Pergamon

Tetrahedron Letters, Vol. 35, No. 28, pp. 4985-4988, 1994
Elsevier Science Ltd
Printed in Great Britain
0040-4039/94 $7.00+0.00

0040-4039(94)00925-2

Structure-Enantioselectivity Relationships Using Neural Networks for the Reduction of Carbonyl Compounds with Baker's Yeast

D. Zakarya[a*], L. Farhaoui [b] and S. Fkih-Tetouani[b]

a)-Laboratoire de Chimie Organique Appliquée et de Traitement de l'Information Chimique (LCOATIC), Faculté des Sciences et Techniques, B.P. 146, Mohammedia-Morocco. b)- Laboratoire de Chimie des Plantes et de Synthèse Organique et Bioorganique, Faculté des Sciences, Rabat-Morocco. * Member of the international CAIC group.

Abstract : Structure-enantioselectivity relationships were established for the reduction of a set of 41 carbonyl compounds with baker's yeast. The obtained neural network model, allows the classification of the 41 alcohols (obtained by baker's yeast reduction) as R or S with 85% of success. The implicit rules detected by the neural network were given.

Baker's yeast (BY) reduction of carbonyl compounds (e.g. β-keto-esters) is one of the most extensively studied microbial transformation to obtain chiral compounds or blocks[1]. It was shown that several parameters govern enantio and regioselectivity of the BY reduction, among them structural environment of the carbonyl group seemed to be the most important one. Prelog rule[2] is generally applied in predicting the configuration of the alcohol formed. However, it is empirical and has often proved to be not useful in the case of β-keto esters reduction[3]. Sih[4] has shown that the yeast produces several enzymes which are able to distinguish between small and large groups, adjacent to the carbonyl; thus hydrogen is delivered from both the *re-* and *si-* faces , producing both enantiomers at different rates. The difference in bulk between the two groups should then regulate the kinetics of the reaction leading to selective reduction.

Enantioselectivity of carbonyl reduction depends on intrinsic parameters related to its structural environment (R_1 and R_2), and extrinsic parameters (e.g reaction conditions such as time, temperature, solvent nature and amount, substrat, baker's yeast and glucose concentrations).

In this preliminary work, we attempt to quantify the effect of structural parameters on the enantioselectivity of the reduction of the carbonyl group with baker's yeast. We try to establish a model for the prediction of the enantioselectivity for a sample of miscellaneous carbonyl compounds using the neural network (NN) approach. We also evaluate the contribution of each group to the enantioselectivity and compare the model to that obtained using multidimensional statistical analysis.

Material and Methods: A set of 41 carbonyl compounds is taken from different literature sources (table 1). All the baker's yeast reductions are carried out in relatively similar conditions (e.g. amounts of baker's yeast, glucose, substrat, reaction time). The studied compounds have a common C=O group, then their coding was simplified. Each molecule was described by a set of variables related to radicals R_1 and R_2. This procedure allows us to evaluate the incidence of each radical on the enantioselectivity of the reduction's product. The

group having the priority according to Cahn-Ingold-Prelog rules is considered as R_2 in order to obtain a homogeneous description of the molecules.

Table 1: Chemical structures of the compounds studied

Nr	Chemical structure	Nr	Chemical structure
1	CH_3-CO-CH_2-CO_2CH_3 (93.5[a])[5]	2	CF_3-CO-CH_2-CO_2CH_3 (26)[5]
3	$ClCH_2$-CO-CH_2-CO_2CH_3 (77.5)[5]**	4	$ClCH_2$-CO-CH_2-$CO_2C_2H_5$ (75)[5]**
5	$ClCH_2$-CO-CH_2-$CO_2C_3H_7$ (60)[5]**	6	$ClCH_2$-CO-CH_2-$CO_2C_4H_9$** (60)[5]
7	$ClCH_2$-CO-CH_2-$CO_2C_5H_{11}$ (5)[5]	8	$ClCH_2$-CO-CH_2-$CO_2C_6H_{13}$ (2.5)[5]
9	$ClCH_2$-CO-CH_2-$CO_2C_7H_{15}$ (1.5)[5]	10	$ClCH_2$-CO-CH_2-$CO_2C_8H_{17}$ (1)[5]
11	CH_3-CO-CH_2-CS-SC_2H_5 (98)[6]	12	CH_3-CO-CH_2-CS-SCH_3 (98)[6]
13	CH_3-CO-$CH(CH_3)$-CS-SCH_3 (98)[6]	14	CH_3-CO-CH_2-CO-SC_2H_5 (98)[6]
15	CH_3-CO-$CH(CH_3)$-CO-SC_2H_5 (98)[6]	16	N_3CH_2-CO-CH_2-$OCOCH_3$ (87.5)[7]
17	N_3CH_2-CO-CH_2-$OCOC_6H_5$ (89)[7]	18	N_3CH_2-CO-CH_2-$OCOC(CH_3)_3$ (95)[7]
19	C_2H_5-CO-$(CH_2)_3$-NO_2 (99)[8]	20	C_2H_5-CO-C_3H_7 (83.5)[8]
21	$BrCH_2$-CO-CH_2-$CO_2C_2H_5$ (99)[9]**	22	N_3CH_2-CO-CH_2-$CO_2C_8H_{17}$ (1)[9]**
23	C_4H_9-CO-CH_2-COOH (0)[10]	24	C_6H_5-CO-CH_2-$CO_2C_2H_5$ (100)[11]
25	$C_6H_5CH_2O(CH_2)_2COCH_2CO_2C_5H_{11}$(98)[12]	26	$C_6H_5CH_2O(CH_2)_2COCH_2CO_2C_6H_{13}$ (98)[12]
27	CH_3-CO-CH_2OH (4)[13]	28	C_2H_5-CO-CH_2OH (0)[13]
29	C_4H_9-CO-CH_2OH (0)[13]	30	C_6H_5-CO-CH_2-$OCOCH_3$ (97)[14]
31	$C_6H_5CH_2O$-CH_2-CO-CH_3 (95)[14]	32	CH_3-CO-$CO_2C_2H_5$ (95.5)[15]
33	C_2H_5-CO-$CO_2C_2H_5$ (87.5)[15]	34	C_6H_5-CO-CO_2CH_3 (0)[15]
35	C_6H_5-CO-$CONH_2$ (0)[11]	36	2-Furyl-CO-CO_2CH_3 (4)[16]
37	CH_3-CO-CHO* (98)[17]	38	C_3H_7-CO-CHO* (98)[17]
39	CH_3O-CH_2-CO-CHO* (97.5)[18]	40	CH_3-CO-CH_2-$SO_2C_6H_5$ (97.5)[1]
41	CH_3-CO-CH_2-SCH_3 (97.5)[1]		

a) % S ; * aldehyde protected by $HS(CH_2)_3SH$, ** compounds not correctly classified using our neural network model.

We consider molar refraction (MR)[19] that accounts for both electronic and size of the radicals, and p (Wiener's polarity index)[20] in order to take into account the shape. Enantioselectivity index (E : %S enantiomer) is coded using a binary variable (positive if the reduction lead to S mainly and negative elsewere). E is correlated to the descriptors using Linear Discriminant Analysis[21] and the Neural Network approach [22-24] .

Results and Discussions

Linear Discriminant Analysis: The final data set (41 x 5) is subjected to LDA in order to establish a linear model between the enantioselectivity index E and the descriptors used. The most significant model is :

$$E_{calc.} = 0.018\ MR(R_2) - 0.404 \quad (n = 41,\ r = 0.492,\ s = 0.415) \qquad (1)$$

The coefficient associated with $MR(R_2)$ is statistically significant ($p<1\%$). However, according to the statistical criteria, the equation is not significant. It leads to 50% of good classification of the compounds pro-S, and 92.3% of those pro-R. We think that the good prediction ability of pro-R is only due to the composition of our sample. The average prediction ability is 63%. Then LDA is not appropriate for modelling enantioselectivity of

the reduction for the studied sample. This is due to the non-linearity of the enantioselectivity as a function of the parameters describing R_1 and R_2 groups. Thus, we adopte the neural network approach.

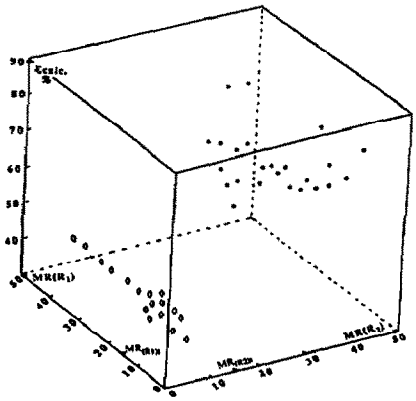
Neural Network: Data is subjected as a training set to the NN (configuration 4-4-1, using a sigmoid as a transfer function with a = 1, b = 0.2 and c = 0.5) in order to evaluate the weights of the connections between all the neurons (Table 2). After several attempts, input values of the enantioselectivity function were taken as 0.2 when the reaction lead to R enantiomer mainly and 0.8 elsewere.

Table 2 : Weight matrix for the connections between neurons

	H1	H2	H3	H4
I1	-8.48	-6.12	-7.84	O.21
I2	4.09	-3.28	1.00	4.00
I3	15.94	5.83	9.19	1.56
I4	27.41	12.31	22.29	1.9
O	9.62	-1.75	-2.37	-2.14

I1->I4 : input neurons representing descriptors.
H1->H4 : hidden neurons. O : output neuron.
Cost : 1.89, $C_{MR(R1)}$ = 43%
$C_{MR(R2)}$ = 31%, C_{P1} = 7%
C_{P2} = 18%, Number of cycles is 2050.
$C_{x(Ri)}$: contribution of a descriptor related to Ri.

After a training phase, the classification ability of the NN is tested by calculating $E_{calc.}$ for each molecule. Thirty five (85.4%) molecules are correctly classified by the NN (82.1% and 92.3% for pro-S and pro-R respectively). Classification ability of the network is higher than that of LDA. However its use, as a tool in the enantioselectivity prediction, is not easy due to the calculations associated with the prediction procedure. In fact, we try to explicit the rules found by the NN model and present them more simply. We focuse our investigation by considering molar refraction of both R_1 and R_2 only, due to the contribution that these parameters account for (43 and 31% for MR_1 and MR_2 respectively). Curve representing Ecalc. = f($MR_{(Ri)}$) (fig.) reveals the behaviour of the two descriptors.

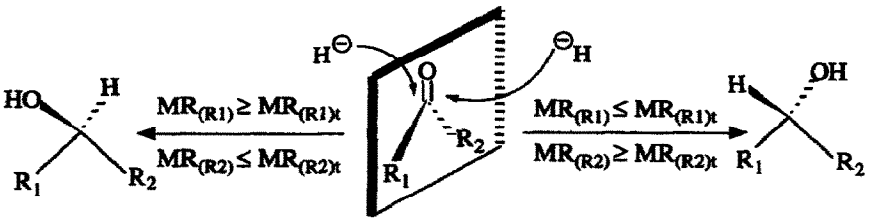


The cloud of points at the top of the cube represents enantiomers of S configuration, with one R enantiomer (not correctly predicted). The other cloud of points represents enantiomers of R configuration, with five S enantiomers (not correctly predicted).

Figure : Evolution of Ecalc. with $MR_{(Ri)}$. $MR_{(Ri)t}$ is approximatively equal to 15 MR units for both radicals.

Figure allows the following conclusions.

$MR_{(R1)} \leq MR_{(R1)t}$ and $MR_{(R2)} \geq MR_{(R2)t}$ => S mainly. $MR_{(R1)} \geq MR_{(R1)t}$ and $MR_{(R2)} \leq MR_{(R2)t}$ => R mainly. $MR_{(R1)t} \sim MR_{(R2)t}$ (threshold) corresponds, approximatively, to the molar refraction of groups e.g. equivalent to $CO_2CH_2CH_3$. For the other cases a mixture of R+S with low enantiomeric excess is obtained.



The model presented is efficient and shows that enantioselectivity may be quantified if data is unambigous. A more imporant set of carbonyl compounds is under study in order to elaborate general rules including some other parameters.

References

1. Servi, S. *Synth. Commun.* **1990**, 1-25.
2. Prelog, V. *Pure Appl. Chem.*, **1964**,*9*, 119.
3. Sih, C.J.; Zhou, B.; Goupalan, A.S.; Shieh, W.R.; Van Middlesworth, F. in *:Selectivity. A goal for Synthetic Efficiency.*, Bartman, W.; Trost, B. Eds., Workshop Conferences Hoechst. vol. 4, VCH, Weinheim, 1983, pp.251-261.
4. Zhou, B.; Goupalan, A.S.; Van Middlesworth, F.; Sih, C.J.; Shieh, W.R. *J. Am. Chem. Soc.*, **1983**, *105*, 5925.
5. Furstos, R. *Actualité Chimique*, **1990** (Janv. Fév.), 6-16
6. Itoh, T.Y.; Yonekawa, Y.; Sato,T.; Fujisawa,T. *Tetrahedron Lett.* **1986**, *27*, 5405-08.
7. Sato,T.; Mizutani,T.; Okumura,Y.; Fujisawa,T. *Tetrahedron Lett.* **1989**, *30*, 3701-02.
8. Nakamura,K.; Inoue,Y.; Shibahara,J.; Oka,S. *Tetrahedron Lett.* **1988**, *29*, 4769-70.
9. Fuganti,C.; Grasselli,P.; casati,P.; Carmeno,M. *Tetrahedron Lett.*. **1985**, *26*, 101.
10. Lemieux, R.U.; Giguère,J. *Can. J. Chem.* **1951**, *29*, 678.
11. Deol, B.S.; Ridley, D.D.; Simpson, G.W.; *Aust. J. Chem.* **1976**, *29*, 2459.
12. Hirama, M.; Shimizu, M.; Iwashita,M. *J. Chem. Soc. Chem. Commun.* **1983**, *2*, 599.
13. Cuette, J.P.; Spassky,N.; Bougherot, D. *Bull. Soc. Chim. Fr.* **1972**, 4217
14. Manzocchi, A.; Fiecchi,A.; Santan, E. *J. Org. Chem.*, **1988**, *53*, 4405.
15. Nakamura, K. ; Ikoue, K.; Ushio, K.; Oka, S.; Ohno, A. *J. Org. Chem.* **1988**, *53*, 2598.
16. Barton, D.H.R.; Brown, B.D.; Ridley, D.D.; Widdowson, D.A.; Keys, A.J.; Leaner, C.J. *J. Chem. Soc. Perkin Trans* **1975**, 2069.
17. Fujisawa, T.; Kojima, E.; Itoh, T.; Sato, T. *Chem. Lett.* **1985**, 1751.
18. Guanti, G.; Banfi, L.; Narrisano,E. *Tetrahedron Lett.* **1986**, *27*, 3547.
19. (a)- Weast, R. C. *Handbook of Chemistry and Physics*, 1st Student Edition; CRS : 1988, pp.E-318. (b)-Pannetier, G.*Chimie Physique Générale- Atomistique, Liaisons Chimiques, et Structure moléculaire*, Masson & Cie, 1969.
20. Wiener, H. *J. Am. Chem. Soc.* **1974**, *69*, 17-20.
21. Statview S.E. Graphics, Brainpower, 1987.
22. De Saint Laumer, J.Y.; Chastrette, M. *Eur. J. Med. Chem.* **1990**, *26*, 829-833.
23. Rumelhart , D.E.; Hinton, C.E.; Williams, R.J. *Nature* , **1986**, *323*, 533-536.
24. Chastrette, M.; Zakarya, D.; Peyraud, J.F. *Eur. J. Med. Chem.* accepted for publication.

(*Received in France* 29 *March* 1994; *accepted* 10 *May* 1994)