Journal of Computer-Aided Molecular Design **18:** 303–308, 2004.

Reverse-docking as a computational tool for the study of asymmetric organocatalysis

D. Joseph Harriman & Ghislain Deslongchamps*
Department of Chemistry, University of New Brunswick, Fredericton, New Brunswick, Canada E3B 5A3

Received 20 May 2004; accepted in revised form 9 July 2004

Key words: asymmetric synthesis, AutoDock, azidation, catalysis, docking, MOE, molecular mechanics, organocatalyst, scoring

Summary

A novel methodology for 'reverse-docking' a cationic peptide-based organocatalyst to a rigid anionic transition state (TS) model for the conjugate addition of azide to α,β-unsaturated carbonyl substrates is described. The resulting docking poses serve as simplified TS models for enantioselective catalysis. Molecular mechanics-based scoring and ranking of the docking poses, followed by clustering and structural analysis, reveal a clear energetic preference for docking to the *S*-enantiomeric azidation TS model, in agreement with experiment. Clear energetic trends emerged from docking the catalyst to both enantiomers of all six azidation TS models of this study. Structural analysis of the most favorable pose suggests a mechanism for enantioselective catalysis that is consistent with principles of molecular recognition, catalysis, and experimental data.

Introduction

Peptide catalysts have re-emerged in the last decade as a viable approach to asymmetric catalysis [1, 2]. The metal-free variants, referred to as organocatalysts, tend to be tolerant to water and aerobic conditions. Of particular interest are the 'Miller catalysts', a series of peptide-based organocatalysts identified through combinatorial synthesis capable of catalyzing different asymmetric reactions including acylations [3–5], phosphorylations [6], and conjugate additions of azide [7, 8]. Although these catalysts display significant enantioselectivities, little is known of their exact mechanism of action. Transition state (TS) models of asymmetric reactions, whether based on energetics or simpler mnemonics, are valuable tools for understanding observed stereoselectivities and for predicting the stereochemical outcome of such reactions. However, devising useful TS models of the Miller catalysts poses a considerable challenge due to their conformational flexibility, as they range from tri- to octapeptides. We wish to report a 'reverse-docking' methodology for developing a transition state model for the peptide-catalyzed asymmetric addition of azide to α,β-unsaturated compounds.

*To whom correspondence should be addressed (Fax: +1-506-453-4981; E-mail: ghislain@unb.ca)

Miller azidation organocatalyst

The amine-catalyzed conjugate addition of hydrazoic acid was first reported in 1997 [9]. Upon reporting the conjugate addition of azide catalyzed by various amine bases, including imidazoles [10], Miller developed a peptide-based organocatalyst that could deliver azide anion (N_3^-) to unsaturated imides enantioselectively (Figure 1). A histidine side-chain, present in this catalyst, is thought to form a corresponding imidazolium azide salt upon addition of hydrazoic acid (*via* $TMSN_3$/pivaloic acid); this would deliver the azide nucleophile to the α,β-unsaturated substrate upon binding to the organocatalyst. Although the reaction shows good enantioselectivity for a variety of substrates, a three-dimensional transition state model consistent with these results has yet to be devised.

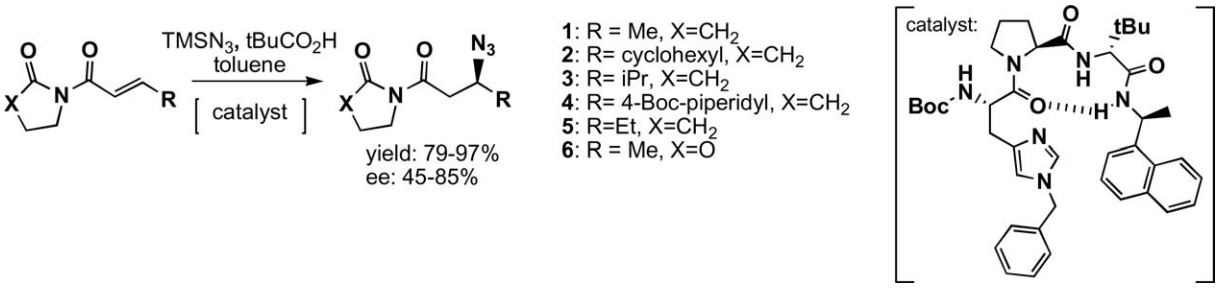

Figure 1. Asymmetric azidation of α,β-unsaturated carbonyl with Miller organocatalyst.

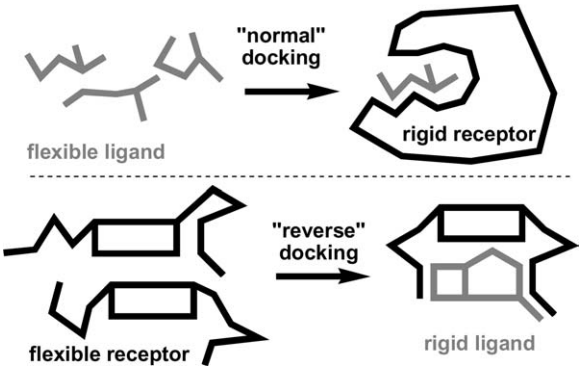

Figure 2. Reverse-docking versus normal docking.

Reverse-docking

We have developed a computational approach for deriving a useful TS model of the Miller azidation reaction based on the concept of 'reverse-docking' (Figure 2). While typical docking involves placing a flexible ligand into a largely rigid receptor, reverse-docking places a flexible receptor around a small, rigid ligand. For the present study, the flexible receptor corresponds to the Miller azidation catalyst (with protonated imidazole) while the ligand is mapped to a rigid representation of the preferred transition state (TS) for addition of N_3^- to the unsaturated substrate.

Reverse-docking would then be used to search the configurational space for the most energetically favored catalyst/TS-model complex. The best docking poses would represent a simplified TS model for the asymmetric reaction. Reverse-docking to a rigid TS of the azide addition, as opposed to the immediate azidation product (i.e. O-enolate produced upon N_3^- addition) should be compatible with the Curtin-Hammett hypothesis. The inherent rigidity of the α,β-unsaturated imides makes them ideal substrates for this study; all bonds would be fixed except for the incipient azide β-carbon bond. With this strategy, the ability to search through a vast configurational space will come at the expense of a simplified computational approach [11, 12]. Notwithstanding, the method should provide insight on the catalysis and a qualitative account of the observed enantioselectivities, especially when comparing docking energies to the two enantiomeric azidation transition states.

Methodology

Catalyst modeling. The peptide catalyst (with protonated histidine) was modeled initially with the MMFF94s [13] force field in MOE [14], then geometry optimized at the HF/6-31G* level with Gaussian 98 [15]. RESP charges [16] were subsequently calculated using the Antechamber function in Amber 7 [17]. Modeling the catalyst with a protonated histidine was justified on the basis that the pKa of hydrazoic acid (formed *in situ* by reaction of $TMSN_3$ with $tBuCO_2H$) is 4.5, as opposed to 6.5 for histidine. Thus, docking the catalyst with a protonated imidazole (total charge $+1$) to the azidation TS model (total charge -1) should account for electrostatic interactions between the two species, provided that an accurate charge model was used. For this reason, RESP calculations were carried out on both catalyst and TS model prior to docking, even if it meant lengthy computation times ($\approx$ 5 weeks for catalyst).

Azidation TS modeling. The unsaturated imide substrates, azide anion, and azidation products (O-enolate form) were initially modeled using MMFF94s in MOE, then geometry optimized in Gaussian 98 (HF/6-31G*). The TS structures were obtained by QST3 (STQN) redundant internal coordinate calculation in Gaussian 98 [18]. RESP charges [16] for ground states and TS were then calculated using Antechamber. Both enantiomers of the TSs leading to S and R azidation products were generated and used in subsequent reverse-docking experiments.

Reverse-docking. Reverse-docking was carried out with AutoDock 3.0.5 [19] by reversing the roles of ligand and receptor. The aqueous solvation term was excluded from the scoring to give a closer mimic of the low-dielectric environment of the toluene solvent. To ensure adequate sampling of the peptide poses by AutoDock's genetic algorithm, the following parameters were used: population size: 50; generations: 27,000; energy evaluations: 1,500,000; all dockings carried out in triplicate. These correspond to the same parameters as those reported for docking the largest flexible ligands in the original AutoDock test set [19]. Increasing the population size and energy evaluations in the reverse-docking simulations showed no significant improvement in the energy or ligand conformations. Following the docking simulations, the resulting poses were re-processed in MOE.

Re-minimization and scoring: The resulting AutoDock pose databases (100 poses each) were imported in MOE, the azidation TS model was fixed and the catalyst poses were energy-minimized (MMFF94s, solvent dielectric: 2.4) and scored based on MOE-Dock's internal and non-bonded energy terms (Coulombic and van der Waals).

Clustering criteria. Reverse-docking poses were clustered on the basis of H-bonding between the catalyst's imidazolium N-H and the azidation TS model (Figure 3).

Results and discussion

Docking. AutoDock was chosen for the docking simulations for its efficient Lamarckian Genetic Algorithm (LGA) search capabilities. However, AutoDock uses a scoring function (as opposed to an energy-based function) trained on a limited number of ligand/receptor complexes in water with Gasteiger-Marsili charges [20] (even for the ligands). For this reason, catalyst docking poses obtained with AutoDock were imported into MOE databases, energy-minimized (MMFF94s) with toluene solvation ($\varepsilon = 2.4$) around rigid azidation TS models and scored in MOE (E = potential energy of catalyst + intermolecular interactions [electrostatic + van der Waals]). This allowed for a molecular mechanics-based geometry optimization of the peptide poses around the fixed azidation model, a necessary energetic correction due to the poor performance of AutoDock's scoring function in this reverse-docking setting.

Pose clustering. Based on principles of general acid/base catalysis, hydrogen bonding distance and angle criteria were used to cluster the databases, as seen in Figure 3. In the 'carbonyl-type' poses, the imidazolium group of the catalyst H-bonds to the incipient enolate oxygen of the azidation TS model, while in the 'azide-type' poses, the imidazolium H-bonds to the terminal azide nitrogen. Poses having N-H$\cdots$X bonding distances ≤ 4Å and corresponding N-H·X angles between 100° and 180° were retained for clustering. This produced databases of catalytically relevant pose clusters with MOE docking energies. In addition, docking ranks determined before the clustering were recorded for each entry. Thus, the overall rank of a clustered pose, relative to all the docking poses (i.e. prior to clustering), was retained for analysis and discussion. Reverse-docking results for the catalyst with both enantiomers of the azidation TS model, carried out in triplicate, are reported in Table 1.

Reverse-docking energies. Table 1 reveals a large energetic discrepancy between carbonyl-type docking poses and their azide-type counterparts. The carbonyl-type poses represent not only the bulk of the output poses from the docking runs, but also consistently give the best energy rank. For all docking experiments (36/36) the carbonyl-type catalytic poses were more energetically favorable than the azide-types. On average, these energetic differences are greater for the experimentally relevant *S*-enantiomer models (12.6 kcal/mol) than for the *R*-enantiomer models (10.2 kcal/mol). All the carbonyl-type poses emerge within the top three ranks prior to clustering.

Enantioselectivity. Since the enantioselectivity of the azidation always favored the formation of the *S*-enantiomer at the β-carbon of the product, one would anticipate that the docking energies for the *S*-enantiomer of the azidation TS model would be lower than those of the corresponding *R*-enantiomer. Based on the energies of the highest-ranked poses, the results agree with the enantioselectivity trend. Indeed, for 18/18 docking runs, the highest ranking carbonyl-type pose of the *S*-enantiomer shows a significantly lower energy than that of the *R*-enantiomer. Only a qualitative correlation of enantioselectivity was obtained, since the average energy difference between *S* and *R* enantiomers for the carbonyl-type docking poses was 10.1 kcal/mol ($\sigma = 3.4$), corresponding essentially to 100% enantiomeric excess (compared to 45–85% ee experimentally).

3D transition state model for asymmetric organocatalysis. Interestingly, many of the carbonyl-type

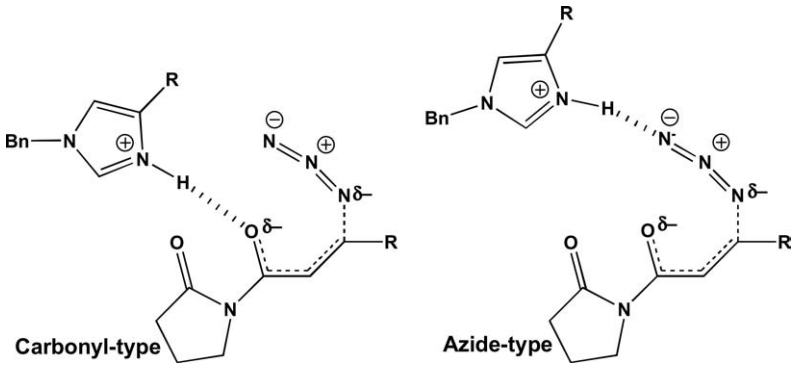

Figure 3. Two types of clusters used to categorize reverse-docking poses of the organocatalyst to the azidation TS model.

Table 1. Reverse-docking results following clustering.

Substrate[a]		*S* Enantiomer[b]				*R* Enantiomer			
		Carbonyl[c]		Azide[c]		Carbonyl		Azide	
		E[d]	(rank)[e]	E	(rank)	E	(rank)	E	(rank)
1	A	56.4	(1)	78.7	(8)	63.8	(1)	79.0	(15)
	B	62.2	(1)	73.4	(3)	77.5	(5)	79.9	(8)
	C	66.7	(1)	77.7	(5)	78.0	(9)	80.7	(14)
2	A	73.9	(1)	83.3	(5)	83.0	(3)	84.4	(4)
	B	82.8	(1)	91.6	(12)	86.9	(5)	95.8	(16)
	C	78.2	(1)	86.9	(9)	86.6	(2)	95.2	(20)
3	A	79.9	(3)	89.1	(12)	91.7	(10)	95.1	(18)
	B	79.4	(3)	83.7	(5)	91.6	(10)	95.1	(18)
	C	78.2	(2)	85.2	(9)	84.6	(5)	97.0	(35)
4	A	71.0	(2)	87.2	(29)	87.2	(15)	89.3	(31)
	B	71.0	(1)	93.8	(31)	82.3	(4)	86.3	(13)
	C	71.6	(1)	92.1	(27)	80.5	(2)	85.1	(11)
5	A	80.8	(1)	90.0	(10)	96.6	(18)	99.2	(21)
	B	74.0	(1)	88.0	(7)	84.3	(3)	90.2	(15)
	C	81.9	(1)	90.1	(10)	89.4	(14)	90.6	(17)
6	A	65.0	(1)	73.9	(5)	77.2	(1)	86.1	(3)
	B	68.4	(2)	89.2	(14)	73.9	(7)	89.5	(22)
	C	68.6	(2)	83.2	(10)	77.8	(2)	88.0	(14)

[a]Substrates according to Figure 1, rows A–C refer to triplicate reverse-docking runs. [b]Enantiomer of azidation TS model leading to *S*-product. [c]Cluster type according to Figure 3. [d]Docking energy of lowest-energy entry as calculated by MOE (kcal/mol). [e]Overall docking rank based on MOE energy prior to clustering (1 = best rank).

poses for all six substrates, including their lowest energy structures, are consistent with a TS model favoring formation of the product *S*-enantiomer where the imidazolium ring could play a dual role (Figure 4a). Indeed, the imidazolium ring is close to the azide ($\approx$ 3.8 Å), suggesting good Coulombic interaction, while the imidazolium NH group is hydrogen bonded to the incipient enolate (N-**H**·**O**: 1.85 Å). Bifurcated hydrogen bonding of the former with the enolate and distal carbonyl oxygens was also observed in some poses. Thus, in addition to delivering azide anion to the α,β-unsaturated substrate *via* electrostatic tethering, the histidine side-chain may further participate in substrate anchoring and ensuing electrophilic catalysis *via* H-bonding to the incipient enolate oxygen of the substrate. H-bonding between the ring carbonyl of the substrate and the C-terminal NH group of the catalyst may provide further substrate anchoring. Finally, the structure depicted in Figure 4a reveals a β-turn pattern between the histidine carbonyl and the C-terminal NH

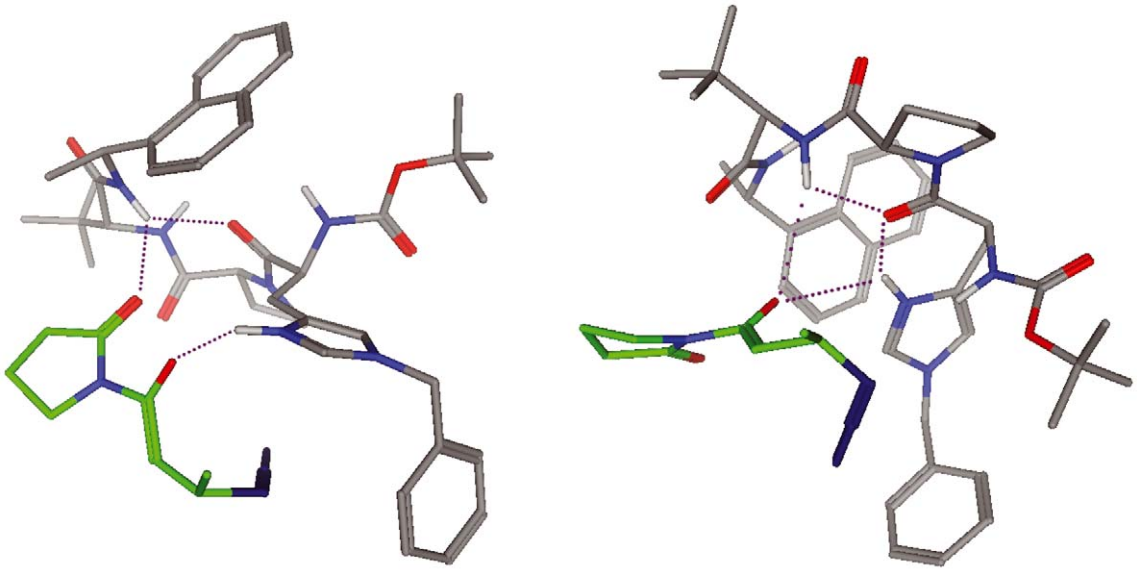

Figure 4. Best reverse-docking poses (carbonyl-type) of the catalyst to the azidation TS model for substrate **1.** (a) *S*-enantiomeric TS model (E = 56.4 kcal/mol, rank = 1); (b) *R*-enantiomeric TS (E = 63.8 kcal/mol, rank = 1). Non-polar hydrogens omitted for clarity.

group which is consistent with solution NMR data reported for the neutral catalyst alone [21]. Interestingly, this intramolecular peptide H-bond was not observed until re-minimization of AutoDock poses in MOE. The binding scenario described herein is fully compatible with the non-competitive nature of the solvent, toluene.

In sharp contrast, the analysis of the docking poses to the *R*-enantiomeric azidation TS model, which were much higher in energy as a whole, revealed a much less consistent picture for catalysis with no apparent structural trend. The poses did not hint to a coherent mechanism for recognition and catalysis. For example, the best pose depicted in Figure 4b suggests no clear electrostatic interaction between azide and imidazolium, yet suggests two intramolecular H-bonds (6- and 7-membered) within the catalyst. Similar conclusions could be drawn from inspection of the azide-type poses, further to the fact that their energies were also much higher than those of their carbonyl-type counterparts.

An important extension of this work will be to correlate docking energies with the enantiomeric excesses. One of the current limitations of this initial modeling approach is that the azidation TS model is completely rigid, which does not allow for rotation of the emerging bond between the azide and the β-carbon of the substrate in response to binding to the catalyst. Select TS models derived from this study can be studied by *ab initio* calculations and mixed-mode simulations to probe structural adjustments at the transition state, in comparison to our semi-rigid models.

We are currently testing a Lamarckian genetic algorithm for docking, written in SVL, for incorporation into MOE. This would obviate having to use AutoDock for generating poses, eliminate the geometry optimization step in MOE prior to scoring, and considerably streamline the computational process.

Conclusions

In summary, a reverse-docking scheme was devised to dock a cationic organocatalyst around a rigid anionic TS model for addition of azide to α,β-unsaturated carbonyl substrates. The resulting reverse-docking poses were used as simplified TS models for enantioselective catalysis. Molecular mechanics-based scoring and ranking of the docking poses, followed by clustering and structural analysis, revealed a clear energetic preference for docking to the *S*-enantiomeric azidation TS model, in agreement with experiment. Clear energetic trends emerged from docking the catalyst to both enantiomers of all six azidation TS models, in triplicate. Structural analysis of the most favorable pose suggests a mechanism for enantioselective catalysis

that is consistent with principles of molecular recognition, catalysis, and experimental data. Ultimately, it is our hope that further development of this reverse-docking approach may produce a useful tool for the rational design of peptide-based organocatalysts and other nano-molecular devices.

Acknowledgements

Financial support from the Natural Sciences and Engineering Research Council, the New Brunswick Innovation Foundation, and the University of New Brunswick is gratefully acknowledged. We are also grateful to the Chemical Computing Group Inc. for software licenses and SVL support.

References

1. Jarvo, E.R. and Miller, S.J., Tetrahedron, 58 (2002) 2481.
2. Dalko, P.I. and Moisan, L., Angew. Chem. Int. Ed. Engl., 40 (2001) 3726.
3. Jarvo, E.R., Copeland, G.T., Papaioannou, N., Bonitatebus Jr., P.J. and Miller, S.J., J. Am. Chem. Soc., 121 (1999) 11638.
4. Copeland, G.T., Papaioannou, N., Horstmann, T.E., Ruel, E.M. and Miller, S.J., J. Am. Chem. Soc., 120 (1998) 1629.
5. Jarvo, E.R., Vasbinder, M.M. and Miller, S.J., Tetrahedron, 56 (2000) 9773.
6. Sculimbrene, B.R. and Miller, S.J., J. Am. Chem. Soc., 123 (2001) 10125.
7. Guerin, D.J. and Miller, S.J., J. Am. Chem. Soc. 124 (2002) 2134.
8. Horstmann, T.E., Guerin, D.J. and Miller, S.J., Angew. Chem. Int. Ed. Engl., 39 (2000) 3635.
9. Lakshmipathi, P. and Rama Rao, A.V., Tetrahedron Lett., 38 (1997) 2551.
10. Guerin, D.J., Horstmann, T.E. and Miller, S.J., Org. Lett., 1 (1999) 1107.
11. Moitessier, N., Henry, C., Len, C. and Chapleur, Y., J. Org. Chem., 67 (2002) 7275.
12. Dudding, T. and Houk K.N., Proc. Natl. Acad. Sci. USA, 101 (2004) 5770.
13. Halgren, T.A., J. Comput. Chem., 20 (1999) 720.
14. Molecular Operating Environment version 2003.04, Chemical Computing Group Inc., Montreal, Canada.
15. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery, J.A. Jr., Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayalo, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J.V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, I., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E.S. and Pople, J.A., Gaussian 98 Rev. A.9, Gaussian Inc., Pittsburgh, PA (1998).
16. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.
17. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A., Amber 7, University of California, San Francisco, CA (2002).
18. Peng, C., Ayala, P.Y. and Schlegel, H.B., J. Comput. Chem., 17 (1996) 49.
19. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comput. Chem., 19 (1998) 1639.
20. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3210.
21. Blank, J.T., Guerin, D.J. and Miller, S.J., Org. Lett., 2 (2000) 1247.