# Structure−Activity Relationship Study of the Metallocene Catalyst Activity in Ethylene Polymerization

Victor L. Cruz,*,† Javier Ramos,‡ Sonia Martinez,‡ Antonio Muñoz-Escalona,‡ and Javier Martinez-Salazar‡

*Centro Tecnico de Informatica, CSIC, Pinar, 19, E-28006, Madrid, Spain, and Instituto de Estructura de la Materia, CSIC, Serrano, 113bis, E-28006, Madrid, Spain*

*Received June 7, 2005*

Three-dimensional quantitative structure−activity relationship (3D-QSAR) methods have been successfully applied in the field of drug design to search for a correlation between detailed molecular structure and bioactivity. Other areas, such as the development of single-site catalysts for homogeneous olefin polymerization, could benefit from the potential advantages associated with the 3D-QSAR technique. Using 3D-QSAR, in particular the comparative molecular field analysis (COMFA) method, the experimental results obtained for polymerization activity have been successfully correlated with 3D structural descriptors such as steric, electrostatic, LUMO, and local softness fields calculated by density functional theory (DFT) methods. The good predictive ability of the models allows one to consider the application of the 3D-QSAR methodology as a valuable tool in the design of better catalysts for olefin polymerization.

## 1. Introduction

In the 1980s metallocene and other single-site catalysts activated by methylaluminoxane compounds (MAO) as cocatalyst were found to be highly active for the polymerization of olefins.[1] Over the last years, this important discovery in polymer science has stimulated the design and synthesis of new organometallic complexes as potential polymerization catalysts. It is widely recognized that the polymer growth takes place in the cationic metallocene alkyl complex, which forms a weakly or noninteracting ion-pair with an anionic MAO cage.[2,3] However, the role of the cocatalyst is still an open question.

The performance of the catalytic olefin polymerization using organometallic compounds is affected by several experimental variables. Among other factors, the temperature, solvent, nature of cocatalyst, cocatalyst/catalyst ratio, monomer pressure, and polymerization time are well-known to have a strong influence on the catalytic activity as well as on the microstructure of the formed polymers.[4] By reviewing the literature, it can be found that each author uses different polymerization conditions, which hinder the possibility of establishing useful relationships between catalyst structure and catalytic performance. By keeping all the experimental conditions constant, the observed differences in the polymerization process could be directly attributed to the molecular structure of the organometallic catalyst. To establish empirical relationships between catalyst structures and polymerization activity or molecular weights of the produced polymers, some research groups have carried out systematic experimental studies of the polymerization of olefins with organometallic compounds under well-controlled experimental conditions. For the ethylene polymerization, Kaminsky's group has analyzed the catalytic activity and molecular weights of the produced polymers by using 31 different unbridged and bridged metallocene catalysts under the same experimental conditions.[5] They have empirically concluded that the electron-donating bisindenyl metallocenes showed a very high activity, while metallocenes substituted with very bulky ligands, such as neomenthyl or methylcyclopentadienyl, yielded lower productivity. Furthermore, it could be inferred that metallocenes based on a carbon bridge are less active for ethylene polymerization than those based on a silicon bridge. This has also been observed by us using different catalyst structures and other experimental conditions.[6] Alt et al.[7] have also reported another systematic study concerning the influence of the metallocene structure on the catalytic performance of the ethylene polymerization process. These authors concluded that there are many factors involved in the polymerization activity and that qualitative trends could be appropriately inferred through molecular modeling or systematic experiments. Among others, the previous examples show that the catalyst structure has a strong influence on the catalytic

* To whom correspondence should be addressed. Fax: +34915616193. E-mail: victor@cti.csic.es.
† Centro Tecnico de Informatica.
‡ Instituto de Estructura de la Materia.

(1) Sinn, H.; Kaminsky, W.; Vollmer, H. J.; Woldt, R. DE Patent Appl. p 3007725, 1980.
(2) Zurek, E.; Ziegler, T. *Prog. Polym. Sci*. **2004**, *29*, 107.
(3) Lanza, G.; Fragala, I. L.; Marks, T. J. *Organometallics* **2002**, *21*, 5594.
(4) Reddy, S. S.; Sivaram, S. *Prog. Polym. Sci.* **1995**, *20*, 309.
(5) Kaminsky, W. *Macromol. Chem. Phys.* **1996**, *197*, 3907.
(6) Haider, N.; Exposito M. T.; Muñoz-Escalona, A.; Ramos, J.; Méndez, L.; Martinez-Salazar, J. *Polym. Sci. Part A* **2004**, submitted for publication.
(7) Alt, H. G.; Koppl, A. *Chem. Rev.* **2000**, *100*, 1205.

10.1021/om050458f CCC: $30.25 
Publication on Web 09/13/2005

performance. Therefore, systematic experiments keeping all experimental conditions but catalyst structure are highly valuable for quantitative structure−activity relationship studies.

The quantitative structure−activity relationship (QSAR) approach is a rational method widely used in medicinal chemistry for the design of new and more efficient drugs.[8] Classical QSAR methods are based on statistical correlation of biological activities of drugs with physicochemical properties or scalar descriptors that encode certain structural features of the ligands.[9] However, they are based on a two-dimensional description of the structures neglecting the three-dimensional (3D) shape of real molecules. Therefore, different 3D-QSAR methods have been developed. This methodology is based on the assumption that drug molecules interact with biological targets accommodating the 3D drug structure in the biological receptor pocket. This accommodation is driven by 3D molecular interaction potentials. The method attempts to identify spatial regions in the drug that could be candidates for complementary molecular properties that match the biomolecule binding site. Among the 3D-QSAR methods, the comparative molecular field analysis (CoMFA) is one of the most frequently used. A database of molecules with known properties, called a "training set", is appropriately aligned in 3D space following an alignment rule. Then, one can calculate a field (steric, electrostatic, molecular orbital fields, and so on) in a grid of points around each molecule belonging to the training set that characterizes its local shape or reactivity. Then, using the partial least squares (PLS) method[10] correlations between these fields and a property of interest are found. This chemometric tool allows the treatment of hundreds or thousands of variables, the field values at each grid node, which might be more or less correlated, by extracting a few independent uncorrelated "latent variables". These "latent variables" are composed by linear combinations of the original variables that can now be used for the regression process. Once a correlation model between molecular structure and activity/property is found, any number of compounds, including those not yet synthesized, could be screened on a computer. This allows one to select the most promising precursors having a set of desired properties. Finally, these compounds could be synthesized and tested in the laboratory.

An area that can benefit from the potential advantages associated with the 3D-QSAR technique is the development of single-site catalysts for homogeneous olefin polymerization. The monomer insertion in the growing polymer chain at the catalyst active site resembles the macromolecule-ligand picture in drug design. In this case the aim is to model the catalyst structure, while in medicinal chemistry one tries to design the drug molecule. A limited number of articles dealing with the application of the QSAR approach to the design of polymerization catalysts have been published.[11−13] However, none of them take into consideration the 3D-QSAR approach.

The first attempt reported using 3D-QSAR CoMFA analysis in metallocene-based catalysts for ethylene polymerization has been recently published by us.[14] In this study, we used a set of metallocene catalysts for determining polymerization activity and polymer molecular weight under the same experimental reaction conditions. Descriptors obtained from DFT (density functional theory) calculations, such as LUMO (lowest unoccupied molecular orbital), electrostatic, steric, and local softness fields, were used in order to describe the structure of the catalysts. The electronic interaction was confirmed by correlations found between activity and LUMO (lowest unoccupied molecular orbital) as well as between activity and local softness. The model revealed that the experimental variance in catalytic activity is well explained in terms of the arrangement of the ligands around the metal center of the aromatic ligands (i.e., Cp or Ind). Furthermore, it was found that in ansa-type metallocene catalysts the bridges were not directly involved in LUMO and local softness fields, but rather in the Cp−Zr−Cp angle, promoting an electronic interaction between the metal center and the atoms of the ligands to a greater or lesser extent. Polymer molecular weight was found to correlate also with those fields, the Cp−Zr−Cp angle being the key geometric variable. However, steric fields were not able to explain the variance in the molecular weight data.

In the present work we have extended the analysis to a larger and more varied type of metallocene catalyst systems extracted from an experimental study conducted by the Kaminsky group.[5] The results obtained in this new 3D-QSAR analysis are in agreement with the correlations found in our previous study[14] with an additional implication of the electrostatic field, not found before. The good predictive ability of the models encourages the application of the 3D-QSAR methodology as a valuable tool for the design of better catalysts for olefin polymerization.

## 2. Methods

**2.1. Source of Catalysts and Polymerization Data.** A training set of 25 metallocenes was taken from the paper by Kaminsky.[5] The compounds considered along with the corresponding activities are shown in Table 1. As it can be seen, this training set covers a wide variability of metallocene structures with different aromatic ligands, bridges, and metals. Furthermore, this set is significantly greater than the set of seven catalysts used in our previous paper.[14]

All polymerization reactions were carried out under the same experimental conditions:[5] a reactor temperature of 30 °C, an ethylene pressure of 2.5 bar, a metallocene concentration of $6.25 \times 10^{-6}$ mol/L, and a molar ratio MAO(methyl aluminoxane)/metallocene of 250.

**2.2. Molecular Modeling.** The active species of the catalyst is a cationic organometallic complex with a vacant coordination site where polymerization takes place and can interact with the cocatalyst. Thus, the DFT calculations are based on the cationic species rather than on the precursor metallocene itself.

(8) Martin, Y. C. *Quantitative Drug Design*; Marcel Dekker: New York, 1978.
(9) Kubinyi H. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 448−460.
(10) Cruciani, G.; Baroni, M.; Clementi, S.; Costantino, G.; Riganelli, D.; Skagerberg, B. *J. Chemom.* **1992**, *6*, 335.
(11) Möring, P. C. *J. Mol. Catal.* **1992**, *77*, 41.
(12) Yao, S.; Shoji, T.; Iwamoto, Y.; Kamei, E. *Comput. Theor. Polym. Sci.* **1999**, *9*, 41.
(13) Linnolahti, M.; Pakkanen, T. A. *Macromolecules* **2000**, *33*, 9205.
(14) Cruz, V.; Ramos, J.; Muñoz-Escalona, A.; Lafuente, P.; Peña, B.; Martínez-Salazar, J. *Polymer* **2004**, *45*, 2061.

**Table 1. Experimental Values**

| no. | catalyst | activity[a] | log activity |
|---|---|---|---|
| 1 | $Cp_2ZrCl_2$ | 60.90 | 4.11 |
| 2 | $Cp_2TiCl_2$ | 34.20 | 3.53 |
| 3 | $Cp_2HfCl_2$ | 4.20 | 1.44 |
| 4 | $(NmCp)_2ZrCl_2$ | 12.20 | 2.50 |
| 5 | $(C_5Me_5)_2ZrCl_2$ | 1.30 | 0.26 |
| 6 | $(C_5Me_4Et)_2ZrCl_2$ | 18.80 | 2.93 |
| 7 | $[O(SiMe_2Cp)_2]ZrCl_2$ | 57.80 | 4.06 |
| 8 | $[O(SiMe_2{}^tBuCp)_2]ZrCl_2$ | 11.70 | 2.46 |
| 9 | $[En(Ind)_2]ZrCl_2$ | 41.10 | 3.72 |
| 10 | $[En(Ind)_2]HfCl_2$ | 2.89 | 1.06 |
| 11 | $[En(2,4,7Me_3Ind)_2]ZrCl_2$ | 78.00 | 4.36 |
| 12 | $[En(IndH_4)_2]ZrCl_2$ | 22.20 | 3.10 |
| 13 | $[Me_2Si(Ind)_2]ZrCl_2$ | 36.90 | 3.61 |
| 14 | $[Ph_2Si(Ind)_2]ZrCl_2$ | 20.20 | 3.01 |
| 15 | $[Bz_2Si(Ind)_2]ZrCl_2$ | 12.20 | 2.50 |
| 16 | $[Me_2Si(2,4,7Me_3Ind)_2]ZrCl_2$ | 111.90 | 4.72 |
| 17 | $[Me_2Si(IndH_4)_2]ZrCl_2$ | 30.20 | 3.41 |
| 18 | $[Me_2Si(2Me\text{-}4,6{}^iPr_2Ind)_2]ZrCl_2$ | 18.60 | 2.92 |
| 19 | $[Me_2Si(2Me\text{-}4Ph\text{-}Ind)_2]ZrCl_2$ | 16.60 | 2.81 |
| 20 | $[Ph_2C(Ind)(Cp)]ZrCl_2$ | 3.33 | 1.20 |
| 21 | $[Me_2C(Ind)(Cp)]ZrCl_2$ | 1.55 | 0.44 |
| 22 | $[Me_2C(Ind)(3MeCp)]ZrCl_2$ | 2.70 | 0.99 |
| 23 | $[Ph_2C(Fluo)(Cp)]ZrCl_2$ | 2.90 | 1.06 |
| 24 | $[Me_2C(Fluo)(Cp)]ZrCl_2$ | 2.00 | 0.69 |
| 25 | $[Me_2C(Fluo)(Cp)]HfCl_2$ | 0.89 | −0.12 |

[a] Catalyst activity in kg PE (mol metallocene × h × [ethylene])$^{-1}$ × $10^{-3}$.

The reactivity descriptors considered in this work reflect some characteristics of the isolated reactants. This implies that the former are relevant only with respect to the initial interaction between the catalyst and the other species in the reaction medium. This type of information can be related to the reactivity only if the reaction has an early transition state, as it is the case with the ethylene insertion into the metal alkyl bond of the metallocene catalyst. Furthermore, the possible effect on the activity and molecular weight is assumed to be the same in all cases, although the cocatalyst is not taken into account explicitly in the present work. This assumption can be valid as far as the same cocatalyst and the same Al/Zr ratio are used for all polymerization reactions. For each organometallic cationic species, geometry optimization at B3LYP[15]/LANL2DZ[16] was performed using the Gaussian98 package.[17] The following 3D fields were evaluated in Cartesian grids: electron densities, electrostatic potential, HOMO and LUMO molecular orbitals. We also calculated electrostatic charges by fitting the electrostatic potential to nuclear positions according to the CHELPG[18] scheme. Steric and electrostatic 3D fields are calculated by the CoMFA method through the interaction between each catalyst and a probe atom. The probe atom should have specific charge and steric properties to evaluate the interaction energy at each particular point in the grid. The probe atom selected was a $sp^3$ C atom with a −1 point charge. This atom corresponds to atom C.3 in the Tripos Force Field[19] which was used to calculate van der Waals (steric) interactions. The value for the probe atom charge was selected to represent the effect of the electrostatic nature of either the ethylene or the anionic cocatalyst.

The so-called local softness is another 3D field used to evaluate catalyst reactivity. This field takes into consideration the change in electron density of the cationic active species under the influence of an incoming reagent. The local softness was defined by Yang and Parr as[20]

$$s(r) = \left(\frac{\delta\rho(r)}{\delta\mu}\right)_{v(r)} \tag{1}$$

where $\rho$ is the electron density, $\mu$ is the chemical potential, $v(r)$ is the external potential, and $r$ is the spatial coordinate. Geerlings[21] and Harbola et al.[22] suggested that local softness could be used as an intermolecular reactivity index. This assumption is based on the fact that the local softness can be expressed in terms of the global softness index weighted by the Fukui function, according to the equation

$$s(r) = f(r)\cdot S \tag{2}$$

where $f(r)$ is the Fukui function defined by Parr and Yang as[23]

$$f(r) = \left(\frac{\delta\rho(r)}{\delta N}\right)_{v(r)} \tag{3}$$

Due to the discontinuity of the first derivate with respect to the number of electrons, $N$, in eq 3, the following three functions have to be defined in a finite difference approximation, given as

$$f^+(r) \approx \rho_{N_o+1} - \rho_{N_o} \tag{4}$$

$$f^-(r) \approx \rho_{N_o} - \rho_{N_o-1} \tag{5}$$

$$f^0(r) \approx 1/2(\rho_{N_o+1} - \rho_{N_o-1}) \tag{6}$$

where $\rho_{N_o}$, $\rho_{N_o+1}$, and $\rho_{N_o-1}$ are the electron density of the system $N$, $N + 1$, and $N - 1$ electrons. These three indices can be calculated for a nucleophilic, an electrophilic, and a radical reaction, respectively. In our study, the $f^+$ index for a nucleophilic attack of the monomer on the cationic species of the catalyst is of special importance, as it is assumed that the ethylene insertion process corresponds to a nucleophilic attack to the active cationic species. The Fukui function has been recognized as an adequate intramolecular reactivity descriptor.

In a similar way, the global softness, $S$, can be calculated using the following approximation:

$$S = \left(\frac{\delta N}{\delta\mu}\right)_{v(r)} \approx \frac{1}{\mathrm{IE} - \mathrm{EA}} \tag{7}$$

where IE and EA are the ionization energy and electron affinity, respectively.[23] The global softness index is a characteristic property of each molecule. Local softness was qualified as the natural DFT concept for characterizing a reactive site.[21] LUMO and local softness can be considered useful representations of the electron density in the analysis of the nucleophilic reactions such as olefin polymerization catalyzed by metallocenes.

(15) Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098.
(16) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, E.; Burant, J. C.; Dapprich, S.; Millan, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuk, A. D.; Raghavachari, K.; Foresman, J. B.; Ciolowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T. A.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Reploge, E. S.; Pople, J. A. *Gaussian98* (Revision A.1); Gaussian Inc.: Pittsburgh, PA, 1998.
(18) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(19) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982.
(20) Yang, W.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6723.
(21) Geerlings, P.; De Proft, F.; Langenaeker, W. *Density Functional Theory: A Source of Chemical Concepts and a Cost-Effective Methodology for their Calculation*; Academic Press: New York, 1999; p 303.
(22) Harbola, M. K.; Chattaraj, P. K.; Parr, R. G. *Isr. J. Chem.* **1991**, *35*, 395.
(23) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford Univertsity Press: New York, 1989.

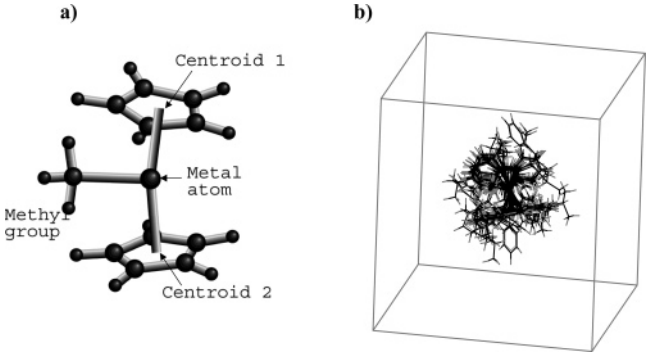


**Figure 1.** (a) Points used for the alignment rule. (b) Alignment of the training set molecules and cubic region used to calculate the 3D fields.

**2.3. COMFA Details.** All the reported 3D-QSAR analyses were done with the CoMFA module[24] implemented in the Sybyl package.[25]

**2.3.1. Alignment Rule.** It is essential in 3D-QSAR to align all structures in a common framework in order to make possible the comparison between all cationic active species of the tested catalysts.[26] In this study the catalyst molecules were aligned in such a way that the active site presented similar orientations. Zirconium atom, cyclopentadienyl centroids $Cp_1$ and $Cp_2$, and the alkyl carbon atom C3 attached to the metal were used for the molecular alignment, as represented in Figure 1a. The resulting alignment is shown in Figure 1b along with the cubic region used to calculate the molecular fields. Several grid spacings were used, but the best results are obtained with a lattice of 1.0 Å spacing. This value represents a compromise in the sense that a higher precision in the evaluation of the 3D field represented by a finer grid increases the so-called "brown noise" due to the sensitivity of the statistical technique applied to generate the models.[27]

**2.3.2. PLS Analysis.** The method used in 3D-QSAR analysis to handle thousands of descriptors to fit a few dependent variables is partial least squares (PLS).[28] Classical methods such as multiple linear regression (MLR), which yield the unique solution, cannot be used for such a situation with many more intercorrelated 3D descriptors than compounds. PLS calculates a few "latent variables" (LV) as linear combinations of the independent variable set.

$$\begin{aligned} \mathrm{LV}_1 &= b_{1,1}x_1 + b_{1,2}x_2 + \ldots \mathrm{b}_{1,n}x_n \\ \mathrm{LV}_2 &= b_{2,1}x_1 + b_{2,2}x_2 + \ldots b_{2,n}x_n \\ \mathrm{LV}_q &= b_{q,1}x_1 + b_{q,2}x_2 + \ldots b_{q,n}x_n \end{aligned} \tag{8}$$

where $\mathrm{LV}_i$ are the latent variables, $x_j$ are the independent variable set, and $b_{i,j}$ are the linear combinations coefficients.

These latent variables are used in the PLS equation:

$$Y = a_1\mathrm{LV}_1 + a_2\mathrm{LV}_2 + \ldots a_m\mathrm{LV}_\mathrm{m} \tag{9}$$

where $Y$ is a dependent variable and $a_1$ to $a_m$ are regression coefficients fitted by the PLS procedure. The calculations of the LVs and the regression step are carried out simultaneously in an iterative procedure. The properties of the LVs calculated in this way are as follows.

(1) The first latent variable explains maximum variance in the independent set; successive latent variables explain successively smaller amounts of variance.

(2) The latent variables conform to 1 with the provision that they are maximally correlated with the dependent variable.

(3) The latent variables are orthogonal to one another.

Finally the QSAR equation relating $Y$ with the 3D descriptors $x_j$ can be obtained by merging eqs 8 and 9.

$$Y = (a_1b_{1,1} + a_2b_{2,1} + \ldots a_mb_{m,1})x_1 + \ldots (a_1b_{1,n} + a_2b_{2,n} + \ldots a_mb_{m,n})x_n \tag{10}$$

This equation has two properties that are usually exploited in the drug design arena: modeling power and explanatory power. With the former the activity of any new catalyst can be predicted by introducing their associated descriptor values in the QSAR equation once the structure has been appropriately aligned. With the latter the molecular details that have more influence on the activity can be explained by looking at their coefficients in the QSAR equation. Due to the large number of descriptors used and their spatial nature, 3D graphical representations are essential to analyze the results.

PLS analyses were performed for different combinations of field descriptors. PLS calculations with the combined field were performed using the so-called autoscaling, where each field is scaled to have unit variance. The software calculates the standard deviation (SD) of each field and divides each value by the corresponding SD. The effect is to give each variable the same prior importance in the analysis. Leave one out (LOO)[29] cross-validated PLS analysis was initially performed to determine both the robustness of the statistical models and the optimal number of components or LVs. This can be achieved by examining the predictive residual sum of squares (PRESS) and the cross-validated regression coefficient ($q^2$) as guidelines. The $q^2$ statistic is defined as

$$\mathrm{PRESS} = \sum_{i=1}^{N}(Y_{\mathrm{obs},i} - Y_{\mathrm{pred},i})^2 \tag{11}$$

$$q^2 = 1 - \mathrm{PRESS/SSD} \tag{12}$$

where $Y_{\mathrm{obs},I}$ and $Y_{\mathrm{pred},I}$ are respectively the actual and predicted dependent variables and SSD is the sum of the squared deviations of each dependent variable from the mean of all dependent variables. It has been estimated by some authors[30] that a $q^2$ value greater than 0.3 has a 95% confidence limit. The usual practice in drug design is to consider valid a model with a $q^2$ greater than 0.5, i.e., halfway between perfect predictions ($q^2 = 1.0$) and no model at all ($q^2 = 0.0$). The optimum number of components was determined by minimizing PRESS while maximizing $q^2$ values. Whenever the increase in $q^2$ with an additional component was less than 5%, the model with fewer components was selected. Addition of more components improves the fitting statistics but has two disadvantages: on one hand, it complicates the model and on the other hand its predictive ability is lost. Finally, subsequent non-cross-validated PLS analysis was carried out for the optimum number of components to obtain a final model.

The CoMFA results are graphically represented as 3D maps, and for the sake of clarity only one catalyst of the set has been depicted in the corresponding figures.

## 3. Results

Different classes of scalar structural descriptors (analysis of accessible surface, electrostatic charges,

(24) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

(25) *SYBYL*, Molecular Modeling System; Tripos Inc.: 1699 S. Hanley Rd., St. Louis, MO 63144.

(26) Martin, Y. C.; Kim, K. H.; Liu, C. T. In *Advances in Quantitative Structure−Property Relationships*; Charton, M., Eds.; JAI Press: Greenwich, CT, 1996; Vol. 1, pp 1−52.

(27) Rännar, S.; Lindgren, F.; Geladi, P.; Wold, S. *J. Chemom.* **1994**, *8*, 111.

(28) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. SIAM *J. Sci. Stat. Comput.* **1984**, *5*, 735.

(29) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18.

**Table 2. PLS Statistics**

| 3D field | $q^2$/no comp. | SEP[a] | $r^2$ | SEE[a] |
|---|---|---|---|---|
| steric | 0.434/3 | 20.3 | 0.907 | 9.5 |
| electrostatic | 0.563/3 | 19.1 | 0.899 | 8.1 |
| local softness | 0.489/4 | 14.8 | 0.966 | 7.7 |
| LUMO | 0.533/5 | 11.4 | 0.973 | 5.6 |

[a] In kg PE (mol metallocene × h × [ethylene])$^{-1}$ × $10^{-3}$. SEP: standard error of prediction. SEE: standard error of estimate.

dipolar moment, and so on) were examined prior to the use of the 3D-QSAR methodology. No correlation was found between any combination of the scalar descriptors considered and the polymerization activity.

The region of points around each molecule where the structural descriptors are calculated is an essential component of the 3D-QSAR work. As was mentioned above, the best results were obtained with a grid spacing of 1.0 Å, which represents a compromise between the statistical noise introduced by the amount of variables considered and the fine details of the calculated field.

Table 2 shows the statistics obtained with the different fields considered in this work. Models built with any combination of fields give slightly poorer results than those models composed by any of the fields alone. This could be due to the different nature of the fields considered even though they are independently scaled to have the same variance.

It is generally assumed in the 3D-QSAR area that a cross-validated correlation coefficient ($q^2$) greater than 0.5 can be associated with reliable models. As can be seen in Table 1, the four models considered in this work give statistic values near or above 0.5. Thus they can be considered useful models with good predictive and modeling power.

The results obtained with each model are presented in the next subsections.

**3.1. Steric Field.** The model including the steric field descriptor, although still significant, gives the less reliable statistics of the set of models, with a $q^2$ value of 0.434 for three components (see Table 2). The final model including the whole training set gives an $r^2$ coefficient of 0.907 with a standard error of estimate (SEE) of $9.5 \times 10^3$ kg PE (mol metallocene × h × [ethylene])$^{-1}$.

One shortcoming that could be anticipated with this descriptor is the conformational variability associated with some compounds in the training set. It is a difficult task to decide which conformations should be included in the final alignment. Although several recommendations can be followed for the selection procedure, we have decided to take the calculated lowest energy conformation to represent each compound.

The PLS analysis gives information about the QSAR equation and its characteristics in terms of 3D fields. To this respect, it is generally accepted that the most informative field is that resulting from the product of the standard deviation at each grid point and the coefficient obtained for the same point in the QSAR equation.

The standard deviation (stdev) times the QSAR coefficient ($\beta$) field gives a rough location where structure−activity relationship statements can be inferred, discriminating areas where the local descriptor is important from those that have no significance.[31] Figure 2 shows isosurfaces for the stdev × $\beta$ field contoured at 0.1 (dark gray, positive values) and −0.1 (light gray, negative values) onto the most active molecule of the set. The positive values indicate where an increase of steric field corresponds to an increase in activity. Two main areas of positive steric influence can be observed, one near substituent positions in the aromatic ligands, indicating that addition of some bulky substituent will enhance activity. This could be interpreted in terms of the catalyst−cocatalyst interaction. A possible interpretation could be that bulky substituents in those positions would prevent the approach of the anionic cocatalyst species to the metal, so that an easier complexation and insertion of the ethylene monomer is facilitated. On the other hand, negative values are associated with locations where it is necessary to release steric impediment to increase activity. The light gray isosurfaces in Figure 2 correspond to those areas with a negative influence of the steric field on polymerization activity. These regions are located in the proximity of the active site, where monomer coordination will take place. Summing up the above results, one could conclude that the design of new catalysts should take into account the incorporation of bulky substituents in the aromatic ligands while leaving enough space at the active site area for the olefin approach.

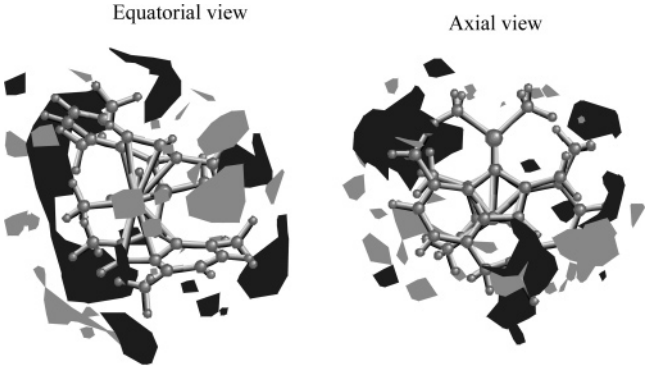


**Figure 2.** Standard deviation times coefficient CoMFA maps for the steric field. Equatorial view corresponds to the plane containing the metal atom and the methyl group representing the growing chain. Axial view corresponds to an axis through the centroid of the aromatic ligands.

**3.2. Electrostatic Field.** The electrostatic field exerted by a molecule outside its van der Waals radius is supposed to be a principal descriptor of intermolecular interaction in 3D-QSAR.[32] The electrostatic nature of the catalyst active species can play an important role in the case of ethylene polymerization with metallocene catalysts, where several molecules (monomer, solvent, or cocatalyst) can interact through electrostatic forces with the metallocene compound.

The PLS regression model obtained with this field gives a good cross-validated $q^2$ coefficient of 0.563 with three components and a standard error of prediction of $19.1 \times 10^3$ kg PE (mol metallocene × h × [ethylene])$^{-1}$. The final regression model with three components gives a correlation coefficient $r^2$ of 0.899 with a standard error

(30) Clark, M.; Cramer, R. D., III. *Quant. Struct.-Act. Relat.* **1993**, *12*, 137.
(31) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Kluwer-ESCOM: Dordrecht, 2000; pp 443−485.
(32) Wade, R. C. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Kluwer/ESCOM: Dordrecht, 2000; pp 486−505.

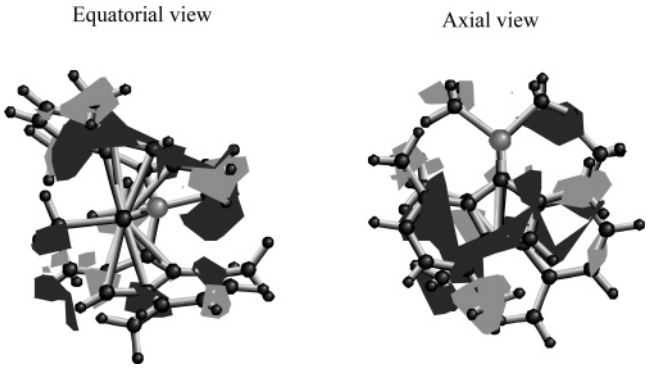


**Figure 3.** Standard deviation times coefficient CoMFA maps for the electrostatic field. Equatorial view corresponds to the plane containing the metal atom and the methyl group representing the growing chain. Axial view corresponds to an axis through the centroid of the aromatic ligands.

of estimate of $8.1 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$ (see Table 2).

Figure 3 shows the stdev $\times$ $\beta$ field derived after the PLS regression model of the polymerization activity versus electrostatic potential. Again, the dark gray areas correspond to those regions where addition of positive charge will enhance activity. The light gray represents areas where an increase in negative charge will also contribute to improve polymerization activity. It could be observed that most of the regions with beneficial positive charge lie around the C atoms of the aromatic ligands. The negative areas are positioned in the vicinity of the aromatic ring substituents. These areas correspond to the most external surface of the catalyst above the active site. These observations suggest that electron-withdrawing or negatively charged substituents in the aromatic rings would be able to facilitate the ion-pair separation, making more room for the ethylene coordination to the metal center. This might result in an improvement of the catalyst activity, as mentioned in the previous section on steric field.

The correlation of the electrostatic field with the activity was not observed in our previous report.[14] This can be explained by the absence of catalyst structures containing substituents on the aromatic ligands in the training set considered in the first work.

**3.3. LUMO Field.** Taking into consideration the frontier molecular orbital theory, the LUMO corresponds to molecular regions where the addition of electronic density is energetically more favorable. Olefin polymerization is considered to be a nucleophilic reaction, so that the incoming ethylene monomer will attack the catalyst species through the most favorable area for electron addition. Thus, the LUMO of the cationic species can be considered a suitable field descriptor of the polymerization activity.

The PLS regression model obtained with this field gives a good cross-validated $q^2$ coefficient of 0.533 with five components and a standard error of prediction of $11.4 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$. The final regression model with five components gives a correlation coefficient $r^2$ of 0.973 with a standard error of estimate of $5.6 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$ (see Table 2).

The stdev $\times$ $\beta$ field derived after the PLS regression model of the polymerization activity versus LUMO field

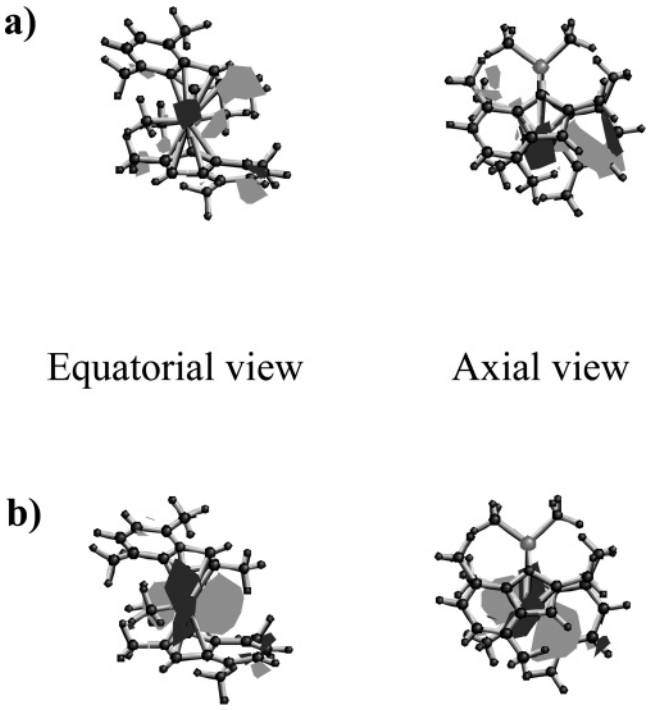


**Figure 4.** (a) Standard deviation times coefficient CoMFA maps for the LUMO field. (b) Three-dimensional shape of the LUMO field. Equatorial view corresponds to the plane containing the metal atom and the methyl group representing the growing chain. Axial view corresponds to an axis through the centroid of the aromatic ligands.

is shown in Figure 4. The observations are very similar to those reported in our previous work for a different set of metallocene catalysts. In that work it was explained what the shape of the LUMO orbital was and how the different parts of the molecule contribute to the formation of the orbital. It has also been observed in the present work that increasing the contribution to the LUMO from atomic orbitals belonging to atoms of the aromatic ligands will enhance the catalytic activity. The light gray areas around some aromatic ligand atoms observed in Figure 4 are indicative that a contribution from these atoms to the negative phase of the LUMO will enhance activity. Those areas correspond precisely to the location of the negative phase of the LUMO orbital, as can be seen in Figure 4b. On the other hand, the dark gray areas are indicative of where an increase of the LUMO positive phase will be beneficial for the polymerization activity. These regions are located in the vicinity of the metal center, which correspond to the positive phase of the LUMO orbital.

**3.4. Local Softness Field.** The concept of local softness has been related with the localization of reactive sites in a molecule where nucleophilic, electrophilic, or radical attack can take place.[20−23] The ethylene insertion into the metal−alkyl bond of the metallocene active species is considered as a nucleophilic reaction. Then, the field of interest should be the nucleophilic local softness, calculated as described in the computational methods section.

The PLS regression model obtained with this field gives a good cross-validated $q^2$ coefficient of 0.489 with four components and a standard error of prediction of $14.8 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$. The final regression model with four components gives

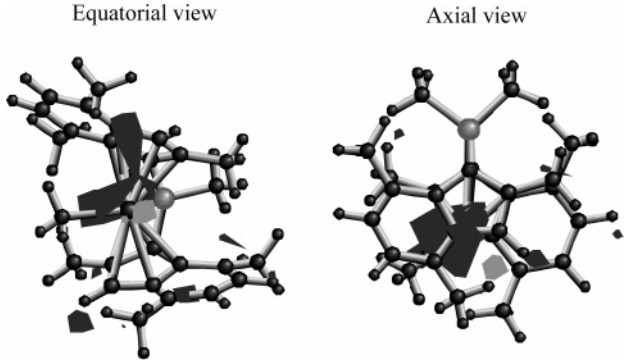


**Figure 5.** Standard deviation times coefficient CoMFA maps for the local softness field. Equatorial view corresponds to the plane containing the metal atom and the methyl group representing the growing chain. Axial view corresponds to an axis through the centroid of the aromatic ligands.

a correlation coefficient $r^2$ of 0.966 with a standard error of estimate of $7.7 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$ (see Table 2).

The stdev $\times$ $\beta$ field obtained for the final model is shown in Figure 5. The dark gray area corresponds to the region where an increase in local softness will enhance activity. It can be observed that this isosurface connects the metal−methyl bond with the aromatic cyclopentadienyl ligand. This is indicative of the important contribution of the aromatic ligand to the increase of the softness of the metal−alkyl bond where the insertion of the olefin occurs. This observation was also discussed in our previous work, so that it can be confirmed with the new training set considered in the present paper. The sensitivity of the active site to a nucleophilic ethylene insertion is increased in the case of indenyl with respect to the Cp ligands, resulting in higher polymerization activity.

**3.5. Combined Field.** The models obtained with different combinations of fields give slightly poorer results than those obtained with each field alone, as mentioned above. However an attempt was made to build a model including all the fields considered in this work just to give an idea of the relative contribution of each field to the overall activity data.

When all fields are considered, the resulting model gives a $q^2$ coefficient of 0.401 with five components and a standard error of prediction of $25.7 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$. The final regression model with five components gives a correlation coefficient $r^2$ of 0.95 with a standard error of estimate of $11.3 \times 10^3$ kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$. The relative contribution of each field was 23% steric, 17% electrostatic, 25% LUMO, and 35% local softness. This result shows a slight preponderance of electronic effects (about 60%) over steric or electrostatic interactions (about 40%) in the olefin insertion process. Figure 6 shows a plot of predicted versus experimental activity values for this model. As can be seen, a good correlation is obtained between both sets. It is worthwhile to mention that there is no catalyst markedly outside of the diagonal line.

This result shows a comparable importance of the contributions related to the catalyst structure used to explain the variance in the ethylene polymerization activity. The steric and electrostatic field descriptors are assumed to model mainly the ion-pair interaction between the catalyst and the cocatalyst. On the other hand, the LUMO and local softness fields try to model the electron transfer nature of the ethylene complexation and insertion processes. The global effect is that electron transfer seems to have more influence that the catalyst/cocatalyst intermolecular interaction.

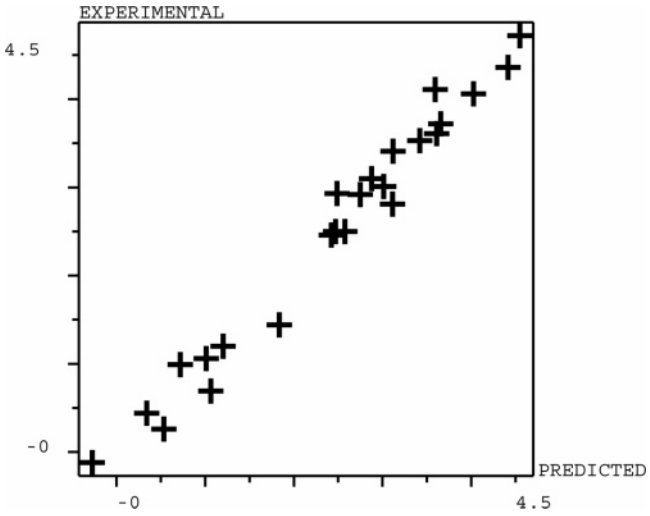


**Figure 6.** Experimental vs predicted activity plot. All values are given in log units.

**Table 3. Actual vs Predicted Activities for the Test Set**

| catalyst | experimental activity[a] | predicted activity[b] |
|---|---|---|
| $[(Me_2Si)(H_2C)(Cp)_2]ZrCl_2$ | 4.0 ± 0.9 | 3.6 (26.5) |
| $[En(Cp)_2]ZrCl_2$ | 2.0 ± 0.9 | 1.2 (8.9) |
| $[Me_2Si(Cp)_2]ZrCl_2$ | 1.0 ± 0.9 | 1.0 (7.4) |

[a] Catalyst activity in kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$ $\times$ $10^{-3}$. [b] Normalized activity. Values predicted by the model in parentheses in kg PE (mol metallocene $\times$ h $\times$ [ethylene])$^{-1}$ $\times$ $10^{-3}$.

**3.6. Predictive Performance.** In addition to the ability of the 3D-QSAR models to reveal important information about structural characteristics of the metallocene active species that influence the polymerization activity, it is possible to use such models to predict the activity of a different set of metallocene catalysts. A new compound can be easily tested as long as it is conveniently aligned with the training set structures and their 3D descriptors calculated.

The set of compounds considered in our previous work was selected as the test set in this case, which is used to check the predictive performance of the models derived with the Kaminsky training set. The CoMFA model used for the predictions was that corresponding to the LUMO field, which gives the best result in terms of standard error of prediction and $r^2$ correlation coefficient. Table 3 shows the formulas and experimental and predicted polymerization activities obtained for the test set compounds. Those catalysts that were already considered in the training set were not tabulated. It should be mentioned that the experimental activities measured at our laboratory were obtained under conditions different from those reported by Kaminsky. For this reason, the predicted activity values have been normalized so that the activity corresponding to the least active catalyst is unity.

An acceptable performance of the predictions can be observed despite the different polymerization conditions used in both sets.

## 4. Conclusions

The present paper is the second contribution to explore the use of the 3D-QSAR methodologies to analyze the polymerization activity in terms of structural descriptors of the metallocene catalyst. The organometallic nature of the system together with the bond-breaking and -forming characteristics of the olefin insertion reaction demands the usage of an adequate quantum mechanical treatment. The DFT framework provides the set of descriptors that characterize the catalyst structure. The hypothesis that the structure of the catalysts is the main factor that influences the polymerization activity validates the usage of quantitative structure−activity relationship (QSAR) methodologies to search those correlations. The localized nature of the active site where polymerization takes place and the relative rigidity of the metallocene structures make these systems suitable for the application of 3D-QSAR tools. The combination of DFT calculations and 3D-QSAR statistical tools applied to the study of a metallocene catalyst system is to our knowledge innovative in the scientific literature.

The selection of the methyl cationic active species as the structural framework responsible for the changes in polymerization activity seems to be reasonable in view of the acceptable PLS cross-validated statistics. However different active species as the $\beta$-agostic cationic ones should be taken into account in future studies. The models found for the steric and electrostatic fields can be interpreted in terms of catalyst/cocatalyst ion-pair interactions. Steric hindrance in specific positions and charge distribution around the aromatic ligands are correlated with an increase in activity. This could be explained by a weakening of the catalyst/cocatalyst interaction, which results in more room for the ethylene insertion reaction.

The models formed by electronic-based descriptors such as LUMO and local softness enhance the influence on the polymerization activity of the electron density redistribution at the metal active site due to the aromatic ligands.

The results obtained in the present work reinforce those presented in our previous paper. An additional correlation between experimental activity and electrostatic field is reported here and was absent in our previous study.[14] This could be due to the fact that the training set used in the first 3D-QSAR study did not contain any catalyst with substituents on the aromatic ring. The presence of such substituents is essential to explain the influence of electrostatic interactions on the polymerization activity.

The predictive performance of the model was assessed by the prediction of activities for the set of metallocene catalysts considered in our previous work.[14] The results were satisfactory despite that the experimental conditions were different in the training and the test sets.

Several issues remain to be investigated in order to improve the application of 3D-QSAR to the study of olefin polymerization catalyzed by metallocenes. The usage of auto- and cross-covariance (ACC)[33] transforms could partially solve the alignment problem. This arrangement of raw data also provides new data that take into account neighbor effects, i.e., the required continuity between grid nodes. On the other hand, the information given by DFT-derived descriptors should be used as complete as possible. This can be achieved by employing finer 3D grids with the subsequent increase in the number of variables entering the PLS analysis. To handle adequately this problem, it would be necessary to take into account any suitable variable selection method, such as GOLPE[34] (generating optimal linear PLS estimations).

**Acknowledgment.** Thanks are due to the CICYT (Grant MAT2002-01242) for financial support. S.M. acknowledges the MEC for the tenure of a fellowship.

OM050458F

(33) Clementi, S.; Cruciani, G.; Baroni, M.; Costantino, G. Series Design. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; Kluwer/ESCOM: Dordrecht, 2000; pp 443−485.

(34) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli. D.; Valigi, R.; Clementi, S. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9.