the interactions of drugs with the biological targets. Concepts for the description of ligand binding site interactions have been developed from the growing amount of information on three-dimensional (3D) structures of proteins. Often, however, the structure of the biological target is not available. In this case, the use of 3D-quantitative structure−activity relationship (QSAR) methods has proven to be useful in rationalizing binding data of sets of drug molecules.[1] When constructing a 3D-QSAR model, two major crucial steps must be taken into account: definition of a selfconsistent alignment rule and choice of the assumed active conformation for each compound. The latter step is by far the most difficult, especially when dealing with highly flexible compounds. Methods that investigate conformational space (e.g., molecular dynamics or Monte Carlo simulation together with cluster analysis[2,3]) are commonly used to find the best match between various ligands.

The automatic identification of pharmacophoric pattern based on atomistic comparison has been extensively described (e.g., Disco, Compair, Family[4]). Another promising approach for defining alignment rules focuses on the consideration of chemical features rather than explicit atoms.[5,6] Within the program CATALYST,[7] feature-based alignments are produced in a three-step procedure: (i) for each molecule, a conformational model is generated using a quasirandom sampling approach with sequential poling[8] that has been shown to produce a good coverage of the conformational space[9,10]; (ii) each conformer is examined for the presence of chemical features; and (iii) a set of chemical features common to the input molecules that correlates best with biological activity is determined. This 3D array of chemical features (called a hypothesis) provides a relative alignment for each input molecule. Considered chemical features can be H-bond acceptors and donors, aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups, and aromatic planes.[11] Multiple hypotheses are produced and ranked according to how well they explain the biological activity of the input molecule set. In a hypothesis, each function is characterized by a location constraint (*x*,*y*,*z* coordinates).

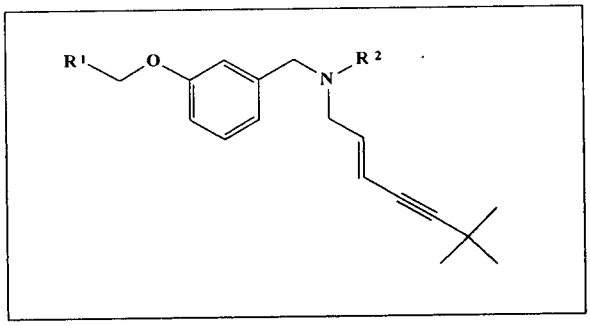
To investigate the utility of feature-based alignments as input for other 3D-QSAR methods, we analyzed a set of 15 highly flexible competitive inhibitors of rat liver squalene epoxidase (EC.1.14.99.7) covering an affinity range of 2 nM to 10 μM (see Figure 1). Examining the two-dimensional (2D) structures of these squalene epoxidase inhibitors shows clearly the functions that may be involved in binding. However, because of their high degree of flexibility and the lack of a 3D structure of the enzyme, it is difficult to determine which 3D arrangement of these functions will be the most suitable to explain the differences in activity within this set of inhibitors. Conformational analysis and hypothesis generation was performed within the CATALYST program. Then, a 3D-QSAR model using the CoMFA[12] technique together with GRID[13,14] was constructed starting from the alignment resulting from a combination of the two best hypotheses.

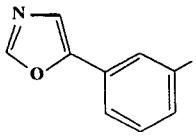
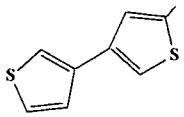
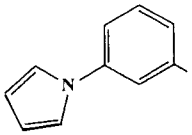
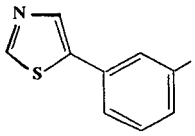
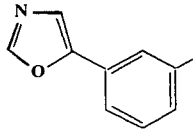
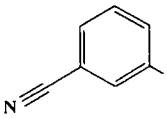
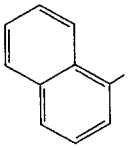
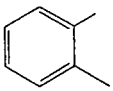
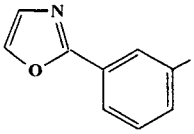
MATERIAL AND METHODS

Alignment Definition. The entire study was performed using the following software packages: CATALYST,[7] SYBYL,[15] and GRIN/GRID[13,14] installed on a SGI Indy 4400 or a SGI PowerChallenge XL. All structures were generated using the 2D/3D editor sketcher in CATALYST and minimized to the closest local minimum (CHARMm force field, energy convergence = 1.10^{-5} kcal/mol/Å, gradient convergence = 1.10^{-4} kcal/mol/Å). Conformational models were calculated using a 5 kcal energy cutoff (minimization convergence criteria during conformational analysis: energy convergence = 1.10^{-4} kcal/mol/Å, gradient convergence = 1.10^{-3} kcal/mol/Å). All molecules with their associated conformational models were regrouped into a spreadsheet including the biological data (IC_{50} values given in nM units). Hypothesis generation was performed using the positive ionizable, hydrophobic, and H-bond acceptor functions.[11] A reduced interfeature distance spacing parameter of 1.5 Å

† University of Innsbruck.
‡ Molecular Simulations Sarl.

S0095-2338(97)00448-4 CCC: $15.00
Published on Web 02/18/1998



R^1	R^2	Double Bond Geometry	Name	Affinity IC_{50} (nM)	Reference
	Ethyl	Trans	Cpd-120	2.7	18
	Ethyl	Trans	NB598	4.4	16
	Ethyl	Trans	Cpd-119	7.1	18
	Ethyl	Trans	Cpd-121	26	18
	Methyl	Trans	Cpd-78	63	18
	Ethyl	Trans	Cpd-46	150	18
	Methyl	Trans	Cpd- E-19	250	18
	n-propyl	Trans	Cpd-37	270	18
	Methyl	Trans	Cpd-112	690	18
	Methyl	Trans	Cpd-76	1200	18

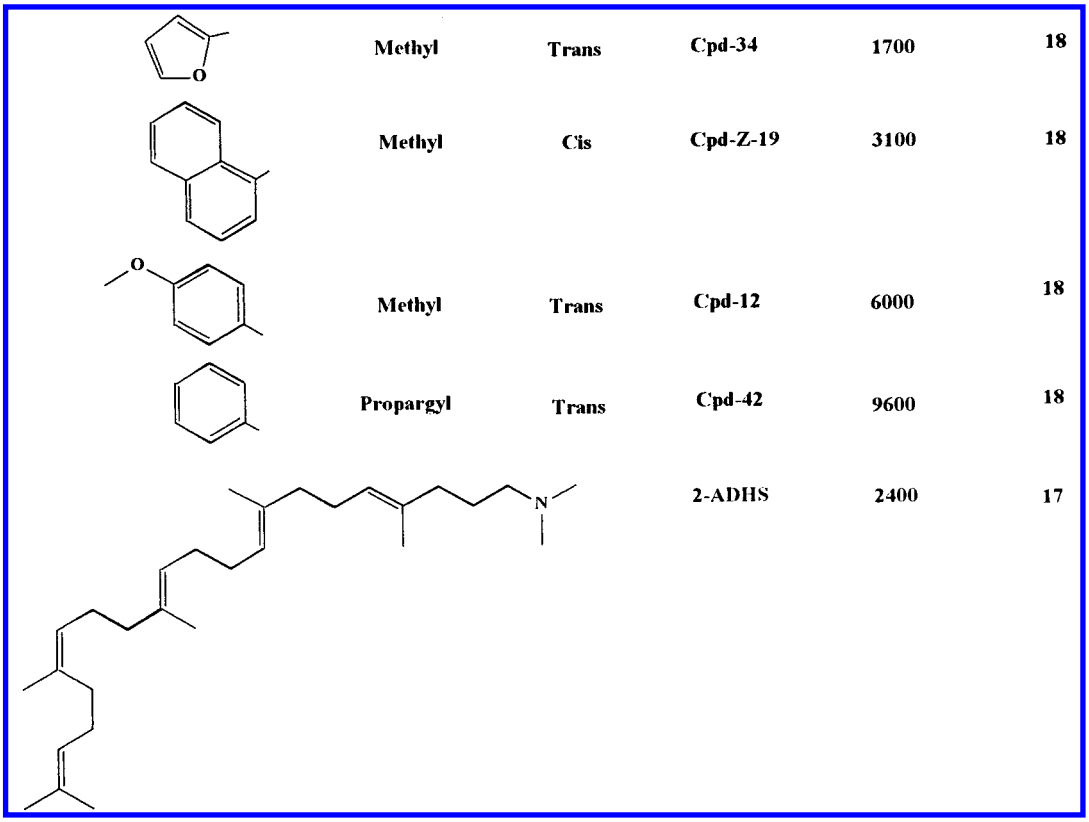


Figure 1. Compounds investigated.

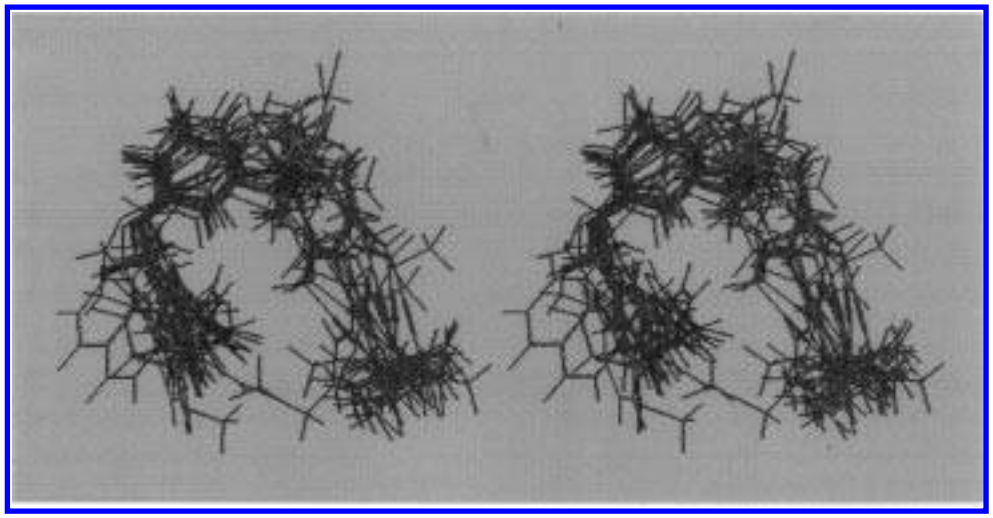


Figure 2. Stereoview of final alignment as used for construction of the 3D-QSAR model.

rather than the default value of 3 Å was used to allow hydrophobic mapping of *N*-alkyl substituents. In this series of molecules, *N*-ethyl derivatives are more active than *N*-methyl ones. The final hypothesis consisting of seven features (see Figure 2) was obtained by merging the two best hypotheses considering three features in each as common anchor points. All molecules were sequentially analyzed using the final hypothesis to determine their alignment mode. The structures obtained using this procedure were exported in their final alignment (see Figure 2).

3D-QSAR: CoMFA/GRID. The CoMFA runs were done with interaction energy matrices generated using the GRID force field. A 3D box was chosen (2 Å grid spacing in *x*, *y*, and *z* directions; box size 22 × 24 × 20 Å; 1320 points) assuring that every grid in all directions protruded at least 4 Å beyond the shape of each molecule. The probes C-sp^3, water, ammonium, and carbonyl oxygen of the GRID parameter set were chosen and the resulting fields were imported into the SYBYL QSAR module without scaling of the energy values. The linear expression of the CoMFA results was calculated with the partial least-squares analysis (PLS) algorithm[19] in conjugation with the cross-validation procedure ('leave-one-out' technique). This method provides a determination of the optimal number of components and permits an evaluation of predictivity of the model as indicated by the highest correlation (predictive r^2) value. The standard deviation threshold for exclusion of columns from the PLS analysis was set to 2.0. The PLS analysis of the descriptors with the optimum number of components but without cross-validation afforded conventional r^2 values.

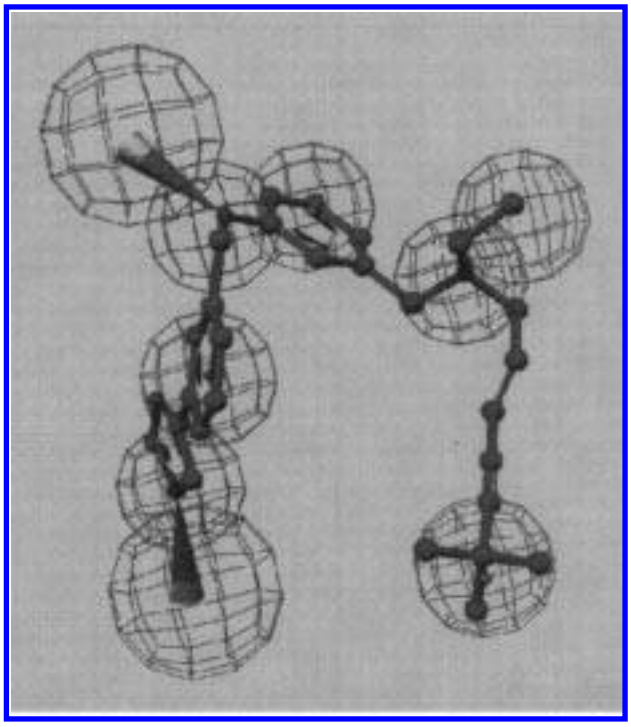


Figure 3. **Cpd-120** mapped on the hypothesis generated for squalene epoxidase inhibitors (red: positive ionizable; cyan: hydrophobic; green: H-bond acceptor; experimental affinity: 2.7 nM; estimated affinity: 7.2 nM; ΔE: 4.62 kcal/mol).

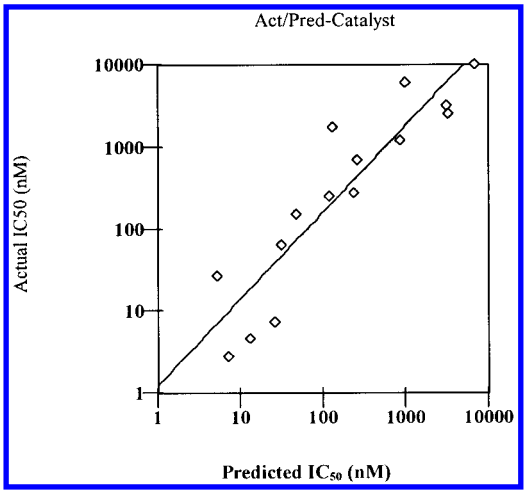


Figure 4. Regression diagram CATALYST.

RESULTS AND DISCUSSION

CATALYST. The CATALYST hypothesis used to generate the molecule alignments consists of seven functions, as depicted in Figure 3: one for positive ionizable groups that maps the basic nitrogen of all the molecules, two H-bond acceptors (the first generally maps an aromatic heteroatom, the second one maps an sp^3 ether oxygen), and four hydrophobic interactions (generally localized on the *t*-butyl group, the *N*-alkyl chain and the two aromatic rings).

Figure 3 shows **Cpd-120**, the most active molecule in the training set, aligned on the hypothesis. For all the molecules in the training set, reasonable low-energy conformers that align on the hypothesis were found (see Table 1). The overall ability of this hypothesis to estimate properly the activities of all molecules within the training set is shown by the the good r^2 value between predicted and estimated affinities ($r^2 = 0.83$, Figure 4).

3D-QSAR: GRID/CoMFA. It has already been pointed out by Folkers et al.[20] that using the GRID force field in a CoMFA study may lead to better results than using the original Tripos parameters (van der Waals radius, Lennard-Jones parameters) for the present resolution (2 Å). PLS analyses were done independently for all four different interaction energy matrices. The difference between the results originating from different probes (single-atom or multi-atom probes), however, was small (Table 2). All models obtained can be considered to be predictive, as shown by a cross-validated r^2 (r^2_{cv}) of >0.5 (Table 3). The best correlation ($r^2_{cv} = 0.542$) in the cross-validated models was found when the interaction energy matrix was calculated with the ammonium probe. The final model calculated was obtained using all interaction energy matrices together ($r^2 = 0.954$).

General Aspects. CATALYST aligns molecules according to their chemical functionality. Therefore, structurally different molecules that present the same 3D chemical features can be aligned together without necessity of atom-by-atom comparison. This method constitutes a different

Table 1. Characteristics of the Aligned Conformers

compound	number of conformers	activity IC_{50} (nM)	predicted activity (nM)	error[a]	conformer used	ΔE (kcal/mol)[b]	fit index[c]
Cpd-120	96	2.7	7.2	2.67	60	4.62	4.25
NB 598	95	4.4	13	2.95	76	2.59	3.98
Cpd-119	88	7.1	26	3.66	55	2.52	3.69
Cpd-121	99	26.0	5.2	−5	3	2.35	4.39
Cpd-78	99	63.0	31	−2.03	38	3.72	3.62
Cpd-46	49	150.0	47	−3.19	6	4.30	3.44
Cpd-E-19	53	250.0	120	−2.08	32	4.50	3.03
Cpd-37	48	270.0	240	−1.13	33	3.60	2.72
Cpd-112	72	690.0	260	−2.65	53	1.55	2.69
Cpd-76	117	1200.0	850	−1.41	29	4.38	2.18
Cpd-34	61	1700.0	130	−13.07	48	3.27	2.98
2-ADHS	50	2400.0	3200	1.33	21	3.89	1.60
Cpd-Z-19	27	3100.0	3100	1	6	4.59	1.62
Cpd-12	50	6000.0	1000	−6	14	4.68	2.73
Cpd-42	77	9600.0	6700	−1.43	67	3.63	1.28

[a] Error: when predicted > experimental: error = predicted/experimental; when predicted < experimental: error = experimental/predicted. [b] ΔE (kcal/mol): Energy difference between the conformer used for mapping and the global minimum calculated by CATALYST. [c] Fit index: fit = $\Sigma_{\text{mapped hypo functions}}\, w[1 - \Sigma_{\text{hypo function}}\, (\text{disp/tol})^2]$, w: weight of the hypothesis function sphere, adaptively determined; disp: distance of the mapping function on the molecule to the hypothesis sphere center; tol: radius of a hypothesis function sphere.

Table 2. GRID/CoMFA: Actual and Predicted Affinities and Residuals

	log(1/IC_{50} nM)		
compound	actual	calculated	residual
Cpd-120	−0.43	−0.61	0.18
NB-598	−0.64	−0.82	0.18
Cpd-119	−0.85	−0.95	0.10
Cpd-121	−1.41	−0.90	−0.51
Cpd-78	−1.80	−1.92	0.12
Cpd-46	−2.18	−2.10	−0.08
Cpd-E-19	−2.40	−2.50	0.10
Cpd-37	−2.43	−2.35	−0.08
Cpd-112	−2.84	−3.07	0.23
Cpd-76	−3.08	−3.17	0.09
Cpd-34	−3.23	−3.19	−0.04
Cpd-Z-19	−3.49	−3.49	0.00
Cpd-12	−3.78	−3.24	−0.54
Cpd-42	−3.98	−3.38	−0.60
2-ADHS	−3.38	−3.48	0.10

Table 3. Statistical Results of GRID /CoMFA Cross-Validated Analysis[a]

		S_{press}				
GRID-probe	r^2_{cv}	**comp1**	**comp2**	**comp3**	**comp4**	**comp5**
C sp^3	0.526	0.783	0.885	0.961	0.990	1.033
water	0.532	0.778	0.816	0.873	0.938	1.003
ammonium	0.542	0.769	0.822	0.874	0.936	0.972
carbonyl O	0.521	0.787	0.843	0.876	0.910	0.972
all together	0.535	0.775	0.837	0.884	0.938	0.953

[a] Final analysis without cross-validation (2 comp.) all together, $r^2 = 0.954$; $s = 0.253$; F ($n_1 = 3$, $n_2 = 11$) = 125.548.

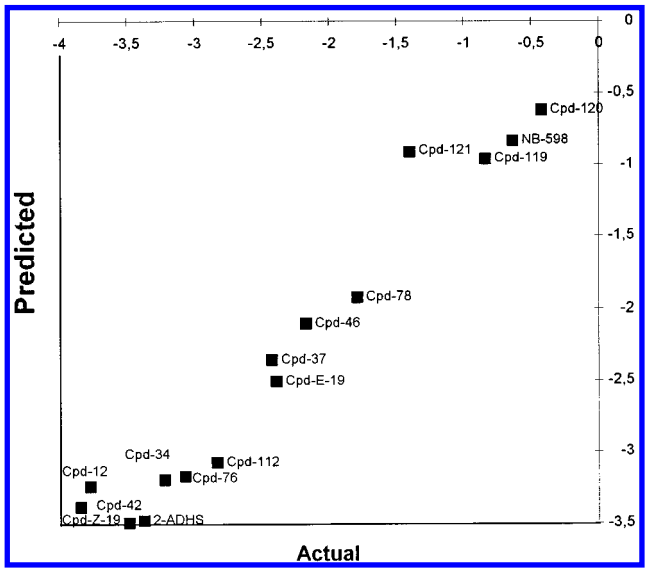


Figure 5. Actual *versus* predicted CoMFA-2: log(1/IC_{50} nM).

but straightforward way of representing ligand receptor/enzyme interactions that avoids the bias that can be introduced by relying on atom-based alignment. Molecules that belong to different structural families can be considered conveniently because different chemical groups or atoms can belong to the definition of a given function.[11] Moreover, because the conformational models are generated over a specified energy range above the computed global minimum, one can expect reasonable conformers. The models derived with this method can also predict the affinities for the molecules in the training set (see Table 1 and Figure 4). Prediction of activity is based on a geometric fit of the functions of the molecules to the chemical features in the hypothesis. This geometric fit procedure is particularly efficient when one considers only 3D pattern recognition, but is limited if analysis of more subtle chemical phenomena is the final goal. The use of additional 3D-QSAR techniques in such cases can be more insightful.

The advantage of the CoMFA/GRID approach is that the results of the analysis can be mapped back into 3D space. Moreover, the descriptor probes used within GRID are more relevant to the structure−activity correlation problem, because they can be considered as a representative selection among the variety of the main interaction modes with amino acids. In this paper, we focused only on the use of feature-based alignments to derive a 3D-QSAR model. One should keep in mind that physicochemical properties (log *P*, dipole, etc.) and entropic factors (that are not taken into consideration here) can also play a role.

CONCLUSION

Two different 3D-QSAR approaches have been tested on a set of 15 highly flexible rat liver squalene epoxidase inhibitors using a chemical feature-based alignment rule. Both models exhibited satisfactory predictivity within an acceptable range of error. The results of this investigation clearly indicate the potential usefulness of feature-based alignments as starting points for a CoMFA study, especially when no 3D structure of the biological target is available.

ACKNOWLEDGMENT

The authors thank P. Goodford and Molecular Discovery Ltd. for making available the GRIN/GRID software. Computational facilities generously provided by the *'EDV-Zentrum der Universität Innsbruck'* (H. Bielowski and O. Wörz) are gratefully acknowledged. P. Sprague is thanked for helpful discussion.

REFERENCES AND NOTES

(1) *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM. Leiden, 1993.

(2) Perkins, T.; Dean, P. An exploration of a novel stratedgy for superposing several flexible molecules *J. Comput.-Aided. Mol. Des.* **1993**, *7*, 155−172.

(3) Wermuth, C. G.; Langer, T. Inhibitors of prolyl endopeptidase: Characterization of the pharmacophore using conformational analysis and 3D-QSAR *J. Comput.-Aided. Mol. Des.* **1993**, *7*, 253−262.

(4) Martin, Y.; Bures, M.; Danaher, E.; DeLazzer, J. New strategies that improve the efficiency of the 3D design of bioactive molecules. In *Trends in QSAR and Molecular Modelling 92*, Wermuth, C. G., Ed.; ESCOM. Leiden, 1993; pp 20−27.

(5) Hoffmann, R. D.; Langer, T. Use of the Catalyst Program as a new alignment tool for 3D QSAR. In *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*; Sanz, F.; Giraldo, J.; Manaut, F., Eds.; Prous Science: Barcelona, 1995; pp 466−469.

(6) Palomer, A.; Giolitti, A.; García, M. L.; Cabré, F.; Mauléon, D.; Carganico, G. Molecular Modeling and CoMFA Investigations on LTD_4 Receptor Antagonists. In *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*; Sanz, F.; Giraldo, J.; Manaut, F., Eds.; Prous Science: Barcelona, 1995; pp 444−450.

(7) *CATALYST 3.1*; Molecular Simulations Inc.: San Diego, CA, 1996.

(8) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting Conformational Coverage *J. Comp. Chem.* **1995**, *16*, 171−187.

(9) Smellie, A.; Kahn, S. D.; Teig, S. L. An Analysis of Conformational Coverage 1. Validation and Estimation of Coverage *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285−294.

(10) Smellie, A.; Kahn, S. D.; Teig, S. L. An Analysis of Conformational Coverage 2. Applications *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 295−304.
(11) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical funtion queries for 3D database search *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 1297−1308.
(12) 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.
(13) Goodford, P. J. A computational procedure for determining energetically favourable binding sites an biologically important macromolecules *J. Med. Chem.* **1985**, *28*, 849−857.
(14) *GRIN/GRID V.11*; Molecular Discovery Ltd.: Oxford, U.K., 1993.
(15) *SYBYL 6.1*; Tripos Association: St. Louis, MO, 1994.
(16) Horie, M.; Tsuchiya, Y.; Hayashi, M.; Iida, Y.; Iwasawa, Y.; Nagata, Y.; Sawasaki, Y.; Fukuzumi, H.; Kitani, K.; Kamei, T. NB-598: a potent competitive inhibitor of squalene epoxidase *J. Biol. Chem.* **1990**, *265*, 18075−18078.
(17) Ryder, N. S.; Dupont, M. C.; Franck, I. Inhibition of Fungal and Mammalian Sterol Biosynthesis by 2-Aza-2,3-dihydrosqualene. *FEBS Lett*. **1986**, *204*, 239−242.
(18) Takezawa, H.; Hayashi, M.; Iwasawa, Y.; Hosoi, M.; Iida, Y.; Tsuchiya, Y.; Horie, M.; Kamei, T. Eur. Pat. App. EP 0 318 860 A2.
(19) Wold, S.; Albano, C.; Dunn,W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*, Kowalski, B., Ed.; Reidel: Dordrecht, Netherlands, 1984.
(20) Folkers, G.; Merz, A.; Rognan, D. *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM. Leiden, 1993; p 583.

CI970448S