JOC Article

Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions. 1. Application to Sharpless Asymmetric Dihydroxylation

Nicolas Moitessier,[†] Christophe Henry,[†] Christophe Len,[‡] and Yves Chapleur*[,†]

Groupe SUCRES, Unité Mixte 7565 CNRS, Université Henri Poincaré-Nancy 1, B.P. 239, F-54506 Nancy-Vandoeuvre, France, and Laboratoire des Glucides, Université de Picardie-Jules Verne, F-80039 Amiens, France

yves.chapleur@persmail.uhp-nancy.fr

Received April 12, 2002

A reliable and computationally tractable protocol directed at the study of the stereochemical outcome of asymmetric reactions and its application to the Sharpless asymmetric dihydroxylation reaction are proposed. This method, based on a genetic algorithm and molecular mechanics, effectively provides qualitative as well as semiquantitative results and explains the origin of the observed enantioselectivity. For instance, the method reliably predicts reversal of selectivity between similar substrates, unexpected isomers, and even enantio- and diastereoisomeric excess with good accuracy. Two binding modes, closely related to those proposed by Sharpless and Corey, are favored. After comparison with Sharpless' mnemonic device, we propose two alternative interpretations.

Introduction

For the last two decades, catalytic asymmetric synthesis has evolved into one of the main focuses in organic chemistry. Beside chemistry itself, one of the major concerns of asymmetric synthesis chemists has been the rationalization of the stereochemical outcome of reactions, which often result from complex formations.[1] Rapidly, modeling with the help of X-ray crystallography has been recognized as a useful tool for this task.[2] However, most of the studies relied on ab initio and semiempirical calculations, which are not suitable for large or flexible systems and are often applied to truncated model systems.

Meanwhile, the ability to dock ligands in protein active sites and to predict binding affinities has become a major concern in computer-aided drug design.[3] Indeed, recent advances in drug design led to reliable and computationally tractable protocols able to position a ligand in the active site of a protein and to evaluate its binding affinity.[3,4] The methods, which must be fast and accurate for further application in virtual screening, allow flexibility of the ligand using, for instance, a genetic algorithm[5] or an incremental construction.[6] Since many catalyst−substrate complexes can be approximated to ligand−receptor complexes, a fast and accurate docking method would be well suited for nicely predicting the binding mode of the reactant into the catalyst "binding site". Involved in both kind of projects, we thought to develop such a tool that would be fast, accurate, and easy-to-use for organic chemists and therefore based on molecular mechanics. We were then facing the intrinsic limitations of molecular mechanics and met the need for a fine energy difference evaluation.

During the last twenty years, Sharpless' asymmetric dihydroxylation (AD) reaction has emerged as one of the most widely used asymmetric reactions in organic synthesis.[7] As a part of d4T and Adenophostin A analogue synthesis programs, we disclosed the AD of styrene derivatives **1** and **2**[8] and allyl xylopyranosides **3** and **4**[9] (Figure 1).

Introduction of a D-xylose moiety onto styrene (**1**) did not affect the stereoselectivity of the reaction since the corresponding dioxane **2** also reacted with very high

* To whom correspondence should be addressed. Phone: +33-383-68-4773. Fax: +33-383-68-4780.

[†] Université Henri Poincaré-Nancy 1.

[‡] Université de Picardie-Jules Verne.

(1) (a) Jacobsen, E. N., Pfaltz, A., Yamamoto, H., Eds. *Comprehensive Asymmetric Catalysis*; Springer: New York, 1999; Vols. 1−3. (b) Noyori, R. *Asymmetric Catalysis in Organic Synthesis*; Wiley & Sons: New York, 1999.

(2) For a recent example, see: Cui, M.; Adam, W.; Shen, J. H.; Luo, X. M.; Tan, X. J.; Chen, K. X.; Ji, R. Y.; Jiang, H. L. *J. Org. Chem.* **2002**, *67*, 1427−1435.

(3) (a) Klebe, G. *J. Mol. Med.* **2000**, *78*, 269−281. (b) Kubinyi, H. *Curr. Opin. Drug. Discov. Dev.* **1998**, *1*, 4−15.

(4) For example, see: (a) Hanessian, S.; Moitessier, N.; Therrien, E. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 873−881, (b) Hanessian, S.; Moitessier, N.; Cantin, L.-D. *Tetrahedron* **2001**, *57*, 6885−6900. (c) Hanessian, S.; MacKay, D. B.; Moitessier, N. *J. Med. Chem.* **2001**, *44*, 3074−3082.

(5) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(6) Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175 and references therein.

(7) Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B. *Chem. Rev.* **1994**, *94*, 2483−2547.

(8) (a) Ewing, D. F.; Fahmi, N.-E.; Len, C.; Mackenzie, G.; Pranzo, A. *J. Chem. Soc., Perkin Trans. 1* **2000**, 3561−3565. (b) Ewing, D. F.; Len, C.; Mackenzie, G.; Ronco, G.; Villa, P. *Tetrahedron: Asymmetry* **2000**, *11*, 4995−5002. (c) Pilard, S.; Riboul, D.; Glaçon, V.; Moitessier, N.; Chapleur, Y.; Postel, D.; Len, C. *Tetrahedron: Asymmetry* **2002**, *13*, 529−537.

(9) (a) Moitessier, N.; Chrétien, F.; Chapleur, Y. *Tetrahedron: Asymmetry* **1997**, *8*, 2889−2892. (b) Roussel, F.; Moitessier, N.; Hilly, M.; Chrétien, F.; Mauger, J.-P.; Chapleur, Y. *Bioorg. Med. Chem.* **2002**, *10*, 759−768.

10.1021/jo0258148 CCC: $22.00
Published on Web 09/17/2002

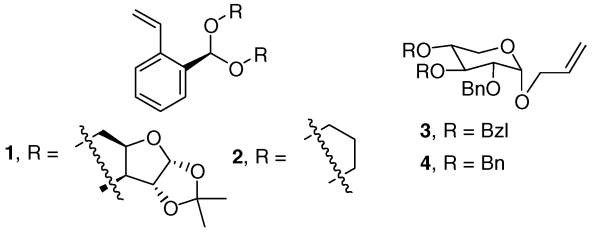


FIGURE 1. Carbohydrate-based substrates for AD.

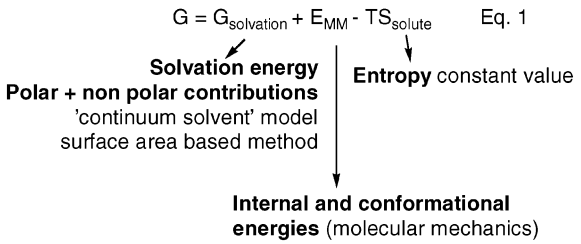


FIGURE 2. Equation describing the free energy.

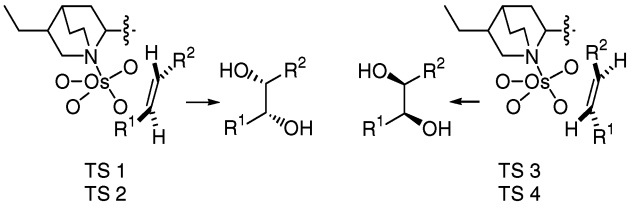


FIGURE 3. Four transition states.

diastereoselectivity. As a result of a strong mismatch effect, the dihydroxylation of **3** and **4** yielded the (*R*)-isomer, while Sharpless' mnemonic device predicted the (*S*)-epimer.[10] This unusual behavior was investigated by molecular modeling, and a match pair effect mostly due to the large size of the olefin was found. In this earlier work, a pure molecular mechanics approach supported by experimental data was used.[10a] We then reasoned that further improvements, including a more efficient conformational search, would provide a tool to rationalize the stereochemical outcome of catalytic reactions that would be of great interest for the organic chemistry community. Herein, we disclose the first semipredictive protocol that exploits a genetic algorithm and applies to rigid and highly flexible substrates.

Results and Discussion

Protocol. Reports of work on AD of rigid olefins (e.g., styrene) using either molecular mechanics (transition state force fields)[11a−c] or more elaborate approaches (QM/MM) are scarce,[11d] and a universal mechanistic picture has yet to be drawn that includes AD of flexible olefins.[12] Although we did not expect to provide an MM protocol with the QM/MM accuracy, we aimed at developing a fast and reliable method that would identify attractive and steric interactions responsible for the facial discrimination.[13] Such an easy-to-use tool would be appropriate for rationalization of the stereochemical outcome of any organic reactions. The major obstacle in using molecular mechanics approaches arises from the difficulty of measuring small differences in the free energy controlling the enantiodiscrimination.[14] Due to the binding-pocket/substrate-like behavior of the olefin/catalyst systems,[15] we assumed that the complex formation was reversible and that the facial selectivity was controlled at the binding stage. With these hypotheses, the Curtin Hammett principle is operative. In other words, the *ratio* of products formed should be determined by the difference between the free activation energies (or the transition state free energies $\Delta\Delta G$). This double-difference approach would result in cancellation of the approximations and the inherent force field errors.

Most of the reported scoring functions, either force-field-based or knowledge-based, encountered in docking programs, rely on the additivity of the different contributions to the free energy change upon binding. On this basis, we chose to use the scoring function depicted in Figure 2.[16] Although the system conformation was described as mainly governed by hydrophobic effects, the solvation free energy was evaluated.[17] Water was used as a model solvent (reaction medium: *t*-BuOH/H_2O). Since the reaction mixture was highly saline, the ionic strength effect was computed as well. Finally, with the intent of estimating the free energy difference between diastereoisomeric transition states, we thought that considering the entropy as a constant would not lead to such a large approximation.

We assumed that the reaction proceeded through a [3 + 2] cycloaddition, since most of the reported investigations support this mechanism.[18] Four models representing the four geometries of attack were built from previously reported transition state geometries[18e] and frozen during the following computations (Figure 3). Fixing the complex was shown to work reasonably well by Houk and

(10) (a) Moitessier, N.; Maigret, B.; Chrétien, F.; Chapleur, Y. *Eur. J. Org. Chem.* **2000**, 995−1005. (b) Mnemonic device predicts the outcome of AD on similar substrates; see, for example: Dominique, R.; Roy, R. *Tetrahedron Lett.* **2002**, *43*, 395−398.

(11) (a) Wu, Y.-D.; Wang, Y.; Houk, K. N. *J. Org. Chem.* **1992**, *57*, 1362−1369. (b) Norrby, P.-O.; Kolb, H. C.; Sharpless, K. B. *J. Am. Chem. Soc.* **1994**, *116*, 8470−8478. (c) Norrby, P.-O.; Rasmussen, T.; Haller, J.; Strassner, T.; Houk, K. N. *J. Am. Chem. Soc.* **1999**, *121*, 10186−10192. (d) Ujaque, G.; Maseras, F.; Lledos, A. *J. Am. Chem. Soc.* **1999**, *121*, 1317−1323.

(12) During the course of our work, synthetic investigations on substrate binding were reported. Bayer, A.; Svendsen, J. S. *Eur. J. Org. Chem.* **2001**, 1769−1780.

(13) QM/MM accuracy can be reached using carefully parametrized force fields. Nonbonded interactions, which are extremely important in differentiating paths in asymmetric synthesis, are problematic in most QM approaches but well described by empirical parameters. Norrby, P.-O. Private communication.

(14) For an excellent account of enantiodiscriminating binding, see: Lipkowitz, K. B. *Acc. Chem. Res.* **2000**, *33*, 555−562.

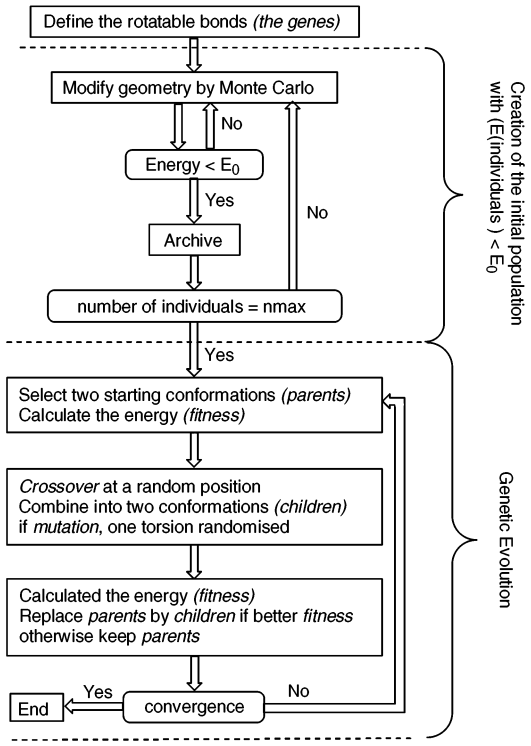
(15) (a) Corey, E. J.; Noe, M. *J. Am. Chem. Soc.* **1993**, *115*, 12579−12580. (b) Corey, E. J.; Noe, M.; Sarshar, S. *Tetrahedron Lett.* **1994**, *35*, 2861−2864. (c) Corey, E. J.; Noe, M. *J. Am. Chem. Soc.* **1996**, *118*, 319−329.

(16) Kollman, P. A.; Masova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889−897.

(17) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187−4225.

(18) (a) Corey, E. J.; Noe, M. C.; Grogan, M. J. *Tetrahedron Lett.* **1996**, *37*, 4899−4902. (b) Pidun, U.; Boehme, C.; Frenking, G. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2817−2820. (c) Dapprich, S.; Ujaque, G.; Maseras, F.; Lledos, A.; Musaev, D. G.; Morokuma, K. *J. Am. Chem. Soc.* **1996**, *118*, 11660−11661. (d) Torrent, M.; Deng, L.; Duran, M.; Sola, M.; Ziegler, T. *Organometallics* **1997**, *199*, 13−19. (e) DelMonte, A. J.; Haller, J.; Houk, K. N.; Sharpless, K. B.; Singleton, D. A.; Strassner, T.; Thomas, A. A. *J. Am. Chem. Soc.* **1997**, *119*, 9907−9908. (f) Haller, J.; Strassner, T.; Houk, K. N. *J. Am. Chem. Soc.* **1997**, *119*, 8031−8034.

SCHEME 1. Flow Chart of the Genetic Algorithm as Implemented in Our Calculations



co-workers who developed a semiquantitative model using a transition state MM2 force field.[11a]

It is noteworthy that the force field (a modified CFF91 parametrized as previously reported[10a]) was able to reproduce the energy differences between transition states initially computed on the basis of ab initio calculations with good accuracy.[18a,f] For instance, the energy differences for simple models of propene transition states[18a] or butene transition states[18f] were predicted within 0.3 kcal/mol compared to ab initio values.

With the conditions set, we turned our attention to evolutionary approaches, which are known to be highly efficient in mapping out the potential energy surface.[19] Simulated annealing and Monte Carlo approaches, which were initially used for conformational studies, were found to be time consuming. A genetic algorithm (GA) script was then developed in which the fitness was the energy and the genes were the rotatable torsion angles. During computations, all torsion angles were free to rotate. Scheme 1 presents the GA as implemented in our calculations (for more details, see the Experimental Section).

The initial population was constructed by randomly rotating the dihedral angles and roughly optimizing the resulting conformations by minimizing their energy function (Monte Carlo). The creation of these initial individuals was followed by a loop over generation, repeated until convergence was reached. A generation consists of three stages: (1) stochastic selection of the parents, (2) generation of the offspring via one-point crossover, random mutations, and quick optimization, and (3) fitness evaluation.

(19) Westhead, D. R.; Clark, D. E.; Murray, C. W. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 209−228.

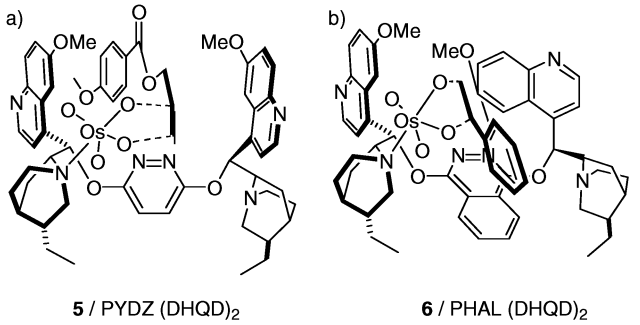


FIGURE 4. Corey- and Sharpless-type transition states. Geometries were proposed by Ujaque et al. and Norrby et al.[11c,d]

So as to secure and optimize the method, several computations were performed on AD of **5** and **6** (Figure 4). Working models with these two olefins were previously proposed by Corey and Sharpless and will be compared with our structures (Figure 4).[20] These two models suggested two different "pockets" for the binding. Although the Sharpless model was based on the [2 + 2] model, which has been proven unlikely, it was also supported by NMR studies. In addition, Norrby and Houk confirmed the existence of such an arrangement with the [3 + 2] model.[11c] This mechanistic proposal was therefore a good basis for our study.

First, adding a cutoff in energy for the construction of the initial population was found to fasten the computations. Thus, high-energy conformations were excluded from the initial population. Repetitive energy-minimization steps on every offspring were crucial to accelerating the evolution process (which resembles the Lamarckian GA described by Olson and co-workers[5]). However, they were, by far, the most time consuming and had to be finely tuned. Typically, 200 minimization steps, 100 individuals in the initial population, and mutation and crossover rates of 0.05 and 1, respectively, were optimal. From 40 to 60 generations were required to attain convergence, depending on the flexibility of the olefin under study. The quality of GA was corroborated by its reproducibility (the same minima was located regardless of the initial population). Reevaluation of the charge distribution (by Mulliken population computation) and full energy-minimization refined the resulting four conformations. Computation of the solvation contribution to the free energy was next performed using DELPHI module implemented in Insight II. These calculations included the polar contribution, which also accounts for the ionic strength of the solution, and nonpolar contributions by a solvent-accessible surface area-based method. These calculations, performed on $PHAL(DHQD)_2$/**5**, led to a sandwichlike model in agreement with that proposed by Corey (Figure 4). A transition state arrangement for $PHAL(DHQD)_2$/**6** was next proposed, which corroborated the high quality of the developed protocol.

Application and General Models. With this optimization achieved, we next applied the method to a large panel of olefins (chiral, achiral, aromatic, aliphatic, mono-, di-, and trisubstituted, see Figure 5). Cis olefins

(20) For a discussion on these models, see: Becker, H.; Ho, P. T.; Kolb, H. C.; Loren, S.; Norrby, P.-O.; Sharpless, K. B. *Tetrahedron Lett.* **1994**, *35*, 7315−7318.

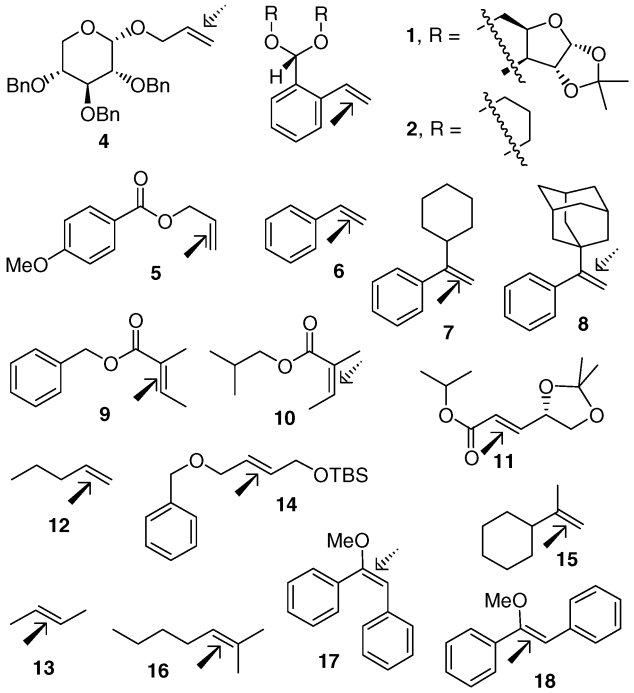


FIGURE 5. Computed olefins and reported senses of attack with AD-mix *β*.

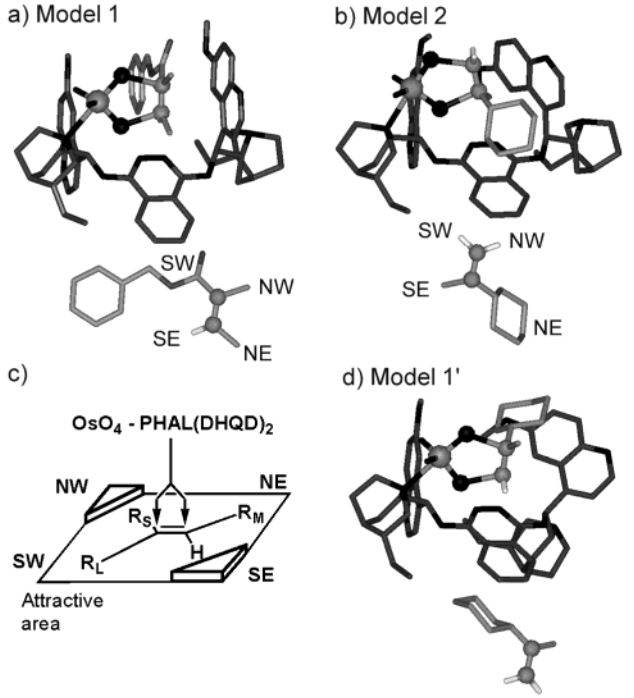


FIGURE 6. Two distinct transition state models as exemplified with (a) benzyl tiglate **9** and (b) 2-cyclohexylpropene **15**, (c) the empirical mnemonic device, and (d) the additional hybrid conformation with **15**.

were not studied because they have previously been shown to not work so well.[21] An α,*β*-unsaturated ester and two enol ethers were also studied. The olefin **4** was chosen for the observed enantioselectivity, which was not expected by the Sharpless mnemonic device. In addition, unique reversals of selectivity were reported between **7** and **8**[22] and between **9** and **10**.[23] Understanding these data would be of utmost interest. In addition, predicting such observations using our modeling protocol would also be a nice validation. Finally, upon treatment with AD-mix-*β*, both (*E*)- and (*Z*)-methyl enol ethers of benzoin (**17** and **18**) were found to give the same (*R*)-α-hydroxy ketone. Again these results "appear to violate Sharpless selectivity mnemonic by placing a phenyl subsituent in the quadrant where the smallest non-hydrogen substituent would lie".[24] In addition, this implies that the relative position of the phenyl rings does not influence the sense of attack, while the monosubstituted olefin part acts as styrene (**6**) does and controls the facial selectivity. There remains the question of the role of the disubstituted part in the binding.

First of all, we correlated the experimental and computational data and concluded that the computations predicted the correct isomer in all cases. Throughout the study, only two main possible transition states were found to be energetically accessible (Figure 6a,b). Although the catalyst conformations are essentially identical, the models differ dramatically in the orientation of the incoming olefin. Model 1 is similar to that proposed in our previous study and also similar to the model initially suggested by Corey and then by Ujaque[11d] using QM/MM methods (U-shaped catalyst). A closer look reveals that the flat group (**5**, **9**, **11**) is stacked between both of the methoxyquinoline walls and that the terminal carbon is attacked by equatorial osmium tetroxide oxygen. In model 2, which is equivalent to the Sharpless model,[25] a more bulky group (cyclohexyl of **15** in Figure 6b) lies on the phthalazine ring and the axial oxygen attacks the terminal carbon. An additional model (model 1′) was also found and depicted in Figure 6d. However, it can be seen as a hybrid of models 1 and 2. In this particular case, the catalyst adopts a conformation similar to model 2, which is the structure found by Sharpless based on NOE data, whereas the olefin adopts an orientation similar to model 1. The bystander (eastern) quinoline provides significant stabilizing interactions, while the working (western) quinoline is too far to interact with the olefin. Ideally, to adopt the "sandwich-like" conformation invoked in model 1, the double bond has to be separated from a hydrophobic group (aromaticity is not a prerequisite, see **11**) by a flat three-atom spacer (**5**, **9**, and **11**). Otherwise, the western (working) quinoline is too far away to provide a significant wall, and models 2 or 1′ are therefore preferred.

Our results are in agreement with those of Norrby and co-workers, who also found that the preference for Corey-type and Sharpless-type models varies from one olefin to another.[11c] Looking at the source of the interactions, two ligand pockets for binding are observed and probably govern the facial discrimination (Figure 7). These two

(21) Wang, Z.-M.; Kakiuchi, K.; Sharpless, K. B. *J. Org. Chem*. **1994**, *59*, 6895−6897.
(22) Vanhessche, K. P. M.; Sharpless, K. B. *J. Org. Chem*. **1996**, *61*, 7978−7979.
(23) Shao, H.; Rueter, J. K.; Goodman, M. *J. Org. Chem*. **1998**, *63*, 5240−5244.
(24) Hashiyama, T.; Morikawa, K.; Sharpless, K. B. *J. Org. Chem*. **1992**, *57*, 5067−5068.
(25) Kolb, H. C.; Andersson, P. G.; Sharpless, K. B. *J. Am. Chem. Soc*. **1994**, *116*, 1278−1291.

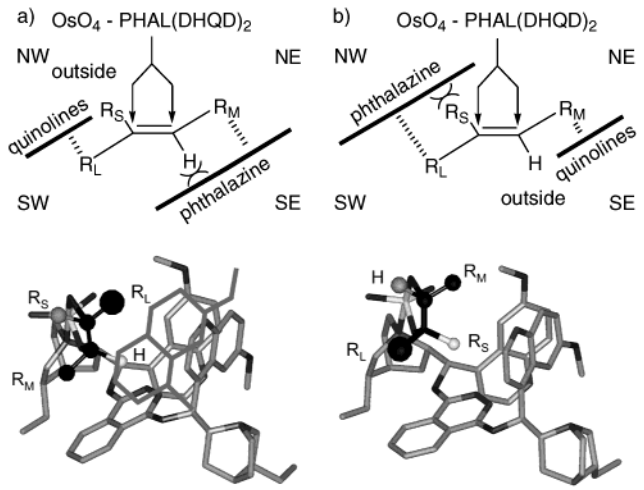


FIGURE 7. Rationalization of Sharpless' device. (a) Model 1, (b) model 2.

models shed light on critical stabilizing interactions favored by a polar solvent (1:1 *t*-BuOH/H_2O). Besides, taking into account the free energy of solvation led to refined predicted stereoisomeric excesses and a more predictive protocol. This shows that the solvent plays a decisive role in the binding process.

On the basis of a large body of work, Sharpless proposed a reliable empirical mnemonic device to predict the stereochemical outcome of the reaction. Our structural models were tentatively used to rationalize this empirical model, which works in most cases. We thus identified the role of each part as shown in Figure 7. The attractive area defined by Sharpless can be related to hydrophobic or π-stacking interactions with either the methoxyquinoline rings (model 1) or the phthalazine ring (model 2) of the catalyst. The phthalazine has a second role: while R_L can favorably interact with this "floor" in model 2, H (model 1) and R_S (model 2) clash with it. It is noteworthy that *both models should be considered* to explain the two sterically unfavored corners of the Sharpless empirical model.

Rationalization of the Experimental Data. In five cases (computations with olefins **2**, **6**, **8**, **10**, and **15**), models 1′ and 2 were found to have energy differences of less than 1 kcal/mol (Table 1). 1-Pentene (**12**) and 2-methyl-2-heptene (**16**) adopted the conformation in model 2. This was mainly due to the aliphatic character in conjunction with the length of these substrates that disfavored the first model, which is expected for shorter and/or aromatic olefins. Although Figure 6a,d present only the transition states with 2-cyclohexylpropene **15**, complexes with **2**, **6**, **8**, **12**, and **16** are highly similar. Compounds **5**, **9**, and **11** adopted the conformation in model 1. These three olefins share a common pattern: a double bond and a hydrophobic group linked by an ester moiety.

Compound **4**, which was dihydroxylated to yield the opposite diol to that predicted by Sharpless' device, approached the catalyst as in model 2 but with the terminal olefin carbon reacting with an equatorial oxygen (Figure 8). The tribenzylated xylopyranoside core was found to be too large to fit into the binding site. As a result, the location of the olefin is very different from both above-mentioned models. The approaching olefin nestles onto the catalyst with a concomitant decrease in the nonpolar solvent-accessible surface area and stabilizing interaction with both quinuclidine moieties and with the phthalazine ring (Figure 8a).

Upon treatment with AD mix β, both (*E*)- and (*Z*)-methyl enol ethers of benzoin gave the same corresponding (*R*)-α-hydroxyketone.[20] This particular data and the surprisingly high ee values (these two olefins cannot match with Sharpless mnemonic, vide supra) can be easily explained if one considers the proposed models. As already pointed out in the last section, two binding pockets, which lead to the same isomer, are offered by the ligand that consequently prefers trans olefins.[7,11c] The high enantioselectivity therefore arises from the interplay of the two binding pockets where both phenyl rings of **18** (as stilbene[11c]) exploit stabilizing interactions (with the phthalazine as in model 2 and the bystander quinoline rings as in model 1, see Figure 9b). This quantitative observation is in perfect agreement with the observed major isomer and high ee (99%). More surprisingly, **17** led also to the same isomer, albeit with lower ee (90%). This can be understood from Figure 9. Comparing the binding modes of **17** to that of styrene **6**, it can be seen that the least hindered phenyl ring of **17** is positioned in the catalyst in the same fashion as **6** (model 2). Therefore, the methoxy group in the trans configuration is not involved in the orientation of the olefin (Figure 9a).

Quantitative Results and Reversal of Selectivity. At the outset of this project, we had to make assumptions that resulted from a literature survey. For instance, the oxide was much more reactive in the presence of a ligand leading to a highly reactive axial oxygen. If these hypotheses would not influence any qualitative results, it may be useful to list them before entering a quantitative study. Thus, we made the following assumptions: (1) the reaction takes place only in the presence of the ligand, (2) no reaction happens in solution or outside the pocket, and (3) the olefin reacts with one equatorial oxygen atom and the axial oxygen atoms. At last, this "four transition state" approach assumes that only one arrangement led to the major isomer and one to the minor isomer. In other words, the Boltzmann distribution of conformers, which was considered by Norrby et al,[11c] was ignored. This might be a shortcoming when a quantitative protocol is expected. However, this was not a major fault when considering that a predictive but not quantitative method was our primary concern. In addition, this protocol was later found to be nearly quantitative (vide infra).

Since both paths leading to the minor and major isomers were considered, the energy computation allowed for the evaluation of theoretical enantio/diastereomeric excesses. So far, the qualitative analysis depicted the path to the expected isomer. Indeed, to understand and predict stereoselectivity, it is also necessary to examine the path to the minor isomer. For instance, the four proposed transition states for **6** (as defined in Figure 3) are illustrated in Figure 10. The first two (Figure 10a,b) would lead to the observed isomer, whereas the last two, which were higher in energy, would lead to the minor isomer. Examination of these four models indicates that this discrepancy (TS 3 and 4 lie more than 1.5 kcal/mol above TS 1) arises from a loss of interaction of the phenyl ring with the catalyst and an increase in the solvent-

TABLE 1. Experimental and Computed ee or de Values (AD-mix *β*)

olefin	observed ee or de	Sharpless device	models ($\Delta\Delta G$)[a]	predicted ee or de	comment
1	98 (*R*)[8c]	(*R*)	distorted 1′	49 (*R*)	chiral
2	98 (*R*)[8b]	(*R*)	2, 1′ (0.2)	74 (*R*)	
4	78 (*S*)[10a]	(*R*)	see text	66 (*S*)	chiral, unexpected isomer
6	97 (*R*)[7]	(*R*)	2, 1′ (0.9)	90 (*R*)	
7	57 (*R*)[22]	(*R*)	1′	53 (*R*)	
8	53 (*S*)[22]	**(*S*)/(*R*)**[b]	2, 1′ (0.2)	59 (*S*)	isomer opposite to **7**
5	98 (*R*)[28]	(*R*)	1	>99 (*R*)	
9	98 (2*S*,3*R*)[23]	(2*S*,3*R*)	1	98 (2*S*,3*R*)	
10	60 (2*R*,3*R*),[23] (2*S*,3*S*)[c]	**(*S*,*S*)/(*R*,*R*)**[d]	1′, 2 (0.8)	48 (2*S*,3*S*)	revised sense of attack (reported as opposite to **9**)
11	97 (*S*,*S*)[7]	(*S*,*S*)	1	>99 (*S*,*S*)	chiral olefin
12	79 (*R*)[7]	(*R*)	2	99 (*R*)	terminal olefin
13	72 (*R*,*R*)[7]	(*R*,*R*)	1′,2	99 (*R*,*R*)	trans-disubstituted olefin
14	90 (*R*,*R*)[7]	(*R*,*R*)	distorted 1′	95 (*R*,*R*)	trans-disubstituted olefin
15	69 (*R*)[7]	(*R*)	2, 1′ (0.4)	9 (*R*)	gem-disubstituted olefin
16	98 (*R*)[7]	(*R*)	2	>99 (*R*)	trisubstituted olefin
17	90 (*R*)[24]	(*R*)/(*S*)[d]	2	83 (*R*)	(*E*)-enol ether
18	99 (*R*)[24]	(*R*)/(*S*)[d]	1,2	>99 (*R*)	(*Z*)-enol ether

[a] $\Delta\Delta G = (\Delta G\text{model1}' - (\Delta G\text{model2})$, kcal/mol. [b] See ref 20 for discussion. [c] Revised configuration: see text. [d] These olefins cannot match with the Sharpless mnemonic.

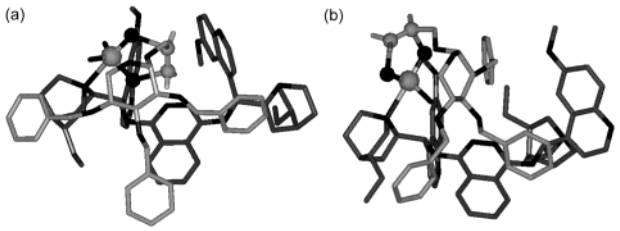


FIGURE 8. (a) Proposed transition state model for **4**. (b) Proposed transition state leading to the minor isomer.

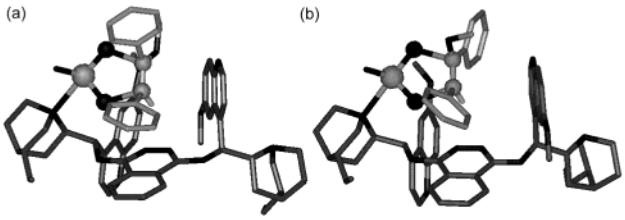


FIGURE 9. Proposed transition state models with **17** and **18**.

accessible surface area. In these last two cases, the aromatic ring was surrounded in water leading to unfavored interactions. This observation confirms the decisive role of the hydrophobic and aromatic interactions with the catalyst and the role of the aqueous medium (at the origin of the hydrophobic interactions).

Our approach allowed us to explain reversals of selectivity as well. As highlighted in Figure 11 compared to Figure 10, the most energetically favored transition state for **8** (Figure 11c) is not similar to the model proposed for styrene **6** (Figure 10a). Steric and hydrophobic interactions with the adamantyl moiety compete with hydrophobic and aromatic interactions with the phenyl ring. In any case (**6**−**8**), the interactions between the phenyl ring and the phthalazine floor were favored over other hydrophobic or aromatic stacking. However, the bulky adamantyl ring induced a new attack of the OsO_4. The adamantyl group is located in the NE binding pocket (as in model 2, see Figure 6b), while the phenyl ring lies on the phthalazine. However, steric clash at that position is known to destabilize the complex. Indeed the

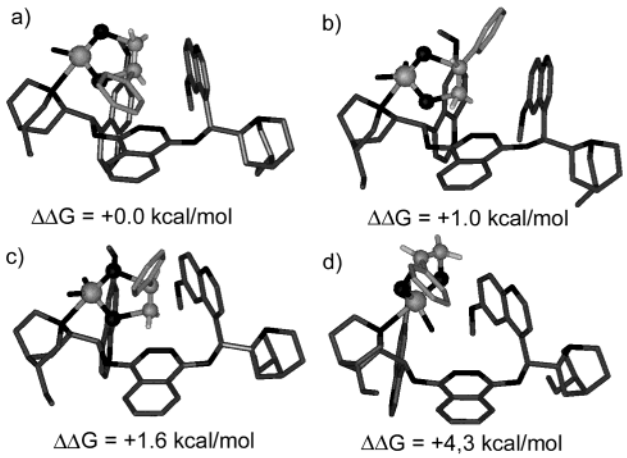


FIGURE 10. Four computed transition states for styrene **6**.

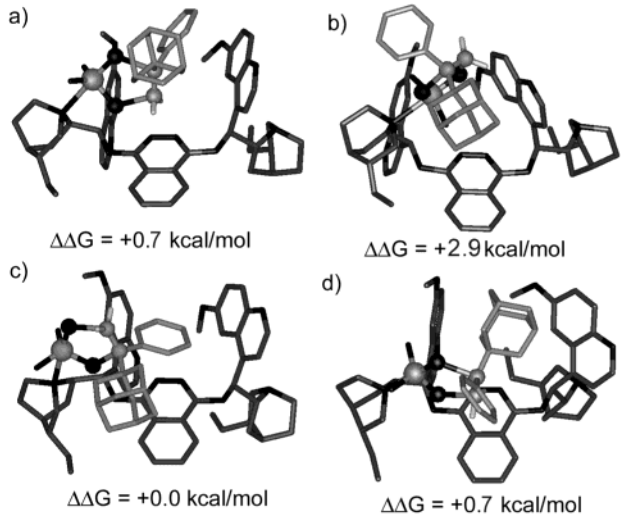


FIGURE 11. Four computed transition states for α-adamantyl-styrene **8**.

olefin/OsO_4 transition state is pushed away to allow the aromatic ring to fit in the binding site. Examination of the four proposed transition states revealed that a second

complex should exist in solution (Figure 11d) that also would lead to the observed (but unexpected) isomer. This model resembles the model 1′ shown in Figures 6c and 10a for styrene. However, in this particular case, the participating hydrophobic interactions with the adamantyl moiety in the binding pockets overcome the aromatic and hydrophobic interactions with the phenyl ring and dictate the dominance of the experimentally observed (*S*)-isomer.

From that data, we can draw preliminary conclusions. First, the solvent-driven competition between the possible aromatic and hydrophobic interaction dictates the facial selection. Second, bulky groups (see **4** and **8**) may introduce new steric clash and induce significant deviation from the "ideal" structure.

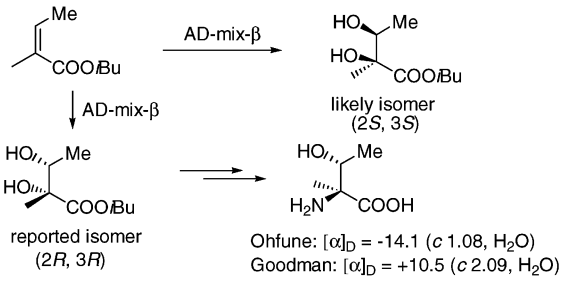
Careful examination of all the data would make possible the analysis of the path to the minor isomer. As addressed in the protocol section, all these computations assumed that the reaction takes place only with the proposed transition states (in the presence of the ligand and with the axial oxygen). We therefore did not pay attention to any other transition state arrangement (such as reaction with two equatorial oxygen atoms, which is unlikely). For the mechanistic path minor isomers, no general scheme (for a general scheme to the major isomer, see Figure 7) can be drawn. Either steric clash with the catalyst, lack of interaction or unfavorable interaction with the aqueous medium disfavored the attack of the oxide by the other face of the olefin (examples are shown in Figures 10 and 11).

Comparing the binding of benzyltiglate **9** and *iso*-butylangelate **10** was of interest since Goodman and coworkers reported an opposite sense of attack between **9** and **10**.[23] Indeed, the latter olefin does not easily match with the Sharpless mnemonic device (**10**, R_L and H are trans; mnemonic device, R_L and H are cis). The SE quarter of the mnemonic is the most sterically crowded, and steric clash at that position appeared as detrimental to the binding and to the control of the enantioselectivity.

To our surprise, the proposed protocol showed that olefin **9** binds to the catalyst according to model 1 (Figure 7a), whereas olefin **10** would bind according to model 1′. However, both models lead to the same (2*S*)-isomer. Thus, our computational model was in agreement with experimental results for olefin **9** but predicted for AD with olefin **10** the (2*S*,3*S*)-configuration that is opposite to the one reported. Indeed, Sharpless AD was the key step in the enantioselective preparation of (2*S*,3*R*)-α-methylthreonines. Looking for reference optical rotation values,[26] we found out that Moon and Ohfune had already published these four enantiomers and diastereomers of α-methylthreonines.[26b] Corroborating the optical rotation values, we proposed that the observed diol was not (2*R*,3*R*) but the initially expected (2*S*,3*S*) (Scheme 2).[27] This last example highlights the high predictability of our protocol and the role it can play in such a project.

Quantitative Results and Predicted Excesses. To fulfill the convergence criterion, two runs were performed

SCHEME 2. Revised Configuration of Goodman's Diol



on each system. Indeed, a slightly more energetically favored transition state was found in only two cases (**8**, first run 51%, second run 59%; **9**, first run 97.7%, second run 97.9%). Table 1 summarizes the data computed for the 17 olefins. The computation predicted the enantio-(diastereo-)selectivity with less than 25% error in 14 cases and less than 10% in 10 out of 17 cases (Table 1). These results, when compared to those of Norrby et al. (25%, 13/15; 10%, 12/15)[11c] indicated that the lack of relaxation at the frozen reacting center was not too much of an approximation. Decomposition of the total free energy difference into the molecular mechanics and solvation energies also suggests that the latter plays a role in the catalytic process. So far, this role was ignored in the other mechanistic studies and might be at the origin of a part of the error.

The reversal of selectivity from **8** to **9** and the unexpected outcome with **4** have been predicted with excellent accuracy. Surprisingly, the data in Table 1 revealed that the bigger the olefins, the more accurate the predictions (**4**, **5**, **7**, **8**, **14**, **16** vs **13**, **15**). Similarly, the accuracy decreases for the aliphatic olefins (**12**, **13**, **15**). This might arise from an inefficient treatment of the hydrophobic interactions.

Conclusions

Given the poor description of the stacking geometries of the force fields,[29] the presumed aggregate formation, the heterogeneous medium, and the reaction condition requirements,[30] the results described in the present work are highly reliable. The use of a GA as a conformational search algorithm allowed for a fast study of AD with flexible olefins. The full computational protocol quickly provided three-dimensional models for AD that are in good agreement with Norrby's models based on calculations using an advanced transition state force field and a Monte Carlo/Low Mode search algorithm. Moreover, a semiprediction on the stereoisomeric excess was made possible by a refined computation of the free energy difference including the solvation contribution. Finally, two alternative models were proposed and subsequently used to rationalize Sharpless' device.

Further improvements such as including a comparison step in the GA would provide a final generation with individuals that would all be different (for instance, more

(26) (a) Seebach, D.; Aebi, J. D.; Gander-Coquoz, M.; Naef, R. *Helv. Chim. Acta*. **1987**, *70*, 1194−1216. (b) Moon, S.-H.; Ohfune, Y. *J. Am. Chem. Soc.* **1994**, *116*, 7405−7406.

(27) The authors compared the optical rotation of the final compounds to Seebach's previously reported values.[26a] However, Seebach had disclosed the (2*R*,3*R*)- and (2*S*,3*S*)-α-methylthreonines but neither (2*R*,3*S*) nor (2*S*,3*R*) that were expected by Goodman's approach.

(28) For ee obtained with $PYDZ(DHQD)_2$: Corey, E. J.; Guzman-Perez, A.; Noe, M. C. *J. Am. Chem. Soc.* **1994**, *116*, 12109−12110.

(29) Kamishima, M.; Kojima, M.; Yoshikawa, Y. *J. Comput. Chem.* **2001**, *22*, 835−845.

(30) Mehltretter, G. M.; Döbler, C.; Sundermeier, U.; Beller, M. *Tetrahedron Lett.* **2000**, *41*, 8083−8087.

than a user-defined value of the root mean square deviation). This protocol would ultimately account for the Boltzmann distribution. In addition, combination of refined parameters for the transition states (such as those developed by Norrby et al.) in the force field with the described GA should provide a more quantitative tool. Application to other catalytic asymmetric reactions is underway.

Experimental Section

General Remarks. Computational simulations were performed with the Insight II 2000 package using the CFF91 force field. The osmium set of CFF91 parameters have been derived from the corresponding CVFF parameters reported previously.[10a] Moreover, the possible poor reliability of the implemented osmium parameters has been circumvented by constraining the complex in the starting conformation (frozen transition state). The used genetic algorithm was written in BTCL language (Insight II User Guide). Graphical displays were printed out from the Insight II molecular-modeling system.

Construction of the Systems. The initial systems were built from the ab initio transition states reported in the literature.[18e] Standard atom partial charges were assigned to the systems that were optimized with the modified force field. Atom partial charges were next generated using the MNDO semiempirical method for the determination of the Mulliken electronic population (implemented in the AMPAC/MOPAC module).

Genetic Algorithm. The algorithm was written following the flowchart below (see also Figure 4):

1. Create initial population (100 individuals) by Monte Carlo: (i) randomly rotate the torsions, (ii) calculate the fitness (energy) of the solutions, and (iii) keep those with fitness less than a user-defined value (in this case $E < 300$ kcal/mol). This value depends on the system and the force field in use.

2. Begin a user-defined number of genetic operations (population usually converges within 40−60 generations)

2.1. Select two parents within the 100 individuals: the first 50 individuals were successively chosen as the father, and the mother was randomly chosen among the last 50. Then the last 50 were chosen successively as the father, and the mother was randomly chosen among the first 50. This allows an efficient mixing of the genes.

2.2. Produce two children by one-point crossover: (i) choose a random position, (ii) divide parents at this point, and (iii) obtain children by combining the first piece of one parent with the second piece of the other parent.

2.3. Apply mutation: randomization of a gene (torsion) with a user defined rate (0.05 in this case).

2.4. Optimize the solution (local search): optimizations were performed by conjugate gradients energy minimization with a convergence criterion of 0.001 kcal/mol. Solvent conditions were represented implicitly using a distance-dependent dielectric constant $\epsilon = r$.

2.5. Replace the least fit parent if the child's fitness (energy) is lower. Since the steady-state form was applied (continual update of the population), the concept of generation may not be obvious. However, the successive selection of the father (see 2.1) is iterative over the population. A generation is therefore considered every 100 selections.

2.6. If the population has converged, then end; otherwise use 2.1.

This algorithm was applied to the four transition states built up previously and led to four models.

Measure of the Total Free Energy Differences. Again, atom partial charges were generated using the MNDO semiempirical method for the determination of the Mulliken electronic population (implemented in the AMPAC/MOPAC module). The resulting structures were optimized by conjugate gradients energy minimization with a convergence criterion of 0.001 kcal/mol and a distance-dependent dielectric constant $\epsilon = r$. This force field computation gave the values of E_{MM} presented in Figure 2. The solvation free energy was next computed using the DELPHI module (Insight II User Guide). This allows for the calculation of the electrostatic potential in and around the complex using a finite difference solution of the nonlinear Poisson−Boltzmann equation and therefore of the electrostatic contribution to the solvation energy. In addition, the program allows specification of the ionic strength. Using the same module allowed computation of the nonpolar contribution to be carried out using a surface-based method. The solvation free energy was calculated by the difference of these values computed in water ($\epsilon = 80$, ionic strength $= 2$) and in a vacuum ($\epsilon = 1$, ionic strength $= 0$).

The free energy solvation was next added to E_{MM} to provide an estimation of ΔG. The four models were compared, and a value of $\Delta\Delta G$ was attributed to each of them.

Acknowledgment. We thank Dr. Bernard Magret and Dr. Shawn Johnstone for critical reading of this manuscript and Institut Nancéien de Chimie Moléculaire for financial support.

Supporting Information Available: Calculated energies and predicted enantio- and diastereoisomeric excesses (Table S1) and structures and Cartesian coordinates of representative transition states with **5** (model 1) and **6** (model 2 and 1′). This material is available free of charge via the Internet at http://pubs.acs.org.

JO0258148