QUANTITATIVE STRUCTURE–ENANTIOSELECTIVITY RELATIONSHIPS USING NEURAL NETWORKS. BIOCONVERSION OF CARBONYL COMPOUNDS USING BAKER'S YEAST

DRISS ZAKARYA

Faculté des Sciences et Techniques, B.P. 146, Mohammedia, Morocco

LAHBIB FARHAOUI

Faculté des Sciences et Techniques, B.P. 146, Mohammedia and Laboratoire de Chimie des Plantes et de Synthèse Organique et Bioorganique, Faculté des Sciences, Rabat, Morocco

AND

SOUÂD FKIH-TÉTOUANI

Laboratoire de Chimie des Plantes et de Synthèse Organique et Bioorganique, Faculté des Sciences, Rabat, Morocco

Quantitative structure–enantioselectivity relationships were studied for the reduction of a set of 73 carbonyl compounds with baker's yeast. The established model, using a neural network, allowed the prediction of the reduction selectivity ($\% S$ enantiomer) with success. The correlation coefficient between the observed and calculated $\% S$ was 0·99. The model was also used to predict the enantioselectivity of the reduction of α-diketones using baker's yeast and different microorganisms.

INTRODUCTION

Baker's yeast is cheap and easily available and is often used as reagent in the conversion of carbonyl compounds to the corresponding alcohols. The reduction with baker's yeast does not always afford alcohols with the desired configuration in satisfactory enantiomeric excess. For example, the reduction of ethyl 3-oxobutanoate and ethyl 3-oxopentanoate under the same conditions leads to very different results: the former gives a good enantiomeric excess (70–97%)[1] whereas the latter gives unsatisfactory results (40% enantiomeric excess).[2] Hence methods for controlling the stereochemical course in yeast reduction are required. It seemed of interest to elaborate structure-enantioselectivity relationships.

In previous work,[3] we studied the effect of structural parameters on enantioselectivity control in the reduction of 41 carbonyl compounds by baker's yeast under relatively similar conditions (e.g. amounts of yeast and glucose, reaction time). In the present work, we attempted to extend the approach to a more important set of carbonyls, in order to establish quantitative structure–enantioselectivity relationships. The best model obtained was tested in the prediction of the stereochemistry of alcohols resulting from bioconversion of carbonyls.

---

* Author for correspondence.

EXPERIMENTAL

Data set and structure coding

A set of 73 carbonyl compounds was taken from different literature sources (Table 1). Bioconversions of these compounds were carried out under the following conditions: no cofactors such as haloesters,[32] allylic alcohols[33] α, β-ethylenic ketones[34] were added; solvent: water; 10–20% glucose in the solvent; amounts of substrate 0·1–0·5 mmol/g of baker's yeast; temperature: *ca* room temperature; reaction time: 12–72 h. Only the structural parameters were taken into account in the enantioselectivity modelling.

As all the compounds studied have a common $C\!=\!O$ group, the description of the molecules was simplified, and each was described by a set of variables coding radicals linked to the $C\!=\!O$ group. In order to obtain a homogeneous description for all the molecules studied, R_2 was considered as the group which had priority according to the Cahn–Ingold–Prelog rules (CIP). In addition to this hypothesis, substituents were assigned both R_1 and R_2 for each molecule. The generated set of molecules was treated.

In order to take into account all aspects of the molecule, parameters related to steric effects (V_i, Van der Waals volume of R_i)[35] and lipophilicity (Lip$_i$, estimated according to Rekker's method[36]) were selected.

Table 1. Compounds studied

No.	Structure[a]	No.	Structure[a]
1	$CH_3COCH_2COCH_3$ (99·5; 99·5)[4]	38	$CH_3OCOCH_2COCH_2Cl$ (77·5; 85)[20]
2	$CH_3COCH_2COCH_2CH_3$ (99·5; 99·9)[4]	39	$C_2H_5OCOCH_2COCH_2Cl$ (75; 77·9)[20]
3	$CH_3COCH_2CO(CH_2)_3CH=CH_2$ (99·5; 99)[4]	40	$n\text{-}C_3H_7OCOCH_2COCH_2Cl$ (60; 99·8)[20]
4	$CH_3COCH(CH_3)COCH_3$ (97·5; 99·3)[4]	41	$n\text{-}C_4H_9OCOCH_2COCH_2Cl$ (60; 45·6)[20]
5	$CH_3COCH(CH_3)CN$ (91; 94·6)[5]	42	$n\text{-}C_5H_{11}OCOCH_2COCH_2Cl$ (5; 19)[20]
6	$CH_3COCH(C_2H_5)CN$ (99; 99·8)[5]	43	$n\text{-}C_6H_{13}OCOCH_2COCH_2Cl$ (2·5; 05)[20]
7	$CH_3COCH(C_4H_9)CN$ (99; 99·9)[5]	44	$n\text{-}C_7H_{15}OCOCH_2COCH_2Cl$ (1·5; 1·5)[20]
8	$CH_3COCH(C_6H_5)CN$ (99; 99·9)[5]	45	$n\text{-}C_8H_{17}OCOCH_2COCH_2Cl$ (1; 0·7)[20]
9	$CH_3COCH_2COOC_2H_5$ (97·5; 99·9)[6]	46	$CH_3COCH_2CSSCH_3$ (98; 99·5)[21]
10	$CH_3COCH_2CH_2CN$ (99; 96·4)[7]	47	$CH_3COCH_2CSSC_2H_5$ (98; 99·9)[21]
11	$CH_3CH_2COCH_2CH_2CN$ (73; 71·7)[7]	48	$CH_3COCH(CH_3)CSSCH_3$ (98; 99·9)[21]
12	$CH_3COCH_2CH_2CN$ (98·5; 99·8)[7]	49	$CH_3COCH_2COSC_2H_5$ (98; 99·9)[21]
13	$C_5H_{11}COCH_2CH_2CN$ (14; 15·9)[7]	50	$CH_3COCH(CH_3)COSC_2H_5$ (98; 99·9)[21]
14	$NCCH_2CH_2COC_6H_5$ (92·5; 97·8)[7]	51	$N_3CH_2COCH_2OCOCH_3$ (87·9; 87·3)[22]
15	CH_3COCH_2COOH (93; 98)[8]	52	$N_3CH_2COCH_2OCOC_6H_5$ (89; 88)[22]
16	$C_2H_5COCH_2COOH$ (1; 14)[8]	53	$N_3CH_2COCH_2OCOC(CH_3)_3$ (95; 96·9)[22]
17	$n\text{-}C_3H_7COCH_2COOH$ (1; 2)[8]	54	$CH_3CO(CH_2)_3NO_2$ (99; 99·8)[23]
18	$n\text{-}C_5H_{11}COCH_2COOH$ (1; 0·8)[8]	55	$C_2H_5COCH_2CH_2CH_3$ (83·5; 99·8)[23]
19	$n\text{-}C_8H_{17}COCH_2COOH$ (1; 0·5)[8]	56	$C_2H_5OCOCH_2COCH_2Br$ (99; 95)[24]
20	$p\text{-}(CO_2H)C_6H_4COCF_3$ (10; 11·7)[9]	57	$n\text{-}C_4H_9COCH_2COOH$ (0; 1·2)[8]
21	$p\text{-}(CH_3OOC)C_6H_4COCF_3$ (10; 7·3)[9]	58	$n\text{-}C_3H_7COCH_2CO_2C_2H_5$ (98; 99·8)[25]
22	$CH_3COCH(Cl)COOC_2H_5$ (98; 99·9)[10]	59	$C_6H_5CH_2OCH_2CH_2COCH_2CO_2\text{-}i\text{-}C_5H_{11}$ (98; 92·7)[26]
23	$C_6H_5COCH_2OH$ (4; 0·4)[11]	60	$C_6H_5CH_2O(CH_2)_2COCH_2CO_2C_6H_{13}$ (98; 96·2)[26]
24	$C_6H_5COCH_2OCOC_2H_5$ (97; 95·8)[11]	61	CH_3COCH_2OH (04·5; 12·6)[27]
25	$CH_3COCH_2OCH_2C_6H_5$ (95; 99·9)[11]	62	$CH_3CH_2CH_2COCH_2OH$ (0; 0·8)[27]
26	$C_2H_5OCOCH_2COCF_3$ (27·5; 21·2)[12]	63	$C_4H_9COCH_2OH$ (0; 0·3)[27]
27	$CH_3CH_2COCH_2COOCH_3$ (94·5; 92·6)[13]	64	$CH_3COCOOC_2H_5$ (95·5; 99)[18]
28	$CH_3CH_2COCH_2COO\text{-}n\text{-}C_8H_{17}$ (99; 100)[14]	65	$CH_3CH_2COCOOC_2H_5$ (87·5; 92·5)[18]
29	$C_2H_5COOCH_2COC_6H_5$ (100; 99·8)[15]	66	$C_6H_5COCOOCH_3$ (0; 0·8)[18]
30	$CH_3COCH_2CH_2CH=C(CH_3)_2$ (97; 100)[16]	67	$C_6H_5COCONH_2$ (0; 0·4)[15]
31	$CH_3COC_6H_5$ (84·5; 99·9)[17]	68	2-Furyl-$COCOOCH_3$ (4; 2·3)[28]
32	$p\text{-}(CH_3CO)C_6H_4I$ (98; 100)[17]	69	CH_3COCH^b (98; 99·9)[29]
33	$n\text{-}C_4H_9COCOOC_2H_5$ (75; 72·9)[18]	70	$C_3H_7COCH^b$ (98; 99·6)[29]
34	$n\text{-}C_5H_{11}COCOOC_2H_5$ (65; 66·7)[18]	71	$CH_3OCH_2COCH^b$ (97·5; 99·9)[30]
35	$CH_3COCH_2COO(CH_2)_3CH_3$ (98; 99·9)[19]	72	$CH_3COCH_2SOOC_6H_5$ (97·5; 99·9)[31]
36	$CH_3COCH_2COOCH_3$ (93; 98·7)[20]	73	$CH_3COCH_2SCH_3$ (97·5; 99·6)[31]
37	$CH_3OCOCH_2COCF_3$ (26; 29·5)[20]		

[a] In parentheses: $\%S_{obs.}$ and $\%S_{calc.}$ using neural network [equation (7)].
[b] Aldehyde protected by $HS(CH_2)_3SH$.

Structure–enantioselectivity relationships were established using regression analysis and the neural network approach.

Neural network

The neural network[37,38] used has a configuration with three layers and complete connections between neurons (Figure 1). The input layer (layer I) is constituted by the descriptors used, the hidden layer (layer H) has five neurons (determined after several trials) and the output layer (layer O) is constituted by one neuron and represents the calculated $\%S$ (indicating the enantioselectivity).

Estimation of the weights was made using the back propagation algorithm. Each neuron of a given layer (except the input one) takes a value Y calculated using the following transfer function.

$$\%S_{calc} = [1 + \exp(-bZ_o + c)]^{-1} \qquad Z_o = \sum W_{ko} Y_k \quad (1)$$

Y_j = value of neuron j in the input layer I; Y_k = value of neuron k in the hidden layer H; Y_o = value of the output neuron O; W_{jk} = weight of the connection between neuron j in layer I and neuron k in layer H; W_{ko} = weight of the connection between neuron k in layer H and the output neuron O.

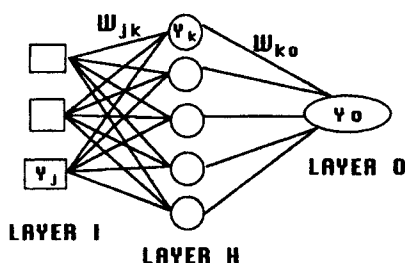After several attempts, b was taken equal to 1 and c equal to 0. The network was trained and weights

Figure 1. Configuration of neural network used. Y_j = value of neuron j in the input layer I; Y_k = value of neuron k in the hidden layer H; Y_o = value of the output neuron O; W_{jk} = weight of the connection between neuron j in layer I and neuron k in layer H; W_{ko} = weight of the connection between neuron k in layer H and the output neuron O

modified until minimizing the cost C defined as:

$$C = (1/n) \sum_1^n (\%S_{calc} - \%S_{obs})^2;$$

$$n = \text{number of compounds} \quad (2)$$

Contribution of each descriptor was calculated using the following relation:[39]

$$Cont\%(j) = 100 C_j^{av} \Big/ \sum C_j^{av} \quad (3)$$

$$C_j^{av} = \sum_k |W_{jk}| Y_j^{av} \quad (4)$$

$$Y_j^{av} = (1/n) \sum Y_j \quad (5)$$

RESULTS AND DISCUSSION

The two methods of description were tested and that based on the CIP rules was found the most adequate.

Regression analysis

Data set subjected to regression analysis, in order to establish a linear model between the percentage of the S enantiomer and the descriptors used, leads to the following model.

$$(\%S)_{obs.} = (-20\cdot21 \pm 2\cdot45)V_1/V_2$$
$$+ (14\cdot84 \pm 2\cdot70)Lip_2 + 84\cdot65 \quad (6)$$
$$n = 73; \qquad r = 0\cdot82; \qquad s = 23\cdot76;$$

V_1 and V_2 are the Van der Waals volumes of R_1 and R_2 respectively. Lip_2 is the estimated contribution of R_2 to log P. Correlation coefficient between V_1/V_2 and Lip_2 is equal to -0.30.

The Coefficients associated with descriptors were statistically significant at more than 99%. The calculated contribution for descriptors (V_1/V_2 and Lip_2) were 64% and 36% respectively.

Table 2. Weight matrix for the connections between neurons[a]

	H_1	H_2	H_3	H_4	H_5
I_1	$-0\cdot19$	$-3\cdot40$	$0\cdot65$	$0\cdot98$	$-2\cdot22$
I_2	$-1\cdot38$	$1\cdot17$	$-1\cdot05$	$4\cdot08$	$0\cdot27$
I_3	$-2\cdot96$	$1\cdot81$	$-3\cdot37$	$-1\cdot69$	$5\cdot05$
O	$5\cdot62$	$5\cdot71$	$5\cdot85$	$6\cdot44$	$-5\cdot84$

[a]I_1, I_2 and I_3: input neurons representing descriptors (Lip_1, Lip_2 and V_1/V_2 respectively); H_1-H_5, hidden neurons; O, output neuron. Number of cycles = 6000.
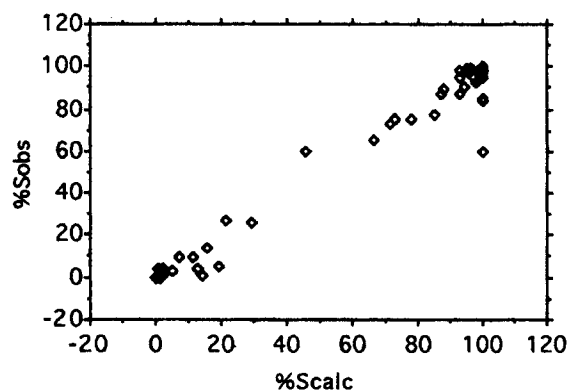

Figure 2. Correlation between $\%S_{obs.}$ and $\%S_{calc.}$ using neural network (equation 7)

Considering the correlation coefficient, equation (6) is not of good quality, then regression analysis is not appropriate for enantioselectivity modelling for the studied sample. This is due to the nonlinearity of the phenomenon vs the structural features of R_1 and R_2 groups. To overcome this problem, we used the neural network approach recently known as a promising technique. Lip_1, Lip_2 and V_1/V_2 were considered as inputs. The correlation coefficient between Lip_1 and V_1/V_2 is 0·41 and that between Lip_1 and Lip_2 is -0.13.

Neural network (NN)

Data were subjected as a training set to the NN having a 3−5−1 configuration in order to determine the weights of the connections between all the neurons (Table 2). The best connection's weights are summarized in Table 2.

As shown in Table 1(or Figure 2), the calculated $\%S$ using the neural network is generally in good agreement with that observed.

The correlation coefficient between observed and calculated $\%S$ is 0·99:

$$\%S_{obs.} = (0\cdot98 \pm 0\cdot02)\%S_{calc.} - 0\cdot70$$
$$n = 73; \qquad r = 0\cdot99; \qquad s = 6\cdot29 \quad (7)$$

Table 3. Correlation coefficient and contribution of each variable for all trials

	No. of	Contribution (%)[39]			
Trial	of compounds	r^a	V_1/V_2	Lip_1	Lip_2
1	65	0·988	64	25	11
2	65	0·982	55	27	18
3	65	0·984	44	34	22
4	65	0·980	50	38	12
5	65	0·994	48	34	18
6	65	0·982	43	39	18
7	6S	0·985	47	38	15
8	65	0·983	67	30	13
9	64	0·984	60	28	12

[a] Correlation coefficient.

Figure 3. Correlation between $\%S_{obs.}$ and $\%S_{calc.}$ using neural network (equation 8)

The quality of equation (7) showed that the NN approach is more efficient than regression analysis in the enantioselectivity modelling. Any clouds of points corresponding to particular experimental parts were observed.

Prediction ability of the model

To test the prediction ability of the model, we elaborated a series of new models on the basis of 65 compounds only. Each time, the remaining compounds were used as a test. The compounds tested were chosen with respect to the composition of the basic sample (e.g. two compounds leading to R mainly and six leading to S mainly).

The procedure was repeated nine different times in order to predict the enantioselectivity for all compounds of the basic training set (see Table 3).

The correlation coefficient between the calculated (using the models established on the basis of sets of 65 compounds) and observed $\%S$ is relatively similar than that observed in equation (7) [equation (8), Figure 3]

$$\%S_{obs.} = (0·94 \pm 0·03)\%S_{calc.} + 0·89$$
$$n = 73; \qquad r = 0·97; \qquad s = 10·67 \tag{8}$$

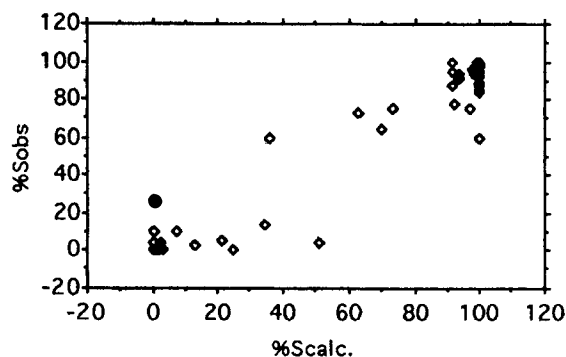The correlation coefficients observed for all trials (Table 3) showed that the models established were in

good agreement with the global model. Descriptors were classified, according to their contributions for all trials, as follow: $V_1/V_2 > Lip_1 > Lip_2$. All these results revealed good stability of the studied structure-enantioselectivity relationships, and confirm the fact that enantioselectivity depends essentially on the structural features of the molecule.

Prediction of the enantioselectivity of the bioconversion of other type of compound

Recently, Bel Rhlid[40] achieved the bioconversion of α-diketones using different microorganisms, including baker's yeast (B.Y.), such as *Beauveria sulfurescens* (B.S.), *Geotrichum candidum* (G.C.), *Rhodotorula rubra* (R.R.) and *Dipodascopsis uninucleata* (D.U.). We attempted to predict the enantioselectivity of the bioconversions that he carried out using the NN model. Table 4 shows a comparison between the observed configuration and predicted $\%S$. All the predicted results were in good agreement with the observed values except for compound **78**, which was predicted to be near a racemic mixture.

In conclusion, we have shown that the enantioselectivity of the bioconversion of carbonyl compounds by baker's yeast can be quantitatively estimated using a neural network model. Reaction conditions are certainly

Table 4. Observed configuration and predicted $\%S$ for C=O group bioconversion

Compound	B.S.	B.Y.	G.C.	R.R.	D.U.	$\%S^a$
74: $H_3CCOCOCH_3$	2S	2S	2S	2S	2S	78·5 (S)
75: $H_3CCOCOC_2H_5$	2S	2S	—	—	3S	93·1 (S)
76: $H_3CCOCOC_5H_{11}$	2S	2S	2S	2S	3S	99·9 (S)
77: $H_3CCOCOC_6H_5$	2S	2S	2S	2S	—	99·9 (S)
78: $H_5C_2COCOC_2H_5$	3S	3S	3S	—	—	52·2 (S)
79: $H_5C_2COCO(CH_2)_3CH=CH_2$	3S	3S	3S	—	—	99·4 (S)

[a] $\%S$ calculated using the NN model (Table 2).

important features; however, we demonstrated that the enantioselectivity was essentially due to the structural features of the molecule. The obtained model was tested with success in the prediction of the enantioselectivity for the studied compounds and also for other type of molecules.

REFERENCES

1. E. Hungerbuheler, D. Seebach and D. Wasmuth, *Helv. Chim. Acta* **64**, 1467–1487 (1981).
2. G. Frater, *Helv. Chim. Acta* **62**, 2829–2832 (1979).
3. D. Zakarya, L. Farhaoui and S. Fkih-Tétouani, *Tetrahedron Lett.* **35**, 4985–4988 (1994).
4. J. Colte, G. J. Gouray and H. Veschambre, *Tetrahedron Lett.* **27**, 565–568 (1986).
5. T. Itoh, T. Fukuda and T. Fujisawa, *Bull. Chem. Soc. Jpn.* **62**, 3851–3855 (1989).
6. Y. Naoshima, J. Maeda, Y. Munkata, T. Nishiyama, M. Kamezawa and H. Tachibana, *J. Chem. Soc., Chem. Commun.* 964–965 (1990).
7. A. Gopalan, R. Lacero, H. Jacobs and K. Bevyama, *Synth. Commun.* **21**, 1321–1329 (1991).
8. M. Utaka, H. Watabu, H. Higashi, T. Sakai, S. Tsuboi and S. Torii, *J. Org. Chem.* **55**, 3917–3921 (1990).
9. T. Fujisawa, H. Ichikawa and M. Shimizu, *Tetrahedron: Asymmetry* **4**, 1237–1240 (1993).
10. M. Hamdani, B. Jero, H. Deleuze, A. Saux and B. Maillard, *Tetrahedron: Asymmetry* **4**, 1233–1236 (1993).
11. A. Manzocchi, A. Fiecchi and E. Santaniello, *Synthesis* 1007–1009 (1987).
12. D. Seebach, P. Renaud, W. B. Schweizer and M. Zueger, *Helv. Chim. Acta* **62**, 1843–1846 (1984).
13. K. Nakamura, Y. Kawai, S. Oka and A. Ohno, *Tetrahedron Lett.* **30**, 2245–2246 (1989).
14. C. J. Sih, B. Zhou, A. S. Gopalan, W. R. Shieh and F. Van Middlesworth, in *Selectivity. A Goal for Synthetic Efficiency*, edited by W. Bartmann and B. Trost, Hoechst Workshop Conferences, Vol. 4, pp. 251–261. VCH, Weinheim (1983).
15. B. S. Deol, D. D. Ridlay and G. W. Simpson, *Aust. J. Chem.* **29**, 2459–2467 (1976).
16. A. Archelas and R. Furstoss, *Tetrahedron Lett.* **33**, 5241–5242 (1992).
17. K. Nakamura, K. Ushio, S. Oka and A. Ohno, *Tetrahedron Lett.* **25**, 3979–3982 (1984).
18. K. Nakamura, K. Inoue, K. U. Shio and A. Ohno, *J. Org. Chem.* **53**, 2589–2593 (1988).
19. V. Spilioti, D. Papanatjis and N. Ragaoussis, *Tetrahedron Lett.* **31**, 1615–1616 (1990).
20. R. Furstoss, *Actual. Chim.* Jan.–Feb., 6–16 (1990).
21. T. Y. Itoh, Y. Yonekawa, T. Sato and T. Fujisawa, *Tetrahedron Lett.* **27**, 5405–5408 (1986).
22. T. Sato, T. Mizutani, Y. Okumura and T. Fujisawa, *Tetrahedron Lett.* **30**, 3701–3702 (1989).
23. K. Nakamura, Y. Inoue, J. Shibahara and S. Oka, *Tetrahedron Lett.* **29**, 4769–4770 (1988).
24. C. Fuganti, P. Grasselli, P. Casati and M. Carmeno, *Tetrahedron Lett.* **26**, 101–104 (1985).
25. D. R. Crump, *Aust. J. Chem.* **35**, 1945–1948 (1982).
26. M. Hirama, M. Shimizu and M. Iwashita, *J. Chem. Soc., Chem. Commun.* 599–602 (1983).
27. I. P. Guette, N. Spassky and D. Bougerot, *Bull. Soc. Chim. Fr.* 4217–4224 (1972).
28. D. H. R. Barton, B. D. Brown, D. D. Ridley, D. A. Widdowson, A. J. Keys and C. J. Leaner, *J. Chem. Soc., Perkin Trans.* 2069–2072 (1975).
29. T. Fujisawa, E. Kojima, T. Itoh and T. Sato, *Chem. Lett.* 1751–1754 (1985).
30. G. Guanti, L. Banfi and E. Narisano, *Tetrahedron Lett.* **27**, 3547–3550 (1986).
31. S. Servi, *Synth. Commun.* 1–25 (1990).
32. K. Nakamura, Y. Kawal and A. Ohno, *Tetrahedron Lett.* **31**, 267–270 (1990).
33. K. Nakamura, K. Inoue, K. Ushio, S. Oka and A. Ohno, *Chem. Lett.* 679–682 (1987).
34. K. Nakamura, Y. Kawal, S. Oka and A. Ohno, *Bull. Chem. Soc. Jpn.* **62**, 875–879 (1989).
35. A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
36. G. Nys and R. F. Rekker, *Eur. J. Med. Chem.* **9**, 361–375 (1974).
37. D. E. Rumelhart, C. E. Hinton and R. J. Williams, *Nature (London)* **323**, 533–536 (1986).
38. J. Y. De Saint Laumer and M. Chastrette, *Eur. J. Med. Chem.* **26**, 829–833 (1991).
39. M. Chastrette, D. Zakarya and J. F. Peyraud, *Eur. J. Med. Chem.* **29**, 343–348 (1994).
40. R. Bel Rhlid, Thése, Université Blaise Pascal, Clermont Ferrand (1990).