Computational catalysis

# Quantum molecular interaction field models of substrate enantioselection in asymmetric processes☆

Marisa C. Kozlowski *, James C. Ianni

*Department of Chemistry, Roy and Diana Vagelos Laboratories, University of Pennsylvania, Philadelphia, PA 19104, USA*

### ABSTRACT

Computational models correlating substrate structure to enantioselection with asymmetric catalysts using the QMQSAR program are described. In addition to rapidly providing predictions that could be used to facilitate the screening of catalysts for novel substrates, the QMQSAR program identifies the portions of the substrate that most directly influence the enantioselectivity. The lack of an underlying relationship between all the substrates in one case, requires two quantitative structure selectivity relationships (QSSR) models to describe all of the experimental results.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

The ultimate goals of asymmetric catalysis are the discovery of reactions that provide desired products with high enantioselectivities and yields. Given the immense effort applied in this endeavor, computational tools are a logical resource to facilitate the design and optimization of asymmetric reactions [1,2]. However, even with massive increases in computer speed, the number of variables in any asymmetric transformation makes modeling of discrete transition states often challenging. The use of other computational tools encompassing linear free energy relationships, such as quantitative structure activity relationships (QSAR) [3,4], has been shown to provide useful information. In prior reports [5–8] we introduced the development of grid-based quantitative structure selectivity relationship (QSSR) models using quantum mechanical molecular interaction fields for correlating enantiomeric excess with catalyst structure. Other workers have embraced this approach with success [9–13]. In addition, intriguing reports utilizing QSSR models to predict substrate enantioselection with chiral catalysts have appeared [14,15]. Here, we report our independent efforts toward generating models to predict enantioselection for substrates which has also led to a tool to identify potential mechanistic differences between substrate classes.

## 2. Methods and experimental data

### 2.1. Quantum molecular interaction fields

The enantiomeric excesses arising from the aldehyde substrates were analyzed with 3D-QSSR (quantitative structural selectivity relationship) methods employing quantum molecular interaction fields as implemented in the program QMQSAR [16]. Although, detailed descriptions on how this program computes and utilizes molecular fields are described elsewhere [5–8,16], a short introduction is provided here.

For the aldehydes, generation of a model to explain the enantioselection commences with calculation of the lowest energy substrate conformers. Initially, conformers of the compounds were constructed and computed using the semi-empirical method PM3 in Spartan [17]. The lowest energy conformers of the substrates were aligned and were then used with the QMQSAR program. The important feature here is that all the substrates are treated and aligned in the same way so that their relationships can be interrogated; similar relationships (i.e. *t*-Bu is larger than Me, methoxy is more electron rich than trifluoromethyl, etc.) would likely hold even if a higher energy ground state accounts for the predominant reaction pathway. The requisite quantum mechanical interaction fields were computed using single-point PM3 semi-empirical calculations with Divcon [18] and were in the form of electrostatic potential field (EPF) values at ordered grid points encompassing the substrates.

### 2.2. Model generation

Grid spacing in the EPF was initially 0.35 Å and was adjusted during the course of the model building to a finer grid around cor-

---

☆ This paper is part of a special issue on Computational Catalysis.
* Corresponding author.
 *E-mail address:* marisa@sas.upenn.edu (M.C. Kozlowski).

**Table 1**
Experimental and calculated substrate enantioselection for the reaction in Eq. (1).
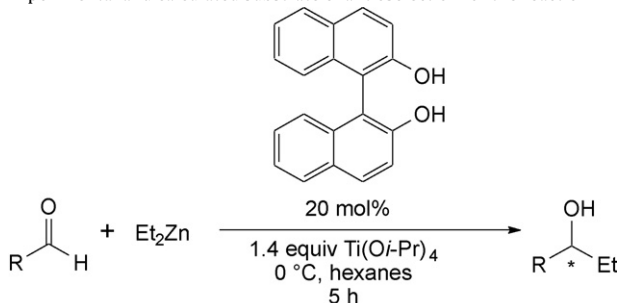


| Entry | RCHO | Calculated | | Experimental | | Entry | RCHO | Calculated | | Experimental | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Delta G_{ee}$[a,b] (kcal/mol) | ee (%) | $\Delta G_{ee}$[a,c] (kcal/mol) | ee (%)[d] | | | $\Delta G_{ee}$[a,b] (kcal/mol) | ee (%) | $\Delta G_{ee}$[a,c] (kcal/mol) | ee (%)[d] |
| 1 | Ph | 1.49 | 88 | 1.72 | 92 | 7 | p-MePh | 1.64 | 91 | 1.49 | 88 |
| 2 | β-Nap | 1.39 | 86 | 1.22 | 81 | 8 | p-MeOPh | 1.56 | 89 | 1.16 | 79 |
| 3 | m-MeOPh | 2.09 | 96 | 1.88 | 94 | 9 | o-ClPh | 0.71 | 57 | 0.92 | 69 |
| 4 | m-ClPh | 1.45 | 87 | 1.49 | 88 | 10 | o-BrPh | 0.96 | 71 | 0.73 | 59 |
| 5 | m-O₂NPh | 1.24 | 82 | 0.94 | 70 | 11 | p-FPh | 1.67 | 91 | 1.40 | 86 |
| 6 | p-ClPh | 1.37 | 85 | 1.49 | 88 | | | | | | |

[a] $\Delta G_{ee}$ refers to the energy difference between the transition structures of the pathways leading to the S and R enantiomeric products.
[b] Calculated values from a leave-one-out model constructed from the 10 remaining substrates.
[c] Values obtained from ee results by using $\Delta G = -RT \ln[(S)/(R)]$.
[d] From Ref. [21].

related EPF points according to a MAXMIN diversity algorithm [19]. The EPF values represent the pool of independent variables from which the multi-linear regression (MLR) models were built. The MLR models between the EPF points and the $\Delta G_{ee}$ values of the substrates were optimized by a simulated Monte Carlo approach [16]. Initial models were constructed using all substrates and from 1 to 4 EPF points generating expressions according to the following equation: $\Delta G_{calc} = a + b \times EPF_1 + c \times EPF_2 + d \times EPF_3 + e \times EPF_4$ where $a$–$e$ are constants/coefficients calculated by the program and where $EPF_{1-4}$ are the values of the EPF at gridpoints 1–4 selected by the program. Models were evaluated by their goodness-of-fit ($r^2$) and the standard deviation (SD) with respect to the experimental data. Models were further refined using a leave-one-out [20] analysis, where n models containing all the combinations of $n-1$ substrates were constructed; for each model the substrate absent in the parameterization set was then calculated giving rise to a predicted enantioselection. The goodness-of-fit of these leave-one-out (LOO) cross-validated predictions is summarized in the term $r^2_{LOO}$.

### 2.3. Experimental data

All experimental data was obtained from Refs. [21,22]. Substrates were compared only for reactions conducted in the same

solvent, at the same temperature, and with the same catalyst and catalyst loading. For all the analyses, the enantioselectivities are converted to $\Delta G_{ee}$ using the relationship: $\Delta G_{ee} = -RT \ln[(S)/(R)]$ so that the variables used in the correlation possess an underlying linear relationship.

### 3. Results and discussion

In this study, instead of correlating catalyst structure to selectivity, we correlate the substrate structure to selectivity for reactions with the same catalyst. A prerequisite for this process is data for $\geq 10$ substrates that encompass a wide enantioselectivity range. One candidate is outlined in Eq. (1) [21]. These calculations are even simpler than those described for the chiral catalysts [5–8], since only the structures of the substrates (no metals, etc.) need to be calculated. In this case, after the rapid calculation of the ground state structures, the substrates were aligned about the aryl ring of the aldehyde and the EFP fields were calculated quickly (seconds to minutes). Models with 1–4 EPF points per substrate were then constructed and evaluated. Two EPF points were sufficient to generate models with high $r^2$ and SD values. Furthermore, a leave-one-out analysis [20] indicated that the models were highly predictive for the 11 substrates listed in Table 1 with a $SD_{LOO}$ of
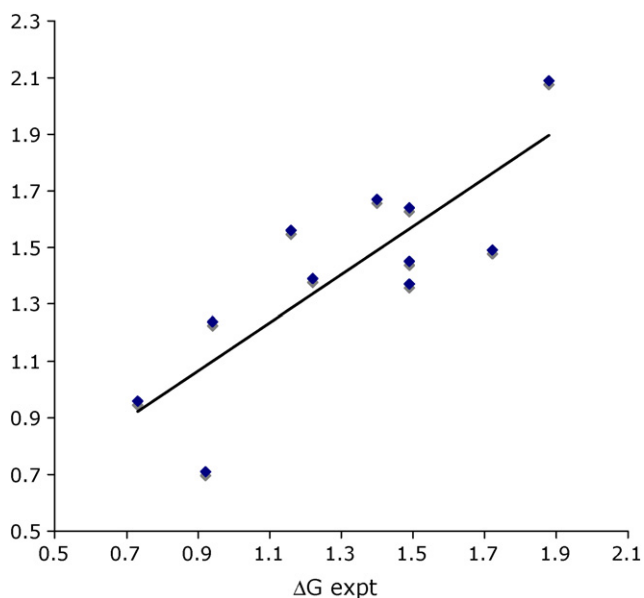
**Fig. 1.** Cross-validation results of substrates from entries 1 to 11 (Table 1): plot of predictions from leave-one-out models constructed from the 10 remaining substrates ($y = 0.85x + 0.30$, $r^2_{LOO} = 0.67$, CC = 0.82).



**Fig. 2.** Superposition of the EPF points from the 11 leave-one-out models around the 11 aldehyde substrates listed in Table 1 (blue = positive EFP value correlates to higher ee, red = positive EFP value correlates to lower ee).

0.35 kcal/mol.The plot of experimental and calculated $\Delta G_{ee}$ values in Fig. 1 confirms the ability of the leave-one-out models to predict the selectivity of *substrates that were not used in construction of the actual model*. Notably, a good cross-validation correlation was observed with $r^2_{LOO} = 0.67$ and a correlation coefficient (CC) of 0.82 (CC offers a rough gauge of the ability of the program to correctly rank the selectivities of different substrates). This tool would be invaluable in the many instances when asymmetric transformations of novel substrates through a well-known process (i.e. asymmetric hydrogenation) are required. With the wealth of data available for a broad range of substrates and catalysts, it would be straightforward to construct models for each catalyst and then screen the novel substrates in silico in a matter of minutes. Such a process would allow the chemist to rapidly refine selection of catalysts for initial screening.

In addition, the EPF points identified in the models (Fig. 2) indicate that two regions of the substrates account for most of the variance in enantiomeric excess; groups of different sizes or electronic aspects at these positions modify the enantioselection. In line with our general understanding of this transformation, the ESP points are primarily near the aldehyde group, probably reflecting electronic differences in the aldehyde carbonyl which coordinates the catalyst, or the point of variation on the aromatic ring, most likely reflecting steric biases between these sites and the catalyst.

With this success in hand, a second more challenging case with a broader range of substrates (Eq. (2), Table 2, R = alkyl and aryl) was analyzed [22]. After aligning the substrates using the CHO atoms of the aldehyde, these structures were analyzed with the QMQSAR program as described above.
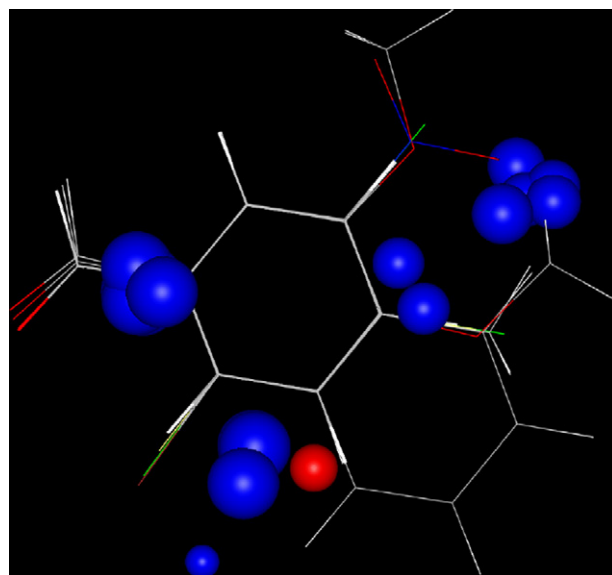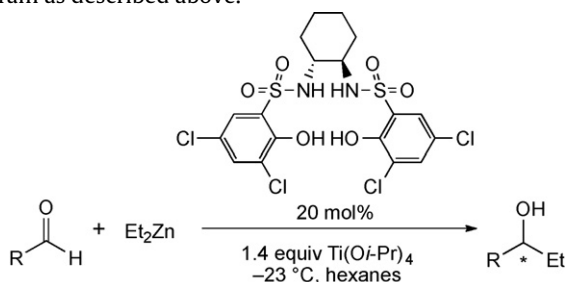
Interestingly, one model does not account for all of the results. In undertaking the model construction, models with 1–4 EPF points per substrate were evaluated again. Fitted models employing two points initially appeared promising and took on the form: $\Delta G_{calc} = 2.35 + 0.55 \times EPF_1 - 0.15 \times EPF_2$ (SD = 0.43, $r^2 = 0.68$). However, the high standard deviation was troublesome. Furthermore, despite a thorough evaluation of the parameters, one model could not be identified that yielded highly cross-validated results in a leave-one-out analysis [20].

To locate the source of this problem, partial substrate sets were employed. When two sets of models were generated, two-point models with much lower standard deviations could be identified. For example, with entries 11–20 of Table 2 the model displayed a standard deviation of 0.09 kcal/mol and an $r^2$ value of 0.98. In addition, satisfactorily cross-validated models were obtained.
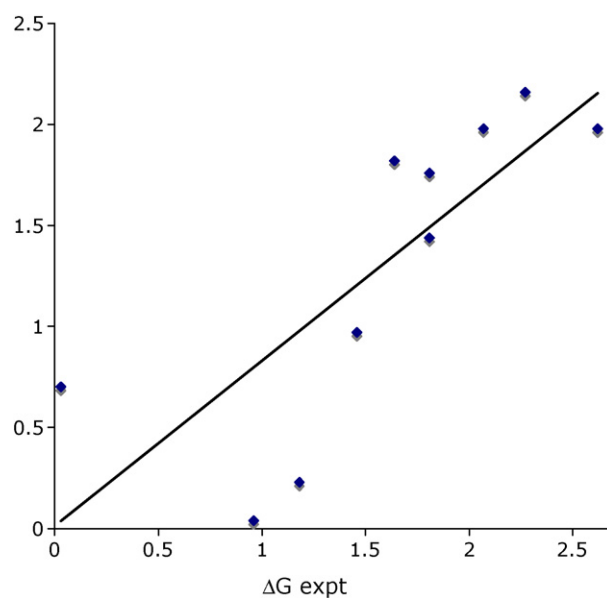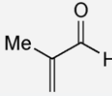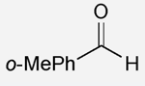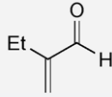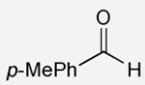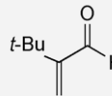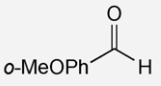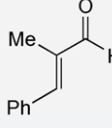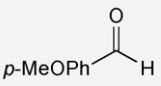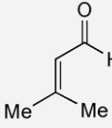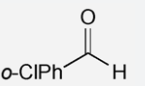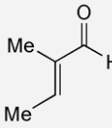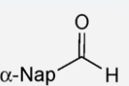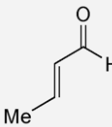


**Fig. 3.** Cross-validation results of substrates from entries 1 to 10 (Table 2): plot of predictions from leave-one-out models constructed from the 9 remaining substrates ($y = 0.82x + 0.01$, $r^2_{LOO} = 0.61$, $CC_{LOO} = 0.77$).

**Table 2**
Experimental and calculated substrate enantioselection for the reaction in Eq. (2).

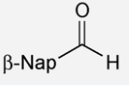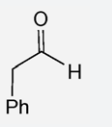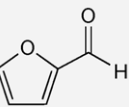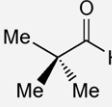| Entry | RCHO | Calculated | | Experimental | | Entry | RCHO | Calculated | | Experimental | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Delta G_{ee}$ [a,b] (kcal/mol) | ee (%) | $\Delta G_{ee}$ [a,c] (kcal/mol) | ee (%) [d] | | | $\Delta G_{ee}$ [a,e] (kcal/mol) | ee (%) | $\Delta G_{ee}$ [a,c] (kcal/mol) | ee (%) [d] |
| 1 | Ph—CHO | 1.98 | 96 | 2.62 | 99 | 11 | Me (methacrolein) | 1.63 | 93 | 1.81 | 95 |
| 2 | o-MePh—CHO | 1.98 | 96 | 2.07 | 97 | 12 | Et | 1.69 | 94 | 1.72 | 94 |
| 3 | p-MePh—CHO | 1.82 | 95 | 1.64 | 93 | 13 | t-Bu | 2.72 | 99 | 2.07 | 97 |
| 4 | o-MeOPh—CHO | 0.04 | 4 | 0.96 | 75 | 14 | Me / Ph | 1.80 | 95 | 1.93 | 96 |
| 5 | p-MeOPh—CHO | 1.44 | 90 | 1.81 | 95 | 15 | Me Me | 0.03 | 3 | 0.02 | 2 |
| 6 | o-ClPh—CHO | 0.23 | 23 | 1.18 | 83 | 16 | Me Me | 1.00 | 76 | 0.64 | 57 |
| 7 | α-Nap—CHO | 2.16 | 97 | 2.27 | 98 | 17 | Me | 0.10 | 10 | 0.22 | 22 |
| 9 | β-Nap—CHO | 1.76 | 94 | 1.81 | 95 | 19 | Ph—CH2—CHO | 0.69 | 60 | 0.56 | 51 |
| 10 | furan-2-CHO | 0.97 | 75 | 1.46 | 90 | 20 | Me Me Me | 2.19 | 97 | 0.56 | 51 |

[a] $\Delta G_{ee}$ refers to the energy difference between the transition structures of the pathways leading to the S and R enantiomeric products.
[b] Calculated values from a leave-one-out model constructed from the 9 remaining substrates in entries 1–10.
[c] Values obtained from ee results by using $\Delta G = -RT \ln[(S)/(R)]$.
[d] From Ref. [22].
[e] Calculated values from a leave-one-out model constructed from the 9 remaining substrates in entries 11–20.

For entries 1–10, one model was obtained with $r^2_{LOO} = 0.61$ and $CC_{LOO} = 0.77$ (Fig. 3). For entries 11–20, a second model was obtained with $r^2_{LOO} = 0.61$ and $CC_{LOO} = 0.78$ (Fig. 4). The fact that no single model could accommodate all the results in spite of the large number of potential EFP variables explored and the success of this method with highly different substrates and catalysts in the past [5–15] suggests that there is no underlying relationship between all the substrates in this instance. One of the assumptions in constructing these type of QSSR models is that each substrate (or catalyst) interacts with its corresponding catalyst (or substrate) in

the same way. The presence of two distinct models here may indicate different mechanistic regimes. For example, interaction of the substrate and catalyst may not be constant. The aryls of the catalyst may undergo a π-stacking interaction with the aryl groups in the first group of substrates in Table 2 resulting in a different set of low energy transition states compared to those from the in the second group of substrates. Similar breaks are observed with other linear free energy relationships [23]. As a consequence, this MLR QSSR method can probe potential mechanism changes in silico and indicate when further experimental study is warranted.
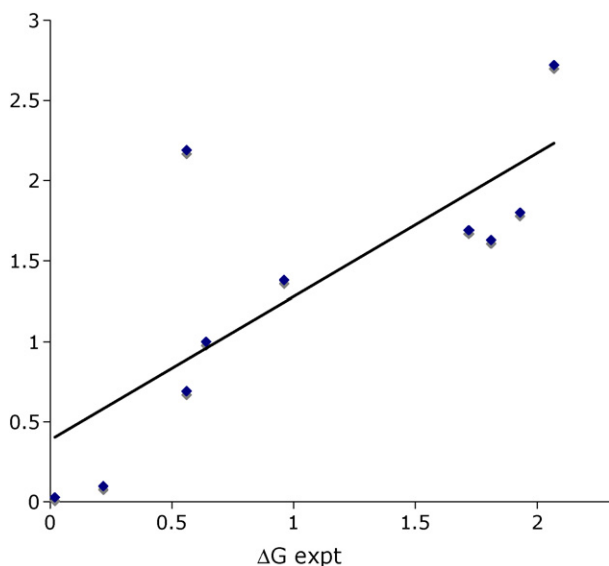
**Fig. 4.** Cross-validation results of substrates from entries 11 to 20 (Table 2): plot of predictions from leave-one-out models constructed from the 9 remaining substrates ($y = 0.89x + 0.38$, $r^2_{LOO} = 0.61$, $CC_{LOO} = 0.78$).

## 4. Conclusion

The ability to predict the enantioselectivity of novel substrates in asymmetric transformations as demonstrated above with QMQSAR, would be invaluable in the many instances when a well-known process (i.e. asymmetric hydrogenation) needs to be applied. Such a process would allow the chemist to rapidly refine selection of catalysts to optimize productivity in reaction screening. Furthermore, the output of the QMQSAR program provides direct information to the user as to which portions of the substrates are most relevant to enantioselection. In addition to predicting the selectivity of new substrates, the QSSR method described herein may be a simple, rapid means of assessing mechanism shifts in many experimental systems. Further mechanism experiments to investigate this proposed mechanism shift will be reported in due course.

## References

[1] K.N. Houk, P.H.Y. Cheong, Nature 455 (2008) 309.
[2] J.M. Brown, R.J. Deeth, Angew. Chem. Int. Ed. 48 (2009) 4476.
[3] For the first QSSR work in a selective reaction, see: J.D. Oslob, B. Åkermark, P. Helquist, P.-O. Norrby, Organometallics 16 (1997) 3015.
[4] M.S. Sigman, J.J. Miller, J. Org. Chem. 74 (2009) 7633.
[5] M.C. Kozlowski, S. Dixon, M. Panda, G. Lauri, J. Am. Chem. Soc. 125 (2003) 6614.
[6] P.-W. Phuan, J.C. Ianni, M.C. Kozlowski, J. Am. Chem. Soc. 126 (2004) 15473.
[7] J.C. Ianni, V. Annamalai, P.-W. Phuan, M.C. Kozlowski, Angew. Chem., Int. Ed. 45 (2006) 5502.
[8] J. Huang, J.C. Ianni, J.E. Antoline, R.P. Hsung, M.C. Kozlowski, Org. Lett. 8 (2006) 1565.
[9] (a) K.B. Lipkowitz, M. Pradhan, J. Org. Chem. 68 (2003) 4648;
(b) S. Alvarez, S. Schefzick, K. Lipkowitz, D. Avnir, Chem.-Eur. J. 9 (2003) 5832.
[10] M. Hoogenraad, G.M. Klaus, N. Elders, S.M. Hooijschuur, B. McKay, A.A. Smith, E.W.P. Damen, Tetrahedron: Asymmetry 15 (2004) 519.
[11] J.L. Melville, K.R.J. Lovelock, C. Wilson, B. Allbutt, E.K. Burke, B. Lygo, J.D. Hirst, J. Chem. Inf. Model. 45 (2005) 971.
[12] S. Sciabola, A. Alex, P.D. Higginson, J.C. Mitchell, M.J. Snowden, I. Morao, J. Org. Chem. 70 (2005) 9025.
[13] M. Urbano-Cuadrado, J.J. Carbó, A.G. Maldonado, C. Bo, J. Chem. Inf. Model. 47 (2007) 2228.
[14] J.B. Van der Linden, E.-J. Ras, S.M. Hooijschuur, G.M. Klaus, N.T. Luchters, P. Dani, G. Verspui, A.A. Smith, E.W.P. Damen, B. McKay, M. Hoogenraad, QSAR Comb. Sci. 24 (2005) 94.
[15] J. Chen, W. Jiwu, L. Mingzong, T. You, J. Mol. Catal. A 258 (2006) 191.
[16] S. Dixon Jr., K.M. Merz, G. Lauri, J.C. Ianni, J. Comput. Chem. 26 (2005) 23.
[17] Spartan'02; Wavefunction Inc., Irvine, CA, 2002.
[18] S.L. Dixon, K.M. Merz, J. Chem. Phys. 107 (1997) 879.
[19] S. Kirkpatrick Jr., C.D. Gelatt, M.P. Vecchi, Science 220 (1983) 671.
[20] P. Gramatica, QSAR Comb. Sci. 26 (2007) 694.
[21] (a) F.-Y. Zhang, C.-W. Yip, R. Cao, A.S.C. Chan, Tetrahedron: Asymmetry 8 (1997) 585;
(b) F.-Y. Zhang, A.S.C. Chan, Tetrahedron: Asymmetry 8 (1997) 3651.
[22] (a) X. Zhang, C. Guo, Tetrahedron Lett. 36 (1995) 4947;
(b) J. Qiu, C. Guo, X. Zhang, J. Org. Chem. 62 (1997) 2665.
[23] (a) E.V. Anslyn, D.A. Dougherty, Modern Physical Organic Chemistry, University Science Books, Mill Valley, CA, 2006;
(b) C. Hansch, A. Leo, R.W. Taft, Chem. Rev. 91 (1991) 165;
(c) J.P. Richards, W.P. Jencks, J. Am. Chem. Soc. 104 (1982) 4689.