New Quantum Mechanics-Based Three-Dimensional Molecular Descriptors for Use in QSSR Approaches: Application to Asymmetric Catalysis

Manuel Urbano-Cuadrado,[†] Jorge J. Carbó,[‡] Ana G. Maldonado,[§,||] and Carles Bo*[,†,‡]

Institute of Chemical Research of Catalonia (ICIQ), Avinguda Països Catalans, 16, E-43007 Tarragona, Spain, Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, Marcelí Domingo, s/n, E-43007 Tarragona, Spain, Institut de Topologie et de Dynamique des Systèmes, CNRS UMR-7086, University Paris-7, 1, rue Guy de la Brosse, 75005 Paris, France, and van't Hoff Institute of Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Received May 28, 2007

This paper presents a new protocol based on 3D molecular descriptors using QM calculations for use in CoMFA-like 3D-QSSR. The new method was developed and then applied to predict catalytic selectivity in the asymmetric alkylation of aldehydes catalyzed by Zn-aminoalcohols. The molecular descriptors are obtained straightforwardly from the electronic charge density function, $\rho(r)$, and the molecular electrostatic potential (MEP) distributions. The chemically meaningful Molecular Shape Field (MSF) descriptor that accounts for the shape properties of the catalyst is defined from $\rho(r)$. Alignment independence was achieved by computing the product of the MSF and MEP values of pairs of points over a given distance range on a molecular isosurface and then selecting the product with the highest value. The new QSSR method demonstrated good predictive ability ($q^2 = 0.79$) when full cross-validation procedures were carried out. Accurate predictions were made for a larger data set, although some deviations occurred in the predictions for catalytic systems with low enantiodiscrimination. Analysis of this QSSR model allows for the following: (1) evaluation of the contribution of each functional group to enantioselectivity and (2) the molecular descriptors to be related to previously proposed stereochemical models for the reaction under study.

1. INTRODUCTION

Nowadays enormous efforts are being dedicated to obtaining enantiomerically pure compounds to use as building blocks in the synthesis of bioactive and pharmaceutical agents.[1] In silico tools, including those based on quantum mechanics, are widely used in fundamental studies aimed at rationalizing the selectivity outcome of catalyzed transformations.[2a] Other approaches known as QSAR or QSSR (Quantitative Structure−Activity/Selectivity Relationships) seek to identify quantitative relationships between catalyst structures and activity or selectivity. Their predictive and modeling techniques, usually applied to biological systems, are emerging as useful in asymmetric catalysis,[2b] and they are computationally fast and inexpensive.

To identify such quantitative relationships, molecular descriptors are produced, from either the molecular topology and related geometrical properties,[2c,3] or from 3D molecular interaction fields. These latter procedures, known as CoMFA[4]-related methods or 3D-QSSR, have recently demonstrated their predictive ability for a number of reactions.[5] Most CoMFA and CoMFA-like approaches use empirical molecular fields or semiempirical quantum mechanical methods to compute the interaction fields of a molecule under study, by means of a probe positioned at different points in a grid enclosing the molecular system. CoMFA is the most popular 3D-QSAR method, and CoMFA modeling has produced fairly good results in many pharmacological and fine chemistry contexts; a molecular interaction field (MIF) describes how a receptor responds to the chemical characteristics of a compound, making MIFs useful descriptors for correlating molecular properties with activity. The generation of CoMFA descriptors involves three stages: optimization of geometries designed to apply to stable molecular conformations; alignment of structures in order to obtain reproducible predictive spaces (a 3D grid point must represent the same chemical meaning for all data set elements); and computation of the interaction fields at each of the grid points.

We looked at several modifications of the classical CoMFA procedure for the transition-metal homogeneous asymmetric catalysis field and decided to use molecular fields determined directly from a quantum mechanics calculation in order to deal with the transition-metal complexes often involved in asymmetric catalysis. The molecular descriptors are assumed to be chemically meaningful and to account for electrostatic and steric properties of the molecules. Our new protocol uses an alignment-free approach which simplifies the classical procedure by skipping the alignment step. This allows the range of potential applications to be expanded and include diverse families of catalysts for which there is no clear alignment hypothesis. Sciabola et al.[6] have already shown that the use of grid-independent descriptors (GRIND)—introduced by Pastor et al.[7]—combined with empirical molecular fields can be successfully applied in asymmetric catalysis.

* Corresponding author e-mail: cbo@iciq.es.
[†] Institute of Chemical Reserach of Catalonia (ICIQ).
[‡] Universitat Rovira i Virgili.
[§] University Paris-7.
[||] University of Amsterdam.

10.1021/ci700181v CCC: $37.00
Published on Web 10/26/2007

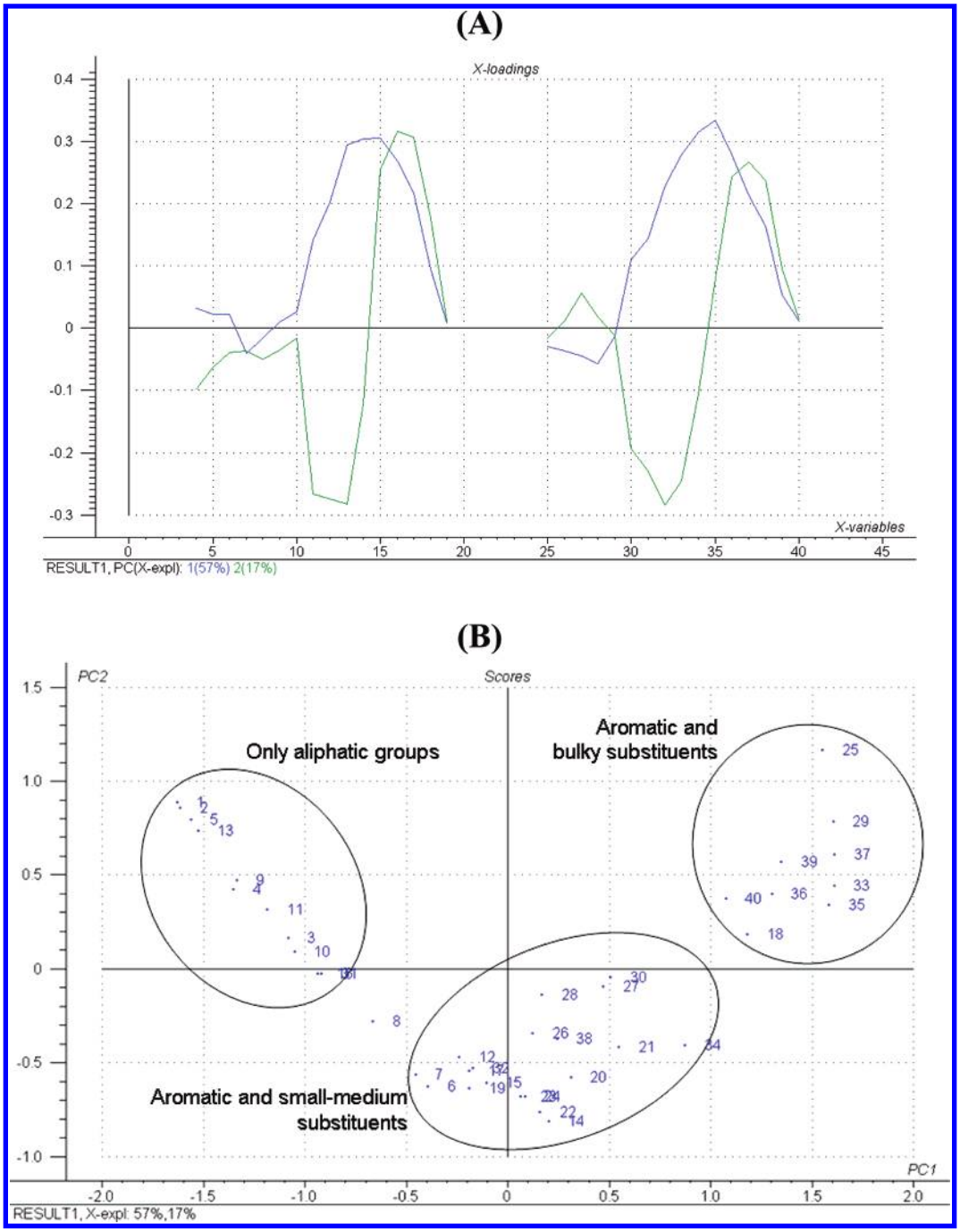
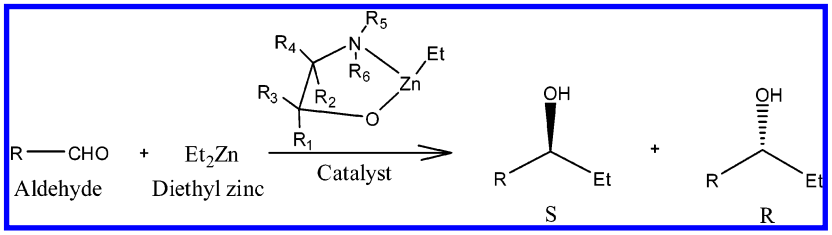


Figure 1. Loadings (A) and scores (B) for the two first principal components obtained in the PCA of MSF−MSF and MEP−MEP autocorrelograms for all catalysts studied. For the loadings, the first principal component (57% of the data variance) is shown in blue, and the second principal component (17% of the data variance) is shown in green; MSF−MSF variables are shown on the left side, and MEP−MEP variables are shown on the right side.

Scheme 1



Asymmetric alkylation of aldehydes catalyzed by Zn-aminoalcohols (see Scheme 1) was the first application tested, selected because of the high number of experimentally tested systems[8] and also the availability in the literature of both previous QSSR studies[5c,g,h,6] and fundamental theoretical studies which offer a rationale for the origin of enantioselectivity.[9] We initially chose a set containing the same 18 distinct catalytic systems used in previous QSSR studies and

Table 1. Statistical Parameters for Fitting ($N = 14$) and Test ($N = 4$) Stages and Experimental vs Predicted ees[a]

QSSR approach	r^2	q^2	ee (exp, pred)
ref 5c	0.96		(3,5), (86,75), (98,99), (63,75)
ref 6	0.99	0.69	(3,18), (86,83), (98,92), (63,66)
this work	0.96	0.63	(3,0), (86,47), (98,96), (63,82)

[a] $r^2 = 0.96$, fitting error = 0.22 kcal·mol^{-1}, slope = 1.00, bias = 0.00 kcal·mol^{-1}.

then expanded this to form a separate set containing 40 catalysts, to examine Zn-aminoalcohol complexes with a larger range of substituent diversity.

2. METHODOLOGY

The working protocol consisted of five stages: (1) generation of molecular structures; (2) generation of quantum molecular fields; (3) filtering of the grid points; (4) computation of molecular descriptors using GRIND methodology; and (5) multivariate analysis.

We first optimized the metal−ligand structures in their dimeric form in order to mimic the tetrahedral environment of the Zn atoms found in the geometries of transition states. These structures provide suitable models because their metal atom coordination sphere closely resembles that of the transition state, as Kozlowski and co-workers[5h] reported. Although the catalytic activity originates in the catalyst's monomeric form, to a major extent it adopts a dimeric structure, a further justification for optimizing dimers. This was done by testing a molecular mechanics force field[10] and using a DFT method;[11,12] both methods gave similar results for the first set of 18 catalysts (outlined above). For subsequent analysis (descriptor generation, and construction and validation of different QSSR equations) we used a DFT method to investigate the monomeric structure. The electronic density distribution was obtained by computing the Kohn−Sham molecular orbitals on the monomeric moiety of the optimized dimmer at the DFT level. From this we then generated the steric and electrostatic 3D descriptors.

The next step was to generate a molecular surface from the DFT-derived electronic density distribution.[13] The molecular surface was defined by an electronic density isosurface at a value of 0.002 au, and it was generated using the values computed in a rectangular grid of points 0.2 Å apart and extending 5 Å beyond molecular boundaries. The shape of the catalyst is thought to be crucial in discriminating between the two enantiomers, so we selected the grid points that compose the molecular surface and defined the Molecular Shape Field (MSF) to take this into account. To obtain data suitable for later analysis, the local curvature values at each of the grid points of the molecular surface were computed using a cosine expression similar to that used by Pastor et al.[7] The MSF values ranged between 0 and −1 for convex areas and 0 and +1 for concave ones. We then recorded the Molecular Electrostatic Potential (MEP) at each of the grid points of the defined molecular surface. These descriptors can be related to chemical meaningful quantities and allow evaluation of the steric and electrostatic interaction fields generated by the molecular structure.

The MSF and MEP fields were filtered to reduce the number of surface points and to develop the subsequent alignment-free strategy. Taking into account MSF values,

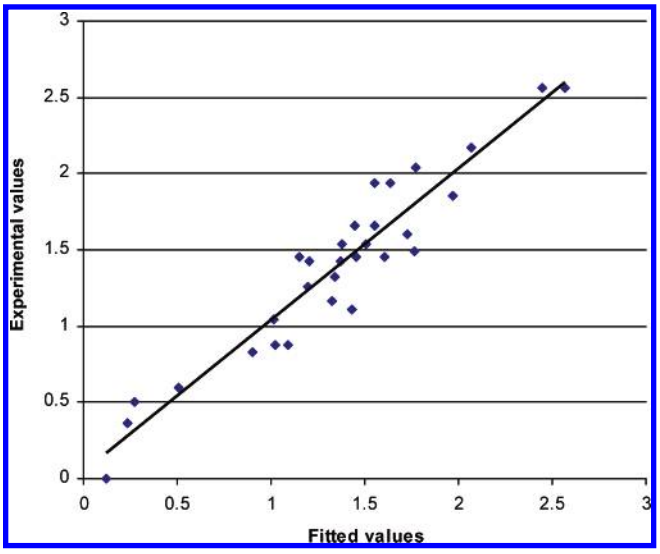


Figure 2. Experimental vs fitted DG values for the fitting set. Values are in kcal·mol^{-1}.

Table 2. Experimental and Predicted ee Values (in %) for the 10 Catalytic Systems Employed for Testing the QSSR Model[a]

catalytic system	experimental	predicted
2	3	49
3	59	68
14	86	81
15	98	97
18	97	95
23	91	87
27	77	87
31	45	21
32	69	74
37	92	87

[a] Catalytic system entries refer to Table S1 in the Supporting Information.

we selected the most convex areas of the isosurface as being the most descriptive of the molecular shape and the most suitable for representing structural diversity. The spatial extent of the selected area depends on the number of filtered points, a value that was optimized (q^2 values in full cross-validations were larger when one node per 15 surface nodes was selected). A low number would preclude the description of important areas of large catalysts, whereas a higher number adds a noise factor. The independence of the grid alignment was achieved from all these 3D data by computing the product of the MSF and the MEP values for each pair of filtered points and selecting the highest values of these products for each distance range. The optimal distance range was found to be 1 Å, and products were computed between 3 and 20 Å, an interval that contains the key interactions without noise components. In this way, only the highest values of these products were kept and used to build the MSF−MSF and MEP−MEP vs distance autocorrelograms,[7a−c] which were then simply concatenated. These grid-independent descriptors constituted the predictive vectors for subsequent multivariate analysis. We also tested the use of MSF-MEP cross-correlograms built from the products of curvature and potential and selection of their highest values, which was useful in the study of biological systems by Pastor et al.[7] However, the addition of this new variable did not improve the results of multivariate analysis in our study.

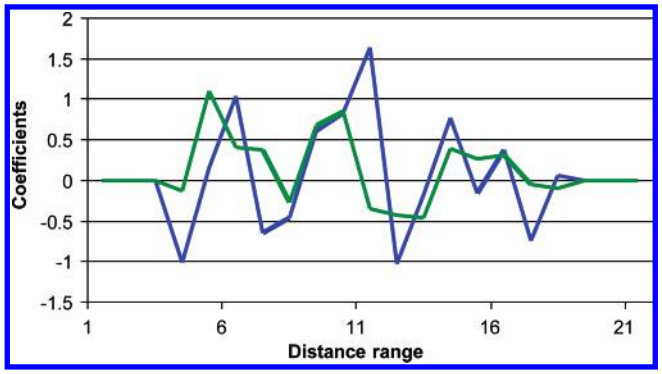


Figure 3. Equation coefficients of the QSSR model built in the fitting stage (30 compounds). MSF−MSF and MEP−MEP coefficients are shown in blue and green, respectively.

Finally, we used Principal Components Analysis (PCA) and Partial Least-Squares Regression (PLSR)[14] as exploratory and multivariate regression techniques, respectively. Different statistical parameters were employed to evaluate the predictive ability of models during the fitting and test stages, namely the correlation coefficient (r^2), the standard error, and the slope and bias of the fitted/predicted vs the observed values. The determination coefficient (q^2) was also computed when full cross-validations were carried out.

3. RESULTS AND DISCUSSION

In the present study, 40 catalytic systems were used to construct and validate the models. Wide structural diversity in the catalysts was encompassed by means of varying the structure and the position of all the R groups (R_1−R_6) in Scheme 1—namely aromatic and aliphatic substituents, small and bulky groups, rigid and flexible structures, and cyclic and acyclic amines. The catalysts and their structures are listed in Table S.1 in the Supporting Information.

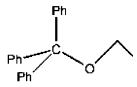
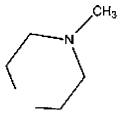
PCA was applied to the 40-catalyst data set in order to analyze the chemical information provided by the combined MSF−MSF and MEP−MEP autocorrelograms. As can be observed in the loading plot in Figure 1A, both kinds of 3D field have the same influence on the first two components, which explains the variance occurring in 74% of the data. Figure 1B shows the projection of the 40 catalysts onto the score plot mapping the two first main components; it demonstrates a clear clustering of the catalysts that was obtained. This clustering is closely related to the chemical nature of the amino alcohol substituents; a visual analysis of the systems allows identification of the clusters according to the size and aromatic characteristics of their substituents. Thus the autocorrelograms generated as outlined above retain information of chemical significance. An autocorrelogram itself does not retain explicit chiral information, but molecular descriptors of this sort are perfectly capable of providing information about the size, shape, and electrostatic characteristics of the catalyst. It is therefore reasonable to investigate their use to correlate molecular structure with activity.[6]

To allow comparison, we first selected a set consisting of the same 18 catalytic systems already considered by other authors.[5c,6] The combination MSF−MSF and MEP−MEP correlograms were built for the selected 18 catalysts and full cross-validation carried out. Comparison of predicted enantiomeric excess (ee) values and observed quantities yielded good statistical parameters (see the Supporting Information, Figure S.1). The optimal number of PLS factors—in this case 4—was arrived at by selecting the number of components that showed the first minimum in the prediction error plotted against the number of PLS factors for full cross-validation. This way we looked at prediction and not at fitting error, thereby avoiding overfitting phenomena. We also evaluated differences arising from the method for optimizing the molecular structures and found that DFT optimizations led to somewhat better results ($q^2 = 0.79$) than using molecular mechanics ($q^2 = 0.65$).

Although the number of samples was relatively low, we divided the data set into a fitting set containing 14 compounds and a test set of 4 compounds, respectively.[5c,6] As Table 1 shows, the statistical parameters obtained in the present study are similar to those of previously published QSSR models. Using molecular descriptors derived from a DFT calculation led to correct predictions for three of the four external compounds, although there was a large discrepancy (39%) in one case. But despite the discrepancies observed, the model is still able to discriminate between low, medium, and high enantioselective catalysts.

Second, we considered an enlarged data set of 40 catalysts (see the Supporting Information, Table S.1) corresponding to different Zn-aminoalcohols catalyzing the transformation of benzaldehyde. Molecular diversity is a subtle question as it has been recently pointed out by Westerhuis et al.[15] Consequently, the diversity of this enlarged data set was analyzed. Structural fingerprint descriptors[16] were computed for the original and the enlarged data sets (18 and 40 catalysts), and similarity indices were calculated using the MolDiA software.[17] Three kinds of indices were used: Tanimoto, Simpson, and Cosine coefficients. A square similarity matrix was obtained for each data set, and diversity plots were built for each pair data set/index (see Figures S.3 and S.4 in the Supporting Information). No significant differences were observed in the structural diversity distribution of both data sets. Molecules belonging to the original 18 catalysts data set are as diverse as those of the expanded 40 catalysts data set. Limitations regarding the selection of catalysts (lacking of available experimental values for ee) and the position/nesting of the R groups (R_1−R_6) have not strongly affected the diversity of the 40 catalysts data set. We then developed a QSSR model for predicting ee values in this larger set of catalysts. Molecular mechanics force field optimization gave a good full cross-validation of catalysts 1−18, so in order to shorten the time required the remaining systems (19−40) were optimized using molecular mechanics. After full cross-validation of all 40 catalysts (see the Supporting Information, Figure S.2), the q^2 value was 0.67 for the optimal number of PLS factors (5 components). The data set was divided into fitting and test subsets of 30 and 10 catalysts, respectively, to develop an externally validated QSSR model. The test set was randomly chosen and covered a wide enantioselectivity range (standard deviations for the fitting and test sets were 22% and 29%, respectively). The fitting stage was carried out using the optimal number of PLS factors obtained in the full cross-validation with the following results: $N = 30$, $r = 0.95$, standard error $= 0.18$ kcal·mol^{-1}, slope $= 0.99$, and bias $= 0.05$ kcal·mol^{-1}. Figure 2 shows the experimentally measured plotted against the fitted ee values. Once the QSSR model was established, the following predictions were made for the test set: $N = 10$, r

Table 3. Influence of the Different Regions of the Catalyst in the Enantiodiscrimination Achieved According to Their Correlogram Values Weighted by the QSSR Equation Coefficients

Catalyst[a]	R_1	R_2	R_3	R_4	R_5	R_6	Ethyl	e.e. (%)
37	H	H	 **13 %**		 **37 %**		**50 %**	92

[a] Label of the catalytic system according to Table S1 included in the Supporting Information.

= 0.92, standard error = 0.32 kcal·mol^{-1}, slope = 1.11, bias = −0.20 kcal·mol^{-1} (see Table 2). The data in Table 2 collected for catalysts in the test set show agreement between the experimental and predicted ee values in cases with medium and high enantioselectivities (>50%). On the other hand, the predicted enantioselectivity for catalysts showing low ee values (<50%), such as systems 2 and 31, is rather poor. The reason for these discrepancies could very well be that the ee values in the available fitting set do not form a uniformly distributed data set. Only 4 of the catalysts exhibit enantioselectivities lower than 50%.

The final QSSR equation contains both positive and negative coefficients which, combined with the values of the molecular descriptor for a particular system, produces the corresponding predicted ee value. The QSSR coefficients for the MSF−MSF and MEP−MEP variables are shown in Figure 3 (in blue and green, respectively) for the QSSR equation obtained in the fitting stage (with 30 compounds). The weights of the QSSR coefficients corresponding to the MSF−MSF and MEP−MEP variables are 61% and 39%, respectively; these values highlight the relative importance of each descriptor in predicting enantioselectivity. The relatively small difference between the weights of the descriptors may seem to contradict the idea that the enantiomeric excess is mainly determined by molecular shape recognition. But we think that as well as electrostatic properties the MEP−MEP values contain a second level of steric information which derives from the selection of the nodes based on MSF values.

The ultimate goal of this analysis is to gain insight into catalyst performance and to derive rules to guide catalyst design. To assess our results we analyzed the influence which each selected node on the isosurface had on the enantiomeric discrimination ability of a particular catalytic system and assigned the nodes to a catalyst substituent. Evaluating the individual influence of each node is not straightforward, given that the QSSR equation variables (correlogram values) involve the product of values of the field (MSF or MEP) in two nodes. This, however, can be approached by assigning to each node the individual relative weight of the molecular field in the correlogram product. In this way the contribution of a single node is computed as the product of three terms: the QSSR equation coefficient, the correlogram value, and the individual percentage of the product at the node. One node may be involved in several interaction products, so all the individual contributions of a given node in the whole correlogram are then added, keeping their sign. The final step was to assign the total contribution to a single structural moiety, such as a catalyst substituent (R1−R6) or a functional group (ethyl), by applying a proximity-based criterion. This allowed us to define a new parameter, the Group Structural Influence (GSI) ratio, which accounts for the total contribution of each substituent or each functional group to overall enantioselectivity. The GSI ratio is defined as that of the aggregated contributions of the nodes assigned to a given substituent, to the total contribution of all the nodes, and is expressed in relative terms. In this way we can estimate the relative global intensity of each functional group's contribution to enantiodifferentiation and consequently extract useful information to rationalize and guide catalyst design. Table 3 shows the results of such an analysis for catalyst 37, one of the extreme cases among the Zn-aminoalcohol catalysts in the test set. The weight of hydrogen atoms is negligible in determining enatioselectivity, whereas the ethyl ligand, the amine substituent, and substituents at position R_3 all play an important role (GSI ratios of 50%, 37%, and 13%, respectively).

At this point we realized that a relationship can be established between the information derived from the QSSR modeling and previous mechanistic knowledge. Previous authors proposed a stereochemical model for the reaction under study, based on the relative energies of four transition states: *anti-pro-R, anti-pro-S, syn-pro-R, and syn-pro-S*.[9] With this model, the anti path is favored over the syn one, the *anti-pro-S* species being the most stable of the anti transition states. In the *anti-pro-R* transition state, previous authors identified a nonfavorable steric interaction between the ethyl group of Zn and the aryl group of benzaldehyde as being the key factor for the destabilization of the *pro-R* path. In addition, the influence of the ethyl group will be greater if the amine substituents are bulky groups that sterically hinder the rotation of the ethyl group. In agreement with this model, the ethyl bonded to the Zn center is also in the region that the QSSR model identified as the region with the largest contribution to enantioselectivity for catalyst 37 (Table 3). Moreover, the next two larger GSI values correspond to those of the substituents of the amine group (R5 and R6). Note also that the structure of catalyst 37 is very similar to that in the system used by Maseras and coworkers[9e] in their detailed mechanistic study on the addition of diethylzinc to benzaldehyde. Figure 4 shows the molecular structure of catalyst 37 and the isosurface points (nodes) selected according to the MSF values. Figure 4 also represents the two nodes involved in the largest interaction

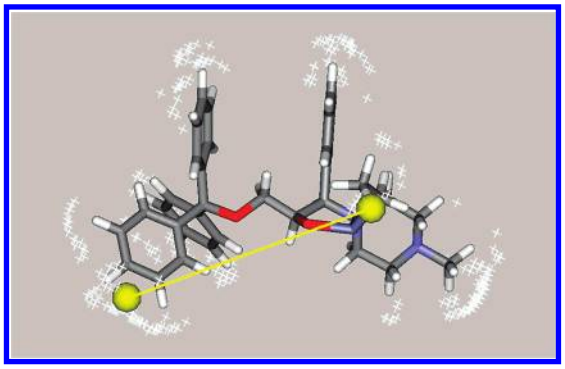


Figure 4. Isosurface points selected according to MSF values for catalyst 37. Nodes involved in the interaction product computed at the distance range with the highest QSSR equation coefficient are highlighted in yellow.

product (in yellow) computed for the distance range with the highest QSSR coefficient. One of the two nodes was located precisely in the vicinity of the ethyl group, confirming the reliability of both the QSSR model and the analysis outlined above. Other correlogram variables with high QSSR equation coefficients consist of vectors joining nodes located in the vicinity of amine substituents and of the ethyl group.

4. CONCLUSIONS

We have designed and tested a new CoMFA-like QSSR protocol based on the use of 3D molecular descriptors. This new method predicts the enantiomeric excess in the ethylation of aldehydes using Zn-aminoalcohols. We also investigated a grid alignment-free approach.

This new method is slower than the CoMFA approaches that use empirical and semiemprical descriptors, but it can represent any element in the periodic table. We put forward a molecular descriptor based on the electronic charge density and the molecular electrostatic potential, both easily obtained from standard DFT codes. Given current computing resources, the method is not computationally demanding: once built, it takes just a few minutes to compute the data for medium-sized molecules and predict their catalytic performance.

The method shows a predictive ability similar to that observed for a set of 18 catalysts with benzaldehyde as substrate in previous empirical QSSR approaches although a significant deviation was observed in one of the four test compounds employed by previous researchers. The predictive ability was maintained for catalytic systems showing ee values higher than 50% when more complex catalyst structures were used to fit and test the new model, but some deviations were observed in systems with low enantiodiscrimination power.

Finally, we developed analysis procedures for determining the influence of the substituents and the ethyl group on enantioselectivity and obtained agreement with TS modeling. Application of this new methodology to other catalytic transformations is currently under way in our laboratory.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Education of Spain (project CTQ2005-06909-C02-01/BQU and CTQ2005-06909-C02-02/BQU), AGAUR of the Generalitat de Catalunya (2005SGR00715 and 2005SGR00104), and the ICIQ Foundation and Consolider Ingenio 2010 (Grant CSD2006-0003). M.U. thanks the Torres Quevedo program of the Ministerio de Educación y Ciencia, Spain.

Supporting Information Available: Detailed list of the 40 catalytic systems employed in this study, statistical parameters for the full cross-validation of the models for catalysts 1−18 and 1−40, and diversity plots for the two data sets used. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) (a) *Comprehensive Asymmetric Catalysis*; Jacobsen, E. N., Pfaltz, A., Yamamoto, H., Eds.; Springer-Verlag: New York, 1999. (b) van Leeuwen, P. W. N. M. *Homogeneous Catalysis*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2004.

(2) (a) *Computational Modeling of Homogeneous Catalysis*; Maseras, F., Lledós, A., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002. (b) Burello, E.; Rothenberg, G. In Silico Design in Homogeneous Catalysis Using Descriptor Modelling. *Int. J. Mol. Sci.* **2006**, *7*, 375−404. (c) Burello, E.; Marion, P.; Galland, J.-C.; Chamard, A.; Rothenberg, G. Ligand Descriptor Analysis in Nickel-Catalysed Hydrocyanation: A Combined Experimental and Theoretical Study. *Adv. Synth. Catal.* **2005**, *347*, 803−810.

(3) (a) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147−158. (b) Golbraikh, A.; Tropsha, A. QSAR Modeling Using Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 144−154. (c) Jiang, C.; Li, Y.; Tian, Q.; You, T. QSAR Study of Asymmetric Reactions with Topological Indices. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1876−1881. (d) Hoogenraad, M.; Klaus, G. M.; Elders, N.; Hooijschuur, S. M.; McKay, B.; Smith, A. A.; Damen, E. W. P. Oxazaborolidine Mediated Asymmetric Ketone Reduction: Prediction of Enantiomeric Excess Based on Catalyst Structure. *Tetrahedron: Asymmetry* **2004**, *15*, 519−523.

(4) Cramer, R. D., III; Paterson, 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.

(5) (a) Lipkowitz, K. B.; Kozlowski, M. C. Understanding Stereoiduction in Catalysis via Computer: New Tools for Asymmetric Synthesis. *Synlett* **2003**, *125*, 1547−1565. (b) Lipkowitz, K. B.; Pradham, M. Computational Studies of Chiral Catalysts: A Comparative Molecular Field Analysis of an Asymmetric Diels-Alder Reaction with Catalysts Containing Bisoxazoline or Phosphinooxazoline Ligands. *J. Org. Chem.* **2003**, *68*, 4648−4656. (c) Kozlowski, M. C.; Dixon, S. L.; Panda, M.; Lauri, G. Quantum Mechanical Models Correlating Structure with Selectivity: Predicting the Enantioselectivity of β-Amino Alcohols Catalysts in Aldehyde Alkylation. *J. Am. Chem. Soc.* **2003**, *125*, 6614−6615. (d) Melville, J. L.; Andrews, B. I.; Lygo, B.; Hirst, J. D. Computational Screening of Combinatorial Catalyst Libraries. *Chem. Commun.* **2004**, 1410−1411. (e) Dixon, S.; Merz, K. M., Jr.; Lauri, G.; Ianni, J. C. QMQSAR: Utilization of a Semiempirical Probe Potential in a Field-based QSAR Method. *J. Comput. Chem.* **2004**, *26*, 23−34. (f) Cruz, V. L.; Ramos, J.; Martinez, S.; Muñoz-Escalona, A.; Martinez-Salazar, J. Structure-Activity Relationship Study of the Metallocene Catalyst Activity in Ethylene Polymerization. *Organometallics* **2005**, *24*, 5095−5102. (g) Huang, J.; Ianni, J. C.; Antoline, J. E.; Hsung, R. P.; Kozlowski, M. C. De Novo Chiral Amino Alcohols in Catalyzing Asymmetric Additions to Aryl Aldehydes. *Org. Lett.* **2006**, *8*, 1565−1568. (h) Ianni, J. C.; Annamalai, V. P.; Phuan, W.; Panda, M.; Kozlowski, M. C. A Priori Theoretical Prediction of Selectivity in Asymmetric Catalysis: Design of Chiral Catalysts by Using Quantum Molecular Interaction Fields. *Angew. Chem., Int. Ed.* **2006**, *45*, 5502−5505.

(6) Sciabola, S.; Alex, A.; Higginson, P. D.; Mitchell, J. C.; Snowden, M. J.; Morao, I. Theoretical Prediction in Asymmetric Catalysis. An Alignment-Independent Molecular Interaction Field Based Approach. *J. Org. Chem.* **2005**, *70*, 9025−9027.

(7) (a) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243. (b) Fontaine, F.; Pastor, M.; Sanz, F. Incorporating Molecular Shape into the Alignment-free GRid-INdependent Descriptors. *J. Med. Chem.* **2004**, *47*, 2805−2815. (c) Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F. Anchor-GRIND: Filling the Gap between Standard 3D QSAR and the GRid-INdependent Descriptors. *J. Med. Chem.* **2005**, *48*, 2687−2694.

(8) (a) Pu, L.; Yu, H. B. Catalytic Asymmetric Organozinc Additions to Carbonyl Compounds. *Chem. Rev.* **2001**, *101*, 757−824. (b) Soai, K.;

Niwa, S. Enantioselective Addition of Organozinc Reagents to Aldehydes. *Chem. Rev.* **1992**, *92*, 833−856.

(9) (a) Oguni, N.; Omi, T. Enantioselective Addition of Diethylzinc to Benzaldehyde Catalyzed by a Small Amount of Chiral 2-Amino-1-Alcohols. *Tetrahedron Lett.* **1984**, *25*, 2823−2824. (b) Kitamura, M.; Suga, S.; Kawai, K.; Noyori, R. Catalytic Asymmetric Induction. Highly Enantioselective Addition of Dialkylzincs to Aldehydes. *J. Am. Chem. Soc.* **1986**, *108*, 6071−6072. (c) Yamakawa, M.; Noyori, R. An Ab Initio Molecular Orbital Study on the Amino Alcohol-Promoted Reaction of Dialkylzincs and Aldehydes. *J. Am. Chem. Soc.* **1995**, *117*, 6327−6335. (d) Yamakawa, M.; Noyori, R. Asymmetric Addition of Dimethylzinc to Benzaldehyde Catalyzed by (2*S*)-3-*exo*-(Dimethylamino)isobornenol. A Theoretical Study on the Origin of Enantioselection. *Organometallics* **1999**, *18*, 128−133. (e) Vázquez, J.; Pericás, M. A.; Maseras, F.; Lledós, A. A Quantum Mechanics/Molecular Mechanics Study of the Highly Enantioselective Addition of Diethylzinc to Benzaldehyde Promoted by (*R*)-2-Piperidino-1,1,2-triphenylethanol. *J. Org. Chem.* **2000**, *65*, 7303−7309. (f) Rasmussen, T.; Norrby, P. O. Modeling the Stereoselectivity of the α-Amino Alcohol-Promoted Addition of Dialkylzinc to Aldehydes. *J. Am. Chem. Soc.* **2003**, *125*, 5130−5138. (g) Rudolph, J.; Bolm, C.; Norrby, P. O. New Insights into the Stereoselectivity of the Aryl Zinc Addition to Aldehydes. *J. Am. Chem. Soc.* **2005**, *127*, 1548−1552.

(10) (a) UFF as implemented in ArgusLab 4.0. Thompson, M. A. Planaria Software LLC: Seattle, WA. http://www.planaria-software.com (accessed Jul 24, 2007). (b) Casewit, C. J.; Colwell, K. S.; Rappé, A. K. Application of a Universal Force Field to Organic Molecules. *J. Am. Chem. Soc.* **1992**, *114*, 10035−10046. (c) Casewit, C. J.; Colwell, K. S.; Rappé, A. K. Application of a Universal Force Field to Main Group Compounds. *J. Am. Chem. Soc.* **1992**, *114*, 10046−10053. (d) Rappé, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(11) The calculations were carried out using the ADF code, version 2005.01. (a) te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Fonseca-Guerra, C.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931−967. (b) Fonseca-Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. Towards an Order-N DFT Method. *Theor. Chem. Acc.* **1998**, *99*, 391−403. In all calculations the BP86 functional was employed. (c) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100. (d) Perdew, J. P. Erratum: Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *34*, 7406. (e) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33*, 8822−8827. (f) A TZP Slater type basis set was used for the geometry optimizations as well as single point calculation.

(12) The initial set of 18 catalyst structures was fully optimized at a DFT level (ref 11). The same set optimized at the UFF level did not cause significant differences in the QSSR model, so the larger set of 40 catalytic systems was optimized at the least time-consuming UFF level. In some cases, the lowest energy conformer was located using molecular dynamics sampling.

(13) The electronic charge density and molecular electrostatic potential distributions were evaluated using the DENSF code included in the ADF package. A new code we developed was used to read these data and to generate the isosurface, compute the local curvature, filter the grid points, and generate the autocorrelograms. Although this code is adapted to ADF data structures it would be very easy to modify it to read files generated by other computational packages (e.g. Gaussian cube files). This code, still in development, is available upon request.

(14) (a) Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130. (b) Geladi, P.; Kowlaski, B. Partial Least Square Regression: A Tutorial. *Anal. Chim. Acta* **1985**, *35*, 1−17.

(15) Westerhuis, J. A.; Hageman, J. A.; Frühauf, H.-W.; Rothenberg, G. Understanding Ligand Diversity. *Chim. Oggi - Chem. Today* **2007**, *25*, 28.

(16) Maldonado, A.; Petitjean, M.; Doucet, J.-P.; Panaye, A.; Fan, B. T. MolDIA: XML based system of molecular diversity analysis towards virtual screening and QSPR. *SAR QSAR Environ. Res.* **2006**, *17*, 11−23.

(17) Maldonado, A.; Petitjean, M.; Doucet, J.-P.; Fan, B. T. Molecular Similarity and Diversity: Concepts and Applications. *Mol. Diversity* **2006**, *10*, 39−79.

CI700181V