JOC*Note*

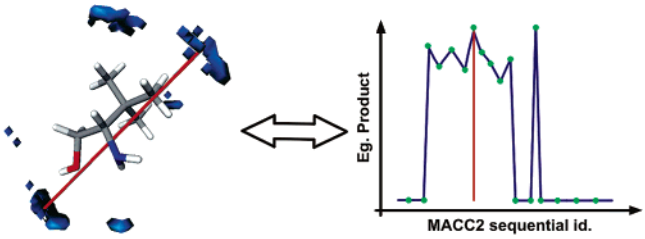
Theoretical Prediction of the Enantiomeric Excess in Asymmetric Catalysis. An Alignment-Independent Molecular Interaction Field Based Approach

Simone Sciabola,[†,‡] Alexander Alex,[†] Paul D. Higginson,[*,†] John C. Mitchell,[§] Martin J. Snowden,[§] and Iñaki Morao[*,†,§]

Pfizer Ltd., Global Research and Development, Sandwich Laboratories, Sandwich, Kent CT13 9NJ, U.K., Laboratorio di Chemiometria, Universita di Perugia, Via Elce di Sotto, 10, 1-06123 Perugia, Italy, and University of Greenwich, Medway Sciences, Chatham Maritime ME4 4TB, U.K.

inaki.morao@pfizer.com

Received July 19, 2005



The enantiomeric excess of three different asymmetric catalyses has been predicted in excellent agreement with the experiments using a 3D-QSPR approach. In particular, GRid INdependent Descriptors generated from molecular interaction fields together with a simple partial least-squares method were found to be adequate to describe the enantioselectivity induced by these metal−ligand complexes.

The application of asymmetric catalysis to the large-scale synthesis of medicines is a desired strategy within the pharmaceutical chemistry.[1] Despite its relevance, the economic constraints and the lack of process robustness are still key issues to solve. To minimize these difficulties, high-throughput experimentation (HTE) becomes a practical industrial approach for catalyst optimization.[2] On the other hand, computational chemistry has shown to be a very useful and complementary tool to rationalize the experimental outcome in the organometallics field.[3] However, their complex mechanisms together with the need for high levels of theory have hindered a marriage between theory and experiment. Moreover, this marital situation still has not changed in practical terms even with the present-day computing speed and with much faster QM/MM procedures.[4] Alternatively, quantitative structure property relationship (QSPR) methodology is a chemoinformatic technique successfully applied to a wide range of chemical and biological problems.[5] The generation of a mathematical model, based on previously tested molecules, capable of predicting the behavior of new compounds can optimize synthetic efforts. To study chirality as a property, 3D-descriptors become the obvious choice.[6] Of these, molecular interaction field (MIF) based descriptors have shown to work particularly well in drug design.[7]

To the best of our knowledge, only two papers have been recently published to predict the enantiomeric excess of chiral catalysts using MIF-based 3D-QSPR methods.[8−12] Lipkowitz et al. developed a CoMFA analysis of an asymmetric Diels−Alder reaction.[8] Although the results are excellent, the disadvantage of this approach is the superimposition step for all 3D structures, which is always subjective and difficult to manage for very diverse compounds.[9] In the other study, by Kozlowski et al., the stereochemical induction of the Et_2Zn addition to benzaldehyde was analyzed and predicted.[10] This in-house grid based method again implies the alignment scheme and also the optimization of the corresponding transition structures. Here we present a general and fast alignment-free QSPR methodology to predict the enantiomeric excess in asymmetric catalysis. This procedure has been tested in three different reactions (Scheme 1)[8,10,11] with comparable or better results to the above-mentioned approaches.

[†] Pfizer, Ltd.
[‡] Universita di Perugia.
[§] Medway Sciences.

(1) (a) Hawkins, J. M.; Watson, T. J. N. *Angew. Chem., Int. Ed.* **2004**, *43*, 3224−3228. (b) *Asymmetric catalysis on industrial scale: challenges, approaches and solutions;* Blaser, H. U., Schmidt, E., Eds.; Wiley-VCH: Weinheim, 2004.

(2) (a) Boussie, T. R.; Diamond, G. M.; Goh, C.; Hall, K. A.; LaPointe, A. M.; Leclerc, M.; Lund, C.; Murphy, V.; Shoemaker, J. A. W.; Tracht, U.; Turner, H.; Zhang, J.; Uno, T.; Rosen, R. K.; Stevens, J. C. *J. Am. Chem. Soc.* **2003**, *125,* 4306−4317. (b) Hoogenraad, M.; van der Linden, J. B.; Smith, A. A.; Hughes, B.; Derrick, A. M.; Harris, L. J.; Higginson, P. D.; Pettman, Alan J. *Org. Process. Res. Dev.* **2004**, *8*, 469−476.

(3) Niu, S.; Hall, M. B. *Chem Rev.* **2000**, *100*, 353−405. For a review about new computational tools for asymmetric synthesis, see: Lipkowitz, K. B.; Kozlowski, M. C. *Synlett* **2003**, 1547−1565.

(4) (a) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179. (b) Mikami, K.; Aikawa, K.; Yamanaka, M. *Pure Appl. Chem.* **2004**, *76*, 537−540.

(5) *QSPR / QSAR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Nova Science Publishers: New York, 2001.

(6) *Handbook of Molecular Descriptors*; Todeschini, R., Consonni, V., Eds.; Wiley-VCH: Weinheim, 2000.

(7) (a) GRID method: Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849−857. (b) CoMFA method: Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1998**, *110*, 5959−5967. (c) CoMSIA method: Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130−4146. (d) Spatial autocorrelation method: Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775. (e) Volsurf method: Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. *J. Med. Chem.* **2000**, *43*, 2204−2216. (f) GRIND method: Fontaine, F.; Pastor, M.; Sanz, F. *J. Med. Chem.* **2004**, *47*, 2805−2815

(8) Lipkowitz, K. B.; Pradhan, M. *J. Org. Chem.* **2003**, *68*, 4648−4656.

(9) ComFA method can be carried out using *Sybyl 6.9.* software from Tripos, St. Louis, MO.

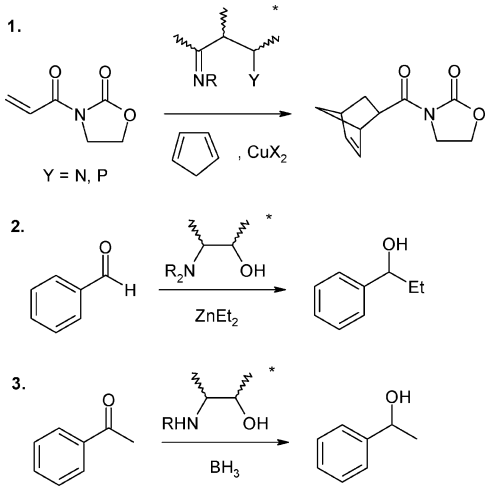
(10) Kozlowski, M. C.; Dixon, S. L.; Panda, M.; Lauri, G. *J. Am. Chem. Soc.* **2003**, *125*, 6614−6615.

(11) A combined theoretical and experimental study for the prediction of enantiomeric excess in asymmetric ketone reduction. Unfortunately, the identity of the descriptors used was not disclosed: Hoogenraad, M.; Klaus, G. M.; Elders, N.; Hooijschuur, S. M.; McKay, B.; Smith, A. A.; Damen, E. W. P. *Tetrahedron: Asymmetry* **2004**, *15*, 519−523.

(12) During the preparation of this manuscript other CoMFA analysis has been published. The predicted values of this communication have not been compared with our methodology. See: Melville, J. L.; Andrews, B. I.; Lygo, B.; Hirst, J. D. *Chem. Commun.* **2004**, 1410−1411.

10.1021/jo051496b CCC: $30.25
Published on Web 09/27/2005

SCHEME 1. Reactions under Examination



Our working protocol is graphically represented in Figure 1. The first step consisted of the geometry optimization of the ligand−metal complexes (L^0-$CuCl_2$, L^{-1}−ZnEt, and L^{-2}−BH). The conformation of the ligand was found to be crucial to build the diagnostic model. As expected, predictions made from metal-bound ligand geometries were much better than those from free conformations. The semiempirical level of theory PM3-tm was the best compromise between accuracy and computational speed.[13,14] The optimized structures of the coordinated ligands (L) were imported into ALMOND software.[15] The inclusion of the whole complex hindered the effect of the ligands showing slightly worse results. The GRID method was used to calculate the interaction fields between the molecule and probes (MIFs), which evaluate electrostatic, van der Vaals, and hydrogen-bond interactions.[16]

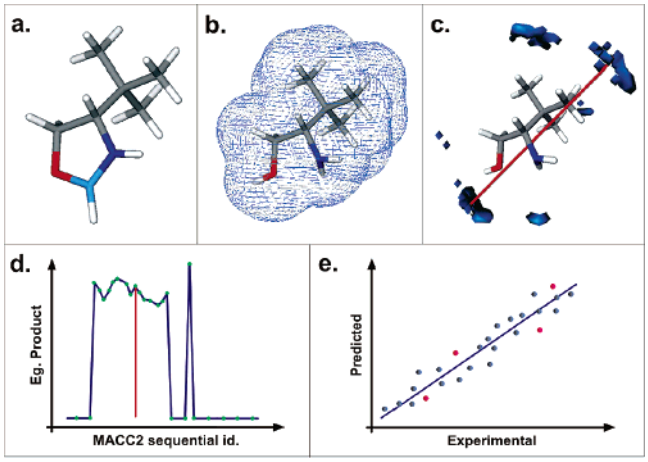


FIGURE 1. Flow chart of the working procedure carried out: (a) geometry optimization; (b) MIF calculation; (c) filtered points derivation; (d) energy vs distance plot; (e) PLS model.

Then, the calculated MIFs were filtered and only the regions in which the intensity of the field is maximized at relative distances were considered. The alignment independence was achieved computing the product of the interaction energy for each pair of filtered points (nodes). Only the highest values of these products are encoded using an autocorrelation transform and reported as a function of the distance separating the nodes. These computed GRid INdependent Descriptors (GRIND) are usually plotted in diagrams called correlograms. Opposite to other alignment-independent molecular methods, the original descriptors (molecular interaction fields) can be regenerated from the autocorrelation transform and the results of the analysis can be represented graphically in 3D plots together with the original molecular structures.[17] We used these descriptors as the X-matrix for the partial least squares (PLS) diagnostic model.[18] PLS regression analysis is particularly useful because it can handle a system where the number of highly correlated descriptors is much higher than the number of compounds.

TABLE 1. Information Summary of the QSPR Models Applied to Reactions 1−3

rxn[a]	ref[b]	no.[c]	LV[d]	$r^{2\,e}$	$q^{2\,f}$	ee of the prediction set[g]
1	8	18	4	0.95	0.52	(46,53); (54,67); (75,83); (30,39)
2	10	14	3	0.99	0.69	(3,18); (86,83); (98,92); (63,66)
3	11	24	3	0.94	0.80	(73,73); (96,97); (59,71); (2,0)

[a] See Scheme 1. [b] Reference number for comparison. [c] Number of catalysts used in the training set to build the model. [d] Number of latent variables projected. [e] Cross-correlation coefficient (r^2). [f] Cross-validated correlation coefficient (q^2). [g] (Experimental, predicted) ee values set of the prediction set.

The most significant results obtained for the three studied reactions are summarized in Table 1. The use of GRIND descriptors was appropriate to fit the training set in all three reactions (cross correlation, r^2, is over 0.9). The predictive ability (q^2) was calculated exclusively from the training set data using the leave-one-out (LOO) cross-validation method. The optimum number of latent variables in the PLS analysis was found to be 3−4 in order to get the maximum prediction power. This value ranges from 0.5 to 0.8, which shows the robustness of the internally cross-validated mathematical models generated.[19] It is also noteworthy that the better dispersed the experimental data points are along the activity range the higher is the prediction power value obtained. Nevertheless, the validity of any model is ultimately confirmed after external validation. We selected the same four compounds that were chosen in each of the previous studies to serve as the external prediction sets for the three reactions studied.[20] Obviously, these compounds

(13) Spartan′02 from Wavefunction, Inc.: Irvine, CA.

(14) PM3tm gave us reasonable geometries of the metal complexes structures and it has been used successfully in combined QM(DFT)/QM(PM3tm) methods; see: Morao, I.; McNamara, J. P.; Hillier, I. H. *J. Am. Chem. Soc.* **2003**, *125*, 628−629.

(15) ALMOND 3.3.0 from Molecular Discovery, London, UK.

(16) In the present work, we have used standard probes: DRY (hydrophobic), N1 (hydrogen-bond acceptor), O (hydrogen-bond donor), and TIP (shape related and therefore steric).

(17) For the reversibility properties of GRIND descriptors see ref 7f.

(18) ALMOND software includes the PLS module and therefore there is no need of other statistical programs. For a practical introduction in Chemometrics: *Multi- and Megavariate Data Analysis: Principles and Applications;* Eriksson, L., Johansson, E., Wold, N. K., Wold, S., Eds.; Umetrics AB: Umea, 2001.

(19) Our models satisfy the three criteria for an appropriate validation, see: Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Model.* **2002**, *20*, 269−276.

(20) For structural details of the ligands, see Supporting Information.

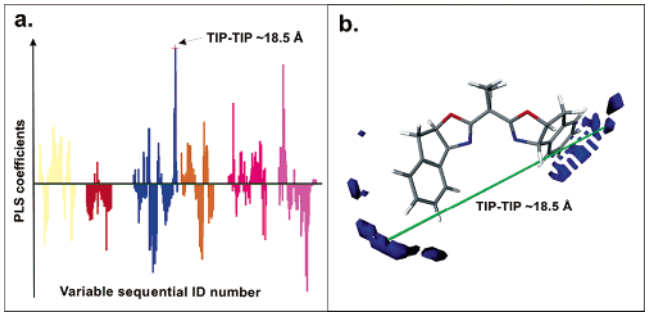


FIGURE 2. (a) PLS coefficients plot for the reaction **1** using GRIND descriptors. (b) The highest positive coefficient of this efficient catalyst (ee 96%) corresponds to a TIP−TIP nodal interaction at 18.5 Å.

were not used to build the model. In all three cases, the predicted values for the test set were very close to the experimental ones (the last column of Table 1), being the worst discrepancy between theory-experiment (residual) only 15 units.

Additional information about the stereochemical induction can be retrieved from the PLS coefficient profile. This plot (Figure 2a) allows us to identify those X variables (GRIND descriptors) that are directly (+) or inversely (−) correlated to the dependent variable *Y* (ee) in a quantitative manner. GRIND variables are grouped into blocks representing the interactions between couples of nodes generated by the same probe or by combinations of the probes (auto- and cross-correlograms, respectively). We here only analyze the copper dataset (reaction 1). In this particular case, the highest positive PLS coefficient corresponds to a TIP−TIP nodal interaction at 18.5 Å. This shape-related descriptor is consistently present in those ligands that show high enantioselectivity (Figure 2b) but completely absent in the weakly active catalysts. Given that interactions from the molecular shape field encode the geometrical relationships between the spatial extents of the molecule a high contribution of the steric effects to the total stereo-induction is totally expected. In addition, structural variations around these regions (the extremes of this nodal interaction) should significantly modify the enantioselective outcome. These results are in concordance with the conclusions drawn from the previously reported CoMFA coefficient contour maps. [8,21]

A practical application of this methodology is for instance the virtual screening of new chiral ligands. The general objective is to replace the efficient but expensive/hazardous/patented ligand by a much cheaper/safer/royalty-free alternative with the same enantioselective performance. In conclusion, we propose this fast and straightforward alignment-independent QSPR methodology for the prediction of the enantiomeric excess in asymmetric catalysis.

Acknowledgment. Pfizer, Inc. is acknowledged for continuing support to this research. This paper is dedicated to Profs. Manuel Pastor and Sergio Clementi for their contributions to the development of GRIND methodology.

Supporting Information Available: ISIS pictures and numerical data of the training and prediction sets for the three reactions studied. This material is available free of charge via the Internet at http://pubs.acs.org.

JO051496B

(21) Steric and electronic STDEV*COEFF contour plots derived from a CoMFA analysis define regions in the space where steric bulk and electron density respectively will influence the stereoinduction. For applications of this methodology, see, for example, refs 7b and 8.