PREDICTION OF PREDOMINANT CONFIGURATION IN THE ENANTIOSELECTIVE REDUCTION OF β-KETOACID DERIVATES WITH BAKER'S YEAST USING NEURAL NETWORKS

Didier Villemin,[a]* Driss Cherqaoui[a] and Abdelhalim Mesbah[b]

Keywords: β-ketoacid; reduction; enantioselectivity; prediction; neural networks.

Correlation between chemical structure and enantioselectivity in baker's yeast reduction of a set of carbonyl compounds was constructed by means of a multi-layer neural network using the back-propagation algorithm. To evaluate the predictive power of the neural network (NN) model, the cross-validation procedure was used, 88 % of the reactions were correctly predicted.

* Corresponding Authors
E-Mail: villemin@ensicaen.fr
[a] Normandie Université, France, ENSICAEN, LCMT, UMR CNRS 6507, INC3M, FR 3038, Labex EMC3, Labex SynOrg, 14050 Caen, France.
[b] University Cadi Ayyad, Faculty of Sciences, Depart. Chemistry, BP 2390, Marrakech, Marocco.

Introduction

Concerning the enantioselective reduction of carbonylated compounds, the prediction of the preferred alcohol enantiomer in synthesis is a difficult task. For the prediction of the R/S configuration in an asymmetric reduction of carbonyl compounds, Prelog, Cram and Felkin have developed models of limited applicability based on steric criteria.[1]

J. Aires de Sousa *et al* have studied the enantioselective reduction of ketones by DIP-chloride,[2] the addition of diethylzinc on benzaldehyde[2] and the enantioselective hydrolysis of ester by *Pseudomonas*,[3] all predictions were made using neural network. Using also neural networks for prediction of enantioselectivity, W. M. F. Fabian *et al* [4] have described the ring opening of epoxide by hydrolases.

In our study, the baker's yeast (*Saccharomyces cerevisiae*) enantioselective reduction of β-ketoacid derivates was chosen because of its importance in preparation of chiral alcohols (Fig. 1). Numerous enzymatic systems which are present, can perform such a reduction, but different experimental conditions do not generally influence the resulted configuration.[5]

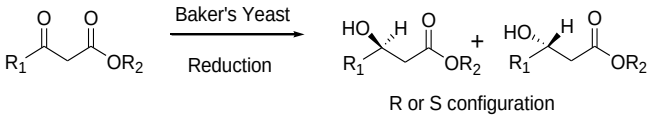

Figure 1. Bakers' yeast reduction of β-ketoacid derivates

Neural Networks (NNs) have successfully been used in organic chemistry,[7] particularly in QSAR studies, where numerous enzymatic systems and metabolic ways are implicated.[8] NNs are mathematical models of biological neural systems which fit non-linear problems and give better correlations than the multiple linear regression (MLR) ones. A description of the back-propagation algorithm was given previously[9] as well a more extensive description.[7] An attempt of structure-enantioselectivity relationships using neural networks was already described,[10] but the authors were unable to make predictions with molecular refraction[22] and finally proposed a Prelog model.

Methods

For our study, 35 reactions were extracted from reviews by Servi and Czuk et al.[6], where we can find the nearby experimental conditions (Table 1). Under these reactions the enantioselectivity was generally total, therefore our attention was focused on the prediction of R/S configuration of the produced enantiomer.

Previous studies[3,11] have shown that the correlation between the structure of the starting ketone and the alcohol obtained depends on steric criteria but other effects can also occur such as electronic ones. Therefore, the substituting groups R_1 and R_2 can be described by 2 kinds of parameters, electronic and steric parameters.

Electronic parameters: Hammett[12] assigned to every substituent a constant σ which represents its electronic effects on the reaction site. Taft *et al*.[13] have suggested two models where inductive and resonance contributions are quantitatively separated.

Steric parameters: Several studies[16] could be found in literature concerning the steric effects of substituent groups in organic reactions.[17]

Table 1: Chemical structures of the compounds studied

No	R_1	R_2	No	R_1	R_2	No	R_1	R_2
1	CH_3	C_2H_5	13	C_2H_5	C_8H_{17}	25	$CH_2CH_2CH=CH_2$	tC_4H_9
2	CH_3	nC_3H_7	14	NC_3H_7	C_8H_{17}	26	$CH_2CH_2C(CH_3)=CH_2$	CH_3
3	CH_3	nC_5H_{11}	15	NC_4H_9	C_2H_5	27	$CH_2CH_2C(CH_3)=CH_2$	C_2H_5
4	CH_3	nC_8H_{17}	16	CH_2Cl	CH_3	28	$CH_2CH_2CH=CH_2$	H
5	CH_3	iC_3H_7	17	CH_2Cl	C_2H_5	29	$CH_2CH_2CH=CH_2$	CH_3
6	CH_3	Ph	18	CH_2Br	C_2H_5	30	$CH_2CH_2CH=CH_2$	C_2H_5
7	CH_3	CH_3	19	CH_2Br	nC_3H_7	31	CH_2Cl	nC_8H_{17}
8	CH_3 CH_3	nC_4H_9	20	CH_2Br	nC_7H_{15}	32	C_2H_5	nC_8H_{17}
9	CH_3	tC_4H_9	21	CH_2Br	nC_8H_{17}	33	NC_3H_7	H
10		CH_3	22	CCl_3	C_2H_5	34	NC_3H_7	C_2H_5
11	C_2H_5	C_2H_5	23	CF_3	C_2H_5	35	NC_4H_9	H
12	Ph	C_2H_5	24	CH_3	C_2H_5			

Hansch *et al.*[18] proposed molecular refraction and molecular mass as sample measures of steric effects of substituting groups. In this latter study, the steric parameter is the volume of the substituents (V) computed using the Gavezzotti method.[19]

Results and discussions

A three layers Neural Network (NN) was used with the back-propagation (BP) algorithm for the prediction of predominant configuration (R or S) of the final product. Two methods were used to describe the reactions:

First method: Every reaction is described by 6 parameters σ_m, σ_p and V for R_1 and the same for R_2, this represents the input layer (6 neurons). The output layer contains only one neuron, which takes the value of 1 if the predominant configuration is S and 0 if it is R.

Second method: In this case, four parameters were used to describe each substituent (σ_m, σ_p, *V* and *L*), where *L* is the Verloop[20] parameter *L* represents the length of the substituent along the bond axis between the substituent atom and the parent compound, it was chosen because it permits to distinguish isomers.

We used a network with 6 or 8 units and a bias in the input layer, a variable hidden layer including bias, and one unit in the output layer. Input and output data were normalized between 0.1 and 0.9. The weights were initialized to random values between -0.5 and +0.5 and no momentum was added. The learning rate was initially set to 1 and was gradually decreased until the error function could no longer be minimized. All computations were performed using our own programs, written in the C language.

Learning. In order to determine the best architecture, different NNs have been tried using the two description systems [6-*x*-1 and 8-*x*-1; *x* = 1, 2, 3, 4, 5, 6, ...) with the all 35 reactions as a training set. The criteria used for the comparison of the architectures is the percentage of reactions correctly classified. We consider that we have a correct classification for a reaction if the output neuron was greater than 0.6 for S configuration and less than 0.4 for R configuration. After 2000 iterations, the NNs of structure [6-*x*-1] (*x*=2,3,4,5) were able to classify 34 of the 35 reactions studied.

Prediction. The predictive ability of an NN is its ability to give a satisfactory output for a molecule not included in the examples the NN learned. To determine that predictive ability, cross-validation has been used.[21] After 1400 iterations in the cross validation procedure, 29 of the 35 reactions were correctly predicted (Table 3) with an NN [6-3-1] and [6-4-1] and 31 reactions of the 35 reactions (88 %) are correctly predicted with an NN (8-3-1). Clearly the parameter L provides new information to the NN.[22] The use of percent (% of S) or 2 neurons (0,1) as output does not improve the results.

We have used σ_m and σ_p parameters introduced by Taft for electronic effects of meta and para positions. These parameters are available in the literature[14] and Kvasnicka has shown[15] that these parameters can be computed by a neural network using simple structural data as inputs.

Table 2. Prediction results using the first method

NN architecture	Number of reactions correctly predicted/35
6-1-1	26
6-2-1	27
6-3-1	29
6-4-1	29

Table 3. Prediction results using the second method

NN architecture	Number of reactions correctly predicted/35
8-1-1	26
8-2-1	30
8-3-1	31
8-4-1	30

Conclusion

After the neural network has been fully trained, it has shown that the network was able to form reliable generalizations to predict R/S configuration in baker's yeast reduction of the carbonyl compounds presented to it. This shows that steric and electronic parameters (σ_m, σ_p, *V* and *L*) provide sufficient information to a neural network for prediction of the reactivity of the compounds studied. This study represents a first approach to the prediction of enatioselectivity, the efficiency of induction depends on the concentrations of substrats, and for a quantitative prediction this factor must be added to steric and electronic parameters.

Acknowledgments:

We thank Dr. J.M. Cense (Chimie ParisTech) for computation of the volumes.

References and notes

[1](a)Cram, D. J., Abd Elhafez, F. A., *J. Am. Chem. Soc.*, **1952**, *74*, 5828-5835; (b) Chérest, M., Felkin, H., Prudent, N., *Tetrahedron Lett.*, **1968**, 2199-2204; (c) Prelog, V., *Pure Appl. Chem*., **1964**, *9*, 119-130.

[2]Joao Aires-de-Sousa, J., Gasteiger, J., *J. Chem. Inf. Comput. Sci*., **2001,** *41*, 369-375

[3]Zhang, Q. Y., Aires-de-Sousa, J., *J. Chem. Inf. Model.*, **2006**, *46*, 2278-2287

[4]Funar-Timofei, S., Suzuki, T., Paier, J. A. , Steinreiber, A., Faber, K., Fabian, W. M. F., *J. Chem. Inf. Comput. Sci.*, **2003,** *43*, 934-940

[5]Zhou, B., Gopalan, A. S., Van Middelsworth, F., Shieh, W., Sih, C. J., *J. Am. Chem. Soc*., **1983**, *105*, 5925-5928.

[6](a)Servi, S., *Synthesis*, **1991**, 1-25; (b) Csuk, R., Glänzer, B. I., *Chem. Rev*., **1991**, *91*, 49-97; (c) Santaniello, E., Ferraboschi, P., Grisenti P., Manzocchi, A., *Chem. Rev.*, **1992**, *92*, 1071-1140; (d) Theil, F., *Chem. Rev*., **1995**, *95*, 2203-2227; (e) Hall, M., Bommarius, A. S., *Chem. Rev.*, **2011**, *111*, 4088-4110; (f) Strohmeier, G. A., Pichler, H., May, O., Gruber-Khadjawi, M., *Chem. Rev.*, **2011**, *111*, 4141-4164.

[7](a)Zupan, J., Gasteiger, J., *Neural Networks for Chemists, An Introduction*, VCH editor, New York, **1993**; (b) Zupan, J., Gasteiger, J., *Neural Networks in Chemistry and Drug Design: An Introduction*, 2nd Edition**,** Wiley editor, **1999**; ISBN: 978-3-527-29779-5.

[8](a)Andrea. T. A., Kalayeh, H., *J. Med. Chem.* **1991**, *34*, 2824-2836; (b)So, S. S., Richards, W. G., *J. Med. Chem*. **1992**, *35*, 3201-7; (c) Villemin, D., Cherqaoui D., Cense, J. M., *J. Chim. Phys.*, **1993**, *90*, 1505-1519; (d)Ajay, A., *J. Med. Chem.*, **1993**, *36*, 3565-3571.

[9]Cherqaoui, D., Villemin, D., *J. Chem. Soc. Faraday Trans.*, **1994**, *90*, 97-102.

[10]Zakarya, D., Farhaoui, L., Fikh-Tetouani, S., *Tetrahedron Lett.*, **1994**, *35*, 4985- 4988.

[11]Shieh, W., Gopalan, A. S., Sih, C. J., *J. Am. Chem. Soc.*, **1985**, *107*, 2993-2994.

[12](a)Hammett. L. P., *J. Am. Chem. Soc*., **1937**, *59*, 96-103; (b) Hammett. L. P., *Physical Organic Chemistry*, 2nd edition, McGraw-Hill, New York, **1970**.

[13](a)Taft, R. W., Lewis, I. C., *J. Am. Chem. Soc.*, **1958**, *80*, 2436-2436; (b) Taft, R. W., Lewis, I. C., *J. Am. Chem. Soc.*, **1959**, *81*, 5343-5352.

[14]Hansch, C., Leo, A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley editor, New York, **1979**.

[15]Kvasnicka, V., Sklenak, S., Pospichal, J., *J. Am. Chem. Soc.*, **1993**, *115*, 1495-1500.

[16](a)Taft, R. W., Newman, M. S., *Steric Effects in Organic Chemistry*, John Wiley editor, New York, **1956**, 556-565; (b) Charton, M., *Prog. Phys. Org. Chem.*, **1989**, *13*, 119-251; (c) Gallo, R. *Prog. Phys. Org. Chem.*, **1983**, *14*, 115-163; (d) Kutter, E., Hansch, C., *J. Med. Chem.*, **1969**, *12*, 647-652.

[17]Unger, S. H., Hansch, C., *Prog. Phys. Org. Chem.*, **1976**, *12*, 91-118.

[18]Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikaitani, D., Lien, E. J., *J. Med. Chem.*, **1973**, *16*, 1207-1216.

[19]Gavezzotti, A., *J. Am. Chem. Soc.*, **1983**, *105*, 5220-5225.

[20]Verloop, A., Hoogenstraaten, W., Tipker, J., *Drug Design*, **1976**, *7*, 165-207.

[21]Cherqaoui, D., Villemin, D., Mesbah, A., Cense, J. M., Kvasnicka, V., *J. Chem. Soc. Faraday Trans*, **1994**, 90, 2015-2019.

[22]For example:[*L*(n-Pr) = 5.05 Å and *L*(iPr) = 4.11 Å], contrary to volume (*V*) and molecular refraction (*MR*) which indicates only the bulk [*V*(nPr) = 60.67Å^3 and *V*(iPr) = 60.63Å^3 ; *MR*(nPr) = *MR*(iPr)].

Received: 31.03.2016.
Accepted: 06.06.2016.